Undergraduate Real Analysis

Lecture Notes

Thomas Laetsch

These notes are **copyright 2014** by **Thomas Laetsch**. They may be used for personal or classroom purposes, but *not* for commercial purposes.

Contents

Part 1. An Introduction to Ordered Fields 7

Chapter 0. Warm Up 9
1. Sets 9
2. Proof Strategy 19
3. Induction 20
4. Rational Numbers 21
5. Exercises 22

Chapter 1. Ordered Fields with an Introduction to the Real Numbers 27
1. Fields 27
2. Ordered Fields 31
3. Absolute Value and Distance on Ordered Fields 36
4. Common Notation and the Inclusion of $\mathbb{Q}$ 39
5. Archimedes and the Density of $\mathbb{Q}$ 40
6. Supremums and Infimums 43
7. The Real Numbers 46
8. Exercises 49

Part 2. Exploring the Topological Structure of $\mathbb{R}$ 53

Chapter 2. Sequences and Sequential Convergence in $\mathbb{R}$ 55
1. General Sequences 55
2. Convergence of Sequences in $\mathbb{R}$ 56
3. Properties of the Limit 59
4. Divergent Sequences 62
5. Monotone Sequences 63
6. Subsequences 64

7. Cauchy Sequences 68
8. Sequences and Convergence Within Subsets of $\mathbb{R}$ 69
9. Exercises 71

Chapter 3. The Topology on $\mathbb{R}$ 77
1. Open Sets 77
2. Closed Sets 79
3. Classifications of Points in a Set 81
4. Classification of Closed Sets 82
5. Compactness 84
6. Relative Topology 86
7. Exercises 88

Chapter 4. Continuous Functions on $\mathbb{R}$ 91
1. Continuous Functions and Sequences 92
2. Continuous Functions and Topology 94
3. Continual Limits of Functions 95
4. Continuous Images of Sets 97
5. Pointwise Algebraic Constructions and Continuity 100
6. Stronger Forms of Continuity 103
7. Exercises 106
8. Other Problems 115

Chapter 5. Differential Calculus, Riemann Integration, and Function Spaces 117
1. Differential Calculus 117
2. Riemann Integration 126
3. Series 143
4. More on Limits of Functions 156
5. Taylor Expansions and Approximation 159
6. Exercises 163

Part 3. Real Analysis II: Abstractions and functions of several real variables. 171

Chapter 6. Metric Spaces 173

1. The definition of a metric space and first examples 173
2. Sequences on metric spaces 174
3. Open and Closed Sets 175
4. Further Set Classifications 176
5. More on Closed Sets 180
6. Compactness 182
7. Continuous Functions Between Metric Spaces 187
8. The Contraction Mapping Theorem and an Example 192
9. Exercises 197

Chapter 7. Introduction to Functional Analysis 207
1. Basics of Elementary Linear Algebra 207
2. Normed Vector Spaces 213
3. Banach Spaces 218
4. Inner Product Spaces 225
5. Hilbert Spaces 231
6. Bounded linear operators between Banach spaces 236
7. Bounded linear operators between Hilbert spaces 241
8. A first look at product spaces 244
9. Exercises 250

Chapter 8. Differential Calculus in $\mathbb{R}^n$ 257
1. Some Notes on the Structure of $\mathbb{R}^n$ 257
2. The Derivative 259
3. Directional and Partial Derivatives 266
4. Inverse and Implicit Function Theorems 276
5. Exercises 280

Chapter 9. Introduction to Measure Theory 285
1. Motivation 285
2. Measure Spaces 286
3. Integration of Measurable Functions 296
4. Exercises 312

Part 1

An Introduction to Ordered Fields

CHAPTER 0

Warm Up

1. Sets

Roughly speaking, a *set* is a collection of *elements* which one would like to house together, so that the set might be considered as an object unto itself when judicious. If S is a set (note that labeling the entire set with the symbol S is a manifestation of our considering the set as an object unto itself worth "naming") and a is an element of S, then we write $a \in S$; synonymously, we can say that S *contains* a. If S does not contain a, then we write $a \notin S$. The *empty set*, denoted $\varnothing$, is the set which contains no elements.

1.1. Cursory Material.

Subsets. If S and T are sets such that every element in S is also in T, then we say that S is a *subset* of T, in symbols we write $S \subseteq T$ (or, equivalently in these notes, $S \subset T$); in this case, we can also say that T is a *superset* of S. Two sets S and T are said to be equal when $S \subseteq T$ and $T \subseteq S$; intuitively, this means that S and T contain all the same elements. Further, if S is a subset of T, but T is not also a subset of S (in symbols, $S \subset T$ and $T \not\subset S$), then we might write $S \subsetneq T$ to emphasize that S is a subset of T, but does not equal T; often a subset $S \subsetneq T$ is called a *proper subset* of T. Note that if S is a subset of T, but whether or not S is a proper subset of T is unknown or unimportant, it is still correct to write $S \subset T$ or $S \subseteq T$, even if it happens that S is actually a proper subset of T. (The point being, for these notes the symbols $\subset$ and $\subseteq$ are synonymous and can be translated into the english phrase "is a subset of"; however, $\subsetneq$ must be translated as "is a proper subset of").

Set Builder Notation. There are several notations commonly used when describing a set via the elements contained within the set. The most common general theme of such a set description follows the paradigm $S = \{\text{elements within the set}\}$; here S is the label we are giving to the set whose elements are those described within the curly brackets $\{\ \}$. An explicit form of this notation might be $S_1 = \{a, b, c\}$ or $S_2 = \{\text{all even integers}\}$; here S_1 is the set whose elements are a, b, and c, whereas S_2 is the set whose elements are all even integers.

While this explicit description is fine when practical, it is often intractable to list out all the elements of a set, in which case we frequently rely on a more metaphoric description. This metaphoric description has the general form $S = \{x \text{ such that } Q(x)\}$. In this case, $Q(x)$ is some proposition based on x which can be determined as true or false, we can then read the expression $S = \{x \text{ such that } Q(x)\}$ as "S is the set containing the elements x such that $Q(x)$ is true". It's important to note here that the symbol x in this notation is just a dummy variable; i.e., $\{x \text{ such that } Q(x)\}$ is the same set as $\{A \text{ such that } Q(A)\}$ is the same set as $\{\xi \text{ such that } Q(\xi)\}$. As an example of such a description, let $S = \{x \text{ such that } x \text{ is a real number and } x^2 < 2\}$; here S is the set containing all real numbers whose square is less than 2, in interval notation of the real line, this is the same as $S = (-\sqrt{2}, \sqrt{2})$.

In the metaphoric description, it becomes tedious to write "such that" and the shorthand notation | or : is often substituted; e.g. the set $\{x \text{ such that } x \text{ is a real number and } x^2 < 2\} = \{x : x \text{ is a real number and } x^2 < 2\} = \{x \mid x \text{ is a real number and } x^2 < 2\}$. Less common, though preferred by the author, the acronisation s.t. for "such that" can also be used; e.g., $\{x \text{ such that } x \text{ is a real number and } x^2 < 2\} = \{x \text{ s.t. } x \text{ is a real number and } x^2 < 2\}$. We can be a bit loose on the ordering as well, as long as the metaphoric description is still clear and correct. Indeed, note that $\{x \text{ such that } x \text{ is a real number and } x^2 < 2\}$ and $\{\text{real numbers } x \text{ such that } x^2 < 2\}$ should describe the same sets. Using our shorthand, we would arrive at $\{x \mid x \text{ is a real number and } x^2 < 2\}$ for the first and $\{\text{real numbers } x \mid x^2 < 2\}$ for the second; in fact, both variations are perfectly good and commonly used to describe the same set. Further, if $\mathbb{R}$ is the set of all real numbers, we may as well have written $\{x \in \mathbb{R} \mid x^2 < 2\}$; in words, this would translate to "the set of elements x in the real numbers such that $x^2 < 2$".

1.1.1. *Disjointness.* If S_1 and S_2 are sets which have no elements in common, then we say that S_1 and S_2 are *disjoint*. If $S_1, S_2, ..., S_n$ are all sets (for some $n \geq 2$), then we say that the sets S_k (for $k \in \{1, 2, ..., n\}$) are *mutually disjoint*, also called *mutually exclusive*, when no two sets share any elements with each other. Described more rigorously, $S_1, S_2, ..., S_n$ are mutually disjoint when for all $m, k \in \{1, 2, ..., n\}$ with $m \neq k$, the sets S_m and S_k are disjoint. More generally, if I is a set (with possibly infinitely many elements) and for each $k \in I$ there is an assigned set S_k, then these sets are said to be *mutually disjoint* when for all $m, k \in I$ with $m \neq k$, the sets S_m and S_k are disjoint.

Russell's Paradox. We have ignored a rigorous definition of a set to this point and will continue to sweep many of the details aside for the sake of brevity and intuition. However, it is true that there are restrictions to what one can define as a set. We now present a famous paradox that can arise from ignoring such details, it is called *Russell's Paradox*:

Suppose there were no restrictions on the types of elements we could assign as elements within a set. Hence let S be a set which contains as elements all sets which don't contain themselves; notationally $S = \{A : A \text{ is a set and } A \notin A\}$ (for example, if $A = \{a, B, c\}$ then $A \in S$ since $A \notin A$; however, if $B = \{a, B, c\}$, then $B \notin S$ since $B \in B$). We now arrive at the paradox when we ask: is $S \in S$? You see, if $S \in S$, then S contains itself, and by the definition of S we must have $S \notin S$; this shows that $S \in S$ implies $S \notin S$. On the other hand, if $S \notin S$, then by the definition of S, we must have $S \in S$ since S contains sets which don't contain themselves; this shows that $S \notin S$ implies $S \in S$. We conclude that $S \in S$ if and only if $S \notin S$, our paradox.

While such a paradoxes do exist, all examples that follow which are claimed to be sets are, in fact, sets. If the reader finds themselves skeptical of this, let this skepticism be a motivation to further explore a rigorous account of set theory (of which the literature is plentiful), rather than an impedance to appreciating the work built on top of the foundational notion of a set.

1.2. Set Operations: Building New Sets From Old.

Union. Given two sets S_1 and S_2, the *union* of S_1 and S_2 is the set

$$S_1 \cup S_2 = \{x \text{ s.t. } x \in S_1 \text{ or } x \in S_2\}$$

Though we needn't restrict ourselves to two sets. Indeed, suppose that $S_1, S_2, ..., S_n$ are all sets for some $n \geq 2$, then the union of the sets $S_1, S_2, ..., S_n$ results in the set

$$\begin{aligned} S_1 \cup S_2 \cup \cdots \cup S_n &= \{x \text{ s.t. } x \in S_1 \text{ or } x \in S_2 \text{ or } \cdots \text{ or } x \in S_n\} \\ &= \{x \text{ s.t. } x \in S_k \text{ for some } k \in \{1, 2, ..., n\}\} \end{aligned}$$

For notational convenience, the latter union is typically written as $\bigcup_{k=1}^{n} S_k$; i.e.,

$$\bigcup_{k=1}^{n} S_k = S_1 \cup S_2 \cup \cdots \cup S_n$$

where the symbol k is just a "dummy variable" informing us of the indexes we are to union over; e.g., $\bigcup_{l=1}^{n} S_l$, $\bigcup_{k=1}^{n} S_k$, and $\bigcup_{\alpha=1}^{n} S_\alpha$ all represent the same union $S_1 \cup S_2 \cup \cdots \cup S_n$.

Continuing on in this fashion, if I is a set (with potentially infinitely many elements), and for every $k \in I$ we assign a set S_k, then the union of the sets S_k with $k \in I$ results in the set

$$\bigcup_{k \in I} S_k = \{x \text{ s.t. } x \in S_k \text{ for some } k \in I\}.$$

Note that if $I = \{1, 2, ..., n\}$, then $\bigcup_{k \in I} S_k = \bigcup_{k=1}^{n} S_k = S_1 \cup S_2 \cup \cdots \cup S_n$.

1.2.1. *Intersection.* Given two sets S_1 and S_2, the *intersection* of S_1 and S_2 is the set

$$S_1 \cap S_2 = \{x \text{ s.t. } x \in S_1 \text{ and } x \in S_2\}$$

As before, we needn't restrict ourselves to two sets. Indeed, suppose that $S_1, S_2, ..., S_n$ are all sets for some $n \geq 2$, then the intersection of the sets $S_1, S_2, ..., S_n$ results in the set

$$\begin{aligned} S_1 \cap S_2 \cap \cdots \cap S_n &= \{x \text{ s.t. } x \in S_1 \text{ and } x \in S_2 \text{ and } \cdots \text{ and } x \in S_n\} \\ &= \{x \text{ s.t. } x \in S_k \text{ for all } k \in \{1, 2, ..., n\}\} \end{aligned}$$

Notational convenience again insists that we frequently denote the latter intersection as $\bigcap_{k=1}^{n} S_k$; i.e.,

$$\bigcap_{k=1}^{n} S_k = S_1 \cap S_2 \cap \cdots \cap S_n$$

here (as in the case with unions) the symbol k is just a "dummy variable" informing us of the indexes we are to intersect over; e.g., $\bigcap_{l=1}^{n} S_l$, $\bigcap_{k=1}^{n} S_k$, and $\bigcap_{\alpha=1}^{n} S_\alpha$ all represent the same intersection $S_1 \cap S_2 \cap \cdots \cap S_n$.

Most generally, if I is any set, and for every $k \in I$ we assign a set S_k, then the intersection of the sets S_k (for $k \in I$) results in the set

$$\bigcap_{k \in I} S_k = \{x \text{ s.t. } x \in S_k \text{ for all } k \in I\}.$$

If $I = \{1, 2, ..., n\}$, then $\bigcap_{k \in I} S_k = \bigcap_{k=1}^{n} S_k = S_1 \cap S_2 \cap \cdots \cap S_n$.

1.2.2. *Complimentation.* Given two sets S and T, the *relative complement* of S in T, denoted $T \backslash S$ or $T - S$, is the set

$$T \backslash S = T - S = \{x \text{ s.t. } x \in T \text{ and } x \notin S\}.$$

In words, the relative complement of S in T is the set of elements in T which are not also in S; i.e., those elements in T which are "outside" of S.

Often times, there exists some *ambient* superset R such that every set accounted for is understood to be considered as a subset of R. In this case, if $S \subset R$, we can drop the adjective "relative" and simply say the *complement* of S, denoted S^c, noting that $S^c = R \backslash S$. Notice that the notation S^c does not include R, so it is vital that the set R is understood from context as the ambient superset, otherwise S^c is meaningless or misleading. Indeed, if we were to let $S = \{a\}$ and ask, "what is S^c?", there would be no way to answer.

1.2.3. *Power Set.* Given a set S, the *power set* of S, denoted 2^S, is the set of all subsets of S. That is,

$$2^S = \{A \text{ s.t. } A \subset S\}.$$

It's important to note that an *element* of 2^S is a *subset* of S, a concept often confused. To further batter this point, if $S = \{a, b\}$ then $2^S = \big\{\varnothing, \{a\}, \{b\}, \{a, b\}\big\}$; note that, for example, $a \notin 2^S$, though $\{a\} \in 2^S$.

THEOREM 0.1.1. *Let S be any set. If S has n elements for some $n \in \mathbb{N}$, then 2^S has 2^n elements. Moreover, for any set S, the cardinality of 2^S is larger than that of S.*

1.3. Functions. Given two sets S and T, a *function* from S to T is an assignment of each element in S to a *unique* element of T. The notation $f : S \to T$ is used to represent a function from S to T; here f is a symbolic label representing the assignment of each element in S, where if $s \in S$, then $f(s)$ is the (unique!) assigned element in T. Given a function $f : S \to T$, S is called the *domain* and T is called the *codomain*. The *image* of f, denoted $f(S)$, is the subset of T defined by

$$f(S) = \{f(s) \text{ s.t. } s \in S\}.$$

A function is not a formula! One very common misconception is that of a function as a formula; for example, you might see some literature say "let $f(x) = x^2$" with no reference to the domain or codomain of the function; in mathematical terms, if you don't know the domain where x is being taken from, nor the codomain where $f(x)$ is landing in, then you do not have a function. To stress the point, a function requires three pieces of information: 1) the domain set S, 2) the codomain set T, and 3) the assignment f of each element $s \in S$ to a unique element $f(s)$ in T. However, we in the mathematics community do often abuse

notation so that if $f : S \to T$ has been defined as a function and we reference this function later, then we often only use the symbol f to represent the function $f : S \to T$; to exacerbate this abuse, we will very commonly call f the function! Moreover, it is *less* offensive if you give only the domain S and the assignment f without direct mention of the codomain T, since presumably you know what element $f(S)$ is, so there is some implicit understanding of T; however, this is still sloppy (see Exercise 6 below).

Suppose that $f : S \to T$ is a function. We say that the function is *injective* (also called *one-to-one*) when for all $s, t \in S$, it holds that if $f(s) = f(t)$, then $s = t$. We say that f is *surjective* when $f(S) = T$; that is, $f : S \to T$ is surjective when for all $t \in T$ there exists an $s \in S$ such that $f(s) = t$. We say that the function is *bijective* when it is both injective and surjective.

Composition of functions. Suppose that $f : S \to T$ is a function and that U is a set such that $f(S) \subset U$. Suppose further that $g : U \to V$ is a function. The *composition* of $f : S \to T$ with $g : U \to V$ results in the new function $g \circ f : S \to V$ defined by $g \circ f(s) = g\big(f(s)\big)$.

If $f : S \to T$ and $g : T \to S$ are functions then both compositions $g \circ f : S \to S$ and $f \circ g : T \to T$ make sense. If for every $s \in S$ it holds that $g \circ f(s) = s$, then $g : T \to S$ is called the *left inverse* of $f : S \to T$; analogously, if for every $t \in T$ it holds that $f \circ g(t) = t$ then $g : T \to S$ is called a *right inverse* of $f : S \to T$. If $g : T \to S$ is both a left inverse and right inverse of $f : S \to T$, then we say that $f : S \to T$ is *invertible* and call $g : T \to S$ the *inverse* of $f : S \to T$. It is standard to use the notation $f^{-1} : T \to S$ for the inverse of $f : S \to T$. The following proposition gives us some relationships between injectivity, subjectivity, and invertibility.

LEMMA 0.1.2. *Suppose that S and T are non-empty sets and $f : S \to T$ is a bijection.*

(1) *Suppose that $f : S \to T$ has both left and right inverses. Let $g : T \to S$ be a left inverse of $f : S \to T$, and $h : T \to S$ be a right inverse of $f : S \to T$. Then $g = h$.*
(2) *If $f : S \to T$ has an inverse $f^{-1} : T \to S$, then the inverse is unique.*

Proof.

(1) Let $t \in T$ be arbitrary. We will that $g(t) = h(t)$ which will prove that $g : T \to S$ and $h : T \to S$ are the same function. Since f is a bijection, there is a unique element $s_t \in S$ such that $f(s_t) = t$. Since g is the left inverse of f, this implies that $g(t) = s_t$. On the other hand, since h is a right inverse of f, it is true that

$f \circ h(t) = t$, and hence $f(s_t) = f(h(t)) = t$. Since f is a bijection, this implies that $h(t) = s_t$. Therefore $h(t) = g(t) = s_t$, finishing the proof.

(2) Suppose that $f_1^{-1} : T \to S$ and $f_2^{-1} : T \to S$ are both inverses of $f : S \to T$. In particular, this means that $f_1^{-1} : T \to S$ is a left inverse of $f : S \to T$ and $f_2^{-1} : T \to S$ is a right inverse of $f : S \to T$, and we just proved that this means $f_1^{-1} = f_2^{-1}$. Hence the inverse is unique.

■

PROPOSITION 0.1.3. *Suppose that S and T are non-empty sets. Let $f : S \to T$ be a function. Then:*

(1) *A left inverse of $f : S \to T$ exists if and only if $f : S \to T$ is injective.*
(2) *A right inverse of $f : S \to T$ exists if and only if $f : S \to T$ is surjective.*
(3) *An inverse of $f : S \to T$ exists if and only if $f : S \to T$ is bijective.*

Proof.

(1) Suppose a left inverse $g : T \to S$ of $f : S \to T$ exists. Let $s_1, s_2 \in S$ be arbitrary. If $f(s_1) = f(s_2)$, the $g \circ f(s_1) = g \circ f(s_2)$ implying that $s_1 = s_2$ since $g : T \to S$ is the left inverse of $f : S \to T$. This proves that if $f : S \to T$ has a left inverse, it is injective.

Conversely, suppose that $f : S \to T$ is injective. Fix some point $s_0 \in S$. We now construct a right inverse $g : T \to S$ of $f : S \to T$. For each $t \in f(S)$, there is a unique point $s_t \in S$ such that $f(s_t) = t$; we know such an element s_t exists since t is in the image of f, and is unique since f is injective. Hence for $t \in f(S)$ define $g(t) = s_t$. For all $t \in T \setminus f(S)$, let $g(t) = s_0$. We now have the following function $g : T \to S$ defined by

$$g(t) = \begin{cases} s_t & t \in f(S) \\ s_0 & t \in T \setminus f(S) \end{cases}$$

From here, if $s \in S$ then $g \circ f(s) = s$ since if $f(s) = t$, then $s = s_t$ and hence $g(f(s)) = g(t) = s_t = s$. This shows that if $f : S \to T$ is injective, it has a left inverse.

(2) Suppose that there exists a right inverse $g : T \to S$ of $f : S \to T$. Let $t \in T$ be arbitrary. We need to show that there exists some $s \in S$ such that $f(s) = t$, from

which we will conclude that f is surjective. Let $s = g(t)$, then $f(s) = f \circ g(t) = t$ since g is the right inverse. This proves that if $f : S \to T$ has a right inverse, then it is surjective.

Conversely, suppose that $f : S \to T$ is surjective. For each $t \in T$, there exists a element s_t such that $f(s_t) = t$; note that there are possibly many elements in S such which f assigns to t, but we are fixing our favorite one and calling it s_t. Define the function $g : T \to S$ by $g(t) = s_t$. Then for all $t \in T$, $f \circ g(t) = f(s_t) = t$. This proves that if $f : S \to T$ is surjective, then it has a right inverse.

(3) This is an exercise.

■

Inverse Image. A general function $f : S \to T$ is not necessarily invertible (which by Proposition 0.1.3 is the same as the statement that $f : S \to T$ is not necessarily bijective). Regardless, given any function $f : S \to T$, we can define the *inverse image* function (also called the *pre image* function), denoted $f^{-1} : 2^T \to 2^S$, which is defined by

$$f^{-1}(V) = \{s \in S \text{ s.t. } f(s) \in V\}.$$

Some caution is in order here. Indeed, we have seen the symbol f^{-1} before when defining inverse functions, but notice that the inverse image function $f^{-1} : 2^T \to 2^S$ is defined with domain 2^T and codomain 2^S and does not depend on the invertibility of $f : S \to T$. On the other hand, the inverse function $f^{-1} : T \to S$ only exists if $f : S \to T$ is invertible and is defined with domain T and codomain S.

1.4. Indexed Collections vs. Sets. The elements contained within a general set have no particular order, nor do redundancies of elements augment the set. For example, the sets $S_1 = \{a, b, c\}$, $S_2 = \{b, a, c\}$, and $S_3 = \{a, a, b, b, c\}$ are all the same set! Indeed, it is clear that each of these sets is comprised of exactly the elements a, b, and c. Indeed, if $x \in S_1$ then $x = a$ or $x = b$ or $x = c$; if $x \in S_2$ then $x = a$ or $x = b$ or $x = c$; if $x \in S_3$ then then $x = a$ or $x = b$ or $x = c$. This shows that $x \in S_1 \iff x \in S_2 \iff x \in S_3$, from which we can deduce that $S_1 = S_2 = S_3$.

One can imagine a situation where we want to concern ourselves with the order of the elements listed within a set or where we may need or want to repeat elements being considered. One construct that is extensively used to handle this situation is an *indexed collection*. To wit,

an indexed collection (also frequently called an *indexed family*) requires an *indexing set* I and an *underlying set* S. From here, we create an indexed collection by assigning to each element in I an element of S. To clarify, let $I = \{1, 2, 3, 4, 5\}$ and $S = \{a, b, c\}$. Then we can create the indexed collection $\{x_i\}_{i\in I}$ with $x_1 = a$, $x_2 = a$, $x_3 = b$, $x_4 = b$, and $x_5 = c$. In this way, we have created collection of elements that behaves like $\{a, a, b, b, c\}$ and distinguishes this collection from, say, $\{a, b, c\}$. Note also, we could instead have created an indexed collection $\{y_i\}_{i\in I}$ with $y_1 = b, y_2 = a, y_3 = b, y_4 = c, y_5 = a$, resulting in the collection behaving like $\{b, a, b, c, a\}$ but distinguished from $\{x_i\}_{i\in I} \approx \{a, a, b, b, c\}$ by the ordering of the elements; that is, x_i and y_i do not agree for every $i \in I$ (e.g., $x_1 = a$ whereas $y_1 = b$).

A concrete and useful perspective of an indexed collection with indexing set I and underlying set S is to view the indexed collection as a function $x : I \to S$. Indeed, if for all $i \in I$ we equate the function value $x(i)$ with the symbol x_i, we can identify the function x with the indexed collection $\{x_i\}_{i\in I}$. More on this is discussed when we start into sequences (see Section 1).

1.5. Cartesian Product. The *cartesian product* of two sets S_1 and S_2, denoted $S_1 \times S_2$, results in the set of all 2-tuples (s, t) where $s \in S_1$ and $t \in S_2$; i.e.,

$$S_1 \times S_2 = \{(s, t) \text{ s.t. } s \in S_1 \text{ and } t \in S_2\}.$$

Slightly generalizing this notion, the cartesian product of the sets S_1, S_2, ..., S_n, denoted $S_1 \times S_2 \times \cdots \times S_n$ or more conveniently $\prod_{i=1}^{n} S_i$, results in the set of all n-tuples $(s_1, s_2, ..., s_n)$ where $s_i \in S_i$ for every i in the range $1 \leq i \leq n$; i.e.,

$$S_1 \times S_2 \cdots \times S_n = \prod_{i=1}^{n} S_i = \{(s_1, s_2, ..., s_n) \text{ s.t. } s_i \in S_i \text{ for every } i \in \{1, 2, ..., n\}\}.$$

More generally, let I be a set and $\{S_i\}_{i\in I}$ be an indexed collection of sets. The Cartesian product of the sets $\{S_i\}_{i\in I}$, denoted $\prod_{i\in I} S_i$ is the of all indexed collections $\{s_i\}_{i\in I}$ where $s_i \in S_i$ for every $i \in I$; i.e.,

$$\prod_{i\in I} S_i = \{\{s_i\}_{i\in I} \text{ s.t. } s_i \in S_i \text{ for all } i \in I\}.$$

We conclude this section with a venerable question in mathematics, "for a general indexing set I, is $\prod_{i\in I} S_i$ non-empty?" The answer is... we don't know! If I is finite, the answer is in the affirmative; i.e., whenever I is finite, $\prod_{i\in I} S_i$ is non-empty (which can be shown by an

induction argument – see Section 3). Yet no proof of the non-emptiness of the product $\prod_{i\in I} S_i$ is known if I has infinitely many elements. That said, the general math community (particularly those whose work is in analysis) by and large accept that this set is non-empty as an *axiom*; this is the infamous *Axiom of Choice*.

1.6. Relations, Equivalence Relations, and Partitions. Suppose that S and T are sets. A subset $R \subset S \times T$ is called a *relation*. If $S = T$, then we say that R is a relation on S. When a particular relation R on S is in consideration, we typically denote the inclusion $(s,t) \in R$ by $s \sim t$, and say s *is related to* t. Often the relation R is defined by how the elements relate to each other, so it is common to introduce a relation R through the symbol $\sim$, because of this it is common to introduce the symbol $\sim$ (or whatever symbol we choose to use in place of $\sim$) and refer to R as the equivalence relation on S defined by $\sim$. We say that R is a *reflexive* relation on S when for all $s \in S$ it holds that $s \sim s$; we say that R is a *symmetric* relation on S when for all $s,t \in S$ it holds that $s \sim t \implies t \sim s$; we say that R is a transitive relation on S when for all $r,s,t \in S$, it holds that if $s \sim t$ and $t \sim r$ then $s \sim r$. If R is a reflexive, symmetric, and transitive relation on S, we say that R is an *equivalence* relation on S.

One point of view is that the notion of an equivalence relation on a set is the generalization of the notion of equality. One can readily check that the relation defined shorthand by $s \sim t \iff s = t$ defines an equivalence relation on any set. If R is an equivalence relation on a set S, and $s \in S$, then the *equivalence class* of s is the set $\{t \in S \text{ s.t. } s \sim t\}$; we will denote the equivalence class of an element s by $[s]$. The collection of all equivalence classes of S, denoted by $S/\sim$. That is,

$$S/\sim = \{[s] \text{ s.t. } s \in S\}.$$

A *partition* of a set S is any indexed collection of mutually disjoint subsets $\{T_i\}_{i\in I}$ of S such that $\bigcup_{i\in I} T_i = S$. Each element T_i in the partition is called a *piece* of the partition.

LEMMA 0.1.4. *Any partition of S defines induces an equivalence relation on S via the relation that s is related to t if and only if both s and t belong to the same piece of the partition. Conversely, any equivalence relation on S induces a partition of S, where each piece of the partition is one of the equivalence classes.*

Proof. Exercise. ∎

2. Proof Strategy

Suppose that P is a proposition, let us consider the statement: *Prove* P. Since P is a proposition, it can either be *true* or *false*. When asked to *Prove* P, the goal is to create an argument which implies that P must take the value *true*. In the discussion that follows, it is often useful to interpret a proposition P as an implication. Indeed, any proposition P is logically equivalent to the implication $\textit{true} \implies P$. To convince yourself of this, note that $\textit{true} \implies P$ is *true* if and only if P is true, which is enough to conclude that they must be equivalent propositions.

The two basic strategies for "*Prove* P":

- (Directly) Create a direct argument that P is *true*.
- (Contradiction) Conclude that $\neg P$ is *false* by assuming that $\neg P$ is *true* and deriving a logical contradiction.

Note that P is true if and only if $\neg P$ is false, and therefore there is no logical difference whether your argument shows that P is *true* or $\neg P$ is *false*. In particular, if you can find a direct argument that $\neg P$ is *false*, this is the same as a direct proof that P is *true*.

Example 0.2.1. Consider the following: *Suppose that x is an integer such that $x < 2$. Prove that $x < 3$.* Although a bit silly and contrived, let's argue in two ways. Note here that the proposition P is that $x < 3$, and hence $\neg P$ is the proposition that $x \geq 3$.

– *Directly (P is true)*: Since $x < 2$ and $2 < 3$, transitivity of the inequality implies that $x < 3$.

– *Contradiction*: Assume that $x \geq 3$. Since $x < 2$ we have that $3 \leq x < 2$ which, by transitivity of the inequality, implies that $3 < 2$, a contradiction. Therefore, x is not ≥ 3, implying that $x < 3$.

From this vantage point, we can start building up proof strategies for more intricate propositions P. We list several here which is by no means an exhaustive list. However, we trust that understanding how to build these up will allow one to do so for most propositions not included.

2.1. Proofs with Implications.

We now start adding structure to the propositions we will try to prove. We start with a familiar lemma to anyone who has had some experience with logic.

LEMMA 0.2.2. *Let P and Q be propositions. The following are equivalent.*

(1) $P \implies Q$.

(2) $\neg Q \implies \neg P$.

(3) $(P \wedge \neg Q) \implies \text{false}$.

Proof. The equivalence between (1) and (3) is standard and easily deduced using, for example, a truth table. For the equivalence between (1) and (3), we could also argue with a truth table. However, it may be more instructive to argue as follows. Moreover, $P \implies Q$ is equivalent to $\mathit{truth} \implies (P \implies Q)$. From here, the equivalence of (1) and (3) implies that $\mathit{true} \implies (P \implies Q)$ is equivalent to $\neg(P \implies Q) \implies \neg \mathit{true}$. With $\neg(P \implies Q)$ equivalent to $P \wedge \neg Q$, and with $\neg \mathit{true}$ equivalent to *false*, we discover that $P \implies Q$ is equivalent to $(P \wedge \neg Q) \implies \mathit{false}$. ■

The two basic strategies applied to "*Prove* $P \implies Q$":

- (Directly)
  - Create a direct argument showing that $P \implies Q$ is *true*. It suffices to assume that P is *true* and argue that this assumption implies that Q is *true*.
  - (Contrapositive) Create a direct argument that $\neg Q \implies \neg P$ is *true*. In doing so, it suffices to assume that $\neg Q$ is *true* and argue that this assumption implies that $\neg P$ is *true*.
- (Contradiction) Conclude that $P \wedge \neg Q$ is *false* by assuming that $P \wedge \neg Q$ is *true* and deriving a contradiction.

$$(\exists x \in S) P(x) \iff \bigvee_{x \in S} P(x)$$

$$(\forall x \in S) P(x) \iff \bigwedge_{x \in S} P(x)$$

3. Induction

THEOREM 0.3.1. *The following are equivalent:*

1) (The Well-Ordering Principle) If $S \subset \mathbb{N}$ is non-empty, then S has a smallest element. That is, if $S \subset \mathbb{N}$ is non-empty, there exists some $n_0 \in S$ such that for all $n \in S$, it holds that $n_0 \leq n$.

2) (Weak Induction) If $S \subset \mathbb{N}$ is such that $1 \in S$, and $n+1 \in S$ whenever $n \in S$, then $S = \mathbb{N}$.

3) (Strong Induction) If $S \subset \mathbb{N}$ *such that* $1 \in S$ *and* $n+1 \in S$ *whenever* $k \in S$ *for all integers* k *in the range* $1 \leq k \leq n$*, then* $S = \mathbb{N}$.

Proof. (1 $\implies$ 2) Suppose that 1) holds and that $S \subset \mathbb{N}$ satisfying $1 \in S$ and $n \in S \implies n+1 \in S$. Let $T = \mathbb{N} \setminus S$. We want to show that $T = \varnothing$, which will imply that $S = \mathbb{N}$. We will argue by contradiction. Assume $T \neq \varnothing$. By the Well Ordering Principle there is some least element $n_0 \in T$. Since $1 \in S$, $n_0 \geq 2$; this implies $n_0 - 1 \in \mathbb{N} \setminus T$. Since $\mathbb{N} \setminus T = S$, we conclude that $n_0 - 1 \in S$. By the assumptions on S, we have that $n_0 - 1 \in S \implies (n_0 - 1) + 1 \in S$, showing that $n_0 \in S$. This is a contradiction since $n_0 \in T$ and T was defined such that S and T are disjoint. Therefore, $T = \varnothing$.

(2 $\implies$ 3) Exercise.

(3 $\implies$ 1) Assume that 3) holds. Let $S \subset \mathbb{N}$. We will prove that the well ordering principle holds by proving that if S does not have a least element, then $S = \varnothing$; that is, we will prove the contrapositive. To this end, assume that S does not have a least element. Let $T = \mathbb{N} \setminus S$. Since S doesn't have a least element, certainly $1 \notin S$, since any subset of $\mathbb{N}$ containing 1 has a least element (namely, 1). Therefore $1 \in T$. Now, suppose that $n \geq 1$ and $k \in T$ for all integers k in the range $1 \leq k \leq n$. If we can show that this implies that $n+1 \in T$, then by 3) we will conclude that $T = \mathbb{N}$ and hence $S = \varnothing$, finishing the proof. If $n+1 \notin T$ then $n+1 \in S$; however, this means that $n+1$ is the least element in S since $k \in T$ for all k in the range $1 \leq k \leq n$, yet S is assumed not to have a least element. Therefore $n+1 \in T$, and we're done. ■

4. Rational Numbers

Roughly stated, a rational number is any number which can be expressed as the ratio of any integer p by any non-zero integer q; i.e., $r \in \mathbb{Q}$ when $r = p/q$ for some $p \in \mathbb{Z}$ and some $q \in \mathbb{Z} \setminus \{0\}$. From a different, but equivalent, perspective, the set of rational numbers is a field (see Definition 1.1.1 below for the definition of a field) which is the "smallest" field which contains the integers as a subset.

To give a rigorous construction of the rational numbers, we can do the following. Put the equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ defined such that $(p, q) \sim (r, s)$ if and only if $ps = qr$. We then define the rational numbers $\mathbb{Q}$ to be the set of equivalence classes $\mathbb{Q} = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\}) / \sim$, with addition $+$ defined by $[(p, q)] + [(r, s)] = [(ps + rq, qs)]$, multiplication $\cdot$ defined by $[(p, q)] \cdot [(r, s)] = [(pr, qs)]$, and ordering defined by $[(p, q)] < [(r, s)]$ if and only if either q

and s have the same sign and $ps < qr$, or q and s have different signs and $ps > qr$. Common notation dictates that the equivalence class $[(p,q)]$ be denoted $\frac{p}{q}$ or p/q. Note that, for example, $(3,6) \sim (1,2)$, implying $[(3,6)] = [(1,2)]$, which is the same statement as $\frac{3}{6} = \frac{1}{2}$ in common notation. The integers are considered a subset of $\mathbb{Z}$ by identifying the element $p \in \mathbb{Z}$ with the equivalence class $\frac{p}{1}$.

We will not spend much time in these notes working rigorously to develop all our well-known properties and manipulations of rational numbers. However, most of these are proved more generally for ordered fields in the next chapter, so long as we believe that the rational numbers are an ordered field; fear not, they are. That said, the interested reader can work through Exercise 5 (13) to get delve deeper into these properties which we will henceforth take for granted.

5. Exercises

(1) Prove D'Morgan's Laws: For any set I and indexed collection of sets $\{S_i\}_{i\in I}$,
   (a) $\left[\bigcup_{i\in I} S_i\right]^c = \bigcap_{i\in I} S_i^c$.
   (b) $\left[\bigcap_{i\in I} S_i\right]^c = \bigcup_{i\in I} S_i^c$.

(2) What is $2^{2^\varnothing}$? That is, what is the power set of the power set of the empty set?

(3) Given sets S and T and a function $f : S \to T$, define the *graph* of f to be the set $G_f \subset S \times T$ given by $G_f = \{(s, f(s)) \text{ s.t. } s \in S\}$.
   (a) Is it true in general that for all $s \in S$, if $(s,t) \in G_f$ and $(s,r) \in G_f$, then $t = r$. Prove this if it is true, or give a counter example if not.
   (b) When is it true that $G_f = S \times f(S)$?

(4) Suppose that S is a set with n elements (where $n \in \mathbb{N}$). Prove that the number of elements in 2^S is 2^n. *Hint*: You might try an induction on the number of elements in S.

(5) Let $f : \mathbb{Z} \to \mathbb{Z}$ be the function defined by $f(p) = p^2$.
   (a) What is $f^{-1}(\{1\})$?
   (b) What is $f^{-1}(\{0\})$?
   (c) What is $f^{-1}(\{-1\})$?

   Make sure to briefly justify your answers.

(6) Let $S \subset \mathbb{Z}$ and $T \subset \mathbb{Z}$ and $f : S \to T$ be defined by $f(s) = s^2$. With brief justification, decide if f is injective, surjective, or bijective when:

(a) $S = T = \mathbb{Z}$.

(b) $S = \mathbb{Z}$, $T = \mathbb{N} \cup \{0\}$.

(c) $S = \mathbb{N}$, $T = \mathbb{Z}$.

(d) $S = \{-n \text{ s.t. } n \in \mathbb{N}\}$, $T = \mathbb{N}$.

(e) $S = \{-1, 1\}$, $T = \{1\}$.

Note that in all these problems, the assignment f remained the same, but your answers changed depending on S and T. This shows why to define a function, you really need to know the domain and codomain, and not just the assignment f.

(7) Finish the proof of Proposition 0.1.3.

(8) It often happens that an object is defined that at a first glance seems to be a function, but in reality it does not satisfy the definition of a function. When this is case, we say that *the function is not well defined*; otherwise, if the object does satisfy the definition of a function (and therefore *is* a function), we say that the function is *well defined*. In the following parts, we give "functions" which are not well defined. Describe why the function is not well defined.

(a) $f : \mathbb{Z} \to \mathbb{N}$ is defined by $f(p) = p^2$.

(b) $f : \mathbb{Q} \to \mathbb{Z}$ is defined by $f\left(\frac{p}{q}\right) = p + q$.

(c) $f : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ is defined by $f\left(\frac{p}{q}, \frac{r}{s}\right) = \frac{p+r}{q+s}$.

(9) Prove the implication $(2 \implies 3)$ in Theorem 0.3.1.

(10) For all $r, s \in \mathbb{Q}$ with $r < s$, let $((r, s))$ denote the set $\{t \in \mathbb{Q} \text{ s.t. } r < t < s\}$. Let $I = ((0, 1))$. Consider the indexed collection of sets $\{S_r\}_{r \in I}$, where for every $r \in I$, $S_r = \left(\left(r - \frac{1}{2}, r + \frac{1}{2}\right)\right)$. Prove that $\bigcup_{r \in I} S_r = ((-1/2, 3/2))$ and $\bigcap_{r \in I} S_r = \{1/2\}$.

(11) For any pair of sets S and T, let us write $S \sim T$ when there is a bijection $f : S \to T$. Show that $\sim$ is reflexive, symmetric, and transitive; i.e., for all sets S it holds that $S \sim S$; for all sets S and T it holds that $S \sim T$ implies $T \sim S$; for all sets S, T, and U it holds that $S \sim T$ and $T \sim U$ implies $S \sim U$. Does this prove that $\sim$ defines an equivalence relation on the collection of all sets?

(12) Suppose that S is a set and R is an equivalence relation on S defined by $\sim$.

(a) Show that if $s, t \in S$ then $s \sim t$ if and only if $[s] = [t]$.

(b) Show that $S/\sim$ is a partition of S.

(13) Referring to the construction of $\mathbb{Q}$ given in Section 4.

(a) Show that $(p,q) \sim (r,s) \iff ps = qr$ does, in fact, define an equivalence relation on $\mathbb{Z} \times (\mathbb{Z}\backslash\{0\})$.

(b) Show that the addition $[(p,q)] + [(r,s)] = [(ps+rq, qs)]$ is well defined in $\mathbb{Q}$. That is, suppose that $(p_1,q_1) \sim (p_2,q_2)$ and $(r_1,s_1) \sim (r_2,s_2)$, show that $[(p_1,q_1)]+[(r_1,s_1)] = [(p_2,q_2)]+[(r_2,s_2)]$. How does this notion of "well defined" agree with that introduced in Exercise 8?

(c) Show that the multiplication $[(p,q)]\cdot[(r,s)] = [(pr,qs)]$ is well defined in $\mathbb{Q}$. That is, suppose that $(p_1,q_1) \sim (p_2,q_2)$ and $(r_1,s_1) \sim (r_2,s_2)$, show that $[(p_1,q_1)] \cdot [(r_1,s_1)] = [(p_2,q_2)] \cdot [(r_2,s_2)]$. How does this notion of "well defined" agree with that introduced in Exercise 8?

(d) Define the *absolute value* function $|\cdot| : \mathbb{Q} \to \mathbb{Q}$ by $\big|[(p,q)]\big| = [(|p|,|q|)]$. Show that the absolute value function is well defined. That is, suppose that $(p,q) \sim (r,s)$, show that $\big|[(p,q)]\big| = \big|[(r,s)]\big|$. How does this notion of "well defined" agree with that introduced in Exercise 8?

(e) Suppose that p, q, and r are positive integers such that $q \leq r$. Prove that $[(p,r)] \leq [(p,q)]$. Use this to deduce that for any rational number $\frac{p}{q}$, it holds that $\frac{p}{q} \leq \frac{|p|}{|q|} \leq |p|$.

(14) This problem is to help understand the *ordered field* structure of $\mathbb{Q}$. Define the positive rationals as the set $P_{\mathbb{Q}} = \{r \in \mathbb{Q} \text{ s.t. } r > 0\}$.

(a) Let $r, s \in P_{\mathbb{Q}}$. Prove that $r + s \in P_{\mathbb{Q}}$.

(b) Define the set $-P_{\mathbb{Q}}$ as the negative rationals, i.e., $-P_{\mathbb{Q}} = \{r \in \mathbb{Q} \text{ s.t. } r < 0\}$. Prove that the sets $P_{\mathbb{Q}}, -P_{\mathbb{Q}}$, and $\{0\}$ are mutually disjoint.

(c) Prove that $\mathbb{Q} = -P_{\mathbb{Q}} \cup \{0\} \cup P_{\mathbb{Q}}$.

(15) Prove the following (useful) equalities by induction.

(a) $\displaystyle\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.

(b) $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.

(c) $\displaystyle\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}$.

(16) Prove the following (useful) inequalities by induction.

(a) For every integer $n \geq 4$, $n! > 2^n$.

(b) If $a \geq 0$, then for every $n \in \mathbb{N}$, $(1+a)^n \geq 1 + na$.

(17) Prove that $\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}$ in two ways:
   (a) By induction.
   (b) Using partial fractions.

CHAPTER 1

Ordered Fields with an Introduction to the Real Numbers

In this chapter we introduce the basic theory of ordered fields, culminating at the definition of the real numbers; the unique (up to relabelling) supremum complete ordered field. Up to this point, the rational numbers have been constructed, and according to Exercise 14 in Section 5, the rational numbers are themselves an ordered field. In fact, we will learn that any ordered field has a copy (read: potentially relabeled version) of the rational numbers as a subset; this is a rigorous interpretation of the statement that the rational numbers are the "smallest" ordered field. We will also see that the rational numbers are incomplete insofar as every bounded subset of rational numbers does not necessarily have a *least upper bound* or *greatest lower bound*; in essence, the absence of these special bounds creates "gaps" in the rational numbers, leading to another perspective of the real numbers as the ordered field which "fills the gaps" in the rational numbers.

1. Fields

DEFINITION 1.1.1. A non-empty set $\mathbb{F}$ closed under two binary operations, addition $+$ and multiplication $\times$, is called a *field* whenever the following conditions are met:

(F1) (Associativity of Addition and Multiplication) For all $a, b, c \in \mathbb{F}$, $(a + b) + c = a + (b + c)$ and $(a \times b) \times c = a \times (b \times c)$.

(F2) (Commutativity of Addition and Multiplication) For all $a, b \in \mathbb{F}$, $a + b = b + a$ and $a \times b = b \times a$.

(F3) (Additive and Multiplicative Identities) There exist elements $0_{\mathbb{F}}, 1_{\mathbb{F}} \in \mathbb{F}$ such that for every $a \in \mathbb{F}$, $a + 0_{\mathbb{F}} = a$ and $a \times 1_{\mathbb{F}} = a$.

(F4) (Additive and Multiplicative Inverses) For every $a \in \mathbb{F}$ with $a \neq 0_{\mathbb{F}}$ there exist $b, c \in \mathbb{F}$ such that $a + b = 0_{\mathbb{F}}$ and $a \times c = 1_{\mathbb{F}}$. We will write $b = -a$ and $c = a^{-1}$ or $c = 1/a$.

(F5) (Distributive Property) For all $a, b, c \in \mathbb{F}$, $a \times (b + c) = (a \times b) + (a \times c)$.

REMARKS 1.1.2.

(1) The assumption of the existence of binary operations of addition and multiplication on $\mathbb{F}$ can be understood rigorously as the existence of functions $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ and $\times : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ where we write $+(a, b) = a + b$ and $\times(a, b) = a \times b$.

(2) The distributive property (F5) is what actually differentiates addition and multiplication; it is also the only property which exposes how addition and multiplication interact.

(3) Given a set with a single element $\mathbb{F} = \{a\}$, we can define addition and multiplication as $a + a = a$ and $a \times a = a$. With these operations, we will find that $\mathbb{F}$ is a field with $0_{\mathbb{F}} = 1_{\mathbb{F}} = a$. However, this field isn't interesting. We call a field with only one element a *trivial field*. From here on, we will assume that all fields are not trivial fields, hence all fields considered will have at least two elements.

NOTATION 1.1.3. We will begin to use familiar notation. For one, we will drop the subscript $\mathbb{F}$ from the identity elements, writing 0 and 1 instead of $0_{\mathbb{F}}$ or $1_{\mathbb{F}}$; we also reserve the symbols 0 and 1 as the additive and multiplicative identities, respectively, with or without mention. For another, we won't continue to use the multiplication symbol $\times$; rather, we use juxtaposition ab or "dot" notation $a \cdot b$ to symbolize multiplication. Further, we will will often write $a - b$ instead of $a + (-b)$ and say that we are subtracting b; and we will write a/b instead of ab^{-1} and say that we are dividing by b.

We will start exploring the algebraic manipulations on a field $\mathbb{F}$ with a quick proposition about how the additive identity $0 \in \mathbb{F}$ interacts with multiplication, which matches our experience with the zero we are familiar with.

PROPOSITION 1.1.4. *Let $\mathbb{F}$ be a field and $a \in \mathbb{F}$. Then $a \cdot 0 = 0$. Moreover, if $b \in \mathbb{F}$, then $ab = 0$ if and only if $a = 0$ or $b = 0$.*

Proof. Let $a \in \mathbb{F}$. Then, using the definition of 0 and the distributive property, $a \cdot 0 = a \cdot (0 + 0) = (a \cdot 0) + (a \cdot 0)$. Subtracting $a \cdot 0$ from both sides results in the equality

$$(a \cdot 0) - (a \cdot 0) = ((a \cdot 0) + (a \cdot 0)) - (a \cdot 0)$$

The lefthand side is 0; by associativity, the righthand side becomes $(a \cdot 0) + ((a \cdot 0) - (a \cdot 0)) = a \cdot 0 + 0 = a \cdot 0$. Therefore, we find $0 = a \cdot 0$, proving the first claim. Next, suppose that $b \in \mathbb{F}$.

If either $a = 0$ or $b = 0$, then $a \cdot b = 0$ by what we have just proved. Conversely, if $ab = 0$ and $a \neq 0$, then we can multiply both sides of the equality by a^{-1} to find $a^{-1}(ab) = a^{-1} \cdot 0$. The righthand side is 0; by associativity, the lefthand side is $(a^{-1}a)b = 1 \cdot b = b$. Thus $b = a^{-1}$, proving the second claim. ■

COROLLARY 1.1.5. *Let $\mathbb{F}$ be a field, $0 \in \mathbb{F}$ the additive identity and $1 \in \mathbb{F}$ the multiplicative identity. It holds that $0 \neq 1$.*

Proof. Since we always assume that $\mathbb{F}$ is not the trivial field, there exists at least one element $a \in \mathbb{F}$ with $a \neq 0$. By the definition of the multiplicative identity, $1 \cdot a = a \neq 0$. Proposition 1.1.4 then implies that $1 \neq 0$. ■

The importance of the next result – the uniqueness of identities and inverses – in the development of the algebraic behavior of a field $\mathbb{F}$ can not be overstated.

THEOREM 1.1.6 (Uniqueness of Identities and Inverses). *Let $\mathbb{F}$ be a field. and $a, b \in \mathbb{F}$ with $a \neq 0$. Then*

(1) *For any $a, b \in \mathbb{F}$, $a + b = a$ if and only if $b = 0$.*
(2) *For any $a, b \in \mathbb{F}$, $ab = a$ if and only if either $a = 0$ or $a \neq 0$ and $b = 1$.*
(3) *For any $a, b \in \mathbb{F}$, $a + b = 0$ if and only if $b = -a$.*
(4) *For any $a, b \in \mathbb{F}$, $a \cdot b = 1$ if and only if $a \neq 0$ and $b = a^{-1}$.*

Proof. (1) If $a + b = a$, then subtracting a from both sides leaves the righthand side 0. The lefthand side becomes $-a + (a + b) = (-a + a) + b = 0 + b = b$, where we used associativity along the way. By equating these, we have shown that $a + b = a$ implies $b = 0$. Conversely, if $b = 0$, then $a + b = a + 0 = a$; hence $b = 0$ implies $a + b = a$.

(2) If $ab = a$ and $a \neq 0$, then multiplying a^{-1} to both sides leaves the righthand side 1. The lefthand side becomes $a^{-1}(ab) = (a^{-1}a)b = 1 \cdot b = b$, where, as before, we used associativity. By equating these, we have shown that $ab = a$ implies $b = 1$. Conversely, if $a = 0$, then $ab = a$ for any $b \in \mathbb{F}$ by Proposition 1.1.4. If $a \neq 0$ and $b = 1$, then $ab = a \cdot 1 = a$. Hence either $a = 0$ or $a \neq 0$ and $b = 1$ implies $ab = a$.

(3) If $a + b = 0$, then subtracting a from both sides leaves the righthand side $-a$. As in the proof of (1), the lefthand side of the equality becomes b. By equating these, we have shown that $a + b = 0$ implies that $b = -a$. Conversely, if $b = -a$, then $a + b = a - a = 0$.

(4) If $ab = 1$, then Proposition 1.1.4 implies that $a \neq 0$. Multiplying a^{-1} to both sides leaves the righthand side a^{-1}. As in the proof of (2), the lefthand side becomes b. By

equating these, we have shown that $ab = 1$ implies that $a \neq 0$ and $b = a^{-1}$. Conversely, if $a \neq 0$ and $b = a^{-1}$, then $ab = aa^{-1} = 1$. ■

COROLLARY 1.1.7. *Let $\mathbb{F}$ be a field and $1 \in \mathbb{F}$ the multiplicative identity. Then $1^{-1} = 1$.*

Proof. Corollary 1.1.5 implies that $1 \neq 0$, and hence 1^{-1} exists. By the definition of multiplicative inverse, $1^{-1} \cdot 1 = 1$. But this implies that $1^{-1} = 1$ by the uniqueness of the multiplicative identity. ■

COROLLARY 1.1.8. *For every $a \in \mathbb{F}$, it holds that $-(-a) = a$ and, if $a \neq 0$, $(a^{-1})^{-1} = a$.*

Proof. Let $a \in \mathbb{F}$. Then, since $-(-a)$ is the additive inverse of $-a$ by definition, we have that $-(-a) + (-a) = 0$. However, $a + (-a) = 0$ since $-a$ is the additive inverse of a. This shows that $-(-a)$ and a both are additive inverses of a, which by uniqueness implies that $a = -(-a)$.

If $a \neq 0$, then a^{-1} exists and $a^{-1} \neq 0$ (because if $a^{-1} = 0$, $1 = aa^{-1} = a \cdot 0 = 0$, a contradiction). Therefore $(a^{-1})^{-1}$ exists and is, by definition, the multiplicative inverse of a. But then, $aa^{-1} = 1$ and $(a^{-1})^{-1}a^{-1} = 1$, which implies that $a = (a^{-1})^{-1}$ by the uniqueness of multiplicative inverses. ■

COROLLARY 1.1.9. *Let $\mathbb{F}$ be a field and $a, b \in \mathbb{F}$. The following hold.*

(1) $-a = (-1) \cdot a$.
(2) $-(a + b) = (-a) + (-b)$.
(3) $-(ab) = (-a)b = a(-b)$.
(4) $(-a)(-b) = ab$.

Proof. (1) Let $a \in \mathbb{F}$. Then $a + (-1) \cdot a = 1 \cdot a + (-1) \cdot a = (1 + (-1)) \cdot a = 0 \cdot a = 0$. By the uniqueness of additive inverses, this implies that $(-1) \cdot a = -a$.

(2) Let $a, b \in \mathbb{F}$. By what we have just proved and the distributive property,

$$-(a + b) = (-1)(a + b) = (-1) \cdot a + (-1) \cdot b = (-a) + (-b).$$

(3) Let $a, b \in \mathbb{F}$. Then, by the distributive property, $ab + (-a)b = (a + (-a))b = 0 \cdot b = 0$. By the uniqueness of additive inverses, this proves that $(-a)b = -(ab)$. The symmetric argument also shows that $a(-b) = -(ab)$.

(4) Let $a, b \in \mathbb{F}$. Then by what we have just proved and Corollary 1.1.8, $(-a)(-b) = -(a(-b)) = -(-(ab)) = ab$. ■

COROLLARY 1.1.10. *Suppose that $\mathbb{F}$ is a field. The following hold.*

(1) *If $a, b \in \mathbb{F}$ are such that $ab \neq 0$, then $a \neq 0$, $b \neq 0$, and $(ab)^{-1} = a^{-1}b^{-1}$.*

(2) *If $a \in \mathbb{F}$ with $a \neq 0$, then $(-a)^{-1} = -(a^{-1})$.*

Proof. (1) Let $a, b \in \mathbb{F}$. If $ab \neq 0$, then $a \neq 0$ and $b \neq 0$ by Proposition 1.1.4. By associativity and commutativity, we have $(ab)(a^{-1}b^{-1}) = (aa^{-1})(bb^{-1}) = 1 \cdot 1 = 1$. By the uniqueness of multiplicative inverses, this implies that $(ab)^{-1} = a^{-1}b^{-1}$.

(2) Let $a \in \mathbb{F} \setminus \{0\}$. Arguing similar to Corollary 1.1.9 (4), we have that $(-a)(-(a^{-1}) = -((-a)(a^{-1})) = -(-(aa^{-1})) = aa^{-1} = 1$. By the uniqueness of multiplicative identities, this proves that $(-a)^{-1} = -(a^{-1})$. ■

One final collection of algebraic manipulations we mention here are a simple consequence of associativity and commutativity.

PROPOSITION 1.1.11. *Let $\mathbb{F}$ be a field and suppose that $a, b, c, d \in \mathbb{F}$.*

(1) *It holds that $(a+b)+(c+d) = ((a+b)+c)+d = (a+c)+(b+d) = (a+d)+(b+c)$.*

(2) *It holds that $(a \cdot b) \cdot (c \cdot d) = ((a \cdot b) \cdot c) \cdot d = (a \cdot c) \cdot (b \cdot d) = (a \cdot d) \cdot (b \cdot c)$.*

(3) *If $b \neq 0$, $c \neq 0$, and $d \neq 0$, then $(a/b)/(c/d) = (a \cdot d)/(b \cdot c)$.*

Proof. Exercise 8 (1). ■

At this point we have proved many results which, put roughly, allow us to use "highschool algebra" on fields. Henceforth we will take advantage of these familiar manipulations and will often do so without reference to the proof. In reality, the most important result of this section is the uniqueness of inverses and identities, Theorem 1.1.6; whereas, the results follow uniqueness are there to build an arsenal of familiar manipulations which we can use without reserve.

2. Ordered Fields

DEFINITION 1.2.1. Let $\mathbb{F}$ be a field and $P \subset \mathbb{F}$. The pair $(\mathbb{F}, P)$ is called and *ordered field* whenever the following hold:

(1) P is closed under addition and multiplication. That is, for every $a, b \in \mathbb{P}$, both $a + b \in \mathbb{P}$ and $a \cdot b \in \mathbb{P}$.

(2) The set $P, \{0\}$, and $-P$ partition $\mathbb{F}$, where $-P = \{-a \text{ s.t. } a \in P\}$.

If $(\mathbb{F}, P)$ is an ordered field, then we call the elements of $\mathbb{P}$ the *positive elements* or *positive numbers* of $\mathbb{F}$, and call the elements of $-P$ the *negative elements* or *negative numbers* of $\mathbb{F}$.

NOTATION 1.2.2. It is common to omit explicit mention of the set P and simply say that $\mathbb{F}$ is an ordered field, which implicitly suggests that there is some subset of positive numbers in consideration, even without explicit mention; however, to be extra careful, we sometimes write the pair $(\mathbb{F}, P)$ in what follows.

PROPOSITION 1.2.3. *The rational numbers $\mathbb{Q}$ are an ordered field with positive elements $P = \{r \in \mathbb{Q} \text{ s.t. } r > 0\}$.*

Proof. Exercise 5 (14). ■

DEFINITION 1.2.4. Consistent with our intuition at this point, for two elements $a, b \in \mathbb{F}$, we say that a and b have the *same sign* when $a, b \in P \cup \{0\}$ or $a, b \in -P \cup \{0\}$. That is, a and b have the same sign when either they are both positive, or both negative, or at least one of them is 0. Further, we will say that a and b have *opposite signs* when either $a \in P$ and $b \in -P$, or $a \in -P$ and $b \in P$.

THEOREM 1.2.5. *Let $(\mathbb{F}, P)$ be an ordered field.*

(1) *Let $a \in \mathbb{F}$. Then*
   i) $a \in P \iff -a \in -P$.
   ii) $a \in -P \iff -a \in P$.
(2) $1_{\mathbb{F}} \in P$.
(3) *If $a, b \in -P$ then $a + b \in -P$.*
(4) *Let $a \in P$ and $b \in \mathbb{F}$. Then*
   i) $b \in P \iff a \cdot b \in P$.
   ii) $b \in -P \iff a \cdot b \in -P$.
(5) *Let $a \in -P$ and $b \in \mathbb{F}$. Then*
   i) $b \in P \iff a \cdot b \in -P$.
   ii) $b \in -P \iff a \cdot b \in P$.
(6) *Let $a \in \mathbb{F}$. Then*
   i) $a \in P \iff a^{-1} \in P$.
   ii) $a \in -P \iff a^{-1} \in -P$.
(7) *If $a, b \in \mathbb{F}$ then one and only one of the following hold: $a - b \in P$, $b - a \in P$, or $a - b = 0$.*
(8) *If $a, b \in \mathbb{F}$ have the same sign and are non-zero, then $b - a \in P$ if and only if $a^{-1} - b^{-1} \in P$.*

Proof. (1) i) If $a \in P$, then $-a \in -P$ by definition. If $-a \in -P$, then by the definition of P, there exists some $b \in P$ such that $-a = -b$, this then implies that $a = b$, so $a \in P$.
ii) If $a \in -P$, then there exists some $b \in P$ such that $a = -b$. Then $-a = -(-b) = b$ by Theorem **??**. Therefore $-a = b \in P$.

(2) By (1), $1 \in P$ if and only if $-1 \in -P$, so it suffices to prove that $-1 \in -P$. Since $-1 \neq 0$, and $-P$, $\{0\}$, P forms a partition of $\mathbb{F}$, then either $-1 \in -P$ or $-1 \in P$. If $a \in P$, then $(-1) \cdot a = -a \in -P$. Therefore, $-1 \notin P$ (since by the definition of an ordered field, if $-1 \in P$ and $a \in P$, then $(-1) \cdot a \in P$, which we just showed was not true); hence $-1 \in -P$.

(3) If $a, b \in -P$, then $-a, -b \in P$ by (1). Therefore, $(-a) + (-b) \in P$. We have that $(-a) + (-b) = -(a + b)$ by Theorem **??** (**??**). Moreover, $-(-(a + b)) = a + b$ by Theorem **??** (**??**). From this we deduce that $-((-a) + (-b)) = a + b$, and since $(-a) + (-b) \in P$, then $-((-a) + (-b)) = a + b \in -P$.

(4) If $b \in P$, then $a \cdot b \in P$ by the definition of an ordered field. We will prove the converse by contradiction. Suppose that $a \cdot b \in P$ and $b \notin P$. Then either $b = 0$ or $b \in -P$. If $b = 0$ then $a \cdot b = 0 \notin P$, so $b \neq 0$. This leaves $b \in -P$. If $b \in -P$, the $-b \in P$ by (1). Since $a \in P$ by assumption and $-b \in P$, then by the definition of an ordered field, $a \cdot (-b) \in P$. However, $a \cdot (-b) = -(a \cdot b)$ by Theorem **??** (**??**). Since we assumed that $a \cdot b \in P$, then $-(a \cdot b) \in -P$. From this we deduce that $-(a \cdot b) \in (-P) \cap P = \varnothing$, which is our contradiction. Therefore, $a \cdot b \in P \implies b \in P$.

(5), (6), and (7) are exercises.

(8) If $a, b \in P$ the $a^{-1}, b^{-1} \in P$, or if $a, b \in -P$, then $a^{-1}, b^{-1} \in -P$. In either case, $a^{-1} \cdot b^{-1} \in P$. Therefore, $a - b \in P$ if and only if $a^{-1}b^{-1}(a - b) \in P$. Using the same methods as we have many times above, $a^{-1}b^{-1}(a - b) = (a^{-1}a)b^{-1} - (b^{-1}b)a^{-1} = b^{-1} - a^{-1}$. Therefore, $a - b \in P$ if and only if $b^{-1} - a^{-1} \in P$. ∎

2.1. Inequalities on Ordered Fields.

Definition 1.2.6. Let $(\mathbb{F}, P)$ be an ordered field. Given $a, b \in \mathbb{F}$, we say that a is *less than* b, written $a < b$, whenever $b - a \in P$. We say that a is *greater than* b, written $a > b$, whenever $a - b \in P$. Along these same lines, we say that a is *less than or equal to* b, written $a \leq b$, whenever $b - a \in P$ or $b - a = 0$. We say that a is *greater than or equal to* b, written $a \geq b$, whenever $a - b \in P$ or $a - b = 0$.

Lemma 1.2.7. *Let $(\mathbb{F}, P)$ be an ordered field and suppose $a \in \mathbb{F}$. Then*

(1) $a \in P \iff 0 < a$.

(2) $a \in -P \iff a < 0$.

Proof. (1) The notation $0 < a$ is defined to mean that $a - 0 \in P$. But, $a - 0 = a$, so $a \in P$. On the other hand, if $a \in P$, then $a - 0 \in P$ since $a = a - 0$, therefore $0 < a$.

(2) If $a \in -P$, then $-a = -a + 0 \in P$. But, $-a + 0 = 0 - a$, so since $0 - a \in P$, then $a < 0$. On the other hand, if $a < 0$ then $0 - a \in P$, which is equivalent to $-a \in P$, but then $a \in -P$ by Theorem 1.2.5 (1). ■

Let us spend a moment to rewrite Theorem 1.2.5 in the inequality notation. We have

THEOREM 1.2.8. *Let $(\mathbb{F}, P)$ be an ordered field.*

(1) *Let $a \in \mathbb{F}$. Then*

i) $0 < a \iff -a < 0$.

ii) $a < 0 \iff 0 < -a$.

(2) $0 < 1_{\mathbb{F}}$.

(3) *If $a < 0$ and $b < 0$ then $a + b < 0$.*

(4) *Let $0 < a$ and $b \in \mathbb{F}$. Then*

i) $0 < b \iff 0 < a \cdot b$.

ii) $b < 0 \iff a \cdot b < 0$.

(5) *Let $a < 0$ and $b \in \mathbb{F}$. Then*

i) $0 < b \iff a \cdot b < 0$.

ii) $b < 0 \iff 0 < a \cdot b$.

(6) *Let $a \in \mathbb{F}$. Then*

i) $0 < a \iff 0 < a^{-1}$.

ii) $a < 0 \iff a^{-1} < 0$.

(7) *If $a, b \in P$ then one and only one of the following hold: $b < a$, $a < b$, or $a = b$.*

(8) *If $a, b \in \mathbb{F}$ have the same sign and are non-zero, then $a < b$ if and only if $a^{-1} > b^{-1}$.*

Proof. The proof of this is just using our inequality notation and applying the results of Theorem 1.2.5. ■

PROPOSITION 1.2.9. *Let $(\mathbb{F}, P)$ be an ordered field. Then*

(1) *If $a, b, c \in \mathbb{F}$ are such that $a < b$ and $b < c$, then $a < c$. The same can be said for $\leq$. That is, the transitive property holds for the relations $<$ and $\leq$ on $(\mathbb{F}, P)$. Moreover, if either $a < b$ and $c \leq d$ or $a \leq b$ and $c < d$, then $a < d$.*

(2) *$a \leq b$ and $b \leq a$ if and only if $a = b$.*

(3) *Let $a, b, c \in \mathbb{F}$ and suppose $a < b$. Then*

i) $a \cdot c < b \cdot c \iff c > 0$.

ii) $a \cdot c > b \cdot c \iff c < 0$.

iii) $a \cdot c = b \cdot c \iff c = 0$.

(4) *If $a, b, c, d \in \mathbb{F}$ are such that $a \leq b$ and $c < d$, then $a + c < b + d$. If instead we assumed that $c \leq d$, then $a + c \leq b + d$.*

Proof. (1) If $a < b$ and $b < c$, then $b - a \in P$ and $c - b \in P$. Therefore, $c - a = c - b + b - a = (c - b) + (b - a) \in P$, since the addition of two positive numbers is a positive number. But $c - a \in P$ is equivalent to $a < c$ by the definition of $<$. The analogous proof works for $\leq$ by interchanging $P \cup \{0\}$ rather than P. For the case that $a < b$ and $b \leq c$, then either $b < c$, in which case we have already shown that if $a < c$, or $b = c$, in which case $a < b$ is equivalent to $a < c$ as $b = c$. The analogous argument works when $a \leq b$ and $b < c$.

(2) If $a \leq b$ then $b - a \in P \cup \{0\}$. If also $b \leq a$ then $a - b \in P \cup \{0\}$, further implying that $b - a = -(a - b) \in -P \cup \{0\}$. Therefore if both $a \leq b$ and $b \leq a$ we have that $b - a$ is either in both P and $-P$, or $b - a = 0$. Since $-P \cap P = \varnothing$, it must be that $b - a = 0$, which concludes the proof.

(3) Exercise.

(4) If $a \leq b$ and $c < d$, then $b - a \in P \cup \{0\}$ and $d - c \in P$. Therefore $(d - c) + (b - a) \in P$, and since $(d - c) + (b - a) = (b + d) - (a + c)$; this means $a + c < b + d$. If instead $c \leq d$, then $d - c \in P \cup \{0\}$, and arguing analogously, we find that $(b + d) - (a + c) \in P \cup \{0\}$; this means $a + c \leq b + d$. ∎

2.2. Interval Notation. Armed with inequalities, we introduce the familiar notation of intervals. Let $\mathbb{F}$ be an ordered field with $a, b \in \mathbb{F}$. We have

$$(a, b) = \{r \in \mathbb{F} \text{ s.t. } a < r < b\}, \qquad [a, b) = \{r \in \mathbb{F} \text{ s.t. } a \leq r < b\},$$

$$(a, b] = \{r \in \mathbb{F} \text{ s.t. } a < r \leq b\}, \qquad [a, b] = \{r \in \mathbb{F} \text{ s.t. } a \leq r \leq b\}$$

Notice that we did not assume that $a < b$ here; however, if $a > b$, then by definition, each of these intervals is $\varnothing$. The same is true if $a = b$, except for $[a, a] = \{a\}$.

We also define infinite intervals

$$(-\infty, a) = \{r \in \mathbb{F} \text{ s.t. } r < a\}, \qquad (-\infty, a] = \{r \in \mathbb{F} \text{ s.t. } r \leq a\},$$
$$(a, \infty) = \{r \in \mathbb{F} \text{ s.t. } a < r\}, \qquad [a, \infty) = \{r \in \mathbb{F} \text{ s.t. } a \leq r\}$$

3. Absolute Value and Distance on Ordered Fields

3.1. Absolute Value. Once given an ordered field, we get the notion of an *absolute value* freely. The absolute value is not only analytically useful, but is so intuitively as well. Indeed, the absolute value gives us a notion of length or magnitude, as well as further letting us define a distance between elements within the field.

DEFINITION 1.3.1. Let $(\mathbb{F}, P)$ be an ordered field. The *absolute value* is a function $|\cdot| : \mathbb{F} \to \mathbb{F}$ defined as

$$|a| = \begin{cases} a & 0 \leq a \\ -a & a \leq 0 \end{cases} \tag{3.1}$$

THEOREM 1.3.2. *The following hold.*

(1) *For all $a \in \mathbb{F}$, $|a| \geq 0$ with equality if and only if $a = 0$.*
(2) *For all $a, b \in \mathbb{F}$, $|a \cdot b| = |a| \cdot |b|$.*
(3) *(Triangle Inequality) For all $a, b \in \mathbb{F}$, $|a + b| \leq |a| + |b|$ with equality if and only if either $a \geq 0$ and $b \geq 0$, or $a \leq 0$ and $b \leq 0$.*

Proof. (1) Let $a \in \mathbb{F}$. If $a > 0$, then $|a| = a > 0$. If $a = 0$, then $|a| = a = 0$. If $a < 0$, then $|a| = -a$, and $-a > 0$. This shows that if $a \neq 0$ then $|a| > 0$; otherwise, if $a = 0$ then $|a| = 0$. In particular, we see that $|a| \geq 0$ with equality if and only if $a = 0$.

(2) Let $a, b \in \mathbb{F}$ be arbitrary. First consider the case that a and b have the same sign. Then $a \cdot b \geq 0$ implying that $|a \cdot b| = a \cdot b$. If $a \geq 0$ and $b \geq 0$, then $|a| = a$ and $|b| = b$ implying that $|a \cdot b| = a \cdot b = |a| \cdot |b|$. If $a \leq 0$ and $b \leq 0$, then $|a| = -a$ and $|b| = -b$ implying that $|a| \cdot |b| = (-a) \cdot (-b) = a \cdot b = |a \cdot b|$. Alternatively, suppose the case that a and b have opposite signs. Then $a \cdot b \leq 0$, implying that $|a \cdot b| = -(a \cdot b)$. Without losing generality, assume that $a < 0$ and $b > 0$. Then $|a| = -a$ and $|b| = b$, and hence $|a \cdot b| = -(a \cdot b) = (-a) \cdot b = |a| \cdot |b|$. Having considered all cases, we conclude that $|a \cdot b| = |a| \cdot |b|$.

(3) Let $a, b \in \mathbb{F}$. Consider first the case that a and b have the same sign. If $a \geq 0$ and $b \geq 0$, then $a + b \geq 0$, implying that $|a + b| = a + b = |a| + |b|$. If $a \leq 0$ and $b \leq 0$, then $a + b \leq 0$, implying that $|a + b| = -(a + b) = (-a) + (-b) = |a| + |b|$. Next, consider the case that a and b have opposite signs. Without losing generality, assume that $a < 0$ and $b > 0$. In this case $a < |a|$ implying $a + b < |a| + b = |a| + |b|$; also $-b \leq b$ implying $(-a) + (-b) < (-a) + b = |a| + |b|$. Since $|a+b| = a+b < |a|+|b|$ or $|a+b| = -(a+b) = (-a)+(-b) < |a|+|b|$, we have deduced that if a and b have opposite signs, then $|a + b| < |a| + |b|$. Since if a and b have the same sign, we found $|a + b| = |a| + |b|$, or if a and b have opposite signs, we found $|a + b| < |a| + |b|$, we conclude that $|a + b| \leq |a| + |b|$ with equality if and only if a and b have the same sign. ■

PROPOSITION 1.3.3. *Let $(\mathbb{F}, P)$ be an ordered field and $a \in \mathbb{F}$. The following hold.*

(1) *If $r \in \mathbb{F}$ with $r > 0$, then $|a| < r$ if and only if $-r < a < r$. Similarly, $|a| \leq r$ if and only if $-r \leq a \leq r$.*
(2) *If $r, s \in \mathbb{F}$ with $r \geq 0$ and $s \geq 0$, then $-r < a < s$ implies that $|a| < \max\{r, s\}$. If either $-r \leq a < s$ or $-r < a \leq s$ or $-r \leq a \leq s$, then $|a| \leq \max\{r, s\}$.*

Proof. (1) Suppose that $r > 0$ and $|a| < r$. Then either $a = a$, in which case $0 \leq a < r$. Otherwise, if $a = -a$, then $0 < -a < r$, which rearranges to $-r < a < 0$. In either case, this shows that $|a| < r$ implies $-r < a < r$. Conversely, suppose that $-r < a < r$. This implies immediately that $a < r$, and by taking the negative through the inequality $-a < r$. Since $|a|$ is either equal to a or $-a$, we find that $-r < a < r$ implies $|a| < r$. We have finished the proof for the strict inequality $<$, the proof for the non-strict inequality $\leq$ is done the same by simply replacing the inequalities throughout the argument.

(2) Suppose that r and s are non-negative and that $-r < a < s$. Then this implies immediately that $a < s \leq \max\{r, s\}$, and by taking the negative through the inequality $-a < r \leq \max\{r, s\}$. Since $|a|$ is either a or $-a$, we conclude that $|a| < \max\{r, s\}$. The argument with any of the strict inequalities swapped with a non-strict inequality is argued analogously. ■

3.2. Distance. The absolute value on an ordered field $\mathbb{F}$ is an extremely useful tool, in large part because it induces a distance. Much of the results we prove concerning the structure in $\mathbb{R}$ only require a notion of distance, rather than the full power of an absolute value, so it is prudent to introduce and learn to work with the distance.

DEFINITION 1.3.4. Let $(\mathbb{F}, P)$ be an ordered field. The *distance* induced from the absolute value is a function $\text{dist} : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ defined by $\text{dist}(a, b) = |a - b|$.

Since the definition of distance is completely determined by the absolute value, we can get an analogous theorem as Theorem 1.3.2.

THEOREM 1.3.5. *The following hold.*

(1) *For all* $a, b \in \mathbb{F}$, $\text{dist}(a, b) \geq 0$ *with equality if and only if* $a = b$.
(2) *(Symmetry) For all* $a, b \in \mathbb{F}$, $\text{dist}(a, b) = \text{dist}(b, a)$.
(3) *(Triangle Inequality) For all* $a, b, c \in \mathbb{F}$, $\text{dist}(a, c) \leq \text{dist}(a, b) + \text{dist}(b, c)$.

Proof. Exercise 8 (14). ■

REMARK 1.3.6. One proof of the triangle inequality for the distance emphasizes a commonly used manipulation when bounded the distance between elements. Indeed, if we want to get an estimate on the distance between two elements a and c based on their relative distances to some third element b, we can do so by a judicious addition of 0,

$$\text{dist}(a, b) = |a - c| = |a - b + b - c| \leq |a - b| + |b - c| = \text{dist}(a, b) + \text{dist}(b, c).$$

□

Throughout our work in $\mathbb{R}$ and later in a more general setting, we will alternate between considering sets of elements within some distance of one another and intervals. We now present two results that will help us with this manipulations in ordered fields.

LEMMA 1.3.7. *Suppose that* $\mathbb{F}$ *is an ordered field and* $x, a \in \mathbb{F}$. *If there exist* $r \geq 0$ *and* $s \geq 0$ *such that* $x \in (a - r, a + s)$, *then* $\text{dist}(x, a) < \max\{r, s\}$. *If instead we assumed that either* $x \in (a - r, a + s]$ *or* $x \in [a - r, a + s)$ *or* $x \in [a - r, a + s]$, *then* $\text{dist}(x, a) \leq \max\{r, s\}$.

Proof. By definition $x \in (a - r, a + s)$ if and only if $a - r < x < a + s$. By subtracting a from each term in the latter inequality, $-r < x - a < s$. Now, by Proposition 1.3.3, $\text{dist}(x, a) = |x - a| < \max\{r, s\}$. The immediate analogue follows for the $\leq$ case. ■

LEMMA 1.3.8. *Let* $x, \varepsilon \in \mathbb{F}$ *with* $\varepsilon > 0$. *The interval* $(x - \varepsilon, x + \varepsilon)$ *equals the set* $\{y \in \mathbb{F}$ *s.t.* $\text{dist}(x, y) < \varepsilon\}$. *In other words,* $y \in (x - \varepsilon, x + \varepsilon)$ *if and only if* $\text{dist}(x, y) < \varepsilon$.

Proof. Let $y \in (x-\varepsilon, x+\varepsilon)$. Then, by Lemma 1.3.7, $\text{dist}(x,y) < \varepsilon$. Conversely, if $\text{dist}(x,y) < \varepsilon$, then $-\varepsilon < x - y < \varepsilon$, implying that $x - \varepsilon < y < x + \varepsilon$ and hence $y \in (x-\varepsilon, x+\varepsilon)$. ■

We finish this section with one final result which illustrates a common technique in analysis of proving that two elements are equal using distance estimations.

LEMMA 1.3.9. *Let $\mathbb{F}$ be an ordered field and $a, b \in \mathbb{F}$. Then $a = b$ if and only if for every positive $\varepsilon \in \mathbb{F}$, it holds that* $\text{dist}(a,b) < \varepsilon$.

Proof. Suppose first that $a = b$. Then $\text{dist}(a,b) = 0$, and therefore $\text{dist}(a,b) < \varepsilon$ for any positive $\varepsilon \in \mathbb{F}$. We will prove the converse by proving the contrapositive statement. Suppose that there exists some positive $\varepsilon \in \mathbb{F}$ such that $\text{dist}(a,b) \geq \varepsilon$. Then, since $\varepsilon > 0$, we find that $\text{dist}(a,b) > 0$, implying that $a \neq b$ by Theorem 1.3.5 (1). ■

4. Common Notation and the Inclusion of $\mathbb{Q}$

In this section we argue that within any ordered field $\mathbb{F}$, there exists a copy of the rationals $\mathbb{Q}$. Are argument is correct in essence, but skips some of the detailed technicalities that would blur to confusion the overall theme.

DEFINITION 1.4.1. We say that an ordered field $\mathbb{F}$ contains a *copy* of the rational numbers $\mathbb{Q}$ when there is a subset $\mathbb{Q}_{\mathbb{F}}$ of $\mathbb{F}$ such that there is a bijection between the rationals $\mathbb{Q}$ and the copy $\mathbb{Q}_{\mathbb{F}}$ which preserves the structure of $\mathbb{Q}$. That is, for every element $r \in \mathbb{Q}$, we can associate an element $r_{\mathbb{F}} \in \mathbb{Q}_{\mathbb{F}}$ in such a way that:

(1) (Preserves addition) For all $t, r, s \in \mathbb{Q}$, it holds that $t = r+s$ if and only if $t_{\mathbb{F}} = r_{\mathbb{F}} + s_{\mathbb{F}}$.
(2) (Preserves multiplication) For all $t, r, s \in \mathbb{Q}$, it holds that $t = r \cdot s$ if and only if $t_{\mathbb{F}} = r_{\mathbb{F}} \cdot s_{\mathbb{F}}$.
(3) (Preserves ordering) For all $r, s \in \mathbb{Q}$, it holds that $r < s$ if and only if $r_{\mathbb{F}} < s_{\mathbb{F}}$.

In the case that $\mathbb{F}$ contains a copy of $\mathbb{Q}$, we will often say that $\mathbb{Q}$ is a subset of $\mathbb{F}$ *up to relabelling* of elements.

THEOREM 1.4.2. *Every ordered field contains a copy of* $\mathbb{Q}$.

Proof. The argument goes as follows, we will argue that if $\mathbb{F}$ is an ordered field, then $\mathbb{F}$ must contain a copy of $\mathbb{N}$ and hence a copy of $\mathbb{Z}$. (Here we are using the word *copy* analogously as it is defined above). From here, we will show that there must be a copy of $\mathbb{Q}$ in $\mathbb{F}$.

To start, we know that $1_{\mathbb{F}} \in \mathbb{F}$. Hence, for any $n \in \mathbb{N}$, we have that $\underbrace{1_{\mathbb{F}} + \cdots + 1_{\mathbb{F}}}_{n \text{ terms}}$ is an element of $\mathbb{F}$. Define $n_{\mathbb{F}}$ as this sum of $1_{\mathbb{F}}$ with itself n times. This gives us the copy of $\mathbb{N}$ in $\mathbb{F}$ since the set $\mathbb{N}_{\mathbb{F}}$ defined by $\mathbb{N}_{\mathbb{F}} = \{n_{\mathbb{F}} \text{ s.t. } n \in \mathbb{N}\}$ preserves all the structure (addition, multiplication, ordering) of $\mathbb{N}$. (Note that the ordering of $\mathbb{F}$ is being used here since this ordering implies that for $n, m \in \mathbb{N}$, it holds that $n < m$ if and only if $n_{\mathbb{F}} < m_{\mathbb{F}}$.)

Next, since $\mathbb{F}$ is a field, each element in $\mathbb{N}_{\mathbb{F}}$ admits an additive inverse. Therefore, we can define a copy of the negative integers by $(-n)_{\mathbb{F}} = -n_{\mathbb{F}}$ for any $n \in \mathbb{N}$. With a copy of the negative and positive integers in $\mathbb{F}$, we then have a copy of the integers in $\mathbb{F}$ by letting $\mathbb{Z}_{\mathbb{F}}$ be the union of the negative and positive integers along with the zero element $0_{\mathbb{F}}$. As before, $\mathbb{Z}_{\mathbb{F}}$ preserves all the structure (addition, multiplication, ordering) of $\mathbb{Z}$.

At this point we have argued that every ordered field contains a copy of the integers. Finally, to conclude that a copy of the rationals exist in $\mathbb{F}$, let us note that for all non-zero elements in $\mathbb{Z}_{\mathbb{F}}$, there exists a multiplicative inverse. So, for every $r \in \mathbb{Q}$, we have that $r = n/m$ for integers $n \in \mathbb{Z}$ and $m \in \mathbb{Z} \setminus \{0\}$. We then define the element $r_{\mathbb{F}} \in \mathbb{F}$ by $r_{\mathbb{F}} = n_{\mathbb{F}}/m_{\mathbb{F}}$, which makes sense because we have already developed a copy of the integers in $\mathbb{F}$. Finally, we create the copy of the rationals $\mathbb{Q}_{\mathbb{F}}$ by collecting all such $r_{\mathbb{F}}$; that is, define $\mathbb{Q}_{\mathbb{F}} = \{r_{\mathbb{F}} \text{ s.t. } r \in \mathbb{Q}\}$. This gives us the copy a copy of $\mathbb{Q}$ in $\mathbb{F}$, which you can readily confirm preserves the structure (addition, multiplication, ordering) of $\mathbb{Q}$ ■

NOTATION 1.4.3. As before when we dropped the subscript $\mathbb{F}$ from the 0 and 1 elements, we will do the same for elements in $\mathbb{Q}_{\mathbb{F}}$. In particular, since we have shown that $\mathbb{F}$ contains $\mathbb{Q}$ up to relabelling of elements, we will assume that the copy of $\mathbb{Q}$ inside of $\mathbb{F}$ uses the same labels we are used to and write statements such as $2/3 \in \mathbb{F}$, rather than $2_{\mathbb{F}}/3_{\mathbb{F}} \in \mathbb{F}$.

5. Archimedes and the Density of $\mathbb{Q}$

Throughout this section, $(\mathbb{F}, P)$ is an ordered field.

5.1. Boundedness and the Archimedean Property.

DEFINITION 1.5.1. Let $U \subset \mathbb{F}$ be non-empty. We say that U is *bounded above* if there exists some $a \in \mathbb{F}$ such that for all $u \in U$, it holds that $u \leq a$. In this case we call a an *upper bound* of U. Similarly, we say that U is *bounded below* if there exists some $b \in \mathbb{F}$ such that for all $u \in U$, it holds that $b \leq u$. In this case, we call b a *lower bound* of U. If either U is not bounded above, or not bounded below, we can say that U is *unbounded*.

For any $p/q \in \mathbb{Q}$ where $p, q \in \mathbb{Z}$, the inequality $p/q \leq |p|$ implies that the set of natural numbers are an unbounded subset of $\mathbb{Q}$. Indeed, for any $p/q \in \mathbb{Q}$, let $n = |p| + 1 \in \mathbb{N}$, then $p/q < n$. This is an extremely useful property that we explore throughout this section and with implications throughout the remainder of this text.

DEFINITION 1.5.2. The ordered field $(\mathbb{F}, P)$ satisfies the *Archimedean property* when $\mathbb{N} \subset \mathbb{F}$ is unbounded above. That is, $(\mathbb{F}, P)$ satisfies the Archimedean property when for all $a \in \mathbb{F}$, there exists some $n \in \mathbb{N}$ such that $a < n$.

THEOREM 1.5.3 (Some Equivalences of the Archimedean Property). *Let $(\mathbb{F}, P)$ be an ordered field. Then the following are equivalent.*

(1) *$(\mathbb{F}, P)$ satisfies the Archimedean Property. That is, for all $a \in \mathbb{F}$, there exists an $n \in \mathbb{N}$ such that $a < n$.*
(2) *For all $a \in \mathbb{F}$ with $a > 0$, there exits an $n \in \mathbb{N}$ such that $\frac{1}{n} < a$.*
(3) *For all $a, b \in \mathbb{F}$, if $a \leq b + \frac{1}{n}$ for every $n \in \mathbb{N}$, then $a \leq b$.*
(4) *For any $a \in \mathbb{F}$, if $|a| \leq \frac{1}{n}$ for every $n \in \mathbb{N}$, then $a = 0$.*
(5) *For all $\varepsilon, a \in \mathbb{F}$ such that $\varepsilon > 0$ and $a > 0$, there exists some $n \in \mathbb{N}$ such that $a < n \cdot \varepsilon$.*

Proof. In this proof, we will show that (1) $\implies$ (2), (2) $\implies$ (3), (3) $\implies$ (4), and (1) $\iff$ (5). In Exercise 8 (21), the implication (4) $\implies$ (1) is proved, completing the proof.

(1) $\implies$ (2). Assume that (1) holds. If $a > 0$, then a^{-1} exists, so there exists some $n \in \mathbb{N}$ such that $a^{-1} < n$ by the assumption of the Archimedean property. But then, $a > n^{-1} = 1/n$.

(2) $\implies$ (3). Assume that (2) holds. Let $a, b \in \mathbb{F}$ be such that for all $n \in \mathbb{N}$ it holds that $a \leq b + 1/n$. We then deduce that $a - b \leq 1/n$ for every $n \in \mathbb{N}$. If $a - b > 0$, then by (2), we can find some $n \in \mathbb{N}$ such that $1/n < a - b$. But then, $1/n < a - b \leq 1/n$, which as we argued previously, is impossible. Therefore, $a - b \not> 0$, leaving $a - b \leq 0$. Thus, $a \leq b$.

(3) $\implies$ (4). Let $a \in \mathbb{F}$ such that $|a| \leq 1/n$ for every $n \in \mathbb{N}$. By assuming (3) with $b = 0$, this implies that $|a| \leq 0$. By the definition of the absolute value, $|a| \geq 0$, and hence $|a| = 0$. This implies $a = 0$, finishing the proof.

(1) $\iff$ (5). Assume that (1) holds. Let $\varepsilon, a \in \mathbb{F}$ be positive arbitrary elements. Then there exists some $n \in \mathbb{N}$ such that $n > a \cdot \varepsilon^{-1}$; multiplying on both sides of the inequality by ε (leaving the direction of the inequality unchanged, since $\varepsilon > 0$), this rearranges to $n \cdot \varepsilon > a$.

Conversely, assume that (5) holds. Let $a \in \mathbb{F}$. Using (5), we can find some $n \in \mathbb{N}$ such that $|a| < n \cdot 1$ (by choosing $\varepsilon = 1$). Since $a \leq |a|$ and $n \cdot 1 = n$, we have that $a < n$. Since $a \in \mathbb{F}$ was arbitrary, we have proved that the Archimedean Property holds. ■

This next proposition is quite intuitive and will be used later. It requires the Well Ordering Principle of the natural numbers, and the resulting consequences which you will prove in the exercises below; namely, that a subset of $\mathbb{Z}$ which is bounded below has a minimal element.

PROPOSITION 1.5.4. *Suppose that $(\mathbb{F}, P)$ satisfies that Archimedean property. Then, for all $a, b \in \mathbb{F}$ such that $a + 1 < b$, there exists some $p \in \mathbb{Z}$ such that $a < p < b$. Colloquially, when the Archimedean principle is satisfied, any two elements which are further apart that 1 must have an integer between them.*

Proof. Suppose that $a, b \in \mathbb{F}$ are such that $a + 1 < b$. Let $U = \{p \in \mathbb{Z} \text{ s.t. } b \leq p\}$. Since b is a lower bound of U and U is non-empty by the Archimedean property (i.e., there is some $n \in \mathbb{N}$ such that $b < n$, implying that $n \in U$), Exercises 8 (20) implies there is a minimal element $p_0 \in U$. Set $p = p_0 - 1$. Since p_0 is the minimal element in U and $p < p_0$, then $p \notin U$, meaning that $p < b$. On the other hand, $a + 1 < b \leq p_0$, so $a < p_0 - 1 = p$. Therefore $a < p < b$. ■

5.2. Density. Here we will briefly introduce the notion of density within an ordered field, and prove that the density of the rational numbers is equivalent to the Archimedean property.

DEFINITION 1.5.5. Let $S \subset \mathbb{F}$. We say that S is *dense* in $\mathbb{F}$ when for every $a, b \in \mathbb{F}$, if $a < b$ then there exists some $r \in S$ such that $a < r < b$.

LEMMA 1.5.6. *Any ordered field is a dense subset of itself.*

Proof. As $\mathbb{F}$ is an arbitrary ordered field, let $a, b \in \mathbb{F}$ and assume $a < b$. Since $\mathbb{Q} \subset \mathbb{F}$, and $0 < 1/2 < 1$, we have that $a < a + (a + b)/2 < a + (a + b) = b$. Therefore, by setting $r = (a + b)/2$, we have found a point $r \in \mathbb{F}$ such that $a < r < b$. ■

THEOREM 1.5.7 (Density of the Rationals). *The ordered field $\mathbb{F}$ satisfies the Archimedean property if and only if $\mathbb{Q}$ is dense in $\mathbb{F}$. That is, $\mathbb{F}$ satisfies the Archimedean property if and only if for all $a, b \in \mathbb{F}$, if $a < b$, then there exists some rational number $r \in \mathbb{F}$ such that $a < r < b$.*

Proof. Assume first that $(\mathbb{F}, P)$ satisfies the Archimedean Property. Let $a, b \in \mathbb{F}$ be such that $a < b$; in particular, $b - a > 0$. By the assumption of the Archimedean Property and Theorem 1.5.3 (5), there exists some $n \in \mathbb{N}$ such that $n \cdot (b - a) > 1$; in particular, $a \cdot n + 1 < b \cdot n$. By Proposition 1.5.4, we can find some $p \in \mathbb{Z}$ such that $n \cdot a < p < n \cdot b$. Dividing through by n, we have $a < p/n < b$, and by construction, $p/n \in \mathbb{Q}$. ■

COROLLARY 1.5.8. *Let $\tilde{\mathbb{F}}$ be another ordered field such that $\tilde{\mathbb{F}} \subset \mathbb{F}$. If $\mathbb{F}$ satisfies the Archimedean property, then $\tilde{\mathbb{F}}$ is dense in $\mathbb{F}$.*

Proof. Since $\tilde{\mathbb{F}}$ is an ordered field, we have that $\mathbb{Q} \subset \tilde{\mathbb{F}}$. If $\mathbb{F}$ satisfies the Archimedean property, then $\mathbb{Q}$ is a dense subset of $\mathbb{F}$, which implies that $\tilde{\mathbb{F}}$ is also by Exercise 8 (32). ■

6. Supremums and Infimums

DEFINITION 1.6.1. Let $(\mathbb{F}, P)$ be an ordered field and suppose that $U \subset \mathbb{F}$ is non-empty. If U is bounded above and there exists some $a \in \mathbb{F}$ such that a is an upper bound of U and whenever $c \in \mathbb{F}$ is any other upper bound of U, it holds that $a \leq c$, then we call a the *least upper bound* of U, or more commonly, the *supremum* of U. Analogously, if U is bounded below and there exists some $b \in \mathbb{F}$ such that b is a lower bound of U and whenever $c \in \mathbb{F}$ is any other lower bound of U, it holds that $c \leq b$, then we say that b is the *greatest lower bound* of U, or more commonly, the *infimum* of U.

There is nothing unique about being an upper bound to a set U. Indeed, if a is an upper bound of U, then so is $a + b$ for every non-negative b. However, there is uniqueness of supremums and infimums.

PROPOSITION 1.6.2. *Let $(\mathbb{F}, P)$ be an ordered field. Suppose that $U \subset \mathbb{F}$ is non-empty. If the supremum of U exists, then it is unique. Similarly, if the infimum of U exists, then it is unique.*

Proof. Suppose that $a, a' \in \mathbb{F}$ are both supremums of U. Then $a \leq a'$ since a' is an upper bound of U and a is the supremum; symmetrically, $a' \leq a$ since a is an upper bound of U and a' is the supremum. Since $a \leq a'$ and $a' \leq a$, then $a = a'$.

The analogous argument work for the infimum and is left as an exercise. ■

NOTATION 1.6.3. If the supremum of some sete U exists, we write $\sup(U)$ as the supremum. Similarly, if the infimum of a set U exists, then we write $\inf(U)$ for the infimum. Since

the supremum and infimum are both unique (if they exist), there is no ambiguity about which element $\sup(U)$ or $\inf(U)$ represent.

EXAMPLE 1.6.4. For every integer $n \geq 2$, let $U_n = \left\{r \in \mathbb{Q} \text{ s.t. } -\frac{1}{n} \leq r < 1 - \frac{1}{n}\right\}$. Define $U = \bigcup_{n=2}^{\infty} U_n$. We will show that the infimum and supremum of U both exist in $\mathbb{Q}$, and that $\inf(U) \in U$ while $\sup(U) \notin U$.

We first show that $\inf(U) = -\frac{1}{2}$ and $-\frac{1}{2} \in U$. To proof this, let us observe that $-1/2 \in U_2$, which further implies that $-1/2 \in U$. Moreover, if b is any lower bound of U, then $b \leq -1/2$ since $-1/2 \in U$. Notice that for any integer $n \geq 2$, $-1/2 \leq -1/n$, from which we deduce that $-1/2$ is a lower bound of U_n, and therefore $-1/2$ is a lower bound of U. Since $-1/2$ is a lower bound of U, and for any other lower bound b of U, it holds that $b \leq -1/2$, we have proved that $-1/2 = \inf(U)$.

We next prove that $\sup(U) = 1$ and $1 \notin U$. By design, for any integer $n \geq 2$, $1 - 1/n$ is an upper bound of U_n and $1 - 1/n < 1$, so $1 \notin U_n$. Since $1 \notin U_n$ for any integer $n \geq 2$, then $1 \notin U$ by the definition of U. This also shows that for every integer $n \geq 2$, 1 is an upper bound of U_n and hence an upper bound of U. To show that 1 is the least upper bound, we will argue by contradiction. Suppose that $a \in \mathbb{Q}$ is an upper bound of U such that $a < 1$. Then $0 < 1 - a$, so, by the Archimedean property (which we have proved holds in $\mathbb{Q}$), there exists some $n \in \mathbb{N}$ such that $1/n < 1 - a$. But then, $a < 1 - 1/n$. By the density of the rational numbers, there exists some $r \in \mathbb{Q}$ such that $a < r < 1 - 1/n$. In particular, $r \in U_n$ and thusly in U, but a is smaller than r, contradicting that a is an upper bound of U. So, $1 \leq a$, proving that 1 is the least upper bound, and therefore we conclude $\sup(U) = 1$. □

EXAMPLE 1.6.5 (Supremums don't always exist). Consider the ordered field $\mathbb{Q}$, and let $U = \left\{r \in \mathbb{Q} \text{ s.t. } r^2 \leq 2\right\}$. Then we claim that the supremum of U does not exist as an element of $\mathbb{Q}$. We readily check that U is bounded above (by 2, say) and is non-empty (for example, $1 \in U$). We will prove that $\sup(U)$ does not exist in $\mathbb{Q}$. We break this into two pieces.

We first prove that if $s = \sup(U) \in \mathbb{Q}$ exists, then $s^2 = 2$. We will argue by contradiction. Suppose that the supremum s exists, but $s^2 \neq 2$. Immediately we know that $s \geq 1$ since $1 \in U$ and s is an upper bound of U. If $s^2 > 2$, then using the Archimedean property, let $n \in \mathbb{N}$ be large enough such that $n \cdot (s^2 - 2) > 2s$. Rearranging this, we get $s^2 - \frac{2s}{n} > 2$. In particular, by adding $\frac{1}{n^2}$ to the lefthand side (making it larger), $s^2 - 2s + \frac{1}{n^2} > 2$. We then have $\left(s - \frac{1}{n}\right)^2 > 2$. This shows that $s - \frac{1}{n}$ must also be an upper bound of U, but $s - \frac{1}{n} < s$,

contradicting that s was the least upper bound. This contradiction shows that $s^2 \not> 2$. If $s^2 < 2$, then using a similar argument, we can apply the Archimedean property to find an $n \in \mathbb{N}$ large enough such that $\left(s + \frac{1}{n}\right)^2 < 2$. But this contradicts that s is an upper bound, since $s < s + \frac{1}{n}$, and $s + \frac{1}{n} \in U$ (because its square is less than 2). This contradiction shows that $s^2 \not< 2$. Therefore, since $s^2 \not< 2$ and $s^2 \not> 2$, which contradicts the fact that $s^2 \neq 2$. So, $s^2 = 2$.

We next finish this example by proving that there is no element $s \in \mathbb{Q}$ such that $s^2 = 2$. Again, we argue by contradiction. Suppose that there is some $p/q \in \mathbb{Q}$ such that $(p/q)^2 = 2$. We can assume that either p is even and q is odd, or p is odd and q is even (since we can divide off all common factors of 2). Now, if $(p/q)^2 = 2$, then $p^2 = 2q^2$, which shows that p^2 is even. However, p^2 is even if and only if p is even. Therefore, p is even. In particular, there is some $m \in \mathbb{Z}$ such that $p = 2m$. Now, $2q^2 = p^2 = 4m^2$, so $q^2 = 2m^2$. This shows that q^2 is even, which happens if and only if q is even. We have thus showed that if $(p/q)^2 = 2$, then p and q must both be even, which contradicts that we can assume that p or q is odd.

Hence we conclude that $\sup(U)$ does not exist in $\mathbb{Q}$. Indeed, we proved that if $s = \sup(U)$ is an element in $\mathbb{Q}$, then $s^2 = 2$, but we've also prove that there is no element $s \in \mathbb{Q}$ such that $s^2 = 2$. □

PROPOSITION 1.6.6. *Let $(\mathbb{F}, P)$ be an ordered field and $U \subset \mathbb{F}$ be non-empty. Define $-U = \{-u \text{ s.t. } u \in U\}$. Then $\sup(U)$ exists in $\mathbb{F}$ if and only if $\inf(-U)$ exists in $\mathbb{F}$. Moreover, if $\sup(U)$ exists, then $\inf(-U) = -\sup(U)$.*

Proof. Suppose that $s = \sup(U)$ exists in $\mathbb{F}$. Then for all $u \in U$, $u \leq s$. But then, for all $u \in U$, $-s \leq -u$, showing that $-s$ is a lower bound of $-U$. Moreover, if $l \in \mathbb{F}$ is any lower bound of $-U$, then for all $u \in U$, $l \leq -u$, showing that $u \leq -l$. This shows that $-l$ is an upper bound of U, and since s is the supremum of U, $s \leq -l$. This then implies that $l \leq -s$. Since $-s$ has been shown to be a lower bound of $-U$ such that for any other arbitrary lower bound l of $-U$ it holds that $l \leq -s$, we have proved that $-s = \inf(-U)$.

The analogous proof works when assuming that $\inf(-U)$ exists in $\mathbb{F}$ and is left as an exercise. ■

6.1. Infinite Supremums and Infimums. We briefly introduce intuitive notation here which builds a symbolic language we can use to describe supremums and infimums when the sets are unbounded.

NOTATION 1.6.7 (Sup and Inf for Unbounded Sets). Let $(\mathbb{F}, P)$ be an ordered field and $S \subset \mathbb{F}$. If S is unbounded above, we write $\sup(S) = \infty$. If S is unbounded below, we write $\inf(S) = -\infty$.

LEMMA 1.6.8. *Let $(\mathbb{F}, P)$ be an ordered field and $S \subset \mathbb{F}$. The following hold.*

(1) $\sup(S) = \infty$ *if and only if for every $r \in \mathbb{F}$ with $r > 0$, there exists some $s \in \mathbb{S}$ such that $r < s$.*
(2) $\inf(S) = -\infty$ *if and only if for every $r \in \mathbb{F}$ with $r > 0$, there exists some $s \in \mathbb{S}$ such that $s < -r$.*

Proof. These are directly from the definition of $\sup(S) = \infty$ and $\inf(S) = -\infty$, where we simply have explicitly written out what it means for S to be unbounded above or unbounded below. ■

NOTATION 1.6.9 (Sup and Inf for the Empty Set). The standard convention (for good reason) requires that we define $\sup(\varnothing) = -\infty$ and $\inf(\varnothing) = \infty$. Note that this is the opposite convention as for unbounded sets.

The intuition behind these notations for supremums and infimums goes as follows. To find the supremum of a set S, a subset of some ordered field $\mathbb{F}$, we start out to the left at $-\infty$ and move to the right towards $+\infty$ until we just hit the point at which no members of S can be found any further to the right. If S is unbounded above, then we must keep moving to the right until we hit $+\infty$. On the other hand, if $S = \varnothing$, then we're already done when we started at $-\infty$. The mirror intuition holds for the infimum, where in the infimum case, you are moving from the right starting at $+\infty$ to the left towards $-\infty$.

7. The Real Numbers

DEFINITION 1.7.1. An ordered field $(\mathbb{F}, P)$ is called *supremum complete* when for all non-empty subsets $U \subset \mathbb{F}$ which are bounded above, the supremum of U exists in $\mathbb{F}$.

Example 1.6.5 shows us that the ordered field $\mathbb{Q}$ is not supremum complete. Incredibly though, there does exist a supremum complete ordered field which satisfying the Archimedean Property. Moreover, such a field is essentially unique. The main result for this section is Theorem 1.7.4, proved below, after a few other interesting results that we need to build up along the way. The proofs of some of these results are not as technically complete as they

should be to make the argument truly rigorous; however, the outline and theme of each proof is correct and captures the truth and intuition of the results without obfuscating them with technicalities.

LEMMA 1.7.2. *Let $(\mathbb{F}, P)$ be an ordered field satisfying the Archimedean property. If $T \subset \mathbb{F}$ is any subset which is bounded above, there exists a subset $U \subset \mathbb{Q}$ such that the set of all upper bounds of T is the same as the set of all upper bounds of U, which further implies that $\sup(U)$ exists if and only if $\sup(T)$ exists, in which case $\sup(U) = \sup(T)$. This further implies that for every $a \in \mathbb{F}$, there exists a subset $U \subset \mathbb{Q}$ such that $\sup(U) = a$. (Note that, as always, we are using the copy of the rationals $\mathbb{Q}$ in the field $\mathbb{F}$ in our consideration).*

Proof. Let $T \subset \mathbb{F}$ be non-empty and bounded above. For each $a \in T$, define $U_a = \{r \in \mathbb{Q} \text{ s.t. } r \leq a\}$. Now, let $U = \bigcup_{a \in T} U_a$. We need to show that an element $b \in \mathbb{F}$ is an upper bound of $\mathbb{F}$ if and only if it is an upper bound of U. If b is an upper bound of T, then for every $a \in T$, $a \leq b$. But this means that for every $r \in U_a$, we have $r \leq a \leq b$, so b is an upper bound of U_a. Since this is true for every $a \in T$, it holds for every element in the union U of the sets U_a, and hence b is an upper bound of U. Conversely, suppose b is an upper bound of U. Let $a \in T$. For a contradiction, suppose that $a > b$. By the density of the rationals, there is some $r \in \mathbb{Q}$ such that $b < r < a$. But this implies that $r \in U_a$ since $r \leq a$, and hence $r \in U$. But this contradicts that b is an upper bound of U, since $b < r$. Therefore, $a \leq b$. Since $a \in T$ was arbitrary, this proves that b is an upper bound of T.

The remaining parts of the proof are an easy corollary to what we have just proved and left as Exercise (25). ■

THEOREM 1.7.3. *Let $\mathbb{F}$ be an ordered field satisfying the Archimedean property and suppose that $\mathbb{R}$ is a supremum complete ordered field satisfying the Archimedean property. Then there exists a copy of $\mathbb{F}$ as a subset of $\mathbb{R}$.*

Proof. After relabelling elements, we can assume that $\mathbb{Q} \subset \mathbb{F} \cap \mathbb{R}$. Let $a \in \mathbb{F}$. Define $T = \{a\}$, then clearly $\sup(T) = a$. By Lemma 1.7.2, there exists some set $U \subset \mathbb{Q}$ such that $\sup(U) = a$. Since we have assumed that $\mathbb{R}$ is supremum complete, then $\sup(U)$ exists in $\mathbb{R}$ and hence, after possibly relabelling the element, $a \in \mathbb{R}$. Therefore, $a \in \mathbb{F}$ implies that $a \in \mathbb{R}$, proving that $\mathbb{F} \subset \mathbb{R}$. ■

THEOREM 1.7.4. *Up to relabelling of elements, there exists exactly one supremum complete ordered field which satisfies the Archimedean Property. We call this field the* real numbers*, denoted by* $\mathbb{R}$.

Proof. Suppose that $\mathbb{R}'$ is another supremum complete ordered field satisfying the Archimedean Property. From Theorem 1.7.3 we have, after relabelling, $\mathbb{R}' \subset \mathbb{R}$. But the result of Theorem 1.7.3 will work just as well for $\mathbb{R}'$, and therefore $\mathbb{R} \subset \mathbb{R}'$. ■

A picture to have in mind is that if $\mathbb{F}$ is any ordered field, then, up to relabelling elements, $\mathbb{Q} \subseteq \mathbb{F}$. If further $\mathbb{F}$ satisfies the Archimedean property, then, up to relabelling of elements, $\mathbb{Q} \subseteq \mathbb{F} \subseteq \mathbb{R}$. In intuitive poetry, $\mathbb{Q}$ is the smallest ordered field, and $\mathbb{R}$ is the largest ordered field satisfying the Archimedean property.

7.1. Inherited Properties on $\mathbb{R}$.

THEOREM 1.7.5. *The rational numbers* $\mathbb{Q}$ *are a dense subset of* $\mathbb{R}$.

Proof. $\mathbb{R}$ is an ordered field satisfying the Archimedean property, therefore $\mathbb{Q}$ is dense in $\mathbb{R}$ by Theorem 1.5.7. ■

DEFINITION 1.7.6. The set $\mathbb{R} \setminus \mathbb{Q}$ is called the *irrational numbers*. If a is an element of the irrational numbers, we say that a is *irrational*.

THEOREM 1.7.7. *The set of irrational numbers is dense in* $\mathbb{R}$. *That is, given any two elements* $s, t \in \mathbb{R}$ *such that* $s < t$, *there exists an irrational number* $a \in \mathbb{R} \setminus \mathbb{Q}$ *such that* $s < a < t$.

Proof. In Exercises 8 (31), you prove that given any two rational numbers $p, q \in \mathbb{Q}$ such that $p < q$, the number $a = p + \frac{1}{\sqrt{2}}(q - p)$ is irrational and $p < a < q$. If $s, t \in \mathbb{R}$ with $s < t$, then by the density of the rationals, we can choose $p, q \in \mathbb{Q}$ such that $s < p < q < t$. Therefore, defining $a = p + \frac{1}{\sqrt{2}}(q - p)$ we have $s < p < a < q < t$, and since a is irrational, the proof is finished. ■

THEOREM 1.7.8. *Let* $S \subset \mathbb{R}$ *be a non-empty subset. Then* $\sup(S)$ *exists in* $\mathbb{R}$ *if and only if* S *is bounded above. Analogously,* $\inf(S)$ *exists in* $\mathbb{R}$ *if and only if* S *is bounded below.*

Proof. By definition, $\mathbb{R}$ is supremum complete. Therefore any set is bounded above (resp. bounded below) has a supremum (resp. infimum) in $\mathbb{R}$. On the other hand, if S is

not bounded above (resp. below), then there are no upper bounds (resp. lower bounds) of S, and hence has no supremum (resp. infimum). ■

DEFINITION 1.7.9. Let $x \in \mathbb{R}$ and $n \in \mathbb{N}$. Define the set $S = \{r \in \mathbb{Q} \text{ s.t. } r^n \leq x\}$. If S is non-empty, we define the nth root of x by $x^{1/n} = \sup(S)$. Otherwise, if S is empty, we say that the nth root of x does not exist in $\mathbb{R}$. As standard notation dictates, if $n = 2$ and $x^{1/2}$ exists, we will often write $\sqrt{x}$ instead of $x^{1/2}$; more generally, we can write $\sqrt[n]{x}$ for $x^{1/n}$, but rarely will within these notes.

7.2. Construction of $\mathbb{R}$.

8. Exercises

(1) (a) Prove Proposition 1.1.11 (1).
    (b) Prove Proposition 1.1.11 (2).
    (c) Prove Proposition 1.1.11 (3).
(2) Prove Theorem **??** (**??**).
(3) Prove Theorem **??** (**??**).
(4) Prove Theorem **??** (**??**).
(5) Prove Theorem **??** (**??**).
(6) Prove Theorem **??** (**??**).
(7) Prove Theorem **??** (**??**).
(8) Prove Proposition 1.2.9 (3).
(9) (a) If $a, b, c, d \in \mathbb{F}$ are all positive such that $a \leq b$ and $c < d$, then $a \cdot c < b \cdot d$. If instead we assumed that $c \leq d$, then $a \cdot c \leq b \cdot d$.
    (b) If $a, b, c, d \in \mathbb{F}$ are all negative such that $a \leq b$ and $c < d$, then $b \cdot d < a \cdot c$. If instead we assumed that $c \leq d$, then $b \cdot d \leq a \cdot c$.
(10) Let $(\mathbb{F}, P)$ be an ordered field.
    i) Prove that if $a \in \mathbb{F}$ and $b \in P$, then $a < a + b$.
    ii) Prove that if $a \in \mathbb{F}$ and $b \in -P$, then $a + b < a$.
    iii) Use i) (or ii) to prove that there are infinitely many elements in $\mathbb{F}$.
(11) Let $(\mathbb{F}, P)$. Suppose that $S \subset \mathbb{F}$ and $T \subset \mathbb{F}$ are such that for all $s \in S$ and for all $t \in T$, it holds that $s < t$. Prove that S and T are disjoint.
(12) Let $(\mathbb{F}, P)$ be an ordered field.

(a) Prove that for every $a \in \mathbb{F}$, it holds that $|a| = |-a|$.

(b) Prove that for every $a \in \mathbb{F}$, it holds that $a \leq 0$ if and only if $a = -|a|$.

(13) Let $(\mathbb{F}, P)$ be an ordered field. Prove the *reverse triangle inequality* for the absolute value: For all $a, b \in \mathbb{F}$, it holds that $\big||a| - |b|\big| \leq |a + b|$.

(14) Prove Theorem 1.3.5.

(15) Let $(\mathbb{F}, P)$ be an ordered field. Prove the *reverse triangle inequality* for the distance: For all $a, b \in \mathbb{F}$, it holds that $|\operatorname{dist}(a, b) - \operatorname{dist}(b, c)| \leq \operatorname{dist}(a, c)$.

(16) Let $(\mathbb{F}, P)$ be an ordered field and suppose that $r \in \mathbb{F}$ is positive. Prove that the function $|\cdot|_r : \mathbb{F} \to \mathbb{F}$ defined by $|a|_r = r \cdot |a|$ satisfies the claims of Theorem 1.3.2 (1) & (3), but that Theorem 1.3.2 (2) holds if and only if $r = 1$.

(17) Let $(\mathbb{F}, P)$ be an ordered field and suppose that $r \in \mathbb{F}$ is positive. Prove that the function $\operatorname{dist}_r : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ defined by $\operatorname{dist}_r(a, b) = r \cdot \operatorname{dist}(a, b)$ satisfies all the claims of Theorem 1.3.5.

(18) Fill in the missing details of Theorem **??**.

(19) Fill in the missing details of Theorem **??**.

(20) In this problem, you will prove the following proposition: *Let $(\mathbb{F}, P)$ be an ordered field satisfying the Archimedean property and $U \subset \mathbb{Z}$ is bounded below in $\mathbb{F}$ (i.e., there exists some $b \in \mathbb{F}$ such that for all $p \in U$, it holds that $b \leq p$). Then U has a minimal element.* You will prove this in a few steps.

(a) Prove that if $(\mathbb{F}, P)$ satisfies the Archimedean property, then for all $a \in \mathbb{F}$ there exists an $n \in \mathbb{N}$ such that $-n < a$.

(b) If $U \subset \mathbb{Z}$ is bounded below in $\mathbb{F}$, prove that there is some $n_0 \in \mathbb{N}$ such that for all $p \in U$, it holds that $-n_0 < p$.

(c) Using n_0 from part ii), let $\tilde{U} = \{p + n_0 \text{ s.t. } p \in U\}$ (basically, you are "moving U to the right" by n_0). Use the Well Ordering Principle on $\mathbb{N}$ to prove that $\tilde{U}$ has a minimal element $\tilde{p}_0$.

(d) Using n_0 and $\tilde{p}_0$ from part iii), let $p_0 = \tilde{p}_0 - n_0$. Prove that $p_0 \in U$ is the minimal element of U.

(21) Let $(\mathbb{F}, P)$ be an ordered field.

(a) In Theorem 1.5.3, prove that (4) $\implies$ (1), completing the proof of the theorem.

(b) In Theorem 1.5.3, prove directly that (2) $\implies$ (1).

(c) In Theorem 1.5.3, prove directly that (3) $\implies$ (1).

(22) Let $(\mathbb{F}, P)$ be an ordered field satisfying the Archimedean Property. Let $a \in \mathbb{F}$. Prove that the following four statements are equivalent:

(a) $a = 0$.

(b) For all $n \in \mathbb{N}$, $|a| \leq 1/n$.

(c) For all $\varepsilon \in \mathbb{Q}$ such that $\varepsilon > 0$, $|a| < \varepsilon$.

(d) For all $\varepsilon \in \mathbb{F}$ such that $\varepsilon > 0$, $|a| < \varepsilon$.

(23) In this problem you will prove that 3 is an upper bound to the set $U = \left\{ \sum_{k=0}^{n} \frac{1}{k!} \text{ s.t. } n \in \mathbb{N} \right\}$. You will do this in steps.

(a) Using Exercise 5 (16), show that for any integer $n \geq 4$,

$$\sum_{k=0}^{n} \frac{1}{k!} < \frac{8}{3} + \sum_{k=4}^{n} 2^{-k}.$$

(b) Prove that for all $n \geq 4$,

$$\sum_{k=4}^{n} 2^{-k} = 2\left(2^{-4} - 2^{-(n+1)}\right).$$

Hint: Note that

$$\frac{1}{2}\sum_{k=4}^{n} 2^{-k} = \sum_{k=4}^{n} 2^{-k} - \frac{1}{2}\sum_{k=4}^{n} 2^{-k} = \sum_{k=4}^{n} 2^{-k} - \sum_{k=5}^{n+1} 2^{-k}.$$

Many terms on the righthand side cancel.

(c) Deduce from part b that for all $n \geq 4$, $\sum_{k=4}^{n} 2^{-k} < \frac{1}{8}$. Hence, with the help of part a, conclude that 3 is an upper bound of the set U.

(24) Finish the proof of Proposition 1.6.6.

(25) Finish the proof of Lemma 1.7.2.

(26) Suppose that S and T are both subsets of the ordered field $\mathbb{F}$ such that $S \subset T$.

(a) If $\sup(S)$ and $\sup(T)$ exist, prove that $\sup(S) \leq \sup(T)$.

(b) If $\inf(S)$ and $\inf(T)$ exist, prove that $\inf(T) \leq \inf(S)$.

(27) Suppose that $(\mathbb{F}, P)$ is an ordered field and $S \subset \mathbb{F}$ is bounded above. Assume further that $\sup(S)$ exists in $\mathbb{F}$. Let $T \subset \mathbb{F}$ be the set of upper bounds of S (i.e., $b \in T$ if and only if b is an upper bound of S). Prove that $\inf(T)$ exists in $\mathbb{F}$ by proving that $\inf(T) = \sup(S)$.

(28) Let $(\mathbb{F}, P)$ be an ordered field and suppose that $S \subset \mathbb{F}$ and $T \subset \mathbb{F}$ are subsets such that $\inf(S), \sup(T)$, and $\inf(T)$ exist.

(a) Let $S+T$ be defined as the set $S+T=\{s+t \text{ s.t. } s\in S, t\in T\}$. Prove that $\inf(S+T)=\inf(S)+\inf(T)$. (In particular, $\inf(S+T)$ exists).

(b) Let $S-T$ be defined as the set $S-T=\{s-t \text{ s.t. } s\in S, t\in T\}$. Prove that $\inf(S-T)=\inf(S)-\sup(T)$. *Hint*: You could do this directly, but it might be easier to notice that $S-T=S+(-T)$ where $-T=\{-t \text{ s.t. } t\in T\}$, and then use part (a) along with Proposition 1.6.6.

(c) Let $a\in\mathbb{F}$. Define the set $a\cdot T$ as $a\cdot T=\{a\cdot t \text{ s.t. } t\in T\}$. Prove that if $a\geq 0$, then $\inf(a\cdot T)=a\cdot\inf(T)$; otherwise, if $a<0$, then $\inf(a\cdot T)=a\cdot\sup(T)$.

(29) For every $n\in\mathbb{N}$ with $n\geq 2$, let $U_n\subset\mathbb{Q}$ be defined by $U_n=\left\{r\in\mathbb{Q} \text{ s.t. } -\frac{1}{n}<r<1-\frac{1}{n}\right\}$. Let $U=\bigcap_{n=2}^{\infty}U_n$. Show that $\inf(U)$ and $\sup(U)$ both exist in $\mathbb{Q}$ by finding their values, and further show that $\inf(U)\in U$, whereas $\sup(U)\notin U$. *Hint*: It may help to look at Example 1.6.4.

(30) Let $(\mathbb{F},P)$ be an ordered field and suppose that $U\subset\mathbb{F}$ is such that $\sup(U)$ and $\inf(U)$ exist; in particular, U is non-empty. Prove that $\inf(U)=\sup(U)$ if and only if U is a singleton set (i.e., there exists some $u\in\mathbb{F}$ such that $U=\{u\}$).

(31) Let $a,b\in\mathbb{Q}$ with $a<b$ and define $\kappa=a+\frac{1}{\sqrt{2}}(b-a)$. Prove that $a<\kappa<b$ and $\kappa\in\mathbb{I}$ (the irrational numbers, see Definition 1.7.6).

(32) Suppose that $\mathbb{F}$ is an ordered field with subsets S and T. Suppose further that $S\subset T$. Prove that if S is dense in $\mathbb{F}$, then so is T.

(33) Let $\mathbb{F}$ be an ordered field. Suppose that $a,b\in\mathbb{F}$ with $a<b$.

(a) Prove that a is the infimum and b is the supremum of each of the intervals (a,b), $(a,b]$, $[a,b)$, and $[a,b]$. *Hint*: First prove this for (a,b). The remaining three cases are then easily deduced.

(b) What analogous statements can be said about infinite intervals (a,∞), $[a,\infty)$, $(-\infty,a)$, and $(-\infty,a]$?

Part 2

Exploring the Topological Structure of $\mathbb{R}$

CHAPTER 2

Sequences and Sequential Convergence in $\mathbb{R}$

1. General Sequences

This section introduces us to a fundamental concept in analysis, a *sequence*. Although we will concern ourselves mostly with sequences consisting of elements from a field, the concept of a sequence is more general and deserves attention. (If the reader feels unfamiliar with the terms *indexed collection*, *indexing set*, or *underlying set* [in the context of indexed collections], see Section 1.4).

DEFINITION 2.1.1 (Sequences: Indexed Collection Version). Let S be a set. A *sequence* in S, also called a *sequence of elements* in S, is an indexed collection of elements with indexing set $\mathbb{N}$ and underlying set S. That is, a collection of elements $\{x_n\}_{n\in\mathbb{N}}$ such that $x_n \in S$ for every $n \in \mathbb{N}$, is called a *sequence* in S.

DEFINITION 2.1.2 (Sequences: Function Version). Let S be a set. A *sequence* in S, also called a sequence of elements in S, is a function whose domain is the natural numbers $\mathbb{N}$ and whose codomain is S. That is, a function $\mathbf{x} : \mathbb{N} \to S$ is called a *sequence* in S.

REMARK 2.1.3. These two definitions for sequences agree. Indeed, if $\mathbf{x} : \mathbb{N} \to S$ is a sequence given by the function definition, then we have the corresponding sequence $\{x_n\}_{n\in\mathbb{N}}$ in S from the indexed collection definition by defining $x_n = \mathbf{x}(n)$ for all $n \in \mathbb{N}$; conversely, if $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in S given by the indexed collection version, then we can define the sequence $\mathbf{x} : \mathbb{N} \to S$ from the function definition by defining $\mathbf{x}(n) = x_n$ for all $n \in \mathbb{N}$. Therefore, you can see that the difference in definitions is not a fundamental mathematical divide, but rather two different perspectives referring to the same object of study. □

It is most common notationally to take the perspective of a sequence being an indexed collection. For this reason, we take a moment to review some notational conventions.

NOTATION 2.1.4. When indexing over $\mathbb{N}$, it is common practice to write $\overset{\infty}{n=1}$ rather than $_{n\in\mathbb{N}}$. For example, the sequence $\{x_n\}_{n\in\mathbb{N}}$ is more commonly denoted $\{x_n\}_{n=1}^{\infty}$. It is

also important to note that the symbol n is a dummy variable; the sequence $\{x_n\}_{n=1}^{\infty}$ can harmlessly be denoted $\{x_k\}_{k=1}^{\infty}$ or $\{x_\xi\}_{\xi=1}^{\infty}$, for example. Moreover, the symbol x is not sacred, we could write, "let $\{y_n\}_{n=1}^{\infty}$ be a sequence in S", or "let $\{A_n\}_{n=1}^{\infty}$ be a sequence in S", etc., and have no trouble in understanding what is being defined. However, unlike the "dummy" indexing character n, once the symbol x (or y or A or ...) has been introduced for a sequence, then it is the label for that sequence from then on unless explicitly changed. These appropriate uses of symbols within sequence notation is one arena where taking the function definition perspective is useful; indeed, if you have a function $x : \mathbb{N} \to S$, then you know that x is the label for that function (you wouldn't suddenly change the label to f, say), and you wouldn't worry about the variable symbol you use as the argument for x (e.g., $x(n)$ or $x(k)$ or $x(\xi)$), as long as you did so consistently within the appropriate context (e.g., writing "$x(n) = n^2$ for all $n \in \mathbb{N}$" is the same description as writing "$x(k) = k^2$ for all $k \in \mathbb{N}$", but is different than "$x(n) = k^2$ for all $n \in \mathbb{N}$"). Moreover, these notational conventions can be easily analogized to the setting of more general indexed collections.

2. Convergence of Sequences in $\mathbb{R}$

We move to the setting of the ordered field $\mathbb{R}$. We will frequently use the symbols $|\cdot|$ and dist as the absolute value and distance introduced in Sections 3.1 and 3.2.

DEFINITION 2.2.1. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. We say that the sequence $\{x_n\}_{n=1}^{\infty}$ *converges* when there exists an element $x_0 \in \mathbb{R}$ such that for all $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$, it holds that $\text{dist}(x_n, x_0) < \varepsilon$. In this case we say that the sequence $\{x_n\}_{n=1}^{\infty}$ converges to x_0; we call x_0 the *limit* of the sequence; and we write $\lim_{n\to\infty} x_n = x_0$, or $x_n \to x_0$ as $n \to \infty$.

While the initial look at the definition of convergence might seem contrived, it matches our intuition for what convergence should mean. To wit, a rough and intuitive definition of convergence would go as: the sequence $\{x_n\}_{n=1}^{\infty}$ converges to x_0 when the sequence gets gradually closer to x_0 as n gets larger in such a way that as $n \to \infty$, the sequence settles on x_0. With this intuition, we can consider $\varepsilon > 0$ as an arbitrary "tolerance" to test how close the sequence is to x_0, and hence n_ε tells us at what point in the sequence, the subsequent elements $x_{n_\varepsilon}, x_{n_\varepsilon+1}, x_{n_\varepsilon+2}, ...$ are all closer to x_0 than the tolerance ε.

THEOREM 2.2.2. *The limit of a sequence is unique. That is, if $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ which converges to both the elements x_0 and y_0, then $x_0 = y_0$.*

Proof. Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence which converges to the elements $x_0 \in \mathbb{R}$ and $y_0 \in \mathbb{R}$. We will prove that $x_0 = y_0$ using Lemma 1.3.9, by proving that for every $\varepsilon > 0$ it holds that $\text{dist}(x_0, y_0) < \varepsilon$. To do so, let $\varepsilon > 0$. Let $n_1, n_2 \in \mathbb{N}$ such that for every $n \geq n_1$, it holds that $\text{dist}(x_n, x_0) < \varepsilon/2$, and for every $n \geq n_2$, it holds that $\text{dist}(x_n, y_0) < \varepsilon/2$. Let $N = \max\{n_1, n_2\}$. We then have

$$\text{dist}(x_0, y_0) \leq \text{dist}(x_0, x_N) + \text{dist}(x_N, y_0) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

where the first inequality uses the triangle inequality and the second is from the our choice of n_1 and n_2 along with the observation that $N \geq n_1$ and $N \geq n_2$. Therefore, for any $\varepsilon > 0$, we find $\text{dist}(x_0, y_0) < \varepsilon$, proving that $x_0 = y_0$. ■

EXAMPLE 2.2.3. Let $\{r_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$ defined by $r_n = \frac{2n-3}{5n}$. We will prove that $\lim_{n\to\infty} r_n = \frac{2}{5}$.

Background Work: Given some $\varepsilon > 0$, it's our job to find some $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$, $\left|\frac{2n-3}{5n} - \frac{2}{5}\right| < \varepsilon$. Let's note that $\frac{2n-3}{5n} = \frac{2}{5} - \frac{3}{5n}$; hence, $\frac{2n-3}{5n} - \frac{2}{5} = -\frac{3}{5n}$; in particular, $\left|\frac{2n-3}{5n} - \frac{2}{5}\right| = \frac{3}{5n}$. Now, we want $\frac{3}{5n} < \varepsilon$, and since we're searching for a good choice of n, we can rearrange the last inequality to find $n > \frac{3}{5\varepsilon}$.

Proof: Let $\varepsilon > 0$ be arbitrary. Using the Archimedean property, let $n_\varepsilon \in \mathbb{N}$ be any integer such that $n_\varepsilon > 3/\varepsilon$. In particular, if $n \geq n_\varepsilon$ then $n > \frac{3}{\varepsilon}$, implying $\varepsilon > \frac{3}{n}$. Therefore, for all $n \geq n_\varepsilon$,

$$\text{dist}\left(r_n, \frac{2}{5}\right) = \left|\frac{2n-3}{5n} - \frac{2}{5}\right| = \frac{3}{5n} < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have proved that $\lim_{n\to\infty} \frac{2n-3}{5n} = \frac{2}{5}$. □

EXAMPLE 2.2.4. Let $\{s_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)}.$$

We will prove that $\lim_{n\to\infty} s_n = 1$.

Background Work: We first use partial fractions to realize that for any $k \in \mathbb{N}$, $\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$. Therefore,

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left\{ \frac{1}{k} - \frac{1}{k+1} \right\} = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k+1} = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=2}^{n+1} \frac{1}{k}$$

Now

$$\sum_{k=1}^{n} \frac{1}{k} - \sum_{k=2}^{n+1} \frac{1}{k} = \left(1 + \sum_{k=2}^{n} \frac{1}{k}\right) - \left(\frac{1}{n+1} + \sum_{k=2}^{n} \frac{1}{k}\right) = 1 - \frac{1}{n+1}$$

from which we can now see that $s_n = 1 - \frac{1}{n+1}$. Therefore, our job simplifies to showing that $1 - \frac{1}{n+1} \to 1$, and it's useful to note that $\text{dist}(s_n, 1) = \left|1 - \frac{1}{n+1} - 1\right| = \frac{1}{n+1}$.

Proof: Let $\varepsilon > 0$. Let $n_\varepsilon \in \mathbb{N}$ be large enough that $\frac{1}{n_\varepsilon} < \varepsilon$. In particular, for every $n \geq n_\varepsilon$, it holds that $\frac{1}{n+1} < \varepsilon$ since $\frac{1}{n+1} < \frac{1}{n} \leq \frac{1}{n_\varepsilon} < \varepsilon$. Notice that for any $n \in \mathbb{N}$,

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left\{ \frac{1}{k} - \frac{1}{k+1} \right\} = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=2}^{n+1} \frac{1}{k} = 1 - \frac{1}{n+1}.$$

Therefore, for any $n \geq n_\varepsilon$,

$$\text{dist}(s_n, 1) = \left|1 - \frac{1}{n+1} - 1\right| = \frac{1}{n+1} < \varepsilon.$$

Since ε was arbitrary, we have proved that $\lim_{n\to\infty} s_n = 1$. □

EXAMPLE 2.2.5. Let $\{x_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by $x_n = (-1)^n$. We will prove that the sequence $\{x_n\}_{n=1}^{\infty}$ does not converge.

Background Work: To prove that $\{x_n\}_{n=1}^{\infty}$ doesn't converge we need to show that for all $x \in \mathbb{R}$ there exists some $\varepsilon > 0$ such that for all $n \in \mathbb{N}$ there exists an $m \geq n$ such that $\text{dist}(x_m, x) \geq \varepsilon$. We have that $x_n = 1$ if n is even, however $x_n = -1$ when n is odd. If $x \in \mathbb{R}$, then either $x \neq 1$ or $x \neq -1$ (potentially x isn't equal to either). Let's say $d = \max\{\text{dist}(x, 1), \text{dist}(x, -1)\}$. Then, we have that for any $n \in \mathbb{N}$ either $\text{dist}(x_n, x) = d$ or $\text{dist}(x_{n+1}, x) = d$, showing that $\varepsilon = d$ is an judicious tolerance to use.

Proof: Let $x \in \mathbb{R}$ be arbitrary. Define $\varepsilon = \max\{\text{dist}(x, 1), \text{dist}(x, -1)\}$; note that $\varepsilon > 0$. Now, for any $n \in \mathbb{N}$ we have that either $x_n = 1$ and $x_{n+1} = -1$ or $x_n = -1$ and $x_{n+1} = 1$. In particular, $\text{dist}(x_n, x) = \varepsilon$ or $\text{dist}(x_{n+1}, x) = \varepsilon$. In either case, for any $n \in \mathbb{N}$, there exists some $m \geq n$ such that $\text{dist}(x_m, x) = \varepsilon$. This proves that $\{x_n\}_{n=1}^{\infty}$ does not converge to x, and since x was arbitrary, we have proven that $\{x_n\}_{n=1}^{\infty}$ does not converge. □

3. Properties of the Limit

Let us start here by proving an intuitive result that, stated roughly, illustrates that for a sequence to converge and the limit of the convergence depends only on the behavior of the sequence as n approaches infinity.

THEOREM 2.3.1 (Only the Tail Matters). *Let k be an arbitrary non-negative integer. Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$. Define the sequence $\{y_n\}_{n=1}^{\infty}$ by $y_n = y_{k+n}$. Then $\{x_n\}_{n=1}^{\infty}$ converges if and only if $\{y_n\}_{n=1}^{\infty}$ converges, and in that case $\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n$.*

Proof. Assume first that $\{x_n\}_{n=1}^{\infty}$ converges and suppose that x_0 is the limit. Let $\varepsilon > 0$. Let $n_\varepsilon \in \mathbb{N}$ be large enough that for all $n \geq n_\varepsilon$, it holds that $\text{dist}(x_n, x_0) < \varepsilon$. Now, if $n \geq n_\varepsilon$, then $n + k \geq n_\varepsilon$ and so $\text{dist}(x_{n+k}, x_0) = \text{dist}(y_n, x_0) < \varepsilon$. Therefore, we have found an n_ε such that for all $n \geq n_\varepsilon$, it holds that $\text{dist}(y_n, x_0) < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we have proved that that $\lim_{n\to\infty} y_n = x_0$. Thus we have proved that if $\lim_{n\to\infty} x_n = x_0$, then $\lim_{n\to\infty} y_n = x_0$.

Now assume that $\{y_n\}_{n=1}^{\infty}$ converges and suppose that y_0 is the limit. Let $\varepsilon > 0$. Let $\tilde{n}_\varepsilon$ be large enough that for all $n \geq \tilde{n}_\varepsilon$, it holds that $\text{dist}(y_n, y_0) < \varepsilon$. Define $n_\varepsilon = \tilde{n}_\varepsilon + k$. Then for all $n \geq n_\varepsilon$, it holds that $n - k \geq \tilde{n}_\varepsilon$ and so $\text{dist}(y_{n-k}, y_0) = \text{dist}(x_n, y_0) < \varepsilon$. Therefore, we have found some n_ε such that for all $n \geq n_\varepsilon$, it holds that $\text{dist}(x_n, y_0) < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we have proved that $\lim_{n\to\infty} x_n = y_0$. Thus we have proved that if $\lim_{n\to\infty} y_n = y_0$, then $\lim x_n = y_0$. ■

NOTATION 2.3.2. In Theorem 2.3.1, it is most common to denote the sequence $\{y_n\}_{n=1}^{\infty}$ by $\{x_{n+k}\}_{n=1}^{\infty}$ or $\{x_n\}_{n=k+1}^{\infty}$ since this matches our intuition.

DEFINITION 2.3.3. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. We say that the sequence $\{x_n\}_{n=1}^{\infty}$ is *bounded below* when there exists some $l \in \mathbb{R}$ such that $l \leq x_n$ for every $n \in \mathbb{N}$. Analogously, we say that the sequence is *bounded above* with there exists some $b \in \mathbb{R}$ such that $x_n \leq b$ for every $n \in \mathbb{N}$. If the sequence is both bounded above and below, we simply say that the sequence is *bounded*.

PROPOSITION 2.3.4. *Every convergent sequence in $\mathbb{R}$ is bounded.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a convergent sequence with limit $x_0 \in \mathbb{R}$. Let $n_1 \in \mathbb{N}$ be chosen such that for every $n \geq n_1$, it holds that $\text{dist}(x_n, x_{n_1}) = |x_n - x_0| < 1$; such an n_1 must exist since this sequence is assumed to converge. By the reverse triangle inequality, this implies

that $|x_n| - |x_0| < 1$ whenever $n \geq n_1$, which quickly rearranges to $|x_n| < 1 + |x_0|$ for every $n \geq n_1$. Let $r = \max\{|x_1|, ..., |x_{n_1}|, |x_0| + 1\}$. Then, for every $n \in \mathbb{N}$, we have that $|x_n| \leq r$, proving that the sequence is bounded. ■

LEMMA 2.3.5. *Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ which converges such that $\lim_{n\to\infty} x_n = x_0 \neq 0$. Then there exists some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, it holds that $|x_0|/2 < |x_n| < 3|x_0|/2$.*

Proof. Since it is assumed that $x_0 \neq 0$, then $|x_0| > 0$. Let $n_0 \in \mathbb{N}$ be chosen such that for all $n \geq n_0$, it holds that $|x_n - x_0| < |x_0|/2$. By the reverse triangle inequality we have both $|x_n| - |x_0| < |x_0|/2$ and $|x_0| - |x_n| < |x_0|/2$. Rearranging the first of these inequalities, we get that $|x_n| < 3|x_0|/2$ whenever $n \geq n_0$; rearranging the second gives $|x_n| > |x_0|/2$ whenever $n \geq n_0$. ■

THEOREM 2.3.6. *Suppose that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are sequences in $\mathbb{R}$ which converge to x_0 and y_0, respectively. The following hold.*

1. *For any $\alpha \in \mathbb{R}$, the sequence $\{x_n + \alpha y_n\}_{n=1}^{\infty}$ converges to $x_0 + \alpha y_0$.*
2. *The sequence $\{x_n y_n\}_{n=1}^{\infty}$ converges to $x_0 y_0$.*
3. *If $y_0 \neq 0$ and $y_n \neq 0$ for any n, then the sequence $\{x_n/y_n\}_{n=1}^{\infty}$ converges to x_0/y_0.*

Proof. (1) Let $\varepsilon > 0$. Find $n_1, n_2 \in \mathbb{N}$ such that for every $n \geq n_1$ it holds that $|x_n - x_0| < \frac{\varepsilon}{2}$, and for every $n \geq n_2$ it holds that $|y_n - y_0| < \frac{\varepsilon}{2(|\alpha|+1)}$. Set $n_\varepsilon = \max\{n_1, n_2\}$. Then for all $n \geq n_\varepsilon$, we have

$$\begin{aligned} |(x_n - \alpha y_n) - (x_0 - \alpha y_0)| = |(x_n - x_0) + \alpha(y_n - y_0)| &\leq |x_n - x_0| + |\alpha|\,|y_n - y_0| \\ &< \frac{\varepsilon}{2} + |\alpha|\frac{\varepsilon}{2|\alpha| + 1} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Thus, $\lim_{n\to\infty} (x_n + \alpha y_n) = x_0 + \alpha y_0$.

(2) Since the sequence $\{x_n\}_{n=1}^{\infty}$ converges, it is bounded. Let $r > 0$ be chosen such that $|x_n| < r$ for every $n \in \mathbb{N}$. Let $n_1, n_2 \in \mathbb{N}$ be chosen such that for every $n \geq n_1$ it holds that $|x_n - x_0| < \frac{\varepsilon}{2r}$, and for every $n \geq n_2$ it holds that $|y_n - y_0| < \frac{\varepsilon}{2|y_0|+1}$. Define $n_\varepsilon = \max\{n_1, n_2\}$.

Then, for every $n \geq n_\varepsilon$, we have

$$\begin{aligned} |x_n y_n - x_0 y_0| &= |x_n y_n - x_n y_0 + x_n y_0 - x_0 y_0| = |x_n(y_n - y_0) + y_0(x_n - x_0)| \\ &\leq |x_n|\,|y_n - y_0| + |y_0|\,|x_n - x_0| < r|y_n - y_0| + |y_0|\,|x_n - x_0| \\ &< r\frac{\varepsilon}{2r} + |y_0|\frac{\varepsilon}{2|y_0|+1} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Thus, $\lim_{n\to\infty}(x_n y_n) = x_0 y_0$.

(3) It suffices to prove that $1/y_n \to 1/y_0$, since (2) then implies $x_n/y_n = x_n(1/y_n) \to x_0(1/y_0) = x_0/y_0$. To this end, use Lemma 2.3.5 to choose $n_1 \in \mathbb{N}$ such that for every $n \geq n_1$ it holds that $|y_n| > |y_0|/2$. Choose $n_2 \in \mathbb{N}$ such that for every $n \geq n_2$ it holds that $|y_n - y_0| < \frac{|y_0|^2}{2}\varepsilon$. Define $n_\varepsilon = \max\{n_1, n_2\}$. Then,

$$\left|\frac{1}{y_n} - \frac{1}{y_0}\right| = \left|\frac{y_0 - y_n}{y_n y_0}\right| = \frac{|y_n - y_0|}{|y_n|\,|y_0|} < \frac{|y_n - y_0|}{|y_0|^2/2} < \frac{|y_0|^2\varepsilon/2}{|y_0|^2/2} = \varepsilon.$$

Thus, $\lim_{n\to\infty}\frac{1}{y_n} = \frac{1}{y_0}$, which further implies $\lim_{n\to\infty}\frac{x_n}{y_n} = \frac{x_0}{y_0}$. ■

REMARKS 2.3.7.

(1) Theorem 2.3.6 can be interpreted as a confirmation that the limit operation respects pointwise-defined algebraic operations on convergent sequences. In words, as long as the sequences in consideration converge, the limit of a sum is the sum of the limits; the limit of a product is the product of the limits; and the limit of the ratio is the ratio of the limits.

(2) In Theorem 2.3.6 (1) and (2) you will notice when we had a scalar multiplied by the limit terms, we scaled ε by dividing off by the absolute value of the scalar and added 1 in the denominator. The reason that we added 1 in the denominator was to avoid trouble when the scalar happens to be 0.

THEOREM 2.3.8 (Squeezing Principle). *Suppose that $\{x_n\}_{n=1}^\infty$, $\{y_n\}_{n=1}^\infty$, and $\{a_n\}_{n=1}^\infty$ are sequences in $\mathbb{R}$ satisfying $x_n \leq a_n \leq y_n$ for every $n \in \mathbb{N}$. If there exists some $l \in \mathbb{R}$ such that $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} y_n = l$, then the sequence $\{a_n\}_{n=1}^\infty$ also converges and $\lim_{n\to\infty} a_n = l$.*

Proof. Suppose that $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} y_n = l$. Let $\varepsilon > 0$. By assumption, it holds that $x_n - l \leq a_n - l \leq y_n - l$ for every n, and therefore $|a_n - l| \leq \max\{|x_n - l|, |y_n - l|\}$ also holds for every n. Let $n_1, n_2 \in \mathbb{N}$ be chosen such that for every $n \geq n_1$ it holds that $|x_n - l| < \varepsilon$, and for every $n \geq n_2$ it holds that $|y_n - l| < \varepsilon$. Then, with $n_\varepsilon = \max\{n_1, n_2\}$,

for every $n \geq n_\varepsilon$, we have that $|a_n - l| \leq \max\{|x_n - l|, |y_n - l|\} < \varepsilon$. Thus, the sequence $\{a_n\}_{n=1}^{\infty}$ also converges and $\lim_{n\to\infty} a_n = l$. ■

EXAMPLE 2.3.9. Suppose that $\{y_n\}_{n=1}^{\infty}$ is a sequence which is bounded, but does not necessarily converge. For every $n \geq 1$ define $x_n = \frac{y_n}{n^2-n+1}$. We will prove that $\lim_{n\to\infty} x_n = 0$. By assumption there is some $r > 0$ such that $-r < y_n < r$ for every n. This implies that

$$-\frac{r}{n^2} < x_n < \frac{r}{n^2 - n + 1} < \frac{r}{n}$$

The rightmost inequality follows from the identity $n^2 - n + 1 \geq n$, which is easily proved via induction. From here, it is an easy exercise to prove that $1/n^2$ converges to 0 as $n \to \infty$, and we have proved that $1/n \to 0$ as $n \to \infty$ (in fact, knowing that $1/n \to 0$, you can use the Squeezing Principle to prove that $1/n^2 \to 0$, since $0 \leq 1/n^2 \leq 1/n$). By Theorem 2.3.6 (1), both $-r/n^2$ and r/n converge to 0 as $n \to \infty$. By the Squeezing Principle, this implies that $\lim_{n\to\infty} x_n = 0$. □

THEOREM 2.3.10 (Limits Respect Non-Strict Inequalities). *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be convergent sequences in $\mathbb{R}$ such that for all $n \in \mathbb{N}$, $a_n \leq b_n$. Then $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.*

Proof. Suppose that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. We want to show that $a \leq b$. We will argue by contradiction. Suppose that $b < a$. Let $\varepsilon = a - b$; note that $\varepsilon > 0$ by the contradiction assumption. Let $N \in \mathbb{N}$ be large enough that both $|a_N - a| < \varepsilon/2$ and $|b_N - b| < \varepsilon/2$ (we are using Exercises 9 (13) here). By assumption $0 \leq b_N - a_N = (b_N - b) + (b - a) + (a - a_N)$. We have that $b_N - b < \varepsilon/2$, $a - a_N < \varepsilon/2$, and $b - a = -\varepsilon$. Therefore, $0 \leq b_N - a_N < \varepsilon/2 - \varepsilon + \varepsilon/2 = 0$. Since $0 \leq b_N - a_N < 0$ is clearly a contradiction, we must have $a \leq b$. ■

4. Divergent Sequences

DEFINITION 2.4.1. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. We say that this sequence diverges to ∞, denoted $\lim_{n\to\infty} x_n = \infty$, when given any $r > 0$, there exists some $n_r \in \mathbb{N}$ such that for all $n \geq n_r$, it holds that $x_n > r$. Analogously, we say that the sequence diverges to $-\infty$, denoted $\lim_{n\to\infty} x_n = -\infty$, when given any $r > 0$, there exists some $n_r \in \mathbb{N}$ such that for all $n \geq n_r$, it holds that $x_n < -r$.

REMARK 2.4.2. Compare Definition 2.4.1 to Notation 1.6.7. They are quite similar. □

LEMMA 2.4.3. *Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$. The following hold.*

(1) *If* $\lim_{n\to\infty} x_n = \infty$*, then the set of elements* $\{x_n : n \in \mathbb{N}\}$ *comprising the sequence is unbounded above.*

(2) *If* $\lim_{n\to\infty} x_n = -\infty$*, then the set of elements* $\{x_n : n \in \mathbb{N}\}$ *comprising the sequence is unbounded below.*

In particular, by considering the contrapositive statements, this implies that if a sequence is bounded it does not diverge (though this still does not imply that it converges).

Proof. Exercise 9 (15). ■

5. Monotone Sequences

DEFINITION 2.5.1. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. We say the sequence is *non-decreasing* when for every $n \in \mathbb{N}$ it holds that $x_n \leq x_{n+1}$; if $\leq$ is exchanged with $<$, then we can say the sequence is *strictly increasing*. We say that the sequence is *non-increasing* when for every $n \in \mathbb{N}$ it holds that $x_n \geq x_{n+1}$; if $\geq$ is exchanged with $>$, we can say the sequence is *strictly decreasing*. We say that the sequence is *monotone* when it is either non-decreasing or non-increasing.

THEOREM 2.5.2 (Monotone Convergence for Sequences). *Suppose* $\{x_n\}_{n=1}^{\infty}$ *is a monotone sequence in* $\mathbb{R}$*. Then the following hold.*

(1) *If* $\{x_n\}_{n=1}^{\infty}$ *is non-decreasing, then* $\lim_{n\to\infty} x_n = \sup\{x_n : n \in \mathbb{N}\}$*, including the possibility that the sequence diverges to* ∞ *when* $\sup\{x_n : n \in \mathbb{N}\} = \infty$*.*

(2) *If* $\{x_n\}_{n=1}^{\infty}$ *is non-increasing, then* $\lim_{n\to\infty} x_n = \inf\{x_n : n \in \mathbb{N}\}$*, including the possibility that the sequence diverges to* $-\infty$ *when* $\inf\{x_n : n \in \mathbb{N}\} = -\infty$*.*

In particular, this implies that a monotone sequence will converge if and only if it is bounded.

Proof. (1) Suppose $\{x_n\}_{n=1}^{\infty}$ is bounded and non-decreasing. Since the sequence is bounded, the *set* $\{x_n \text{ s.t. } n \in \mathbb{N}\}$ is also bounded. Hence let $b = \sup\{x_n \text{ s.t. } n \in \mathbb{N}\}$, which exists in $\mathbb{R}$. We now show that $\lim_{n\to\infty} x_n = b$. Let $\varepsilon > 0$. Since $b - \varepsilon$ is no longer an upper bound of $\sup\{x_n \text{ s.t. } n \in \mathbb{N}\}$, there exists some $n_\varepsilon \in \mathbb{N}$ such that $b - \varepsilon < x_{n_\varepsilon}$. For all $n \geq n_\varepsilon$, $x_{n_\varepsilon} \leq x_n$, so $b - \varepsilon < x_n$. Since b is an upper bound of the set $\sup\{x_n \text{ s.t. } n \in \mathbb{N}\}$, we must then have that for all $n \geq n_\varepsilon$, it holds that $b - \varepsilon < x_n \leq b$, and hence $\text{dist}(x_n, b) = |x_n - b| = b - x_n < b - (b - \varepsilon) = \varepsilon$. Since ε was arbitrary, we proved that $\lim_{n\to\infty} x_n = b$.

Alternatively, assume that the sequence is non-decreasing but unbounded. Then, since $x_1 \leq x_n$ for every n, we see that the sequence is bounded below, so the assumption of unboundedness implies that the sequence is unbounded above. We will prove that the sequence diverges to ∞, which will finish the proof since $\{x_n \text{ s.t. } n \in \mathbb{N}\}$ is unbounded above and therefore $\sup\{x_n \text{ s.t. } n \in \mathbb{N}\} = \infty$. Let $r > 0$ be arbitrary, and use the fact that the sequence is unbounded above to find some $n_r \in \mathbb{N}$ such that $r < x_{n_r}$. By the non-decreasing assumption, this implies that for every $n \geq n_r$, it holds that $r < x_{n_r} \leq x_n$, which proves that the sequence diverges to ∞. So, as before, this proves that $\lim_{n\to\infty} x_n = \infty = \sup\{x_n \text{ s.t. } n \in \mathbb{N}\}$.

(2) Exercise 9 (16). ■

6. Subsequences

DEFINITION 2.6.1. A function $\mathbf{n} : \mathbb{N} \to \mathbb{N}$ will be called a *subsequence index* when it is strictly increasing. That is, we will say that $\mathbf{n} : \mathbb{N} \to \mathbb{N}$ is a subsequence index when $\mathbf{n}(k) < \mathbf{n}(m)$ whenever $k < m$. It is most common to write n_k for $\mathbf{n}(k)$, and viewing $\mathbf{n}$ as a sequence of natural numbers $\{n_k\}_{k=1}^{\infty} = \{n_1, n_2, n_3, ...\}$.

LEMMA 2.6.2. *Suppose that $\{n_k\}_{k=1}^{\infty}$ is a subsequence index. Then for every $k \in \mathbb{N}$ it holds that $k \leq n_k$.*

Proof. Exercise 9 (17). ■

DEFINITION 2.6.3. Let S be a set (not necessarily a subset of $\mathbb{R}$) and $\{x_n\}_{n=1}^{\infty}$ a sequence in S. A *subsequence* of $\{x_n\}_{n=1}^{\infty}$ is a sequence $\{x_{n_k}\}_{k=1}^{\infty}$ where $\{n_1, n_2, n_3, ...\}$ is a subsequence index.

REMARK 2.6.4. The definition of subsequence we give above takes the perspective of a sequence being an indexed collection. However, we might gain some intuition looking at this through the function interpretation of a sequence. If $\mathbf{x} : \mathbb{N} \to S$ is a sequence, then a subsequence is a composition $\mathbf{x} \circ \mathbf{n} : \mathbb{N} \to S$ where $\mathbf{n} : \mathbb{N} \to \mathbb{N}$ is a subsequence index. From this perspective, we have the identification $\mathbf{x}(\mathbf{n}(k)) \leftrightarrow x_{n_k}$. □

NOTATION 2.6.5. It is not uncommon to have an explicit formula for a subsequence index, such as $n_k = k^2 + 1$, say. In this case, it is common to subscript a subsequence with the formula for n_k, such as $\{x_{k^2+1}\}_{k=1}^{\infty}$.

EXAMPLE 2.6.6. Let $\{x_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by $x_n = (-1)^n$ for every $n \in \mathbb{N}$. Consider the subsequence $\{x_{2k}\}_{k=1}^{\infty} = \{x_2, x_4, x_6, ...\}$. Now $x_{2k} = (-1)^{2k} = 1$, so this subsequence $\{x_{2k}\}_{k=1}^{\infty}$ is $\{1, 1, 1, ...\}$. In particular, this subsequence converges to 1, even though the overall sequence does not converge. □

DEFINITION 2.6.7. Suppose that $\{n_k\}_{k=1}^{\infty}$ and $\{m_k\}_{k=1}^{\infty}$ are both subsequence indices. We say that these *eventually saturate* $\mathbb{N}$ when there is some $N \in \mathbb{N}$ such that for every $n \geq N$, it holds that $n \in \{n_k \text{ s.t. } k \in \mathbb{N}\}$ or $n \in \{m_k \text{ s.t. } k \in \mathbb{N}\}$. More generally, suppose that $\{n_k^{(1)}\}_{k=1}^{\infty}, \{n_k^{(2)}\}_{k=1}^{\infty}, ..., \{n_k^{(r)}\}_{k=1}^{\infty}$ are all subsequence indices. We say that these *eventually saturate* $\mathbb{N}$ when there exists some $N \in \mathbb{N}$ such that for every $n \geq N$, it holds that $n \in \bigcup_{j=1}^{r} \left\{ n_k^{(j)} \text{ s.t. } k \in \mathbb{N} \right\}$.

EXAMPLE 2.6.8. Define the subsequence index $\{n_k\}_{k=1}^{\infty}$ by $n_k = 2k$. Define another subsequence index $\{m_k\}_{k=1}^{\infty}$ by $m_k = 2k+5$. Then $\{n_k\}_{k=1}^{\infty}$ and $\{m_k\}_{k=1}^{\infty}$ eventually saturate $\mathbb{N}$ since for all $n \geq 6$, either $n = 2k$ for some k or $n = 2k + 5$ for some k. Notice that $\{n_k \text{ s.t. } k \in \mathbb{N}\} \cup \{m_k \text{ s.t. } k \in \mathbb{N}\} = \{2, 4, 6, 7, 8, 9, 10, ...\}$, so starting at 6 and beyond every natural number lands in one of the two subsequence indices, which is what we need to conclude the saturation. □

THEOREM 2.6.9. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$ and $x \in \mathbb{R}$. The following are equivalent.*

(1) *The sequence $\{x_n\}_{n=1}^{\infty}$ converges to x.*
(2) *Every subsequence of $\{x_n\}_{n=1}^{\infty}$ converges to x.*
(3) *Every subsequence of $\{x_n\}_{n=1}^{\infty}$ has a further subsequence which converges to x.*
(4) *There exist subsequence indices $\{n_k^{(1)}\}_{k=1}^{\infty}, ..., \{n_k^{(r)}\}_{k=1}^{\infty}$ which eventually saturate $\mathbb{N}$ such that for $j = 1, ..., r$, all subsequences $\{x_{n_k^{(j)}}\}_{k=1}^{\infty}$ converge to x.*

Proof. (1) $\implies$ (2): Suppose that (1) holds. Let $\{x_{n_k}\}_{k=1}^{\infty}$ be any subsequence of $\{x_n\}_{n=1}^{\infty}$. We need to prove that $\lim_{k\to\infty} x_{n_k} = x$. To that end, let $\varepsilon > 0$. By (1), there exists some $k_\varepsilon \in \mathbb{N}$ such that for all $n \geq k_\varepsilon$, it holds that $\text{dist}(x_n, x) < \varepsilon$. Then for all $k \geq k_\varepsilon$, $n_k \geq n_{k_\varepsilon} \geq k_\varepsilon$, so $\text{dist}(x_{n_k}, x) < \varepsilon$. This shows that given an arbitrary $\varepsilon > 0$, there exists some $k_\varepsilon \in \mathbb{N}$ such that for all $k \geq k_\varepsilon$, it holds that $\text{dist}(x_{n_k}, x) < \varepsilon$, and hence $\lim_{k\to\infty} x_{n_k} = x$. Since the subsequence was arbitrary, this proves that all subsequences converge to x.

(2) $\implies$ (3): Assume that every subsequence converges to x. Then any further subsequence must also converge to x by our last proof.

(3) $\implies$ (1): By proving the contrapositive. Suppose that the sequence itself does not converge to x. This means that there exists some $\varepsilon > 0$ such that for every $n \in \mathbb{N}$, we can find some $m > n$ such that $\text{dist}(x_m, x) \geq \varepsilon$. This allows us to build a subsequence inductively as follows. Let $n_1 \geq 1$ be any index such that $\text{dist}(x_{n_1}, x) \geq \varepsilon$. Define $n_2 > n_1$ to be any index such that $\text{dist}(x_{n_2}, x) \geq \varepsilon$. Continuing inductively, for every $k \in \mathbb{N}$, we define $n_{k+1} > n_k$ such that $\text{dist}(x_{n_k}, x) \geq \varepsilon$. Thus we have created a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ where for every $k \in \mathbb{N}$, it holds that $\text{dist}(x_{n_k}, x) \geq \varepsilon$. But this prevents any further subsequence from converging to x as the distance from any element in the subsequence to x is bounded below by ε. Hence there exists a subsequence which does not have a further subsequence converging to x, proving the contrapositive implication.

(1) $\implies$ (4): Immediate. Indeed, (1) says that $\{x_n\}_{n=1}^{\infty}$ is a sequence converging to x; since the indices of this sequence saturate $\mathbb{N}$, it must be that (4) holds.

(4) $\implies$ (1): Suppose that for $j = 1, ..., r$, the subsequence indices $\{n(j)_k\}_{k=1}^{\infty}$ eventually saturate $\mathbb{N}$ and that each of the subsequences $\{x_{n_k^{(j)}}\}_{k=1}^{\infty}$ converge to x. Define $N \in \mathbb{N}$ such that for all $n \geq N$, it holds that n is in one of the subsetting indices; such an N must exist by the saturation assumption. Take $\varepsilon > 0$. We need to find some $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$, it holds that $\text{dist}(x_n, x) < \varepsilon$. To start, for each $j = 1, 2, ..., r$, let $k_j \in \mathbb{N}$ be chosen such that for all $k \geq k_j$, it holds that $\text{dist}(x_{n_k^{(j)}}, x) < \varepsilon$. Let $k_\varepsilon = \max\{k_1, ..., k_r\}$ and define $n_\varepsilon = \max\{n_{k_\varepsilon}^{(1)}, ..., n_{k_\varepsilon}^{(r)}, N\}$. If $n \geq n_\varepsilon$, then $n \geq N$, so $n = n_k^{(j)}$ for some k and j. Moreover, by the construction of n_ε, if $n = n_k^{(j)}$, then $k \geq k_j$, and therefore $\text{dist}(x_n, x) = \text{dist}(x_{n_k^{(j)}}, x) < \varepsilon$. Putting these pieces together implies that for every $n \geq n_\varepsilon$, it holds that $|x_n - x| < \varepsilon$, proving that the original sequence converges to x. ■

EXAMPLE 2.6.10. Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ such that there is an $x \in \mathbb{R}$ where the subsequence of odd terms $\{x_{2k-1}\}_{k=1}^{\infty}$ and the subsequence of even terms $\{x_{2k}\}_{k=1}^{\infty}$ both converge to x. Since the subsequence indices of odd and even terms eventually saturate $\mathbb{N}$, the sequence $\{x_n\}_{n=1}^{\infty}$ must also converge to x. □

EXAMPLE 2.6.11. Once again, consider the sequence $\{x_n\}_{n=1}^{\infty}$ defined by $x_n = (-1)^n$. The sequence of even terms $\{x_{2k}\}_{k=1}^{\infty} = \{1, 1, 1, ...\}$ clearly converges to 1, whereas the sequence of odd terms $\{x_{2k-1}\}_{k=1}^{\infty} = \{-1, -1, -1, ...\}$ clearly converges to -1. Since these two subsequences do not converge to the same point, we conclude the overall sequence $\{x_n\}_{n=1}^{\infty}$ does not converge. □

EXAMPLE 2.6.12. Let p be a positive integer. The sequence $\{1/n^p\}_{n=1}^{\infty}$ converges to 0. We have multiple ways to prove this at our disposal. One is to notice that $n \leq n^p$, implying that $0 \leq 1/n^p \leq 1/n$ and hence apply the Squeezing Principle armed with the already-proved knowledge $1/n \to 0$. Another method is to notice that the sequence $\{1/n^p\}_{n=1}^{\infty}$ is, in fact, a subsequence of $\{1/n\}_{n=1}^{\infty}$, and since $1/n \to 0$, all its subsequences must also converge to 0. (To confirm that $\{1/n^p\}_{n=1}^{\infty}$ is a subsequence of $\{1/n\}_{n=1}^{\infty}$, if we define $n_k = k^p$ for every $k \in \mathbb{N}$, then $\{n_1, n_2, ...\}$ is a subsequence index and $\{1/n_k\}_{k=1}^{\infty}$ is the sequence $\{1/n^p\}_{n=1}^{\infty}$). □

THEOREM 2.6.13. *Every sequence in $\mathbb{R}$ has a monotone subsequence.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. Define a *peak index* of the sequence to be an index $n \in \mathbb{N}$ such that $x_n \geq x_m$ for all $m \geq n$. Let T be the subset of $\mathbb{N}$ containing all the peak indexes of the sequence arranged in ascending order. We now break the proof into two cases.

The first case is when there are infinitely many elements in T. In this case, we can apply the Well Ordering Principle to label each element in T in ascending order $T = \{n_1, n_2, n_3, ...\}$ with $n_k < n_{k+1}$ for every k. Hence $\{n_1, n_2, ...\}$ is a subsequence index and, by the definition of peak index, we find $x_{n_k} \geq x_{n_{k+1}}$ for each k. Thus $\{x_{n_k}\}_{k=1}^{\infty}$ is a non-increasing sequence (and thus monotone).

In the second case, T has at most finitely many elements (including the possibility that $T = \varnothing$). Let $n_1 \in \mathbb{N} \setminus T$ be chosen to be an upper bound of T. By definition, $n_1 \notin T$ and is thusly not a peak index. This means that we can find some $n_2 > n_1$ such that $x_{n_2} > x_{n_1}$. Now, since $n_2 > n_1$ and $n_1 > p$ for every $p \in T$, we have that n_2 is not a peak index, so we can find some $n_3 > n_2$ such that $x_{n_3} > x_{n_2}$. Continuing this way inductively, we can create a subsequence index $\{n_1, n_2, n_3, ...\}$ such that the subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ is strictly increasing (and thus monotone). ■

COROLLARY 2.6.14. *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a bounded sequence. Since the sequence is bounded, every subsequence is also bounded. Since every sequence has a monotone subsequence, we conclude that the sequence has a bounded monotone subsequence. By the Monotone Convergence Theorem for sequences, this implies that there is a subsequence which converges. ■

7. Cauchy Sequences

We begin discussion of Cauchy sequences in this section. It turns out that the usage Cauchy sequences have become indispensable in the study of analysis and the definition is easily generalized to much more general spaces than $\mathbb{R}$. The main theorem of this sequence is Theorem 2.7.6 below.

DEFINITION 2.7.1. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. We say that that the sequence is *Cauchy* when given any $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ such that for all $n, m \geq n_\varepsilon$, it holds that $\text{dist}(x_n, x_m) < \varepsilon$. A sequence is said to satisfy the *Cauchy criteria* when it is a Cauchy sequence.

REMARK 2.7.2. Given a way to measure distances between points, the determination of whether a sequence is Cauchy or not is an intrinsic property of the sequence; it only requires examination of the elements of the sequence itself. This is in contrast to the determination of whether or not a sequence converges. Indeed, given a notion of distance, to determine whether or not a sequence converges, we need to consider a candidate limit to critique distances between the tail of the sequence and this candidate limit; generally the candidate limit is extrinsic information, found in the ambient set in which the sequence is being considered, rather than information within the sequence itself. □

PROPOSITION 2.7.3. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$ which converges. Then $\{x_n\}_{n=1}^{\infty}$ is Cauchy.*

Proof. Suppose that x_0 is the limit of the sequence $\{x_n\}_{n=1}^{\infty}$. Let $\varepsilon > 0$ and let $n_\varepsilon \in \mathbb{N}$ be chosen such that for all $n \geq n_\varepsilon$, it holds that $\text{dist}(x_n, x_0) < \varepsilon/2$. Now, if $n, m \geq n_\varepsilon$, then using the triangle inequality, $\text{dist}(x_n, x_m) \leq \text{dist}(x_n, x_0) + \text{dist}(x_0, x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. Since $\varepsilon > 0$ was arbitrary, and we found some $n_\varepsilon \in \mathbb{N}$ such that for all $n, m \geq n_\varepsilon$ it holds that $\text{dist}(x_n, x_m) < \varepsilon$, we have proved that the sequence $\{x_n\}_{n=1}^{\infty}$ is Cauchy. ■

PROPOSITION 2.7.4. *Every Cauchy sequence in $\mathbb{R}$ is bounded.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence. Let $n_1 \in \mathbb{N}$ be chosen such that for every $n \geq n_1$, it holds that $\text{dist}(x_n, x_{n_1}) = |x_n - x_{n_1}| < 1$; such an n_1 must exist since this sequence is assumed Cauchy. By the reverse triangle inequality, this implies that $|x_n| - |x_{n_1}| < 1$ whenever $n \geq n_1$, which quickly rearranges to $|x_n| < 1 + |x_{n_1}|$ for every $n \geq n_1$. Let

$r = \max\{|x_1|, ..., |x_{n_1}|, |x_{n_1}| + 1\}$. Then, for every $n \in \mathbb{N}$, we have that $|x_n| \leq r$, proving that the sequence is bounded. ■

LEMMA 2.7.5. *A Cauchy sequence in $\mathbb{R}$ converges if and only if it has a convergent subsequence.*

Proof. If a Cauchy sequence converges, then certainly it has a convergent subsequence; just take the subsequence to be the entire sequence itself.

Let us now prove the converse (and more interesting) direction. Let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $\mathbb{R}$ and assume that the subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ converges. Let $x_0 \in \mathbb{R}$ be the limit of this subsequence. We will prove that the entire sequence converges to x_0. Let $\varepsilon > 0$. Using the Cauchy assumption, choose $n_\varepsilon \in \mathbb{N}$ such that for every $n, m \geq n_\varepsilon$ it holds that $\text{dist}(x_n, x_m) < \varepsilon/2$. By the assumption of the convergent subsequence, let $k_\varepsilon \in \mathbb{N}$ be chosen such that for every $k \geq k_\varepsilon$ it holds that $\text{dist}(x_{n_k}, x_0) < \varepsilon/2$. Now, by the definition of a subsequence index, for any $n \geq n_\varepsilon$, we can find some $k \geq k_\varepsilon$ such that $n \leq n_k$, which then implies

$$\text{dist}(x_n, x_0) \leq \text{dist}(x_n, x_{n_k}) + \text{dist}(x_{n_k}, x_0) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Here $\text{dist}(x_n, x_{n_k}) < \varepsilon/2$ since $n, n_k \geq n_\varepsilon$, and $\text{dist}(x_{n_k}, x_0) < \varepsilon/2$ since $k \geq k_\varepsilon$. Therefore, for any $n \geq n_\varepsilon$ it holds that $\text{dist}(x_n, x_0) < \varepsilon$, proving that if a Cauchy sequence has a convergent subsequence, then the entire sequence also converges. ■

THEOREM 2.7.6. *A sequence in $\mathbb{R}$ converges in $\mathbb{R}$ if and only if it is Cauchy.*

Proof. We proved in Proposition 2.7.3 that a convergent sequence is Cauchy.

Conversely, let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $\mathbb{R}$. Theorem 2.6.13 tells us that this sequence has a monotone subsequence $\{x_{n_k}\}_{k=1}^{\infty}$. This monotone subsequence must be bounded, since the entire sequence is bounded by Proposition 2.7.4. By the monotone convergence theorem for sequences, this further implies that the subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ converges. By Lemma 2.7.5, this then implies that the sequence itself converges. ■

8. Sequences and Convergence Within Subsets of $\mathbb{R}$

Up to now, we've concerned ourselves with sequences in $\mathbb{R}$ which converge in $\mathbb{R}$. However, there is often good reason to consider not only when a sequence converges in $\mathbb{R}$, but when it converges within a particular subset of $\mathbb{R}$. We begin some exploration in this direction in

this section; however, this paradigm arises time and again throughout the remainder of these notes.

DEFINITION 2.8.1. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$, and let $U \subset \mathbb{R}$. We say that the sequence *converges in* U when the sequence converges and the limit is an element of U.

REMARK 2.8.2. Suppose $U \subset \mathbb{R}$. Most commonly, when we are interested in whether a sequence $\{x_n\}_{n=1}^{\infty}$ converges in U, then we have assumed that the sequence was in U to start with. That is, we usually start with the assumption that some sequence $\{x_n\}_{n=1}^{\infty}$ is in U (i.e., $x_n \in U$ for every n), make some further assumptions on the sequence, and then ask whether or not the sequence converges in U, if it converges at all. □

THEOREM 2.8.3. *Let $\mathbb{F}$ be an ordered field satisfying the Archimedean Property. Then, up to relabelling elements, we can assume that $\mathbb{F} \subset \mathbb{R}$. The field $\mathbb{F}$ is supremum complete, if and only if every Cauchy sequence in $\mathbb{F}$ converges in $\mathbb{F}$.*

Proof. Suppose first that $\mathbb{F}$ is supremum complete. Then we know that, by possibly relabelling elements, $\mathbb{F} = \mathbb{R}$. In this case Theorem 2.7.6 proves that every Cauchy sequence in $\mathbb{F}$ converges in $\mathbb{F}$.

Conversely, suppose that $\mathbb{F}$ has the property that every Cauchy sequence in $\mathbb{F}$ converges in $\mathbb{F}$. Let $U \subset \mathbb{F}$ be non-empty and bounded above. Since $\mathbb{F} \subset \mathbb{R}$, then U is also a non-empty subset of $\mathbb{R}$ which is bounded above. Hence $\sup(U)$ exists in $\mathbb{R}$; let $b = \sup(U)$. We need to prove that $b \in \mathbb{F}$ to complete the proof. We have proved that there is a sequence $\{x_n\}_{n=1}^{\infty}$ in U which converges to $\sup(U)$. Since this sequence converges, it is Cauchy, and since $U \subset \mathbb{F}$, this sequence is also in $\mathbb{F}$. By assumption, every Cauchy sequence in $\mathbb{F}$ converges in $\mathbb{F}$ implying that the sequence $\{x_n\}_{n=1}^{\infty}$ converges in $\mathbb{F}$. Since limits are unique, and this sequence was chosen such that $\lim_{n\to\infty} x_n = b$, we have proved that $b \in \mathbb{F}$. This finishes the proof since given an arbitrary non-empty subset in $\mathbb{F}$ which is bounded above, the supremum of that set is and element of $\mathbb{F}$, and hence $\mathbb{F}$ is supremum complete. ■

There are two noteworthy ways to interpret the result of Theorem 2.8.3. The first is that if $\mathbb{F} \subset \mathbb{R}$ is any sub-field, then $\mathbb{F} = \mathbb{R}$ if and only if every Cauchy sequence in $\mathbb{F}$ converges in $\mathbb{F}$. The second is that the property of supremum completeness is equivalent to the property that every Cauchy sequence converges. This latter perspective warrants the following definition, which turns out to be more general and useful in many settings than the notion of supremum completeness.

DEFINITION 2.8.4. Let $U \subset \mathbb{R}$. We way that the set U is *complete* whenever any Cauchy sequence in U converges in U.

EXAMPLE 2.8.5. The half-open interval $(0,1]$ in $\mathbb{R}$ is not complete. Indeed, the sequence $\{1/n\}_{n=1}^{\infty}$ is in $(0,1]$ (since $1/n \in (0,1]$ for every n) and is Cauchy since it converges to 0. However, $0 \notin (0,1]$, and therefore there exists a Cauchy sequence in $(0,1]$ which does not converge in $(0,1]$. □

EXAMPLE 2.8.6. $\mathbb{R}$ is complete. Indeed, Theorem 2.7.6 proves that every Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$. □

PROPOSITION 2.8.7. *Let $U \subset \mathbb{R}$ be a non-empty set with only finitely many elements. Then U is complete.*

Proof. Assume that U has only finitely many elements, say $U = \{u_1, ..., u_N\}$ for some $N \in \mathbb{N}$. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in U which is Cauchy. Let ε be chosen as the minimal distance between any two consecutive points in U; i.e.,

$$\varepsilon = \min \{\text{dist}(u_i, u_j) \text{ s.t. } i, j \in \{1, ..., N\} \text{ and } i \neq j\}.$$

Since we have assumed the sequence $\{x_n\}_{n=1}^{\infty}$ is Cauchy, this means that there is some $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$ it holds that $\text{dist}(x_n, x_{n_\varepsilon}) < \varepsilon$. Since $x_{n_\varepsilon} = u_k$ for some $k \in \{1, 2, ..., N\}$, this implies that for all $n \geq n_\varepsilon$, it holds that $\text{dist}(x_n, u_k) < \varepsilon$. By the choice of ε, if $u \in U$ and $\text{dist}(u, u_k) < \varepsilon$ then $u = u_k$; otherwise, $\text{dist}(u, u_k) \geq \varepsilon$. This means that when $n \geq n_\varepsilon$, it holds that $x_n = u_k$. This proves that if $\{x_n\}_{n=1}^{\infty}$ is Cauchy in U, it must be the that the tail of the sequence is eventually a constant element in U, which converges to that constant element in U. Hence, any Cauchy sequence in U converges in U, proving that U is complete. ■

9. Exercises

(1) Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$ and let $x_0 \in \mathbb{R}$. Prove that $\{x_n\}_{n=1}^{\infty}$ converges to x_0 if and only if the sequence $\{d_n\}_{n=1}^{\infty}$ in $\mathbb{R}$ defined by $d_n = \text{dist}(x_n, x_0)$ converges to 0.

(2) Let $\{a_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$, let $a \in \mathbb{R}$, and let $k > 0$. Prove that $\{a_n\}_{n=1}^{\infty}$ converges to a if and only if given any $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ such that for

all $n \geq n_\varepsilon$, $|a_n - a| < k\varepsilon$. (Basically, you are proving that for some constant k, the tolerance $k\varepsilon$ is just as arbitrary as ε itself).

(3) Reprove Theorem 2.2.2 using the distance function dist rather than the absolute value. Look at Remark **??** for some clarity, if needed.

(4) In Example 2.2.3 we proved that the sequence $\{r_n\}_{n=1}^{\infty}$ converges to $\frac{2}{5}$, where for every $n \in \mathbb{N}$ we define $r_n = \frac{2n-3}{5n}$. Since we now know the sequence converges, find the minimal positive integer n_0 such that for all $n \geq n_0$, it holds that $\left|r_n - \frac{2}{5}\right| < \frac{1}{100}$. Compare how you found this n_0 to how we found n_ε in the example.

(5) Let $\{r_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$ defined by

$$r_n = \frac{3n^2 - 5n}{5n^2 + 1}.$$

Prove that $\{r_n\}_{n=1}^{\infty}$ converges and find the limit.

(6) Let $\{s_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)(k+2)}.$$

Prove that $\{s_n\}_{n=1}^{\infty}$ converges and find the limit of the sequence. *Hint*: It might help to look at Example 2.2.4.

(7) Let $\{x_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by

$$x_n = \begin{cases} \frac{1}{n} & \text{if } n \neq 2^k \text{ for any } k \in \mathbb{N} \\ 1 & \text{if } n = 2^k \text{ for some } k \in \mathbb{N} \end{cases}$$

For clarity, the first few terms in this sequence are $1, 1, \frac{1}{3}, 1, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, 1, \frac{1}{9}, ..., \frac{1}{15}, 1, \frac{1}{17}, ...$ Prove or disprove that $\{x_n\}_{n=1}^{\infty}$ converges.

(8) Assume that $\{h_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R} \setminus \{0\}$ which converges to 0. Let $x \in \mathbb{R}$ and define the sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathbb{R}$ by

$$x_n = \frac{(x + h_n)^3 - x^3}{h_n}.$$

Prove that $\lim_{n\to\infty} x_n = 3x^2$.

(9) Let $a, c \in \mathbb{R} \setminus \{0\}$, $b, d \in \mathbb{R}$ where we assume that $d \neq -cn$ for any $n \in \mathbb{N}$, and $k, k' \in \mathbb{N}$. Define the sequence $\{r_n\}_{n=1}^{\infty}$ in $\mathbb{R}$ by

$$r_n = \frac{a\,n^k + b}{c\,n^{k'} + d}.$$

(a) If $k = k'$, prove that $\lim_{n\to\infty} r_n = \frac{a}{c}$.

(b) If $k > k'$, prove that $\lim_{n\to\infty} r_n = +\infty$ when a and c have the same sign, otherwise $\lim_{n\to\infty} r_n = -\infty$.

(c) If $k < k'$, prove that $\lim_{n\to\infty} r_n = 0$.

(10) Let us extend Theorem 2.3.6. Suppose that $\{x_n^1\}_{n=1}^{\infty}$, $\{x_n^2\}_{n=1}^{\infty}$, ..., $\{x_n^k\}_{n=1}^{\infty}$ are all sequences in $\mathbb{R}$ which converge. Use induction to prove the following.

(a) The sequence $\left\{ \sum_{r=1}^{k} x_n^r \right\}_{n=1}^{\infty}$ converges and

$$\lim_{n\to\infty} \left(\sum_{r=1}^{k} x_n^r \right) = \sum_{r=1}^{k} \left(\lim_{n\to\infty} x_n^r \right).$$

(b) The sequence $\left\{ \prod_{r=1}^{k} x_n^r \right\}_{n=1}^{\infty}$ converges and

$$\lim_{n\to\infty} \left(\prod_{r=1}^{k} x_n^r \right) = \prod_{r=1}^{k} \left(\lim_{n\to\infty} x_n^r \right).$$

(11) Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers. Prove that $\lim_{n\to\infty} x_n = 0$ if and only if $\lim_{n\to\infty} \frac{1}{x_n} = +\infty$. (For the meaning of a limit diverging to $+\infty$, see Definition 2.4.1). Could we conclude something similar if we only assumed that $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R} \setminus \{0\}$, but not necessarily all positive?

(12) Let $x \in \mathbb{R}$ be fixed. Suppose that $\{a_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ which converges to a. Define the sequence $\{x_n\}_{n=1}^{\infty}$ by $x_n = x + a_n$ for every $n \in \mathbb{N}$.

(a) Prove that $\lim_{n\to\infty} x_n = x + a$.

(b) By considering $a_n = 0$ for every n, use this to give meaning to taking the limit of a constant $\lim_{n\to\infty} x = x$.

(13) Suppose that $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ are convergent sequences in $\mathbb{R}$ such that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$.

(a) Prove that for any $\varepsilon > 0$ there exists $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$, it holds that both $|a_n - a| < \varepsilon$ and $|b_n - b| < \varepsilon$.

(b) Extend this result by induction. That is, suppose that $\{x_n^1\}_{n=1}^{\infty}$, $\{x_n^2\}_{n=1}^{\infty}$, ..., $\{x_n^k\}_{n=1}^{\infty}$ are convergent sequences such that $\lim_{n\to\infty} x_n^i = r_i$ for every $i \in \{1, 2, ..., k\}$. Prove that given any $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ it holds that $|x_n^1 - r_1| < \varepsilon$, and $|x_n^2 - r_2| < \varepsilon$, and ..., and $|x_n^k - r_k| < \varepsilon$.

(14) Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers and let $a \in \mathbb{R}$. Prove that $\lim_{n\to\infty} a_n = a$ if and only if there exists a sequence $\{b_n\}_{n=1}^{\infty}$ which converges to 0 such that for all $n \in \mathbb{N}$, $|a_n - a| \leq b_n$.

(15) (a) Prove Lemma 2.4.3 (1).

(b) Prove Lemma 2.4.3 (2).

(16) Prove part (2) in the Monotone Convergence Theorem for sequences, Theorem 2.5.2.

(17) Prove Lemma 2.6.2.

(18) Let $\{x_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by

$$x_n = \begin{cases} 1 & \text{if } n = 3k \text{ for some } k \in \mathbb{N} \\ \frac{2k^2}{2k^2 - k} & \text{if } n = 3k + 1 \text{ for some } k \in \mathbb{N} \\ \frac{1}{2}(1 + 2^{-k}) + \sum_{m=1}^{k} \frac{m}{k^2} & \text{if } n = 3k + 2 \text{ for some } k \in \mathbb{N} \end{cases}$$

Prove that $\lim_{n\to\infty} x_n = 1$.

(19) In *Case 2* of the proof of Theorem 2.6.13, we claim that we can inductively created a sequence of integers $n_1 < n_2 < n_3 < \cdots$ such that $x_{n_1} < x_{n_2} < x_{n_3} < \cdots$. Prove this is true. That is, use induction to prove that we can build such a sequence of positive integers as described. *Hint*: With P the set of peak indices described in *Case 2*, let n_0 be the minimal element in $\mathbb{N}$ such that for all $m \in P$, $m \leq n_0$. Then define the set T by

$$T = \big\{k \in \mathbb{N} : \text{there exists a collection positive integers } n_1, n_2, ..., n_k \text{ such that } n_0 = n_1 < n_2 < ... < n_k \text{ and } x_{n_1} < x_{n_2} < ... < x_{n_k}\big\}$$

Use induction to show that $T = \mathbb{N}$.

(20) Let $T = \{r \in \mathbb{Q} \text{ s.t. } 0 < r < 1\}$. As T contains an infinite number of rational numbers and the rational numbers are countable, then T is also countable. In particular, there exists a bijection $\varphi : \mathbb{N} \to T$. Let $\{x_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by $x_n = \varphi(n)$. (In this way, we have "arranged" the rational numbers between 0 and 1

into a sequence). Prove that $\{x_n\}_{n=1}^{\infty}$ has a convergent subsequence.

(21) Let $\{s_n\}_{n=1}^{\infty}$ be the sequence in $\mathbb{R}$ defined by $s_n = \sum_{k=1}^{n} \frac{1}{k}$.

(a) Prove by induction that for every $n \in \mathbb{N}$, $\sum_{k=1}^{2^n} \frac{1}{k} \geq 1 + \frac{n}{2}$.

Hint: Here is the intuition behind this inequality. Suppose that $n = 4$, then

$$\sum_{k=1}^{2^4} \frac{1}{k} = 1 + \Big(\frac{1}{2}\Big) + \Big(\frac{1}{3} + \frac{1}{4}\Big) + \Big(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\Big) + \Big(\frac{1}{9} + \cdots + \frac{1}{16}\Big)$$
$$\geq 1 + \frac{1}{2} + \Big(\underbrace{\frac{1}{4} + \frac{1}{4}}_{=1/2}\Big) + \Big(\underbrace{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}}_{=1/2}\Big) + \Big(\underbrace{\frac{1}{16} + \cdots + \frac{1}{16}}_{=1/2}\Big) = 1 + \frac{4}{2}.$$

(b) Prove that $\lim_{n\to\infty} s_n = +\infty$.

(c) Prove that given any $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$ and $m = n + 1$, $|s_n - s_m| < \varepsilon$.

(d) Is $\{s_n\}_{n=1}^{\infty}$ Cauchy? Justify your answer.

(22) The following claim is false and hence the given proof has some fault in it. Find the fault in the proof and provide a counter example to show that the claim is false.

Claim: *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be sequences in $\mathbb{R}$ such that $\{a_n\}_{n=1}^{\infty}$ is Cauchy and $\{b_n\}_{n=1}^{\infty}$ is bounded. Then the sequence $\{a_n b_n\}_{n=1}^{\infty}$ is Cauchy.*

Proof: Let $B > 0$ be such that for all $n \in \mathbb{N}$, $|b_n| \leq B$; we can find such a B since the sequence $\{b_n\}_{n=1}^{\infty}$ is bounded by assumption. Let $\varepsilon > 0$. Let $n_\varepsilon \in \mathbb{N}$ be large enough such that for all $n \geq n_\varepsilon$ and for all $m \geq n_\varepsilon$, it holds that $|a_n - a_m| < \varepsilon/B$; we can find such a n_ε since the sequence $\{a_n\}_{n=1}^{\infty}$ is assumed to be Cauchy. Now, for all $n \geq n_\varepsilon$ and all $m \geq n_\varepsilon$, we have

$$|a_n b_n - a_m b_m| \leq \big|B(a_n - a_m)\big| = B|a_n - a_m| < B(\varepsilon/B) = \varepsilon.$$

Therefore $\{a_n b_n\}_{n=1}^{\infty}$ is Cauchy.

(23) Give yet another proof that the sequence $\{1/n\}_{n=1}^{\infty}$ converges in $\mathbb{R}$ by proving directly that it is Cauchy. (That is, don't use the fact that it converges to then imply that it is Cauchy, instead show that the sequence is Cauchy directly by using the definition of what it means for a sequence to be Cauchy).

(24) Suppose that $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$. Prove directly from the definition of a Cauchy sequence that the sequence $\{x_n^2\}_{n=1}^{\infty}$ is a Cauchy sequence. *Hint*: Note that $|x_n^2 - x_m^2| = |x_n + x_m|\,|x_n - x_m|$.

(25) Suppose that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are Cauchy sequences. Prove directly from the definition of a Cauchy sequence that the sequence $\{x_n y_n\}_{n=1}^{\infty}$ is Cauchy. *Hint*: Following the argument of Theorem 2.3.6 (2), it may help be helpful to start your manipulations with the observation that $|x_n y_n - x_m y_m| = |x_n y_n - x_m y_n + x_m y_n - x_m y_m|$...

(26) Consider the following proposition: *There exists a constant $c \in \mathbb{R}$ such that for any $n, m \in \mathbb{N}$ with $m \leq n$, the bound $\sum_{k=m}^{n} \frac{1}{k^2} \leq \frac{c}{\sqrt{m}}$ holds.* Although we have not yet proved this proposition, assume it is true for this problem. Define the sequence $\{s_n\}_{n=1}^{\infty}$ by $s_n = \sum_{k=1}^{n} \frac{1}{k^2}$. Using the proposition, prove that the sequence $\{s_n\}_{n=1}^{\infty}$ converges in $\mathbb{R}$.

(27) Let $a, b \in \mathbb{R}$ be chosen so that $a < b$. Prove that the interval $[a, b] = \{r \in \mathbb{R} \text{ s.t. } a \leq r \leq b\}$ is complete.

(28) Let $S = \{1/n \text{ s.t. } n \in \mathbb{N}\}$.

(a) Prove that S is not complete.

(b) Prove that $S \cup \{0\}$ is complete.

CHAPTER 3

The Topology on $\mathbb{R}$

We will exhaustively use interval notation from here on, so if you find yourself unfamiliar with it, refer back to Section 2.2. This chapter is to introduce us to several topological notions on the real number line; such as, open and closed sets, compact sets, and how such sets help us learn from and about structures we have already become familiar with on $\mathbb{R}$.

1. Open Sets

DEFINITION 3.1.1. A subset $U \subset \mathbb{R}$ is said to be *open* when for every $x \in U$ there exists some $\varepsilon_x > 0$ such that $(x - \varepsilon_x, x + \varepsilon_x) \subset U$. Stated equivalently, U is open when for every $x \in U$ there exists some $\varepsilon_x > 0$ such that for every $y \in \mathbb{R}$ with $\text{dist}(x, y) < \varepsilon_x$, it holds that $y \in U$.

REMARK 3.1.2. The equivalence of the two stated definitions in Definition 3.1.1 follows immediately from Lemma 1.3.8. □

PROPOSITION 3.1.3 (Open Intervals are Open). *The following are open subsets in* $\mathbb{R}$.

(1) $\mathbb{R}$ *and* $\varnothing$.
(2) *Any interval of the form* (x, y) *where* $x < y$.
(3) *Any interval of the form* $(-\infty, y)$ *or* (x, ∞).

Proof. (1) If $x \in \mathbb{R}$, then for any $\varepsilon > 0$, the interval $(x - \varepsilon, x + \varepsilon)$ is a subset of $\mathbb{R}$, implying that $\mathbb{R}$ is open. On the other hand, there are no elements in φ to check, hence the proposition that $\varnothing$ is open is *vacuously true*.

(2) Let $x, y \in \mathbb{R}$ with $x < y$. Take $z \in (x, y)$ and let $\varepsilon_z = \min\{z - x, y - z\}$; that is $\varepsilon_z = \min\{\text{dist}(x, z), \text{dist}(y, z)\}$. We will be finished once we show that $(z - \varepsilon_z, z + \varepsilon_z) \subset (x, y)$. To this end, since $\varepsilon_z \leq z - x$, we find that $z - \varepsilon_z \geq z - (z - x) = x$; since $\varepsilon_z \leq y - z$, we find that $z + \varepsilon_z \leq z + (y - z) = y$. This shows that $x \leq z - \varepsilon_z < z + \varepsilon_z \leq y$ and hence $(z - \varepsilon_z, z + \varepsilon_z) \subset (x, y)$.

(3) This follows the same outline as the previous proof, where if $z \in (-\infty, y)$, then take $\varepsilon_z = y - z$; if $z \in (x, \infty)$, then take $\varepsilon_z = z - x$. The details are left as an Exercise. ■

THEOREM 3.1.4. *Let $\{U_i\}_{i\in I}$ be an indexed collection of open subsets in $\mathbb{R}$. Then*

(1) *The set $\bigcup_{i\in I} U_i$ is open in $\mathbb{R}$.*
(2) *If I is has only finitely many values, then the set $\bigcap_{i\in I} U_i$ is open in $\mathbb{R}$.*

Put poetically, the union of arbitrarily many open sets is open and the intersection of finitely many open sets is open.

Proof. (1) If $\bigcup_{i\in I} U_i = \varnothing$ then it is open. Otherwise, if $\bigcup_{i\in I} U_i \neq \varnothing$, let $x \in \bigcup_{i\in I} U_i$. Since x is in the union, there exists some $i_0 \in I$ such that $x \in U_{i_0}$. Since U_{i_0} is open by assumption, there exists some $\varepsilon_x > 0$ such that $(x - \varepsilon_x, x + \varepsilon_x) \subset U_{i_0}$, implying further that $(x - \varepsilon_x, x + \varepsilon_x) \subset \bigcup_{i\in I} U_i$. Thus, $\bigcup_{i\in I} U_i$ is open.

(2) If $\bigcap_{i\in I} U_i = \varnothing$ then it is open. Otherwise, if $\bigcap_{i\in I} U_i \neq \varnothing$ then let $x \in \bigcap_{i\in I} U_i$. For every $i \in I$, U_i is open, and therefore for each $i \in I$ there exists some $\varepsilon_i > 0$ such that $(x - \varepsilon_i, x + \varepsilon_i) \subset U_i$. Let $\varepsilon_x = \min\{\varepsilon_i \text{ s.t. } i \in I\}$; since this minimum is over only finitely many positive elements, it holds that $\varepsilon_x > 0$. Now, for every $i \in I$, $(x - \varepsilon_x, x + \varepsilon_x) \subset (x - \varepsilon_i, x + \varepsilon_i) \subset U_i$, implying $(x - \varepsilon_x, x + \varepsilon_x) \subset U_i$. Therefore, since $(x - \varepsilon_x, x + \varepsilon_x) \subset U_i$ for every $i \in I$, we deduce that $(x - \varepsilon_x, x + \varepsilon_x) \subset \bigcap_{i\in I} U_i$. ■

1.1. Neighborhoods.

DEFINITION 3.1.5. Given some $x \in \mathbb{R}$, we say that a set U is a *neighborhood* of x whenever there exists some open set V with $x \in V$ and $V \subset U$. If U is a neighborhood of x and happens to be open, then we say that U is an *open neighborhood* of x. In the case that U is a neighborhood of x, we say that x is an *interior point* of U; we denote the set of interior points of U by $\text{int}(U)$, or commonly U° in other texts.

LEMMA 3.1.6. *For any subset $U \subset \mathbb{R}$, it holds that $\text{int}(U)$ is open.*

Proof. Let U be an arbitrary subset of $\mathbb{R}$ and let $x \in \text{int}(U)$ be arbitrary. Then there exists some $\varepsilon_x > 0$ such that $(x - \varepsilon_x, x + \varepsilon_x) \subset U$. We will show that $(x - \varepsilon_x, x + \varepsilon_x) \subset \text{int}(U)$, finishing the proof. To this end, let $y \in (x - \varepsilon_x, x + \varepsilon_x)$. Since the interval $(x - \varepsilon_x, x + \varepsilon_x)$ is open, there is some $\varepsilon_y > 0$ such that $(y - \varepsilon_y, y + \varepsilon_y) \subset (x - \varepsilon_x, x + \varepsilon_x)$, which by transitivity implies

that $(y-\varepsilon_y, y+\varepsilon_y) \subset U$. This implies that $y \in \text{int}(U)$, and since $y \in (x-\varepsilon_x, x+\varepsilon_x) \implies y \in \text{int}(U)$, we conclude that $(x-\varepsilon_x, x+\varepsilon_x) \subset \text{int}(U)$. ■

DEFINITION 3.1.7. Given some $x \in \mathbb{R}$, we say that the set $\tilde{U}$ is a *deleted neighborhood* of x whenever there exists a neighborhood U of x such that $\tilde{U} = U \setminus \{x\}$.

LEMMA 3.1.8. *A set $\tilde{U}$ is a deleted neighborhood of some point x if and only if $x \notin \tilde{U}$ and there is some open set U containing x such that $U \setminus \{x\} \subset \tilde{U}$.*

Proof. $\tilde{U}$ is a deleted neighborhood of x if and only if $\tilde{U} \cup \{x\}$ is a neighborhood of x. From here, $\tilde{U} \cup \{x\}$ is a neighborhood of x if and only if there is an open set V containing x such that $V \subset \tilde{U} \cup \{x\}$. This proves that $\tilde{U}$ is a deleted neighborhood of x if and only if there is an open set V containing x such that $V \subset \tilde{U} \cup \{x\}$, and it is clear that $V \subset \tilde{U} \cup \{x\}$ if and only if $V \setminus \{x\} \subset \tilde{U}$. This is enough to deduce the result. ■

EXAMPLE 3.1.9. Any open set U is a neighborhood of all of its elements. Alternatively, if U is any neighborhood of x, then the set $U \setminus \{x\}$ is a deleted neighborhood of x (created by "deleting" x from the neighborhood U). In fact, □

PROPOSITION 3.1.10. *Let $U \subset \mathbb{R}$. The following are equaivalent.*

(1) *U is open.*
(2) *Every element of U is an interior point of U. That is, $\text{int}(U) = U$.*
(3) *For every $x \in U$, there is a neighborhood V of x such that $V \subset U$.*
(4) *For every $x \in U$, there is a deleted neighborhood V of x such that $V \subset U$.*

Proof. (1) $\iff$ (2): Suppose that U is open. Then U is a neighborhood of every point in U by the definition of an open set, and hence $U \subset \text{int}(U)$. However, by definition, $\text{int}(U) \subset U$, and thus $\text{int}(U) = U$. Conversely, assume that $\text{int}(U) = U$. Lemma 3.1.6 shows that $\text{int}(U)$ is open, and with $\text{int}(U) = U$, then U is open.

The remaining proofs are the content of Exercise 7 (5). ■

2. Closed Sets

DEFINITION 3.2.1. A subset $U \subset \mathbb{R}$ is called *closed* when the complement $\mathbb{R} \setminus U$ is open.

EXAMPLE 3.2.2. Any finite subset of $\mathbb{R}$ is closed. That is, if $U \subset \mathbb{R}$ has only finitely many elements, then U is closed. Indeed, let $U \subset \mathbb{R}$ be assumed to only have finitely many

elements. The case that $U = \varnothing$ is dealt with below in Proposition 3.2.3, so we will assume here that U has n elements for some $n \in \mathbb{N}$. Label U such that $U = \{u_1, u_2, ..., u_n\}$ where $u_1 < u_2 < \cdots < u_n$. It is easy to show that

$$\mathbb{R} \setminus U = (-\infty, u_1) \cup (u_1, u_2) \cup \cdots \cup (u_{n-1}, u_n) \cup (u_n, \infty)$$

which is a union of sets already proved to be open, and hence is open. Therefore, the complement $\mathbb{R} \setminus U$ is open, implying that U is closed. □

PROPOSITION 3.2.3 (Closed Intervals are Closed). *The following are closed subsets in* $\mathbb{R}$.

(1) $\mathbb{R}$ *and* $\varnothing$.

(2) *Any interval of the form* $[x, y]$ *where* $x \leq y$.

(3) *Any interval of the form* $(-\infty, y]$ *or* $[x, \infty)$.

Proof. (1) We have that $\mathbb{R} \setminus \mathbb{R} = \varnothing$ and $\mathbb{R} \setminus \varnothing = \mathbb{R}$, both of which are open by Proposition 3.1.3. Hence $\mathbb{R}$ and $\varnothing$ are closed.

(2) Let $x, y \in \mathbb{R}$ with $x \leq y$. Then $\mathbb{R} \setminus [x, y] = (-\infty, x) \cup (y, \infty)$. Both $(-\infty, x)$ and (y, ∞) are open by Proposition 3.1.3, hence so is there union by Proposition 3.1.4. Therefore $[x, y]$ is closed.

(3) If $x, y \in \mathbb{R}$, then $\mathbb{R} \setminus (-\infty, x] = (x, \infty)$ and $\mathbb{R} \setminus [y, \infty) = (-\infty, y)$. Since (x, ∞) and $(-\infty, y)$ are both open by Proposition 3.1.3, we conclude that $(-\infty, x]$ and $[y, \infty)$ are both closed. ■

Let us conclude with this brief introduction to closed sets by providing the analogue of Theorem 3.1.4 above.

THEOREM 3.2.4. *Let* $\{C_i\}_{i \in I}$ *be an indexed collection of closed subsets in* $\mathbb{R}$. *Then*

(1) *The set* $\bigcap_{i \in I} C_i$ *is closed in* $\mathbb{R}$.

(2) *If* I *has only finitely many elements, then the set* $\bigcup_{i \in I} C_i$ *is closed in* $\mathbb{R}$.

Put poetically, the intersection of arbitrarily many closed sets is closed and the union of finitely *many closed sets is closed.*

Proof. (1) By D'Morgan's Laws, $\mathbb{R} \setminus \left[\bigcap_{i \in I} C_i \right] = \bigcup_{i \in I} (\mathbb{R} \setminus C)$. Since $\mathbb{R} \setminus C_i$ is open for each $i \in I$, the union $\bigcup_{i \in I} (\mathbb{R} \setminus C_i)$ is open by Proposition 3.1.4. This shows that $\mathbb{R} \setminus \left[\bigcap_{i \in I} C_i \right]$ is open, hence $\bigcap_{i \in I} C_i$ is closed.

(2) By D'Morgan's Laws, $\mathbb{R} \setminus \left[\bigcup_{i \in I} C_i \right] = \bigcap_{i \in I} (\mathbb{R} \setminus C_i)$. Since $\mathbb{R} \setminus C_i$ is open for each $i \in I$ and I is finite, the intersection $\bigcap_{i \in I} (\mathbb{R} \setminus C_i)$ is open by Proposition 3.1.4. This shows that $\mathbb{R} \setminus \left[\bigcup_{i \in I} C_i \right]$ is open, and hence $\bigcup_{i \in I} C_i$ is closed. ■

EXAMPLE 3.2.5. To get a feel for why an infinite union of closed sets may not itself be closed, consider the sets $C_n = [1/n, 1 - 1/n]$ for every $n \geq 2$. Then, for each n, the interval C_n is closed; however, $\bigcup_{n \geq 2} C_n = (0, 1)$, which is not closed. □

3. Classifications of Points in a Set

In Definition 3.1.5 above, we already defined the notion of an interior point of a set. In this section we give some other classification of points within a set.

DEFINITION 3.3.1. Let $U \subset \mathbb{R}$. A point $x \in \mathbb{R}$ is called a *boundary point* of U when for all neighborhood V of x, it holds that $V \cap U \neq \varnothing$ and $V \cap (\mathbb{R} \setminus U) \neq \varnothing$. The set of all boundary points of U is denoted $\text{bdry}(U)$, or commonly ∂U in other texts.

EXAMPLE 3.3.2. Suppose that J is an interval in $\mathbb{R}$ with left endpoint $a \in \mathbb{R}$. Then $a \in \text{bdry}(J)$. Indeed, notice that since J is an interval with left endpoint a, there is some $0 < \delta < 1$ such that $(a, a + \delta) \subset J$; we also note that $(-\infty, a) \subset \mathbb{R} \setminus J$. Now, $(a - 1, a + 1)$ is a neighborhood of a and $\varnothing \neq (a, a + \delta) = (a - 1, a + 1) \cap (a, a + \delta) \subset (a - 1, a + 1) \cap J$ and $\varnothing \neq (a - 1, a) = (a - 1, a + 1) \cap (-\infty, a) \subset (a - 1, a + 1) \cap (\mathbb{R} \setminus J)$. Thus $a \in \text{bdry}(J)$. □

LEMMA 3.3.3. *Let $U \subset \mathbb{R}$. The sets $\text{int}(U)$ and $\text{bdry}(U)$ are disjoint. Moreover, every point $x \in U$ is either in $\text{int}(U)$ or $\text{bdry}(U)$; that is, $U \subset \text{int}(U) \cup \text{bdry}(U)$.*

Proof. By definition, it should be clear that $\text{int}(U)$ and $\text{bdry}(U)$ are disjoint since both definitions can not be simultaneously satisfied without deriving a contradiction. So, we need only show that every element of U is either a boundary point, or an interior point. To this end, suppose that $x \in U$. If $x \notin \text{int}(U)$, then this means that for every neighborhood V of x, it must be that $V \not\subset U$. This is equivalent to stating that if $x \notin \text{int}(U)$, every neighborhood V of x has the property that $V \cap (\mathbb{R} \setminus U) \neq \varnothing$; but since $x \in V \cap U$, it must also be that $V \cap U \neq \varnothing$. Thus x is a boundary point of U. ■

DEFINITION 3.3.4. Let $U \subset \mathbb{R}$. A point $x \in \mathbb{R}$ is called a *limit point* of U when for all deleted neighborhoods V of x, it holds that $U \cap V \neq \varnothing$. The set of limit points of U will be denoted $\text{limpts}(U)$.

EXAMPLE 3.3.5. The set of limit points of $\mathbb{Q}$ in $\mathbb{R}$ is all of $\mathbb{R}$. That is, $\text{limpts}(\mathbb{Q}) = \mathbb{R}$. Indeed, let $x \in \mathbb{R}$ and V be any deleted neighborhood of x. According to Exercise 7 (4), there is some $\varepsilon > 0$ such that $(x-\varepsilon, x) \subset \mathbb{V}$. But then, by the density of the rationals, there is some $r \in \mathbb{Q}$ such that $r \in (x-\varepsilon, x)$, and thus $V \cap \mathbb{Q} \neq \varnothing$. Since V was an arbitrary deleted neighborhood of x and we showed that $V \cap \mathbb{Q} \neq \varnothing$, this proves that $x \in \text{limpts}(\mathbb{Q})$. However, x was an arbitrary element of $\mathbb{R}$, proving that $\mathbb{R} = \text{limpts}(\mathbb{Q})$. □

LEMMA 3.3.6. *Let $U \subset \mathbb{R}$. The element $x \in \mathbb{R}$ is a limit point of U if and only if there is a sequence $\{x_n\}_{n=1}^{\infty}$ in $U \setminus \{x\}$ such that $\lim_{n\to\infty} x_n = x$.*

Proof. Exercise 7 (7). ■

We end this section with one of the manifestations of a famous theorem by Bolzano and Weierstrass.

THEOREM 3.3.7 (Bolzano-Weierstrass Theorem). *If $S \subset \mathbb{R}$ is bounded and has infinitely many elements, then S has a limit point.*

Proof. If S has infinitely many elements, then we can create a sequence $\{x_n\}_{n=1}^{\infty}$ such that $x_n = x_m \iff n = m$; this is equivalent to saying that there is an injection $\mathbf{x} : \mathbb{N} \to S$ and defining $x_n = \mathbf{x}(n)$. Since S is bounded, the sequence $\{x_n\}_{n=1}^{\infty}$ is bounded. Therefore, by Corollary 2.6.14, $\{x_n\}_{n=1}^{\infty}$ has a convergent subsequence $\{x_{n_k}\}$ which converges to some $x \in \mathbb{R}$. We claim that $x \in \text{limpts}(S)$. If for every $k \in \mathbb{N}$, $x_{n_k} \neq x$, then $\{x_{n_k}\}_{k=1}^{\infty}$ is a sequence in $S \setminus \{x\}$ converging to x, and hence x is a limit point of S by Lemma 3.3.6. Otherwise, suppose that $K \in \mathbb{N}$ is such that $x_{n_K} = x$. Since no two points of the sequence $\{x_n\}_{n=1}^{\infty}$ are equal, it holds that at most one of the points in the subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ is equal to x. In particular, for all $k > K$, $x_{n_k} \neq x$. Therefore the sequence $\{x_{n_{K+k}}\}_{k=1}^{\infty}$ is in $S \setminus \{x\}$ and converges to x, once again showing that x is a limit point of S. ■

4. Classification of Closed Sets

We have seen above a few equivalent ways to recognize open sets. In this brief section, we present an extremely useful ways to recognize closed sets. To start, we present a useful result.

LEMMA 3.4.1. *Let $U \subset \mathbb{R}$. Then for every element $x \in \text{bdry}(U)$ there is a sequence $\{x_n\}_{n=1}^{\infty}$ in U which converges to x. In particular, $\text{bdry}(U) \setminus U \subset \text{limpts}(U)$.*

Proof. Suppose $x \in \text{bdry}(U)$. Then, by the definition of a boundary point, for every $n \in \mathbb{N}$ there is an element $x_n \in (x - 1/n, x + 1/n) \cap U$. This defines a sequence $\{x_n\}_{n=1}^{\infty}$ in U such that the estimate $\text{dist}(x_n, x) < 1/n$ holds. Hence $\lim_{n\to\infty} = x$. To see now that $\text{bdry}(U) \setminus U \subset \text{limpts}(U)$, suppose that $x \in \text{bdry}(U) \setminus U$. If $\{x_n\}_{n=1}^{\infty}$ is a sequence in U converging to x (and we just proved such a sequence exists), then since $x \notin U$, this sequence is in $U \setminus \{x\}$ and we hence conclude that $x \in \text{limpts}(U)$. ■

THEOREM 3.4.2. *Let $C \subset \mathbb{R}$. The following are equivalent.*

(1) *C is closed.*
(2) $\text{bdry}(C) \subset C$.
(3) $\text{limpts}(C) \subset C$.
(4) *Every sequence in C which converges in $\mathbb{R}$, converges in C.*

Proof. (1) $\implies$ (2). By contradiction. Suppose that $b \in \text{bdry}\, C$ and $b \notin C$, implying that $b \in C^c$. Since C^c is open as C is closed, there exists some $\varepsilon > 0$ such that $(b-\varepsilon, b+\varepsilon) \subset C^c$. But then, $(b - \varepsilon, b + \varepsilon)$ is a neighborhood of b which does not intersect C, contradicting that b is a boundary point of C.

(2) $\implies$ (3). In Exercises 9 (8), we prove that if $b \in \mathbb{R}$ is a limit point of C which is not an element of C, then $b \in \text{bdry}(C)$. Therefore, if $\text{limpts}(C) \not\subset C$, there exists some $b \in \text{limpts}(C) \setminus C$, implying that $b \in \text{bdry}(C)$, further implying that $\text{bdry}(C) \not\subset C$. This proves the contrapositive statement.

(3) $\implies$ (4). Assume (3). Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in C which converges in $\mathbb{R}$. Suppose that $\lim_{n\to\infty} x_n = x$. We need to prove that $x \in C$, which we will do by contradiction. Suppose that $x \notin C$, then $\{x_n\}_{n=1}^{\infty}$ is a sequence in $C = C \setminus \{x\}$ which converges to x. By Lemma 3.3.6, this implies that x is a limit point of C, and by assumption $\text{limpts}(C) \subset C$. This is our contradiction since we assumed $x \notin C$ and hence proved $x \in C$.

(4) $\implies$ (1). By the contrapositive. Assume that C is not closed. Then C^c is not open. Therefore, there exists some $x \in C^c$ such that for every $\varepsilon > 0$, $(x - \varepsilon, x + \varepsilon) \cap C \neq \varnothing$. In particular, for every $n \in \mathbb{N}$, there exists some $x_n \in C$ such that $x_n \in (x - 1/n, x + 1/n)$. This implies that for every $n \in \mathbb{N}$, it holds that $\text{dist}(x_n, x) < 1/n$ and hence $\lim_{n\to\infty} x_n = x$. Therefore, $\{x_n\}_{n=1}^{\infty}$ is a sequence in C which converges in $\mathbb{R}$, but does not converge in C, proving the contrapositive. ■

Because $\mathbb{R}$ is complete, property (4) and completeness are identical. We state this rigorously in the following lemma, but understand that in spaces other than $\mathbb{R}$, spaces which may not be complete, this is not necessarily the case (see, for example, Exercise 7 (21)).

LEMMA 3.4.3. *Let $C \subset \mathbb{R}$. Then C is closed if and only if it is complete.*

Proof. Exercise 7 (14) ■

5. Compactness

DEFINITION 3.5.1. Let $K \subset \mathbb{R}$. We say that K is *sequentially compact* when given any sequence $\{x_n\}_{n=1}^{\infty}$ in K, there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ which converges in K. Briefly, K is sequentially compact when every sequence in K has a subsequence which converges in K.

THEOREM 3.5.2. *Let $K \subset \mathbb{R}$. K is sequentially compact if and only if K is closed and bounded.*

Proof. Suppose first that K is closed and bounded. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in K. Since K is bounded, so is the sequence, and therefore by Corollary 2.6.14, there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ which converges in $\mathbb{R}$. But since $\{x_{n_k}\}_{k=1}^{\infty}$ is a sequence in K which converges, then it converges in K, since K is closed. This proves that K is sequentially compact.

To prove the converse, we will prove the contrapositive. Assume that K is either not closed or not bounded. In Exercise 9 (15), we prove that if K is not closed, there exists a sequence $\{x_n\}_{n=1}^{\infty}$ which converges to a point $x \notin K$, which by Theorem **??** implies that every subsequence of $\{x_n\}_{n=1}^{\infty}$ converges to x, precluding the possibility that $\{x_n\}_{n=1}^{\infty}$ has a subsequence which converges in K. In the same exercise, we show that if K is not bounded, then there exists a sequence $\{y_n\}_{n=1}^{\infty}$ which diverges, in which case every subsequence of $\{y_n\}_{n=1}^{\infty}$ diverges, precluding the possibility that $\{y_n\}_{n=1}^{\infty}$ has a convergent subsequence. In either case, if K is either not closed or not bounded, then K is not sequentially compact. ■

DEFINITION 3.5.3. Let $K \subset \mathbb{R}$. An *open cover* of K is a collection $\{U_i\}_{i \in I}$ of open subsets in $\mathbb{R}$ such that $K \subset \bigcup_{i \in I} U_i$. We say that the open cover $\{U_i\}_{i \in I}$ of K *admits a finite subcover* of K when there is a finite subset $A \subset I$ such that $K \subset \bigcup_{i \in A} U_i$. The set K is called *compact* when every open cover of K admits a finite subcover of K.

EXAMPLE 3.5.4. The interval $(0,1]$ is not compact. To show this, consider the collection of open sets $\{U_n\}_{n=2}^{\infty}$ where $U_n = \left(\frac{1}{n}, 2\right)$. We have that $(0,1] \subset \bigcup_{n=2}^{\infty} U_n$, and but there does not exist a finite subcover. Indeed, suppose to arrive at a contradiction we assume that there is a finite subcover, say $\{U_{n_1}, U_{n_2}, ..., U_{n_m}\}$ is an open cover of $(0,1]$. Let $N = \max\{n_1, ..., n_m\}$; this implies that $U_{n_i} \subset U_N$. Thus $\bigcup_{i=1}^{m} U_{n_i} = U_N = \left(\frac{1}{N}, 2\right)$. However, U_N certainly does not cover $(0,1]$ since, for example, $\frac{1}{N+1} \in (0,1]$ but $\frac{1}{N+1} \notin U_N$. □

THEOREM 3.5.5 (Heine-Borel Theorem). *Let $K \subset \mathbb{R}$. Then K is compact if and only if K is closed and bounded.*

Proof. The proof that if K is compact, then it is closed and bounded is done with a contrapositive argument in Exercise 9 (17).

To prove the converse, we first prove a simpler case. Suppose that $a, b \in \mathbb{R}$ with $a \leq b$. We will now prove that $[a,b]$ is compact. Let $\{U_i\}_{i \in I}$ be an open cover of $[a,b]$. Define the set $T = \{t \in [a,b] \text{ s.t. } \{U_i\}_{i \in I} \text{ admits a finite subcover of } [a,t]\}$. Our argument will show that $b \in T$ which will finish the proof of this simpler case. Moving forward, let us make a useful observation. If $t \in T$, then $s \in T$ for every $a \leq s \leq t$, since any finite cover of $[a,t]$ is also a finite cover of a subset $[a,s]$ of $[a,t]$. We next observe that T is non-empty since $a \in T$, which is clear since there must be some $i \in I$ such that $a \in U_i$ (and hence there is a finite cover of $[a,a]$). Since T is non-empty and bounded above by definition, $\sup(T)$ exists; define $k = \sup(T)$. We will finish the proof by showing that $k \in T$ and $k = b$. To this end, we note that $a \leq k \leq b$ since b is an upper bound of T and $a \in T$. Therefore, there is some $i_0 \in I$ such that $k \in U_{i_0}$, which further implies that there is some $\varepsilon > 0$ such that $(k - \varepsilon, k + \varepsilon) \subset U_{i_0}$. By our first observation, we deduce that $t \in T$ for every $a \leq t < k$. In particular, there exists a $t \in T$ with $k - \varepsilon < t < k$. By the definition of T, there exists some $n \in \mathbb{N}$ and indices $i_1, ..., i_n \in I$ such that $[a,t] \subset \bigcup_{k=1}^{n} U_{i_k}$. But then, for any s with $k < s < k - \varepsilon$, we have that $[a,s] \subset U_{i_0} \cup \left[\bigcup_{k=1}^{n} U_{i_k}\right] = \bigcup_{k=0}^{n} U_{i_k}$. This shows that there is a finite subcover of $[a,s]$ which first shows that $k \in T$ since $k < s$ and second implies that $k = b$, since otherwise we could have chosen $s \in T$ with $s > k$, contradicting that k is an upper bound of T. Hence $k = b \in T$, proving that $[a,b]$ is compact.

To prove the general case, assume that $K \subset \mathbb{R}$ is closed and bounded. If $K = \varnothing$ then K is trivially compact. Otherwise, let $a = \inf(K)$ and $b = \sup(K)$, which exist since K is bounded and non-empty. Note here that $K \subset [a,b]$. Let $\{U_i\}_{i=1}^{\infty}$ be an arbitrary open cover

of K. Define $V = \mathbb{R} \setminus K$, which is open since K is closed. Then $\{U_i\}_{i=1}^{\infty} \cup \{V\}$ is an open cover of $\mathbb{R}$ and hence of $[a, b]$. Therefore, there exists a finite subcover $\{U_{i_1}, ..., U_{i_n}, V\}$ of $[a, b]$ which in turn is a finite subcover of K, since $K \subset [a, b]$. However, with $K \cap V = \varnothing$, we deduce that $\{U_{i_1}, ..., U_{i_n}\}$ covers K. Thus we have proved that any open cover of K admits a finite subcover, proving that K is compact. ■

From Theorem 3.5.2 and the Heine-Borel Theorem 3.5.5, we arrive at an immediate corollary.

COROLLARY 3.5.6. *Let $K \subset \mathbb{R}$. The following are equivalent.*

(1) *K is compact.*
(2) *K is sequentially compact.*
(3) *K is closed and bounded.*

6. Relative Topology

DEFINITION 3.6.1. Let $S \subset \mathbb{R}$. A set $V \subset S$ is called *relatively open* in S when there is an open subset $U \subset \mathbb{R}$ such that $V = S \cap U$. Analogously, a set $C \subset S$ is called *relatively closed* in S when there is a closed subset $K \subset \mathbb{R}$ such that $C = S \cap K$.

PROPOSITION 3.6.2. *Let $S \subset \mathbb{R}$. A subset $U \subset S$ is relatively open in S if and only if for every $x \in U$ there exists some $\varepsilon_x > 0$ such that $(x - \varepsilon_x, x + \varepsilon_x) \cap S \subset U$. Equivalently U is relatively open in S if and only if given any $x \in U$ there is some $\varepsilon_x > 0$ such that if $y \in S$ and $\operatorname{dist}(x, y) < \varepsilon_x$.*

Proof. Suppose that $U \subset S$ is relatively open and let $V \subset \mathbb{R}$ be open such that $U = S \cap V$. Let $x \in U$. Then $x \in V$ and since V is open, there is some $\varepsilon_x > 0$ such that $(x - \varepsilon_x, x + \varepsilon_x) \subset V$. Hence $(x - \varepsilon, x + \varepsilon) \cap S \subset V \cap S = U$.

Conversely, suppose that $U \subset S$ has the property that for every $x \in U$ there is some $\varepsilon_x > 0$ such that $(x - \varepsilon_x, x + \varepsilon_x) \cap S \subset U$. Define the open set $V = \bigcup_{x \in U} (x - \varepsilon_x, x + \varepsilon_x)$. By the definition of V, $U \subset V \cap S$ and $V \cap S = \bigcup_{x \in U} [(x - \varepsilon_x, x + \varepsilon_x) \cap S] \subset U$. This proves that $V \cap S = U$, and thus U is relatively open in S. ■

PROPOSITION 3.6.3. *Let $S \subset \mathbb{R}$. A subset $C \subset S$ is relatively closed in S if and only if the set $S \setminus C$ is relatively open in S.*

Proof. Exercise 7 (18). ■

If interpreted corrected, Theorem 3.4.2 is still valid for relatively closed sets (we simply need to define *relative limit points* and *relative boundary points*, which can be done by taking the intersection of the limit points or boundary points with the set we are working relative to). However, Lemma 3.4.3 does not generally hold and needed the fact that $\mathbb{R}$ was complete. We explore this fact in Exercise 7 (21).

DEFINITION 3.6.4. Let $S \subset \mathbb{R}$ and $x \in S$. We call a subset $V \subset S$ a *relative neighborhood* of x in S when there is a neighborhood U of x such that $U \cap S = V$.

PROPOSITION 3.6.5. *Let $S \subset \mathbb{R}$, $x \in S$, and $V \subset S$. The set V is a relative neighborhood of x in S if and only if there is some $\varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \cap S \subset V$. Put differently, V is a relative neighborhood of x in S if and only if there exists some $\varepsilon > 0$ such that for every $y \in S$ with $\operatorname{dist}(x,y) < \varepsilon$, it holds that $y \in V$.*

Proof. Suppose first that V is a relative neighborhood of x in S. Let U be a neighborhood of x. Then there exists some $\varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \subset U$. Hence $(x-\varepsilon, x+\varepsilon) \cap S \subset U \cap S = V$. Conversely, suppose that there is some $\varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \cap S \subset V$. Define the set U as $U = V \cup (x-\varepsilon, x+\varepsilon)$. Then U is a neighborhood of x and by how we defined U, $U \cap S = V$. ■

DEFINITION 3.6.6. Let $S \subset \mathbb{R}$ and $K \subset \mathbb{S}$. A collection $\{U_i\}_{i \in I}$ is called a *relative open cover* of K in S when U_i is relatively open in S for each $i \in I$, and when $K \subset \bigcup_{i \in I} U_i$. We way that a relative open cover of K in S *admits a finite subcover* when, as before, there is a subset $A \subset K$ with finitely many elements such that $K \subset \bigcup_{i \in A} U_i$. Finally, we say that K is *relatively compact* in S when every relative open cover of K in S admits a finite subcover.

LEMMA 3.6.7. *Let $S \subset \mathbb{R}$ and suppose that $K \subset S$. Then K is relatively compact in S if and only if K is a compact subset of $\mathbb{R}$.*

Proof. Suppose that $K \subset S$ is relatively compact in S. Let $\{U_i\}_{i \in I}$ be an open cover of K. Then $\{S \cap U_i\}_{i \in I}$ is a relatively open cover of K in S and, by assumption, admits a finite subcover $\{S \cap U_i\}_{i \in A}$, where $A \subset I$ is some finite subset. But then, as $S \cap U_i \subset U_i$ for each i, it is clear that $\{U_i\}_{i \in A}$ also covers K, and hence $\{U_i\}_{i \in I}$ admits a finite subcover of K. Since $\{U_i\}_{i \in I}$ was an arbitrary open cover of K and we have proved that it admits a finite subcover of K, we conclude that K is compact.

For the converse, suppose that $K \subset S$ and K is compact. Let $\{V_i\}_{i\in I}$ be a relatively open cover of K. By the definition of relatively open, for every $i \in I$ there exists an open set U_i such that $U_i \cap S = V_i$. Hence $K \subset \bigcup_{i\in I} V_i \subset \bigcup_{i\in I} U_i$. Since K is assumed to be compact, there exist $i_1, ..., i_n \in I$ such that $K \subset U_{i_1} \cup \cdots \cup U_{i_n}$. Since $K \subset S$, then $K \subset [U_{i_1} \cup \cdots \cup U_{i_n}] \cap S = [U_{i_1} \cap S] \cup \cdots \cup [U_{i_n} \cap S] = V_{i_1} \cup \cdots \cup V_{i_n}$. Therefore, $\{V_i\}_{i\in I}$ admits a finite subcover of K, proving that K is relatively compact. ■

One point Lemma 3.6.7 illustrates is that we don't actually need to define *relatively compact* since a subset K is relatively compact in S if and only if it is compact. However, unlike the case with open and closed sets, it is not generally true that the intersection of an arbitrary compact set with S results in a relatively compact subset. Also, the Heine-Borel theorem does not generally hold either, and really depended heavily on the completeness of $\mathbb{R}$. We explore these issues in Exercise 7 (20).

7. Exercises

(1) Let $J \subset \mathbb{R}$. Prove that J is an interval if and only if for all $x, y \in J$, it holds that $(x, y) \subset J$. *Note*: If $x \geq y$, the interval $(x, y) = \varnothing$, so we don't need to specify whether or not $x < y$ to make sense of this statement.

(2) Let $J \subset \mathbb{R}$. We say that J is *not connected* when there exists some $r \in \mathbb{R}$ such that $J \subset (-\infty, r) \cup (r, \infty)$, but $J \not\subset (-\infty, r)$ and $J \not\subset (r, \infty)$. If this does not hold, then J is called *connected*. Prove that J is connected if and only if J is an interval.

(3) Prove Proposition 3.1.3 (3).

(4) Let $x \in \mathbb{R}$.
   (a) Suppose that $U \subset \mathbb{R}$. Prove that U is a neighborhood of x if and only if there exists some $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subset U$.
   (b) Suppose that $U \subset \mathbb{R}$. Prove that U is a deleted neighborhood of x if and only if $x \notin U$ and there exists some $\varepsilon > 0$ such that $(x - \varepsilon, x) \cup (x, x + \varepsilon) \subset U$.

(5) Finish the proof of Proposition 3.1.10.

(6) Let $U \subset \mathbb{R}$ which is bounded. Prove that $\sup(U)$ and $\inf(U)$ are boundary points of U.

(7) Prove Lemma 3.3.6.

(8) For any set $U \subset \mathbb{R}$, prove that $\text{limpts}(U) \setminus U = \text{bdry}(U) \setminus U$.

(9) Let $U \subset \mathbb{R}$. An element $x \in U$ is called an *isolated point* of U when there exists an $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \cap U = \{x\}$.
   (a) Prove that if x is an isolated point of U, then $x \in \text{bdry}(U)$.
   (b) Prove that if x is an isolated point of U, then $x \notin \text{limpts}(U)$.
   (c) Prove that if $x \in \text{bdry}(U) \cup U$ and x is not an isolated point of U, then x is a limit point of U.

(10) Let $U = \left\{\frac{1}{n} \text{ s.t. } n \in \mathbb{N}\right\} \cup (2, 3]$.
   (a) What is $\text{bdry}(U)$? Justify your answer.
   (b) What is $\text{limpts}(U)$? Justify your answer.

(11) Let $U \subset \mathbb{R}$.
   (a) Prove that U is open if and only if U and the boundary of U are disjoint; i.e., U is open if and only if $\text{bdry}(U) \cap U = \varnothing$.
   (b) Prove that if U is open, then $U \subset \text{limpts}(U)$.

(12) Prove that the only two subsets of $\mathbb{R}$ which are both open and closed are, in fact, $\mathbb{R}$ and $\varnothing$. One way to do this is to show that if $U \subset \mathbb{R}$ then U is both open and closed if and only if $\text{bdry}(U) = \varnothing$ (look at Exercise (11) and Theorem 3.4.2 to establish this) and then show that $\text{bdry}(U) = \varnothing$ if and only if $U = \mathbb{R}$ or $U = \varnothing$.

(13) Let $C \subset \mathbb{R}$. In Example 3.2.2, we showed that such a set C is closed. Here you will prove this fact in two different ways.
   (a) Prove that $C = \text{bdry}(C)$ and apply Theorem 3.4.2.
   (b) Prove that $\text{limpts}(C) = \varnothing$ and apply Theorem 3.4.2.

(14) Prove Lemma 3.4.3.

(15) Suppose that $C \subset \mathbb{R}$.
   (a) Suppose that C is not closed. Prove that there exists a sequence $\{x_n\}_{n=1}^{\infty}$ in C and an element $x \notin C$ such that $\lim_{n\to\infty} x_n = x$. *Hint*: Try $x \in \text{limpts}(C) \setminus C$.
   (b) Suppose that C is not bounded. Prove that there is a sequence $\{x_n\}_{n=1}^{\infty}$ in C which diverges.

(16) Let U and V be disjoint open subsets of $\mathbb{R}$. Suppose that $x \in U$ and $y \in V$ satisfy $x < y$. Prove that there exists some $r \in \mathbb{R}$ such that $x < r < y$ with $r \notin U$ and $r \notin V$. *Hint*: The value $r = \sup\{u \in U : u < y\}$ should work.

(17) Suppose that $K \subset \mathbb{R}$.

(a) Suppose that K is not closed, then by Theorem 3.4.2, there must be some limit point x of K such that $x \notin K$. Prove that $\{U_n\}_{n=1}^{\infty}$, with U_n defined by $U_n = (-\infty, x - 1/n) \cup (x + 1/n, \infty)$, is an open cover of K which does not admit a finite subcover.

(b) Suppose that K is not bounded. Prove that $\{U_n\}_{n=1}^{\infty}$, with U_n defined by $U_n = (-n, n)$, is an open cover of K with no finite subcover.

(c) Prove the "other direction" of the Heine-Borel Theorem 3.5.5. That is, prove that if $K \subset \mathbb{R}$ is compact, then it is closed and bounded. *Hint*: The two previous parts should be a huge help in proving the contrapositive.

(18) Prove Proposition 3.6.3.

(19) Let $U \subset \mathbb{R}$ be open and $C \subset \mathbb{R}$ be closed.

(a) Prove that a subset $V \subset U$ is relatively open in U if and only if V is open in $\mathbb{R}$. However, find a counter example to show that there exists a closed set C and a relatively open set $V \subset C$, but V is not open as a subset of $\mathbb{R}$.

(b) Prove that a subset $K \subset C$ is relatively closed in C if and only if K is closed in $\mathbb{R}$. However, find a counter example to show that there is an open set U and a relatively closed set $K \subset U$, but K is not closed as a subset of $\mathbb{R}$.

(20) Consider the set $C = [-\sqrt{2}, \sqrt{2}] \cap \mathbb{Q}$.

(a) Prove that C is both relatively open and relatively closed in $\mathbb{Q}$.

(b) Prove that C is not relatively compact, which shows that the Heine-Borel theorem generally hold when working in relative topologies (we really did need the completeness of $\mathbb{R}$, which is not available in $\mathbb{Q}$).

(21) Let $S \subset \mathbb{R}$. Prove that a subset $C \subset S$ is relatively closed in S if and only if every sequence in C which converges in S actually converges in C. However, show that this does not imply that a relatively closed set is complete by giving an example of a set C which is relatively closed in $\mathbb{Q}$, but is not complete.

CHAPTER 4

Continuous Functions on $\mathbb{R}$

We now start the exploration of functions $\mathbb{R} \to \mathbb{R}$. More specifically, we will start looking at continuous functions.

DEFINITION 4.0.1. Let U be a non-empty subset in $\mathbb{R}$ and let $f : U \to \mathbb{R}$ be a function. We say that f is *continuous* at a point $x_0 \in U$ when for every $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that if $x \in U$ and $\text{dist}(x_0, x) < \delta_\varepsilon$ then $\text{dist}(f(x_0), f(x)) < \varepsilon$. If it so happens that $f : U \to \mathbb{R}$ is continuous at every point in U, then we simply say that f is *continuous* (or sometimes *continuous on* U).

EXAMPLE 4.0.2. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & x \in \mathbb{Q} \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

Then f is continuous at the point 0, but is not continuous at any other point $x_0 \in \mathbb{R} \setminus \{0\}$. To prove that f is continuous at 0, let $\varepsilon > 0$. Let us make two observations. The first is that for every $x \in \mathbb{R}$, we have $|f(x)| \leq |x|$ (since either $|f(x)| = |x|$ when $x \in \mathbb{Q}$, or $|f(x)| = 0$ when $x \notin \mathbb{Q}$). The second is that $\text{dist}(x, 0) = |x - 0| = |x|$ and $\text{dist}(f(x), f(0)) = |f(x) - f(0)| = |f(x) - 0| = |f(x)| \leq |x|$; this shows that $\text{dist}(f(x), f(0)) \leq \text{dist}(x, 0)$. Armed with these observations, define $\delta = \varepsilon$. For any $x \in \mathbb{R}$ such that $\text{dist}(x, 0) < \delta$ we find $\text{dist}(f(x), f(0)) \leq \text{dist}(x, 0) < \delta = \varepsilon$. This proves that f is continuous at 0.

To prove that f is not continuous at any other point, let $x_0 \in \mathbb{R} \setminus \{0\}$. We will prove that for any $\delta > 0$, there exists some $x \in \mathbb{R}$ such that $\text{dist}(x, x_0) < \delta$ and $\text{dist}(f(x), f(x_0)) \geq |x_0|/2$. To do so, let $\delta > 0$. Suppose first that $x_0 \in \mathbb{Q}$. Then $f(x_0) = x_0$. By the density of the irrational numbers, there is some element $x \in \mathbb{R} \setminus \mathbb{Q}$ such that $\text{dist}(x, x_0) < \delta$. Since $f(x) = 0$ for this x, we find $\text{dist}(f(x), f(x_0)) = |f(x) - f(x_0)| = |0 - x_0| = |x_0| > |x_0|/2$. Alternatively, suppose that $x_0 \in \mathbb{R} \setminus \mathbb{Q}$. Then $f(x_0) = 0$. By the density of the rational numbers, we can find some rational number x with $|x| \geq |x_0|/2$ and $\text{dist}(x, x_0) < \delta$; Indeed, there must be

some rational $x \in (x_0 - \delta, x_0 + \delta) \cap (x_0 - |x_0|/2, x_0 + |x_0|/2)$. Now, since $f(x) = x$, we find $\text{dist}(f(x), f(x_0)) = |f(x) - f(x_0)| = |x - 0| = |x| \geq |x_0|/2$. In either case, we find that given any $\delta > 0$ there exist some $x \in \mathbb{R}$ with $\text{dist}(x, x_0) < \delta$ and $\text{dist}(f(x), f(x_0)) \geq |x_0|/2$, proving that f is not continuous at x_0. □

1. Continuous Functions and Sequences

There are many ways to recognize continuity of functions, one of the most useful is how a function maps sequences. We spend this section exploring some of these ideas, starting with the main theorem.

THEOREM 4.1.1. *Let $U \subset \mathbb{R}$ and let $f : U \to \mathbb{R}$ be a function. The function $f : U \to \mathbb{R}$ is continuous at some point $x_0 \in U$ if and only if given every sequence $\{x_n\}_{n=1}^{\infty}$ in U converging to x_0, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges to $f(x_0)$.*

Proof. Suppose first that f is continuous at x_0. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in U converging to x_0. We need to prove that $\lim_{n\to\infty} f(x_n) = f(x_0)$. To this end, let $\varepsilon > 0$. By the assumption of continuity, there exists some $\delta > 0$ such that if $x \in U$ and $\text{dist}(x, x_0) < \delta$ then $\text{dist}(f(x), f(x_0)) < \varepsilon$. Since the sequence $\{x_n\}_{n=1}^{\infty}$ converges to x_0, there is some $n_\delta \in \mathbb{N}$ such that for every $n \geq n_\delta$, it holds that $\text{dist}(x_n, x_0) < \delta$. But then, by our choice of δ, for every $n \geq n_\delta$, it holds that $\text{dist}(f(x_n), f(x_0)) < \varepsilon$. This proves that $\lim_{n\to\infty} f(x_n) = f(x_0)$.

Conversely, assume that for every sequence $\{x_n\}_{n=1}^{\infty}$ in U converging to $x_0 \in U$, it holds that $\lim_{n\to\infty} f(x_n) = f(x_0)$. We need to prove that f is continuous at x_0. We will argue by contradiction. Assume that f is not continuous at x_0. Then, there exists some $\varepsilon > 0$ such that for every $\delta > 0$ there is some $x \in U$ such that $\text{dist}(x, x_0) < \delta$ and $\text{dist}(f(x), f(x_0)) \geq \varepsilon$. In particular, with this choice of ε, for every $n \in \mathbb{N}$ we can find an element $x_n \in U$ such that $\text{dist}(x_n, x_0) < 1/n$ and $\text{dist}(f(x_n), f(x_0)) \geq \varepsilon$. However, this sequence $\{x_n\}_{n=1}^{\infty}$ that we have defined converges to x_0 (since $\text{dist}(x_n, x_0) < 1/n$ for every n), but the sequence $\{f(x_n)\}_{n=1}^{\infty}$ does not converge to x_0 (since $\text{dist}(f(x_n), f(x_0)) \geq \varepsilon$ for every n), which contradicts our assumption. This contradiction implies that f is continuous at x_0. ■

REMARK 4.1.2. Notice that Theorem 4.1.1 says that if $\{x_n\}_{n=1}^{\infty}$ is a sequence in U which converges in U at a point of continuity of a function $f : U \to \mathbb{R}$, then we can *commute* the limit and the function: $\lim_{n\to\infty} f(x_n) = f\big(\lim_{n\to\infty} x_n\big)$. □

COROLLARY 4.1.3. *Let $U \subset \mathbb{R}$ and let $f : U \to \mathbb{R}$ be a function. The function f is continuous at $x_0 \in U$ if and only if for all sequences $\{x_n\}_{n=1}^{\infty}$ in U which converges to x_0, it holds that the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges in $\mathbb{R}$.*

Proof. The difference between what we are currently trying to prove here and what we have proved in Theorem 4.1.1 is that here we show that we don't add the assumption that the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges to $f(x_0)$, but rather, we simply assume convergence. If we assume that f is continuous at x_0, then given any sequence $\{x_n\}_{n=1}^{\infty}$ in U converging to x_0, it holds that $\{f(x_n)\}_{n=1}^{\infty}$ converges by Theorem 4.1.1. Conversely, suppose that for every sequence $\{x_n\}_{n=1}^{\infty}$ in U converging to x_0, it holds that $\{f(x_n)\}_{n=1}^{\infty}$ converges. We will prove that this implies that $\{f(x_n)\}_{n=1}^{\infty}$ converges to $f(x_0)$. To do so, let $\{x_n\}_{n=1}^{\infty}$ be any sequence in U converging to x_0. Define the sequence $\{\tilde{x}_n\}_{n=1}^{\infty}$ as follows

$$\tilde{x}_n = \begin{cases} x_{n/2} & n \text{ even} \\ x_0 & n \text{ odd} \end{cases}$$

That is, $\{\tilde{x}_n\}_{n=1}^{\infty} = \{x_0, x_1, x_0, x_2, x_0, x_3, ...\}$. Now, the subsequence of even terms of $\{\tilde{x}_n\}_{n=1}^{\infty}$ is the original sequence $\{x_n\}_{n=1}^{\infty}$, which converges to x_0 by assumption; the subsequence of odd terms of $\{\tilde{x}_n\}_{n=1}^{\infty}$ is constantly equal to x_0, and hence converges to x_0. Hence $\{\tilde{x}_n\}_{n=1}^{\infty}$ is a sequence in U which converges to x_0. By assumption, this implies that the sequence $\{f(\tilde{x}_n)\}_{n=1}^{\infty}$ converges. However, the subsequence of odd terms of $\{f(\tilde{x}_n)\}_{n=1}^{\infty}$ is constant $f(x_0)$, and hence coverges to $f(x_0)$, which implies that the entire sequence converges to $f(x_0)$. This then implies that the subsequence of even terms of $\{f(\tilde{x}_n)\}_{n=1}^{\infty}$ also converges to $f(x_0)$, and since this subsequence of even terms is the same sequence as $\{f(x_n)\}_{n=1}^{\infty}$, we conclude that $\lim_{n\to\infty} f(x_n) = f(x_0)$. By Theorem 4.1.1, we have proved that f is continuous at x_0. ■

COROLLARY 4.1.4. *Let $U \subset \mathbb{R}$ and let $f : U \to \mathbb{R}$ be a function. The function f is continuous on U if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ in U which converges in U, it holds that $\{f(x_n)\}_{n=1}^{\infty}$ is a convergent sequence.*

Proof. This is an immediate consequence of Corollary 4.1.5. ■

COROLLARY 4.1.5. *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a function. Then f is continuous on $\mathbb{R}$ if and only if given any convergent sequence $\{x_n\}_{n=1}^{\infty}$, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges. That is, f is continuous on $\mathbb{R}$ if and only if f maps convergent sequences to convergent sequences.*

Proof. This is an immediate consequence of Corollary 4.1.5. ■

EXAMPLE 4.1.6 (A Non-Example). Consider the function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined by $f(x) = 1/x$. The sequence $\{1/n\}_{n=1}^{\infty}$ is in the domain of f and converges in $\mathbb{R}$. However, the sequence $\{f(1/n)\}_{n=1}^{\infty} = \{n\}_{n=1}^{\infty}$ does not converge in $\mathbb{R}$. Although this may seem to contradict the results in this section, it does not. The reason this does not contradict our previous results is that, while the sequence $\{1/n\}_{n=1}^{\infty}$ is in the domain of f and converges in $\mathbb{R}$, it does not converge in the domain of f.

EXAMPLE 4.1.7. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = a\,x^2$ for some constant $a \in \mathbb{R}$. Then f is continuous. To prove this, Theorem 2.3.6 implies that if $\{x_n\}_{n=1}^{\infty}$ is a convergent sequence, then so does the sequence $\{a\,x_n^2\}_{n=1}^{\infty}$. That is, if $\{x_n\}_{n=1}^{\infty}$ converges, so does $\{f(x_n)\}_{n=1}^{\infty}$. Since f maps convergent sequences to convergent sequences, we can conclude that f is continuous. □

2. Continuous Functions and Topology

THEOREM 4.2.1. *Suppose that $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$ is a function. The following hold.*

(1) *The function f is continuous at x_0 if and only if for all neighborhoods V of $f(x_0)$, it holds that $f^{-1}(V)$ is a relative neighborhood of x_0 in U.*
(2) *The function f is continuous on U if and only if for all open subsets $V \subset \mathbb{R}$, the inverse image $f^{-1}(V)$ is relatively open in U.*

Proof. (1) Assume first that $f : U \to \mathbb{R}$ is continuous at x_0. Let V be a neighborhood of $f(x_0)$. We can find some $\varepsilon > 0$ so that $(f(x_0) - \varepsilon, f(x_0) + \varepsilon) \subset V$. For notational simplicity, let $V_\varepsilon = (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$. By the continuity of f at x_0, we can find some $\delta_\varepsilon > 0$ such that for all $x \in U$ with $\operatorname{dist}(x_0, x) < \delta_\varepsilon$ it holds that $\operatorname{dist}(f(x), f(x_0)) < \varepsilon$. That is, $x \in (x_0 - \delta_\varepsilon, x_0 + \delta_\varepsilon) \cap U \implies f(x) \in V_\varepsilon$, which is equivalent to $(x_0 - \delta, x_0 + \delta) \cap U \subset f^{-1}(V_\varepsilon)$. Since $V_\varepsilon \subset V$, we have $f^{-1}(V_\varepsilon) \subset f^{-1}(V)$, which, by Proposition 3.6.5, proves that $f^{-1}(V)$ is a relative neighborhood of x in U. The converse is proved in Exercise 9 (9).

(2) First assume that $f : U \to \mathbb{R}$ is continuous. Let $V \subset \mathbb{R}$ be open. We must prove that $f^{-1}(V)$ is relatively open in U. If $f^{-1}(V) = \varnothing$, then we're done since $\varnothing$ is open and $\varnothing = \varnothing \cap U$. If $f^{-1}(V) \neq \varnothing$, then for all $x \in f^{-1}(V)$, it holds that V is a neighborhood of $f(x)$ and from (1), this implies that $f^{-1}(V)$ is a relative neighborhood of x in U. This proves that $f^{-1}(V)$ is a relative neighborhood of all elements in U, implying that $f^{-1}(V)$ is open.

For the converse, assume that $f^{-1}(V)$ is relatively open in U whenever V is open. We will to prove that f is continuous. Let $x_0 \in U$ be arbitrary. Let V be a neighborhood of $f(x_0)$ and find some $\varepsilon > 0$ such that $(f(x_0)-\varepsilon, f(x_0)+\varepsilon) \subset V$). As before, let $V_\varepsilon = (f(x_0)-\varepsilon, f(x_0)+\varepsilon)$. By assumption, $f^{-1}(V_\varepsilon)$ is relatively open in U and, by how V_ε was defined, $x_0 \in f^{-1}(V_\varepsilon)$; in particular, $f^{-1}(V_\varepsilon)$ is a relative neighborhood of x_0 in U. Since $V_\varepsilon \subset V$, we find that $f^{-1}(V_\varepsilon) \subset f^{-1}(V)$, implying that $f^{-1}(V)$ is also a relative neighborhood of x_0 in U. By (1), we deduce that f is continuous at x_0, and since x_0 was chosen arbitrarily, we conclude that f is continuous. ■

While the previous theorem works in relative topologies on $\mathbb{R}$, when the domain set U of a function $f : U \to \mathbb{R}$ is open, the manifestation of this theorem is much more simply stated, which we record here.

COROLLARY 4.2.2. *Suppose that $U \subset \mathbb{R}$ is open and $f : U \to \mathbb{R}$ is a function. The following hold.*

(1) *The function f is continuous at $x_0 \in U$ if and only if given any neighborhood V of $f(x_0)$, the inverse image $f^{-1}(V)$ is a neighborhood of x_0.*
(2) *The function f is continuous in U if and only if given any open set V, the inverse image $f^{-1}(V)$ is open.*

Proof. These follows immediately from Theorem 4.2.1, since if U is open, then any relative neighborhood of x_0 in U is a neighborhood of x_0, and any relatively open set in U is open. ■

3. Continual Limits of Functions

While considering continuity, we have so far allowed the domain of a function to be a general subset of $\mathbb{R}$. When considering "continual limits" we must be a bit more selective about the domains we consider, since to take a continual limit, we really want to have the function variable move along a continuum of values. Namely, this requires that we assume our domain sets contain an interval.

DEFINITION 4.3.1. Let U be a subset in $\mathbb{R}$ containing the interval (a, b) for some $a < b$. Let $f : U \to \mathbb{R}$ be a function. We say that f has a *lefthand limit* at b, or that a lefthand limit of f exists at b, when there exists some $y_b \in \mathbb{R}$ such that for all $\varepsilon > 0$ there exists some $\delta_\varepsilon > 0$ such that for any $x \in U$ with $x < b$ and $\text{dist}(x, b) < \delta_\varepsilon$, it holds that $\text{dist}(f(x), y_b) < \varepsilon$. In this

case, we say that y_b is the lefthand limit of f at b and often denote this by $\lim_{x\to b-} f(x) = y_b$. Symmetrically, we say that f has a *righthand limit* at a, or that a righthand limit of f exists at a, when there is some $y_a \in \mathbb{R}$ such that for all $\varepsilon > 0$ there exists some $\delta_\varepsilon > 0$ such that for any $x \in U$ with $a < x$ and $dist(x, a) < \delta_\varepsilon$, it holds that $\text{dist}(f(x), y_a) < \varepsilon$. In this case, we say that y_a is the righthand limit of f at a and often denote this by $\lim_{x\to a+} f(x) = y_a$.

NOTATION 4.3.2. Let $\{x_n\}_{n=1}^\infty$ be a sequence and $x_0 \in \mathbb{R}$. We will write $x_n \uparrow x_0$ when $x_n < x_0$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = x_0$. If $x_n \uparrow x_0$, we will say that the sequence $\{x_n\}_{n=1}^\infty$ converges to x_0 *from below*. Analogously, we will write $x_n \downarrow x_0$ when $x_n > x_0$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = x_0$. If $x_n \downarrow x_0$, we will say that the sequence $\{x_n\}_{n=1}^\infty$ converges to x_0 *from above*.

LEMMA 4.3.3. *Suppose that $U \subset \mathbb{R}$ contains an the interval (a, b) for some $a < b$. Suppose that $f : U \to \mathbb{R}$ is a function. Then there exists some $y_b \in \mathbb{R}$ such that $\lim_{x\to b-} f(x) = y_b$ if and only if for all sequences $\{x_n\}_{n=1}^\infty$ in U converging to y_b from below, it hold that $\{f(x_n)\}_{n=1}^\infty$ converges to y_b. Symmetrically, there exists some $y_a \in \mathbb{R}$ such that $\lim_{x\to a+} f(x) = y_a$ if and only if for all sequences $\{x_n\}_{n=1}^\infty$ converging to a from above, it holds that $\{f(x_n)\}_{n=1}^\infty$ converges to y_a.*

Proof. Suppose first that the lefthand limit of f exists at b and is equal to y_b; i.e., $\lim_{x\to b-} f(x) = y_b$. Let $\{x_n\}_{n=1}^\infty$ be a sequence in U converging to b from below. We need to prove that $\{f(x_n)\}_{n=1}^\infty$ converges to y_b. Let $\varepsilon > 0$. Let $\delta_\varepsilon > 0$ be chosen such that if $x \in U$ with $x < b$ and $\text{dist}(x, b) < \delta_\varepsilon$, it holds that $\text{dist}(f(x), y_b) < \varepsilon$. Since $\lim_{n\to\infty} x_n = b$, there exists some $n_\varepsilon \in \mathbb{N}$ such that for every $n \geq n_\varepsilon$ it holds that $\text{dist}(x_n, b) < \varepsilon$. Hence, for every $n \geq n_\varepsilon$, we have that $x_n \in U$, $x_n < b$, and $\text{dist}(x_n, b) < \varepsilon$, which implies that $\text{dist}(f(x_n), y_b) < \varepsilon$. This proves that $\lim_{n\to\infty} f(x_n) = y_b$.

We now argue the converse direction by a contrapositive argument. Assume that y_b is not the lefthand limit of f at b. This means that there exists some $\varepsilon > 0$ such that for all $\delta > 0$, there exists some $x \in U$ with $x < b$ and $\text{dist}(x, b) < \delta$, but $\text{dist}(f(x), y_b) \geq \varepsilon$. In particular, for every $n \in \mathbb{N}$, there exists some $x_n \in U$ with $x_n < b$ and $\text{dist}(x_n, b) < 1/n$ such that $\text{dist}(f(x_n), y_n) \geq \varepsilon$. By definition, the sequence we have defined $\{x_n\}_{n=1}^\infty$ in U converges to b from below, but $\{f(x_n)\}_{n=1}^\infty$ does not converge to y_b since $\text{dist}(f(x_n), y_b) \geq \varepsilon$ for every n. This finishes the contrapositive argument.

The symmetric arguments for the righthand limit are done in Exercise 7 (**??**). ∎

DEFINITION 4.3.4. Suppose that $U \subset \mathbb{R}$ is a neighborhood or a deleted neighborhood of some point $x_0 \in \mathbb{R}$. Let $f : U \to \mathbb{R}$ be a function. We say that f has a *limit* at x_0, or that a limit of f exists at x_0, when there exists some $y_0 \in \mathbb{R}$ such that $\lim_{x \to x_0-} f(x) = y_0$ and $\lim_{x \to x_0+} f(x) = y_0$. That is, f has a limit at x_0 when f has righthand and lefthand limits at x_0 and these limits are equal. In this case, we say that y_0 is the limit of f at x_0 and often denote this by $\lim_{x \to x_0} f(x) = y_0$.

PROPOSITION 4.3.5. *Let $U \subset \mathbb{R}$ be a neighborhood of some point x_0 and $f : U \to \mathbb{R}$ be a function. Then f is continuous at x_0 if and only if $f : U \to \mathbb{R}$ has a limit at x_0 and the limit equals $f(x_0)$. Symbolically, $f : U \to \mathbb{R}$ is continuous at x_0 if and only if $\lim_{x \to x_0} f(x) = f(x_0)$.*

Proof. Suppose first that f is continuous at x_0. Let $\varepsilon > 0$. By the definition of continuity, there is some $\delta_\varepsilon > 0$ such that for every $x \in U$ with $\text{dist}(x, x_0) < \delta_\varepsilon$ it holds that $\text{dist}(f(x), f(x_0)) < \varepsilon$. Then certainly for any $x \in U$ with $x < x_0$ and $dist(x, x_0) < \delta_\varepsilon$ it holds that $\text{dist}(f(x), f(x_0)) < \varepsilon$, showing that $\lim_{x \to x_0-} f(x) = f(x_0)$. The symmetric argument works to show that $\lim_{x \to x_0+} f(x) = f(x_0)$. Hence, if f is continuous at x_0, then $\lim_{x \to x_0-} f(x) = \lim_{x \to x_0+} f(x) = f(x_0)$.

For the converse, assume that $\lim_{x \to x_0-} f(x) = \lim_{x \to x_0+} f(x) = f(x_0)$. Let $\varepsilon > 0$ be arbitrary. Using the definition of lefthand limits, find $\delta_1 > 0$ such that for every $x \in U$ with $x < x_0$ and $\text{dist}(x, x_0) < \delta_1$ it holds that $\text{dist}(f(x), f(x_0)) < \varepsilon$. Symmetrically, using the definition of righthand limits, take $\delta_2 > 0$ such that for every $x \in U$ with $x_0 < x$ and $\text{dist}(x, x_0) < \delta$ it holds that $\text{dist}(f(x), f(x_0)) < \varepsilon$. Define $\delta_\varepsilon = \min\{\delta_1, \delta_2\}$. By our definition of δ_ε, we then have that if $x \in U$ with $\text{dist}(x, x_0) < \delta_\varepsilon$, it holds that $\text{dist}(f(x), f(x_0)) < \varepsilon$, proving that f is continuous at x_0. ■

4. Continuous Images of Sets

4.1. Continuous Images of Compact Sets.

THEOREM 4.4.1. *Suppose that $U \subset \mathbb{R}$ is open, and $K \subset \mathbb{R}$ is compact with $K \subset U$. If $f : U \to \mathbb{R}$ is continuous, then the image $f(K)$ is compact. Poetically, the continuous image of a compact set is compact.*

Proof. Suppose that $\{U_i\}_{i \in I}$ is an open cover of $f(K)$. Then by Exercise 9 (14), $\{f^{-1}(U_i)\}_{i \in I}^{\infty}$ is an open cover of K. Since K is compact, there exists a finite subcover

$\{f^{-1}(U_{i_k})\}_{k=1}^N$ of K. But then $\{U_{i_k}\}_{k=1}^N$ is a finite subcover of $f(K)$. Hence any open cover of $f(K)$ admits a finite subcover of $f(K)$ proving that $f(K)$ is compact. ■

COROLLARY 4.4.2. *Suppose $U \subset \mathbb{R}$ is open, $K \subset U$ is non-empty and compact, and $f : U \to \mathbb{R}$ is continuous. Set $y_{\max} = \sup\{f(x) \text{ s.t. } x \in K\} = \sup f(K)$ and $y_{\min} = \inf\{f(x) \text{ s.t. } x \in K\} = \inf f(K)$. Then there exists (not necessarily unique) elements $x_{\min}, x_{\max} \in K$ such that $f(x_{\max}) = y_{\max}$ and $f(x_{\min}) = y_{\min}$.*

Proof. By Theorem 4.4.1, we know that $f(K)$ is compact, in particular it's closed and bounded. In Exercise 9 (6) we have shown that $y_{\max}$ and $y_{\min}$ are in the boundary of $f(K)$. Since $f(K)$ is closed, this implies that $y_{\max}, y_{\min} \in f(K)$, which further implies there exist some points $x_{\max}, x_{\min} \in K$ such that $f(x_{\min}) = y_{\min}$ and $f(x_{\max}) = y_{\max}$. ■

COROLLARY 4.4.3. *Suppose that $U \subset \mathbb{R}$ is open and $K \subset U$ is compact. If $f : U \to \mathbb{R}$ is continuous, then the restriction $f : K \to \mathbb{R}$ is bounded. That is, there exists some $R > 0$ such that for all $x \in K$, it holds that $|f(x)| \leq R$.*

Proof. Since $f(K)$ is compact, it is closed and bounded. In particular, the set $f(K) = \{f(x) \text{ s.t. } x \in K\}$ is bounded, which is equivalent to the existence of some $R > 0$ such that for all $x \in K$ it holds that $|f(x)| \leq R$. ■

4.2. Continuous Images of Intervals: The Intermediate Value Theorem. We now work to rigorously prove a concept that in illustration is obvious. The Intermediate Value Theorem (IVT) roughly states that if a real-valued function f is continuous on some interval $[a, b]$, then it can not "jump" from $f(a)$ to $f(b)$ without hitting all values between (all intermediate values). Before jumping right into the statement, we first prove a lemma that slightly generalizes one direction of Exercise 7 (2).

LEMMA 4.4.4. *Let $J \subset \mathbb{R}$ be a non-empty set. If J is an interval, then there do not exist disjoint open sets A and B such that $J \subset U \cup V$ with $J \cap U \neq \varnothing$ and $J \cap V \neq \varnothing$.*

Proof. We will prove this by the contrapositive. Suppose that J is a non-empty subset such that there disjoint exist open sets U and V such that $J \subset U \cup V$ with $J \cap U \neq \varnothing$ and $J \cap V \neq \varnothing$. Let $x \in J \cap U$ and $y \in J \cap V$, and without loss of generality, assume that $x < y$. Since $x \in U$ and $y \in V$, and U and V are disjoint open sets, Exercise 7 (16) implies that there exists some $r \in \mathbb{R}$ with $x < r < y$ and $r \notin U$ and $r \notin V$. Since $r \notin U \cup V$ and $J \subset U \cup V$, we

have that $J \subset (-\infty, r) \cup (r, \infty)$. Moreover, $x \in J \cap (-\infty, r)$ and $y \in J \cap (r, \infty)$. By Exercise 7 (2), this proves that J is not an interval. ■

DEFINITION 4.4.5. Let $x, y \in \mathbb{R}$. A value $c \in \mathbb{R}$ is said to be an *intermediate value* between x and y when either $x \leq c \leq y$ or $x \geq c \geq y$.

THEOREM 4.4.6 (Intermediate Value Theorem). *Suppose that $U \subset \mathbb{R}$ is a set containing the interval $[a, b]$ for some $a \leq b$. If $f : U \to \mathbb{R}$ is continuous at all points in $[a, b]$, then for all intermediate values $c \in \mathbb{R}$ between $f(a)$ and $f(b)$, there exists some $x_c \in [a, b]$ such that $f(x_c) = c$.*

Proof. Let $f : U \to \mathbb{R}$ be a function continuous on the interval $[a, b] \subset U$ for some $a \leq b$. For this proof, we do not lose generality by assuming $a < b$ and $f(a) < f(b)$, so we will make these assumptions. Suppose that $c \in \mathbb{R}$ is some intermediate value $f(a) < c < f(b)$. Let $U = (a, b) \cap f^{-1}(-\infty, c)$ and $V = (a, b) \cap f^{-1}(c, \infty)$. If we assume that f is continuous at all points in $[a, b]$, then certainly f is continuous on the open interval (a, b). Because of this, U and V are open sets, since they are the inverse images of open sets by a continuous function f restricted to the open interval (a, b). Moreover, U and V are disjoint since f is well-defined. We will now prove that $U \neq \varnothing$ and $V \neq \varnothing$, which by Lemma 4.4.4 will imply that $(a, b) \not\subset U \cup V$, further implying that there is some $x_c \in (a, b) \setminus (U \cup V)$, meaning $f(x_c) = c$.

We give an almost full proof here with some details relegated to Exercise 9 (**??**). Let us restrict to the case that the domain of f is the interval J, since we are not critiquing the behavior of f outside of J; this restriction means that we are assuming $f : J \to \mathbb{R}$ is continuous. If $f(x_1) = f(x_2)$ there is no c between $f(x_1)$ and $f(x_2)$ to check and hence nothing to prove. So, we assume that $f(x_1) \neq f(x_2)$ and that the value c lands between $f(x_1)$ and $f(x_2)$. Define $U_1 = f^{-1}\big((-\infty, c)\big)$ and $U_2 = f^{-1}\big((c, \infty)\big)$; note that U_1 and U_2 are both open since f is continuous, and disjoint by construction. We claim that $U_1 \cap (x_1, x_2) \neq \varnothing$ and $U_2 \cap (x_1, x_2) \neq \varnothing$. To show this, let us assume that $f(x_1) < c < f(x_2)$; the proof in the case that $f(x_1) > c > f(x_2)$ follows by direct analogy. Since $f(x_1) < c$ and f is continuous, there is some $\delta_1 > 0$ such that for all $x \in (x_1, x_1 + \delta_1)$ it holds that $f(x) < c$, so $(x_1, x_1 + \delta_1) \subset (x_1, x_2) \cap U_1$; similarly, since $f(x_2) > c$ and f is continuous, there is some $\delta_2 > 0$ such that for all $x \in (x_2 - \delta_2, x_2)$ it holds that $f(x) > c$, so $(x_2 - \delta_2, x_2) \subset (x_1, x_2) \cap U_2$; this proves that $(x_1, x_2) \cap U_1 \neq \varnothing$ and $(x_1, x_2) \cap U_2 \neq \varnothing$. Now, Proposition 4.4.4 implies

that $(x_1, x_2) \not\subset U_1 \cup U_2$, so there exists some $x_c \in (x_1, x_2)$ such that $x_c \notin U_1 \cup U_2$. But then, $f(x_c) = c$. ■

5. Pointwise Algebraic Constructions and Continuity

DEFINITION 4.5.1. Suppose that S_1 and S_2 are sets (not necessarily subsets of $\mathbb{R}$), and $f : S_1 \to \mathbb{R}$ and $g : S_2 \to \mathbb{R}$ are functions. Because of the structure on the codomain $\mathbb{R}$, we can create new functions out of f and g via *pointwise* constructions. Let $S = S_1 \cap S_2$ (which very well might be empty).

(1) A *linear combination* of f and g is a function $h : S \to \mathbb{R}$ defined by $h(x) = \alpha f(x) + \beta g(x)$ where $\alpha, \beta \in \mathbb{R}$ are fixed (i.e., α and β remain fixed independent of $x \in S$). The function $h : S \to \mathbb{R}$ is typically denoted by $h = \alpha f + \beta g$.
(2) The *product* of f and g is a function $h : S \to \mathbb{R}$ defined by $h(x) = f(x)g(x)$. The function $h : S \to \mathbb{R}$ is typically denoted by $h = fg$.
(3) If we further restrict S to the set $\tilde{S} = \{x \in S \text{ s.t. } g(x) \neq 0\}$, we then define the *quotient* of f and g as the function $h : \tilde{S} \to \mathbb{R}$ defined by $h(x) = f(x)/g(x)$. The function $h : \tilde{S} \to \mathbb{R}$ is typically denoted $h = f/g$.

Before we start considering some of the continuity properties of linear combinations, products, and quotients of functions, let's first prove a necessary lemma.

LEMMA 4.5.2. *Suppose that $U \subset \mathbb{R}$ is a neighborhood of some point x_0. If $f : U \to \mathbb{R}$ is continuous at x_0 and $f(x_0) = y_0 \neq 0$, there exists some $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta)$ it holds that $|f(x)| > |y_0|/2$. In particular, if $f : U \to \mathbb{R}$ is continuous at x_0 and $f(x_0) \neq 0$, there is a neighborhood N of x_0 such that $f(x) \neq 0$ for any $x \in N$.*

Proof. Suppose that $f : U \to \mathbb{R}$ is continuous at x_0 and $f(x_0) = y_0 \neq 0$. Since f is continuous at x_0 and $|y_0|/2 > 0$, there exists some $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta)$ it holds that $|f(x) - f(x_0)| < |y_0|/2$. The reverse triangle inequality implies that for all $x \in (x_0 - \delta, x_0 + \delta)$, we have $|f(x_0)| - |f(x)| < |y_0|/2$, which rearranges to $|y_0|/2 < |f(x)|$. ■

THEOREM 4.5.3. *Suppose that U_1 and U_2 are both subsets of $\mathbb{R}$ which are both neighborhoods of some point x_0. Further suppose that $f : U_1 \to \mathbb{R}$ and $g : U_2 \to \mathbb{R}$ are both continuous at x_0. Set $U = U_1 \cap U_2$. Then*

(1) *For any $\alpha, \beta \in \mathbb{R}$, the function $\alpha f + \beta g : U \to \mathbb{R}$ is continuous at x_0.*

(2) *The function* $fg : U \to \mathbb{R}$ *is continuous at* x_0.

(3) *If* $g(x_0) \neq 0$, *then there exists some* $\delta > 0$ *such that* $f/g : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ *is well-defined and continuous at* x_0. *Moreover, if* $g(x) \neq 0$ *for all* $x \in U$, *then* $f/g : U \to \mathbb{R}$ *is well-defined and continuous at* x_0.

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be an arbitrary sequence in U converging to x_0. By Theorem **??**, the sequences $\{f(x_n)\}_{n=1}^{\infty}$ and $\{g(x_n)\}_{n=1}^{\infty}$ converge to $f(x_0)$ and $g(x_0)$, respectively.

(1) Let $\alpha, \beta \in \mathbb{R}$ and $h : U \to \mathbb{R}$ be defined by $h(x) = \alpha f(x) + \beta g(x)$. Then the sequence $\{h(x_n)\}_{n=1}^{\infty}$ is given by $h(x_n) = \alpha f(x_n) + \beta g(x_n)$ and, by Theorem 2.3.6, converges to $\alpha f(x_0) + \beta g(x_0) = h(x_0)$. Therefore, $\lim_{n\to\infty} h(x_n) = h(x_0)$. Since $\{x_n\}_{n=1}^{\infty}$ was chosen arbitrarily, we conclude by Theorem **??** that h is continuous at x_0.

(2) Let $k : U \to \mathbb{R}$ be defined by $k(x) = f(x)g(x)$. Then the sequence $\{k(x_n)\}_{n=1}^{\infty}$ is given by $k(x_n) = f(x_n)g(x_n)$ and, by Theorem 2.3.6, converges to $f(x_0)g(x_0) = k(x_0)$. Therefore, $\lim_{n\to\infty} k(x_n) = k(x_0)$. Since $\{x_n\}_{n=1}^{\infty}$ was chosen arbitrarily, we conclude by Theorem **??** that k is continuous at x_0.

(3) If $g(x_0) \neq 0$, by Lemma 4.5.2 there exists some $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subset U$ and $g(x) \neq 0$ for any $x \in (x_0 - \delta, x_0 + \delta)$. Let $\{\tilde{x}_n\}_{n=1}^{\infty}$ be an arbitrary sequence in $(x_0 - \delta, x_0 + \delta)$ which converges to x_0. By Theorem **??** we again know that $\{f(\tilde{x}_n)\}_{n=1}^{\infty}$ and $\{g(\tilde{x}_n)\}_{n=1}^{\infty}$ converge to $f(x_0)$ and $g(x_0)$ respectively. Define the function $l : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ by $l(x) = f(x)/g(x)$. Since for all n, $\tilde{x}_n \in (x_0 - \delta, x_0 + \delta)$ we find that $g(x_n) \neq 0$. Therefore, by Theorem , we have that the sequence $\{l(\tilde{x}_n)\}_{n=1}^{\infty}$ given by $l(\tilde{x}_n) = f(\tilde{x}_n)/g(\tilde{x}_n)$ converges to $f(x_0)/g(x_0) = l(x_0)$. Therefore, $\lim_{n\to\infty} l(\tilde{x}_n) = l(x_0)$. Since $\{\tilde{x}_n\}_{n=1}^{\infty}$ was chosen arbitrarily in $(x_0 - \delta, x_0 + \delta)$, we conclude by Theorem **??** that l is continuous at x_0. Had we originally known that $g(x) \neq 0$ for all $x \in U$, the same proof would have worked for l defined on the domain U, replacing the arbitrary sequence $\{\tilde{x}_n\}_{n=1}^{\infty}$ in $(x_0 - \delta, x_0 + \delta)$ with the arbitrary sequence $\{x_n\}_{n=1}^{\infty}$ in U.

■

5.1. Composition and Continuity. When being introduced to composition of functions $f : A \to B$ and $g : C \to D$, to create the composition $g \circ f : A \to D$ it must be true that $f(A) \subset C$, since otherwise, if some $x \in A$ is such that $f(x) \notin C$, then $g(f(x))$ is not defined. However, we can consider composing these functions when $f(A) \not\subset C$ by further restricting the domain. Indeed, let $S = \{x \in A \text{ s.t. } f(x) \in C\} = f^{-1}(C)$. Then $g \circ f : S \to D$ makes

sense; however, you may notice that it could so happen that $S = \varnothing$, which isn't all that interesting. However, if it happens that $f(A) \subset C$, then $S = A$ and we find ourselves in the original scenario.

DEFINITION 4.5.4. Given functions $f : A \to B$ and $g : C \to D$, the set $S = f^{-1}(C) = \{x \in A \text{ s.t. } f(x) \in C\}$ is called the *implicit domain* of the composition $g \circ f : S \to D$. Without explicitly mentioning otherwise, we will always assume the domain of a composition is the implicit domain.

One final point to make is that since we will be using the implicit domain of composition, *any* two functions can be composed.

THEOREM 4.5.5. *Let U_1 and U_2 be subsets of $\mathbb{R}$, and $f : U_1 \to \mathbb{R}$ and $g : U_2 \to \mathbb{R}$ be functions. Suppose that U_1 is a neighborhood of some point $x_0 \in \mathbb{R}$ and U_2 is a neighborhood of $f(x_0)$. If f is continuous at x_0 and g is continuous at $f(x_0)$, then $g \circ f : S \to \mathbb{R}$ is continuous at x_0 where $S = f^{-1}(U_2)$ is the implicit domain of composition.*

Proof. *Method 1*: Let $\varepsilon > 0$ be arbitrary. Let $\delta_1 > 0$ such that if $y \in U_2$ and $|y - f(x_0)| < \delta_1$ the $|g(y) - g(f(x_0))| < \varepsilon$; we can find such a δ_1 since g is assumed to be continuous at $f(x_0)$. Let $\delta > 0$ be such that if $x \in S$ and $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \delta_1$; we can find such a $\delta > 0$ since f is assumed to be continuous at x_0. Now, for any $x \in S$ such that $|x - x_0| < \delta$ we have that $|f(x) - f(x_0)| < \delta$ implying that $|g(f(x)) - g(f(x_0))| < \varepsilon$. That is, for all $x \in S$ such that $|x - x_0| < \delta$ it holds that $|g \circ f(x) - g \circ f(x_0)| < \varepsilon$, proving that $g \circ f$ is continuous at x_0.

Method 2: Let $\{x_n\}_{n=1}^{\infty}$ be an arbitrary sequence in S converging to x_0. Then $\{f(x_n)\}_{n=1}^{\infty}$ is a sequence converging to $f(x_0)$ since f is continuous at x_0. But then, since $\{f(x_n)\}_{n=1}^{\infty}$ is a sequence converging to $f(x_0)$ and g is continuous at $f(x_0)$, it holds that $\{g(f(x_n))\}_{n=1}^{\infty}$ is a sequence converging to $g(f(x_0))$. That is, given an arbitrary sequence $\{x_n\}_{n=1}^{\infty}$ converging to x_0, it holds that $\{g \circ f(x_n)\}_{n=1}^{\infty}$ converges to $g \circ f(x_0)$, proving that $g \circ f$ is continuous at x_0.

Method 3: Let V be an arbitrary neighborhood of $g \circ f(x_0)$. Note that $(g \circ f)^{-1}(V) = f^{-1}(g^{-1}(V))$. Since g is continuous at $f(x_0)$ and V is a neighborhood of $g(f(x_0))$, it holds that $g^{-1}(V)$ is a neighborhood of $f(x_0)$. Since f is continuous at x_0 and $g^{-1}(V)$ is a neighborhood of $f(x_0)$, it holds that $f^{-1}(g^{-1}(V))$ is a neighborhood of x_0; that is, $(g \circ f)^{-1}(V)$ is a

neighborhood of x_0. We have thus shown that given an arbitrary neighborhood V of $g \circ f(x_0)$, it holds that $(g \circ f)^{-1}(V)$ is a neighborhood of x_0, proving that $g \circ f$ is continuous at x_0. ■

6. Stronger Forms of Continuity

6.1. Uniform Continuity.

DEFINITION 4.6.1. Let U be a subset of $\mathbb{R}$ and $f : U \to \mathbb{R}$ be a function. We say that f is *uniformly continuous* when for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in U$ with $|x - y| < \delta$, it holds that $|f(x) - f(y)| < \varepsilon$.

REMARK 4.6.2. Note that the definition of uniformly continuous critiques the behavior of f "uniformly" over its entire domain. Indeed, the definition of uniform continuity requires that the choice of δ depends only on ε, but not on any specific point in U. This is in contrast to first having a point $x_0 \in U$ in mind and then finding some δ depending on ε and x_0, as we would to prove continuity. You'll notice that because of this we do not define uniform continuity at a point, but rather over the entire domain; that is, the statement "$f : U \to \mathbb{R}$ is uniformly continuous" makes sense, but the statement "$f : U \to \mathbb{R}$ is uniformly continuous at x_0" does not. □

EXAMPLE 4.6.3. Let $f : [1, \infty) \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$. Then f is uniformly continuous. To prove this, let $\varepsilon > 0$ and set $\delta = \varepsilon$. Note that if $x, y \in [1, \infty)$ then $|\sqrt{x} + \sqrt{y}| \geq |1 + 1| = 2$, and further that $x - y = (\sqrt{x} - \sqrt{y})(\sqrt{x} + \sqrt{y})$. Therefore, for any $x, y \in [1, \infty)$ with $|x - y| < \delta$ it holds that $|f(x) - f(y)| = |\sqrt{x} - \sqrt{y}| = \dfrac{|x - y|}{|\sqrt{x} + \sqrt{y}|} < \dfrac{\delta}{2} = \dfrac{\varepsilon}{2} < \varepsilon$. □

EXAMPLE 4.6.4. Let $f : (0, 1) \to \mathbb{R}$ be defined by $f(x) = \frac{1}{x}$. Note that $f : (0, 1) \to \mathbb{R}$ is continuous by Theorem 4.5.3 (3) However, f is not uniformly continuous. We prove this by contradiction. Assume that $f : (0, 1) \to \mathbb{R}$ is uniformly continuous. In this case, there exists some $\delta > 0$ such that for all $x, y \in (0, 1)$ with $|x - y| < \delta$ it holds that $|f(x) - f(y)| < 1$. Now, $|f(x) - f(y)| = \left|\frac{1}{x} - \frac{1}{y}\right| = \frac{|x-y|}{|xy|}$. In particular, suppose $N \in \mathbb{N}$ is large enough such that $1/N < \delta$ and $N > 2$, then with $x = \frac{1}{N}$ and $y = \frac{2}{N}$ we have that $|x - y| = \frac{1}{N} < \delta$ and $|f(x) - f(y)| = \frac{1/N}{2/N^2} = \frac{N}{2} > 1$. This contradicts that $|x - y| < \delta \implies |f(x) - f(y)| < 1$, showing that f is not uniformly continuous. □

PROPOSITION 4.6.5. *Suppose that $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$ is uniformly continuous. Then f is continuous with respect to the relative topology on U. In particular, if U is open and $f : U \to \mathbb{R}$ is uniformly continuous, then f is continuous.*

Proof. Exericse. ∎

THEOREM 4.6.6. *Suppose that $K \subset \mathbb{R}$ is compact and $f : K \to \mathbb{R}$ is continuous. Then $f : K \to \mathbb{R}$ is uniformly continuous. That is, a continuous function on a compact set is uniformly continuous.*

Proof. Let $\varepsilon > 0$. Since f is continuous, for each point $x \in K$, there exists some $\delta_x > 0$ such that for all $y \in (x - \delta_x, x + \delta_x) \cap K$ it holds that $|f(x) - f(y)| < \varepsilon/2$. For every $x \in K$ let $U_x = (x - \delta_x/2, x + \delta_x/2)$. Then $\{U_x\}_{x \in K}$ is an open cover of K, and since K is compact, it admits a finite subcover. Hence, let $\{x_1, ..., x_n\} \subset K$ such that $\{U_{x_1}, ..., U_{x_n}\}$ covers K. Define $\delta = \min\{\delta_{x_1}/2, ..., \delta_{x_n}/2\}$. Suppose that $x, y \in K$ with $|x - y| < \delta$. Since $\{U_{x_1}, ..., U_{x_n}\}$ is an open cover of K, it holds that $x \in U_{x_i}$ for some $i \in \{1, 2, ..., n\}$, particularly, this implies that $|x - x_i| < \delta_{x_i}/2$. Now, by the triangle inequality, $|y - x_i| \leq |y - x| + |x - x_i| < \delta + \delta_{x_i}/2 \leq \delta_{x_i}/2 + \delta_{x_i}/2 = \delta_{x_i}$. So, $|y - x_i| < \delta_{x_i}$ and $|x - x_i| < \delta_{x_i}$, which by the way we chose δ_{x_i} implies that $|f(y) - f(x_i)| < \varepsilon/2$ and $|f(x) - f(x_i)| < \varepsilon/2$. In particular, $|f(x) - f(y)| \leq |f(x) - f(x_i)| + |f(y) - f(x_i)| < \varepsilon/2 + \varepsilon/2 = \varepsilon$. What we have thus proved is that if $x, y \in K$ with $|x - y| < \delta$, it holds that $|f(x) - f(y)| < \varepsilon$, and since δ was chosen depending only on ε, but not on the points x nor y, this proves that $f : K \to \mathbb{R}$ is uniformly continuous. ∎

THEOREM 4.6.7. *Suppose that U_1 and U_2 are subsets of $\mathbb{R}$ and $f : U_1 \to \mathbb{R}$ and $g : U_2 \to \mathbb{R}$ are functions. Let $U = U_1 \cap U_2$ and $S = f^{-1}(U_2)$. If f and g are uniformly continuous then*

(1) *For any $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g : U \to \mathbb{R}$ is uniformly continuous.*
(2) *$g \circ f : S \to \mathbb{R}$ is uniformly continuous.*

Proof. (1) Let $\varepsilon > 0$. Define $\delta_1 > 0$ such that for all $x, y \in U_1$ such that $|x - y| < \delta_1$ it holds that $|f(x) - f(y)| < \frac{\varepsilon}{2(|\alpha|+1)}$. Define $\delta_2 > 0$ such that for all $x, y \in U_2$ such that $|x - y| < \delta_2$ it holds that $|g(x) - g(y)| < \frac{\varepsilon}{2(|\beta|+1)}$. Set $\delta = \min(\delta_1, \delta_2)$. Then for all $x, y \in U$

such that $|x-y|<\delta$,

$$\begin{aligned}|\alpha f(x)+\beta g(x)-\alpha f(y)-\beta g(y)| &\leq |\alpha||f(x)-f(y)|+|\beta||g(x)-g(y)| \\ &< |\alpha|\frac{\varepsilon}{2(|\alpha|+1)}+|\beta|\frac{\varepsilon}{2(|\beta|+1)} < \frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.\end{aligned}$$

(2) Let $\varepsilon>0$. Let $\delta_1>0$ be chosen so that for all $x,y\in U_2$ with $|x-y|<\delta_1$ it holds that $|g(x)-g(y)|<\varepsilon$. Let $\delta>0$ such that for all $x,y\in U_1$ with $|x-y|<\delta$ it hlds that $|f(x)-f(y)|<\delta_1$. Now, if $x,y\in S$ with $|x-y|<\delta$, then $|f(x)-f(y)|<\delta_1$, further implying that $|g(f(x))-g(f(y))|<\varepsilon$. That is, for all $x,y\in S$ with $|x-y|<\delta$ it holds that $|g\circ f(x)-g\circ f(y)|<\varepsilon$, proving that $g\circ f:S\to\mathbb{R}$ is uniformly continuous. ■

PROPOSITION 4.6.8. *Suppose that $U\subset\mathbb{R}$ is bounded and $f:U\to\mathbb{R}$ is uniformly continuous. Then $f:U\to\mathbb{R}$ is bounded. That is, there exists some $R>0$ such that for all $x\in U$, it holds that $|f(x)|\leq R$.*

6.2. Lipschitz Continuity.

DEFINITION 4.6.9. Let $U\subset\mathbb{R}$. A function $f:U\to\mathbb{R}$ is called *Lipschitz continuous* when there exists some $c>0$ such that for all $x,y\in U$ it holds that $|f(x)-f(y)|\leq c|x-y|$. Letting $k=\inf\{c>0 \text{ s.t. } |f(x)-f(y)|\leq c|x-y| \text{ for all } x,y\in U\}$, then k is called the *Lipschitz bound* or *Lipschitz constant* of $f:U\to\mathbb{R}$.

PROPOSITION 4.6.10. *Let $U\subset\mathbb{R}$. Suppose that $f:U\to\mathbb{R}$ is Lipschitz continuous with Lipschitz bound k, and define $C=\{c\geq 0 \text{ s.t. } |f(x)-f(y)|\leq c|x-y| \text{ for all } x,y\in U\}$. Then $C=[k,\infty)$. That is, for all $c\geq k$, it holds that $|f(x)-f(y)|\leq c|x-y|$ for all $x,y\in U$.*

Proof. The proof is outlined here, the details are left as an exercise. First note that if $c\in C$, then for any $b\geq c$ we have that $|f(x)-f(y)|\leq c|x-y|\leq b|x-y|$ for all $x,y\in U$; this shows that $b\in C$, from which we conclude that if $c\in C$, then $[c,\infty)\subset C$. Since k is a lower bound of C, this further implies that $k\in C$ implies $[k,\infty)=C$. To prove that $k\in C$, we argue by contradiction. Assume that $k\notin C$; this means that there exists some $x,y\in K$ such that $|f(x)-f(y)|>k|x-y|$. In particular there exists some $n\in\mathbb{N}$ such that $|f(x)-f(y)|-k|x-y|>\frac{1}{n}|x-y|$ and hence $|f(x)-f(y)|>(k+1/n)|x-y|$. But this implies that $k+1/n$ is a lower bound of C, contradicting that k is the infimum of C. ■

EXAMPLE 4.6.11. Let $f:[0,1]\to\mathbb{R}$ be defined by $f(x)=\sqrt{x}$. Then f is not Lipschitz continuous. **Proof.** By contradiction. Suppose that k is the Lipschitz bound of $f:[0,1]\to\mathbb{R}$.

Then for all $x, y \in [0,1]$ it holds that $|f(x) - f(y)| \leq k|x - y|$. In particular, setting $y = 0$ and $x = 1/n^2$ for some $n \in \mathbb{N}$ we see that $|1/n| \leq k|1/n^2|$ implying that $n \leq k$. Since we could have originally chosen $n \in \mathbb{N}$ large enough that $n > k$, we see our contradiction. ■ □

THEOREM 4.6.12. *Let $U \subset \mathbb{R}$ and suppose that $f : U \to \mathbb{R}$ is a function. If f is Lipschitz continuous, then it is uniformly continuous.*

Proof. Exercise. See Exercise 9 (37). ■

THEOREM 4.6.13. *Suppose that U_1 and U_2 are subsets of $\mathbb{R}$ and $f : U_1 \to \mathbb{R}$ and $g : U_2 \to \mathbb{R}$ are functions. Let $U = U_1 \cap U_2$ and $S = f^{-1}(U_2)$. If f and g are Lipschitz continuous then*

(1) *For any $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g : U \to \mathbb{R}$ is Lipschitz continuous.*
(2) *$g \circ f : S \to \mathbb{R}$ is Lipschitz continuous.*

Proof. (1) Suppose that there is some $c_1 > 0$ such that $|f(x) - f(y)| \leq c_1|x - y|$ for all $x, y \in U$, and $c_2 > 0$ such that $|g(x) - g(y)| \leq c_2|x - y|$ for all $x, y \in U$. Then $|\alpha f(x)+\beta g(x)-\alpha f(y)-\beta g(y)| \leq |\alpha||f(x)-f(y)|+|\beta||g(x)-g(y)| \leq (|\alpha|c_1+|\beta|c_2)|x-y|$ for all $x, y \in U$. Setting $c = (|\alpha|c_1 + |\beta|c_2)$, this says that $|\alpha f(x) + \beta g(x) - \alpha f(y) - \beta g(y)| \leq c|x - y|$ for all $x, y \in U$.

(2) Suppose that there is some $c_1 > 0$ such that $|f(x) - f(y)| \leq c_1|x - y|$ for all $x, y \in S$, and $c_2 > 0$ such that $|g(x) - g(y)| \leq c_2|x - y|$ for all $x, y \in U_2$. Then for all $x, y \in S$, we have $|g(f(x)) - g(f(y))| \leq c_2|f(x) - f(y)| \leq c_2c_2|x - y|$. Setting $c = c_1c_2$, we have shown that $|g(f(x)) - g(f(y))| \leq c|x - y|$ for all $x, y \in S$. ■

7. Exercises

(1) Let $S \subset \mathbb{R}$ and let $x \in S$ be an isolated point of S. That is, there exists some $\delta > 0$ such that $(x - \delta, x + \delta) \cap S = \{x\}$ (see Exercise 7 (9) for more on isolated points). Prove that any function $f : S \to \mathbb{R}$ is continuous at x. In particular, what does this say about any sequence $f : \mathbb{N} \to \mathbb{R}$?
(2) Let $S \subset \mathbb{R}$ be defined as $S = \{1/n \text{ s.t. } n \in \mathbb{N}\}$. Exercise (1) proves that any function $f : S \to \mathbb{R}$ is continuous since S is a discrete set (i.e., all elements in S are isolated points of S). However, this is not necessarily true if the domain is $S \cup \{0\}$. Find an example $f : S \cup \{0\} \to \mathbb{R}$ which is not continuous at 0.

(3) Let $U = [0,2] \setminus \{1\}$. Define the function $f : U \to \mathbb{R}$ as

$$f(x) = \begin{cases} -1 & 0 \leq x < 1 \\ 1 & 1 < x \leq 2 \end{cases}$$

(a) Using Definition 4.0.1, prove that f is continuous.

(b) Let $y \in \mathbb{R}$ be any value. Extend f to the interval $[0,2]$ by defining $f(1) = y$. That is,

$$f(x) = \begin{cases} -1 & 0 \leq x < 1 \\ y & x = 1 \\ 1 & 1 < x \leq 2 \end{cases}$$

Using Definition 4.0.1, prove that $f : [0,2] \to \mathbb{R}$ is not continuous. (This proves that there is no way to "continuously extend" f to the entire interval $[0,2]$).

(4) Let $U \subset \mathbb{R}$ and let $f : U \to \mathbb{R}$ be a function.

(a) Suppose that f has the property that for all Cauchy sequences $\{x_n\}_{n=1}^{\infty}$ in U, it holds that the sequence $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy. Prove that f is continuous.

(b) In the case $U = \mathbb{R}$, prove that a function $f : U \to \mathbb{R}$ is continuous if and only if it satisfies the previous property. That is, prove that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only if for all Cauchy sequences $\{x_n\}_{n=1}^{\infty}$ in $\mathbb{R}$, it holds that the sequence $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy.

(c) Find an example of a set $U \subsetneq \mathbb{R}$ and a function $f : U \to \mathbb{R}$ which is continuous, but f does not satisfy the previous property. That is, give an example of a continuous function $f : U \to \mathbb{R}$ such that there is some Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ in U where $\{f(x_n)\}_{n=1}^{\infty}$ is not Cauchy.

(5) Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x & x \in \mathbb{Q} \\ -x & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

(a) Using Definition 4.0.1, prove that f is continous at 0, but not continuous at any other point.

(b) Using Theorem 4.1.1, prove that f is continuous at 0, but not continuous at any other point.

(c) Using Theorem 4.2.1, prove that f is continuous at 0, but not continuous at any other point.

(6) Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 & x \geq 0 \\ -x^2 & x < 0 \end{cases}$$

(a) Using Definition 4.0.1, prove that f is continous.

(b) Using Theorem 4.1.1, prove that f is continuous.

(c) Using Theorem 4.2.1, prove that f is continuous.

(7) Let $U \subset \mathbb{R}$ and suppose that $f : U \to \mathbb{R}$ is continuous. Prove that if $\{V_i\}_{i \in I}$ is an open cover of $f(U)$, then $\{f^{-1}(V_i)\}_{i \in I}$ is a relatively open cover of U (relative to U).

(8) Let $U \subset \mathbb{R}$ be a neighborhood of some point x_0 and let $f : U \to \mathbb{R}$ be a function. Prove that f is continuous at x_0 if and only if for all $\varepsilon > 0$ there exists some $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subset U$ and for all $x \in (x_0 - \delta, x_0 + \delta)$, it holds that $f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$.

(9) Prove the "other direction" of Theorem 4.2.1 (1). That is, prove that if a function $f : U \to \mathbb{R}$ (where U is a neighborhood of x_0) has the property that for all neighborhoods V of $f(x_0)$ it holds that $f^{-1}(V)$ is a neighborhood of x_0, then $f : U \to \mathbb{R}$ is continuous at x_0. *Hint*: To show continuity at x_0, let $\varepsilon > 0$ be arbitrary and set $V = (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$; from here let $\delta > 0$ be such that $(x_0 - \delta, x_0 + \delta) \subset f^{-1}(V)$.

(10) Let $U \subset \mathbb{R}$. Prove that U is open if and only if U is a neighborhood for each of its elements.

(11) Suppose that $U \subset \mathbb{R}$ is open and $f : U \to \mathbb{R}$ be a function.

(a) Suppose that for some $d > 0$, f has the property that for every $x_0 \in U$ there is some $k > 0$ (which might depend on x_0 and d) such that $|f(x) - f(x_0)| \leq k|x - x_0|$ whenever $|x - x_0| < d$. Prove that f is continuous.

(b) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3$. Prove that f is continuous. Could have you used part (a) here to interpret your solution? *Hint*: Don't forget that $x^3 - x_0^3 = (x - x_0)(x^2 + xx_0 + x_0^2)$; in particular, if $|x - x_0| < 1$ (here 1 is just as good as any other positive choice), then justify that

$$|x^3 - x_0^3| \leq \left((x_0 + 1)^2 + |(x_0 + 1)x_0| + x_0^2\right) |x - x_0|.$$

(c) Let $f : [0, \infty) \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$. Prove that f is continuous on $(0, \infty)$, and continuous from the right at 0. Could you interpret your solution through the setup of part (a) to show that f is continuous on $(0, \infty)$?

(12) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{otherwise} \end{cases}$$

Prove that f is continuous at 0, but not continuous at any other point.

(13) Let $U \subset \mathbb{R}$ be open, let $f : U \to \mathbb{R}$, and let $x_0 \in U$. Prove that f is continuous at x_0 if and only if f is continuous from the right and from the left at x_0.

(14) Let $U \subset \mathbb{R}$ be open and $f : U \to \mathbb{R}$ is continuous. Also, suppose that $K \subset U$.

(a) Suppose that $\{U_i\}_{i \in I}$ is an open cover of $f(K)$. Prove that $\{f^{-1}(U_i)\}_{i \in I}$ is an open cover of K.

(b) Continuing with the notation from the previous part, suppose that $\{f^{-1}(U_{i_k})\}_{k=1}^{N}$ is a subcover of K. Prove that $\{U_{i_k}\}_{k=1}^{N}$ is a finite cover of $f(K)$.

(15) Rewrite Definition 4.0.1 with only the distance function dist; no absolute values should be involved.

(16) A more general (topological) definition of continuity goes as follows. Let $S \subset \mathbb{R}$ and let $f : S \to \mathbb{R}$ be a function. We say that f is *continuous in the relative topology* on S when given any open subset $V \subset R$, the inverse image $f^{-1}(V)$ is relatively open in S. Prove that $f : S \to \mathbb{R}$ is continuous in the relative topology on S if and only if for all $x_0 \in S$ there exists some $\delta > 0$ such that if $x \in S$ and $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$.

(17) Prove Proposition **??**.

(18) Prove (**??**) of Theorem **??**.

(19) Write out the detailed proof of Theorem **??**.

(20) Let $f : \mathbb{R} \to \mathbb{R}$ be a function.

(a) Prove that f is continuous if and only if for every Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathbb{R}$ it holds that $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy.

(b) Part a is not generally true if we pick a proper open subset $U \subsetneq \mathbb{R}$ as the domain of f. Suppose that the domain of f is $U = (0, 1)$ and define $f(x) = 1/x$. We haven't proved that $f(x) = 1/x$ is continuous yet, but we will, so you can use

this fact here. Show that there exists a Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ in U such that $\{f(x_n)\}_{n=1}^{\infty}$ is not Cauchy.

(21) Let $U = (-1, 0) \cup (0, 2) \cup (2, 4)$ and $f : U \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x + 1 & -1 < x < 0 \\ x^2 & 0 < x < 2 \\ 4/(x-1) & 2 < x < 3 \\ x & 3 \leq x < 4 \end{cases}$$

For the given x_0 below, find the left- and righthand limits of f at x_0 if they exist, and prove that these are, in fact, the limits in two ways: i) By the definition of left- or righthand limit, ii) using Theorem **??**.

(a) $x_0 = -1$.

(b) $x_0 = 0$.

(c) $x_0 = 2$.

(d) $x_0 = 3$. Also, prove that f is continuous from the right at $x_0 = 3$, but not continuous from the left.

(22) Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ such that for all points $x_0 \in \mathbb{R}$ it holds that neither $\lim_{x \to x_0-} f(x)$ nor $\lim_{x \to x_0+} f(x)$ exist. Make sure to prove your function truly meets this criteria.

(23) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \frac{x^2-9}{x-3} & x \neq 3 \\ 6 & x = 3 \end{cases}$$

(a) Prove that f is continuous at 3.

(b) From part (a), we know that f is continuous at 3, and so we can find some $\delta > 0$ such that if $|x - 3| < \delta$ it holds that $|f(x) - 6| < .01$. Find such a δ.

(24) Prove that if U is an open subset of $\mathbb{R}$, then $U \cap \text{bdry}(U) = \varnothing$; however, $U \subset \text{limpts}(U)$.

(25) In this exercise, you will fill in the missing details of Proposition 4.4.4. Continuing the notation from this proposition, U_1 and U_2 are non-empty disjoint open subsets of $\mathbb{R}$ such that $a \in U_1 \cap (x, y)$, $b \in U_2 \cap (x, y)$, and $a < b$. We defined $T =$

$\{s \in (a,b) \text{ s.t. } (a,s) \subset U_1 \cap (x,y)\}$ and showed that T was non-empty and bounded above by b; in particular, $c = \sup(T)$ is such that $c \leq b$.

(a) Prove that if $c \in U_1 \cap (x,y)$ and $\varepsilon > 0$ such that $(c-\varepsilon, c+\varepsilon) \subset U_1 \cap (x,y)$, then $c + \varepsilon \in T$.

(b) Prove that if $c \in U_2 \cap (x,y)$ and $\varepsilon > 0$ is such that $(c-\varepsilon, c+\varepsilon) \subset U_2 \cap (x,y)$, then $c - \varepsilon/2 \in T$.

(26) In this exercise, you will fill in the missing details of Theorem 4.4.6. We will use the notation of Theorem and its proof above.

(a) Suppose that U is a neighborhood of some element z. Prove that if f is continuous at z and $f(z) < c$ then there exists some $\delta > 0$ such that for all $x \in (z-\delta, z+\delta)$, it holds that $f(x) < c$.

(b) Suppose that U is a neighborhood of some element z. Prove that if f is continuous at z and $f(z) > c$ then there exists some $\delta > 0$ such that for all $x \in (z-\delta, z+\delta)$, it holds that $f(x) > c$.

(c) Prove that if $U_1 \cap (x,y) \neq \varnothing$ and $U_2 \cap (x,y) \neq \varnothing$ using the previous two parts to fill in the missing details of this statement in the proof given.

(d) Prove that for any $x \in (x_1, x_2) \setminus (U_1 \cup U_2)$, it holds that $f(x) = c$.

(e) At the beginning of the proof, we claimed that it is sufficient to consider the case that J is the domain of f, so that $f : J \to \mathbb{R}$ is continuous. Prove that this is true. In other words, we proved the intermediate value theorem for all continuous functions whose domains are intervals; as a quick corollary to this, prove the more general statement of the theorem (i.e., the domain of f is U not J).

(27) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x^5 + x^2 + 2x + 1$. Prove that f has a root in the interval $(-1, -1/2)$. That is, prove that there is some $x_0 \in (-1, -1/2)$ such that $f(x_0) = 0$.

(28) Suppose that U_1 and U_2 are both open and contain some common point x_0 (i.e., $x_0 \in U_1 \cap U_2$). If $f : U_1 \to \mathbb{R}$ and $g : U_2 \to \mathbb{R}$ are both continuous and $f(x_0) > g(x_0)$, prove that there is some $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \in U_1 \cap U_2$ and for all $x \in (x_0 - \delta, x_0 + \delta)$ it holds that $f(x) > g(x)$.

(29) Suppose that U_1 and U_2 are both open and contain some common interval $[a,b]$ (i.e., $[a,b] \in U_1 \cap U_2$). Suppose further that $f : U_1 \to \mathbb{R}$ and $g : U_2 \to \mathbb{R}$ are continuous

with $f(a) < g(a)$ and $f(b) > g(b)$. Prove the following: $h : [a, b] \to \mathbb{R}$ given by $h(x) = \frac{1}{f(x)-g(x)}$ is not a well-defined function, but if we restrict the domain to $V \subset [a, b]$ where it is well-defined, then it is continuous with respect to the relative topology on V.

(30) Suppose that U is open and $K \subset U$ is compact. Let $\alpha, \beta \in \mathbb{R}$. If $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}$ are continuous. Let $\{x_n\}_{n=1}^{\infty}$ be any sequence in K and define $\{y_n\}_{n=1}^{\infty}$ as the sequence $y_n = \alpha f(x_n) + \beta g(x_n)$.

(a) Prove that $\{y_n\}_{n=1}^{\infty}$ has a convergent subsequence.

(b) Provide an example where the sequence $\{y_n\}_{n=1}^{\infty}$ does not converge.

(31) For this problem, you can assume the normal facts about $\sin : \mathbb{R} \to \mathbb{R}$; in particular, that sin is continuous, that $\sin(x) = \sin(x + 2\pi k)$ for every $x \in \mathbb{R}$ and $k \in \mathbb{Z}$, and that $\sin(\pi/2) = -\sin(-\pi/2) = 1$. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

(a) Prove that f is continuous for all points $x \neq 0$.

(b) Prove that f is not continuous at $x = 0$ in two ways: i) By the (negation of the) definition of continuity, ii) Using Theorem **??**.

(32) Let $U \subset \mathbb{R}$ be open.

(a) For all $n \in \mathbb{N}$, prove that the function $f_n : U \to \mathbb{R}$ defined by $f_n(x) = x^n$ is continuous.

(b) Prove that any polynomial $p : U \to \mathbb{R}$ is continuous. Recall that a *polynomial* is a function of the form $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ where $a_0, ..., a_n \in \mathbb{R}$, and $n \in \mathbb{N}$.

(33) Let $K \subset \mathbb{R}$ be a set with no limit points. Let $f : K \to \mathbb{R}$ be a function.

(a) Prove that $f : K \to \mathbb{R}$ is continuous with respect to the relative topology on K.

(b) Prove that any real-valued function (i.e., any function with codomain $\mathbb{R}$) whose domain is a subset of $\mathbb{Z}$ is continuous with respect to the relative topology on the domain.

(34) Suppose that $a, b \in \mathbb{R}$ and $U \subset \mathbb{R}$ is an open set containing the interval $[a, b] \subset U$. Suppose further that $f : U \to \mathbb{R}$ is a continuous function such that $f(a) < 0$ and $f(b) > 0$. We define the sequence $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, and $\{c_n\}_{n=1}^{\infty}$ in the following

way. Let $a_1 = a$, $b_1 = b$, and c_1 be the midpoint $c_1 = (a_1 + b_1)/2$. If $f(c_1) \leq 0$, set $a_2 = c_1$ and $b_2 = b_1$, otherwise if $f(c_1) > 0$, set $a_2 = a_1$ and $b_2 = c_1$. In this way, on this iteration we have created a new interval $[a_2, b_2] \subset [a_1, b_1]$ with half the length and such that $f(a_2) \leq 0$ and $f(b_2) > 0$. Continuing in this fashion, we let $c_2 = (a_2 + b_2)/2$. If $f(c_2) \leq 0$ we set $a_3 = c_2$ and $b_3 = b_2$, otherwise if $f(c_2) > 0$ then we set $a_3 = a_2$ and $b_3 = c_2$. We continue this process to get our sequences $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, and $\{c_n\}_{n=1}^{\infty}$; note that, as before, $\cdots \subset [a_n, b_n] \subset \cdots \subset [a_2, b_2] \subset [a_1, b_1]$ and that for any $n \in \mathbb{N}$, $f(a_n) \leq 0$ and $f(b_n) > 0$.

(a) Prove that the sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ converge to the same point x_0.

(b) Prove that $f(x_0) = 0$.

(c) Suppose that $a = 0$ and $b = 1$. You are a programer who wants to write a program to find a root x_0 of f on this interval $[0, 1]$ (where as before $f(0) < 0$ and $f(1) > 0$ and f is continuous). So, you write a program which starts building the sequences $a_1, a_2, a_3, ...$; $b_1, b_2, b_3...$; $c_1, c_2, c_3....$ Of course, you can't really iterate infinitely many times since this would take infinite time, so instead you have to approximate the root x_0 of f. How many iterations n should you run so that c_n is within 2^{-10} of the actually root x_0?

(35) Let $U = \mathbb{R} \setminus \{0\}$ and let $f : U \to \mathbb{R}$ be defined by

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

Prove that f is continuous, but not uniformly continuous.

(36) Let $b > 0$ and defined the function $f_b : (0, b) \to \mathbb{R}$ by $f_b(x) = x^2$; also define $f : (0, \infty) \to \mathbb{R}$ by $f(x) = x^2$. Prove that f_b is uniformly continuous, but that f is not.

(37) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous. For $a, b \in \mathbb{R}$, define the function $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = f(ax + b)$. Prove that g is uniformly continuous.

(38) Let $U \subset \mathbb{R}$ and suppose that $f : U \to \mathbb{R}$ is a function.

(a) If $f : U \to \mathbb{R}$ is uniformly continuous, prove that for any $V \subset U$, the restriction $f : V \to \mathbb{R}$ is uniformly continuous.

(b) If there exists some $V \subset U$ such that the restriction $f : V \to \mathbb{R}$ is not uniformly continuous, prove that $f : U \to \mathbb{R}$ is not uniformly continuous.

(39) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous.

(a) Suppose that $U \subset \mathbb{R}$ is bounded. Prove that the restriction $f : U \to \mathbb{R}$ is uniformly continuous. *Hint*: If U is bounded, then for some $R > 0$, $U \subset [-R, R]$, so if you can prove that $f : [-R, R] \to \mathbb{R}$ is uniformly continuous, then you can use the Exercise 9 (34).

(b) Find an example of a function $f : (0, \infty) \to \mathbb{R}$ which is continuous, but where there exists a bounded subset $U \subset (0, \infty)$ such that $f : U \to \mathbb{R}$ is not uniformly continuous.

(40) In this problem, you will fill in the details of the proof of Proposition 4.6.10. Let $f : U \to \mathbb{R}$, $C = \{c \geq 0 \text{ s.t. } |f(x) - f(y)| \leq c|x - y|\}$, and $k = \inf C$, be as in the Proposition.

(a) Suppose that $b \geq 0$. Prove that if there exist $x, y \in U$ such that $|f(x) - f(y)| \geq b|x - y|$ if and only if b is a lower bound of C.

(b) Prove that if $x, y \in U$ with $|f(x) - f(y)| > k|x - y|$ there exists some $n \in \mathbb{N}$ such that $|f(x) - f(y)| > (k + 1/n)|x - y|$.

(41) Let $f : [0, \infty) \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$. Is f uniformly continuous? Prove it.

(42) In Exercise 9 (20)b, we found an example of a domain $U \subset \mathbb{R}$ and a function $f : U \to \mathbb{R}$ where f is continuous, but there exists a Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ in U such that $\{f(x_n)\}_{n=1}^{\infty}$ is not a Cauchy sequence. The issue was that although $\{x_n\}_{n=1}^{\infty}$ was a sequence in U which converged, it converged to a point $x_0 \notin U$, and thus does not contradict Theorem **??**. You can imagine that it would be nice to know if a function $f : U \to \mathbb{R}$ has the property that given any Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ in U, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy. As just mentioned, it is not enough just to assume that f is continuous; however, it is enough to assume that f is uniformly continuous.

(a) Prove that if $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$ is uniformly continuous, then for all Cauchy sequences $\{x_n\}_{n=1}^{\infty}$ in U, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy.

(b) Suppose that $U \subsetneq \mathbb{R}$ and $x_0 \notin U$. Suppose further that $\{x_n\}_{n=1}^{\infty}$ is a sequence in U converging to x_0. Prove that if a function $f : U \to \mathbb{R}$ is such that the sequence $\{f(x_n)\}_{n=1}^{\infty}$ does not converge, then f is not uniformly continuous.

(c) Use the previous part to give a quick proof of the fact that the function $f : (0, 1) \to \mathbb{R}$ defined by $f(x) = 1/x$ is not uniformly continuous.

(43) Prove Theorem 4.6.12. That is, if $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$ is Lipschitz continuous, then it is uniformly continuous.

(44) Let $U \subset \mathbb{R}$ and suppose that $f : U \to \mathbb{R}$ is a function.

(a) If $f : U \to \mathbb{R}$ is Lipschitz continuous, prove that for any $V \subset U$, the restriction $f : V \to \mathbb{R}$ is Lipschitz continuous.

(b) If there exists some $V \subset U$ such that the restriction $f : V \to \mathbb{R}$ is not Lipschitz continuous, prove that $f : U \to \mathbb{R}$ is not Lipschitz continuous.

(45) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \frac{1}{x^2+1}$. Prove that f is Lipschitz and that the Lipschitz constant k of f is such that $1/2 \leq k < 2$. *Hint*: For any $x, y \in \mathbb{R}$, we have $\frac{|x|}{(x^2+1)(y^2+1)} \leq \frac{|x|}{x^2+1} < 1$.

(46) Suppose that $U \subset \mathbb{R}$ is bounded.

(a) If $\{x_n\}_{n=1}^{\infty}$ is any sequence in U, prove that it has a convergent subsequence.

(b) Suppose that $f : U \to \mathbb{R}$ is not bounded above; that is, for any $R > 0$ there exists some $x \in U$ such that $f(x) > R$. Prove that there exists sequence $\{x_n\}_{n=1}^{\infty}$ in U which converges to some point $x_0 \in \mathbb{R}$ such that $\lim_{n\to\infty} f(x_n) = +\infty$. Why do we know that $x_0 \notin U$?

8. Other Problems

(1) Suppose that $\{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty}$, and $\{c_n\}_{n=1}^{\infty}$ are sequences in $\mathbb{R}$.

(a) (Generalizing the Squeezing Principle) Suppose that there exists some $N \in \mathbb{N}$ such that for all $n \geq N$, $a_n \leq b_n \leq c_n$. If $\{a_n\}_{n=1}^{\infty}$ and $\{c_n\}_{n=1}^{\infty}$ both converge to the same limit l, then $\{b_n\}_{n=1}^{\infty}$ converges and $\lim_{n\to\infty} b_n = l$.

(b) If there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $a_n \leq b_n$. Prove that if $\lim_{n\to\infty} a_n = +\infty$, then $\lim_{n\to\infty} b_n = +\infty$.

(c) If there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $b_n \leq c_n$. Prove that if $\lim_{n\to\infty} c_n = -\infty$, then $\lim_{n\to\infty} b_n = -\infty$.

(2) (Generalized "Saturating Subsequences") Let $\mathbf{n}^1, ..., \mathbf{n}^m$ be strictly increasing functions $\mathbb{N} \to \mathbb{N}$. We say that these functions *eventually saturate* $\mathbb{N}$ when there exists some $N \in \mathbb{N}$ such that $\{n \in \mathbb{N} \text{ s.t. } n \geq N\} = \{N, N+1, N+2, ...\} \subset \mathbf{n}^1(\mathbb{N}) \cup \cdots \cup \mathbf{n}^m(\mathbb{N})$. Note that the case $N = 1$ is exactly the definition of these functions saturating $\mathbb{N}$. Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ and $x \in \mathbb{R}$. Prove that

$\lim_{n\to\infty} x_n = x$ if and only if there exists $\mathbf{n}^1, ..., \mathbf{n}^m$ which eventually saturate $\mathbb{N}$ such that the subsequences $\{x_{\mathbf{n}^1(k)}\}_{k=1}^{\infty}, ..., \{x_{\mathbf{n}^m(k)}\}_{k=1}^{\infty}$ all converge to x.

(3) Prove that a sequence in $\mathbb{R}$ diverges to $+\infty$ if and only every subsequence diverges to $+\infty$. While you're at it, prove the analogous statement for sequences diverging to $-\infty$.

(4) Prove that any subset of $\mathbb{R}$ containing only finitely many points is compact by using the definition of compactness.

CHAPTER 5

Differential Calculus, Riemann Integration, and Function Spaces

1. Differential Calculus

1.1. Basics.

DEFINITION 5.1.1. Let $U \subset \mathbb{R}$ be a neighborhood of some point x_0. Let $f : U \to \mathbb{R}$ be a function such that the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists. Then we say that f is *differentiable at* x_0 and define $f'(x_0)$ as this limit; i.e.,

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

If U is open and f is differentiable at all points $x \in U$, then we say that f is *differentiable on* U and let the symbol f' represent the function $f' : U \to \mathbb{R}$ that for each $x \in U$ gives the value $f'(x)$.

REMARK 5.1.2. We have defined what it means for a function $g : U \to \mathbb{R}$ to have a limit at x_0. So, when we write

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

what this means is that we are taking the limit of the function $g : U \to \mathbb{R}$ defined by

$$g(x) = \frac{f(x) - f(x_0)}{x - x_0}$$

at x_0. □

PROPOSITION 5.1.3. *Let $U \subset \mathbb{R}$ be a neighborhood of x_0 and let $\delta > 0$ be such that $(x_0 - \delta, x_0 + \delta) \subset U$. Let $f : U \to \mathbb{R}$ be a function and $\lambda \in \mathbb{R}$. The following are equivalent:*

(1) *f is differentiable at x_0 and $f'(x_0) = \lambda$.*

(2) *Let* $g : (-\delta, \delta) \setminus \{0\} \to \mathbb{R}$ *be defined by*

$$g(h) = \frac{f(x_0 + h) - f(x_0)}{h}$$

Then $\lim_{h\to 0} g(h) = \lambda$. *Written more instructively,*

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lambda.$$

(3) *Given any sequence* $\{x_n\}_{n=1}^{\infty}$ *of elements from* $U \setminus \{x_0\}$ *converging to* x_0, *it holds that*

$$\lim_{n\to\infty} \frac{f(x_n) - f(x_0)}{x_n - x_0} = \lambda.$$

(4) *If* $\{h_n\}_{n=1}^{\infty}$ *is any sequence in* $(-\delta, \delta) \setminus \{0\}$ *converging to* 0, *then*

$$\lim_{n\to\infty} \frac{f(x_0 + h_n) - f(x_0)}{h_n} = \lambda.$$

Proof. (1) $\iff$ (3) and (2) $\iff$ (4): Immediately from Theorem **??**.

(3) $\iff$ (4): Exercise 6 (1). ■

THEOREM 5.1.4. *Let* U *be a neighborhood of some point* $x_0 \in \mathbb{R}$. *Suppose that* $f : U \to \mathbb{R}$ *is differentiable at* x_0. *Then* f *is continuous at* x_0.

Proof. Since

$$\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$

we can find some $d > 0$ such that for all $x \in (x_0 - d, x_0) \cup (x_0, x_0 + d)$, we have

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < 1.$$

By a standard application of the reverse triangle inequality, this implies for any $x \in (x_0 - d, x_0) \cup (x_0, x_0 + d)$,

$$\left| \frac{f(x) - f(x_0)}{x - x_0} \right| < |f'(x_0)| + 1 \implies |f(x) - f(x_0)| \leq (|f'(x_0)| + 1)|x - x_0|.$$

Therefore, setting $\delta = \min\left(d, \frac{\varepsilon}{|f'(x_0)|+1}\right)$ we have that if $x \in (x_0 - \delta, x_0 + \delta)$ then

$$|f(x) - f(x_0)| < (|f'(x_0)| + 1)|x - x_0| \leq (|f'(x_0)| + 1)\frac{\varepsilon}{|f'(x_0)| + 1} = \varepsilon$$

from which we deduce that f is continuous at x_0. ■

THEOREM 5.1.5. *Let U_1 and U_2 be neighborhoods of some point $x_0 \in \mathbb{R}$. Suppose $f : U_1 \to \mathbb{R}$ and $g : U_2 \to \mathbb{R}$ are both differentiable at x_0. Let $U = U_1 \cap U_2$. Then the following hold.*

(1) *For any $\alpha, \beta \in \mathbb{R}$, it holds that $\alpha f + \beta g : U \to \mathbb{R}$ is differentiable at x_0 and*

$$(\alpha f + \beta g)'(x_0) = \alpha f'(x_0) + \beta g'(x_0).$$

(2) *(Product Rule) $fg : U \to \mathbb{R}$ is differentiable at x_0 and*

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0).$$

(3) *(Quotient Rule) If $g(x_0) \neq 0$, then there is some $\delta > 0$ such that $f/g : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ is differentiable at x_0 and*

$$(f/g)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$$

Proof. (1) Exercise 6 (2)

(2) For any $x \in U \setminus \{x_0\}$ we have

$$\begin{aligned}\frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} &= \frac{f(x)g(x) - f(x_0)g(x) + f(x_0)g(x) - f(x_0)g(x_0)}{x - x_0} \\ &= g(x)\frac{f(x) - f(x_0)}{x - x_0} + f(x_0)\frac{g(x) - g(x_0)}{x - x_0}.\end{aligned}$$

Now,

$$\begin{aligned}\lim_{x\to x_0} \frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} &= \lim_{x\to x_0} \left[g(x)\frac{f(x) - f(x_0)}{x - x_0} + f(x_0)\frac{g(x) - g(x_0)}{x - x_0}\right] \\ &= \left[\lim_{x\to x_0} g(x)\right]\left[\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}\right] + f(x_0)\left[\lim_{x\to x_0} \frac{g(x) - g(x_0)}{x - x_0}\right] \\ &= g(x_0)f'(x_0) + f(x_0)g'(x_0).\end{aligned}$$

Note that the last equality held since all limits in the bracketed terms exist; in particular, the derivatives exist by assumption and $\lim_{x\to x_0} g(x) = g(x_0)$ since, by Theorem 5.1.4, g is continuous at x_0.

(3) Suppose $g(x_0) \neq 0$. By Lemma 4.5.2, there is some $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subset U$ and for all $x \in (x_0 - \delta, x_0 + \delta)$ it holds that $|g(x)| \neq 0$. It suffices to consider the function $1/g : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$, since then the general case can be proved easily by an application

of the product rule. For any $x \in (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$ we have

$$\frac{1/g(x) - 1/g(x_0)}{x - x_0} = \frac{g(x_0) - g(x)}{g(x)g(x_0)(x - x_0)} = -\left[\frac{1}{g(x)g(x_0)}\right]\left[\frac{g(x) - g(x_0)}{x - x_0}\right]$$

As before, applying the limit to both sides of the equality results in the desired conclusion. ■

LEMMA 5.1.6. *Suppose that $U \subset \mathbb{R}$ is some neighborhood of a point $x_0 \in \mathbb{R}$. Suppose that $f : U \to \mathbb{R}$ is differentiable at x_0 and there exists a sequence $\{x_n\}_{n=1}^{\infty}$ in $U \setminus \{x_0\}$ such that $\lim_{n\to\infty} x_n = x_0$ and $f(x_n) = f(x_0)$ for infinitely many n; i.e., the set $T = \{n \in \mathbb{N} \text{ s.t. } f(x_n) = f(x_0)\}$ has infinitely many elements. Then $f'(x_0) = 0$.*

Proof. We give an outline here and leave some details as an exercise. Note that if $f(x_n) = f(x_0)$ for infinitely many n, then there is a subsequence $\{x_{n_k}\}_{k=1}^{n}$ such that $f(x_{n_k}) = f(x_0)$ for every $k \in \mathbb{N}$, and therefore $\lim_{k\to\infty} f(x_{n_k}) = 0$ which implies that $f(x_0) = 0$ since f is continuous at x_0, and from here,

$$\lim_{k\to\infty} \frac{f(x_{n_k}) - f(x_0)}{x_{n_k} - x_0} = \lim_{k\to\infty} \frac{f(x_0) - f(x_0)}{x_{n_k} - x_0} = 0.$$

Since the above limit must give the value of $f'(x_0)$ because f is assumed to be differentiable at x_0, we conclude that $f'(x_0) = 0$. ■

THEOREM 5.1.7 (Chain Rule). *Let $U_1 \subset \mathbb{R}$ be a neighborhood of some point x_0 and suppose $f : U_1 \to \mathbb{R}$ is differentiable at x_0. Let $U_2 \subset \mathbb{R}$ be a neighborhood of $f(x_0)$ and suppose that $g : U_2 \to \mathbb{R}$ is differentiable at $f(x_0)$. Define $S = f^{-1}(U_2)$. The function $g \circ f : S \to \mathbb{R}$ is differentiable at x_0 and*

$$(g \circ f)'(x_0) = g'(f(x_0)) \cdot f'(x_0).$$

Proof. Since $f : U_1 \to \mathbb{R}$ is continuous at x_0 by Theorem 5.1.4 and U_2 is a neighborhood of $f(x_0)$, we know that $S = f^{-1}(U_2)$ is a neighborhood of x_0 (by Theorem **??**), so it makes sense to consider whether or not $g \circ f : S \to \mathbb{R}$ is differentiable at x_0. Let $\{x_n\}_{n=1}^{\infty}$ be any sequence in $S \setminus \{x_0\}$ which converges to x_0. Define the function $h : S \to \mathbb{R}$ by

$$h(x) = \frac{g(f(x)) - g(f(x_0))}{x - x_0}.$$

We need to show that $\lim_{n\to\infty} h(x_n) = g'(f(x_0)) \cdot f'(x_0)$.

To do so, let's first partition the sequence $\{x_n\}_{n=1}^{\infty}$ into two subsequences: $\{y_n\}_{n=1}^{N_1}$ containing all elements of the sequence such that $f(y_n) \neq f(x_0)$, and $\{z_n\}_{n=1}^{N_2}$ containing all the elements of the sequence where $f(z_n) = f(x_0)$. Note that it can happen that $N_1 < \infty$ or $N_2 < \infty$, but not both can be finite. By definition, it is clear that for all $1 \leq n \leq N_2$, it holds that $h(z_n) = 0$. On the other hand, since $f(y_n) - f(x_0) \neq 0$, we also have

$$h(y_n) = \frac{g(f(y_n)) - g(f(x_0))}{f(y_n) - f(x_0)} \cdot \frac{f(y_n) - f(x_0)}{y_n - x_0}$$

Suppose $N_1 = \infty$. Since f is continuous at x_0, the sequence $\{f(y_n)\}_{n=1}^{\infty}$ is in $U_2 \setminus \{f(x_0)\}$ converging to $f(x_0)$ and therefore, since g is assumed differentiable at x_0

$$\lim_{n\to\infty} \frac{g(f(y_n)) - g(f(x_0))}{f(y_n) - f(x_0)} = g'(f(x_0))$$

and since f is assumed differentiable at x_0 and $\{y_n\}_{n=1}^{\infty}$ is a sequence in $U_1 \setminus x_0$ converging to x_0,

$$\lim_{n\to\infty} \frac{f(y_n) - f(x_0)}{x_n - x_0} = f'(x_0).$$

This shows that $\lim_{n\to\infty} h(y_n) = g'(f(x_0)) \cdot f'(x_0)$.

Suppose that $N_2 = \infty$. Then by Lemma 5.1.6 $f'(x_0) = 0$. Therefore $\lim_{n\to\infty} h(z_n) = 0 = g'(f(x_0)) \cdot 0 = g'(f(x_0))f'(x_0)$.

We see now that whenever $\{h(y_n)\}_{n=1}^{N_1}$ or $\{h(z_n)\}_{n=1}^{N_2}$ are subsequences of $\{h(x_n)\}_{n=1}^{\infty}$, they converge to $g'(f(x_0))f'(x_0)$. Since these subsequences partition $\{h(x_n)\}_{n=1}^{\infty}$, we conclude that $\lim_{n\to\infty} h(x_n) = g'(f(x_0))f'(x_0)$. ■

DEFINITION 5.1.8. Suppose that $U \subset \mathbb{R}$ is a neighborhood of some point x_0, and $f : U \to \mathbb{R}$ is a function. We say that f has a *local maximum* at x_0 when there exists some $\delta > 0$ such that for all $x \in U$ with $|x - x_0| < \delta$ it holds that $f(x) \leq f(x_0)$. If it happens that for all $x \in U$, $f(x) \leq f(x_0)$ then we say that f has a *global maximum* at x_0. Conversely, we say that f has a *local minimum* at x_0 when there exists some $\delta > 0$ such that for all $x \in U$ with $|x - x_0| < \delta$ it holds that $f(x) \geq f(x_0)$. If it happens that for all $x \in U$, $f(x) \geq f(x_0)$ then we say that f has a *global minimum* at x_0. If f has either a local maximum or minimum at x_0, we say that f has an *extremum* at x_0.

DEFINITION 5.1.9. Suppose $U \subset \mathbb{R}$ is a neighborhood of some point x_0 and $f : U \to \mathbb{R}$ is differentiable at x_0. If $f'(x_0) = 0$, we say that x_0 is a *critical point* of f.

THEOREM 5.1.10 (Zero Derivative at Extrema). *Suppose that $U \subset \mathbb{R}$ is a neighborhood of some point x_0 and $f : U \to \mathbb{R}$ is differentiable at x_0. If f has an extremum at x_0, then x_0 is a critical point of f.*

Proof. Suppose first that f has a local maximum at x_0. Take $\delta > 0$ be such that $(x_0 - \delta, x_0 + \delta) \subset U$ and for all $x \in (x_0 - \delta, x_0 + \delta)$ we have $f(x) \leq f(x_0)$. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of elements in $(x_0, x_0 + \delta)$ converging to x_0. Then for every n, let y_n be defined by

$$y_n = \frac{f(x_n) - f(x_0)}{x_n - x_0}$$

Note that $y_n \leq 0$ since the numerator is ≤ 0 and the denominator is positive. On the other hand, let $\{\tilde{x}_n\}_{n=1}^{\infty}$ be a sequence of elements in $(x_0 - \delta, x_0)$ converging to x_0. Then for every n, let $\tilde{y}_n$ be defined by

$$\tilde{y}_n = \frac{f(\tilde{x}_n) - f(x_0)}{\tilde{x}_n - x_0}$$

Note that for all n, $\tilde{y}_n \geq 0$ since the numerator is ≤ 0 and the denominator is negative. By assumption of the differentiability of f, we have $\lim_{n\to\infty} y_n = f'(x_0)$ implying that $f'(x_0) \leq 0$ and $\lim_{n\to\infty} \tilde{y}_n \to f'(x_0)$ implying that $f'(x_0) \geq 0$. Hence $f'(x_0) = 0$. The argument for a local minimum is done analogously and left as an exercise. ■

NOTATION 5.1.11. Any function whose codomain is $\mathbb{R}$ is commonly referred to as *real-valued*. Moreover, it is common to neglect the entire domain of a function to focus on a subset of the domain which is in consideration. For example, we may say something such as: let f be a real-valued function defined on the set S. What this means is that the domain of f contains S as a subset, and the codomain of f is $\mathbb{R}$. This is most commonly done when what follows only concerns the restriction $f : S \to \mathbb{R}$ and we can ignore what the true larger domain of f truly is. Another example of such phrasing is: let f be a real-valued function which is differentiable on the open set $U \subset \mathbb{R}$. Again, this means that U is a subset of the true domain of f, the codomain of f is $\mathbb{R}$, and for all $x \in U$, $f'(x)$ exists.

COROLLARY 5.1.12. *Suppose that f is a real-valued function defined on some non-empty open interval $J \subset \mathbb{R}$ and suppose further $a, b \in J$ with $a < b$. If f is constant on $[a, b]$ then $f'(x) = 0$ for every $x \in [a, b]$.*

Proof. Exercise 6 (3). ■

THEOREM 5.1.13 (Rolle's Theorem). *Suppose that $J \subset \mathbb{R}$ is a non-empty open interval and f is a real-valued function which is differentiable on J. Let $a, b \in J$ with $a < b$ and suppose that $f(a) = f(b)$. Then there is some $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. If f is constant on $[a, b]$ then take any $c \in (a, b)$ and $f'(c) = 0$ by Corollary 5.1.12. So, let's assume that f is not constant on $[a, b]$. Since f is differentiable on J, it is continuous on J. Therefore, f is continuous on the compact subset $[a, b]$, meaning that we can find $c_1, c_2 \in [a, b]$ such that $f(c_1)$ gives a maximum value of $f([a, b])$ and $f(c_2)$ is the minimum value of $f([a, b])$. Since f is non-constant, then either $c_1 \in (a, b)$ or $c_2 \in (a, b)$. Let $c = c_1$ if $c_1 \in (a, b)$ or $c = c_2$ if $c_2 \in (a, b)$. Applying the Zero Derivative at Extrema Theorem 5.1.10, we're done. ■

THEOREM 5.1.14 (Mean Value Theorem). *Suppose that $J \subset \mathbb{R}$ is a non-empty open interval and f is a real-valued function which is differentiable on J. Let $a, b \in J$ with $a < b$. Then there is some $c \in (a, b)$ such that*

$$f'(c) \cdot (b - a) = f(b) - f(a).$$

Equivalently, this can be written as

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

Proof. Define the line

$$h(x) = \left(\frac{f(b) - f(a)}{b - a} \right) (x - a) + f(a)$$

for $x \in J$. Now, let $g : J \to \mathbb{R}$ be defined as $g(x) = f(x) - h(x)$. Then g is differentiable since both f and h are. Moreover $g(a) = g(b) = 0$. So, applying Rolle's theorem to g let's us find some $c \in (a, b)$ such that $g'(c) = 0$. But then,

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

which concludes the proof. ■

COROLLARY 5.1.15. *Suppose that f is a real-valued function which is differentiable on some open interval $J \subset \mathbb{R}$ containing the point x_0. Then for every $x \in J$, there exists some*

c between x and x_0 such that

$$f(x) = f(x_0) + f'(c) \cdot (x - x_0).$$

Proof. Exercise 6 (4) ■

THEOREM 5.1.16 (Some Consequences of the MVT). *Suppose that f is a real-valued function which is differentiable on some interval $(a, b) \subset \mathbb{R}$. The following hold.*

(1) *If for every $x \in (a, b)$ it holds that $f'(x) > 0$, then f is strictly increasing on (a, b).*
(2) *If for every $x \in (a, b)$ it holds that $f'(x) < 0$, then f is strictly decreasing on (a, b).*
(3) *If for every $x \in (a, b)$ it holds that $f'(x) = 0$, then f is constant on (a, b).*
(4) *Suppose that for some $k > 0$ it holds that $\sup\{|f'(x)| \text{ s.t. } x \in (a, b)\} \leq k$ (in words, this says that f' is bounded on (a, b)). Then f is Lipschitz continuous on (a, b) and $|f(x) - f(y)| \leq k|x - y|$ for every $x, y \in (a, b)$.*

Proof. (1) Let $x, y \in (a, b)$ with $x < y$. We need to prove that $f(x) < f(y)$. To that end, by the Mean Value Theorem, there is some $c \in (x, y)$ such that $f(y) = f(x) + f'(c)(y - x)$. By assumption $f'(c) > 0$ and $(y - x) > 0$, and therefore $f'(c)(y - x) > 0$. This implies that $f(x) + f'(c)(y - x) > f(x)$, proving that $f(y) > f(x)$.

(2) Exercise 6 (5).

(3) Let $x_0 \in (a, b)$ and define $\alpha = f(x_0)$. We will prove that for every $x \in (a, b)$, $f(x) = \alpha$. Indeed, let $x \in (a, b) \setminus \{x_0\}$ and using the Mean Value Theorem, let c land between x and x_0 such that $f(x) = f(x_0) + f'(c)(x - x_0)$. By assumption $f'(c) = 0$, and so $f(x) = f(x_0) = \alpha$. Since x was arbitrary, we conclude that $f(x) = \alpha$ for all $x \in (a, b)$.

(4) Let $k > 0$ be such that $\sup\{|f'(x)| \text{ s.t. } x \in (a, b)\} \leq k$. Let $x, y \in (a, b)$ be arbitrary, but not equal. Using the Mean Value Theorem, let c land between x and y such that $f(x) - f(y) = f'(c)(x - y)$. Applying absolute values to both sides of the previous equality and using the definition of k gives $|f(x) - f(y)| = |f'(c)|\,|x - y| \leq k|x - y|$. Since x and y were arbitrary, we conclude that f is Lipschitz on (a, b) and for all $x, y \in (a, b)$, $|f(x) - f(y)| \leq k|x - y|$. ■

1.2. L'Hopital's Rule.

LEMMA 5.1.17. *Suppose that $U \subset \mathbb{R}$ is a neighborhood of some point x_0 and $f : U \to \mathbb{R}$ is differentiable at x_0. If $f'(x_0) \neq 0$, then there exists some $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$ it hold that $f(x) \neq f'(x_0)$.*

Proof. Let that $\varepsilon = |f'(x_0)|$. Then $\varepsilon > 0$ by the assumption that $f'(x_0) \neq 0$. Since

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$

the definition of limits implies that there exists some $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$,

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < \varepsilon$$

which from the reverse triangle inequality implies

$$|f'(x_0)| - \varepsilon < \left| \frac{f(x) - f(x_0)}{x - x_0} \right|$$

By the definition of ε and by multiplying through by $|x - x_0|$, we have $0 < |f(x) - f(x_0)|$ implying that $f(x) \neq f(x_0)$ for any $x \in (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$. ∎

THEOREM 5.1.18 (L'Hopital's Rule - Version 1). *Suppose that f and g are real-valued functions on some common interval (a, b) containing the point $c \in (a, b)$. Suppose that $f(c) = g(c) = 0$ and also that $f'(c)$ and $g'(c)$ exist and that $g'(c) \neq 0$. Then*

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

Proof. By Lemma 5.1.17, if $g'(c) \neq 0$ then there is some $\delta > 0$ such that $(c - \delta, c + \delta) \subset (a, b)$ such that for every $x \in (c - \delta, c + \delta) \setminus \{c\}$, it holds that $g(x) \neq g(c)$. Now, since $f(c) = 0 = g(c)$ we have that for every $x \in (c - \delta, c + \delta) \setminus \{c\}$ that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(c)}{g(x) - g(c)} = \left(\frac{f(x) - f(c)}{x - c} \right) \left(\frac{x - c}{g(x) - g(c)} \right)$$

By assumption

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

and

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

But, we have proved that

$$\frac{g(x)-g(c)}{x-c} = g'(c) \neq 0 \implies \frac{x-c}{g(x)-g(c)} = \frac{1}{g'(c)}.$$

So, since the limit of a product is the product of the limits

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \Big[\lim_{x\to c}\frac{f(x)-f(c)}{x-c}\Big]\Big[\lim_{x\to c}\frac{x-c}{g(x)-g(c)}\Big] = \frac{f'(c)}{g'(c)}$$

which is what we wanted to prove. ■

THEOREM 5.1.19 (L'Hopital's Rule - Version 2). *Suppose that $f(x)$ and $g(x)$ are both defined and differentiable on an interval (a,b) and c is a limit point of (a,b); equivalently, $c \in [a,b]$. Suppose that $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ (where these limits are understood as one-sided when $c = a$ or $c = b$) and that for every $x \in (a,b)\backslash\{c\}$ it holds that $g(x) \neq 0$. If $\lim_{x\to c}\frac{f'(x)}{g'(x)} = \lambda$ exists, then*

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)} = \lambda.$$

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $(a,b)\setminus\{c\}$ converging to c. By Exercise 6 (13) (Cauchy's mean value theorem), there exists a sequence $\{\xi_n\}_{n=1}^{\infty}$ such that for every n, ξ_n lands between x_n and c and, moreover, that $\frac{f(x_n)}{g(x_n)} = \frac{f(x_n)-f(c)}{x_n-c}\cdot\frac{x_n-c}{g(x_n)-g(c)} = \frac{f'(\xi_n)}{g'(\xi_n)}$. Our choice of $\{\xi_n\}_{n=1}^{\infty}$ implies that for every n, $|\xi_n - c| < |x_n - c|$, proving that $\xi_n \to c$ as $n \to \infty$ (by Exercise 9 (1), for example). Hence

$$\lim_{n\to\infty}\frac{f(x_n)}{g(x_n)} = \lim_{n\to\infty}\frac{f'(\xi_n)}{g'(\xi_n)} = \lambda.$$

Since the sequence $\{x_n\}_{n=1}^{\infty}$ was arbitrarily chosen in $(a,b)\setminus\{c\}$ converging to c, this proves that

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)} = \lambda.$$

■

2. Riemann Integration

2.1. Partitions, Upper and Lower Sums.

DEFINITIONS 5.2.1. Let $a,b \in \mathbb{R}$ with $a < b$. A *partition* $\mathcal{P}$ of $[a,b]$ is a finite ordered set of distinct real numbers $\{x_0, x_1, ..., x_n\}$ such that

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

If $\mathcal{P}$ and $\mathcal{Q}$ are both partitions of $[a,b]$, then we say that $\mathcal{P}$ is a *refinement* of $\mathcal{Q}$ when $\mathcal{Q} \subset \mathcal{P}$. If $\mathcal{P}$ is a partition of $[a,b]$ then the *mesh* of $\mathcal{P}$ is defined to be the largest distance between any two consecutive elements of $\mathcal{P}$. That is, if $\mathcal{P} = \{x_0, x_1, ..., x_n\}$, then the mesh of $\mathcal{P}$ is $\max\{x_1 - x_0, x_2 - x_1, ..., x_n - x_{n-1}\}$. Another common way to refer to the mesh of $\mathcal{P}$ is as the *gap* of $\mathcal{P}$. We will denote this by $\text{mesh}(\mathcal{P})$. If $\mathcal{P}$ is a partition of $[a,b]$, then we say that $\mathcal{P}$ is *evenly spaced* when the distance between any consecutive elements in $\mathcal{P}$ is constant.

REMARK 5.2.2. By the definition of $\text{mesh}(\mathcal{P})$, you know that $0 < \text{mesh}(\mathcal{P}) \leq b-a$. When $\mathcal{P}$ is evenly spaced and $\mathcal{P} = \{x_0, x_1, ..., x_n\}$ then $x_k = \frac{k}{n}(b-a) + a$ for each k and moreover $\text{mesh}(\mathcal{P}) = \frac{b-a}{n}$. □

NOTATION 5.2.3. Suppose that $c, d \in \mathbb{R}$ with $a < b$ and $c < d$. If $\mathcal{P}$ is a partition of $[a,b]$ and $\mathcal{Q}$ is a partition of $[c,d]$ then we write $\mathcal{P} \cup \mathcal{Q}$ as the partition of $[\min(a,c), \max(b,d)]$ containing all the elements of $\mathcal{P}$ and $\mathcal{Q}$. (Note this is just a set-theoretic union from the perspective of ordered sets).

EXAMPLE 5.2.4. If $\mathcal{P} = \{0, 1, 2, 2.5, 2.7, 3\}$ is a partition of $[0,3]$ and $\mathcal{Q} = \{.25, 1, 1.5, e, \pi\}$ is the partition of $[.25, \pi]$ then

$$\mathcal{P} \cup \mathcal{Q} = \{0, .25, 1, 1.5, 2, 2.5, 2.7, e, 3, \pi\}$$

is a partition of $[0, \pi]$. □

DEFINITIONS 5.2.5. Let f be a bounded real-valued function defined on the interval $[a,b]$ and $\mathcal{P} = \{x_0, x_1, ..., x_n\}$ be a partition of $[a,b]$. For each i in the rangle $1 \leq i \leq n$, define $M_i^f = \sup\{f(x) \text{ s.t. } x \in [x_{i-1}, x_i]\}$ and $m_i^f = \inf\{f(x) \text{ s.t. } x \in [x_{i-1}, x_i]\}$. The *upper Riemann sum* of f on $[a,b]$ with respect to $\mathcal{P}$ is

$$U(f, \mathcal{P}) = \sum_{i=1}^{n} M_i^f \, (x_i - x_{i-1})$$

The *lower Riemann sum* of f on $[a,b]$ with respect to $\mathcal{P}$ is

$$L(f, \mathcal{P}) = \sum_{i=1}^{n} m_i^f \, (x_i - x_{i-1})$$

REMARK 5.2.6. In Exercise 6 (16) we will explore what happens to the upper or lower sums if we neglect the assumption that f is bounded on $[a,b]$. □

PROPOSITION 5.2.7. *Let f be a real-valued function on $[a,b]$ and $\mathcal{P},\mathcal{Q}$ partitions of $[a,b]$ such that $\mathcal{P}$ is a refinement of $\mathcal{Q}$. Then $U(f,\mathcal{Q}) \geq U(f,\mathcal{P})$ and $L(f,\mathcal{Q}) \leq L(f,\mathcal{P})$.*

Proof. Here we prove that $U(f,\mathcal{Q}) \geq U(f,\mathcal{P})$ whenever $\mathcal{P}$ is a refinement of $\mathcal{Q}$ and leave the lower sum inequality as an exercise. We will prove this by induction on $n = \#(\mathcal{P} \setminus \mathcal{Q})$. Let T be the subset of $\mathbb{N}$ where $n \in T$ whenever for any partition $\mathcal{Q}$ of $[a,b]$ and any refinement $\mathcal{P}$ of $\mathcal{Q}$ such that $n = \#(\mathcal{P} \setminus \mathcal{Q})$ it holds that $U(f,\mathcal{Q}) \geq U(f,\mathcal{P})$. We first show that $1 \in T$. Let's label the elements of $\mathcal{Q} = \{x_0, x_1, ..., x_m\}$ for some $m \in \mathbb{N}$. Let $\mathcal{P}$ be any refinement of $\mathcal{Q}$ containing exactly one more element that $\mathcal{Q}$; label this element y. Then, $\mathcal{P} = \{x_0, x_1, ..., x_{i-1}, y, x_i, ..., x_m\}$ for some $i \in \{1, 2, ..., m\}$. Now, notice that if $M_i^f = \sup\{f(x) \text{ s.t. } x \in [x_{i-1}, x_i]\}$ then

$$M_i^f(x_i - x_{i-1}) \geq N_1(x_i - y) + N_2(y - x_{i-1})$$

where we define $N_1 = \sup\{f(x) \text{ s.t. } x \in [y, x_i]\}$, and $N_2 = \sup\{f(x) \text{ s.t. } x \in [x_{i-1}, y]\}$. The reason these inequalities are true is because $N_1 \leq M_i$ and $N_2 \leq M_i$. But then

$$U(f,\mathcal{Q}) = \sum_{k=1}^{m} M_k^f(x_k - x_{k-1}) = \sum_{\substack{k=1 \\ k\neq i}}^{m} M_k^f(x_k - x_{k-1}) + M_i^f(x_i - x_{k-1})$$

$$\geq \sum_{\substack{k=1 \\ k\neq i}}^{m} M_k^f(x_k - x_{k-1}) + N_1(x_i - y) + N_2(y - x_{i-1}) = U(f,\mathcal{P}).$$

which concludes the proof for upper sums. The proof for lower sums is left as exercise. ∎

COROLLARY 5.2.8. *Let f be a real-valued function on $[a,b]$ and $\mathcal{P},\mathcal{Q}$ partitions of $[a,b]$. Then*

$$L(f,\mathcal{Q}) \leq U(f,\mathcal{P}).$$

Proof. Let $\mathcal{P}^*$ be the refinement of both $\mathcal{P}$ and $\mathcal{Q}$ given by their union $\mathcal{P}^* = \mathcal{P} \cup \mathcal{Q}$. From Proposition 5.2.7,

$$L(f,\mathcal{Q}) \leq L(f,\mathcal{P}^*) \leq U(f,\mathcal{P}^*) \leq U(f,\mathcal{P})$$

where the inequality $L(f,\mathcal{P}^*) \leq U(f,\mathcal{P}^*)$ is proved in Exercise 6 (17). ∎

2.2. Riemann Integral.

DEFINITION 5.2.9. Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that f is a bounded real-valued function defined on $[a, b]$. The *upper Riemann integral* of f on $[a, b]$ is denoted $\overline{\int_a^b} f$ and defined by

$$\overline{\int_a^b} f = \inf \{U(f, \mathcal{P}) \text{ s.t. } \mathcal{P} \text{ is a partition of } [a, b]\}$$

Alternatively, the *lower Riemann integral* of f on $[a, b]$ is denoted $\underline{\int_a^b} f$ and defined by

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

Finally, if $\overline{\int_a^b} f = \underline{\int_a^b} f$ then we say that f is *Riemann integrable* and in this case, and define the *Riemann integral of* f, denoted $\int_a^b f$, as

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

NOTATION 5.2.10. It is very common in notation to indicate the *variable of integration*. You will often see $\int_a^b f$ written as $\int_a^b f(x)\,dx$ where x is the variable of integration; it is also perfectly common and reasonable to integrate with respect to any other variable suiting the situation, such as $\int_a^b f(t)\,dt$. It's important to note that $\int_a^b f$, $\int_a^b f(x)\,dx$, $\int_a^b f(t)\,dt$, etc. are all representing the same integral. The notation $\int_a^b f(x)\,dx$ is quite useful in many situations: it agrees with the intuition of how we developed the integral (the intuition is $\sum_i f(x_i)(x_i - x_{i-1}) = \sum_i f(x_i)\Delta_i x \to \int_a^b f(x)\,dx$ as the mesh tends to 0); we often have functions of the form $x \mapsto f(x)+c$ where c remains constant as x varies, so it is less ambiguous to interpret what $\int_a^b [f(x) + c]\,dx$ is asking for rather than $\int_a^b [f + c]$. These statements hold analogously for upper and lower integrals as well.

LEMMA 5.2.11. *Let f be a bounded real valued function defined on $[a, b]$, and let $\mathcal{Q}$ and $\mathcal{P}$ be any partitions of $[a, b]$. Then the following inequalities hold.*

$$L(f, \mathcal{Q}) \leq \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx \leq U(f, \mathcal{P}).$$

Proof. Let $S = \{L(f, \mathcal{P}) \text{ s.t. } \mathcal{P} \text{ is a partition of } [a, b]\}$ and $T = \{U(f, \mathcal{P}) \text{ s.t. } \mathcal{P} \text{ is a partition of } [a, b]\}$. Using the definition of the upper and lower Riemann integrals, what we want to prove is that for any $s \in S$ and $t \in T$ we have $s \leq \sup(S) \leq \inf(T) \leq t$. By the definition of supremum and infimum, we only need to prove that $\sup(S) \leq \inf(T)$, since $s \leq \sup(S)$ and $\inf(T) \leq t$ are immediate. Now, by Corollary 5.2.8, for any $s \in S$ we have that $s \leq t$ for all t, implying

that s is a lower bound of the set T and therefore $s \leq \inf(T)$. However, since s was arbitrary, this last inequality proves that $\inf(T)$ is an upper bound of S, and therefore $\sup(S) \leq \inf(T)$. ■

PROPOSITION 5.2.12. *Suppose that f is a bounded real-valued function defined on $[a,b]$ with bound $R > 0$ (that is, for all $x \in [a,b]$, $|f(x)| \leq R$). Suppose that $\mathcal{Q}$ and $\mathcal{P}$ are partitions of $[a,b]$ such that $\mathcal{P}$ is a refinement of $\mathcal{Q}$. If $\#(\mathcal{P} \setminus \mathcal{Q}) = k$ then both*

$$0 \leq U(f,\mathcal{Q}) - U(f,\mathcal{P}) \leq 2Rk\,\mathrm{mesh}(\mathcal{Q})$$

and

$$0 \leq L(f,\mathcal{P}) - L(f,\mathcal{Q}) \leq 2Rk\,\mathrm{mesh}(\mathcal{Q}).$$

Proof. We will do this by induction on k. Let $T \subset \mathbb{N}$ be the set of integers such that $k \in T$ if and only if for any partitions $\mathcal{Q} \subset \mathcal{P}$ of $[a,b]$ with $k = \#(\mathcal{P} \setminus \mathcal{Q})$, the claimed inequality holds for the upper sum. The proof that $1 \in T$ is left as an exercise. Now, suppose that $k \in T$. Let $\mathcal{Q} \subset \mathcal{P}$ be any partitions of $[a,b]$ where $k+1 = \#(\mathcal{P} \setminus \mathcal{Q})$. Let $y \in \mathcal{P} \setminus \mathcal{Q}$. Set $\mathcal{Q}^* = \mathcal{Q} \cup \{y\}$ so that $\mathcal{Q}^*$ is a refinement of $\mathcal{Q}$ with $\mathcal{Q}^* \setminus \mathcal{Q} = 1$, and $\mathcal{P}$ is a refinement of $\mathcal{Q}^*$ with $\mathcal{P} \setminus \mathcal{Q}^* = k$. By Exercise 6 (18), $1 \in T$, and by the induction hypothesis $k \in T$. Therefore

$$0 \leq U(f,\mathcal{Q}) - U(f,\mathcal{Q}^*) \leq 2R\,\mathrm{mesh}(\mathcal{Q})$$
$$0 \leq U(f,\mathcal{Q}^*) - U(f,\mathcal{P}) \leq 2Rk\,\mathrm{mesh}(\mathcal{Q}^*)$$

Therefore,

$$\begin{aligned} 0 \leq U(f,\mathcal{Q}) - U(f,\mathcal{P}) &= U(f,\mathcal{Q}) - U(f,\mathcal{Q}^*) + U(f,\mathcal{Q}^*) - U(f,\mathcal{P}) \\ &\leq 2R\,\mathrm{mesh}(\mathcal{Q}) + 2Rk\,\mathrm{mesh}(\mathcal{Q}^*) \leq 2R(k+1)\,\mathrm{mesh}(\mathcal{Q}) \end{aligned}$$

where we used that $\mathrm{mesh}(\mathcal{Q}^*) \leq \mathrm{mesh}(\mathcal{Q})$ (since $\mathcal{Q}^*$ is a refinement of $\mathcal{Q}$) in the last equality. This proves that $k \in T \implies k+1 \in T$ and concludes the proof for the upper sum. The proof for the lower sum is done analogously and left as an exercise. ■

PROPOSITION 5.2.13. *Suppose that f is a bounded real-valued function defined on $[a,b]$ with bound $R > 0$. Given any $\varepsilon > 0$ there exists an $\delta > 0$ such that if $\mathcal{P}$ is a partition of $[a,b]$*

with $\operatorname{mesh}(\mathcal{P}) < \delta$ *then*

$$U(f,\mathcal{P}) - \overline{\int_a^b} f(x)\,dx < \varepsilon$$

and

$$\underline{\int_a^b} f(x)\,dx - L(f,\mathcal{P}) < \varepsilon.$$

Proof. By Exercise 6 (20) there is some partition $\mathcal{P}_0$ such that $U(f,\mathcal{P}_0) - \overline{\int_a^b} f < \varepsilon/2$ and $\underline{\int_a^b} f - L(f,\mathcal{P}_0) < \varepsilon/2$. Let $k = \#(\mathcal{P}_0)$ and set $\delta = \varepsilon/(4Rk)$. If $\mathcal{P}$ is any partition of $[a,b]$ such that $\operatorname{mesh}(\mathcal{P}) < \delta$. Using Lemma 5.2.12 we have

$$0 \leq U(f,\mathcal{P}) - U(f,\mathcal{P}\cup\mathcal{P}_0) \leq 2Rk\operatorname{mesh}(\mathcal{P}) < 2Rk\delta = \varepsilon/2$$

and

$$0 \leq L(f,\mathcal{P}\cup\mathcal{P}_0) - L(f,\mathcal{P}) \leq 2Rk\operatorname{mesh}(\mathcal{P}) < 2rk\delta = \varepsilon/2$$

Therefore

$$\begin{aligned} U(f,\mathcal{P}) - \overline{\int_a^b} f &= U(f,\mathcal{P}) - U(f,\mathcal{P}\cup\mathcal{P}_0) + U(f,\mathcal{P}\cup\mathcal{P}_0) - \overline{\int_a^b} f \\ &\leq \frac{\varepsilon}{2} + U(f,\mathcal{P}_0) - \overline{\int_a^b} f \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

and

$$\begin{aligned} \underline{\int_a^b} f - L(f,\mathcal{P}) &= \underline{\int_a^b} f - L(f,\mathcal{P}\cup\mathcal{P}_0) + L(f,\mathcal{P}\cup\mathcal{P}_0) - L(f,\mathcal{P}) \\ &\leq \underline{\int_a^b} f - L(f,\mathcal{P}_0) + \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

This concludes the proof. ■

THEOREM 5.2.14. *Suppose that* f *is a bounded real-valued function defined on* $[a,b]$*. Then the following are equialent:*

(1) f *is Riemann integrable on* $[a,b]$*.*
(2) *Given any* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that for any partition* $\mathcal{P}$ *of* $[a,b]$ *with* $\operatorname{mesh}(\mathcal{P}) < \delta$ *then* $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$*.*
(3) *Given any* $\varepsilon > 0$ *there exists a partition* $\mathcal{P}$ *of* $[a,b]$ *such that* $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$*.*

Proof. (1 $\implies$ 2) By Proposition 5.2.13, given any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any partition $\mathcal{P}$ with $\operatorname{mesh}(\mathcal{P}) < \delta$, $U(f,\mathcal{P}) - \overline{\int_a^b} f < \varepsilon/2$ and $\underline{\int_a^b} f - L(f,\mathcal{P}) < \varepsilon/2$. By

assumption $\overline{\int_a^b} f = \underline{\int_a^b} f$. Therefore,

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) = U(f,\mathcal{P}) - \overline{\int_a^b} f + \underline{\int_a^b} f - L(f,\mathcal{P}) < \varepsilon.$$

(2 $\implies$ 3) Immediate.

(3 $\implies$ 1) Take $\varepsilon > 0$. Let $\mathcal{P}$ be a partition of $[a,b]$ such that $U(f,\mathcal{P}) - \overline{\int_a^b} f < \varepsilon/3$, $\underline{\int_a^b} f - L(f,\mathcal{P}) < \varepsilon/3$, and $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon/3$. The existence of such a partition $\mathcal{P}$ is given in Exercise 6 (23). By the triangle inequality

$$\begin{aligned}\overline{\int_a^b} f - \underline{\int_a^b} f &= \overline{\int_a^b} f - U(f,\mathcal{P}) + U(f,\mathcal{P}) - L(f,\mathcal{P}) + L(f,\mathcal{P}) - \underline{\int_a^b} f \\ &\leq \left| U(f,\mathcal{P}) - \overline{\int_a^b} f \right| + \left| U(f,\mathcal{P}) - L(f,\mathcal{P}) \right| + \left| \underline{\int_a^b} f - L(f,\mathcal{P}) \right| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\end{aligned}$$

This shows that, given any $\varepsilon > 0$, $\left| \overline{\int_a^b} f - \underline{\int_a^b} f \right| < \varepsilon$. Hence $\overline{\int_a^b} f = \underline{\int_a^b} f$, proving that f is Riemann integrable. ■

COROLLARY 5.2.15. *Suppose that f is a bounded real-valued function on $[a,b]$ and suppose that $\{\mathcal{P}_n\}_{n=1}^{\infty}$ is a sequence of partitions on $[a,b]$ such that $\lim_{n\to\infty} \text{mesh}(\mathcal{P}_n) = 0$. Then f is Riemann integrable on $[a,b]$ if and only if*

$$\lim_{n\to\infty} L(f,\mathcal{P}_n) = \lim_{n\to\infty} U(f,\mathcal{P}_n).$$

Moreover, if f is Riemann integrable, then

$$\lim_{n\to\infty} L(f,\mathcal{P}_n) = \lim_{n\to\infty} U(f,\mathcal{P}_n) = \int_a^b f(x)\,dx.$$

Proof. First suppose f is Riemann integrable on $[a,b]$. Let $\varepsilon > 0$. Let $\delta > 0$ be chosen such that for any partition $\mathcal{P}$ such that if $\text{mesh}(\mathcal{P}) < \delta$ then $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$. In particular, this implies that $U(f,\mathcal{P}) - \int_a^b f < \varepsilon$ and $\int_a^b f - L(f,\mathcal{P}) < \varepsilon$ since $U(f,\mathcal{P}) \geq \int_a^b f \geq L(f,\mathcal{P})$. Let $n_\varepsilon \in \mathbb{N}$ be large enough that $\text{mesh}(\mathcal{P}_n) < \delta$ for every $n \geq n_\varepsilon$ (we can find such an n_ε because we assumed that $\text{mesh}(\mathcal{P}_n) \to 0$). Then, by how we chose δ, for every $n \geq n_\varepsilon$ we have $U(f,\mathcal{P}_n) - \int_a^b f < \varepsilon$ and $\int_a^b f - L(f,\mathcal{P}_n) < \varepsilon$. Since ε was arbitrary, we have proved the convergence we wanted.

For the converse, suppose that $\lim_{n\to\infty} L(f,\mathcal{P}_n) = \lim_{n\to\infty} U(f,\mathcal{P}_n)$. Let $\varepsilon > 0$. By assumption, there is some $N \in \mathbb{N}$ such that $U(f,\mathcal{P}_N) - L(f,\mathcal{P}_N) < \varepsilon$. Therefore, by Theorem 5.2.14 (3), f is Riemann integrable on $[a,b]$. ■

2.3. Properties of the Riemann integral.

THEOREM 5.2.16. *Let f and g be bounded real-valued functions which are Riemann integrable on $[a,b]$. Then the following hold.*

(1) *(The integral is linear) For all $\alpha,\beta \in \mathbb{R}$, the function $\alpha f + \beta g$ is Riemann integrable on $[a,b]$ and*

$$\int_a^b \big[\alpha f(x) + \beta g(x)\big]\,dx = \alpha \int_a^b f(x)\,dx + \beta \int_a^b g(x)\,dx$$

(2) *(Products are integrable & Hölder's inequality) fg is Riemann integrable on $[a,b]$. Further,*

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \Big(\int_a^b f(x)^2\,dx\Big)^{1/2}\Big(\int_a^b g(x)^2\,dx\Big)^{1/2}$$

(3) *(Triangle inequality for integrals) The absolute value $|f|$ (defined by $|f|(x) = |f(x)|$ for every x in the domain of f) is Riemann integrable on $[a,b]$ and*

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx$$

(4) *(Positivity of the integral) If $g \le f$ on $[a,b]$ (that is, $g(x) \le f(x)$ for every $x \in [a,b]$), then*

$$\int_a^b g(x)\,dx \le \int_a^b f(x)\,dx.$$

Proof. (1) The proof here is mostly complete, but the final step is Exercise 6 (34). Suppose that f and g are Riemann integrable on $[a,b]$. If $\mathcal{P}$ is any partition of $[a,b]$, then

$$L(f,\mathcal{P}) + L(g,\mathcal{P}) \le L(f+g,\mathcal{P}) \le U(f+g,\mathcal{P}) \le U(f,\mathcal{P}) + U(g,\mathcal{P}).$$

Let $\{\mathcal{P}_n\}_{n=1}^{\infty}$ be a sequence of partitions of $[a,b]$ such that $\operatorname{mesh}(\mathcal{P}_n) \to 0$. Replacing $\mathcal{P}$ in the above inequalities with $\mathcal{P}_n$ and using the Riemannan integrability assumption, we have that

$$\lim_{n\to\infty} L(f,\mathcal{P}_n) + L(g,\mathcal{P}_n) = \int_a^b f + \int_a^b g$$

and

$$\lim_{n\to\infty} U(f,\mathcal{P}_n)+U(g,\mathcal{P}_n)=\int_a^b f+\int_a^b g.$$

Therefore, using the Squeezing Principle, both $L(f+g,\mathcal{P}_n)$ and $U(f+g,\mathcal{P}_n)$ converge to $\int_a^b f+\int_a^b g$. This proves that both $f+g$ is Riemann integrable and that $\int_a^b(f+g)=\int_a^b f+\int_a^b g$.

Now, let $\alpha\in\mathbb{R}$. By the properties of supremum and infimum, we have that if $\alpha\geq 0$, then $U(\alpha f,\mathcal{P})=\alpha U(f,\mathcal{P})$ and $L(\alpha f,\mathcal{P})=\alpha L(f,\mathcal{P})$; if $\alpha<0$ then $U(\alpha f,\mathcal{P})=\alpha L(f,\mathcal{P})$ and $L(\alpha f,\mathcal{P})=\alpha U(f,\mathcal{P})$. In either case, once again by replacing $\mathcal{P}$ with a sequence of partitions $\{\mathcal{P}_n\}_{n=1}^{\infty}$ whose mesh tends to 0, we discover that $\lim_{n\to\infty} U(\alpha f,\mathcal{P}_n)=\lim_{n\to\infty} L(\alpha f,\mathcal{P}_n)=\alpha\int_a^b f$. This proves that αf is Riemann integrable and that $\int_a^b \alpha f=\alpha\int_a^b f$. From this it is left as an exercise to deduce the result.

(2) In Exercise 6 (26), you prove that if a function is Riemann integrable on $[a,b]$, then the square of that function is also Riemann integrable on $[a,b]$. Now, if f and g are Riemann integrable on $[a,b]$, the so is $f+g$ by the previous part. But then, by the exercise and the previous part, $\frac{1}{2}f^2$, $\frac{1}{2}g^2$, and $\frac{1}{2}(f+g)^2$, are also Riemann integrable on $[a,b]$. From here, we have

$$fg=\frac{1}{2}(f+g)^2-\frac{1}{2}f^2-\frac{1}{2}g^2.$$

By the previous part, the righthand side of this equality is Riemann integrable on $[a,b]$, and therefore fg is Riemann integrable on $[a,b]$.

For the inequality $\left|\int_a^b fg\right|^2\leq\int_a^b f^2\int_a^b g^2$, we first note that below, in Theorem **??**, we prove that for any positive numbers $a_1,...,a_n$, and $b_1,...,b_n$ we have

$$\sum_{i=1}^n a_ib_i\leq\Big(\sum_{i=1}^n a_i^2\Big)^{1/2}\Big(\sum_{i=1}^n b_i^2\Big)^{1/2}$$

Suppose first that f and g are both non-negative on $[a,b]$. Let $\mathcal{P}=\{x_0,x_1,...,x_n\}$ be any partition and let's note that, using the same notation as in Definition 5.2.5, $M_i^{f^2}=(M_i^f)^2$, $M_i^{g^2}=(M_i^g)^2$, and $M_i^{fg}\leq M_i^fM_i^g$. Therefore

$$\begin{aligned}U(fg,\mathcal{P})&=\sum_{i=1}^n M_i^{fg}(x_i-x_{i-1})\leq\sum_{i=1}^n M_i^fM_i^g(x_i-x_{i-1})\\&\leq\Big(\sum_{i=1}^n(M_i^f)^2(x_i-x_{i-1})\Big)^{1/2}\Big(\sum_{i=1}^n(M_i^g)^2(x_i-x_{i-1})\Big)^{1/2}\\&=\Big(U(f^2,\mathcal{P})\Big)^{1/2}\Big(U(g^2,\mathcal{P})\Big)^{1/2}.\end{aligned}$$

The second inequality came from Theorem **??** mentioned above with $a_i = M_i^f \sqrt{(x_i - x_{i-1})}$ and $b_i = M_i^g \sqrt{(x_i - x_{i-1})}$. What we now see is that, replacing $\mathcal{P}$ with a sequence of partitions $\mathcal{P}_n$ whose mesh tends to 0, we get

$$\int_a^b fg \leq \Big(\int_a^b f^2\Big)^{1/2} \Big(\int_a^b g^2\Big)^{1/2}$$

which finishes the proof in the case that f and g are both non-negative on $[a, b]$. For the general case, we have that

$$\left|\int_a^b fg\right| \leq \int_a^b |fg| \leq \Big(\int_a^b |f|^2\Big)^{1/2} \Big(\int_a^b |g|^2\Big)^{1/2} = \Big(\int_a^b f^2\Big)^{1/2} \Big(\int_a^b g^2\Big)^{1/2}$$

where the first inequality is from (3) (which is proved in Exercise 6 (24)) and the second inequality is deduced from what we have just shown for non-negative functions $|f|$ and $|g|$.

(3) Exercise 6 (24)

(4) If f and g are both Riemann integrable on $[a, b]$ and $f(x) \geq g(x)$ for all $x \in [a, b]$, then $f - g$ is integrable on $[a, b]$ (from (1) as the linear combination of integrable functions) and $f(x) - g(x) \geq 0$ for all $x \in [a, b]$. But since this implies that for any partition $\mathcal{P}$ of $[a, b]$ we have $L(f - g, \mathcal{P}) \geq 0$ we deduce that $\int_a^b (f - g) \geq 0$. Using (1), this is the same as $\int_a^b f - \int_a^b g \geq 0$, which is one quick rearrangement from the desired inequality. ■

PROPOSITION 5.2.17. *Let $a, b \in \mathbb{R}$ with $a < b$ and let J be any of the intervals $[a, b], (a, b], [a, b), (a, b)$. The* indicator function *on J, denoted $1_J : \mathbb{R} \to \mathbb{R}$, defined by*

$$1_J(x) = \begin{cases} 1 & x \in J \\ 0 & x \notin J \end{cases}$$

is Riemann integrable on any interval $[a', b']$ containing J and

$$\int_{a'}^{b'} 1_J(x)\, dx = b - a.$$

Moreover, if $[a', b']$ is any interval containing J and f is Riemann integrable on $[a, b]$ then $f1_J$ is Riemann integrable on $[a', b']$ and

$$\int_{a'}^{b'} f(x)1_J(x)\, dx = \int_a^b f(x)\, dx.$$

Proof. Exercise 6 (28). ■

PROPOSITION 5.2.18. *Let f be a real-valued function which is Riemann integrable on $[a,b]$. For any sub-interval $[c,b] \subset [a,b]$, it holds that f is Riemann integrable on $[c,d]$.*

Proof. Exercise 6 (27). ■

THEOREM 5.2.19. *Let f be a real-valued function which is Riemann integrable on $[a,b]$. Then if $c \in (a,b)$, then f is Riemann integrable on $[a,c]$ and on $[c,b]$ and*

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

Proof. Let $1_{[a,c)}$ and $1_{[c,b]}$ be indicator functions as defined in Proposition 5.2.17. Let f be any Riemann integrable function on $[a,b]$ and notice that for every $x \in [a,b]$, $1_{[a,b]}(x) = 1_{[a,c)}(x) + 1_{[c,d]}(x)$. In particular, it holds that $f1_{[a,b]} = f1_{[a,c)} + f1_{[c,b]}$. Therefore,

$$\int_a^b f = \int_a^b \left[f1_{[a,c)} + f1_{[c,b]}\right] = \int_a^b f1_{[a,c)} + \int_a^b f1_{[c,b]} = \int_a^c f + \int_c^b f$$

where the second equality follows from the linearity of the integral, and the last equality comes from Proposition 5.2.17. ■

THEOREM 5.2.20. *Suppose that f is a real-valued function defined and* continuous *on $[a,b]$. Then f is Riemann integrable on $[a,b]$.*

Proof. Since f is continuous on $[a,b]$, it is uniformly continuous on $[a,b]$ since $[a,b]$ is compact. Take $\varepsilon > 0$ and find a $\delta > 0$ such that for all $x, y \in [a,b]$ with $|x-y| < \delta$ then $|f(x) - f(y)| < \varepsilon/(b-a)$. Let $\mathcal{P}$ be any partition such that $\text{mesh}(\mathcal{P}) < \delta$. Let's write $\mathcal{P} = \{x_0, x_1, ..., x_n\}$. Then

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) = \sum_{i=1}^n M_i^f (x_i - x_{i-1}) - \sum_{i=1}^n m_i^f (x_i - x_{i-1}) = \sum_{i=1}^n (M_i^f - m_i^f)(x_i - x_{i-1})$$

where, as before, $M_i^f = \sup\{f(x) \text{ s.t. } x \in [x_{i-1}, x_i]\}$ and $m_i^f = \inf\{f(x) \text{ s.t. } x \in [x_{i-1}, x_i]\}$. Since for each i, f is continuous on the compact interval $[x_{i-1}, x_i]$, Corollary 4.4.2 ensures that there exist points $C_i, c_i \in [x_{i-1}, x_i]$ such that $M_i^f = f(C_i)$ and $m_i^f = f(c_i)$. Moreover, $|C_i - c_i| < \delta$ since $|C_i - c_i| \leq |x_i - x_{i-1}| \leq \text{mesh}(\mathcal{P}) < \delta$. By our choice of δ, this means

$f(C_i) - f(c_i) < \varepsilon/(b-a)$. Therefore

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) = \sum_{i=1}^{n}(M_i^f - m_i^f)(x_i - x_{i-1}) = \sum_{i=1}^{n}\big[f(C_i) - f(c_i)\big](x_i - x_{i-1})$$
$$< \sum_{i=1}^{n}\frac{\varepsilon}{b-a}(x_i - x_{i-1}) = \frac{\varepsilon}{b-a}\sum_{i=1}^{n}(x_i - x_{i-1}) = \frac{\varepsilon}{b-a}(b-a) = \varepsilon.$$

The last equality came from the fact that $\sum_{i=1}^{n}(x_i - x_{i-1}) = x_n - x_0 = b - a$. What we have shown is that given any $\varepsilon > 0$, there exists a partition $\mathcal{P}$ such that $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$, which means that f is Riemann integrable on $[a,b]$ (see Theorem 5.2.14). ■

DEFINITION 5.2.21. Thusfar we have only defined the integral $\int_a^b f$ when $a < b$. However it is meaningful at times to consider $\int_b^a f$. To do so, we define $\int_b^a f = -\int_a^b f$. In this way we give an *orientation* to the real numbers which we integrate with respect to (from left to right).

While defining $\int_b^a f = -\int_a^b f$ may seem a bit arbitrary, compare this next theorem with Theorem 5.2.19, and consider it as a good argument as to why this is a sensible definition.

THEOREM 5.2.22. *Suppose that f is a real-valued function which is Riemann integrable on some interval containing the points a, b, c. Then*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Proof. Exercise 6 (31) ■

2.4. Derivatives and Integrals – Fundamental Theorem of Calculus.

DEFINITIONS 5.2.23. Suppose that $U \subset \mathbb{R}$ is an open set and $f : U \to \mathbb{R}$ is a function. We say that F is an *anti-derivative* of f when F is differentiable on U and $F'(x) = f(x)$ for every $x \in U$. Suppose instead that f is a real-valued function defined on $[a,b]$. We say that F is an anti-derivative of f on $[a,b]$ when F is an anti-derivative of f on (a,b) and F has a right-hand derivative at a, a left-hand derivative at b, and $F'(a+) = f(a)$, $F'(b-) = f(b)$.

THEOREM 5.2.24. *Suppose that f is a real-valued function which is Riemann integrable on $[a,b]$. For each $x \in [a,b]$, let $F(x) = \int_a^x f(t)\,dt$. Then the following hold.*

(1) *F is Lipshitz continuous on $[a,b]$.*

(2) *If there exists an* $x_0 \in (a,b)$ *such that* f *is right continuous at* x_0*, then* F *is differentiable from the right at* x_0 *and* $F'(x_0+) = f(x_0)$*. Similarly, if instead* f *is left continuous at* x_0*, then* F *is differentiable from the left at* x_0 *and* $F'(x_0-) = f(x_0)$*. Finally, if* f *is continuous at* x_0*, the* F *is differentiable at* x_0 *and* $F'(x_0) = f(x_0)$*.*

Proof. (1) Since f is Riemann integrable, it is bounded. Therefore, there is some $k > 0$ such that $|f(x)| \leq k$ for every $x \in [a,b]$. Let $x, y \in [a,b]$ with $x < y$. Using Theorem 5.2.19, we have that

$$F(y) - F(x) = \int_a^y f - \int_a^x f = \int_x^y f$$

Now, using the triangle inequality in Theorem 5.2.19, and the positivity of the integral in Theorem 5.2.16, we have

$$|F(y) - F(x)| = \left| \int_x^y f(t)\,dt \right| \leq \int_x^y |f(t)|\,dt \leq \int_x^y k\,dt.$$

Since k is constant (and hence a polynomial), using Exercise 6 (29), we have $\int_x^y k\,dt = k(y-x)$. Hence

$$|F(y) - F(x)| \leq k|y - x|.$$

(2) Suppose that f is right continuous at x_0. Take $\varepsilon > 0$ and find some $\delta > 0$ such that if $x > x_0$ and $x - x_0 < \delta$, then $|f(x) - f(x_0)| < \varepsilon$. Hence, if $x > x_0$ and $x - x_0 < \delta$

$$\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| = \left| \frac{1}{x - x_0} \int_{x_0}^x \big[f(t) - f(x_0)\big]\,dt \right| \leq \frac{1}{x - x_0} \int_{x_0}^x |f(t) - f(x_0)|\,dt$$
$$\leq \frac{1}{x - x_0} \int_{x_0}^x \varepsilon\,dx = \varepsilon.$$

Therefore, what we have is that given any $\varepsilon > 0$ there is a $\delta > 0$ such that if $x > x_0$ and $x - x_0 < \delta$, then

$$\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| < \varepsilon$$

proving that $F'(x_0+)$ exists and $F'(x_0+) = f(x_0)$. The analogous argument works for the left handed derivative. Further, if f is continuous at x_0, then $F'(x_0+) = f(x_0) = F'(x_0-)$ which means that $F'(x_0)$ exists and $F'(x_0) = f(x_0)$. ∎

THEOREM 5.2.25 (Fundamental Theorem of Calculus - Part 1). *If* f *is continuous on* $[a,b]$*, then* $F : [a,b] \to \mathbb{R}$ *defined by* $F(x) = \int_a^x f$ *is an antiderivative of* f *on* $[a,b]$*.*

Proof. If f is continuous on $[a,b]$, then we can apply Theorem 5.2.24 (2) to each $x \in [a,b]$ to deduce that $F'(x) = f(x)$ (including the one-sided derivatives at the endpoints), and therefore F is an antiderivative of f on $[a,b]$. ■

THEOREM 5.2.26 (Fundamental Theorem of Calculus – Part 2). *Suppose that f is a continuous real-valued function on $[a,b]$. Then there exists an anti-derivative of f on $[a,b]$. Moreover, if F is any anti-derivative of f on $[a,b]$, then*

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

Proof. From the Fundamental Theorem of Calculus – Part 1, we know that an anti-derivative of f exists on $[a,b]$; in fact, if for each $x \in [a,b]$, we set $G(x) = \int_a^x f(t)\,dt$, then $G(x)$ is an anti-derivative of f on $[a,b]$. If F is any anti-derivative of f on $[a,b]$, then let $h(x) = F(x) - G(x)$. We have $h'(x) = F'(x) - G'(x) = f(x) - f(x) = 0$ for every $x \in [a,b]$. By Theorem 5.1.16 (3), h is a constant on $[a,b]$ (as its derivative is 0 everywhere). So, there exists some $c \in \mathbb{R}$ such that for every $x \in [a,b]$, $c = F(x) - G(x) \implies G(x) = F(x) - c$. In particular, $G(b) - G(a) = (F(b) - c) - (F(a) - c) = F(b) - F(a)$. It comes down to proving that $G(b) - G(a) = \int_a^b f(t)\,dt$, but this is true by the definition of G. ■

NOTATION 5.2.27. Because of the relationship between integrals and derivatives, if f is a function admitting an antiderivative (on some or all of its domain), it is common to write write $\int f$ or $\int f(x)\,dx$ (or, as before, integrating with respect to whatever variable suits the situation) as a generic symbol representing the antiderivative of f. It is important to note that when using this notation, we do not include bounds on the integral.

REMARK 5.2.28. Note that if f is differentiable and f' is continuous (being intentionally vague about domains here), then Theorem 5.2.26 roughly states $\int f' = f$. □

NOTATION 5.2.29. We will often employ the standard calculus shorthand where if F is any real-valued function defined on the points a, b, then $F\big|_a^b$ means $F(b) - F(a)$. That is, $F\big|_a^b = F(b) - F(a)$.

COROLLARY 5.2.30 (Integration by parts). *Suppose that f and g are differentiable on $[a,b]$ (with right and left derivatives at the boundary) such that f' and g' are continuous on $[a,b]$. Then*

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

In calculus short hand, this is

$$\int_a^b f'g = fg\Big|_a^b - \int_a^b fg'.$$

Proof. By the product rule,

$$(fg)' = f'g + fg'$$

and by assumption the right hand side of this equality is continuous (as the sum and product of continuous functions). Therefore, using the fundamental theorem of calculus and the linearity of the integral,

$$f(b)g(b) - f(a)g(a) = \int_a^b \big(fg\big)' = \int_a^b \big(f'g + fg'\big) = \int_a^b f'g + \int_a^b fg'$$

Now, subtract $\int_a^b fg'$ from each side of the equality. ■

COROLLARY 5.2.31 (u-Substitution). *Suppose that g is differentiable on $[a,b]$ (with left and right handed derivatives at a and b) such that g' continuous on $[a,b]$. Further suppose that f is a real-valued function defined on the image $g([a,b])$ and differentiable on some open set U containing $g([a,b])$, with f' continuous on U. Then,*

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

Written another way,

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

Proof. Let $h = f \circ g$. As the composition of differentiable functions, h is differentiable on $[a,b]$ with $h'(x) = f'(g(x))g'(x)$. Since f', g, and g' are continuous on $[a,b]$ by assumption, h' is continuous on $[a,b]$ and is therefore Riemann integrable on $[a,b]$. The fundamental theorem of calculus says

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

■

2.5. Remarks on Definite Integrals and Antiderivatives. We start with the following observation, which is a quick corollary of the fundamental theorem of calculus.

THEOREM 5.2.32. *Suppose that f is continuous on $[a,b]$ and that there exists some function F defined on $[a,b]$ such that $\int_c^d f = F\big|_c^d$ for every $c, d \in [a,b]$. Then F is an antiderivative*

of f on $[a,b]$. In particular, a function G defined $[a,b]$ is an antiderivative if and only if there exists some constant C such that $G(x) = F(x) + C$ for every $x \in [a,b]$.

Proof. Since for every $x \in [a,b]$, $\int_a^x f = F(x) - F(a)$ by assumption, the fundamental theorem of calculus tells us that $F'(x) = f(x)$ for every $x \in [a,b]$. The fact that any two antiderivatives only differ by a constant part of the proof of Theorem 5.2.26. For the converse of this statement, suppose that $C \in \mathbb{R}$ is arbitrary and $G(x) = F(x)+C$ for every $x \in J$. Then $G'(x) = F'(x) + 0 = F'(x) = f(x)$ for every $x \in [a,b]$ (including the one-sided derivatives at the boundary). Therefore G is an antiderivative of f on $[a,b]$. ■

REMARK 5.2.33. Often in elementary calculus courses students are given a problem where they are given some easily integrable continuous function f and asked to find the general antiderivative $\int f(x)\,dx$. On doing so they typcially find a specific antiderivative F of f and then need to remember to "add C." Their full correct answer is then $F(x) + C$. This "$+C$" is a manifestation of the last statement of Theorem 5.2.32, illuminating the fact that all antiderivatives of f differ by only a constant. □

Another fact that is used frequently in calculus and we can recover easily here is the following.

THEOREM 5.2.34 (Fundamental Theorem of Calculus - More General Version). *Let $J \subset \mathbb{R}$ be some non-empty interval and suppose that $f : J \to \mathbb{R}$ is continuous. Then f has an antiderivative on J. Moreover, if $a \in J$, then the function $F : J \to \mathbb{R}$ defined by $F(x) = \int_a^x f$ is such an antiderivative. Further, $G : J \to \mathbb{R}$ is an antiderivative of f on J if and only if there exist some $C \in \mathbb{R}$ such that $G(x) = F(x) + C$ for every $x \in J$. Finally, if G is any antiderivative of f on J, then for all $c, d \in J$, $\int_c^d f = G(d) - G(c)$.*

Proof. Let $a, x \in J$. Find $b, c \in J$ such that $a, x \in [b,c]$ (we can find such a b and c since J is an interval). Now, define $G(y) = \int_b^y f$ for every $y \in [b,c]$. In particular, by Theorem 5.2.22, $G(y) - G(a) = \int_a^y f = F(y)$. This shows that F and G differ by only the constant $G(a)$ on $[b,c]$ which, using Theorem 5.2.32 proves that F is an antiderivative of f on $[b,c]$. In particular, since $x \in [b,c]$, $F'(x) = f(x)$. Since x was arbitrarily chosen in J, we have proved that F is an antiderivative of f on J. The proofs of the remaining claims follow similarly using the methods of Theorems 5.2.25, 5.2.26, and 5.2.32. ■

REMARK 5.2.35. Recalling the techniques used in elementary calculus to find antiderivatives $\int f$, they matched the techniques one would use to calculate definite integrals $\int_a^b f$. Theorems 5.2.32 and 5.2.34 give a rigorous basis for this. Indeed, if $J \subset \mathbb{R}$ is some interval, f is a continuous function defined on J, and F is a function such that $\int_a^b f = F\big|_a^b$ for any choices of $a, b \in J$, then you can be confident that F is an antiderivative of f on J. In particular, many of the techniques you used (e.g. u-substitution, integration by parts) to find F when calculating $\int_a^b f$ also work as tools to help find F more generally on J. □

EXAMPLE 5.2.36. Let $p \in \mathbb{N}$. Define $f_p : \mathbb{R} \to \mathbb{R}$ be defined by $f_p(x) = x^p$. Then the function $F_p(x) = \frac{1}{p+1} x^{p+1}$ is an antiderivative of f_p on $\mathbb{R}$. We could, of course, prove this by simply checking that $F_p'(x) = f_p(x)$ for every $x \in \mathbb{R}$. However, let's instead only assume that we know $F_1'(x) = f_1(x)$ (i.e., $\frac{d}{dx}\frac{1}{2}x^2 = x$) and show the result for general F_p through integrating f_p inductively. Our inductive assumption is that for some $p \in \mathbb{N}$ it holds that $F_p'(x) = f_p(x)$ for every $x \in \mathbb{R}$. Now, let $a, b \in \mathbb{R}$ be arbitrary. Using integration by parts (with "$u = x$" and "$dv = x^p$", which means that "$v = \frac{1}{p+1} x^{p+1}$" by the inductive assumption)

$$\begin{aligned}\int_a^b x^{p+1}\,dx = \int_a^b x \cdot x^p\,dx &= \frac{1}{p+1} x \cdot x^{p+1}\Big|_a^b - \frac{1}{p+1}\int_a^b x^{p+1}\,dx \\ &= \frac{1}{p+1} x^{p+2}\Big|_a^b - \frac{1}{p+1}\int_a^b x^{p+1}\,dx\end{aligned}$$

Moving $\frac{1}{p+1}\int_a^b x^{p+1}\,dx$ to the lefthand side we have

$$\frac{1}{p+1} x^{p+2}\Big|_a^b = \int_a^b x^{p+1}\,dx + \frac{1}{p+1}\int_a^b x^{p+1}\,dx = \frac{p+2}{p+1}\int_a^b x^{p+1}\,dx$$

Therefore,

$$\int_a^b x^{p+1}\,dx = \frac{1}{p+2} x^{p+1}\Big|_a^b.$$

Since $a, b \in \mathbb{R}$ were arbitrary, Theorems 5.2.32 and 5.2.34 imply that $F_{p+1}(x) = \frac{1}{p+2}x^{p+1}$ is an antiderivative of $f_{p+1}(x) = x^{p+1}$ on $\mathbb{R}$. This concludes the inductive step. In particular, we have found that, in calculus notation, $\int x^p\,dx = \frac{1}{p+1}x^{p+1} + C$. It is worth noting that the introduction of general a, b as bounds was superfluous as we never actually used them, and is why such calculations are typically left in antiderivative notation; for example, the integration by parts step is usually written

$$\int x^{p+1}\,dx = \frac{1}{p+1} x \cdot x^{p+1} - \frac{1}{p+1}\int x^{p+1}\,dx = \frac{1}{p+1}x^{p+2} - \frac{1}{p+1}\int x^{p+1}\,dx.$$

□

3. Series

Formally, a *series* $\sum_{n=1}^{\infty} x_n$ is the sum of all terms of a sequence $\{x_n\}_{n=1}^{\infty}$. However, $\sum_{n=1}^{\infty} x_n$ is by no means guaranteed to exist (imagine if for every n we define $x_n = 1$, or $x_n = (-1)^n$, etc.). Instead, we use the symbol $\sum_{n=1}^{\infty} x_n$ in much the same way as we might write $\lim_{n\to\infty} y_n$ for some sequence $\{y_n\}_{n=1}^{\infty}$; this limit $\lim_{n\to\infty} y_n$ may or may not exist, but it is convenient notation to use anyway.

DEFINITIONS 5.3.1. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. For each $N \in \mathbb{N}$, the Nth *partial sum* of this sequence (often said to be the Nth partial sum of the series $\sum_{n=1}^{\infty} x_n$) is

$$\sum_{n=1}^{N} x_n = x_1 + x_2 + \cdots + x_N$$

We say that the *series* $\sum_{n=1}^{\infty} x_n$ converges (resp. diverges, resp. does not converge) when the sequence of partial sums converges (resp. diverges, resp. does not converge). That is, let $s_N = \sum_{n=1}^{N} x_n$ be the Nth partial sum, we critique the convergence of the series $\sum_{n=1}^{\infty} x_n$ by the convergence of the sequence $\{s_N\}_{N=1}^{\infty}$ of partial sums. If $\sum_{n=1}^{\infty} x_n$ converges then the symbol $\sum_{n=1}^{\infty} x_n$ represents the limit of the partial sums; i.e.,

$$\lim_{N\to\infty} \sum_{n=1}^{N} x_n = \sum_{n=1}^{\infty} x_n$$

Moreover, if this series converges, we often just say $\sum_{n=1}^{\infty} x_n$ *exists*.

NOTATION 5.3.2. It is common to write statements of the form "Consider the series $\sum_{n=1}^{\infty} x_n$" or "Assume that the series $\sum_{n=1}^{\infty} x_n$ converges" etc. without explicitly introducing the sequence $\{x_n\}_{n=1}^{\infty}$ beforehand. Implicitly the sequence $\{x_n\}_{n=1}^{\infty}$ is given in such statements and, in fact, it is usually the case that these statements are critiquing the partial sums $\{s_N\}_{N=1}^{\infty}$ of the sequence rather than the sequence itself, although this is not always true.

3.1. Properties of Convergent Series.

THEOREM 5.3.3. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. The following are equivalent.*

(1) *The series* $\sum_{n=1}^{\infty} x_n$ *converges.*

(2) *There exists some* $S \in \mathbb{R}$ *such that given any* $\varepsilon > 0$ *there exists some* $N_\varepsilon \in \mathbb{N}$ *such that for every* $N \geq N_\varepsilon$,

$$\left| \sum_{n=1}^{N} x_n - S \right| < \varepsilon.$$

(3) *(Cauchy's Criteria for Series) Given any* $\varepsilon > 0$ *there eixsts some* $N_\varepsilon \in \mathbb{N}$ *such that for every* $N \geq N_\varepsilon$ *and* $M \geq N_\varepsilon$,

$$\left| \sum_{n=1}^{N} x_n - \sum_{n=1}^{M} x_n \right| < \varepsilon.$$

Proof. This proof is basically just a reworking of notation. Let $\{s_N\}_{N=1}^{\infty}$ be the sequence of partial sums of the sequence $\{x_n\}_{n=1}^{\infty}$. Note that by definition $\sum_{n=1}^{\infty} x_n$ exists if and only if the sequence $\{s_N\}_{N=1}^{\infty}$ converges, in which case the limit of $\{s_N\}_{N=1}^{\infty}$ is denoted $\sum_{n=1}^{\infty} x_n$.

(1) $\iff$ (2): By definition, the sequence $\{s_N\}_{N=1}^{\infty}$ converges with $\lim_{N\to\infty} s_N = S$ if and only if for every $\varepsilon > 0$ there exists some N_ε such that $|s_N - S| < \varepsilon$ for every $N \geq N_\varepsilon$.

(1) $\iff$ (3): We have proved that the sequence $\{s_N\}_{N=1}^{\infty}$ converges if and only if it is Cauchy. ∎

Example 5.3.4. Let $x \in \mathbb{R}$ and $|x| < 1$. Then

$$\sum_{n=0}^{\infty} x^n = 1 + \sum_{n=1}^{\infty} x^n = \frac{1}{1-x}$$

Proof. Let $\{s_N\}_{N=0}^{\infty}$ be the sequence of partial sums; i.e., $s_N = \sum_{n=0}^{N} x^n$. We have that

$$xs_N = x \sum_{n=0}^{N} x^n = \sum_{n=0}^{N} x^{n+1} = \sum_{n=1}^{N+1} x^n.$$

Therefore

$$(1-x)s_N = s_N - xs_N = x^0 - x^{N+1} = 1 - x^{N+1}.$$

Dividing through by $1 - x$ we have

$$s_N = \frac{1 - x^{N+1}}{1-x}.$$

Since $|x| < 1$, we have proved that $\lim_{N\to\infty} x^N = 0$, which implies that

$$\lim_{N\to\infty} s_N = \frac{1 - x \lim_{N\to\infty} x^N}{1-x} = \frac{1}{1-x}.$$

This proves that the sequence of partial sums converges and converges to $\frac{1}{1-x}$, concluding the proof. ■ □

EXAMPLE 5.3.5. The *harmonic series* $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges to $+\infty$.

Proof. This was proved in Exercise 9 (21). ■ □

PROPOSITION 5.3.6. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of real numbers. If $\sum_{n=1}^{\infty} x_n$ converges then $\lim_{n\to\infty} x_n = 0$. However, the converse needn't be true.*

Proof. Let $\varepsilon > 0$. Let $N_\varepsilon \in \mathbb{N}$ be large enough such that for any $N, M \geq N_\varepsilon$ it holds that $\left| \sum_{n=1}^{N} x_n - \sum_{n=1}^{M} x_n \right| < \varepsilon$. In particular, for any $N \geq N_\varepsilon + 1$,

$$\left| \sum_{n=1}^{N} x_n - \sum_{n=1}^{N-1} x_n \right| = |x_N| < \varepsilon.$$

This shows that given any $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ (take $n_\varepsilon = N_\varepsilon + 1$ for example) such that for every $n \geq n_\varepsilon$, it holds that $|x_n - 0| < \varepsilon$, proving that $\lim_{n\to\infty} x_n = 0$.

To see that the converse needn't be true, look at Example 5.3.5. ■

REMARK 5.3.7. Note that the contrapositive of Proposition 5.3.6: If $\{x_n\}_{n=1}^{\infty}$ does not converge to 0 then the series $\sum_{n=1}^{\infty} x_n$ does not converge. □

NOTATION 5.3.8. Similar to the case of sequences, it is often convenient to consider series starting the sum at some integer other than 1. That is, it is often convenient to consider a series of the form $\sum_{n=p}^{\infty} x_n$ where $p \in \mathbb{Z}$. In this case, it is (implicitly or explicitly) understood that the sequence being summed is defined starting for all integers $n \geq p$. Moreover, the partial sums $s_N = \sum_{n=p}^{N} x_n$ will also start at p (i.e., s_N is defined for all $N \geq p$) unless otherwise explicitly mentioned.

A quick lemma emphasizing how the above notation is often used is as follows.

LEMMA 5.3.9. *Let $p \in \mathbb{Z}$ and $k \in \mathbb{N}$. Then, as long as the sequence being summed over is indexed within the given ranges and each of the series in consideration converge,*

$$\sum_{n=p}^{\infty} x_{n+k} = \sum_{n=p+k}^{\infty} x_n, \qquad \sum_{n=p}^{\infty} x_{n-k} = \sum_{n=p-k}^{\infty} x_n,$$

and for any $N \geq p$,

$$\sum_{n=p}^{\infty} x_n - \sum_{n=p}^{N} x_n = \sum_{n=N+1}^{\infty} x_n.$$

Proof. Exercise 6 (38) ■

REMARK 5.3.10. As noted with sequences, we actually lose no generality by proving our general results with the sum starting at $p = 1$. The reason for this is that if we are considering the series $\sum_{n=p}^{\infty} x_n$, it is the same as considering the series $\sum_{n=1}^{\infty} y_n$ where $y_n = x_{n+p-1}$ is simply reindexing x_n. Therefore we will continue to state our results in terms of series starting at $p = 1$, but these results hold for more general p. □

PROPOSITION 5.3.11. *The series $\sum_{n=1}^{\infty} x_n$ exists if and only if for every $k \geq 1$, the series $\sum_{n=k}^{\infty} x_n$ exists.*

Proof. In terms of the partial sums, the claim of this proposition is that the sequence of partial sums converges if and only if the tail ends of the sequence converges, which we know is true by Theorem 2.3.1. ■

DEFINITION 5.3.12. We say that the series $\sum_{n=1}^{\infty} x_n$ *converges absolutely* when the series $\sum_{n=1}^{\infty} |x_n|$ converges.

LEMMA 5.3.13. *Let $\{s_N\}_{N=1}^{\infty}$ be the sequence of partial sums of the series $\sum_{n=1}^{\infty} |x_n|$. Then $\{s_N\}_{N=1}^{\infty}$ is monotonically increasing and therefore either $\lim_{N\to\infty} s_N$ exists or $\lim_{N\to\infty} s_N = +\infty$. Particularly, $\sum_{n=1}^{\infty} |x_n|$ converges if and only if there is some $R > 0$ such that for all $N \in \mathbb{N}$, it holds that $\sum_{n=1}^{N} |x_n| \leq R$.*

Proof. Exercise 6 (39). ■

THEOREM 5.3.14. *Suppose that $\sum_{n=1}^{\infty} x_n$ converges absolutely. Then the following hold.*

(1) *The series* $\sum_{n=1}^{\infty} x_n$ *converges and* $\left|\sum_{n=1}^{\infty} x_n\right| \leq \sum_{n=1}^{\infty} |x_n|$.

(2) *If* $\sigma : \mathbb{N} \to \mathbb{N}$ *is a bijection (a rearrangement) then* $\sum_{n=1}^{\infty} x_{\sigma(n)}$ *exists and* $\sum_{n=1}^{\infty} x_n = \sum_{n=1}^{\infty} x_{\sigma(n)}$. *This is the rigorous way to say that if* $\sum_{n=1}^{\infty} x_n$ *converges absolutely, then we can rearrange the order of the sum without changing the outcome.*

Proof. (1) Let $\varepsilon > 0$ be arbitrary. Find $N_\varepsilon \in \mathbb{N}$ such that if $N \geq N_\varepsilon$ then $\sum_{n=N}^{\infty} |x_n| < \varepsilon$ (which we can do since by assumption the series $\sum_{n=1}^{\infty} |x_n|$ converges). Now, if $N, M \geq N_\varepsilon$ then

$$\left|\sum_{n=1}^{N} x_n - \sum_{n=1}^{M} x_n\right| = \left|\sum_{\min(N,M)+1}^{\max(N,M)} x_n\right| \leq \sum_{\min(N,M)+1}^{\max(N,M)} |x_n| \leq \sum_{N_\varepsilon}^{\infty} |x_n| < \varepsilon$$

The first inequality is from the standard triangle inequality, and the second follows since we simply added more non-negative terms to the sum (see Exercise 6 (**??**)). Therefore, by the Cauchy Criteria Theorem 5.3.3 (3), the series converges. Moreover, we have

$$\left|\sum_{n=1}^{N} x_n\right| \leq \sum_{n=1}^{N} |x_n| \leq \sum_{n=1}^{\infty} |x_n|$$

The first inequality came again by the triangle inequality, and the second because we're adding on more non-negative terms as before. Now, we have the the sequence of partial sums exists and is bounded by $\sum_{n=1}^{\infty} |x_n|$ (which is assumed to exist), therefore

$$\left|\sum_{n=1}^{\infty} x_n\right| = \left|\lim_{N\to\infty} \sum_{n=1}^{N} x_n\right| = \lim_{N\to\infty} \left|\sum_{n=1}^{N} x_n\right| \leq \lim_{N\to\infty} \sum_{n=1}^{\infty} |x_n| = \sum_{n=1}^{\infty} |x_n|.$$

The second equality follows by the continuity of the absolute value (which lets us commute the limit and the absolute value). This concludes the proof.

(2) By (1), we know that $\sum_{n=1}^{\infty} x_n$ converges. Say $\sum_{n=1}^{\infty} x_n = S$. Let $\{s_N\}_{N=1}^{\infty}$ be the partial sums of $\sum_{n=1}^{\infty} x_n$ and let's use the notation $s_N^\sigma = \sum_{n=1}^{N} x_{\sigma(n)}$. We want to show that $\lim_{N\to\infty} s_N^\sigma = S$. Let $\varepsilon > 0$. Let $N_1 \in \mathbb{N}$ be large enough that for every $N \geq N_1$, $\sum_{n=N}^{\infty} |x_n| < \varepsilon$. We can find such an N_1 since it is assumed that $\lim_{M\to\infty} \sum_{n=1}^{M} |x_n| = \sum_{n=1}^{\infty} |x_n|$ converges and $\sum_{n=N}^{\infty} |x_n| = \sum_{n=1}^{\infty} |x_n| - \sum_{n=1}^{N-1} |x_n|$.

Since $\sigma : \mathbb{N} \to \mathbb{N}$ is a bijection, there is some N_ε such that $\{1, 2, ..., N_1\} \subset \{\sigma(1), \sigma(2), ..., \sigma(N_\varepsilon)\}$. Now, if $N \geq N_\varepsilon$ then

$$\left| \sum_{n=1}^{\infty} x_n - \sum_{n=1}^{N} x_{\sigma(n)} \right| \leq \sum_{n=N_0}^{\infty} |x_n| < \varepsilon$$

We have thus proved that given any $\varepsilon > 0$ there exists some $N_\varepsilon \in \mathbb{N}$ such that if $N \geq N_\varepsilon$ then $|s_N^\sigma - S| < \varepsilon$, proving that $\lim_{N\to\infty} s_N^\sigma = S$. ■

DEFINITION 5.3.15. As we will see below, it can certainly happen that a series $\sum_{n=1}^{\infty} x_n$ converges, but does not converge absolutely. When this is the case, we say that the series $\sum_{n=1}^{\infty} x_n$ *conditionally converges* (or is *conditionally convergent*).

3.2. Series Formulas Analogous to Integration Formulas. Many of the calculus integral results that we have so far seen (the fundamental theorem of calculus, integration by parts, etc.) have analogues in series. In this section we explore some of these. Before doing so, let's build up a little intuition about how "calculus" can take place with respect to sequences and series. Recall that a sequence of real numbers can be viewed as a function $f : \mathbb{N} \to \mathbb{R}$. However, there is no way to take a standard derivative of such a function. The best we can do is look at finite differences. That is, look at ratios of the form

$$\frac{f(n) - f(m)}{n - m}$$

where n and m are as close as possible in the domain of f without begin equal. You'll realize that the closest we can make n and m without equality is when $|n - m| = 1$. Let's introduce notation

$$\Delta_n f = f(n + 1) - f(n).$$

Then

$$\frac{f(n + 1) - f(n)}{(n + 1) - n} = \frac{f(n + 1) - f(n)}{1} = \Delta_n f.$$

Now, see that $\Delta_n f$ is a reasonable candidate for a "derivative" of f at n (from the right, whereas $\Delta_{n-1} f$ is a good candidate for a "derivative" of f at n from the left). So, it is not unreasonable to think f' when seeing $\Delta_n f$ (or $\Delta_{n-1} f$).

On the other hand, let's consider the sum $\sum_{n=M}^{N} f(n)$. This plays the analogue of the integral of $f : \mathbb{N} \to \mathbb{R}$ on the range $n = M$ to $n = N$. Indeed, to analogize the integral to sequences, it is sensible to consider sums of the form $\sum_{\text{range of } n} f(n)\Delta n$ in much the same

way as to integrate we consider sums of the form $\sum_{\text{range of } x} g(x)\Delta x$. At the end of the day, to produce an integral from the sum $\sum g(x)\Delta x$, we take the distance between consecutive points x to 0, i.e., $\Delta x \to 0$. With regards to the sum $\sum f(n)\Delta n$, the smallest we can make the difference Δn of integers is $\Delta n = 1$. So, the sum $\sum_{n=M}^{N} f(n) = \sum_{n=M}^{N} f(n)\Delta n$ is the reasonable analogue of integrating the sequence f over the values $n = M$ to $n = N$.

THEOREM 5.3.16 (The Fundamental Theorem of Calculus for Sums – compare to Theorem 5.2.26). *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of real numbers then for any $N > M \geq 1$*

$$\sum_{n=M}^{N-1} \big(f_{n+1} - f_n\big) = f_N - f_M.$$

In our intuitive notation, this says

$$\sum_{n=M}^{N-1} \Delta_n f = f\big|_M^N.$$

Proof. A real rigorous proof would go by induction, however let's forgo the details for now to try to understand why it is true. Notice:

$$f_{M+2} - f_M = (f_{M+2} - f_{M+1}) + (f_{M+1} - f_M) = \sum_{n=M}^{M+1} \big(f_{n+1} - f_n\big).$$

Similarly,

$$f_{M+3} - f_M = (f_{M+3} - f_{M+2}) + (f_{M+2} - f_{M+1}) + (f_{M+1} - f_M) = \sum_{n=M}^{M+2} \big(f_{n+1} - f_n\big).$$

In fact, you can now believe that for any $k \in \mathbb{N}$,

$$f_{M+k} - f_M = \sum_{n=M}^{M+k-1} (f_{n+1} - f_n).$$

So, letting $k = N - M$ in the statement of the theorem gives us the formula we want. ∎

DEFINITION 5.3.17. A sum (or series) of the form $\sum \big(f_{n+1} - f_n\big)$ as introduced in the previous theorem is called a *telescoping sum* (or *telescoping series*) since the cross terms cancel and the sum "collapses" to just the beginning and ending terms, heuristically in the same way a personal telescope often collapses.

THEOREM 5.3.18 (Summation By Parts – compare to Corollary 5.2.30). *Let* $\{f_n\}_{n_1}^{\infty}$ *and* $\{g_n\}_{n=1}^{\infty}$ *be sequences of real numbers. Suppose that* $M, N \in \mathbb{N}$ *with* $M < N$. *Then*

$$\sum_{k=M}^{N} f_k(g_{k+1} - g_k) = \left[f_{N+1}g_{N+1} - f_M g_M\right] - \sum_{k=M}^{N} (f_{k+1} - f_k)g_{k+1}$$

In rough, but intuitive notation, this says

$$\sum_{k=M}^{N} f_k\, \Delta_k g = fg\big|_M^{N+1} - \sum_{k=M}^{N} \Delta_k f\; g_{k+1}.$$

Proof. Let $n = N - M$. We will prove this by induction on n. Suppose that $n = 1$, then $N = M + 1$ and it's an exercise to show that the equality we want holds. Now, assume that for some $n \geq 1$ the equality holds. Then, let's check this implies it also holds for $n + 1$. We have,

$$\begin{aligned}
\sum_{k=M}^{M+n+1} f_k(g_{k+1} - g_k) &= f_{M+n+1}(g_{M+n+2} - g_{M+n+1}) + \sum_{k=M}^{M+n} f_k(g_{k+1} - g_k) \\
&= f_{M+n+1}(g_{M+n+2} - g_{M+n+1}) + \left[f_{M+n+1}g_{M+n+1} - f_M g_M\right] - \sum_{k=M}^{M+n} (f_{k+1} - f_k)g_{k+1} \\
&= f_{M+n+1}g_{M+n+2} - f_M g_M - \sum_{k=M}^{M+n} (f_{k+1} - f_k)g_{k+1} \\
&= f_{M+n+2}g_{M+n+2} - f_M g_M - \left[f_{M+n+2}g_{M+n+2} - f_{M+n+1}g_{M+n+2}\right] - \sum_{k=M}^{M+n} (f_{k+1} - f_k)g_{k+1} \\
&= f_{M+n+2}g_{M+n+2} - f_M g_M - \sum_{k=M}^{M+n+1} (f_{k+1} - f_k)g_{k+1}
\end{aligned}$$

This concludes the inductive step, and hence the proof. ∎

3.3. Tests for convergence – Non-negative sequences. In this section we consider series $\sum_{n=1}^{\infty} x_n$ where $x_n \geq 0$ for every n. Notice that since $x_n = |x_n|$ in this case, the series $\sum_{n=1}^{\infty} x_n = \sum_{n=1}^{\infty} |x_n|$. More importantly, if we want to critique whether or not a series $\sum_{n=1}^{\infty} y_n$ converges absolutely (whether or not y_n is positive or negative), then we want to consider the series $\sum_{n=1}^{\infty} |y_n|$ and certainly $|y_n| \geq 0$, so the techniques of this section can thusly be applied

to the series $\sum_{n=1}^{\infty} |y_n|$, even though they might not be able to be directly applied to the series $\sum_{n=1}^{\infty} y_n$.

THEOREM 5.3.19 (Comparison Test). *Suppose that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are sequences of non-negative real numbers. Also suppose that there exists some $N_0 \in \mathbb{N}$ such that for every $n \geq N_0$, it holds that $x_n \leq y_n$. Then the following hold.*

(1) *If $\sum_{n=1}^{\infty} y_n$ converges, then $\sum_{n=1}^{\infty} x_n$ converges.*
(2) *If $\sum_{n=1}^{\infty} x_n$ does not converge, then both $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ diverge to $+\infty$.*

Proof. By Proposition 5.3.11, the series $\sum_{n=1}^{\infty} x_n$ converges if and only if the series $\sum_{n=N_0}^{\infty} x_n$ converges; the same can be said for the series $\sum_{n=1}^{\infty} y_n$ and $\sum_{n=N_0}^{\infty} y_n$. Therefore it suffices to consider only the series $\sum_{n=N_0}^{\infty} x_n$ and $\sum_{n=N_0}^{\infty} y_n$, which is convenient recalling that for every $n \geq N_0$ it holds that $0 \leq x_n \leq y_n$.

(1) Suppose that $\sum_{n=N_0}^{\infty} y_n$ converges, say $\sum_{n=N_0}^{\infty} y_n = S$. Since $x_n \leq y_n$ for every $n \geq N_0$, then for any $N \geq N_0$,

$$\sum_{n=N_0}^{N} x_n \leq \sum_{n=N_0}^{N} y_n \leq \sum_{n=N_0}^{\infty} y_n = S.$$

The second inequality follows from Lemma 5.3.13. This proves that the sequence of partial sums of the series $\sum_{n=N_0}^{\infty} x_n$ are bounded above by S, which, from another application of Lemma 5.3.13, proves that $\sum_{n=N_0}^{\infty} x_n$ converges.

(2) Assume that $\sum_{n=N_0}^{\infty} x_n$ does not converge. By Lemma 5.3.13, this means that $\sum_{n=N_0}^{\infty} x_n$ diverges to $+\infty$. In particular, given any $R > 0$, there is some $N \geq N_0$ such that $R < \sum_{n=N_0}^{N} x_n$, and since $x_n \leq y_n$ for each $n \geq N_0$, this implies that $R < \sum_{n=N_0}^{N} y_n$. Since $R > 0$ was arbitrary, we have proved the the partial sums of the series $\sum_{n=N_0}^{\infty} y_n$ are not bounded, which, by another application of Lemma 5.3.13, proves that $\sum_{n=N_0}^{\infty} y_n$ diverges to $+\infty$. ■

THEOREM 5.3.20 (Ratio Test). *Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence of positive real numbers. Suppose also that $\lim_{n\to\infty}(x_{n+1}/x_n) = \lambda$ exists. Then if $\lambda < 1$, $\sum_{n=1}^{\infty} x_n$ exists.*

Alternatively, if $\lambda > 1$, then $\sum_{n=1}^{\infty} x_n$ diverges to ∞. For $\lambda = 1$, this test does not lead to a conclusion.

Proof. First suppose that $\lambda < 1$. Then we can find some s such that $\lambda < s < 1$. Set $\varepsilon = (s - \lambda)$. Now, since $\lim_{n\to\infty} \frac{x_{n+1}}{x_n} = \lambda$ we can find an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $|(x_{n+1}/x_n) - \lambda| < \varepsilon$. Now,

$$x_{n+1}/x_n - \lambda \leq |(x_{n+1}/x_n) - \lambda| < s - \lambda$$

which implies that for every $n \geq n_0$, $x_{n+1}/x_n < s$ (by adding λ to the leftmost and rightmost side of the above inequality). This is equivalent to $x_{n+1} < sx_n$. Notice that we can recursively use this inequality to find that if $k \in \mathbb{N}$, then

$$x_{n_0+k} < s(x_{n_0+k-1}) < s\big(sx_{n_0+k-2}\big) = s^2 x_{n_0+k-2} < \cdots < s^k x_{n_0}$$

Therefore, for any $N > n_0$

$$\sum_{n=n_0}^{N} x_n < \sum_{n=n_0}^{N} s^{n-n_0} x_{n_0} = x_{n_0} \sum_{k=0}^{N-n_0} s^k < x_{n_0} \sum_{k=0}^{\infty} s^k$$

Since $0 < s < 1$, then series on the right side converges and therefore, by the comparison test, you will find that $\sum_{n=n_0}^{\infty} x_n$ converges as well. But, as we have already proved, since the tail $\sum_{n=n_0}^{\infty} x_n$ converges, so does the entire series $\sum_{n=1}^{\infty} x_n$.

Conversely, if $\lambda > 1$, let s be chosen so that $\lambda > s > 1$. Here we do a similar calculation to show that for some $n_0 \in \mathbb{N}$ and any $k \in \mathbb{N}$, $x_{n_0+k} > s^k x_{n_0}$. This will then imply that

$$x_{n_0} \sum_{k=0}^{N-n_0} s^k < \sum_{n=n_0}^{N} x_n$$

With $s > 1$ the series on the left side will diverge as $N \to \infty$, which will force the series on the right side to diverge. The details of this divergent direction are left as an exercise (see Exercises **??**). ■

THEOREM 5.3.21 (Gut Feeling Comparison Test). *Suppose that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are sequences of real numbers such that for every n, $x_n \geq 0$ and $y_n > 0$. Then*
1) If $\sum_{n=1}^{\infty} y_n$ exists and $\lim_{n\to\infty} (x_n/y_n) < \infty$ exists, then $\sum_{n=1}^{\infty} x_n$ exists.

2) If $\sum_{n=1}^{\infty} x_n$ *diverges to* ∞ *and* $\lim_{n\to\infty}(y_n/x_n)$ *either diverges to* ∞ *or* $\lim_{n\to\infty}(y_n/x_n)$ *exists and is strictly positive* (>0)*, then* $\sum_{n=1}^{\infty} y_n$ *diverges to* ∞*.*

3) Most importantly, if there is some $\lambda > 0$ *such that* $\lim_{n\to\infty}(x_n/y_n) = \lambda$ *or* $\lim_{n\to\infty}(y_n/x_n) = \lambda$ *then* $\sum_{n=1}^{\infty} x_n$ *converges if and only if* $\sum_{n=1}^{\infty} y_n$ *converges.*

Proof. 1) If the limit of x_n/y_n converges as $n \to \infty$, then the sequence $\{x_n/y_n\}_{n=1}^{\infty}$ is bounded. Suppose that $k > 0$ is a bound such that for all n, $x_n/y_n \leq k$. Then, for all n, $x_n \leq ky_n$. Therefore, for any $N \in \mathbb{N}$

$$\sum_{n=1}^{N} x_n \leq \sum_{n=1}^{N} ky_n \leq k\sum_{n=1}^{\infty} y_n$$

Here, as usual, the last inequality is a result of adding on more non-negative terms. Since $\sum_{n=1}^{\infty} y_n$ converges, this shows that the partial sums $\left\{\sum_{n=1}^{N} x_n\right\}_{N=1}^{\infty}$ are bounded, and since the terms x_n are always ≥ 0, the partial sums are monotonically increasing. Since monotonically increasing, bounded sequences converge, we find that $\sum_{n=1}^{\infty} x_n$ converges.

2) If the limit of y_n/x_n either diverges to ∞ or converges to a value $\lambda > 0$, then there exists some $\delta > 0$ and some $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, $y_n/x_n \geq \delta$ (for example, if $y_n/x_n \to \lambda$ take $\delta = \lambda/2$, if $y_n/x_n \to \infty$, take $\delta = 1$). This means that for every $n \geq n_0$, $y_n \geq \delta x_n$. Therefore, for any $N \geq n_0$,

$$\delta\sum_{n=n_0}^{N} x_n \leq \sum_{n=n_0}^{N} y_n \leq \sum_{n=1}^{N} y_n$$

Now, by assumption the left side of the above inequality diverges to ∞ as $N \to \infty$, and hence so must the right most side, showing that $\sum_{n=1}^{\infty} y_n$ diverges to ∞.

3) This third statement is just an immediate corollary to the previous two. ■

REMARK 5.3.22. The "Gut Feeling Convergence Test" may not at first glance appear intuitive. But in fact, it is! When deciding the convergence of a certain series $\sum x_n$, we might "in our gut" feel like the terms x_n grow like a more familiar sequence y_n, although we may not be able to directly apply a comparison test to the terms x_n and y_n. In this case, we can use the "Gut Feeling Convergence Test" to decide if $\sum x_n$ and $\sum y_n$ behave similarly. □

EXAMPLE 5.3.23. For every $n \in \mathbb{N}$, let $a_n = \dfrac{n}{n^2 + 2\sin(n^2)}$. Let's determine whether $\sum_{n=1}^{\infty} a_n$ converges. We know that for every x, $|\sin(x)| \leq 1$, so as n becomes large, we

imagine that $a_n \approx \dfrac{n}{n^2} = \dfrac{1}{n}$ and $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges. However, since the sign of $\sin(n^2)$ will alternate, we can't immediately say that $a_n \leq \frac{1}{n}$ or $a_n \geq \frac{1}{n}$ to use the comparison test. Fear not! The Gut Feeling Comparison Test lets us compare $b_n = \frac{1}{n}$ to a_n. Indeed,

$$\frac{a_n}{b_n} = \frac{n}{n + \frac{2}{n}\sin(n^2)} \to 1$$

Therefore, $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges. But, $\sum_{n=1}^{\infty} b_n$ diverges to ∞, implying that $\sum_{n=1}^{\infty} a_n$ will diverge to ∞. □

3.4. Other Series Convergence Tests.

THEOREM 5.3.24 (Dirichlet's Test for Convergence). *Suppose that* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *are sequences of real numbers such that the following three properties hold:*

1) $a_n \to 0$ *as* $n \to \infty$.

2) $\sum_{n=1}^{\infty} |a_{n+1} - a_n|$ *converges.*

3) There exists some $R > 0$ *such that for any* $n \in \mathbb{N}$ *and* m *a non-negative integer,*

$$\left|b_n + b_{n+1} + \cdots + b_{n+m}\right| \leq R.$$

Then $\displaystyle\sum_{n=1}^{\infty} a_n b_n$ *converges.*

Proof. Here is the idea of the proof. Define $B_1 = 0$ and for each $k \in \mathbb{N}$ $B_{k+1} = b_1 + b_2 + \cdots + b_k$ (that is B_{k+1} is the kth partial sum of the b_n's). Now, notice that $B_{n+1} - B_n = b_n$. Let $M, N \in \mathbb{N}$ with $M < N$, then

$$\sum_{n=M}^{N} a_n b_n = \sum_{n=M}^{N} a_n (B_{n+1} - B_n)$$

applying summation by parts,

$$\sum_{n=M}^{N} a_n b_n = a_{N+1} B_{N+1} - a_M B_M - \sum_{n=M}^{N} B_{n+1}(a_{n+1} - a_n)$$

Using the triangle inequality,

$$\left|\sum_{n=M}^{N} a_n b_n\right| \leq |a_{N+1}||B_{N+1}| + |a_M||B_M| + \sum_{n=M}^{N} |B_{n+1}||a_{n+1} - a_n|$$

Note that the assumptions of the theorem guarantee that as $M, N \to \infty$, the right hand side of the inequality goes to 0, ensuring that the partial sums of $\sum a_n b_n$ are Cauchy, and hence convergent. ■

COROLLARY 5.3.25 (Abel's Test for Convergence). *Suppose that* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *are sequences of real numbers such that the following three properties hold:*

1) There exists some $n_0 \in \mathbb{N}$ *such that if* $n \geq n_0$ *then* $a_{n+1} \leq a_n$.

2) $a_n \to 0$ *as* $n \to \infty$.

3) There exists some $R > 0$ *such that for any* $n \in \mathbb{N}$ *and* m *a non-negative integer,*

$$\left|b_n + b_{n+1} + \cdots + b_{n+m}\right| \leq R.$$

Then $\displaystyle\sum_{n=1}^{\infty} a_n b_n$ *converges.*

Proof. By Theorem 5.3.24, we only need to show that $\sum_{n=1}^{\infty} |a_{n+1} - a_n|$ converges. For $n \geq n_0$, notice that $|a_{n+1} - a_n| = a_n - a_{n+1}$. Therefore, using a telescoping series, for any $N, M > n_0$

$$\sum_{n=M}^{n=N} |a_{n+1} - a_n| = \sum_{n=M} \{a_n - a_{n+1}\} = a_M - a_N$$

Since, $a_n \to 0$ as $n \to \infty$, we have that given any $\epsilon > 0$, there exists some $m_0 \in \mathbb{N}$ such that if $M > m_0$ then $|a_M| < \varepsilon/2$. Therefore, taking $N_0 = \max(n_0, m_0)$ we have

$$\sum_{n=M}^{n=N} |a_{n+1} - a_n| = \sum_{n=M} \{a_n - a_{n+1}\} = a_M - a_N \leq |a_M| + |a_N| < \varepsilon$$

Therefore the partial sums of $\sum\{|a_{n+1} - a_n|\}$ are Cauchy, hence $\sum_{n=1}^{\infty} |a_{n+1} - a_n|$ converges. ■

COROLLARY 5.3.26 (Alternating Series Test). *Suppose that* $\{a_n\}_{n=1}^{\infty}$ *is a sequence of real numbers such that the following properties hold*

1) There exists some $n_0 \in \mathbb{N}$ *such that for every* $n \geq n_0$, *then* $a_{n+1} \leq a_n$.

2) $a_n \to 0$ *as* $n \to \infty$.

Then $\sum_{n=1}^{\infty} (-1)^n a_n$ *converges.*

Proof. The proof of this is an exercise (see Exercise in Exercises **??**). The idea is to let $b_n = (-1)^n$ and use Abel's convergence test. ■

4. More on Limits of Functions

4.1. Horizontal Assymptotes. With the interpretation of a sequence $\{x_n\}_{n=1}^{\infty}$ as a function $x : \mathbb{N} \to \mathbb{R}$ (with $x(n) = x_n$), we have made sense of what it means for $\lim_{n\to\infty} x(n)$ to exist (this is simply the convergence of the sequence). Intuitively, we understand that if the sequence converges then $x(n)$ tends to some value (the limit of the sequence) as $n \to \infty$. In this section we make sense of this more generally where if, instead of the domain being $\mathbb{N}$, we consider functions $f : U \to \mathbb{R}$ where $U \subset \mathbb{R}$ is unbounded.

Definitions 5.4.1. Let $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$ be a function. If U is unbounded above, we say that the *asymptotic limit* of f at $+\infty$ exists whenever there exists some $y \in \mathbb{R}$ such that for all $\varepsilon > 0$, there exists some $R_\varepsilon \in \mathbb{R}$ such that for all $x \in U$ with $x \geq R_\varepsilon$ it holds that $|f(x) - y| < \varepsilon$. In this case, we say that y is the *righthand horizontal asymptote* of f and write $\lim_{x\to\infty} f(x) = y$.

On the other hand, if U is unbounded below, we say that the *asymptotic limit* of f at $-\infty$ exists whenever there exists some $y \in \mathbb{R}$ such that for all $\varepsilon > 0$, there exists some $R_\varepsilon \in \mathbb{R}$ such that for all $x \in U$ with $x \leq R_\varepsilon$ it holds that $|f(x) - y| < \varepsilon$. In this case we say that y is the *lefthand horizontal asymptote* of f and write $\lim_{x\to-\infty} f(x) = y$.

Remark 5.4.2. In the case $U = \mathbb{N}$, then $f : U \to \mathbb{R}$ is a sequence. Notice that the sequence f converges if and only if it has a righthand asymptote. See Exercise 6 (**??**). □

Theorem 5.4.3 (L'Hopital's Rule – Horizontal Asymptote Version). *Suppose that f and g are differentiable functions on some interval (a, ∞) such that $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = 0$. Further suppose that $g(x) \neq 0$ for any $x \in (a, \infty)$ and that $\lim_{x\to\infty} \frac{f'(x)}{g'(x)}$ exists. Then $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)}$.*

Proof. It suffices to consider $a > 0$, otherwise notice $(1, \infty) \subset (a, \infty)$ and we can restrict our attention to $(1, \infty)$ and set $a = 1$. The function $h : (0, 1/a) \to \mathbb{R}$ defined by $h(x) = \frac{1}{x}$ is differentiable (as the ratio of differentiable functions where the denominator is non-zero on the domain) and $h'(x) \neq 0$ for any $x \in (0, 1/a)$ since $h(x)$ is strictly decreasing. By Exercise 6 (**??**),

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

By using L'Hopital's rule Theorem 5.1.19 applied to the functions $f \circ h$ and $g \circ h$ we have

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

where the last equality was another application of Exercise 6 (**??**). ■

4.2. Vertical Asymptotes.

DEFINITIONS 5.4.4. Let $U \subset \mathbb{R}$, and $f : U \to \mathbb{R}$ be a function.. Let $x_0 \in \mathbb{R}$ or $x_0 = \pm\infty$. We say that the lefthand limit of f at x_0 *diverges* to $+\infty$ when given any $R > 0$, there exists some $\delta > 0$ such that $(x_0 - \delta, x_0) \subset U$ and $f(x) > R$ for all $x \in (x_0 - \delta, x_0)$. In this case we write $\lim_{x\to x_0-} f(x) = +\infty$. We say that the lefthand limit of f at x_0 *diverges* to $-\infty$ when given any $R < 0$, there exists some $\delta > 0$ such that $(x_0 - \delta, x_0) \subset U$ and $f(x) < R$ for all $x \in (x_0 - \delta, x_0)$. In this case we write $\lim_{x\to x_0-} f(x) = -\infty$.

Analogously, we say that the righthand limit of f at x_0 *diverges* to $+\infty$ when given any $R > 0$, there exists some $\delta > 0$ such that $(x_0, x_0+\delta) \subset U$ and $f(x) > R$ for all $x \in (x_0, x_0+\delta)$. In this case we write $\lim_{x\to x_0+} f(x) = +\infty$. We say that the righthand limit of f at x_0 *diverges* to $-\infty$ when given any $R < 0$, there exists some $\delta > 0$ such that $(x_0, x_0 + \delta) \subset U$ and $f(x) < R$ for all $x \in (x_0, x_0 + \delta)$. In this case we write $\lim_{x\to x_0-} f(x) = -\infty$.

EXAMPLE 5.4.5. The function $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$ is such that $\lim_{x\to 0+} f(x) = +\infty$. The function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^3$ is such that $\lim_{x\to-\infty} g(x) = -\infty$ and $\lim_{x\to\infty} g(x) = +\infty$. □

DEFINITION 5.4.6. Suppose $U \subset \mathbb{R}$ and let $x_0 \in \mathbb{R}$ or $x_0 = \pm\infty$. Suppose that $f : U \to \mathbb{R}$ is such that $\lim_{x\to x_0-} f(x) = \pm\infty$ (resp. $\lim_{x\to x_0+} f(x) = \pm\infty$), then we say that f has a *vertical asymptote* from the left (resp. right) at x_0. If it happens that f has a vertical asymptote at x_0 from both the right and the left, then we simply say that f has a vertical asymptote at x_0.

THEOREM 5.4.7 (L'Hopital's Rule – Vertical Asymptote Version). *Suppose that f and g are both real valued functions such that both f and g have vertical asymptote from the left at x_0. Suppose further that are differentiable on some non-empty interval of the form (a, x_0) both f and g are differentiable and that $\lim_{x\to x_0-} \frac{f'(x)}{g'(x)}$ exists. Then $\lim_{x\to x_0-} \frac{f(x)}{g(x)} = \lim_{x\to x_0-} \frac{f'(x)}{g'(x)}$. The analogous statement can be made for vertical asymptotes from the right.*

Proof. Exercise 6 (**??**). ■

4.3. Improper Riemann Integrals. Suppose that f is a real valued function which is defined on some interval (a, b). We have not defined what $\int_a^b f$ means, since this was only defined when f is defined on the interval $[a, b]$. Moreover, we also haven't defined integrals of the form $\int_a^\infty f$, nor $\int_{-\infty}^b f$, nor $\int_{-\infty}^\infty f$. In this section we attempt to make sense of these integrals.

NOTATION 5.4.8. Suppose that f is defined on some non-empty interval $[c, b)$ (we allow for the case $b = +\infty$) such that for all $x \in [c, b)$, it holds that f is Riemann integrable on $[c, x]$. For every $x \in [c, b)$ let $F(x) = \int_c^x f$. Notice that by the assumption of Riemann integrability of f, F is well-defined. If $\lim_{x\to b-} F(x)$ exists, then we write $\int_c^b f$ as this limit; i.e., $\lim_{x\to b-} \int_c^x f = \int_c^b f$.

Instead suppose that f is defined on some non-empty interval $(a, c]$ (we allow for the case $a = -\infty$) such that for all $x \in (a, c]$, it holds that f is Riemann integrable on $[x, c]$. For every $x \in (a, c]$ let $G(x) = \int_x^c f$. If $\lim_{x\to a+} G(x)$ exists, then we write $\int_a^c f$ as this limit; i.e., $\lim_{x\to a+} \int_x^c f = \int_a^c f$.

Further, if f is defined on some interval (a, b) and for some $c \in (a, b)$ both $\int_a^c f$ and $\int_c^b f$ exist, then we write $\int_a^b f$ as $\int_a^c f + \int_c^b f$. In Exercise 6 (**??**) you prove that if $\int_a^c f$ and $\int_c^b f$ exist for some $c \in (a, b)$, then they exist for all $c \in (a, b)$ and result in the same value for $\int_a^b f$.

EXAMPLE 5.4.9. We will prove that $\int_1^\infty \frac{1}{x^2}\, dx = 1$. Using our standard techniques, we know that $F(x) = -\frac{1}{x}$ is an anti-derivative of $f(x) = \frac{1}{x^2}$ on the interval $[1, \infty)$. Therefore, for any $y \in [1, \infty)$ we have $\int_1^y f = F(y) - F(1) = -\frac{1}{y} + 1$. Since by Exercise 6 (**??**) $\lim_{y\to\infty} \left(-\frac{1}{y} + 1\right) = 0 + 1 = 1$, we have proved that $\int_1^\infty f = 1$. □

EXAMPLE 5.4.10. We have proved that if $F(x) = 2\sqrt{x}$ for $x \in (0, 1)$, then $F'(x) = \frac{1}{\sqrt{x}}$ for every $x \in (0, 1)$. Let $f : (0, 1) \to \mathbb{R}$ be defined by $f(x) = \frac{1}{\sqrt{x}}$. Then, by what we just mentioned, F is an antiderivative of f on $(0, 1)$. We will prove that $\int_0^1 f$ exists and that $\int_0^1 f = 1$. Let $c = 1/4 \in (0, 1)$ and $x \in (0, c]$. Then $\int_x^{1/4} f = F(1/4) - F(x) = 1/2 - \sqrt{x}$. Now $\lim_{x\to 0+} (1/2 - \sqrt{x}) = 1/2$ since $\sqrt{x}$ is continuous from the right at $x = 0$ with $\sqrt{0} = 0$. Therefore $\int_0^{1/4} f = 1/2$. On the other hand, let $x \in [1/4, 1)$. Then $\int_{1/4}^x f = F(x) - F(1/4) = \sqrt{x} - 1/2$. Now $\lim_{x\to 1-} (\sqrt{x} - 1/2) = 1 - 1/2 = 1/2$ since $\sqrt{x}$ is continuous at $x = 1$ with $\sqrt{1} = 1$. Therefore $\int_{1/4}^1 f = 1/2$. Putting this together, we have that $\int_0^1 f = \int_0^{1/4} f + \int_{1/4}^1 f = 1/2 + 1/2 = 1$. □

4.4. Limit Supremums and Limit Infimums.

5. Taylor Expansions and Approximation

We begin here by recalling standard notation introduced from calculus for derivatives. Let $U \subset \mathbb{R}$. If $f : U \to \mathbb{R}$ is differentiable at some point $x_0 \in U$, then we've already introduced $f'(x_0)$ as the notation to denote the derivative of f at x_0. We will also use the notation $\frac{df}{dx}(x_0)$ or $\frac{d}{dx}\big|_{x_0} f$ or $\frac{d}{dx}\big|_{x_0} f(x)$ to denote this derivative as well. Further, if the variable we are using to denote the argument of f is given by a different symbol, say we are using the notation $f(t)$ rather than $f(x)$, then $\frac{df}{dt}(x_0)$ or $\frac{d}{dt}\big|_{x_0} f$ or $\frac{d}{dt}\big|_{x_0} f(t)$ are also commonly used. We extend this notation as before and simply write $\frac{df}{dx}$ or $\frac{d}{dx} f$ in place of f' whenever the derivative itself is the function in consideration.

Definition 5.5.1. Let $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$ be a function. The *0th derivative of f* is denoted $f^{(0)}$ and is defined as f itself. That is $f^{(0)} = f$. Suppose that there is some interval $(a, b) \subset U$ and x_0 be a point in (a, b). If f' exists on (a, b), it makes sense then to investigate whether we can take the derivative of f' at any point in (a, b). If we can, say $(f')'(x_0)$ exists for some $x_0 \in (a, b)$, we say that the *second derivative of f at x_0* is $(f')'(x_0)$ and denote this by $f^{(2)}(x_0)$ or $f''(x_0)$. We can continue in this fashion inductively. The nth derivative of f at a point $x_0 \in U$ is defined when there is some interval (c, d) such that the $(n-1)$st derivative is defined for every point in (c, d) and $(f^{(n-1)})'(x_0)$ exists. In this case, the nth derivative of f at x_0 is $(f^{(n-1)})'(x_0)$ and will be denoted by $f^{(n)}(x_0)$.

Remark 5.5.2. Note that for the nth derivative is only defined at a point x_0 for a function f when the $(n-1)$st derivative exists within some (possibly small) interval (a, b) containing x_0 and also must be continuous at x_0. In turn, this then means that the $(n-2)$nd derivative of f also exists at every point in (a, b) and will be differentiable (and thus continuous as well) at every point in (a, b). We continue backwards until we arrive at the original function f. We can conclude that f is differentiable and each higher derivative of $f^{(k)}$ is itself continuous for $0 \leq k \leq n-2$. □

Definition 5.5.3. Let $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$ be a function such that f has n derivatives at some point $x_0 \in U$. That is $f^{(n)}(x_0)$ exists. For $n \in \mathbb{Z}$ with $n \geq 0$, we define the *nth*

Taylor polynomial of f centered at x_0 as the polynomial $p_n^f(\,\cdot\,;x_0):\mathbb{R}\to\mathbb{R}$ given by

$$\begin{aligned}p_n^f(x;x_0) &= \sum_{k=0}^{n}\frac{f^{(k)}(x_0)}{k!}\,(x-x_0)^k\\ &= f(x_0)+f'(x_0)\,(x-x_0)+\frac{f''(x_0)}{2}\,(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}\,(x-x_0)^n.\end{aligned}$$

LEMMA 5.5.4. *Suppose that $U\subset\mathbb{R}$ is open and f is a real valued function which is differentiable on U. Suppose further that for some $n\geq 1$, $f^{(n)}(x_0)$ exists at some point $x_0\in U$. Then*

$$\frac{d}{dx}\,p_n^f(x;x_0)=p_n^{f'}(x;x_0)$$

for every $x\in U$.

Proof. Since $p_n^f(x;x_0)$ is a polynomial (in the variable x), it is differentiable on all of $\mathbb{R}$. Now, by the linearity of the derivative, and chain rule,

$$\begin{aligned}&\frac{d}{dx}p_n^f(x_0,x)=\frac{d}{dx}\left(f(x_0)+f'(x_0)\,(x-x_0)+\frac{f''(x_0)}{2}\,(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}\,(x-x_0)^n\right)\\ &=\frac{d}{dx}\,f(x_0)+f'(x_0)\frac{d}{dx}\,(x-x_0)+\frac{f''(x_0)}{2}\frac{d}{dx}\,(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}\frac{d}{dx}\,(x-x_0)^n\\ &=f'(x_0)+f''(x_0)\,(x-x_0)+\cdots+\frac{f^{(n-1)}}{(n-2)!}\,(x-x_0)^{n-2}+\frac{f^{(n)}(x_0)}{(n-1)!}\,(x-x_0)^{n-1}\\ &=f'(x_0)+(f')'(x_0)\,(x-x_0)+\cdots+\frac{(f')^{(n-2)}(x_0)}{(n-2)!}\,(x-x_0)^{n-2}+\frac{(f')^{(n-1)}(x_0)}{(n-1)!}\,(x-x_0)^{n-1}\\ &=p_{n-1}^{f'}(x;x_0).\end{aligned}$$

This concludes the proof. ∎

THEOREM 5.5.5. *Suppose $U\subset\mathbb{R}$ is an open set and f is a real valued function which has n derivatives at some point $x_0\in U$, for some $n\geq 1$. Define the function $R_n:U\to\mathbb{R}$ by*

$$R_n(x)=\begin{cases}\dfrac{f(x)-p_n^f(x;x_0)}{(x-x_0)^n} & x\neq x_0\\ 0 & x=x_0\end{cases}$$

Then $\lim\limits_{x\to x_0}R_n(x)=0$. Incidentally, since we defined $R_n(x_0)=0$, this implies that R_n is continuous at x_0.

Proof. By induction. Let $T \subset \mathbb{N}$ be defined so that $n \in T$ if and only if every real-valued function f defined on some neighborhood of x_0 with n derivatives at x_0 has $\lim_{x \to x_0} R_n(x) = 0$, where R_n is defined as in the statement of the theorem. The fact that $1 \in T$ is proved in Exercise 6 (11). Now, assume $n \in T$, and we want to prove that $n+1 \in T$. To do so, let f be defined on some neighborhood U of x_0 with $n+1$ derivatives at x_0. Let $R_{n+1} : U \to \mathbb{R}$ be given as above by

$$R_{n+1}(x) = \begin{cases} \dfrac{f(x) - p_{n+1}^f(x; x_0)}{(x - x_0)^{n+1}} & x \neq x_0 \\ 0 & x = x_0 \end{cases}$$

Let $\tilde{U} \subset U$ be the subset of U where $f^{(n)}$ exists (that is $x \in \tilde{U}$ if and only if $f^{(n)}(x)$ exists). Note that $\tilde{U}$ is a neighborhood of x_0, since $f^{(n+1)}(x_0)$ exists, $f^{(n)}$ must be defined in some open set containing x_0. Since f' then must also be defined on $\tilde{U}$ (possibly on a larger subset) and has n derivatives at x_0, we can apply the inductive assumption to conclude that $\lim_{x \to x_0} S_n(x) = 0$ where S_n is defined on the domain of f' by

$$S_n(x) = \begin{cases} \dfrac{f'(x) - p_n^{f'}(x; x_0)}{(x - x_0)^n} & x \neq x_0 \\ 0 & x = x_0 \end{cases}$$

An application of L'Hopital's rule and Lemma 5.5.4 shows that

$$\begin{aligned} \lim_{x \to x_0} R_{n+1}(x) = \lim_{x \to x_0} \frac{\frac{d}{dx}\left(f(x) - p_{n+1}^f(x; x_0)\right)}{\frac{d}{dx}(x - x_0)^{n+1}} &= \lim_{x \to x_0} \frac{f'(x) - p_n^{f'}(x; x_0)}{(n+1)(x - x_0)^n} \\ &= \frac{1}{n+1} \lim_{x \to x_0} S_n(x) = 0. \end{aligned}$$

This shows that $n \in T \implies n+1 \in T$, and finishes the proof. ∎

THEOREM 5.5.6 (Taylor's Remainder Theorem - Differential Version). *Suppose that $J \subset \mathbb{R}$ is a non-empty open interval and f is a function defined and has $n+1$ times differentiable on J. Then given $x, x_0 \in J$ with $x \neq x_0$, there exists some c between x and x_0 such that*

$$f(x) - p_n^f(x; x_0) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1}$$

Alternatively written, there exists some c between x and x_0 such that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1}$$

Proof. By induction. Let T be the set of non-negative integers such that $n \in T$ if and only if for any real valued function f which is n times differentiable on any non-empty open interval, it holds that for every points x_1 and x_0 within that interval, there exists some c between x_1 and x_0 such that

$$f(x_1) - p_{n-1}^f(x_1; x_0) = \frac{f^{(n)}(c)}{n!}(x_1 - x_0)^n.$$

For the case $n = 1$ this is the mean value theorem, and therefore is already proved. Assume that $n \in T$, and we will prove that $n + 1 \in T$. To do so, assume that f is $n + 1$ times differentiable on some non-empty open interval J. Let $x_1, x_0 \in J$. By Theorem 5.5.5 (and using the same notation therein),

$$f(x_1) - p_{n+1}^f(x_1; x_0) = f(x_1) - p_n^f(x_1; x_0) - \frac{f^{(n+1)}(x_0)}{(x_1 - x_0)^{n+1}}(x_1 - x_0)^{n+1} = R_{n+1}(x_1)(x_1 - x_0)^{n+1}$$

From this, we rearrange to find

$$f(x_1) - p_n^f(x_1; x_0) = R_{n+1}(x_1)(x_1 - x_0)^{n+1} + \frac{f^{(n+1)}(x_0)}{(x_1 - x_0)^{n+1}}(x_1 - x_0)^{n+1} = k(x_1 - x_0)^{n+1}$$

where $k = R_{n+1}(x_1) + \frac{f^{(n+1)}(x_0)}{(x_1 - x_0)}$; most importantly for our proof, we have that there is some $k \in \mathbb{R}$ such that $f(x_1) - p_n^f(x_1; x_0) = k(x_1 - x_0)^{n+1}$. We now prove that $k = \frac{f^{(n+1)}(c)}{(n+1)!}$ for some c between x_1 and x_0 to finish the proof.

Define $h : J \to \mathbb{R}$ on J by

$$h(x) = f(x) - p_n^f(x; x_0) - k(x - x_0)^{n+1}.$$

By inspection $h(x_0) = 0$, and by how we chose k, $h(x_1) = 0$. Therefore, we can apply the mean value theorem (in the guise of Rolle's Theorem) to deduce that there is some point x_2 between x_1 and x_0 such that $h'(x_2) = 0$. By how h was defined, and using Lemma 5.5.4, this implies that $f'(x_2) - p_{n-1}^{f'}(x_2, x_0) = (n+1)k(x_2 - x_0)^n$. However, by the induction hypothesis applied to the function f', there is some c between x_0 and x_2 such that $f'(x_2) - p_{n-1}^{f'}(x_2, x_0) = \frac{(f')^{(n)}(c)}{n!}(x_2 - x_0)^n$. This shows that $(n+1)(x_2 - x_0)^n = \frac{f^{(n+1)}(c)}{n!}(x_2 - x_0)^n$, which rearranges to $k = \frac{f^{n+1}(c)}{(n+1)!}$. Since c is between x_0 and x_2, it is between x_0 and x_1, and therefore we have shown that $n \in T \implies n + 1 \in T$. ∎

6. Exercises

(1) Prove (3) $\iff$ (4) in Theorem 5.1.3. Briefly explain why this will finish the proof of the theorem.

(2) Prove Theorem 5.1.5 (1).

(3) Prove Corollary 5.1.12

(4) Prove Corollary 5.1.15.

(5) Prove Theorem 5.1.16 (2).

(6) Let $U \subset \mathbb{R}$ be open. Let $f : U \to \mathbb{R}$ be any polynomial (i.e., $f(x) = a_n x^n + \cdots + a_1 x + a_0$ for some $a_0, ..., a_n \in \mathbb{R}$). Prove that f is differentiable on U and that if $f(x) = a_n x^n + \cdots + a_1 x + a_0$ then $f'(x) = n a_n x^{n-1} + \cdots + 2a_2 x + a_1$. *Hint*: Use induction on the degree of the polynomial and Theorem 5.1.5. See Exercise 9 (32).

(7) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x^2 \sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Prove that f is differentiable, but that the derivative function $f' : \mathbb{R} \to \mathbb{R}$ is not continuous at $x = 0$. In particular, that $\sin : \mathbb{R} \to \mathbb{R}$ is differentiable, $\sin'(x) = \cos(x)$, and $|\sin(x)| \leq 1$ for every x. Moreover, you can assume that $\cos(x) = \cos(x \pm 2\pi k)$ for every $x \in \mathbb{R}$ and $k \in \mathbb{Z}$, $\cos(0) = 1$, and $\cos(\pi) = -1$.

(8) Suppose that $f : (0, 1) \to \mathbb{R}$ is differentiable and that $f'(x) \neq 1$ for all $x \in (0, 1)$. Prove that there is *at most* one $c \in (0, 1)$ such that $f(c) = c$.

(9) Suppose that U is open and $f : U \to \mathbb{R}$ is differentiable such that $f' : U \to \mathbb{R}$ is continuous. If $[a, b] \subset U$, prove that the restriction $f : [a, b] \to \mathbb{R}$ is Lipschitz continuous.

(10) Suppose that U is a neighborhood of x_0 and $f : U \to \mathbb{R}$ is a function. Prove that f is differentiable at x_0 if and only if there exists some λ such that for all $\varepsilon > 0$ there exists some $\delta > 0$ such that $x \in (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$ it holds that

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - \lambda \right| < \varepsilon,$$

and if this is the case, the $f'(x_0) = \lambda$.

(11) Let $\lambda \in \mathbb{R}$ and suppose that U is a neighborhood of some point x_0 and $f : U \to \mathbb{R}$ is a function. Prove that f is differentiable at x_0 and $f'(x_0) = \lambda$ if and only if there exists some neighborhood V of 0 and a function $R : V \to \mathbb{R}$ such that $\lim_{x\to 0} R(x) = R(0) = 0$ and $f(x) = f(x_0) + \lambda(x - x_0) + R(x - x_0) \cdot (x - x_0)$ for every $x \in U$. *Hint*: Notice that $\frac{f(x)-f(x_0)}{x-x_0} = \lambda + R(x - x_0)$ when $x \neq x_0$.

(12) (a) Suppose that $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f : [a, b] \to \mathbb{R}$ is such that f is differentiable on the open interval (a, b), is continuous on $[a, b]$, and $f(a) = f(b)$. Prove that Rolle's Theorem still holds. That is, prove that there exists some $c \in (a, b)$ such that $f'(c) = 0$.

(b) Suppose that $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f : [a, b] \to \mathbb{R}$ is such that f is differentiable on the open interval (a, b), is continuous on $[a, b]$. Prove that the mean value theorem still holds. That is, prove that there exists some $c \in (a, b)$ such that $\frac{f(b)-f(a)}{b-a} = f'(c)$.

(13) (Cauchy's Mean Value Theorem) Suppose that U is an open set containing the interval $[a, b]$ and that f and g are real-valued functions which are differentiable on U. In this problem we will prove that if $g(a) \neq g(b)$ then there exists some $c \in (a, b)$ such that $\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}$.

(a) Let $h : U \to \mathbb{R}$ be defined by $h(x) = [f(x) - f(a)] - k[g(x) - g(a)]$ where $k = \frac{f(b)-f(a)}{g(b)-g(a)}$. Justify that h is differentiable on U and that $h(a) = h(b) = 0$.

(b) Apply Rolle's theorem to justify that there exists some $c \in (a, b)$ such that $h'(c) = 0$.

(c) Show that for the value c found in the previous part, it holds that $\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}$.

(14) For the following functions f, find all points x_0 in the domain of f such that $f(x_0)$ is an extremum. For each of these x_0, state whether or not x_0 is a critical point of f and classify whether $f(x_0)$ is a local maximum, local minimum, global maximum, global minimum, or none of these. Also, if the point x_0 is in the domain of f and $f(x_0)$ is an extremum even though x_0 is not a critical point, make sure to Make sure to prove your assertions.

(a) $f : (0, 1) \to \mathbb{R}$ is defined by $f(x) = x^2$.

(b) $f : (-1, 1) \to \mathbb{R}$ is defined by $f(x) = x^2$.

(c) $f : (0, 1) \to \mathbb{R}$ is defined by $f(x) = x(1 - x)$.

(d) $f : (-1, 1) \to \mathbb{R}$ is defined by $f(x) = x^3$.

(15) Suppose f is a real valued function defined on some set U and $x_0 \in U$. We say that f is differentiable from the right at x_0 when the right hand limit $\lim_{x \to x_0+} \frac{f(x)-f(x_0)}{x-x_0}$ exists; in this case, we write $f'(x_0+) = \lim_{x \to x_0+} \frac{f(x)-f(x_0)}{x-x_0}$. We say that f is differentiable from the left at x_0 when the lefthand limit $\lim_{x \to x_0-} \frac{f(x)-f(x_0)}{x-x_0}$ exists; in this case, we write $f'(x_0-) = \lim_{x \to x_0-} \frac{f(x)-f(x_0)}{x-x_0}$.

(a) Prove that if f is differentiable from the right/left at x_0, then it is continuous from the right/left at x_0.

(b) Prove that f is differentiable from the right at x_0 with $f'(x_0+) = \lambda$ if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ in U such that $x_n \downarrow x_0$, it holds that $\lim_{n \to \infty} \frac{f(x_n)-f(x_0)}{x_n-x_0} = \lambda$.

(c) Prove that f is differentiable from the left at x_0 with $f'(x_0-) = \lambda$ if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ in U such that $x_n \uparrow x_0$, it holds that $\lim_{n \to \infty} \frac{f(x_n)-f(x_0)}{x_n-x_0} = \lambda$.

(16) Suppose that f is a real valued function defined on $[a, b]$ (where $a, b \in \mathbb{R}$ with $a < b$). Let $\mathcal{P}$ be any partition of $[a, b]$. Assume that we had defined upper and lower sums $U(f, \mathcal{P})$ and $L(f, \mathcal{P})$ without the requirement that f is bounded. We will treat $+\infty$ and $-\infty$ as familiar objects such that for every $x \in \mathbb{R}$, $-\infty < x < \infty$, $x + \infty = \infty$, $x - \infty = -\infty$, and $\infty - (-\infty) = \infty$. Further suppose that we define the supremum of a set which is not bounded above as $+\infty$, and the infimum of a set not bounded below as $-\infty$.

(a) Prove that $U(f, \mathcal{P}) > -\infty$ and $L(f, \mathcal{P}) < +\infty$.

(b) If f is unbounded, then prove that either $U(f, \mathcal{P}) = +\infty$ or $L(f, \mathcal{P}) = -\infty$.

(c) Prove that if f is unbounded, then $U(f, \mathcal{P}) - L(f, \mathcal{P}) = \infty$.

(17) Suppose that f is a bounded real-valued function defined on $[a, b]$ (where $a, b \in \mathbb{R}$ with $a < b$). Let $\mathcal{P}$ be a partition of $[a, b]$. Prove that $L(f, \mathcal{P}) \leq U(f, \mathcal{P})$.

(18) We fill in the missing inductive piece of the proof of Proposition 5.2.12 in this problem. Suppose that f is a bounded real-valued function defined on $[a, b]$. Let $R > 0$ be the bound of f. If $\mathcal{Q} \subset \mathcal{P}$ are partitions of $[a, b]$ with $1 = \#(\mathcal{P} \setminus \mathcal{Q})$, prove that

$$U(f, \mathcal{Q}) - U(f, \mathcal{P}) \leq 2R \operatorname{mesh}(\mathcal{Q})$$

and

$$L(f,\mathcal{P}) - L(f,\mathcal{Q}) \leq 2R\,\mathrm{mesh}(\mathcal{Q}).$$

Explain why this finishes the proof that $T = \mathbb{N}$, where T is given in the proof of Proposition 5.2.12.

(19) For the following functions f defined on $[0,1]$ and partitions $\mathcal{P}$ of $[0,1]$, find $U(f,\mathcal{P})$ and $L(f,\mathcal{P})$.

(a) $f(x) = x,\ \mathcal{P} = \{0, 1/4, 3/4, 1\}$.

(b) $f(x) = x(1-x),\ \mathcal{P} = \{0, 1/4, 3/4, 1\}$.

(c) $f(x) = \begin{cases} 0 & x \in \mathbb{Q} \\ 1 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases},\ \mathcal{P} = \{0, x_1, x_2, ..., x_{n-1}, 1\}$.

(20) Let $a, b \in \mathbb{R}$ and suppose that f is a bounded real-valued function on the interval $[a,b]$. Prove that for any $\varepsilon > 0$ there exists a partition $\mathcal{P}$ such that both $U(f,\mathcal{P}) - \overline{\int_a^b} f < \varepsilon$ and $\underline{\int_a^b} f - L(f,\mathcal{P}) < \varepsilon$ hold.

(21) Prove the lower sum inequality of Proposition 5.2.7.

(22) Prove the lower sum inequality of Proposition 5.2.12.

(23) Let $a, b \in \mathbb{R}$ with $a < b$. Let f be a bounded real-valued function on $[a,b]$ which is Riemann integrable. Let $\varepsilon > 0$. Suppose that there exists some partition $\mathcal{Q}$ of $[a,b]$ such that $U(f,\mathcal{Q}) - L(f,\mathcal{Q}) < \varepsilon$. Prove that there exists a partition $\mathcal{P}$ of $[a,b]$ such that $U(f,\mathcal{P}) - \overline{\int_a^b} f < \varepsilon$, $\underline{\int_a^b} f - L(f,\mathcal{P}) < \varepsilon$, and $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$.

(24) Let f be a Riemann integrable function on some interval $[a,b]$. Define the absolute value of f, denoted $|f|$, by $|f|(x) = |f(x)|$ for every x in the domain of f. In this problem, we will prove that $|f|$ is also Riemann integrable and $\left|\int_a^b f\right| \leq \int_a^b |f|$.

(a) Suppose that $x_1, x_2 \in [a,b]$ with $x_1 < x_2$. Let $M = \sup\{f(x) \text{ s.t. } x \in [x_1, x_2]\}$ and $m = \inf\{f(x) \text{ s.t. } x \in [x_1, x_2]\}$. Similarly, let $N = \sup\{|f(x)| \text{ s.t. } x \in [x_1, x_2]\}$ and $n = \inf\{|f(x)| \text{ s.t. } x \in [x_1, x_2]\}$. Prove that $N - n \leq M - m$. *Hint*: In fact, if M and m have the same sign then $N - n = M - m$, otherwise $N - n \leq \max\{M, |m|\} < M - m$.

(b) Using the previous part, prove that if $\mathcal{P}$ is any partition of $[a,b]$, then $U(|f|,\mathcal{P}) - L(|f|,\mathcal{P}) \leq U(f,\mathcal{P}) - L(f,\mathcal{P})$.

(c) Use the previous part and Theorem 5.2.14 to prove that $|f|$ is Riemann integrable on $[a,b]$.

(d) Prove that if $\mathcal{P}$ is any partition of $[a,b]$ then $|U(f,\mathcal{P})| \leq U(|f|,\mathcal{P})$.

(e) Using the previous part, prove that $\left|\int_a^b f\right| \leq \int_a^b |f|$.

(25) (2 out of 3 Lemma for Riemann integration). Let f, g, and h all be real valued functions defined on the interval $[a, b]$ such that for all $x \in [a, b]$, $h(x) = f(x) + g(x)$. Prove that if any two of the three functions are Riemann integrable on $[a, b]$ then so is the third.

(26) In this problem we prove that if f is Riemann integrable on $[a, b]$, then so is f^2. From here, assume that f is Riemann integrable on $[a, b]$.

(a) First suppose that f is non-negative on the interval from $[a, b]$. Let $\mathcal{P} = \{x_0, x_1, ..., x_n\}$. Using the notation from Definition 5.2.5, justify that $M_i^{f^2} = (M_i^f)^2$ and $m_i^{f^2} = (m_i^f)^2$. *Note*: $M_i^{f^2} = \sup\left\{f^2(x) \text{ s.t. } x \in [x_{i-1}, x_i]\right\}$ and $m_i^{f^2} = \inf\left\{f^2(x) \text{ s.t. } x \in [x_{i-1}, x_i]\right\}$.

(b) Continuing from the previous part, show that

$$U(f^2, \mathcal{P}) - L(f^2, \mathcal{P}) = \sum_{i=1}^{n} \left[(M_i^f)^2 - (m_i^f)^2\right](x_i - x_{i-1}).$$

(c) Continuing from the previous two parts, if $R > 0$ is an upper bound of f, then show that

$$U(f^2, \mathcal{P}) - L(f^2, \mathcal{P}) \leq 2R \sum_{i=1}^{n} \left[M_i^f - m_i^f\right](x_i - x_{i-1})$$

Hint: Difference of squares.

(d) Continuing from the previous parts, prove that f^2 is Riemann integrable by showing that if $\{\mathcal{P}_n\}_{n=1}^{\infty}$ is a sequence of partitions of $[a, b]$ such that $\text{mesh}(\mathcal{P}_n) \to 0$, then

$$U(f^2, \mathcal{P}_n) - L(f^2, \mathcal{P}_n) \to 0.$$

(e) Prove the general case where we no longer assume that f is non-negative. This is quite easy if you look at Exercise 6 (24) since if f is Riemann integrable, then so is $|f|$ and you can use your previous work applied to $|f|$.

(27) Prove Proposition 5.2.18. That is, suppose that f is a real-valued function which is Riemann integrable on $[a, b]$. Prove that for any interval $[c, d] \subset [a, b]$ it holds that f is Riemann integrable on $[c, d]$.

(28) Let $a < b$ and let J be an interval with bounds a and b. That is $J = (a, b)$, $J = [a, b)$, $J = (a, b]$, or $J = [a, b]$. Define the indicator function of J, denoted $1_J : \mathbb{R} \to \mathbb{R}$, as

$$1_J(x) = \begin{cases} 1 & x \in J \\ 0 & \text{otherwise} \end{cases}$$

(a) Let $\mathcal{P}$ be any partition of $[a, b]$. Prove that $U(1_J, \mathcal{P}) = b - a$ and

$$U(1_J, \mathcal{P}) - L(1_J, \mathcal{P}) \leq 2\,\text{mesh}(\mathcal{P}).$$

(b) Prove that for any interval $[a', b']$ containing J, 1_J is Riemann integrable and $\int_a^b 1_J = b - a$.

(c) Suppose that f is a real-valued function which is Riemann integrable on $[a, b]$. Although it is not completely consistent with Definition 4.5.1, we consider the pointwise product $f1_J : \mathbb{R} \to \mathbb{R}$ (where we interpret this as $x \mapsto f(x)1_J(x)$ as usual, but isn't completely consistent since we are considering the domain to be $\mathbb{R}$ rather than the intersection of the corresponding domains – this is a common interpretation when using indicator function) defined by

$$f1_J(x) = \begin{cases} f(x) & x \in J \\ 0 & x \notin J \end{cases}$$

Prove that if $[a', b']$ contains J, then $f1_J$ is Riemann integrable on $[a', b']$ and $\int_{a'}^{b'} f1_J = \int_a^b f$. *Hint*: Argue similarly as the previous parts with the inequality $U(f1_J, \mathcal{P}) - L(f1_J, \mathcal{P}) \leq U(f, \mathcal{P}) - L(f, \mathcal{P}) + 2R\,\text{mesh}(\mathcal{P})$ where $\mathcal{P}$ is any partition of $[a, b]$ and $R = \sup\{f(x) : x \in [a, b]\} - \inf\{f(x) : x \in [a, b]\}$.

(29) Suppose that f is a polynomial defined on $[a, b]$ with $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ for some scalars $a_0, a_1, ..., a_n \in \mathbb{R}$. Prove that f is Riemann integrable on $[a, b]$ and that $\int_a^b f = \sum_{k=0}^{n} \frac{a_k}{k+1}\left(b^{k+1} - a^{k+1}\right)$.

(30) (Mean Value Theorem for Integrals) Suppose that f is continuous on $[a, b]$. Prove that there exists some $c \in (a, b)$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f.$$

(31) Prove Theorem 5.2.22. That is, suppose that f is a real-valued function which is Riemann integrable on some interval containing the points a, b, c. Prove that

$$\int_a^c f + \int_c^b f = \int_a^b f.$$

Note: This is quick, but not immediate from Theorem 5.2.19 since it could happen that $a < b < c$ or $c < a < b$.

(32) Suppose that U and V are open subsets of $\mathbb{R}$ and $f : U \to V$ is a bijective function such that f and f^{-1} are differentiable, f' is continuous and $f'(x) \neq 0$ for every $x \in U$, and f^{-1} is Riemann integrable on some interval $[a, b] \subset V$.

(a) For every $y \in V$, prove that $(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}$. That is, $(f^{-1})' = \frac{1}{f' \circ f^{-1}}$.

(b) Using a judicious integration-by-parts followed by u-substitution argument, show that

$$\begin{aligned}\int_a^b f^{-1}(y)\,dy &= yf^{-1}(y)\Big|_a^b - \int_a^b \frac{y}{f'(f^{-1}(y))}\,dy \\ &= yf^{-1}(y)\Big|_a^b - \int_{f^{-1}(a)}^{f^{-1}(b)} \frac{f(x)}{f'(x)}\,dx.\end{aligned}$$

(c) Make the same conclusion as derived in the previous part by showing that $F : [a, b] \to \mathbb{R}$ defined by

$$F(y) = yf^{-1}(y) - \int_a^y \frac{y}{f'(f^{-1}(y))}\,dy$$

is an antiderivative of f^{-1} on $[a, b]$.

(33) In any way you like, find the an antiderivative of the given function f on the given interval $[a, b]$, prove that the antiderivative you found is an antiderivative, and use it to calculate $\int_a^b f$.

(a) $[a, b] = [1, 2]$, $f(x) = x^p$ for $p \in \mathbb{Z}$ with $p \neq -1$.

(b) $[a, b] = [2/3, 1]$, $f(x) = g(3x-2)$ where $g : \mathbb{R} \to \mathbb{R}$ has antiderivative $G : \mathbb{R} \to \mathbb{R}$. You can leave everything in terms of G.

(c) $[a, b] = [0, \pi/4]$, $f(x) = \frac{1}{\cos(x)}$. You are welcome to assume the standard properties of sin and cos.

(34) Let f and g be Riemann integrable on some interval $[a, b]$. Finish the proof of Theorem 5.2.16 (1). That is, using what is already proved for Theorem 5.2.16 (1), prove that for any $\alpha, \beta \in \mathbb{R}$, $\int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g$.

(35) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x = \frac{1}{n} \text{ for some } n \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

Prove that f is Riemann integrable on $[0, 1]$ and that $\int_0^1 f = 0$.

(36) Suppose f is Riemann integrable on $[a, b]$ and that $\{\mathcal{P}_n\}_{n=1}^{\infty}$ is a sequence of partitions where $\text{mesh}(\mathcal{P}_n) \to 0$. For each n, write $\mathcal{P}_n = \{x_0^{(n)}, x_1^{(n)}, x_2^{(n)}, ..., x_{m(n)}^{(n)}\}$ and let $c_i^{(n)}$ be some point $c_i^{(n)} \in [x_{i-1}^{(n)}, x_i^{(n)}]$, for every i.

(a) By showing that

$$L(f, \mathcal{P}_n) \leq \sum_{i=1}^{m(n)} f(c_i^{(n)})(x_i - x_{i-1}) \leq U(f, \mathcal{P}_n)$$

holds for every n, deduce that

$$\lim_{n\to\infty} \sum_{i=1}^{m(n)} f(c_i^{(n)})(x_i - x_{i-1}) = \int_a^b f.$$

(b) For each $n \in \mathbb{N}$, let $\Delta x = \Delta x(n) = \frac{b-a}{n}$, and for each $i \in \{0, 1, ..., n\}$ let $c_i = c_i(n) = a + i\,\Delta x$. If f is Riemann integrable on $[a, b]$, justify the following

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(c_i)\Delta x = \lim_{n\to\infty} \sum_{i=1}^{n} f(c_i)\Delta x = \lim_{n\to\infty} \sum_{i=1}^{n} f\big(\frac{c_i + c_{i-1}}{2}\big)\Delta x = \int_a^b f.$$

What you should notice is that the first term is the (limit of) the lefthand sum, the second term is the (limit of) the righthand sum, and third term is the (limit of) the midpoint sum, all of which you are used to from elementary calculus.

(37) Find the following limits. *Hint*: These will look like left- or righthand sums when interpreted correctly.

(a) $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{n^2}$.

(b) $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{n^3}\big(n + k\big)^2$.

(38) Prove Lemma 5.3.9.

(39) Prove Lemma 5.3.13.

Part 3

Real Analysis II: Abstractions and functions of several real variables.

CHAPTER 6

Metric Spaces

1. The definition of a metric space and first examples

DEFINITION 6.1.1. Let M be a non-empty set and ρ be a function $\rho : M \times M \to \mathbb{R}$ such that the following properties hold:

M1: For all $x, y \in M$, $\rho(x, y) \geq 0$.

M2: For all $x, y \in M$, $\rho(x, y) = 0$ if and only if $x = y$.

M3: (Symmetry) For all $x, y \in M$, $\rho(x, y) = \rho(y, x)$.

M4: (Triangle Inequality) For all $x, y, z \in M$, $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$.

Then ρ is called a *metric* on M, and the pair (M, ρ) is called a *metric space*. It is common to say that M *endowed* with the metric ρ is a metric space.

REMARKS 6.1.2. 1) Property **M1** in the definition of a metric can be rewritten as $\rho(M \times M) \subset [0, \infty)$. That is, the image of ρ is a subset of the non-negative real numbers.

2) It is tedious to write out the pair (M, ρ) every time rather than just the symbol M, so it is common to simply say that M is a metric space when there is a metric understood in the background.

EXAMPLE 6.1.3. The map $\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $\rho(x, y) = |x - y|$ is a metric on $\mathbb{R}$. Notice that ρ is just the distance function. Hence $(\mathbb{R}, \rho)$ is a metric space. □

EXAMPLE 6.1.4. If (M, ρ) is a metric space and $N \subset M$ is any non-empty set. Then (N, ρ) is also a metric space. □

EXAMPLE 6.1.5. Let $S \subset \mathbb{R}$ be non-empty and compact. Define $C(S \to \mathbb{R})$ to be the collection of all continuous, real-valued functions with domain S. Then $C(S \to \mathbb{R})$ endowed with the metric

$$\rho(f, g) = \sup \{|f(x) - g(x)| \text{ s.t. } x \in S\}$$

is a metric space. □

2. Sequences on metric spaces

For the remainder of this chapter, we assume (M, ρ) is a metric space.

DEFINITION 6.2.1. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in M. We say that the sequence *converges* when there exists an $x \in M$ such that for all $\varepsilon > 0$ there exists an $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$ it holds that $\rho(x_n, x) < \varepsilon$.

In the case of convergence, we say that the sequence *converges to* x and call x the *limit* of the sequence. We commonly use any of the following notations for when $\{x_n\}_{n=1}^{\infty}$ converges to x: $\lim_{n\to\infty} x_n = x$, or $x_n \to x$ as $n \to \infty$, or simply $x_n \to x$. The last is just shorthand used when it is understood that we are taking $n \to \infty$.

THEOREM 6.2.2. *Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in M which converges, then the limit is unique.*

Proof. Suppose that $x_n \to x$ as $n \to \infty$ and $x_n \to y$ as $n \to \infty$. Take $\varepsilon > 0$ and find $n_x \in \mathbb{N}$ such that if $n \geq n_x$, then $\rho(x_n, x) < \varepsilon/2$; find $n_y \in \mathbb{N}$ such that if $n \geq n_y$, then $\rho(x_n, y) < \varepsilon/2$. Set $N = \max(n_x, n_y)$. Then, using the triangle inequality,

$$\rho(x, y) \leq \rho(x, x_N) + \rho(y, x_N) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

The last inequality follows from the fact that $N \geq n_x$ and $N \geq n_y$. Now, we have show that given any $\varepsilon > 0$, $\rho(x, y) < \varepsilon$, which implies that $\rho(x, y) = 0$. ■

DEFINITION 6.2.3. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in M. We say that the sequence is *Cauchy* (or is a *Cauchy sequence*) when for all $\varepsilon > 0$ there exists an $n_\varepsilon \in \mathbb{N}$ such that $n \geq n_\varepsilon$ and $m \geq n_\varepsilon$ implies $\rho(x_n, x_m) < \varepsilon$.

REMARK 6.2.4. Notice! The definition of convergence and Cauchy depend on ρ. It is conceivable that the same set has several different metrics such that with respect to one metric a sequence is convergent (or is Cauchy), but with respect to another, that same sequence does not converge (or is not Cauchy). □

THEOREM 6.2.5. *Let (M, ρ) be a metric space and $\{x_n\}_{n=1}^{\infty}$ a sequence in M. Then the following hold.*

(1) *If $\{x_n\}_{n=1}^{\infty}$ converges, it is Cauchy.*

(2) *Suppose* $\{x_n\}_{n=1}^{\infty}$ *is Cauchy and there exists a subsequence* $\{x_{n_k}\}_{k=1}^{\infty}$ *of* $\{x_n\}_{n=1}^{\infty}$ *and a point* $x \in M$ *such that* $\lim_{k\to\infty} x_{n_k} = x$. *Then* $\{x_n\}_{n=1}^{\infty}$ *converges and moreover,* $\lim_{n\to\infty} x_n = x$.

Proof. (1) If $\{x_n\}_{n=1}^{\infty}$ converges, let $x \in M$ be the limit of $\{x_n\}_{n=1}^{\infty}$. Take $\varepsilon > 0$ and find $n_\varepsilon \in \mathbb{N}$ such that for every $n \geq n_\varepsilon$, $\rho(x_n, x) < \varepsilon/2$. Then, for any $n, m \geq n_\varepsilon$, the triangle inequality yields $\rho(x_n, x_m) \leq \rho(x_n, x) + \rho(x, x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. This shows that $\{x_n\}_{n=1}^{\infty}$ are Cauchy.

(2) Let $\varepsilon > 0$. Let $k_\varepsilon \in \mathbb{N}$ be chosen so that for every $k \geq k_\varepsilon$, $\rho(x_{n_k}, x) < \varepsilon/2$. Let $n_\varepsilon \in \mathbb{N}$ be chosen so that for every $n, m \geq n_\varepsilon$, $\rho(x_n, x_m) < \varepsilon/2$. Set $K = \max\{k_\varepsilon, n_\varepsilon\}$ and take $n \geq K$. The triangle inequality yields $\rho(x_n, x) \leq \rho(x_n, x_{n_K}) + \rho(x_{n_K}, x)$. Now, since $K \geq k_\varepsilon$, $\rho(x_{n_K}, x) < \varepsilon/2$, and since $n_K \geq K \geq n_\varepsilon$ and $n \geq K \geq n_\varepsilon$, $\rho(x_n, x_{n_N}) < \varepsilon/2$. Therefore $\rho(x_n, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon$, proving that $x_n \to x$ as $n \to \infty$. ■

Definition 6.2.6. Let (M, ρ) be a metric space. We say that (M, ρ) is *complete* when: Given any Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ there exists an $x \in M$ such that $\lim_{n\to\infty} x_n = x$. Loosely speaking, a metric space is complete when *all Cauchy sequences converge.*

3. Open and Closed Sets

Definition 6.3.1. Let $r > 0$ and $x \in M$. The *open ball of radius* r *centered at* x, denoted $B(x; r)$, is the subset of M defined by

$$B(x; r) = \{y \in M \text{ s.t. } \rho(x, y) < r\}.$$

Similarly, the *closed ball of radius* r *centered at* x, denoted $\overline{B}(x; r)$, is the subset of M defined by

$$\overline{B}(x; r) = \{y \in M \text{ s.t. } \rho(x, y) \leq r\}.$$

Definitions 6.3.2. A subset $U \subset M$ is called *open* when given any $x \in U$, there is an $\varepsilon > 0$ such that $B(x; \varepsilon) \subset U$. A subset $C \subset M$ is called *closed* when the complement set $M \setminus C$ is open.

Proposition 6.3.3 (Common Open Sets). *The following hold.*

(1) M *and* $\varnothing$ *are open sets.*

(2) *If* $r > 0$ *and* $x \in M$, *then* $B(x; r)$ *is open.*

(3) *If A is any indexing set and $\{U_a\}_{a\in A}$ is a collection of open sets, then $\bigcup_{a\in A} U_a$ is an open set. Poetically, the union of arbitrarily many open sets is open.*

(4) *If B is a* finite *indexing set and $\{U_b\}_{b\in B}$ is a collection of open sets, then $\bigcap_{b\in B} U_b$ is open. Poetically, the intersection of finitely many open sets is open.*

Proof. (1) Let $x \in M$ and $r > 0$. By definition, $B(x;r) \subset M$ and hence M is open. On the other hand, since there is no x in $\varnothing$ to check, the open property is vacuously true.

(2) Let $y \in B(x;r)$ and $s = \rho(y,x)$. Set $\varepsilon = r - s$ and we will now check that $B(y;\varepsilon) \subset B(x;r)$. To do so, take $z \in B(y;\varepsilon)$, then by the triangle inequality

$$\rho(x,z) \leq \rho(x,y) + \rho(y,z) = s + \rho(y,z) < s + (r-s) = r.$$

Therefore $z \in B(x;r)$ showing that $B(y;\varepsilon) \subset B(x;r)$ and hence confirming that $B(x;r)$ is open.

(3) Let $U = \cup_{a\in A} U_a$. If $U = \varnothing$ then we're done by (1). Otherwise, suppose that $x \in U$, in which case there is some $a \in A$ such that $x \in U_a$. Since U_a is open, let $\varepsilon > 0$ such that $B(x;\varepsilon) \subset U_a$. Hence $B(x;\varepsilon) \subset U$ since $U_a \subset U$, and we have hence proved U is open.

(4) Let $V = \cap_{b\in B} U_b$. If $V = \varnothing$, then like before, we're done. If not, let $x \in V$, in which case $x \in U_b$ for every $b \in B$. Since for every $b \in B$, U_b is open, let $\varepsilon_b > 0$ be chosen such that $B(x;\varepsilon_b) \subset U_b$. Set $\varepsilon = \min\{\varepsilon_b \text{ s.t. } b \in B\}$. With B a finite set, and each $\varepsilon_b > 0$, we have that $\varepsilon > 0$. Moreover, ε was chosen such that $B(x;\varepsilon) \subset B(x;\varepsilon_b) \subset U_b$ for every $b \in B$, and hence $B(x;\varepsilon) \subset U$. We have therefore proved that U is open. ■

PROPOSITION 6.3.4 (Common Closed Sets). *The following hold.*

(1) *M and $\varnothing$ are closed sets.*

(2) *If B is a* finite *indexing set and $\{C_b\}_{b\in B}$ is a collection of closed sets, then $\cup_{b\in B} C_b$ is a closed set.*

(3) *If A is any indexing set and $\{C_a\}_{a\in A}$ is a collection of closed sets, then $\cap_{a\in A} C_a$ is closed.*

Proof. Exercises 9 (9). ■

4. Further Set Classifications

4.1. Induced Sets. Given any subset $U \subset M$, there are several subsets of M which are induced from U which help us understand the structure of U.

DEFINITIONS 6.4.1. Let $U \subset M$ and let U^c denote the complement $U^c = M \setminus U$.

(1) The *interior* of U is the set

$$\text{int}(U) = \{x \in U \text{ s.t. there exists some open set } V \text{ with } x \in V \text{ and } V \subset U\}$$

A point $x \in \text{int}(U)$ is called an *interior point of* U. It is common in literature to denote the interior of U by U^o.

(2) The *exterior* of U is the set

$$\text{ext}(U) = \{x \in U^c \text{ s.t. there exists some open set } V \text{ with } x \in V \text{ and } V \subset U^c\}$$

A point $x \in \text{ext}(U)$ is called an *exterior point of* U. Notice that $\text{ext}(U) = \text{int}(U^c)$.

(3) The *boundary* of U is the set

$$\text{bdry}(U) = \{x \in M \text{ s.t. for every } \varepsilon > 0,\ B(x;\varepsilon) \cap U \neq \varnothing \text{ and } B(x;\varepsilon) \cap U^c \neq \varnothing\}$$

A point $x \in \text{bdry}(U)$ is called a *boundary point of* U. It is common in literature to denote the boundary of U by ∂U.

(4) The *limit points* or *accumulation points* of U is the set

$$\text{limpts}(U) = \{x \in M \text{ s.t. for every } \varepsilon > 0,\ [B(x;\varepsilon) \setminus \{x\}] \cap U \neq \varnothing\}$$

A point $x \in \text{limpts}(U)$ is called a *limit point* or *accumulation point* of U.

(5) The *closure* of U is the set $\text{cl}(U) = U \cup \text{bdry}(U)$. It is common in literature to denote the closure of U by $\overline{U}$.

REMARK 6.4.2. By the definitions of $\text{int}(U)$ and $\text{cl}(U)$, you have $\text{int}(U) \subset U$ and $U \subset \text{cl}(U)$. More concisely, $\text{int}(U) \subset U \subset \text{cl}(U)$. □

Let's make an observation here. From our definitions, it intuitively seems that the interior of a set is open and the closure of a set is closed. This is in fact true.

LEMMA 6.4.3. *For any set* $U \subset M$, $\text{int}(U)$ *is an open subset of* M *and* $\text{cl}(U)$ *is a closed subset of* M.

Proof. If $\text{int}(U) = \varnothing$, we're done. So, assume $\text{int}(U)$ is non-empty and take $x \in \text{int}(U)$. By definition, there is some $\varepsilon > 0$ such that $B(x;\varepsilon) \subset U$. Now, let's show that $B(x;\varepsilon) \subset \text{int}(U)$, which will imply that $\text{int}(U)$ is open and hence conclude the proof. Let $y \in B(x;\varepsilon)$,

and if we can show that $y \in \text{int}(U)$ we'll be done. Since $B(x;\varepsilon)$ is open, there is some $\varepsilon_2 > 0$ such that $B(y;\varepsilon_2) \subset B(x;\varepsilon)$. With $B(x;\varepsilon) \subset U$, we have $B(y;\varepsilon_2) \subset U$, showing that $y \in \text{int}(U)$.

If $\text{cl}(U) = M$, we're done. Otherwise, to show that $\text{cl}(U)$ is closed, suppose $x \in M \setminus \text{cl}(U)$. Then $x \in M \setminus U$ and x is not a boundary point of U. Therefore, there is some $\varepsilon > 0$ such that $B(x;\varepsilon) \subset M \setminus U$ (if there was no such ε, then x would be a boundary point of U!). We will now show that $B(x;\varepsilon) \subset M \setminus \text{cl}(U)$, implying that $M \setminus \text{cl}(U)$ is open and hence $\text{cl}(U)$ is closed. To this end, let $y \in B(x;\varepsilon)$. Since $B(x;\varepsilon)$ is open, there is some $\varepsilon_2 > 0$ such that $B(y;\varepsilon_2) \subset B(x;\varepsilon)$ and hence $B(y;\varepsilon_2) \in U^c$. Particularly, $y \in U^c$ and $y \notin M \setminus U$, which means $y \in M \setminus \text{cl}(U)$. Therefore $B(x;\varepsilon) \subset M \setminus \text{cl}(U)$, and we're done. ■

Let's also justify the name "limit points" with the following lemma.

LEMMA 6.4.4. *Let $U \subset M$. Then $x \in \text{limpts}(U)$ if and only if there is a sequence $\{x_n\}_{n=1}^{\infty}$ in $U \setminus \{x\}$ such that $\lim_{n\to\infty} x_n = x$.*

Proof. First assume that $x \in \text{limpts}(U)$. Then for every $n \in \mathbb{N}$, $[B(x;1/n)\setminus\{x\}]\cap U \neq \varnothing$. In particular, for every $n \in \mathbb{N}$, we can find some $x_n \in [B(x;1/n) \setminus \{x\}] \cap U$. This lets us define a sequence $\{x_n\}_{n=1}^{\infty}$ in $U \setminus \{x\}$ such that for every $n \in \mathbb{N}$, it holds that $\rho(x_n, x) < 1/n$. From Exercise 9 (5), this implies that $\lim_{n\to\infty} x_n = x$.

Conversely, assume that $x \in M$ and there is a sequence $\{x_n\}_{n=1}^{\infty}$ in $U \setminus \{x\}$ converging to x. Then by the definition of convergence, for any $\varepsilon > 0$, we can find some $n_\varepsilon \in \mathbb{N}$ such that $\rho(x_{n_\varepsilon}, x) < \varepsilon$. This implies that $x_{n_\varepsilon} \in [B(x;\varepsilon) \setminus \{x\}] \cap U$ since $x_{n_\varepsilon} \in U \setminus \{x\}$ by assumption, and $\rho(x_n, x) < \varepsilon$. This shows that for any $\varepsilon > 0$ it must hold that $[B(x;\varepsilon) \setminus \{x\}] \cap U \neq \varnothing$ since $x_{n_\varepsilon} \in [B(x;\varepsilon) \setminus \{x\}] \cap U$. Therefore $x \in \text{limpts}(U)$. ■

PROPOSITION 6.4.5. *Let $U \subset M$. Then*

(1) $\text{int}(U)$ *is the union of all open sets V such that $V \subset U$.*
(2) $\text{cl}(U)$ *is the intersection of all closed sets C such that $U \subset C$.*
(3) *U is open if and only if $U = \text{int}(U)$.*
(4) *U is closed if and only if $U = \text{cl}(U)$.*
(5) $\text{cl}(U) = \text{int}(U) \cup \text{bdry}(U)$.

Proof. (1) Let $\Gamma_o = \{V \subset M \text{ s.t. } V \text{ is open and } V \subset U\}$. We want to prove that $\text{int}(U) = \bigcup_{V\in\Gamma_o} V$. Let $U^o = \bigcup_{V\in\Gamma_o} V$. By Lemma 6.4.3, $\text{int}(U)$ is open, and moreover,

$\text{int}(U) \subset U$. Hence $\text{int}(U) \subset U^o$. It remains to show that $U^o \subset \text{int}(U)$. Since the set U^o is a union of open sets, it is open. Moreover, since for each $V \in \Gamma_o$, it is true that $V \subset U$, we deduce that $U^o \subset U$. From here, take $x \in U^o$. Since U^o is open, there is an $\varepsilon > 0$ such that $B(x;\varepsilon) \subset U^o$, implying that $B(x;\varepsilon) \subset U$ since $U^o \subset U$. Therefore $x \in \text{int}(U)$; since $x \in U^o$ was arbitrary, this proves that $U^o \subset \text{int}(U)$, concluding the proof.

(2) Let $\Gamma_c = \{C \subset M \text{ s.t. } C \text{ is closed and } U \subset C\}$ as above. We want to prove that $\text{cl}(U) = \bigcap_{C \in \Gamma_c} C$. Let $\overline{U} = \bigcap_{C \in \Gamma_c} C$. By Lemma 6.4.3, $\text{cl}(U)$ is closed and, as noted, $U \subset \text{cl}(U)$. Hence $\text{cl}(U) \in \Gamma_c$ and therefore $\overline{U} \subset \text{cl}(U)$. It remains to prove that $\text{cl}(U) \subset \overline{U}$. Let $C \in \Gamma_c$ be any closed set such that $U \subset C$. Our aim is to show that this implies $\text{bdry}(U) \subset C$. To this end we argue by contrapositive. Let $C^c = M \setminus C$ and suppose that $x \in C^c$. Since C^c is open, there is an $\varepsilon > 0$ such that $B(x;\varepsilon) \subset C^c$. But, since $U \subset C$, this means that $B(x;\varepsilon) \subset U^c$ and therefore, $x \notin \text{bdry}(U)$. By the contrapositive argument, this proves that $\text{bdry}(U) \subset C$. With $U \subset C$ by assumption and having just shown $\text{bdry}(U) \subset C$, we have $\text{cl}(U) = U \cup \text{bdry}(U) \subset C$. Since C was any arbitrary closed set containing U, this means that *all* closed sets containing U also contains $\text{cl}(U)$. Therefore $\text{cl}(U) \subset \bigcap_{C \in \Gamma_c} C = \overline{U}$, finishing the proof.

(3) Suppose U is open. If $x \in U$, then there is some $\varepsilon > 0$ such that $B(x;\varepsilon) \subset U$, hence $x \in \text{int}(U)$. This implies that $U \subset \text{int}(U)$. By definition $\text{int}(U) \subset U$ and hence $U = \text{int}(U)$. Conversely, suppose $\text{int}(U) = U$. Since we have proved that $\text{int}(U)$ is open, this implies immediately that U is open, concluding the proof.

(4) In the proof of (2), we learned that every closed set containing U also contains $\text{cl}(U)$. This means that if U is closed then $\text{cl}(U) \subset U$, since U is a closed set containing U. However, $U \subset \text{cl}(U)$ and therefore $U = \text{cl}(U)$. Conversely, suppose $\text{cl}(U) = U$. Since we've proved that $\text{cl}(U)$ is closed, this immediately implies that U is closed.

(5) By definition, $\text{cl}(U) = U \cup \text{bdry}(U)$. So, to show that $\text{cl}(U) = \text{int}(U) \cup \text{bdry}(U)$, it suffices to show that if $x \in U \setminus \text{int}(U)$, then $x \in \text{bdry}(U)$. Suppose $x \in U \setminus \text{int}(U)$. Let $\varepsilon > 0$ be arbitrary. Since $x \in B(x;\varepsilon)$ and $x \in U$, we have $U \cap B(x;\varepsilon) \neq \varnothing$. On the other hand, since $x \notin \text{int}(U)$, it must be that there is some point $y \in U^c$ such the $y \in B(x;\varepsilon)$. This implies that $B(x;\varepsilon) \cap U^c \neq \varnothing$. Hence $x \in \text{bdry}(U)$, concluding the proof. ∎

4.2. Density.

DEFINITION 6.4.6. A subset $U \subset M$ is said to be *dense* in M when $\text{cl}(U) = M$.

LEMMA 6.4.7. *For any subset $U \subset M$, U is dense in* $\mathrm{cl}(U)$.

Proof. This is immediate from the definition of density. ■

REMARK 6.4.8. At first glance you might protest that we have not defined what it means for U to be dense in $\mathrm{cl}(U)$, but we have! Remember that $\mathrm{cl}(U) \subset M$, so $(\mathrm{cl}(U), \rho)$ is again a metric space, and we know what it means for U to be dense in the metric space $\mathrm{cl}(U)$. □

THEOREM 6.4.9. *$U \subset M$. The following are equivalent.*

(1) *U is dense in M.*
(2) *For all $x \in M$, there is a sequence $\{x_n\}_{n=1}^{\infty}$ of elements in U such that $\lim_{n\to\infty} x_n = x$.*
(3) *Given any non-empty open subset $V \subset M$, $V \cap U \neq \varnothing$.*

Proof. (1) $\implies$ (2): If U is dense in M, then $\mathrm{cl}(U) = M$. Now, using Exercises 9 (11), the conclusion follows.

(2) $\implies$ (3): Let V be a non-empty open subset of M. Let $x \in V$ and, by assumption, let $\{x_n\}_{n=1}^{\infty}$ be a sequence of elements in U converging to x. Since V is open, let $\varepsilon > 0$ such that $B(x;\varepsilon) \subset V$. Since $x_n \to x$, there exists an $n \in \mathbb{N}$ such that $\rho(x_n, x) < \varepsilon$. In different notation this says that $x_n \in B(x;\varepsilon)$, implying that $x_n \in V$. Since also $x_n \in U$, we see that $U \cap V \neq \varnothing$.

(3) $\implies$ (1): Let $V = M \setminus \mathrm{cl}(U)$, which is open since $\mathrm{cl}(U)$ is closed. Since $U \subset \mathrm{cl}(U)$, then $V \subset M \setminus U$, implying that $V \cap U = \varnothing$. By assuming (3), since V is an open set such that $V \cap U = \varnothing$, we deduce that $V = \varnothing$. Therefore $M \setminus \mathrm{cl}(U) = V = \varnothing$ implying that $\mathrm{cl}(U) = M$, proving that U is dense. ■

5. More on Closed Sets

Closed sets tend to be particularly nice analytically. For this reason, we dedicate this section to exploring two theorems illuminates several of these useful properties.

THEOREM 6.5.1. *Let $C \subset M$. The following are equivalent.*

(1) *C is closed.*
(2) $C = \mathrm{cl}(C)$.
(3) $\mathrm{bdry}(C) \subset C$.
(4) $\mathrm{limpts}(C) \subset C$.

(5) *Given any sequence* $\{x_n\}_{n=1}^{\infty}$ *of elements in* C *which converges to a point in* M*, then that point is in* C*. That is, if* $\{x_n\}_{n=1}^{\infty}$ *is any sequence in* C *and there exists some* $x \in M$ *such that* $x_n \to x$*, then* $x \in C$*.*

Proof. (1) $\implies$ (2): By definition of closure, we immediately have $C \subset \text{cl}(C)$. So, we only need to show that $\text{cl}(C) \subset C$. However, in Proposition 6.4.5 (2), we learned that $\text{cl}(C)$ is the intersection of all closed sets containing C. By the definition of intersections, this implies that $\text{cl}(C)$ is a subset of all closed sets containing C. If C is closed, then C is clearly a closed set containing C, and therefore $\text{cl}(C) \subset C$.

(2) $\implies$ (3): Since $\text{cl}(C) = C \cup \text{bdry}(C)$, then $\text{bdry}(C) \subset \text{cl}(C)$. So, if $C = \text{cl}(C)$, then $\text{bdry}(C) \subset C$.

(3) $\implies$ (4): If $\text{limpts}(C) = \varnothing$, then there is nothing to prove. Else, suppose $x \in \text{limpts}(C)$. We will use a contradiction argument to show that $x \in C$. If $x \notin C$, then let $\{x_n\}_{n=1}^{\infty}$ be a sequence in C such that $\lim\limits_{n\to\infty} x_n = x$; we are using Lemma 6.4.4 here. Now, for any $\varepsilon > 0$, there is some $n_\varepsilon \in \mathbb{N}$ such that if $n \geq n_\varepsilon$, then $x_n \in B(x;\varepsilon)$ (notice that this is equivalent to saying that if $n \geq n_\varepsilon$, $\rho(x_n, x) < \varepsilon$). Therefore, $B(x;\varepsilon) \cap C \neq \varnothing$. However, since $x \notin C$, $B(x;\varepsilon) \cap C^c \neq \varnothing$. This shows that $x \in \text{bdry}(C)$. Yet, by assuming (3), we've assumed that $\text{bdry}(C) \subset C$, implying that $x \in C$. This contradicts that we chose $x \notin C$ and therefore shows that $x \in \text{limpts}(C) \implies x \in C$. Thus we have proved that assuming (3) implies $\text{limpts}(C) \subset C$.

(4) $\implies$ (5): Suppose that $x \in M$ and that there exists a sequence $\{x_n\}_{n=1}^{\infty}$ in C which converges to x. We want to show that $x \in C$. Let's again argue by contradiction. If $x \notin C$, then $\{x_n\}_{n=1}^{\infty}$ is a sequence in $C \backslash \{x\} = C$ such that $\lim\limits_{n\to\infty} x_n = x$. By Lemma 6.4.4, this means that $x \in \text{limpts}(C)$. However, we assumed that $\text{limpts}(C) \subset C$, and therefore $x \in C$ even though we assumed that $x \notin C$. This contradiction shows that $x \in C$ to start with.

(5) $\implies$ (1): Let C^c denote $M \setminus C$. If $C^c = \varnothing$ then we're done, since $\varnothing$ is open. Now, suppose that $C^c \neq \varnothing$. Our goal is to prove that C is closed by showing that C^c is open. Take $x \in C^c$. Since we are assuming (5), our goal is to show that since there is no sequence $\{x_n\}_{n=1}^{\infty}$ in C such that $\lim\limits_{n\to\infty} x_n = x \in C^c$, there must be some $\varepsilon > 0$ such that $B(x;\varepsilon) \subset C^c$. We will argue the contrapositive and show that if there is no such $\varepsilon > 0$, then there is a sequence $\{x_n\}_{n=1}^{\infty}$ in C which converges to x. To this end, suppose there is no $\varepsilon > 0$ such that $B(x;\varepsilon) \subset C^c$. Then, for each $n \in \mathbb{N}$, let $x_n \in C$ be taken such that $x_n \in B(x;1/n)$.

Then $\{x_n\}_{n=1}^{\infty}$ is a sequence in C and $\rho(x_n, x) < 1/n$, implying that $\lim_{n\to\infty} x_n = x$. What this shows is that since there is no sequence $\{x_n\}_{n=1}^{\infty}$ in C converging to $x \in C^c$ (this is where we are using assumption (5)), there must be an $\varepsilon > 0$ such that $B(x;\varepsilon) \subset C^c$. This proves C^c is open, and hence C is closed, concluding the proof. ■

REMARK 6.5.2. Some of the most interesting and exciting results where closed sets are involved often result from the equivalence (1) $\iff$ (5) in Theorem 6.5.1. □

THEOREM 6.5.3. *Let (M, ρ) be a complete metric space and $C \subset M$. C is a closed subset of M if and only if (C, ρ) is a complete metric space.*

Proof. First assume that C is closed. Let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence in (C, ρ). Then, since $C \subset M$, $\{x_n\}_{n=1}^{\infty}$ is also a Cauchy sequence in M (since we are using the same metric ρ on M). Since M is complete, there is some $x \in M$ such that $\lim_{n\to\infty} x_n = x$. Since C is closed, then, by the (1) $\iff$ (5) equivalence of Theorem 6.5.1, it holds that $x \in C$. Thus, since $\{x_n\}_{n=1}^{\infty}$ was an arbitrarily chosen Cauchy sequence in C which converges to a point in C, we conclude that (C, ρ) is complete.

Conversely, suppose that (C, ρ) is complete. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in C which converges in M to some point x. Since $\{x_n\}_{n=1}^{\infty}$ converges in M, then $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence in (M, ρ). This implies that $\{x_n\}_{n=1}^{\infty}$ is Cauchy in C as well (since we're using the same metric ρ on C), and therefore, since (C, ρ) is complete and $\{x_n\}_{n=1}^{\infty}$ is Cauchy in C, this sequence converges in C, implying that $x \in C$. By the (1) $\iff$ (5) equivalence of Theorem 6.5.1, this proves that C is closed. ■

6. Compactness

DEFINITIONS 6.6.1. Let $U \subset M$. An *open cover* of U is an indexed collection of open subsets of M, say $\{U_a \text{ s.t. } a \in A\}$, such that $U \subset \bigcup_{a\in A} U_a$. If it happens that $\{U_a \text{ s.t. } a \in A\}$ is an open cover of U and B is a finite subset of A such that $U \subset \bigcup_{b\in B} U_b$, then we say that $\{U_a\}_{a\in A}$ *admits a finite subcover* of U.

EXAMPLE 6.6.2. Let $M = \mathbb{R}^n$ for some n and suppose that $x \in M$ and $\varepsilon > 0$. Define $U = B(x;\varepsilon)$. For every $n \in \mathbb{N}$ let $\varepsilon_n = (1 - 1/n)\varepsilon$. Define the set $U_n = B(x;\varepsilon_n)$. We then have that $\{U_n\}_{n=1}^{\infty}$ is an open cover of U since $U \subset \bigcup_{n=1}^{\infty} U_n$. However, the open cover $\{U_n\}_{n=1}^{\infty}$ does *not* admit a finite subcover of U. We explore this more in Exercise 22. □

DEFINITION 6.6.3. A subset $U \subset M$ is called *compact* when every open cover of U admits a finite sub-cover of U.

EXAMPLE 6.6.4. Let U be as in Example 6.6.2. Since there is an open cover of U which does not admit a finite subcover, we conclude that U is not compact. We explore this more in Exercise 22. □

REMARK 6.6.5. It is perfectly possible that the entire metric space M is compact. In such a case we typically say that M is a *compact metric space*. □

In $\mathbb{R}$ (using the standard distance metric), the Heine-Borel theorem says that a subset $K \subset \mathbb{R}$ is compact if and only if it is closed and bounded. We aren't so lucky as to get the Heine-Borel theorem exactly as stated for general metric spaces, but we do get something close. First some definitions.

DEFINITIONS 6.6.6. Let $U \subset M$. We say that U is *bounded* when there exists an $x \in U$ and an $R > 0$ such that $U \subset B(x; R)$. Further, we say that U is *totally bounded* when for all $\varepsilon > 0$, there exist finitely many points $\{x_1, x_2, ..., x_{n_\varepsilon}\} \subset U$ such that $\{B(x_i; \varepsilon) \text{ s.t. } 1 \leq i \leq n_\varepsilon\}$ is an open cover of U.

EXAMPLE 6.6.7. Let $T > 0$. Consider the interval $(0, T) \subset \mathbb{R}$. This interval is totally bounded (using the standard distance metric on $\mathbb{R}$). To prove this, let $\varepsilon > 0$ be arbitrary. We need to find finitely many points $\{x_1, x_2, ..., x_n\}$ in $(0, T)$ such that

$$(0, T) \subset (x_1 - \varepsilon, x_1 + \varepsilon) \cup (x_2 - \varepsilon, x_2 + \varepsilon) \cup \cdots \cup (x_n - \varepsilon, x_n + \varepsilon).$$

To this end, let $N \in \mathbb{N}$ be large enough that $T/N < \varepsilon$. For every $k \in \{0, 1, 2, ..., N\}$ define $x_k = k(T/N)$. This gives us finitely many points $\{x_0, x_1, ..., x_N\}$ such that $|x_{k+1} - x_k| = T/N < \varepsilon$ for every k. In particular, $(x_k - \varepsilon, x_k + \varepsilon) \supset (x_k - T/N, x_k + T/N) = (x_{k-1}, x_{k+1})$ for each $k \in \{1, 2, ..., N - 1\}$. We now claim that $\{x_1, x_2, ..., x_{N-1}\}$ are a collection of finitely many points in U giving us the desired result. Note that we excluded the points $x_0 = 0$ and $x_N = T$ since we want the points to be elements of $(0, T)$. We have

$$\begin{aligned}(0, T) &= (x_0, x_2) \cup (x_1, x_3) \cup \cdots \cup (x_{N-2}, x_N) \\ &\subset (x_1 - \varepsilon, x_1 + \varepsilon) \cup (x_2 - \varepsilon, x_2 + \varepsilon) \cup \cdots \cup (x_{N-1} - \varepsilon, x_{N-1} + \varepsilon)\end{aligned}$$

This proves that $(0, T)$ is totally bounded. □

LEMMA 6.6.8. *Suppose that M is totally bounded and $U \subset M$. Then U is totally bounded.*

Proof. Assume $U \neq \varnothing$, otherwise there is nothing to prove. Take $\varepsilon > 0$. Let $x_1, x_2, ..., x_n \in M$ such that $\{B(x_1;\varepsilon/2), B(x_2;\varepsilon/2), ..., B(x_n;\varepsilon/2)\}$ covers M. In particular, we then have that $U \subset \cup_{i=1}^{n} B(x_i;\varepsilon/2)$. Let us restrict ourselves to the subset

$$\{B(x_{n_1};\varepsilon/2), B(x_{n_2};\varepsilon/2), ..., B(x_{n_m};\varepsilon/2)\} \subset \{B(x_1;\varepsilon/2), B(x_2;\varepsilon/2), ..., B(x_n;\varepsilon/2)\}$$

of only those balls which intersect U. Then for every $1 \leq i \leq m$, $B(x_{n_i};\varepsilon/2) \cap U \neq \varnothing$ and $U \subset \cup_{i=1}^{m} B(x_{n_i};\varepsilon/2)$. Find points $\{u_1, u_2, ..., u_m\}$ by $u_i \in B(x_{n_i};\varepsilon/2) \cap U$ for every $1 \leq i \leq m$. To finish the proof we will show that $\{B(u_1;\varepsilon), B(u_2;\varepsilon), ..., B(u_m;\varepsilon)\}$ is a cover of U. To prove this, let $u \in U$ and let $i \in \{i, 2, ..., m\}$ such that $u \in B(x_{n_i};\varepsilon/2)$. Now, $\rho(u_i, u) \leq \rho(u_i, x_{n_i}) + \rho(x_{n_i}, u) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. This proves that $u \in B(u_i;\varepsilon)$, and therefore $U \subset \cup_{i=1}^{m} B(u_i;\varepsilon)$. ■

THEOREM 6.6.9 (Metric space version of the Heine-Borel theorem). *Let $K \subset M$. Then K is compact if and only if it is complete and totally bounded.*

Proof. A more general theorem is proved below (see Theorem 6.6.11), and therefore we do not prove this here. ■

DEFINITION 6.6.10. Given a subset $U \subset M$, we say that U is *sequentially compact* when: Given any sequence $\{x_n\}_{n=1}^{\infty}$ of elements in U, there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ of $\{x_n\}_{n=1}^{\infty}$ and a point $x \in U$ such that $\lim_{k\to\infty} x_{n_k} = x$.

THEOREM 6.6.11. *The following are equivalent.*

(1) *(M, ρ) is a compact metric space.*

(2) *(M, ρ) is complete and totally bounded.*

(3) *(M, ρ) is sequentially compact.*

(4) *For all subsets $U \subset M$ with infinitely many elements,* $\text{limpts}(U) \neq \varnothing$.

Proof. Here will be our strategy. We start by proving (2) $\implies$ (4) $\implies$ (3) $\implies$ (2). Once this is done, we will have proved that (2), (3), and (4) are equivalent conditions. We next will show that 1 is also equivalent by showing that (1) $\implies$ (4) and that (2) + (3) $\implies$ (1).

(2) $\implies$ (4): Let $U \subset M$ which contains infinitely many elements. Since we are assuming that M is totally bounded, there exist finitely many open balls of radius 1 which

covers M. Then these balls also cover U, and since U has infinitely many elements, there is at least one point $x_1 \in M$ such that $B(x_1; 1)$ contains infinitely many elements of U. Let $U_1 = B(x_1; 1) \cap U$. Now, replacing U with U_1 in this argument, we can find some point $x_2 \in M$ such that $B(x_2; 1/2)$ contains infinitely many elements of U_1 and define $U_2 = B(x_2, 1/2) \cap U_1$. We next find $x_3 \in M$ such that $B(x_3; 1/4)$ has infinitely many elements of U_2, and define $U_3 = B(x_3; 1/4) \cap U_2$. Continuing on in this fashion, for each $n \in \mathbb{N}$, we find $x_{n+1} \in B(x_n; 1/2^{n-1})$ such that $B(x_{n+1}; 1/2^n)$ contains infinitely many elements of U, and define $U_{n+1} = B(x_{n+1}; 1/2^n) \cap U_n$. Note here that by construction, there is some point $u \in U$ such that $u \in B(x_n; 1/2^{n-1}) \cap B(x_{n+1}; 1/2^n)$; hence by the triangle inequality $\rho(x_{n+1}, x_n) \leq \rho(x_n, u) + \rho(u, x_{n+1}) < 2^{-n+1} + 2^{-n} < 2^{-n+2}$. Now, we build the sequence $\{u_n\}_{n=1}^{\infty}$ as follows. Let $u_1 \in U_1$; take $u_2 \in U_2 \backslash \{u_1\}$, which we can do since U_2 has infinitely many elements; take $u_3 \in U_3 \backslash \{u_1, u_2\}$, which we can do since U_3 has infinitely many elements; continuing on, for every $n \in \mathbb{N}$, take $u_n \in U_n \backslash \{u_1, u_2, ..., u_{n-1}\}$, which we can do since U_n has infinitely many elements. Now, for $n, m \in \mathbb{N}$ with $m < n$, repeated use of the triangle inequality shows

$$\begin{aligned}
\rho(u_n, u_m) &\leq \rho(u_m, x_m) + \rho(x_m, x_{m+1}) + \cdots + \rho(x_n, u_n) \\
&< 2^{-m+1} + 2^{-m+2} + \cdots + 2^{-n+1} \\
&< 2^{-m+3} \sum_{k=0}^{n-m} 2^{-k} \\
&< 2^{-m+3} \sum_{k=0}^{\infty} 2^{-k} \\
&= 2^{-m+4}.
\end{aligned}$$

Since $2^{-m+4} \to 0$ as $m \to \infty$, we can conclude that the sequence $\{u_n\}_{n=1}^{\infty}$ is Cauchy. Indeed, let $\varepsilon > 0$ and find $n_\varepsilon \in \mathbb{N}$ such that $2^{-n_\varepsilon+4} < \varepsilon$. Then, by what we have shown above, if $n, m \geq n_\varepsilon$, then $\rho(u_n, u_m) < 2^{-n_\varepsilon+4} < \varepsilon$. Now, using the assumption that M is complete, we know that there is a point $x \in M$ such that $u_n \to x$ as $n \to \infty$.

All we have left to show from here is that $x \in \text{limpts}(U)$ (since we are trying to prove that $\text{limpts}(U) \neq \varnothing$). If $\{u_n\}_{n=1}^{\infty} \subset U \backslash \{x\}$ we're done by Lemma 6.4.4. Otherwise, note that $\{u_n\}_{n=1}^{\infty}$ was constructed such that $u_n = u_m$ if and only if $n = m$. Hence if there exists

some $n_0 \in \mathbb{N}$ such that $u_{n_0} = x$, then for every other $n \neq n_0$, $u_n \neq x$. Hence $\{u_{n_0+k}\}_{k=1}^{\infty}$ is a sequence of elements in $U\backslash\{x\}$ which converges to x.

(4) $\implies$ (3): Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in M. If the sequence takes on only finitely many values in M, then the pigeon hole principle ensures that one of those values is taken infinitely many times by the sequence, which in turn ensures that there is a convergent subsequence (just take the subsequence to be the constant sequence constructed from those points in the sequence which hit the same value infinitely many times). Otherwise, if the sequence takes on infinitely many values, then (4) implies that there is a limit point $x \in M$ of the set $\{x_n\}_{n=1}^{\infty}$. In particular, if x is a limit point of the set $\{x_n\}_{n=1}^{\infty}$, then there is a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ such that $\lim\limits_{k\to\infty} x_{n_k} = x$ by Exercise 9 (15).

(3) $\implies$ (2): Suppose that $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence in M. By assuming (3), this implies that there is a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ of $\{x_n\}_{n=1}^{\infty}$ and an $x \in M$ such that $x_{n_k} \to x$ as $k \to \infty$. But, this implies that $x_n \to x$ as $n \to \infty$ by Theorem 6.2.5 (2), implying that M is complete.

We now need to show that assuming (3) implies that M is totally bounded. Let's prove this by the contrapositive. That is, let's assume that M is not totally bounded and show that this implies that M is not sequentially compact. Indeed, assume there exists $\varepsilon_0 > 0$ such that for any finite collection of point $\{y_1, y_2, ..., y_n\} \subset M$, $M \not\subset \cup_{i=1}^{n} B(y_i; \varepsilon)$. We will now construct a sequence with no convergent subsequence. Let $x_1 \in M$ be any point. Then, find x_2 such that $x_2 \in M\backslash B(x_1; \varepsilon_0)$. Next, find x_3 such that $x_3 \in M\backslash\big(B(x_1; \varepsilon_0) \cup B(x_2; \varepsilon_0)\big)$. Continue on this way (inductively) such that for every $n \in \mathbb{N}$, we have a point $x_n \in M\backslash \cup_{i=1}^{n-1} B(x_i; \varepsilon_0)$. Note that this is possible since we've assumed that for any $n \in \mathbb{N}$, $M \not\subset \cup_{i=1}^{n-1} B(x_i; \varepsilon_0)$. We will show that this sequence we've constructed, $\{x_n\}_{n=1}^{\infty}$, has no convergent subsequence. Let $\{x_{n_k}\}_{k=1}^{\infty}$ be any subsequence of $\{x_n\}_{n=1}^{\infty}$. For every $n, m \in \mathbb{N}$, we chose our sequence so that $\rho(x_n, x_m) \geq \varepsilon_0$ implying that for every $k, l \in \mathbb{N}$, $\rho(x_{n_k}, x_{n_l}) \geq \varepsilon_0$. Therfore $\{x_{n_k}\}_{k=1}^{\infty}$ is not a Cauchy sequence, and hence does not converge by (the contrapositive statement of) Theorem 6.2.5 (1).

(1) $\implies$ (4): Let U be a subset of M. We will prove that if U does not have a limit point, then U has only finitely many elements (and hence be done since this will have proved the contrapositive). It suffices to assume that U is non-empty, since the empty set has finitely many elements. Hence assume that U is non-empty and suppose that $\text{limpts}(U) = \varnothing$. Then U is closed (see Exercise 9 (19)); therefore $V = M \setminus U$ is open. Moreover, if $x \in U$, then there

is an $\varepsilon_x > 0$ such that $B(x;\varepsilon_x) \cap U = \{x\}$, otherwise $x \in \text{limpts}(U)$. That is, $B(x;\varepsilon_x)$ has no other elements in common with U other than x. The collection $\{V\} \cup \{B(x;\varepsilon_x) \text{ s.t. } x \in U\}$ is an open cover of M (see Exercise 9 (20)). Since we have assumed that M is compact, there is a finite sub-cover; say $x_1, x_2, ..., x_n \in U$ such that $\{V, B(x_1;\varepsilon_{x_1}), B(x_2;\varepsilon_{x_2}), ..., B(x;\varepsilon_{x_n})\}$ covers M. Since

$$U \subset M \subset V \cup B(x_1;\varepsilon_{x_1}) \cup B(x_2;\varepsilon_{x_2}) \cup \cdots \cup B(x;\varepsilon_{x_n})$$

this collection also covers U. However, $U \cap V = \varnothing$ which implies that

$$U \subset B(x_1;\varepsilon_{x_1}) \cup B(x_2;\varepsilon_{x_2}) \cup \cdots \cup B(x;\varepsilon_{x_n}).$$

This then implies that $U = \{x_1, x_2, ..., x_n\}$ since for each $i = 1, 2, ..., n$, the only element in U which is also in $B(x_i;\varepsilon_{x_i})$ is x_i. We have now proved that U has only finitely many elements.

(2) + (3) $\implies$ (1): We will do this by contradiction; assume both (2) and (3) hold, but (1) does not. In this case, there exists an open cover $\{U_a\}_{a\in A}$ of M which does not admit a finite sub-cover. In particular, assuming M is totally bounded implies that for every $\varepsilon > 0$ there is some $x \in M$ such that $B(x;\varepsilon)$ is not contained in any one of the U_a for any $a \in A$ (see Exercise 9 (21)). This observation allows us to construct a sequence $\{x_n\}_{n=1}^{\infty}$ in M such that for every $n \in \mathbb{N}$, it holds that $B(x_n; 1/n)$ is not contained in any one of the U_a for any $a \in A$. Under assumption (3), there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ and a point $x \in M$ such that $x_{n_k} \to x$ as $k \to \infty$. Now, there exists an $a_0 \in A$ such that $x \in U_{a_0}$, since $\{U_a\}_{a\in A}$ is a cover of M. Hence there exists some $\varepsilon > 0$ such that $B(x;\varepsilon) \subset U_{a_0}$, since U_{a_0} is open. There exists an $n_1 \in \mathbb{N}$ such that for every $n \geq n_1$, $\rho(x_n, x) < \varepsilon/3$; similarly, there exists an $n_2 \in \mathbb{N}$ such that $1/n_2 < \varepsilon/2$. Set $n_0 = \max\{n_1, n_2\}$. Then $B(x_{n_0}, 1/n_0) \subset B(x;\varepsilon)$. To see this, take $y \in B(x_{n_0}, 1/n_0)$, and by the triangle inequality $\rho(x, y) \leq \rho(x, x_{n_0}) + \rho(x_{n_0}, y) < \varepsilon/2 + 1/n_0 < \varepsilon/2 + \varepsilon/2 = \varepsilon$. Hence $y \in B(x;\varepsilon)$, proving that $B(x_{n_0}, 1/n_0) \subset B(x;\varepsilon)$. Now, $B(x_{n_0}, 1/n_0) \subset B(x;\varepsilon) \subset U_{a_0}$, which contradicts the assumption that there does not exists an $a \in A$ such that $B(x_{n_0}, 1/n_0)$ is contained in U_a. This concludes the proof. ■

7. Continuous Functions Between Metric Spaces

For this section, we will continue to assume that (M, ρ) is a metric space and, since we will be considering functions between metric spaces, let (N, d) be another metric space.

DEFINITION 6.7.1. A function $f : M \to N$ is *continuous* at $x_0 \in M$ when for all $\varepsilon > 0$ there exists a $\delta > 0$ such that if $x \in M$ and $\rho(x, x_0) < \delta$ then $d(f(x), f(x_0)) < \varepsilon$.

REMARK 6.7.2. A couple equivalent ways to define continuity using Definition 6.7.1 come immediately from the definition of an open ball. They are:

(1) A function $f : M \to N$ is continuous at $x_0 \in M$ when for all $\varepsilon > 0$ there exists a $\delta > 0$ such that if $x \in B(x_0; \delta)$ then $f(x) \in B(f(x_0); \varepsilon)$.
(2) A function $f : M \to N$ is continuous at $x_0 \in M$ when for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $f\big(B(x_0; \delta)\big) \subset B(f(x_0); \varepsilon)$.

The slightly subtle point to be careful about using these definitions is that when we write $B(x_0; \delta)$, we are making the open ball around x_0 using the metric ρ; whereas, when we write $B(f(x_0); \varepsilon)$, we are making the open ball around $f(x_0)$ using the metric d. □

THEOREM 6.7.3. *Let $f : M \to N$ and $x_0 \in M$. The following are equivalent:*

(1) *f is continuous at x_0.*
(2) *Given any subset $V \subset N$ which is a neighborhood of $f(x_0)$, the inverse image $f^{-1}(V)$ is a neighborhood of x_0.*
(3) *For all sequences $\{x_n\}_{n=1}^{\infty}$ in M, if $\lim_{n\to\infty} x_n = x_0$ then $\lim_{n\to\infty} f(x_n) = f(x_0)$.*

Proof. (1) $\implies$ (2): Suppose f is continuous at x_0. Let $V \subset N$ be a neighborhood of $f(x_0)$. By Exercise 9 (12), there is some $\varepsilon > 0$ such that $B(f(x_0); \varepsilon) \subset V$. Now, let $\delta > 0$ such that if $x \in M$ and $\rho(x, x_0) < \delta$ then $d(f(x), f(x_0)) < \varepsilon$. This is equivalent to saying $B(x_0; \delta) \subset f^{-1}(V)$, implying that $f^{-1}(V)$ is a neighborhood of x_0.

(2) $\implies$ (1): Take $\varepsilon > 0$. Let $V = B(f(x_0); \varepsilon) \subset N$. Then, by (2), $f^{-1}(V)$ is a neighborhood of x_0. Hence, there exists some $\delta > 0$ such that $B(x_0; \delta) \subset f^{-1}(V)$. In particular, $f\big(B(x_0; \delta)\big) \subset V$, which is equivalent to saying that if $x \in B(x_0; \delta)$ then $f(x) \in B(f(x_0); \varepsilon)$. In turn, this is equivalent to saying that if $x \in M$ and $\rho(x, x_0) < \delta$, then $d(f(x), f(x_0)) < \varepsilon$, proving that f is continuous at x_0.

(1) $\implies$ (3): Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in M such that $\lim_{n\to\infty} x_n = x_0$. We need to show that $\lim_{n\to\infty} f(x_n) = f(x_0)$. Take $\varepsilon > 0$. By assumption, we can find some $\delta > 0$ such that if $\rho(x, x_0) < \delta$, then $d(f(x), f(x_0)) < \varepsilon$. Since $x_n \to x_0$, we can find some $n_\delta \in \mathbb{N}$ such that for all $n \geq n_\delta$, $\rho(x_n, x_0) < \delta$; but then, for all $n \geq n_\delta$, $d(f(x_n), f(x_0)) < \varepsilon$, proving that $\lim_{n\to\infty} f(x_n) = f(x_0)$.

(3) $\implies$ (1): We prove this by the contrapositive. If f is not continuous at x_0, then there exists some $\varepsilon > 0$ such that for all $\delta > 0$, there exists a $x \in M$ with $\rho(x, x_0) < \delta$ and $d(f(x), f(x_0)) \geq \varepsilon$. Thus, for every $n \in \mathbb{N}$, define the point $x_n \in M$ to be found such that $\rho(x_n, x_0) < 1/n$ and $d(f(x_n), f(x_0)) \geq \varepsilon$. This defines a sequence $\{x_n\}_{n=1}^{\infty}$ such that for every n it holds that $\rho(x_n, x_0) < 1/n$, and hence $x_n \to x_0$; however, for every $n \in \mathbb{N}$, $d(f(x_n), f(x_0)) \geq \varepsilon$ and therefore $f(x_n) \not\to f(x_0)$. ■

PROPOSITION 6.7.4. *Suppose that (X, ν) is another metric space, $f : M \to N$, and $g : N \to X$. If $x_0 \in M$, f is continuous at x_0, and g is continuous at $f(x_0)$, then $g \circ f : M \to X$ is continuous at x_0.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be any sequence in M which converges to x_0. Since f is assumed continuous at x_0, Theorem 6.7.3 (3) ensures that the sequence $\{f(x_n)\}_{n=1}^{\infty}$ in N converges to $f(x_0)$. In turn, since g is continuous at $f(x_0)$ and $\{f(x_n)\}_{n=1}^{\infty}$ is a sequence which converges to $f(x_0)$, another application of Theorem 6.7.3 (3) shows that the sequence $\{g(f(x_n))\}_{n=1}^{\infty}$ in X converges to $g(f(x_0))$. What we have then deduced is that given any sequence $\{x_n\}_{n=1}^{\infty}$ which converges to x_0, we have $\lim_{n\to\infty} g(f(x_n)) = g(f(x_0))$, proving that $g \circ f$ is continuous at x_0 (again using the equivalence of (1) and (3) in Theorem 6.7.3). ■

DEFINITION 6.7.5. If $f : M \to N$, then we say that f is *continuous* on M when f is continuous at all points $x \in M$.

Nearly immediately we get two results concerning continuous functions by using the previous arguments at all points in their domain.

THEOREM 6.7.6. *Let $f : M \to N$. The following are equivalent.*

(1) *f is continuous.*
(2) *Given any open subset $V \subset N$, $f^{-1}(V)$ is an open subset of M.*
(3) *Given any closed subset $C \subset N$, $f^{-1}(C)$ is a closed subset of M.*
(4) *Given any sequence $\{x_n\}_{n=1}^{\infty}$ in M and any point $x \in M$, if $\lim_{n\to\infty} x_n = x$ then $\lim_{n\to\infty} f(x_n) = f(x)$.*

Proof. (1) $\iff$ (2) and (1) $\iff$ (4): Exercise 9 (27).

(2) $\iff$ (3): Suppose that $C \subset N$. By definition, $f^{-1}(C)$ is closed if and only if $f^{-1}(C)^c = M \setminus f^{-1}(C)$ is open. Since $f^{-1}(C)^c = f^{-1}(C^c)$ where $C^c = N \setminus C$, what we have

stated so far is $f^{-1}(C)$ is closed if and only if $f^{-1}(C^c)$ is open. Equivalently, for $V \subset N$, it holds that $f^{-1}(V)$ is open if and only if $f^{-1}(V^c)$ is closed. ■

PROPOSITION 6.7.7. *If $f : M \to N$ and $g : N \to X$ are continuous, then $g \circ f : M \to X$ is continuous.*

Proof. By Proposition 6.7.4, $g \circ f$ is continuous at all points in M, which is what it means for $g \circ f$ to be continuous. ■

Motivated by Theorem 6.7.6 (4), we prove that, from a certain perspective, the metric ρ is continuous.

PROPOSITION 6.7.8 (Continuity of the Metric). *Let (M, ρ) be a metric space. Suppose that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are sequences in M which converges to x_0 and y_0, respectively. Then*

$$\lim_{n\to\infty} \rho(x_n, y_n) = \rho(x_0, y_0).$$

Proof. By two applications of the triangle inequality we have

$$\begin{aligned}\rho(x_n, y_n) - \rho(x_0, y_0) &\leq \rho(x_n, x_0) + \rho(x_0, y_0) + \rho(y_0, y_n) - \rho(x_0, y_0) \\ &= \rho(x_n, x_0) + \rho(y_n, y_0)\end{aligned}$$

and similarly, from two applications of the reverse triangle inequality,

$$\begin{aligned}\rho(x_n, y_n) - \rho(x_0, y_0) &\geq \rho(x_0, y_0) - \rho(x_n, x_0) - \rho(y_n, y_0) - \rho(x_0, y_0) \\ &= -[\rho(x_n, x_0) + \rho(y_n, y_0)]\end{aligned}$$

What this shows is that

$$|\rho(x_n, y_n) - \rho(x_0, y_0)| \leq \rho(x_n, x_0) + \rho(y_n, y_0).$$

Since the righthand side of this last inequality converges to zero (and the absolute value is always bounded below by zero), the squeezing principle tells us

$$\lim_{n\to\infty} |\rho(x_n, y_n) - \rho(x_0, y_0)| = 0$$

which proves that $\lim_{n\to\infty} \rho(x_n, y_n) = \rho(x_0, y_0)$. ■

7.1. Uniform and Lipschitz Continuity. As was in the case of continuous functions of one real variable, we can define the general notion of uniform and Lipschitz continuity of

functions between metric spaces. As before, these special types of continuous functions occur frequently in practice and are worth dedicating some attention.

DEFINITION 6.7.9. If $f : M \to N$, we say that f is *uniformly continuous* when: Given any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $x, y \in M$ and $\rho(x, y) < \delta$ then $d(f(x), f(y)) < \varepsilon$.

LEMMA 6.7.10. *If $f : M \to N$ is uniformly continuous, then it is continuous.*

Proof. Exercise 9 (29). ■

THEOREM 6.7.11. *Suppose that M is compact and $f : M \to N$ is continuous. Then f is uniformly continuous.*

Proof. Let $\varepsilon > 0$. By the assumption of continuity, for every $x \in M$ there exists some $\delta_x > 0$ such that if $y \in B(x; \delta_x)$ then $f(y) \in B(f(x); \varepsilon/2)$. Now, $\{B(x; \delta_x/2)\}_{x \in M}$ is an open cover of M, and therefore there exist points $x_1, x_2, ..., x_n \in M$ such that $\{B(x_i; \delta_{x_i}/2)\}_{i=1}^n$ is a finite sub-cover of M, since M is compact. In particular, this means that for every $x \in M$, there exists some $i \in \{1, 2, ..., n\}$ such that $x \in B(x_i; \delta_{x_i}/2)$. Let $\delta = \min\{\delta_{x_1}/2, ..., \delta_{x_n}/2\}$. Take $x, y \in M$ such that $\rho(x, y) < \delta$. Let $i \in \{1, 2, ..., n\}$ such that $x \in B(x_i, \delta_{x_i}/2)$. Then $\rho(x, x_i) \leq \delta_{x_i}/2$ and $\rho(x_i, y) \leq \rho(x_i, x) + \rho(x, y) < \delta_{x_i}/2 + \delta \leq \delta_{x_i}$. Therefore, if $\rho(x, y) < \delta$, then $d(f(x), f(y)) \leq d(f(x), f(x_i)) + d(f(x_i), f(y)) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. This proves that f is uniformly continuous. ■

THEOREM 6.7.12. *Let $K \subset M$ be compact and $f : M \to N$ be continuous. Then $f(K)$ is a compact subset of N. Poetically, this says that the continuous image of a compact set is compact.*

Proof. Let $\{V_a\}_{a \in A}$ be an open cover of $f(K)$. For each $a \in A$ let $U_a = f^{-1}(V_a)$. Since V_a is open in N and $f : M \to N$ is continuous, then U_a is open in M by Theorem 6.7.6. Particularly, $\{U_a\}_{a \in A}$ is an open cover of K. By the assumption that K is compact, there exists a finite sub-cover $\{U_b\}_{b \in B} \subset \{U_a\}_{a \in A}$ of K. We have that $f(U_b) \subset V_b$ holds for every $b \in B$, which implies that $\{V_b\}_{b \in B}$ covers $f(K)$. To see this, if $y \in f(K)$, then there is some point $x \in K$ such that $f(x) = y$. From here we can find some $b \in B$ such that $x \in U_b$, and hence $y \in f(U_b) \subset V_b$. We conclude that $\{V_b\}_{b \in B}$ is therefore a finite sub-cover admitted by $\{V_a\}_{a \in A}$, proving that $f(K)$ is compact. ■

DEFINITION 6.7.13. A function $f : M \to N$ is called *Lipschitz continuous* if there exists some $k > 0$ such that for all $x, y \in M$, it holds that $d(f(x), f(y)) \leq k\,\rho(x, y)$. In the case that $f : M \to N$ is Lipschitz, the value $k_f = \inf\{k > 0 \text{ s.t. for all } x, y \in M,\ d(f(x), f(y)) < k\,\rho(x, y)\}$ is called the Lipschitz constant of f.

REMARK 6.7.14. In Exercise 9 (33) you prove that for all $x, y \in M$, it holds that $d(f(x), f(y)) \leq k_f\,\rho(x, y)$, which isn't immediately true from the definition of the Lipschitz bound k_f. However, what is immediately true is that if $k > k_f$, then $d(f(x), f(y)) < k\,\rho(x, y)$; we use this fact setting $k = k_f + 1$ below in Lemma 6.7.15. □

LEMMA 6.7.15. . *If $f : M \to N$ is Lipschitz continuous, then it is uniformly continuous.*

Proof. Let k_f be the Lipschitz constant of f. Let $\varepsilon > 0$ and define $\delta = \frac{\varepsilon}{k_f+1}$; we are adding 1 in the denominator to deal with the pesky possibility that $k_f = 0$. Then for any $x, y \in M$, if $\rho(x, y) < \delta$ it holds that

$$d(f(x), f(y)) < (k_f + 1)\,\rho(x, y) < (k_f + 1)\frac{\varepsilon}{k_f + 1} = \varepsilon.$$

Since δ was chosen without regard to the arbitrarily chosen $x, y \in M$, we conclude that f is uniformly continuous. ■

8. The Contraction Mapping Theorem and an Example

This section will introduce the *Contraction Mapping Theorem* with proof. This is a very powerful theorem, and is one of several "fixed point theorems" used frequently in mathematics. To illustrate the power of this theorem we are going to give a proof of existence and uniqueness of a certain family of differential equations. Let $D \subset \mathbb{R}^2$ and suppose that $f : D \to \mathbb{R}$. Let $t_0, y_0 \in \mathbb{R}$ and consider the differential equation

$$y'(t) = f(t, y(t)),\ y(t_0) = y_0.$$

Such an equation is called an initial value problem, since we interpret t_0 as the *initial time* and if $y(t)$ solves this differential equation, the *initial value* of y is then $y(t_0) = y_0$. However, it is not always true that a solution to a differential equation of this form exists, or if a solution exists, it can happen that there are many solutions! The existence and uniqueness of a solution is governed by the function f. If f is nice enough, then we will not only be guaranteed a solution, but we we be guaranteed that there is one and only one solution. The

theorem we will prove through an application of the Contraction Mapping Theorem is the following:

THEOREM 6.8.1. *Let $a, b, r \in \mathbb{R}$ with $t_0 \in (a, b)$, $0 < r$, and $\mathscr{R} = [a, b] \times [y_0 - r, y_0 + r]$. Suppose that $\mathscr{R} \subset D \subset \mathbb{R}^2$. Assume that $f : D \to \mathbb{R}$ is continuous and assume further that f is Lipschitz in the second variable uniformly in the first variable on $\mathscr{R}$; that is, there is some $\kappa > 0$ such that $|f(t, x) - f(t, y)| < \kappa |x - y|$ holds for every $t \in [a, b]$ and $x, y \in [y_0 - r, y_0 + r]$. Then there exists some $\delta > 0$ such that $[t_0 - \delta, t_0 + \delta] \subset [a, b]$ and there is a unique solution to the initial value problem*

$$y'(t) = f(t, y(t))$$
$$y(t_0) = y_0$$

for every t such that $|t - t_0| < \delta$.

8.1. Contraction Maps. Before proving Theorem 6.8.1, we need to build up the machinery. We will let (M, ρ) be a metric space.

DEFINITION 6.8.2. A function $T : M \to M$ is called a contraction when T is Lipschitz continuous with Lipschitz constant $k_T < 1$.

THEOREM 6.8.3 (Contraction Mapping Theorem). *Suppose that (M, ρ) is a complete metric space and $T : M \to M$ is a contraction. Then there is a unique point $x^* \in M$ such $T(x^*) = x^*$. Moreover, the point x^* can be found by the following routine:*
Step 1) Start with any point $x_0 \in M$.
Step 2) For every $n \geq 0$, let $x_{n+1} = T(x_n)$.
Step 3) Then $x^* = \lim_{n \to \infty} x_n$.

Proof. We let $q \in (0, 1)$ be chosen such that $k_T \leq q$ where k_T is the Lipschitz constant of T; we could just use the Lipschitz constant k_T instead of q, but that would be more notationally taxing. Let $x_0 \in M$ and for every $n \geq 0$, let $x_{n+1} = T(x_n)$. Notice that this is iterating applications of T: $x_1 = T(x_0)$, $x_2 = T(x_1) = T(T(x_0))$, $x_3 = T(x_2) = T(T(T(x_0)))$, etc. This defines a sequence $\{x_n\}_{n=0}^{\infty}$ of elements in M. We will now break the proof into three parts: Part 1 will be to show that $\{x_n\}_{n=0}^{\infty}$ is Cauchy and hence converges to some point $x^* \in M$ since M is complete; Part 2 will be to show that $T(x^*) = x^*$; Part 3 will show that

x^* is the unique element in M such that $T(x^*) = x^*$.

Part 1: Notice that for any $k \in \mathbb{N}$,

$$\rho(x_{k+1}, x_k) = \rho(T(x_k), T(x_{k-1})) \leq q\,\rho(x_k, x_{k-1}) \leq q\big(q\,\rho(x_{k-1}, x_{k-2})\big) = q^2\rho(x_{k-1}, x_{k-2}).$$

Continuing this inductively, we see that $\rho(x_{k+1}, x_k) = q^k\rho(x_1, x_0)$. Define $R = \rho(x_1, x_0)$. Now, let $m, n \in \mathbb{N}$ with $m < n$. Then by the triangle inequality and our previous observation

$$\begin{aligned}\rho(x_n, x_m) &\leq \rho(x_n, x_{n-1}) + \rho(x_{n-1}, x_{n-2}) + \cdots + \rho(x_{m+1}, x_m) \leq q^{n-1}R + q^{n-2}R + \cdots + q^mR \\ &= R\sum_{k=m}^{n-1} q^k = Rq^m \sum_{k=0}^{n-m-1} q^k \leq Rq^m \sum_{k=0}^{\infty} q^k = \frac{Rq^m}{1-q}.\end{aligned}$$

Since $|q| < 1$ by assumption, $\frac{Rq^m}{1-q} \to 0$ as $m \to \infty$. In particular, $\{x_n\}_{n=0}^{\infty}$ is a Cauchy sequence. Therefore, since M is complete, there exists some $x^* \in M$ such that $\lim\limits_{n\to\infty} x_n = x^*$.

Part 2: By definition, $T(x_n) = x_{n+1}$ for every $n \in \mathbb{N}$. Now, applying the limit to both sides of this equality yields, $\lim\limits_{n\to\infty} T(x_n) = \lim\limits_{n\to\infty} x_{n+1} = x^*$. However, since T is Lipschitz continuous, it is continuous, so $\lim\limits_{n\to\infty} T(x_n) = T\big(\lim\limits_{n\to\infty} x_n\big) = T(x^*)$. Therefore $T(x^*) = x^*$.

Part 3: Suppose that $y \in M$ is such that $T(y) = y$. We will show that $\rho(y, x^*) = 0$ which will imply that $y = x^*$, proving that x^* is unique. To this end,

$$\rho(y, x^*) = \rho(T(y), T(x^*)) \leq q\,\rho(y, x^*).$$

The first equality came from the fact that $T(y) = y$ and $T(x^*) = x^*$, and the second because T is a contraction. Now, since $q < 1$, the only solution to $\rho(y, x) \leq q\rho(y, x)$ is $\rho(y, x) = 0$. ∎

REMARKS 6.8.4.

(1) If we define $T^n = \overbrace{T \circ T \circ \cdots \circ T}^{n \text{ applications}}$, we could have equivalently defined the sequence in the previous theorem by $x_n = T^n(x_0)$. This is common notation for iterating compositions of a function.

(2) Consider any function f defined from a set U to itself, $f : U \to U$. If there is an element $x \in U$ such that $f(x) = x$, we call x a *fixed point* of f.

(3) The contraction mapping theorem was a part of Banach's PhD thesis. It is therefore commonly called *Banach's Fixed Point Theorem*. Our proof closely followed the proof in his thesis.

8.2. Proof of Theorem 6.8.1. We are now ready to prove Theorem 6.8.1.

Proof. For convenience, let $J = [y_0 - r, y_0 + r]$ and recall that $\mathscr{R} = [a, b] \times J$. Note that $\mathscr{R}$ is a closed and bounded subset of $\mathbb{R}^2$, which by the Euclidean Heine-Borel theorem (proved in Chapter 5) means that $\mathscr{R}$ is compact. Since f is assumed continuous, then $f(\mathscr{R})$ is a compact subset of $\mathbb{R}$, and hence is bounded. This implies that there exists some $B > 0$ such that $\sup_{\omega \in \mathscr{R}} |f(\omega)| \leq B$. Also, by the Lipschitz continuity of f in the second variable, there is some $\kappa > 0$ such that $|f(t, x) - f(t, y)| \leq \kappa |x - y|$ for every $t \in [a, b]$. We define δ as any positive value

$$\delta \leq \min \left\{ \frac{r}{B}, \frac{1}{2\kappa} \right\}$$

such that $[t_0 - \delta, t_0 + \delta] \subset [a, b]$. For convenience, let $I = [t_0 - \delta, t_0 + \delta]$.

As in Example 6.1.5, let $C(I \to \mathbb{R})$ be the collections of continuous functions $h : I \to \mathbb{R}$, and let ρ be a metric on $C(I \to \mathbb{R})$ be defined by

$$\rho(h, g) = \sup_{x \in I} |h(x) - g(x)|.$$

In the next chapter we will prove that $C(I \to \mathbb{R})$ is, in fact, a complete metric space. We will use this fact here without proof. Define the set $\mathscr{M} \subset C(I \to \mathbb{R})$ by

$$\mathscr{M} = \{h \in C(I \to \mathbb{R}) \text{ s.t. } h(I) \subset J\}.$$

$\mathscr{M}$ is a closed subset of $C(I \to \mathbb{R})$. Indeed, let $\{h_n\}_{n=1}^{\infty}$ be a sequence of elements in $\mathscr{M}$ such that $\lim_{n \to \infty} h_n = h$ for some $h \in C(I \to \mathbb{R})$. Since this implies that for every $t \in I$, the sequence $\{h_n(t)\}_{n=1}^{\infty}$ in J converges to $h(t)$ in $\mathbb{R}$, and J is a closed subset of $\mathbb{R}$, we must have $h(t) \in J$. This proves that $h \in \mathscr{M}$, from which we conclude that $\mathscr{M}$ is closed by Theorem 6.5.1 (5). Since $C(I \to \mathbb{R})$ is complete and $\mathscr{M}$ is a closed subset of $C(I \to \mathbb{R})$, we can conclude that $(\mathscr{M}, \rho)$ is a complete metric space by Theorem 6.5.3. Moreover, $\mathscr{M}$ is non-empty because the constant function h_0 is in $\mathscr{M}$ where $h_0(t) = y_0$ for every $t \in I$.

Define the function $T : \mathscr{M} \to \mathscr{M}$ by

$$T(h)(t) = y_0 + \int_{t_0}^{t} f(s, h(s))\, ds$$

for every $t \in I$. We need to be a bit careful here. Simply because we declared that the image of T lands in $\mathscr{M}$, doesn't mean it truly does. To make sure T is well-defined, we need to show that for every $h \in \mathscr{M}$, the function $T(h)$ is in $\mathscr{M}$. Indeed, let $t \in I$, then

$$T(h)(t) = y_0 + \int_{t_0}^{t} f(s, h(s))\, ds \implies \big|T(h)(t) - y_0\big| = \left|\int_{t_0}^{t} f(s, h(s))\, ds\right| \leq B|t - t_0|.$$

Since $|t - t_0| \leq \delta \leq r/B$ we conclude that $|T(h)(t) - y_0| < r$ implying that $T(h)(t) \in [y_0 - r, y_0 + r] = J$. Moreover, since f is continuous, and h is continuous, we can conclude that $T(h)$ is continuous. Therefore, $T(h) \in \mathscr{M}$.

Next we show that T is a contraction. Let $h_1, h_2 \in \mathscr{M}$. Then

$$\begin{aligned}\rho(T(h_1), T(h_2)) &= \sup_{t\in I} \left|\int_{t_0}^{t} \big\{f(s, h_1(s)) - f(s, h_2(s))\big\}\, ds\right| \\ &\leq \sup_{t\in I} \left|\int_{t_0}^{t} \big|f(s, h_1(s)) - f(s, h_2(s))\big|\, ds\right|\end{aligned}$$

By assumption that f is Lipschitz continuous in the second variable, we have

$$\big|f(s, h_1(s)) - f(s, h_2(s))\big| \leq \kappa |h_1(s) - h_2(s)| \leq \kappa \sup_{u\in I} |h_1(u) - h_2(u)| = \kappa\, \rho(h_1, h_2)$$

Therefore,

$$\rho(T(h_1), T(h_2)) \leq \sup_{t\in I} \left|\int_{t_0}^{t} k\rho(h_1, h_2)\, ds\right| = k\, \rho(h_1, h_2) \sup_{t\in I} |t - t_0| = k\delta\, \rho(h_1, h_2).$$

Now, we originally chose δ such that $\delta \leq 1/(2\kappa)$, which means that

$$\rho(T(h_1), T(h_2)) \leq \frac{1}{2}\, \rho(h_1, h_2)$$

and hence T is a contraction. Therefore, by the Contraction Mapping Theorem, T has a unique fixed point in $\mathscr{M}$. To finish the proof, notice that h is a fixed point of T if and only if

$$h(t) = T(h)(t) = y_0 + \int_{t_0}^{t} f(s, h(s))\, ds$$

Using the fundamental theorem of calculus, this implies that h is a fixed point of T if and only if h satisfies the initial value problem. Since the fixed point of T exists and is unique, we have proved that the initial value problem has a unique solution. ■

9. Exercises

For these exercises, let (M, ρ) and (N, d) be metric spaces.

(1) Prove that examples 6.1.3, 6.1.4, and 6.1.5 are, in fact, metric spaces.

(2) Prove the reverse triangle inequality: For all $x, y, z \in M$,

$$|\rho(x, z) - \rho(z, y)| \leq \rho(x, y)$$

(3) Give an example of a complete metric space. Also, give an example of a metric space which is not complete.

(4) Let $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$ be the real numbers with an extra point ∞ extending the reals. For any $x, y \in \mathbb{R}_\infty$, if $x, y \in \mathbb{R}$ define $x < y$ (resp. $x \leq y$) if and only if $x < y$ (resp. $x \leq y$) as numbers in $\mathbb{R}$. If $x \in \mathbb{R}$ and $y = \infty$ then define $x < y$. This lets us have an ordering on $\mathbb{R}_\infty$. Let M be a set and $\rho : M \times M \to \mathbb{R}_\infty$ a function satisfying the four properties of a metric with the ordering replaced by that on $\mathbb{R}_\infty$. (Basically, all we're doing here is allowing a metric to take on the value ∞ giving a notion of two points in M being infinitely far apart). Prove that M can be partitioned into disjoint open sets of the form $U_x = \{y \in M \text{ s.t. } \rho(x, y) < \infty\}$ (in other words, prove that the relation $x \sim y \iff \rho(x, y) < \infty$ is an equivalence relation). Moreover, for each $x \in M$, (U_x, ρ) is itself a metric space as in Definition 6.1.1.

(5) Suppose that $\{x_n\}_{n=1}^\infty$ is a sequence in some metric space M with metric ρ. Let $x \in M$. Define the sequence $\{b_n\}_{n=1}^\infty$ in $\mathbb{R}$ by $b_n = \rho(x_n, x)$. Prove that the sequence $\{x_n\}_{n=1}^\infty$ converges to x if and only if the the sequence $\{b_n\}_{n=1}^\infty$ converges to 0. What you are proving is that $\lim_{n\to\infty} \rho(x_n, x) = 0$ if and only $\lim_{n\to\infty} x_n = x$.

(6) Let $U \subset M$ and $x \in M$. Define the distance between x and U by

$$\operatorname{dist}(x, U) = \inf \{\rho(x, y) \text{ s.t. } y \in U\}.$$

Prove the following: a) If U is closed and $x \notin U$, then $\operatorname{dist}(x, U) > 0$.

b) If U is not closed, then there exists an $x \notin U$ such that $\operatorname{dist}(x, U) = 0$.

(7) Define the metric ρ on $\mathbb{R}^2$ by

$$\rho\big((x_1, x_2), (y_1, y_2)\big) = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

Let $\varepsilon = 1$ and draw $B\big((0, 0); \varepsilon\big)$.

(8) Let (M, ρ) be a metric space and N be a set. Suppose that $f : N \to M$ is a bijection. Define the function $d : N \times N \to \mathbb{R}$ by $d(a, b) = \rho(f(a), f(b))$. Prove that d is a metric on N. Do we need the full strength of f being a bijection; i.e., did you actually need that it is both injective and surjective?

(9) Prove Proposition 6.3.4.

(10) Let $U \subset M$.

a) Prove that $\text{ext}(U)$ is an open subset of M.

b) Prove that $\text{cl}\big(\text{cl}(U)\big) = \text{cl}(U)$ and $\text{int}\big(\text{int}(U)\big) = \text{int}(U)$. Can we say the same for the boundary of U?

c) Prove that $M \setminus \text{cl}(U) = \text{int}(M \setminus U) = \text{ext}(U)$.

(11) Let $U \subset M$ and suppose $x \in \text{cl}(U)$. Prove that there is a sequence $\{x_n\}_{n=1}^{\infty}$ of elements in U such that $\lim_{n\to\infty} x_n = x$.

(12) Consider the following definition.

DEFINITION 6.9.1. A set $N \subset M$ is called a *neighborhood* of some point $x \in M$ when there an open set U with $x \in U$ such that $U \subset N$. A set $X \subset M$ is called a *deleted neighborhood* of x when $x \notin X$ and the set $N = X \cup \{x\}$ is a neighborhood of x.

(a) Prove the following proposition.

PROPOSITION 6.9.2. *Let $x \in M$ and $N \subset M$. The following are equivalent.*

(i) *N is a neighborhood of x.*

(ii) *There exists some $\varepsilon > 0$ such that $B(x; \varepsilon) \subset N$.*

(iii) *There exists some $\varepsilon > 0$ such that $\overline{B}(x; \varepsilon) \subset N$.*

(b) Prove the following proposition.

PROPOSITION 6.9.3. *Let $x \in M$ and $N \subset M$. The following are equivalent.*

(i) *X is a deleted neighborhood of x.*

(ii) *$x \notin X$ and there exists some $\varepsilon > 0$ such that $[B(x; \varepsilon) \setminus \{x\}] \subset X$.*

(iii) *$x \notin X$ and there exists some $\varepsilon > 0$ such that $[\overline{B}(x; \varepsilon) \setminus \{x\}] \subset X$.*

(13) Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in the metric space (M, ρ). Prove that this sequence converges to some limit x if and only if every subsequence converges to x.

(14) Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in the metric space (M, ρ) and let $n_0 \in \mathbb{N}$. Prove that $\{x_n\}_{n=1}^{\infty}$ converges to some limit x if and only if the sequence $\{x_{n_0+k}\}_{k=1}^{\infty}$ converges to x.

(15) Suppose that $\{x_n\}_{n=1}^{\infty}$ is a sequence in the metric space (M, ρ). Suppose further that the set of elements of this sequence, $\{x_n \text{ s.t. } n \in \mathbb{N}\}$, has a limit point x. Prove that there exists a subsequence of $\{x_{n_k}\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} x_{n_k} = x$.

(16) Let (M, ρ) be a metric space. Suppose $U \subset M$ is non-empty. Let $\overline{U}$ be defined as the set $\overline{U} = \{x \in M \text{ s.t. for every } \varepsilon > 0 \text{ it holds that } B(x;\varepsilon) \cap U \neq \varnothing\}$. Prove that $\overline{U} = \text{cl}(U)$.

(17) Let U be a subset of the metric space (M, ρ). Prove that $\text{bdry}(U)$ is a closed subset of M. *Hint:* Let $B = \text{bdry}(U)$. One way to do this is to show that if $x \in \text{bdry}(B)$ then $x \in \text{bdry}(U)$; another would be to show that if $\{x_n\}_{n=1}^{\infty}$ is a sequence in B which converges to some point x in M, then $x \in B$; either of these methods use Theorem 6.5.1.

(18) This exercise it to add a clarification to a subtle detail to the proof of Theorem 6.5.3.

(a) Let (M, ρ) be any metric space and $C \subset M$. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of elements in C. Suppose that there is some $x \in C$ and $y \in M$ such that $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} x_n = y$. Prove that $x = y$.

(b) Discuss why, if the previous part was not true, why we would not be able to use the proof Theorem 6.5.3 in the "converse" direction.

(19) Prove that if U is a subset of the metric space (M, ρ) has no limit points, then U is closed.

(20) Let $U \subset M$ be any closed set and let $V = M \setminus U$. For every $x \in U$, let ε_x be any positive real number. Prove that $\{V\} \cup \{B(x;\varepsilon_x) \text{ s.t. } x \in U\}$ is an open cover of M. That is, prove

$$M \subset V \cup \Big(\bigcup_{x\in U} B(x;\varepsilon_x) \Big).$$

(21) Suppose that (M, ρ) is a metric space and $\{U_a\}_{a\in A}$ is an open cover of M. Prove that if you assume that $\{U_a\}_{a\in A}$ does *not* admit a finite subcover and you assume that M is totally bounded, then for any $\varepsilon > 0$ there exists some $x \in M$ such that $B(x;\varepsilon)$ is not contained in any one of the U_a for any $a \in A$. *Hint*: It might be easiest to argue by contradiction.

(22) For this problem, consider the following claim and its *almost correct* proof followed by a counter example to the claim. Once you've done this, answer the question at the end.

Claim: Let (M, ρ) be a metric space and $x \in M$. Given any $r > 0$, it holds that $B(x; r)$ is not compact.

Proof. Define the collection of open sets $\{U_n\}_{n=2}^{\infty}$ by

$$U_n = B\big(x; (1 - 1/n)r\big).$$

Notice that $\{U_n\}_{n=2}^{\infty}$ is, in fact, an open cover of $B(x; r)$. Indeed, for any $y \in B(x; r)$ we have that $\rho(x, y) < r$ and therefore $\frac{\rho(x,y)}{r} < 1$. Let $n_0 \in \mathbb{N}$ be large enough that $\frac{\rho(x,y)}{r} < 1 - 1/n_0$; we can find such an n_0 be the Archimedean property. Then, for this choice of n_0, we have $\rho(x, y) < (1-1/n_0)r$, implying that $y \in U_{n_0}$. In particular, this implies that $y \in \bigcup_{n=2}^{\infty} U_n$ and, since $y \in B(x; r)$ was chosen arbitrarily, we deduce $B(x; r) \subset \bigcup_{n=2}^{\infty} U_n$. This proves that $\{U_n\}_{n=2}^{\infty}$ is an open cover of $B(x; r)$.

To finish the proof that $B(x; r)$ is not compact, we will show that $\{U_n\}_{n=2}^{\infty}$ does not admit a finite subcover. We do this by contradiction. Suppose that $\{U_n\}_{n=2}^{\infty}$ does admit a finite subcover of $B(x; r)$. Then we can find some finite collection $\{U_{n_1}, ..., U_{n_k}\}$ which covers $B(x; r)$. Assume without loss of generality that $n_1 < n_2 < \cdots < n_k$; note that this implies $U_{n_1} \subset U_{n_2} \subset \cdots \subset U_{n_k}$. We then have $\bigcup_{l=1}^{k} U_{n_l} = U_{n_k} = B(x; (1 - 1/n_k)r)$. Since $(1 - 1/n_k)r < r$, it must be that $B(x; (1 - 1/n_k)r) \subsetneq B(x; r)$, contradicting that our finite collection covered $B(x; r)$. ■

Counter Example: Let M be the interval in $\mathbb{R}$ given by $M = [-1/2, 1/2]$ and assume that ρ is the standard distance in $\mathbb{R}$: $\rho(x, y) = |x - y|$. Notice that the open ball in M centered at 0 of radius 1 equals M itself, i.e., $B(0; 1) = \{x \in M \text{ s.t. } \rho(x, 0) < 1\} = M$. Since $M = B(0; 1)$ then the previous claim would prove that M is not compact; however, since M is closed and bounded, the Heine-Borel theorem tells us that M is compact.

Question: What went wrong with the proof of the claim so that we were able to find this contradictive counter example and what assumption could have we made in the statement of the claim to have made our proof correct and excluded the counter

example as a possible space to apply the claim (for instance, this claim and its proof works fine in the case of Examples 6.6.2 and 6.6.4)?

(23) A metric space (M, ρ) is called *separable* when there exists a subset $U \subset M$ such that U is dense in M and U has at most countably many elements. In this case, we say that U is a *countably dense* subset of M. For example, $\mathbb{R}$ is separable since $\mathbb{Q}$ is a countably dense subset of $\mathbb{R}$.

(a) Let (M, ρ) be any metric space and V a non-empty open subset of M. For all $x \in V$, define $r_x = \sup\{r > 0 \text{ s.t. } B(x; r) \subset V\}$. Prove that $B(x; r_x) \subset V$.

(b) We continue with the same notation as part a. Suppose further that (M, ρ) is separable and let U be a countably dense subset of M. Given any $y \in V$, prove that there exists $x \in V \cap U$ such that $y \in B(x; r_x)$. From here deduce that there exist countably many points $\{x_1, x_2, x_3, ...\} \subset V$ such that

$$\bigcup_{n=1}^{\infty} B(x_n; r_{x_n}) = V.$$

Note: In part (a) you are showing that given a point x in a open subset V, it makes sense to talk about the "largest" open ball around x which is contained in V. In part (b) you are showing that if you assume that M is separable, then any open subset of M can be written as the union of only countably many open balls. This translated into the situation $M = \mathbb{R}$, we realize that any open subset of $\mathbb{R}$ can be written as the countable union of open intervals (assuming we are using the standard distance metric on $\mathbb{R}$).

(24) Let U and V be sets and $f : U \to V$ be a function. Suppose that $\{V_a\}_{a \in A}$ is a cover of the image $f(U)$ of f; i.e., $f(U) \subset \bigcup_{a \in A} V_a$. Prove that $\{U_a\}_{a \in A}$ is a cover of U where $U_a = f^{-1}(V_a)$ for every $a \in A$. (Recall that $f^{-1}(V_a) = \{x \in U \text{ s.t. } f(x) \in V_a\}$ is the inverse image – also called pre-image – of f).

(25) If (M, ρ) is compact and $U \subset M$ is closed, prove that (U, ρ) is compact. *Hint:* Theorem 6.5.3 might help you here.

(26) The standard *Euclidean* metric ρ on $\mathbb{R}^n$ is defined by the usual distance formula:

$$\rho\big((x_1, x_2, ..., x_n), (y_1, y_2, ..., y_n)\big) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2}.$$

Note that when $n = 1$, $\rho(x, y) = \sqrt{(x-y)^2} = |x - y|$, giving the standard distance on $\mathbb{R}$.

a) Consider $\mathbb{R}^2$ and $\mathbb{R}$ as metric spaces endowed with the standard Euclidean metric. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = x + y$. Prove that f is continuous at all points in $\mathbb{R}^2$.

b) Consider $\mathbb{R}^2$ and $\mathbb{R}$ as metric spaces endowed with the standard Euclidean metric. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be given by $h(x, y) = xy$. Prove that h is continuous at all points in $\mathbb{R}^2$.

Hint for b: $x_1y_1 - x_2y_2 = x_1y_1 - x_1y_2 + x_1y_2 - x_2y_2 = x_1(y_1 - y_2) + y_2(x_1 - x_2)$.

Note: In the past, you've probably proved that for a fixed $a \in \mathbb{R}$, the functions $x \mapsto a + x$ and $x \mapsto ax$ are continuous. However, this was for a fixed a. What you prove here is stronger! Indeed, what you prove is that that $(x, y) \mapsto x + y$ and $(x, y) \mapsto xy$ are simultaneously continuous in both x and y variables.

(27) (a) Prove (1) $\iff$ (2) in Theorem 6.7.6.

(b) Prove (1) $\iff$ (4) in Theorem 6.7.6.

(28) Consider the metric spaces $M = (-1, 1)$ and $N = \mathbb{R}$, both endowed with the standard Euclidean distance on $\mathbb{R}$. The function $f : M \to N$ defined simply by $f(x) = x$ is certainly continuous.

(a) By Theorem 6.7.6 (3), we should have $f^{-1}([0, 1])$ is a closed subset of M. Realize that $f^{-1}([0, 1]) = [0, 1)$. Is there a contradiction here? If so, what's wrong. If not, give a brief justification that $[0, 1)$ is a closed subset of M.

(b) Consider the sequence $\{x_n\}_{n=1}^{\infty}$ in M defined by $x_n = 1 - 1/n$. Does this sequence converge? Does the sequence $\{f(x_n)\}_{n=1}^{\infty}$ in N converge?

(29) Prove Lemma 6.7.10.

(30) Prove Proposition 6.7.7 using the equivalent description of continuity given in Theorem 6.7.6 (2).

(31) Let (M, ρ) and (N, d) be metric spaces and suppose that $f : M \to N$ is continuous. Prove that if $U \subset M$ then the restriction $f : U \to N$ is continuous a function between the metric spaces (U, ρ) and (N, d). Conversely, suppose that $x_0 \in M$ and prove that if U is a neighborhood containing x_0 and the restriction $f : U \to N$ is continuous at x_0, then the ambient function $f : M \to N$ is also continuous at x_0.

(32) Consider the metric space $M = \left\{ \frac{1}{n} \text{ s.t. } n \in \mathbb{N} \right\}$ endowed with the standard Euclidean distance ρ on $\mathbb{R}$. In this problem we give two proofs that (M, ρ) is not compact.

(a) Show that for every $x \in M$, the set $U_x = \{x\}$ is open in M. Therefore prove that $\{U_x\}_{x \in M}$ is an open cover of M that does not admit a finite subcover; this proves M is not compact.

(b) Let $f : M \to \mathbb{R}$ be defined by $f(x) = 1/x$. You know that f is continuous since it is the restriction of the map $x \mapsto 1/x$ defined on $(0, 1)$ restricted to M (see Exercise 9 (31)). Prove that $f : M \to \mathbb{R}$ is not uniformly continuous and hence, use Theorem 6.7.11 to conclude that M is not compact.

(33) Let (M, ρ) and (N, d) be metric spaces. Suppose that $f : M \to N$ is Lipschitz continuous with Lipschitz bound k_f. Prove that for all $x, y \in M$, it holds that $d(f(x), f(y)) \leq k_f \, \rho(x, y)$.

(34) In this problem you will approximate the point $\sqrt{5}$ using the Contraction Mapping Theorem 6.8.3. To start, the computational root finding algorithm known as Newton's method is used to find the roots of a differentiable function f by creating the sequence $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$. Roughly, the derivation comes by assuming x is near the root x^*, taking the first Taylor polynomial centered at the root $0 = f(x^*) \approx f(x) + f'(x)(x^* - x)$, and solving for x^*. So, if we want to apply this to find $\sqrt{5}$, a judicious choice of function is $f(x) = x^2 - 5$. Applying Newton's method, our sequence will look like

$$x_{n+1} = x_n - \frac{1}{2}\left(\frac{x_n^2 - 5}{x_n}\right) = \frac{1}{2}\left(x_n + \frac{5}{x_n}\right)$$

(a) Since $2^2 = 4 < 5 < 9 = 3^2$, let's agree that $2 < \sqrt{5} < 3$. Consider $M = [2, 3]$ a the metric space endowed with the standard Euclidean metric on $\mathbb{R}$. Define the map $T : M \to M$ by

$$T(x) = \frac{1}{2}\left(x + \frac{5}{x}\right).$$

Prove that T is, in fact, a map from M to M (that is, prove that if $x \in M$, then $T(x) \in M$); if this was not the case, then as presented, T would not be well defined.

(b) We will apply the contraction mapping theorem to $T : M \to M$, but we need to check two things first: i) M is complete, and ii) T is a contraction. Prove that both i) and ii) are true.

(c) From the previous part, you know that $T : M \to M$ is a contraction on a complete metric space, and by Theorem 6.8.3, must have a fixed point x^*. Show that the fixed point of T is $x^* = \sqrt{5}$. Then, starting with any $x_0 \in M$ you like, write out the first five iterations (using a calculator) of the sequence $x_1 = T(x_0)$, $x_2 = T(x_1)$, ..., $x_5 = T(x_4)$. Note that $\sqrt{5} = 2.236068 \pm 10^{-6}$. (If you know how to do some basic programming, it shouldn't take long to write a code to calculate x_n for large values of n, giving you very good approximations to $\sqrt{5}$).

(35) This problem lets you generalize the intermediate value theorem. A subset of a metric space $U \subset M$ is called *disconnected* or *not connected* when: There exist two non-empty disjoint open subsets $A \subset M$ and $B \subset M$ such that $U \cap A \neq \varnothing$, $U \cap B \neq \varnothing$ and $U \subset A \cup B$. Make sure to note that the assumption is that both A and B are non-empty, and that they are disjoint. Draw a picture of such a scenario and you will easily see why such a subset U is called *not* connected. On the other hand, if there do not exist such open subsets A and B, then U is called *connected*. That is, U is called connected when: For all open subsets $A \subset M$ and $B \subset M$ such that $U \cap A \neq \varnothing$ and $U \cap B \neq \varnothing$ and $U \subset A \cup B$, either $A \cap B \neq \varnothing$ or at least one of A or B is empty.

a) Let $f : M \to N$ be continuous. Prove that if M is connected (considering M as a subset of itself), then the image $f(M)$ is a connected subset of N. *Hint:* It might help to note that if $f(M) \subset A \cup B$, then $M \subset f^{-1}(A) \cup f^{-1}(B)$.

b) Endow $\mathbb{R}$ with the standard Euclidean metric, and suppose that $V \subset \mathbb{R}$. Prove that V is connected if and only if it is an interval. *Hint for the hard direction:* If V is an interval and A and B are open subsets such that $V \cap A \neq \varnothing$ and $V \cap B \neq \varnothing$, then let $a \in V \cap A$ and $b \in V \cap B$. Without loss of generality, assume that $a < b$. Let $c = \sup\{x \in V \cap A \text{ s.t. } x < b\}$. Argue that since $a < c \leq b$, then $c \in V$, and that by how we found c, it must be true that $c \in B$. From here use the fact that B is open so that $(c - \varepsilon, c + \varepsilon) \subset B$ and conclude that $A \cap B \neq \varnothing$.

c) Combining parts a and b, prove that if $f : M \to \mathbb{R}$ is continuous and M is

connected, then the image $f(M)$ is an interval in $\mathbb{R}$. Once you do this, use this result to prove the intermediate value theorem. *Note:* You might want to convince yourself the the intermediate value theorem is equivalent to: If $U \subset \mathbb{R}$ is an interval and $f : U \to \mathbb{R}$ is continuous, then the image $f(U)$ is an interval.

(36) Let $x \in M$ and $r > 0$. Prove that $\text{bdry}(B(x;r)) \subset \{y \in M \text{ s.t. } \rho(x,y) = r\}$. Hence prove that $\overline{B}(x;r)$ is closed.

(37) Let M be a set and assume that ρ_1 and ρ_2 are both metrics on M. We say that ρ_1 and ρ_2 are *equivalent metrics* when there exist constants $c, C > 0$ such that for all $x, y \in M$, $c\rho_2(x,y) \leq \rho_1(x,y) \leq C\rho_2(x,y)$. Let $U \subset M$. If ρ_1 and ρ_2 are equivalent metrics on M, prove:

a) U is open in (M, ρ_1) if and only if it is open in (M, ρ_2).

b) The sequence $\{x_n\}_{n=1}^{\infty}$ in M converges in (M, ρ_1) if and only if it converges in (M, ρ_2). *Note*: There are example of two different metrics ν_1 and ν_2 on the same set X where a convergent sequence in (X, ν_1) does not imply the same sequence converges in (X, ν_2); however, as you prove in this problem, this can never happens when the two metrics are equivalent.

CHAPTER 7

Introduction to Functional Analysis

1. Basics of Elementary Linear Algebra

DEFINITION 7.1.1. Let X be a set and $\mathbb{F}$ a field. We say that X is a *vector space* over $\mathbb{F}$ when:

Vector Addition. There exists a *vector addition* $+ : X \times X \to X$ satisfying:

i) For all $x, y \in X$, $x + y = y + x$.

ii) For all $x, y, z \in X$, $(x + y) + z = x + (y + z)$.

iii) There exists an element $z \in X$ such that for all $x \in X$, $z + x = x$. We call z the *zero vector* of X and denote it by 0_X.

iv) For all $x \in X$ there exists an element $y \in X$ such that $x + y = 0_X$. We call y the *additive inverse* of x and denote it by $-x$.

Scalar Multiplication. There exists a *scalar multiplication* $\cdot : \mathbb{F} \times X \to X$ satisfying:

i) For a $a, b \in \mathbb{F}$ and $x \in X$, $(ab) \cdot x = a \cdot (b \cdot x)$.

ii) For all $a, b \in \mathbb{F}$, $(a + b) \cdot x = (a \cdot x) + (b \cdot x)$.

iii) For all $a \in \mathbb{F}$ and $x, y \in X$, $a \cdot (x + y) = (a \cdot x) + (a \cdot y)$.

iv) If 1 is the multiplicative identity in $\mathbb{F}$, then for all $x \in X$, $1 \cdot x = x$.

v) If 0 is the additive identity in $\mathbb{F}$, then for all $x \in X$, $0 \cdot x = 0_X$.

For the remainder of this section, we will assume that X is a vector space over $\mathbb{F}$. The elements of X are called *vectors* and the elements of $\mathbb{F}$ are called *scalars*.

REMARK 7.1.2. One can easily check that a field $\mathbb{F}$ is, in fact, a vector space over itself. □

DEFINITION 7.1.3. If $Y \subset X$, then we say that Y is a *subspace* of X when Y is itself a vector space when endowed with the same vector addition and scalar multiplication as X.

PROPOSITION 7.1.4. *If $Y \subset X$, then Y is a subspace of X if and only if the following three properties hold:*

(1) $0_X \in Y$,

(2) $x, y \in Y$ *implies* $x + y \in Y$,

(3) $a \in \mathbb{F}$ *and* $x \in Y$ *implies* $a \cdot x \in Y$.

Proof. Exercise 9 (1). ■

DEFINITION 7.1.5. A finite set of vectors $\{x_1, x_2, ..., x_n\}$ is said to be *linearly independent* when: Given any collection $c_1, c_2, ..., c_d \in \mathbb{F}$ of scalars then $\sum_{i=1}^{n} c_i x_i = 0_X$ if and only if $c_i = 0$ for every $1 \leq i \leq n$. An arbitrary set of vectors $\mathscr{A} \subset X$ is said to be *linearly independent* when every finite subset of $\mathscr{A}$ is linearly independent. If a set of vectors is not linearly independent, then we say that it is *linearly dependent*.

DEFINITION 7.1.6. If $\mathscr{A} \subset X$ is a set of vectors, then the *span* of $\mathscr{A}$, denoted $\text{span}(\mathscr{A})$, is the collection of all *finite linear combinations* of the vectors in $\mathscr{A}$. That is $y \in \text{span}(\mathscr{A})$ if and only if there exist finitely many vectors $x_1, x_2, ..., x_n \in \mathscr{A}$ and scalars $c_1, c_2, ..., c_n \in \mathbb{F}$ such that $y = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$.

PROPOSITION 7.1.7. *The span of any subset of vectors is a subspace. More explicitly, for any subset $U \subset X$, the set $Y = \text{span}(U)$ is a subspace of X.*

Proof. Exercise 9 (2). ■

DEFINITION 7.1.8. A collection of vectors $\mathscr{A} \subset X$ is called a *linear algebraic basis*, or simply *basis*, for X when $\mathscr{A}$ is linearly independent and $\text{span}(\mathscr{A}) = X$.

PROPOSITION 7.1.9. *Let X be a vector space over $\mathbb{F}$ and $\mathscr{A} \subset X$. Then $\mathscr{A}$ is a basis for X if and only if $\mathscr{A}$ is a linearly independent set and for every $y \in X$, the set $\mathscr{A} \cup \{y\}$ is linearly dependent.*

Proof. Suppose that $\mathscr{A}$ is linearly independent such that for every $y \in X$, it holds that $\mathscr{A} \cup \{y\}$ is linearly dependent. We need to prove that $\text{span}(\mathscr{A}) = X$. To do so, let $y \in X$ be arbitrary. Since $\mathscr{A} \cup \{y\}$ is linearly dependent, there exists some $k \in \mathbb{N}$, elements $x_1, ..., x_k \in \mathscr{A}$, and non-zero scalars $c_1, ..., c_k, a \in \mathbb{F}$ such that $c_1 x_1 + \cdots + c_k x_k + ay = 0_X$. Now, we must be able to choose $a \neq 0$ since otherwise $c_1 x_1 + \cdots + c_k x_k = 0_X$, contradicting that $\mathscr{A}$ was linearly independent. But this means $y = \frac{1}{a}(c_1 x_1 + \cdots + c_k x_k)$, showing that $y \in \text{span}(\mathscr{A})$. Since $y \in X$ was arbitrary, this proves that $\text{span}(\mathscr{A}) = X$.

The other direction is left as Exercise 9 (3). ■

Proposition 7.1.9 admits an immediate corollary.

COROLLARY 7.1.10. *Let X be a vector space over $\mathbb{F}$. If $\mathscr{A} \subset X$ is a basis of X. Then, for any $y \in X$, the set $\mathscr{A} \cup \{y\}$ is not linearly independent.*

LEMMA 7.1.11. *Suppose that X is a vector space over $\mathbb{F}$ and $\mathscr{A} \subset X$ is a basis. For every non-zero $y \in X$ there exists a unique integer $k \in \mathbb{N}$, unique vectors $x_1, .., x_k \in \mathscr{A}$, and unique non-zero scalars $c_1, ..., c_k$ such that $c_1x_1 + c_2x_2 + \cdots + c_kx_k = y$.*

Proof. Exercise 9 (4). ■

THEOREM 7.1.12. *Let X be a vector space over $\mathbb{F}$. Suppose that $\mathscr{A} \subset X$ is a basis for X. If $\#(\mathscr{A}) = d < \infty$, then any subset $\mathscr{B} \subset X$ which is linearly independent is a basis if and only if $\#(\mathscr{B}) = d$.*

DEFINITION 7.1.13. Suppose $\mathscr{A}$ is a basis of X. If $\#(\mathscr{A}) = d < \infty$, then we say that X is a *finite-dimensional* vector space, and say that the *dimension* of X is d. If $\#(\mathscr{A}) = \infty$, then we say that X is an *infinite-dimensional* vector space.

DEFINITION 7.1.14. Suppose that Y is a vector space and $f : X \to Y$ is a map between vector spaces. We say that f is *linear* when: For all $x, y \in X$ and $a \in \mathbb{F}$, $f(a \cdot x + y) = a \cdot f(x) + f(y)$. The *null space* of a linear map $f : X \to Y$ is the subset of X given by $f^{-1}(\{0_Y\}) = \{x \in X \text{ s.t. } f(x) = 0_Y\}$. The null space of f is commonly denoted $\text{Null}(f)$. If $f : X \to Y$ is linear and also a bijection, then we say that f is a *linear isomorphism* between X and Y. If a linear isomorphism exists between X and Y, we say that X and Y are *isomorphic* as vector spaces.

LEMMA 7.1.15. *Suppose X and Y are vector spaces and that $f : X \to Y$ is linear. Then the following hold.*

(1) *The null space of f, $\text{Null}(f)$, is a subspace of X.*
(2) *The range of f, $f(X)$, is a subspace of Y.*

Proof. Exercise 9 (5). ■

LEMMA 7.1.16. *Suppose that X, Y, and Z are vector spaces over $\mathbb{F}$; and $f : X \to Y$, $g : X \to Y$, and $h : Y \to Z$ are linear. Then the following hold.*

(1) *For any* $\alpha \in \mathbb{F}$*, it holds that* $f + \alpha \cdot g : X \to Y$ *is linear. As usual,* $(f + \alpha \cdot g)(x) = f(x) + \alpha \cdot g(x)$.
(2) $h \circ f : X \to Z$ *is linear.*

Proof. Exercise 9 (6). ■

REMARK 7.1.17. Lemma 7.1.16 shows that if X and Y are vector spaces over $\mathbb{F}$, then the space $\mathscr{L}(X \to Y)$ of linear functions $f : X \to Y$ is, in fact, a vector space over $\mathbb{F}$. Note that the zero-vector in $\mathscr{L}(X \to Y)$ is the linear map $f_0 : X \to Y$ defined by $f_0(x) = 0_Y$ for every $x \in X$. □

THEOREM 7.1.18. *Suppose that* X *and* Y *are vector spaces over* $\mathbb{F}$ *and that* $\mathscr{A} \subset X$ *is a basis. Given every function* $f : \mathscr{A} \to Y$ *there exists a unique linear function* $F : X \to Y$ *such that* $F(x) = f(x)$ *for every* $x \in \mathscr{A}$.

Proof. Suppose that $f : \mathscr{A} \to Y$. Let $x \in X$ be arbitrary and let $x_1, ..., x_k \in \mathscr{A}$ and $c_1, ..., c_k \in \mathbb{F}$ be such that $c_1 \cdot x_1 + \cdots + c_k x_k = x$. Define $F(x)$ as

$$F(x) = c_1 \cdot f(x_1) + \cdots + c_k \cdot f(x_k).$$

It is easy an easy exercise to check that $F : X \to Y$ is well-defined using Lemma 7.1.11, and also that F is linear. For uniqueness, suppose that $G : X \to Y$ is also linear such that $G(x) = f(x)$ for every $x \in \mathscr{A}$. Then, by linearity, for any $x \in X$

$$G(x) = c_1 \cdot G(x_1) + \cdots + c_k \cdot G(x_k) = c_1 \cdot f(x_1) + \cdots + c_k \cdot f(x_k) = F(x)$$

where, as before, $c_1, ..., c_k \in \mathbb{F}$ and $x_1, ..., x_k \in \mathscr{A}$ were chosen such that $x = c_1 \cdot x_1 + \cdots + c_k \cdot x_k$. This shows that $G = F$. ■

REMARK 7.1.19. Because of Theorem 7.1.18, it is common practice to only define a linear map by describing how it maps basis elements, and then *extend it linearly* to a function on the whole space. For example, we extended the map $f : \mathscr{A} \to Y$ linearly to the map $F : X \to Y$ in the proof of Theorem 7.1.18. □

THEOREM 7.1.20. *Suppose that* X *and* Y *are vector spaces over* $\mathbb{F}$*. If* X *has dimension* $d < \infty$*, then* X *and* Y *are isomorphic as vector spaces if and only if the dimension of* Y *is* d*. In this case,* $f : X \to Y$ *is a linear isomorphism if and only if* $f(\mathscr{A})$ *is a basis for* Y *any*

time $\mathscr{A}$ is a basis for X; in turn, this happens if and only if there exists some basis $\mathscr{A}$ of X such that $f(\mathscr{A})$ is a basis for Y. Moreover, if $f : X \to Y$ is a linear isomorphism, then f is invertible (since it is bijective) and $f^{-1} : Y \to X$ is also a linear isomorphism.

1.1. Linear Dual Spaces.

DEFINITION 7.1.21. Let X be a vector space over $\mathbb{F}$. The *linear dual* of X, denoted X', is the collection of all linear functions from X to $\mathbb{F}$. That is,

$$X' = \{\varphi : X \to \mathbb{F} \text{ s.t. } \varphi \text{ is linear}\}$$

Each map $\varphi \in X'$ is called a *linear functional*, or simply *functional*, on X.

REMARK 7.1.22. In the notation of Remark 7.1.17, $X' = \mathscr{L}(X \to \mathbb{F})$. Lemma 7.1.16 and Remark 7.1.17 both illustrate that X' is a vector space over $\mathbb{F}$. It is worth noting again that the zero-vector $0_{X'}$ in X' is the map $0_{X'} : X \to \mathbb{F}$ given by $0_{X'}(x) = 0$ for every $x \in X$. □

DEFINITION 7.1.23. Suppose that X is finite dimensional and $\mathscr{A} \subset X$ be a basis for X. Then $\mathscr{A} = \{x_1, x_2, ..., x_d\}$ is a linear independent set with $d = \dim(X)$. The *dual basis* $\mathscr{A}' \subset X'$ corresponding to $\mathscr{A}$ is defined as $\mathscr{A}' = \{\varphi^1, \varphi^2, ..., \varphi^d\}$ where

$$\varphi^i(x_j) = \delta^i{}_j = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

for every i and j. Here $\delta^i{}_j$ is the *Dirac delta* function.

REMARK 7.1.24. Although we have only defined how φ^i in $\mathscr{A}'$ maps each basis element x_j in $\mathscr{A}$, this uniquely defines a linear function $\varphi^i : X \to \mathbb{F}$ by Theorem 7.1.18. □

THEOREM 7.1.25. *Suppose that X is a finite dimensional vector space over $\mathbb{F}$ and $\mathscr{A} \subset X$ is a bases. Then the dual basis $\mathscr{A}'$ is a basis for X'. In particular, if X is finite dimensional, then X' is also finite dimensional and is isomorphic to X. This result does not necessarily hold true if X if infinite dimensional.*

Proof. Let $\mathscr{A} = \{x_1, ..., x_d\}$ be a basis of X and $\mathscr{A}' = \{\varphi^1, ..., \varphi^d\} \subset X'$ the dual basis. Suppose that $\alpha_1, ..., \alpha_d \in \mathbb{F}$ are scalars such that

$$\alpha_1 \cdot \varphi_1 + \cdots + \alpha_d \cdot \varphi_d = 0_{X'}$$

Then, for any $i = 1, 2, ..., d$, we have

$$0 = 0_{X'}(x_i) = (\alpha_1 \cdot \varphi_1 + \cdots + \alpha_d \cdot \varphi_d)(x_i) = \alpha_1 \cdot \varphi^1(x_i) + \cdots + \alpha_i \cdot \varphi^i(x_i) + \cdots + \alpha_d \cdot \varphi^d(x_i)$$
$$= \alpha_1 \cdot 0 + \cdots + \alpha_i \cdot 1 + \cdots + \alpha_d \cdot 0 = \alpha_i.$$

Since i was arbitrary and we found that $\alpha_i = 0$, this implies that $\mathscr{A}'$ is linearly independent.

We now show that $\text{span}(\mathscr{A}') = X'$. Let $\psi \in X'$ be arbitrary. For each $i = 1, 2, ..., d$, let $c_i \in \mathbb{F}$ be defined by $c_i = \psi(x_i)$. We claim that $\psi = c_1 \cdot \varphi^i + \cdots + c_d \cdot \varphi^d$. Once we show this, then we will have proved that $\psi \in \text{span}(\mathscr{A}')$, from which the result will follow. Define $\tilde{\psi} = c_1 \cdot \varphi^i + \cdots + c_d \cdot \varphi^d$; we will show that $\psi = \tilde{\psi}$ by showing $\psi(y) = \tilde{\psi}(y)$ for every $y \in X$. The crux of this proof comes from the easily verified fact that $\tilde{\psi}(x_i) = c_i$ for every $i = 1, 2, ..., d$. Now, let $y \in X$ be arbitrary. Since $\mathscr{A}$ is a basis of X, we can find scalars $a^1, ..., a^d \in \mathbb{F}$ such that $y = a^1 x_1 + \cdots + a^d x_d$. We then have

$$\psi(y) = a^1\psi(x_1) + \cdots + a^d\psi(x_d) = a^1 c_1 + \cdots a^d c_d = a^1\tilde{\psi}(x_1) + \cdots + a^d\tilde{\psi}(x_d) = \tilde{\psi}(y).$$

This finishes the proof. ■

1.2. Function Spaces as Vector Spaces. Let X be a vector space over $\mathbb{F}$. Let U be any set and $\mathscr{F}(U \to X)$ be the collection of all functions with domain U and codomain X. Since X is a vector space, there is an induced vector space structure on $\mathscr{F}(U \to X)$.

THEOREM 7.1.26. *Let X be a vector space over $\mathbb{F}$. Then $\mathscr{F}(U \to X)$ is a vector space over $\mathbb{F}$ under vector addition and scalar multiplication defined pointwise. That is, for all $f, g \in \mathscr{F}(U \to X)$ and $a \in \mathbb{F}$, we define $f + g \in \mathscr{F}(U \to X)$ and $a \cdot f \in \mathscr{F}(U \to X)$ by*

$$(f + g)(u) = f(u) + g(u) \quad \text{and} \quad (a \cdot f)(u) = a \cdot f(u)$$

for all $u \in U$.

Proof. Exercise 9 (7). ■

REMARK 7.1.27. Note that in Theorem 7.1.26, the domain U has no extra structure beyond being a set, whereas the codomain space X is a vector space. Therefore, you will recognize that the fact that $\mathscr{F}(U \to X)$ is a vector space is wholly induced from the vector space structure on X. □

EXAMPLE 7.1.28. Let $n \in \mathbb{N}$. Elementary linear algebra tells us that $\mathbb{R}^n$ is a vector space over the field $\mathbb{R}$ with vector addition and scalar multiplication given componentwise. Indeed, if $(x_1, x_2, ..., x_n), (y_1, y_2, ..., y_n) \in \mathbb{R}^n$ and $a \in \mathbb{R}$, then

$$(x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$$
$$\text{and} \quad a \cdot (x_1, x_2, ..., x_n) = (ax_1, ax_2, ..., ax_n).$$

However, with the right perspective, this is nothing more than a vector space of functions as in Theorem 7.1.26. To see this, consider $\mathscr{F}(\{1, 2, ..., n\} \to \mathbb{R})$; a function $f \in \mathscr{F}(\{1, 2, ..., n\} \to \mathbb{R})$ is easily identified with the vector $(f(1), f(2), ..., f(n)) \in \mathbb{R}^n$. Similarly, a vector $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$ is easily identified with the function $x \in \mathscr{F}(\{1, 2, ..., n\} \to \mathbb{R})$ defined by $x(i) = x_i$ for all $1 \leq i \leq n$. Putting this together, we can identify $\mathbb{R}^n$ with the function space $\mathscr{F}(\{1, 2, ..., n\} \to \mathbb{R})$. Moreover, giving $\mathscr{F}(\{1, 2, ..., n\} \to \mathbb{R})$ a vector space structure as in Theorem 7.1.26, we again arrive at the componentwise addition and multiplication that is familiar to us at this point. An insight to take away from this example is that a good portion of vector spaces you will encounter (if not *all* vector spaces) can be interpreted as function spaces given a vector space structure as in Theorem 7.1.26. □

2. Normed Vector Spaces

Let X be a vector space over $\mathbb{F}$. A norm on X is a function that endows X with a notion of "length" of a vector. One can also interpret a norm as the abstraction of the absolute value. Closer to the abstraction of an absolute value, we first describe what a norm is on the field $\mathbb{F}$, which is then used to define a norm on X.

DEFINITION 7.2.1. A function $|\cdot| : \mathbb{F} \to \mathbb{R}$ is called a *norm* or an *absolute value* on $\mathbb{F}$ when it satisfies:

(1) For all $a \in \mathbb{F}$, $|a| \geq 0$.
(2) For all $a \in \mathbb{F}$, $|a| = 0$ if and only if $a = 0$.
(3) (Triangle Inequality) For all $a, b \in \mathbb{F}$, $|a + b| \leq |a| + |b|$.
(4) For all $a, b \in \mathbb{F}$, $|ab| = |a|\,|b|$.

When $\mathbb{F}$ is endowed with a norm $|\cdot|$, we say that $(\mathbb{F}, |\cdot|)$ is a *normed field*. As usual, we often omit the tuple $(\mathbb{F}, |\cdot|)$ from the notation and simply call $\mathbb{F}$ a normed vector space, implicitly identifying a norm on $\mathbb{F}$.

EXAMPLE 7.2.2. If $(\mathbb{F}, P)$ is any ordered field (here $P \subset \mathbb{F}$ is the collection of "positive elements" in $\mathbb{F}$), then there is a natural notion of absolute value defined such that if $r \in \mathbb{F}$ then

$$|r| = \begin{cases} r & \text{if } r \in P \\ -r & \text{if } r \notin P \end{cases}$$

□

EXAMPLE 7.2.3. The absolute value on $\mathbb{R}$ makes $\mathbb{R}$ into a normed field as a special case of the previous example. The non-ordered field of complex numbers $\mathbb{C}$ is a normed field with the standard norm $|z| = \sqrt{z\overline{z}}$, where $\overline{z}$ is the complex conjugate of z. □

REMARK 7.2.4. Although we are momentarily keeping our field $\mathbb{F}$ with norm $|\cdot|$ general, all of our examples will come from either $\mathbb{R}$ or $\mathbb{C}$ normed as in the previous example. □

DEFINITION 7.2.5. Let X be a vector space over the *normed* field $(\mathbb{F}, |\cdot|)$. A function $\|\cdot\| : X \to \mathbb{R}$ is called a *norm* on X when it satisfies:

(1) For all $x \in X$, $\|x\| \geq 0$.
(2) For all $x \in X$, $\|x\| = 0$ if and only if $x = 0_X$.
(3) (Triangle Inequality) For all $x, y \in X$, $\|x + y\| \leq \|x\| + \|y\|$.
(4) For all $a \in \mathbb{F}$ and $x \in X$, $\|a \cdot x\| = |a| \, \|x\|$.

When X is endowed with a norm $\|\cdot\|$, we say that $(X, \|\cdot\|)$ is a *normed vector space*. As before, we often omit the tuple $(X, \|\cdot\|)$ from the notation and simply call X a normed vector space, implicitly identifying a norm on X.

EXAMPLE 7.2.6. Let $n \in \mathbb{N}$. The *Euclidean norm* $\|\cdot\|$ on $\mathbb{R}^n$ is defined by the standard formula

$$\|(x_1, x_2, ..., x_n)\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Typically, when we want to imply that we have endowed $\mathbb{R}^n$ with the Euclidean norm, we simply call $\mathbb{R}^n$ *Euclidean space*. In fact, older texts used to use the notation E^n instead of $\mathbb{R}^n$ when they wanted to emphasize that they have equipped $\mathbb{R}^n$ with the Euclidean structure. □

EXAMPLE 7.2.7. Let $(\mathbb{F}, |\cdot|)$ be normed field. We can view $\mathbb{F}$ as a vector space over itself. Then, in comparison of Definitions 7.2.1 and 7.2.5, it turns out that $|\cdot|$ is a norm on $\mathbb{F}$ viewed as a vector space. In particular, we won't find ourselves tripping over different concepts of norms. □

2.1. Normed Vector Spaces as Metric Spaces.

THEOREM 7.2.8 (Normed spaces are metric spaces). *Suppose that $(X, \|\cdot\|)$ is a normed vector space. Let $\rho : X \times X \to \mathbb{R}$ be defined by $\rho(x, y) = \|x - y\|$. Then ρ is a metric on X. We say that ρ is the metric* induced *from the norm $\|\cdot\|$.*

Proof. Exercise 9 (9). ■

EXAMPLE 7.2.9. Endowing $\mathbb{R}^n$ with the Euclidean norm, the induced metric is exactly the Euclidean metric on $\mathbb{R}^n$. □

COROLLARY 7.2.10. *Let $(X, \|\cdot\|)$ be a normed vector space and ρ the induced metric. A sequence $\{x_n\}_{n=1}^{\infty}$ of elements from X converges to $x \in X$ with respect to ρ if and only if $\lim_{n\to\infty} \|x_n - x\| = 0$.*

Proof. This is immediate since $x_n \to x$ if and only if $\rho(x_n, x) \to 0$, but $\rho(x_n, x) = \|x_n - x\|$. ■

REMARK 7.2.11. While all normed vector spaces are metric spaces with a *canonical* metric induced from the norm, the opposite need not be true. In fact, all normed vector spaces are, in fact, vector spaces; there is no reason that a general metric space is even a vector space. So, metric spaces are certainly a more general concept that normed vector spaces, although normed vector spaces are extremely common and an important object of study. □

Given any normed vector space $(X, \|\cdot\|)$ and armed with Theorem 7.2.8, there is always a metric on X induced from the norm. We always assume that when we are given a norm on X, that X is a metric space endowed with the induced metric. From here we can inject all we've learned about metric spaces. In particular, we know what it means for sequences to converge in X; we know about the topology on X including what defines and open set, closed set, compact set, and density; we know about continuous functions between normed vector spaces. However, one question you might have is the following: how do the metric and norm "interact"? The following proposition is in this vein. The proof of the proposition will explicitly point out where the metric is being used in conjunction with the norm; however, after this we will implicitly understand this to be the case without necessarily giving it explicit mention.

PROPOSITION 7.2.12. *Let ρ be the metric induced on X from the norm $\|\cdot\|$. Then $\|\cdot\|$ is a continuous function viewed as a map between the metric space (X, ρ) and $(\mathbb{R}$, Euclidean metric$)$.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be an arbitary sequence in X which converges to some point $x \in X$. If we can show that $\lim_{n\to\infty} \|x_n\| = \|x\|$, we will conclude that $\|\cdot\| : X \to \mathbb{R}$ is continuous using Theorem 6.7.6. To do this, the reverse triangle inequality (see Exercise 9 (10)) gives

$$\big|\|x_n\| - \|x\|\big| \leq \|x_n - x\|.$$

By the assumption that $x_n \to x$, it holds that $\lim_{n\to\infty} \|x_n - x\| = 0$ using Corollary 7.2.10. With $0 \leq \big|\|x_n\| - \|x\|\big| \leq \|x_n - x\|$, a squeezing argument lets us conclude that $\lim_{n\to\infty} \big|\|x_n\| - \|x\|\big| = 0$, proving that $\lim_{n\to\infty} \|x_n\| = \|x\|$. Hence we deduce that $\|\cdot\|$ is continuous, which concludes the proof. ■

A next reasonable question is to ask how do the linear operations of vector addition and scalar multiplication interact with the induced metric. Here is a proposition giving some insight to that question.

PROPOSITION 7.2.13. *Let $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ be sequences in the normed vector space X which converge to $x \in X$ and $y \in X$, respectively. Then the sequence $\{x_n + y_n\}_{n=1}^{\infty}$ converges to $x + y$. Further, if $a \in \mathbb{F}$, then the sequence $\{a \cdot x_n\}_{n=1}^{\infty}$ converges to $a \cdot x$.*

Proof. For the first claim,

$$\|x_n + y_n - (x + y)\| = \|(x_n - x) + (y_n - y)\| \leq \|x_n - x\| + \|y_n - y\|.$$

For the second,

$$\|a \cdot x_n - a \cdot x\| = \|a(x_n - x)\| = |a| \, \|x_n - x\|.$$

The remainder of the proof is now from Exercise 9 (11). ■

2.2. Series. Working on a vector space X, we have addition defined on the vectors, so it makes sense to look at sums of finitely many of the vectors. If $(X, \|\cdot\|)$ is a normed vector space, then we can consider sums of infinitely many vectors in a similar way as we did in $\mathbb{R}$. Formally, given a sequence $\{x_n\}_{n=1}^{\infty}$ in X, a *series* $\sum_{n=1}^{\infty} x_n$ is the sum of all terms of a sequence $\{x_n\}_{n=1}^{\infty}$. However, as was the case in $\mathbb{R}$, the series $\sum_{n=1}^{\infty} x_n$ is by no means guaranteed to exist (imagine if for every n we define $x_n = (-1)^n \cdot x$ for some fixed $0_X \neq x \in X$). Instead, we

use the symbol $\sum_{n=1}^{\infty} x_n$ in much the same way as we might write $\lim_{n\to\infty} y_n$ for some sequence $\{y_n\}_{n=1}^{\infty}$; this limit $\lim_{n\to\infty} y_n$ may or may not exist, but it is convenient notation to use anyway.

DEFINITIONS 7.2.14. Let $(X, \|\cdot\|)$ be a normed vector space and let $\{x_n\}_{n=1}^{\infty}$ be a in X.

(1) For each $n \in \mathbb{N}$, the sum $s_n = \sum_{k=1}^{n} x_k$ is called the the *nth partial sum* of the sequence. The *series* $\sum_{n=1}^{\infty} x_n$ is said to *converge* when the sequence of partial sums $\{s_n\}_{n=1}^{\infty}$ converges, in which case we let the symbol $\sum_{k=1}^{\infty} x_k$ be the limit of the partial sums $\sum_{k=1}^{\infty} x_k = \lim_{n\to\infty} s_n$.

(2) We say that the series $\sum_{n=1}^{\infty} x_n$ *converges absolutely* when the series $\sum_{n=1}^{\infty} \|x_n\|$ converges in $\mathbb{R}$.

REMARK 7.2.15. Although we have only defined series to be a sum over $n \in \mathbb{N}$, we can easily generalize this notion to be a sum of the form $n \geq k$ for $k \in \mathbb{Z}$. Indeed, suppose that $k \in \mathbb{Z}$ and $\{x_n\}_{n=k}^{\infty}$ is a sequence in X, then the series $\sum_{n=k}^{\infty} x_n$ can be defined by $\sum_{n=1}^{\infty} x_{k+n-1}$. □

LEMMA 7.2.16 (Sigma Notation Properties). *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in the normed vector space X. For $n, m \in \mathbb{N}$, the following hold.*

(1) *If $m < n$ then $\sum_{k=1}^{n} x_k - \sum_{k=1}^{m} x_k = \sum_{k=m+1}^{n} x_k$.*

(2) *If $\sum_{k=1}^{\infty} x_n$ converges then $\sum_{k=1}^{\infty} x_k - \sum_{k=n+1}^{\infty} x_k = \sum_{k=1}^{n} x_k$.*

Proof. Exercise 9 (13). ■

THEOREM 7.2.17 (Cauchy Criteria). *Let $\{x_n\}_{n=1}^{\infty}$ a sequence in the normed vector space X. Let $\{s_n\}_{n=1}^{\infty}$ be the sequence of partial sums of the series $\sum_{k=1}^{\infty} x_k$. The sequence of partial sums $\{s_n\}_{n=1}^{\infty}$ is Cauchy if and only if for any $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$ and $m \geq n$, it holds that $\left\| \sum_{k=n}^{m} x_k \right\| < \varepsilon$.*

Proof. If $\{s_n\}_{n=1}^{\infty}$ is Cauchy, then given $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$ such that for all $n, m \geq n_\varepsilon$, it holds that $\|s_m - s_n\| < \varepsilon$. But then, for all $n \geq n_\varepsilon + 1$ and all $m \geq n$, we have that $\varepsilon > \|s_m - s_{n-1}\| = \left\| \sum_{k=n}^{m} x_k \right\|$ proving one direction of the claim. For the other direction, let $\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ be such that for all $n \geq n_\varepsilon$ and all $m \geq n$ it holds that $\left\| \sum_{k=n}^{m} x_k \right\| < \varepsilon$. Let $n, m \geq n_\varepsilon$ be arbitrary, and without losing generality, assume that $m > n$. Then $n + 1 \geq n_\varepsilon$ and $m \geq n + 1$, so $\|s_m - s_n\| = \left\| \sum_{k=n+1}^{m} x_k \right\| < \varepsilon$. ■

COROLLARY 7.2.18. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in the normed vector space X. If the series $\sum_{k=1}^{\infty} x_k$ converges absolutely, the the sequence of partial sums is Cauchy. This further implies that if X is complete, then any absolutely convergent series $\sum_{k=1}^{\infty} x_k$ converges.*

Proof. If $m \geq n$, then by the triangle inequality

$$\Big\| \sum_{k=n}^{m} x_k \Big\| \leq \sum_{k=n}^{m} \|x_k\|$$

Since $\sum_{n=1}^{\infty} \|x_n\|$ is a series of non-negative real numbers which is assumed to converge, given an $\varepsilon > 0$ we can find some $n_\varepsilon \in \mathbb{N}$ such that if $n \geq n_\varepsilon$ then $\sum_{k=n}^{\infty} \|x_k\| < \varepsilon$. Therefore, if $n \geq n_\varepsilon$ and $m \geq n$,

$$\sum_{k=n}^{m} \|x_k\| \leq \sum_{k=n}^{\infty} \|x_k\| < \varepsilon.$$

By Theorem 7.2.17, the sequence of partial sums is Cauchy. If we assume X is complete, this implies that the sequence of partial sums converges, which is equivalent to the convergence of the series $\sum_{k=1}^{\infty} x_k$. ■

3. Banach Spaces

DEFINITION 7.3.1. A normed vector space $(X, \|\cdot\|)$ which is complete (using the metric induced by the norm) is called a *Banach space*.

THEOREM 7.3.2. *A normed vector space $(X, \|\cdot\|)$ is a Banach space if and only if every absolutely convergent series converges.*

Proof. If we assume that $(X, \|\cdot\|)$ is a Banach space, then by Corollary 7.2.18, any absolutely convergent series converges. On the other hand, suppose that we only assume that $(X, \|\cdot\|)$ is a normed vector space and that every absolutely convergent series in $(X, \|\cdot\|)$ converges. We then need to show that $(X, \|\cdot\|)$ is complete. To this end, let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence in X. We build a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ of $\{x_n\}_{n=1}^{\infty}$ in the following way. First, let $n_1 \in \mathbb{N}$ be chosen such that for every $m \geq n_1$, $\|x_m - x_{n_1}\| < 1/2$ (we can do this since the sequence is assumed Cauchy). Next, choose $n_2 > n_1$ such that for every $m \geq n_2$, $\|x_m - x_{n_2}\| < 1/4$. Continuing on this way, for every $k \in \mathbb{N}$ we let $n_{k+1} > n_k$ be chosen such that for all $m \geq n_{k+1}$, $\|x_m - x_{n_{k+1}}\| < 2^{-(k+1)}$. This defines a subsequence of the natural numbers $n_1 < n_2 < n_3 < \cdots$ and hence a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ of $\{x_n\}_{n=1}^{\infty}$. Let's

now consider the alternate sequence $\{y_k\}_{k=1}^{\infty}$ defined by $y_k = x_{n_{k+1}} - x_{n_k}$. Note that by our choices of the n_k, $\|y_k\| < 2^{-k}$. In particular, $\sum_{k=1}^{\infty} \|y_k\| < \sum_{k=1}^{\infty} 2^{-k} < \infty$. By assumption, this means that $\sum_{k=1}^{\infty} y_k$ converges, since it converges absolutely. Now, by a telescoping series argument, $\sum_{k=1}^{K} y_k = x_{n_K} - x_{n_1}$. So, if $S = \sum_{k=1}^{\infty} y_k$, then $\lim_{K\to\infty} (x_{n_K} - x_1) = S$ which implies that $\lim_{K\to\infty} x_{n_K} = S + x_{n_1}$. Most importantly, this shows that the the subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ converges. Since $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence with a convergent subsequence, then we can conclude that $\{x_n\}_{n=1}^{\infty}$ converges as well. ■

EXAMPLE 7.3.3. Let X be a Banach space over the field $\mathbb{R}$. Suppose that $M : X \to X$ is a linear function. For any non-negative integer k, define M^k as the k-fold composition $M^k = \overbrace{M \circ M \circ \cdots \circ M}^{k}$, where we define $M^0 = I$, with I the identity map $I(x) = x$. Assume that there exists some $\lambda > 0$ such that $\|M(x)\| \leq \lambda \|x\|$ for every $x \in X$; inductively this implies that $\|M^k(x)\| \leq \lambda^k \|x\|$. Then, for any $t \in \mathbb{R}$, the function $e^{tM} : X \to X$ is well-defined pointwise by $e^{tM}(x) = \sum_{k=0}^{\infty} \frac{t^k}{k!} M^k(x)$. Indeed, for any $x \in X$ we have

$$\sum_{k=0}^{\infty} \left\| \frac{t^k}{k!} M^k(x) \right\| = \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \|M^k(x)\| \leq \sum_{k=0}^{\infty} \frac{\lambda^k |t|^k}{k!} \|x\| = \|x\| e^{\lambda |t|}.$$

This shows that for each fixed $x \in X$, the series defining $e^{tM}(x)$ converges absolutely. Since we have assumed that X is complete, Theorem 7.3.2 implies that $e^{tM}(x)$ converges pointwise. □

3.1. Aside: Pointwise and Uniform Convergence of Functions. Let U be an arbitrary set. There are many ways to consider convergence of a sequence of functions. In this section we explore two extremely common and useful such convergences.

DEFINITION 7.3.4. Let (M, ρ) be a metric space and $\{f_n\}_{n=1}^{\infty}$ a sequence of functions $f_n : U \to M$. We say that this sequence of functions converges *pointwise* to a function $f : U \to M$ when for every fixed $x \in U$, it holds that $\lim_{n\to\infty} \rho(f_n(x), f(x)) = 0$. We say that the sequence of functions converges to $f : U \to M$ *uniformly* when $\lim_{n\to\infty} \left[\sup_{x\in U} \rho(f_n(x), f(x)) \right] = 0$.

While much of the theory of pointwise and uniform convergence can be done in the setting of functions taking values in metric spaces, we will specialize to the case that the metric space is a normed vector space. Because of this, we will repeat Definition 7.3.4 using the notation from normed vector spaces.

DEFINITION 7.3.5 (Definition 7.3.4 in Normed Vector Spaces). Let $(X, \|\cdot\|)$ a normed vector space and $\{f_n\}_{n=1}^{\infty}$ a sequence of functions $f_n : U \to X$. Then this sequence of functions converges to a function $f : U \to X$ pointwise when for every fixed $x \in U$,

$$\lim_{n\to\infty} \|f_n(x) - f(x)\| = 0.$$

The sequence of functions converges to $f : U \to X$ uniformly when

$$\lim_{n\to\infty} \left[\sup_{x\in U} \|f_n(x) - f(x)\|\right] = 0.$$

EXAMPLE 7.3.6 (Pointwise and Not Uniform). For every $n \in \mathbb{N}$, let $f_n : [0,1] \to \mathbb{R}$ be defined by $f_n(x) = x^n$. We will show that f_n converges *pointwise* to the function $f : [0,1] \to \mathbb{R}$ given by

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

We will also show directly that f_n does not converge to f uniformly (even though we already know this since f is not continuous, so f_n can not converge uniformly to f otherwise we would be in violation of Corollary 7.3.17).

f_n converges to f pointwise: If $x \in [0,1)$, then we know that $x^n \to 0$ as $n \to \infty$ (since $|x| < 1$). This shows that for a fixed $x \in [0,1)$, $f_n(x) \to 0$ as $n \to \infty$; equivalently $f_n(x) \to f(x)$ as $n \to \infty$ since for $x \in [0,1)$, $f(x) = 0$. If $x = 1$, then for every n, $f_n(1) = 1^n = 1 = f(1)$, and again we can conclude that $f_n(1) \to f(1)$ as $n \to \infty$. Therefore, for every fixed $x \in [0,1]$, $f_n(x) \to f(x)$, meaning that f_n converges to f pointwise.

f_n does not converge to f uniformly: We will prove that for any $n \in \mathbb{N}$, $\sup_{x\in[0,1]} |f_n(x) - f(x)| = 1$. Using notation that will be introduced below, we will write $\|f_n - f\|_\infty$ to mean $\sup_{x\in[0,1]} |f_n(x) - f(x)|$. Hence, what we are going to show is that for every $n \in \mathbb{N}$, $\|f_n - f\|_\infty = 1$. Once we do this, we can conclude that $f_n \not\to f$ uniformly (if $f_n \to f$ uniformly, $\|f_n - f\|_\infty \to 0$, but since $\|f_n - f\|_\infty = 1$ for every n, this can not happen). For any $n \in \mathbb{N}$, $\lim_{x\to 1^-} x^n = 1$; this is true because the polynomial $g(x) = x^n$ is a continuous function. This implies that given any $\varepsilon > 0$, we can find some $x \in (0,1)$ such that $\left|1 - x^n\right| < \varepsilon$, which rearranges to $1 - \varepsilon < x^n$. Therefore, for any $n \in \mathbb{N}$, there exists some $x \in (0,1)$ such that $1 - \varepsilon < x^n = |f(x) - f_n(x)|$. Since ε was arbitrary, we conclude that $1 \leq \|f_n - f\|_\infty$. Moreover, for any $x \in [0,1]$,

$|f_n(x) - f(x)| \leq x^n \leq 1$. Hence $\|f(x) - f_n(x)\|_\infty \leq 1$. What we have then shown is that $1 \leq \|f_n - f\|_\infty \leq 1$, which concludes the proof. □

PROPOSITION 7.3.7. *Let $(X, \|\cdot\|)$ be a normed vector space. Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of functions from U to X which converge pointwise to a function $f : U \to X$ and converge uniformly to a function $g : U \to X$. Then $f = g$.*

Proof. We will prove that $f(x) = g(x)$ for every $x \in U$, which will imply that $f = g$. Let $x \in U$ be arbitrary and take $\varepsilon > 0$. Let $n \in \mathbb{N}$ be large enough that $\|f_n(x) - f(x)\| < \varepsilon/2$ and $\|f_n - g\|_\infty < \varepsilon/2$; such an n exists by our convergence assumptions. Then

$$\begin{aligned}\|f(x) - g(x)\| \leq \|f(x) - f_n(x)\| + \|f_n(x) - g(x)\| &\leq \|f(x) - f_n(x)\| + \|f_n - g\|_\infty \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we have shown that $\|f(x) - g(x)\| = 0$, proving that $f(x) = g(x)$. Since $x \in U$ was arbitrary, we're done. ■

3.2. Bounded Function Spaces as Banach Spaces in the Uniform Norm.

NOTATION 7.3.8. Define $\mathscr{F}_b(U \to X)$ as the set of all X-valued *bounded* functions with domain U. That is, $f \in \mathscr{F}_b(U \to X)$ if and only if $f : U \to X$ and $\sup_{x \in U} \|f(x)\| < \infty$.

THEOREM 7.3.9. . *$\mathscr{F}_b(U \to X)$ is a vector space. Moreover, define the function*

$$\|\cdot\|_\infty : \mathscr{F}_b(U \to X) \to \mathbb{R} \quad \text{by} \quad \|f\|_\infty = \sup_{x \in U} |f(x)|.$$

Then $\|\cdot\|_\infty$ is a norm on $\mathscr{F}_b(U \to X)$. In other words, $\big(\mathscr{F}_b(U \to X), \|\cdot\|_\infty\big)$ is a normed vector space.

Proof. Exercise 9 (19). ■

DEFINITION 7.3.10. The norm $\|\cdot\|_\infty$ defined on $\mathscr{F}_b(U \to X)$ is called the *sup-norm* or the *uniform norm*.

REMARK 7.3.11. Based on Definition 7.3.5, a sequence of functions $\{f_n\}_{n=1}^\infty$ in $\mathscr{F}_b(U \to X)$ will converge uniformly to a function $f \in \mathscr{F}_b(U \to X)$ if and only if $\{f_n\}_{n=1}^\infty$ converges to f in the uniform norm; i.e.,

$$\lim_{n \to \infty} \|f_n - f\|_\infty = 0.$$

□

THEOREM 7.3.12. *If $(X, \|\cdot\|)$ is Banach, then $\left(\mathscr{F}_b(U \to X), \|\cdot\|_\infty\right)$ is a Banach space.*

Proof. Let $\{f_n\}_{n=1}^\infty$ be a Cauchy sequence in $\mathscr{F}_b(U \to X)$ with respect to the uniform norm. We need to prove that there exists some $f \in \mathscr{F}_b(U \to X)$ such that $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$ (equivalently, we could write $\lim_{n\to\infty} f_n = f$ since we know we're using the uniform norm). Here we will find the candidate function f; in Exercise 9 (16) you will prove that $f \in \mathscr{F}_b(U \to X)$ and that $f_n \to f$ uniformly. To find our candidate $f : U \to X$, we note that for every $x \in U$, the sequence $\{f_n(x)\}_{n=1}^\infty$ is a Cauchy sequence in X. Indeed,

$$\|f_n(x) - f_m(x)\| \leq \|f_n - f_m\|_\infty$$

which implies that $\{f_n(x)\}_{n=1}^\infty$ is Cauchy in X since $\{f_n\}_{n=1}^\infty$ is Cauchy in $\mathscr{F}_b(U \to X)$. Since X is assumed complete, this means for every $x \in U$, there exists a limit of the sequence $\{f_n(x)\}_{n=1}^\infty$. Define $f : U \to X$ as this limit; i.e., $f(x) = \lim_{n\to\infty} f_n(x)$. We now have our candidate function $f : U \to X$. What remains for Exercise 9 (16) is to show that $f_n \to f$ uniformly and that f is in $\mathscr{F}_b(U \to X)$ (i.e., show that f is bounded). ■

3.3. Bounded and Continuous Function Spaces in the Uniform Norm. Let (M, ρ) be a metric space and $(X, \|\cdot\|)$ be a normed vector space over $(\mathbb{F}, |\cdot|)$. Recall that we use the notation $C(M \to X)$ to mean the continuous functions from M to X. Here continuity is implicitly using the metric ρ on M and the metric induced from the norm on X, and we know what defines a continuous function from one metric space M to another X. We will also continue with the notation $\mathscr{F}_b(M \to X)$ to mean the bounded functions from M to X, as in the previous section.

DEFINITION 7.3.13. The *bounded continuous functions* from M to X is the set

$$C_b(M \to X) = C(M \to X) \cap \mathscr{F}_b(M \to X).$$

THEOREM 7.3.14. *The set of bounded continuous functions $C_b(M \to X)$ is a vector subspace of $\mathscr{F}_b(M \to X)$. Moreover $C_b(M \to X)$ is a closed subset of $\mathscr{F}_b(M \to X)$ in the uniform norm.*

Proof. The proof that $C_b(M \to X)$ is a vector subspace of $\mathscr{F}_b(M \to X)$ is left as an exercise. We will prove here that $C_b(M \to X)$ is a closed subset of $\mathscr{F}_b(M \to X)$ in the

uniform norm. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence in $C_b(M \to X)$ which converges uniformly to some $f \in \mathscr{F}_b(M \to X)$. If we can deduce that $f \in C_b(M \to X)$, then we can conclude that $C_b(M \to X)$ is closed by Theorem 6.5.1. To this end, we need to prove is that f is continuous. Let $x_0 \in M$ and $\varepsilon > 0$. Since $\{f_n\}_{n=1}^{\infty}$ is uniformly convergent to f, there exists some $n \in \mathbb{N}$ such that $\|f_n - f\|_\infty < \varepsilon/3$. Since f_n is continuous by assumption, there exists some $\delta > 0$ such that for all $x \in B(x_0; \delta)$, $\|f_n(x) - f_n(x_0)\| < \varepsilon/3$. Therefore, for every $x \in B(x_0; \delta)$,

$$\begin{aligned}
\|f(x) - f(x_0)\| &= \|f(x) - f_n(x) + f_n(x) - f_n(x_0) + f_n(x_0) - f(x_0)\| \\
&\leq \|f(x) - f_n(x)\| + \|f_n(x) - f_n(x_0)\| + \|f_n(x_0) - f(x_0)\| \\
&\leq \|f - f_n\|_\infty + \|f_n(x) - f_n(x_0)\| + \|f_n - f\|_\infty \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.
\end{aligned}$$

Therefore, $x \in B(x_0; \varepsilon)$ implies that $\|f(x) - f(x_0)\| < \varepsilon$, proving that f is continuous at x_0. Since x_0 was arbitrary, we can conclude that f is continuous. In symbols, $f \in C(M \to X)$. Since $f \in C(M \to X)$ and $f \in \mathscr{F}_b(M \to X)$ (by assumption), then $f \in C_b(M \to X)$ and the proof is done. ■

COROLLARY 7.3.15. *If $(X, \|\cdot\|)$ is a Banach space, then $C_b(M \to X)$ is a Banach space with respect to the uniform norm.*

Proof. If $(X, \|\cdot\|)$ is a Banach space, then $\mathscr{F}_b(M \to X)$ is also a Banach space in the uniform norm. Since $C_b(M \to X)$ is a closed subset of $\mathscr{F}_b(M \to X)$, then $C_b(M \to X)$ is also complete. ■

COROLLARY 7.3.16. *Suppose that (M, ρ) is a compact metric space and $(X, \|\cdot\|)$ is a Banach space. Then the continuous functions from M to X, $C(M \to X)$, is a complete normed vector space when endowed with the uniform norm $\|\cdot\|_\infty$. In particular, if $\{f_n\}_{n=1}^{\infty}$ is a sequence of continuous functions in $C(M \to X)$ which converges to some function f uniformly, then f is a continuous function from M to X; i.e., $f \in C(M \to X)$.*

Proof. If M is compact and $f : M \to X$ is continuous, the f is bounded since $f(M)$ is a compact subset of X. What this implies is that $C(M \to X) = C_b(M \to X)$. The result now follows from Corollary 7.3.15. ■

COROLLARY 7.3.17. *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that for every $n \in \mathbb{N}$, $f_n : [a, b] \to \mathbb{R}$ is continuous and there is some function $f : [a, b] \to \mathbb{R}$ such that $f_n \to f$ uniformly. Then f is continuous.*

Proof. This is a special case of Corollary 7.3.16 with $M = [a, b]$. ∎

EXAMPLE 7.3.18. Let us revisit Example 7.3.3. In this example, we defined $e^{tM} = \sum_{k=0}^{\infty} \frac{t^k}{k!} M^k$ as a function $e^{tM} : X \to X$. Recall that $M : X \to X$ is a bounded linear function with bound $\lambda > 0$; this means that $\|M^k(x)\| \leq \lambda^k \|x\|$ holds for every $x \in X$. In Example 7.3.3 we proved that $e^{tM} : X \to X$ was well defined pointwise $e^{tM}(x) = \sum_{k=0}^{\infty} \frac{t^k}{k!} M^k(x)$, since the sequence of partial sums $f_n = \sum_{k=0}^{n} \frac{t^k}{k!} M^k$ converged pointwise (and we defined that pointwise limit to be e^{tM}). Here, we will prove that if X is a Banach space, then $e^{tM} : X \to X$ is, in fact, continuous. We will do this by showing that for every $x \in M$, there is a neighborhood U of X such that the partial sums f_n are in $C_b(U \to X)$ and are Cauchy in the uniform norm. Once this is done, since we have assumed that X is a Banach space, then by Theorem 7.3.14, $C_b(U \to X)$ is a Banach space and thusly the partial sums converge to some continuous function $f \in C_b(U \to X)$; however, since $f_n \to e^{tM}$ pointwise on X, it must be that $f = e^{tM}$ on U by Proposition 7.3.7.

To this end, let $x, y \in X$. Then $\|M(x) - M(y)\| = \|M(x - y)\| \leq \lambda \|x - y\|$, where the first equality used that M is linear. This shows that M is Lipschitz continuous and therefore continuous. Since the composition, addition, and scalar multiplication of continuous functions results in a continuous function, it follows that for each n, the partial sum $f_n = \sum_{k=0}^{n} \frac{t^k}{k!} M^k$ is continuous. Now, let $x_0 \in X$ be arbitrary. Let U be the open ball around x_0 of radius 1; i.e., $U = B(x_0; 1)$. (Note that since X is a normed vector space, it is a metric space, so any subset of X inherits the metric space structure, though not necessarily the linear [vector space] structure). For every $x \in U$, an application of the triangle inequality shows that $\|x\| \leq \|x_0\| + 1$. For convenience, define $\kappa = \|x_0\| + 1$, and note that $\|x\| \leq \kappa$ for every $x \in U$. Now, using a similar estimate as we did in Example 7.3.3, for every $x \in U$,

$$\|f_n(x)\| = \left\| \sum_{k=0}^{n} \frac{t^k}{k!} M^k(x) \right\| \leq \sum_{k=0}^{n} \frac{|t|^k \lambda^k}{k!} \|x\| \leq e^{|t|\lambda} \|x\| \leq \kappa e^{|t|\lambda}.$$

In particular, since the rightmost inequality is independent of $x \in U$, we have shown that $\|f_n\|_\infty \leq \kappa e^{|t|\lambda}$, where $\|\cdot\|_\infty$ is the uniform norm on $\mathscr{F}_b(U \to X)$. We therefore have that

each partial sum is both continuous and bounded as functions from U to X, showing that $f_n \in C_b(U \to X)$ for every n. We have left to show that $\{f_n\}_{n=1}^{\infty}$ is Cauchy in $C_b(U \to X)$ using the uniform norm. For this, let $x \in U$ be arbitrary. Let $n \leq m$ be integers. Then,

$$\|f_m(x) - f_n(x)\| \leq \sum_{k=n+1}^{m} \frac{|t|^k \lambda^k}{k!} \|x\| \leq \kappa \sum_{k=n+1}^{\infty} \frac{|t|^k \lambda^k}{k!}.$$

Since the right hand side is independent of $x \in U$, this shows that $\|f_m - f_n\|_\infty \leq \kappa \sum_{k=n+1}^{\infty} \frac{|t|^k \lambda^k}{k!}$ With this estimate in mind, let $\varepsilon > 0$ be arbitrary. Since $\sum_{k=0}^{\infty} \frac{|t|^k \lambda^k}{k!} = e^{|t|\,\lambda}$ converges, let $n_\varepsilon \in \mathbb{N}$ be large enough that for any $n \geq n_\varepsilon$ we have that $\sum_{k=n+1}^{\infty} \frac{|t|^k \lambda^k}{k!} < \varepsilon/\kappa$. Then, for any $n, m \geq n_\varepsilon$ with $m \geq n$, our previous estimate implies that

$$\|f_m - f_n\|_\infty \leq \kappa \sum_{k=n+1}^{\infty} \frac{|t|^k \lambda}{k!} < \varepsilon.$$

This proves that $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $C_b(U \to X)$ with respect to the uniform norm, and hence converges uniformly to some function $f \in C_b(U \to X)$. As previously mentioned, this function must agree with the pointwise limit e^{tM}. So, we have proved that the restriction $e^{tM} : U \to X$ is continuous. Since U is a neighborhood of x_0, this further proves that the unrestricted function $e^{tM} : X \to X$ is continuous at x_0 (see Chapter 6 Exercise (31)). Since $x_0 \in X$ was arbitrary, we have proved that $e^{tM} : X \to X$ is continuous. □

4. Inner Product Spaces

We begin our discussion of inner product and Hilbert spaces. For this, we specialize to the case that the underlying field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. For $z \in \mathbb{C}$, we will write $\overline{z}$ as the complex conjugate of z. Moreover, we consider $\mathbb{R}$ to be a subspace of $\mathbb{C}$; particularly, if $r \in \mathbb{R}$, then $\overline{r} = r$, since the complex conjugate of a real number is just that number itself. As is always the case unless otherwise explicitly mentioned, we assume that $\mathbb{R}$ and $\mathbb{C}$ are endowed with their respective standard Euclidean norm $|\cdot|$.

4.1. Basics.

Definition 7.4.1. Let X be a vector space over the field $\mathbb{F}$, which is either $\mathbb{R}$ or $\mathbb{C}$. An *inner product* on X is a function $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{F}$ which satisfies the following.

(1) For all $x, y \in X$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$.

(2) For all $a \in \mathbb{F}$ and $x, y, z \in X$, $\langle x + \alpha \cdot y, z\rangle = \langle x, z\rangle + \alpha \langle y, z\rangle$.

(3) For all $x \in X$, $\langle x, x\rangle \geq 0$.

(4) For all $x \in X$, $\langle x, x\rangle = 0$ if and only if $x = 0_X$.

When we endow X with a such an inner product, we call $(X, \langle \cdot, \cdot \rangle)$ an *inner product space*; implicitly this assumes that the field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. If $\mathbb{F} = \mathbb{R}$, we often explicitly state that $(X, \langle \cdot, \cdot \rangle)$ is a *real inner product space*, whereas if $\mathbb{F} = \mathbb{C}$, then we often state that it is a *complex inner product space* (also known as a *Hermitian inner product space* in some circles). As always, we often drop the tuple $(X, \langle \cdot, \cdot \rangle)$ and just state that X is an inner product space.

REMARK 7.4.2. The case we will become most familiar with in these notes is when X is a real inner product space. When X is a real inner product space, then property (1) of Definition 7.4.1 says that the inner product $\langle \cdot, \cdot \rangle$ is *symmetric*; that is, $\langle x, y\rangle = \langle y, x\rangle$ for any $x, y \in X$. □

LEMMA 7.4.3. *Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{F}$, which is either $\mathbb{R}$ or $\mathbb{C}$. Then for every $\alpha \in \mathbb{F}$ and $x, y, z \in X$, it holds that*

$$\langle x, y + \alpha \cdot z\rangle = \langle x, y\rangle + \overline{\alpha} \langle x, z\rangle.$$

Proof. Exercise 9 (21). ■

DEFINITION 7.4.4. Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. The function $\| \cdot \| : X \to \mathbb{R}$ defined by $\|x\| = \sqrt{\langle x, x\rangle}$. is called the *induced norm* on X.

REMARK 7.4.5. Though obvious by definition, let's point out that $\|x\|^2 = \langle x, x\rangle$. In many calculations, it is easier to consider $\|x\|^2$ rather that $\|x\|$. □

THEOREM 7.4.6 (Cauchy-Schwarz Inequality). *Let X be an inner-product space. For all $x, y \in X$,*

$$\big|\langle x, y\rangle\big| \leq \|x\| \, \|y\|.$$

Moreover, $\big|\langle x, y\rangle\big| = \|x\| \, \|y\|$ if and only if x and y are linearly dependent.

Proof. Let $x, y \in X$. Let's first prove this in the case that $\langle x, y\rangle \in \mathbb{R}$; in particular, $\langle x, y\rangle = \langle y, x\rangle$. Let $\lambda \in \mathbb{R}$. Then

$$0 \leq \langle x - \lambda y, x - \lambda y\rangle = \|x\|^2 + \lambda^2 \|y\|^2 - \lambda \langle y, x\rangle - \lambda \langle x, y\rangle = \|x\|^2 + \lambda^2 \|y\|^2 - 2\lambda \langle x, y\rangle$$

This is a concave up parabola in the variable λ with minimum found by

$$2\lambda\|y\|^2 - 2\langle x, y\rangle = 0.$$

Therefore, the minimum is $\lambda = \frac{\langle x,y\rangle}{\|y\|^2}$. Evaluating the above inequality at the minimum,

$$0 \leq \|x\|^2 + \frac{\langle x, y\rangle^2}{\|y\|^2} - 2\frac{\langle x, y\rangle^2}{\|y\|^2} = \|x\|^2 - \frac{\langle x, y\rangle^2}{\|y\|^2}.$$

Rearranging this, we arrive at

$$\langle x, y\rangle^2 \leq \|x\|^2\|y\|^2.$$

Taking the square root on both sides finishes the proof in the symmetric case $\langle x, y\rangle = \langle y, x\rangle$ (in particular, if X is a real inner product space, then we're done).

Now we tackle the case that $\langle x, y\rangle \in \mathbb{C}$. In this case, there is some $r \geq 0$ and $\theta \in [0, 2\pi)$ such that $\langle x, y\rangle = re^{i\theta}$. Let $z = e^{i\theta}y$. Then $\langle x, z\rangle = e^{-i\theta}\langle x, y\rangle = e^{-i\theta}re^{i\theta} = r$, showing that $\langle x, z\rangle \in \mathbb{R}$. Moreover $\big|\langle x, z\rangle\big| = \big|e^{-i\theta}\big|\big|\langle x, y\rangle\big| = \big|\langle x, y\rangle\big|$ and $\|z\| = |e^{i\theta}|\,\|y\| = \|y\|$. Putting this all together, since $\langle x, z\rangle \in \mathbb{R}$, we can apply the first case to find that $\big|\langle x, z\rangle\big| \leq \|x\|\,\|z\|$. But then, $\big|\langle x, y\rangle\big| = \big|\langle x, z\rangle\big| \leq \|x\|\,\|z\| = \|x\|\,\|y\|$.

The proof of equivalence between $\big|\langle x, y\rangle\big| = \|x\|\,\|y\|$ and linear dependence of x and y is an exercise (Exercise 9 (22)). ■

We next show that the induced norm is called a norm for good reason.

THEOREM 7.4.7. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner-product space. Then the induced norm is a norm on X.*

Proof. We will prove the triangle inequality holds, the other properties of a norm are left as an exercise (see Exercise 9 (23)). Before going head deep into the proof, let's make a quick observation: if $a \in \mathbb{C}$ is any complex number, then $a + \overline{a} \leq 2\,|a|$. Now, if $x, y \in X$, then

$$\begin{aligned}\|x + y\|^2 = \langle x + y, x + y\rangle &= \|x\|^2 + \|y\|^2 + \langle x, y\rangle + \langle y, x\rangle = \|x\|^2 + \|y\|^2 + \langle x, y\rangle + \overline{\langle x, y\rangle} \\ &\leq \|x\|^2 + \|y\|^2 + 2\big|\langle x, y\rangle\big| \leq \|x\|^2 + \|y\|^2 + 2\|x\|^2\|y\|^2 = \big(\|x\| + \|y\|\big)^2.\end{aligned}$$

The first inequality came from our observation, the second from the Cauchy-Schwarz inequality. By taking the square root, we arrive at

$$\|x + y\| \leq \|x\| + \|y\|$$

which is the triangle inequality. ■

REMARK 7.4.8. We now know that every inner product space is a normed vector space using the induced norm, which in turn is a metric space using the induced metric (the metric is induced from the norm, and hence induced from the inner product). Therefore, whenever one encounters an inner product space, it is implicitly understood that any results from metric spaces and normed vector spaces can be carried over, and that the norm and metric being used are those which are induced unless explicitly mentioned otherwise. Note that the induced metric $\rho : X \times X \to \mathbb{R}$ can be written in terms of the inner product by $\rho(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y\rangle}$. □

PROPOSITION 7.4.9. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space over $\mathbb{F}$. Let $x, y \in X$ and suppose that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are sequences in X such that $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$. Then $\lim_{n\to\infty} \langle x_n, y_n\rangle = \langle x, y\rangle$.*

Proof. We have

$$\begin{aligned}\big|\langle x_n, y_n\rangle - \langle x, y\rangle\big| &= \big|\langle x_n, y_n\rangle - \langle x_n, y\rangle + \langle x_n, y\rangle - \langle x, y\rangle\big| = \big|\langle x_n, y_n - y\rangle + \langle x_n - x, y\rangle\big| \\ &\leq \big|\langle x_n, y_n - y\rangle\big| + \big|\langle x_n - x, y\rangle\big| \leq \|x_n\|\,\|y_n - y\| + \|x_n - x\|\,\|y\|.\end{aligned}$$

In Proposition 7.2.12, we have proved that $\|x_n\| \to \|x\|$ as $n \to \infty$, and by assumption $\|x_n - x\| \to 0$ and $\|y_n - y\| \to 0$. Therefore $\|x_n\|\,\|y_n - y\| + \|x_n - x\|\,\|y\| \to 0$ as $n \to \infty$. With

$$0 \leq \big|\langle x_n, y_n\rangle - \langle x, y\rangle\big| \leq \|x_n\|\,\|y_n - y\| + \|x_n - x\|\,\|y\|$$

and the rightmost side converging to 0, a squeezing argument proves that

$$\big|\langle x_n, y_n\rangle - \langle x, y\rangle\big| \to 0 \text{ as } n \to \infty$$

which proves that $\lim_{n\to\infty} \langle x_n, y_n\rangle = \langle x, y\rangle$. ■

4.2. Orthogonality.

DEFINITIONS 7.4.10. Given vectors $x, y \in X$, we say that x and y are *orthogonal* when $\langle x, y\rangle = 0$. In this case it is common to write $x \perp y$. A collection of vectors $U \subset X$ are said to be *orthogonal* when $x \perp y$ for all $x, y \in U$. Further if $V \subset X$ is any collection of vectors and $x \in X$ is orthogonal to every vector in V, then it is common to write $x \perp V$ and say that x is *orthogonal to* V.

THEOREM 7.4.11. *Suppose that X is an inner product space. Let $U \subset X$ be non-empty. Define the set $U^{\perp}$ by*

$$U^{\perp} = \{x \in X \text{ s.t. } x \perp y \text{ for every } y \in U\} = \{x \in X \text{ s.t. } x \perp U\}.$$

Then

(1) $U^{\perp} \cap U = \{0_X\}$.
(2) *$U^{\perp}$ is a closed subspace of X. That is, $U^{\perp}$ is a vector subspace of X which is closed as a subset of X.*

Proof. Suppose that $x \in U \cap U^{\perp}$. Then $\langle x, x\rangle = 0$, implying that $x = 0_X$. Therefore, if such an x exists, it must be 0_X, otherwise, there is no such x.

The proof that $U^{\perp}$ is a subspace is left as an exercise (see Exercises **??** below). We now prove that $U^{\perp}$ is closed. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $U^{\perp}$ which converge to some point $x \in X$. By Proposition 7.4.9, for any $y \in U$, $\lim_{n\to\infty} \langle y, x_n\rangle = \langle y, x\rangle$, but by assumption, $\langle y, x_n\rangle = 0$ for every n, implying that $\langle y, x\rangle = 0$. Since this is true for any arbitrary $y \in U$, we conclude that $x \in U^{\perp}$ and therefore $U^{\perp}$ is closed. ■

DEFINITION 7.4.12. The set $U^{\perp}$ defined in Theorem 7.4.11 is called the *orthogonal complement* of U.

THEOREM 7.4.13 (Pythagorean Theorem). *Suppose that $x_1, x_2, ..., x_n \in X$ are orthogonal. Then*

$$\Big\| \sum_{k=1}^{n} x_k \Big\|^2 = \sum_{k=1}^{n} \|x_k\|^2$$

Proof. We have

$$\Big\| \sum_{k=1}^{n} x_k \Big\|^2 = \Big\langle \sum_{k=1}^{n} x_k, \sum_{j=1}^{n} x_j \Big\rangle = \sum_{k=1}^{n}\sum_{j=1}^{n} \langle x_k, x_j\rangle = \sum_{k=1}^{n}\sum_{j=1}^{n} \delta_{kj}\|x_k\|^2 = \sum_{k=1}^{n} \|x_k\|^2$$

where the penultimate equality comes from the orthogonality assumption and we used the *Kronecker delta* notation

$$\delta_{kj} = \begin{cases} 1 & k = j \\ 0 & \text{otherwise} \end{cases}$$

This concludes the proof. ■

THEOREM 7.4.14. *Suppose that Y is a subspace of X. Let $x \in X$ and $w_x \in Y$. Then $\|x - w_x\| \leq \|x - y\|$ for every $y \in Y$ if and only if $x - w_x \in Y^\perp$. Moreover, if such an element $w_x \in Y$ exists, then it is unique.*

Proof. Suppose first that $\|x - w_x\| \leq \|x - y\|$ for every $y \in Y$. Then for any $y \in Y$, the function $f(t) = \|x - w_x - ty\|^2$ has a minimum when $t = 0$ by assumption. Therefore $f'(0) = 0$. However, you can also calculate that $f'(0) = -\big(\langle x - w_x, y\rangle + \langle y, x - w_x\rangle\big)$, implying that $\langle x - w_x, y\rangle + \langle y, x - w_x\rangle = 0$. If $\mathbb{F} = \mathbb{R}$, we're done since $\langle x - w_x, y\rangle + \langle y, x - w_x\rangle = 2\langle x - w_x, y\rangle$ from which we can deduce that $\langle x - w_x, y\rangle = 0$. If $\mathbb{F} = \mathbb{C}$, then by arguing analogously for the function $g(t) = \|x - w_x - ity\|^2$, we will find that $\big(\langle x - w_x, y\rangle - \langle y, x - w_x\rangle\big) = 0$ also. We then add these together $\big(\langle x - w_x, y\rangle + \langle y, x - w_x\rangle\big) + \big(\langle x - w_x, y\rangle - \langle y, x - w_x\rangle\big) = 2\langle x - w_x, y\rangle$, which again implies that $\langle x - w_x, y\rangle = 0$.

Now, assume that $x - w_x$ is orthogonal to every element in Y. Then for any $y \in Y$, $\|x - y\|^2 = \|x - w_x + w_x - y\|^2 = \|x - w_x\|^2 + \|w_x - y\|^2$ by the Pythagorean theorem since $w_x - y \in Y$. Therefore $\|x - w_x\|^2 = \|x - y\|^2 - \|w_x - y\|^2 \leq \|x - y\|^2$. Taking square roots shows that $\|x - w_x\| \leq \|x - y\|$.

Finally, to prove uniqueness, let's assume that $w_1, w_2 \in Y$ are given such that $\langle x - w_1, y\rangle = 0$ and $\langle x - w_2, y\rangle = 0$ for every $y \in Y$. Then

$$\|w_1 - w_2\|^2 = \langle w_1 - x + x - w_2, w_1 - w_2\rangle = -\langle x - w_1, w_1 - w_2\rangle + \langle x - w_2, w_1 - w_2\rangle.$$

Since $w_1 - w_2 \in Y$, both the terms on the right side of the equality are zero by assumption. Therefore $\|w_1 - w_2\|^2 = 0$ implying that $w_1 = w_2$. ■

DEFINITION 7.4.15. Let Y be a subspace of X, $x \in X$ and $w_x \in Y$ be as in the previous theorem. We say that w_x is the *orthogonal projection of x onto Y*.

REMARKS 7.4.16.

(1) If we define $\operatorname{dist}(x, Y) = \inf\{\|x - y\| \text{ s.t. } y \in Y\}$ (which is standard), then Theorem 7.4.14 says that that if w_x exists, then it is the unique element in Y such that $\operatorname{dist}(x, Y) = \|x - w_x\|$. Colloquially this says that w_x is the "closest" element to x in the subspace Y. Moreover, we can interpret the equivalence of $\operatorname{dist}(x, Y) = \|x - w_x\|$ and $\langle x - w_x, y\rangle = 0$ for every $y \in Y$, as the statement: the distance from x to Y can be found by starting at x and moving perpendicularly towards Y; w_x will be the first element in Y we hit.

(2) Be careful! This theorem does *not* claim nor prove that the orthogonal projection of x onto Y actually exists. In fact, it might not! Do not fear; in the next section we will give sufficient conditions for existence.

5. Hilbert Spaces

DEFINITION 7.5.1. An inner product space which is complete (as a metric space with the induced metric) is called a *Hilbert space*.

THEOREM 7.5.2. *Assume that X is a Hilbert space and let Y be a* closed *subspace of X. Then for every $x \in X$, there exists an orthogonal projection $w_x \in Y$ of x onto Y.*

Proof. Let $x \in X$ and $d = \operatorname{dist}(x, Y) = \inf\{\|x - y\| \text{ s.t. } y \in Y\}$. We can therefore find a sequence $\{y_n\}_{n=1}^{\infty}$ in Y such that $\lim_{n\to\infty} \|y_m - x\| = d$ (see Exercises 9 (27)). By the parallelogram law (see Exercises 9 (28)),

$$\|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2 = 2\big(\|y_n - x\|^2 + \|y_m - x\|^2\big)$$

which rearranges to

$$\|y_n - y_m\|^2 = 2\big(\|y_n - x\|^2 + \|y_m - x\|^2\big) - 4\left\|\frac{1}{2}(y_n + y_m) - x\right\|$$

Since Y is a vector space, $\frac{1}{2}(y_n + y_m) \in Y$, and therefore $\|\frac{1}{2}(y_n + y_m) - x\| \leq d$ by the definition of d. Hence

$$\|y_n - y_m\|^2 \leq 2\big(\|y_n - x\|^2 + \|y_m - x\|^2\big) - 4d$$

By how we chose the sequence $\{y_n\}_{n=1}^{\infty}$, we know that $\|y_n - x\| \to d$ and $\|y_m - x\| \to d$ as $n \to \infty$ and $m \to \infty$. We therefore can deduce that $\|y_n - y_m\| \to 0$ as $n, m \to \infty$, from which we can conclude that $\{y_n\}_{n=1}^{\infty}$ is Cauchy. Since X is complete and Y is closed, Y is also complete. Thus the sequence $\{y_n\}_{n=1}^{\infty}$ converges to some point $w_x \in Y$ and we have

$$\|w_x - x\| = \lim_{n\to\infty} \|y_n - x\| = \operatorname{dist}(x, Y).$$

■

THEOREM 7.5.3 (Least Squares). *Let $Y \subset X$ be a closed subspace. Then for every $x \in X$, there exist unique elements $x_1 \in Y$ and $x_2 \in Y^{\perp}$ such that $x = x_1 + x_2$. Moreover, x_1 is the orthogonal projection of x onto Y.*

Proof. Theorem 7.5.2 tells us that the orthogonal projection of x onto Y exists. Set x_1 as the orthogonal projection of x onto Y, and let $x_2 = x - x_1$. Clearly $x = x_1 + x_2$, and by Theorem 7.5.2, $x_2 = x - x_1 \in Y^\perp$. To show that x_1 and x_2 are unique, not that if $y_1 \in Y$ and $y_2 \in Y^\perp$ such that $x = y_1 + y_2$, then $x - y_1 = y_2 \in Y^\perp$ which means that y_1 is the unique orthogonal projection of x onto Y, from which we conclude $x_1 = y_1$, and thusly $x_2 = y_2$. ■

REMARK 7.5.4. Given a vector space W, if there exist two subspaces V_1 and V_2 such that $V_1 \cap V_2 = \{0_W\}$ and every element $w \in W$ can be written uniquely as $w = w_1 + w_2$ for some $w_1 \in V_1$ and $w_2 \in V_2$, then we say that W is the *direct sum* of V_1 and V_2, denoted $W = V_1 \oplus V_2$. Hence, Theorem 7.5.3 can be restated in linear algebraic terms as: If $Y \subset X$ is a closed subspace, then $X = Y \oplus Y^\perp$. □

5.1. Least Squares in $\mathbb{R}^n$. The dot product is a real inner product on $\mathbb{R}^n$ making $\mathbb{R}^n$ into a Hilbert space. Viewing the vectors in $\mathbb{R}^n$ as column vectors, we can define the dot product via matrix multiplication as

$$\langle x, y \rangle = y^{tr} x$$

where y^{tr} is the transpose of the vector $y \in \mathbb{R}^n$. Given a matrix $A : \mathbb{R}^n \to \mathbb{R}^m$ and a vector $b \in \mathbb{R}^m$, one often wants to solve an equation of the form $Ax = b$ for the unknown variable $x \in \mathbb{R}^n$. However, if b is not in the range of A (the column space), then there is no solution to this equation. However, it is arguable that the "best" solution is the value $\tilde{x} \in \mathbb{R}^n$ such that $\|A\tilde{x} - b\| \leq \|Ax - b\|$ for all $x \in \mathbb{R}^n$. We will denote the image of A by $\text{Ran}(A)$, which is a closed subspace of $\mathbb{R}^n$.

Conducting some non-rigorous background work,

$$\begin{aligned} Ax = b \implies A^{tr}Ax = A^{tr}b &\implies x = \left(A^{tr}A\right)^{-1} A^{tr}b \\ &\implies Ax = A\left(A^{tr}A\right)^{-1} A^{tr}b. \end{aligned}$$

We needed to assume that $A^{tr}A$ was invertible, and we will continue with this assumption here. Now, certainly if we set $\tilde{x} = \left(A^{tr}A\right)^{-1} A^{tr}b$, then $A\tilde{x} \in \text{Ran}(A)$. We now show that $A\tilde{x}$ gives us the orthogonal projection of b onto the range of A. An arbitrary element in $\text{Ran}(A)$

can be written as Ax for some arbitrary $x \in X$. Then

$$\langle b - A\tilde{x}, Ax\rangle = (Ax)^{tr}(b - A\tilde{x}) = x^{tr}\big(A^{tr}b - A^{tr}A\tilde{x}\big) = x^{tr}\big(A^{tr}b - \overbrace{A^{tr}A\big(A^{tr}A\big)^{-1}}^{=I}A^{tr}b\big)$$
$$= x^{tr}\big(A^{tr}b - A^{tr}b\big) = 0.$$

Therefore $b - A\tilde{x} \in \mathrm{Ran}(A)^{\perp}$, proving that $A\tilde{x}$ is the orthogonal projection of b onto $\mathrm{Ran}(A)$. In particular, for any $Ax \in \mathrm{Ran}(A)$, $\|A\tilde{x} - b\| \leq \|Ax - b\|$ by Theorem 7.4.14.

PROPOSITION 7.5.5. *Consider Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ as Hilbert spaces endowed with the standard dot product. Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator such that $A^{tr}A$ is invertible. For any $b \in \mathbb{R}^m$, the vector $\tilde{x} = (A^{tr}A)^{-1}A^{tr}b$ is an element in $\mathbb{R}^n$ such that $\|A\tilde{x} - b\| \leq \|Ax - b\|$ for every $x \in \mathbb{R}^n$. This further implies that $\tilde{x}$ will be a true solution in the case $b \in \mathrm{Ran}(A)$.*

5.2. Orthonormal Bases.

DEFINITION 7.5.6. A set of vectors $\{x_a\}_{a\in A} \subset X$ is called *orthonormal* when they are orthogonal and $\|x_a\| = 1$ for every $a \in A$.

THEOREM 7.5.7 (Bessel's Inequality). *Suppose that for some $N = 1, 2, ..., \infty$, the collection $\{x_k\}_{k=1}^N$ is an orthonormal set in the inner product space $(X, \langle\cdot,\cdot\rangle)$. Then for every $x \in X$,*

$$\sum_{k=1}^{N} |\langle x, x_k\rangle|^2 \leq \|x\|^2.$$

In particular, the series $\sum_{k=1}^{N} |\langle x, x_k\rangle|^2$ converges in $\mathbb{R}$.

Proof. Let $n \in \mathbb{N}$ be any value $n \leq N$. Then,

$$\Big\|x - \sum_{k=1}^{n}\langle x, x_k\rangle x_k\Big\|^2 = \|x\|^2 - 2\sum_{k=1}^{n}|\langle x, x_k\rangle|^2 + \Big\|\sum_{k=1}^{n}\langle x, x_k\rangle x_k\Big\|^2$$
$$= \|x\|^2 - 2\sum_{k=1}^{n}|\langle x, x_k\rangle|^2 + \sum_{k=1}^{n}|\langle x, x_k\rangle|^2 = \|x\|^2 - \sum_{k=1}^{n}|\langle x, x_k\rangle|^2$$

Since the leftmost expression in this equality is clearly ≥ 0 we get that

$$\sum_{k=1}^{n} |\langle x, x_k\rangle|^2 \leq \|x\|^2.$$

If $N < \infty$, then we can set $n = N$ and this argument still holds, hence we are done. If $N = \infty$, what this tells us is that the partial sums of $\sum_{k=1}^{\infty} |\langle x, x_k\rangle|^2$ are bounded above by $\|x\|^2$. Since the series $\sum_{k=1}^{\infty} |\langle x, x_k\rangle|^2$ has non-negative terms, the partial sums are a monotonically increasing sequence and hence converge as $n \to \infty$ since bounded. Since limits respect non-strict inequalities,

$$\sum_{k=1}^{\infty} |\langle x, x_k\rangle|^2 = \lim_{n\to\infty} \sum_{k=1}^{n} |\langle x, x_k\rangle|^2 \leq \|x\|^2$$

which finishes the proof. ■

COROLLARY 7.5.8. *Suppose that for some $N = 1, 2, ..., \infty$, the collection $\{x_k\}_{k=1}^N$ is an orthonormal set in the Hilbert space X. Then for every $x \in X$, the sum $\sum_{k=1}^{N} \langle x, x_k\rangle x_k$ converges and*

$$\Big\| \sum_{k=1}^{N} \langle x, x_k\rangle x_k \Big\|^2 \leq \|x\|^2.$$

Proof. Let us assume that $N = \infty$ since this is the meaningful case. Take $x \in X$. Let $n, m \in \mathbb{N}$ with $n < m$. We then have

$$\Big\| \sum_{k=n}^{m} \langle x, x_k\rangle x_k \Big\|^2 = \sum_{k=n}^{m} |\langle x, x_k\rangle|^2 \|x_k\|^2 = \sum_{k=n}^{m} |\langle x, x_k\rangle|^2$$

where the first equality was using the Pythagorean theorem, and the second was by the assumption that $\|x_k\|^2 = 1$ for each k. Since Theorem 7.5.7 proves that $\sum_{k=1}^{\infty} |\langle x, x_k\rangle|^2$ converges we can apply the equality we just deduced and the Cauchy criteria for series $\sum_{k=1}^{\infty} |\langle x, x_k\rangle|^2$ to conclude that, in fact, the partial sums of $\sum_{k=1}^{\infty} \langle x, x_k\rangle x_k$ are Cauchy. Since X is assumed complete (as it is a Hilbert space), this implies that the series converges. The inequality $\big\| \sum \langle x, x_k\rangle x_k \big\|^2 \leq \|x\|^2$ is simply another application of the Pythagorean theorem and Bessel's inequality. ■

THEOREM 7.5.9 (Parseval's Identity). *Suppose that $N = 1, 2, 3, ..., \infty$ and $\{x_k\}_{k=1}^N$ is an orthonormal set in the inner product space X. For every $x \in X$, the equality $\sum_{k=1}^{N} \big|\langle x, x_k\rangle\big|^2 = \|x\|^2$ holds if and only if $x = \sum_{k=1}^{N} \langle x, x_k\rangle x_k$.*

Proof. In the proof of Bessel's inequality, we deduced

$$\left\| x - \sum_{k=1}^{n} \langle x, x_k \rangle x_k \right\|^2 = \|x\|^2 - \sum_{k=1}^{n} |\langle x, x_k \rangle|^2$$

for any integer $n \leq N$. From here it is obvious that the righthand side of the equality converges to 0 if and only if the partial sums of the series $\sum_{k=1}^{n} \langle x, x_k \rangle x_k$ converge to x, which is what we needed to show. ∎

FACT 7.5.10. Although unproven at this point (we will prove it in a Section **??** below), it turns out that given any orthonormal set $\{x_a\}_{a \in A}$ and $x \in X$, there exist *at most* countably many of the $a \in A$ such that $\langle x, x_a \rangle \neq 0$. That is, given any $x \in X$, we can find a countable subset $B \subset A$ such that $\langle x, x_a \rangle \neq 0$ if and only if $a \in B$.

Let us simply summarize the results stemming from Bessel's Inequality and Parseval's identity with Fact 7.5.10 in tow.

THEOREM 7.5.11. *Suppose that $(X, \langle \cdot, \cdot \rangle)$ is an inner product space and $\{x_a\}_{a \in A}$ is an orthonormal set. Then for every $x \in X$, it holds that $\sum_{a \in A} |\langle x, x_a \rangle|^2 \leq \|x\|^2$ with equality if and only if $x = \sum_{a \in A} \langle x, x_a \rangle$. Moreover, for every $x \in X$, the series $\sum_{a \in A} |\langle x, x_a \rangle|^2$ converges in $\mathbb{R}$, and if we further assume that X is a Hilbert space, then the series $\sum_{a \in A} \langle x, x_a \rangle x_a$ converges in X.*

DEFINITION 7.5.12. Let $\beta = \{x_a\}_{a \in A}$ be an orthonormal set in X indexed by some set A. We say that β is an *orthonormal basis* for X when for every $x \in X$

$$x = \sum_{a \in A} \langle x, x_a \rangle x_a.$$

As noted in Fact 7.5.10, even if A is uncountable, for each $x \in X$, at most countably many of the $a \in A$ leave $\langle x, x_a \rangle$ non-zero, so the sum over A truly is just a sum over those (at most countable) non-zero terms.

LEMMA 7.5.13. *Suppose that $(X, \langle \cdot, \cdot \rangle)$ is an inner product space. Let $\{x_a\}_{a \in A}$ be an orthonormal set in X. For each $a \in A$, let λ_a be scalars such that at most countably many are non-zero and $\sum_{a \in A} \lambda_a x_a$ converges. Then for any $b \in A$,*

$$\left\langle \sum_{a \in A} \lambda_a x_a, x_b \right\rangle = \sum_{a \in A} \lambda_a \langle x_a, x_b \rangle = \lambda_b.$$

Proof. Since at most countably many of the λ_a are non-zero, let's denumerate the collection $\{\lambda_k\}_{k=0}^{\infty}$ of all non-zero elements (if there are only finitely many, just augment that collection by setting the remainder to 0), and without losing generality, assume that $b = 0$. For any $n \in \mathbb{N}$, the linearity of the first component of the inner product guarantees that

$$\left\langle \sum_{k=0}^{n} \lambda_k \, x_k, x_0 \right\rangle = \sum_{k=0}^{n} \lambda_k \, \langle x_k, x_0 \rangle = \sum_{k=0}^{n} \lambda_0 \, \delta_{k,0} = \lambda_0$$

where we are using the Dirac delta function $\delta_{k,0}$ which results in 0 unless $k = 0$, in which case the result is 1. (The Dirac delta function arises by the assumption of orthonormality). Since we have assumed that the sequence of partial sums $\sum_{k=0}^{n} \lambda_k x_k$ converges as $n \to \infty$, and we've proved that we can interchange the limit and inner-product, the result now follows. ∎

THEOREM 7.5.14. *Let $\{x_a\}_{a\in A}$ be an orthonormal set in X. The following are equivalent.*

(1) *$\{x_a\}_{a\in A}$ is an orthonormal basis.*
(2) *For every $x \in X$, $\|x\|^2 = \sum_{a\in A} |\langle x, x_a\rangle|^2$.*
(3) *If $\langle x, x_a\rangle = 0$ for every $a \in A$, then $x = 0_X$.*

Proof. (1) $\iff$ (2). The equivalence of 1 and 2 follows from Theorem 7.5.9.

(2) $\implies$ (3). Exercise 9 (**??**).

(3) $\implies$ (1). Let $x \in X$ and define $y = x - \sum_{a\in A} \langle x, x_a\rangle x_a$. Using Lemma 7.5.13, for any $b \in A$,

$$\langle y, x_b\rangle = \langle x, x_b\rangle - \left\langle \sum_{a\in A} \langle x, x_a\rangle x_a, x_b \right\rangle = \langle x, x_b\rangle - \sum_{a\in A} \langle x, x_a\rangle \underbrace{\langle x_a, x_b\rangle}_{\delta_{ab}} = \langle x, x_b\rangle - \langle x, x_b\rangle = 0.$$

Therefore, by assuming (3), we deduce $y = 0_X$, implying that $x = \sum_{a\in A} \langle x, x_a\rangle x_a$. ∎

6. Bounded linear operators between Banach spaces

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces over the common field $(\mathbb{F}, |\cdot|)$ which is either $\mathbb{R}$ or $\mathbb{C}$ with the standard absolute value. Recall from linear algebra that a *linear operator* between X and Y is a function $A : X \to Y$ which satisfies

$$A(a \cdot x + y) = a \cdot A(x) + A(y)$$

for all $a \in \mathbb{F}$ and $x, y \in X$. Of course this definition makes sense as X and Y are both vector spaces. However, since X and Y are both normed spaces, we can explore more than just linear properties of linear operators.

NOTATION 7.6.1. If $A : X \to Y$ is linear, it is common to omit parenthesis around the arguments of A; that is, we often write Ax rather than $A(x)$. This notation holds over from matrix theory, since if X and Y are finite dimensional, Ax can be interpreted as matrix vector multiplication.

DEFINITION 7.6.2. Let $A : X \to Y$ be a linear operator. The *operator norm* of A is

$$\|A\|_{op} = \inf \{C > 0 \text{ s.t. } \|Ax\|_Y \leq C\|x\|_X \text{ for all } x \in X\}. \tag{6.1}$$

where, as usual, we define $\inf\{\varnothing\} = \infty$. If $\|A\|_{op} < \infty$, we say that A is a *bounded linear operator*, also commonly called a *continuous linear operator*.

THEOREM 7.6.3. *Let $A : X \to Y$ be a linear operator. The following are equivalent.*

(1) *A is a bounded linear operator.*
(2) *A is Lipschitz continuous.*
(3) *A is uniformly continuous.*
(4) *A is continuous.*
(5) *A is continuous at any point in X.*
(6) *A is continuous at 0_X.*

Proof. (1) $\implies$ (2). Suppose that $C \in \mathbb{R}$ is some positive number such that for every $x \in X$, it holds that $\|Ax\|_Y \leq C\|x\|_X$. Then, for any $x, y \in X$, it holds by linearity that $\|Ax - Ay\| = \|A(x - y)\| \leq C\|x - y\|$. This proves that if there exists such a C, then A is Lipschitz continuous.

(2) $\implies$ (3) $\implies$ (4) $\implies$ (5). Immediate.

(5) $\implies$ (6). Suppose that A is continuous at x_0. Let $\varepsilon > 0$. Take $\delta > 0$ such that for and x with $\|x - x_0\|_X < \delta$, it holds that $\|Ax - Ax_0\|_Y < \varepsilon$. Now, suppose $y \in X$ is some element such that $\|y\|_X < \delta$. Then, defining $x = y + x_0$, we have $\|x - x_0\|_X < \delta$ so $\|Ax - Ax_0\|_Y < \varepsilon$. Since A is linear, $Ax - Ax_0 = A(x - x_0) = Ay$. So, for every $y \in X$ such that $\|y\|_X < \delta$ it holds that $\|Ay\|_Y < \varepsilon$, proving that A is continuous at 0_X.

(6) $\implies$ (1). Take $\varepsilon > 0$ and let $\delta > 0$ such that if $\|x\|_X < \delta$ then $\|Ax\|_Y < \varepsilon$. Then, for any $x \in X$, by the linearity of A (and dropping all norm subscripts for sanity)

$$\|Ax\| = \frac{2\|x\|}{\delta} \left\| A\left(\tfrac{\delta}{2\|x\|}\, x\right) \right\| < \frac{2\varepsilon}{\delta}\|x\|$$

where the last inequality follows from the fact that $\|x_\delta\|_X = \delta/2 < \delta$ and hence $\|Ax_\delta\|_Y < \varepsilon$; here $x_\delta = \frac{\delta}{2\|x\|}\, x$. But then, setting $C = 2\varepsilon/\delta$, we have that for all $x \in X$, it holds that $\|Ax\|_Y \leq C\|x\|_X$, proving that A is a bounded linear operator. ∎

6.1. The Operator Norm and the Uniform Norm on the Sphere. We will assume that $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are normed spaces (over the same field $\mathbb{F}$) and often omit the distinguishing subscripts on the norms when it is clear from context with which space the norm is being used.

Definition 7.6.4. Let X be a normed vector space and define $\mathcal{S}_X$ as the *unit sphere* in X. That is, $\mathcal{S}_X = \{x \in X \text{ s.t. } \|x\| = 1\}$.

Theorem 7.6.5. *Let $A : X \to Y$ be a linear operator. By $\|A\|_\infty^{\mathcal{S}}$ let us denote the uniform norm of A over the unit sphere $\|A\|_\infty^{\mathcal{S}} = \sup\{\|Ax\|_Y \text{ s.t. } x \in \mathcal{S}_X\}$. Then $\|A\|_{op} = \|A\|_\infty^{\mathcal{S}}$. That is, the operator norm of A is equal to the uniform norm of A restricted to the unit sphere.*

Proof. Let us first assume that $\|A\|_{op} = \infty$. Then for any $k > 0$ there exists some $x \in X$ such that $\|Ax\| > k\|x\|$. In particular, this implies that $x \neq 0_X$, and so we can divide by $\|x\|$ and use the linearity of A to deduce that $\|A\hat{x}\| > k$ where $\hat{x} = x/\|x\| \in \mathcal{S}_X$. Since k was arbitrarily large, we conclude that $\|A\|_\infty^{\mathcal{S}} = \infty$. Conversely, if $\|A\|_\infty^{\mathcal{S}} = \infty$, then given any $k > 0$ there is some point $x \in \mathcal{S}_X$ such that $\|Ax\| > k$. But since $\|x\| = 1$, then $\|Ax\| > k\|x\|$. Since k was arbitrarily large, this implies that $\|A\|_{op} = \infty$. What we have thus shown is that $\|A\|_{op} = \infty$ if and only if $\|A\|_\infty^{\mathcal{S}} = \infty$.

Now, suppose that $\|A\|_\infty^{\mathcal{S}} = s < \infty$. We will show that $\|A\|_{op} = s$. Take $\varepsilon > 0$. Then for every $x \in \mathcal{S}_X$, it holds that $\|Ax\| < (s + \varepsilon)$ by the definition of s. Hence for any $x \in X$, we must have $\|A\hat{x}\| < (s + \varepsilon)$ with $\hat{x} = x/\|x\|$. Using the linearity of A, this implies that $\|Ax\| < (s + \varepsilon)\|x\|$, and since x was arbitrary, this shows that $\|A\|_{op} \leq (s + \varepsilon)$. On the other hand, there exists some $x \in \mathcal{S}_X$ such that $\|Ax\| > (s - \varepsilon)$ by the definition of s; since $\|x\| = 1$, this is equivalent to $\|Ax\| > (s - \varepsilon)\|x\|$. From this we deduce that $\|A\|_{op} > (s - \varepsilon)$. Therefore, $s - \varepsilon < \|A\|_{op} \leq s + \varepsilon$, which implies that $\|A\|_{op} = s$ since ε was arbitrary. ∎

NOTATION 7.6.6. For a function $f : X \to Y$, we will continue to use the notation $\|f\|_\infty^{\mathcal{S}}$ to denote the uniform norm restricted to the unit sphere

$$\|f\|_\infty^{\mathcal{S}} = \sup \{\|f(x)\|_Y \text{ s.t. } x \in \mathcal{S}_X\}.$$

COROLLARY 7.6.7. *The operator norm $\|\cdot\|_{op}$ behaves like a norm on bounded linear functions. That is, if A and B are bounded linear functions from X to Y, then*

(1) $\|A\|_{op} \geq 0$.
(2) $\|A\|_{op} = 0$ *if and only if* $A = \mathbf{0}$, *where* $\mathbf{0} : X \to Y$ *is the zero map* $\mathbf{0}(x) = 0_Y$.
(3) $\|A + B\|_{op} \leq \|A\|_{op} + \|B\|_{op}$.
(4) $\|\alpha A\|_{op} = |\alpha| \, \|A\|_{op}$ *for any* $\alpha \in \mathbb{F}$, *where as usual,* $\alpha A : X \to Y$ *is defined by* $(\alpha A)(x) = \alpha \cdot A(x)$.

Proof. We know that the uniform norm $\|\cdot\|_\infty^{\mathcal{S}}$ is a norm on the continuous functions $f : \mathcal{S}_X \to Y$. Since A and B are bounded linear functions $X \to Y$, they are continuous. Hence their restrictions to the unit sphere are also continuous. Since $\|\cdot\|_{op}$ and $\|\cdot\|_\infty^{\mathcal{S}}$ agree on bounded linear functions, and $\|\cdot\|_\infty^{\mathcal{S}}$ is a norm, then $\|\cdot\|_{op}$ must also behave like a norm. ■

COROLLARY 7.6.8. *Suppose that $\{A_n\}_{n=1}^\infty$ is a sequence of bounded linear functions from X to Y, and $A : X \to Y$ is another bounded linear function. Then the sequence $\{A_n\}_{n=1}^\infty$ is Cauchy in the operator norm $\|\cdot\|_{op}$ if and only if it is Cauchy with respect to the uniform norm on the sphere $\|\cdot\|_\infty^{\mathcal{S}}$. Moreover, the sequence converges to A in the operator norm if and only if the sequence converges to A uniformly on the unit sphere $\mathcal{S}_X$.*

Proof. From Theorem 7.6.5, the norms $\|\cdot\|_{op}$ and $\|\cdot\|_\infty^{\mathcal{S}}$ are the same, so they must determine the same Cauchy sequences and convergences. ■

6.2. Bounded Linear Functions as a Normed Vector Space.

DEFINITION 7.6.9. The space $\mathcal{L}_b(X \to Y)$ of *bounded linear operators* from X to Y is the set of all linear operators from X to Y which are bounded. That is,

$$\mathcal{L}_b(X \to Y) = \{A : X \to Y \text{ s.t. } A \text{ is linear and } \|A\|_{op} < \infty\}.$$

In consideration of Theorem 7.6.3, $\mathcal{L}_b(X \to Y)$ is also commonly called the *continuous linear operators* from X to Y.

THEOREM 7.6.10. *$\mathcal{L}_b(X \to Y)$ endowed with $\|\cdot\|_{op}$ is a normed vector space. Moreover, if Y is a Banach space, then so is $(\mathcal{L}_b(X \to Y), \|\cdot\|_{op})$.*

Proof. The fact $\mathcal{L}_b(X \to Y)$ is a vector space and that $\|\cdot\|_{op}$ is a norm on the vector space is left as an exercise (see Corollary 7.6.7 and Exercise 9 (34)). We will prove here that if Y is complete, then so is $\mathcal{L}_b(X \to Y)$. To that end, suppose that Y is a Banach space. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of bounded linear operators which is Cauchy with respect to $\|\cdot\|_{op}$. We need to prove that there is a bounded linear function $A : X \to Y$ such that $\lim_{n\to\infty} A_n = A$, where the limit taken in the operator norm. We will break this proof up into three claims.

Claim 1. There exists a function $A : X \to Y$ such that $A_n \to A$ pointwise.

Proof of Claim 1. Let $x \in X$. Then by the definition of the operator norm,

$$\|A_n x - A_m x\| = \|(A_n - A_m)x\| \leq \|A_n - A_m\|_{op}\,\|x\|$$

From this, since $\{A_n\}_{n=1}^{\infty}$ is Cauchy in the operator norm, it is clear that for each fixed $x \in X$, the sequence $\{A_n x\}_{n=1}^{\infty}$ is a Cauchy in Y. Since Y is complete by assumption, this sequence must converge. So, define the function $A : X \to Y$ to be defined by $Ax = \lim_{n\to\infty} A_n x$. Hence by construction, $A_n \to A$ pointwise.

Claim 2. The pointwise limit $A : X \to Y$ is bounded and linear.

Proof of Claim 2. Let $x, y \in X$ and $\alpha \in \mathbb{F}$. Then, by the definition of A, we have

$$A(x + \alpha\, y) = \lim_{n\to\infty} \big[A_n(x + \alpha y)\big] = \lim_{n\to\infty} \big[A_n x + \alpha A_n y\big] = Ax + \alpha Ay$$

where the second to last equality follows from the assumption that A_n is linear for each n, and the last equality follows from the fact that the limit of the sum of convergent sequences is equal to the sum of the limits of those sequences. This has shown that $A : X \to Y$ is linear. As for bounded, we could make a quick argument using the fact that $\|\cdot\|_{op}$ and $\|\cdot\|_{\infty}^{\mathcal{S}}$ agree, but let us instead give a more illustrative proof. The reverse triangle inequality tells us that $\big|\|A_n\|_{op} - \|A_m\|_{op}\big| \leq \|A_n - A_m\|_{op}$, which implies that the sequence of norms $\{\|A_n\|_{op}\}_{n=1}^{\infty}$ is Cauchy in our field $\mathbb{F}$ (which is either $\mathbb{R}$ or $\mathbb{C}$, and hence complete). This ensures that $\lim_{n\to\infty} \|A_n\|_{op}$ exists and moreover that $\sup_{n\in\mathbb{N}}\{\|A_n\|_{op}\} < \infty$ (see e.g., Exercise 9 (15)). It is easy to see then that $\lim_{n\to\infty} \|A_n\|_{op} \leq \sup_{n\in\mathbb{N}}\{\|A_n\|_{op}\}$. With this in mind, for any $x \in X$ we have

$$\|Ax\| = \Big\| \lim_{n\to\infty} A_n x\Big\| = \lim_{n\to\infty} \|A_n x\| \leq \lim_{n\to\infty} \|A_n\|_{op}\|x\| \leq \sup_{n\in\mathbb{N}}\{\|A_n\|_{op}\}\|x\|.$$

The second equality comes from the fact that we have shown that norms are continuous, so we can interchange the norm and the limit. By defining $C = \sup_{n\in\mathbb{N}}\{\|A_n\|_{op}\} < \infty$, we see that for any $x \in X$, it holds that $\|Ax\| \leq C\|x\|$, which proves that A is a bounded linear operator.

Claim 3. The sequence $\{A_n\}_{n=1}^{\infty}$ converges to A in the operator norm.

Proof of Claim 3. Since $\{A_n\}_{n=1}^{\infty}$ is Cauchy in the operator norm $\|\cdot\|_{op}$, then by Corollary 7.6.8, it is Cauchy in the uniform norm on the sphere $\|\cdot\|_{\infty}^{\mathcal{S}}$. From Claim 1, we have proved that $A_n \to A$ pointwise (in particular, the convergence is pointwise on $\mathcal{S}_X$), and in Exercise 9 (16) it is proved that this implies $\{A_n\}_{n=1}^{\infty}$ converges uniformly to A on the sphere $\mathcal{S}_X$. But then, since $A_n \to A$ in the norm $\|\cdot\|_{\infty}^{\mathcal{S}}$, then $A_n \to A$ in the operator norm (again, by Corollary 7.6.8). ∎

6.3. Dual Spaces. We continue to let $(X, \|\cdot\|)$ be a normed vector space over $\mathbb{F}$, which is either $\mathbb{R}$ or $\mathbb{C}$.

Definitions 7.6.11. 1) A linear operator $f : X \to \mathbb{F}$ is called a *linear functional*.
2) The space of all linear functionals on X is called the *algebraic dual* of X.
3) The space of all bounded linear functionals on X, denoted X^*, is called the *continuous dual* of X.

Remark 7.6.12. With our previous notation, since $\mathbb{F}$ is itself a normed vector space, we have $X^* = \mathcal{L}(X \to \mathbb{F})$. Often times the algebraic dual of X is denoted by X'; however, X' is not as "nice" as X^* since we know that every linear functional $f \in X^*$ is continuous, whereas we can't assume continuity for an arbitrary element in X'. From Theorem 7.6.10 in the previous section, we know that X^* is a Banach space when endowed with the operator norm, since both $\mathbb{R}$ and $\mathbb{C}$ are complete (Banach spaces). However, the algebraic dual X' makes sense when only assuming that X is a vector space without any notion of a norm, whereas the continuous dual X^* uses the norm structure on X to "choose" which linear functionals are continuous. □

7. Bounded linear operators between Hilbert spaces

7.1. Reisz Representation on Hilbert Spaces.

Proposition 7.7.1. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. Given any $x \in X$, the map $\varphi_x : X \to \mathbb{F}$ given by $\varphi_x(y) = \langle x, y\rangle$ is a linear functional on X. Moreover, $\varphi_x \in X^*$ and $\|\varphi_x\|_{op} = \|x\|$.*

Proof. That φ_x is a linear functional on X follows directly from the definition of an inner product. To prove the upper bound for $\|\varphi_x\|_{op}$, let $y \in X$ with $y \neq 0$, then

$$\big|\varphi_x(y)\big| = \big|\langle x, y\rangle\big| \leq \|x\|\|y\|.$$

The last inequality was an application of the Cauchy-Schwarz inequality. Since $y \in X$ was arbitrary, (6.1) implies that $\|\varphi_x\|_{op} \leq \|x\|$. However, $\varphi_x(x) = \langle x, x\rangle = \|x\|^2$, which means that $\|\varphi_x\|_{op} \geq \|x\|$. We thusly conclude that $\|\varphi_x\|_{op} = \|x\|$. ∎

THEOREM 7.7.2 (Riesz Representation – Hilbert space version). *Suppose that $(X, \langle\cdot,\cdot\rangle)$ is a Hilbert space and $f : X \to \mathbb{F}$. Then $f \in X^*$ if and only if there is some element $z_f \in X$ such that $f(x) = \langle x, z_f\rangle$ for every $x \in X$. Moreover, if $f \in X^*$ and $z_f \in X$ is such that $f(x) = \langle x, z_f\rangle$, then z_f is the unique and $\|f\|_{op} = \|z_f\|$.*

Proof. For any $z \in X$, Proposition 7.7.1 ensures that $f : X \to \mathbb{F}$ defined by $f(x) = \langle x, z\rangle$ is bounded linear functional (i.e. $f \in X^*$). Conversely, suppose that $f \in X^*$. Let's find an elements $z_f \in X$ such that $f(x) = \langle x, z_f\rangle$. If $f(x) = 0$ for every $x \in X$, then set $z_f = 0_X$. If not, then the orthogonal complement of $\text{Null}(f) = f^{-1}(\{0\})$ is a one-dimensional subspace of X. Let $0 \neq x, y \in \text{Null}(f)^\perp$ and we find some $\lambda \in \mathbb{F}$ such that $\lambda \cdot x = y$ (we can find such a λ since x and y are in the same one-dimensional space). Therefore, $f(y) = \lambda f(x)$ and $\langle y, x\rangle = \lambda\|x\|^2$. Solving for lambda we find

$$\frac{f(y)}{f(x)} = \frac{\langle y, x\rangle}{\|x\|^2} \implies f(y) = \left\langle y, \frac{\overline{f(x)}}{\|x\|^2} \cdot x \right\rangle.$$

Therefore, let's try $z_f = \frac{\overline{f(x)}}{\|x\|^2} \cdot x$. Since $\text{Null}(f)$ is a closed subspace of X (see Exercises **??**), by Theorem 7.5.3, $X = \text{Null}(f) \oplus \text{Null}(f)^\perp$. Hence given any $y \in X$, there exist $y_1 \in \text{Null}(f)$ and $y_2 \in \text{Null}(f)^\perp$ such that $y = y_1 + y_2$. Then by our previous calculation,

$$f(y) = f(y_1) + f(y_2) = 0 + \langle y_2, z_f\rangle.$$

Also,

$$\langle y, z_f\rangle = (y_1, z_f) + (y_2, z_f) = 0 + (y_2, z_f).$$

What we have thus shown is that given any $y \in X$, $f(y) = \langle y, z_f\rangle$.

To show that z_f is unique, suppose that $w \in X$ is such that $f(x) = \langle x, w\rangle$ for every $x \in X$. Then given any $x \in X$, $0 = f(x) - f(x) = \langle x, z_f\rangle - \langle x, w\rangle = \langle x, z_f - w\rangle$. Hence for

all $x \in X$, $\langle x, z_f - w\rangle = 0$, which implies that $z_f - w = 0_X$, proving that $z_f = w$. The fact that $\|f\|_{op} = \|z_f\|$ is proved in Proposition 7.7.1. ■

7.2. Adjoints. We continue to let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$, and suppose that X is a vector space over $\mathbb{F}$.

THEOREM 7.7.3. *Let $(X, \langle\cdot,\cdot\rangle)$ be a Hilbert space and suppose that $A \in \mathscr{L}_b(X \to X)$ is a bounded linear operator from X to itself. Then there exists a unique bounded linear operator $A^* \in \mathscr{L}_b(X \to X)$ such that $\langle A^*x, y\rangle = \langle x, Ay\rangle$ holds for all $x, y \in X$.*

Proof. Fix $x \in X$. Define the operator $f_x : X \to \mathbb{F}$ by $f_x(y) = \langle Ay, x\rangle$. As the composition of linear operators, f_x is a linear operator. Moreover, using Proposition 7.7.1, $\|f_x\|_{op} = \|Ax\|$, showing that f_x is a bounded linear operator from X to $\mathbb{F}$. Hence we can apply the Riesz representation theorem to conclude that there is some element $z_x \in X$ such that $f_x(y) = \langle y, z_x\rangle$. Hence for all $x \in X$, there is an element z_x such that $\langle Ay, x\rangle = \langle y, z_x\rangle$. This then defines the map $A^* : X \to X$ by $A^*x = z_x$. By construction $\langle Ay, x\rangle = \langle y, A^*x\rangle$, and by the uniqueness in the Riesz representation theorem A^* is the unique map with this property. To prove that A^* is linear, let $x, y, z \in X$ and $\alpha \in \mathbb{F}$. We then have

$$\langle x, A^*(y+\alpha z)\rangle = \langle Ax, y+\alpha z\rangle = \langle Ax, y\rangle + \overline{\alpha}\langle Ax, z\rangle = \langle x, A^*y\rangle + \overline{\alpha}\langle x, A^*z\rangle = \langle x, A^*y + \alpha A^*z\rangle$$

Since x was arbitrary, this proves that $A^*(y + \alpha z) = A^*y + \alpha A^*z$ and hence A^* is linear. Finally, to prove that A^* is bounded, let $x, y \in X$ be arbitrary. Then

$$|\langle x, A^*y\rangle| = |\langle Ax, y\rangle| \leq \|Ax\|\|y\| \leq \|x\|\,\|A\|_{op}\,\|y\|$$

where the penultimate inequality was due to Cauchy-Schwarz. By Exercise 9 (**??**), this implies that $\|A^*y\| \leq \|A\|_{op}\,\|y\|$ which implies that A^* is a bounded linear operator; in fact, it shows that $\|A^*\|_{op} \leq \|A\|_{op}$. ■

DEFINITION 7.7.4. The operator A^* in Theorem 7.7.3 is called the *adjoint* of A.

THEOREM 7.7.5. *Let $(X, \langle\cdot,\cdot\rangle)$ be a Hilbert space and suppose that $A, B \in \mathscr{L}_b(X \to X)$ are bounded linear operators. Then the following hold.*

(1) $(AB)^* = B^*A^*$.
(2) $(A^*)^* = A$.
(3) *If A is invertible, then so is A^* and, in fact, $(A^*)^{-1} = (A^{-1})^*$.*

(4) $\|A\|_{op} = \|A^*\|_{op}$.

(5) $\|A^*A\|_{op} = \|A\|_{op}^2$.

Proof. The proof of (1), (2), and (3) are Exercise 9 (38).

(4) In the proof of Theorem 7.7.3 we showed that $\|A^*\|_{op} \leq \|A\|_{op}$. Hence $\|(A^*)^*\|_{op} \leq \|A^*\|_{op}$. Since $(A^*)^* = A$, the result easily follows.

(5) Let $x, y \in X$ be arbitrary. An application of the Cauchy-Schwarz inequality shows

$$|\langle x, A^*Ay\rangle = |\langle Ax, Ay\rangle| \leq \|Ax\|\,\|Ay\| \leq \|x\|\,\|A\|_{op}^2\,\|y\|$$

By Exercise 9 (35), this implies that $\|A^*Ay\| \leq \|A\|_{op}^2\|y\|$, which further implies that $\|A^*A\|_{op} \leq \|A\|_{op}^2$. By Exercise 9 (36) and Theorem 7.6.5,

$$\|A^*A\|_{op} \geq \sup_{\|x\|=1} \langle x, A^*Ax\rangle = \sup_{\|x\|=1} \langle Ax, Ax\rangle = \sup_{\|x\|=1} \|Ax\|^2 = \|A\|_{op}^2.$$

Therefore $\|A^*A\|_{op} = \|A\|_{op}^2$. ∎

8. A first look at product spaces

8.1. Products of Metric Spaces.

Definition 7.8.1. Let $(M_1, \rho_1), (M_2, \rho_2), ..., (M_n, \rho_n)$ be metric spaces. The *product metric space* is the metric space (M, ρ) where M is the Cartesian product of the spaces M_k

$$M = M_1 \times M_2 \times \cdots \times M_n = \{(x_1, x_2, ..., x_n) \text{ s.t. } x_k \in M_k \text{ for every } k = 1, 2, ..., n\}$$

and the metric ρ on M, called the *product metric*, is defined by

$$\rho\big((x_1, ..., x_n), (y_1, ..., y_n)\big) = \sqrt{\rho_1(x_1, y_1)^2 + \rho_2(x_2, y_2)^2 + \cdots + \rho_n(x_n, y_n)^2}.$$

The metric ρ is often denoted $\rho = \rho_1 \times \rho_2 \times \cdots \times \rho_n$. If $\mathbf{x} \in M$, then for each $k = 1, 2, ..., n$, there is some $x_k \in M_k$ such that $\mathbf{x} = (x_1, x_2, ..., x_n)$. We say that x_k is the kth *component* of $\mathbf{x}$.

Example 7.8.2. You can consider $\mathbb{R}^n$ with the Euclidean metric as the product metric space of n copies of $\mathbb{R}$ each endowed with the standard distance. Indeed,

$$\mathbb{R}^n = \{(x_1, x_2, ..., x_n) \text{ s.t. } x_k \in \mathbb{R} \text{ for every } k = 1, 2, ..., n\} = \overbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}^{n \text{ copies}}$$

and

$$\rho\big((x_1, x_2, ..., x_n), (y_1, y_2, ..., y_n)\big) = \sqrt{|y_1 - x_1|^2 + |y_2 - x_2|^2 + \cdots + |y_n - x_n|^2}$$

where we certainly recognize that $|y_k - x_k|$ is the standard distance between $x_k, y_k \in \mathbb{R}$. □

NOTATION 7.8.3. For the remainder of this section, let (M, ρ) be the product metric space of $(M_1, \rho_1), (M_2, \rho_2), ..., (M_n, \rho_n)$.

PROPOSITION 7.8.4. *The product metric space (M, ρ) is a metric space. Moreover, the following inequalities hold*

$$\max\{\rho_k(x_k, y_k) : 1 \leq k \leq n\} \leq \rho(\mathbf{x}, \mathbf{y}) \leq \sqrt{n} \cdot \max\{\rho_k(x_k, y_k) : 1 \leq k \leq n\} \tag{8.1}$$

and

$$\rho(\mathbf{x}, \mathbf{y}) \leq \sum_{j=1}^{n} \rho_j(x_j, y_j) \leq n \cdot \max\{\rho_k(x_k, y_k) : 1 \leq k \leq n\}. \tag{8.2}$$

for every $\mathbf{x} = (x_1, x_2, ..., x_n), \mathbf{y} = (y_1, y_2, ..., y_n) \in M$.

THEOREM 7.8.5. *Suppose that $\{\mathbf{x}_m\}_{m=1}^{\infty}$ is a sequence in M, and for each $k = 1, 2, ..., n$, let $x_m^{(k)}$ be the kth component of $\mathbf{x}_m$; i.e. $\mathbf{x}_m = (x_m^{(1)}, x_m^{(1)}, ..., x_m^{(n)})$. Then $\{\mathbf{x}_m\}_{m=1}^{\infty}$ is Cauchy in (M, ρ) if and only if $\{x_m^{(k)}\}_{m=1}^{\infty}$ is Cauchy in (M_k, ρ_k) for every $1 \leq k \leq n$. Moreover, $\{\mathbf{x}_m\}_{m=1}^{\infty}$ converges to some point $\mathbf{x} = (x_1, x_2, ..., x_n) \in M$ if and only if the sequence converges componentwise. That is, $\lim_{m\to\infty} \mathbf{x}_m = \mathbf{x}$ with respect to ρ if and only if $\lim_{m\to\infty} x_m^{(k)} = x_k$ with respect to ρ_k for every $1 \leq k \leq n$.*

Proof. We have

$$\max\{\rho_k(x_m^{(k)}, x_k) : 1 \leq k \leq n\} \leq \rho(\mathbf{x_m}, \mathbf{x}) \leq \sqrt{n} \cdot \max\{\rho_k(x_m^{(k)}, x_k) : 1 \leq k \leq n\}.$$

From this inequality, everything follows. ■

EXAMPLE 7.8.6. Consider the sequence $\{\mathbf{x}_m\}_{m=1}^{\infty}$ in $\mathbb{R}^3$ given by $\mathbf{x}_m = \big((1+2/m)^m, 5^{-m}, 1\big)$. In the notation of the previous theorem, $x_m^{(1)} = (1 + 2/m)^m$, $x_m^{(2)} = 5^{-m}$, and $x_m^{(3)} = 1$. We know that $\lim_{m\to\infty} x_m^{(1)} = e^2$, $\lim_{m\to\infty} x_m^{(2)} = 0$, and $\lim_{m\to\infty} x_m^{(3)} = 1$; we can therefore deduce that $\mathbf{x}_m \to (e^2, 0, 1)$ as $m \to \infty$. On the other hand, consider the sequence $\{\mathbf{y}_m\}_{m=1}^{\infty}$ in $\mathbb{R}^3$ given by $\mathbf{y}_m = \big(1/m, (-1)^m, e + 2^{-m}\big)$. Although, $1/m \to 0$ and $e + 2^{-m} \to e$ both have limits as

$m \to \infty$, $(-1)^m$ does not converge to a limit, which prevents the sequence $\{\mathbf{y}_m\}_{m=1}^{\infty}$ from converging to a limit. □

COROLLARY 7.8.7. *If for all $k = 1, 2, ..., n$, (M_k, ρ_k) is a complete metric space, then so is (M, ρ).*

Proof. If $\{\mathbf{x}_m\}_{m=1}^{\infty}$ is a Cauchy sequence in (M, ρ) and $\mathbf{x}_m = (x_m^{(1)}, ..., x_m^{(n)})$, then for every $1 \leq k \leq n$, $\{x_m^{(k)}\}_{m=1}^{\infty}$ is Cauchy in (M_k, ρ_k). Therefore, if each of the (M_k, ρ_k) are complete, there is some $x_k \in M_k$ such that $\lim_{m\to\infty} x_m^{(k)} = x_k$. This then implies that $\lim_{m\to\infty} \mathbf{x}_m = (x_1, x_2, ..., x_m)$, showing that the arbitrary Cauchy sequence has a limit in M, thus proving completeness. ■

COROLLARY 7.8.8. *Suppose that for each $k = 1, 2, ..., n$, $U_k \subset M_k$ is a closed subset with respect to ρ_k. Then $U = U_1 \times U_2 \times \cdots \times U_n \subset M$ is a closed subset with respect to ρ.*

Proof. Let $\{\mathbf{x}_m\}_{m=1}^{\infty}$ be a sequence of elements in U which converges to some point $\mathbf{x} = (x_1, ..., x_n) \in M$. Then for every $1 \leq k \leq n$, $\lim_{m\to\infty} x_m^{(k)} = x_k$ where, as usual, $\mathbf{x}_m = (x_m^{(1)}, ..., x_m^{(n)})$. Since for each k, U_k is assumed closed, this implies that $x_k \in U_k$, further implying that $(x_1, ..., x_n) \in U_1 \times \cdots \times U_n$. Therefore, $\mathbf{x} \in U$ from which we conclude that U is closed. ■

THEOREM 7.8.9. *For each $k = 1, 2, ..., n$, let $U_k \subset M_k$ be open subsets (with respect to the corresponding metrics). Then $U_1 \times U_2 \times \cdots \times U_n$ is an open subset in M with respect to ρ. Moreover, $V \subset M$ is open with respect to ρ if and only if for every $\mathbf{x} = (x_1, x_2, ..., x_n) \in V$ there exist open subsets $V_k \subset M_k$ such that $x_k \in V_k$ for every $1 \leq k \leq n$ and $V_1 \times V_2 \times \cdots \times V_n \subset V$.*

Proof. Exercise 9 (39). ■

THEOREM 7.8.10. *Suppose that for every $k = 1, 2, ..., n$, M_k is compact with respect to ρ_k, then M is compact with respect to ρ.*

Proof. We will prove that M is sequentially compact, which will complete the proof by Theorem 6.6.11. Notice that if $n = 1$, then $(M, \rho) = (M_1, \rho_1)$ and is compact by assumption. Assume $n = 2$, then $M = M_1 \times M_2$ and $\rho = \rho_1 \times \rho_2$. Let $\{\mathbf{x}_m\}_{m=1}^{\infty}$ be any sequence in M. For each $m \in \mathbb{N}$, let $y_m \in M_1$ and $z_m \in M_2$ such that $\mathbf{x}_m = (y_m, z_m)$. Since

M_1 is compact, it is sequentially compact, and therefore there is a subsequence $\{y_{m_k}\}_{k=1}^{\infty}$ of $\{y_m\}_{m=1}^{\infty}$ which converges to some $y \in M_1$. Now, $\{\mathbf{x}_{m_k}\}_{k=1}^{\infty} = \{(y_{m_k}, z_{m_k})\}_{k=1}^{\infty}$ is a subsequence of $\{\mathbf{x}_m\}_{m=1}^{\infty}$ in which the first component converges. We have that $\{z_{m_k}\}_{k=1}^{\infty}$ is a sequence in M_2, which is compact, so there exists a further subsequence $\{z_{m_{k_l}}\}_{l=1}^{\infty}$ which converges to some point $z \in M_2$. Moreover, since $\{y_{m_{k_l}}\}_{l=1}^{\infty}$ is a subsequence of the convergent sequence $\{y_{m_k}\}_{k=1}^{\infty}$, then $y_{m_{k_l}} \to y$ as $l \to \infty$. We now have constructed a subsequence $\{\mathbf{x}_{m_{k_l}}\}_{l=1}^{\infty} = \{(y_{m_{k_l}}, z_{m_{k_l}})\}_{l=1}^{\infty}$ in which both components converge, $y_{m_{k_l}} \to y$ and $z_{m_{k_l}} \to z$ as $l \to \infty$. By Theorem 7.8.5, this implies that $\mathbf{x}_{m_{k_l}} \to (y, z)$ as $l \to \infty$. What we have then shown is that given an arbitrary sequence $\{\mathbf{x}_m\}_{m=1}^{\infty}$, we can find a subsequence which converges in M, proving that M is sequentially compact and thus compact.

We finish up the case for general $n \geq 2$ by induction. Suppose that given some $n \geq 2$ and and n compact metric spaces $(M_1, \rho_1), ..., (M_n, \rho_n)$, the product metric space is compact. Let $(M_1, \rho_1), ..., (M_{n+1}, \rho_{n+1})$ be compact metric spaces and let (M, ρ) be the product metric space. Let (N, d) be the product metric space: $N = M_1 \times \cdots \times M_n$ and $d = \rho_1 \times \cdots \times \rho_n$. By the inductive hypothesis, N is compact. Now $M = N \times M_{n+1}$ and $\rho_{n+1} = d \times \rho_{n+1}$. Hence we can realize that M is the product metric space of two compact metric spaces, and we just proved that the product space of two compact metric spaces is compact. We thus conclude that (M, ρ) is compact. ■

DEFINITION 7.8.11. Suppose that U is a set and $f : U \to M = M_1 \times \cdots \times M_n$ is a function. Then there are functions $f_k : U \to M_k$ such that $f(u) = (f_1(u), f_2(u), ..., f_n(u))$ for every $u \in U$. The f_k are called the *component functions* (or simply *components*) of f.

THEOREM 7.8.12. *Let (N, d) be another metric space and $f : N \to M$. Then f is continuous at $x_0 \in N$ as a map between the metric spaces (N, d) and (M, ρ) if and only if each component $f_k : N \to M_k$ is continuous x_0 as a map between the metric spaces (N, d) and (M_k, ρ_k). In particular, f is continuous if and only if each component is continuous.*

Proof. f is continuous at x_0 if and only if $\{f(x_m)\}_{m=1}^{\infty}$ converges to $f(x_0)$ whenever $\{x_m\}_{m=1}^{\infty}$ converges to x_0. By Theorem 7.8.5, $\{f(x_m)\}_{m=1}^{\infty}$ converges to $f(x_0)$ if and only if each component converges $\lim_{m \to \infty} f_k(x_m) = f_k(x_0)$. In turn, the component sequences $\{f_k(x_m)\}_{m=1}^{\infty}$ converge to $f_k(x_0)$ whenever $x_m \to x_0$ if and only if each component is continuous. Following this trail of equivalences, we arrive at the desired result. ■

EXAMPLE 7.8.13. Endow $[0, 2\pi]$ and $\mathbb{R}^3$ with the appropriate Euclidean metrics. Let $f : [0, 2\pi] \to \mathbb{R}^3$ be defined as the spiral $f(t) = (\cos(t), \sin(t), t)$. The first and second components of $f(t)$ are the trigonometric functions $\cos(t)$ and $\sin(t)$, the third is the linear map t. Since each component of f is continuous (using the respective Euclidean metrics), we can be assured that f itself is continuous. □

8.2. Products of Normed and Inner Product Spaces.

DEFINITION 7.8.14. Suppose that $X_1, ..., X_n$ are all vector spaces over the common field $\mathbb{F}$. The product space X is the Cartesian product $X = X_1 \times \cdots \times X_n$ made into a vector space (over $\mathbb{F}$) by component wise addition and scalar multiplication. That is, if $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ are both in X, and $\alpha \in \mathbb{F}$, then

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, ..., x_n + y_n) \quad \text{and} \quad \alpha\,\mathbf{x} = (\alpha\, x_1, ..., \alpha\, x_n).$$

Moreover, for any $x_k \in X_k$, the *canonical extension* of x_k to the vector $\mathbf{x}_k \in X$ is $\mathbf{x}_k = (0, ..., x_k, 0, ..., 0)$, where x_k is in the kth component.

DEFINITION 7.8.15. Suppose that $(X_1, \|\cdot\|_1), ..., (X_n, \|\cdot\|_n)$ are normed vector spaces over the common field $\mathbb{F}$. The *product normed space* is the normed vector space $(X, \|\cdot\|)$ where X is the product space and the *product norm* $\|\cdot\|$ is defined as

$$\|\mathbf{x}\| = \sqrt{\|x_1\|_1^2 + \cdots + \|x_n\|_n^2}$$

for every $\mathbf{x} = (x_1, ..., x_n) \in X$.

PROPOSITION 7.8.16. *Let $\rho_1, ..., \rho_n$ be the induced metrics on $(X_1, \|\cdot\|_1), ..., (X_n, \|\cdot\|_n)$, respectively, and let $(X, \|\cdot\|)$ be the product normed space. The product metric $\rho = \rho_1 \times \cdots \times \rho_n$ on the product space X is the same as the induced metric from the norm $\|\cdot\|$. In other words, the product metric space (X, ρ) created from $(X_1, \rho_1), ..., (X_n, \rho_n)$ is the same metric space as induced from the product normed space $(X, \|\cdot\|)$.*

Proof. Exercise 9 (40). ■

At this point, everything we know from product metric spaces can now be applied to product normed spaces. One particularly nice result is the following.

COROLLARY 7.8.17. *If $(X_1, \|\cdot\|_1), ..., (X_n, \|\cdot\|_n)$ are all Banach spaces, then so is the product normed space $(X, \|\cdot\|)$.*

DEFINITION 7.8.18. Suppose that $(X_1, \langle\cdot,\cdot\rangle_1), ..., (X_n, \langle\cdot,\cdot\rangle_n)$ are all inner product spaces over the common field $\mathbb{F}$. The *product inner product space* is the inner product space $(X, \langle\cdot,\cdot\rangle)$ where X is the product space and the inner product $\langle\cdot,\cdot\rangle$ is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle x_1, y_1 \rangle_1 + \cdots + \langle x_n, y_n \rangle_n$$

for $\mathbf{x} = (x_1, ..., x_n), \mathbf{y} = (y_1, ..., y_n) \in X$.

PROPOSITION 7.8.19. *Let $\|\cdot\|_1, ..., \|\cdot\|_n$ be the induced norms on $(X_1, \langle\cdot,\cdot\rangle_1), ..., (X_n, \langle\cdot,\cdot\rangle_n)$, respectively, and let $(X, \langle\cdot,\cdot\rangle)$ be the product inner product space. The product norm $\|\cdot\|$ on X defined in Definition 7.8.15 is the same as the induced norm from the inner product $\langle\cdot,\cdot\rangle$. In other words, the product normed space $(X, \|\cdot\|)$ created from $(X_1, \|\cdot\|_1), ..., (X_n, \|\cdot\|_n)$ is the same normed space as induced from the product inner product space $(X, \langle\cdot,\cdot\rangle)$. In light of Proposition 7.8.16, this further implies that the product metric on X and induced metric on X are the same.*

Proof. Exercise 9 (41). ■

PROPOSITION 7.8.20. *Suppose that $U_1, ..., U_n$ are orthonormal bases for the Hilbert spaces $(X_1, \langle\cdot,\cdot\rangle_1), ..., (X_n, \langle\cdot,\cdot\rangle_n)$, respectively. For each k, let $\tilde{U}_k$ be the vectors in U_k canonically extended to the product space X as in Definition 7.8.14. Then the set $U = \bigcup_{j=1}^{n} \tilde{U}_j$ is an orthonormal basis for the product Hilbert space $(X, \langle\cdot,\cdot\rangle)$.*

Proof. Exercise 9 (**??**) proves that U is an orthonormal set. So, using Parseval's identity, if we can show that for every $\mathbf{x} \in X$, it holds that $\mathbf{x} = \sum_{\mathbf{u}\in U} \langle \mathbf{x}, \mathbf{u} \rangle \mathbf{u}$, then we'll be done. To this end, let $\mathbf{x} = (x_1, ..., x_n) \in X$. Note that by assumption, we have that

$$\mathbf{x}_1 = (x_1, 0, ..., 0) = \sum_{u\in U_1} \langle x_1, u \rangle_1 (u, 0, ..., 0) = \sum_{\mathbf{u}\in \tilde{U}_1} \langle \mathbf{x}_1, \mathbf{u} \rangle \mathbf{u},$$

$$\mathbf{x}_2 = (0, x_2, ..., 0) = \sum_{u\in U_2} \langle x_2, u \rangle_1 (0, u, ..., 0) = \sum_{\mathbf{u}\in \tilde{U}_2} \langle \mathbf{x}_2, \mathbf{u} \rangle \mathbf{u},$$

$$\cdots$$

$$\mathbf{x}_n = (0, 0, ..., x_n) = \sum_{u\in U_n} \langle x_n, u \rangle_n (0, 0, ..., u) = \sum_{\mathbf{u}\in \tilde{U}_n} \langle \mathbf{x}_n, \mathbf{u} \rangle \mathbf{u}.$$

Since $\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_n$, we see that

$$\mathbf{x} = \sum_{k=1}^{n} \sum_{\mathbf{u} \in \tilde{U}_k} \langle \mathbf{x}_k, \mathbf{u} \rangle \, \mathbf{u} = \sum_{\mathbf{u} \in U} \langle \mathbf{x}, \mathbf{u} \rangle \, \mathbf{u}$$

which concludes the proof. ■

EXAMPLE 7.8.21. Consider $\mathbb{R}^n$ as the Hilbert space created as the product space of n copies of $\mathbb{R}$, where each copy of $\mathbb{R}$ is endowed with the standard inner product (multiplication). Since $\{1\} \subset \mathbb{R}$ is an orthonormal basis of $\mathbb{R}$, Proposition 7.8.20 tells us that $\{(1, 0, ..., 0)\} \cup \{(0, 1, 0, ..., 0)\} \cup \cdots \cup \{(0, 0, ..., 1)\}$ is an orthonormal basis for $\mathbb{R}^n$; a fact we are very familiar with. □

9. Exercises

(1) Prove Proposition 7.1.4.

(2) Prove Proposition 7.1.7.

(3) Prove the "other direction" in Proposition 7.1.9.

(4) Prove Lemma 7.1.11.

(5) Prove Lemma 7.1.15.

(6) Prove Lemma 7.1.16.

(7) Prove Theorem 7.1.26.

(8) Let $\ell^2(\mathbb{R})$ be the collection of all sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathbb{R}$ such that $\sum_{n=1}^{\infty} x_n^2 < \infty$. Interpret this space $\ell^2(\mathbb{R})$ as a function space by identifying the sequence $\{x_n\}_{n=1}^{\infty}$ with the function $f : \mathbb{N} \to \mathbb{R}$ under the identification $f(n) = x_n$. For each $k \in \mathbb{N}$, let $f_k \in \ell^2(\mathbb{R})$ be given by

$$f_k(n) = \delta_{kn} = \begin{cases} 1 & k = n \\ 0 & \text{otherwise} \end{cases}$$

Show that $\{f_k\}_{k=1}^{\infty}$ has the property that for every $f \in \ell^2(\mathbb{R})$ there are countably many scalars $x_1, x_2, ... \in \mathbb{R}$ such that

$$f = \sum_{k=1}^{\infty} x_k \cdot f_k$$

However, prove that $\{f_k\}_{k=1}^{\infty}$ is not a basis of $\ell^2(\mathbb{R})$.

(9) Prove Theorem 7.2.8.

(10) Prove the reverse triangle inequality for norms: If $(X, \|\cdot\|)$ is a normed vector space and $x, y \in X$, then $\big|\|x\| - \|y\|\big| \leq \|x - y\|$.

(11) Finish the proof of Proposition 7.2.13

(12) Prove the claims of Examples 7.2.6, 7.2.7, and 7.2.9. *Hint* for Example 7.2.6: the hardest part is the triangle inequality; for $n = 2$, it might help to show that $\|x + y\|^2 \leq \big(\|x\| + \|y\|\big)^2$ from which the inequality will follow; for $n \geq 2$, try induction.

(13) Prove Lemma 7.2.16.

(14) Prove that if $\sum_{k=1}^{\infty} x_k$ is a convergent series in the normed vector space $(X, \|\cdot\|)$, then given any $\varepsilon > 0$ there exists an $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$ it holds that $\Big\| \sum_{k=n}^{\infty} x_k \Big\| < \varepsilon$. *Hint*: Note that if $\{s_n\}_{n=1}^{\infty}$ are the partial sums whose limit is s, then by Lemma 7.2.16, $\|s - s_n\| = \Big\| \sum_{k=n+1}^{\infty} x_k \Big\|$.

(15) Suppose that $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence in the normed vector space X. Prove that this sequence is bounded. That is, prove that there is some $k > 0$ such that for every $n \in \mathbb{N}$, we have $\|x_n\| < k$.

(16) Assume that X is a Banach space, let $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $\mathscr{F}_b(U \to X)$, and assume $f : U \to X$ is the pointwise limit of this sequence; i.e., $f(x) = \lim_{n\to\infty} f_n(x)$. We will prove that this implies $f_n \to f$ uniformly and that $f \in \mathscr{F}_b(U \to X)$. This will prove that a uniformly Cauchy sequence which converges pointwise must, in fact, converge uniformly.

   (a) Prove that $f_n \to f$ uniformly. That is, prove that $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$.

   (b) Prove that $f \in \mathscr{F}_b(U \to X)$. Note that you already know that $f : U \to X$ exists, so you only need to prove that $f : U \to X$ is bounded.

   Possible Outline: For (a), assume that $\|f_n - f\|_\infty$ does not converge to 0 as $n \to \infty$. Therefore there exists some $\varepsilon > 0$ such that for any $N \in \mathbb{N}$ we can find some $n \geq N$ such that $\|f_n - f\|_\infty > \varepsilon$. Let $n_\varepsilon \in \mathbb{N}$ be chosen such that for every $n, m \geq n_\varepsilon$ it holds that $\|f_n - f_m\|_\infty < \varepsilon/2$ (remember we are assuming the sequence is Cauchy in the uniform norm). By the contradictive assumption, we can find some $n_1 \geq n_\varepsilon$ such that $\|f_{n_1} - f\|_\infty > \varepsilon$. Justify that this implies that there is some $x_0 \in U$ such that $\|f_{n_1}(x_0) - f(x_0)\| > \varepsilon$. From here, justify that there exists some $n_2 \geq n_\varepsilon$ such that $\|f_{n_2}(x_0) - f(x_0)\| < \varepsilon/2$. Arrive at the contradiction we've been working towards in

the previous two parts by proving that $\varepsilon < \|f_{n_1}(x_0) - f(x_0)\| \leq \varepsilon$. For (b), notice $\sup_{x \in U} \|f(x)\| \leq \|f - f_n\|_\infty + \|f_n\|_\infty$.

(17) In a paper by the famous mathematician Weierstrass, the function $f : \mathbb{R} \to \mathbb{R}$ was introduced with

$$f(x) = \sum_{k=0}^{\infty} r^k \cos(s^k \pi\, x)$$

where $r \in (0, 1)$ and s is an odd integer. Prove that $f : \mathbb{R} \to \mathbb{R}$ exists and is continuous. *Hint*: Let $a_k(x) = r^k \cos(s^k \pi\, x)$. Show that the sequence of functions $\{a_k\}$ is in $C_b(\mathbb{R} \to \mathbb{R})$ and that the series $\sum_{k=0}^{\infty} a_k$ converges converges absolutely in the $\|\cdot\|_\infty$ norm; hence use Corollary 7.3.15 and Theorem 7.3.2. Alternatively, prove Exercise 9 (18) and apply that result here.

Note: An interesting thing about the function $f : \mathbb{R} \to \mathbb{R}$ is that it is continuous, but nowhere differentiable!

(18) Suppose that $\{a_k\}_{k=1}^{\infty}$ are a sequence of continuous functions $a_k : M \to X$, where (M, ρ) is some metric space and $(X, \|\cdot\|)$ is a Banach space. Suppose further that the series $\sum_{k=1}^{\infty} \|a_k\|_\infty$ converges. Show that the function $f : M \to X$ defined by $f = \sum_{k=1}^{\infty} a_k$ exists and is continuous. *Hint*: Show that the sequence of functions $\{a_k\}_{k=1}^{\infty}$ is in $C_b(M \to X)$. Hence use Corollary 7.3.15 and Theorem 7.3.2.

(19) Prove Theorem 7.3.9.

(20) (a) For every $n \in \mathbb{N}$, let $f_n : \mathbb{R} \to \mathbb{R}$ be defined by $f_n(x) = \sin(x/n)$. Prove that the sequence $\{f_n\}_{n=1}^{\infty}$ converges pointwise to the zero function, but not uniformly. (The "zero function" is the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 0$).

(b) For every $n \in \mathbb{N}$, let $f_n : \mathbb{R} \to \mathbb{R}$ be defined by $f_n(x) = \sin(x)/n$. Prove that the function $f : \mathbb{R} \to \mathbb{R}$ converges uniformly to the zero function.

(21) Prove Proposition 7.4.3.

(22) Finish the proof of Theorem 7.4.6. That is, prove that if $x, y \in X$ such that $|\langle x, y\rangle| = \|x\|\,\|y\|$, then x are y are linearly dependent. Note that x and y are linearly dependent if and only if there are scalars a and b such that $a \cdot x = b \cdot y$.

(23) Finish the proof of Theorem 7.4.7. That is, prove that the induced norm $\|\cdot\|$ on an inner product space is a norm (the proof of the triangle inequality is given in Theorem 7.4.7).

(24) Consider $\mathbb{R}^n$ as a vector space over $\mathbb{R}$. For $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, ..., y_n) \in \mathbb{R}^n$, define

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Show that $\langle \cdot, \cdot \rangle$ is a (real) inner product on $\mathbb{R}^n$ and find an explicit formula for the induced norm. (Note that the inner product here is just the standard dot product you're used to).

(25) Let $a, b \in \mathbb{R}$ with $a < b$. For $f, g \in C([a, b] \to \mathbb{R})$ (which is a vector space over $\mathbb{R}$), define

$$\langle f, g \rangle_2 = \int_a^b f(x) g(x)\, dx.$$

Prove that $\langle \cdot, \cdot \rangle_2$ is a real inner product on $C([a, b] \to \mathbb{R})$ and find an explicit formula for the induced norm.

(26) Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. Let $U \subsetneq X$ be a non-empty proper subset. Prove that $U^\perp$ is a closed subspace of X.

(27) Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. Let $Y \subset X$ be a non-empty subset and $x \in X$. Let $d = \operatorname{dist}(x, Y) = \inf \{\|x - y\| \text{ s.t. } y \in Y\}$. Prove that there exists a sequence $\{y_n\}_{n=1}^\infty$ in Y such that $\lim_{n \to \infty} \|y_n - x\| = d$.

(28) Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. Prove the *parallelogram law*: For any $x, y \in X$,

$$\|x + y\|^2 + \|x - y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big).$$

Also, explicitly show how the parallelogram law was used in the proof of Theorem 7.5.2.

(29) Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. Let $v \in X$ be a non-zero vector and $Y = \operatorname{span}\{v\}$. Then Y is a 1-dimensional subspace of X (you can picture Y as the line passing through the origin in the direction of v). Let $x \in X$ and show that the orthogonal projection of x onto Y is given by

$$w_x = \frac{\langle x, v \rangle}{\langle v, v \rangle}\, v.$$

(30) Let $\ell^2 = \ell^2(\mathbb{R})$ be the set of all sequences $(x_1, x_2, x_3, ...)$ in $\mathbb{R}$ such that $\sum_{k=1}^\infty x_n^2$ converges. (We are intentionally writing sequences here to look like vectors, since

we conceptualize ℓ^2 as a subspace of $\mathbb{R}^n$ with $n = \infty$). We can identify a sequence $(x_1, x_2, x_3, ...)$ with a function $x : \mathbb{N} \to \mathbb{R}$ by $x(k) = x_k$. Under this identification, an element $x \in \ell^2$ is a function $x : \mathbb{N} \to \mathbb{R}$ such that $\sum_{k=1}^{\infty} x(k)^2 < \infty$. As with most function spaces, we consider ℓ^2 a vector space (in this case over $\mathbb{R}$) with vector addition and scalar multiplication done componentwise.

(a) Prove by induction that for any $n \in \mathbb{N}$ and real numbers $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$, that

$$\Big(\sum_{k=1}^{n} x_k y_k\Big)^2 \leq \Big(\sum_{k=1}^{n} x_k^2\Big)\Big(\sum_{k=1}^{n} y_k^2\Big)$$

It might help to make the base case $n = 2$.

(b) Prove that ℓ^2 is closed under scalar multiplication and under vector addition. Note: ℓ^2 is closed under scalar multiplication when for every $a \in \mathbb{R}$ and $x \in \ell^2$, then $a \cdot x \in \ell^2$; similarly, ℓ^2 is closed under vector addition when for all $x, y \in \ell^2$, $x + y \in \ell^2$.

(c) Define the function $\langle \cdot, \cdot \rangle : \ell^2 \times \ell^2 \to \mathbb{R}$ by

$$\langle x, y \rangle = \sum_{k=1}^{\infty} x(k) y(k)$$

Prove that $\langle \cdot, \cdot \rangle$ is an inner product on ℓ^2. *Note:* This is the same formula as the dot product on $\mathbb{R}^n$ in the case $n = \infty$.

(d) For $x, y \in \ell^2$, write an explicit formula for the Cauchy-Schwarz inequality

$$\big|\langle x, y \rangle\big| \leq \|x\| \, \|y\|.$$

(e) For every $k \in \mathbb{N}$, let $f_k \in \ell^2(\mathbb{R})$ defined by

$$f_k(m) = \delta_{km} = \begin{cases} 1 & k = m \\ 0 & k \neq m \end{cases}$$

for every $m \in \mathbb{N}$. Prove that $\{f_k\}_{k=1}^{\infty}$ is an orthonormal basis for ℓ^2 with respect to the inner product given in part c.

(31) Suppose that X is a Hilbert space. Prove that if A is a finite set, then an orthonormal basis $\{x_a\}_{a \in A}$ for X is a vector space basis for X. On the other hand, prove that if

A is a countable set, then an orthonormal basis $\{x_a\}_{a\in A}$ can not be a vector space basis of X. *Hint:* If A is countable, then we can denumerate the elements in A as $\{a_1, a_2, a_3, ...\} = A$; define the element $x = \sum_{k=1}^{\infty} 2^{-k} x_{a_k}$. Show that $x \in X$, but $x \notin \text{span}\{x_a\}_{a\in A}$.

(32) Suppose that X is a Hilbert space and that $\{x_k\}_{k=1}^{\infty}$ is an orthonormal basis. Prove that the set $\{x_k\}_{k=1}^{\infty}$ is *total*. That is, prove that the subspace $Y = \text{span}\{x_k\}_{k=1}^{\infty}$ is dense in X. *Hint*: Consider the partial sums in Parseval's identity and show that every element in X is the limit of elements in Y.

(33) Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces. Suppose that $A : X \to Y$ is a bounded linear function. Prove that

$$\|A\|_{op} = \sup\left\{ \frac{\|Ax\|_Y}{\|x\|_X} \text{ s.t. } x \in X, x \neq 0_X \right\}.$$

Hint. It might be easier to show that $\|A\|_{\infty}^{\mathcal{S}}$ is equal to the righthand side and hence use Theorem 7.6.5.

(34) In reference to Theorem 7.6.10, prove that $\mathcal{L}_b(X \to Y)$ is a normed vector space when endowed with the norm $\|\cdot\|_{op}$. For ease, look at Corollary 7.6.7.

(35) Let $(X, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Let $y \in X$. Prove the following.

(a) If $k > 0$ is such that for all $x \in X$, it holds that $|\langle x, y\rangle| \leq k\|x\|$, then $\|y\| \leq k$.

(b) $\|y\| = \sup\{|\langle x, y\rangle| \text{ s.t. } \|x\| = 1\}$.

(36) Let $(X, \langle\cdot,\cdot\rangle)$ be a Hilbert space and suppose that $A \in \mathscr{L}_b(X \to X)$ is a bounded linear operator. Prove that

$$\sup\{|\langle Ax, x\rangle| \text{ s.t. } \|x\| = 1\} \leq \|A\|_{op}.$$

(37) In the proof of Theorem 7.7.3, we proved that given a bounded linear operator $A \in \mathscr{L}_b(X \to X)$, there is a unique bounded linear operator $A^* \in \mathscr{L}_b(X \to X)$ such that $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for every $x, y \in X$. However, in the statement of the theorem the claim was that $\langle A^*x, y\rangle = \langle x, Ay\rangle$ for every x, y. Prove that these are equivalent. That is, prove that $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for all x and y if and only if $\langle A^*x, y\rangle = \langle x, Ay\rangle$ for all x and y.

(38) Prove (1), (2), and (3) of Theorem 7.7.5.

(39) Prove Theorem 7.8.9.

(40) Prove Proposition 7.8.16.

(41) Prove Proposition 7.8.19.

CHAPTER 8

Differential Calculus in $\mathbb{R}^n$

1. Some Notes on the Structure of $\mathbb{R}^n$

Throughout this chapter we will be working in the Euclidean spaces $\mathbb{R}^n$ for $n \in \mathbb{N}$. We consider $\mathbb{R}^n$ as the n-fold product space $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$. Since each copy of $\mathbb{R}$ in the product caries with it an inner-product (simple multiplication), a norm (the standard absolute value), and a distance (the absolute value of the difference), in turn $\mathbb{R}^n$ inherits an inner product, norm, and metric as constructed in Chapter 7 Section 8. The formula for these structures on $\mathbb{R}^n$ should be familiar; they give rise to the standard dot-product, and sum of squares distance formula we are used to from elementary calculus classes. Moreover, since each copy of $\mathbb{R}$ is complete, Corollary 7.8.7 implies that $\mathbb{R}^n$ is complete. We summarize these statements here.

THEOREM 8.1.1. *$\mathbb{R}^n$ is a Hilbert space (over the field $\mathbb{R}$) when endowed with the inner product*

$$\langle \mathbf{x}, \mathbf{y} \rangle_n = \sum_{k=1}^{n} x_k y_k = x_1 y_1 + \cdots + x_n y_n$$

where $\mathbf{x} = (x_1, ..., x_n), \mathbf{y} = (y_1, ..., y_n) \in \mathbb{R}^n$. The corresponding norm $\| \cdot \|_n$ on $\mathbb{R}^n$ is thusly

$$\|\mathbf{x}\|_n = \sqrt{\sum_{k=1}^{n} x_k^2} = \sqrt{x_1^2 + \cdots + x_n^2}$$

and hence the metric ρ_n on $\mathbb{R}^n$ is

$$\rho_n(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_n = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2} = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

NOTATION 8.1.2. We will often be considering situations where we are working with several Euclidean spaces simultaneously, such as considering functions $f : \mathbb{R}^n \to \mathbb{R}^m$ for some $n, m \in \mathbb{N}$. Because of this, we will sometimes write the inner product, norm, and metric in $\mathbb{R}^n$ as $\langle \cdot, \cdot \rangle_n$, $\| \cdot \|_n$, and ρ_n respectively; we used this notation in Theorem 8.1.1. The subscript n

tells us which dimension Euclidean space we are conducting the operation in. For example, if $f : \mathbb{R}^n \to \mathbb{R}^m$, a statement such as $\|\mathbf{x}\|_n < \delta \implies \|f(\mathbf{x})\|_m < \varepsilon$ accentuates where the norms are taking place. For similar reasons, we will also frequently write 0_n or 0_m to be the zero vector in $\mathbb{R}^n$ or $\mathbb{R}^m$, respectively.

DEFINITION 8.1.3. Let $n \in \mathbb{N}$. For each $k \in \{1, ..., n\}$, we let $\mathbf{e}_k = (0, ..., 0, 1, 0, ..., 0) \in \mathbb{R}^n$ be the vector in $\mathbb{R}^n$ with 1 in the kth component and 0 elsewhere. The set $\{\mathbf{e}_k\}_{k=1}^n$ is called the *standard basis* in $\mathbb{R}^n$.

PROPOSITION 8.1.4. *The standard basis $\{\mathbf{e}_k\}_{k=1}^n$ is a orthonormal basis of $\mathbb{R}^n$.*

Proof. One could notice that $\{1\}$ is clearly an orthonormal basis of $\mathbb{R}$ and then use Proposition 7.8.20 to conclude this result. However, it is an easy matter to check this directly (see Exercise 1). ■

Before beginning, let us make a quick note about the components of a function f with codomain $\mathbb{R}^m$.

PROPOSITION 8.1.5. *Let U be a set and $f : U \to \mathbb{R}^m$. For each $k = 1, 2, ..., m$ let $f_k : U \to \mathbb{R}$ be the kth component of f; i.e. $f(x) = (f_1(x), f_2(x), ..., f_m(x))$ for every $x \in U$. Let $\{\mathbf{e}_k\}_{k=1}^m$ be the standard basis in $\mathbb{R}^m$. Then*

$$f(x) = \sum_{k=1}^{m} f_k(x)\, \mathbf{e}_k.$$

Moreover, each component f_k can be found by

$$f_k(x) = \langle f(u), \mathbf{e}_k \rangle_m$$

which further implies that

$$f(x) = \sum_{k=1}^{m} \langle f(x), \mathbf{e}_k \rangle \cdot \mathbf{e}_k.$$

Proof. Exercise 5 (2) ■

1.1. Heine-Borel in $\mathbb{R}^n$. In the general theory of metric spaces, we learned that a subset of a metric space is compact if and only if it is complete and totally bounded; this was Theorem 6.6.9. However, in $\mathbb{R}$ we learned a simpler (and easier to check) criteria for compactness through the Heine-Borel Theorem, which said that a subset of $\mathbb{R}$ is compact if

and only if it is closed and bounded. Fortunately, this familiar version of the Heine-Borel Theorem holds over in general Euclidean space.

THEOREM 8.1.6 (Heine-Borel Theorem - Several Dimensions). *For any $n \in \mathbb{N}$, a subset $U \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Proof. Since $\mathbb{R}^n$ is complete, any subset of $\mathbb{R}^n$ is complete if and only if it is closed by Theorem 6.5.3. If $U \subset \mathbb{R}^n$ is bounded then there exists some $R > 0$ such that for every $(x_1, ..., x_n) \in U$, $\max\{|x_i| \text{ s.t. } 1 \leq i \leq n\} < R$. In particular, U is a subset of the rectangle $\mathcal{R} = [-R, R] \times \cdots \times [-R, R]$. Since $[-R, R]$ is a compact subset of $\mathbb{R}$, $\mathcal{R}$ is compact by Theorem 7.8.10. In particular, since U is a subset of a totally bounded set, it is totally bounded. Since totally bounded implies bounded, we have shown that U is bounded if and only if it is totally bounded. To summarize, U is complete if and only if it is closed and bounded if and only if it is totally bounded. By Theorem 6.6.9, the proof is finished. ■

REMARK 8.1.7. In the proof of the Heine-Borel theorem, we actually proved two facts that could stand alone as their own propositions:

PROPOSITION 8.1.8. *A subset $U \subset \mathbb{R}^n$ is complete if and only if it is closed.*

and

PROPOSITION 8.1.9. *A subset $U \subset \mathbb{R}^n$ is bounded if and only if it is totally bounded.*

These propositions themselves are a manifestation of how "nice" the structure of $\mathbb{R}^n$ is compared to more general metric spaces. □

2. The Derivative

DEFINITION 8.2.1. Let $n, m \in \mathbb{N}$ and U be an open subset in $\mathbb{R}^n$. A function $f : U \to \mathbb{R}^m$ is said to be differentiable at $\mathbf{x}_0 \in U$ when there exists a linear operator $\Lambda_{\mathbf{x}_0} : \mathbb{R}^n \to \mathbb{R}^m$ such that for any sequence $\{\mathbf{h}_k\}_{k=1}^{\infty}$ in $\mathbb{R}^n \setminus \{0_n\}$ converging to 0_n,

$$\lim_{k\to\infty} \frac{\|f(\mathbf{x}_0 + \mathbf{h}_k) - f(\mathbf{x_0}) - \Lambda_{\mathbf{x}_0}\mathbf{h}_k\|_m}{\|\mathbf{h}_k\|_n} = 0. \tag{2.1}$$

In this case we denote $\Lambda_{\mathbf{x}_x}$ by $f'(\mathbf{x}_0)$, and call $f'(\mathbf{x}_0)$ the *derivative* of f at $\mathbf{x}_0$. If f is differentiable at all points in $\mathbb{R}^n$, we say that f is *differentiable*.

REMARKS 8.2.2. (1) Note well! If f is differentiable, it seems sensible to write f' withholding the argument until later used. This is OK, but you must realize that as a function,

$$f' : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m).$$

Particularly, f' is *not* generally a linear operator; only once f' is evaluated at a point $\mathbf{x} \in \mathbb{R}^n$ are we guaranteed a linear operator $f'(\mathbf{x}) \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$.

(2) It is more common to use "continuous limit" notation rather than "sequential limit" notation when defining the limit. That is, we say that $\Lambda_{\mathbf{x}_0}$ is the derivative of f at x_0 when

$$\lim_{\mathbf{h}\to 0} \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x_0}) - \Lambda_{\mathbf{x}_0}\mathbf{h}\|_m}{\|\mathbf{h}\|_n} = 0.$$

To rigorously understand this continuous limit perspective, realize that by defining $g : \mathbb{R}^n \to \mathbb{R}$ as

$$g(\mathbf{h}) = \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x_0}) - \Lambda_{\mathbf{x}_0}\mathbf{h}\|_m}{\|\mathbf{h}\|_n},$$

the continuous limit perspective is simply the limit of the function g as $\mathbf{h} \to 0$, if that limit exists. However, both perspectives agree, so you are free to consider whichever suits you.

PROPOSITION 8.2.3. *Let U be an open subset of $\mathbb{R}^n$ and suppose $f : U \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0$. Then f is continuous at $\mathbf{x}_0$.*

THEOREM 8.2.4. *Let U be an open subset of $\mathbb{R}^n$ and suppose $f : U \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0$. Then the following hold.*

(1) *Suppose that $g : U \to \mathbb{R}^m$ is also differentiable at $\mathbf{x}_0$. Let $h : U \to \mathbb{R}^m$ be defined as $h(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$. Then h is differentiable at $\mathbf{x}_0$ and*

$$h'(\mathbf{x}_0) = f'(\mathbf{x}_0) + g'(\mathbf{x}_0).$$

(2) *([Inner] Product Rule) Suppose that $g : U \to \mathbb{R}^m$ is also differentiable at $\mathbf{x}_0$. Let $h : U \to \mathbb{R}$ be defined as $h(\mathbf{x}) = \langle g(\mathbf{x}), f(\mathbf{x})\rangle_m$. Then h is differentiable at $\mathbf{x}_0$ and*

$$h'(\mathbf{x}_0)(\cdot) = \langle g'(\mathbf{x}_0)(\cdot), f(\mathbf{x}_0)\rangle_m + \langle g(\mathbf{x}_0), f'(\mathbf{x}_0)(\cdot)\rangle_m.$$

This notation means that for any $\mathbf{y} \in \mathbb{R}^n$,

$$h'(\mathbf{x}_0)\mathbf{y} = \langle g'(\mathbf{x}_0)\mathbf{y}, f(\mathbf{x}_0)\rangle_m + \langle g(\mathbf{x}_0), f'(\mathbf{x}_0)\mathbf{y}\rangle_m$$

In particular, in the case $\mathbb{R}^m = \mathbb{R}$,

$$h'(\mathbf{x}_0)\mathbf{y} = [g'(\mathbf{x}_0)\mathbf{y}]\, f(\mathbf{x}_0) + g(\mathbf{x}_0)\, [f'(\mathbf{x}_0)\mathbf{y}].$$

(3) *(Chain Rule) Suppose that* V *is an open subset of* $\mathbb{R}^m$ *containing* $f(\mathbf{x}_0)$ *and* $g : V \to \mathbb{R}^p$ *is differentiable at* $f(\mathbf{x}_0)$. *Let* $h = g \circ f$. *Then* h *is differentiable at* $\mathbf{x}_0$ *and*

$$h'(\mathbf{x}_0) = g'(f(\mathbf{x}_0)) \circ f'(\mathbf{x}_0)(\cdot).$$

This notation means that for any $\mathbf{y} \in \mathbb{R}^n$,

$$h'(\mathbf{x}_0)\mathbf{y} = g'(f(\mathbf{x}_0))f'(\mathbf{x}_0)\mathbf{y}$$

Proof. (1) Define $\Lambda_{\mathbf{x}_0} : \mathbb{R}^n \to \mathbb{R}^m$ by

$$\Lambda_{\mathbf{x}_0} = f'(\mathbf{x}_0) + g'(\mathbf{x}_0).$$

That is, for any $\mathbf{y} \in \mathbb{R}^n$, $\Lambda_{\mathbf{x}_0}\mathbf{y} = f'(\mathbf{x}_0)\mathbf{y} + g'(\mathbf{x}_0)\mathbf{y}$. Moreover, $\Lambda_{\mathbf{x}_0}$ is linear as it is the sum of linear operators. For any $\mathbf{h} \in \mathbb{R}^n$ sufficiently near 0_n such that $\mathbf{x}_0 + h \in U$, the triangle inequality yields

$$\begin{aligned}
&\frac{\|f(\mathbf{x}_0 + \mathbf{h}) + g(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - g(\mathbf{x}_0) - \Lambda_{\mathbf{x}_0}\mathbf{h}\|_m}{\|\mathbf{h}\|_n} \\
&\leq \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)\mathbf{h}\|_m}{\|\mathbf{h}\|_n} + \frac{\|g(\mathbf{x}_0 + \mathbf{h}) - g(\mathbf{x}_0) - g'(\mathbf{x}_0)\mathbf{h}\|_m}{\|\mathbf{h}\|_n}.
\end{aligned}$$

By assumption, the latter two terms converge to 0 as $\mathbf{h} \to 0_n$. This proves that $h = f + g$ is differentiable at $\mathbf{x}_0$ and $h'(\mathbf{x}_0) = \Lambda_{\mathbf{x}_0}$.

(2) Let $\mathbf{h} \in \mathbb{R}^n$ be near enough to 0_n such that $\mathbf{x}_0 + \mathbf{h} \in U$. By appropriately adding 0,

$$\begin{aligned}
&\big\langle g(\mathbf{x}_0 + \mathbf{h}), f(\mathbf{x}_0 + \mathbf{h})\big\rangle - \big\langle g(\mathbf{x}_0), f(\mathbf{x}_0)\big\rangle \\
&\qquad = \big\langle g(\mathbf{x}_0 + \mathbf{h}) - g(\mathbf{x}_0), f(\mathbf{x}_0 + \mathbf{h})\big\rangle + \big\langle g(\mathbf{x}_0), f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\big\rangle.
\end{aligned}$$

Define $\Lambda_{\mathbf{x}_0} : \mathbb{R}^n \to \mathbb{R}$ by

$$\Lambda_{\mathbf{x}_0}(\cdot) = \langle g'(\mathbf{x}_0)(\cdot), f(\mathbf{x}_0)\rangle + \langle g(\mathbf{x}_0), f'(\mathbf{x}_0)(\cdot)\rangle$$

which is linear by the (bi-) linearity of the real inner product. Hence

$$\begin{aligned}
&\frac{|h(\mathbf{x}_0+\mathbf{h})-h(\mathbf{x}_0)-\Lambda_{\mathbf{x}_0}\mathbf{h}|}{\|\mathbf{h}\|} \\
&\quad= \frac{\left|\left\langle g(\mathbf{x}_0+\mathbf{h})-g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})\right\rangle + \left\langle g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)\right\rangle - \Lambda_{\mathbf{x}_0}\mathbf{h}\right|}{\|\mathbf{h}\|} \\
&\quad\leq \frac{\left|\left\langle g(\mathbf{x}_0+\mathbf{h})-g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})\right\rangle - \left\langle g'(\mathbf{x}_0)\mathbf{h}, f(\mathbf{x}_0)\right\rangle\right|}{\|\mathbf{h}\|} \\
&\qquad\qquad + \frac{\left|\left\langle g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)\right\rangle - \left\langle g(\mathbf{x}_0), f'(\mathbf{x}_0)\mathbf{h}\right\rangle\right|}{\|\mathbf{h}\|}
\end{aligned}$$

where the last inequality was an application of the triangle inequality. Our job now will be to show each term in the final sum converges to 0 as $\mathbf{h}\to 0_n$. The second term will be easiest to deal with. Indeed, by the (bi-)linearity of the inner product and the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\frac{\left|\left\langle g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)\right\rangle - \left\langle g(\mathbf{x}_0), f'(\mathbf{x}_0)\mathbf{h}\right\rangle\right|}{\|\mathbf{h}\|} = \frac{\left|\left\langle g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-f'(\mathbf{x}_0)\mathbf{h}\right\rangle\right|}{\|\mathbf{h}\|} \\
&\leq \frac{\|g(\mathbf{x}_0)\|\,\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-f'(\mathbf{x}_0)\mathbf{h}\|}{\|\mathbf{h}\|} = \|g(\mathbf{x}_0)\|\left(\frac{\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-f'(\mathbf{x}_0)\mathbf{h}\|}{\|\mathbf{h}\|}\right).
\end{aligned}$$

By the assumption that f is differentiable at $\mathbf{x}_0$, the term in parentheses approaches 0 as $\mathbf{h}\to 0_n$. It remains to show that

$$\frac{\left|\left\langle g(\mathbf{x}_0+\mathbf{h})-g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})\right\rangle - \left\langle g'(\mathbf{x}_0)\mathbf{h}, f(\mathbf{x}_0)\right\rangle\right|}{\|\mathbf{h}\|}\to 0$$

as $\mathbf{h}\to 0_n$. To ease notation, let's define $\Delta_{\mathbf{h}}g(\mathbf{x}_0)=g(\mathbf{x}_0+\mathbf{h})-g(\mathbf{x}_0)$ and similarly for $\Delta_{\mathbf{h}}f(\mathbf{x}_0)=f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)$. By once again adding 0,

$$\begin{aligned}
&\left\langle \Delta_{\mathbf{h}}g(\mathbf{x}_0), f(\mathbf{x}_0+\mathbf{h})\right\rangle - \left\langle g'(\mathbf{x}_0)\mathbf{h}, f(\mathbf{x}_0)\right\rangle \\
&\qquad= \left\langle \Delta_{\mathbf{h}}g(\mathbf{x}_0)-g'(\mathbf{x}_0)\mathbf{h}, f(\mathbf{x}_0+\mathbf{h})\right\rangle + \left\langle g'(\mathbf{x}_0)\mathbf{h}, \Delta_{\mathbf{h}}f(\mathbf{x}_0)\right\rangle
\end{aligned}$$

An application of the Cauchy-Schwarz inequality on the first term of this sum gives

$$\frac{\left|\left\langle \Delta_{\mathbf{h}}g(\mathbf{x}_0)-g'(\mathbf{x}_0)\mathbf{h}, f(\mathbf{x}_0+\mathbf{h})\right\rangle\right|}{\|\mathbf{h}\|}\leq \|f(\mathbf{x}_0+\mathbf{h})\|\left(\frac{\|\Delta_{\mathbf{h}}g(\mathbf{x}_0)-g'(\mathbf{x}_0)\|}{\|k\|}\right)$$

The right side of this inequality approaches 0 as $\mathbf{h} \to 0_n$ by the assumption of the differentiability of g. Another application of the Cauchy-Schwarz inequality yields

$$\frac{\left|\left\langle g'(\mathbf{x}_0)\mathbf{h}, \Delta_{\mathbf{h}} f(\mathbf{x}_0)\right\rangle\right|}{\|\mathbf{h}\|} \leq \frac{\|\Delta_{\mathbf{h}} f(\mathbf{x}_0)\| \, \|g'(\mathbf{x}_0)\mathbf{h}\|}{\|\mathbf{h}\|} \leq \left(\frac{\|\Delta_{\mathbf{h}} f(\mathbf{x}_0)\| \, \|g'(\mathbf{x}_0)\|_{op} \, \|\mathbf{h}\|}{\|\mathbf{h}\|}\right)$$
$$= \|\Delta_{\mathbf{h}} f(\mathbf{x}_0)\| \, \|g'(\mathbf{x}_0)\|_{op}.$$

Since $g'(\mathbf{x}_0)$ is a linear operator defined between finite dimensional vector spaces, the operator norm $\|g'(\mathbf{x}_0)\|_{op}$ is finite. Moreover, since f is continuous at $\mathbf{x}_0$ (since it is differentiable there), $\Delta_{\mathbf{h}} f(\mathbf{x}_0) \to 0_m$ as $\mathbf{h} \to 0_n$, which further implies that $\|\Delta_{\mathbf{h}} f(\mathbf{x}_0)\| \to \|0_m\| = 0$ as $\mathbf{h} \to 0_n$ since the norm is continuous. Putting these pieces together, we finally conclude that

$$\frac{\left|\left\langle g(\mathbf{x}_0 + \mathbf{h}) - g(\mathbf{x}_0), f(\mathbf{x}_0 + \mathbf{h})\right\rangle - \left\langle g'(\mathbf{x}_0)\mathbf{h}, f(\mathbf{x}_0)\right\rangle\right|}{\|\mathbf{h}\|} \to 0$$

as $\mathbf{h} \to 0_n$, which concludes the proof.

(3) Let us start by making the following observation,

$$\begin{aligned}
&g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\big(f'(\mathbf{x}_0)\mathbf{h}\big) \\
&= \Big[g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\big(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\big)\Big] \\
&\qquad + \Big[g'(f(\mathbf{x}_0))\big(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\big) - g'(f(\mathbf{x}_0))\big(f'(\mathbf{x}_0)\mathbf{h}\big)\Big]
\end{aligned}$$

Therefore, by applying the norm and using the triangle inequality

$$\begin{aligned}
&\big\|g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\big(f'(\mathbf{x}_0)\mathbf{h}\big)\big\|_m \\
&\leq \big\|g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\big(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\big)\big\|_m \\
&\qquad + \big\|g'(f(\mathbf{x}_0))\big(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\big) - g'(f(\mathbf{x}_0))\big(f'(\mathbf{x}_0)\mathbf{h}\big)\big\|_m \\
&= \big\|g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\big(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\big)\big\|_m \\
&\qquad + \big\|g'(f(\mathbf{x}_0))\big(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)\mathbf{h}\big)\big\|_m
\end{aligned}$$

and the fact that $g'(f(\mathbf{x}_0))$ is a bounded linear operator,

$$\begin{aligned}
&\leq \big\|g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\big(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\big)\big\|_m \\
&\qquad + \big\|g'(f(\mathbf{x}_0))\big\|_{op} \big\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)\mathbf{h}\big\|_m
\end{aligned}$$

Let $\mathbf{k} = f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)$; note that by the continuity of f at $\mathbf{x}_0$, $\mathbf{k} \to 0_m$ as $\mathbf{h} \to 0_n$. Rewriting the last inequality with this notation and dividing by $\|\mathbf{h}\|_n$,

$$\begin{aligned}
&\frac{\big\|g(f(\mathbf{x}_0+\mathbf{h})) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\big(f'(\mathbf{x}_0)\mathbf{h}\big)\big\|_m}{\|\mathbf{h}\|_n} \\
&\quad\le \frac{\big\|g(f(\mathbf{x}_0)+\mathbf{k}) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\mathbf{k}\big\|_m}{\|\mathbf{h}\|_n} \\
&\qquad\qquad + \big\|g'(f(\mathbf{x}_0))\big\|_{op} \frac{\big\|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)\mathbf{h}\big\|_m}{\|\mathbf{h}\|_n}
\end{aligned}$$

By the assumption that f is differentiable at $\mathbf{x}_0$, the second term on the right hand side converges to 0 as $\mathbf{h} \to 0_n$. So it remains to show that

$$\frac{\big\|g(f(\mathbf{x}_0)+\mathbf{k}) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\mathbf{k}\big\|_m}{\|\mathbf{h}\|_n} \to 0 \quad \text{as } \mathbf{h} \to 0_n.$$

If $\mathbf{k} = 0_m$, then there is nothing to show. If not, we have

$$\begin{aligned}
&\frac{\big\|g(f(\mathbf{x}_0)+\mathbf{k}) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\mathbf{k}\big\|_m}{\|\mathbf{h}\|_n} \\
&\quad= \left(\frac{\big\|g(f(\mathbf{x}_0)+\mathbf{k}) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\mathbf{k}\big\|_m}{\|\mathbf{k}\|_m}\right) \frac{\|\mathbf{k}\|_m}{\|\mathbf{h}\|_n}.
\end{aligned} \tag{2.2}$$

Using the triangle inequality

$$\frac{\|\mathbf{k}\|_m}{\|\mathbf{h}\|_n} = \frac{\|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0)\|_m}{\|\mathbf{h}\|_n} \le \frac{\|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)\mathbf{h}\|_m}{\|\mathbf{h}\|_n} + \frac{\|f'(\mathbf{x}_0)\mathbf{h}\|_m}{\|\mathbf{h}\|_n}$$

and that $f'(\mathbf{x}_0)$ is a bounded linear operator

$$\begin{aligned}
&\le \frac{\|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)\mathbf{h}\|_m}{\|\mathbf{h}\|_n} + \frac{\|f'(\mathbf{x}_0)\|_{op}\,\|\mathbf{h}\|_n}{\|\mathbf{h}\|_n} \\
&= \frac{\|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)\mathbf{h}\|_m}{\|\mathbf{h}\|_n} + \|f'(\mathbf{x}_0)\|_{op}
\end{aligned}$$

What is important to note is that in the last equality, the first term converges to 0 as $\mathbf{h} \to 0_n$, and the second remains bounded; particularly, the expression is bounded as $\mathbf{h}$ approaches 0_n. This shows that for $\mathbf{h}$ sufficiently close to 0_n, there is some bound $C > 0$ such that

$$\frac{\|\mathbf{k}\|_m}{\|\mathbf{h}\|_n} \le C.$$

Since $\mathbf{h} \to 0_n \implies \mathbf{k} \to 0_m$ and g is assumed to be differentiable at $f(\mathbf{x}_0)$, from (2.2)

$$\lim_{\mathbf{h}\to 0_n} \frac{\left\| g(f(\mathbf{x}_0)+\mathbf{k}) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\mathbf{k} \right\|_m}{\|\mathbf{h}\|_n}$$
$$\leq C \lim_{\mathbf{k}\to 0_m} \left(\frac{\left\| g(f(\mathbf{x}_0)+\mathbf{k}) - g(f(\mathbf{x}_0)) - g'(f(\mathbf{x}_0))\mathbf{k} \right\|_m}{\|\mathbf{k}\|_m} \right) = 0$$

which finally finishes the proof. ■

PROPOSITION 8.2.5. *The following are common examples of differentiable functions.*

(1) *If $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function, then A is differentiable at every $\mathbf{x} \in \mathbb{R}^n$ and $A'(\mathbf{x}) = A$.*

(2) *Let U be an open subset of $\mathbb{R}^n$. If $f : U \to \mathbb{R}^m$ is a constant function (i.e. there is some $\mathbf{a} \in \mathbb{R}^m$ such that $f(\mathbf{x}) = \mathbf{a}$ for every $\mathbf{x} \in U$), then for every $\mathbf{x} \in U$, $f'(\mathbf{x}) = \mathbf{0}$. Here $\mathbf{0} : \mathbb{R}^n \to \mathbb{R}^m$ is the zero map $\mathbf{0}(\mathbf{x}) = 0_m$.*

(3) *Suppose that U is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$. Let $\mathbf{v} \in \mathbb{R}^m$. Then the function $g : U \to \mathbb{R}^m$ defined by $g(\mathbf{x}) = f(\mathbf{x})\,\mathbf{v}$ is differentiable at $\mathbf{x}_0$ and $g'(\mathbf{x}_0)(\cdot) = f'(\mathbf{x}_0)(\cdot)\,\mathbf{v}$.*

(4) *Suppose that U is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^m$ be differentiable at $\mathbf{x}_0$. Let $\mathbf{v} \in \mathbb{R}^m$. Then the function $g : U \to \mathbb{R}$ defined by $g(\mathbf{x}) = \langle f(\mathbf{x}), \mathbf{v}\rangle_m$ is differentiable at $\mathbf{x}_0$ and $g'(\mathbf{x}_0)(\cdot) = \langle f'(\mathbf{x}_0)(\cdot), \mathbf{v}\rangle_m$.*

(5) *The function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\mathbf{x}) = \|\mathbf{x}\|^2$ is differentiable and $f'(\mathbf{x})(\cdot) = 2\langle \mathbf{x}, \cdot\rangle$. This means that for any $\mathbf{y} \in \mathbb{R}^n$, $f'(\mathbf{x})\mathbf{y} = 2\langle \mathbf{x}, \mathbf{y}\rangle$.*

Proof. Exercise 5 (7). ■

PROPOSITION 8.2.6. *Let U be an open subset of $\mathbb{R}^n$ and suppose $f : U \to \mathbb{R}^m$. For each $k = 1, 2, ..., m$, let $f_k : U \to \mathbb{R}$ be the kth component of f, then $f = (f_1, f_2, ..., f_m)$. Then f is differentiable at $\mathbf{x}_0 \in U$ if and only if each component f_k is differentiable at $\mathbf{x}_0$. In that case, $f'(\mathbf{x}_0) = (f_1'(\mathbf{x}_0), f_2'(\mathbf{x}_0), ..., f_m'(\mathbf{x}_0))$. This notation means that for any $\mathbf{y} \in \mathbb{R}^n$,*

$$f'(\mathbf{x}_0)\mathbf{y} = (f_1'(\mathbf{x}_0)\mathbf{y}, f_2'(\mathbf{x}_0)\mathbf{y}, ..., f_m'(\mathbf{x}_0)\mathbf{y})$$

Proof. Let $\{e_k\}_{k=1}^m$ be the standard (orthonormal) basis of $\mathbb{R}^m$. Then by Proposition 8.1.5,

$$f(\mathbf{x}) = \big(f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_m(\mathbf{x})\big) = \sum_{k=1}^m f_k(\mathbf{x})\,\mathbf{e}_k = \sum_{k=1}^m \langle f(\mathbf{x}), \mathbf{e}_k\rangle_m\,\mathbf{e}_k.$$

If each component f_k is differentiable at $\mathbf{x}_0$, then using Theorem 8.2.4 and Proposition 8.2.5, we have that f is differentiable at $\mathbf{x}_0$ and $f'(\mathbf{x}_0)(\cdot) = \sum_{k=1}^{m} f_k'(\mathbf{x}_0)(\cdot)\,\mathbf{e}_k$. Rewriting this in different terms, this shows that $f'(\mathbf{x}_0) = (f_1'(\mathbf{x}_0), f_2'(\mathbf{x}_0), ..., f_m'(\mathbf{x}_0))$.

If we assume that f is differentiable at $\mathbf{x}_0$, then for any $k = 1, 2, ..., m$, $f_k(\mathbf{x}) = \langle f(\mathbf{x}), \mathbf{e}_k \rangle_m$. By Proposition 8.2.5, we have that f_k is differentiable at $\mathbf{x}_0$ and

$$f_k'(\mathbf{x}_0)(\cdot) = \langle f'(\mathbf{x}_0)(\cdot), \mathbf{e}_k \rangle_m.$$

Moreover, by another application of Proposition 8.2.5, since $f(\mathbf{x}) = \sum_{k=1}^{m} \langle f(\mathbf{x}), \mathbf{e}_k \rangle\, \mathbf{e}_k$,

$$f'(\mathbf{x}_0)(\cdot) = \sum_{k=1}^{m} \langle f'(\mathbf{x}_0)(\cdot), \mathbf{e}_k \rangle\, \mathbf{e}_k.$$

Therefore, comparing terms, we have $f'(\mathbf{x}_0)(\cdot) = \sum_{k=1}^{m} f_k'(\mathbf{x}_0)(\cdot)\,\mathbf{e}_k$. Rewriting in different notation, this shows that $f'(\mathbf{x}_0) = (f_1'(\mathbf{x}_0), f_2'(\mathbf{x}_0), ..., f_m'(\mathbf{x}_0))$. We have thus shown that f is differentiable at $\mathbf{x}_0$ if and only if each component is differentiable, and that $f'(\mathbf{x}_0) = (f_1'(\mathbf{x}_0), f_2'(\mathbf{x}_0), ..., f_m'(\mathbf{x}_0))$. ■

3. Directional and Partial Derivatives

When first exposed to vectors, students often learn to interpret them as an object with two quantities: direction and magnitude. While this concept may be hard to picture in more general vector spaces than $\mathbb{R}^n$, it is an apt description for the points (vectors) $\mathbf{v} \in \mathbb{R}^n$ for many practical purposes; to name just a few: describing velocities, accelerations, forces, and momenta. What we will be interested in for this section is how to find the "rate of change of a function in the direction of a given vector." Particularly, picture $f : \mathbb{R}^2 \to \mathbb{R}$ as an "elevation map" of some terrain; here $\mathbb{R}^2$ can be understood as the latitude and longitude coordinates, and for a particular $(x, y) \in \mathbb{R}^2$, $f(x, y)$ gives the elevation at that point. Given a point (vector) $\mathbf{v} \in \mathbb{R}^2$, imagine at the location (x, y) you are moving over the terrain with velocity $\mathbf{v}$ (this velocity is with respect to the latitude-longitude plane $\mathbb{R}^2$). If you ask the reasonable question, "what is the rate of change in my elevation as I pass through the point (x, y) at velocity $\mathbf{v}$?" then the answer is given by the *directional derivative* of f at (x, y) with respect to $\mathbf{v}$. We now work to make this a rigorous mathematical theory.

DEFINITION 8.3.1. Let U be an open subset of $\mathbb{R}^n$, $\mathbf{x}_0 \in U$, $\mathbf{v} \in \mathbb{R}^n$, and $f : U \to \mathbb{R}^m$. The *directional derivative* of f at $\mathbf{x}_0$ with respect to $\mathbf{v}$ is defined a

$$\lim_{t\to 0} \frac{f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0)}{t}$$

whenever this limit exists. Note that t is running through values in $\mathbb{R}$.

NOTATION 8.3.2. Because of the similarity to a real derivative, we often shorthand the limit as

$$\frac{d}{dt}\Big|_0 f(\mathbf{x}_0 + t \cdot \mathbf{v}) = \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0)}{t}$$

If the directional derivative does exist, we will denote it by $\dfrac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$.

For the remainder of this section, U will be an open subset of $\mathbb{R}^n$, $\mathbf{x}_0 \in U$, and $f : U \to \mathbb{R}^m$.

REMARK 8.3.3. Note well! If the directional derivative $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$ exists, then it is in the codomain of f, i.e. $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0) \in \mathbb{R}^m$. Indeed, another way to write the limit which might make this more apparent is

$$\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0) = \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0)}{t} = \lim_{t\to 0} \frac{1}{t} \cdot \big(f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0)\big).$$

The fact that we are using scalar multiplication here means that we can get away without using norms as we were forced to do in the case of a general derivative. Moreover, what this means is that if $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x})$ exists for every $\mathbf{x} \in U$, then we consider $\frac{\partial f}{\partial \mathbf{v}}$ a function $\frac{\partial f}{\partial \mathbf{v}} : U \to \mathbb{R}^m$ in the familiar way. □

PROPOSITION 8.3.4. *Suppose that U is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^m$ is has components f_i for $i = 1, 2, ..., m$. Then $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$ exists if and only if $\frac{\partial f_i}{\partial \mathbf{v}}(\mathbf{x}_0)$ exists for each i. If this is the case, then jth component of $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$ is $\frac{\partial f_j}{\partial \mathbf{v}}(\mathbf{x}_0)$; that is,*

$$\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0) = \Big(\frac{\partial f_1}{\partial \mathbf{v}}(\mathbf{x}_0), ..., \frac{\partial f_m}{\partial \mathbf{v}}(\mathbf{x}_0)\Big). \tag{3.1}$$

Proof. This is immediate once one realizes that

$$\frac{f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0)}{t} = \Big(\frac{f_1(\mathbf{x}_0 + t \cdot \mathbf{v}) - f_1(\mathbf{x}_0)}{t}, ..., \frac{f_m(\mathbf{x}_0 + t \cdot \mathbf{v}) - f_m(\mathbf{x}_0)}{t}\Big).$$

Therefore the limit as $t \to 0$ on the lefthand side exists if and only if it does for each component on the righthand side, and when the limit does exist, it results in (3.1). ■

THEOREM 8.3.5. *If f is differentiable at $\mathbf{x}_0$, then given any $\mathbf{v} \in \mathbb{R}^n$, the directional derivative $\frac{\partial f}{\partial \mathbf{v}}(x_0)$ exists. Moreover,*

$$\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0) = f'(\mathbf{x}_0)\mathbf{v}.$$

Proof. By the assumption that f is differentiable at $\mathbf{x}_0$, we have that $f'(\mathbf{x}_0) \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$ exists and

$$\lim_{t\to 0} \frac{\|f(\mathbf{x}_0 + t\cdot\mathbf{v}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)(t\cdot\mathbf{v})\|_m}{\|t\cdot\mathbf{v}\|_n} = 0.$$

Using the properties of norms and the linearity of $f'(\mathbf{x}_0)$, for any $t \neq 0$,

$$\begin{aligned}
&\frac{\|f(\mathbf{x}_0 + t\cdot\mathbf{v}) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)(t\cdot\mathbf{v})\|_m}{\|t\cdot\mathbf{v}\|_n} \\
&\quad = \frac{1}{\|\mathbf{v}\|_n}\left\|\frac{f(\mathbf{x}_0 + t\cdot\mathbf{v}) - f(\mathbf{x}_0) - t\cdot f'(\mathbf{x}_0)\mathbf{v}}{t}\right\|_m \\
&\quad = \frac{1}{\|\mathbf{v}\|_n}\left\|\frac{f(\mathbf{x}_0 + t\cdot\mathbf{v}) - f(\mathbf{x}_0)}{t} - f'(\mathbf{x}_0)\mathbf{v}\right\|_m
\end{aligned}$$

Since as $t \to 0$ the first term goes to 0, this forces all terms in the equality to converge to 0, which in particular means

$$\lim_{t\to 0}\left\|\frac{f(\mathbf{x}_0 + t\cdot\mathbf{v}) - f(\mathbf{x}_0)}{t} - f'(\mathbf{x}_0)\mathbf{v}\right\|_m = 0$$

which is equivalent to

$$\lim_{t\to 0}\frac{f(\mathbf{x}_0 + t\cdot\mathbf{v}) - f(\mathbf{x}_0)}{t} = f'(\mathbf{x}_0)\mathbf{v}.$$

This proves that $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$ exists and equals $f'(\mathbf{x}_0)\mathbf{v}$. ■

REMARK 8.3.6. A calculational observation that is often utilized is the following: Suppose that U is an open subset of $\mathbb{R}^n$, $f : U \to \mathbb{R}$ and $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$ exists. Let $J \subset \mathbb{R}$ be some interval containing 0 which is sufficiently small that for every $t \in J$, $\mathbf{x}_0 + t\cdot\mathbf{v} \in U$. We then define the auxilery function $\varphi : J \to \mathbb{R}$ by $\varphi(t) = f(\mathbf{x}_0 + t\cdot\mathbf{v})$. What's to notice now is that the existence of $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$ is equivalent to the existence of $\frac{d\varphi}{dt}(0)$ and

$$\frac{d\varphi}{dt}(0) = \lim_{t\to 0}\frac{\varphi(t) - \varphi(0)}{t} = \lim_{t\to 0}\frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0).$$

(Remember here that we use $\frac{d\varphi}{dt}$ to give the scalar-valued derivative of φ). □

THEOREM 8.3.7 (Mean Value Theorem). *Let U be an open subset of $\mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{x} \in U$, and $f : U \to \mathbb{R}^m$. Suppose that $\frac{\partial f}{\partial \mathbf{v}}$ exists at all points in some open neighborhood V of*

U. If $t > 0$ is small enough that $\mathbf{x} + s \cdot \mathbf{v} \in V$ for all $s \in [0, t]$, then there exists some $\theta = (\theta_1, ..., \theta_m) \in \mathbb{R}^m$ with $\theta_i \in [0, t]$ for each i and

$$f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x}) = t \cdot \Big(\frac{\partial f_1}{\partial \mathbf{v}}(\mathbf{x} + \theta_1 \mathbf{v}), ..., \frac{\partial f_m}{\partial \mathbf{v}}(\mathbf{x} + \theta_m \mathbf{v}) \Big).$$

In particular, if $m = 1$ (i.e. $f : U \to \mathbb{R}$), then there is some $\theta \in [0, t]$ such that

$$f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x}) = t \cdot \frac{\partial f}{\partial \mathbf{v}}(\mathbf{x} + \theta \mathbf{v}).$$

Proof. Let $k \in \{1, 2, ..., m\}$ and let $J = [0, t] \subset \mathbb{R}$. Define $\varphi : J \to \mathbb{R}$ as $\varphi_k(s) = f_k(\mathbf{x}_0 + s \cdot \mathbf{v})$. An assumption of the problem is that $\frac{\partial f}{\partial \mathbf{v}}$ exists at all points in V which implies that $\frac{d\varphi}{dt}$ exists for all points $s \in J$ (since $\mathbf{x}_0 + s \cdot \mathbf{v} \in V$ for every $s \in J$). Moreover,

$$\frac{d\varphi}{dt}(s) = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0 + s \cdot \mathbf{v}).$$

Using the standard mean value theorem, we know that there exists some $\theta_k \in [0, t]$ such that

$$\varphi(t) - \varphi(0) = t \cdot \frac{d\varphi}{dt}(\theta_k)$$

which is equivalent to

$$f_k(\mathbf{x}_0 + t \cdot \mathbf{v}) - f_k(\mathbf{x}_0) = t \cdot \frac{\partial f_k}{\partial \mathbf{v}}(\mathbf{x}_0 + \theta_k \cdot \mathbf{v}).$$

Since k was arbitrary, we have shown that there exist $\theta_1, \theta_2, ..., \theta_m \in [0, t]$ such that

$$\begin{aligned} &f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0) \\ &\quad = \big(f_1(\mathbf{x}_0 + t \cdot \mathbf{v}) - f_1(\mathbf{x}_0), f_2(\mathbf{x}_0 + t \cdot \mathbf{v}) - f_2(\mathbf{x}_0), ..., f_n(\mathbf{x}_0 + t \cdot \mathbf{v}) - f_n(\mathbf{x}_0) \big) \\ &\quad = t \cdot \Big(\frac{\partial f_1}{\partial \mathbf{v}}(\mathbf{x} + \theta_1 \mathbf{v}), ..., \frac{\partial f_m}{\partial \mathbf{v}}(\mathbf{x} + \theta_m \mathbf{v}) \Big). \end{aligned}$$

■

NOTATION 8.3.8. Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. It is perfectly possible that the directional derivative $\frac{\partial f}{\partial \mathbf{w}}$ might be differentiable with respect to $\mathbf{v}$ at some point $\mathbf{x}_0$ in the domain. In this case we typically write

$$\frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0)$$

to mean

$$\frac{\partial}{\partial \mathbf{v}} \left(\frac{\partial f}{\partial \mathbf{w}} \right) (\mathbf{x}_0).$$

PROPOSITION 8.3.9. *Let $U \subset \mathbb{R}^n$, $\mathbf{x}_0 \in U$, $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, and $f : U \to \mathbb{R}^m$. Suppose that $\frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}$ exists in some neighborhood of $\mathbf{x}_0$ and is continuous at $\mathbf{x}_0$. Then*

$$\frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0) = \lim_{t \to 0} \frac{f(\mathbf{x}_0 + t \cdot (\mathbf{v} + \mathbf{w})) - f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0 + t \cdot \mathbf{w}) + f(\mathbf{x}_0)}{t^2}. \tag{3.2}$$

Proof. As in the proof of the mean value theorem, it suffices to show that (3.2) holds for an arbitrary component f_k of f, so it suffices to just assume that $f : U \to \mathbb{R}$; we assume this for the remainder of the proof. Let's also start with the assumption that $U = \mathbb{R}^n$ solely for scaling reasons. For a fixed $t \in \mathbb{R}$, we also define $\psi : \mathbb{R} \to \mathbb{R}$ by

$$\psi(s) = f(\mathbf{x}_0 + t \cdot \mathbf{v} + s \cdot \mathbf{w}) - f(\mathbf{x}_0 + s \cdot \mathbf{w}).$$

Then

$$\psi(t) - \psi(0) = f(\mathbf{x}_0 + t \cdot (\mathbf{v} + \mathbf{w})) - f(\mathbf{x}_0 + t \cdot \mathbf{v}) - f(\mathbf{x}_0 + t \cdot \mathbf{w}) + f(\mathbf{x}_0),$$

which we recognize is the numerator of the righthand side of (3.2). Using the mean value theorem in $\mathbb{R}$, there exists $s_t \in [0, t]$ such that $\psi(t) - \psi(0) = t \cdot \frac{d\psi}{dt}(s_t)$. Moreover,

$$\frac{d\psi}{dt}(s_t) = \frac{\partial f}{\partial \mathbf{w}}(\mathbf{x}_0 + t \cdot \mathbf{v} + s_t \cdot \mathbf{w}) - \frac{\partial f}{\partial \mathbf{w}}(\mathbf{x}_0 + s_t \cdot \mathbf{w}).$$

By another application of the mean value theorem applied to the righthand side of the last equality viewed as a function of in the variable t with s_t held fixed, there is some $r_t \in [0, t]$ such that

$$\frac{d\psi}{dt}(s_t) = t \cdot \frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0 + r_t \cdot \mathbf{v} + s_t \cdot \mathbf{w}).$$

Putting these pieces together, we have

$$\psi(t) - \psi(0) = t \cdot \frac{d\psi}{dt}(s) = t^2 \cdot \frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0 + r_t \cdot \mathbf{v} + s_t \cdot \mathbf{w}).$$

Therefore

$$\frac{\psi(t) - \psi(0)}{t^2} = \frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0 + r_t \cdot \mathbf{v} + s_t \cdot \mathbf{w}) \implies \lim_{t \to 0} \frac{\psi(t) - \psi(0)}{t^2} = \frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0).$$

The last equality used the fact that both $r_t \to 0$ and $s_t \to 0$ as $t \to 0$, and the continuity assumption of the second partial derivative implies that since $(\mathbf{x}_0 + r_t \cdot \mathbf{v} + s_t \cdot \mathbf{w}) \to \mathbf{x}_0$, then $\frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0 + r_t \cdot \mathbf{v} + s_t \cdot \mathbf{w}) \to \frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0)$. ∎

THEOREM 8.3.10 (Mixed Partials Commute). *Let $U \subset \mathbb{R}^n$, $\mathbf{x}_0 \in U$, $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, and $f : U \to \mathbb{R}^m$. Suppose that both $\frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}$ and $\frac{\partial^2 f}{\partial \mathbf{w} \partial \mathbf{v}}$ exist in some neighborhood of $\mathbf{x}_0$ are are continuous at $\mathbf{x}_0$. Then*

$$\frac{\partial^2 f}{\partial \mathbf{v} \partial \mathbf{w}}(\mathbf{x}_0) = \frac{\partial^2 f}{\partial \mathbf{w} \partial \mathbf{v}}(\mathbf{x}_0).$$

Proof. Equation (3.2) is symmetric in $\mathbf{v}$ and $\mathbf{w}$, which implies the claim. Details are left as an exercise (see Exercise 5 (9)). ■

3.1. Gradient and Jacobian Matrix. For this section, U will be an open subset of $\mathbb{R}^n$, $\mathbf{x}_0 \in U$.

DEFINITION 8.3.11. Let $\{\mathbf{e}_k\}_{k=1}^n$ be the standard basis in $\mathbb{R}^n$. Suppose $f : U \to \mathbb{R}$. For each $k = 1, 2, ..., n$, The directional derivative $\frac{\partial f}{\partial \mathbf{e}_k}(\mathbf{x}_0)$ is called a *partial derivative* of f at $\mathbf{x}_0$ when it exists. Often, to be more explicit, we say that $\frac{\partial f}{\partial \mathbf{e}_k}$ is the partial derivative of f in the $\mathbf{e}_k$ direction.

NOTATION 8.3.12. It is much more common to use a slightly altered notation when dealing with partial derivatives, here is a list of some of the most common for $\frac{\partial f}{\partial \mathbf{e}_k}$:

$$\frac{\partial f}{\partial x_k}, \ f_{x_k}, \ f_{,k}$$

(the comma before the k in the last expression is intentional!) If $n = 2$ or 3 the dominant notation is

$$\begin{aligned} \frac{\partial f}{\partial x} \text{ or } f_x \text{ for } \frac{\partial f}{\partial \mathbf{e}_1} \\ \frac{\partial f}{\partial y} \text{ or } f_y \text{ for } \frac{\partial f}{\partial \mathbf{e}_2} \\ \frac{\partial f}{\partial z} \text{ or } f_z \text{ for } \frac{\partial f}{\partial \mathbf{e}_3} \end{aligned}$$

It is also extremely common to let the notation be governed by the choice of symbol for independent variables. For example, if you consider f to be a function of the variables (r, θ), i.e. you write $f(r, \theta)$, then it is common to use $\frac{\partial f}{\partial r}$ or f_r for $\frac{\partial f}{\partial \mathbf{e}_1}$ and $\frac{\partial f}{\partial \theta}$ or f_θ for $\frac{\partial f}{\partial \mathbf{e}_2}$.

DEFINITION 8.3.13. Suppose that $f : U \to \mathbb{R}$ is differentiable at the point $\mathbf{x}_0$. Then the *gradient* of f at $\mathbf{x}_0$, denoted $\nabla f(\mathbf{x}_0)$, is the unique vector in $\mathbb{R}^n$ such that

$$f'(\mathbf{x}_0)\mathbf{v} = \langle \nabla f(\mathbf{x}_0), \mathbf{v} \rangle$$

for every $\mathbf{v} \in \mathbb{R}^n$. As usual, if f is differentiable within some subset $V \subset U$, we consider ∇f to be the function $\nabla f : V \to \mathbb{R}^n$ given by $\mathbf{x} \mapsto \nabla f(\mathbf{x})$.

LEMMA 8.3.14. *The gradient is well-defined.*

Proof. If $f : U \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$, then $f'(\mathbf{x}_0)$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}$. In other words, $f'(\mathbf{x}_0)$ is a bounded linear functional on $\mathbb{R}^n$ (i.e. $f'(\mathbf{x}_0) \in \left(\mathbb{R}^n\right)^*$). Since $\mathbb{R}^n$ is a Hilbert space, the Riesz representation theorem (see Theorem 7.7.2) implies that there exists a unique vector in $\nabla f(\mathbf{x}_0) \in \mathbb{R}^n$ such that

$$f'(\mathbf{x}_0)\mathbf{v} = \langle \nabla f(\mathbf{x}_0), \mathbf{v} \rangle$$

for every $\mathbf{v} \in \mathbb{R}^n$. ■

THEOREM 8.3.15. *Suppose that $f : U \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$. Then*

$$\nabla f(\mathbf{x}_0) = \Big(\frac{\partial f}{\partial x_1}(\mathbf{x}_0), ..., \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \Big) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{x}_0)\, \mathbf{e}_k.$$

where $\{\mathbf{e}_k\}_{k=1}^n$ is the standard basis in $\mathbb{R}^n$. In particular, for any $\mathbf{v} = (v_1, ..., v_n) \in \mathbb{R}^n$,

$$f'(\mathbf{x}_0)\mathbf{v} = \langle \nabla f(\mathbf{x}_0), \mathbf{v} \rangle = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{x}_0)\, v_k.$$

Proof. Let $\{\mathbf{e}_k\}_{k=1}^n$ be the standard basis in $\mathbb{R}^n$. Let $\mathbf{v} \in \mathbb{R}^n$ be arbitrary; we can then write $\mathbf{v} = \sum\limits_{k=1}^{n} v_k \cdot \mathbf{e}_k$, where for each k, v_k is the kth component of $\mathbf{v}$. Using the linearity of $f'(\mathbf{x}_0)$, we have $f'(\mathbf{x}_0)\mathbf{v} = \sum\limits_{k=1}^{m} v_k \cdot f'(\mathbf{x}_0)\mathbf{e}_k$. Using Theorem 8.3.5,

$$f'(\mathbf{x}_0)\mathbf{e}_k = \frac{\partial f}{\partial \mathbf{e}_k}(\mathbf{x}_0)$$

and the right hand side is just different notation for $f_{x_k}(\mathbf{x}_0)$. Therefore,

$$f'(\mathbf{x}_0)\mathbf{v} = \sum_{k=1}^{n} f_{x_k}(\mathbf{x}_0) v_k = \sum_{k=1}^{n} f_{x_k}(\mathbf{x}_0) \langle \mathbf{e}_k, \mathbf{v} \rangle = \Big\langle \sum_{k=1}^{n} f_{x_k}(\mathbf{x}_0)\mathbf{e}_k, \mathbf{v} \Big\rangle$$

By the definition of the gradient, we conclude that $\nabla f(\mathbf{x}_0) = \sum\limits_{k=1}^{n} f_{x_k}(\mathbf{x}_0) \cdot \mathbf{e}_k$ ■

COROLLARY 8.3.16. *Suppose that $g : U \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0$ and for each $k = 1, 2, ..., m$, g_k is the kth component of g. Define the matrix $Dg(\mathbf{x}_0)$ such that for each $k =$*

$1, 2, ..., m$, *the kth row of* $Dg(\mathbf{x}_0)$ *is given by* $\nabla g_k(\mathbf{x}_0)$. *Explicitly*

$$Dg(\mathbf{x}_0) = \begin{pmatrix} \nabla g_1(\mathbf{x}_0) \\ \nabla g_2(\mathbf{x}_0) \\ \vdots \\ \nabla g_n(\mathbf{x}_0) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}_0) & \frac{\partial g_1}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}_0) \\ \frac{\partial g_2}{\partial x_1}(\mathbf{x}_0) & \frac{\partial g_2}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial g_2}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\mathbf{x}_0) & \frac{\partial g_m}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial g_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}$$

For any $\mathbf{v} = (v_1, ..., v_n) \in \mathbb{R}^n$, *written as a column vector,*

$$g'(\mathbf{x}_0)\mathbf{v} = Dg(\mathbf{x}_0) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}_0) & \frac{\partial g_1}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}_0) \\ \frac{\partial g_2}{\partial x_1}(\mathbf{x}_0) & \frac{\partial g_2}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial g_2}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1}(\mathbf{x}_0) & \frac{\partial g_n}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial g_n}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

where the right hand side is to be interpreted using matrix multiplication. Put another way, using the standard bases in $\mathbb{R}^n$ *and* $\mathbb{R}^m$, *the matrix representation of the linear operator* $g'(\mathbf{x}_0)$ *is* $Dg(\mathbf{x}_0)$.

Proof. By Proposition 8.2.6, $g'(\mathbf{x}_0)\mathbf{v} = (g_1'(\mathbf{x}_0)\mathbf{v}, ..., g_m'(\mathbf{x}_0)\mathbf{v})$ for any $\mathbf{v} \in \mathbb{R}^n$. By Theorem 8.3.15, $g_k'(\mathbf{x}_0)\mathbf{v} = \langle \nabla g_k(\mathbf{x}_0), \mathbf{v} \rangle$ for each $k = 1, 2, ..., m$. The definition of matrix multiplication yields that the matrix $Dg(\mathbf{x}_0)$ multiplied on the right by $\mathbf{v}$ results in the vector $\big(\langle \nabla g_k(\mathbf{x}_0), \mathbf{v} \rangle, ..., \langle \nabla g_n(\mathbf{x}_0), \mathbf{v} \rangle\big)$. Putting these pieces together,

$$g'(\mathbf{x}_0)\mathbf{v} = \big(g_1'(\mathbf{x}_0)\mathbf{v}, ..., g_m'(\mathbf{x}_0)\mathbf{v}\big) = \big(\langle \nabla g_k(\mathbf{x}_0), \mathbf{v} \rangle, ..., \langle \nabla g_n(\mathbf{x}_0), \mathbf{v} \rangle\big) = Dg(\mathbf{x}_0)\mathbf{v}$$

which concludes the proof. ■

Definition 8.3.17. The matrix Dg is called the *Jacobian* (or Jacobian matrix) of g. Another common name is the *total derivative* matrix of g. When $Dg(\mathbf{x})$ exists for all $\mathbf{x} \in V \subset U$, then we consider Dg as a function with domain V. From this perspective, then components of the matrix Dg can also be viewed as functions from V to $\mathbb{R}$, which we will call the *component functions* of Dg.

Remark 8.3.18. Note that interpreting the gradient ∇f as a row vector, we have ∇f and Df are the same. A nice perspective to take away from this is the following. When

originally introduced to derivatives of functions $f : \mathbb{R} \to \mathbb{R}$, we find a scalar value $\frac{df}{dx}(x_0) \in \mathbb{R}$ for each fixed $x_0 \in \mathbb{R}$. However, if the derivative exists at multiple points, we consider $\frac{df}{dx}$ as a function itself. Now, the "best" analogue of this object $\frac{df}{dx}$ in higher dimensions $f : \mathbb{R}^n \to \mathbb{R}^m$ is the Jacobian matrix Df (or the gradient when $m = 1$). As in the one-dimensional case, we can write Df explicitly and easily see how it is a function on all values which it is defined. Further, just like in the one-dimensional case where we have $f'(x_0)(v) = \frac{df}{dx}(x_0)v$ where the righthand side is scalar multiplication, in higher dimensions we have $f'(\mathbf{x}_0)(\mathbf{v}) = Df(\mathbf{x}_0)\mathbf{v}$ where the righthand side is matrix multiplication. □

THEOREM 8.3.19. *Suppose that the partial derivatives $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}$ of $f : U \to \mathbb{R}$ exist in some neighborhood of $\mathbf{x}_0$ and are continuous at $\mathbf{x}_0$. Then f is differentiable at $\mathbf{x}_0$.*

Proof. We will prove this is true in the case $n = 2$ from which it will be clear that it holds in higher dimensions. As is standard for $n = 2$, we will write f_x for $\frac{\partial f}{\partial x_1}$ and f_y for $\frac{\partial f}{\partial x_2}$. Let $\mathbf{x}_0 = (x, y)$ and $\mathbf{h} = (h_1, h_2) \neq (0, 0)$. Define the linear operator $\Lambda_{\mathbf{x}_0} : \mathbb{R}^2 \to \mathbb{R}$ by $\Lambda_{\mathbf{x}_0}(v_1, v_2) = f_x(\mathbf{x}_0)v_1 + f_y(\mathbf{x}_0)v_2$. We will show that

$$\lim_{\mathbf{h}\to(0,0)} \frac{|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \Lambda_{\mathbf{x}_0}\mathbf{h}|}{\|\mathbf{h}\|} = 0.$$

First considering the numerator for sufficiently small $\mathbf{h}$,

$$\begin{aligned}
&f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \Lambda_{\mathbf{x}_0}\mathbf{h} = f(x + h_1, y + h_2) - f(x, y) - f_x(x, y)h_1 - f_y(x, y)h_2 \\
&= \big(f(x + h_1, y + h_2) - f(x, y + h_2) - f_x(x, y)h_1\big) + \big(f(x, y + h_2) - f(x, y) - f_y(x, y)h_2\big) \\
&= \big(f_x(x + \theta_1, y + h_2)h_1 - f_x(x, y)h_1\big) + \big(f_y(x, y + \theta_2)h_2 - f_y(x, y)h_2\big) \\
&= \big(f_x(x + \theta_1, y + h_2) - f_x(x, y)\big)h_1 + \big(f_y(x, y + \theta_2) - f_y(x, y)\big)h_2.
\end{aligned}$$

where we arrived at the second equality by adding 0 and the third equality by applying the mean value theorem with $\theta_1 \in [0, h_1]$ and $\theta_2 \in [0, h_2]$. Applying the triangle inequality,

$$\begin{aligned}
&\frac{|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \Lambda_{\mathbf{x}_0}\mathbf{h}|}{\|\mathbf{h}\|} \\
&\leq \frac{|f_x(x + \theta_1, y + h_2) - f_x(x, y)|\,|h_1|}{\|\mathbf{h}\|} + \frac{|f_y(x, y + \theta_2) - f_y(x, y)|\,|h_2|}{\|\mathbf{h}\|} \\
&= \frac{|h_1|}{\|\mathbf{h}\|}\big|f_x(x + \theta_1, y + h_2) - f_x(x, y)\big| + \frac{|h_2|}{\|\mathbf{h}\|}\big|f_y(x, y + \theta_2) - f_y(x, y)\big|.
\end{aligned}$$

By the assumption of continuity of f_x and f_y at $\mathbf{x}_0 = (x, y)$, we have that $|f_x(x+\theta_1, y+h_2) - f_x(x, y)| \to 0$ as $\mathbf{h} \to (0, 0)$ (since this implies that $\theta_1 \to 0$ and $h_2 \to 0$), and similarly $|f_y(x, y+\theta_2) - f_x(x, y)| \to 0$ as $\mathbf{h} \to (0, 0)$. Moreover $|h_1|/\|\mathbf{h}\| \leq 1$ and $|h_2|/\|\mathbf{h}\| \leq 1$. Therefore, both terms in the final sum converge to 0 as $\mathbf{h} \to (0, 0)$, proving that f is differentiable at $\mathbf{x}_0$ (and $f'(\mathbf{x}_0) = \Lambda_{\mathbf{x}_0}$, but we already knew that from Theorem 8.3.15). ■

COROLLARY 8.3.20. *Suppose that $g : U \to \mathbb{R}^m$ and for each $k = 1, 2, ..., m$ the partial derivatives $\frac{\partial g_k}{\partial x_1}, ..., \frac{\partial g_k}{\partial x_n}$ exist within some neighborhood of $\mathbf{x}_0$ and are continuous at $\mathbf{x}_0$. Then g is differentiable at $\mathbf{x}_0$.*

Proof. By Proposition 8.2.6, g is differentiable at $\mathbf{x}_0$ if and only if g_k is differentiable at $\mathbf{x}_0$ for each $k = 1, 2, ..., m$. By the assumption of continuous partial derivatives on the components of g, we can apply Theorem 8.3.19 to each component g_k to conclude that each componenet g_k is differentiable at $\mathbf{x}_0$, which concludes the proof. ■

3.2. An Extremal Property of the Derivative. For this section we let U be an open subset of $\mathbb{R}^n$, $\mathbf{x}_0 \in U$, $f : U \to \mathbb{R}^m$, and for each $k = 1, 2, ..., m$, let f_k be the kth component of f.

PROPOSITION 8.3.21. *Suppose that f is differentiable at $\mathbf{x}_0$. Then the following are equivalent:*

(1) *$f'(\mathbf{x}_0) = \mathbf{0}$ (here, as usual, $\mathbf{0} : \mathbb{R}^n \to \mathbb{R}^m$ is the zero map $\mathbf{0}(\mathbf{x}) = 0_m$).*
(2) *For all $k = 1, 2, ..., m$ and $x = 1, 2, ..., n$, $\frac{\partial f_k}{\partial x_i}(\mathbf{x}_0) = 0$.*
(3) *For all $k = 1, 2, ..., m$, $\nabla f_k(\mathbf{x}_0) = 0_n$, the zero vector in $\mathbb{R}^n$.*
(4) *$Df(\mathbf{x}_0)$ is the $m \times n$ zero matrix.*

DEFINITION 8.3.22. If f is differentiable at $\mathbf{x}_0$ such that $f'(\mathbf{x}_0) = \mathbf{0}$ (or any of the equivalent conditions of Proposition 8.3.21), then $\mathbf{x}_0$ is called a *critical point* of f.

DEFINITION 8.3.23. A function $g : U \to \mathbb{R}$ is said to have a *local maximum* at $\mathbf{x}_0$ when there exists some neighborhood $V \subset U$ of $\mathbf{x}_0$ such that $g(\mathbf{x}) \leq g(\mathbf{x}_0)$ holds for all $\mathbf{x} \in V$. Similarly, we say that g has a *local minimum* at $\mathbf{x}_0$ when there exists some neighborhood $S \subset U$ of $\mathbf{x}_0$) such that $g(\mathbf{x}) \geq g(\mathbf{x}_0)$ holds for all $\mathbf{x} \in S$.

THEOREM 8.3.24. *Suppose that $g : U \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$ and has a local maximum or minimum at $\mathbf{x}_0$. Then $\mathbf{x}_0$ is a critical point of g.*

3.3. Some notes about the derivative $g' : V \to \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$ as a function. Suppose $g : U \to \mathbb{R}^m$ is differentiable at all points in some open subset $V \subset U$. Then $g' : V \to \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$ is a function from the metric space V to the Banach space $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$. At this point, we have much technology built up to analyze functions between metric spaces and Banach spaces, so we can apply this technology to g'. For this section we will focus mostly on continuity properties.

PROPOSITION 8.3.25. *Suppose that $g : U \to \mathbb{R}^m$ is differentiable in some open subset $V \subset U$, and $\mathbf{x}_0 \in V$. Then the following are equivalent.*

1. *The derivative $g' : V \to \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$ is continuous at $\mathbf{x}_0$ when endowing $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$ with the operator norm. That is, given any $\varepsilon > 0$ there exists some $\delta > 0$ such that if $\|\mathbf{x}_0 - \mathbf{x}\|_n < \delta$ then $\|f'(\mathbf{x}_0) - f'(\mathbf{x})\|_{op} < \varepsilon$.*
2. *For all $k = 1, 2, ..., m$ and all $i = 1, 2, ..., n$, the partial derivatives $\frac{\partial g_k}{\partial x_i} : V \to \mathbb{R}$ are all continuous at $\mathbf{x}_0$.*
3. *For all $k = 1, 2, ..., m$, the gradients of each component $\nabla g_k : V \to \mathbb{R}^m$ are continuous at $\mathbf{x}_0$.*
4. *All component functions of the matrix Dg are continuous at $\mathbf{x}_0$.*

DEFINITION 8.3.26. Suppose that f is differentiable at $\mathbf{x}_0 \in U$. Then we say f is *continuously differentiable* at $\mathbf{x}_0$ when there exists a neighborhood V of $\mathbf{x}_0$ such that any of the equivalent conditions of Proposition 8.3.25 hold.

4. Inverse and Implicit Function Theorems

Suppose that A is an open subset of $\mathbb{R}^n$ and B is an open subset of $\mathbb{R}^m$. Then $A \times B \subset \mathbb{R}^n \times \mathbb{R}^m$ (identified with $\mathbb{R}^{n+m}$). We will denote points in $A \times B$ as $(\mathbf{x}, \mathbf{y}) = (x_1, ..., x_n, y_1, ..., y_m)$ with $\mathbf{x} \in A$ and $\mathbf{y} \in B$. Let $g : A \times B \to \mathbb{R}^p$ with components g_i for $i = 1, 2, ..., p$. For each fixed $\mathbf{x}$, we have the induced function $g^{\mathbf{x}} : B \to \mathbb{R}^p$ defined by $g^{\mathbf{x}}(\mathbf{y}) = g(\mathbf{x}, \mathbf{y})$; similarly, for each fixed $\mathbf{y} \in B$, we have the induced function $g^{\mathbf{y}} : A \to \mathbb{R}^p$ defined by $g^{\mathbf{y}}(\mathbf{x}) = g(\mathbf{x}, \mathbf{y})$. It is often useful to consider the properties of these induced functions. To that end, we make the following definition.

NOTATION 8.4.1. Suppose that for some $\mathbf{x}_0 \in A$, $g^{\mathbf{x}_0}$ is differentiable at the point $\mathbf{y}_0 \in B$, we write $D_{\mathbf{y}}g(\mathbf{x}_0, \mathbf{y}_0)$ as the Jacobian matrix of $g^{\mathbf{x}_0}$ at the point $\mathbf{y}_0$. That is,

$$D_{\mathbf{y}}g(\mathbf{x}_0, \mathbf{y}_0) = Dg^{\mathbf{x}_0}(\mathbf{y}_0) = \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(\mathbf{x}_0, \mathbf{y}_0) & \frac{\partial g_1}{\partial y_2}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial g_1}{\partial y_m}(\mathbf{x}_0, \mathbf{y}_0) \\ \frac{\partial g_2}{\partial y_1}(\mathbf{x}_0, \mathbf{y}_0) & \frac{\partial g_2}{\partial y_2}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial g_2}{\partial y_m}(\mathbf{x}_0, \mathbf{y}_0) \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial g_p}{\partial y_1}(\mathbf{x}_0, \mathbf{y}_0) & \frac{\partial g_p}{\partial y_2}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial g_p}{\partial y_m}(\mathbf{x}_0, \mathbf{y}_0) \end{pmatrix}$$

Similarly, when there is some $\mathbf{y}_0 \in B$ such that $g^{\mathbf{y}_0}$ is differentiable at $\mathbf{x}_0 \in A$, then we write

$$D_{\mathbf{x}}g(\mathbf{x}_0, \mathbf{y}_0) = Dg^{\mathbf{y}_0}(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}_0, \mathbf{y}_0) & \frac{\partial g_1}{\partial x_2}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}_0, \mathbf{y}_0) \\ \frac{\partial g_2}{\partial x_1}(\mathbf{x}_0, \mathbf{y}_0) & \frac{\partial g_2}{\partial x_2}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial g_2}{\partial x_m}(\mathbf{x}_0, \mathbf{y}_0) \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial g_p}{\partial x_1}(\mathbf{x}_0, \mathbf{y}_0) & \frac{\partial g_p}{\partial x_2}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial g_p}{\partial x_n}(\mathbf{x}_0, \mathbf{y}_0) \end{pmatrix}$$

Intuitively, we think of $D_{\mathbf{y}}$ and $D_{\mathbf{x}}$ as taking the derivatives in the $\mathbf{y}$ and $\mathbf{x}$ variables respectively.

THEOREM 8.4.2 (Inverse Function Theorem). *Let U be an open subset of $\mathbb{R}^n$, $\mathbf{x}_0 \in U$, and $f : U \to \mathbb{R}^n$. Suppose that f is continuously differentiable in U and $f'(\mathbf{x}_0) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Then there exists a neighborhood $\tilde{U} \subset U$ of $\mathbf{x}_0$ and a neighborhood $V \subset \mathbb{R}^n$ of $f(\mathbf{x}_0)$ such that the following hold:*

(1) *the restriction $f : \tilde{U} \to V$ is a bijection,*
(2) *the inverse $f^{-1} : V \to \tilde{U}$ is continuously differentiable in V, and*
(3) $(f^{-1})'(\mathbf{y}) = f'\big(f^{-1}(\mathbf{y})\big)^{-1}$ *for every $\mathbf{y} \in V$.*

Proof. We will start by proving a simpler case, and from there deduce the general statement.

Case 1: Assume that $f'(\mathbf{x}_0) = I$, where I is the identity operator: $I(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. Find $\delta > 0$ small enough that $\overline{B}(\mathbf{x}_0, 2\delta) \subset U$ and such that for all $\mathbf{x} \in \overline{B}(\mathbf{x}_0, 2\delta)$, both $\|f(\mathbf{x}) - f(\mathbf{x}_0)\| \leq 1/4$ and $\|f'(\mathbf{x}) - I\|_{op} \leq 1/4$ (we can find such a δ since U is open and f is assumed continuously differentiable with $f'(\mathbf{x}_0) = I$). For $\mathbf{x} \in U$, define $\psi(\mathbf{x}) = f(\mathbf{x}) - \mathbf{x}$;

then $\psi'(\mathbf{x}) = f'(\mathbf{x}) - I$. This implies that for any $\mathbf{x}, \mathbf{x}+\mathbf{h} \in \overline{B}(\mathbf{x}_0, 2\delta)$,

$$\|\psi(\mathbf{x}+\mathbf{h}) - \psi(\mathbf{x})\| \leq \frac{1}{2}\|\mathbf{h}\|. \tag{4.1}$$

Define $V = B(f(\mathbf{x}_0), \delta)$ and $\tilde{U} = \left\{\mathbf{x} \in \overline{B}(\mathbf{x}_0, 2\delta) \text{ s.t. } f(\mathbf{x}) \in \tilde{V}\right\} = f^{-1}(V) \cap \overline{B}(\mathbf{x}_0, 2\delta)$. We will show that $f : \tilde{U} \to V$ is a bijection. Given the definition of $\tilde{U}$, this boils down to proving that given any $\mathbf{y} \in V$ there exists a unique $\mathbf{x} \in \overline{B}(\mathbf{x}_0, 2\delta)$ such that $f(\mathbf{x}) = \mathbf{y}$. We prove this using the contraction mapping theorem 6.8.3. Indeed, fix $\mathbf{y} \in V$. For every $\mathbf{x} \in \overline{B}(\mathbf{x}_0, 2\delta)$, let $T(\mathbf{x}) = \mathbf{x} + (\mathbf{y} - f(\mathbf{x}))$. For $\mathbf{x}^* \in \overline{B}(\mathbf{x}_0, 2\delta)$, $f(\mathbf{x}^*) = \mathbf{y}$ if and only if $T(\mathbf{x}^*) = \mathbf{x}^*$, i.e. if and only if $\mathbf{x}^*$ is a fixed point of T. Since $\overline{B}(\mathbf{x}_0, 2\delta)$ is a closed subset of the complete metric space $\mathbb{R}^n$, it is a complete metric space. So, if we can show i) that $T : \overline{B}(\mathbf{x}_0, 2\delta) \to \overline{B}(\mathbf{x}_0, 2\delta)$, and ii) there exists some $k \in (0,1)$ such that for all $\mathbf{x}, \mathbf{y} \in \overline{B}(\mathbf{x}_0, 2\delta)$, $\|T(\mathbf{x}) - T(\mathbf{y})\| \leq k\|\mathbf{x} - \mathbf{y}\|$, then we can apply the contraction mapping theorem to conclude that there is a unique fixed point $\mathbf{x}^* \in \overline{B}(\mathbf{x}_0, 2\delta)$ such that $T(\mathbf{x}^*) = \mathbf{x}^*$; equivalently, there is a unique $\mathbf{x}^* \in \overline{B}(\mathbf{x}_0, 2\delta)$ such that $f(\mathbf{x}^*) = \mathbf{y}$.

For i) Let $\mathbf{x} \in \overline{B}(\mathbf{x}_0, 2\delta)$, and set $\mathbf{h} = \mathbf{x}_0 - \mathbf{x}$

$$\begin{aligned}\|T(\mathbf{x}) - \mathbf{x}_0\| &= \|\mathbf{x} + (\mathbf{y} - f(\mathbf{x})) - \mathbf{x}_0\| = \|\big(\mathbf{x} - f(\mathbf{x}) + f(\mathbf{x}_0) - \mathbf{x}_0\big) + \big(\mathbf{y} - f(\mathbf{x}_0)\big)\| \\ &\leq \|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \mathbf{h}\| + \|\mathbf{y} - f(\mathbf{x}_0)\| < \frac{1}{2}\|h\| + \delta \leq \delta + \delta = 2\delta.\end{aligned}$$

The final two inequalities come from (4.1), the fact that $\mathbf{y} \in V = B(f(\mathbf{x}_0), \delta)$, and that $\|\mathbf{h}\| = \|\mathbf{x} - \mathbf{x}_0\| \leq 2\delta$. What this shows is that for any $\mathbf{x} \in \overline{B}(\mathbf{x}_0, 2\delta)$, $T(\mathbf{x}) \in \overline{B}(\mathbf{x}_0, 2\delta)$.

For ii) let $\mathbf{x}, \mathbf{y} \in \overline{B}(\mathbf{x}_0, 2\delta)$ and set $\mathbf{h} = \mathbf{y} - \mathbf{x}$. Then we have $T(\mathbf{y}) - T(\mathbf{x}) = -(\psi(\mathbf{x} + h) - \psi(\mathbf{x}))$. By (4.1), this implies

$$\|T(\mathbf{y}) - T(\mathbf{x})\| = \|\psi(\mathbf{x}+\mathbf{h}) - \psi(\mathbf{x})\| \leq \frac{1}{2}\|\mathbf{h}\| = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|.$$

This proves that $T : \overline{B}(\mathbf{x}_0, 2\delta) \to \overline{B}(\mathbf{x}_0, 2\delta)$ is a contraction mapping on the complete metric space $\overline{B}(\mathbf{x}_0, 2\delta)$, and hence the claim of bijectivity of $f : \tilde{U} \to V$. ■

One application of the Inverse Function Theorem is the Implicit Function Theorem, which we have all many times in our past, unaware of the details surrounding seemingly algebraic manipulation.

EXAMPLE 8.4.3. We present this example formally as one might see in an elementary calculus class. Using implicit differentiation, let's find the slope $\frac{dy}{dx}$ of the graph $x^2 + y^2 = 25$

at the point $(x, y) = (4, 3)$. To do so, let $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by $g(x, y) = x^2 + y^2$ and notice we are looking at the equation $g(x, y) = 25$. Using the chain rule

$$\frac{d}{dx}\big(g(x, y) = 25\big) \implies g_x(x, y) + g_y(x, y)\frac{dy}{dx} = 0.$$

Solving for $\frac{dy}{dx}$,

$$\frac{dy}{dx}(x, y) = -\frac{g_x(x, y)}{g_y(x, y)}. \tag{4.2}$$

From here, $g_x(4, 3) = 2(4) = 8$ and $g_y(4, 3) = 2(3) = 6$. Therefore,

$$\frac{dy}{dx}(4, 3) = -\frac{4}{3}.$$

□

REMARK 8.4.4. Let's make some notes about some of the manipulations used in the last example. The first is that when looking at the equation $g(x, y) = 25$, we *implicitly* assume there is some function $h(x)$ such that $y = h(x)$; at least near the point $(4, 3)$ we can use the equation $x^2 + y^2 = 25$ to "solve for" $y = \sqrt{25 - x^2} = h(x)$. Particularly, near $(x, y) = (4, 3)$, we are assuming we can write $g(x, y) = 25$ as $g(x, h(x)) = 25$.

Another something to point out is that when solving for dh/dx (aka dy/dx), we arrived at (4.2),

$$\frac{dh}{dx}(x, y) = -\frac{g_x(x, y)}{g_y(x, y)}.$$

Particularly, we need that $g_y(x, y) \neq 0$ for (x, y) near $(4, 3)$; this requirement is equivalent to saying $g_y(x, y)$ is invertible near $(x, y) = (4, 3)$. □

THEOREM 8.4.5 (Implicit Function Theorem). *Let A and B be open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ and $g : A \times B \to \mathbb{R}^m$. Let $(\mathbf{x}_0, \mathbf{y}_0) \in A \times B$ and $\mathbf{a} = g(\mathbf{x}_0, \mathbf{y}_0)$. Suppose that g is continuously differentiable in $A \times B$ and $D_{\mathbf{y}} g(\mathbf{x}_0, \mathbf{y}_0)$ is invertible. Then there exists a neighborhood $\tilde{A} \subset A$ of $\mathbf{x}_0$ and a function $h : \tilde{A} \to \mathbb{R}^m$ such that the following hold:*

(1) $h(\mathbf{x}_0) = \mathbf{y}_0$,
(2) $g(\mathbf{x}, h(\mathbf{x})) = \mathbf{a}$ *for all* $\mathbf{x} \in \tilde{A}$, *and*
(3) h *is continuously differentiable in* $\tilde{A}$.

Proof. Let $f : A \times B \to \mathbb{R}^{n+m}$ be defined by $f(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, g(\mathbf{x}, \mathbf{y}))$. Then f is continuously differentiable on $A \times B$ and $f'(\mathbf{x}_0, \mathbf{y}_0)$ is invertible. Indeed, we that $Df(\mathbf{x}_0, \mathbf{y}_0)$ is the

block matrix

$$Df(\mathbf{x}_0,\mathbf{y}_0)=\begin{pmatrix} I_{n\times n} & 0_{n\times m} \\ 0_{m\times n} & D_{\mathbf{y}}g(\mathbf{x}_0,\mathbf{y}_0)\end{pmatrix}$$

where $I_{n\times n}$ is the $n\times n$ identity, $0_{n\times m}$ is the $n\times m$ zero matrix, and similarly for $0_{m\times n}$. The assumption that $D_{\mathbf{y}}g(\mathbf{x}_0,\mathbf{y}_0)$ is invertible implies that $Df(\mathbf{x}_0,\mathbf{y}_0)$ is an invertible matrix, which is equivalent to the invertibility of $f'(\mathbf{x}_0,\mathbf{y}_0)$. Therefore, the inverse function theorem tells us that there are open subsets $U\subset A\times B$ and $V\subset\mathbb{R}^{n+m}$ such that $(\mathbf{x}_0,\mathbf{y}_0)\in U$, $f:U\to V$ is a bijection, and $f^{-1}:V\to U$ is continuously differentiable. Since $U\subset A\times B$, there are component functions $k_1:V\to A$ and $k_2:V\to B$ such that $f^{-1}=(k_1,k_2)$. For any $(\mathbf{x},\mathbf{v})\in V$, $f(f^{-1}(\mathbf{x},\mathbf{v}))=(\mathbf{x},\mathbf{v})$; however, $f(f^{-1}(\mathbf{x},\mathbf{v}))=f(k_1(\mathbf{x},\mathbf{v}),k_2(\mathbf{x},\mathbf{v}))=(k_1(\mathbf{x},\mathbf{v}),g(k_1(\mathbf{x},\mathbf{v}),k_2(\mathbf{x},\mathbf{v})))$. By equating these two expressions for $f(f^{-1}(\mathbf{x},\mathbf{v}))$, we learn that i) $k_1(\mathbf{x},\mathbf{v})=\mathbf{x}$ and ii) every $\mathbf{x}$ component of the vectors in V must coincide with the $\mathbf{x}$ components of the vectors in U; that is,

$$\{\mathbf{x}\in\mathbb{R}^n \text{ s.t. } (\mathbf{x},\mathbf{v})\in V \text{ for some } \mathbf{v}\in\mathbb{R}^m\}=\{\mathbf{x}\in\mathbb{R}^n \text{ s.t. } (\mathbf{x},\mathbf{y})\in U \text{ for some } \mathbf{y}\in\mathbb{R}^m\}.$$

From here, $(\mathbf{x},\mathbf{v})=f(\mathbf{x},k_2(\mathbf{x},\mathbf{v}))=(\mathbf{x},g(\mathbf{x},k_2(\mathbf{x},\mathbf{v})))$ from which we learn the $g(\mathbf{x},k_2(\mathbf{x},\mathbf{v}))=\mathbf{v}$. Let $(\mathbf{x}_0,\mathbf{v}_0)\in V$ such that $f^{-1}(\mathbf{x}_0,\mathbf{v}_0)=(\mathbf{x}_0,\mathbf{y}_0)$ (we can find such a $\mathbf{v}_0$ since $f^{-1}:V\to U$ is a bijection and $(\mathbf{x}_0,\mathbf{y}_0)\in U$); particularly, $k_2(\mathbf{x}_0,\mathbf{v}_0)=\mathbf{y}_0$. Let $\tilde{A}$ be defined by $\tilde{A}=\{\mathbf{x}\in A \text{ s.t. } (\mathbf{x},\mathbf{v}_0)\in V\}$. For every $\mathbf{x}\in\tilde{A}$, $g(\mathbf{x},k_2(\mathbf{x},\mathbf{v}_0))=\mathbf{v}_0$. However, $g(\mathbf{x}_0,k_2(\mathbf{x}_0,\mathbf{v}_0))=g(\mathbf{x}_0,\mathbf{y}_0)=\mathbf{a}$. Hence $\mathbf{v}_0=\mathbf{a}$ and for every $\mathbf{x}\in\tilde{A}$, $g(\mathbf{x},k_2(\mathbf{x},\mathbf{a}))=\mathbf{a}$. Define $h:\tilde{A}\to\mathbb{R}^m$ by $h(\mathbf{x})=k_2(\mathbf{x},\mathbf{a})$. We now have that $h(\mathbf{x}_0)=\mathbf{y}_0$ and for every $\mathbf{x}\in\tilde{A}$, $g(\mathbf{x},h(\mathbf{x}))=\mathbf{a}$. It remains to prove that h is continuously differentiable in $\tilde{A}$. However, when we appealed to the inverse function theorem, the inverse function f^{-1} we found was continuously differentiable. In turn this implies the continuous differentiability of the components k_1 and k_2, and hence of h since $h(\mathbf{x})=k_2(\mathbf{x},\mathbf{a})$. ■

5. Exercises

(1) Prove Proposition 8.1.4 directly by showing that $\{\mathbf{e}_k\}_{k=1}^n$ is an orthonormal set such that for every $\mathbf{x}\in\mathbb{R}^n$ we have $\mathbf{x}=\sum_{k=1}^n\langle\mathbf{x},\mathbf{e}_k\rangle\,\mathbf{e}_k$. Justify that once you prove this, you can conclude that $\{\mathbf{e}_k\}_{k=1}^n$ is an orthonormal basis of $\mathbb{R}^n$.

(2) Prove Proposition 8.1.5

(3) Let $f:\mathbb{R}\to\mathbb{R}$ and $x_0\in\mathbb{R}$.

(a) Prove that f is differentiable at x_0 in the sense of Definition 8.2.1 if and only if f is differentiable at x_0 in the way defined in elementary calculus (Definition 5.1.1).

(b) In the case that f is differentiable at x_0, let λ_{x_0} be the scalar value you would assign to $f'(x_0)$ in a calculus course and let Λ_{x_0} be the linear operator you would assign to $f'(x_0)$ through Definition 8.2.1. For each $x \in \mathbb{R}$, give an explicit formula for $\Lambda_{x_0} x$ in terms of λ_{x_0} and x.

(4) Let $m : \mathbb{R}^2 \to \mathbb{R}$ be the multiplication function: $m(x, y) = xy$. For any $(x_0, y_0) \in \mathbb{R}$ show that the derivative $m'(x_0, y_0) \in \mathcal{L}(\mathbb{R}^2 \to \mathbb{R})$ is given by

$$m'(x_0, y_0)(x, y) = y_0 x + x_0 y \quad \text{for all } (x, y) \in \mathbb{R}^2.$$

(5) Let $a : \mathbb{R}^2 \to \mathbb{R}$ be the addition function: $a(x, y) = x + y$. For any $(x_0, y_0) \in \mathbb{R}$, find the derivative $a'(x_0, y_0)$ and evaluate $a'(\pi, e)(1, 0)$.

(6) Let $f : \mathbb{R} \to \mathbb{R}^3$ be defined by

$$f(x) = \Big(e^{x^2+1}, \int_0^x e^{-t^2}\, dt,\, 5\Big).$$

For any $x_0 \in \mathbb{R}$, find the derivative $f'(x_0) \in \mathcal{L}(\mathbb{R} \to \mathbb{R}^2)$ and evaluate $f'\big(\sqrt{\ln(2)}\,\big)\frac{1}{2}$.

(7) (a) Prove Proposition 8.2.5 (1).

(b) Prove Proposition 8.2.5 (2).

(c) Prove Proposition 8.2.5 (3).

(d) Prove Proposition 8.2.5 (4).

(e) Prove Proposition 8.2.5 (5).

(8) For the given function f and given values of $\mathbf{x}_0$ and $\mathbf{v}$, calculate $\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0)$:

(a) $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $f(s, t) = (\cos(s + t), e^{s^2 t}, t)$; $\mathbf{v} = (1, 2)$; $\mathbf{x}_0 = (1, 0)$.

(b) $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(r, s, t) = r^2 \sin(st)$; $\mathbf{v} = (1, 0, 0)$; $\mathbf{x}_0 = (-1, 3\pi/2, 1)$.

(c) $f : \mathbb{R} \to \mathbb{R}$ defined by $f(t) = e^{2\pi t}$; $\mathbf{v} = 1$; $\mathbf{x}_0 = t$. In fact, what you show here is that $\frac{df}{dt}(t) = \frac{\partial f}{\partial 1}(t)$ (this is true in general for differentiable functions $\mathbb{R} \to \mathbb{R}$ with our notation), though the notation $\frac{\partial f}{\partial 1}(t)$ will never again be used in these notes.

(9) Fill in the details of the proof of Theorem 8.3.10.

(10) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(s,t) = \begin{cases} \frac{st(t^2-s^2)}{s^2+t^2} & (s,t) \neq (0,0) \\ 0 & (s,t) = 0 \end{cases}$$

Let $\mathbf{e}_1 = (1,0)$ and $\mathbf{e}_2 = (0,1)$. Show that $\frac{\partial^2 f}{\partial \mathbf{e}_1 \partial \mathbf{e}_2}(0,0)$ and $\frac{\partial^2 f}{\partial \mathbf{e}_2 \partial \mathbf{e}_1}(0,0)$ exist, but are unequal.

(11) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined as $f(s,t) = (s^2+t^2, s^2-t^2)$. Prove that there exist $\theta_1, \theta_2 \in [0,2]$ such that

$$f(3,5) - f(1,1) = 2 \cdot \left(\frac{\partial f_1}{\partial (1,2)}(1+\theta_1, 1+2\theta_1),\ \frac{\partial f_2}{\partial (1,2)}(1+\theta_2, 1+2\theta_2) \right)$$

and find such values θ_1 and θ_2.

(12) Prove Corollary 8.3.16.

(13) Prove Corollary 8.3.20. *Hint:* Look at Proposition 8.2.6.

(14) In the proof of Theorem 8.3.19, we write, "First considering the numerator for sufficiently small $\mathbf{h}$.." What this means is that there is some $\varepsilon > 0$ such that the statement holds for $\|\mathbf{h}\| < \varepsilon$. Why do we need sufficiently small $\mathbf{h}$? How would you choose this epsilon?

(15) Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$. Suppose that f is differentiable at $\mathbf{x}_0 = (a_1, a_2, ..., a_n) \in U$. For a fixed k, let $\varphi(t) = f(a_1, ..., a_{k-1}, t, a_{k+1}, ..., x_n)$ for t in a sufficiently small neighborhood J of a_k; we can choose $J = (a_k - \varepsilon, a_k + \varepsilon)$ for some $\varepsilon > 0$ such that for all $t \in J$, $(a_1, ..., a_{k-1}, t, a_{k+1}, ..., x_n) \in U$.

(a) Prove that $\frac{d\varphi}{dt}(a_k) = \frac{\partial f}{\partial x_k}(\mathbf{x}_0)$. (Recall that we use $\frac{d\varphi}{dt}(a_k)$ as the scalar value identification of the derivative).

(b) In calculus, you mechanically learn that the partial derivative f_{x_k} is "taking the derivative in the x_k variable with all others held fixed." Briefly explain why part a is a proof of this statement.

(c) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = e^{x+y}\sin(\cos(xy)+1)$. Find $f_x(x,y)$ and $f_y(x,y)$ and make sure to justify how you arrive at your solution. You may assume that the standard functions e^x, $\sin(x)$, and $\cos(x)$ are differentiable on $\mathbb{R}$ with derivatives $\frac{de^x}{dx} = e^x$, $\frac{d\sin(x)}{dx} = \cos(x)$, and $\frac{d\cos(x)}{dx} = -\sin(x)$.

(16) Suppose that U is an open subset of $\mathbb{R}$. Let $f : U \to \mathbb{R}$ and assume f is differentiable at $x_0 \in \mathbb{R}$. We have made a "big deal" about the identification of the derivative of

f at x_0 as a linear operator $f'(x_0)$ as opposed to the identification of the derivative as a scalar $\frac{df}{dt}(x_0)$. More generally, if V is an open subset of $\mathbb{R}^n$ and $g : V \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$, this same type of identification can be made between the linear operator $g'(\mathbf{x}_0)$ and the vector $\nabla g(\mathbf{x}_0)$. For this problem, show that this is a reasonable generalized identification by proving that $\frac{df}{dt}(x_0) = \nabla f(x_0)$ (note that by definition $\nabla f(x_0)$ is a vector in $\mathbb{R}$, but a vector in $\mathbb{R}$ is just a scalar!).

(17) We continue along the same lines as the previous problem. Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$. Assume f is differentiable at $\mathbf{x}_0$. Show that $Df(\mathbf{x}_0) = \nabla f(\mathbf{x}_0)$ (identifying the $1 \times n$ matrix $Df(\mathbf{x}_0)$ as a vector and $\nabla f(\mathbf{x}_0)$ as a row vector).

(18) Let $U = (0, \infty) \times (0, 2\pi) \times (-\pi, \pi) \subset \mathbb{R}^3$ and $g : U \to \mathbb{R}^3$ be defined as

$$g(r, \theta, \phi) = \big(r\cos(\theta)\sin(\phi), r\sin(\theta)\sin(\phi), r\cos(\phi)\big).$$

a) Find $Dg(r, \theta, \phi)$.

b) Let $\{\mathbf{e}_k\}_{k=1}^3$ be the standard basis in $\mathbb{R}^3$. Find $\dfrac{\partial g}{\partial \mathbf{e}_k}(r, \theta, \phi)$ for $k = 1, 2, 3$.

c) Let $\mathbf{v} = (-1, 10, 5)$ and $\mathbf{x}_0 = (1, \pi, 0)$. Find the directional derivative of g at the point $\mathbf{x}_0$ with respect to $\mathbf{v}$. That is, find $\dfrac{\partial g}{\partial \mathbf{v}}(\mathbf{x}_0)$.

(19) Let $U \subset \mathbb{R}^n$ be an open subset and suppose that $f : U \to \mathbb{R}^m$ is differentiable. Let $V \subset \mathbb{R}^m$ be an open subset containing the image $f(U)$ and suppose $g : V \to \mathbb{R}^p$ is differentiable. Prove that for any $\mathbf{x} \in U$, it holds that $D(g \circ f)(\mathbf{x}) = Dg(f(\mathbf{x}))Df(\mathbf{x})$, where the righthand side is matrix multiplication. Use this to further prove that if $p = 1$ (so that $g : \mathbb{R}^m \to \mathbb{R}$), then $\nabla(g \circ f)(\mathbf{x}) = \nabla g(f(\mathbf{x}))Df(\mathbf{x})$, where the righthand side is matrix multiplication of the row vector $\nabla g(f(\mathbf{x}))$ and the matrix $Df(\mathbf{x})$.

(20) Let $x : \mathbb{R} \to \mathbb{R}$ and $y : \mathbb{R} \to \mathbb{R}$ be continuously differentiable functions for $i = 1, 2$, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable. Define the function $g : \mathbb{R} \to \mathbb{R}$ by $g(t) = f(x(t), y(t))$. Prove

$$\frac{dg}{dt}(t) = \frac{\partial f}{\partial x}(x(t), y(t))\,\frac{dx}{dt}(t) + \frac{\partial f}{\partial y}(x(t), y(t))\,\frac{dy}{dt}(t).$$

This is a justification of the chain rule abusively written (but oft used) as

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\,\frac{dy}{dt}.$$

Generalize this chain rule to a function $f : \mathbb{R}^n \to \mathbb{R}$ and prove its validity.

(21) Let $U \subset \mathbb{R}^2$ be the right half plane $U = \big\{(x,y) \in \mathbb{R}^2 \text{ s.t. } x > 0\big\}$. Define the functions $r : U \to \mathbb{R}$ and $\theta : U \to \mathbb{R}$ by $r(x,y) = \sqrt{x^2+y^2}$ and $\theta(x,y) = \arctan(y/x)$. Define the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x,y) = x^2\,(\cos(y), \sin(y))$. Find the total derivative matrix $Df(r,\theta)$ at all points in U. That is, find the total derivative matrix for the function $f(r,\theta) : U \to \mathbb{R}^2$ with $f(r,\theta)(x,y) = f(r(x,y),\theta(x,y))$. *Hint*: Let $g(x,y) = (r(x,y),\theta(x,y))$ and notice you are finding $D(f \circ g)$.

(22) Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is a function such that $f'(\mathbf{x})$ is invertible for every $\mathbf{x} \in \mathbb{R}^n$. Prove that for any open subset $U \subset \mathbb{R}^n$, the image $f(U)$ is an open subset of $\mathbb{R}^n$. (Such maps where the image of open sets are open are called *open maps*). *Hint*: Inverse function theorem.

(23) For some $a, b \in (0,\infty)$, let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x,y) = \frac{x^2}{a^2} + \frac{y^2}{b^2}.$$

Using the implicit function theorem, find all points $(x,y) \in \mathbb{R}^2$ such that there is an interval J containing x and a continuously differentiable function $h : J \to \mathbb{R}$ such that $f(x,h(x)) = 1$ for every $x \in J$. Further, find distinct formulae for the function h.

(24) In this problem we prove that if a function $f : \mathbb{R}^2 \to \mathbb{R}$ is continuously differentiable and there exists some $\delta > 0$ such that $\frac{\partial f}{\partial y}(x,y) \geq \delta$ for every $(x,y) \in \mathbb{R}^2$, then there exists a continuously differentiable function $h : \mathbb{R} \to \mathbb{R}$ such that $f(x,h(x)) = 0$ for every $x \in \mathbb{R}$, and moreover, if $f(x,y) = 0$, then $h(x) = y$.

(a) Prove that if $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $g'(x) \geq \delta$ for every $x \in \mathbb{R}$, then there is a unique point $x_0 \in \mathbb{R}$ such that $g(x_0) = 0$.

(b) Prove the existence of a function $h : \mathbb{R} \to \mathbb{R}$ such that $f(x,h(x)) = 0$ for every $x \in \mathbb{R}$. *Hint:* For each fixed $x \in \mathbb{R}$, apply the previous part to the function $g(y) = f(x,y)$.

(c) Show that the function $h : \mathbb{R} \to \mathbb{R}$ you found in the previous part is continuously differentiable and argue that if $f(x,y) = 0$, then $h(x) = y$. *Hint*: You may find the implicit function theorem helpful.

(d) You've finished the proof. Now, use it to prove that there exists a function $h : \mathbb{R} \to \mathbb{R}$ which solves the equation $h(x)^3 + h(x) = x$.

CHAPTER 9

Introduction to Measure Theory

Unless stated otherwise, we will assume that $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$ for this chapter.

1. Motivation

If someone were to ask you how you measure the length (i.e., one-dimensional volume) of generic subset $E \subset \mathbb{R}$, what would you answer? A reasonable start would be to say that if $E = (a, b)$ is an interval, then the measure of E is given by $b - a$. By investigating further, you would likely stumble upon the following properties as desirable for such a measure:

i) The measure of $[0, 1]$ is 1.

ii) If $E \subset \mathbb{R}$ and $F \subset \mathbb{R}$ such that F is a *translation* of E (i.e. there is some $t_0 \in \mathbb{R}$ such that $F = E + t_0 = \{x + t_0 : x \in E\}$), then the length of E and the length of F are the same.

iii) If $E_1, E_2, ...$ are all disjoint subsets of $\mathbb{R}$, then the length of their union $\cup E_i$ should equal the sum of each of their individual lengths.

So, after much thought, you come up with the following idea to study how to measure subsets of $\mathbb{R}$: Let $2^{\mathbb{R}}$ be the power set of $\mathbb{R}$ and let $\mu : 2^{\mathbb{R}} \to [0, \infty]$ be called a *measure* on $\mathbb{R}$ when:

i) $\mu([0, 1]) = 1$,

ii) $\mu(E + t_0) = \mu(E)$ for any $E \in 2^{\mathbb{R}}$ and $t_0 \in \mathbb{R}$, and

iii) If $E_1, E_2, ...$ are disjoint subsets in $2^{\mathbb{R}}$, then $\mu(\cup E_i) = \sum \mu(E_i)$.

Do we get to go pick up our fields medal now? Unfortunately not. The problem is that the requirements we have on μ are inconsistent: there is no function $\mu : 2^{\mathbb{R}} \to [0, \infty]$ satisfying those three properties. The outline of how we would show such an inconsistency is as follows: We create a countable collection of (pretty strange looking) disjoint subsets $\{E_i\}_{i=1}^{\infty}$ such that each E_i is a translation of E_1 and $\cup_{i=1}^{\infty} E_i = [0, 1]$. Then, $\mu([0, 1]) = \mu\big(\cup_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} \mu(E_i)$ by property iii). However, $\mu(E_i) = \mu(E_1)$ by property ii), so if $\mu(E_i) > 0$ then then $\sum_{i=1}^{\infty} \mu(E_i) = \sum_{i=1}^{\infty} \mu(E_1) = \infty$; if $\mu(E_1) = 0$ then $\sum_{i=1}^{\infty} \mu(E_i) = \sum_{i=1}^{\infty} 0 = 0$. You conclude that $\mu([0, 1]) = 0$ or ∞, which breaks property i).

However, not all is lost. We could just forgo property i) and declare that the measure of every set is 0 (or that the measure of every non-empty interval is ∞), but for the case we want to assign a positive finite measure to non-empty intervals, we need to do some work.

2. Measure Spaces

2.1. Algebras.

NOTATION 9.2.1. We continue to denote the power set of a set Ω by 2^{Ω}.

DEFINITION 9.2.2. Let Ω be a set and $\mathcal{A} \subset 2^{\Omega}$. We say that $\mathcal{A}$ is an *algebra* (or as probabilists say, *field*) on Ω when the following three properties hold.

(1) $\varnothing \in \mathcal{A}$.

(2) If $E_1, E_2 \in \mathcal{A}$, then $E_1 \cup E_2 \in \mathcal{A}$.

(3) If $E \in \mathcal{A}$, then $E^c \in \mathcal{A}$.

If $\mathcal{F} \subset 2^{\Omega}$ is called an algebra, we further say that $\mathcal{F}$ is a *σ–algebra* (or for probabilistic, a *σ–field*) on Ω when $\mathcal{F}$ also satisfies the following.

(2') If $E_1, E_2, E_3, ... \in \mathcal{F}$ is any countable collection of sets in $\mathcal{F}$, then $\cup_{i=1}^{\infty} E_i \in \mathcal{F}$.

If $\mathcal{A}$ satisfies Property 2, we say that $\mathcal{A}$ is *closed under finite unions*; if it satisfies Property 3, we say that $\mathcal{A}$ is *closed under compliments*. Similarly, if $\mathcal{F}$ satisfies Property 9.2.2, we say that $\mathcal{F}$ is *closed under countable unions*. Analogous uses of the phrase "*closed under*" when referencing a collection of subsets are used frequently throughout this chapter and are common in literature.

PROPOSITION 9.2.3. *Algebras are closed under finite intersections and σ-algebras are closed under countable intersections.*

Proof. Suppose $\mathcal{A}$ is an algebra or σ-algebra. If $E_1, E_2, ..., E_n \in \mathcal{A}$ (where $n = \infty$ is allowed when $\mathcal{A}$ is a σ-algebra), then by Definition 9.2.2 (3), $E_1^c, E_2^c, ..., E_n^c \in \mathcal{A}$ and hence $\bigcup_{i=1}^{n} E_i^c \in \mathcal{A}$. Taking the compliment of this union, we have $\left[\bigcup_{i=1}^{n} E_i^c\right]^c = \bigcap_{i=1}^{n} E_i$ by D'Morgan's Laws. By another application of Definition 9.2.2 (3), we then deduce $\bigcap_{i=1}^{n} E_i \in \mathcal{A}$. ■

LEMMA 9.2.4. *Let Ω be a set, then $\{\varnothing, \Omega\}$ and 2^{Ω} are both σ–algebras on Ω.*

PROPOSITION 9.2.5. *The intersection of arbitrarily many algebras on Ω is again an algebra on Ω. The same is true for σ–algebras.*

Proof. Exercise 4 (3). ■

THEOREM 9.2.6. *Let $\mathcal{E} \subset 2^\Omega$ be any arbitrary subset of 2^Ω. Then there exists a unique σ–algebra, denoted $\sigma(\mathcal{E})$, such that $\mathcal{E} \subset \sigma(\mathcal{E})$ and such that if $\mathcal{F}$ is any other σ–algebra with $\mathcal{E} \subset \mathcal{F}$, then $\sigma(\mathcal{E}) \subset \mathcal{F}$.*

Proof. Define $\sigma(\mathcal{E})$ as the intersection of all σ–algebras on Ω containing $\mathcal{E}$. We know that there exists at least one such σ–algebra, since 2^Ω is σ–algebra. By Proposition 9.2.5, $\sigma(\mathcal{E})$ is a σ–algebra. Moreover, since every σ–algebra in the intersection contains $\mathcal{E}$, then $\mathcal{E} \subset \sigma(\mathcal{E})$. Particularly, if $\{\mathcal{F}_j\}_{j\in J}$ is the collection of all σ–algebras containing $\mathcal{E}$ (indexed by some set J), then $\sigma(\mathcal{E}) = \cap_{j\in J}\mathcal{F}_j$, from which we can conclude that if $\mathcal{F}$ is any σ–algebra containing $\mathcal{E}$, $\sigma(\mathcal{E}) \subset \mathcal{F}$. If $\tilde{\sigma}(\mathcal{E})$ is any another sigma algebra satisfying the conclusions of this theorem, then since $\mathcal{E} \subset \tilde{\sigma}(\mathcal{E})$ we have shown $\sigma(\mathcal{E}) \subset \tilde{\sigma}(\mathcal{E})$; since $\mathcal{E} \subset \sigma(\mathcal{E})$, then by assumption $\tilde{\sigma}(\mathcal{E}) \subset \sigma(\mathcal{E})$. Thus $\tilde{\sigma}(\mathcal{E}) = \sigma(\mathcal{E})$ and uniqueness is proved. ■

DEFINITION 9.2.7. Given $\mathcal{E} \subset 2^\Omega$, the σ–algebra $\sigma(\mathcal{E})$ is called the *σ–algebra generated by $\mathcal{E}$*.

One main reason we care about "generating" an (σ–)algebra is because working with arbitrary sets within the algebra might not be tractable since they don't need to be nice. However, we can often get away with working only with a generating collection of sets which we better understand.

LEMMA 9.2.8. *Suppose that $\mathcal{F}, \mathcal{E} \subset 2^\Omega$. If $\mathcal{F} \subset \sigma(\mathcal{E})$, then $\sigma(\mathcal{F}) \subset \sigma(\mathcal{E})$.*

Proof. Exercise 4 (4). ■

DEFINITION 9.2.9. Let Ω be a metric space and $\tau \subset 2^\Omega$ be the collection of all open subsets of Ω. The *Borel σ–algebra* on Ω is defined to be $\sigma(\tau)$. We will denote the Borel σ–algebra by $\mathcal{B}_\Omega = \sigma(\tau)$.

LEMMA 9.2.10. *If (M, ρ) is a metric space and $\mathcal{B}_M$ is the Borel σ-algebra on M, then for every $x \in M$, the singleton set $\{x\}$ is an element in $\mathcal{B}_M$. That is, for every $x \in M$, it holds that $\{x\} \in \mathcal{B}_M$.*

Proof. Let $x \in M$. Then for every $n \in \mathbb{N}$, we have that $B(x; 1/n) \in \mathcal{B}_X$. Hence $\{x\} = \bigcap_{n=1}^{\infty} B(x; 1/n) \in \mathcal{B}_X$. ■

THEOREM 9.2.11. *Suppose that Ω is a separable metric space. That is, Ω is a metric space and there exists a subset $Q \subset \Omega$ which is countable and dense (with respect to the given metric). Let $\mathcal{E} \subset 2^\Omega$ be the collection of all open balls in Ω. Then $\mathcal{B}_\Omega = \sigma(\mathcal{E})$. Poetically, the Borel σ–algebra is generated by the open balls.*

Proof. We have proved that any open set in a separable metric space can be written as the union of countably many open balls. Therefore every open set must be in the σ-algebra generated by the open balls. This shows that if τ is the collection of open sets and $\mathcal{E}$ is the collection of all open balls, then $\tau \subset \sigma(\mathcal{E})$, implying that $\sigma(\tau) \subset \sigma(\mathcal{E})$ by Lemma 9.2.8. Similarly, it is clear that $\mathcal{E} \subset \tau \subset \sigma(\tau)$, and hence $\sigma(\mathcal{E}) \subset \sigma(\tau)$, giving the equality we need. ■

COROLLARY 9.2.12. *The Borel σ–algebra on $\mathbb{R}^n$ is generated by the open balls.*

COROLLARY 9.2.13. *The Borel σ-algebra on $\mathbb{R}$ is generated by each of the following.*

(1) *Open intervals:* $\mathcal{E}_1 = \{(a, b) \text{ s.t. } a < b\}$
(2) *Closed intervals:* $\mathcal{E}_2 = \{[a, b] \text{ s.t. } a < b\}$
(3) *Half open intervals:* $\mathcal{E}_3 = \{[a, b) \text{ s.t. } a < b\}$ *or* $\mathcal{E}_4 = \{(a, b] \text{ s.t. } a < b\}$
(4) *Open half lines:* $\mathcal{E}_5 = \{(a, \infty) \text{ s.t. } a \in \mathbb{R}\}$ *or* $\mathcal{E}_6 = \{(-\infty, a) \text{ s.t. } a \in \mathbb{R}\}$
(5) *Closed half lines:* $\mathcal{E}_7 = \{[a, \infty) \text{ s.t. } a \in \mathbb{R}\}$ *or* $\mathcal{E}_8 = \{(-\infty, a] \text{ s.t. } a \in \mathbb{R}\}$.

Proof. Let τ be the collection of all open subsets of $\mathbb{R}$.

(1) The open intervals in $\mathbb{R}$ are the open balls in $\mathbb{R}$, so this claim is a direct result of Theorem 9.2.11.

(2) For any $a, b \in \mathbb{R}$ with $a < b$, we have the equalities

$$[a, b] = \bigcap_{n=1}^{\infty} (a - 1/n, b + 1/n) \quad \text{and} \quad (a, b) = \bigcup_{n=1}^{\infty} [a - 1/n, b + 1/n]$$

from which we can deduce that $\mathcal{E}_2 \subset \sigma(\mathcal{E}_1)$ and $\mathcal{E}_1 \subset \sigma(\mathcal{E}_2)$. From Lemma 9.2.8, this implies that $\sigma(\mathcal{E}_1) = \sigma(\mathcal{E}_2)$. Since $\sigma(\mathcal{E}_1) = \mathcal{B}_\mathbb{R}$ by (1), we conclude that $\sigma(\mathcal{E}_2) = \mathcal{B}_\mathbb{R}$.

(3), (4), and (5) Exercise 4 (5). ■

2.2. Measures.

DEFINITIONS 9.2.14. Let $\mathcal{A} \subset 2^\Omega$ be an algebra and suppose $\mu : \mathcal{A} \to [0, \infty]$.

(1) μ is called a *finitely additive measure* on $\mathcal{A}$ when:

i) $\mu(\varnothing) = 0$,

ii) (finite additivity) if $E_1, E_2 \in \mathcal{A}$ are disjoint, then $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$.

(2) μ is called a *pre-measure* on $\mathcal{A}$ when:

i) μ is a finitely additive measure,

ii) (countable additivity) if $\{E_i\}_{i=1}^{\infty} \subset \mathcal{A}$ is a countable collection of disjoint sets such that $\cup_{i=1}^{\infty} E_i \in \mathcal{A}$, then $\mu\big(\cup_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} \mu(E_i)$.

(3) μ is called a *measure* when:

i) $\mathcal{A}$ is a σ-algebra,

ii) μ is a pre-measure on $\mathcal{A}$.

EXAMPLE 9.2.15. Let Ω be endowed with any σ-algebra $\mathcal{F}$, we define the *Dirac delta measure* at a point $\omega_0 \in \Omega$, denoted δ_{ω_0}, by

$$\delta_{\omega_0}(A) = \begin{cases} 1 & \omega_0 \in A \\ 0 & \text{otherwise} \end{cases}$$

We now check to ensure that δ_{ω_0} is, in fact, a measure on $\mathcal{F}$. First, $\delta_{\omega_0}(\varnothing) = 0$ since $\omega_0 \notin \varnothing$. Second, if $\{E_i\}_{i=1}^{\infty}$ are disjoint measurable subsets of Ω, we want to show that $\delta_{\omega_0}(\cup_i E_i) = \sum_i \delta_{\omega_0}(E_i)$. We consider two cases. Case 1 is that $\omega_0 \in \cup_{i=1}^{\infty} E_i$. Hence, there exists exactly one $i_0 \in \mathbb{N}$ such that $\omega_0 \in E_{i_0}$ and (by the disjoint assumption) this implies that $\omega_0 \notin E_i$ for all $i \neq i_0$. We therefore have that $1 = \delta_{\omega_0}(\cup_{i=1}^{\infty} E_i)$ and $\sum_{i=1}^{\infty} \delta_{\omega_0}(E_i) = \delta_{\omega_0}(E_{i_0}) + \sum_{i \neq i_0} \delta_{w_0}(E_i) = 1 + 0 = 1$. So, in this case, $\delta_{\omega_0}(\cup_{i=1}^{\infty} E_i) = 1 = \sum_{i=1}^{\infty} \delta_{\omega_0}(E_i)$. Case 2 is that $\omega_0 \notin \cup_{i=1}^{\infty} E_i$. In this case $\omega_0 \notin E_i$ for all $i \in \mathbb{N}$ and therefore $\delta_{\omega_0}(E_i) = 0$ for all i. Hence $\delta(\cup_{i=1}^{\infty} E_i) = 0 = \sum_{i=1}^{\infty} 0 = \sum_{i=1}^{\infty} \delta_{\omega_0}(E_i)$. □

DEFINITIONS 9.2.16. Let $\mathcal{F}$ be a σ-algebra on Ω.

(1) When we endow Ω with $\mathcal{F}$, we say that $(\Omega, \mathcal{F})$ is a *measurable space*. Any subset $E \subset \Omega$ such that $E \in \mathcal{F}$ is called a *measurable* set, or if we need to be extra explicit, an *$\mathcal{F}$-measurable* set.

(2) If further, $\mu : \mathcal{F} \to [0, \infty]$ is a measure on $\mathcal{F}$, then when endowing Ω with $\mathcal{F}$ and μ, we say that $(\Omega, \mathcal{F}, \mu)$ is a *measure space*.

(3) If $\mu(\Omega) < \infty$, we call μ a *finite measure*.

(4) In the special case that $\mu(\Omega) = 1$, we call μ a *probability measure*, or often just a *probability*.

EXAMPLE 9.2.17. Let Ω be the set of all sequences of 0s and 1s of length n. That is $\Omega = \{(x_1, ..., x_n) \text{ s.t. } x_i \in \{0, 1\} \text{ for each } i\}$. Let $\mathcal{F} = 2^\Omega$, which we have shown is a σ-algebra on Ω. Let $p \in (0, 1)$, and define the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ by

$$\mathbb{P}(\{(x_1, ..., x_n)\}) = p^{(\#\text{ of 1s})}(1-p)^{(\#\text{ of 0s})}.$$

For example, if $n = 5$ and $\omega = (1, 0, 0, 1, 0)$, then $\mathbb{P}(\omega) = p^2(1-p)^3$. In Exercise 4 (10), you prove that $\mathbb{P}$ does indeed define a probability measure on $(\Omega, \mathcal{F})$. Now, consider the function $X : \Omega \to \mathbb{R}$ defined by $X(\omega) = \sum_{i=1}^{n} x_i$, where $\omega = (x_1, ..., x_n)$; effectively, $X(\omega)$ counts the number of 1s in the sequence ω. The image of X, we will denote S_X, is $S_X = X(\Omega) = \{0, 1, 2, ..., n\}$. Let $\mathcal{B} = 2^{S_X}$. Define the probability μ on $(S_X, \mathcal{B})$ by $\mu(\{k\}) = \mathbb{P}(\{\omega \in \Omega \text{ s.t. } X(\omega) = k\}) = \mathbb{P}(X^{-1}(\{k\}))$ for each $k \in S_X$. By Exercise 4 (10), you are guaranteed that μ is a measure on Ω. By simple counting arguments, we find that $\mu(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$. The intuition in this example is that you flip a coin n times with a probability p of achieving heads on any single flip. You numerically keep track of the results by assigning a 1 for the ith flip if it resulted in heads, otherwise you assign a 0. In this case, X represents the number of heads achieved in n flips, and the measure μ is defined (in probabilistic notation) by $\mu(\{k\}) = \mathbb{P}(X = k)$. □

PROPOSITION 9.2.18. *Suppose that $(\Omega, \mathcal{F}, \mu)$ is a measure space, E is a measurable set and $F \subset E$ is also measurable. Then $\mu(E) = \mu(F) + \mu(E \setminus F)$. In particular, if $\mu(F) < \infty$, then $\mu(E \setminus F) = \mu(E) - \mu(F)$. Moreover, if μ is a finite measure, then for any measurable sets E_1, E_2, we have $\mu(E_1 \cup E_2) = \mu(E_1 \cap E_2) + \mu(E_1 \Delta E_2)$, where $E_1 \Delta E_2 = (E_1 \setminus E_2) \cup (E_2 \setminus E_1)$ is the symmetric difference of the sets E_1 and E_2. In particular, if $\mu(E_1 \Delta E_2) < \infty$, then $\mu(E_1 \cap E_2) = \mu(E_1 \cup E_2) - \mu(E_1 \Delta E_2)$.*

LEMMA 9.2.19. *Suppose that $\{E_n\}_{n=1}^\infty$ is a sequence of subsets of Ω. Then there exists a sequence $\{F_n\}_{n=1}^\infty$ of subsets of Ω which are disjoint, $F_n \subset E_n$ for each n, and $\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} F_n$. Moreover, if Ω is endowed with a σ-algebra $\mathcal{F}$ and each of the E_n is measurable, then the disjoint sequence $\{F_n\}_{n=1}^\infty$ can be found such that F_n is also measurable for each n.*

Proof. The basic idea is the following. Let $F_1 = E_1$, then recursively define $F_{n+1} = E_{n+1} \setminus \left(\bigcup_{k=1}^{n} E_k \right)$. The details are left for Exercise 4 (**??**). ■

THEOREM 9.2.20 (Continuity of Measure). *Let $(\Omega, \mathcal{F}, \mu)$ is a measurable space.*

(1) *If $E_1 \subset E_2 \subset E_3 \subset \cdots$ is an increasing sequence of measurable sets, then*

$$\mu\Big(\bigcup_{n=1}^{\infty} E_n \Big) = \lim_{n\to\infty} \mu(E_n)$$

(2) *If $E_1 \supset E_2 \subset E_3 \supset \cdots$ is a decreasing sequence of measurable sets and $\mu(E_n) < \infty$ for some n, then*

$$\mu\Big(\bigcap_{n=1}^{\infty} E_n \Big) = \lim_{n\to\infty} \mu(E_n).$$

Proof.

■

PROPOSITION 9.2.21 (Sub-Additivity). *Let $\mathcal{A}$ be an algebra on Ω and μ a finitely additive measure on $\mathcal{A}$. Then the following hold.*

(1) *If $E, F \in \mathcal{A}$ and $F \subset E$, then $\mu(F) \leq \mu(E)$.*

(2) *If $\{E_i\}_{i=1}^{n} \subset \mathcal{A}$ then $\mu(\cup_{i=1}^{n} E_i) \leq \sum_{i=1}^{n} \mu(E_i)$.*

If we further assume that $\mathcal{A}$ is a σ-algebra and μ is a measure, then we can replace (2) with

(2') *If $\{E_i\}_{i=1}^{\infty} \subset \mathcal{A}$ then $\mu(\cup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu(E_i)$.*

Proof. (1) In Proposition 9.2.18, we see that $\mu(E) = \mu(F) + \mu(E \setminus F)$. Since $\mu(E \setminus F) \geq 0$, we must have $\mu(F) \leq \mu(E)$.

(1) Let $\{F_i\}_{i=1}^{n}$ be the disjoint sequence of measurable sets from Lemma **??**. Then $F_i \subset E_i$ and $\cup F_i = \cup E_i$. Hence,

$$\mu\Big(\bigcup_{i=1}^{\infty} E_i \Big) = \mu\Big(\bigcup_{i=1}^{\infty} F_i \Big) = \sum_{i=1}^{n} \mu(F_i) \leq \sum_{i=1}^{n} \mu(E_i).$$

■

2.3. The Extended Real Line.

DEFINITION 9.2.22. Find two elements which are not elements of $\mathbb{R}$, denote them by ∞ and $-\infty$. The *extended real numbers*, denoted $\overline{\mathbb{R}}$, is the set $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$, endowed with additional ordering and metric type structures we lightly develop in this section.

While there are many ways to (equivalently) define the structure on $\overline{\mathbb{R}}$ so that it behaves as our intuition tells us it should, we will develop it in the following way. Define the bijection $h : [-1, 1] \to \overline{\mathbb{R}}$ by

$$h(x) = \begin{cases} -\infty & x = -1 \\ \dfrac{x}{(1+x)(1-x)} & x \in (-1, 1). \\ \infty & x = 1 \end{cases}$$

You can easily check that h does, in fact, define a bijection. Effectively, what h allows us to do is "stretch" the interval $[-1, 1]$ to the extended real numbers $\overline{\mathbb{R}}$ in such a way that $(-1, 1) \mapsto \mathbb{R}$, $-1 \mapsto -\infty$, and $1 \mapsto \infty$. From this, we export all the standard Euclidean structure from $[-1, 1]$ to $\overline{\mathbb{R}}$. In particular, we have the following

(1) (Ordering) For $x, y \in \overline{\mathbb{R}}$, we have $x < y$ if and only if $h^{-1}(x) < h^{-1}(y)$.
(2) (Alternative Metric) Define a metric ρ_{alt} on $\overline{\mathbb{R}}$ by $\rho_{alt}(x, y) = |h^{-1}(x) - h^{-1}(y)|$.

THEOREM 9.2.23. *Let ρ be the standard Euclidean metric on $\mathbb{R}$, let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$, and let $x \in \overline{\mathbb{R}}$.*

(1) *If $x \in \mathbb{R}$, then the sequence converges to x with respect to ρ if and only if it converges to x with respect to ρ_{alt}.*
(2) *The sequence diverges to ∞ (resp. $-\infty$) in the usual sense if and only if it converges to ∞ (resp. $-\infty$) in $\overline{\mathbb{R}}$ with respect to ρ_{alt}.*
(3) *A subset $U \subset \mathbb{R}$ is open in $\mathbb{R}$ with respect to ρ if and only if it is open in $\overline{\mathbb{R}}$ with respect to ρ_{alt}. Moreover, a subset $U \subset \overline{\mathbb{R}}$ is open with respect to ρ_{alt} if and only if $h^{-1}(U)$ is open as a subset of $[-1, 1]$ viewed as a metric space with respect to ρ.*

Theorem 9.2.23 is useful because it tells us that we can use ρ_{alt} as a tool to make rigorous metric structures on $\overline{\mathbb{R}}$ which stay consistent with those that are already extant on $\mathbb{R}$. However, realize that the existence of ρ_{alt} is just as a rigorous tool. In practice, we still like to endow $\overline{\mathbb{R}}$ with the standard Euclidean structure the best we can, so for $x, y \in \overline{\mathbb{R}}$, we still use the intuitive Euclidean like distance between x and y when we can. In particular, we put ordered field-like operations on $\overline{\mathbb{R}}$ in the following way.

(1) For every $x, y \in \mathbb{R} \subset \overline{\mathbb{R}}$, $x + y$ and xy are defined in the usual way as elements of $\mathbb{R}$.
(2) For every $x \in \mathbb{R} \cup \{\infty\}$, we define $x + \infty = \infty$. If $x > 0$, then $x \cdot \infty = \infty$. If $x < 0$ then $x \cdot \infty = -\infty$. We leave $0 \cdot \infty$ undefined.

(3) For every $x \in \mathbb{R} \cup \{-\infty\}$, we define $x + (-\infty) = x - \infty = -\infty$. If $x > 0$, then $x \cdot (-\infty) = -\infty$. If $x < 0$ then $x \cdot (-\infty) = \infty$. We leave $0 \cdot (-\infty)$ undefined.

(4) We leave $\infty + (-\infty)$ undefined.

(5) For every $x \in \overline{\mathbb{R}}$, we define

$$|x| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}$$

(6) For any subset $U \subset \overline{\mathbb{R}}$, we define the *supremum* of U as

$$\sup(U) = \begin{cases} \infty & \text{if either } \infty \in U \text{ or } U \cap \mathbb{R} \text{ is non-empty and not bounded above} \\ \sup(U \cap \mathbb{R}) & \text{if } \infty \notin U \text{ and } U \cap \mathbb{R} \text{ is non-empty and bounded above} \\ -\infty & U = \{-\infty\} \text{ or } U = \varnothing \end{cases}$$

Similarly, we define the *infimum* of U as

$$\inf(U) = \begin{cases} -\infty & \text{if either } -\infty \in U \text{ or } U \cap \mathbb{R} \text{ is non-empty and not bounded below} \\ \inf(U \cap \mathbb{R}) & \text{if } -\infty \notin U \text{ and } U \cap \mathbb{R} \text{ is non-empty and bounded below} \\ \infty & U = \{\infty\} \text{ or } U = \varnothing \end{cases}$$

From here, it makes sense to talk about the "Euclidean distance" on $\overline{\mathbb{R}}$ by requiring the distance between x and y to be $|x - y|$ whenever this is defined. Notice that $|x - y|$ will take values in $\overline{\mathbb{R}}$ (for instance if one of the x or y is an infinite value, then this distance is either not defined, or ∞). Because this version of Euclidean distance on $\overline{\mathbb{R}}$ is consistent with the true Euclidean distance on $\mathbb{R}$ when restricted to $\mathbb{R}$, Theorem 9.2.23 ensures that we can use this distance to critique converge and of sequences (even to the points ∞ and $-\infty$ in the divergent cases). We do this henceforth without explicitly referenced ρ_{alt} unless the need arises. Before moving to define the Borel σ-algebra on $\overline{\mathbb{R}}$, we state the monotone convergence theorem for sequences in $\overline{\mathbb{R}}$ and apply it to limit supremums and limit infimums.

THEOREM 9.2.24. *All monotone sequences in $\overline{\mathbb{R}}$ converge.*

Proof. In Theorem **??**, we proved that every monotone sequence in $\mathbb{R}$ either converges in $\mathbb{R}$ if and only if it is bounded, otherwise it diverges. From Theorem 9.2.23 we learn that a sequence in $\mathbb{R}$ diverges if and only if it converges to an infinite value in $\overline{\mathbb{R}}$. Therefore, if

$\{x_n\}_{n=1}^{\infty}$ is a monotone sequence in $\mathbb{R}$, then it converges in $\overline{\mathbb{R}}$. The case that $\{x_n\}_{n=1}^{\infty}$ takes on some infinite values is covered in Exercise 4 (**??**). ■

DEFINITION 9.2.25. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in in $\overline{\mathbb{R}}$. The *limit supremum* of this sequence, denoted $\limsup_{n\to\infty} x_n$, is defined as

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \left(\sup\{x_n, x_{n+1}, x_{n+2}, ...\} \right) = \lim_{n\to\infty} \sup\{x_k \text{ s.t. } k \geq n\}.$$

Analogously, the *limit infimum*, denoted $\liminf_{n\to\infty} x_n$, is defined as

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \left(\inf\{x_n, x_{n+1}, x_{n+2}, ...\} \right) = \lim_{n\to\infty} \inf\{x_k \text{ s.t. } k \geq n\}.$$

THEOREM 9.2.26. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\overline{\mathbb{R}}$. The following hold.*

(1) *Both $\limsup_{n\to\infty} x_n$ and $\liminf_{n\to\infty} x_n$ exist in $\overline{\mathbb{R}}$.*
(2) *$\limsup_{n\to\infty} x_n \geq \liminf_{n\to\infty} x_n$ with equality if and only if $\lim_{n\to\infty} x_n$ exists. In this case, $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \lim_{n\to\infty} x_n$.*

Proof. (1) By defining $y_n = \sup\{x_k \text{ s.t. } k \geq n\}$, the sequence $\{y_n\}_{n=1}^{\infty}$ is monotonically decreasing (as it is taking the sup of smaller and smaller subsets of the original sequence). Since monotone sequences always converge in $\overline{\mathbb{R}}$, we know that $\lim_{n\to\infty} y_n$ convereges. However, $\lim_{n\to\infty} y_n = \limsup_{n\to\infty} x_n$, and therefore $\limsup_{n\to\infty} x_n$ exists. The symmetric argument proves the case for $\liminf$.

(2) It is easily shown that $\limsup_{n\to\infty} x_n \geq \liminf_{n\to\infty} x_n$. Suppose that equality holds and let $x = \limsup_{n\to\infty} x_n$. We will assume that $x \in \mathbb{R}$, where the case that $x \in \{-\infty, \infty\}$ is the substance of Exercise 4 (**??**). It is clear then that for any n,

$$\inf\{x_k \text{ s.t. } k \geq n\} - x \leq x_n - x \leq \sup\{x_k \text{ s.t. } k \geq n\} - x$$

If $\liminf x_n = \limsup x_n = x$, a squeezing argument shows that $x_n - x \to 0$ as $n \to \infty$, proving that $\lim_{n\to\infty} x_n = x$. Conversely, if $x_n \to x$, then given any $\varepsilon > 0$, we can find some n_ε such that for every $n \geq n_\varepsilon$, it holds that $|x_n - x| < \varepsilon$, implying that for $n \geq n_\varepsilon$

$$x_n - x - \varepsilon \leq \inf\{x_k \text{ s.t. } k \geq n\} - x \leq x_n - x \leq \sup\{x_k \text{ s.t. } k \geq n\} - x \leq x_n - x + \varepsilon$$

which, by the arbitrarity of $\varepsilon > 0$ and a squeezing argument, proves that if $x_n \to x$, then $\liminf x_n = x$ and $\limsup x_n = x$. ■

We now arrive at the motivation for our defining of the extended real numbers. We will consider $\overline{\mathbb{R}}$ as a measure space. Since we have defined $\overline{\mathbb{R}}$ as a metric space, we can define the Borel σ-algebra on $\overline{\mathbb{R}}$.

PROPOSITION 9.2.27. *Let* $\mathcal{J} = \{\varnothing, \{-\infty\}, \{\infty\}, \{-\infty, \infty\}\}$. *Then*

$$\mathcal{B}_{\overline{\mathbb{R}}} = \bigcup_{J \in \mathcal{J}} \{F \cup J \text{ s.t. } F \in \mathcal{B}_{\mathbb{R}}\}.$$

In other words, a set E *is in* $\mathcal{B}_{\overline{\mathbb{R}}}$ *if and only if there is some set* $\tilde{E} \in \mathcal{B}_{\mathbb{R}}$ *such that either* $E = \tilde{E}$, *or* $E = \tilde{E} \cup \{-\infty\}$, *or* $E = \tilde{E} \cup \{\infty\}$, *or* $E = \tilde{E} \cup \{-\infty, \infty\}$.

Proof Outline. The idea of the proof goes as follows. Every open set U of $\overline{\mathbb{R}}$ is either a standard open subset of $\mathbb{R}$ or the union of a standard open subset of $\mathbb{R}$ with an interval of the form $(a, \infty]$ or $[-\infty, b)$. This leads us to conclude that $\mathcal{B}_{\overline{\mathbb{R}}}$ is generated by the collection $\mathcal{E} = \{\mathcal{B}_{\mathbb{R}}, \{\infty\}, \{-\infty\}\}$. From here, define $\mathcal{R} = \left\{\tilde{E} \cup F \text{ s.t. } \tilde{E} \in \mathcal{B}_{\mathbb{R}} \text{ and } F \in \mathcal{J}\right\}$. By the definition of $\mathcal{R}$, $\mathcal{E} \subset \mathcal{R}$. As is a common tactic, we now prove that $\mathcal{R}$ is a σ-algebra, from which we will deduce that $\sigma(\mathcal{E}) \subset \mathcal{R}$. Further, by the definition of $\mathcal{R}$, it is clear that $\mathcal{R} \subset \sigma(\mathcal{E})$, concluding the proof. ∎

PROPOSITION 9.2.28. *Suppose that* $f : \Omega \to \overline{\mathbb{R}}$ *is a function on the measurable space* $(\Omega, \mathcal{F})$ *such that* $f(\Omega) \subset \mathbb{R}$. *Then* f *is* $(\mathcal{F}/\mathcal{B}_{\overline{\mathbb{R}}})$*-measurable when considered a function* $\Omega \to \overline{\mathbb{R}}$ *if and only if* f *is* $(\mathcal{F}/\mathcal{B}_{\mathbb{R}})$*-measurable when considered a function* $\Omega \to \mathbb{R}$.

Proof. This is a Corollary of Proposition 9.3.39 below, so we leave it unproven here. ∎

PROPOSITION 9.2.29. *The following sets generate* $\mathcal{B}_{\overline{\mathbb{R}}}$.

(1) $\mathcal{E}_1 = \{(a, b) \text{ s.t. } a, b \in \mathbb{R} \text{ with } a < b\} \cup \{\{\infty\}\}$
(2) $\mathcal{E}_2 = \{[a, b] \text{ s.t. } a, b \in \overline{\mathbb{R}} \text{ with } a < b\}$
(3) $\mathcal{E}_3 = \{[a, b) \text{ s.t. } a \in \overline{\mathbb{R}}, b \in \mathbb{R}, \text{ and } a < b\}$ *or* $\mathcal{E}_4 = \{(a, b] \text{ s.t. } a \in \mathbb{R}, b \in \overline{\mathbb{R}}, \text{ and } a < b\}$
(4) $\mathcal{E}_5 = \{(a, \infty) \text{ s.t. } a \in \mathbb{R}\} \cup \{\{\infty\}\}$ *or* $\mathcal{E}_6 = \{(-\infty, a) \text{ s.t. } a \in \mathbb{R}\} \cup \{\{\infty\}\}$
(5) $\mathcal{E}_7 = \{[a, \infty) \text{ s.t. } a \in \overline{\mathbb{R}}\}$ *or* $\mathcal{E}_8 = \{(-\infty, a] \text{ s.t. } a \in \overline{\mathbb{R}}\}$
(6) $\mathcal{E}_9 = \{[-\infty, a) \text{ s.t. } a \in \mathbb{R}\}$ *or* $\mathcal{E}_{10} = \{(a, \infty] \text{ s.t. } a \in \mathbb{R}\}$.

Proof. We will prove (1) and leave the rest as an exercise for those interested. Using Corollary 9.2.13, $\mathcal{B}_{\mathbb{R}} \subset \sigma(\mathcal{E}_1)$ since the open intervals are a subset of $\mathcal{E}_1$. Since $\{\infty\} \in \mathcal{E}_1$, this further implies that $\mathbb{R} \cup \{\infty\} = (-\infty, \infty] \in \sigma(\mathcal{E}_1)$, and thusly $(-\infty, \infty]^c = \{-\infty\} \in \sigma(\mathcal{E}_1)$.

This shows that $\mathcal{B}_{\mathbb{R}} \cup \{\{-\infty\}, \{\infty\}\} \subset \sigma(\mathcal{E}_1)$ and therefore $\mathcal{B}_{\overline{\mathbb{R}}} \subset \sigma(\mathcal{E}_1)$. However, since $\mathcal{E}_1 \subset \mathcal{B}_{\mathbb{R}} \cup \{\{-\infty\}, \{\infty\}\}$, implying that $\sigma(\mathcal{E}_1) \subset \mathcal{B}_{\overline{\mathbb{R}}}$, concluding the proof. ■

REMARK 9.2.30. In what follows, we are often going to be considering monotone sequences of functions $\{f_n\}_{n=1}^{\infty}$ whose domain is Ω and whose range is a subset of $[0, \infty]$. (By monotone, we mean $f_n(\omega) \leq f_{n+1}(\omega)$ for every $n \in \mathbb{N}$ and $\omega \in \Omega$). By including the possibility of ∞ as an acceptable range value, then we know that every monotone sequence of real numbers converge (possibly to the value of ∞ if monotonically increasing, and $-\infty$ if monotonically decreasing). So, by allowing ourselves to consider the extended real numbers, we can assuredly state that every sequence of monotone, non-negative functions $\{f_n\}_{n=1}^{\infty}$ converges pointwise to some f whose range is a subset of $[0, \infty]$. □

3. Integration of Measurable Functions

3.1. Measurable Functions.

DEFINITION 9.3.1. Let $(\Omega, \mathcal{F}_\Omega)$ and $(Y, \mathcal{F}_Y)$ be measurable spaces and $f : \Omega \to Y$. We say that f is $(\mathcal{F}_\Omega/\mathcal{F}_Y)$-measurable when for every $f \in \mathcal{F}_Y$ it holds that $f^{-1}(F) \in \mathcal{F}_\Omega$.

NOTATION 9.3.2. The notation $f^{-1}(F) = \{\omega \in \Omega \text{ s.t. } f(\omega) \in F\}$ is common for the pre-image of F with respect to f. Also $\{f \in F\}$ is another common way to denote this pre-image; $\{f \in F\} = f^{-1}(F)$. This notation is intuitive insofar as $\{f \in F\}$ is the set of $\omega \in \Omega$ such that $f(\omega) \in F$. Moreover, if $F = \{y\}$ is a singleton set, then it is common to write $\{f = y\}$ for $f^{-1}(\{y\})$; again this is intuitive since $\{f = y\}$ is the set of $\omega \in \Omega$ such that $f(\omega) = y$.

Moreover, given some measure μ on $\mathcal{F}_\Omega$, we will often be measuring sets such as $\{f \in F\}$. It is common to write $\mu(f \in F)$ for $\mu(\{f \in F\})$ or $\mu(f = y)$ for $\mu(\{f = y\})$. If $Y = \mathbb{R}$, expressions such as $\mu(a < f < b)$ or $\mu(f \geq a)$ etc. are often used in place of $\mu(f \in (a, b))$ or $\mu(f \in [a, \infty))$ etc.

EXAMPLE 9.3.3. If either Ω is endowed with the σ–algebra 2^Ω *or* Y is endowed with the σ–algebra $\{\varnothing, Y\}$, then every function $\Omega \to Y$ is measurable. On the other hand, if Ω is endowed with the σ–algebra $\{\varnothing, \Omega\}$ *and* Y is endowed with any σ-algebra containing all singleton sets, then the only measurable functions $\Omega \to Y$ are constant functions. □

THEOREM 9.3.4. *Let $(\Omega, \mathcal{F}_\Omega)$ and $(Y, \mathcal{F}_Y)$ be measurable spaces and $f : \Omega \to Y$. Suppose that $\mathcal{E} \subset 2^Y$ and $\mathcal{F}_Y = \sigma(\mathcal{E})$. Then f is measurable if and only if $f^{-1}(E) \in \mathcal{F}_\Omega$ for every $E \in \mathcal{E}$.*

Proof. If f is measurable, then the fact that $\mathcal{E} \subset \mathcal{F}_Y$ implies that $f^{-1}(E) \in \mathcal{F}$ for every $E \in \mathcal{E}$. So, the only real substance is the claim that if $f^{-1}(E) \in \mathcal{F}$ for every $E \in \mathcal{E}$, then f is measurable; we prove this presently. Let $\mathcal{R} = \left\{E \in 2^\Omega \text{ s.t. } f^{-1}(E) \in \mathcal{F}_\Omega\right\}$. By assumption $\mathcal{E} \subset \mathcal{R}$. We will show that $\mathcal{R}$ is a σ-algebra, which will finish the proof because $\mathcal{E} \subset \mathcal{R}$ implies that $\sigma(\mathcal{E}) \subset \mathcal{R}$; with $\mathcal{F}_Y = \sigma(\mathcal{E})$ by assumption, we can then conclude that $f^{-1}(E) \in \mathcal{F}_\Omega$ for every $E \in \mathcal{F}_Y$, proving the measurability of f. To start, $\varnothing \in \mathcal{R}$ since $f^{-1}(\varnothing) = \varnothing \in \mathcal{F}_\Omega$. Next, if $E \in \mathcal{R}$, then $f^{-1}(E) \in \mathcal{F}_\Omega$ by the definition of $\mathcal{R}$ and therefore $\left(f^{-1}(E)\right)^c = f^{-1}(E^c) \in \mathcal{F}_\Omega$, showing that $E^c \in \mathcal{R}$. Finally, if $\{E_n\}_{n=1}^\infty \subset \mathcal{R}$, then $f^{-1}(E_n) \in \mathcal{F}_\Omega$ for every n, implying that $\cup_n f^{-1}(E_n) \in \mathcal{F}_\Omega$, but $\cup_n f^{-1}(E_n) = f^{-1}\big(\cup_n E_n\big)$, and therefore $\cup_n E_n \in \mathcal{R}$. This proves that $\mathcal{R}$ is a σ-algebra and concludes the proof. ∎

THEOREM 9.3.5. *Suppose that $(\Omega, \mathcal{F}_\Omega)$, $(X, \mathcal{F}_X)$, and $(Y, \mathcal{F}_Y)$ are measurable spaces, $f : \Omega \to X$ and $g : X \to Y$ are measurable functions. Then $g \circ f : \Omega \to Y$ is a measurable function.*

Proof. Let $E \in \mathcal{F}_Y$ be arbitrary. Then $g^{-1}(E) \in \mathcal{F}_X$ by the assumption that g is measurable, and hence $f^{-1}(g^{-1}(E)) \in \mathcal{F}_\Omega$ by the assumption that f is measurable. This shows that for any $E \in \mathcal{F}_Y$, it holds that $(g \circ f)^{-1}(E) = f^{-1}(g^{-1}(E)) \in \mathcal{F}_\Omega$, proving that $g \circ f$ is measurable. ∎

3.2. Simple Functions.

DEFINITION 9.3.6. If $f : \Omega \to Y$ is a function between sets, we say that f is a *simple function* whenever $\#f(\Omega) < \infty$. That is, whenever the range of f is a finite subset of its codomain.

DEFINITION 9.3.7. Let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$. Given a subset $E \subset \Omega$, the *indicator function of E* is the simple function $1_E : \Omega \to \mathbb{F}$ defined by

$$1_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \text{otherwise} \end{cases}$$

REMARK 9.3.8. Without spending much time looking at the indicator function 1_E and the Dirac delta measure δ_{ω_0}, it might at first seem that we are roughly redefining the same function twice. This is not true! For the indicator function 1_E, the subset E is fixed while

the points $\omega \in \Omega$ are variable $1_E(\omega)$. For the Dirac delta function δ_{ω_0}, the point $\omega_0 \in \Omega$ stays fixed, while the measurable subsets E are variable $\delta_{\omega_0}(E)$. □

REMARK 9.3.9. Before continuing forward, it is worth noting the following. Whenever using the field $\mathbb{F}$ as a measure space, unless explicitly stated otherwise, it is typically assumed that we endow $\mathbb{F}$ with its Borel σ-algebra. (Remember that $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, and both are Euclidean spaces and have a natural metric defined (the standard distance), and therefore we can easily talk about the Borel σ-algebra of $\mathbb{F}$). We will make no exception here; if we do not explicitly mention a different σ-algebra on $\mathbb{F}$, then we assume that we are using the Borel σ-algebra $\mathcal{B}_{\mathbb{F}}$ on $\mathbb{F}$ resulting from the Euclidean structure on $\mathbb{F}$. □

THEOREM 9.3.10. *Suppose Ω is a set and $f : \Omega \to \mathbb{F}$ is a simple function. Then the following hold.*

(1) *There exists a unique finite partition $E_1, E_2, ..., E_n$ of Ω and distinct scalars $y_1, y_2, ..., y_n \in \mathbb{F}$ such that $f = \sum_{i=1}^{n} y_i 1_{E_i}$.*
(2) *If Ω is a measurable space, then a simple function $f : \Omega \to \mathbb{F}$ is $(\mathcal{F}/\mathcal{B}_{\mathbb{F}})$-measurable if and only if the partition sets E_i in (1) are measurable subsets of Ω for $1 \leq i \leq n$.*
(3) *If there exist any collection of measurable sets $F_1, ..., F_m$ and scalars $z_1, ..., z_m$ (not necessarily distinct) such that a function $f : \Omega \to \mathbb{F}$ has the form $f = \sum_{i=1}^{m} z_i 1_{F_i}$, then f is a measurable simple function.*

Proof. (1) Let $f : \Omega \to \mathbb{F}$ be simple and define the scalars $y_1, y_2, ..., y_n \in \mathbb{F}$ as the distinct elements in the image of f; i.e., $\{y_1, y_2, ..., y_n\} = f(\Omega)$. For each i, let $E_i = \{f = y_i\}$. By construction, it is clear that the E_i partition Ω and are unique and that $f = \sum_{i=1}^{n} y_i 1_{E_i}$.

(1) If $f = \sum_{i=1}^{n} y_i 1_{E_i}$ is a simple function represented as the sum from (1), then for any open set $U \subset \mathbb{F}$,

$$f^{-1}(U) = \bigcup_{y_i \in U} E_i.$$

Since the union on the righthand side is of finitely many of E_i and the open subset of $\mathbb{F}$ generate the Borel σ-algebra, it also follows that f is measurable if and only if each of the E_i is measurable.

(3) This follows a similar argument as (2) and is left as Exercise 4 (12). ■

NOTATION 9.3.11. When $(\Omega, \mathcal{F})$ and $(\mathcal{X}, \mathcal{B})$ are measurable spaces, we will define $\mathscr{S}((\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{B}))$ as the collection of all $(\mathcal{F}/\mathcal{B})$-measurable simple functions $f : \Omega \to \mathcal{X}$. Often, when

(one or both of) the σ-algebras $\mathcal{F}$ or $\mathcal{B}$ is fixed and implicitly understood, we will omit them from the notation, simply writing $\mathscr{S}(\Omega \to \mathcal{X})$. In particular, when $\mathcal{X} = \mathbb{F}$, we will implicitly assume that $\mathbb{F}$ is endowed with its Borel σ-algebra, and hence write $\mathscr{S}((\Omega, \mathcal{F}) \to \mathbb{F})$ or $\mathscr{S}(\Omega \to \mathbb{F})$. Further, we define $\mathscr{S}^+((\Omega, \mathcal{F}) \to \mathbb{R})$ (or $\mathscr{S}^+(\Omega \to \mathbb{R})$ when $\mathcal{F}$ is understood) as the real-valued simple functions $f : \Omega \to \mathbb{R}$ which are non-negative.

THEOREM 9.3.12. *The spaces $\mathscr{S}(\Omega \to \mathbb{F})$ and $\mathscr{S}^+(\Omega \to \mathbb{R})$ are vector spaces over $\mathbb{F}$ and $\mathbb{R}$, respectively. Moreover, both spaces are also closed under point-wise multiplication. Explicitly, if $f, g \in \mathscr{S}(\Omega \to \mathbb{F})$, then*

(1) *$f + \alpha\, g \in \mathscr{S}(\Omega \to \mathbb{F})$ for every $\alpha \in \mathbb{F}$.*
(2) *$fg \in \mathscr{S}(\Omega \to \mathbb{F})$ where, as usual, we define $(fg)(x) = f(x)g(x)$.*

The same can be said for $\mathscr{S}^+(\Omega \to \mathbb{R})$ as long as we exchange $\mathbb{F}$ with $\mathbb{R}$.

Proof. Let us first assume that $f, g \in \mathscr{S}(\Omega \to \mathbb{F})$ are given by $f = \sum_{i=1}^{n} y_i \, 1_{E_i}$ and $g = \beta \, 1_F$ where, by Theorem 9.3.10, the sets E_i and F are measurable. Then

$$f + \alpha g = \sum_{i=1}^{n} \left\{ y_i \, 1_{E_i \setminus F} + (y_i + \alpha\beta) \, 1_{E_i \cap F} \right\} + \alpha\beta \, 1_{F \setminus \cup_i E_i}$$

which is again a measurable simple function by Theorem 9.3.10. The more general case is now a simple matter of induction on the number of terms in the sum representing g. For multiplication, the point to notice that if A and B are sets, then the pointwise multiplication of their indicators is $1_A 1_B = 1_{A \cap B}$. Therefore, since finitely many intersections of measurable sets results in a measurable set, the conclusion follows. The details of this are left as Exercise 4 (13). ∎

DEFINITION 9.3.13. Let $f \in \mathscr{S}^+(\Omega \to \mathbb{R})$. We define the *integral* of f with respect to μ as the value in $[0, \infty]$ by

$$\int_\Omega f \, d\mu = \sum_{y \in f(\Omega)} y \cdot \mu(f = y) = \sum_{y \in f(\Omega)} y \cdot \mu\big(f^{-1}(\{y\})\big)$$

where we define the term $0 \cdot \mu(f = 0) = 0$, preemptively avoiding complications in the case $\mu(f = 0) = \infty$.

EXAMPLE 9.3.14. Consider a scenario where you flip a biased coin three times. The bias of the coin gives a probability of $p \in (0, 1)$ for flipping heads, and $q = 1 - p$ for

flipping tails. You model the possible outcomes of three flips of the coin as the set $\Omega = \{(x, y, z) \text{ s.t. } x, y, z \in \{0, 1\}\}$; here, an entry of 1 for x, y, or z represents the outcome of heads on the first, second, or third flip respectively; analogously, an entry of 0 represents the outcome of a tail on that respective flip. As Exercise 4 (**??**) suggests, since Ω contains only finitely many elements, it makes sense to define the σ-algebra on Ω as the full power set 2^{Ω}. From here, we define the probability measure μ defined on the singleton sets by

$$\mu(\{(x, y, z)\}) = p^{x+y+z} q^{3-(x+y+z)}$$

Note $x+y+z$ gives the number of heads flipped in the outcome (x, y, z), whereas $3-(x+y+z)$ yields the number of tails. Thus, the measure of μ is designed to match the probabilities of flipping heads and tails which our original scenario suggested. For example, $\mu(\{(1, 0, 0)\}) = p\, q^2$, which agrees with the naive probability we'd assign to the event of first flipping a heads, and then flipping two tails. It is also worth noting at this point that, again due to Exercise 4 (**??**), we have now completely defined μ since we have defined it on every singleton set within the finite set Ω.

Now, consider the measurable function $f : \Omega \to \mathbb{R}$ defined by $f(x, y, z) = x + y + z$; intuitively, for each $\omega \in \Omega$, the value $f(\omega)$ is the number of heads flipped in outcome ω. Note that the image of f is $f(\Omega) = \{0, 1, 2, 3\}$, and therefore $f \in \mathscr{S}^+(\Omega \to \mathbb{R})$. Therefore

$$\int_{\Omega} f \, d\mu = \sum_{k=0}^{3} k \, \mu(f = k) = \mu(f = 1) + 2\, \mu(f = 2) + 3\, \mu(f = 3).$$

To finish the calculation, we find

$$\mu(f = 1) = \mu(\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}) = \mu(\{(1, 0, 0)\} + \mu(\{(0, 1, 0)\}) + \mu(\{(0, 0, 1)\}) = 3\, p\, q^2,$$

$$\mu(f = 2) = \mu(\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}) = \mu(\{(1, 1, 0)\} + \mu(\{(1, 0, 1)\}) + \mu(\{(0, 1, 1)\}) = 3\, p^2\, q,$$

$$\mu(f = 3) = \mu(\{(1, 1, 1)\}) = p^3$$

and hence

$$\int_{\Omega} f \, d\mu = 3\, p\, q^2 + 6\, p^2\, q + 3\, p^3 = 3p$$

where the last equality was after minimal simplification after replacing q with the value $1 - p$.

It is of interest to note that, setting n, using standard counting notation, we found $\mu(f = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the choose function. Such a measurable

function f is typically defined as a *binomial random variable* with probability of success p with n trials. □

LEMMA 9.3.15. *"The integral is well defined on $\mathscr{S}^+(\Omega \to \mathbb{R})$." If $f \in \mathscr{S}^+(\Omega \to \mathbb{R})$ and $\{E_i\}_{i=1}^m$ is any partition of Ω comprised of measurable subsets such that $f = \sum_{i=1}^m y_i \, 1_{E_i}$, then*

$$\int_\Omega f \, d\mu = \sum_{i=1}^m y_i \, \mu(E_i).$$

Proof. Let $f(\Omega) = \{x_1, x_2, ..., x_n\} \subset \mathbb{R}$ and $F_j = \{f = x_j\} = f^{-1}(\{x_j\})$ for each $j = 1, 2, ..., n$. Since $\{E_i\}_{i=1}^m$ and $\{F_j\}_{j=1}^n$ are both partitions of Ω and when f is constantly equal to y_i when restricted to any E_i, we must have that given any i, E_i is a subset of exactly one of the F_j and disjoint from the rest. This means that for any i and j, either $E_i \cap F_j = \varnothing$ or $E_i \cap F_j = E_i$. Now consider the new partition (of which many of the subsets are likely $\varnothing$):

$$\{E_i \cap F_j \text{ s.t. } 1 \le i \le m, \ 1 \le j \le n\}.$$

We have $\cup_{i=1}^m (E_i \cap F_j) = F_j$ and $\cup_{j=1}^n (E_i \cap F_j) = E_i$. Using the additivity of μ, this implies

$$\sum_{i=1}^m \mu(E_i \cap F_j) = \mu(F_j) \ \text{ and } \ \sum_{j=1}^n \mu(E_i \cap F_j) = \mu(E_i).$$

From here, we also claim that $x_j \, \mu(E_i \cap F_j) = y_i \, \mu(E_i \cap F_j)$; this is true since either $E_i \cap F_j = \varnothing$, in which case both sides are 0, otherwise $E_i \subset F_j$ and f is constantly y_i on E_i and x_j on F_j, hence it must be that $x_i = y_j$. Therefore,

$$\begin{aligned}\int_\Omega f \, d\mu = \sum_{j=1}^n x_j \, \mu(F_j) = \sum_{j=1}^n x_j \sum_{i=1}^m \mu(E_i \cap F_j) = \sum_{i=1}^m \sum_{j=1}^n x_j \, \mu(E_i \cap F_j) = \sum_{i=1}^m \sum_{j=1}^n y_i \, \mu(E_i \cap F_j) \\ = \sum_{i=1}^m y_i \sum_{j=1}^n \mu(E_i \cap F_j) = \sum_{i=1}^m y_i \, \mu(E_i)\end{aligned}$$

This concludes the proof. ■

PROPOSITION 9.3.16. *The following properties of integration hold.*

(1) *If $f : \Omega \to \mathbb{R}$ with $f \equiv 0$, then $\int_\Omega f \, d\mu = 0$.*
(2) *For all $f, g \in \mathscr{S}^+(\Omega \to \mathbb{R})$ and for all $a \in [0, \infty)$,*

$$\int_\Omega (f + a\, g) \, d\mu = \int_\Omega f \, d\mu + a \int_\Omega g \, d\mu.$$

(3) *For all* $f, g \in \mathscr{S}^+(\Omega \to \mathbb{R})$ *such that* $f(\omega) \leq g(\omega)$ *for all* $\omega \in \Omega$, $\int_\Omega f\,d\mu \leq \int_\Omega g\,d\mu$.

Proof. (1) This is straight from the definition of the integral.

(2) Note that if $g \in \mathscr{S}^+(\Omega \to \mathbb{R})$, then so is $a\,g$ for any $a \in [0, \infty)$. Now, as a direct result of the distributive property on finite sums of real numbers. If $g = \sum_{j=1}^{n} x_j\,1_{\{g=x_j\}}$, then

$$\int ag\,d\mu = \sum_{j=1}^{n} ax_j\,\mu(g = x_j) = a\sum_{j=1}^{n} x_j\,\mu(g = x_j) = a\int_\Omega g\,d\mu.$$

It remains to show that $\int(f+g) = \int f + \int g$. Suppose that $f = \sum_{i=1}^{m} y_i\,1_{E_i}$ and $g = \sum_{j=1}^{n} x_j\,1_{F_j}$ are standard representations of f and g. Then the collection

$$\{E_i \cap F_j \text{ s.t. } 1 \leq i \leq m, 1 \leq j \leq n\}$$

is another partition of measurable subsets of Ω. We can then write

$$f = \sum_{i=1}^{m}\sum_{j=1}^{n} y_i\,1_{E_i\cap F_j} \quad \text{and} \quad g = \sum_{i=1}^{m}\sum_{j=1}^{n} x_j\,1_{E_i\cap F_j}.$$

From this, we find

$$f + g = \sum_{i=1}^{m}\sum_{j=1}^{n}(y_i + x_j)\,1_{E_i\cap F_j} \implies \int_\Omega (f+g)\,d\mu = \sum_{i=1}^{m}\sum_{j=1}^{n}(y_i + x_j)\mu(E_i \cap F_j)$$

Now, $\sum_{j=1}^{n}\mu(E_i\cap F_j) = \mu(E_i)$ since $E_i = \cup_{j=1}^{n}(E_i\cap F_j)$ where this union is disjoint. Similarly, $\sum_{i=1}^{m}\mu(E_i \cap F_j) = \mu(F_j)$. Therefore,

$$\begin{aligned}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_i + x_j)\mu(E_i \cap F_j) &= \sum_{i=1}^{m} x_i \sum_{j=1}^{n}\mu(E_i \cap F_j) + \sum_{j=1}^{n} y_j \sum_{i=1}^{m}\mu(E_i \cap F_j)\\ &= \sum_{i=1}^{m} x_i\,\mu(E_i) + \sum_{j=1}^{n} y_j\,\mu(F_j) = \int_\Omega f\,d\mu + \int_\Omega g\,d\mu.\end{aligned}$$

(3) (See Notation 9.3.17 below for some of the notation used in this argument). By the linear properties of the integral, $\int f \leq \int g$ if and only if $\int(g - f) \geq 0$. Therefore, we need to only show that if $g \geq 0$, then $\int g \geq 0$. However, this is immediate since the sum definite $\int g$ is of only non-negative terms. ∎

3.3. Non-negative Measurable Functions.

NOTATION 9.3.17. If f, g are two functions $\Omega \to \mathbb{R}$ (or $\overline{\mathbb{R}}$) such that $f(\omega) \leq g(\omega)$ for every $\omega \in \Omega$, we will write $f \leq g$. Similarly, $f < g$, $f > g$, and $f \geq g$ are to be interpreted as pointwise true statements.

THEOREM 9.3.18. *Let $f : \Omega \to \overline{\mathbb{R}}$ be non-negative and $(\mathcal{F}/\mathcal{B}_{\overline{\mathbb{R}}})$-measurable. Then there exists a sequence of functions $\{f_n\}_{n=1}^{\infty} \subset \mathscr{S}^+(\Omega \to \mathbb{R})$ such that for every $n \in \mathbb{N}$, $f_n \leq f_{n+1}$, and $f_n \to f$ pointwise.*

Proof. For each $n \in \mathbb{N}$, define

$$f_n = 2^n \cdot 1_{\{2^n < f\}} + \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} \cdot 1_{\left\{\frac{k}{2^n} < f \leq \frac{k+1}{2^n}\right\}}$$

This sequence $\{f_n\}_{n=1}^{\infty}$ will do the trick. ■

NOTATION 9.3.19. We let $L^+(\Omega, \mathcal{F}, \mu)$ be the collection of $(\mathcal{F}/\mathcal{B}_{\overline{\mathbb{R}}})$-measurable functions $f : \Omega \to \overline{\mathbb{R}}$ which are non-negative μ-a.s. We will often simplify this notation by writing $L^+(\mu)$ when Ω and $\mathcal{F}$ are understood.

PROPOSITION 9.3.20. *The space $L^+(\Omega, \mathcal{F}, \mu)$ is closed under addition, multiplication by non-negative scalars, and pointwise multiplication.*

Proof. This proof is left as an exercise. ■

DEFINITION 9.3.21. Let $f \in L^+(\mu)$. The *Lebegue integral* of f with respect to μ is

$$\int_\Omega f \, d\mu = \sup \left\{ \int_\Omega \varphi \, d\mu \text{ s.t. } \varphi \in \mathscr{S}^+(\Omega \to \mathbb{R}) \text{ and } \varphi \leq f \right\} \tag{3.1}$$

THEOREM 9.3.22 (Monotone Convergence Theorem). *Let $f : \Omega \to \overline{\mathbb{R}}$. Suppose that $\{f_n\}_{n=1}^{\infty} \subset L^+(\mu)$ is a sequence such that for all $n \geq 1$, $f_n \leq f_{n+1}$, and that $f_n \to f$ pointwise. Then $f \in L^+(\mu)$ and*

$$\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega f_n \, d\mu.$$

COROLLARY 9.3.23. *Given any $f \in L^+(\mu)$ there exists a sequence $\{f_n\}_{n=1}^{\infty} \subset \mathscr{S}^+(\Omega \to \mathbb{R})$ such that $f_n \leq f_{n+1}$ for every n, $f_n \to f$ pointwise, and*

$$\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega f_n \, d\mu.$$

Proof. Combine the Monotone Convergence Theorem with Theorem 9.3.18. ■

COROLLARY 9.3.24. *The following properties of integration on $L^+(\mu)$ hold.*

(1) *For all $f, g \in L^+(\mu)$ and for all $a \in [0, \infty)$, it holds that*

$$\int_\Omega (f + ag)\, d\mu = \int_\Omega f\, d\mu + a \int_\Omega g\, d\mu.$$

(2) *For all $f, g \in L^+(\mu)$ such that $f \leq g$ for all $\omega \in \Omega$, $\int_\Omega f\, d\mu \leq \int_\Omega g\, d\mu$.*

Proof. (1) We give a thorough outline of the proof here; the technical details are easily filled in by those interested. If $f, g \in L^+(\mu)$ and $a \in [0, \infty)$, then $f + ag \in L^+(\mu)$. Moreover, by Corollary 9.3.23, we can find monotonically increasing sequences of simple functions $\{f_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$ such that $f_n \to f$, $g_n \to g$, $\int f_n \to \int f$, and $\int g_n \to \int g$. Using Proposition 9.3.16

$$\int (f_n + ag_n) = \int f_n + a \int g_n$$

for every n, and hence $\int (f_n + ag_n) \to \int f + a \int g$. However, $\{f_n + ag_n\}_{n=1}^\infty$ is also a monotonically increasing sequence of simple functions which converges to $f + ag$ pointwise. Therefore, by the Monotone Convergence Theorem

$$\int (f_n + ag_n) \to \int (f + ag).$$

This shows that $\int (f + ag) = \int f + a \int g$ and finishes the proof.

(2) Given that for any $\varphi \in \mathscr{S}^+(\Omega \to \mathbb{R})$ such that $\varphi \leq f$, we also have $\varphi \leq g$ since $f \leq g$. Therefore,

$$\left\{ \int_\Omega \varphi\, d\mu \text{ s.t. } \varphi \in \mathscr{S}^+(\Omega \to \mathbb{R}) \text{ and } \varphi \leq f \right\} \subset \left\{ \int_\Omega \varphi\, d\mu \text{ s.t. } \varphi \in \mathscr{S}^+(\Omega \to \mathbb{R}) \text{ and } \varphi \leq g \right\}$$

Taking the supremum of both sets, this claim now follows by the definition of the integral and properties of the supremum. ∎

COROLLARY 9.3.25. *Let $N \in \mathbb{N} \cup \{\infty\}$. Assume that $\{f_n\}_{n=1}^N \subset L^+(\mu)$ and $\sum_{n=1}^N f_n$ is a pointwise convergent series such that $\sum_{n=1}^N f_n \in L^+(\mu)$ (if $N < \infty$, then the assumption of pointwise convergence and belonging to $L^+(\mu)$ is vacuous). Then*

$$\int \sum_{n=1}^N f_n\, d\mu = \sum_{n=1}^N \int f_n\, d\mu.$$

Proof. If $N < \infty$, then this is simply a consequence of Corollary 9.3.24 (1). So, we assume $N = \infty$. The partial sums $F_m = \sum_{n=1}^{m} f_n$ are a monotonically increasing sequence of non-negative measurable functions. Using Corollary 9.3.24 (1), we have that

$$\int F_m = \sum_{n=1}^{m} \int f_n.$$

By the Monotone Convergence Theorem $\lim_{m\to\infty} \int F_m = \int \lim_{m\to\infty} F_m$ and thus,

$$\sum_{n=1}^{\infty} \int f_n = \lim_{m\to\infty} \sum_{n=1}^{m} \int f_n = \lim_{m\to\infty} \int F_m = \int \lim_{m\to\infty} F_m = \int \sum_{n=1}^{\infty} f_n.$$

This concludes the proof. ∎

3.4. Integration of Measurable Functions. We continue to assume that $(\Omega, \mathcal{F}, \mu)$ is a measure space.

Definition 9.3.26. Let $f : \Omega \to \mathbb{R}$ be a function. Define the *positive portion* of f as the non-negative function $f_+ : \Omega \to \mathbb{R}$ by $f_+(\omega) = \max\{f(\omega), 0\}$. Define the *negative portion* of f as the non-negative function $f_- : \Omega \to \mathbb{R}$ by $f_-(\omega) = \max\{-f(\omega), 0\}$.

Proposition 9.3.27. *Let $f : \Omega \to \mathbb{R}$. The following hold.*

(1) $f = f_+ - f_-$.
(2) $|f| = f_+ + f_-$. *In particular,* $f_+ \leq |f|$ *and* $f_- \leq |f|$.
(3) *If f is $(\mathcal{F}/\mathcal{B}_{\overline{\mathbb{R}}})$- measurable, then $f_+, f_- \in L^+(\Omega, \mathcal{F}, \mu)$.*

Definition 9.3.28. Let $f : \Omega \to \overline{\mathbb{R}}$ be measurable. We define the *integral* of f, denote $\int_\Omega f \, d\mu$, as

$$\int_\Omega f \, d\mu = \int_\Omega f_+ \, d\mu - \int_\Omega f_- \, d\mu$$

whenever the righthand side is defined in $\overline{\mathbb{R}}$. Equivalently, we define the integral of f by the above equality whenever either $\int f_+ < \infty$ or $\int f_- < \infty$.

Definition 9.3.29. For every $p \in \mathbb{R}$ with $p \geq 1$, we define the space $\mathscr{L}^p(\Omega, \mathcal{F}, \mu)$, or simply $\mathscr{L}^p(\mu)$, as the space of $(\mathcal{F}, \mathcal{B}_{\overline{\mathbb{R}}})$–measurable functions $f : \Omega \to \overline{\mathbb{R}}$ such that $\int_\Omega |f|^p \, d\mu < \infty$. Any function $f \in \mathscr{L}^1(\mu)$ is called *integrable*.

Remark 9.3.30. If $f : \Omega \to \overline{\mathbb{R}}$ is measurable, then so is $|f|^p$, since the function $|\cdot|^p : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ is a continuous function, and hence measurable. Thus $|f|^p$ is the composition of the measurable functions $|\cdot|^p$ and f, which is again measurable. □

THEOREM 9.3.31. *The space $\mathscr{L}^1(\mu)$ is a vector space over $\mathbb{R}$. Moreover, for every $f, g \in \mathscr{L}^1(\mu)$ and $a \in \mathbb{R}$, it holds that*

(1) $\left| \int_\Omega f \, d\mu \right| \leq \int_\Omega |f| \, d\mu$.

(2) $\int_\Omega (f + ag) d\mu = \int_\Omega f \, d\mu + a \int_\Omega g \, d\mu$.

(3) *If $f \leq g$, then* $\int_\Omega f \, d\mu \leq \int_\Omega g \, d\mu$.

Proof. By the triangle inequality of the absolute value on $\mathbb{R}$, we have that $|f + ag| \leq |f| + |a|\,|g|$ for any $f, g \in \mathscr{L}^1(\mu)$ and $a \in \mathbb{R}$. Hence, by the properties of integration on $L^+(\mu)$, we have that $\int |f + ag| \leq \int (|f| + |a|\,|g|) = \int |f| + |a| \int |g| < \infty$, where the finiteness of righthand is by the assumption that $f, g \in \mathscr{L}^1(\mu)$. From this it is easy to deduce that $\mathscr{L}^1(\mu)$ is a vector space over $\mathbb{R}$.

(1) We have

$$\left| \int f \right| = \left| \int f_+ - \int f_- \right| \leq \left| \int f_+ \right| + \left| \int f_- \right| = \int f_+ + \int f_- = \int (f_+ + f_-) = \int |f|.$$

(2) Assume first that $a \geq 0$. Then $(f + ag)_+ = (f_+ + ag_+)$ and $(f + ag)_- = (f_- + ag_-$. Therefore, using properties of the integrals on $L^+(\mu)$,

$$\begin{aligned} \int (f + ag) = \int (f + ag)_+ - \int (f + ag)_- &= \Big(\int f_+ + a \int g_+ \Big) - \Big(\int f_- + a \int g_- \Big) \\ &= \Big(\int f_+ - \int f_- \Big) + a \Big(\int g_+ - \int g_- \Big) = \int f + a \int g. \end{aligned}$$

(3) Again, $f \leq g$ if and only if $g - f \geq 0$. So, if $f, g \in \mathscr{L}^1(\mu)$ with $f \leq g$, then $g - f \in L^+(\mu)$. Therefore $0 \leq \int (g - f) = \int g - \int f$. The last equality came from (2). This finishes the proof. ■

REMARK 9.3.32. Theorem 9.3.31 proves that the integral $\int$ is a linear function on vector space $\mathscr{L}^1(\mu)$. That is, $\int : \mathscr{L}^1(\mu) \to \mathbb{R}$ is a linear function. □

THEOREM 9.3.33. *For each $p \geq 1$, define the function $\| \cdot \|_p : \mathscr{L}^p(\mu) \to \mathbb{R}$ by*

$$\|f\|_p = \Big(\int_\Omega |f|^p \, d\mu \Big)^{1/p}.$$

The following properties hold.

(1) *For every $f \in \mathscr{L}^p(\mu)$, $\|f\|_p \geq 0$.*

(2) *For every $a \in \mathbb{R}$ and $f \in \mathscr{L}^p(\mu)$, $\|af\|_p = |a|\,\|f\|_p$.*

(3) *For every* $f, g \in \mathscr{L}^p(\mu)$, $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.

However, if there exists a non-empty measurable set N *such that* $\mu(N) = 0$, *then there exists a function* $f \in \mathscr{L}^p(\mu)$ *such that* $\|f\|_p = 0$, *but* $f \neq 0$, *showing that* $\|\cdot\|_p$ *is not a true norm on* $\mathscr{L}^p$.

Proof. (1) and (2) are immediate from the definition and the linear properties of the integral.

(3)

Now, suppose that $N \subset \Omega$ is measurable and non-empty. Let $f = 1_N$ be the indicator function of N. Then $f(\omega) = 1$ for every $\omega \in N$, and hence $f \neq 0$. However, $\int |f|^p = \int |1_N|^p = \int 1_N = 1 \cdot \mu(N) = 0$. Therefore $\|f\|_p = 0$. ∎

Definition 9.3.34. The function $\|\cdot\|_p$ on $\mathscr{L}^p(\mu)$ is called the *p-norm*.

Remarks 9.3.35.

(1) Although, Theorem 9.3.33 shows that $\|\cdot\|_p$ is not a true norm on $\mathscr{L}^p(\mu)$, we will make an adjustment to the space $\mathscr{L}^p(\mu)$ below which forces $\|\cdot\|_p$ to be a true norm, and therefore then name is not a misnomer.

(2) It is also worth noting that $f \in \mathscr{L}^p(\mu)$ if and only if $f : \Omega \to \overline{\mathbb{R}}$ is measurable and $\|f\|_p < \infty$.

Theorem 9.3.36 (Hölder's Inequality). *Suppose* $p, q \in (1, \infty)$ *are such that* $\frac{1}{p} + \frac{1}{q} = 1$. *For any* $f \in \mathscr{L}^p(\mu)$ *and* $g \in \mathscr{L}^q(\mu)$, *the pointwise product* fg *is in* $\mathscr{L}^1(\mu)$. *In fact,*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

3.5. Restricted σ-algebras.

Definition 9.3.37. Let $\Omega_0 \subset \Omega$ and define $\mathcal{F}_0 = \{E \cap \Omega_0 \text{ s.t. } E \in \mathcal{F}\}$. The $\mathcal{F}_0$ is called the *restriction* of $\mathcal{F}$ to Ω_0.

Theorem 9.3.38. *Let* $\mathcal{F}_0$ *be the restriction of* $\mathcal{F}$ *to* Ω_0. *Then* $\mathcal{F}_0$ *is a* σ*-algebra on* Ω_0. *Moreover, if* Ω_0 *is measurable, then* $\mathcal{F}_0 \subset \mathcal{F}$ *and* μ *is a measure on* $\mathcal{F}_0$.

Proof. Since $\varnothing \cap \Omega_0 = \varnothing$, $\varnothing \in \mathcal{F}_0$. Moreover, since $(\Omega \setminus E) \cap \Omega_0 = \Omega_0 \setminus (\Omega_0 \cap E)$ and $(\cup_n E_n) \cap \Omega_0 = \cup_n (E_n \cap \Omega_0)$, the properties of a σ-algebra easily follow. If Ω_0 is measurable, then $E \cap \Omega_0 \in \mathcal{F}$ for any $E \in \mathcal{F}$, which shows that $\mathcal{F}_0 \subset \mathcal{F}$; since μ is a measure on $\mathcal{F}$, all

the properties of μ defining it as a measure transcend to $\mathcal{F}_0$, and is therefore a measure on $\mathcal{F}_0$. ■

PROPOSITION 9.3.39. *Let $f : \Omega \to \overline{\mathbb{R}}$ and define $\Omega_0 = f^{-1}(\mathbb{R})$. Then f is $(\mathcal{F}/\mathcal{B}_{\overline{\mathbb{R}}})$-measurable if and only if $f^{-1}(\{-\infty\}), f^{-1}(\{\infty\}) \in \mathcal{F}$ and the restriction $f : \Omega_0 \to \mathbb{R}$ is $(\mathcal{F}_0/\mathcal{B}_{\mathbb{R}})$-measurable, where $\mathcal{F}_0$ is the restriction of $\mathcal{F}$ to Ω_0.*

Proof. If f is $(\mathcal{F}/\mathcal{B}_{\overline{\mathbb{R}}})$-measurable, then $f^{-1}(\{-\infty\}), f^{-1}(\{\infty\}) \in \mathcal{F}$ because both $\{-\infty\}$ and $\{\infty\}$ are in $\mathcal{B}_{\overline{\mathbb{R}}}$. Further, since every set $E \in \mathcal{B}_{\mathbb{R}}$ is in $\mathcal{B}_{\overline{\mathbb{R}}}$, this implies that $f^{-1}(E) \in \mathcal{F}$. We need to show that $f^{-1}(E)$ is, in fact, in $\mathcal{F}_0$ to conclude that $f : \Omega_0 \to \mathbb{R}$ is measurable. Since $f^{-1}(E) \cap f^{-1}(\mathbb{R}) = f^{-1}(E)$ since $E \subset \mathbb{R}$ by assumption, and $\mathcal{F}_0 = \big\{F \cap f^{-1}(\mathbb{R}) \text{ s.t. } F \in \mathcal{F}\big\}$, we see now that $E \in \mathcal{F}_0$.

Let's now assume that $f^{-1}(\{-\infty\}), f^{-1}(\{\infty\}) \in \mathcal{F}$, and the restriction $f : \Omega_0 \to \mathbb{R}$ is measurable. By Proposition 9.2.29 and Theorem 9.3.4, it suffices to show that $f^{-1}((a,\infty]) \in \mathcal{F}$ for any $a \in \mathbb{R}$. To do so, we have that $(a,\infty] = (a,\infty) \cup \{\infty\}$, so $f^{-1}((a,\infty]) = f^{-1}\big((a,\infty)\big) \cup f^{-1}(\{\infty\})$. By assumption $f^{-1}(\{\infty\}) \in \mathcal{F}$, hence it remains only to show that $f^{-1}((a,\infty)) \in \mathcal{F}$ (since $\mathcal{F}$ is closed under unions). We have that $(a,\infty) \in \mathcal{B}_{\mathbb{R}}$, so $f^{-1}((a,\infty)) \in \mathcal{F}_0$ by assumption of measurability of the restriction; since $\mathcal{F}_0 \subset \mathcal{F}$ by Theorem 9.3.38 (noticing that $\Omega_0 \in \mathcal{F}$), we can now easily conclude this direction. ■

3.6. Null Sets and Almost Sure Statements. Suppose that $(\Omega, \mathcal{F}, \mu)$ is a measure space.

DEFINITION 9.3.40. A set $N \in \mathcal{F}$ is called a *μ-null set* (or simply *μ-null*) when $\mu(N) = 0$.

LEMMA 9.3.41. *Let N be a μ-null set. Then,*

1) If $F \subset N$ is in $\mathcal{F}$, then $\mu(F) = 0$.

2) If $E \in \mathcal{F}$, then $\mu(E \cup N) = \mu(E \backslash N) = \mu(E)$.

Proof. 1) This follows immediately from the inequality $0 \leq \mu(F) \leq \mu(N) = 0$.

2) We have $E \cup N = \big(E \backslash N\big) \cup (E \cap N) \cup \big(N \backslash E\big)$. Both $E \cap N$ and $N \backslash E$ are measurable subsets of N, so by additivity and 1),

$$\mu(E \cup N) = \mu(E \backslash N) + \mu(E \cap N) + \mu(N \backslash E) = \mu(E \backslash N).$$

However, $E \backslash N \subset E \subset E \cup N$, so $\mu(E \backslash N) \leq \mu(E) \leq \mu(E \cup N)$. From here, the conclusion easily follows. ■

Let's assume for the sake of discussion that μ is a finite measure; i.e. $\mu(\Omega) < \infty$. Suppose that we want to talk about some property P holding true for certain elements in our space Ω. If we want to know how likely it is the that property P holds for an element in Ω, then a reasonable way to do this (in fact, the only way we have to measure this without knowing more structure on Ω) is to consider $\mu(E)/\mu(\Omega)$ where $E = \{\omega \in \Omega \text{ s.t. property } P \text{ holds for } \omega\}$ (we are, of course, assuming that E is measurable here); the expression $\mu(E)/\mu(\Omega)$ gives the (μ-weighted) proportion of elements in Ω satisfying property P. We see that we'd get $\mu(E)/\mu(\Omega) = 1$ whenever $\mu(E) = \mu(\Omega)$. Intuitively, this says that if we were to randomly select an element $\omega \in \Omega$ where our selection is weighted by the measure μ, there is a 100% chance that this randomly chosen element satisfies property P whenever $\mu(E) = \mu(\Omega)$. However, this does not necessarily mean that $E = \Omega$. Instead, what this means is that the set $N = \Omega \backslash E$ is a μ-null set (since $\mu(\Omega) = \mu(E) + \mu(N)$). This technicality, that $\mu(E) = \mu(\Omega)$ and $E \subsetneq \Omega$, requires some addition to our vocabulary.

Definition 9.3.42 (Almost Sure Paradigm). Let P be some property such that we can check if each $\omega \in \Omega$ satisfies P (you could take $P : \Omega \to \{0, 1\}$ where $P(\omega) = 1$ means ω satisfies the property). If there exists a μ-null set N such that

$$\{\omega \in \Omega \text{ s.t. } \omega \text{ does not satisfy property } P\} \subset N,$$

then we say that property P holds *μ-almost surely*, abbreviated μ-a.s.

This definition holds for any measure μ, not just finite measures. You might ask why did we not make the definition to be something of the form: let $E = \{\omega \in \Omega : \omega \text{ satisfies } P\}$, if $\mu(E) = \mu(\Omega)$ then we say that P holds μ-a.s. (instead of requiring the existence of a μ-null set N such that $\Omega \setminus E \subset N$)? There are a couple reasons. The first reason is that these two statements don't necessarily agree if μ is not finite. In the case that $\mu(\Omega) = \infty$, we can have that $\mu(E) = \infty$ and $\mu(\Omega) = \infty$, but $\mu(\Omega \backslash E) \neq 0$; we don't want to consider E to be the same as Ω almost surely if $\mu(\Omega \backslash E) \neq 0$, since the elements in $\Omega \backslash E$ (those not satisfying property P) have a non-zero measure. The second more technical reason is that it could so happen that E is not measurable (and hence $\Omega \setminus E$ is also not measurable). Therefore, it might not make any sense to even write the expression $\mu(E)$ or $\mu(E^c)$; by first requiring the existence of the μ-null set N (which is measurable by definition), we can get a scope of the size of E without requiring E itself to be measurable.

We can also compare two measurable sets this way.

DEFINITION 9.3.43. Let $E, F \in 2^\Omega$. We say that $E = F$ μ-almost surely when there exists a μ-null set N such that $E\Delta F \subset N$; equivalently, $E \cap F \subset N^c$. Here $E\Delta F = (E\backslash F) \cup (F\backslash E) = (E \cup F)\backslash(E \cap F)$, the collection of all elements in E or F which they do not share; $E\Delta F$ is called the *symmetric difference* of E and F.

Here are some examples of common almost sure statements regarding functions.

DEFINITIONS 9.3.44. 1) Suppose that Y is some set and $A \subset Y$. If $f : \Omega \to Y$, we say that f is in A μ-almost surely when there is a μ-null set N such that $\{\omega \in \Omega \text{ s.t. } f(\omega) \notin A\} \subset N$. We denote this by $f \in A$ μ-a.s.
2) Suppose that Y is some set. If $f : \Omega \to Y$ and $g : \Omega \to Y$, then we say that f equals g μ-almost surely when there is a μ-null set N such that $\{\omega \in \Omega \text{ s.t. } f(\omega) \neq g(\omega)\} \subset N$. We denote this by $f = g$ μ-a.s.
3) Suppose that Y is a metric space and $f : \Omega \to Y$. For every $n \in \mathbb{N}$, suppose that $f_n : \Omega \to Y$. We say that the sequence $\{f_n\}_{n=1}^\infty$ converges to f μ-almost surely when there exists a μ-null set N such that $\left\{\omega \in \Omega \text{ s.t. } \lim_{n\to\infty} f_n(\omega) \neq f(\omega)\right\} \subset N$. We denote this by $f_n \to f$ μ-a.s.

3.7. Complete σ-algebras.

DEFINITION 9.3.45. Given $(\Omega, \mathcal{F}, \mu)$, we say that μ is complete with respect to $\mathcal{F}$ when: Given any μ-null set N, then every subset $U \subset N$ is in $\mathcal{F}$. That is, μ is complete whenever $N \in \mathcal{F}$ and $\mu(N) = 0$ implies $2^N \subset \mathcal{F}$.

LEMMA 9.3.46. *Let $\{N_n\}_{n=1}^\infty \subset \mathcal{F}$ be a collection of μ-null sets. That is, $\mu(N_n) = 0$ for all n. Then $\mu(\cup_{n=1}^\infty N_n) = 0$; i.e., the countable union of μ-null sets is a μ-null set.*

Proof. This proof is an exercise (see Exercises **??** below). ■

THEOREM 9.3.47. *Given the measure space $(\Omega, \mathcal{F}, \mu)$, there is a σ-algebra $\overline{\mathcal{F}}$ and a measure $\overline{\mu}$, which is complete with respect to $\overline{\mathcal{F}}$ such that $\mathcal{F} \subset \overline{\mathcal{F}}$ and $\mu(E) = \overline{\mu}(E)$ for every $E \in \mathcal{F}$. Moreover, we can construct $\overline{\mathcal{F}}$ and $\overline{\mu}$ in the following way:*

$$\overline{\mathcal{F}} = \{E \cup U \text{ s.t. } E \in \mathcal{F} \text{ and } U \text{ is a subset of a } \mu\text{-null set}\} \text{ and } \overline{\mu}(E \cup U) = \mu(E).$$

Proof. We start by showing $\overline{\mathcal{F}}$ is a σ-algebra. To begin with, $\varnothing \in \mathcal{F}$ is a μ-null set, so $\varnothing \cup \varnothing = \varnothing \in \overline{\mathcal{F}}$. Now, assume that $E \in \mathcal{F}$ and U is a subset of a μ-null set, we want to show that $(E \cup U)^c \in \overline{\mathcal{F}}$. We start with some assistance from D'Morgan: $(E \cup U)^c = E^c \cap U^c$. If N is a μ-null set such that $U \subset N$ (such an N must exist by the definition of U), then we have that $N^c \subset U^c$ and therefore $U^c \cap N^c = N^c$. Hence $E^c \cap U^c = E^c \cap (U^c \cap (N^c \cup N)) = (E^c \cap (U^c \cap N^c)) \cup (E^c \cap U^c \cap N) = (E^c \cap N^c) \cup (E^c \cap U^c \cap N) = \tilde{E} \cup \tilde{U}$, where $\tilde{E} = E^c \cap N^c$ which is in $\mathcal{F}$ as E and N are in $\mathcal{F}$, and $\tilde{U} = E^c \cap U^c \cap N$ is a subset of the μ-null set N. This means that $(E \cup U)^c = \tilde{E} \cup \tilde{U} \in \overline{\mathcal{F}}$, hence $\overline{\mathcal{F}}$ is closed under complements. Finally, if $\{(E_n \cup U_n)\}_{n=1}^{\infty}$ is a countable collection of sets in $\overline{\mathcal{F}}$ with $E_n \in \mathcal{F}$ and U_n a subset of a μ-null set for every n, then $\cup_{n=1}^{\infty}(E_n \cup U_n) = (\cup_{n=1}^{\infty} E_n) \cup (\cup_{n=1}^{\infty} U_n)$. Since $\cup_n E_n \in \mathcal{F}$ and $\cup_n U_n$ is a subset of a μ-null set by Lemma 9.3.46, we again see that $\cup_n (E_n \cup U_n) \in \overline{\mathcal{F}}$, proving that $\overline{\mathcal{F}}$ is closed under countable unions, and thusly a σ-algebra. ■

DEFINITION 9.3.48. The σ-algebra $\overline{\mathcal{F}}$ is called the *completion* of $\mathcal{F}$ with respect to μ, and $\overline{\mu}$ is called the *completion* of μ.

The reason we care about complete measures is loosely stated as, "we don't want to worry about what happens on null sets." There are many arguments where we have functions which are equal μ-a.s., or two sets are equal μ-a.s., etc., but we don't want to care (or can't know!) what happens on the μ-null set where they might disagree. Completeness often let's us ignore these problems with relative ease. We conclude this section with useful results that give examples of such arguments.

LEMMA 9.3.49. *Assume that μ is complete with respect to $\mathcal{F}$. Let $E \in \mathcal{F}$ and N be a μ-null set. If $U \subset N$, then $E \cup U \in \mathcal{F}$, $E \cap U \in \mathcal{F}$, and $E \backslash U \in \mathcal{F}$.*

Proof. By the assumption of completeness of μ, we know that $U \in \mathcal{F}$. Therefore, the result follows since $\mathcal{F}$ is closed under unions, complements, and intersections. (Remember that $E \backslash U = E \cap U^c$). ■

PROPOSITION 9.3.50. *Assume that μ is complete with respect to $\mathcal{F}$. Let $E \in \mathcal{F}$ and $F \subset \Omega$. If $E = F$ μ-a.s., then $F \in \mathcal{F}$.*

Proof. Since $E = F$ μ-a.s., there is some μ-null set N such that $E \Delta F \subset N$ (recall that $E \Delta F = (E \setminus F) \cup (F \setminus E) = (E \cup F) \setminus (E \cap F)$ is the symmetric difference of the sets E and

F). This implies that $E \cap N^c = F \cap N^c$; in particular, since $E \cap N^c$ is in $\mathcal{F}$, then $F \cap N^c$ must be also. Since $F \cap N \subset N$, and μ is assumed complete with respect to $\mathcal{F}$, then $F \cap N$ is in $\mathcal{F}$. Therefore, $(F \cap N^c) \cup (F \cap N) \in \mathcal{F}$, since $\mathcal{F}$ is closed under intersections. We are done since $F = (F \cap N^c) \cup (F \cap N)$. ■

PROPOSITION 9.3.51. *Suppose that μ is complete with respect to $\mathcal{F}$ and $f : \Omega \to \overline{\mathbb{R}}$ is a measurable function. If $g : \Omega \to \overline{\mathbb{R}}$ is a function such that $f = g$ μ-a.s., then g is measurable.*

Proof. Let us start with an easier case, and work up to the general. Suppose that $h = 0$ μ-a.e. This means that there is some μ-null set N, such that $\{\omega \in \Omega \text{ s.t. } h(\omega) \neq 0\} \subset N$. Let $E = N^c$ and $U = \{\omega \in \Omega \text{ s.t. } h(\omega) = 0\} \cap N$. Then $h^{-1}(\{0\}) = E \cup U$, which is in $\mathcal{F}$ since E is the complement of a measurable set, and the completeness of μ implies that U is in $\mathcal{F}$; it is a subset of N, which is a μ-null set. From here, if $F \in \mathcal{B}_{\mathbb{R}}$, then $h^{-1}(F)$ is a subset of N if $0 \notin F$, and hence in $\mathcal{F}$ as a subset of a μ-null set. On the other hand, if $0 \in F$, then $h^{-1}(F) = h^{-1}(F \setminus \{0\}) \cup h^{-1}(\{0\})$, which is in $\mathcal{F}$ as the union of measurable sets. Therefore, h is measurable.

In the general case, assume that $f = g$ μ-a.s. and f is measurable. Then set $h = f - g$, in which case $h = 0$ μ-a.s., and hence measurable by what we have thus far proved. We are done now after invoking the result of Proposition **??** below to conclude that since h and f are measurable, so is $h + f$, which equals g. ■

4. Exercises

(1) Prove that for any non-empty set Ω, both $\{\varnothing, \Omega\}$ and 2^{Ω} are σ-algebras on Ω.
(2) Let Ω be a non-empty set and $A \subset \Omega$. Describe $\sigma(\{A\})$.
(3) Prove Proposition 9.2.5
(4) Prove Lemma 9.2.8.
(5) (a) Prove Corollary 9.2.13 (3).
    (b) Prove Corollary 9.2.13 (4).
    (c) Prove Corollary 9.2.13 (5).
(6) Prove that if Ω is a metric space and $\mathcal{B}_{\Omega}$ is the Borel σ-algebra on Ω, then all singleton sets are in $\mathcal{B}_{\Omega}$; that is, for all $\omega \in \Omega$, $\{\omega\} \in \mathcal{B}_{\Omega}$.

(7) Suppose that $\mathcal{A}$ is an algebra on Ω and that $\mu_1, \mu_2, ..., \mu_n$ are all pre-measures on $\mathcal{A}$. Let $\lambda_1, ..., \lambda_n$ all be non-negative real numbers. Show that $\mu = \sum_{i=1}^{n} \lambda_i \mu_i$ is a pre-measure on $\mathcal{A}$. Note that we are defining μ as $\mu(A) = \sum_{i=1}^{n} \lambda_i \mu_i(A)$. Also note that once you've proved this, it must also be true when assuming $\mathcal{A}$ with any σ-algebra and the μ_i are measures, since every σ-algebra is an algebra and every measure is a pre-measure.

(8) Suppose that $\mathcal{A}$ is an algebra on Ω and μ is a finitely additive measure on Ω such that $\mu(\Omega) < \infty$. For every $E \in \mathcal{A}$, prove that $\mu(E) = \mu(\Omega) - \mu(E^c)$. In particular, use this to deduce that if $\mathcal{A}$ is a σ-algebra and μ is a probability measure, then $\mu(E) = 1 - \mu(E^c)$ for every $E \in \mathcal{A}$.

(9) Suppose that $(\Omega, \mathcal{F})$ is a measurable space. Let $W \subset \Omega$ be a non-empty subset and define $\mathcal{F}_W = \{E \cap W \text{ s.t. } E \in \mathcal{F}\}$. Prove that $\mathcal{F}_W$ is a σ-algebra on W. We say that $\mathcal{F}_W$ is the σ-algebra $\mathcal{F}$ *restricted* to W.

(10) Let Ω be a non-empty finite or countably infinite set. Unless explicitly told otherwise, it is implicitly understood that the σ-algebra on Ω is the power set 2^Ω. Here we prove some of the results which motivate this default assumption.

   (a) Prove that if $\mathcal{F}$ is a σ-algebra on Ω containing all singleton sets (i.e., for every $\omega \in \Omega$, it holds that $\{\omega\} \in \mathcal{F}$) then $\mathcal{F} = 2^\Omega$.

   (b) Use Exercises (6) and (9) to deduce that if Ω is a subset of a metric space, then the Borel σ-algebra restricted to Ω is the power set 2^Ω.

   (c) Prove that if μ is a measure on $(\Omega, 2^\Omega)$, then μ is completely defined by its action on singleton sets. That is, prove that if $\{c_\omega\}_{\omega \in \Omega}$ are any values in $[0, \infty)$, then there is one unique measure μ on $(\Omega, 2^\Omega)$ such that $\mu(\{\omega\}) = c_\omega$. *Hint*: For any $E \subset \Omega$, you have that $E = \bigcup_{\omega \in E} \{\omega\}$, which is a finite or countable disjoint union. Therefore any measure must have the property $\mu(E) = \sum_{\omega \in E} \mu(\{\omega\})$ by definition.

(11) Suppose that $f : \Omega \to \mathcal{X}$ is a simple function between the measurable spaces $(\Omega, \mathcal{F})$ and $(\mathcal{X}, \mathcal{B})$. Suppose further that $\mathcal{B}$ contains all singleton sets; i.e., $\{x\} \in \mathcal{B}$ for every $x \in \mathcal{X}$. Prove that $f \in \mathcal{S}(\Omega \to \mathcal{X})$ if and only if $\{f = x\}$ is a measurable subset of $\mathcal{X}$ for every $x \in \mathcal{X}$.

(12) Prove Theorem 9.3.10 (3).

(13) Fill in the details of the proof of Theorem 9.3.12.

(14) Let $f \in \mathscr{S}^+(\mathbb{N} \to \mathbb{R})$. Note that (considering Exercise 10) the default σ-algebra on $\mathbb{N}$ is $2^{\mathbb{N}}$. Let ν be the measure on $\mathbb{N}$ defined by $\nu(\{n\}) = 1$ for every n. Prove that

$$\int f \, d\nu = \sum_{n=1}^{\infty} f(n).$$

Remark: This is a first step towards interpreting a series as an integral. The measure ν is called the *counting measure* on $\mathbb{N}$ since $\nu(E) = \#(E)$, which counts the number of elements in E.