# Honors Advanced Calculus

Conrad Plaut

2005-6

# Contents

# Chapter 1

# Logic, Set Theory, and the Real Numbers

In this chapter we will briefly cover the fundamentals necessary to study real analysis, including the basics of logic and set theory, and all of the axioms for the real numbers. We would like to begin by reminding the reader of the axiomatic method on which mathematics (and through it science) is based. It is impossible to define all concepts in any logical system, any more than a dictionary can define all words in a language without relying on some prior linguistic knowledge of the user (or pictures or other non-lingual aids). After all, what words would one use to write the "first" definition? The fact that any mathematical system must begin with terms, such as "set" or "element", that must always remain undefined is actually an advantage, not a disadvantage. After all, if we don't specify exactly what elements and sets must be, then we can apply our set theory to sets of numbers, sets of rabbits, sets of matrices, or sets of atoms.

Likewise, not all statements in a mathematical system can be proved. Indeed a proof is only a tool to determine that one statement follows logically from one or more other given statements; one must start somewhere! The starting point is a collection of statements involving the undefined terms, called axioms, that are assumed to be true. Given a particular set of axioms, a whole area of mathematics can be built up by proving statements that follow logically from those axioms, using these statements to prove more statements (adding new definitions when needed), *ad infinitum.* The axioms themselves are intended to be as few and simple and natural as possible, making it as easy as possible to check whether they are applicable in a given situation. Even a small number of axioms can give rise to a powerful collection of theorems. For example, all of the calculus that you have already learned, and much that you have not yet learned, is deduced from only nine axioms (actually between eight and twelve, depending on how you state them). While from the standpoint of mathematics these axioms are simply assumed, and cannot be "verified," their validity in science has been more than adequately justified by the powerful applications that arise from them.

Traditionally, theorems are named "lemmas", "propositions", "theorems" or "corollaries." While there is no fixed convention about the use of these names, a lemma is typically a statement with a relatively simple proof that is primarily used (often frequently) to prove other, deeper statements. A proposition is of greater importance as a statement by itself, and a theorem is usually a statement of major depth and importance. A corollary is always a statement that follows with only a small amount of proof from a theorem, proposition, or lemma.

Beyond their applicability to science, there is another reason why the axioms for the real numbers are likely the "right" ones. If one takes a subset of the axioms for the real numbers, then there are many different systems that satisfy them. For example, if one considers only the axioms concerning addition of numbers, without the commutative law, then the axioms describe something known in mathematics as a *group*. Groups are ubiquitous in mathematics and science, and come in many forms, such as finite groups, matrix groups, and vector spaces. Other subsets of the real number axioms give rise to other general mathematical objects, such as semigroups, rings, fields, and ordered fields. However, it can be proved that the real number system is the *only* number system that satisfies all of the real number axioms together.

We will not actually show that the real numbers exist. As mentioned above, we must start somewhere, and while it is possible to start with certain basic axioms of set theory, construct the real numbers, and show that they satisfy the axioms given later in this chapter, the methods used in this process are not applied further in the study of basic real analysis. Therefore after a few basic preliminaries about logic and set theory we will simply take the axioms for the real numbers as our starting point, and assume that there is a set satisfying those axioms. To reiterate: the real numbers can be constructed in a natural way beginning with everyday observations about counting, they are uniquely determined by their axioms, and they have powerful applications. They are certainly worth studying.

## 1.1 Basic Logic

We will informally discuss the basics of logic and set theory. In fact we will make no effort to develop set theory or logic in a rigorous fashion; as interesting as these subjects are, a thorough discussion here would detract from the subject at hand, not contribute to it. Such a "naive" approach to set theory will not lead to trouble so long as we stick to the real numbers, its subsets, sets (or, as we will often call them, collections) of such subsets, and sets constructed in very concrete ways from the real numbers. An example of the sort of troublemaker that we will avoid considering is "the set of all sets", which confounds the distinction between sets and their elements that our naive set theory would like to maintain.

If $A$ and $B$ represent statements, then the statement "*$A$ and $B$*" is true precisely if $A$ and $B$ are both true. For example, if $A$ is the statement "in the 20th century a meteor hit the earth" and $B$ is the statement "in the 20th century millions of people died," both of which are true, then the statement "in the 20th century a meteor hit the earth and millions of people died" is true. (Our tendency to connect these two statements and read a cause-and-effect relationship into this statement is not logical, and the mathematics student needs to learn to be very careful to avoid this and other kinds of illogic.)

The statement "$A$ or $B$" is true precisely if $A$, $B$, or both $A$ and $B$ are true. For example, the statement "that animal is brown or that animal is a cow" is true if the animal is a brown cat, a green cow, or a brown cow, but is false if the animal is a black sheep. The latter example shows how one negates an "or" statement. The negation of the statement "$A$ or $B$" is "not $A$ *and* not $B$" (in the previous example, not brown and not a cow). Likewise the negation of the statement "$A$ and $B$" is "not $A$ *or* not $B$".

Logic has two *quantifiers*, "for all" and "there exists". There are of course logically equivalent grammatical variations of these quantifiers, such as "for every" and "for some", respectively. "For all" means for every one, without exception, and "there exists" means for at least one (maybe many–or even all). The negation of a "for all" statement is always a "there exists" statement, and vice versa. For example, the negation of "all of my pencils are yellow" is "there exists one of my pencils that is not yellow," and the negation of "some French movies are boring" is "all French movies are not boring". Note that "for all" is a kind of "super *and*" while "there exists" is a kind of "super *or*" in the following sense. Consider statements $A$ and $B$. The statement "$A$ and $B$" is true precisely if all of the statements in question ($A$ and $B$) are true. The statement "$A$ or $B$" is true precisely if there exists a statement (among $A$ and $B$) that is true. While "and" and "or" can connect only two statements, "for all" and "there exist" have no such restrictions. For example, if we let $P(i)$ be the statement $1/i \leq 1/2$, where $i$ is a natural number, then we could write "$P(2)$ is true and $P(3)$ is true and $P(4)$ is true,..." but it is much simpler (and more precise–what does "..." really mean in this statement?) to write "$P(i)$ is true for all natural numbers $i \geq 2$". The parallels between "or" and "for some", and "and" and "for all" will appear again in the discussion of unions and intersections of sets in the next section.

The statement "if $A$ then $B$" is referred to as an implication, where $A$ is the hypothesis and $B$ is the conclusion. The statement "if $A$ then $B$" is false exactly when $A$ is true and $B$ is not true. For example, the statement "if a car is in lot B then the car is red" is false only if there is a car in lot B that is not red. In particular, it is true if the hypothesis is "vacuous" in the sense that the parking lot is empty! This bothers some students at first, but it really is consistent: if the only way to show the above statement is false is to produce a nonred car in lot B then why should it matter whether the failure to do so is the result of an empty parking lot? It is also important to remember that an implication is true whenever the hypothesis is false–regardless of whether the conclusion is true. For example, the statement "if $a$ is even then $3a$ is even" is true even though $3a$ is odd when $a$ is odd. This discussion can be summarized by stating that the negation of "if $A$ then $B$" is "$A$ and not $B$". Very often implications are equivalent to "for all" statements. For example, the above statement "if $a$ is even then $3a$ is even" is equivalent to the statement, "for all even $a$, $3a$ is even." The negation of this statement is "there exists some even $a$ such that $3a$ is odd," which makes more sense grammatically than "$a$ is even and $3a$ is odd." Getting back to the general statement "if $A$ then $B$", which may also be stated "$A$ implies $B$" or "$B$ only if $A$", there are two associated statements, the *converse* statement "if $B$ then $A$" and the *contrapositive* statement "if not $B$ then not $A$". For example, given the statement "if $x = y$ then $x^2 = y^2$", the converse is "if $x^2 = y^2$ then $x = y$" and the contrapositive is "if $x^2 \neq y^2$ then $x \neq y$." The contrapositive of a given statement is true if and only if the given statement is true; more succinctly, a statement and its contrapositive are logically equivalent. Proving or disproving one proves or disproves the other. On the other hand, there is generally no logical relationship between a statement and its converse. In the preceding example (assuming $x$ and $y$ are real numbers) the original statement is true, but its converse is false. Proving the converse of a statement rather than a statement is one of the most common abuses of logic, and is used routinely in politics and advertising. If both "$A$ implies $B$" and "$B$ implies $A$" are true, then we say that $A$ and $B$ are (logically) equivalent; we may also express this as "$A$ if and only if $B$" or "$A$ iff $B$".

The previously stated basic rules for negation of statements are few and simple, but actually negating complex statements can take some practice. Consider, for example, the statement

$$\text{"For every natural number } n \text{ there exist } \varepsilon > 0 \text{ and } \delta > 0 \text{ such that } \frac{1}{\varepsilon} \geq n - \delta.\text{"} \tag{1.1}$$

which negates as "There exists an natural number $n$ such that for every $\varepsilon > 0$ and $\delta > 0$, $\frac{1}{\varepsilon} < n - \delta$." Why did we negate "$\frac{1}{\varepsilon} \geq n$", but not "$\varepsilon > 0$" or "$\delta > 0$"? Why didn't we change the "and" to an "or"? Note that the symbol "$\geq$" stands for "greater than *or* equal to" and so the negation of "$\frac{1}{\varepsilon} \geq n - \delta$" is "$\frac{1}{\varepsilon}$ is not greater than $n$ *and* not equal to $n - \delta$" which leaves as the only possibility "$\frac{1}{\varepsilon} < n - \delta$". With a little practice, negation of statements can become fairly instinctive, as it must be to properly do mathematics.

**Exercise 1** *For each of the following, write down the negation, converse and contrapositive. Based on your prior knowledge, state whether each statement and its converse are true or false. You do not need to supply a proof!*

1. *If $n > M$ then there exists some $\varepsilon > 0$ such that $M + \varepsilon < n$.*
2. *If $n$ and $m$ are integers then there exists some rational $r$ such that $n < r < m$.*
3. *If $a_i < k$ for all $i$ and $\lim_{i\to\infty} a_i$ exists, then $\lim_{i\to\infty} a_i < k$.*
4. *If $f'(x) \leq g'(x)$ for all $x$ such that $0 \leq x \leq 1$ and $f(0) = g(0)$ then $f(x) \leq g(x)$ for all such $x$.*

There are certain commonly used symbols that represent basic components of logic. “For all” is normally written as “$\forall$,” “there exists” as “$\exists$”, “such that” as “$\ni$”, “if $A$ then $B$” as “$A \Rightarrow B$”, “$A$ if and only if $B$” as “$A \Leftrightarrow B$”, “$A$ or $B$” as “$A \vee B$”, and “$A$ and $B$” as “$A \wedge B$”. For example, the statement (1.1) is abbreviated as “$(\forall n \in \mathbb{N})\exists\ (\varepsilon > 0 \wedge \delta > 0)\ \ni (\frac{1}{\varepsilon} \geq n - \delta)$”. The parentheses are included for clarity. Often the “$\ni$” symbol is not used; the presumption being that the phrase following a “there exist” phrase is the “such that” phrase. This makes the statement harder to parse at first, but one advantage of doing this is that the process of negation is simplified (try it!), because the “such that” doesn’t have to be relocated. With the exception of arrows for implications, this simplified notation will rarely be used in this text, but every mathematics student should be familiar with it.

## 1.2 Basic Set Theory

Sets and elements of sets are undefined terms. We write “$x \in X$” to express “$x$ is an element of the set $X$” and “$x \notin X$” to express the negation of this statement. Given sets $A$ and $B$, the *union* of $A$ and $B$ is defined to be the set of all $x$ such that $x \in A$ or $x \in B$; we will use the notation $A \cup B := \{x : x \in A$ or $x \in B\}$. The *intersection* of $A$ and $B$ is defined by $A \cap B := \{x : x \in A$ and $x \in B\}$, and the complement of $B$ in $A$ (more briefly “$A$ take away $B$”) is defined by $A \backslash B := \{x : x \in A \text{ and } x \notin B\}$. The *empty set*, which is the set with no elements, is denoted by $\varnothing$. If $A \cap B = \varnothing$ then $A$ and $B$ are said to be *disjoint*. We say that $A$ is a *subset* of $B$ (written $A \subset B$) if whenever $x \in A$, $x \in B$. We say that $A$ is a *proper subset* of $B$ if $A \subset B$ and there exists some $x \in B$ such that $x \notin A$. We denote this by $A \subsetneq B$. If $X$ is a set and $A$ is a subset of $X$, we define the *complement of* $A$ in $X$ by $A^c := \{x \in X : x \notin A\}$.

To show two sets are equal, it is often convenient to show that one is a subset of the other, and vice versa. We will illustrate this by showing that if $A$ and $B$ are disjoint then $A \backslash B = A$. We will begin by showing that $A \backslash B \subset A$. Let $x \in A \backslash B$. Then by definition $x \in A$ and $x \notin B$. In particular, $x \in A$, which is what we needed to prove. To show that $A \subset A \backslash B$, suppose that $x \in A$. If $x$ were also an element of $B$ we would have $x \in A \cap B = \varnothing$, which is impossible. Therefore $x \notin B$. Since we already had $x \in A$ this shows that $x \in A \backslash B$. The above proof that $x \notin B$ is an example of a *proof by contradiction*, in which one assumes the negation of what one wishes to prove, and then show that this negation logically implies a false statement and therefore itself must have been false. The usefulness of proof by contradiction is one of many reasons why the mathematics student must be good at negating statements.

Elementary set theory is completely analogous to elementary logic, and basic statements in logic have equivalents in set theory, and vice versa. Note the similarity of the symbols for “and” and “intersect”, and “or” and “union”. For example, the fact that the negation of an “and” statement is an “or” statement is expressed in set theory by one of *de Morgan’s laws:*

$$A \backslash (B \cap C) = (A \backslash B) \cup (A \backslash C) \tag{1.2}$$

We will only prove the inclusion $\subset$. Let $x \in A \backslash (B \cap C)$. By definition $x \in A$ and $x \notin B \cap C$. That is, it is not true that $x \in B$ and $x \in C$, which means $x \notin B$ or $x \notin C$. When a proof involves an "or" statement it is often useful to consider each of the two statements as a separate statement. For the present proof suppose that $x \notin B$. Since we already have $x \in A$, then by definition $x \in A \backslash B$, and therefore $x \in (A \backslash B) \cup (A \backslash C)$. In the preceding argument, if we simply replace the letter "$B$" by the letter "$C$" then the exact same proof will show that $x \in A \backslash C$, and therefore $x \in (A \backslash B) \cup (A \backslash C)$. Rather than wasting time writing this down we simply state that the proof of the case $x \notin C$ is similar. We leave it to the reader to prove the opposite inclusion ($\supset$) and the second de Morgan law:

$$A \backslash (B \cup C) = (A \backslash B) \cap (A \backslash C) \tag{1.3}$$

**Exercise 2** *Let $A$ and $B$ be sets with $B \subset A$. Prove that $A \backslash (A \backslash B) = B$. Give an example to show that the statement is not true if we do not assume that $B \subset A$.*

There are other statements that we will use without going through the proofs, mainly because they reduce to statements in logic that are best proved using what are known as "truth tables," which, although very basic, are outside the scope of this text. The interested reader is urged to consult a book on basic logic. Statements we will need are $A \cap B = B \cap A$, $A \cap (B \cap C) = (A \cap B) \cap C$ and the identical statements with "$\cap$" replaced by "$\cup$". It follows (although the proof is a tedious inductive argument that we will skip) that the order of parentheses and sets themselves are irrelevant for any combination of unions or any combination of intersections of sets. For example, it is unambiguous to write $A \cap B \cap C$. Mixing intersections and unions is a different story, however; for example it is not generally true that $(A \cap B) \cup C = A \cap (B \cup C)$.

Although one can prove that a statement is false by showing that it logically implies another statement already known to be false, often the best and simplest way to show that a statement is false is to provide a concrete counterexample. For example, to see why it is false that $(A \cap B) \cup C = A \cap (B \cup C)$ for all sets $A$, $B$, $C$, let $A = \{1\}$, $B = \{2\} = C$. Then $(A \cap B) \cup C = \{2\}$, but $A \cap (B \cup C) = \varnothing$.

A common elementary mistake in trying to disprove a statement is to point out that a particular attempt at a proof fails to work. This is like trying to prove that it is impossible to drive from place $P$ to place $Q$ by showing that one particular road from $P$ leads to a dead end! Nonetheless, the failure of an attempt at a proof, if carefully analyzed, can sometimes lead to a concrete counterexample.

Another common mistake is to give a "counterexample" that is not concrete. For example, let's "disprove" the statement that $x^2 = -1$ has no real solution. Let $x$ and $y$ be real numbers such that $y = 2x$ and $y = x^2 + 2x + 1$. Then we have $2x = x^2 + 2x + 1$ or $x^2 + 1 = 0$, or $x^2 = -1$. Since $x$ is a real number, the statement is false. The error, of course, is that the graphs of the real functions $y = 2x$ and $y = x^2 + 2x + 1$ do not intersect, so no such $x$ and $y$ exist. Trying to find *concrete* real $x$ and $y$ with those properties would quickly have exposed the error. As a rule, counterexamples, even simple ones, must be concrete.

We do have the following relations involving both unions and intersections (called the set theory distributive laws) for all sets $A, B, C$:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \text{ and } A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

**Exercise 3** *Prove or disprove for all sets $A, B$:*

1. $A \backslash B = \varnothing$ if and only if $A = B$

2. $A = (A \backslash B) \cup (A \cap B)$

3. $A \backslash (A \backslash (A \backslash A)) = \varnothing$

Let $X$ and $Y$ be sets. The *Cartesian product* of $X$ and $Y$ is $X \times Y := \{(x, y) : x \in X \text{ and } y \in Y\}$. The elements of a Cartesian product are called *ordered pairs*; the order of the elements in the pair is essential, and this distinguishes the ordered pair $(x, y)$ from the set $\{x, y\}$, in which order is unimportant. In particular, the cartesian product is not "commutative"; $X \times Y$ and $Y \times X$ are different sets unless $X = Y$. The cartesian product is one of the most basic and important ways to construct a new set from existing sets. For example, as the reader knows, the Euclidean plane is the cartesian product of the real line with itself.

Cartesian products can be used to make precise the idea of "labeling" a collection of sets. An *indexing set* $\Lambda$ for a collection $\mathcal{A}$ of sets is a subset of $\Lambda \times \mathcal{A}$ such that for each $\lambda \in \Lambda$ there is exactly one ordered pair $(\lambda, A)$. We normally write $A_\lambda$ rather than $(\lambda, A)$ and denote the indexed collection by $\{A_\lambda\}_{\lambda \in \Lambda}$. Thus the indexing of $\mathcal{A}$ "assigns" to each $\lambda$ a unique set $A_\lambda$ in the collection $\mathcal{A}$. For example, we can consider the collection $\{(-n, n)\}_{n \in \mathbb{N}}$ of all real intervals $(-n, n)$, where $n$ is a natural number. Written more explicitly this is the set $\{(-1, 1), (-2, 2), (-3, 3), ...\}$. In this example the indexing set is $\mathbb{N}$. However, we can consider sets indexed over any arbitrary set–e.g. $\{[-r, r]\}_{r \in (0,1)}$, which consists of all closed intervals having endpoints $[-r, r]$, where $0 < r < 1$.

Given an indexed collection of sets $\{A_\lambda\}_{\lambda \in \Lambda}$, we define the *intersection* of the collection to be

$$\bigcap_{\lambda \in \Lambda} A_\lambda := \{x : x \in A_\lambda \text{ for all } \lambda \in \Lambda\}$$

and the *union* of the collection to be

$$\bigcup_{\lambda \in \Lambda} A_\lambda := \{x : x \in A_\lambda \text{ for some } \lambda \in \Lambda\}.$$

**Example 1** *We will prove that for sets $A$ and $\{A_\lambda\}_{\lambda\in\Lambda}$ we have the following distributive law:*

$$A \cap \left( \bigcup_{\lambda\in\Lambda} A_\lambda \right) = \bigcup_{\lambda\in\Lambda} (A \cap A_\lambda)$$

*Now $x \in A \cap \left(\bigcup_{\lambda\in\Lambda} A_\lambda\right)$ if and only if*

$$x \in A \textit{ and } x \in \bigcup_{\lambda\in\Lambda} A_\lambda$$

$$\Leftrightarrow x \in A \textit{ and } x \in A_\lambda \textit{ for some } \lambda \in \Lambda$$

$$\Leftrightarrow x \in A \cap A_\lambda \textit{ for some } \lambda \in \Lambda$$

$$\Leftrightarrow x \in \bigcup_{\lambda\in\Lambda} (A \cap A_\lambda)\,.$$

**Exercise 4** *Verify that if we have only two sets $A_1$ and $A_2$ (i.e. the indexing set is $\Lambda = \{1, 2\}$) then $\bigcap_{\lambda\in\{1,2\}} A_\lambda = A_1 \cap A_2$ and $\bigcup_{\lambda\in\{1,2\}} A_\lambda = A_1 \cup A_2$. Therefore our new more general definitions are consistent with the old ones.*

**Exercise 5** *Let $\{A_\lambda\}_{\lambda\in\Lambda}$ be a collection of sets. Show that for each $\lambda_0 \in \Lambda$, $\bigcap_{\lambda\in\Lambda} A_\lambda \subset A_{\lambda_0} \subset \bigcup_{\lambda\in\Lambda} A_\lambda$.*

When $\Lambda = \mathbb{N}$ (or $\Lambda$ is finite) we will often write, for example, $\bigcup_{i=1}^{\infty} A_i$ (or $\bigcap_{i=1}^{m} A_i$).

**Exercise 6** *Find the intersections and unions of the following collections*

1. $\{(-n, n)\}_{n=1}^{\infty}$
2. $\{[-r, r]\}_{r\in(0,1)}$
3. $\{(0, \frac{1}{i}]\}_{i=1}^{\infty}$

**Exercise 7** *Prove the following generalization of the de Morgan law (1.3) for arbitrary intersections $A\backslash \left(\bigcup_{\lambda\in\Lambda} A_\lambda\right) = \bigcap_{\lambda\in\Lambda}(A\backslash A_\lambda)$. Formulate the appropriate generalizations of (1.2) and the remaining distributive law, but don't bother to write down proofs.*

Consider now the cartesian product of three sets $A, B, C$. Using our definition of the cartesian product of two sets we can form the products $(A \times B) \times C$ and $A \times (B \times C)$. Elements of the first set are of the form $((a, b), c)$ and elements of the second are of the form $(a, (b, c))$. Strictly speaking these two sets are different, and yet for all purposes it is useful to consider them as being the same–moreover, we would like to use the simpler notation of triples $(a, b, c)$, or, for cartesian product of $n$ sets, $n$-tuples.

**Definition 2** *Let $A_1, ..., A_n$ be sets. The cartesian product of the sets is defined to be*

$$A_1 \times \cdots \times A_n = \prod_{i=1}^{n} A_i := \{(a_1, ..., a_n) : a_i \in A_i\}.$$

*For any $(a_1, ..., a_n) \in A_1 \times \cdots \times A_n$, $a_i$ will be called the $i^{th}$ component (or coordinate) of $(a_1, ..., a_n)$.*

It is possible to consider cartesian products of arbitrary collections $\{A_\lambda\}_{\lambda \in \Lambda}$, but this important construction will not be used in this introductory text. If the sets $A_i$ are all the same set, i.e. $A_i = B$ for some set $B$ and all $i$, we will write $B^n$ rather than $\prod_{i=1}^{n} A_i$. It will be useful in the future to be able to make statements like $A \times (B \times C) = A \times B \times C$, which, as we have pointed out above, is strictly speaking not true. But each of these two sets is simply a relabeling of the other in a very natural way; the "difference" between these sets is purely a matter of notation. We resolve this situation by stating that we will henceforth "identify" the one set with the other, using the notation established in Definition 2. Later, when we add further structures to the sets $A_i$, we should pause for a moment to be sure that this identification that we have made is consistent with the new structure. For example, in linear algebra you have already worked with vectors expressed as $n$-tuples of real numbers, and likely made this identification without even thinking about it when stating, for example, that $\mathbb{R}^4$ "is" the direct sum of $\mathbb{R}^2$ and $\mathbb{R}^2$. Strictly speaking these two vector spaces are different, but isomorphic as vector spaces via the natural function $(a_1, a_2, a_3, a_4) \mapsto ((a_1, a_2), (a_3, a_4))$. These kinds of identifications in mathematics are inevitable and frequently encountered.

## 1.3 Functions

**Definition 3** *Let $X$ and $Y$ be sets. A function $f$ from (or on) $X$ to (or into) $Y$ is a subset of $X \times Y$ such that if $(x, y)$ and $(x, z)$ are in $f$ then $y = z$.*

Normally one assumes that $X$ is the domain of the function, meaning that for every $x \in X$ there is some $(x, y) \in f$. Without further ado we will adopt the standard notation for functions, writing $f : X \to Y$ (where $X$ is the domain of the function) $y = f(x)$ rather than $(x, y) \in f$. The last condition in the above definition then is clearly the familiar requirement that a function be single-valued in the sense that if $f(x) = y$ and $f(x) = z$, then $y = z$. Note that the indexing of a collection of sets discussed in the previous section is in fact a function. The set $X$ is called the domain of $f$ and the set of all $y \in Y$ such that $y = f(x)$ for some $x \in X$ is called the range of $f$. The function $f$ is called *one-to-one (or 1-1, or an injection)* if whenever $f(x) = f(y)$, $x = y$, and *onto (or a surjection)* if for every $y \in Y$ there exists some $x \in X$ such that $y = f(x)$, i.e. if $Y$ is the range of $X$. A function that is both 1-1 and onto is called a *bijection (or one-to-one correspondence).*

Given functions $f : X \to Y$ and $g : Y \to Z$, we define the *composition* of $f$ and $g$ to be $g \circ f : X \to Z$, where $(g \circ f)(x) = g(f(x))$ for all $x \in X$.

**Exercise 8** *Let $f : X \to Y$ and $g : Y \to Z$ be functions.*

1. *Show that if both $f$ and $g$ are 1-1 (respectively onto) then $g \circ f$ is 1-1 (respectively onto). Note: the "(respectively onto)" indicates that this is really two statements, one without these parenthetical phrases, which is about 1-1 functions, and a second statement in which "onto" replaces "1-1" in each instance. You should prove both. In the future we will often abbreviate, using "resp."*

2. *Show that if $g$ is 1-1 and $g \circ f$ is onto then $f$ is onto.*

If $f : X \to Y$ is both 1-1 and onto then there exists a uniquely determined function $f^{-1} : Y \to X$, called "$f$ inverse" defined by $f^{-1}(y) = x$ where $x$ is the unique element of $X$ such that $f(x) = y$. By definition, $f^{-1} \circ f(x) = x$ and $f \circ f^{-1}(x) = x$, and according to Exercise 8, $f^{-1}$ is 1-1 and onto. In addition, $(f^{-1})^{-1} = f$. The function $id_X : X \to X$ defined by $id_X(x) = x$ is called the *identity function* on $X$. In this notation we can write $f^{-1} \circ f = id_X$ and $f \circ f^{-1} = id_Y$.

If $f : X \to Y$ is a function and $A \subset X$ we define $f(A) = \{y \in Y : y = f(x)$ for some $x \in A\}$, called the *image of* $A$. If $B \subset Y$, we define $f^{-1}(B) = \{x \in X : f(x) \in B\}$, called the *inverse image* of $B$. Note that $f^{-1}(B)$ makes sense even if $f$ is not 1-1 or onto.

**Example 4** *Let $f(x) = x^2$. Determine the following sets (no proof needed): $f(\{1, -1\})$, $f^{-1}((0, 1))$, $f(f^{-1}(\{-1\}))$, $f^{-1}(f(\{-1\}))$.*

**Exercise 9** *Let $f : X \to Y$ be a function. Prove that*

1. *$A \subset f^{-1}(f(A))$ and $f(f^{-1}(B)) \subset B$*

2. *$f$ is 1-1 if and only if for every $A \subset X$, $f^{-1}(f(A)) = A$.*

3. *$f$ is onto if and only of for every $B \subset Y$, $f(f^{-1}(B)) = B$.*

We list now list various statements involving images and inverse images of functions.

$$f\left(\bigcup_{\lambda\in\Lambda} A_\lambda\right) = \bigcup_{\lambda\in\Lambda} f(A_\lambda) \text{ and } f\left(\bigcap_{\lambda\in\Lambda} A_\lambda\right) \subset \bigcap_{\lambda\in\Lambda} f(A_\lambda) \tag{1.4}$$

$$f^{-1}\left(\bigcup_{\lambda\in\Lambda} A_\lambda\right) = \bigcup_{\lambda\in\Lambda} f^{-1}(A_\lambda) \text{ and } f^{-1}\left(\bigcap_{\lambda\in\Lambda} A_\lambda\right) = \bigcap_{\lambda\in\Lambda} f^{-1}(A_\lambda) \tag{1.5}$$

$$f^{-1}(A \backslash B) = f^{-1}(A) \backslash f^{-1}(B) \text{ and } f(A)/f(B) \subset f(A/B) \tag{1.6}$$

**Exercise 10** *Prove the statements (1.4-1.6).*

Note that it is easy to check that the inclusions in Formulas 1.4 and 1.6, are equalities when $f$ is $1-1$. .Let $f : X \to Y$ be a function and $A \subset X$. The *restriction* of $f$ to $A$ is the function $f\mid_A: A \to Y$ defined by $f\mid_A (x) = f(x)$ for all $x \in A$. That is, $f\mid_A$is simply $f$, but its domain has been restricted to $A$. This simple concept is more useful than it may seem at first, and will be used frequently.

Finally, we consider functions into cartesian products of sets. Let $A, A_1, ..., A_n$ be sets and $f : A \to \prod_{i=1}^{n} A_i$ be a function. For any $a \in A$, $f(a)$ is of the form $(a_1, ..., a_n)$ for some $a_i \in A_i$. In other words, to each $a \in A$ we assign, using $f$, some $a_i \in A_i$. That is, the function $f$ gives rise to a uniquely determined set of functions $f_i : A \to A_i$, where $f(a) = (f_1(a), ..., f_n(a))$; the function $f_i$ is called the $i^{th}$ component function of $f$. For example, the first component of $f(t) = (t^2 + 1, t, 2)$ is $f_1(t) = t^2 + 1$ and the third component is the constant function $f_3(t) = 2$. Conversely, if we are given functions $g_i : A \to A_i$ for all $i$, then there is a uniquely determined function $g : A \to \prod_{i=1}^{n} A_i$ defined by $g(a) = (g_1(a), ..., g_n(a))$.

Now suppose that $f : \prod_{i=1}^{n} A_i \to A$ is a function. We will adopt the notation commonly used in calculus, writing $f(a_1, ..., a_n)$ rather than the strictly correct notation $f((a_1, ..., a_n))$.

**Definition 5** *For any $j = 1, ..., n$, the function $\pi_j : \prod_{i=1}^{n} A_i \to A_j$ defined by $\pi_j(a_1, ..., a_n) = a_j$ is called the $j^{th}$ projection of $\prod_{i=1}^{n} A_i$.*

**Exercise 11** *Let $f : A \to \prod_{i=1}^{n} A_i$ be a function.*

1. *Prove that, for every $i$, $f_i = \pi_i \circ f$.*

2. *Prove that if every component function $f_i : A \to A_i$ is 1-1 then $f$ is 1-1. Verify that the function $f(t) = (t, t^2)$ from $\mathbb{R}$ into $\mathbb{R}^2$ is a counterexample to the converse of this statement.*

3. *Prove that if $f$ is onto then every component $f_i : A \to A_i$ is onto. Prove or disprove the converse of this statement.*

## 1.4 The Field and Order Axioms

The axioms in this section are familiar to most students from before high school. They are quite natural. Many of them, such as the distributive and associative laws, can be observed in simple concrete situations, such as counting apples, and have been understood and accepted for thousands of years. Others, such as the existence of 0, or the existence of additive inverses, which leads to negative numbers, have been readily accepted only more recently. After all, even if you accept that one can physically have 0 apples, $-5$ apples is a little harder to visualize. However, whether or not $-5$ apples makes physical sense, we now know that negative numbers are extremely useful in many situations. Rather than denying their existence generally, as some mathematicians and scientists (and religious leaders!) have tried to do in previous centuries, we simply recognize that, in the process of translating from an application to mathematics, solving the problem, and translating back, we must be on the lookout for extraneous solutions, or solutions that don't make sense in terms of the original problem.

A *field* is a set $\mathbb{F}$ having two operations $+$ and $\cdot$ (addition and multiplication) satisfying the following axioms:

1. (Associative Laws) $(a + b) + c = a + (b + c)$ and $a(bc) = (ab)c$ for all $a, b, c \in \mathbb{F}$

2. (Commutative Laws) $a + b = b + a$ and $ab = ba$ for all $a$ and $b$ in $\mathbb{F}$

3. (Additive and Multiplicative Identities) There exist elements 0 and 1 in $\mathbb{F}$ with $1 \neq 0$, such that $a + 0 = a$ and $1 \cdot a = a$ for all $a \in \mathbb{F}$

4. (Additive and Multiplicative Inverses) For all $a \in \mathbb{F}$ there is some $-a \in \mathbb{F}$ such that $a + (-a) = 0$ and if $a \neq 0$ there exists some $a^{-1} \in \mathbb{F}$ such that $aa^{-1} = 1$

5. (Distributive Law) $a(b + c) = ab + ac$ for all $a, b, c \in \mathbb{F}$.

Although there are many important fields, in this text we will be only concerned with the real and complex numbers. The complex numbers will be introduced later. From these five axioms can be deduced all the algebraic theorems that the reader has used since childhood. For example, a quite tedious induction argument can be used to prove that the associative law implies that sums and products of any length are independent of arrangement of parentheses–e.g. $a(b(cd)) = (ab)(cd)$ for any real numbers $a, b, c, d$. There is also a "right-sided" distributive law in which the single element is multiplied on the right, which can be proved using the distributive and commutative laws. We will not dwell further on these kinds of theorems, which are fairly simple to prove, and which are found in a typical abstract algebra course (in the more general setting of groups and rings). The reader is not expected to cite each axiom or even show more steps than would be expected in a calculus course. However, the idea of uniqueness has a fairly important role in mathematics, and we will spend a little time proving some basic uniqueness results.

**Proposition 6** *For all $a, b$ in a field, the equation $a + x = b$ has a unique solution $x$.*

By now basic algebra has become fairly "automatic" to the reader, and he or she might give the following argument

$$\begin{aligned} a + x &= b \qquad (1.7) \\ 0 + x &= (-a) + b \\ x &= b - a \end{aligned}$$

Here we have used the standard convention of denoting $b+(-a)$ by $b-a$. Does this mean that $x$ really is a solution to our original equation? Does this show that $x$ is unique? A careful look at the axioms shows that what we in fact have here is three *equivalent* statements. It is clear that the top line implies the second one by adding $-a$ to each side and using the axiom about the additive inverse. But we can also go from the second line to the first by adding $a$ to each side and applying a couple of axioms as well. In other words, using our axioms we can show that the first line implies the second, and that the second line implies the first, hence they are equivalent. The reader should also check that the second and third lines are equivalent, and therefore the first and third lines are equivalent.

What have we proved? Consider first the statement "If $a+x=b$ then $x=b-a$." This is actually the uniqueness of the solution–if there is a solution $x$ then it must be of the form $b-a$. The converse statement "If $x=b-a$ then $a+x=b$" is actually the statement that a solution exists, namely $b-a$. Therefore we have proved that there is a unique solution, and it is of the form $x=b-a$. One can actually check that $x=b-a$ is a solution by "plugging it in" but this is not needed. Likewise one can prove uniqueness in the following way: Suppose that $x_1$ and $x_2$ are solutions to $a+x=b$. Then $a+x_1=b=a+x_2$. Adding $-a$ to each side we obtain that $x_1=x_2$. While this is the standard framework for proving uniqueness of solutions, again this is unneeded in the present case because we have carefully checked our logic. Therefore careful analysis of one's logic can be a great time saver. Such analysis can also prevent errors. Consider, for example, the following "proof" that $1=2$.

$$\begin{aligned} 1 &= 2 \\ 1\cdot 0 &= 2\cdot 0 \\ 0 &= 0 \end{aligned} \tag{1.8}$$

This example plays on the common (and logically incorrect) strategy learned in high school algebra of starting with the statement that one wants to verify, and manipulating it algebraically to obtain a true statement. In high school algebra this plan generally worked because the algebraic manipulation always produced logically equivalent statements. However, we will more and more consider steps that do not produce logically equivalent statements, and in these cases the strategy is clearly backwards. To prove a statement one must start with something known to be true and deduce the statement, not the other way around. In the above example the first line implies the second line, which implies the third line–but that only serves to illustrate that a false statement can logically imply a true statement! Of course the second statement does not imply the first; in order to go this way one would have to multiply by the multiplicative inverse of 0, and our axioms do not provide for such a thing. From this point on, all arguments of this sort, which we will refer to as "parallel" arguments because they involve a list of statements in parallel, must have an indication of which way the logic goes. Therefore the argument 1.7 should be

$$\begin{aligned} a + x &= b \\ \Leftrightarrow 0 + x &= (-a) + b \\ \Leftrightarrow x &= b - a \end{aligned}$$

and the argument 1.8 should be

$$\begin{aligned} 1 &= 2 \\ \Rightarrow 1 \cdot 0 &= 2 \cdot 0 \\ \Leftrightarrow 0 &= 0 \end{aligned}$$

Applying Proposition 6 to the equation $a + x = a$ we see that 0 is the unique additive identity, and applying it to $a + x = 0$ we see that each $a$ has a unique additive inverse. In fact, we could have replaced the two axioms concerning the additive inverse and identity by a single one asserting the existence of a unique solution to all equations of the form $a + x = b$ (see Exercise 13 below).

**Exercise 12** *State and prove a theorem about the existence and uniqueness of solutions of equations of the form $ax = b$ for $a, b, x$ in a field.*

**Exercise 13** *Suppose we are given that $a + x = b$ has a unique solution for all $a, b \in \mathbb{R}$. Use this fact, and only the associative and commutative laws, to show that there exists an additive identity. Hint: Let $0_a$ be the unique solution to $a + x = a$. Next show that for any $b \in \mathbb{R}$, $b + 0_a$ is a solution to $a + x = a + b$.*

Henceforth we will take for granted all algebraic theorems about the real numbers, using them without reference. We finish the discussion of algebraic properties with a few comments about powers of real numbers. For quite some time we will not need to know anything about powers except for integer powers and square roots. For any $x \neq 0$ we define $x^n$ for $n \geq 0$ iteratively: $x^0$ is defined to be 1 and for any $n$, $x^n$ is defined to be $x^{n-1} \cdot x$. For $n < 0$ we define $x^n = (x^{-n})^{-1}$. Some rather tedious checking verifies the usual powers rules: $x^n x^m = x^{n+m}$ and $(x^m)^n = x^{mn}$. We will discuss square roots in the next section.

We will now add the order axioms to our collection. We assume that there is a subset $L$ of $\mathbb{R} \times \mathbb{R}$, the elements of which will be denoted by $a < b$ (rather than $(a, b)$), such that the following axioms are satisfied for all $a, b, c \in \mathbb{R}$:

1. (Transitivity) If $a < b$ and $b < c$ then $a < c$.
2. (Trichotomy) Exactly one of the following is true: $a < b$, $b < a$ or $a = b$.
3. (Additive Property) If $a < b$ then $a + c < b + c$.
4. (Multiplicative Property) If $a < b$ and $c > 0$ then $ac < bc$.

From these three axioms and the algebraic axioms follow all the basic algebraic theorems that the reader has learned previously, including, for example, the fact that multiplication by a negative number reverses inequalities, $0 < 1$, and so on. We will not give the proofs of such statements, which are elementary (if sometimes tricky), and will use all of these statements without further reference.

A very common method for showing that $a = b$ is to show both $a \leq b$ and $b \leq a$. Then by the trichotomy, the only possibility is $a = b$. A useful basic technique for showing that $a \leq b$ is given by the following exercise.

**Exercise 14** *Let $a, b \in \mathbb{R}$.*

1. *Show that if $a \leq b + \varepsilon$ for every $\varepsilon > 0$, then $a \leq b$.*
2. *Prove or disprove: If $a < b + \varepsilon$ for every $\varepsilon > 0$ then $a \leq b$.*
3. *Prove or disprove: If $a < b + \varepsilon$ for every $\varepsilon > 0$ then $a < b$.*

## 1.5 Completeness

For any $x \in \mathbb{R}$ we define the absolute value of $x$ to be $|x| = x$ if $x \geq 0$ and $|x| = -x$ if $x < 0$. It is easy to check the four necessary cases to verify that $|xy| = |x|\,|y|$ for all $x, y \in \mathbb{R}$.

**Proposition 7** *(Triangle Inequality) For any $x, y \in \mathbb{R}$ we have*

$$|x + y| \leq |x| + |y|$$

The proof of the triangle inequality involves a series of simple steps, the first of which is the following useful lemma:

**Lemma 8** *For all $x \in \mathbb{R}$ and $M \geq 0$, $|x| \leq M$ if and only if $-M \leq x \leq M$.*

**Exercise 15** *Prove the above lemma and use it to show that for all $x, y \in \mathbb{R}$, $-(|x| + |y|) \leq x + y \leq |x| + |y|$. Use the lemma again to finish the proof of the triangle inequality.*

**Definition 9** *A subset $A$ of $\mathbb{R}$ is bounded above if there exists some $M \in \mathbb{R}$ (called an upper bound of $A$) such that $x \leq M$ for all $x \in A$. If there exists an upper bound $S$ of $A$ such that $S \leq M$ for all upper bounds $M$ of $A$ then $S$ is called the supremum (or sup–with a long "u") of $A$, denoted* $\sup A$. *If* $\sup A \in A$ *then* $\sup A$ *is called the maximum of $A$, denoted* $\max A$. *Lower bound, infimum and minimum are defined similarly. A real subset that is bounded above and below is simply called bounded.*

**Exercise 16** *Show that $A \subset \mathbb{R}$ is bounded if and only if there exists some $M \geq 0$ such that for all $x \in A$, $|x| \leq M$.*

**Definition 10** *The extended reals are defined to be the set* $\mathbb{E} := \mathbb{R} \cup \{\infty, -\infty\}$*, where* $\infty$ *and* $-\infty$ *are symbols that satisfy the following conventions:*

1. *For every real* $x$*,* $\infty > x$ *(resp.* $-\infty < x$*) and* $\infty$ *(resp.* $-\infty$*) is an upper (resp. lower) bound for any subset of* $\mathbb{E}$*.*
2. *If* $A \subset \mathbb{R}$ *is not bounded above (resp. below) or* $A \subset \mathbb{E}$ *contains* $\infty$ *(resp.* $-\infty$*) then we write* $\sup A = \infty$ *(resp.* $\inf A = -\infty$*).*
3. $x + \infty = \infty$ *for any extended real* $x > -\infty$ *and* $x \cdot \infty = \infty$ *(resp.* $-\infty$*) for any extended real* $x > 0$ *(resp.* $x < 0$*), with analogous conventions for* $-\infty + x$ *and* $x \cdot (-\infty)$*.*

We are not actually defining operations of addition and multiplication on $\mathbb{E}$ (for example we will not consider the combination $\infty + (-\infty)$), but it will be useful to have statements about sums and products of infs and sups (and, at a later time, limits and integrals) that are true also for infinite values. Of course when proving such statements we will have to check the infinite cases separately (see for example Lemmas 13 and 14 below). It is important to realize that these theorems only apply to those operations defined in Definition 10 and they tell us nothing, for example, if one encounters something like $0 \cdot \infty$.

**Exercise 17** *Show that every real number is an upper bound for* $\varnothing$ *and explain why it would be reasonable to define* $\sup \varnothing = -\infty$*. We will not do so, mainly because it is simpler to only consider non-empty sets when dealing with suprema and infima than it is to treat the empty set as a separate case in every proof involving suprema.*

**Exercise 18** *Prove that if* $A \neq \varnothing$ *then* $\sup A$ *is unique. Be sure to address the case when* $A$ *is not bounded above.*

The final axiom for the real numbers states:

**Axiom 11 (Completeness)** *Every nonempty subset of the real numbers has a supremum.*

The above formulation includes our convention about unbounded sets: a subset of the reals is bounded above or not. If it is bounded above then according to the completeness axiom it has a supremum; if it is unbounded then it has a supremum of $\infty$ by definition. One can argue without much difficulty that the completeness axiom also implies that every subset of the real numbers has an infimum. In fact, a set $A$ is bounded above if and only if $-A$ is bounded below; a number $M \geq 0$ is an upper bound for $A$ if and only if $-M$ is a lower bound for $-A$; and the previous statement follows from lots of negating and "un-negating" of sets and numbers. We will see later that a much more useful kind of completeness is true: the completeness axiom implies that every real sequence that "should" converge to something because the terms get "closer and closer together" (we will define this precisely later!) actually does have a limit. This is a statement that is taken for granted in modern elementary calculus and was taken for granted in a limited sense by mathematicians for centuries prior to the $19^{th}$ century when the need to formulate the axiom was realized. Without the completeness axiom, familiar calculus–and even Euclidean geometry–break down. For example, the rational numbers are an *ordered field* in the sense that they satisfy all the axioms for the real numbers except completeness. But if one works only with the rational numbers then a square with sides of length 1 has a diagonal of undefined length; lines in the "rational plane" with different slopes may not intersect; the polynomial $x^2 - 2$ has no zeros. The real numbers can be constructed by "completing" the rational numbers–essentially inserting all the missing suprema in a process known as Dedekind cuts–but there are also many other ordered fields that are quite unlike the rational numbers. These "non-Archimedean" fields contain the integers, just like the rational numbers, but the integers are bounded! However, as we have mentioned previously there is one and only one complete ordered field, namely $\mathbb{R}$.

The following lemma formalizes what the reader may already have instinctively realized about a finite supremum $s$ of a set $A$: that $s$ is an upper bound of $A$ and there are elements of $A$ that are arbitrarily close to $s$. This "analytical description" of the supremum, and an analogous description of the infimum that we will not state, is extremely useful.

**Lemma 12** *(Approximation Property for the Supremum) Let $A$ be a nonempty set of real numbers such that $\sup A < \infty$. Then $s = \sup A$ if and only if $s$ is an upper bound of $A$ and for every $\varepsilon > 0$ there exists some $x \in A$ such that $x > s - \varepsilon$.*

**Proof.** First suppose that $s = \sup A$. Then by definition $s$ is an upper bound of $A$. If the second statement is false then there exists some $\varepsilon > 0$ such that for all $x \in A$, $x \leq s - \varepsilon$. But then $s - \varepsilon$ is an upper bound of $A$ and $s - \varepsilon < s = \sup A$, a contradiction. Conversely, assume that $s$ is an upper bound of $A$ and for every $\varepsilon > 0$ there exists some $x \in A$ such that $x > s - \varepsilon$. If $s \neq \sup A$ then by definition there exists some upper bound $M$ of $A$ such that $M < s$. Let $\varepsilon := s - M$. Then there is some $x \in A$ such that $x > s - \varepsilon = s - (s - M) = M$, which contradicts the fact that $M$ is an upper bound of $A$. ■

Our first application of the Approximation Property is the following statement, the proof of which uses of Lemma 12 four times. The reader is invited to try to come up with a proof that does not use Lemma 12 or some version of its proof (several times)!

**Lemma 13** *Let $\{A_\lambda\}_{\lambda\in\Lambda}$ be a collection of nonempty sets of real numbers. Then*

$$\sup\left\{\bigcup_{\lambda\in\Lambda} A_\lambda\right\} = \sup\left\{\sup A_\lambda\right\}_{\lambda\in\Lambda}.$$

**Proof.** Let $s := \sup\left\{\sup A_\lambda\right\}_{\lambda\in\Lambda}$. Suppose first that $s = \infty$. If $\infty \notin \left\{\sup A_\lambda\right\}_{\lambda\in\Lambda}$ then for every $M \geq 0$ there exists some $\lambda \in \Lambda$ such that $\sup A_\lambda \geq$

$M + \frac{1}{2}$. By Lemma 12 there exists some $x \in A_\lambda$ such that $x \geq M$. Since $x \in \bigcup_{\lambda\in\Lambda} A_\lambda$ this proves that $\bigcup_{\lambda\in\Lambda} A_\lambda$ is not bounded above, so $\sup\left\{\bigcup_{\lambda\in\Lambda} A_\lambda\right\} = \infty$. On the other hand, if $\infty = \sup A_\lambda$ for some $\lambda \in \Lambda$ then for any $M \geq 0$ there is some $x \in A_\lambda$ such that $x \geq M$, and the proof follows as in the previous case.

If $s < \infty$, let $x \in \bigcup_{\lambda\in\Lambda} A_\lambda$. Then $x \in A_\lambda$ for some $\lambda$, and so $x \leq \sup A_\lambda \leq \sup\{\sup A_\lambda\}_{\lambda\in\Lambda}$. Therefore $s := \sup\{\sup A_\lambda\}_{\lambda\in\Lambda}$ is an upper bound for $\bigcup_{\lambda\in\Lambda} A_\lambda$. Let $\varepsilon > 0$; then $\varepsilon/2 > 0$. By the Lemma 12 there exists some $x \in \{\sup A_\lambda\}_{\lambda\in\Lambda}$, i.e. $x = \sup A_\lambda$ for some $\lambda \in \Lambda$, such that $x > s - \varepsilon/2$. Now by Lemma 12 there exists some $y \in A_\lambda$ such that

$$y > x - \varepsilon/2 > s - \varepsilon/2 - \varepsilon/2 = s - \varepsilon.$$

Since $y \in \bigcup_{\lambda\in\Lambda} A_\lambda$ it now follows from Lemma 12 that $s = \sup\left\{\bigcup_{\lambda\in\Lambda} A_\lambda\right\}$. ■

**Exercise 19** *Let* $\{A_\lambda\}_{\lambda\in\Lambda}$ *be a collection of sets of real numbers such that* $\bigcap_{\lambda\in\Lambda} A_\lambda \neq \varnothing$.

1. *Show that* $\sup\left\{\bigcap_{\lambda\in\Lambda} A_\lambda\right\} \leq \inf\{\sup A_\lambda\}_{\lambda\in\Lambda}$ *and give an example to show that these two numbers may not be equal.*
2. *Prove or disprove the following statement. If each* $A_\lambda$ *has a maximum then* $\bigcup_{\lambda\in\Lambda} A_\lambda$ *has a maxumum and* $\max\left\{\bigcup_{\lambda\in\Lambda} A_\lambda\right\} = \max\{\max A_\lambda\}_{\lambda\in\Lambda}$.

**Exercise 20** *Write down, but do not prove, the corresponding statements for infima to Lemmas 12 and 13, as well as the first part of Exercise 19.*

Let $X$ be a set, $f : X \to \mathbb{R}$ be a function, and $A \subset X$. Then $f$ is said to be *bounded* on $A$ if the set $f(A)$ is bounded. Define $\sup_A f := \sup f(A)$. If $f(A)$ has a maximum then we denote the maximum by $\max_A f$. When $X = A$ we may simply use the terms $\sup f$ and $\max f$.

**Exercise 21** *Prove or disprove the following statements:*

1. *If* $f(x) < g(x)$ *for all* $x \in A$ *then* $\sup_A f < \sup_A g$.
2. *If* $f(x) \leq g(x)$ *for all* $x \in A$ *then* $\sup_A f \leq \sup_A g$.
3. *If* $f$ *and* $g$ *have a maximum on* $A$ *and* $f(x) < g(x)$ *for all* $x \in A$ *then* $\max_A f < \max_A g$.

If $f, g : A \to \mathbb{R}$ are functions, we define $f + g : A \to \mathbb{R}$ by $(f + g)(x) = f(x) + g(x)$.

**Lemma 14** *Let* $X$ *be a set,* $f, g : X \to \mathbb{R}$ *be functions, and* $A \subset X$. *Then* $\sup_A(f + g) \leq \sup_A f + \sup_A g$.

**Proof.** If $\sup_A f$ or $\sup_A g$ is $\infty$ then it doesn't matter whether $\sup_A(f+g)$ is finite or infinite; the inequality still holds. Therefore we can suppose that $\sup_A f$ and $\sup_A g$ are finite; i.e., $f$ and $g$ are bounded. Then $f + g$ is also bounded. By definition of supremum we need only show that $\sup_A f + \sup_A g$ is an upper bound of $\{f(x) + g(x) : x \in A\}$. Suppose $x \in A$. Then $f(x) \leq \sup_A f$ and $g(x) \leq \sup_A g$, so $f(x) + g(x) \leq \sup_A f + \sup_A g$. ■

**Exercise 22** *State and prove the analog of Lemma 14 for infima and show by example that* $\sup_A(f+g) < \sup_A f + \sup_A g$ *can occur even if all quantities are finite.*

**Proposition 15** *For any* $a > 0$*, there is a unique real number* $b > 0$ *such that* $b^2 = a$*.*

**Proof.** First suppose that $a > 1$ and consider the set $A := \{x > 0 : x^2 < a\}$. Then $1 \in A \neq \varnothing$. If $x \geq a > 1$ then $x^2 > a^2 > a$; in particular, $A$ is bounded above and $s := \sup A$ satisfies $1 \leq s < a$. Suppose $s^2 > a$. Setting $\varepsilon := \frac{a-s}{2} > 0$ we have that

$$(s-\varepsilon)^2 = s^2 - 2s\varepsilon + \varepsilon^2 > s^2 - 2s(\frac{a-s}{2}) = sa \geq a.$$

Since $s - \varepsilon < s$, this contradicts $s = \sup A$.

Now suppose $s^2 < a$. Let $\varepsilon > 0$ be less than 1 and $\frac{a-s^2}{2s+1}$. In particular, $\varepsilon^2 < \varepsilon$. Now

$$(s+\varepsilon)^2 = s^2 + 2s\varepsilon + \varepsilon^2 < s^2 + \varepsilon(2s+1) < a^2$$

which also contradicts $s = \sup A$. Therefore $s^2 = a$. Uniqueness follows from the field and order axioms; for example if $0 < x < s$ then $x^2 < s^2 = a$. If $a = 1$ there is nothing to prove, and if $a < 1$ then by what we proved above there is some $c$ such that $c^2 = \frac{1}{a}$. Let $b = \frac{1}{c}$ and apply the field axioms to see that $b^2 = a$. ∎

As usual we will denote the number $b$ in the above proposition by $\sqrt{a}$ or $a^{1/2}$. It follows from the field and order axioms that if $a < b$ then $\sqrt{a} < \sqrt{b}$. With more effort one can prove the existence of $n^{th}$ roots of positive numbers, and use this to define rational powers of positive numbers, followed by very tedious checking to verify the powers rule. However, we will have no need for these more general powers until much later in the text. Arbitrary real powers of positive numbers can be quite easily defined using the exponential function, once we have the mathematical machinary built to understand it (see the end of Section 3.8).

We conclude with what may seem at first like a surprising consequence of completeness, namely the *Archimedean Principle*.

**Theorem 16** *For every positive* $a, b \in \mathbb{R}$ *there exists a natural number* $n$ *such that* $na > b$*.*

**Proof.** If $a > b$ then we are done, letting $n = 1$. Otherwise, the set $E$ of all $n \in \mathbb{N}$ such that $na \leq b$ contains 1 and is therefore non-empty. In addition, any $n \in E$ satisfies $n \leq \frac{b}{a}$, and so $E$ is bounded above. Therefore $s := \sup E$ exists and is finite. By the approximation property there exists some $m \in E$ such that $m > \sup E - 1$. Now $n := m + 1$ is a natural number that is greater than $\sup E$, which implies that $n \notin E$. But by definition that means $na > b$, which is what we wanted to prove. ∎

**Corollary 17** *The natural numbers are not bounded above.*

Frequently beginning mathematics students, if given the problem of proving that the natural numbers are not bounded above, will attempt to prove this "obvious" fact without using the completeness axiom. All such attempts are doomed to fail because there exist so-called non-Archimedean ordered fields, which satisfy all the real number axioms except completeness, but for which the Archimedean principle and unboundedness of the natural numbers fail to be true.

**Exercise 23** *Prove that for any two real numbers $r_1 < r_2$ there exits a rational number $\frac{p}{q}$ such that $r_1 < \frac{p}{q} < r_2$. Hint: For $q$ large enough, $q(r_2 - r_1) > 1$.*

## 1.6 Sequences and subsequences

**Definition 18** *If $A$ is a set, a sequence in $A$ is a function $a : \mathbb{N} \to A$.*

We use the following notation for sequences, denoting $a(n)$ by $a_n$ and indicating an entire sequence by $(a_1, a_2, ...)$ or $(a_n)_{n=1}^{\infty}$, or simply $(a_n)$ if no confusion will result. We use this notation to distinguish between the sequence $(a_n)$ and the image of the sequence, which is the set $\{a_n\}$. In a sequence the order is essential and duplicate values are listed. For example the sequence $(-1, 1, -1, 1, ...)$ has image set $\{1, -1\}$. While we have quickly abandoned the "function notation" for sequences, the fact that they are functions allows us to use all definitions and theorems that we have previously been given concerning functions. For example, we know automatically what it means for a sequence to be bounded since we know what it means for a function to be bounded.

The number $a_n$ is called the $n^{th}$ *term* of the sequence. We will not be strict about the requirement that the domain of a sequence be $\mathbb{N}$. In fact we will frequently consider sequences starting with the $0^{th}$ term, or the $k^{th}$ term for some $k > 1$.

**Example 19** *The simplest sequence is a constant sequence defined by $a_n := a$ for some fixed $a \in A$ and all $n \in \mathbb{N}$. Note that this sequence still has infinitely many terms even though they all have the same value.*

Sequences are often defined *iteratively.* One is given a "starting term", often $s_0$ or $s_1$, and a procedure for determining $s_n$ with $n > 1$ based on $s_1, ..., s_{n-1}$. For example, one may let $s_1 := 1$, $s_2 := 1$ and let $s_n := s_{n-2} + s_{n-1}$ for $n > 2$. We have $s_3 := 1 + 1 = 2$, $s_4 = 1 + 2 = 3$,$s_5 := 2 + 3 = 5$ and so on. This sequence is called the sequence of Fibonacci numbers. As another example, let $s_1 := 1$ and $s_n := s_{n-1} + n$ for $n > 1$. That is, $s_n = 1 + 2 + ... + n = \frac{n(n+1)}{2}$ (the latter formula is easily proved by induction). The formula $s_n = \frac{n(n+1)}{2}$ is called the "closed form" for the sequence; it allows one to compute $s_n$ directly from $n$ without using the values of $s_1, ..., s_{n-1}$.

**Exercise 24** *Define a sequence iteratively by $s_1 := \sqrt{3}$ and $s_n := \sqrt{3 + \sqrt{s_{n-1}}}$.*

1. *Write out decimal expansions for the first 4 terms of this sequence.*

2. *Prove that* $s_n < 3$ *for all* $n$.

3. *Prove that* $s_n > s_{n-1}$ *for all* $n > 1$.

**Example 20** *Consider the sequence* $(\frac{1}{n})_{n=1}^{\infty}$. *This sequence has the property that for every* $\varepsilon > 0$ *there exists some natural number* $N$ *such that if* $n \geq N$ *then* $\frac{1}{n} < \varepsilon$. *In fact, for any* $\varepsilon > 0$, $\frac{1}{\varepsilon}$ *is also positive. Applying the Archimedean principle to the numbers* $1$ *and* $\frac{1}{\varepsilon}$, *there exists some* $N \in \mathbb{N}$ *such that* $N = N \cdot 1 > \frac{1}{\varepsilon}$. *But then if* $n \geq N$ *then* $n > \frac{1}{\varepsilon}$ *and* $\frac{1}{n} < \varepsilon$.

**Exercise 25** *Show that the sequence* $(\frac{n+1}{n})_{n=1}^{\infty}$ *has the following property: For every* $\varepsilon > 0$ *there exists some* $N \in \mathbb{N}$ *such that if* $n \geq N$ *then* $\left|\frac{n+1}{n} - 1\right| < \varepsilon$. *You may not use theorems from calculus that have not been proved in this text so far!*

***Example 21*** *Let* $A$ *be an unbounded subset of* $\mathbb{R}$. *Then there is a sequence* $(a_i)$ *in* $A$ *such that* $|a_i| \geq i$ *for all* $i \in \mathbb{N}$. *In fact, since* $A$ *is unbounded, for any* $i$ *there exists some* $x \in A$ *such that* $|x| \geq i$. *Let* $a_i := x$. *Then* $(a_i)_{i=1}^{\infty}$ *is certainly a sequence and satisfies* $|a_i| \geq i$ *by construction. In the future, for constructions of this sort we will eliminate the intermediate naming of the point* $x$. *In this particular case we would shorten the above argument to: "...by definition for any* $i$ *there exists some* $a_i \in A$ *such that* $|a_i| \geq i$*..."*

A *subsequence* $(a_{n_k})_{k=1}^{\infty}$ of $(a_n)_{n=1}^{\infty}$ is the restriction of the function $(a_n)_{n=1}^{\infty}$ to an infinite subset $\{n_1, n_2, ...\} = \{n_k\}_{k=1}^{\infty}$ of the natural numbers such that $n_1 < n_2 < ....$. So a subsequence is an *infinite* selection of terms of the original sequence. Note that every sequence is a subsequence of itself: let $n_1 := 1, n_2 := 2$, etc. In addition, a subsequence is a sequence in its own right; the $k^{th}$ term of the subsequence is the element $a_{n_k}$ in $A$. We can also take a subsequence of a subsequence (which one might call a subsubsequence!), which is of the form $(a_{n_{k_j}})_{j=1}^{\infty}$. One may continue this process indefinitely.

**Example 22** *Consider the sequence* $(\frac{1}{n})_{n=1}^{\infty} = (1, \frac{1}{2}, \frac{1}{3}, ...)$. *If we let* $n_1 := 2$, $n_2 := 4$, $n_3 := 6$, *and in general* $n_k := 2k$, *then the corresponding subsequence is* $(\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, ...)$. *If we define* $n_{k_j} := 2n_k = 4k$ *then we have defined the subsubsequence* $(\frac{1}{4}, \frac{1}{8}, \frac{1}{12}, ...)$.

**Example 23** *The sequence* $(-1, 1, -1, 1, ...)$ *has a subsequence* $(-1, -1, ...)$ *(odd terms) and a subsequence* $(1, 1, 1, ...)$ *(even terms).*

**Example 24** *Suppose that* $a_n \geq n$ *for every* $n \in \mathbb{N}$. *Then* $(a_n)$ *has no bounded subsequence. Suppose that* $(a_{n_k})$ *were a bounded above by* $M$. *By the archimedian principle there is some* $n$ *such that* $a_n \geq n \geq M$. *Since the set of* $\{n_k\}$ *is infinite there is some* $n_k > n \geq M$ *and therefore* $a_{n_k} > M$, *a contradition.*

Iterative constructions are not only used to construct concrete sequences. In a proof an iterative construction can be used to construct a sequence $(x_n)$ such that each $x_n$ has some property $P(n)$ in a process very reminiscent of mathematical induction. We first show how to obtain $x_1$ satisfying $P(1)$ and then suppose that we have constructed $x_1, ..., x_m$ for some $m \leq 1$ such that $x_n$ has property $P(n)$ for all $1 \leq n \leq m$. We then show how to construct $x_{m+1}$ having property $P(m+1)$ and it follows by induction that there exists a sequence $(x_n)_{n=1}^{\infty}$ such that $x_n$ has property $P(n)$ for all $n$. We will illustrate this process in proofs below.

**Lemma 25** *If $(a_i)_{i=1}^{\infty}$ is an unbounded sequence of real numbers then there is a subsequence $(a_{i_k})_{k=1}^{\infty}$ such that $|a_{i_k}| \geq k$ for all $k \in \mathbb{N}$.*

**Proof.** To prove this we must proceed more carefully than in Example 21. We cannot simply construct our sequence by stating that for every $k$ there exists some $a_{i_k}$ such that $|a_{i_k}| \geq k$. This is a true statement, of course, but we are ignoring the requirement that $i_1 < i_2 < i_3 < \cdots$. While the indices $i_k$ must eventually get large, without more effort it is impossible to be sure that they are chosen in the proper order. To make sure of this we will proceed iteratively. Since $\{a_i\}$ is unbounded there exists some $a_n$ such that $|a_n| \geq 1$. Let $i_1 := n$. Before doing the iterative step we will consider the construction of $a_{i_2}$. The problem is that, for example, it might also be true that $|a_n| \geq 2$, but we certainly don't want to choose $i_2 = n = i_1$. To be sure that $i_2 > i_1$ we do the following. Let $M_1 := \max_{j \leq i_1}\{|a_j|\} + 1$ and let $M_2 := \max\{M_1, 2\}$. Since $\{a_i\}$ is unbounded there exists some $a_m$ such that $|a_m| \geq M_2$. Since $|a_m| \geq M_2 > \max_{j \leq i_1}\{|a_j|\}$, we know that $m \neq j$ for any $j \leq i_1$ and therefore $m > i_1$. Let $i_2 := m$. Then $i_2 > i_1$ and by construction $|a_{i_2}| \geq M_2 \geq 2$. To finish the proof suppose that we have chosen $i_1 < \cdots < i_k$ such that for all $1 \leq j \leq k$ we have $\left|a_{i_j}\right| \geq j$. Let $M_3 := \max_{j \leq i_k}\{|a_j|\}$ and $M_4 := \max\{M_3, k\} + 1$. Then as in the above proof of $k = 2$ there exists some $a_{i_{k+1}}$ such that $\left|a_{i_{k+1}}\right| \geq M_4$ and hence $i_{k+1} > i_k$ and $\left|a_{i_{k+1}}\right| \geq k + 1$. This completes the iterative construction of the subsequence $(a_{i_k})$. ■

**Exercise 26** *Show that if $(a_n)$ is a real sequence and every subsequence of $(a_n)$ has a bounded subsequence then $(a_n)$ is bounded.*

**Lemma 26** *Let $(a_i)$ be a sequence in a set $A$ and let $B \subset A$. If $B$ contains infinitely many terms of $(a_i)$ then there is some subsequence $(a_{i_k})$ of $(a_i)$ such that $a_{i_k} \in B$ for all $k$.*

**Proof.** Let $a_{i_1}$ be any term in $B$. Now suppose we have found $i_1 < \cdots < i_k$ such that $a_{i_j} \in B$ for all $1 \leq j \leq k$. Since $B$ contains infinitely many terms there must be some $a_{i_{k+1}} \in B$ with $i_{k+1} > i_k$. This completes the iterative construction. ■

**Corollary 27** *Let $(a_i)$ be a sequence in a set $A$ and suppose that $A = \cup_{j=1}^{k} A_j$. Then there is some $j$ such that a subsequence of $(a_i)$ lies entirely in $A_j$.*

**Proof.** Since there are only finitely many sets $A_j$, at least one of them must contain infinitely many terms of the sequence and hence must contain a subsequence of $(a_i)$ by Lemma 26. ■

Given a sequence $(a_i)_{i=1}^{\infty}$ there is a special kind of subsequence of $(a_i)$ called a *tail*, namely a subsequence $(a_k, a_{k+1}, ...) = (a_i)_{k=1}^{\infty}$ for some $k \geq 1$. We will sometimes use the notation $k$-tail to specify the index of the first term. So the 1-tail of $(a_i)$ is simply the entire sequence $(a_i)$. Although tails are subsequences, we will rarely use double subscripts to index them. For example, strictly speaking, the 3-tail of a sequence is constructed using $i_1 := 3$, $i_2 := 4$, and in general $i_k := k + 2$, i.e., the first term in the tail is $a_3$, the second is $a_4$, and so on. But it is usually easier to just use the original indices starting at $k$ instead of 1.

**Exercise 27** *Let $X = A \cup B$ and suppose $(a_i)$ is a sequence in $X$. Prove or disprove:*

1. *There is some tail of $(a_i)$ that lies entirely in $A$ or entirely in $B$.*
2. *If a tail of $(a_i)$ lies entirely in $A$ then at most finitely many terms of $(a_i)$ lie in $B$.*

# Chapter 2

# Metric spaces

In this chapter we will introduce a class of abstract mathematical objects called metric spaces. The reader who is familiar with abstract algebra will recognize the general idea. We suppose that we have a set satisfying some basic axioms, then prove theorems using those axioms. At the same time we will discuss specific examples, such as the real numbers and Euclidean spaces, and find some applications in these special cases.

## 2.1 Basic Definitions and Examples

**Definition 28** *A metric space consists of a pair $(X, d)$, where $X$ is a set and $d : X \times X \to \mathbb{R}$ is a function, called the metric or distance function, such that the following hold for all $x, y, z \in X$*

1. *(Symmetry)* $d(x, y) = d(y, x)$
2. *(Positive Definiteness)* $d(x, y) \geq 0$, *and* $d(x, y) = 0$ *if and only if* $x = y$
3. *(Triangle Inequality)* $d(x, z) \leq d(x, y) + d(y, z)$

The letter $d$ is the default for metric spaces; often we will simply state "let $X$ be a metric space" rather than specifying $(X, d)$. In addition, we will sometimes even use $d$ to represent distances in different metric spaces in the same discussion; when there is danger of confusion we will be more careful, e.g. using $d_X$ for the distance on $X$ and $d_Y$ for the distance on $Y$. We will often refer to elements of a metric space as "points".

The definition of a "metric" captures the most important and basic elements of what a "distance" should be. The distance from $x$ to $y$ should be the same as that from $y$ to $x$; different points should be at positive distance from one another, but a point should be at distance 0 from itself, and travelling between two points via an arbitrary third point should not be shorter than the distance between the original two. Many metrics have stronger properties–for example, between every two distinct points in the real line (or Euclidean space) there is a unique midpoint. However, the theory of spaces satisfying only these three axioms is so powerful and pervasive in mathematics that it is worthwhile to study them in their own right. In addition, nearly all of the proofs in this chapter would be no simpler (and would sometimes be notationally more complicated) if we proved them in the very restricted setting of Euclidean spaces. Therefore nothing is lost and much is gained by proceeding in this very general setting. We begin with a familiar example:

**Example 29** *We define the distance between points $x, y \in \mathbb{R}$ by $d(x, y) = |x - y|$. Let's check that this makes $\mathbb{R}$ a metric space. Symmetry follows from the definition: $d(x, y) = |x - y| = |y - x| = d(y, x)$. We also know from the definition of absolute value that $|x - y| \geq 0$. For the second part of positive definiteness note that by the definitions of distance and absolute value,*

$$d(x, y) = 0 \Leftrightarrow |x - y| = 0 \Leftrightarrow x - y = 0 \Leftrightarrow x = y.$$

*For the triangle inequality we use the "other" triangle inequality (Proposition 7):*

$$d(x, z) = |x - z| = |(x - y) + (y - z)| \leq |x - y| + |y - z| = d(x, y) + d(y, z)$$

**Exercise 28** *Use the triangle inequality for the distance to prove the triangle inequality for the absolute value. Hint: Use the fact that $x + y = x - (-y)$. This, together with the proof in Example 29 shows that, for the real numbers, the two triangle inequalities are logically equivalent.*

**Example 30** *Let $X$ be any set. For all $x, y \in X$, define $d(x, y) = 1$ if $x$ and $y$ are distinct (i.e. different), and $d(x, x) = 0$. It is easy to check that $d$ is a metric, called the trivial or discrete metric. This metric is of little interest except as a very simple example and counterexample to certain statements. It is also useful to help strip away any preconceived notions about metric spaces that one might have picked up from working with the real numbers or the plane. For example, there are no "midpoints" in a discrete metric space.*

**Example 31** *Let $\mathcal{C}$ be the set of continuous functions $f : [0, 1] \to \mathbb{R}$ and define $d(f, g) := \max_{[0,1]}\{|f(x) - g(x)|\}$. We will check that $(\mathcal{C}, d)$ is a metric space using some facts from elementary calculus that we will later prove in greater generality. First, since $f$ and $g$ are continuous, so is $|f - g|$, and therefore $|f - g|$ does have a maximum on the closed bounded interval $[0, 1]$, and the maximum is clearly non-negative. If $f \neq g$ then for some $x_0$, $f(x_0) \neq g(x_0)$, which means $|f(x_0) - g(x_0)| > 0$. But then*

$$\max_{[0,1]}\{|f(x) - g(x)|\} \geq |f(x_0) - g(x_0)| > 0.$$

*Since $d(f, f)$ is clearly 0, we have proved positive definiteness. Symmetry is an immediate consequence of the definition. To prove the triangle inequality, note that for $f, g, h \in \mathcal{C}$ we have*

$$d(f, h) = \max_{[0,1]}\{|f(x) - h(x)|\} = \max_{[0,1]}\{|f(x) - g(x) + g(x) - h(x)|\}$$

$$\leq \max_{[0,1]}\{|f(x)-g(x)|+|g(x)-h(x)|\}$$

$$\leq \max_{[0,1]}\{|f(x)-g(x)|\}+\max_{[0,1]}\{|g(x)-h(x)|\}=d(f,g)+d(g,h)$$

*by Lemma 14.*

**Exercise 29** *Let $f(x) = x^2$ and $g(x) = x$, both defined on $[0,1]$. Find two different continuous functions that are "midpoints" between $f$ and $g$; a midpoint is a function $h$ such that $d(f,h) = d(g,h) = \frac{1}{2}d(f,g)$ (there are actually infinitely many midpoint functions). For this exercise you may use your knowlege about continuous real functions from calculus.*

**Exercise 30** *Let $f$ be a function that has derivatives of all orders at all points. Suppose that there exists some $M > 0$ such that every $n^{th}$ derivative $f^{(n)}$ satisfies $\left|f^{(n)}(x)\right| \leq M$ for all $x \in [0,1]$. Use Maclaurin's formula with remainder to prove that for every $\varepsilon > 0$ there is a polynomial $g$ such that $d(g,f) < \varepsilon$. You may use facts that you have learned in basic calculus.*

**Definition 32** *Let $x$ be a point in a metric space $X$, and $r > 0$. We define the (open) ball centered at $x$ of radius $r$ to be $B(x,r) := \{y \in X : d(x,y) < r\}$.*

For the metric space $\mathbb{R}$, we have

$$B(x,r) = \{y \in X : d(x,y) < r\}$$

$$= \{y \in X : |x-y| < r\} = \{y \in X : x-r < y < x+r\}$$

which is simply the open interval $(x-r, x+r)$ of length $2r$ centered at $x$. (An interval doesn't "look" like a ball, but such preconceived images need to be abandoned–or replaced with other intuitive images.) Conversely, given any bounded open interval $(a,b)$ in $\mathbb{R}$,

$$(a,b) = \left(\frac{a+b}{2} - \frac{b-a}{2}, \frac{a+b}{2} + \frac{b-a}{2}\right)$$

and so $(a,b) = B(x,r)$, where $x = \frac{a+b}{2}$ and $r = \frac{b-a}{2}$.

**Exercise 31** *Let $x$ be a point in a metric space $X$, and suppose $r > r_0 > 0$. Show that $B(x,r_0) \subset B(x,r)$.*

**Exercise 32** *Let $x, y$ be distinct points in a metric space $X$, and let $r = d(x,y)$. Use the triangle inequality to prove by contradiction that $B(x,r/2)$ and $B(y,r/2)$ are disjoint.*

**Exercise 33** *Let $X$ be a non-empty set with the trivial metric (see Example 30), and let $x \in X$. Describe the following: $B(x,2), B(x,\frac{1}{2}), B(x,1)$.*

**Definition 33** *A subset $A$ of a metric space $X$ is said to be bounded if there exists some $x \in X$ and $r > 0$ such that $A \subset B(x,r)$.*

Note that, in the special case when the metric space in question is $\mathbb{R}$, our new definition of "bounded" is equivalent to our old one. In fact,we have already noted that open balls in $\mathbb{R}$ are simply open intervals; therefore $A \subset \mathbb{R}$ is bounded if and only if $A \subset (a, b)$ for some open interval $(a, b)$, which is equivalent to $a < x < b$ for some $a, b \in \mathbb{R}$ and all $x \in A$, which is equivalent to $A$ being bounded in the sense of Definition 9. However, in a general metric space $X$, "$x < b$" makes no sense, so a new definition is required.

**Exercise 34** *Show that a nonempty set $A \subset X$ is bounded if and only if there exists some $r > 0$ such that for any $x, y \in A$, $d(x, y) < r$.*

**Exercise 35** *Show that if $A \subset X$ is bounded and nonempty then for any $x \in X$ there exists an $r > 0$ such that $A \subset B(x, r)$.*

## 2.2 Open and Closed Sets

**Definition 34** *A subset $A$ of a metric space $X$ is called open if for every $x \in A$ there exists some $r > 0$ such that $B(x, r) \subset A$. A subset $C$ of $X$ is called closed if $C^c$ is open.*

**Exercise 36** *Prove that, for a metric space $X$, $X$ itself and the empty set are subsets of $X$ that are both open and closed.*

In other words, unlike a door, a set can be both open and closed. As we will see later, sets can also be neither. We will frequently use the next lemma, the proof of which is an exercise. Since Definition 34 is logically equivalent to the second statement in the lemma, we could have used either of these for our definition; some texts use the statement of Lemma 35 as the starting point. We will use this lemma so often that we will often not refer to it by number.

**Lemma 35** *A subset $A$ of a metric space $X$ is open if and only if for every $x \in A$ there exists some $\varepsilon > 0$ such that if $d(x, y) < \varepsilon$ then $y \in A$.*

**Exercise 37** *Prove Lemma 35.*

**Lemma 36** *If $X$ is a metric space and $x \in X$ then $\{x\}$ is closed.*

**Proof.** We need to show that $\{x\}^c$ is open. If $y \in \{x\}^c$ then $y \neq x$. By the positive definiteness of the metric, $d(y, x) = r$ for some $r > 0$. But then $x \notin B(y, r)$, and hence $B(y, r) \subset \{x\}^c$. This proves that $\{x\}^c$ is open and so $\{x\}$ is closed. ■

A set containing a single point is often called a *singleton set*. Note that the number $r > 0$ in Lemma 35 may depend on $x$–that is, if $y$ is closer to $x$ then we need to use a smaller $r$. Similarly, it is not completely trivial that $B(x, r)$, which we have already named an "open ball", is in fact open according to our definition. That is, every $y \in B(x, r)$ is contained in an open ball centered at $x$, namely $B(x, r)$, but is every $y \in B(x, r)$ contained in an open ball centered at

$y$ that is contained in $B(x, r)$? This is what we need to show in order to verify that $B(x, r)$ is open. We will check this now. Let $y \in B(x, r)$. By definition, $d(x, y) < r$. Let $\varepsilon := r - d(x, y) > 0$ and suppose that $d(z, y) < \varepsilon$. Then by the triangle inequality

$$d(z, x) \leq d(x, y) + d(y, z) < d(x, y) + (r - d(x, y)) = r,$$

and by definition $z \in B(x, r)$. That is, $B(y, \varepsilon) \subset B(x, r)$ and $B(x, r)$ is an open set. Therefore "open ball" is an appropriate name.

**Example 37** *From the preceding discussion we know that in* $\mathbb{R}$ *open intervals, which we have already observed are open balls, are open sets. Are closed intervals closed according to the above definition? Let* $a \leq b$*. We need to show that* $[a, b]^c$ *is open. Let* $x \notin [a, b]$ *and suppose first that* $x < a$*. Define* $r := a - x > 0$*. If* $y \in B(x, r)$*,* $y < x + r < a$*, hence* $y \notin [a, b]$*. That is,* $B(x, r) \cap [a, b] = \varnothing$*, or* $B(x, r) \subset [a, b]^c$*. A similar argument holds if* $x > b$*.*

***Exercise 38*** *For* $x$ *an element of a metric space* $X$ *and* $r > 0$*, define the closed ball of radius* $r$ *centered at* $x$ *by* $C(x, r) := \{y \in X : d(x, y) \leq r\}$*. Prove that* $C(x, r)$ *is a closed set. The preceding example about closed intervals is a special case of this statement.*

**Proposition 38** *Let* $X$ *be a metric space and* $\{A_\lambda\}_{\lambda \in \Lambda}$ *be a collection of open sets in* $X$*. Then* $\bigcup_{\lambda \in \Lambda} A_\lambda$ *is open in* $X$*.*

**Proof.** Let $x \in \bigcup_{\lambda \in \Lambda} A_\lambda$. Then $x \in A_\lambda$ for some $\lambda$. Since $A_\lambda$ is open there exists some $r > 0$ such that $B(x, r) \subset A_\lambda \subset \bigcup_{\lambda \in \Lambda} A_\lambda$. Hence $\bigcup_{\lambda \in \Lambda} A_\lambda$ is open. ∎

Conversely, every open set $A$ in $X$ is a union of open sets–open balls, in fact. To see this, note that by definition, for each $x \in A$ there is some $r_x > 0$ and $A_x := B(x, r_x)$ such that $A_x \subset A$. Clearly $A = \bigcup_{x \in A} A_x$.

**Exercise 39** *Let* $X$ *be a set with the trivial metric. Show that every subset of* $X$ *is both open and closed. Hint: show that every subset* $\{x\}$ *with* $x \in X$ *is open.*

**Proposition 39** *If* $A_1, ..., A_n$ *are open sets in a metric space* $X$ *then* $\cap_{i=1}^n A_i$ *is open.*

**Proof.** Let $x \in \cap_{i=1}^n A_i$. For each $i$ there exists some $\varepsilon_i > 0$ such that if $d(x, y) < \varepsilon_i$ then $y \in A_i$. Then $\varepsilon := \min\{\varepsilon_1, ..., \varepsilon_n\}$ is positive. If $d(x, y) < \varepsilon$ then $d(x, y) < \varepsilon_i$ for all $i$, and hence $y \in A_i$ for all $i$. By definition, $y \in \cap_{i=1}^n A_i$. ∎

**Example 40** *The above proof fails for infinitely many sets* $A_i$*, since infinite sets may not have minima–and even if we took the infimum of some infinite collection of epsilons, the infimum could well be* $0$*. To see concretely that Proposition 39 is only valid for finitely many sets, consider the collection of open intervals* $\{(-1/n, 1/n)\}_{n=1}^\infty$*. The intersection of this collection is* $\{0\}$*, which we already know is closed. But it is also not open, because it is nonempty and does not contain any open interval at all!*

**Exercise 40** *Use de Morgan's laws to show that the intersection of any collection of closed sets is closed, and the union of finitely many closed sets is closed.*

**Definition 41** *Let $A$ be a subset of a metric space $X$. The closure $\overline{A}$ of $A$ is defined to be the set of all $x \in X$ such that for every $r > 0$, $B(x,r) \cap A \neq \varnothing$.*

Put another way, $\overline{A}$ is the set of all points $x$ in $X$ such that there are points in $A$ that are arbitrarily close to $x$. Note that certainly $A \subset \overline{A}$ since for every $r > 0$, if $x \in A$ then $x \in B(x,r) \cap A$. The next lemma has a satisfying ring to it. It says that closed sets are precisely those that are equal to their closures.

**Lemma 42** *$A$ is a closed subset of a metric space $X$ if and only if $\overline{A} = A$.*

**Proof.** Suppose $A$ is closed. Since $A \subset \overline{A}$ we need only show the opposite inclusion, which we will prove by contrapositive: Suppose $x \notin A$. Since $A^c$ is open there exists an $r_0 > 0$ such that $B(x, r_0) \cap A = \varnothing$. But then by definition $x \notin \overline{A}$ and $\overline{A} \subset A$ is proved.

To prove the converse, suppose that $\overline{A} = A$ and let $x \in A^c = \overline{A}^c$. By definition of closure there exists some $r > 0$ such that $B(x,r) \subset \overline{A}^c = A^c$ and $A^c$ is open; hence $A$ is closed. ■

**Remark 43** *Since $A$ is always a subset of $\overline{A}$, a very useful strategy to prove that a set $A$ is closed is to choose a point $x \in \overline{A}$ and show that $x \in A$.*

**Exercise 41** *Let $A$ be a subset of a metric space $X$.*

1. *Show that if $C$ is any closed set in $X$ containing $A$ then $\overline{A} \subset C$.*
2. *Show that $\overline{A}$ is the intersection of all closed sets containing $A$. In other words, the closure of $A$ is the "smallest" closed set containing $A$.*

**Exercise 42** *Let $X$ be a trivial metric space with at least two points. Show that $\overline{B(x,1)} \subsetneq C(x,1)$; that is the closure of an open ball may not be equal to the closed ball of the same radius.*

**Definition 44** *Let $A$ be a subset of a metric space $X$. The interior $\dot{A}$ of $A$ is defined to be the set of all $x \in A$ such that for some $r > 0$, $B(x,r) \subset A$.*

**Lemma 45** *$A$ is an open subset of a metric space $X$ if and only if $\dot{A} = A$.*

**Exercise 43** *Prove the above lemma.*

**Exercise 44** *State and prove two statements for interior that are analogous to those for closure in Exercise 41.*

**Exercise 45** *Prove that if $A$ is a subset of a metric space, $\dot{A} = (\overline{A^c})^c$.*

Note that the interior of any subset of a metric space always exists, although it may well be the empty set. In fact we have already observed that the set $\{0\}$ in $\mathbb{R}$ contains no open interval and hence has empty interior.

## 2.3 Sequences in Metric Spaces

The following definition of limit points should be familiar from elementary calculus.

**Definition 46** *If $X$ is a metric space, we say that $x \in X$ is the limit point of $(x_n)$, written $x = \lim x_n$ or $x_n \to x$, if for every $\varepsilon > 0$ there exists some natural number $N$ such that for all $n \geq N$, $d(x, x_n) < \varepsilon$, or equivalently, $x_n \in B(x, \varepsilon)$. If $(x_n)$ has a limit point then $(x_n)$ is said to be convergent. If necessary for clarity (e.g. if there are several variables involved) we will write $x = \lim_{n\to\infty} x_n$.*

We will now justify the article "the" of the preceding definition. Suppose that $x$ and $x'$ are limit points of $(x_n)$, and $x \neq x'$. By positive definiteness, $d(x, x') = r > 0$. By definition of limit there exist natural numbers $N$ and $N'$ such that for all $n \geq N$, $d(x, x_n) < r/2$ and for all $n \geq N'$, $d(x', x_n) < r/2$. Now suppose $n \geq \max\{N, N'\}$. Then we have

$$r = d(x, x') \leq d(x, x_n) + d(x_n, x') < r/2 + r/2 = r$$

which is a contradiction. Therefore $x = x'$; this shows that a sequence can have at most one limit point.

**Example 47** *Let $X$ be a metric space and let $x_n := a$ for all $n$. Then for any $\varepsilon > 0$, $d(x_n, a) = d(a, a) = 0 < \varepsilon$ and hence $x_n \to a$.*

**Exercise 46** *Prove the following lemma:*

**Lemma 48** *Let $(x_n)$ be a sequence in a metric space $X$. Then $x \in X$ is a limit point of $(x_n)$ if and only if $x$ is a limit point of every subsequence of $(x_n)$.*

Given any sequence $(x_n)$ in a metric space $X$ and $x \in X$ there is a sequence of real numbers, the $n^{th}$ term of which is the number $d(x_n, x)$. By definition, $x_n \to x$ if and only if for every $\varepsilon > 0$ there exists some $N \in \mathbb{N}$ such that for all $n \geq N$, $d(x, x_n) < \varepsilon$. Equivalently, for every $\varepsilon > 0$ there exists some $N \in \mathbb{N}$ such that for all $n \geq N$, $|d(x, x_n) - 0| < \varepsilon$. But this is precisely what it means for $d(x, x_n) \to 0$ in $\mathbb{R}$. We will use this observation frequently, and so state it as a lemma:

**Lemma 49** *Let $(x_n)$ be a sequence in a metric space $X$. Then $x \in X$ is a limit point of $(x_n)$ if and only if $d(x, x_n) \to 0$.*

**Lemma 50** *Let $(x_n)$ be a sequence in a metric space $X$. Then $x \in X$ is a limit point of $(x_n)$ if and only if for every open set $U$ containing $x$ there exists some $N$ such that for all $i \geq N$, $x_i \in U$.*

**Proof.** If we know that for every open set $U$ containing $x$ there exists some $N$ such that for all $i \geq N$, $x_i \in U$, then in particular letting $U$ be the open set $B(x, \varepsilon)$, it follows from the definition that $x_i \to x$. Conversely, let $x_i \to x$ and let $U$ be an open set containing $x$. Then by definition of open there exists an $\varepsilon > 0$ such that $B(x, \varepsilon) \subset U$. Since $x_i \to x$ there exists an $N$ such that for all $i \geq N$, $x_i \subset B(x, \varepsilon) \subset U$. ■

**Exercise 47** *Prove the "Sandwich Theorem" for real sequences: Let* $(x_i), (y_i), (z_i)$ *be sequences of real numbers such that* $x_i \leq y_i \leq z_i$ *for all* $i$*. If* $x_i \to x$ *and* $z_i \to x$ *then* $y_i \to x$*.*

The above exercise will be used frequently without specific reference. The reader will recall from calculus that not every sequence, even a bounded sequence, has a limit point, for example e.g. $(1, -1, 1, -1, ...)$. A related, weaker concept is obtained by modifying the definition of limit point to require not that every $x_n$ with sufficiently large $n$ be be close to $x$, but only that there exist $x_n$ with arbitrarily large $n$ that are close to $x$. More precisely:

**Definition 51** *Let* $(x_n)$ *be a sequence in a metric space* $X$*. A point* $y \in X$ *is called a cluster (or accumulation) point of* $(x_n)$ *if for every* $\varepsilon > 0$ *and* $N \in \mathbb{N}$ *there exists some* $n \geq N$ *such that* $d(y, x_n) < \varepsilon$*.*

**Example 52** *The sequence* $(1, -1, 1, -1, ...)$ *has two cluster points, but no limit point. The number* $1$ *is a cluster point: given any* $\varepsilon > 0$ *and* $N \in \mathbb{N}$ *let* $n$ *be the next odd number greater than* $N$*. Then* $x_n = 1$ *and* $d(x_n, 1) = d(1, 1) = 0 < \varepsilon$*. Likewise* $-1$ *is a cluster point. However, the sequence has no limit point. In fact, if* $a$ *were a limit point of the sequence then by Lemma 48* $a$ *would have to be a limit point of the subsequence of even terms and the subsequence of odd terms, which is impossible.*

**Exercise 48** *Show the following:*

1. *The sequence* $(1, 2, 3, ...)$ *has no cluster points. We will see later that every bounded real sequence must have at least one cluster point.*

2. *The sequence* $(1, \frac{1}{2}, 2, \frac{1}{3}, 3, \frac{1}{4}, ...)$ *has exactly one cluster point, and that cluster point is not a limit point.*

**Example 53** *There exists a real sequence having every real number as a cluster point. In fact, since the rational numbers are countably infinite, by definition there exists a surjective function* $f : \mathbb{N} \to \mathbb{Q}$*, which is by definition a sequence such that for every rational number* $q$ *there is some* $n$ *such that* $x_n = q$*. Let* $r$ *be any real number, and* $\varepsilon > 0$ *and* $N \in \mathbb{N}$ *be given. Choose* $\delta > 0$ *such that* $\delta \leq \varepsilon$ *and*

$$\delta \leq \min\{d(x_k, r) : k < N \text{ and } x_k \neq r\}$$

*According to Exercise 23, there exists some* $q \in \mathbb{Q}$ *such that* $r - \delta < q < r$ *and* $q = x_n$ *for some* $n$*. By the choice of* $\delta$*,* $n \geq N$ *and we have* $d(x_n, r) < \delta \leq \varepsilon$*. This proves that* $r$ *is a cluster point of* $(x_n)$*.*

The relationship between cluster points and limit points is further clarified by the next lemma.

**Lemma 54** *Let* $(x_n)$ *be a sequence in a metric space* $X$*. A point* $x \in X$ *is a cluster point of* $(x_n)$ *if and only if there is a subsequence* $(x_{n_k})$ *of* $(x_n)$ *such that* $x_{n_k} \to x$*.*

**Proof.** Suppose that $x$ is a cluster point of $(x_n)$. We will iteratively construct a subsequence $(x_{n_k})$ such that $d(x_{n_k}, x) < \frac{1}{k}$ for all $k$. Since $x$ is a cluster point of $(x_n)$ there exists some $x_j$ such that $d(x_j, x) < 1$. Let $n_1 := j$. The $x_{n_1}$ satisfies $d(x_{n_1}, x) < 1 = \frac{1}{1}$. Now suppose that we have constructed $x_{n_1}, ..., x_{n_m}$ for some $m \geq 1$ such that $n_1 < \cdots < n_m$ and $d(x_{n_k}, x) < \frac{1}{k}$ for all $1 \leq n \leq m$. Since $x$ is a cluster point, letting $N := n_k + 1$ there exists some $j \geq N = n_k + 1 > n_k$ such that $d(x, x_j) < \frac{1}{k+1}$. Letting $n_{k+1} := j$ finishes the construction of the subsequence, which converges to $x$ since $d(x_{n_k}, x) \to 0$ by the Sandwich Theorem. The converse is an exercise. ■

**Exercise 49** *Finish the proof of the above lemma.*

From Lemmas 48 and 54 we immediately obtain:

**Corollary 55** *Let $(x_n)$ be a convergent sequence in a metric space $X$. Then $x$ is the limit point of $(x_n)$ if and only if $x$ is the only cluster point of $(x_n)$.*

**Exercise 50** *Give an example of a sequence having a single cluster point that is not convergent.*

**Corollary 56** *Let $(x_n)$ be a sequence in a metric space $X$. If $a$ is a cluster point of a subsequence of $(x_n)$ then $a$ is a cluster point of $(x_n)$.*

**Proposition 57** *Let $A$ be a subset of a metric space $X$. Then*

$$\overline{A} = \{x \in X : \text{there exists a sequence } (x_n) \text{ in } A \text{ such that } x_n \to x\}.$$

**Proof.** Suppose that $x \in \overline{A}$. Then for every $i \in \mathbb{N}$ we can choose some point $x_i \in B(x, \frac{1}{i}) \cap A$. Then $0 \leq d(x, x_i) \leq \frac{1}{i}$, and by the Sandwich Theorem $d(x, x_i) \to 0$ and therefore $x_i \to x$ .

Conversely, suppose that there is a sequence $(x_i)$ in $A$ such that $x_i \to x \in X$. If $r > 0$ then by definition of convergence there exists some $x_i \in B(x, r)$ and therefore $x \in \overline{A}$. ■

**Corollary 58** *A subset $A$ of a metric space is closed if and only if $A$ contains every cluster point of every sequence in $A$.*

**Exercise 51** *Let $A \subset \mathbb{R}$ be bounded above and non-empty. Prove that $\sup A \in \overline{A}$. Therefore if $A$ is closed, $A$ has a maximum. (Of course a similar statement is true when $A$ is bounded below.)*

**Exercise 52** *Let $(x_i)$ be a convergent sequence in $\mathbb{R}$.*

1. *Show that if for some $k \in \mathbb{R}$, $x_i \leq k$ for all $i$, then $\lim x_i \leq k$. Hint: Prove this by contradiction.*

2. *Prove or disprove the statement in the first part if each "$\leq$" is replaced by a "$<$".*

We will now prove a couple of theorems about real sequences that we will use later. These results are more specialized because they involve the ordering of the reals, and so do not apply to metric spaces in general. Nonetheless they are very important for analysis.

**Definition 59** *A sequence* $(x_i)$ *of real numbers is called increasing (resp. decreasing) if for every* $i \geq 1$, $x_i \leq x_{i+1}$ *(resp.* $x_i \geq x_{i+1}$*). If* $(x_i)$ *is either increasing or decreasing then* $(x_i)$ *is simply called monotone.*

**Exercise 53** *Prove that* $(x_i)$ *is increasing if and only if whenever* $1 \leq i \leq j$ *we have* $x_i \leq x_j$*. We can write this shorthand as* $x_1 \leq x_2 \leq \ldots$*. A similar statement holds for decreasing sequences.*

**Proposition 60** *If* $(x_i)$ *is an increasing (resp. decreasing) real sequence that is bounded above (resp. below) then* $(x_i)$ *converges to* $\sup\{x_i\}$ *(resp.* $\inf\{x_i\}$*). More succinctly, every bounded, monotone real sequence is convergent.*

**Proof.** We will only consider the increasing case; the other case is similar. Since $(x_i)$ is bounded above, $s := \sup\{x_i\} < \infty$. Let $\varepsilon > 0$. By the Approximation Property there exists some $N$ such that $x_N > s - \varepsilon$. But since $(x_i)$ is increasing it must be true that $x_i \geq x_N > s - \varepsilon$ for all $i \geq N$. Since we also have $x_i \leq s$ for all $i$, we have that $d(x_i, s) = |x_i - s| < \varepsilon$ for all $i \geq N$. ■

**Notation 61** *If* $(x_i)$ *is a decreasing sequence converging to* $x \in \mathbb{R}$*, we will write* $x_i \searrow x$*, with similar notation for an increasing convergent sequence.*

Note that a sequence is both increasing and decreasing if and only if it is constant.

**Proposition 62** *Let* $A \subset \mathbb{R}$ *be bounded above. Then* $s = \sup A$ *if and only if* $s$ *is an upper bound for* $A$ *and there exists some sequence* $(x_i)$ *in* $A$ *such that* $x_i \to s$*. A similar statement holds for the infimum.*

**Proof.** That $x_i \to s$ means, by definition, that for every $\varepsilon > 0$ there exists some $i$ such that $|x_i - s| < \varepsilon$, which is equivalent to $s - \varepsilon < x_i < s + \varepsilon$. If $s$ is an upper bound for $A$ then in fact $x_i \leq s$ and the proof is done by Lemma Lemma 12. Conversely, suppose that the Approximation Property holds and $s$ is an upper bound of $A$. Using $\varepsilon := \frac{1}{i}$ we have that for every $i$ there $x_i \in A$ such that $s - \frac{1}{i} < x_i \leq s$, which implies $|x_i - s| < \frac{1}{i}$. By Lemma 49, $x_i \to s$. ■

## 2.4 Limits and continuity of functions

The definitions of limits and continuity for metric spaces are essentially the same as those that the reader encountered in calculus, and as we will see, many ideas and theorems involving continuity from basic calculus–such as the intermediate value and max-min theorems–are true in a more general form for metric spaces.

**Definition 63** *Let $f : A \to Y$ be a function, where $A$ is a subset of a metric space $X$ and $Y$ is another metric space. Suppose $x_0 \in \overline{A}$. We say that $\lim_{x \to x_0} f(x) = y \in Y$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $x \in A$ and $0 < d(x, x_0) < \delta$ then $d(f(x), y) < \varepsilon$. If $x_0 \in A$, we say that $f$ is continuous at $x_0$ if $f(x_0) = \lim_{x \to x_0} f(x)$. If $f$ is continuous at every $x_0 \in A$ then we simply say that $f$ is continuous (on $A$).*

It is easy to check that the limit of a function, if it exists, is unique. Note that $0 < d(x, x_0)$ requires that $x \neq x_0$, which means that $\lim_{x \to x_0} f(x)$ by itself has nothing to do with the functional value of $f$ at $x_0$. In fact $f$ may not even be defined at $x_0$ since $x_0$ is only required to be in $\overline{A}$ and not necessarily in the set $A$ where $f$ is defined. That $f$ not be required to be defined at $x_0$ is particularly important for certain kinds of limits. For example, the derivative in elementary calculus is defined by the limit

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

and the function $\frac{f(x+h)-f(x)}{h}$ is *never* defined at 0. The reader may recall that in the setting of elementary calculus the requirement for defining $\lim_{x \to x_0} f(x)$ is usually that $f$ is defined in some interval $(x_0 - \varepsilon, x_0 + \varepsilon)$, except possibly at $x_0$. This may more simply be expressed that $f$ is defined on $A := (x_0 - \varepsilon, x_0) \cup (x_0, x_0 + \varepsilon)$ (which does not exclude the possibility that $f$ is defined on a bigger set, such as $(x_0 - \varepsilon, x_0 + \varepsilon)$. In either case $x_0 \in \overline{A}$ so the conditions of Definition 63 are also satisfied, and the rest of our definition is equivalent to the one from calculus. However, in the setting of calculus one must start over with a new definition for "one-sided" limits; for a left limit, for example, the function must be defined on some interval $(x_0 - \varepsilon, x_0]$ except possibly at $x_0$. Our definition works just fine with a choice of $A = (x_0 - \varepsilon, x_0)$, and all theorems that we prove for limits with some unspecified $A$ will be valid for both left and right limits. In fact, all of these limit notions from elementary calculus, and many more general limits that are necessary for analysis are taken care of by the single abstract Definition 63, with a suitable choice of $A$.

Why do we not permit $x_0$ to lie outside $\overline{A}$? If $x_0$ does not lie in $\overline{A}$ then by definition there exists a small ball $B(x_0, \delta)$ that does not intersect the set $A$ where $f$ is defined. In this case the statement "if $x \in A$ and $0 < d(x, x_0) < \delta$ then $d(f(x), y) < \varepsilon$" is vacuously true for every $y$! We avoid this unpleasant situation by requiring that $x_0$ be in $\overline{A}$ when discussing limits, so that there is always some such $x$ in this statement. Less formally, if we think of a limit as representing what happens to $f(x)$ when $x$ "approaches" $x_0$, then $x$ should at least be able to approach $x_0$ from within $A$.

When considering continuity the picture is considerably simpler because we require that $x_0$ be in $A$ and so the issues discussed in the previous two paragraphs are not relevant. Therefore we normally will state theorems about continuity only in terms of functions $f : X \to Y$, eliminating $A$ from the picture altogether.

Later, when we discuss the subspace metric, we will see that there is no loss in generality in doing this (see Proposition 100). The simplified picture for continuity is codified in the following lemma, which will often be used without reference. The reader should especially note that the statement includes the inequality $d(x,y) < \delta$ and not $0 < d(x,y) < \delta$.

**Lemma 64** *Let $f : X \to Y$ be a function between metric spaces $X$ and $Y$. Then $f$ is continuous at a point $x \in X$ if and only if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $d(x,y) < \delta$ then $d(f(x), f(y)) < \varepsilon$. Equivalently, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $f(B(x,\delta)) \subset B(f(x),\varepsilon)$.*

**Exercise 54** *Prove Lemma 64.*

**Exercise 55** *Prove that if $f(x) = a + bx$, where $a, b, x$ are real numbers, then $f$ is continuous at every point $x$.*

There are two other equivalent formulations of continuity that are less trivial to prove, but quite important. First we need another characterization of limits:

**Proposition 65** *Let $f : A \to Y$ be a function, where $A$ is a subset of a metric space $X$ and $Y$ is another metric space. Suppose $x_0 \in \overline{A}$. Then $\lim\limits_{x\to x_0} f(x) = y \in Y$ if and only if for every sequence $(x_i)$ in $A\backslash\{x_0\}$ such that $x_i \to x_0$, $f(x_i) \to y$.*

**Proof.** Suppose first that $\lim\limits_{x\to x_0} f(x) = y$ and let $x_i \to x_0$ where $x_i \in A\backslash\{x_0\}$. Suppose $\varepsilon > 0$. By definition there exists some $\delta > 0$ such that if $x \in A$ and $0 < d(x, x_0) < \delta$ then $d(f(x), y) < \varepsilon$. Since $x_i \to x_0$ there exists some $N$ such that for all $i \geq N$, $0 < d(x_i, x_0) < \delta$. But then $d(f(x_i), y) < \varepsilon$ for all $i \geq N$ and we have shown $f(x_i) \to y$.

The converse is a bit more tricky; we will suppose that it is *not* true that $\lim\limits_{x\to x_0} f(x) = y$ and construct a sequence $(x_i)$ in $A\backslash\{x_0\}$ such that $x_i \to x_0$ but $(f(x_i))$ does not converge to $y$. If it is not true that $\lim\limits_{x\to x_0} f(x) = y$ then there exists an $\varepsilon > 0$ such that for all $\delta > 0$ there is some $z \in X$ such that $0 < d(x_0, z) < \delta$ and $d(f(z), y) \geq \varepsilon$. Since $\frac{1}{i} > 0$ for any $i \in \mathbb{N}$, we can use $\frac{1}{i}$ for $\delta$ and for each $i \in \mathbb{N}$ we can find some $x_i$ such that $0 < d(x_i, x_0) < \frac{1}{i}$ and $d(f(x_i), y) \geq \varepsilon$. Since $d(x_i, x_0) \to 0$ it follows that $x_i \to x_0$, but since $d(f(x_i), y) \geq \varepsilon > 0$ for all $i$ it is not possible that $f(x_i) \to y$. ■

**Corollary 66** *Let $f : X \to Y$ be a function, where $X$ and $Y$ are metric spaces. For any $x \in X$, $f$ is continuous at $x$ if and only if for every sequence $x_i \to x$ in $X$, $f(x_i) \to f(x)$, or, stated another way, $f(\lim x_i) = \lim f(x_i)$.*

**Corollary 67** *Let $f : A \to Y$ and $g : Y \to Z$ be functions, where $X$, $Y$, and $Z$ are metric spaces and $A \subset X$, and let $x_0 \in \overline{A}$. If $\lim_{x\to x_0} f(x) = y$ and $g$ is continuous at $y$, then*

$$g(\lim_{x\to x_0} f(x)) = g(y) = \lim_{x\to x_0} g(f(x)).$$

*In other words, the limit can be "pulled through" the continuous function $g$.*

**Proof.** Let $x_i \to x_0$, where $x_i \in A\backslash\{x_0\}$. Then applying the Proposition 65 and Corollary 66 we have

$$\lim_{i\to\infty} g(f(x_i)) = g(\lim_{i\to\infty} f(x_i)) = g(\lim_{x\to x_0} f(x)) = g(y).$$

Proposition 65 now implies that

$$\lim_{x\to x_0} g(f(x)) = \lim_{i\to\infty} g(f(x_i)) = g(y).$$

■

If $f$ is also continuous at $y$ in the above Corollary then we may replace $y$ by $f(x_0)$ and we obtain the following corollary:

**Corollary 68** *Let $f : X \to Y$ and $g : Y \to Z$ be functions, where $X$, $Y$, and $Z$ are metric spaces. Suppose that $f$ is continuous at $x \in X$ and $g$ is continuous at $y := f(x)$. Then $g \circ f : X \to Z$ is continuous at $x$.*

**Corollary 69** *Let $f : X \to Y$ and $g : Y \to Z$ be continuous functions, where $X$, $Y$, and $Z$ are metric spaces. Then $g \circ f$ is continuous.*

**Proposition 70** *A function $f : X \to Y$ between metric spaces $X$ and $Y$ is continuous if and only if for every open set $U$ in $Y$, $f^{-1}(U)$ is open in $X$.*

**Proof.** First suppose that $f$ is continuous and let $U$ be open in $Y$. Let $x \in f^{-1}(U)$; by definition $f(x) \in U$. Since $U$ is open there exists some $\varepsilon > 0$ such that $B(f(x), \varepsilon) \subset U$. Since $f$ is continuous at $x$, by Lemma 64 there exists some $\delta > 0$ such that

$$f(B(x, \delta)) \subset B(f(x), \varepsilon) \subset U.$$

By definition $B(x, \delta) \subset f^{-1}(U)$, which proves $f^{-1}(U)$ is open.

Conversely, suppose that $f^{-1}(U)$ is open for every open $U$ in $Y$. Let $x \in X$ and $\varepsilon > 0$. Then $B(f(x), \varepsilon)$ is open in $Y$ and therefore $V := f^{-1}(B(f(x), \varepsilon))$ is open in $X$ and contains $x$ by definition. Since $V$ is open and contains $x$ there is some $\delta > 0$ such that $B(x, \delta) \subset V$. But then

$$f(B(x, \delta)) \subset f(V) = f(f^{-1}(B(f(x), \varepsilon))) \subset B(f(x), \varepsilon)$$

which proves that $f$ is continuous at $x$. ■

**Corollary 71** *A function $f : X \to Y$ between metric spaces $X$ and $Y$ is continuous if and only if for every closed set $A$ in $Y$, $f^{-1}(A)$ is closed in $X$.*

**Proof.** First note that $X\backslash f^{-1}(A) = f^{-1}(Y)\backslash f^{-1}(A) = f^{-1}(Y\backslash A)$. Therefore $f^{-1}(A)$ is closed for every closed set $A$ in $Y$ if and only if $f^{-1}(Y\backslash A)$ is open for every closed set $A$ in $Y$, which is true if and only if $f^{-1}(U)$ is open for every open set $U$ in $Y$. ■

**Example 72** *Any function $f : X \to Y$, where $X$ has the trivial metric and $Y$ is any metric space, is continuous. In fact, in Exercise it was shown that every subset of $X$ is open. Therefore for any set $U$ in $Y$ (whether it is open or not!), $f^{-1}(U)$ is open. On the other hand, functions from a metric space into a trivial metric space are often not continuous because there are simply "too many" open sets in the trivial metric. For example, suppose that $f : X \to Y$ is a continuous 1-1 function between metric spaces, where $Y$ has the trivial metric. Let $x \in X$. Then since $Y$ has the trivial metric, $\{f(x)\}$ is an open set in $Y$, and hence $f^{-1}(\{f(x)\})$ is open in $X$. But since $f$ is 1-1, $f^{-1}(\{f(x)\}) = \{x\}$, which shows that every set $\{x\}$ in $X$ is open. Since every subset $A$ of $X$ can be written as $A = \cup_{x\in A}\{x\}$ it follows from Proposition 38 that every subset of $X$ is open, and hence every subset of $X$ is closed. So while $X$ may not have the trivial metric, like the trivial metric every subset of $X$ is both open and closed. So, for example, there can be no continuous 1-1 function from $\mathbb{R}$ into a trivial metric space.*

**Exercise 56** *Prove that a function $f : X \to Y$ between metric spaces $X$ and $Y$ is continuous if and only if for every $y \in Y$ and $r > 0$, $f^{-1}(B(y, r))$ is open in $X$.*

**Exercise 57** *Let $f : X \to Y$ be a constant function $f(x) = y_0$, where $y_0$ is a fixed element of $Y$. Prove that $f$ is continuous in the following ways:*

1. *Using the definition of continuity.*
2. *Using Corollary 66.*
3. *Using Proposition 70.*

**Exercise 58** *Prove that the identity function $f : X \to X$ given by $f(x) = x$, where $X$ is a metric space, is continuous.*

We consider one final function in this section. This function is obtained by fixing a single point $x$ in a metric space $X$. Then the distance from this fixed point assigns to each point $y \in X$ a non-negative real number. The next proposition shows that this function is continuous; that is, as a point $y$ "moves about" in $X$, the distance from the fixed point $x$ varies continuously.

**Proposition 73** *Let $X$ be a metric space and $x \in X$. The function $d_x : X \to \mathbb{R}$ defined by $d_x(y) = d(x, y)$ is continuous.*

**Proof.** Let $y \in X$ and $\varepsilon > 0$. Let $\delta := \varepsilon > 0$. If $d(y, z) < \delta$ then by the triangle inequality,

$$d(x, y) \leq d(x, z) + d(y, z)$$

or

$$d_x(y) - d_x(z) \leq d(y, z) < \varepsilon.$$

Likewise from

$$d(x, z) \leq d(x, y) + d(y, z)$$

we obtain

$$d_x(y) - d_x(z) \geq -d(y,z) > -\varepsilon$$

and it follows that

$$d(d_x(y), d_x(z) = |d_x(y) - d_x(z)| < \varepsilon.$$

■

**Corollary 74** *If $y_i \to y$ in a metric space $X$ then for any $x \in X$, $d(y_i, x) \to d(y, x)$.*

**Exercise 59** *Prove or disprove the converse to Corollary 74.*

## 2.5 Compactness

**Definition 75** *A subset $A$ of a metric space $X$ is called compact if every sequence in $A$ has a cluster point in $A$.*

It is important to note that the above definition of compactness is not the traditional definition (see Remark 86). For metric spaces the two definitions are logically equivalent (although it requires some work to prove this) but the definition we are using, normally referred to as "sequential compactness", is much more intuitive and simple to work with. The traditional definition is necessary in the more general setting of topological spaces, where these two definitions are *not* equivalent, and sequential compactness is too weak to prove the most important theorems.

Note that the empty set is compact since the hypothesis is vacuous. Recall that a sequence has a cluster point if and only if it has a convergent subsequence. Therefore $A$ is compact if and only if every sequence in $A$ has a subsequence that converges to a point in $A$. If $A = X$ we can more succinctly state: $X$ is compact if and only if every sequence in $X$ has a convergent subsequence.

**Proposition 76** *If $A$ is a compact subset of a metric space $X$ then $A$ is closed and bounded.*

**Proof.** Note that the statement is vacuously true if $A$ is the empty set. Suppose $A$ is compact and $x \in \overline{A}$. By Proposition 57 there is a sequence $(x_i)$ in $A$ such that $x_i \to x$. Since $A$ is compact there is a cluster point $y \in A$ of $(x_i)$. But since $(x_i)$ is convergent, the cluster point $y$ must be equal to the limit point $x$ (by Corollary 55); i.e., $x \in A$. This shows $\overline{A} \subset A$ and so $A$ is closed.

Suppose now that $A$ is not bounded, and fix a point $x \in X$. Then for every $i \in \mathbb{N}$ there exists some point $x_i \in A$ such that $x_i \notin B(x, i)$, i.e., $d(x, x_i) \geq i$. We claim that the sequence $(x_i)$ cannot have a convergent subsequence, which will finish the proof by showing that $A$ is not compact. Suppose, to the contrary that there is a subsequence $x_{i_k} \to y \in A$. But $d(y, x_{i_k}) \geq d(x, x_{i_k}) - d(x, y) \geq i_k - d(x, y) \to \infty$, a contradiction. ■

The converse of the above proposition is false in general (as we will see an example later) but is true if $X$ is the real numbers. This theorem is called the Heine-Borel Theorem, and will be proved after a couple of preliminaries that are useful in their own right.

**Proposition 77** *If $A$ is a compact subset of a metric space $X$ and $B \subset A$ is closed then $B$ is compact.*

**Proof.** Let $(x_i)$ be a sequence in $B$. Then $(x_i)$ is also a sequence in $A$, so has a cluster point $y$ in $A$. But $B$ is closed and hence contains all cluster points of $(x_i)$, and so $y \in B$. ■

**Exercise 60** *Let $\{A_\lambda\}_{\lambda\in\Lambda}$ be a collection of compact subsets of a metric space $X$. Show $\bigcap_{\lambda\in\Lambda} A_\lambda$ is compact.*

**Exercise 61** *Let $X$ be a metric space and $x \in X$. Show any singleton set $\{x\}$ is compact.*

**Proposition 78** *(Bolzano-Weierstrass Theorem) Every bounded sequence of real numbers has a cluster point.*

**Proof.** Let $(x_i)$ be a bounded sequence of real numbers. For all $n$ let $A_n := \{x_i\}_{i\geq n}$; then $A_n$ is non-empty and bounded. Let $s_n := \sup A_n$. Now $\{s_n\}$ is bounded below. In fact, if $L$ is a lower bound of $\{x_i\}$ then we have $s_n \geq x_n \geq L$. In addition, since $A_{n+1} \subset A_n$ for all $n$, $\{s_n\}$ is decreasing, and hence by Proposition 60, $s_n \to x := \inf\{s_n\}$. We will show that $x$ is a cluster point of $(x_i)$. Let $\varepsilon > 0$ and $N \in \mathbb{N}$. Since $s_n \searrow x$ there exists some $N'$ such that

$$\text{for all } n \geq N',\ x \leq s_n < x + \varepsilon. \tag{2.1}$$

Let $M = \max\{N, N'\}$. By the approximation property there exists some $x_i \in A_M$ such that

$$s_M - \varepsilon < x_i \leq s_M. \tag{2.2}$$

Since $M \geq N'$ we can combine Formulas (2.1) and (2.2) to obtain

$$x - \varepsilon \leq s_M - \varepsilon < x_i \leq s_M < x + \varepsilon$$

which is equivalent to $|x - x_i| < \varepsilon$. Finally, since $x_i \in A_M$, by definition $i \geq M \geq N$ and we have shown that $x$ is a cluster point of $(x_i)$. ■

**Theorem 79** *(Heine-Borel Theorem) A subset of the reals is compact if and only if it is closed and bounded.*

**Proof.** Proposition 76 is one direction. Conversely, let $A \subset \mathbb{R}$ be closed and bounded. Then any sequence $(x_i)$ in $A$ is a bounded sequence, so by the Bolzano-Weierstrass Theorem $(x_i)$ has a cluster point. Since $A$ is closed, the cluster point must be in $A$. ■

The Bolzano-Weierstrass and Heine-Borel theorems are actually logically equivalent (in fact we will later, in the setting of $\mathbb{R}^n$, prove the Heine-Borel Theorem first and use it to prove the Bolzano-Weierstrass theorem). In fact, these theorems are logically so close to one another that they are given separate, named statements only for historical reasons. The reader will therefore be forgiven for always referring to the Heine-Borel Theorem (which is the "cleaner" of the two statements) when extracting a convergent subsequence from a bounded real sequence. Some texts do not mention the names "Bolzano-Weierstrass" at all.

**Theorem 80** *Let $X$ and $Y$ be metric spaces, $A$ be a compact subset of $X$, and $f : X \to Y$ be continuous. Then $f(A)$ is a compact subset of $Y$.*

**Proof.** Let $(y_i)$ be a sequence in $f(A)$. By definition, for each $i$ there exists some $x_i \in A$ such that $f(x_i) = y_i$. Since $A$ is compact, there is a subsequence $(x_{i_j})$ of $(x_i)$ such that $x_{i_j} \to x$ for some $x \in A$. By Proposition 66, we have that $y_{i_j} = f(x_{i_j}) \to f(x)$ and by definition $f(x) \in f(A)$, so $(y_i)$ has a convergent subsequence in $f(A)$. ■

We will verify later that the function $f(x) = e^x : \mathbb{R} \to \mathbb{R}$ takes $\mathbb{R}$ (which is closed) to $(0, \infty)$, which is not closed in $\mathbb{R}$. That is, a real continuous function need not take closed sets to closed sets. We will also later show that the natural log function is continuous, but it takes $(0, 1)$, which is bounded, to $(-\infty, 0)$, which is not. However, the following is an immediate corollary of Theorem 80 and the Heine-Borel Theorem:

**Corollary 81** *If $f : \mathbb{R} \to \mathbb{R}$ is continuous and $A \subset \mathbb{R}$ is closed and bounded then $f(A)$ is closed and bounded.*

Here is a powerful application of our results so far:

**Theorem 82** *(Max-min Theorem) Let $A$ be a nonempty compact subset of a metric space $X$ and $f : X \to \mathbb{R}$ be continuous. Then $f$ has a maximum and a minimum on $A$.*

**Proof.** By Theorem 80, $f(A)$ is compact, hence a closed, bounded nonempty set. But then by Exercise 51, $f(A)$ has a maximum $M$. In other words, for some $x \in A$, $f(x) = M$; by definition $M$ is the maximum of $f$ on $A$ The proof that $f$ has a minimum is similar. ■

The next theorem is also useful.

**Theorem 83** *Let $\{A_i\}_{i=1}^{\infty}$ be a collection of non-empty compact subsets of a metric space $X$ that is nested, in the sense that if $i < j$ then $A_j \subset A_i$. Then $\bigcap_{i=1}^{\infty} A_i \neq \varnothing$.*

**Proof.** Since each $A_i$ is nonempty there is some $x_i \in A_i$. Since $x_i \in A_i \subset A_1$ for all $i$ and $A_1$ is compact, $(x_i)$ has a cluster point $x$ in $A_1$. We will show that $x \in A_n$ for all $n$. By definition of cluster point, for each $n$, $x$ is also a cluster point of the tail $(x_i)_{i=n}^{\infty}$ of the sequence $(x_i)$. Since the sets are nested, the tail $(x_i)_{i=n}^{\infty}$ lies entirely in the compact, hence closed, set $A_n$, and hence all of its cluster points, including $x$, are contained in $A_n$. ■

**Example 84** *The collection* $\left\{(0,\frac{1}{i})\right\}$ *consists of bounded, nested sets, but* $\bigcap_{i=1}^{\infty}(0,\frac{1}{i})=\varnothing$*. The collection* $\{[n,\infty)\}$ *consists of closed, nested sets, but also has empty intersection.*

**Exercise 62** *Prove or disprove: If* $f:X\to Y$ *is continuous, where* $X$ *and* $Y$ *are metric spaces, and* $A\subset Y$ *is compact, then* $f^{-1}(A)$ *is compact.*

**Definition 85** *Let* $A$ *be a subset of a metric space* $X$*. An open cover of* $A$ *is a collection of open subsets* $\{U_\lambda\}_{\lambda\in\Lambda}$ *in* $X$ *such that* $X\subset\cup_{\lambda\in\Lambda}U_\lambda$*.*

**Remark 86** *The usual definition of compactness is: Every open cover of* $A$ *has a finite subcover–that is a finite subcollection, the union of which contains* $A$*. A significant part of the proof that these two definitions are equivalent for metric spaces is the following proposition, which will also be very useful for us. A full proof of the equivalence may be found in most topology texts (e.g. [1]).*

**Proposition 87** *Let* $X$ *be a compact metric space and suppose that* $\{U_\lambda\}_{\lambda\in\Lambda}$ *is an open cover of* $X$*. Then there exists an* $\varepsilon>0$*, called the Lebesgue number for the open cover, such that if* $x,y\in X$ *and* $d(x,y)<\varepsilon$ *then there is some* $U_\lambda$ *such that both* $x$ *and* $y$ *are elements of* $U_\lambda$*.*

**Proof.** Suppose not. Using $\varepsilon=\frac{1}{i}$ for all $i$ we can find sequences $(x_i)$ and $(y_i)$ in $X$ such that $d(x_i,y_i)<\frac{1}{i}$, and for all $\lambda$ and $i$, $x_i\notin U_\lambda$ or $y_i\notin U_\lambda$. Since $X$ is compact there is a cluster point $x\in X$ of $(x_i)$. Since $\{U_\lambda\}_{\lambda\in\Lambda}$ is an open cover, $x\in U_\lambda$ for some $\lambda$. Since $U_\lambda$ is open there exists some $r>0$ such that $B(x,r)\subset U_\lambda$. Since $x$ is a cluster point of $(x_i)$ there exists some $i>\frac{2}{r}$ such that $x_i\in B(x,\frac{r}{2})$. But then we have $d(x_i,y_i)<\frac{r}{2}$ and $d(x_i,x)<\frac{r}{2}$. It follows from the triangle inequality that $y_i\in B(x,r)\subset U_\lambda$. Since $x_i$ is also in $U_\lambda$, this is a contradiction. ■

**Lemma 88** *Let* $X$ *be a metric space and* $\{A_j\}_{j=1}^{n}$ *be a finite collection of compact subsets of* $X$*. Then* $\bigcup_{j=1}^{n}A_j$ *is compact.*

**Proof.** Let $(x_i)$ be a sequence in $\bigcup_{j=1}^{n}A_j$. By Corollary 27 there are some $j$ and a subsequence $(x_{i_k})$ of $(x_i)$ such that $(x_{i_k})$ lies entirely in $A_j$. Since $A_j$ is compact, $(x_{i_k})$ has a cluster point in $A_j$, and hence in $\bigcup_{j=1}^{n}A_j$. But this cluster point is also a cluster point of the original sequence by Corollary 56. ■

**Exercise 63** *Show by example that the union of infinitely many compact subsets of* $X$ *need not be compact.*

We finish with some theorems that will be useful later. First we need a definition.

**Definition 89** *Let $A$ be a subset of a metric space $X$ and $\varepsilon > 0$. The $\varepsilon$-neighborhood of $A$ is the set*

$$N(A,\varepsilon) := \{x \in X : d(x,y) < \varepsilon \text{ for some } y \in A\}.$$

Note that $N(\{x\},\varepsilon) = B(x,\varepsilon)$. It follows from basic set theory that $N(A,\varepsilon)$ is the union of all open balls of radius $\varepsilon$ centered at points in $A$; that is,

$$N(A,\varepsilon) = \bigcup_{x\in A} B(x,\varepsilon),$$

and therefore $N(A,\varepsilon)$ is open.

**Proposition 90** *If $C$ is a compact subset of a metric space $X$ and $U$ is an open set in $X$ containing $C$ then for some $\varepsilon > 0$, $N(C,\varepsilon) \subset U$.*

**Proof.** Suppose not. Using $\varepsilon = \frac{1}{i}$ for all $i$ there exists a sequence $\{x_i\}$ such that $x_i \notin U$ and $x_i \in N(C, \frac{1}{i})$. That is, for each $i$ there is some $y_i \in C$ such that $d(x_i, y_i) < \frac{1}{i}$. Since $C$ is compact $\{y_i\}$ has a cluster point $y \in C$. Since $U$ is open there is some $B(y,r) \subset U$. But there also exist $i > \frac{2}{r}$ such that $y_i \in B(y, \frac{r}{2})$. We have $d(x_i,y) \le d(x_i,y_i) + d(y_i,y) < \frac{r}{2} + \frac{r}{2} = r$. In other words, $x_i \in B(y,r) \subset U$, a contradiction. ■

**Exercise 64** *Use the interval $(0,1)$ to show that Proposition 90 is false if $C$ is not compact.*

**Proposition 91** *Let $C$ be a compact subset of a metric space $X$ and suppose that $x \notin C$. Then there exists some $\varepsilon > 0$ such that $x \notin N(C,\varepsilon)$.*

**Proof.** According to Proposition 73 the function $d_x$ is continuous and hence achieves a minimum on the compact set $C$. That is, there exists some $y \in C$ such that $d_x(y)$ is the minimum of $d(x,y)$. Now if $d(x,y) = 0$ then $x = y \in C$, a contradiction. Therefore $d(x,y) = \varepsilon$ for some $\varepsilon > 0$. Since $y$ is at minimal distance from $x$ in $C$, $d(x,z) \ge \varepsilon$ for all $z \in C$ and $x \notin N(C,\varepsilon)$. ■

**Exercise 65** *Let $C$ be a compact subset of a metric space $X$.*

1. *Show that $N(C,\varepsilon)$ is bounded for any $\varepsilon > 0$.*
2. *Show that $C = \bigcap_{\varepsilon>0} N(C,\varepsilon)$.*
3. *Show that if $U$ is any open set containing $C$ then there is some $\varepsilon > 0$ such that $\overline{N(C,\varepsilon)} \subset U$.*

**Definition 92** *A function $f : X \to Y$ between metric spaces is called uniformly continuous if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in X$ such that $d(x,y) < \delta$, $d(f(x), f(y)) < \varepsilon$.*

As we have seen, for continuity the number $\delta > 0$ in the definition may depend on both $\varepsilon$ and the point in question, whereas for uniform continuity a single $\delta > 0$ will do the job "uniformly", i.e. for every $x \in X$.

**Proposition 93** *Let $f : X \to Y$ be a continuous function between metric spaces. If $X$ is compact then $f$ is uniformly continuous.*

**Proof.** Let $\varepsilon > 0$. Consider the collection of all sets of the form $U_y := f^{-1}(B(y, \frac{\varepsilon}{2}))$, where $y \in Y$. Each of these sets is open since $f$ is continuous, and these sets form an open cover of $X$. In fact, if $x \in X$ then $x \in U_y$ where $y = f(x)$. Let $\delta > 0$ be the Lebesgue number for this open cover. By definition, if $d(x, w) < \delta$, then $x$ and $w$ both lie in some $U_y$ and therefore $f(x)$ and $f(w)$ both lie in $B(y, \frac{\varepsilon}{2})$. From the triangle inequality we obtain $d(f(x), f(w)) < \varepsilon$. ■

**Exercise 66** *Let $f(x) := \frac{1}{x}$ on the interval $(0, 1]$. Show that $f$ is not uniformly continuous.*

## 2.6 Subspaces and isometries

**Definition 94** *Let $(X, d)$ be a metric space and $A$ be a subset of $X$. The subspace metric on $A$ is the restriction of the metric $d$ to the set $A \times A \subset X \times X$.*

In other words, we are considering $A$ as a metric space in its own right, measuring the distance between two points $x$ and $y$ in $A$ in the same way it is measured in $X$. Since we are restricting the function $d$ to $A$, this function still satisfies all the requirements of a metric, and so does make $A$ a metric space. Moreover, if $\{x_i\}$ is a sequence in $A$ and $x \in A$ then $x_i \to x$ in $X$ if and only if $d(x_i, x) \to 0$, which is true if and only if $x_i \to x$ in $A$. That is, as long as a given sequence and point both lie in $A$, then convergence in $A$ means the same thing as convergence in $X$.

While the subspace metric may seem at first to be a trivial concept, it is in fact quite important, and the properties of the metric space $A$ can be quite distinct from those of the larger space $X$. The most basic difference has to do with open balls. Given $x \in A$ and $r > 0$, we have two possibilities: the open ball of radius $r$ centered at $x$ in the metric space $X$, and the open ball of radius $r$ centered at $x$ in the metric space $A$. In order to keep them straight we will refer these two balls as $B_X(x, r)$ and $B_A(x, r)$, respectively. How are these open balls related?

$$B_A(x, r) = \{y \in A : d(y, x) < r)\} = \{y \in X : y \in A \text{ and } d(x, y) < r\} = A \cap B_X(x, r)$$

That is, the open ball in the subspace is the intersection of subspace with the open ball of the same radius in the ambient space.

**Example 95** *Consider the interval $A := (0,1] \subset \mathbb{R}$ with the subspace metric. What do some metric balls look like? Certainly*

$$B_A(\frac{1}{2}, \frac{1}{4}) = (0,1] \cap (\frac{1}{4}, \frac{3}{4}) = (\frac{1}{4}, \frac{3}{4}) = B_{\mathbb{R}}(\frac{1}{2}, \frac{1}{4}),$$

*but*

$$B_A(\frac{1}{4}, \frac{1}{2}) = (0,1] \cap (-\frac{1}{4}, \frac{3}{4}) = (0, \frac{3}{4}) \neq B_{\mathbb{R}}(\frac{1}{4}, \frac{1}{2}).$$

*At least this open ball in $A$ is an open interval, but this is not always the case; for example the reader can easily check that $B_A(1, \frac{1}{2}) = (\frac{1}{2}, 1]$. Now $(\frac{1}{2}, 1]$, being an open ball in the metric space $A$, is an open set in the metric space $A$, which means that its complement in $A$, $(\frac{1}{2}, 1]^c = (0, \frac{1}{2}]$ is a closed subset of the metric space $A$. In addition, $(0, \frac{1}{2}]$ is certainly a bounded subset of $A$, and yet it is not compact, since the sequence $(\frac{1}{i})$ has no cluster point in $(0, \frac{1}{2}]$. This shows that the Heine-Borel and Bolzano-Weierstrass Theorems are not true for metric spaces in general.*

The above example illustrates that for subsets of $A$, being "open" is a relative concept; a subset of $A$ can be open in $A$ with the subspace metric but not open in the larger metric space $X$. Therefore we need to distinguish notationally between open (resp. closed) subsets of the metric space $X$ that happen to be contained in $A$ and subsets of $A$ that are open (resp. closed) with respect to the subspace metric of $A$. From now on, the former will be referred to as *open (resp. closed) in $X$* and the latter will be referred to as *open (resp. closed) in $A$*. Some authors use the notation *open (resp. closed) relative to $A$*, which admittedly is less likely to lead to confusion, but is a bit cumbersome and may contribute the the student avoiding the idea that $A$ is a metric space in its own right. Likewise, if $B \subset A$ we will refer to the closure of $B$ as a subset of the metric space $A$ as the closure of $B$ in $A$.

**Example 96** *Revisiting Example 95 with our new terminology we have: $(\frac{1}{4}, \frac{3}{4})$ and $(0, \frac{3}{4})$ are both open in $\mathbb{R}$ and in $A$. In fact, both sets are open balls in both metric spaces, although $(0, \frac{3}{4}) = B_A(\frac{1}{4}, \frac{1}{2})$, while $(0, \frac{3}{4}) = B_{\mathbb{R}}(\frac{3}{8}, \frac{3}{8})$. The set $(0,1]$ is both open and closed in $A$, but neither open nor closed in $\mathbb{R}$.*

The next proposition completely characterizes sets that are open or closed in the subspace metric.

**Proposition 97** *Let $X$ be a metric space and $A \subset X$ have the subspace metric. Then a set $B \subset A$ is open (resp. closed) in $A$ if and only if $B = V \cap A$ for some $V \subset X$ that is open (resp. closed) in $X$.*

**Proof.** Suppose first that $B = V \cap A$ where $V$ is closed in $X$. Let $x$ be in the closure of $B$ in $A$. So $x \in A$ and there is some sequence $(x_i)$ in $B$ with $x_i \to x$. But $(x_i)$ is also a convergent sequence in $V$, and since $V$ is closed in $X$, $x \in V$. That is, $x \in V \cap A = B$ and so $B$ is closed in $A$. Conversely, suppose that $B$ is closed in $A$. Let $V$ be the closure of $B$ in $X$. We need to show that

$B = V \cap A$. Certainly $B \subset V \cap A$; to prove the opposite inclusion let $x \in V \cap A$. By definition, $x \in A$ and there is some sequence $(x_i)$ in $B$ such that $x_i \to x$. But since $x \in A$ and $B$ is closed in $A$, $x \in B$.

We have proved the proposition in the case of closed sets. Now suppose that $B$ is open in $A$. Then $A \backslash B$ is closed in $A$, so (by what we proved above) $A \backslash B = W \cap A$ for some $W$ that is closed in $X$ and if we let $V := X \backslash W$, $V$ is open in $X$. Then

$$B = A \backslash (A \backslash B) = A \backslash (W \cap A) = (A \backslash W) \cup (A \backslash A) = A \backslash W = A \cap (X \backslash W) = A \cap V$$

The converse statement is an exercise. ■

**Exercise 67** *Let $X$ be a metric space and $A \subset X$ have the subspace metric.*

1. *Use de Morgan's law to finish the proof of Proposition 97.*

2. *Prove using only the definitions that if $V$ is open in $X$ then $V \cap A$ is open in $A$.*

**Remark 98** *Give a few moments' thought to the idea of directly proving that if $B$ is open in $A$ then $B = V \cap A$ for some set $V$ open in $X$. While certainly possible, constructing such an open set $V$ requires dealing with complements and requires more steps than the construction of the appropriate closed set in the proof of Proposition 97. This is why we chose to work first with closed sets and use de Morgan's law to obtain the result for open sets, rather than the other way around.*

**Example 99** *Let the integers $\mathbb{Z} \subset \mathbb{R}$ have the subspace metric, and let $n \in \mathbb{Z}$. Then $B_{\mathbb{Z}}(n, \frac{1}{2}) = (n - \frac{1}{2}, n + \frac{1}{2}) \cap \mathbb{Z} = \{n\}$; in other words every set containing a single integer is open in $\mathbb{Z}$. It follows (as in the case of the trivial metric) that every subset of $\mathbb{Z}$ is both open and closed. Note that the subspace metric on $\mathbb{Z}$ is not the trivial metric (e.g. $d(1, 3) = 2$), but still has the same open and closed sets as the trivial metric.*

**Exercise 68** *Prove or disprove: Let $\mathbb{Q} \subset \mathbb{R}$ be the rational numbers with the subspace metric. Then every subset of $\mathbb{Q}$ is open and closed.*

Another question that needs to be sorted out is the following. If a continuous function $f : X \to Y$ is restricted to a subset $A$ of $X$, is it continuous with respect to the subspace metric on $A$? This simple question is important because we have already defined what it means for $f$ to be continuous on $A$, and we would like to be sure that this means the same thing as being continuous when restricted to the metric space $A$. The next proposition verifies this.

**Proposition 100** *Let $f : X \to Y$ be a function between metric spaces and let $x \in A \subset X$. If $f$ is continuous at $x$ then then $f|_A : A \to Y$ is continuous at $x$, when $A$ is given the subspace metric.*

**Proof.** Let $(x_i)$ be a sequence in $A$ with $x_i \to x \in A$. Then $(x_i)$ is also a sequence in $X$ and $x_i \to x$ in $X$. Since $f$ is continuous at $x$ we have $f(x_i) \to f(x)$, which is what we needed to prove. ■

**Exercise 69** *Let $X$ be a metric space and $A \subset X$. We can consider $A$ as a metric space with the subspace metric. The inclusion of $A$ into $X$ is the function $i : A \to X$ defined by $i(x) = x$ for all $x \in A$. Show that $i$ is continuous in three different ways: Using Lemma 64, using Proposition 65, and using Proposition 70.*

The next lemma can be useful if we happen to know that the image of the function $f$ lies in some set $B$ that is smaller than $Y$. Then without affecting continuity we may replace $Y$ by $B$ with the subspace metric.

**Lemma 101** *Let $f : X \to Y$ be a function between metric spaces and $B$ be a subset of $Y$ such that $f(X) \subset B$. Then $f$ is continuous if and only if for every set $A$ that is open (resp. closed) in $B$, $f^{-1}(A)$ is open (resp. closed) in $X$.*

**Exercise 70** *Prove the above lemma. You need only prove it for open sets; the other case is similar.*

We know already that "closed" is a relative concept; a set can be closed in $A \subset X$ but not closed in $X$. Is compactness similarly relative? As the next proposition shows, the answer is "no", and we don't need to bother with phrases like "compact in $A$".

**Proposition 102** *Let $X$ be a metric space and $A \subset X$ have the subspace metric. Then $B \subset A$ is a compact subset of $A$ if and only if $B$ is a compact subset of $X$. In particular, $A$ is a compact subset of $X$ if and only if $A$ is a compact metric space with the subspace metric.*

**Proof.** $B$ is compact as a subset of $A$ if and only if every sequence $(x_i)$ in $B$ has a cluster point in $B$. But this is exactly what it means for $B$ to be a compact subset of $X$! ■

**Definition 103** *Let $X$ and $Y$ be metric spaces. A function $f : X \to Y$ is called an isometry if $f$ is onto and for every $x, y \in X$, $d_X(x, y) = d_Y(f(x), f(y))$. If there exists an isometry $f : X \to Y$ then $X$ and $Y$ are said to be isometric.*

**Exercise 71** *Let $f : X \to Y$ be an isometry between metric spaces.*

1. Prove that $f$ is $1 - 1$.

2. Prove that $f^{-1}$ is an isometry.

3. Prove that $f$ is continuous.

An isometry, therefore, "identifies" the metric space $X$ with the metric space $Y$; they are indistinguishable except for the names of their elements. In general, bijective functions that show that two sets have an identical structure are very important in mathematics. For example, in abstract algebra a function that identifies two groups as having the exact same group structure is called a group isomorphism (and there are also ring and field isomorphisms, and linear isomorphisms that show that two vector spaces have the exact same vector space structure). Given metric spaces $X$ and $Y$, we can ask whether $X$ and $Y$ are "the same" in terms of the open sets they possess, even if they have "different" metrics. For example, we have already seen that the integers can be given two different metrics–the trivial metric and the subspace metric–that give rise to exactly the same open sets (see Example 99). Since many questions, like convergence and continuity, are really determined by open sets and not specifically what the metric is, the following concept is very important:

**Definition 104** *Let $X$ and $Y$ be metric spaces. A function $f : X \to Y$ is called a homeomorphism (or topological isomorphism) if $f$ is a bijection and both $f$ and $f^{-1}$ are continuous. If there exists a homeomorphism $f : X \to Y$ then $X$ and $Y$ are said to be homeomorphic.*

**Lemma 105** *If $f : X \to Y$ is a continuous bijection and $X$ is compact then $f$ is a homeomorphism.*

**Proof.** To show $f^{-1}$ is continuous we need to show that for every closed subset $A$ of $X$, $(f^{-1})^{-1}(A)$ is closed in $Y$. But $(f^{-1})^{-1}(A) = f(A)$ and $A$ is a closed subset of the compact space $X$, and therefore is compact. But by Theorem 80 $f(A)$ is compact and hence closed in $Y$. ■

**Exercise 72** *Prove that $\mathbb{Z}$ with the trivial metric is homeomorphic to $\mathbb{Z}$ with the subspace metric.*

**Exercise 73** *Show that if $f : X \to Y$ is a continuous bijection then $f$ is a homeomorphism if and only if for every open $U$ in $X$, $f(U)$ is open in $Y$.*

**Exercise 74** *Show that*

1. *every isometry is a homeomorphism,*
2. *the inverse of any homeomorphism is a homeomorphism,*
3. *the composition of two homeomorphisms is a homeomorphism.*

## 2.7 Product Metrics

The reader should recognize that the next construction is a generalization of how the usual distance in Euclidean space $\mathbb{R}^n$ is constructed from the distance on $\mathbb{R}$. Some topology texts call the "product metric" what we later refer to as the "max" metric. As we will see, the max metric has the same open and closed sets as the product metric and is simpler to use in some situations, although the max metric is geometrically less familiar. For example, the product metric on $\mathbb{R}^n$ gives rise to classical Euclidean geometry, whereas the max metric gives rise to a different geometry in which, for example, there are infinitely many midpoints between any two points (see Exercise 77).

**Definition 106** *Let $(X_1, d_1), ..., (X_n, d_n)$ be a collection of metric spaces. We define the product metric on the cartesian product $X_1 \times \cdots \times X_n$ by*

$$d((x_1, ..., x_n), (y_1, ..., y_n)) = \left(d_1(x_1, y_1)^2 + \cdots + d_n(x_n, y_n)^2\right)^{\frac{1}{2}}. \tag{2.3}$$

Of course we need to check that this formula really defines a metric. But first we point out that, if $X_i = \mathbb{R}$ with the standard metric for all $i$ then the metric defined in this way on $\mathbb{R}^n$ is given by the familiar formula

$$d((x_1, ..., x_n), (y_1, ..., y_n)) = \left((x_1 - y_1)^2 + \cdots + (x_n - y_n)^2\right)^{\frac{1}{2}}. \tag{2.4}$$

Therefore the product metric on the cartesian product of metric spaces is a generalization of the usual distance formula for Euclidean spaces. We will refer to this metric as the *standard* or *Euclidean metric.*

Positive definiteness of (2.3) is easily checked: Certainly the distance is non-negative, and is 0 if and only if all of the distances $d_i(x_i, y_i)$ are 0, which is equivalent to $x_i = y_i$ for all $i$ (since each $d_i$ is positive definite), which is equivalent to $(x_1, ..., x_n) = (y_1, ..., y_n)$.

**Exercise 75** *Prove the symmetry of (2.3).*

Proving the triangle inequality is more difficult. We will show by induction on $n$ that if we are given

$$x = (x_1, ..., x_n), y = (y_1, ..., y_n), z = (z_1, ..., z_n) \in X_1 \times \cdots \times X_n$$

then $d(x, z) \leq d(x, y) + d(y, z)$. The case $n = 1$ is simply the triangle inequality in the metric space $X_1$. We will now prove the statement for $n = 2$. To simplify the proof, let $a := d(x_1, y_1)$, $b := d(y_1, z_1)$, $c := d(x_2, y_2)$, $d := d(y_2, z_2)$. Then by the triangle inequality for $d_1$ and $d_2$,

$$d(x, z)^2 = d_1(x_1, z_1)^2 + d_2(x_2, z_2)^2 \leq (a + b)^2 + (c + d)^2$$
$$= a^2 + b^2 + c^2 + d^2 + 2(ab + cd).$$

On the other hand,

$$(d(x, y) + d(y, z))^2 = \left(\sqrt{a^2 + c^2} + \sqrt{b^2 + d^2}\right)^2$$
$$= a^2 + c^2 + b^2 + d^2 + 2\sqrt{(a^2 + c^2)(b^2 + d^2)}$$

and the problem reduces to showing that

$$(ab+cd)^2 \leq \left(a^2+c^2\right)\left(b^2+d^2\right) \Leftrightarrow 2abcd \leq c^2b^2+a^2d^2 \Leftrightarrow (ab-cd)^2 \geq 0$$

which is always true. Now suppose that the triangle inequality has been proved for some $n \geq 2$. In particular, we know that the product metric is in fact a metric on $X_1 \times \cdots \times X_n$. Since we have shown the case $n = 2$, we also know that the triangle inequality is valid in the product metric space $(X_1 \times \cdots \times X_n) \times X_{n+1}$, the metric for which we will denote by $\Delta$. But $(X_1 \times \cdots \times X_n) \times X_{n+1}$ is the same set as $X_1 \times \cdots \times X_{n+1}$ (see the comments surrounding to Definition 2). Moreover, the function $d$ on $X_1 \times \cdots \times X_{n+1}$ is computed in the exact same way as $\Delta$: Given

$$x = (x_1, ..., x_n, x_{n+1}) \text{ and } z = (z_1, ..., z_n, z_{n+1}) \text{ in } X_1 \times \cdots \times X_{n+1}.$$

$$\begin{aligned} d(x,z)^2 &= \sum_{i=1}^{n+1} d_i(x_i,z_i)^2 = \sum_{i=1}^{n} d_i(x_i,z_i)^2 + d_{n+1}(x_{n+1},z_{n+1})^2 \\ &= \left(\sqrt{\sum_{i=1}^{n} d_i(x_i,z_i)^2}\right)^2 + d_{n+1}(x_{n+1},z_{n+1})^2 \\ &= \Delta\left(((x_1,...,x_n),x_{n+1}),((z_1,...,z_n),z_{n+1})\right)^2. \end{aligned}$$

That is, the two functions are the same and hence $d$ also satisfies the triangle inequality.

Note that as a consequence of this computation we see that, for example, $(A \times B) \times C$ with the product metric is naturally identified with $A \times (B \times C)$ via the function $f((a, b), c)) = (a, (b, c))$, which is in fact an isometry between these two metric spaces.

There is a useful alternative metric on the cartesian product of metric spaces, defined as follows:

**Definition 107** *Let $(X_1, d_1), ..., (X_n, d_n)$ be a collection of metric spaces. We define the max metric on the cartesian product $X_1 \times \cdots \times X_n$ by*

$$d_{\max}((x_1, ..., x_n), (y_1, ..., y_n)) = \max\{d_1(x_1, y_1), ..., d_n(x_n, y_n)\}.$$

It is somewhat easier to check that this metric is indeed a metric.

**Exercise 76** *Check that the max metric is symmetric and positive definite.*

For the triangle inequality we use the triangle inequality for the spaces $X_i$ and Lemma 14:

$$\begin{aligned} d_{\max}((x_1,...,x_n),(z_1,...,z_n)) &= \max\{d_1(x_1,z_1),...,d_n(x_n,z_n)\} \\ &\leq \max\{d_1(x_1,y_1)+d_1(y_1,z_1),...,d_n(x_n,y_n)+d_n(y_n,z_n)\} \\ &\leq \max\{d_1(x_1,y_1),...,d_n(x_n,y_n)\} + \max\{d_1(y_1,z_1),...,d_n(y_n,z_n)\} \\ &= d_{\max}((x_1,...,x_n),(y_1,...,y_n)) + d_{\max}((y_1,...,y_n),(z_1,...,z_n)). \end{aligned}$$

**Example 108** *The product and max metrics are not the same in general. For example, in the plane we have* $d((1,1),(0,0)) = \sqrt{2}$ *and* $d_{\max}((1,1),(0,0)) = 1$.

**Exercise 77** *Find the set of all "midpoints" between the points* $(-1,0)$ *and* $(1,0)$ *in the plane with the max metric. That is, find all points* $(x,y)$ *in the plane such that*

$$d_{\max}((x,y),(-1,0)) = d_{\max}((x,y),(1,0)) = 1$$

*This shows that, geometrically speaking, the max metric and the usual Euclidean metric (with which every pair of points has a unique midpoint), are quite different.*

Why do we need a second metric on the cartesian product of metric spaces? In many ways the max metric is simpler to work with than the product metric and, as we will see shortly, from standpoint of convergence and continuity they can be used interchangeably. To make this idea precise, we need the following notion:

**Definition 109** *Let* $X$ *be a space with two metrics,* $d_1$ *and* $d_2$. *We say that the two metrics are topologically equivalent if the identity function* $id_X : (X, d_1) \to (X, d_2)$ *is a homeomorphism.*

As an immediate consequence of the exercises in the preceding section, we have the following two lemmas:

**Lemma 110** $(X, d_1)$ *and* $(X, d_2)$ *are topologically equivalent metric spaces if and only if the following statement holds: For every sequence* $(x_i)$ *in* $X$ *and* $x \in X$, $x_i \to x$ *with respect to the metric* $d_1$ *if and only if* $x_i \to x$ *with respect to the metric* $d_2$.

**Lemma 111** $(X, d_1)$ *and* $(X, d_2)$ *are topologically equivalent metric spaces if and only if given any* $U \subset X$, $U$ *is open with respect to* $d_1$ *if and only if* $U$ *is open with respect to* $d_2$.

The above two lemmas show that two metrics on the same set are equivalent if and only if convergence of sequences is the same in both metrics; equivalently the two metrics define the same open (and hence closed) subsets. This means that any definition that depends only on convergence or openness of sets is either satisfied or not satisfied simultaneously for both metrics For example, if $X$ has two topologically equivalent metrics $d_1, d_2$ then a function from $X$ into another metric space (or from another metric space into $X$) is continuous with respect to $d_1$ if and only if it is continuous with respect to $d_2$. And a subset of $X$ is compact with respect to $d_1$ if and only if it is compact with respect to $d_2$.

**Example 112** *Consider the metric space* $\mathbb{R}$ *with the usual metric* $d$ *and the trivial metric, which we will denote for this exercise by* $d_t$. *In a previous exercise it was shown that every subset of a set with the trivial metric is open, whereas* *there are certainly subsets of $\mathbb{R}$ with the usual metric that are not. This means that if $id_X$ is the identity map from $(\mathbb{R}, d)$ to $(\mathbb{R}, d_t)$ then $id_X$ is not continuous, although $id_X^{-1}$ is continuous. So $d$ is not topologically equivalent to $d_t$. On the other hand, the trivial metric and the subspace metric on $\mathbb{Z}$ are topologically equivalent (cf. Exercise 72).*

**Definition 113** *A function $f : X \to Y$ is called Lipschitz if there exists some $\lambda > 0$ such that for all $x, y \in X$, $d_Y(f(x), f(y)) \leq \lambda d(x, y)$.*

A Lipschitz function does not "stretch" any distances by a factor of more than $\lambda$. Note that to prove that a function is Lipschitz, we do not need to consider the case when $x = y$, since then $f(x) = f(y)$ and $d(f(x), f(y)) = 0$.

**Example 114** *Any constant function is Lipschitz for every $\lambda$.*

**Exercise 78** *Prove that every Lipschitz function is uniformly continuous.*

**Exercise 79** *Let $f : X \to Y$ be Lipschitz. Prove that if $A$ is a bounded subset of $X$ then $f(A)$ is a bounded subset of $Y$.*

**Exercise 80** *Prove or disprove whether or not each of the following functions is Lipschitz.*

1. $f(x) = x^2$ *from* $\mathbb{R}$ *to* $\mathbb{R}$.
2. $f(x) = x^2$ *from* $[0, 1]$ *to* $[0, 1]$.
3. $f(x) = \sqrt{x}$ *from* $[0, 1]$ *to* $[0, 1]$.

**Proposition 115** *Let $(X_1, d_1), ..., (X_n, d_n)$ be a collection of metric spaces. Then for every $x, y \in X := X_1 \times ... \times X_n$, $d_{\max}(x, y) \leq d(x, y) \leq \sqrt{n} d_{\max}(x, y)$.*

**Proof.** Let $(x_1, ..., x_n), (y_1, ..., y_n) \in X_1 \times \cdots \times X_n$. Then

$$d_{\max}((x_1, ..., x_n), (y_1, ..., y_n)) = \max\{d_1(x_1, y_1), ..., d_n(x_n, y_n)\}$$

$$= (\max\{d_1(x_1, y_1)^2, ..., d_n(x_n, y_n)^2\})^{\frac{1}{2}} \leq \left(d_1(x_1, y_1)^2 + \cdots + d_n(x_n, y_n)^2\right)^{\frac{1}{2}}$$

$$= d((x_1, ..., x_n), (y_1, ..., y_n)) = \left(d_1(x_1, y_1)^2 + \cdots + d_n(x_n, y_n)^2\right)^{\frac{1}{2}}$$

$$\leq (n \max_i\{d_i(x_i, y_i)^2\})^{\frac{1}{2}} = \sqrt{n} d_{\max}((x_1, ..., x_n), (y_1, ..., y_n)).$$

■

This proposition shows that the identity map from $(X, d)$ to $(X, d_{\max})$ is Lipschitz with $\lambda = 1$ and so is continuous. Likewise, inverse, which is the identity map from $(X, d_{\max})$ to $(X, d)$ is Lipschitz with $\lambda = \sqrt{n}$. Therefore the identity map from $(X, d)$ to $(X, d_{\max})$ is a homeomorphism. In other words,

**Corollary 116** *Let $(X_1, d_1), ..., (X_n, d_n)$ be a collection of metric spaces. Then the product and max metrics on $X := X_1 \times ... \times X_n$ are topologically equivalent.*

**Corollary 117** *Let* $(X_1, d_1), ..., (X_n, d_n)$ *be a collection of metric spaces. Then* $A \subset X := X_1 \times ... \times X_n$ *is bounded with respect to the max metric if and only if* $A$ *is bounded with the product metric.*

**Definition 118** *A function between metric spaces having the property that both it and its inverse are Lipschitz is called bilipschitz, and the metric spaces are said to be bilipschitz equivalent.*

Note that a bilipschitz function is always a homeomorphism. In general, if we are given a collection of metric spaces $(X_1, d_1)$, ...,$(X_n, d_n)$ then unless otherwise stated we will always assume that $X_1 \times \cdots \times X_n$ has the product metric. However, in proofs of purely topological properties (depending only on convergence or open sets) we will sometimes use the max metric because it is considerably simpler in certain situations. The next lemma shows that the open and closed balls in the max metric have a particularly simple structure:

**Lemma 119** *Let* $(X_1, d_1), ..., (X_n, d_n)$ *be a collection of metric spaces and* $X := X_1 \times \cdots X_n$ *with the max metric. Then for every* $(x_1, ..., x_n) \in X$ *and* $r > 0$,

$$B_{\max}(x, r) = B_{X_1}(x_1, r) \times \cdots \times B_{X_n}(x_n, r)$$

*and*

$$C_{\max}(x, r) = C_{X_1}(x_1, r) \times \cdots \times C_{X_n}(x_n, r).$$

**Proof.** We prove only the case for the open ball; the case for the closed ball is similar.

$$B_{\max}(x, r) = \{(y_1, ..., y_n) \in X : d_{\max}((y_1, ..., y_n), (x_1, ..., x_n)) < r\}$$

$$= \{(y_1, ..., y_n) \in X : \max\{d_1(y_1, x_1), ..., d_n(y_n, x_n)\} < r\}$$

$$= \{(y_1, ..., y_n) \in X : d_i(y_i, x_i) < r \text{ for all } i\}$$

$$= \{(y_1, ..., y_n) \in X : y_i \in B_{X_i}(x_i, r) \text{ for all } i\}$$

$$B_{X_1}(x_1, r) \times \cdots \times B_{X_n}(x_n, r)$$

■

**Exercise 81** *Let* $(X_1, d_1), ..., (X_n, d_n)$ *be a collection of metric spaces and* $U_i \subset X_i$ *be open, for all* $i$*. Show that* $U := U_1 \times \cdots \times U_n$ *is open in* $X := X_1 \times \cdots \times X_n$ *with the max metric, and hence with the product metric.*

We will see later (Exercise 85) that an open ball in $X$ with the product metric need not be the cartesian product of open sets $U_i$ in $X_i$.

**Exercise 82** *Let* $(X_1, d_1), ..., (X_n, d_n)$ *be a collection of metric spaces and let* $X := X_1 \times \cdots \times X_n$*.Let* $U \subset X$*. Show that* $U$ *is open in* $X$ *if and only if* $U$ *is a union of open sets of the form* $U_1 \times \cdots \times U_n$ *for some open* $U_i \subset X_i$*.*

**Exercise 83** *Let $X$ be a metric space with distance $d$. Prove that the distance function $d : X \times X \to \mathbb{R}$ is continuous. Hint: Use the max metric on $X \times X$ and Proposition 73.*

**Exercise 84** *Let $(X_1, d_1), ..., (X_n, d_n)$ be a collection of metric spaces. Define a metric on the cartesian product $X_1 \times \cdots \times X_n$ by*

$$d_+((x_1, ..., x_n), (y_1, ..., y_n)) = d_1(x_1, y_1) + \cdots + d(x_n, y_n).$$

*Prove that $d_+$ is a metric, and that it is topologically equivalent to the product and max metrics. Hint: Being topologically equivalent is transitive; decide which of the product or max metrics would be the easiest to work with.*

## 2.8 Euclidean spaces

We now consider the special case of Euclidean space $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$, with the Euclidean metric. The elements of $\mathbb{R}^n$ are $n$-tuples $(x_1, ..., x_n)$ where each $x_i$ is a real number and we will refer to these elements as *vectors or $n$-vectors*. Note that $\mathbb{R}^1 = \mathbb{R}$, and we do not denote elements of $\mathbb{R}$ as "1-tuples". To simplify notation we will often denote elements of $\mathbb{R}^n$ using a single boldface character (e.g. $\mathbf{x} = (x_1, ..., x_n)$). As is traditional, will often denote vectors in $\mathbb{R}^2$ by $(x, y)$ and those in $\mathbb{R}^3$ by $(x, y, z)$.

When $n = 1$, as we have already observed, open balls are simply open intervals. When $n = 2$, the open ball centered at $(x, y)$ of radius $r$ is

$$B((x, y), r) = \{(x_1, y_1) : (x_1 - x)^2 + (y_1 - y)^2 < r^2\}$$

which of course is the region inside the circle of radius $r$ centered at $(x, y)$. Likewise, when $n = 3$ open balls are the interiors of spheres, and so on. In $\mathbb{R}^n$ we will denote $(0, 0, ..., 0)$ simply by $\mathbf{0}$ (the reader needs to be vigilant because the same (unbold) symbol is used to represent the real number 0).

**Exercise 85** *Prove that $B(\mathbf{0}, 1) \subset \mathbb{R}^2$ is not the cartesian product of two subsets of $\mathbb{R}$. Hint: Show that if $B(\mathbf{0}, 1)$ were equal to some set $U_1 \times U_2$ then the point $(\frac{3}{4}, \frac{3}{4})$ would have to lie in $B(\mathbf{0}, 1)$.*

Recall that $\mathbb{R}^n$ is a vector space with addition given by

$$(x_1, ...x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n)$$

and scalar multiplication given by

$$k(x_1, ..., x_n) = (kx_1, ..., kx_n) \text{ for } k \in \mathbb{R}.$$

We will not discuss the basic properties of vector spaces; these properties are very natural and easy to check for $\mathbb{R}^n$ and the reader likely uses them instinctively at this point.

We have already seen that there are two (equivalent) versions of the triangle inequality for $\mathbb{R}$: the triangle inequality for the metric and that for the absolute value (Proposition 7). For $\mathbb{R}^n$ the situation is analogous; first, the absolute value is replaced by the norm, which is defined as follows:

**Definition 120** *For any* $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ *we define the norm of* $\mathbf{x}$ *by*

$$\|\mathbf{x}\| := d(\mathbf{x}, \mathbf{0}) = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Note that in $\mathbb{R}^n$

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

and that for $n = 1$ the norm is the same as the absolute value. A vector $\mathbf{v}$ such that $\|\mathbf{v}\| = 1$ is called a *unit vector.* We will continue to use the absolute value notation in connection with real numbers (especially when they are used for scalar multiplication of vectors), but the norm notation is also legitimate for $\mathbb{R}$ and will be used when making statements about $\mathbb{R}^n$ in general, as in the next exercise.

**Exercise 86** *Show that, for* $\mathbf{x} \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$*,*

1. $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$
2. $\|c\mathbf{x}\| = |c| \, \|\mathbf{x}\|$

**Theorem 121** *(Triangle inequality) For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*,*

$$\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|.$$

**Proof.** Applying the triangle inequality for the distance to the points $\mathbf{x}, \mathbf{y}, \mathbf{0} \in \mathbb{R}^n$ we obtain

$$\begin{aligned}\|\mathbf{x} + \mathbf{y}\| &= \|\mathbf{x} - (-\mathbf{y})\| = d(\mathbf{x}, -\mathbf{y}) \le d(\mathbf{x}, \mathbf{0}) + d(\mathbf{0}, -\mathbf{y}) \\ &= \|\mathbf{x} - \mathbf{0}\| + \|\mathbf{0} - (-\mathbf{y})\| = \|\mathbf{x}\| + \|\mathbf{y}\|.\end{aligned}$$

■

Recall that the dot product in $\mathbb{R}^n$ is defined, for $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ by

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^{n} x_i y_i.$$

The dot product satisfies the following properties, which can be proved simply using the definition:

1. (Symmetry) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ for all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$
2. (Bilinearity) $(k\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = k(\mathbf{x} \cdot \mathbf{z}) + \mathbf{y} \cdot \mathbf{z}$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in $\mathbb{R}^n$ and $k \in \mathbb{R}$

3. (Positive definiteness) $\mathbf{x} \cdot \mathbf{x} \geq 0$ and $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}$

4. $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}$.

The following important theorem is logically equivalent to the triangle inequality (see Exercise 87).

**Theorem 122** *(Cauchy-Schwarz Inequality) For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have*

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$$

**Proof.** By the triangle inequality,

$$(\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \geq \|\mathbf{x} + \mathbf{y}\|^2 = (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2$$

while we also have

$$(\|\mathbf{x}\| + \|\mathbf{y}\|)^2 = \|\mathbf{x}\|^2 + 2\,\|\mathbf{x}\|\,\|\mathbf{y}\| + \|\mathbf{y}\|^2 .$$

Combining these formulas leads to

$$\mathbf{x} \cdot \mathbf{y} \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$$

Applying this formula to $\mathbf{x}$ and $-\mathbf{y}$ leads to

$$-(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot (-\mathbf{y}) \leq \|\mathbf{x}\| \, \|-\mathbf{y}\| = \|\mathbf{x}\| \, \|\mathbf{y}\|$$

The theorem is now proved by Lemma 8. ■

**Exercise 87** *Use the Cauchy-Schwarz inequality to prove the triangle inequality for the norm in $\mathbb{R}^n$. Hint: Start with $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2$ as in the above proof and use the fact that $\mathbf{x} \cdot \mathbf{y} \leq |\mathbf{x} \cdot \mathbf{y}|$.*

Let's take a moment to examine the max metric on $\mathbb{R}^n$. We have the formula

$$d_{\max}((x_1, ..., x_n), (y_1, ..., y_n)) = \max\{|x_1 - y_1|, ..., |x_n - y_n|\}$$

For $\mathbb{R}$ this formula simply reduces to the standard metric.

**Example 123** *In $\mathbb{R}^2$ we see from Lemma 119 that the max metric open ball centered at $(x, y)$ of radius $r > 0$ is the interior of a square with sides parallel to the $x$- and $y$-axes having side length $2r$. In $\mathbb{R}^3$ such an open ball is the interior of a cube, and so on. That is, the open ball of radius $r > 0$ centered at $(x_1, ..., x_n) \in \mathbb{R}^n$ with respect to the max metric is equal to the cartesian product*

$$(x_1 - r, x_1 + r) \times \cdots \times (x_n - r, x_n + r).$$

*For closed balls we simply replace the open intervals in the product with closed intervals of the same radius.*

**Exercise 88** *Describe the open balls in* $\mathbb{R}^2$ *with the metric* $d_+$ *(see Exercise 84).*

**Definition 124** *A norm on a (real) vector space* $V$ *is a function* $\|\|$ *that assigns to each* $\mathbf{v} \in V$ *a real number* $\|\mathbf{v}\|$ *such that*

1. $\|\mathbf{v}\| \geq 0$ and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
2. for any $c \in \mathbb{R}$, $\|c\mathbf{v}\| = |c| \, \|\mathbf{v}\|$
3. for any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.

In Definition 120 we defined the standard norm on $\mathbb{R}^n$, showing in Exercise 86 and Theorem 121 that it satisfies the above definition. As another example, let $\|\mathbf{v}\|_{\max}$ denote $d_{\max}(\mathbf{v}, \mathbf{0})$. It is easy to check that the first two conditions of Definition 124 are satisfied; the proof of the triangle inequality is identical to the proof of Theorem 121. Likewise the metric $d_+$ from Example 84 defines a norm, but not every metric on $\mathbb{R}^n$ gives rise to a norm in this way (see Exercise 90). As we will see below every norm on a vector space does give rise to a metric, moreover a very special kind of metric.

**Definition 125** *A metric* $d$ *on a vector space* $V$ *is called (translation) invariant if for all* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z})$.

**Exercise 89** *Prove that a metric* $d$ *on a vector space* $V$ *is invariant if and only if for every* $\mathbf{y} \in V$, *the function* $T_{\mathbf{y}} : V \to V$ *defined by* $T_{\mathbf{y}}(\mathbf{x}) := \mathbf{y} + \mathbf{x}$ *is an isometry.*

Invariance of metrics on vector spaces is an extremely important concept because it establishes a compatibility between the algebraic (vector space) structure and metric structure. Not all metrics on $\mathbb{R}^n$ are invariant, but in this text we will not be concerned with metrics that are not. Some of the most important metrics on $\mathbb{R}^n$ are given by the following proposition:

**Proposition 126** *Let* $\|\|$ *be a norm on a vector space* $V$. *Defining* $d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|$ *always defines an invariant metric, called the metric induced by the norm.*

**Proof.** Symmetry and positive definiteness are immediate consequences of the definitions of metric and norm. For any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$,

$$d(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\| = \|\mathbf{x} - \mathbf{y} + \mathbf{y} - \mathbf{z}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathbf{z}\| = d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}).$$

Finally,

$$d(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z}) = \|\mathbf{x} + \mathbf{z} - (\mathbf{y} + \mathbf{z})\| = \|\mathbf{x} - \mathbf{y}\| = d(\mathbf{x}, \mathbf{y}).$$

■

Since $\|\mathbf{x}\| = d(x, 0)$, from Proposition 73 we now obtain:

**Corollary 127** *Let $\|\|$ be a norm on a vector space $V$. Then the function $f : V \to \mathbb{R}$ defined by $f(\mathbf{x}) = \|\mathbf{x}\|$ is continuous.*

**Corollary 128** *Let $Y$ be a metric space, $V$ be a normed vector space, and $f : Y \to V$ be a continuous function. If $f(y) \neq \mathbf{0}$ then there exists some $r > 0$, such that if $d(x, y) < r$ then $f(x) \neq \mathbf{0}$.*

**Proof.** The function $g$ defined by $g(x) := \|f(x)\| : Y \to \mathbb{R}$ is the composition of continuous functions and is hence continuous, and $g(y) = \|f(y)\| > 0$. Then there exists some $r > 0$ such that if $d(x, y) < r$ then $|\|f(x)\| - \|f(y)\|| = |g(x) - g(y)| < \frac{\|f(y)\|}{2}$, which implies $\|f(x)\| > \|f(y)\| - \frac{\|f(y)\|}{2} = \frac{\|f(y)\|}{2} > 0$, and hence $f(x) \neq \mathbf{0}$. ■

**Exercise 90** *Show that if $d_t$ denotes the trivial metric on $\mathbb{R}^n$ then $d_t$ is invariant, but defining $\|\mathbf{v}\|_t := d_t(\mathbf{v}, \mathbf{0})$ does not define a norm. In other words, every norm on $\mathbb{R}^n$ defines an invariant metric, but the converse is not true.*

**Exercise 91** *Let $\|\|$ be any norm on $\mathbb{R}^n$. Show that the identity map*

$$i : (\mathbb{R}^n, \|\|_{\max}) \to (\mathbb{R}^n, \|\|)$$

*is Lipschitz. Hint: Let $n_i := \|\mathbf{e}_i\|$, where $\mathbf{e}_i := (0, ..., 0, 1, 0, ..., 0)$ with "1" in the $i^{th}$ coordinate. Express any $\mathbf{v} \in \mathbb{R}^n$ as $\mathbf{v} = \sum_{i=1}^{n} v_i e_i$.*

We will need the following proposition somewhat later.

**Proposition 129** *If $U \subset \mathbb{R}^n$ is open then $U$ is a countable union of open balls in the product metric (resp. the max metric).*

**Proof.** First note that the set

$$Q(U) := \{(y_1, ..., y_n) \in U : y_i \in \mathbb{Q} \text{ for all } i\}$$

is countable. Now consider the countable collection

$$\mathcal{B}_{\max} := \{B_{\max}(\mathbf{q}, r) : \mathbf{q} \in Q(U),\ r \in \mathbb{Q} \text{ and } B_{\max}(\mathbf{q}, r) \subset U\}.$$

Certainly the union of all such open balls (in either metric) is contained in $U$. To show that $U$ is contained in this union, let $\mathbf{x} = (x_1, ..., x_n) \in U$. Since $U$ is open there exists some $B_{\max}(\mathbf{x}, \rho) \subset U$. Let $r$ be rational such that $0 < r < \frac{\rho}{2}$. For all $i$, choose some rational $q_i$ such that $|q_i - x_i| < r$. Then $\mathbf{y} := (y_1, ..., y_n)$ satisfies $d_{\max}(\mathbf{x}, \mathbf{y}) < r$ and so $\mathbf{y} \in B_{\max}(\mathbf{x}, r) \subset B_{\max}(\mathbf{x}, \rho) \subset U$. Therefore $\mathbf{y} \in Q(U)$ and $B(\mathbf{y}, r) \in \mathcal{B}_{\max}$. On the other hand, since $d_{\max}(\mathbf{x}, \mathbf{y}) < r$ implies $\mathbf{x} \in B(\mathbf{y}, r) \in \mathcal{B}_{\max}$, and so $\mathbf{x}$ lies in the union of all balls in $\mathcal{B}_{\max}$. The proof for the product metric is an exercise. ■

**Exercise 92** *Finish the proof of Proposition 129.*

## 2.9 Sequences and compactness in product spaces

Throughout this section, let $X_1, ..., X_n$ be metric spaces and denote the product metric space by $X := X_1 \times ... \times X_n$. At this point there should be no confusion if we denote the metrics in all of these spaces, as well as the product metric, by $d$. Note that all results in this section are equally valid for the max metric, since it is topologically equivalent to the product metric. We will also have occasion to use the max metric in certain proofs.

The first thing that we need to consider is the behavior of sequences, and here there *is* potential for confusion, since we are using subscript notation for both sequences and components of elements of $X$. We will resolve this problem by using double subscripts. Each term in a sequence $(y_j)_{j=1}^{\infty}$ in $X$ has $n$ components; we will write $y_j = (y_{j1}, y_{j2}, ..., y_{jn})$ where each $y_{ji} \in X_i$. If we fix some $i$, then we obtain a sequence $(y_{ji})_{j=1}^{\infty}$ in $X_i$, which is called the $i^{th}$ component sequence of $(y_j)_{j=1}^{\infty}$. Many of the properties of a sequence in $X$ are completely determined by the properties of its components. For example, the following proposition generalizes a familiar theorem from calculus of three variables.

**Proposition 130** *If $(y_j)_{j=1}^{\infty}$ is a sequence in $X = X_1 \times \cdots \times X_n$, then $y_j \to a = (a_1, ..., a_n)$ in $X$ if and only if for all $i$, $y_{ji} \to a_i$. In other words the components of the limit are the limits of the components.*

**Proof.** Since the product metric and the max metric are topologically equivalent, we can prove this proposition for convergence in the max metric, which is simpler to work with. Suppose first that $y_j \to a$. Equivalently, this means that $d_{\max}(y_j, a) \to 0$. But for all $i$,

$$0 \leq d(y_{ji}, a_i) \leq \max\{d(y_{ji}, a_i)\} = d_{\max}(y_j, a)$$

and so $d(y_{ji}, a_i) \to 0$ for all $i$ by the Sandwich Theorem. Conversely, suppose that $d(y_{ji}, a_i) \to 0$ for all $i$. Then for every $\varepsilon > 0$ and every $i$ there exists an $N_i$ such that if $j \geq N_i$ then $d(y_{ji}, a_i) < \varepsilon$. Therefore if $N = \max\{N_1, ..., N_n\}$ and $j \geq N$ then

$$d(y_j, a) = \max\{d(y_{ji}, a_i)\} < \varepsilon.$$

This proves that $d_{\max}(y_j, a) \to 0$ and therefore $y_j \to a$. ■

**Example 131** *In $\mathbb{R}^2$, the sequence $(\frac{1}{i}, 1 - \frac{1}{i})$ converges to $(0, 1)$ because $\frac{1}{i} \to 0$ and $1 - \frac{1}{i} \to 1$.*

The reader has already learned a special case of the next corollary in calculus– that a vector valued function is continuous if and only the component functions are continuous.

**Corollary 132** *Let $Y$ be a metric space and $f : Y \to X := X_1 \times \cdots \times X_n$ be a function. Then $f$ is continuous at $y \in Y$ if and only if each component $f_i : Y \to X_i$ is continuous at $y$.*

**Proof.** Let $(y_j)_{j=1}^{\infty}$ be a sequence in $Y$ such that $y_j \to y$. If each $f_i$ is continuous at $y$, then for all $i$, $(f_i(y_j)) \to f_i(y)$. Since $f_i(y_j)$ is the $i^{th}$ component of $f(y_j)$, this implies $f(y_j) \to f(y)$ and $f$ is continuous at $y$. The converse is an exercise. ■

**Exercise 93** *Finish the proof of Corollary 132.*

**Corollary 133** *For all $i$, the projections $\pi_i : X \to X_i$ are continuous.*

**Exercise 94** *Apply Corollary 132 to the identity map $id_X : X \to X$ to prove the above corollary.*

**Exercise 95** *Let $(y_j)_{j=1}^{\infty}$ be a sequence in $X = X_1 \times \cdots \times X_n$. Prove or disprove the following:*

1. A point $a = (a_1, ..., a_n)$ is a cluster point of $(y_j)_{j=1}^{\infty}$ in $X$ if and only if $a_i$ is a cluster point of $(y_{ji})_{j=1}^{\infty}$ in $X_i$ for all $i$.

2. The sequence $(y_j)_{j=1}^{\infty}$ is bounded in $X$ if and only if $(y_{ji})_{j=1}^{\infty}$ is bounded in $X_i$ for all $i$.

3. The sequence $(y_j)_{j=1}^{\infty}$ is unbounded in $X$ if and only if $(y_{ji})_{j=1}^{\infty}$ is unbounded in $X_i$ for every $i$.

We would like to take this opportunity to introduce a notational shortcut used frequently in mathematics. From now on the phrase "for all large $n$" will replace the statement "there exists an $N$ such that for all $n \geq N$." This notation can save quite a bit of effort. For example, in the proof of Proposition 130, rather than choosing all those $N_i$'s and then taking the maximum, we could simply have stated the following: "Then for every $\varepsilon > 0$ and every $i$, $d(y_{ji}, a_i) < \varepsilon$ for all large $j$." The idea is that if finitely many statements are each true for all large $i$ then all the statements are simultaneously true for all large $i$. As another example, we can restate the definition of $x_i \to x$ as "for any $\varepsilon > 0$, $d(x_i, x) < \varepsilon$ for all large $i$". Likewise, we will replace the statement "for every $N \in \mathbb{N}$ there exists an $i \geq N$" by "for some large $i$". Therefore the definition that $x$ is a cluster point of a sequence $(x_i)$ can be stated as "for every $\varepsilon > 0$, $d(x, x_i) < \varepsilon$ for some large $i$". Now the difference between limit point and cluster point is much more clear: one simply changes an "all" to a "some". Warning: unlike the case with all large $i$, it is possible that finitely many statements may be true for some large $i$, but they are not all simultaneously true for some large $i$. For example, if $x_n = (-1)^n$ then $x_n = 1$ for some large $n$ and $x_n = -1$ for some large $n$. But certainly it is not true that $x_n = 1$ and $x_n = -1$ for some large $n$.

**Theorem 134** *If each $X_i$ is a compact metric space then $X = X_1 \times \cdots \times X_n$ is compact.*

**Proof.** We will prove the theorem by induction in $n$. For $n = 1$ there is nothing to prove; suppose $n = 2$ and $(y_j)_{j=1}^{\infty}$ be a sequence in $X = X_1 \times X_2$.

We will construct a convergent subsequence of $(y_j)_{j=1}^{\infty}$. Since $X_1$ is compact, there is a subsequence $(y_{j_k 1})$ of $(y_{j1})$ such that $y_{j_k 1} \to a_1 \in X_1$. Now consider the subsequence $(y_{j_k})_{k=1}^{\infty}$. Since $X_2$ is compact, $(y_{j_k 2})_{k=1}^{\infty}$ has a subsequence $y_{j_{k_m} 2} \to a_2$ in $X_2$, and since $(y_{j_{k_m} 1})$ is a subsequence of $(y_{j_k 1})$, we still have $y_{j_{k_m} 1} \to a_1$. Finally, the sequence $(y_{j_{k_m}})_{m=1}^{\infty}$ is a subsequence of $(y_j)$ such that $y_{j_{k_m} 1} \to a_1$ and $y_{j_{k_m} 2} \to a_2$, and by Proposition 130, $y_{j_{k_m}} \to (a_1, a_2)$.

Now suppose that we have proved the theorem for some $n \geq 2$. As we have previously observed, $X_1 \times \cdots \times X_{n+1}$ is naturally isometric to $(X_1 \times \cdots \times X_n) \times X_{n+1}$. By the inductive hypothesis, $X_1 \times \cdots \times X_n$ is compact, and by the case $n = 2$ that we have just proved, $(X_1 \times \cdots \times X_n) \times X_{n+1}$ is also compact. ■

**Exercise 96** *Prove or disprove the converse of Theorem 134.*

**Theorem 135** *(Heine-Borel Theorem) A subset $A$ of $\mathbb{R}^n$ is compact if and only if $A$ is closed and bounded.*

**Proof.** If $A$ is bounded in the product metric then $A$ bounded in the max metric and so by definition is contained in a closed ball $B$ in the max metric. But such a ball, according to Example 123 is the cartesian product of closed bounded intervals, and by the Heine-Borel Theorem for the real numbers and Theorem 134 $B$ is compact. Therefore $A$ is a closed subset of a compact set and hence compact. The converse is Proposition 76. ■

**Corollary 136** *If $f : A \to \mathbb{R}$ is a continuous function defined on a subset $A$ of $\mathbb{R}^n$ and $B \subset A$ is closed and bounded then $f$ has a maximum and a minimum on $B$.*

Since every bounded sequence in $\mathbb{R}^n$ is contained in a closed bounded (hence compact) ball, we immediately obtain:

**Corollary 137** *(Bolzano-Weierstrass Theorem) Every bounded sequence in $\mathbb{R}^n$ has a cluster point.*

**Definition 138** *The unit sphere in $\mathbb{R}^n$ is the set $\mathbb{S}^{n-1} := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$.*

Note that the sphere is, in a sense that is intuitively clear but more difficult to define, of one dimension lower than the space in which it lives, hence the "$n - 1$" superscript.

**Exercise 97** *For any $n$, prove $\mathbb{S}^{n-1}$ is compact. You may not use any sequences in your proof! A similar argument shows that $S_{\max}^{n-1} := \{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\|_{\max} = 1\}$ is compact. Hint: $\{1\}$ is closed in $\mathbb{R}$.*

**Exercise 98** *For this exercise you will need Exercise 91. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$.*

1. *Show that there is some $m > 0$ such that $\|\mathbf{v}\| \geq m$ for all $\mathbf{v} \in \mathbb{R}^n$ such that $\|\mathbf{v}\|_{\max} = 1$.*

2. *Prove that for every* $\mathbf{v} \in \mathbb{R}^n$, $\|\mathbf{v}\| \geq m \|\mathbf{v}\|_{\max}$.

3. *Show that the metric spaces of these two norms are bilipschitz equivalent and hence topologically equivalent.*

The above exercise in fact proves the following important statement, which shows that when it comes to topological issues in $\mathbb{R}^n$, one may choose the most convenient norm for a given purpose. We will see later that the same statement is not true for norms on an infinite dimensional vector space. As the proof shows, what is essential about $\mathbb{R}^n$ is that the set of unit vectors is compact.

**Theorem 139** *Any two norms on* $\mathbb{R}^n$ *are bilipschitz equivalent (hence topologically equivalent), in the sense that they induce bilipschitz equivalent metrics.*

## 2.10 Connected Metric Spaces

The definition of connected set may seem a bit strange and abstract at first; however, applications such as the Intermediate Value Theorem should increase the reader's appreciation of this definition.

**Definition 140** *A metric space* $X$ *is said to be connected if* $X$ *is not the union of two disjoint non-empty open sets.*

Note that the empty set is connected. Definitions involving negatives can be tricky to work with; often one proceeds by contradiction. There are various direct approaches to showing that a space $X$ is connected. For example, one can suppose that $X = U \cup V$, suppose that $U$ and $V$ are disjoint, and prove that $U$ or $V$ is empty. Or one can use an approach given by the following exercises:

**Exercise 99** *Prove that a metric space* $X$ *is connected if and only if the only subsets of* $X$ *that are both open and closed are* $X$ *and* $\varnothing$. *As a consequence we see that any metric space having at least two points, with the trivial metric, is not connected.*

**Exercise 100** *Prove that a metric space* $X$ *is connected if and only if whenever* $X = U \cup V$ *where* $U$ *and* $V$ *are disjoint and open, if* $U$ *is non-empty then* $X \subset U$.

**Proposition 141** *Let* $f : X \to Y$ *be a continuous function between metric spaces. If* $X$ *is connected then* $f(X)$ *is connected (with the subspace metric).*

**Proof.** We will use Exercise 99. Let $Z$ be a closed and open subset of $f(X)$. Since $f$ is continuous, $f^{-1}(Z)$ is an open and closed subset of $X$ (see Lemma 101) and hence is either $X$ or $\varnothing$. In the first case, $Z = f(X)$ and in the second case $Z \subset f(f^{-1}(Z)) = \varnothing$. ■

**Proposition 142** *Let* $\{A_\lambda\}_{\lambda \in \Lambda}$ *be a collection of connected subsets of a metric space* $X$ *(with the subspace metric) such that* $X = \bigcup_{\lambda \in \Lambda} A_\lambda$ *and* $\bigcap_{\lambda \in \Lambda} A_\lambda \neq \varnothing$. *Then* $X$ *is connected.*

**Proof.** Let $x \in \bigcap_{\lambda\in\Lambda} A_\lambda$ and suppose that $X = U \cup V$, where $U$ and $V$ are open in $X$ and disjoint; then $x$ lies in one of these sets, say $x \in U$. We will show that $X \subset U$ and apply Exercise 100. Since $X = \bigcup_{\lambda\in\Lambda} A_\lambda$ we need only show that for every $\lambda \in \Lambda$, $A_\lambda \subset U$. This follows from Exercise 100, since $U \cap A_\lambda$ and $V \cap A_\lambda$ are disjoint, open in $A_\lambda$, their union is $A_\lambda$, and $x \in U \cap A_\lambda$. ∎

**Lemma 143** *If $A$ is a connected subset of a metric space $X$ and $B$ is a subset of $X$ such that $A \subset B \subset \overline{A}$ then $B$ is connected.*

**Proof.** Suppose $B = U \cup V$ where $U$ and $V$ are open in $B$, disjoint and nonempty. Then $U \cap A$ and $V \cap A$ are open in $A$ and disjoint. Since $A$ is connected, one of them is empty; say $V \cap A = \varnothing$. But $V$ is non-empty and so there is some $x \in V \subset B$. But $x \in \overline{A}$ implies $V \cap A \neq \varnothing$, a contradiction. ∎

**Example 144** *The set $A := (0,1) \cup (1,2)$ is a union of two disjoint open sets, and hence is not connected. However, $\overline{A} = (0,2)$ is an interval and, as we will see below, is connected. Therefore the converse of the above lemma is false.*

We are now in a position to completely characterize the connected subsets of $\mathbb{R}$.

**Theorem 145** *A nonempty subset $S$ of $\mathbb{R}$ is connected if and only if $S$ is an interval.*

**Proof.** Consider first an open interval $(a,b)$. Suppose that $(a,b) = U \cup V$ where $U$ and $V$ are open in $(a,b)$ (and hence in $\mathbb{R}$), disjoint, and nonempty. Without loss of generality we can suppose that there exist $x \in U$ and $y \in V$ with $x < y$. Let $A = U \cap [x,y]$ and let $s = \sup A$. Then $a < x \leq s \leq y < b$, so $s \in (a,b)$. Moreover, since $V$ is open there exists some $\varepsilon > 0$ such $(y-\varepsilon, y+\varepsilon) \subset V$ and therefore $(y-\varepsilon, y+\varepsilon) \cap U = \varnothing$. In particular, $y \neq s$, i.e. $s < y$. We will obtain a contradiction by showing both $s \notin U$ and $s \notin V$. Suppose $s \in U$. Then since $U$ is open and $s < y$ there exists some $s_0 \in U$ such that $s < s_0 < y$. This contradicts the fact that $s$ is an upper bound of $A$ and proves $s \notin U$. Now suppose $s \in V$. Since $V$ is open there is some $\varepsilon > 0$ such that $(s-\varepsilon, s+\varepsilon) \subset V$ and therefore $(s-\varepsilon, s+\varepsilon) \cap U = \varnothing$, which contradicts the approximation property, so $s \notin V$ and the proof that $(a,b)$ is connected is finished.

Every bounded interval is contained in the closure of some interval $(a,b)$ and so is connected by Lemma 143. Unbounded intervals can be handled by Proposition 142. For example, an interval $(a,\infty)$ is a union of the intervals $(a, a+n)$, for $n \in 2, 3, ...$, which are all connected and contain the point $a+1$.

Conversely, suppose that $S$ is a nonempty connected subset of $\mathbb{R}$. Let $a := \inf S$ and $b := \sup S$, where $a$ and $b$ are extended real numbers. If $a = b$ then $S = \{a\}$ and we are finished. If $a < b$ we will show that $(a,b) \subset S$. Suppose, to the contrary, there is some $c \notin S$ such that $a < c < b$. Letting $U := S \cap (a,c)$ and $V := S \cap (c,b)$ we have that each of $U$ and $V$ is open in $S$, and the two sets are disjoint. In addition, if $U$ were empty then the interval $(a,c)$ would contain no element of $S$, which would contradict $a = \inf S$. Therefore $U \neq \varnothing$. A similar proof shows $V \neq \varnothing$, which contradicts the connectedness of $S$. Since $(a, b) \subset S$ and $a = \sup S$ and $b = \sup S$, $S$ consists of $(a, b)$ together with possibly $a$ and/or $b$ and therefore is an interval. ■

**Theorem 146** *(Intermediate Value Theorem) Let $f : X \to \mathbb{R}$ be a continuous function, where $X$ is a connected metric space. For any $a, b \in X$ such that $f(a) < f(b)$, and for any $k \in (f(a), f(b))$, there exists some $c \in X$ such that $f(c) = k$.*

**Proof.** Since $X$ is connected, $f(X)$ is connected, and hence is an interval $I$. Therefore, since $f(a)$ and $f(b)$ lie in $I$, so does $k$–i.e., $k \in f(X)$. But then by definition $k = f(c)$ for some $c \in X$. ■

**Corollary 147** *(Intermediate Value Theorem for Real Functions) Let $f : A \to \mathbb{R}$ be a continuous function, where $A \subset \mathbb{R}$ is connected. For any $a < b$ (resp. $b < a$) in $A$ such that $f(a) < f(b)$, and for any $k \in (f(a), f(b))$, there exists some $c$ such that $a < c < b$ (resp. $b < c < a$) and $f(c) = k$.*

**Proof.** We suppose $a < b$; the other case is similar. Since $A$ is connected, $A$ is an interval and hence contains the interval $[a, b]$. Applying the Intermediate Value Theorem to the restriction of $f$ to $[a, b]$ means there exists some $c \in [a, b]$ such that $f(c) = k$. Since $k$ is not equal to $f(a)$ or $f(b)$ $c$ cannot be equal to $a$ or $b$, and hence $c \in (a, b)$. ■

**Definition 148** *A real function $f : A \to \mathbb{R}$ is called increasing (resp. decreasing, strictly increasing, strictly decreasing) on $A$ if for every $x < y$ in $A$ we have $f(x) \leq f(y)$ (resp. $f(x) \geq f(y)$, $f(x) < f(y)$, $f(x) > f(y)$).*

**Corollary 149** *Let $f : [a, b] \to \mathbb{R}$ be continuous and one-to-one, where $a < b$. Then $f$ is either strictly increasing or strictly decreasing.*

**Proof.** Since $f$ is one-to-one, either $f(b) > f(a)$ or $f(b) < f(a)$; suppose the first. Let $y \in [a, b)$ and $x \in (a, b]$ with $y < x$. If $f(x) > f(b)$ then by the intermediate value theorem there must be some $c \in (a, x)$ such that $f(c) = f(b)$, a contradiction. This proves $f(x) \leq f(b)$ and similarly $f(a) \leq f(x)$. If $f(y) > f(x)$ then as before there must be some $d \in (a, y)$ such that $f(d) = f(x)$, a contradiction. Since $f(y) \neq f(x)$, the only possibility is $f(y) < f(x)$. ■

Connectedness can be difficult to verify; it is often easier to verify the following condition:

**Definition 150** *A metric space $X$ is called arcwise connected (or path connected) if there exists some $x \in X$ such that for every $y \in X$ there is a continuous function $\alpha : [0, 1] \to X$ (called a curve, path or arc) such that $\alpha(0) = x$ and $\alpha(1) = y$.*

**Proposition 151** *Any arcwise connected metric space is connected.*

**Proof.** Let $X$ be arcwise connected; so there exists an $x \in X$ such that for every $y \in X$ there exists a continuous function $\alpha_y : [0,1] \to X$ such that $\alpha_y(0) = x$ and $\alpha_y(1) = y$. Let $A_y := \alpha_y([0,1])$; then $A_y$ is a connected subset of $X$ that contains both $x$ and $y$. In particular, $X = \bigcup_{y \in X} A_y$ and $x \in \bigcap_{y \in X} A_y$. By Proposition 142, $X$ is connected. ■

**Example 152** *In $\mathbb{R}^n$, every $B(\mathbf{x}, r)$ is an arcwise connected metric space. In fact, given any $\mathbf{y} \in B(\mathbf{x}, r)$ define $f_{\mathbf{y}}(t) = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$. If $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ then in component form*

$$f_{\mathbf{y}}(x_1, ..., x_n) = (x_1 + t(y_1 - x_1), ..., x_n + t(y_n - x_n))$$

*and each component is a real linear function, hence, by Exercise 55, continuous. Therefore $f_{\mathbf{y}}$ is continuous. Certainly $f_{\mathbf{y}}(0) = \mathbf{x}$ and $f_{\mathbf{y}}(1) = \mathbf{y}$. Finally, for all $t$ we have*

$$\|f_{\mathbf{y}}(t) - \mathbf{x}\| = \|\mathbf{x} + t(\mathbf{y} - \mathbf{x}) - \mathbf{x}\| = |t| \, \|\mathbf{y} - \mathbf{x}\| < tr \leq r$$

*Therefore $f_{\mathbf{y}}(t) \in B(\mathbf{x}, r)$ for all $t$ and by definition $f_{\mathbf{y}} : [0,1] \to B(\mathbf{x}, r)$. It follows from Proposition 151 that $B(\mathbf{x}, r)$ is connected; the reader should give a few moments' thought to the difficulty of proving this fact using the definition of connected. Certainly the proof would be significantly more difficult than the first part of the proof of Theorem 145, which is the special case when $n = 1$.*

**Exercise 101** *Prove or disprove: If $f : X \to Y$ is a continuous function between metric spaces and $f(X)$ is connected then $X$ is connected.*

**Exercise 102** *Prove or disprove: If $f : X \to Y$ is a continuous function between metric spaces and $X$ is arcwise connected then $f(Y)$ is arcwise connected.*

**Remark 153** *There exist metric spaces that are connected, but not arcwise connected. One example consists of the union of $\{0\} \times [-1, 1]$ with graph of $f(x) = \sin \frac{1}{x}$ for $x \in (0, 1]$ in the plane (see [1], p. 157 for a proof).*

## 2.11 Metric Completeness

The completeness axiom for the real numbers depends entirely on the existence of an ordering, and yet many frequently encountered metric spaces, including the spaces $\mathbb{R}^n$ for $n > 1$, do not have a natural ordering. Therefore it is useful to have a notion of metric completeness that does not involve an ordering. The general idea is that a metric space is complete if every sequence that "should" have a limit point, really does have a limit point. The following definition makes precise the notion of a sequence that "should" have a limit point.

**Definition 154** *A sequence $(x_i)$ in a metric space $X$ is called Cauchy if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $i, j \geq N$, $d(x_i, x_j) < \varepsilon$.*

More simply we can simply state that a sequence $(x_i)$ is Cauchy if for every $\varepsilon > 0$, $d(x_i, x_j) < \varepsilon$ for all large $i$ and $j$. Convergent sequences behave this way:

**Lemma 155** *If $(x_i)$ is a convergent sequence in a metric space $X$ then $(x_i)$ is Cauchy.*

**Proof.** Suppose $x_i \to x$ and let $\varepsilon > 0$. Then for all large $i$, $d(x_i, x) < \frac{\varepsilon}{2}$. By the triangle inequality, if $i, j$ are large we have that

$$d(x_i, x_j) \leq d(x_i, x) + d(x, x_j) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

■

On the other hand, a Cauchy sequence need not converge. Let $X$ be the metric space $(0, 1]$ with the subspace metric. Note that the sequence $(\frac{1}{i})$ is Cauchy because it is convergent in $\mathbb{R}$. But it is not convergent in the metric space $X$ because $0 \notin X$. Put another way, the sequence $(\frac{1}{i})$ "should" converge to a point, but that point is "missing" from $X$.

**Lemma 156** *If a Cauchy sequence $(x_i)$ has a cluster point $x$ then $x_i \to x$.*

**Proof.** Let $\varepsilon > 0$. Since $(x_i)$ is Cauchy there exists some $N \in \mathbb{N}$ such that if $i, k \geq N$, $d(x_i, x_k) < \frac{\varepsilon}{2}$. Since $x$ is a cluster point of $(x_i)$, for some $k \geq N$, $d(x_k, x) < \frac{\varepsilon}{2}$. By the triangle inequality, for all $i \geq N$ we have

$$d(x_i, x) \leq d(x_i, x_k) + d(x, x_k) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

■

**Exercise 103** *Prove that every Cauchy sequence in a metric space is bounded.*

**Exercise 104** *Let $(x_i)$ and $(y_i)$ be Cauchy sequences in a metric space $X$. Prove that the real sequence $(d_i)$ where $d_i := d(x_i, y_i)$, is convergent.*

**Definition 157** *A metric space $X$ is called complete if every Cauchy sequence $(x_i)$ in $X$ is convergent.*

It is possible to show that every metric space $X$ is isometric to a subset of a complete metric space, called the *metric completion* of $X$ (see [1]), which is also the "smallest" such metric space in a certain sense. For example, the metric completion of $\mathbb{Q}$ is $\mathbb{R}$. However, the general construction of the metric completion of a metric space is beyond the scope of this text. We will be more interested in proving that metric spaces that we are working with either are or are not complete. The following theorem takes us a long way in making such determinations:

**Proposition 158** *Every compact metric space $X$ is complete.*

**Proof.** Let $(x_i)$ be a Cauchy sequence in $X$. Then $(x_i)$ has a cluster point, hence, according to Lemma 156, is convergent. ■

**Corollary 159** *Let $X$ be a metric space such that every closed metric ball of $X$ is compact. Then $X$ is complete.*

**Proof.** Every Cauchy sequence $(x_i)$ in $X$ is bounded by Exercise 103 and therefore contained in some closed metric ball $C$, which is a complete metric space. Hence $(x_i)$ is convergent in $C$ and therefore in $X$. ■

From the Heine-Borel Theorem we now obtain:

**Corollary 160** $\mathbb{R}^n$ *is complete for all $n$ with respect to the product or max metric.*

**Exercise 105** *Prove that any closed subset $A$ of a complete metric space $X$ is complete with the subspace metric.*

**Example 161** *Recall from calculus that the function $e^x : \mathbb{R} \to (0, \infty)$ is continuous and has continuous inverse $\ln x$. Therefore $e^x$ is a homeomorphism. But $\mathbb{R}$ is complete and $(0, \infty)$ is not complete ($(\frac{1}{i})$ is Cauchy but has no limit in $(0, \infty)$). This shows that a non-complete metric space can be homeomorphic to a complete metric space. In other words, completeness is not a "topological" property, but a metric property.*

**Exercise 106** *Suppose that $f : X \to Y$ is a function between metric spaces.*

1. *Prove that if $f$ is Lipschitz and $(x_i)$ is Cauchy then $(f(x_i))$ is Cauchy.*
2. *Prove that if $f$ is bilipschitz and $X$ is complete then $Y$ is complete.*

# Chapter 3

# Complex sequences and Series

## 3.1 Complex Numbers

**Definition 162** *The complex numbers* $\mathbb{C}$ *consist of the metric space* $\mathbb{R}^2$ *having the usual operation of vector addition and a new operation of multiplication defined as follows: For* $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ *we define*

$$(x_1, y_1) \cdot (x_2, y_2) := (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1) \tag{3.1}$$

There is a more standard notation for the complex numbers with which the reader may be familiar, where we write $x + yi$ rather than $(x, y)$. Addition and multiplication are carried out as though the symbols represent real numbers, except that the symbol $i$ has the property that $i^2 = -1$ and otherwise is treated like an independent variable. For example to compute $(x_1 + y_1i) + (x_2 + y_2i)$ we simply add all the terms and "collect" the terms involving $i$ to obtain $(x_1 + x_2) + (y_1 + y_2)i$. For multiplication we distribute and collect terms:

$$(x_1 + y_1i)(x_2 + y_2i) = x_1x_2 + x_1y_2i + y_1x_2i + (i)(i)y_1y_2$$

$$= (x_1x_2 - y_1y_2) + (x_1y_2 + x_2y_1)i$$

Note that the above formula is precisely the same as that in Formula 3.1; only the notation is different. With these two operations, the set $\mathbb{C}$ is a field (see Section 1.4). The details are not difficult, and not too interesting. We simply note the following: The additive identity is $0 + 0i$, which we will simply denote by 0. The multiplicative identity is $1 + 0i$, which we will simply denote by 1. The additive inverse of $x + yi$ is $-x - yi$. The multiplicative inverse of $x + iy$ is more tricky to figure out, but it looks like this:

$$(x + yi)^{-1} = \frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2}i$$

which we may write in the more compact form

$$(x+yi)^{-1} = \frac{x-yi}{x^2+y^2}.$$

**Exercise 107** *Show that* $(x+iy)\left(\frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}\right) = 1.$

One uses the same algebraic definitions and obtains the same basic algebraic theorems that are familiar for $\mathbb{R}$ (or any field). As we did for real numbers, we can define $z^0 := 1$ and $z^n$ iteratively by $z^n := z \cdot z^{n-1}$ for any integer $n > 0$. One then defines $z^n = \frac{1}{z^{-n}}$ for negative integers $n$ whenever $z \neq 0$. In the next section we will prove the existence of $n^{th}$ roots of positive numbers. This will allow us to define the root test, which is part of the logical path to the exponential function and its properties. At the end of this chapter we will use the exponential function to define real powers of positive numbers.

We make two more notational conventions, extending what we did above for 0 and 1. If a complex number $x+yi$ satisfies $y=0$, we write it simply as $x$, and refer to it as real. This makes sense because the set of all such complex numbers is closed under multiplication and addition and for these specific numbers addition and multiplication are identical to addition and multiplication for real numbers. For example

$$(x_1+0i)(x_2+0i) = (x_1x_2 - 0\cdot 0) + (x_1 \cdot 0 + 0 \cdot x_2)i = x_1x_2 + 0i.$$

The student who is familiar with abstract algebra will recognize that what is really going on is that the real numbers are naturally field isomorphic (see the definition below) to the subfield of $\mathbb{C}$ consisting of all complex numbers of the form $x+0i$, but such a formality will not be useful for us now. Note that, when considering the complex numbers as points in the plane, the real numbers correspond to the points having second coordinate 0, i.e. the $x$-axis.

On the other hand if $x=0$ we write $yi$ rather than $0+yi$ and refer to $yi$ as *pure imaginary*. Pure imaginary numbers are closed under addition but, obviously, not under multiplication. We also write $i$ rather than $1i$.

Finally, as one traditionally does in a field, we often will write $\frac{z}{w}$ rather than $zw^{-1}$; one can easily check that the usual formulas for adding and multiplying fractions are still valid.

**Exercise 108** *Show that* $\left(\frac{-1\pm\sqrt{3}i}{2}\right)^3 = 1$, *that is,* 1 *has two cube roots besides* 1 *in* $\mathbb{C}$.

**Definition 163** *If* $\mathbb{F}_1$ *and* $\mathbb{F}_2$ *are fields, a function* $f : \mathbb{F}_1 \to \mathbb{F}_2$ *is called a field isomorphism if* $f$ *is a bijection and for all* $x, y \in \mathbb{F}_1$, $f(x+y) = f(x) + f(y)$ *and* $f(xy) = f(x)f(y)$.

A field isomorphism shows that two fields are algebraically identical.

**Exercise 109** *Consider the set $\mathbb{K}$ of all matrices of the form $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ where $a$ and $b$ are real numbers, with the operations of matrix addition and matrix multiplication.*

1. *Show that $\mathbb{K}$ is closed under addition and multiplication.*
2. *Show that the function $M : \mathbb{C} \to \mathbb{K}$ defined by $M(a+bi) = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ is a field isomorphism. It is not necessary to prove that $\mathbb{K}$ is a field; in fact when you check formally the conditions for a field isomorphism it follows that $\mathbb{K}$ is a field.*

Let us return to considering $\mathbb{C}$ as the metric space $\mathbb{R}^2$. When using the notation $x + yi$ we will use, as in the case of $\mathbb{R}$, absolute value notation to denote the norm, i.e. $|x + yi|$ denotes $\|(x, y)\| = \sqrt{x^2 + y^2}$. We will also use regular roman numerals (not bold) to name elements of $\mathbb{C}$ when we are not using the $x + yi$ notation. For any $z \in \mathbb{C}$, the quantity $|z|$ is often referred to as the *modulus* of $z$. We already have the triangle inequality relating to the modulus of a sum; the following lemma describes the simpler situation for the modulus of a product.

**Lemma 164** *For $z, w \in \mathbb{C}$, $|zw| = |z|\,|w|$.*

**Proof.** Let $z = a + bi$ and $w = c + di$. Then

$$|zw|^2 = |(a+bi)(c+di)|^2 = (ac-bd)^2 + (ad+bc)^2$$
$$= a^2c^2 + b^2d^2 - 2abcd + a^2d^2 + b^2c^2 + 2abcd$$
$$= (a^2+b^2)(c^2+d^2) = |z|^2\,|w|^2 = (|z|\,|w|)^2$$

■

**Exercise 110** *Verify that if $z = a + bi$ then $z$ has the following square roots:*

$$\pm\left(\sqrt{\frac{a+|z|}{2}} + sgn(b)\sqrt{\frac{-a+|z|}{2}}i\right)$$

*where $sgn(b)$ is the sign of the real number $b \neq 0$ and $sgn(b) = 1$ if $b = 0$.*

**Definition 165** *Let $z = a + bi$ be a complex number. The conjugate of $z$ is the complex number $\overline{z} := a - bi$.*

**Proposition 166** *Let $z$ and $w$ be complex numbers. Then*

1. $\overline{\overline{z}} = z$
2. $\overline{z+w} = \overline{z} + \overline{w}$

3. $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$ *(the $\cdot$ refers to multiplication, used for easier reading)*

4. $|z|^2 = z\overline{z}$

**Exercise 111** *Prove the above proposition.*

**Exercise 112** *Let $f : \mathbb{C} \to \mathbb{C}$ be defined by $f(z) = \overline{z}$. Prove that*

1. *$f$ is an isometry, hence a homeomorphism*
2. *$f$ is a field isomorphism*
3. *$f$ is involutive, i.e. $f \circ f = id_{\mathbb{C}}$*
4. *$f \mid_{\mathbb{R}} = id_{\mathbb{R}}$*

**Exercise 113** *Show the following:*

1. *For complex numbers $z$ and $w$, $|z + w|^2 + |z - w|^2 = 2(|z|^2 + |w|^2)$.*
2. *Use the above problem to show that for any parallelogram in the plane the sum of the squares of the lengths of the diagonals is equal to the sum of the squares of the sides. Hint: the sum of the lengths of the diagonals is on the left side of the equation.*

**Exercise 114** *Prove that if $z \neq 0$ then there exists an $N$ such that for all $n \geq N$, $\left|n^2 + \frac{1}{z}\right| \geq n^{\frac{3}{2}}$. Hint: Use the triangle inequality on the three points $0$, $n^2$ and $n^2 + \frac{1}{z}$.*

## 3.2 Complex Sequences

We will quickly review a few basic facts that we know about sequences of complex numbers. Everything we know so far is a consequence of more general theorems about the metric space $\mathbb{R}^2$. We know that $\mathbb{C}$ is a complete metric space, with metric given by $d(z, w) = |z - w|$. In particular, we have $z_n \to z$ if and only if $|z_n - z| \to 0$. Closed and bounded subsets of $\mathbb{C}$ are compact, and every bounded sequence of complex numbers has a convergent subsequence. Setting $z_n := a_i + b_n i$ and $z = a + b$ we have $z_n \to z$ if and only if $a_n \to a$ and $b_n \to b$. Note that the proposition below is valid for real sequences since every real sequence is also a complex sequence.

**Proposition 167** *Let $\{z_n\}$ and $\{w_n\}$ be complex sequences such that $z_n \to z$ and $w_n \to w$. Then*

1. *$z_n + w_n \to z + w$*
2. *$z_n w_n \to zw$*
3. *if $z \neq 0$ then $z_n \neq 0$ for all large $n$ and $z_n^{-1} \to z^{-1}$.*

**Proof.** We leave the proofs of the first two parts as exercises. For the last part note that $|z| > 0$ and for all large $n$ $|z_n - z| < \frac{|z|}{2}$ and therefore by the triangle inequality $|z_n| \geq |z| - |z_n - z| > \frac{|z|}{2} > 0$, so $z_n \neq 0$. Now let $\varepsilon > 0$. For large $n$ we also have $|z_n - z| < \frac{\varepsilon |z|^2}{2}$. Combining the two previous inequalities for large $n$ we have

$$\left| \frac{1}{z_n} - \frac{1}{z} \right| = \left| \frac{z_n - z}{z_n z} \right| < \frac{2\,|z_n - z|}{|z|^2} < \varepsilon.$$

■

**Exercise 115** *Finish the proof of Proposition 167.*

Note that, as a special case of part (2) of Proposition 167 we can let $w_n$ be the constant sequence $w_n = c$, in which case we obtain that $cz_n \to cz$. We can also combine the last two parts to obtain that $\frac{w_n}{z_n} \to \frac{w}{z}$ provided $z \neq 0$.

As in the real case, given a set $A$ and functions $f, g : A \to \mathbb{C}$ and constant $c \in \mathbb{C}$ we define $f + g : A \to \mathbb{C}$ by $(f + g)(x) = f(x) + g(x)$, and $f \cdot g$, $f/g$, $cf$, are defined similarly. All four combinations are defined on $A$ except for $f/g$, which is defined for all $x \in A$ such that $g(x) \neq 0$.

**Corollary 168** *Let $X$ be a metric space and $f, g : X \to \mathbb{C}$ be continuous at $x \in X$ and $c \in \mathbb{C}$. Then the functions $f + g$, $f \cdot g$, $cf$, and (provided $g(x) \neq 0$) $f/g$ are continuous at $x$.*

**Proof.** We will prove only that $f/g$ is continuous at $x$ provided $g(x) \neq 0$; the other proofs are similar and simpler. Let $A := \{x \in X : g(x) \neq 0\}$, which is the domain of definition of $f/g$. Let $x_n \to x$ in $A$. Since $f$ and $g$ are continuous at $x$, $f(x_n) \to f(x)$ and $g(x_n) \to g(x) \neq 0$. By Proposition 167 $f(x_n)/g(x_n) \to f(x)/g(x)$, which proves the continuity of $f/g$ at $x$. ■

**Corollary 169** *Every complex polynomial, i.e., function of the form $f(z) = a_n z^n + a_{n-1} z^{n-1} + ... + a_1 z + a_0$ where each $a_i$ and $z$ are complex numbers, is continuous*

**Proof.** We will prove the statement by induction in the degree $n$ of the polynomial. For $n = 0$, the polynomial is constant, and this case was considered in Exercise 57. Suppose we have proved the statement for $n \geq 0$ and consider $f(z) = a_{n+1} z^{n+1} + a_n z^n + ... + a_1 z + a_0$. We will be finished by the inductive hypothesis and Corollary 168 if we can prove that $a_{n+1} z^{n+1}$ is continuous. However, we know that the identity function $f(z) = z$ is continuous (in any metric space!) and by Corollary 168, $a_{n+1} z$ is continuous. By the inductive hypothesis, $z^n$ is continuous, and by Corollary 168 $a_{n+1} z^{n+1} = (a_{n+1} z) z^n$ is continuous. ■

**Corollary 170** *If $a$ is a positive real number and $n > 1$ is a natural number then there is a unique positive $n^{th}$ root of $a$, that is, a positive real number $x$ such that $x^n = a$, denoted by $\sqrt[n]{a}$ or $a^{\frac{1}{n}}$.*

**Proof.** Consider the real function $f(x) = x^n$ defined on $[0, \infty)$. This function is the restriction of a complex polynomial and hence is continuous. In addition, this function is onto $[0, \infty)$. In fact, $f([0, \infty))$ is connected and contains 0, hence is an interval with left endpoint 0. Since $x > 1$, $x^n > x$, $f([0, \infty))$ is unbounded, hence must be the interval $[0, \infty)$. Since $f$ is onto there exists some $x \in [0, \infty)$ such that $x^n = a \in [0, \infty)$. The uniqueness statement follows from the fact that if $0 < a < b$ then $a^n < b^n$. ■

**Exercise 116** *Prove that the real function $f$ defined by $f(x) := \sqrt[n]{x}$ is continuous for all $n \geq 1$.*

Note that it follows from the field and order axioms that the functions $x \mapsto x^n$ and $x \to x^{\frac{1}{n}}$ are strictly increasing functions, a fact that we will use without further comment.

**Exercise 117** *Let $(z_n)$ be a complex sequence. Prove that if $(z_n)$ is convergent then $(|z_n|)$ is convergent; is the converse true?*

**Example 171** *One immediate consequence of Proposition 167 is that a sequence $(z_n)$ is convergent if and only if $(cz_n)$ is convergent for every $c \neq 0$. However, it is not true that if the sum of two sequences is convergent then the two sequences are convergent. For example, let $a_n = n$ and $b_n = -n$; the sum of the unbounded sequences $(a_n)$ and $(b_n)$ is the constant (hence convergent) sequence whose terms are all 0.*

Problems involving complex sequences and series can often be reduced to problems involving real sequences. For example, we have already observed that $z_n \to z$ if and only if $|z_n - z| \to 0$; the latter is a real sequence. Likewise if $z_n = a_n + b_n i$ then the two sequences $(a_i)$ and $(b_i)$ are real sequences that completely determine the convergence of the complex sequence $(z_n)$. We will therefore consider a few useful theorems about real sequences.

The convergence of several sequences follows from the *binomial formula* for complex numbers $a, b$ and natural number $n$:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \tag{3.2}$$

Here $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ satisfies the following property

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \tag{3.3}$$

that can be verified by direct computation and used to prove the binomial formula by induction in $n$. Formula 3.3 also is the basis for the construction of Pascal's triangle. Note that if $a$ and $b$ are nonnegative then each of the terms in the right side of Formula 3.2 is also nonnegative. In particular, one can obtain many useful inequalities by "throwing away" some of these terms. For example, $(a+b)^n \geq a^n + na^{n-1}b$ (for $a, b \geq 0$) follows by throwing away all terms but the first two. One can also sometimes use clever choices of $a$ and $b$ to get equations and inequalities. For example, by letting $a = b = 1$ one obtains that $2^n = \sum_{k=0}^{n} \binom{n}{k}$.

**Example 172** *If $p > 0$ then $\lim \sqrt[n]{p} = 1$. To prove this, first suppose that $p > 1$ and let $a_n = \sqrt[n]{p} - 1 > 0$ then $p = (1 + a_n)^n$. By the binomial formula (throwing away all but the first two terms) $p \geq 1 + na_n$ and so $0 \leq a_n \leq \frac{p-1}{n}$, and the proof is finished by the sandwich theorem. If $p = 1$ no proof is required, and if $p < 1$ we can apply the previous case to $\frac{1}{p} > 1$ and apply Proposition 167 (3).*

**Exercise 118** *Use the binomial formula to prove that $\lim \sqrt[n]{n} = 1$.*

**Exercise 119** *Prove that if $a > 1$ then $\lim \frac{1}{a^n} \to 0$.*

**Example 173** *Let $z \in \mathbb{C}$ satisfy $0 < |z| < 1$. Then $|z|^n \to 0$ by Exercise 119. Since $|z^n| = |z|^n \to 0$, $z^n \to 0$. In other words, for any $z$ inside open unit ball at $0$ in $\mathbb{C}$, higher and higher powers of $z$ tend to $0$. If $z$ is real then $z^n$ is real for all $n$, and $z^n \to 0$ on the $x$-axis. But other complex numbers do not tend to $0$ in a straight line. See Exercise 120 below.*

**Exercise 120** *Plot the first eight terms of the sequence $z^n$, where $z = \frac{1+i}{2}$.*

**Exercise 121** *Show that if $z$ lies on the unit circle in the complex plane then $z^n$ lies on the unit circle for all natural numbers $n$. What happens if $z$ lies outside the unit circle?*

We finish this section with some theorems about real sequences. First of all we adopt the standard conventions for infinite limits of extended real sequences. Note that we have not defined a metric on the extended real numbers (this is possible, but not in a way that the subspace metric on the reals is the usual metric on the reals). Therefore we must define what we mean by limits involving extended reals without using any kind of metric. We say that $x_i \to \infty$, where $x_i$ is an extended real number, if for every real $M$, $x_i \geq M$ for all large $i$. We similarly define $x_i \to -\infty$. We will use the following notation. We will say that a real sequence $(x_i)$ is convergent if it has a real limit, but when we say that $(x_i)$ has a limit this will mean that $(x_i)$ is convergent or $\lim x_i = \pm\infty$. Note that we can now state that every monotone real sequence $(x_i)$ has a limit (allowing $\pm\infty$). If $(x_i)$ is bounded we have already proved this statement in Proposition 60. The unbounded case is the next exercise.

**Exercise 122** *Show that if $(x_i)$ is an unbounded real sequence then $x_i \to \infty$ (resp. $x_i \to -\infty$) if $(x_i)$ is increasing (resp. decreasing).*

**Proposition 174** *If $(x_i)$ and $(y_i)$ are real sequences having (possibly infinite) limits and $x_i \leq y_i$ for all large $i$ then $\lim x_i \leq \lim y_i$.*

**Proof.** Let $x = \lim x_i$ and $y = \lim y_i$. We will prove the case when both $x$ and $y$ are finite. The remaining cases are an exercise. Suppose that $y < x$. Let $a$ be such that $y < a < x$. Then for all large $i$, $y_i < a < x_i$, a contradiction. ■

**Exercise 123** *Finish the proof of Proposition 174.*

**Definition 175** *Let $(x_i)$ be a real sequence. We define the limit superior (or lim sup) of $(x_i)$ by*

$$\limsup x_i = \lim s_n, \text{ where } s_n := \sup_{k \geq n}\{x_k\}$$

*and the limit inferior (or lim inf) of $(x_i)$ by*

$$\liminf x_i = \lim m_n, \text{ where } m_n := \inf_{k \geq n}\{x_k\}.$$

This definition may take some time to digest. We will discuss the lim sup; a similar discussion applies to the lim inf. First of all we should observe that, unlike the limit of a sequence, the lim sup of a sequence always exists (although it could be $\pm\infty$). To see why, let $S_n := \{x_k\}_{k \geq n}$ and $s_n := S_n$. Since $S_1 \supset S_2 \supset S_3 \cdots$ we have $s_1 \geq s_2 \geq ...$; i.e., the sequence $(s_n)$ is decreasing. Therefore $\limsup x_i = \lim s_n$ exists. Note that it is always true that

$$\liminf x_i \leq \limsup x_i. \tag{3.4}$$

In fact, $m_n \leq s_n$ for all $n$ and the above inequality follows from Proposition 174.

**Lemma 176** *If $(x_i)$ and $(y_i)$ are real sequences and $x_i \leq y_i$ for all large $i$ then $\limsup x_i \leq \limsup y_i$ and $\liminf x_i \leq \liminf y_i$.*

**Proof.** We will prove only the statement for the lim sup; the other case is similar. But this follows from Proposition 174 and the fact that $\sup_{k \geq n}\{x_k\} \leq \sup_{k \geq n}\{y_k\}$. ■

**Lemma 177** *Let $(x_i)$ be a real sequence and let $s := \limsup x_i$ (resp. $s := \liminf x_i$). Then for all extended real numbers $K, M$ such that $K < s < M$, we have $x_i > K$ for some (resp. all) large $i$ and $x_i < M$ for all (resp. some) large $i$.*

**Proof.** We only consider the case of the lim sup; the other case is similar. Since the interval $(K, M)$ is an open set and $s = \limsup\{x_i\}_{i \geq n} \in (K, M)$,

$$K < \sup\{x_i\}_{i \geq n} < M.$$

for all large $n$. Then $x_i \leq \sup\{x_i\}_{i \geq n} < M$ for all $i \geq n$. Moreover, $s_n \geq s > K$ and for any positive $\varepsilon < s_n - K$, the approximation property implies that there exists some $i \geq n$ such that $x_i > s_n - \varepsilon > K$. ■

**Lemma 178** *Let $(x_n)$ be a real sequence. Then $(x_n)$ is convergent (with finite limit) if and only if* $\limsup x_n = \liminf x_n$. *If $(x_n)$ is convergent then* $\lim x_n = \limsup x_n$.

**Proof.** Suppose that $\limsup x_n = \liminf x_n := y$ and let $\varepsilon > 0$. Then $y - \varepsilon < y < y + \varepsilon$ and by Lemma 177, for all large $n$, $y - \varepsilon < x_n < y + \varepsilon$, which implies $x_n \to y$.

Suppose now that $x_n \to x \in \mathbb{R}$. Then for every $\varepsilon > 0$, $x_n < x + \varepsilon$ for all large $n$. By Lemma 176 (using the constant sequence $(x + \varepsilon)$) we have $\limsup x_n \leq x + \varepsilon$ for every $\varepsilon > 0$. Exercise 14 implies $\limsup x_i \leq x$. A similar argument shows $\liminf x_i \geq x$ and the proof is finished by (3.4). ■

**Exercise 124** *Find the lim sup and lim inf of the sequence $(x_n)$, where $x_n = 1 + (-1)^n \frac{n}{n+1}$. You do not need to give a proof for your answer.*

**Exercise 125** *State and prove an analog of Lemma 178 concerning infinite limits.*

**Exercise 126** *Let $(a_n)$ and $(b_n)$ be real sequences.*

1. Prove that $\limsup(a_n + b_n) \leq \limsup(a_n) + \limsup(b_n)$.

2. Give an example that shows that equality may not occur in the above formula.

## 3.3 Series

In this section we will investigate series of complex numbers. We will be concerned with expressions of the form

$$\sum_{n=1}^{\infty} z_n$$

where each $z_n$ is a complex number. The series may also start at $n = 0$ (as in the geometric series below) or some other finite number; theorems below involving divergence and convergence are clearly also true for any no matter what the starting value for $n$.

At first it may not even be clear whether one can make sense of such an expression. One could start adding the $z_n$'s and see what happens. If $z_n = 2^{-n}$ then starting with $n = 1$ one gets $\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, ...$ when adding the first $1, 2, 3, ...$ terms together, and these "partial sums" form the sequence $(\frac{n}{n+1})$, which tends to 1. Therefore it seems clear that $\sum_{n=1}^{\infty} 2^{-n}$ should be 1. (The Ancient Greeks, who had no good way to represent and therefore add fractions, apparently never made this observation and were forever troubled by Zeno's "paradox".) If one tries the same thing with $\sum_{n=1}^{\infty} 1$ then the partial sums are $1, 2, 3, ...$, which converges to $\infty$. Again, it is reasonable to say $\sum_{n=1}^{\infty} 1 = \infty$. What about

$\sum_{n=1}^{\infty} i$? In this case the partial sums are $i, 2i, 3i, 4i, \ldots$ which "diverges" up the $y$-axis in the complex plane. Does it also tend to $\infty$? If it does, is this the same $\infty$ that you "approach" as you go out the $x$-axis? This particular question will have to be left for a later course in complex analysis. At least we can agree that the sum $\sum_{n=1}^{\infty} i$ doesn't converge to any complex number.

Another issue that we will have to contend with is the associative law. Let's consider the series $\sum_{n=1}^{\infty}(-1)^n$. We could compute the partial sums as $-1, 0, -1, 0, -1, \ldots$ which doesn't converge to anything. On the other hand, by combining terms in pairs we could compute

$$(-1+1)+(-1+1)+(-1+1)+\cdots = 0+0+0+\cdots = 0$$

or

$$-1+(1-1)+(1-1)+\cdots = -1+0+0+\cdots = -1$$

Which way is "right"? For the second two computations we have tried to use some kind of "infinite associative law" that allows us to simultaneously rearrange infinitely many parentheses from the arrangement for partial sums, which looks like

$$\cdots((((-1+1)-1)+1)+1)-\cdots$$

Apparently no such law exists, because arbitrary rearrangements of parentheses do change the sum. In particular, if we want to have a well-defined notion of a sum of a series (if it exists) we should stick to finding limits of partial sums, which we will now formally introduce.

A complex *series* is an expression $\sum_{n=1}^{\infty} z_n$ where each $z_n$ is a complex number. Each series has an associated sequence $(\sigma_k)$ called the *sequence of partial sums* defined by

$$\sigma_k := \sum_{n=1}^{k} z_n$$

If the sequence $(\sigma_k)$ converges to a complex number $z$ we say that the series *converges* to $z$ and we will refer to $z$ as the *sum* of the series; otherwise we say the series *diverges.*

**Lemma 179** *(Small Tails Lemma) A series $\sum_{n=1}^{\infty} z_n$ is convergent if and only if for every $\varepsilon > 0$, $|\sum_{n=k}^{\infty} z_n| < \varepsilon$ for all large $k$.*

**Exercise 127** *Prove the above lemma.*

Since $\mathbb{C}$ is a complete metric space, the sequence $(\sigma_k)$ is convergent if and only if it is Cauchy. Since $\sigma_j - \sigma_k = \sum_{n=k+1}^{j} z_n$ for $j \geq k$, this fact can be restated as follows:

**Lemma 180** *(Cauchy Criterion) A series $\sum_{n=1}^{\infty} z_n$ converges if and only if for every $\varepsilon > 0$*

$$\left|\sum_{n=j}^{k} z_n\right| < \varepsilon.$$

*for all large $j \leq k$.*

Letting $k = j$ in the above lemma yields:

**Corollary 181** *(Divergence Test) If $\sum_{n=1}^{\infty} z_n$ converges then $|z_n| \to 0$.*

The above corollary is the complex version of the "divergence test" for real series in elementary calculus. It can be used to prove that a series does not converge (i.e. by showing that $|z_n|$ doesn't converge to 0), but we will see later that there are divergent series such that $|z_n| \to 0$.

**Corollary 182** *If $\sum_{n=1}^{\infty} z_n$ is a series and the real series $\sum_{n=1}^{\infty} |z_n|$ converges then $\sum_{n=1}^{\infty} z_n$ converges.*

**Proof.** Let $\varepsilon > 0$. Then for all large $j \leq k$, the triangle inequality and Cauchy criterion for $\sum_{n=1}^{\infty} |z_n|$ imply

$$\left|\sum_{n=j}^{k} z_n\right| \leq \sum_{n=j}^{k} |z_n| = \left|\sum_{n=j}^{k} |z_n|\right| < \varepsilon.$$

According to the Cauchy criterion, $\sum_{n=1}^{\infty} z_n$ converges. ■

**Definition 183** *A series $\sum_{n=1}^{\infty} z_n$ such that $\sum_{n=1}^{\infty} |z_n|$ converges is called absolutely convergent.*

We will see later that there are series that are convergent but not absolutely convergent.

**Example 184** *The geometric series is the series $\sum_{n=0}^{\infty} z^n$. If $z \neq 1$ then $(1 - z)\sigma_n = 1 - z^{n+1}$ and*

$$\sigma_n = \frac{1 - z^{n+1}}{1 - z}.$$

*According to Example 173, $\sigma_n \to \frac{1}{1-z}$ if $|z| < 1$; that is,*

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \quad (|z| < 1)$$

*On the other hand, if $|z| \geq 1$ then $|z^n| = |z|^n \geq 1$ for all $n$, and $\sum_{n=0}^{\infty} z_n$ diverges by the divergence test. Sometimes it is useful to use $n = 1$ as the starting point for the geometric series, in which case convergence and divergence are unchanged, but the sum, in case of convergence, is $\frac{z}{1-z}$.*

**Proposition 185** *If $\sum_{n=1}^{\infty} z_n$ and $\sum_{n=1}^{\infty} w_n$ are convergent then $\sum_{n=1}^{\infty}(z_n + w_n)$ is convergent and $\sum_{n=1}^{\infty}(z_n + w_n) = \sum_{n=1}^{\infty} z_n + \sum_{n=1}^{\infty} w_n$.*

**Proof.** Let $(\sigma_n)$ and $(\tau_n)$ denote the partial sum sequences for $\sum_{n=1}^{\infty} z_n$ and $\sum_{n=1}^{\infty} w_n$, respectively. Then if $(\eta_n)$ is the partial sum sequence for $\sum_{n=1}^{\infty}(z_n + w_n)$ we have

$$\eta_k = \sum_{n=1}^{k}(z_n + w_n) = \sum_{n=1}^{k} z_n + \sum_{n=1}^{k} w_n = \sigma_k + \tau_k$$

and by Proposition 167, part (1)

$$\sum_{n=1}^{\infty}(z_n + w_n) = \lim \eta_k = \lim(\sigma_k + \tau_k) = \lim \sigma_k + \lim \tau_k = \sum_{n=1}^{\infty} z_n + \sum_{n=1}^{\infty} w_n.$$

■

**Exercise 128** *Let $c \neq 0$ and $\sum_{n=1}^{\infty} z_n$ be a series. Prove that $\sum_{n=1}^{\infty} z_n$ is convergent if and only if $\sum_{n=1}^{\infty} cz_n$ is convergent, and if $\sum_{n=1}^{\infty} z_n$ is convergent then $\sum_{n=1}^{\infty} cz_n = c\sum_{n=1}^{\infty} z_n$. What happens when $c = 0$?*

**Exercise 129** *Prove or disprove: If $\sum_{n=1}^{\infty} z_n$ and $\sum_{n=1}^{\infty} w_n$ are series then $\sum_{n=1}^{\infty}(z_n + w_n)$ is convergent if and only if $\sum_{n=1}^{\infty} z_n$ and $\sum_{n=1}^{\infty} w_n$ are convergent.*

**Exercise 130** *(Telescoping Series) Let $(z_n)$ be a sequence of complex numbers and let $t_n := z_n - z_{n+1}$. Show that the series $\sum_{n=1}^{\infty} t_n$ is convergent if and only if the sequence $(z_n)$ is convergent. When the series is convergent, give a formula for its sum.*

**Exercise 131** *Compute the sum of the following series:*

1. $\sum_{n=1}^{\infty} \frac{3}{(1-i)^n}$
2. $\sum_{n=1}^{\infty} \left(\frac{1}{\sqrt{n^2+2n+1}} - \frac{1}{\sqrt{n^2}}\right)$
3. $\sum_{n=1}^{\infty} \frac{2}{n(n+1)}$ *Hint: Use a partial fractions decomposition (a fraction of the form $\frac{cx+d}{(x+a)(x+b)}$ can always be rewritten in the form $\frac{r}{x+a} + \frac{s}{x+b}$ when $a \neq b$).*

Multiplication of series is a bit more complicated because we will have to contend with the distributive law. Consider what happens to a few terms:

$$(a_0 + a_1 + a_2 + a_3)(b_0 + b_1 + b_2 + b_3)$$

$$= a_0b_0 + a_0b_1 + a_0b_2 + \cdots + a_3b_2 + a_3b_3.$$

We need a way to conveniently collect and enumerate all of these terms. The simplest way to do this is to collect together all terms $a_ib_j$ so that the sum of the indices $i + j$ is a fixed number $n$; that is,

$$c_n := \sum_{k=0}^{n} a_k b_{n-k} \tag{3.5}$$

or alternatively, $c_n := \sum_{i+j=n} a_ib_j$. The first few terms are $c_0 := a_0b_0$, $c_1 := a_0b_1 + a_1b_0$ and $c_3 := a_0b_3 + a_1b_2 + a_3b_0$.

**Proposition 186** *Let $\sum_{n=0}^{\infty} a_n$ be absolutely convergent, $\sum_{n=0}^{\infty} b_n$ be convergent, and $c_n$ be defined according to Formula (3.5). Then*

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = \left(\sum_{n=0}^{\infty} c_n\right).$$

*In particular, the series on the right side is convergent.*

**Exercise 132** *Prove Proposition 186 as follows. Let $(\sigma_k)$, $(\tau_k)$, $(\eta_k)$ be the partial sum sequences of $\sum_{n=0}^{\infty} a_n = a$, $\sum_{n=0}^{\infty} b_n = b$, and $\sum_{n=0}^{\infty} c_n$, respectively. Let $t_k$ be the sum of the $(k+1)$-tail of $\sum_{n=0}^{\infty} b_n$.*

1. *Show that $\eta_n = \sigma_n b + \varepsilon_n$ for all $n$, where*

$$\varepsilon_n = a_0 t_n + a_1 t_{n-1} + \cdots + a_n t_0. \tag{3.6}$$

2. *Suppose $\sum_{n=0}^{\infty} |a_n| = A$, and for given $\varepsilon > 0$, $N$ is such that $|t_n| \leq \varepsilon$ whenever $n \geq N$ (why does such an $N$ exist?). Prove that for any $n \geq N$,*

$$|\varepsilon_n| \leq |a_n t_0 + \cdots + a_{n-N} t_N| + \varepsilon A.$$

3. *Show that $\varepsilon_n \to 0$ and finish the proof.*

## 3.4 Convergence tests

The geometric series and series in Exercise 130 are of the exceedingly rare variety the sum of which can actually be computed; generally the only question that can be answered precisely about a given series is whether or not it converges. This is far more useful than it may seem at first. None only can sums of series be numerically approximated, there are many situations in which knowing the exact sum of a series is not important.

For series with nonnegative (real) terms there is a particularly simple criterion for convergence since in that case the partial sum sequence $(\sigma_n)$ is monotone increasing and converges if and only if it is bounded. We restate this as:

**Lemma 187** *If $\sum_{n=1}^{\infty} x_n$ is a series with nonnegative terms then $\sum_{n=1}^{\infty} x_n$ is convergent if and only if there exists an $M < \infty$ such that for all $k$, $\sum_{n=1}^{k} x_n \leq M$, and in this case $\sum_{n=1}^{\infty} x_n \leq M$.*

**Corollary 188** *(Comparison Test) If $0 \leq x_n \leq c_n$ and $\sum_{n=1}^{\infty} c_n$ is convergent then $\sum_{n=1}^{\infty} x_n$ is convergent and $\sum_{n=1}^{\infty} x_n \leq \sum_{n=1}^{\infty} c_n$*

**Proof.** Clearly the partial sum sequence $(\sigma_n)$ of $\sum_{n=1}^{\infty} x_n$ and the partial sum sequence $(\tau_n)$ of $\sum_{n=1}^{\infty} c_n$ satisfy $\sigma_n \leq \tau_n$ for all $n$; if the latter is bounded then the former is bounded. ■

**Proposition 189** *(Root Test) Let* $\sum_{n=1}^{\infty} z_n$ *be a series and define*

$$r := \limsup \left( \sqrt[n]{|z_n|} \right).$$

1. *If* $r < 1$ *then the series converges absolutely.*
2. *If* $r > 1$ *then the series diverges.*

**Proof.** If $r < 1$ then there is some $\rho$ such that $r < \rho < 1$ and according to Lemma 177 $\sqrt[n]{|z_n|} < \rho$ for all large $n$. Equivalently $|z_n| < \rho^n$ and the series converges absolutely by comparison with the geometric series $\sum_{n=1}^{\infty} \rho^n$. If $r > 1$ then by Lemma 177, for some large $n$, $\sqrt[n]{|z_n|} > 1$ and hence $|z_n| > 1$. Then $|z_n|$ cannot converge to 0 and the series diverges. ■

**Lemma 190** *Let* $(x_n)$ *be a sequence of positive real numbers. Then*

$$\liminf \frac{x_{n+1}}{x_n} \leq \liminf \sqrt[n]{x_n} \leq \limsup \sqrt[n]{x_n} \leq \limsup \frac{x_{n+1}}{x_n}.$$

**Proof.** We will prove the first inequality; the middle inequality is clear and the last is similar to the first. First note that if $\liminf \frac{x_{n+1}}{x_n} = 0$ then there is nothing to prove. Suppose $\liminf \frac{x_{n+1}}{x_n} > 0$ and let $0 < M < \liminf \frac{x_{n+1}}{x_n}$. According to Lemma 177, there is a natural number $N$ such that if $m \geq N$ then $\frac{x_{m+1}}{x_m} > M$, which is equivalent to $x_{m+1} > M x_m$. Iterating this inequality starting with some $n > N$ we obtain

$$x_n > M x_{n-1} > \cdots > M^{n-N} x_N = \frac{x_N}{M^N} M^n$$

or

$$\sqrt[n]{x_n} > M \sqrt[n]{\frac{x_N}{M^N}}.$$

From Example 172 we know that $\sqrt[n]{\frac{x_N}{M^N}} \to 1$ ($N$ is fixed) and by Lemma 176, $\liminf \sqrt[n]{x_n} \geq M$. If $\liminf \frac{x_{n+1}}{x_n}$ is finite we may take $M = \liminf \frac{x_{n+1}}{x_n} - \varepsilon$ for any $\varepsilon > 0$ and conclude that $\liminf \sqrt[n]{x_n} \geq \liminf \frac{x_{n+1}}{x_n}$. If $\liminf \frac{x_{n+1}}{x_n} = \infty$ we may take $M$ to be any natural number to conclude that $\liminf \sqrt[n]{x_n} = \infty$. ■

**Corollary 191** *(Ratio Test) Let* $\sum_{n=1}^{\infty} z_n$ *be a series such that* $z_n \neq 0$ *for all large* $n$.

1. *If* $\limsup \left| \frac{z_{n+1}}{z_n} \right| < 1$ *then the series converges absolutely.*
2. *If* $\liminf \left| \frac{z_{n+1}}{z_n} \right| \geq 1$ *then the series diverges.*

Note that we have used the root test to prove the ratio test. The root test is indeed "stronger" in the sense that it provides a conclusion whenever the ratio test does, while it can be shown that the ratio test may be inconclusive when the root test is not. Therefore the main advantage of the ratio test is that it is sometimes easier to check. Using either of these tests to show divergence may not be the easiest approach because, as the proof reveals, divergence is implied by either test only when the terms of the series do not tend to 0. For concrete series this fact is generally more easily verified directly. A good general strategy for testing series is to first use the divergence test. If this does not reveal divergence, consider the root test. If this seems difficult to check, try the ratio test.

**Proposition 192** *(Alternating Series Test) If $(a_n)$ is a decreasing sequence of positive real numbers such that $a_n \to 0$ then $\sum_{n=1}^{\infty}(-1)^{n-1}a_n$ is convergent.*

**Proof.** For any $m$, since $(a_n)$ is decreasing, we have

$$\sigma_{2m} = (a_1 - a_2) + \cdots + (a_{2m-1} - a_{2m})$$

and

$$\sigma_{2m} = a_1 - (a_2 - a_3) - ... - a_{2m} < a_1$$

The first formula shows that $(\sigma_{2m})$ is a positive increasing sequence, and the second shows $(\sigma_{2m})$ is bounded above, hence $\sigma_{2m} \nearrow s$ for some real number $s$. Let $\varepsilon > 0$. Then for all large $n$ we have $a_n < \frac{\varepsilon}{2}$ and, if $n$ is even, $s - \sigma_n < \varepsilon/2$. If $n$ is odd then $n+1$ is even and

$$|s - \sigma_n| = |s - \sigma_{n+1} - a_{n+1}| \leq |s - \sigma_{n+1}| + |a_{n+1}| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

■

**Example 193** *The harmonic series is the series $\sum_{n=1}^{\infty} \frac{1}{n}$. The harmonic series is a $p$-series, convergence and divergence of which are considered in Example* **??**. *However, one may see directly that the harmonic series diverges. Writing out the terms one sees*

$$1 + \left(\frac{1}{2}\right) + (\frac{1}{3} + \frac{1}{4}) + (\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}) + \cdots$$

*where each of the sets in parentheses consists of $2^{k-1}$ terms each of which is at least $2^{-k}$ and therefore has sum at least $\frac{1}{2}$. It follows that the sequence of partial sums is unbounded. However, the terms of the series decrease to 0, and so by the alternating series test, the alternating harmonic series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ is convergent, but is not absolutely convergent.*

We saw earlier that in general there is no "infinite associative law" for series, and as one might expect there is in general no "infinite commutative" law that allows one shuffle the order of the terms of a series without affecting the sum. Worse yet, it can be shown (see [**?**], Theorem 3.55) that any series that converges but does not converge absolutely (like the alternating harmonic series) can have its terms rearranged so that it converges to any arbitrary real number! On the other hand, as we will see next, the sum any *absolutely* convergent series is unaffected by rearrangement.

**Definition 194** *A rearrangement of a series* $\sum_{n=1}^{\infty} z_n$ *is a series* $\sum_{k=1}^{\infty} w_k$ *such that there is a bijection* $\psi : \mathbb{N} \to \mathbb{N}$ *with the property that* $w_{\psi(k)} = z_k$ *for all* $k$*.*

Note that if $\sum_{k=1}^{\infty} w_k$ is a rearrangement of $\sum_{n=1}^{\infty} z_n$ then $\sum_{n=1}^{\infty} z_n$ is a rearrangement of $\sum_{k=1}^{\infty} w_k$ (via the function $\psi^{-1}$).

**Proposition 195** *Let* $\sum_{n=1}^{\infty} z_n$ *be an absolutely convergent series. If* $\sum_{k=1}^{\infty} w_k$ *is a rearrangement of* $\sum_{n=1}^{\infty} z_n$ *then* $\sum_{k=1}^{\infty} w_k$ *is convergent and* $\sum_{n=1}^{\infty} z_n = \sum_{k=1}^{\infty} w_k$*.*

**Proof.** Let $(\sigma_n)$, $(\tau_k)$ be the partial sum sequences for $\sum_{n=1}^{\infty} z_n = z$ and $\sum_{k=1}^{\infty} w_k$, respectively, $\psi$ be the bijection of the rearrangement, and let $\varepsilon > 0$. Since $\sum_{n=1}^{\infty} z_n$ is absolutely convergent, for all large $m$ $|\sigma_m - z| < \frac{\varepsilon}{2}$ and $\sum_{k=m}^{\infty} |z_k| < \frac{\varepsilon}{2}$. Now let

$$K_1 := \{\psi(1), ..., \psi(m)\}$$

and

$$M := \max K_1 \geq m$$

and

$$K_2 := \{1, ..., M\} \backslash K_1.$$

Since $w_{\psi(k)} = z_k$, $\sigma_m = \sum_{n \in K_1} w_n$ and if $n \in K_2$ then $\psi^{-1}(n) > m$. Therefore

$$|\tau_M - \sigma_m| = \left| \sum_{n \in K_2} w_n \right| \leq \sum_{n \in K_2} |w_n| = \sum_{n \in K_2} \left| z_{\psi^{-1}(n)} \right| \leq \sum_{k=m}^{\infty} |z_k| < \frac{\varepsilon}{2}.$$

Finally

$$|\tau_M - z| \leq |\tau_M - \sigma_m| + |\sigma_m - z| < \varepsilon.$$

■

Note that the above proposition can be applied to the series $\sum_{n=1}^{\infty} |z_n|$ and $\sum_{k=1}^{\infty} |w_k|$ to conclude that $\sum_{k=1}^{\infty} w_k$ is also absolutely convergent. Also, one may conclude that if $\sum_{k=1}^{\infty} z_k$ is not absolutely convergent then neither is any rearrangement of it.

**Exercise 133** *Consider the "double alternating harmonic series"*

$$1 - 1 + \frac{1}{2} - \frac{1}{2} + \frac{1}{3} - \frac{1}{3} + \frac{1}{4} \cdots$$

*Show that this series is convergent and has a rearrangement such that the sequence of partial sums* $(\sigma_k)$ *satisfies* $\sigma_k \to \infty$*. Hint: you will not be able to write down an explicit rearrangement; show one exists by showing that you can always move enough positive terms in front of each negative term to make the partial sums get large.*

## 3.5 Power Series

**Definition 196** *Given a sequence $(c_n)_{n=0}^{\infty}$ of complex numbers, an expression of the form*

$$\sum_{n=0}^{\infty} c_n z^n$$

*is called a power series, and the numbers $c_n$ are called the coefficients of the power series.*

A power series is a kind of "infinite polynomial" except that, because the sum is infinite, there is the question of convergence. Of course the series always converges for $z = 0$ and so there is always some nonempty subset $C$ of $\mathbb{C}$ on which the power series defines a function into $\mathbb{C}$. We will study the properties of these very important functions later, but in the present section we are primarily concerned with understanding the domain of this function.

**Example 197** *The power series $\sum_{n=0}^{\infty} n^n z^n$ converges only for $z = 0$, for if $z \neq 0$ the terms of the series are of the form $(nz)^n$. Since $z$ is fixed, for large $n$ we have that $|nz| > 1$ and therefore the terms of the series do not tend to 0. This series has the smallest possible domain of convergence.*

**Example 198** *Consider the series $\sum_{n=0}^{\infty} \frac{z^n}{n!}$. Applying the ratio test we have*

$$\frac{\left|z^{n+1}\right| n!}{|z^n|\,(n+1)!} = \frac{|z|}{n+1} \to 0 < 1$$

*for all $z$, so this power series converges for all $z$. The reader should recognize that this power series has the coefficients of the Maclaurin series for the real function $e^x$. This motivates the following definition.*

**Definition 199** *Define $\exp : \mathbb{C} \to \mathbb{C}$ by $\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}$. This function is called the (complex) exponential function. We define $e := \exp(1)$.*

Note that if $x$ is positive, the terms in the series are positive. Since the first term is always 1, $\exp(x) > 1$ when $x > 0$. Examining the first two terms shows that $e > 2$.

**Proposition 200** *For any complex numbers $z, w$, we have $\exp(z)\exp(w) = \exp(z + w)$ and*

1. $\exp(0) = 1$.

2. $\exp(-z) = \frac{1}{\exp(z)}$ *for any complex $z$.*

3. $(\exp(z))^n = \exp(nz)$ *for any complex $z$ and natural number $n$.*

4. $\exp(\frac{1}{q}) = \sqrt[q]{e}$ *for any natural number $q$.*

*5.* $\exp(\frac{p}{q}) = (\sqrt[q]{e})^p$ *for any integers* $p$ *and* $q$ *with* $q > 0$*.*

**Exercise 134** *Prove the above proposition.*

**Corollary 201** *The restriction of* exp *to the real numbers is a strictly increasing, positive, unbounded function.*

**Proof.** If $x < y$ then $y = x + k$ for some $k > 0$. We have already observed that $\exp(k) > 1$ if $k > 0$ and therefore $\exp(y) = \exp(x + k) = \exp(x)\exp(k) > \exp(x)$. If $x < 0$ then $\exp(x) = \frac{1}{\exp(-x)} > 0$. Therefore $\exp(x)$ is always positive. It is easy to show by induction that $2^n > n$ for any $n \in \mathbb{N}$ and therefore $\exp(n) = e^n > 2^n > n$, which shows the function is unbounded. ■

**Example 202** *The power series* $\sum_{n=0}^{\infty} z^n$ *is simply the geometric series (Example 184). We have already seen that this series converges for* $|z| < 1$ *and diverges for* $|z| \geq 1$*.*

So far we have seen that power series can converge only at 0, on an open ball, or on all of $\mathbb{C}$. The next theorem shows that the domain is always one of these three possibilities, possibly including part of the boundary circle when the domain is an open ball. We will use the following conventions: If $\tau = \infty$ then define $\frac{1}{\tau} = 0$, if $\tau = 0$ then define $\frac{1}{\tau} = \infty$, and let $B(0, \infty) := \mathbb{C}$. The *open unit ball* refers to $B(0, 1)$.

**Theorem 203** *Let* $\sum_{n=0}^{\infty} c_n z^n$ *be a power series and let*

$$\tau := \limsup \sqrt[n]{|c_n|} \text{ and } R := \frac{1}{\tau}.$$

*If* $R = 0$ *then* $\sum_{n=0}^{\infty} c_n z^n$ *converges only on the set* $\{0\}$*. If* $R > 0$ *then* $\sum_{n=0}^{\infty} c_n z^n$ *converges if* $|z| < R$ *and diverges if* $|z| > R$*.*

**Proof.** Applying the root test we have

$$\limsup \sqrt[n]{|c_n z^n|} = \limsup |z| \sqrt[n]{|c_n|} = |z| \, \tau$$

provided $\tau$ is finite. If $\tau = 0$ then the series converges for all $z \in \mathbb{C}$, otherwise the lim sup is less than 1 if $|z| < R = \frac{1}{\tau}$ and greater than 1 if $|z| > R = \frac{1}{\tau}$. If $\tau = \infty$ then the limit is infinite unless $z = 0$. ■

The extended real number $R$ in the above theorem is called the *radius of convergence*. For $z$ such that $|z| = R$, convergence or divergence is possible; sometimes this can be checked using other methods. Note that although the radius of convergence is defined by a specific formula, the ratio test can be used to find the radius of convergence, provided $\lim \left| \frac{c_{n+1}}{c_n} \right|$ exists (see the next exercise for a more general statement).

**Exercise 135** *Let* $\sum_{n=0}^{\infty} c_n z^n$ *be a series.*

1. *Show that if there exists some* $r \in (0,\infty)$ *such that* $\sum_{n=0}^{\infty} c_n z^n$ *converges for all* $z \in B(0,r)$ *and diverges for all* $z \notin C(0,r)$ *then* $r$ *is the radius of convergence of the series.*

2. *Show that if* $\limsup \left|\frac{c_{n+1}}{c_n}\right| = \tau$ *then the radius of convergence* $R$ *of the series satisfies* $R \geq \frac{1}{\tau}$.

3. *Show that if* $\liminf \left|\frac{c_{n+1}}{c_n}\right| = \sigma$ *then the radius of convergence* $R$ *of the series satisfies* $R \leq \frac{1}{\sigma}$.

**Example 204** *Consider the power series* $\sum_{n=1}^{\infty} \frac{z^n}{n^2}$. *Applying the ratio test from the previous exercise,*

$$\left|\frac{n^2}{(n+1)^2}\right| = \frac{n^2}{(n+1)^2} \to 1$$

*so the series converges absolutely if* $|z| < 1$ *and diverges if* $|z| > 1$.

**Exercise 136** *Show that the power series* $\sum_{n=1}^{\infty} \frac{z^n}{n}$ *converges absolutely on the open unit ball and diverges outside the closed unit ball. Show that it converges at* $z = -1$ *and diverges at* $z = 1$. *It can be shown that this power series in fact converges on the unit circle except at* $z = 1$.

**Exercise 137** *Find the radius of convergence of the following power series* $z - \frac{z^3}{6} + \frac{z^5}{5!} - \cdots$ *Do you recognize the coefficients of this series?*

**Exercise 138** *Prove or disprove: If* $\sum_{n=0}^{\infty} c_n z^n$ *and* $\sum_{n=0}^{\infty} b_n z^n$ *are power series and the radius of convergence of each series is* $R$, *then the radius of convergence of* $\sum_{n=0}^{\infty}(b_n + c_n)z^n$ *is also* $R$.

## 3.6 Pointwise and monotone convergence

At this point, our understanding of convergence involves sequences in metric spaces. We have seen in Example 31 that it is possible to define a metric on some sets of functions, and hence we can discuss convergence of functions in the same way we would any elements of a metric space, applying all of the results in the preceding chapters. We will consider this further later in this chapter. In the meantime we will consider a more basic kind of convergence of functions called pointwise convergence, which is not defined in terms of a metric on the set of functions.

**Definition 205** *Let* $A$ *be a set,* $X$ *be a metric space, and* $(f_i)$ *be a sequence of functions* $f_i : A \to X$. *We say that* $f_i$ *converges pointwise to* $f : A \to X$ *(written* $f_i \xrightarrow{p} f$*) if for every* $x \in A$, $f_i(x) \to f(x)$ *in* $X$.

The word "pointwise" refers to the fact that each point in the domain gives rise to a convergent sequence in the range, but as we will see in the next couple of examples the "rate" of convergence, in some sense, may vary widely from point to point.

**Exercise 139** *Prove that the limit of a pointwise convergent sequence of functions is unique. That is, if* $f_i \xrightarrow{p} f$ *and* $f_i \xrightarrow{p} g$ *then* $f = g$.

**Example 206** *Let* $f_n : \mathbb{R} \to \mathbb{R}$ *be given by* $f_n(x) = \frac{x}{n}$. *Then* $f_n \xrightarrow{p} 0$. *In fact, for any* $x$, *we have* $f_n(x) = \frac{x}{n} \to 0$. *Note that "how fast"* $f_n(x)$ *tends to* $0$ *depends on* $x$ *in the following sense: Given* $x \in \mathbb{R}$ *and* $\varepsilon > 0$, *there exists an* $N_x$ *such that if* $n \geq N_x$ *then* $|f_n(x)| = |f_n(x) - 0| < \varepsilon$. *In fact, we need* $\left|\frac{x}{n}\right| < \varepsilon$ *or* $n > \frac{|x|}{\varepsilon}$, *which is equivalent to* $N_x > \frac{|x|}{\varepsilon}$. *In other words, the larger* $|x|$ *is, the larger we have to choose* $N_x$ *for the same* $\varepsilon$. *This means that for given* $\varepsilon > 0$, *we cannot choose a single* $N$ *such that* $|f_n(x)| < \varepsilon$ *for all* $x \in \mathbb{R}$. *One can see this behavior graphically: Since the line* $y = \frac{x}{n}$ *slopes upwards, the farther out you go the larger* $n$ *must be in order to make* $\frac{x}{n}$ *close to* $0$.

**Example 207** *Let* $f_n : [0,1] \to \mathbb{R}$ *be defined by*

$$f_n(x) = \frac{1}{1+nx}.$$

*We will check that* $f_n$ *converges pointwise to the function* $f$ *defined by* $f(0) = 1$ *and* $f(x) = 0$ *when* $x \neq 0$. *First,* $f_n(0) = 1$ *for all* $n$, *so* $f_n(0) \to 1$. *Now let* $x \in (0,1]$. *Then* $nx \to \infty$ *and therefore* $f_n(x) \to 0$. *Note that the functions* $f_n$ *are all continuous, but* $f$ *is not! This example shows that it is possible for a sequence of continuous real functions defined on a compact set to converge pointwise to a function that is not continuous. This already indicates that pointwise convergence is too weak for some purposes. What goes wrong again in this case is a lack of uniformity of convergence, although it is not as obvious as in the preceding example. As before, fix* $x \in (0,1]$ *and let* $\varepsilon > 0$. *Then* $\left|\frac{1}{1+nx}\right| < \varepsilon$ *is equivalent to* $n > \frac{\frac{1}{\varepsilon}-1}{x}$. *So we need to choose our* $N_x > \frac{\frac{1}{\varepsilon}-1}{x}$. *But* $\frac{\frac{1}{\varepsilon}-1}{x} \to \infty$ *as* $x \to 0$; *in other words if* $x$ *is close to* $0$ *then the* $N_x$ *must be chosen arbitrarily large. Note that for* $x = 0$ *any* $N$ *will work for any* $\varepsilon$.

*This same example illustrates another "problem" with the pointwise convergence that can occur. Note that it follows from our preceding computations that*

$$\lim_{n\to\infty}\left(\lim_{x\to 0} f_n(x)\right) = \lim_{n\to\infty} f_n(0) = \lim_{n\to\infty} 1 = 1$$

*but*

$$\lim_{x\to 0}\left(\lim_{n\to\infty} f_n(x)\right) = \lim_{x\to 0} f(x) = 0.$$

*In other words, "switching the order of limits" changes the value of those limits! In fact, as the reader should notice, the discontinuity of the limit function* $f$ *and the problem with the order of limits are closely related.*

The following lemma is an immediate consequence of the definition and Proposition 167. We note only that if $g_n \xrightarrow{p} g$ and $g(x) \neq 0$ it is possible that for all $n \in \mathbb{N}$ there exists some $x \in A$ such that $g_n(x) = 0$ (the reader is welcome to construct an example) and hence no tail of the sequence $\left(\frac{f_n}{g_n}\right)$ is defined on all of $A$. Hence the stronger requirement for the domain in the third part.

**Lemma 208** *Let $(f_n)$ and $(g_n)$ be sequence of functions from a set $A$ into $\mathbb{C}$. If $f_n \xrightarrow{p} f$ and $g_n \xrightarrow{p} g$ then*

1. $f_n + g_n \xrightarrow{p} f + g$
2. $f_n \cdot g_n \xrightarrow{p} f \cdot g$
3. $\frac{f_n}{g_n} \xrightarrow{p} \frac{f}{g}$ *on* $\{x \in A : g(x) \neq 0$ *and* $g_n(x) \neq 0$ *for all* $n\}$.

Example 207 also illustrates another kind of convergence when the range space is $\mathbb{R}$, defined as follows:

**Definition 209** *Let $A$ be a set and $(f_n)$ be a sequence of functions $f_n : A \to X$, where $X$ is a metric space. We say $(f_n)$ is uniformly bounded if $\bigcup_{n=1}^{\infty} f_n(A)$ is bounded in $X$. If $f_n : A \to \mathbb{R}$ we say that $(f_n)$ is decreasing (resp. increasing) if for all $x \in A$ the sequence $(f_n(x))$ is a decreasing (resp. increasing) sequence in $\mathbb{R}$. If in addition $(f_n)$ converges pointwise to $f : X \to \mathbb{R}$ we write $f_n \searrow f$ (resp. $f_n \nearrow f$). If $(f_n)$ is increasing or decreasing we say that $(f_n)$ is monotone.*

**Exercise 140** *Let $(f_n)$ be a sequence of functions $f_n : A \to \mathbb{C}$. Show that $(f_n)$ is uniformly bounded if and only if there exists some $M \geq 0$ such that for all $x \in A$ and $n$, $|f_n(x)| \leq M$.*

Monotone convergence is "stronger" than pointwise convergence in the sense that there are sequences that converge pointwise but do not converge monotonically, as will be shown in the next exercise:

**Exercise 141** *Give an example of a sequence of real functions that converges pointwise but not monotonically.*

As will be shown in an exercise below, an arbitrary sequence of functions may not converge to any function at all (just like an arbitrary sequence in a metric space need not converge), but uniformly bounded, monotone sequences of real functions always do converge pointwise to a function:

**Lemma 210** *Let $(f_n)$ be decreasing (resp. increasing) uniformly bounded sequence of functions $f_n : A \to \mathbb{R}$. Then there exists some function $f : A \to \mathbb{R}$ such that $f_n \searrow f$ (resp. $f_n \nearrow f$).*

**Proof.** Let $x \in A$. Then by assumption the sequence $(f_n(x))$ is a bounded, decreasing (resp. increasing) sequence in $\mathbb{R}$ and therefore $f_n(x) \to y$ for some $y \in \mathbb{R}$. We define $f(x) := y$. By construction $f_n \searrow f$ (resp. $f_n \nearrow f$). ■

**Exercise 142** *Let $(f_n)$ be a sequence of functions from a set $A$ into a complete metric space $X$. Prove that if for all $x \in A$ the sequence $(f_n(x))$ is Cauchy then there exists some function $f : A \to X$ such that $f_n \xrightarrow{p} f$.*

**Exercise 143** *Let $f_n(x) := x^n$ for $n \in \mathbb{N}$.*

1. *Show that $(f_n)$ converges monotonically on $(-1, 1]$ (i.e. when each $f_n$ is restricted to $(-1, 1]$).*
2. *Show that $(f_n)$ does not converge pointwise to any function on $[-1, 1]$.*
3. *Show that $(f_n)$ has a subsequence that converges pointwise to a function $f$ on $[-1, 1]$ that is not continuous.*
4. *Show that there exists $a \in [-1, 1]$ such that $\lim_{x\to a} (\lim_{n\to\infty} f_n(x))$ exists but $\lim_{n\to\infty} (\lim_{x\to a} f_n(x))$ does not.*

## 3.7 Uniform convergence

In Example 31 we constructed a metric on the space of continuous functions $f : [0, 1] \to \mathbb{R}$ by simply defining

$$d(f, g) = \max_{[0,1]}\{|f(x) - g(x)|\}.$$

This definition is made possible by the Max-Min Theorem for continuous real-valued functions, which depends both on the compactness of $[0, 1]$ and the continuity of the functions in question. If one considers more generally functions defined on metric spaces one might try to simply replace the maximum by the supremum, but it is possible that the supremum is infinite: for example $\sup_{(0,1)}\{\left|x^2 - \frac{1}{x}\right|\} = \infty$. While it is possible to develop a theory of metrics having infinite values, there is a more useful approach to this problem, namely by "truncating" large distances so that the maximum distance between functions is at most 1. If this seems a bit crude, note that what matters, in terms of convergence, is small distances, not large ones.

**Definition 211** *Let $(X, d)$ be a metric space. For every $x, y \in X$, define $d^1(x, y) = \min\{d(x, y), 1\}$.*

**Proposition 212** *If $(X, d)$ is a metric space then $d^1$ is a metric on $X$ that is topologically equivalent to $d$.*

**Proof.** The proof that $d^1$ is a metric is trivial except for the triangle inequality $d^1(x, z) \le d^1(x, y) + d^1(y, z)$, which must be sorted into cases. First note that for any $a, b \in X$, $d^1(a, b) \le 1$ and $d^1(a, b) \le d(a, b)$. If $d^1(x, y) = 1$ or $d^1(y, z) = 1$ then

$$d^1(x, z) \le 1 \le d^1(x, y) + d^1(y, z).$$

Otherwise, $d^1(x,y) = d(x,y)$ and $d^1(y,z) = d(y,z)$, and we have

$$d^1(x,y) + d^1(y,z) = d(x,y) + d(y,z) \geq d(x,z) \geq d^1(x,z).$$

To see that these metrics are topologically equivalent, let $(x_i)$ be a sequence in $X$. If $x_i \to x$ in $(X,d)$ then $d(x_i,x) \to 0$. But for all large $i$, $d(x_i,x) < 1$ and hence $d(x_i,x) = d^1(x_i,x)$, so $d^1(x_i,x) \to 0$ and $x_i \to x$ in $(X,d^1)$. A similar argument shows if $x_i \to x$ in $(X,d^1)$ then $x_i \to x$ in $(X,d)$ and it follows from Lemma 110 that the two metrics are topologically equivalent. ∎

**Definition 213** *Let $A$ be a set and $X$ be a metric space. Let $\mathcal{F}(A,X)$ denote the set of all functions $f : A \to X$ with the metric defined, for $f,g \in \mathcal{F}(A,X)$, by*

$$d(f,g) := \sup_{x\in A} \left\{ d^1(f(x),g(x)) \right\}. \tag{3.7}$$

*If a sequence $(f_n)$ converges to $f$ in $\mathcal{F}(A,X)$ we will say that $f_n$ converges uniformly to $f$ and write $f_n \to f$.*

Using essentially the same proof as in Example 31, we see that the function $d$ defined in the above definition is actually a metric, which satisfies $d(f,g) \leq 1$ for all $f,g \in \mathcal{F}(A,X)$. Note that every subset, and hence any sequence, in $\mathcal{F}(A,X)$ is bounded; this is partly why the separate term "uniformly bounded" was introduced.

**Exercise 144** *Note that the set $\mathcal{C}$ defined in Example 31 is a subset of $\mathcal{F}([0,1],\mathbb{R})$. Show that the metric $d$ defined in Example 31 is different from, but topologically equivalent to, the subspace metric on $\mathcal{C}$ as a subspace of $\mathcal{F}([0,1],\mathbb{R})$ with the metric defined in Definition 213.*

**Lemma 214** *If $f_n \to f$ in $\mathcal{F}(A,X)$ then $f_n \xrightarrow{p} f$.*

**Proof.** Let $x \in A$ and $\varepsilon > 0$. By definition of uniform convergence, for all large $n$ we have $d(f_n,f) < \varepsilon$, which implies $d^1(f_n(x),f(x)) < \varepsilon$ for all large $i$. In other words, $f_n(x) \to f(x)$ in $d^1$ and hence in $d$. ∎

**Exercise 145** *Show that the sequence $(f_n)$ of Exercise 143*

1. *converges pointwise to a continuous function on $[0,1)$*
2. *converges uniformly on $[0,c]$ for any $0 < c < 1$, but*
3. *does not converge uniformly on $[0,1) = \bigcup_{0<c<1} [0,c]$. This shows that a sequence of functions that converges uniformly on each of a collection of sets need not converge uniformly on the union of those sets. We will see other examples of this kind of behavior later.*

**Lemma 215** *Let $A$ be a set, $X$ be a metric space, and $(f_n)$ be a sequence of functions in $\mathcal{F}(A,X)$. Then $f_n \to f$ in $\mathcal{F}(A,X)$ if and only if for every $\varepsilon > 0$ there exists an $N$ such that for all $n \geq N$ and all $x \in A$, $d(f_n(x),f(x)) < \varepsilon$.*

**Proof.** Suppose first that $f_n \to f$ and let $\varepsilon > 0$. Define $\delta := \min\{\varepsilon, 1\}$. By definition there exists an $N$ such that for all $n \geq N$, $d(f_n, f) < \delta$, which implies that $d^1(f_n(x), f(x)) < \delta$ for all such $n$ and $x \in A$. But then

$$d(f_n(x), f(x)) = d^1(f_n(x), f(x)) < \delta \leq \varepsilon$$

for all $n \geq N$ and $x \in A$. The converse is an exercise. ∎

**Corollary 216** *If $(f_n)$ is a sequence in $\mathcal{F}(A, X)$ then $f_n \to f$ if and only if $\sup_{x \in A} \{d(f_n(x), f(x))\} \to 0$.*

The condition given in Lemma 215 is sometimes used as the definition of uniform convergence. We have chosen to define it in terms of the space $\mathcal{F}(A, X)$ because this allows us to use certain theorems that we have proved about metric spaces. Lemma 215, on the other hand, is very often useful in practice and we will generally use it without direct reference to it. Lemma 215 also more clearly reveals the difference between uniform and pointwise convergence: Given a fixed $\varepsilon > 0$, uniform convergence implies there is an $N$ that "works" uniformly, i.e. for every point in $A$ simultaneously, but pointwise convergence only provides an $N_x$ that "works" for an individual $x$ and $N_x$ may vary from point to point. The reader should take a moment to look again at Examples 206 and 207, which, according to our current definitions, are examples of sequences that converge pointwise but not uniformly.

**Exercise 146** *Finish the proof of Lemma 215.*

We have already seen that one of the deficiencies of monotone and pointwise convergence is that switching the order of iterated limits can fail to preserve the limit, or even whether the limit exists. This is not a problem with uniform convergence in complete metric spaces.

**Theorem 217** *Let $(f_n)$ be a sequence of functions $f_n : A \to Y$, where $A$ is a subset of a metric space $X$ and $Y$ is a metric space. Suppose $f_n$ converges uniformly on $A$ to $f : A \to Y$ and for some $x_0 \in \overline{A}$, $y_n := \lim_{x \to x_0} f_n(x)$ exists for all $n$. Then*

1. *$(y_n)$ is Cauchy and*
2. *if $(y_n)$ is convergent (e.g. if $Y$ is complete), then*

$$\lim_{n \to \infty} \lim_{x \to x_0} f_n(x) = \lim_{x \to x_0} f(x) = \lim_{x \to x_0} \lim_{n \to \infty} f_n(x) \tag{3.8}$$

*(in particular all limits in this expression exist).*

**Proof.** For the first part, let $\varepsilon > 0$. Since the sequence $(f_n)$ is Cauchy in $\mathcal{F}(A, X)$, for all large $m$ and $n$ we have $d(f_m(x), f_n(x)) < \frac{\varepsilon}{2}$ for all $x \in A$. Since the distance function is continuous, Corollary 67 implies

$$d(y_m, y_n) = d(\lim_{x \to x_0} f_m(x), \lim_{x \to x_0} f_n(x)) = \lim_{x \to x_0} d(f_m(x), f_n(x)) \leq \frac{\varepsilon}{2} < \varepsilon$$

for all large $m$ and $n$. That is, $(y_n)$ is Cauchy.

For the second part let $y_0 := \lim y_n$. The proof will be complete if we show that $\lim_{x\to x_0} f(x) = y_0$. Let $\varepsilon > 0$. For all large $n$ we have both $d(f_n(x), f(x)) < \frac{\varepsilon}{3}$ for all $x \in A$ and $d(y_n, y_0) < \frac{\varepsilon}{3}$. For any such $n$ the fact that $y_n = \lim_{x\to x_0} f_n(x)$ means, by definition, that there exists a $\delta > 0$ such that if $0 < d(x, x_0) < \delta$ and $x \in A$ then $d(f_n(x), y_n) < \frac{\varepsilon}{3}$. Putting all of these together we have, when $0 < d(x, x_0) < \delta$,

$$d(f(x), y_0) \le d(f(x), f_n(x)) + d(f_n(x), y_n) + d(y_n, y_0) < \varepsilon.$$

That is, $\lim_{x\to x_0} f(x) = y_0$. ■

**Corollary 218** *Let $(f_n)$ be a sequence of functions $f_n : X \to Y$ between metric spaces with $f_n \to f$. If each $f_n$ is continuous at $x_0 \in X$ then $f$ is continuous $x_0$.*

**Proof.** By the continuity of each $f_n$, $y_n := \lim_{x\to x_0} f_n(x) = f_n(x_0)$ and since $f_n \xrightarrow{p} f$, $y_n \to f(x_0)$. We may apply Theorem 217 and obtain

$$\lim_{x\to x_0} f(x) = \lim_{n\to\infty} \left( \lim_{x\to x_0} f_n(x) \right) = \lim_{n\to\infty} f_n(x_0) = f(x_0).$$

■

**Lemma 219** *Let $A$ be a set and $X$ be a metric space. If $(f_i)$ is Cauchy in $\mathcal{F}(A, X)$ and $f_i \xrightarrow{p} f$ then $f_i \to f$.*

**Proof.** Let $\varepsilon > 0$. Since $(f_i)$ is Cauchy there exists some $N$ such that if $i, j \ge N$ then $d(f_i(x), f_j(x)) < \frac{\varepsilon}{2}$ for all $x \in A$. Now fix any $i \ge N$. Then for all $j \ge i$ we have for all $x$

$$d(f_i(x), f_j(x)) < \frac{\varepsilon}{2}.$$

Since $f(x) = \lim\limits_{j\to\infty} f_j(x)$, by Corollary 74 and Exercise 52 we obtain

$$d(f_i(x), f(x)) = \lim_{j\to\infty} d(f_i(x), f_j(x)) \le \frac{\varepsilon}{2}$$

for all $x$, i.e., $d(f_i, f) \le \frac{\varepsilon}{2} < \varepsilon$ for all $i \ge N$. ■

**Proposition 220** *Let $A$ be a set and $X$ be a complete metric space. Then $\mathcal{F}(A, X)$ is complete.*

**Proof.** Let $(f_i)$ be a Cauchy sequence in $\mathcal{F}(A, X)$. Then for each $x \in A$ the sequence $(f_i(x))$ is Cauchy in $X$, hence convergent. Let $f(x) := \lim(f_i(x))$. Then $f_i \xrightarrow{p} f$ and the proof is finished by Lemma 219. ■

**Definition 221** *Let $X$ and $Y$ be metric spaces. By $\mathcal{C}(X, Y)$ we denote the subset of $\mathcal{F}(X, Y)$ of all continuous functions $f : X \to Y$ with the subspace metric.*

As an immediate consequence of Corollary 218 we obtain.

**Theorem 222** *If $X$ and $Y$ are metric spaces with $Y$ complete then $\mathcal{C}(X,Y)$ is a closed subset of $\mathcal{F}(X,Y)$.*

**Corollary 223** *If $X$ and $Y$ are metric spaces and $Y$ is complete then $\mathcal{C}(X,Y)$ is complete.*

**Exercise 147** *Let $f_n, g_n : A \to \mathbb{C}$ be functions for all $n \in \mathbb{N}$ such that $(f_n)$ and $(g_n)$ converge uniformly to functions $f$ and $g$ respectively.*

1. *Prove $(f_n + g_n)$ converges uniformly to $f + g$.*
2. *Prove that if $(f_n)$ and $(g_n)$ are uniformly bounded then $(f_n g_n)$ converges uniformly to $(fg)$.*

**Example 224** *One cannot remove the requirement in the above exercise that the sequences be uniformly bounded. For example, let $f_n(z) := \frac{1}{z}$ and $g_n(z) = \frac{1}{n}$ for all $n \in \mathbb{N}$ and $z \in D := B(0,1)\backslash\{0\}$. For any $z \in D$, $|g_n(z) - 0| = \frac{1}{n}$, and it follows that $g_n$ converges uniformly to the constant function $g(z) = 0$ on $D$. On the other hand, $(f_n)$ is simply a constant sequence and so converges uniformly to $f(z) := \frac{1}{z}$. We have $fg(z) = 0$ for all $z$ in $D$. Now for any $\varepsilon > 0$ and any $n \in \mathbb{N}$, let $z$ be such that $|z| < \frac{1}{n\varepsilon}$. Then*

$$|f_n g_n(z) - fg(z)| = \left|\frac{1}{nz}\right| = \frac{1}{n\,|z|} > \varepsilon.$$

*Therefore $(f_n g_n)$ does not converge uniformly to $0$ on $D$ (although it still converges pointwise to $0$).*

**Exercise 148** *Consider the sequence of real functions $f_n(x) := \frac{x}{nx^2+1}$.*

1. *What is the pointwise limit $f$ of $(f_n)$?*
2. *Prove that $(f_n)$ converges uniformly.*
3. *Using theorems from basic calculus show that $\lim f_n'(0) \neq f'(0)$. That is, a uniformly convergent sequence of differentiable functions need not have even pointwise convergent derivatives.*

## 3.8 Series of Functions

Series of functions are defined in an analogous way to series of complex numbers, namely an expression of the form

$$\sum_{n=0}^{\infty} f_n$$

where each $f_n : A \to \mathbb{C}$ is a function defined on a set $A$. Frequently $A$ will be a subset of complex numbers. As with the case of series of numbers, with series of functions there is a question of convergence. In this case the sequence of partial sums is a sequence of functions $(g_k)_{k=1}^{\infty}$, with $g_k := \sum_{n=0}^{k} f_n$. We can ask on which subsets $B$ of $A$ $(g_k)_{k=1}^{\infty}$ converges pointwise or uniformly, and when it does we will say that $\sum_{n=0}^{\infty} f_n$ converges pointwise or uniformly, respectively, on $B$. In either case we will denote the limiting function by $\sum_{n=0}^{\infty} f_n$, and $\left(\sum_{n=0}^{\infty} f_n\right)(x) = \sum_{n=0}^{\infty} f_n(x)$ by definition. We also say that $\sum_{n=0}^{\infty} f_n$ converges absolutely (pointwise or uniformly) if $\sum_{n=0}^{\infty} |f_n|$ converges (pointwise or uniformly, respectively). As with the case of numerical series, absolute convergence implies convergence, but the converse is not true.

We are primarily interested in uniform convergence because of the stronger properties it provides. For example, if $A$ is a metric space, each of the functions $f_n$ is continuous and the series $\sum_{n=0}^{\infty} f_n$ converges uniformly on $A$ then according to Corollaries 168 and 218 the function $\sum_{n=0}^{\infty} f_n$ is continuous on $A$. It is worth considering the meaning of "Cauchy" in this context. The difference between two terms $\sum_{n=0}^{k} f_n$ and $\sum_{n=0}^{m} f_n$ in the sequence of partial sums with $m > k$ is $\sum_{n=k+1}^{m} f_n$. Therefore the sequence of partial sums is Cauchy if and only if for every $\varepsilon > 0$ we have that $\left|\sum_{n=k}^{m} f_n(x)\right| < \varepsilon$ for all $x \in A$ and all large $k < m$. The next proposition, which is analogous to the Comparison Test for numerical series, is frequently useful.

**Proposition 225** *(Weierstrass M-test) Let $A$ be a set and $f_k : A \to \mathbb{C}$ be a sequence of functions. Suppose there is a convergent real series $\sum_{k=0}^{\infty} M_k$ such that for all $x \in A$ and $k \in \mathbb{N}$, $|f_k(x)| \le M_k$. Then $\sum_{k=0}^{\infty} f_k$ converges absolutely uniformly on $A$.*

**Proof.** To prove uniform convergence we need only prove that the sequence of partial sums is Cauchy. Since the sequence of partial sums of $\sum_{n=0}^{\infty} M_n$ is Cauchy, given $\varepsilon > 0$, for all large $m < k$ and all $x \in A$ we have

$$\left|\sum_{n=0}^{k} |f_n(x)| - \sum_{n=0}^{m} |f_n(x)|\right| = \left|\sum_{n=m+1}^{k} |f_n(x)|\right| = \sum_{n=m+1}^{k} |f_n(x)| \le \sum_{n=m+1}^{k} M_n < \varepsilon.$$

■

The most fundamental type of series is the power series $\sum_{n=0}^{\infty} a_n z^n$. In this case the partial sum functions are polynomials $\sum_{n=0}^{k} a_n z^n$. We have already seen that each power series has a radius of convergence $R \in [0, \infty]$ and for $z \in B(0, R)$ the numerical series $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely. In other words, the series $\sum_{n=0}^{\infty} a_n z^n$ converges pointwise on $B(0, R)$.

**Theorem 226** *If $\sum_{n=0}^{\infty} a_n z^n$ is a power series with radius of convergence $R > 0$ then $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely uniformly on any closed ball $C(0, r)$ such that $0 < r < R$.*

**Proof.** For any $z \in C(0, r)$ we have for all $k \ge 0$,

$$\left|a_k z^k\right| \le |a_k|\, r^k$$

Letting $M_k := |a_k| r^k$ and noting that, since $r < R$, the numerical series $\sum_{n=k+1}^{\infty} |a_n| r^n$ converges, the proof is finished by the Weierstrass M-test. ∎

**Corollary 227** *If $f(z) := \sum_{n=0}^{\infty} a_n z^n$ is a power series with radius of convergence $R > 0$ then $f$ is continuous on $B(0, R)$.*

**Proof.** Let $z \in B(0, R)$; so $|z| = r$, with $r < R$ and suppose $z_i \to z$ in $B(0, R)$. Let $\rho$ be such that $r < \rho < R$. Theorem 226 implies that $f$ is continuous on $C(0, \rho)$. Moreover, for large $i$, $z_i \in C(0, \rho)$. Therefore $f(z_i) \to f(z)$. ∎

**Exercise 149** *A power series centered at $z_0 \in \mathbb{C}$ is an expression of the form $\sum_{n=0}^{\infty} a_n (z - z_0)^n$. Show that $\sum_{n=0}^{\infty} a_n (z - z_0)^n$ converges uniformly for all $z \in B(z_0, r)$, where $r$ is less than the radius of convergence of the series $\sum_{n=0}^{\infty} a_n z^n$.*

**Exercise 150** *Let $f(z) := \sum_{n=0}^{\infty} \frac{1}{1+n^2 z}$.*

1. *Show that $f$ is absolutely and uniformly convergent on any set of the form $\mathbb{C} \backslash (B(0, r)$ where $r > 0$. Hint: Use Exercise 114.*
2. *Determine whether $f$ converges uniformly on $\mathbb{C} \backslash \{0\}$.*

**Exercise 151** *Show that the real function* exp : $\mathbb{R} \to (0, \infty)$ *is a continuous bijection.*

We are now in a position to define powers $a^r$, where $a > 0$ and $r$ is a real number. Since exp is a bijection on$\mathbb{R}$, it has an inverse, which we will give its usual name ln : $(0, \infty) \to \mathbb{R}$. We define $a^r := \exp(r \ln a)$. Note that $e^r = \exp(r \ln e) = \exp(r)$ for any real $r$. Therefore from now on we will denote $\exp(r)$ by $e^r$ when $r$ is real. Using Proposition 200 we have for real $r, s$,

$$a^{r+s} = e^{(r+s)\ln a} = e^{r \ln a + s \ln a} = a^r a^s.$$

Moreover,

$$e^{rs} = e^{(s \ln(e^r))} = (e^r)^s$$

and so

$$a^{rs} = e^{rs \ln a} = \left(e^{r \ln a}\right)^s = (a^r)^s .$$

We need to check that this "new" way to define powers is not really new, at least for rational exponent. But note that For a natural number $n$, induction on the above formula implies

$$a^n = e^{n \ln a} = e^{\ln a} \cdots e^{\ln a} = a \cdots a,$$

where the last two products have $n$ factors. Next $\left(a^{\frac{1}{q}}\right)^q = a^{\left(\frac{q}{q}\right)} = a^1 = a$, so $a^{\frac{1}{q}} = \sqrt[q]{a}$. Similarly it can be checked that for any integer $p$, $a^{\frac{p}{q}} = (\sqrt[q]{a})^p$. We know that the function $a^x = e^{x \ln a}$ is the composition of continuous functions and therefore we have continuously extended rational powers to powers of any real number.

# Chapter 4

# Integration

## 4.1 Riemann Integration

Throughout this section we assume that $a < b$. Recall from basic calculus that the Riemann integral of a real function $f : [a, b] \to \mathbb{R}$ is defined as follows: Take a *partition* $\mathcal{P}$ of the interval $[a, b]$, which is a set $\{x_j\}_{j=0}^{k}$ such that

$$x_0 = a < x_1 < \cdots < x_{k-1} < x_k = b.$$

Choose for each $j = 1, 2, ..., k$ some $c_j$ with $x_{j-1} \le c_j \le x_j$. The corresponding *Riemann sum* is

$$S(\mathcal{P}, \{c_j\}) := \sum_{j=1}^{k} f(c_j)(x_j - x_{j-1}).$$

Note that the value of the Riemann sum depends on both the partition and the choice of the constants $c_j$. The *Riemann integral* exists if these Riemann sums converge to a specific number as the partitions get "fine", regardless of how $c_j$ is chosen. More precisely, for any partition $\mathcal{P}$ of $[a, b]$ we define the *size* of $\mathcal{P}$ to be

$$\sigma(\mathcal{P}) := \max_j \{x_j - x_{j-1}\}.$$

That is, $\sigma(\mathcal{P})$ is the maximum length of the subintervals $[x_{j-1}, x_j]$ in the partition. We say $f$ is *Riemann integrable* on $[a, b]$ if there is some real number $I$ such that for every $\varepsilon > 0$ there exists some $\delta > 0$ such that $|S(\mathcal{P}, \{c_j\}) - I| < \varepsilon$ for any partition $\mathcal{P}$ such that $\sigma(\mathcal{P}) < \delta$ and any choice of $\{c_j\}$. The number $I$ is called the *Riemann integral of* $f$, denoted by $\int_a^b f$ or $\int_a^b f(x)dx$.

**Exercise 152** *Let $f$ be continuous on $[a, b]$. Prove that if $\mathcal{P}$ is any partition of $[a, b]$ then for any choice of $\{c_j\}$,*

$$(b - a) \min_{[a,b]} f \le S(\mathcal{P}, \{c_j\}) \le (b - a) \max_{[a,b]} f.$$

Which functions are Riemann integrable? We will not state the answer to this question later, but without proof. For now we will prove the usual theorem stated in elementary calculus, that continuous functions are Riemann integrable. We need a few preliminaries about partitions. If $\mathcal{P}$ and $\mathcal{P}'$ are partitions, we say that $\mathcal{P}'$ is a *refinement of* $\mathcal{P}$ if $\mathcal{P} \subset \mathcal{P}'$. That is, $\mathcal{P}'$ is obtained by adding more points to $\mathcal{P}$. Note that $\sigma(\mathcal{P}') \leq \sigma(\mathcal{P})$. Note also that every pair of partitions $\mathcal{P}$ and $\mathcal{P}''$ has a *common refinement* $\mathcal{P}' := \mathcal{P} \cup \mathcal{P}''$, which is a refinement of both $\mathcal{P}$ and $\mathcal{P}''$.

Given a continuous function $f : [a, b] \to \mathbb{R}$, we will be interested in the real sequence $(r_i)_{i=1}^{\infty}$ defined by

$$r_i := \sup\{S(\mathcal{P}, \{c_j\})\} : \sigma(\mathcal{P}) \leq \frac{1}{i}\}.$$

That is, $r_i$ is the supremum of all Riemann sums involving partitions of $[a, b]$ into subintervals of length at most $\frac{1}{i}$. According to Exercise 152, $(r_i)_{i=1}^{\infty}$ is bounded below. Moreover, $(r_i)$ is monotone decreasing. In fact if $\sigma(\mathcal{P}) \leq \frac{1}{i}$ then $\sigma(\mathcal{P}) \leq \frac{1}{k}$ for every $k \leq i$ and therefore

$$\left\{S(\mathcal{P}, \{c_j\}) : \sigma(\mathcal{P}) \leq \frac{1}{i}\right\} \subset \left\{S(\mathcal{P}, \{c_j\})\} : \sigma(\mathcal{P}) \leq \frac{1}{k}\right\}.$$

Therefore the supremum of the set on the left is not larger than the supremum of the set on the right.

Next we would like to consider a very specific Riemann sum for a given partition $\mathcal{P}$, namely

$$S_{\max}(\mathcal{P}) := \sum_{j=1}^{k} \left( \max_{[x_{j-1}, x_j]} f(x) \right) (x_j - x_{j-1}).$$

This is the Riemann sum having the largest value for the partition $\mathcal{P}$. One can similarly define $S_{\min}(\mathcal{P})$, and for any partition $\mathcal{P}$ we have $S_{\min}(\mathcal{P}) \leq S(\mathcal{P}, \{c_j\}) \leq S_{\max}(\mathcal{P})$.

**Exercise 153** *Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose $\mathcal{P}$ and $\mathcal{P}'$ are partitions of $[a, b]$.*

1. *Show that if $\mathcal{P}'$ is a refinement of $\mathcal{P}$ then*

$$S_{\min}(\mathcal{P}) \leq S_{\min}(\mathcal{P}') \leq S_{\max}(\mathcal{P}') \leq S_{\max}(\mathcal{P}).$$

2. *Show that $S_{\min}(\mathcal{P}) \leq S_{\max}(\mathcal{P}')$ even if neither partition is a refinement of the other. (Hint: Let $\mathcal{P}''$ be a common refinement of $\mathcal{P}$ and $\mathcal{P}'$.)*

**Theorem 228** *If $f : [a, b] \to \mathbb{R}$ is continuous then $f$ is Riemann integrable.*

**Proof.** Using the notation prior to the statement of this theorem, let $I := \lim r_i$ and fix $\varepsilon > 0$. First, there exists an $N$ such that $I \leq r_N < I + \varepsilon$. According to Proposition 93, $f$ is uniformly continuous; there exists a $\delta > 0$ such that $\delta < \frac{1}{N}$ and if $|x - y| < \delta$ then $|f(x) - f(y)| < \frac{\varepsilon}{3(b-a)}$. Let $\mathcal{P}$ be a partition such that $\sigma(\mathcal{P}) < \delta < \frac{1}{N}$. By definition of $r_N$,

$$S_{\max}(\mathcal{P}) \leq r_N < I + \varepsilon.$$

According to Exercise 153 we will be finished if we can now show

$$S_{\min}(\mathcal{P}) > I - \varepsilon.$$

Next

$$S_{\max}(\mathcal{P}) - S_{\min}(\mathcal{P}) \leq \sum_{j=1}^{k} \left( \max_{[x_{j-1},x_j]} f(x) - \min_{[x_{j-1},x_j]} f(x) \right) (x_j - x_{j-1})$$

$$< \frac{\varepsilon}{3(b-a)} \sum_{j=1}^{k} (x_j - x_{j-1}) = \frac{\varepsilon}{3(b-a)} (b-a) = \frac{\varepsilon}{3}. \tag{4.1}$$

(Here we have used the fact that $\sum_{j=1}^{k}(x_j - x_{j-1})$ "telescopes" and is equal to $b - a$.) Choose any $\frac{1}{i} < \delta$. Since $r_i \geq I > I - \frac{\varepsilon}{3}$ the approximation property for the supremum implies there exists some $\mathcal{P}'$ such that $\sigma(\mathcal{P}') < \frac{1}{i} < \delta$ and choice of $d_j$ such that $S_{\max}(\mathcal{P}') \geq S(\mathcal{P}', \{d_j\}) > I - \frac{\varepsilon}{3}$. Applying Exercise 153 and Formula 4.1 twice we have

$$S_{\min}(\mathcal{P}) \geq S_{\max}(\mathcal{P}) - \frac{\varepsilon}{3} \geq S_{\min}(\mathcal{P}') - \frac{\varepsilon}{3} \geq S_{\max}(\mathcal{P}') - \frac{2\varepsilon}{3} > I - \varepsilon.$$

■

**Proposition 229** *Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable on $[a, b]$ and $g : [a, b] \to \mathbb{R}$ be a function such that $f(x) = g(x)$ for all $x$ except for a single point $y_0 \in [a, b]$. Then $g$ is Riemann integrable on $[a, b]$ and $\int_a^b f = \int_a^b g$.*

**Proof.** Let $\varepsilon > 0$ and $I := \int_a^b f$. Then there exists a $\delta > 0$ such that $\delta < \frac{\varepsilon}{3(|f(y_0) - g(y_0)|)}$ and if a partition $\mathcal{P}$ satisfies $\sigma(\mathcal{P}) < \delta$ then for any choice of $\{c_j\}$ we have $\left| I - \sum_{i=1}^{k} f(c_j)(x_j - x_{j-1}) \right| < \frac{\varepsilon}{3}$. Now consider the corresponding sum $\sum_{j=1}^{k} g(c_j)(x_j - x_{j-1})$. The two sums are identical unless we happen to have chosen $c_j = y_0$ for some $j$. There are two possible cases: we chose only a single $c_j = y_0$ or we chose $c_j = c_{j+1} = y_0$ (in this case $y_0 = x_j$). We will consider the second case; the first is an exercise. Now

$$\left| \sum_{j=1}^{k} f(c_j)(x_j - x_{j-1}) - \sum_{j=1}^{k} g(c_j)(x_j - x_{j-1}) \right| = \left| \sum_{j=1}^{k} (f(c_j) - g(c_j))(x_j - x_{j-1}) \right|$$

$$= |(f(y_0) - g(y_0))((x_j - x_{j-1}) + (x_{j+1} - x_j))| = |(f(y_0) - g(y_0))| \, |x_{j+1} - x_{j-1}|$$

$$< |f(y_0) - g(y_0)| \left( \frac{2\varepsilon}{3\,|g(y_0) - f(y_0)|} \right) = \frac{2\varepsilon}{3}.$$

By the triangle inequality, $\left| I - \sum_{j=1}^{k} g(c_j)(x_j - x_{j-1}) \right| < \varepsilon.$ ■

**Exercise 154** *Prove the remaining case in Proposition 229.*

A simple induction proof yields the following:

**Corollary 230** *If $f : [a, b] \to \mathbb{R}$ is continuous except at finitely many points then $f$ is Riemann integrable.*

**Example 231** *Let $f_\delta : [0,1] \to \mathbb{R}$ be defined by $f_\delta(x) = 1$ if $x$ is rational and $f_\delta(x) = 0$ if $x$ is irrational ($f_\delta$ is sometimes called the Dirichlet function). Certainly this function is not Riemann integrable. In fact, given any partition of $[0, 1]$, if one chooses, for each value of $c_j$ a rational number then the corresponding Riemann sum is always $1$. If one chooses $c_j$ to always be irrational than the corresponding Riemann sum is always $0$, so there can be no limit as the partitions get fine. This shows that a function with countably many discontinuities need not be Riemann integrable. More seriously, $f_\delta$ is a monotone increasing limit of a uniformly bounded sequence of Riemann integrable functions: Let $(q_n)_{n=1}^{\infty}$ be a sequence that is surjective onto the rational numbers. Define $f_i(x)$ to be $1$ if $x \in \{q_1, ..., q_i\}$ and $0$ otherwise. Then each $f_i$ has finitely many discontinuities and hence is Riemann integrable and $f_i \nearrow f_\delta$.*

**Exercise 155** *Prove that the function $f_\delta$ is not continuous at any point in $[0, 1]$.*

The above example illustrates one of the most important limitations of Riemann integration. Riemann integrable functions lack a basic completeness property, which makes them unacceptable for more advanced mathematics. Many problems in pure and applied mathematics are solved using pointwise limits of functions, and a theory of integration that behaves well with respect to such limits is essential. Also, as we will see later in our discussion of iterated integration, the proofs of some very basic theorems about Riemann integration of continuous functions require consideration of non-continuous functions. Without a more general theory these proofs must be carried out using rather *ad hoc* methods that are of little value beyond the proof at hand. In summary, continuous functions, which are central to the topological component of real analysis, do not provide a suitable setting for integration theory. Riemann integrable functions, while more general, are difficult to characterize and still behave badly when it comes to taking limits. In the next section we will define a class of functions, called Borel functions, that is more suitable for integration–in particular having the property that a pointwise limit of Borel functions is a Borel function. Specifically we will show that the function $f_\delta$ is a Borel function, the integral of which is 0. In addition, every Riemann integrable real function $f$ is a Borel function, and we will see that Riemann integration is a special case of the more general Lebesgue integral that we will define. In a sense, Lebesgue integration is a "completion" of Riemann integration. Lebsegue integration enlarges the set of continuous functions in much the way that the completion of the real numbers enlarges the rational numbers–allowing a viable theory of limits to be carried out.

Despite its serious limitations, the Riemann integral should not be completely dismissed. Riemann sums are, at least, easily computable and can provide a way to approximate an integral numerically.

## 4.2 Borel sets and functions

Lebesgue integration theory involves countable processes of the underlying sets, such as countable unions and intersections. This is an important difference from Riemann integration, in which one normally considers only finite unions and intersections. For example, there is a theorem that the integral over the union of two adjacent intervals is the sum of the integrals over the two intervals separately.

**Definition 232** *Let $X$ be a metric space. The collection of Borel sets in $X$ is the smallest collection $\mathcal{B}$ of subsets of $X$ such that*

1. *every open subset of $X$ is in $\mathcal{B}$,*
2. *if $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$*
3. *if $A_i \in \mathcal{B}$ for all $i \in \mathbb{N}$ then $\cup_{i=1}^{\infty} A_i$ is in $\mathcal{B}$.*

*Any nonempty collection of sets satisfying conditions (2) and (3) is called a $\sigma$-algebra.*

Hence a $\sigma$-algebra is a nonempty collection of sets that is closed with respect to complements and countable unions. What do we mean by the "smallest" such $\sigma$-algebra, and how do we know that such a thing exists? First of all there is some collection of subsets of $X$ with the above three properties, namely the collection $\mathcal{P}(X)$ of all subsets of $X$, and $\mathcal{B}$ is formally defined to be the intersection of all such collections. Certainly $\mathcal{B}$ contains all open sets, and an exercise below shows that $\mathcal{B}$ is in fact a $\sigma$-algebra.

**Exercise 156** *Let $\Sigma$ be a $\sigma$-algebra. Show that*

1. *If $A_i$ lies in a $\sigma$-algebra $\Sigma$ for all $i \in \mathbb{N}$ then $\cap_{i=1}^{\infty} A_i$ is in $\Sigma$.*
2. *$\Sigma$ contains the empty set and hence contains all finite unions or intersections of sets in $\Sigma$.*
3. *If $A, B \in \Sigma$ then $A \backslash B \in \Sigma$.*

**Exercise 157** *Let* $\{\Sigma_\alpha\}_{a\in\Lambda}$ *be a collection of* $\sigma$*-algebras. Show that* $\bigcap_{\alpha\in\Lambda}\Sigma_\alpha$ *is a* $\sigma$*-algebra.*

What does $\mathcal{B}$ contain? By definition $\mathcal{B}$ contains all open subsets of $X$, and therefore all closed subsets of $X$ by property (2). It also includes all countable intersections of open sets, all countable unions of closed sets, and so on. Roughly speaking, $\mathcal{B}$ contains any set that can be obtained from open sets by any countable combination of taking unions, intersections, and complements. It turns out that there are subsets of metric spaces, including $\mathbb{R}$, that are not Borel sets, but their construction is not trivial (see, for example, [3], Theorem 3.38).

**Example 233** *In any metric space* $X$*, we know that singleton sets* $\{x\}$ *are closed, hence Borel. Therefore any countable subset of a metric space is Borel, in particular* $\mathbb{Q}$ *is a Borel subset of* $\mathbb{R}$*. On the other hand, this shows that, unlike the case with open sets, uncountable unions of Borel sets need not be Borel sets. In fact any non-Borel subset of* $\mathbb{R}$ *is the uncountable union of the singleton sets containing each of its points and therefore is an uncountable union of Borel sets.*

**Definition 234** *A function* $f : A \to Y$*, where* $X$ *and* $Y$ *are metric spaces and* $A \subset X$ *is a Borel set, is called Borel if for every open set* $U$ *in* $Y$*,* $f^{-1}(U)$ *is a Borel set.*

Recall that $f$ is continuous if and only if for every open set $U$ in $Y$, $f^{-1}(U)$ is open; therefore every continuous function is a Borel function. Borel functions are much more general, however:

**Proposition 235** *Let* $X$ *be a metric space and* $X := \cup_{i=1}^{\infty} E_i$*, where each* $E_i$ *is a Borel set. Suppose* $f : X \to Y$ *is a function such that* $f\mid_{E_i}$ *is Borel for all* $i$*. Then* $f$ *is a Borel function.*

**Proof.** Let $U \subset Y$ be open. Then for any $i$, $f^{-1}(U) \cap E_i = (f\mid_{E_i})^{-1}(U)$ is Borel and hence $f^{-1}(U) = \bigcup_{i=1}^{\infty} f^{-1}(U) \cap E_i)$ is Borel. ■

Note that the above two results also apply to finite unions–if $X = E_1 \cup \cdots \cup E_k$ then we can always let $E_i := E_k$ for all $i > k$ to express $X$ as a countable union $\cup_{i=1}^{\infty} E_i$. Note also that the countability of the union is essential. In fact, $\mathbb{R}$ can be written as an uncountable union of singleton sets, and any function defined on $\mathbb{R}$ is constant, hence continuous, when restricted to a singleton set. Therefore if the above corollary were true for uncountable unions, *every* real function would be a Borel function. Although the proof is beyond the scope of this text, there do exist real functions that are not Borel functions (see [3]).

**Definition 236** *If* $E$ *is a subset of a set* $X$ *we define the characteristic function* $\chi_E : X \to \mathbb{R}$ *by* $\chi_E(x) = 1$ *if* $x \in E$ *and* $\chi_E(x) = 0$ *if* $x \notin E$*.*

The characteristic function, simple as it is, is of considerable importance in integration theory–being, in a sense, the simplest kind of non-trivial function. Note that the characteristic function is rarely continuous–for example, $\chi_{[0,1]} : \mathbb{R} \to \mathbb{R}$ has discontinuities at 0 and 1. However, if $E$ is a Borel set then $X \backslash E$ is a Borel set, and $\chi_E$ is constant, hence continuous, when restricted to $E$ and when restricted to $X \backslash E$. Proposition 235 therefore implies:

**Corollary 237** *If $E$ is a Borel set in a metric space $X$ then $\chi_E : X \to \mathbb{R}$ is a Borel function.*

**Example 238** *In our current notation we now see that $f_\delta = \chi_{\mathbb{Q} \cap [0,1]}$ (see Example 231) and therefore $f_\delta$ is a Borel function that is not continuous at any point.*

**Exercise 158** *Prove that if $f : X \to \mathbb{R}$ is a function defined on a metric space $X$ and $f$ is continuous on $X \backslash A$ where $A$ is countable, then $f$ is Borel.*

**Exercise 159** *Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = 0$ if $x$ is irrational and $g(x) = \frac{1}{q}$ if $x$ is rational and expressed as $x = \frac{p}{q}$ in reduced form (take the reduced form of an integer to have 1 in the denominator).*

1. *Prove that $g$ is continuous at every irrational number and hence is Borel.*

2. *Sketch a graph of $g$ on $[0, 1]$ by indicating the graph above several rational points.*

Recall that the pointwise limit of continuous functions need not be continuous. As we will see shortly, real-valued Borel functions are much more robust. First we need a preliminary lemma.

**Lemma 239** *Let $f : X \to \mathbb{R}$ be a function where $X$ is a metric space. The following are equivalent:*

1. *$f$ is Borel,*

2. *for any closed set $C \subset \mathbb{R}$, $f^{-1}(C)$ is Borel,*

3. *for any closed interval $[a, b] \subset \mathbb{R}$, $f^{-1}([a, b])$ is Borel,*

4. *for any open interval $(a, b) \subset \mathbb{R}$, $f^{-1}((a, b))$ is Borel.*

**Exercise 160** *Prove the above lemma. Hint: Show (1)⇒(2)⇒(3)⇒(4)⇒(1), using Proposition 129 for the last part.*

**Theorem 240** *Let $f_i : X \to \mathbb{R}$ be a sequence of Borel functions defined on a metric space $X$ such that $f_i \xrightarrow{p} f$. Then $f$ is Borel.*

**Proof.** This remarkable theorem is a consequence of some tedious set theory. We will show that for any closed interval $[a,b]$, $f^{-1}([a,b])$ is Borel. For every $j \in \mathbb{N}$, define $U_j := (a - \frac{1}{j}, b + \frac{1}{j})$. Now let $W_{i,j} := f_i^{-1}(U_j)$. Since each $f_i$ is Borel, every $W_{i,j}$ is Borel, as is the set

$$V := \bigcap_{j=1}^{\infty} \left( \bigcup_{N=1}^{\infty} \left( \bigcap_{i=N}^{\infty} W_{i,j} \right) \right).$$

We will be finished if we show that $V = f^{-1}([a,b])$. A quick check of the definitions reveals that $x \in V$ if and only if for every $j \in \mathbb{N}$ we have that

$$a - \frac{1}{j} < f_i(x) < b + \frac{1}{j}$$

for all large $i$. Taking the limit as $i \to \infty$ gives us $f(x) \in \left[a - \frac{1}{j}, b + \frac{1}{j}\right]$ for every $j$, and hence $f(x) \in [a,b]$. Conversely, if $f(x) \in [a,b]$ then for any $j$ we have $|f_i(x) - f(x)| < \frac{1}{j}$ for all large $i$ and hence $f_i(x) \in \left(a - \frac{1}{j}, b + \frac{1}{j}\right)$ for all large $i$. ■

The above theorem shows that a function that is not Borel must be somewhat difficult to construct–after all, it cannot be constructed using any kind of limiting process involving continuous or even Borel functions. At the same the theorem provides perhaps the best way to show that a function is a Borel function, namely to show that it is a pointwise limit of Borel functions. Finally, this theorem also makes Borel functions an excellent (if not the most general) setting for integration theory. We need to know some other ways in which one obtains Borel functions.

**Lemma 241** *If $f : X \to Y$ is a Borel function between metric spaces and $A \subset X$ is a Borel set then $f\mid_A$ is a Borel function.*

**Lemma 242** *If $X, Y, Z$ are metric spaces, $f : Y \to Z$ is continuous and $g : X \to Y$ is Borel then $f \circ g$ is Borel.*

**Exercise 161** *Prove Lemmas 241 and 242.*

**Lemma 243** *If $X$ is a metric space and $h : X \to \mathbb{R}^2$ is a function having Borel components then $h$ is Borel.*

**Proof.** Let $h_1 : X \to \mathbb{R}$ and $h_2 : X \to \mathbb{R}$ be the components of $h$. Suppose first that $U$ and $V$ are open intervals in $\mathbb{R}$. Then

$$h^{-1}(U \times V) = \{x \in X : h_1(x) \in U \text{ and } h_2(x) \in V\}$$
$$= h_1^{-1}(U) \cap h_2^{-1}(V)$$

which is a Borel set since $h_1$ and $h_2$ are Borel functions. Now according to Proposition 129, every open set in $\mathbb{R}^2$ is of the form $W = \cup_{i=1}^{\infty} B_{\max}(\mathbf{x}_i, r_i)$ and each $B_{\max}(\mathbf{x}_i, r_i)$ is of the form $I_i \times J_i$, where $I_i$ and $J_i$ are open intervals (see Lemma 119). From what we proved above we have

$$h^{-1}(W) = \bigcup_{i=1}^{\infty} h^{-1}(I_i \times J_i)$$

is a Borel set. ∎

**Proposition 244** *Let $f, g : X \to \mathbb{R}$ be Borel functions defined on a metric space $X$ and $c$ be any real number. Then the functions $f + g$, $f \cdot g$, $cf$, and $f/g$ (where it is defined) are Borel functions.*

**Proof.** Let $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be the "sum" function defined by $h(x, y) = x + y$. Then $h$ is continuous. In fact, if $(x_i, y_i) \to (x, y)$ then by earlier sequence theorems we have $h(x_i, y_i) = x_i + y_i \to x + y = h(x, y)$. If $k : X \to \mathbb{R} \times \mathbb{R}$ denotes the function having $f$ and $g$ as components then

$$(h \circ k)(x) = h(k(x)) = h(f(x), g(x)) = f(x) + g(x) = (f + g)(x).$$

That is, we have written $f + g$ as a composition of two functions, where $k$ is Borel by Lemma 243 and $h$ is continuous. By Lemma 242, $f + g$ is Borel. The proofs for the other algebraic operations are similar. We only observe that in the case of the quotient, the domain of definition of $f/g$ is $g^{-1}(\mathbb{R}\backslash\{0\})$, which is a Borel set. ∎

## 4.3 Integration of Nonnegative Borel Functions

We will take the existence of the Lebesgue integral on nonnegative Borel functions without proof. In a way this is very similar to our acceptance of the real numbers without actually constructing them. However, we will do a little more in this case. In the course of our discussion we will see how Lebesgue integration can be defined; we will simply not provide the details of the proofs of the properties that we will assume it has in Theorem 247. The arguments involved are primarily set theoretic in nature and are valid in much greater generality in the setting of abstract measure theory, which is normally covered in a first year graduate course in real analysis. We will not even study Lebesgue integration in its most general setting in Euclidean spaces. This would involve introducing the somewhat abstract notions of measurable sets and functions–which again appear in the more general setting of abstract measure theory. For our purposes Lebesgue integration involving Borel sets and functions is adequate.

We will need some preliminary notation.

**Definition 245** *If $A \subset \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$ then we define $\mathbf{v} + A := \{\mathbf{v} + x : x \in A\}$, and refer to $\mathbf{v} + A$ as the translation (or translate) of $A$ by $\mathbf{v}$.*

**Definition 246** *Let $A$ be a set and suppose $\{E_i\}$ is a collection of subsets of $A$. We say that $\{E_i\}$ is pairwise disjoint if whenever $i \neq j$ we have $E_i \cap E_j = \varnothing$.*

We will use below some properties of the closed max metric balls in $\mathbb{R}^n$, which we will refer to as "cubes." The term "cube" will always refer to closed max metric balls (which have edges parallel to the standard axes) and not other geometric cubes (such as the closed balls of the plus metric). Since the max metric is invariant, every cube of a given radius is a translate of any other such cube, including the one centered at $\mathbf{0}$. We will refer to the cube centered at $\mathbf{0}$ in $\mathbb{R}^n$ with radius $r/2$ as $Q^n(r)$ (so this cube has side length $r$), and we will refer to $Q^n(1)$ simply as $Q^n$, and call it the unit cube. Every cube in $\mathbb{R}^n$ also has $2n$ "faces" that come in pairs, each of which is isometric to the cube $Q^{n-1}(r)$. For example, the cube $Q^n(2)$ has two faces perpendicular to the $j^{th}$ axis, namely the sets

$$F_+^j := \{\mathbf{x} = (x_1, ..., x_n) : d_{\max}(\mathbf{0}, \mathbf{x}) = 1 \text{ and } x_j = 1\}$$

and

$$F_-^j := \{\mathbf{x} = (x_1, ..., x_n) : d_{\max}(\mathbf{0}, \mathbf{x}) = 1 \text{ and } x_j = -1\}.$$

The union of the $2n$ faces of a cube of side length $r$ is called the *boundary* of the cube; it is the set of all points at max distance exactly $\frac{r}{2}$ from the center of the cube. Next any cube of side length $r$ can be subdivided into $m^n$ cubes having edge length $\frac{r}{m}$, for any natural number $m$, such that different cubes intersect only their faces. We won't give any details here, which are a bit tedious; the student is encouraged to convince him/herself in lower dimensions with a good picture. Some special cases are considered in the following exercise. Finally, define a *semicube* to be a set of the form $Q \backslash A$ where $Q$ is a cube and $A$ is a compact subset of its boundary.

**Exercise 162** *Sketch the subdivision of $Q^3$ in $\mathbb{R}^3$ into 8 equal sized cubes, and determine for each of them the vector used to translate $Q^3(\frac{1}{2})$ to it. Determine the vector that translates $Q^4(\frac{1}{2})$ to the subdivision cube of $Q^4$ that has $(1,1,1,1)$ as one of its corners. Explicit details are not required in either case.*

**Exercise 163** *Let $B_{jk}(r) := \{(x_1, ..., x_n) \in \mathbb{R}^n : 0 \leq x_i \leq r$ if $i \neq j$ and $0 \leq x_j \leq \frac{1}{k}\}$.*

1. Sketch $B_{24}(3)$ in $\mathbb{R}^2$.
2. Prove that $Q^n(r + \frac{4}{k}) \backslash Q^n(r)$ contains a translate of $B_{jk}(r)$ for all $j$ and $r > 0$.

**Theorem 247** *For each $n \in \mathbb{N}$ there exists a unique function that assigns to each Borel set $A$ in $\mathbb{R}^n$ and nonnegative Borel function $f : A \to \mathbb{R}$ a nonnegative extended real number $\int_A f$ satisfying the following properties for any $f, g : A \to \mathbb{R}$, $c \geq 0$ and $\mathbf{v} \in \mathbb{R}^n$:*

1. *(Positivity) $\int_A f \geq 0$.*
2. *(Linearity) $\int_A (cf + g) = c \int_A f + \int_A g$ (assuming $\int_A f < \infty$ or $c \neq 0$)*
3. *(Countable Set Additivity) If $E = \bigcup_{i=1}^{\infty} E_i$ where $\{E_i\}_{i=1}^{\infty}$ is a pairwise disjoint collection of Borel sets, then $\int_E f = \sum_{i=1}^{\infty} \int_{E_i} f$.*

4. *(Translation Invariance) If $h(\mathbf{x}) = f(\mathbf{x} - \mathbf{v})$, which is defined on $\mathbf{v} + A$, then $\int_A f = \int_{\mathbf{v}+A} h$.*

5. *(Normalization) $\int_{Q^n} 1 = 1$.*

The first two properties are familiar from elementary calculus of the Riemann integral. The "finite" version of the third property is also familiar from calculus–although in calculus one actually considers mainly sets that intersect along some kind of boundary, as in the statement that $\int_a^c f(x)dx = \int_a^b f(x)dx = \int_b^c f(x)dx$ when $a \le b \le c$. In this case the intervals $[a, b]$ and $[b, c]$ intersect in $\{b\}$, but as we will see later, a single point, by itself, contributes nothing to an integral. More important is the fact that the third property is stated not just for two sets, or finitely many sets, but for countably many sets. Note that, as a result of Proposition 195, the order of the sum is not important. The reader may be tempted to believe that condition (3) follows by induction from the case when there are only two sets, but in fact it does not. This situation is explored further in Exercise 166 below.

The fourth property appears in elementary calculus as a special case of change of variables. The fifth condition is necessary for uniqueness and non-triviality. For if one has an integral satisfying only the first four properties, one may multiply it by any non-negative constant (including 0!) and still have an integral satisfying those four properties. Note that the usual conventions involving infinity apply to (2) and (3) (see Definition 10).

We need to establish some very basic properties of the integral.

**Lemma 248** *Let $f, g : A \to \mathbb{R}$ be nonnegative Borel functions, $A \subset \mathbb{R}^n$. If $f(x) \le g(x)$ for all $x \in A$ then $\int_A f \le \int_A g$.*

**Proof.** Let $k(x) := g(x) - f(x) \ge 0$. Then $g(x) = f(x) + k(x)$ and $\int_A k \ge 0$ imply

$$\int_A g = \int_A f + \int_A k \ge \int_A f.$$

■

Our next goal is to understand the integral of a function over the empty set. As annoying as this may seem at first, it is a necessary task. Mainly we need to first understand exactly what a function defined on the empty set is. Recall that a function $f : A \to B$ is a *set* of ordered pairs $(a, b)$ with $a \in A$ and $b \in B$. In particular, if $A = \varnothing$ then the function $f$ is actually the empty set. This is true regardless of what $f$ is called or where $f$ is defined. Therefore, if we show $\int_\varnothing 1 = 0$ then we will know that $\int_\varnothing f = 0$ for any Borel function $f$. But $Q^n$ is the countable pairwise disjoint union of $Q^n \cup \varnothing \cup \varnothing \cup \cdots$ and therefore by countable set additivity

$$\int_{Q^n} 1 = \int_{Q^n} 1 + \sum_{i=1}^{\infty} \int_\varnothing 1.$$

Since $\int_{Q^n} 1 = 1$ we conclude that $\sum_{i=1}^{\infty} \int_\varnothing 1 = 0$, and hence $\int_\varnothing 1 = 0$. We have proved:

**Lemma 249** *If $f : A \to \mathbb{R}$ is a nonnegative Borel function with $A \subset \mathbb{R}^n$ then $\int_{\varnothing} f = 0$.*

If we now let all but finitely many of the sets in Theorem 247 part (3) be the empty set then we conclude from the above lemma:

**Lemma 250** *(Finite Additivity) If $E_1, ..., E_n \subset \mathbb{R}^n$ are pairwise disjoint Borel sets, $E = \bigcup_{i=1}^k E_i$, and $f : E \to \mathbb{R}$ is a nonnegative Borel function, then $\int_E f = \sum_{i=1}^k \int_{E_i} f$.*

We will repeatedly use the following very special case of Finite Additivity: For any Borel sets $E$ and $A$ in $\mathbb{R}^n$ we have that $E = E \cap A \cup E \backslash A$, and this union is a disjoint union. Therefore we have for any nonnegative Borel function $f : E \to \mathbb{R}$,

$$\int_E f = \int_{E \cap A} f + \int_{E \backslash A} f.$$

If it so happens that $A \subset E$ then the above expression becomes $\int_E f = \int_A f + \int_{E \backslash A} f$.

**Lemma 251** *If $f : E \to \mathbb{R}$ is a nonnegative Borel function where $F \subset E \subset \mathbb{R}^n$ are Borel sets then $\int_F f \leq \int_E f$.*

**Proof.** We have $\int_E f = \int_F f + \int_{E \backslash F} f$. Noting that $\int_{E \backslash F} f \geq 0$ finishes the proof. ■

**Exercise 164** *Suppose $f, g : A \to \mathbb{R}$ are nonnegative Borel functions, $A \subset \mathbb{R}^n$, and $f(x) \leq g(x)$ for all $x \in A$. Prove: If $\int_A f < \infty$ then $\int_A (g - f) = \int_A g - \int_A f$.*

**Lemma 252** *(Countable Set Subadditivity) If $E \subset \mathbb{R}^n$ is a Borel set and $E \subset \bigcup_{i=1}^\infty E_i$ where $\{E_i\}_{i=1}^\infty$ is a collection of Borel sets, then $\int_E f \leq \sum_{i=1}^\infty \int_{E_i} f$ for any nonnegative Borel function $f : \bigcup_{i=1}^\infty E_i \to \mathbb{R}$.*

**Proof.** Let $E' := \bigcup_{i=1}^\infty E_i$ and let $F_1 := E_1$, $F_2 := E_2 \backslash E_1$, $F_3 := E_3 \backslash (E_1 \cup E_2)$, and so on. It is an exercise in set theory to see that $E'$ is the pairwise disjoint union of the sets $\{F_i\}_{i=1}^\infty$, and $F_i \subset E_i$ for all $i$. Therefore

$$\sum_{i=1}^\infty \int_{E_i} f \geq \sum_{i=1}^\infty \int_{F_i} f = \int_{E'} f \geq \int_E f.$$

■

**Corollary 253** *If $f(x) = 0$ on a Borel set $A \subset \mathbb{R}^n$ then $\int_A f = 0$.*

**Proof.** We first compute

$$\int_{Q^n} 0 = \int_{Q^n} 0 \cdot 1 = 0 \int_{Q^n} 1 = 0 \cdot 1 = 0.$$

As in the proof of Proposition 129, $\mathbb{R}^n$ is the union of the countably many unit cubes $Q_i$ having centers with rational coordinates. By Translation Invariance $\int_{Q_i} 0 = 0$ and countable subadditivity implies

$$\int_A 0 \le \sum_{i=1}^{\infty} \int_{Q_i} 0 = \sum_{i=1}^{\infty} 0 = 0.$$

■

**Corollary 254** *If $f : E \to \mathbb{R}$ is a nonnegative Borel function and $A \subset E \subset \mathbb{R}^n$ are Borel sets then*

$$\int_A f = \int_E \chi_A \cdot f.$$

**Proof.** By definition,

$$\chi_A \cdot f(x) = \begin{cases} f(x) \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases},$$

and

$$\int_E \chi_A \cdot f = \int_A \chi_A \cdot f + \int_{E \backslash A} \chi_A \cdot f = \int_A f + \int_{E \backslash A} 0 = \int_A f.$$

■

**Exercise 165** *In the proof of the Countable Set Subadditivity lemma, show that $E'$ is the pairwise disjoint union of the sets $\{F_i\}_{i=1}^{\infty}$.*

**Exercise 166** *Prove that the assumption of countable set additivity in Theorem 247 is equivalent to assuming both countable set subadditivity and finite set additivity. Hint: Use Lemma 251 and Lemma 187.*

**Definition 255** *We say that a collection of sets $\{E_i\}_{i=1}^{\infty}$ is increasing (resp. decreasing) if $E_1 \subset E_2 \subset \cdots$ (resp. $E_1 \supset E_2 \supset \cdots$). In this case we say that $E = \bigcup_{i=1}^{\infty} E_i$ is the increasing union (resp. $E = \bigcap_{i=1}^{\infty} E_i$ is the decreasing intersection) of $\{E_i\}_{i=1}^{\infty}$. We will use the notation $E_i \nearrow E$ (resp. $E_i \searrow E$).*

**Proposition 256** *If $\{E_i\}_{i=1}^{\infty}$ is a collection of Borel sets in $\mathbb{R}^n$ such that $E_i \nearrow E$ and $f : E \to \mathbb{R}$ is a nonnegative Borel function then $\int_E f = \lim \int_{E_i} f$.*

**Proof.** Note that $E_1 \subset E_2 \subset \cdots$ implies the sequence $\left(\int_{E_i} f\right)_{i=1}^{\infty}$ is monotone increasing. Let $F_1 := E_1$, $F_2 := E_2 \backslash E_1$, and so on. Then the sets $F_i$ are pairwise disjoint and $E_k = \bigcup_{i=1}^{k} F_i$ and $E = \bigcup_{i=1}^{\infty} F_i$. We have

$$\int_E f = \sum_{i=1}^{\infty} \int_{F_i} f = \lim \sum_{i=1}^{k} \int_{F_i} f = \lim \int_{E_k} f.$$

■

**Corollary 257** *If $\{E_i\}_{i=1}^{\infty}$ is a collection of Borel sets in $\mathbb{R}^n$ such that $E_i \searrow E$ and $f : E_1 \to \mathbb{R}$ is a nonnegative Borel function with $\int_{E_1} f < \infty$ then $\int_E f = \lim \int_{E_i} f$.*

**Exercise 167** *Prove Corollary 257, stating explicitly how you use the fact that $\int_{E_1} f < \infty$. Hint: Consider the sets $F_i := E_1 \backslash E_i$ and use de Morgan's laws to apply Proposition 256.*

**Example 258** *The interval $E_n := [n, \infty)$ contains countably many pairwise disjoint intervals $I_m := [2m, 2m+1]$, each of which is a translate of the unit interval and hence satisfies $\int_{I_m} 1 = 1$. Therefore $\int_{E_n} 1 = \infty$ for all $n$. Then the collection $\{E_n\}_{n=1}^{\infty}$ is decreasing and $\bigcap_{n=1}^{\infty} E_n = \varnothing$, but $\int_{\varnothing} 1 = 0 \neq \infty = \lim \int_{E_n} 1$. This shows the necessity of the requirement that $\int_{E_1} f < \infty$ in Corollary 257 (actually, assuming that some $\int_{E_i} f < \infty$ will do).*

## 4.4 Lebesgue Measure

In elementary calculus one defines the area of a two dimensional region $A$ to be the double integral over $A$ of the function $f(x) = 1$ (of course the regions $A$ were always very simple so that one could actually compute the integral by iterated integration). We will do precisely the same thing now, although, to avoid the dimensional connotations of words like "area" and "volume" we will use the term "measure" for all dimensions.

**Definition 259** *Let $E$ be a Borel subset of $\mathbb{R}^n$. We define the (Lebesgue) measure of $E$ to be*

$$\mu(E) := \int_{\mathbb{R}^n} \chi_E = \int_E 1.$$

From the properties for the Lebesgue integral we immediately obtain the following:

**Theorem 260** *Lebesgue measure in $\mathbb{R}^n$ satisfies the following properties:*

1. *(Positivity) For any Borel set $E$, $\mu(E) \geq 0$.*
2. *(Countable Set Additivity) If $E = \bigcup_{i=1}^{\infty} E_i$ where $\{E_i\}_{i=1}^{\infty}$ is a pairwise disjoint collection of Borel sets, then $\mu(E) = \sum_{i=1}^{\infty} \mu(E_i)$.*
3. *(Translation Invariance) For any Borel set $E$ and $\mathbf{x} \in \mathbb{R}^n$, $\mu(E) = \mu(\mathbf{x} + E)$.*
4. *(Normalization) $\mu(Q^n) = 1$.*

Note that translation invariance implies that every cube of side length 1 in $\mathbb{R}^n$ has measure 1. Other consequences of our theorems about integration include: $\mu(\varnothing) = 0$, finite set additivity, countable set subadditivity, and the fact that if $E \subset F$ then $\mu(E) \leq \mu(F)$. In addition, if $f(x) = c > 0$ then for any Borel set $E$, $\int_E f = c\mu(E)$ (by linearity). The next lemma now follows from Lemma 248.

**Lemma 261** *Suppose that for some positive $c \in \mathbb{R}$, $c \leq f(x)$ (resp. $f(x) \geq c$) for all $x$ in a Borel set $E \subset \mathbb{R}^n$, where $f : E \to \mathbb{R}$ is a Borel function. Then*

$$c\mu(E) \leq \int_E f \ (resp.\ \leq \int_E f \leq c\mu(E)).$$

This simple but important lemma will be used repeatedly in the following way. Consider, for example, the set

$$E = f^{-1}([a,b]) = \{x : a \leq f(x) \leq b\} \text{ with } a > 0.$$

The above lemma implies

$$a\mu(E) \leq \int_E f \leq b\mu(E).$$

It will be very useful to divide the range of a nonnegative Borel function into intervals on which $f$ has specific bounds, above, below, or both. Sometimes it is beneficial to actually partition the range–for example later we will use the partition of $\mathbb{R}$ into the pairwise disjoint collection $\{[n, n+1)\}_{n=0}^{\infty}$; if $f : A \to \mathbb{R}$ is any function then one can easily check that $A$ is the pairwise disjoint union of the collection $\left\{f^{-1}([n, n+1))\right\}_{n=0}^{\infty}$. In other cases, such as the proof of the next lemma, we will use countable subadditivity and do not need a pairwise disjoint collection.

**Proposition 262** *Let $f : A \to \mathbb{R}$ be a nonnegative Borel function defined on a Borel set $A$ such that $\mu(A) = 0$. Then $\int_A f = 0$.*

**Proof.** If $f$ were bounded then the proof would be complete by Lemma 261, but $f$ may not be bounded. To resolve this problem, let $A_n := f^{-1}([0,n])$. Then $f(x) \leq n$ for all $x \in A_n$. In addition, $A_n \subset A$ and therefore $\mu(A_n) = 0$. Note that $A_n \nearrow A$ and $\int_{A_n} f \leq n \cdot \mu(A_n) = 0$. Therefore $\int_A f = \lim \int_{A_n} f = 0$. ■

**Exercise 168** *Let $f : E \to \mathbb{R}$ be a nonnegative Borel function, defined on $E \subset \mathbb{R}^n$ and $\alpha$ be any positive real number.*

1. *Prove the Tschebyshev Inequality:*

$$\mu(f^{-1}((\alpha, \infty))) \leq \frac{1}{\alpha} \int_E f.$$

   *Hint: $f^{-1}((\alpha, \infty)) = \{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}$.*

2. *Illustrate this inequality in the case of a continuous real function with a sketch of a graph, interpreting the integral as area under the graph.*

**Exercise 169** *Let $(f_i)$ be an increasing sequence of Borel functions $f_i : E \to \mathbb{R}$, with $E \subset \mathbb{R}^n$, and suppose that there exists some $M > 0$ such that for all $i$, $\int_E f_i \leq M$. Let $F$ be the set of all $\mathbf{x} \in E$ such that the sequence $(f_i(\mathbf{x}))$ is unbounded. Prove that $\mu(F) = 0$. Hint: For any fixed $N$, consider the measure of the union of the sets $F_k := f_k^{-1}((N, \infty))$.*

**Exercise 170** *A collection $\{E_i\}$ (finite or countably infinite) of Borel sets is called nonoverlapping if for every $i \neq j$, $\mu(E_i \cap E_j) = 0$. Prove that if $E = \bigcup_{i=1}^{\infty} E_i$ where $\{E_i\}_{i=1}^{\infty}$ is a nonoverlapping collection of Borel sets, then for any nonnegative Borel function $f : E \to \mathbb{R}$, $\int_E f = \sum_{i=1}^{\infty} \int_{E_i} f$. Hint: Let $F_i := E_i \backslash \bigcup_{j<i}(E_i \cap E_j)$.*

A direct translation of Proposition 256 and Corollary 257 provides:

**Proposition 263** *If $E$ is the increasing union (resp. decreasing intersection with $\mu(E_1) < \infty$) of a collection $\{E_i\}_{i=1}^{\infty}$ of Borel subsets of $\mathbb{R}^n$ then $\mu(E) = \lim \mu(E_i)$.*

Our next goal is to prove the expected formula that $\mu(Q^n(r)) = r^n$. This is harder than it may seem at first. Remember: at the moment we have absolutely no knowledge of how to actually compute a Lebesgue integral (in fact this formula is the first stage in understanding how to compute them). We only know how to compute the measure of a unit cube, but remarkably this, together with the properties we have proved so far, is enough. But it will take a few steps. First of all we need to check that cubes have finite measure. In fact, by Translation Invariance we need only prove that $Q^n(r)$ has finite measure for all $r$. By the Archimedean Principle we need only check that $\mu(Q^n(m)) < \infty$ for any $m \in \mathbb{N}$. But $Q^n(m)$ is a union of $m^n$ unit cubes and therefore $\mu(Q^n(m)) \leq m^n$.

The next step is to prove that the faces of a cube have measure 0; this is proved by showing that each face is contained in a decreasing sequence of "boxes" the measures of which go to 0.

**Lemma 264** *Let*

$$B_{jk}(r) := \{(x_1, ..., x_n) \in \mathbb{R}^n : 0 \leq x_i \leq r \text{ if } i \neq j \text{ and } 0 \leq x_j \leq \frac{1}{k}\}.$$

*Then for any $j$ and $r > 0$, $\mu(B_{jk}(r)) \to 0$.*

**Proof.** Note that $Q^n(r)$ is the decreasing intersection of the cubes $Q^n(r+\frac{1}{i})$ and (since cubes have finite measure) $\lim \mu(Q^n(r + \frac{1}{i})) = \mu(Q^n(r))$. But finite additivity implies

$$\lim \mu(Q^n(r + \frac{1}{i}) \backslash Q^n(r)) = 0.$$

According to Exercise 163 each of the sets $Q^n(r + \frac{4}{k}) \backslash Q^n(r)$ contains a translate of $B_{jk}(r)$ for all $j$ and $k$ and hence $\mu(B_{jk}(r)) \to 0$. ■

**Corollary 265** *For every $n$ and $r > 0$, $\mu(Q^n(r) \backslash B(\mathbf{0}, \frac{r}{2})) = 0$. In particular the faces of $Q^n(r)$ have measure 0 and the measure of any semicube $Q \backslash A$ is equal to the measure of $Q$.*

**Proof.** Each of the faces of $Q^n(r)$ is the intersection of translates of sets $B_{jk}(r)$ for some fixed $j$, and therefore has measure 0. Therefore the union of all the faces has measure 0, and so $A$ has measure 0. By finite additivity $\mu(Q) = \mu(Q \backslash A)$. ■

**Lemma 266** *For any $r > 0$ and natural numbers $n$ and $m$,*

$$\mu(Q^n(mr)) = m^n \mu(Q^n(r)).$$

**Proof.** The cube $Q^n(rm)$ can be subdivided into $m^n$ cubes that are translates of $Q^n(r)$ and therefore each have measure $\mu(Q^n(r))$. These cubes meet only in faces, which have measure 0 by Corollary 265, and the proof is finished by finite set additivity. ■

**Proposition 267** *For any $r > 0$, $\mu(Q^n(r)) = r^n$.*

**Proof.** First suppose that $r = \frac{p}{q}$, where $p, q \in \mathbb{N}$. Applying Lemma 266 a couple of times yields

$$q^n \mu\left(Q^n(\frac{p}{q})\right) = \mu\left(Q^n(p)\right) = p^n \mu\left(Q(1)\right) = p^n$$

which implies

$$\mu\left(Q^n(\frac{p}{q})\right) = \frac{p^n}{q^n} = r^n.$$

Finally, if $r > 0$ is real, let $r_i \searrow r$, where each $r_i \in \mathbb{Q}$. Then $Q^n(r)$ is the decreasing intersection of the cubes $Q^n(r_i)$ and therefore

$$\mu(Q^n(r)) = \lim \mu(Q^n(r_i)) = \lim r_i^n = r^n.$$

■

**Corollary 268** *For any $\mathbf{x} \in \mathbb{R}^n$, $\mu(\{\mathbf{x}\}) = 0$.*

**Proof.** The set $\{\mathbf{0}\}$ the decreasing intersection of the cubes $Q^n(\frac{1}{i})$, and in general $\{\mathbf{x}\} = \mathbf{x} + \{\mathbf{0}\}$. ■

**Corollary 269** *Any countable subset of $\mathbb{R}^n$ is of measure 0.*

**Corollary 270** $\int_{[0,1]} f_\delta = 0$.

**Proof.**

$$\int_{[0,1]} f_\delta = \int_{[0,1]} \chi_{\mathbb{Q} \cap [0,1]} = \mu(\mathbb{Q} \cap [0,1]) = 0.$$

■

**Exercise 171** *Show that the y-axis (and hence any vertical line) in $\mathbb{R}^2$ has measure 0.*

**Example 271** *Although we have shown that the measure of the closure of a max metric ball is the same as the measure of the ball, it is not in general true that the measure of the closure of a set is equal to the measure of the set. For example, we know that $[0, 1] \cap \mathbb{Q}$ is countable and hence has measure 0, but the closure of $[0, 1] \cap \mathbb{Q}$ is $[0, 1]$, which has measure 1.*

**Lemma 272** *If $f : [a, b] \to \mathbb{R}$ is a continuous nonnegative function and $\int_{[a,b]} =$ 0 then $f(x) = 0$ for all $x \in [a, b]$.*

**Proof.** We will prove the contrapositive statement. If $f(x_0) = \delta > 0$ for some $x_0 \in (a, b)$ then since $f$ is continuous $f(x) > \frac{\delta}{2}$ on some interval $(x_0 - \varepsilon, x_0 + \varepsilon) \subset (a, b)$, where $\varepsilon > 0$. But then $\int_{[a,b]} f \geq \int_{(x_0-\varepsilon,x_0+\varepsilon)} f \geq \frac{\delta}{2}(2\varepsilon) = \delta\varepsilon > 0$. The proof if $x_0$ is an endpoint is similar. ■

We already know that Lemma 272 is not true if we simply replace continuous by Borel (e.g. the function $f_\delta$, which has integral 0 but is positive at countably many points). But something almost like it is true.

**Lemma 273** *If $f : E \to \mathbb{R}$ is a nonnegative Borel function defined on a Borel set $E \subset \mathbb{R}^n$ such that $\int_E f = 0$ then*

$$\mu(f^{-1}((0, \infty))) = \mu(\{x \in E : f(x) > 0\}) = 0.$$

*That is, $f(x) = 0$ except for $x$ in a set of measure 0.*

**Proof.** We will prove the theorem by contrapositive. Letting $F := f^{-1}((0, \infty))$, suppose that $\mu(F) > 0$. As in the previous proof for continuous functions we would be done if we knew that $f$ had a positive lower bound on $F$, but there is no reason why this should be true. Instead, let $F_i := f^{-1}([\frac{1}{i}, \infty))$. Then $F$ is the increasing union of the collection $\{F_i\}$, and so $\mu(F_i) \to \mu(F)$ and for some $i$, $\mu(F_i) > 0$. On $F_i$, we have $f(x) \geq \frac{1}{i}$ and therefore $\int_E f \geq \int_{F_i} f \geq \frac{\mu(F_i)}{i} > 0$. ■

A statement that is true in the complement of a set of measure 0 is said to be true "almost everywhere," abbreviated "a.e." For example, the last statement of the above lemma can be restated as "That is, $f = 0$ a.e."

**Exercise 172** *Prove that if $f$ and $g$ are nonnegative Borel functions such that $f \leq g$ a.e. on a Borel set $E$ then $\int_E f \leq \int_E g$.*

We will conclude this section by showing that one can compute the measure of a compact set using covers by cubes (or semicubes) and obtaining some important consequences of this fact.

**Theorem 274** *If $C \subset \mathbb{R}^n$ is compact then there exists a sequence $\{\mathcal{K}_m\}_{m=1}^{\infty}$, where $\mathcal{K}_m$ is a collection $\{Q_{mi}\}_{i=1}^{n_m}$ of disjoint semicubes such that*

1. *for $m \geq 2$, each $Q_{mi}$ is contained in precisely one $Q_{(m-1)j}$*
2. *if $K_m = \bigcup_{i=1}^{n_m} Q_{mi}$ then $C$ is the decreasing intersection of $\{K_m\}$.*

*In particular, $\mu(C) = \lim \mu(K_m)$ and*

$$\mu(C) = \inf \left\{ \sum_{i=1}^{k} \mu(Q_i) \right\}$$

*where the infimum is over the collection of all $Q_i$ is a semicube (resp. cube) and $C \subset \bigcup_{i=1}^{k} Q_i$.*

**Proof.** Since $C$ is compact, $C$ is closed, hence Borel, and bounded, so $C$ lies in some $Q^n(r)$. We will successively subdivide the cube $Q^n(r)$ dividing the side length by half at each stage. We first subdivide $Q^n(r)$ into $2^n$ cubes $\{Q'_{1k}\}$ of side length $\frac{r}{2}$, any two of which meet only in a face. We then may change this into a pairwise disjoint collection of semicubes $\{Q_{1k}\}$ by letting $Q_{1k} := Q'_{1k} \setminus \bigcup_{i<k} Q_{1i}$. Each $Q'_{1k}$ may be subdivided into $2^n$ smaller cubes, each of which intersects $Q_{1k}$ in a semicube $Q'_{2j}$. Again these meet only in faces, and we may, as in the previous set, replace these semicubes by a pairwise disjoint collection. We will denote by $\mathcal{K}_m$ the collection of all semicubes constructed in this way of side length $r2^{-m}$ that intersect $C$, each of which has measure $(\frac{r}{2^m})^n$.

The corresponding collection $\{K_m\}$ is decreasing, since any semicube $Q$ in $\mathcal{K}_m(C)$ is contained in a larger semicube $Q'$ in the previous subdivision. Since $Q$ intersects $C$, so must $Q'$, which implies $Q' \in \mathcal{K}_{m-1}(C)$. This proves the first condition.

Next, observe that $C \subset K_m$ for all $m$. In fact, the semicubes subdivide all of $Q^n(r)$, which contains $C$ and therefore each $x \in C$ is contained in some semicube in $\mathcal{K}_m(C)$. Therefore $C \subset \bigcap_{m=1}^{\infty} K_m$. On the other hand, we also have that $K_m \subset N\left(C, \left(\frac{r}{2^{m-1}}\right)^n\right)$. According to Exercise 65 we have

$$\bigcap_{m=1}^{\infty} K_m \subset \bigcap_{\varepsilon>0} N(C,\varepsilon) = C.$$

Finally, since $C = \bigcap_{m=1}^{\infty} K_m$ and $\mu(K_1) \leq \mu(Q^n(r)) < \infty$, Proposition 263 implies that $\mu(C) = \lim \mu(K_m)$. The last statement follows from Proposition 62. In fact, $\mu(C)$ is certainly a lower bound for the set in question, and $\mu(K_m)$ is a sequence converging to $\mu(C)$. For any such collection of semicubes we may replace each semicube by the cube with the same interior, which has the same measure, and hence the last statement is also true for cubes (but the resulting cubes are not generally pairwise disjoint). ■

A related proposition, whose proof is an exercise, is the following.

**Proposition 275** *Let $f : U \to \mathbb{R}$ be a nonnegative Borel function defined on an open set $U \subset \mathbb{R}^n$, and suppose that $C \subset U$ is compact and $\int_U f < \infty$. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that $N(C,\delta) \subset U$ and $\int_{N(C,\delta)} f < \int_C f + \varepsilon$.*

**Corollary 276** *If $C \subset \mathbb{R}^n$ is compact then for every $\varepsilon > 0$ there exists an open set $U$ such that $C \subset U$ and $\mu(U) < \mu(C) + \varepsilon$.*

**Exercise 173** *Prove Proposition 275. Explain how the above corollary follows.*

## 4.5 Convergence Theorems

We already know that pointwise limits of Borel functions are Borel functions and therefore we can consider what happens when we integrate a sequence of Borel functions. We already know from Example 258 that the limit of the integrals need not be the integral of the limit. The following example shows this may happen even when the integrals in question are all finite.

**Example 277** *Let $f_n : [0,1] \to \mathbb{R}$ be defined by $f_n := \left(n^2+n\right)\chi_{[\frac{1}{n+1},\frac{1}{n}]}$. Now $f_n(0) = 0$ for all $n \in \mathbb{N}$, and for any $x \in (0,1]$, $f_n(x) = 0$ for all $n$ such that $\frac{1}{n} < x$. In other words, $(f_n)$ converges pointwise to $0$ and $\int \lim f_n = 0$. But*

$$\int_{[0,1]} f_n = \left(n^2+n\right)\left(\frac{1}{n} - \frac{1}{n+1}\right) = 1$$

*for all $n$, and so $\lim \int_{[0,1]} f_n = 1 \neq 0$. We will prove in this section two theorems about when the limit of the integrals is the integral of the limit, and the required conditions are quite simple–either that the sequence of functions is monotone increasing, or that the entire sequence is "dominated," i.e. there is a single function $g$ of finite integral that is greater than or equal to every function in the sequence. This function $g$ prevents the kind of "shifting around" that one sees in this example. In fact, for a function $g$ to be larger than all of the functions $f_n$, $g$ would have to have to have under its graph all of the rectangles having base $[\frac{1}{n+1},\frac{1}{n}]$ and height $n^2+n$. That would force $g$ to have infinite integral.*

We need some preliminary lemmas. The next lemma states, roughly, that if a Borel function has a finite integral then its integral over small sets must be small.

**Lemma 278** *Suppose $f : E \to \mathbb{R}$ is a nonnegative Borel function, $E \subset \mathbb{R}^n$, such that $\int_E f < \infty$. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $F \subset E$ is a Borel set with $\mu(F) < \delta$ then $\int_F f < \varepsilon$.*

**Proof.** Note that if $f$ were bounded by some $M$ the proof would be easy. In that case, we need only choose $\delta = \frac{\varepsilon}{M}$ to get $\int_F f \leq M\mu(f) < \varepsilon$. But $f$ may not be bounded, so we proceed by partitioning the range into bounded pieces. Define $E_m := f^{-1}([m,m+1))$ and $E^m := f^{-1}([m,\infty)) = \bigcup_{k=m}^{\infty} E_k$. Then the $E$ is the pairwise disjoint union of $\{E_k\}_{k=0}^{\infty}$ and

$$\int_E f = \sum_{k=0}^{\infty} \int_{E_k} f.$$

Fix $\varepsilon > 0$. By the Small Tails Lemma (179) there exists some $m$ such that

$$\int_{E^m} f = \sum_{k=m}^{\infty} \int_{E_k} f < \frac{\varepsilon}{2}.$$

Let $\delta := \frac{\varepsilon}{2m}$. Suppose that $F \subset E$ is a Borel set such that $\mu(F) < \delta$. Since $f(x) < m$ for all $x \notin E^m$ we have

$$\int_F f = \int_{E^m} f + \int_{F\backslash E^m} f \leq \int_{E^m} f + m\mu(F\backslash E^m)$$

$$\leq \frac{\varepsilon}{2} + m\mu(F) < \varepsilon.$$

■

**Exercise 174** *Use the sum of the functions $(f_n)$ in Example 277 to show that the assumption that $\int_E f < \infty$ cannot be removed from Lemma 278.*

The next lemma deals with the potential situation that a function could have finite integral on a set of infinite measure (an example of this from basic calculus is $\int_1^\infty \frac{1}{x^2} dx = 1$). Sets with finite measure are much easier to work with, and the next lemma shows that one can always approximate a finite integral over a set of infinite measure by an integral over a set of finite measure.

**Lemma 279** *Suppose $f : E \to \mathbb{R}$ is a nonnegative Borel function, $E \subset \mathbb{R}^n$, such that $\int_E f < \infty$. For every $\varepsilon > 0$ there exists a Borel set $F \subset E$ such that $\mu(F) < \infty$ and $\int_E f - \int_F f < \varepsilon$.*

**Proof.** In this case the proof would be simple if we had $f \geq c > 0$ for some $c$, for then $E$ would already have to have finite measure. The proof would also be trivial if $f$ were 0. This suggests dividing up the range in a different way from the last proof. Let $F_i := f^{-1}([\frac{1}{i}, \infty))$ for all $i \in \mathbb{N}$, $F_\infty = f^{-1}(0,\infty)$ and $F_0 := f^{-1}(\{0\})$. For any $x \in F_i$ we have $f(x) \geq \frac{1}{i}$ and

$$\infty > \int_E f \geq \int_{F_i} f \geq \frac{1}{i}\mu(F_i).$$

Hence $\mu(F_i) < \infty$ for all $i$. Now $F_i \nearrow F_\infty$ and since $\int_{F_0} f = 0$ and $E = F_0 \cup F_\infty$, $\int_{F_i} f \nearrow \int_{F_\infty} f = \int_E f$. Letting $F := F_i$ for sufficiently large $i$ finishes the proof. ■

The analogous statement for infinite integrals is the following:

**Lemma 280** *Suppose $f : E \to \mathbb{R}$ is a nonnegative Borel function, $E \subset \mathbb{R}^n$,such that $\int_E f = \infty$. For every $M > 0$ there exists a Borel set $F \subset E$ such that $\mu(F) < \infty$ and $M \leq \int_F f < \infty$.*

**Proof.** According to Exercise 175 below there is a Borel set $F' \subset E$ such that $\mu(F') < \infty$ and $\int_{F'} f \geq 2M$ (the problem is that $\int_{F'} f$ could still be infinite). Let $F_i := f^{-1}(([0,i])) \cap F'$. For every $i$, $\int_{F_i} f \leq i \cdot \mu(F') < \infty$. Also, $F'$ is the increasing union of the collection $\{F_i\}$ and so $\int_{F_i} f \to \int_{F'} f \geq 2M$. Therefore for some large $j$ $\int_{Fj} f \geq M$; let $F := F_j$. ■

**Exercise 175** *Finish the proof of Lemma 280. Hint: let $K_i := Q^n(i) \cap E$.*

**Theorem 281** *(Monotone Convergence Theorem) Let $(f_i)$ be a sequence of nonnegative Borel functions defined on a Borel set $E \subset \mathbb{R}^n$ such that $f_i \nearrow f$ (pointwise) on $E$. Then $\int_E f = \lim \int_E f_i$.*

**Proof.** Suppose first that $\int_E f < \infty$ and let $\varepsilon > 0$. Since $f \geq f_i$, the proof will be finished by Exercise 164 if we can show that for all large $i$,

$$\int_E f - \int_E f_i = \int_E (f - f_i) < \varepsilon.$$

First of all, by Lemma 279 there exists some Borel set $E' \subset E$ such that $\mu(E') < \infty$ and $\int_E f - \int_{E'} f < \frac{\varepsilon}{3}$. So for all $i$ we have

$$\int_E (f - f_i) - \int_{E'} (f - f_i) = \int_E f - \int_{E'} f - \left( \int_E f_i - \int_{E'} f_i \right) < \frac{\varepsilon}{3}.$$

If $\mu(E') = 0$ then $\int_E (f - f_i) < \frac{\varepsilon}{3} < \varepsilon$ as needed. Otherwise, according to Lemma 278 there exists a $\delta > 0$ such that if $F \subset E$ is a Borel set with $\mu(F) < \delta$ then $\int_F f < \frac{\varepsilon}{3}$. Let

$$F_i := \left\{ x \in E' : f(x) - f_i(x) > \frac{\varepsilon}{3\mu(E')} \right\}$$

Since $f_i \nearrow f$, $\{F_i\}_{i=1}^{\infty}$ is a decreasing sequence with $\bigcap_{i=1}^{\infty} F_i = \varnothing$. Since $\mu(F_i) \leq \mu(E') < \infty$, according to Proposition 263 $\mu(F_i) < \delta$ for all large $i$, and for such $i$ we have:

$$\int_E (f - f_i) \leq \int_{E'} (f - f_i) + \frac{\varepsilon}{3} = \int_{E' \cap F_i} (f - f_i) + \int_{E' \setminus F_i} (f - f_i) + \frac{\varepsilon}{3}$$

$$< \frac{\varepsilon}{3} + \left( \frac{\varepsilon}{3\mu(E')} \right) \mu(E') + \frac{\varepsilon}{3} = \varepsilon.$$

Now suppose that $\int_E f = \infty$ and let $M > 0$. We will prove that for all large $i$, $\int_E f_i \geq M$. According to Lemma 280 there is a Borel set $F \subset E$ such that $2M \leq \int_F f < \infty$. But $f_i \nearrow f$ on $F$ and so we can apply the above case for finite integrals to conclude that $\int_F f_i \to \int_F f \geq 2M$. But then for all large $i$,

$$\int_E f_i \geq \int_F f_i \geq M.$$

■

**Exercise 176** *Let $(f_i)$ be a monotone increasing sequence of nonnegative Borel functions $f_i : E \to \mathbb{R}$, where $E \subset \mathbb{R}^n$ is Borel. Suppose that there exists some $M$ such that for all $i$, $\int_E f_i \leq M$. Show that there exists a subset $F$ of $E$ such that the following statements are true:*

1. $\mu(F) = 0$
2. *if $g_i$ denotes the restriction of $f_i$ to $H := E \setminus F$ then $g_i \nearrow g$ for some Borel function $g : H \to \mathbb{R}$*
3. $\int_E f_i \nearrow \int_H g$.

The Monotone Convergence Theorem is very useful, but it is also important to have theorems about convergence that is not monotone increasing. The following definition provides a useful tool to translate pointwise convergence into monotone convergence.

**Definition 282** *Let $\{f_i\}_{i=1}^{\infty}$ be a collection of real-valued functions defined on a set A. The function $\inf_{k\geq m}(f_k)$ (resp. $\sup_{k\geq m}(f_k)$) is defined by*

$$\inf_{k\geq m}(f_k)(x) := \inf_{k\geq m}\{f_k(x)\} \text{ (resp. } \sup_{k\geq m}(f_k)(x) := \sup_{k\geq m}\{f_k(x)\}\text{)}$$

*for all $x \in A$.*

When $m = 1$ we will omit the subscript $k \geq m$ in this notation. Note that in general the functions $\inf_{k\geq m}(f_k)$ and $\sup_{k\geq m}(f_k)$ have extended real values. Below we will only consider the case when these functions are finite, but a completely general integration theory requires some consideration of possibly infinite valued functions.

**Lemma 283** *Let $f_k : X \to \mathbb{R}$ be Borel functions defined on a metric space $X$. If $\inf_{k\geq m}(f_k)$ (resp. $\sup_{k\geq m}(f_k)$) is finite then $\inf_{k\geq m}(f_k)$ (resp. $\sup_{k\geq m}(f_k)$) is Borel.*

**Proof.** We prove only that $g := \inf(f_k)$ is Borel; the proofs for $\inf_{k\geq m}(f_k)$ and $\sup_{k\geq m}(f_k)$ are similar. According to Lemma 239 we need only show that for any $a < b$ in $\mathbb{R}$, $g^{-1}([a, b])$ is a Borel set. Since $[a, b] = [a, \infty) \cap (-\infty, b]$ we need only show that the inverse images of the latter two sets are Borel sets. We will prove only the first one, leaving the other one for the reader. But $x \in g^{-1}([a, \infty))$ is equivalent to $\inf(f_k)(x) \geq a$, which is equivalent to $a$ being a lower bound for $\{f_k(x)\}_{k=1}^{\infty}$, which means $a \leq f_k(x)$ for all $k$, which means $x \in \bigcap_{k=1}^{\infty} f_k^{-1}([a, \infty))$. Putting these together we have $g^{-1}([a, \infty)) = \bigcap_{k=1}^{\infty} f_k^{-1}([a, \infty))$, which is a Borel set. ■

**Lemma 284** *Let $f_i : A \to \mathbb{R}$ be a sequence of functions such that $f_i \xrightarrow{p} f$ for some $f : A \to \mathbb{R}$. Define $g_k(x) := \inf_{i\geq k} f_i(x)$ (resp. $g_k(x) := \sup_{i\geq k} f_i(x)$). Then $g_k$ is Borel for all $k$ and $g_k \nearrow f$ (resp. $g_k \searrow f$).*

**Proof.** Lemma 283 implies that each $g_k$ is Borel. We prove only the case $g_k \nearrow f$; the other case is similar. But for any $x \in A$, $g_k(x) = \inf_{i\geq k} f_i(x)$ and therefore $(g_k(x))$ is increasing and by definition $\lim g_k(x) = \liminf f_k(x) = \lim f_k(x) = f(x)$. ■

**Lemma 285** *(Fatou's Lemma) Let $(f_i)$ be a sequence of nonnegative Borel functions defined on a Borel set $E$ such that $f_i \xrightarrow{p} f$ on $E$. Then $\int_E f \leq \liminf \int_E f_i$.*

**Proof.** For all $x \in E$ and every $i$, let $g_k(x) := \inf_{i \geq k} f_i(x)$. Lemma 284 implies each $g_k$ is Borel and $g_k \nearrow f$ and so $\int_E g_k \to \int_E f$ by the Monotone Convergence Theorem. Now $g_k(x) \leq f_k(x)$ for all $x \in E$ and therefore

$$\int_E f = \lim \int_E g_i \leq \liminf \int_E f_i.$$

■

**Corollary 286** *Let $(f_i)$ be a sequence of nonnegative Borel functions defined on a Borel set $E$ such that $f_i \xrightarrow{p} f$ on $E$. If $\int_E f = \infty$ then $\lim \int_E f_i = \int_E f = \infty$.*

**Theorem 287** *(Lebesgue Dominated Convergence Theorem) Let $f_i : E \to \mathbb{R}$ be a sequence of nonnegative Borel functions such that $f_i \xrightarrow{p} f$ for some $f : E \to \mathbb{R}$, where $E$ is a Borel subset of $\mathbb{R}^n$. If there exists a Borel function $g : E \to \mathbb{R}$ such that $\int_E g < \infty$ and $g(x) \geq f_i(x)$ for all $i$ and $x \in E$, then*

$$\int_E f = \lim \int_E f_i.$$

**Proof.** By Fatou's Lemma we have $\int_E f \leq \liminf \int_E f_i$. The functions $h_i := g - f_i$ are nonnegative and $h_i \xrightarrow{p} (g - f)$. Fatou's Lemma now implies

$$\int_E g - \int_E f = \int_E g - f \leq \liminf \int_E g - f_i$$

$$= \liminf \left( \int_E g - \int_E f_i \right) = \int_E g - \limsup \int_E f_i$$

or

$$\int_E f \geq \limsup \int_E f_i.$$

■

Recall that Example 277 shows that the "domination" assumption cannot be removed. Example 258 shows that one cannot remove this assumption even for monotone decreasing sequences, however, the next exercise does address this situation.

**Exercise 177** *Prove that if $f_i : E \to \mathbb{R}$ are Borel functions, $E \subset \mathbb{R}^n$, $\int_E f_1 < \infty$, and $(f_i)$ is monotone decreasing to a nonnegative function $f : E \to \mathbb{R}$ then $\int_E f = \lim \int_E f_i$.*

**Exercise 178** *The purpose of this exercise is to study the Uniform Convergence Theorem: Let $E \subset \mathbb{R}^n$ have finite measure and suppose $f_i : E \to \mathbb{R}$ are nonnegative Borel functions functions and $(f_i)$ converges uniformly to $f : E \to \mathbb{R}$. Then $\int_E f_i \to \int_E f$.*

1. *Prove the theorem. Hint: don't forget Corollary 286 and note that the Lebesgue Dominated Convergence Theorem is still valid if "all i" is replaced by "all large i."*

2. *Construct a decreasing sequence of functions* $(f_i)$ *converging uniformly to* 0 *on* $\mathbb{R}$ *such that* $\int_{\mathbb{R}} f_i = \infty$ *for all* $i$*. This shows that the assumption that* $\mu(E) < \infty$ *cannot be removed.*

3. *Construct a sequence of continuous functions* $(f_i)$ *converging uniformly to* 0 *on* $\mathbb{R}$ *such that* $\int_{\mathbb{R}} f_i = 1$ *for all* $i$*. Hint: You may use the fact, proved later, that the integral of a function whose graph is the upper part of a triangle sitting on the* $x$*-axis is the area of the triangle.*

**Exercise 179** *Modify Example 277 to make the functions* $f_n$ *continuous. As in the previous exercise you may use elementary calculus to compute the integrals. Hint: Replace rectangles by triangles.*

## 4.6 Simple Functions

**Definition 288** *A simple function* $f : \mathbb{R}^n \to \mathbb{R}$ *is a function of the form*

$$f = \sum_{i=1}^{k} c_i \chi_{E_i}$$

*where each* $c_i \geq 0$ *and* $E_i$ *is a Borel set.*

**Exercise 180** *Prove that the sum and product of two simple functions are simple functions.*

**Exercise 181** *Let* $\sum_{i=1}^{k} c_i \chi_{E_i}$ *be a simple function. Show that the function can be expressed as a finite linear combination of characteristic functions of pairwise disjoint Borel sets. Hint. For any subset* $S$ *of* $\{1, ..., k\}$*, let* $E_S := \bigcap\limits_{i \in S} E_i$*, and use the collection of all such subsets to index the sum.*

**Exercise 182** *Prove that if* $f = \sum_{i=1}^{k} c_i \chi_{E_i}$ *and* $g : \mathbb{R}^n \to \mathbb{R}^n$ *is a function* $f \circ g = \sum_{i=1}^{k} c_i \chi_{g^{-1}(E_i)}$.

Simple functions are linear combinations of Borel functions and hence are Borel functions. Given a nonnegative continuous function $f$ defined on an interval $[a, b]$, let $\mathcal{P}(n)$ be the partition $a = x_0 < x_1 < \cdots < x_{2^n} = b$ of $[a, b]$ into $2^n$ intervals of length $2^{-n}$. Define a simple function $g_n := \sum_{i=1}^{2^n} m_i \chi_{E_i}$, where $E_i = [x_{i-1}, x_i)$ for $1 \leq i < 2^n$, $E_{2^n} = [x_{2^n-1}, x_{2^n}]$, and for all $i$, $m_i := f(x_{i-1})$.

**Exercise 183** *Show that if* $g_n$ *is the function defined above then* $g_n \xrightarrow{p} f$.

Since $f$ is continuous on $[a, b]$, $f$ is bounded above by some $M < \infty$. By definition we also have $g_n \leq M$ for all $n$. Letting $g(x) = M$, we have $\int_{[a,b]} g = M(b - a) < \infty$ and $g_n \leq g$ for all $n$. The Lebesgue Dominated Converence Theorem applies and we have $\int_{[a,b]} g_n \to \int_{[a,b]} f$. But $\int_{[a,b]} g_n = \sum_{i=1}^{2^n} m_i(2^{-i})$ is also a Riemann sum for $f$ and since these converge to $\int_a^b f$, we have proved:

**Theorem 289** *If $f : [a,b] \to \mathbb{R}$ is nonnegative and continuous then $\int_{[a,b]} f = \int_a^b f$.*

Thus the Riemann integral of a continuous function on a compact interval is the same as the Lebesgue integral, which means that all of the theorems we have proved about Lebesgue integration hold for Riemann integration of continuous functions. However, the statements require many caveats–for example, we already know that pointwise or even monotone limits of continuous functions may not be continuous, so the convergence theorems may only be used when the limit function can otherwise be proved to be continuous, as in the case of uniform convergence of continuous functions. It is not hard to extend the above argument to prove that a continuous nonnegative function defined on a compact set can also be computed using higher-dimensional analogs of Riemann sums, although the region in question generally cannot be subdivided into higher dimensional semicubes.

What about an arbitrary nonnegative Riemann integrable function? It can be shown that a real valued function defined on $[a,b]$ is Riemann integrable if and only if it is continuous except on a subset of a Borel set of measure 0 (with an analogous statement for appropriately defined regions in higher dimensions). The fact that there are such subsets that are not Borel sets points to the necessity of "completing" the collection of Borel sets by adding all subsets of Borel sets having measure 0. The smallest $\sigma$-algebra of sets containing all open sets and all subsets of Borel sets of measure zero is called the $\sigma$-algebra of *measurable sets.* One might say that Lebesgue integration of Borel functions, while not the most general possible, differs from the general theory only by subsets of sets of measure 0–close enough for our purposes.

In elementary calculus one also learns about "improper" integrals, such as $\int_a^\infty f(x)dx$, which is defined to be $\lim_{t\to\infty} \int_a^t f(x)dx$. However, it follows from Proposition 256 that if $f$ is nonnegative and this limit exists then the limit must be equal to the Lebesgue integral $\int_{[a,\infty)} f$.

We will next prove the interesting fact that the Lebesgue integral of an arbitrary nonnegative Borel function can be computed in essentially the same way as the Riemann integral, except that rather than using only simple functions that are defined on partitions, we must use arbitrary simple functions. The key difference is that the approximations (Riemann sums) for the Riemann integral are obtained by subdividing the domain of the interval, while the approximations of the Lebesgue integral are obtained by subdividing the range and "pulling back" the subdivision to the domain. The advantage of the latter becomes clear when the function in question is defined on a subset of $\mathbb{R}^n$ for $n > 1$. While an interval is relatively easy to subdivide into smaller intervals, an arbitrary closed, bounded, connected set in higher dimensional space is not so easy to subdivide nicely, and certainly cannot always be subdivided into cubes. On the other hand, since all of our functions are real-valued, subdividing the range, as is done with the Lebesgue integral, remains easy to do regardless of the dimension of the domain space.

**Definition 290** *A subset $A$ of a metric space $X$ is called an $F_\sigma$ set if $A$ is the countable union of closed sets.*

Certainly $F_\sigma$ sets are Borel sets, and closed sets are $F_\sigma$ sets–but for example any half-open interval is an $F_\sigma$ that is not closed. If $A = \bigcup_{i=1}^{\infty} A_i$, where each $i$ is closed then one can always write $A$ as an increasing union $A = \bigcup_{i=1}^{\infty} (A_1 \cup \cdots \cup A_i)$. If $A$ is an $F_\sigma$ subset of $\mathbb{R}^n$ then we can further write

$$A = \bigcup_{i=1}^{\infty} [(A_1 \cup \cdots \cup A_i) \cap Q^n(i)]$$

where each $(A_1 \cup \cdots \cup A_i) \cap Q^n(i)$ is compact. In other words:

**Lemma 291** *If $A \subset \mathbb{R}^n$ is an $F_\sigma$ then $A$ is the increasing union of countably many compact sets.*

Sets that are countable increasing unions of compact sets are sometimes called *$\sigma$-compact*.

**Proposition 292** *Let $f : E \to \mathbb{R}$ be a nonnegative Borel function defined on a Borel set $E$ in $\mathbb{R}^n$. There exists a sequence $(f_i)$ of simple functions such that $f_i \nearrow f$ on $E$. If $f$ is continuous and $E$ is closed then each $f_i$ is of the form $\sum_{n=1}^{j} c_j \chi_{E_j}$ where each $E_j$ is an $F_\sigma$.*

**Proof.** For every $k \in \mathbb{N}$, subdivide $[0, k)$ into $k \cdot 2^k$ intervals of equal length $2^{-k}$; each interval is of the form $[(j-1)2^{-k}, j2^{-k})$ with $1 \le j \le k \cdot 2^k$. We know that each of the sets $E_j^k := f^{-1}([(j-1)2^{-k}, j2^{-k}))$ is a Borel set. Define $f_k := \sum_{j=1}^{k \cdot 2^k} (j-1)2^{-k} \chi_{E_j^k}$. For $x \in E_j^k$ we have, by definition, $(j-1)2^{-k} \le f(x) < j2^{-k}$ and therefore if $E^k := f^{-1}([0, k))$ we have for all $x \in E^k$, $0 \le f(x) - f_k(x) < 2^{-k}$. Now for any $x \in E$ we have $f(x) \in [0, k)$ for all sufficiently large $k$ and therefore $|f(x) - f_k(x)| \le 2^{-k}$ for all such $k$, so $f_k(x) \to f(x)$. It is an exercise to prove that the sequence $(f_k)$ is monotone increasing.

Note that each interval $[(j-1)2^{-k}, j2^{-k})$ is a countable union of closed intervals $[(j-1)2^{-k}, a_i]$, where $a_i$ is some sequence in the interval $((j-1)2^{-k}, j2^{-k})$ converging to $j2^{-k}$. If $f$ is continuous then since the inverse image of closed sets is closed, each $E_j^k$ is a countable union of sets that are closed in $E$. If $E$ is closed in $\mathbb{R}^n$ then these sets are also closed in $\mathbb{R}^n$. ■

**Corollary 293** *If $f : E \to \mathbb{R}$ is a nonnegative Borel function then $\int_E f = \sup \left\{ \int_E g \right\}$ where the supremum is over all simple functions $g$ such that $g \le f$ on $E$.*

**Exercise 184** *Finish the proof of Proposition 292.*

## 4.7 Fubini's Theorem

In elementary calculus one learns that double and triple integrals can be computed using iterated integration. We will now show that the same procedure works for the Lebesgue integral of a nonnegative continuous function defined on a compact set. This theorem, known as Fubini's Theorem, is true under weaker assumptions, but the proof of the most general version is beyond the scope of this text. The version we present here will be adequate for our purposes; in particular we will not be combining iterated integration with limits of functions and will not need our general convergence theorems when using iterated integration. Even though we will prove the theorem for continuous functions there is still a significant advantage over the theorem for Riemann integration, namely that we can prove it for continuous functions defined on any compact set, and not simply on certain "nice" regions (generally called Jordan regions), which are essentially regions having boundary of measure 0. Such restrictions on the domain of a function are highly problematic in certain kinds of analysis, for example geometric measure theory, which includes analysis of fractal sets.

First, let's examine an exercise from calculus to see how it works in our present terminology. Suppose we wish to integrate a function $f(x, y, z) = xyz$ over the region $R$ in $\mathbb{R}^3$ that is bounded by the parabolic cylinder $y = x^2$, the plane $y + z = 4$, and the $x, y$-plane. We can set up the integral

$$\int_{-2}^{2} \int_{x^2}^{4} \int_{0}^{4-y} (xyz)\, dzdydx.$$

How do we evaluate it? The usual description from elementary calculus is something like "treat $x$ and $y$ like constants and do the inside integral with respect to $z$." This results in the iterated integral $\int_{-2}^{2} \int_{x^2}^{4} \frac{1}{2} xy(4-y)^2 dydx$, which is equal to the integral $\int_A \frac{1}{2} xy(4-y)^2 dydx$ where $A$ is the region in the $(x, y)$-plane bounded by the parabola $y = x^2$ and the line $y = 4$. The region $A$ is also the projection of the region $R$ onto the $(x, y)$-plane. Really what we have done in this first stage is reduce the triple integral to a double integral, and this is the step we will focus on. Reducing the resulting double integral to a single integral is similar.

We consider $\mathbb{R}^3$ as $\mathbb{R}^2 \times \mathbb{R}$, where the first factor is the $(x, y)$-plane and the latter factor is the $z$-axis. We let $\pi_1 : \mathbb{R}^3 \to \mathbb{R}^2$ and $\pi_2 : \mathbb{R}^3 \to \mathbb{R}$ be the projections. So $A = \pi_1(R)$. If we fix any $(x, y) \in A$, the vertical line above $(x, y)$ is $L_{(x,y)} = \pi_1^{-1}((x, y))$. Because $R$ is a "reasonable" set, this line intersects $R$ in a line segment (possibly a single point). We will be integrating a function of one variable, so we project this segment onto the $z$-axis. That gives us a set

$$R_{(x,y)} := \pi_2(\pi_1^{-1}((x, y)) \cap R) = \{z \in \mathbb{R} : (x, y, z) \in R\}.$$

Again, because $R$ is "reasonable" each $R_{(x,y)}$ is an interval in $\mathbb{R}$, which we will refer to as the "slice" of $R$ at the point $(x, y)$. What are the endpoints of this interval? The lower endpoint is the projection of the point where the line $L_{(x,y)}$ first meets the region $R$, which always has $z$-coordinate 0. This line exits $R$

through the plane $y + z = 4$, at the point with $z$-coordinate $4 - y$. Therefore $R_{(x,y)} = [0, 4 - y]$. We may now consider the function of one variable $f_{(x,y)} : R_{(x,y)} \to \mathbb{R}$ defined by $f_{(x,y)}(z) = f(x, y, z) = xyz$. This is what it means to "treat $x$ and $y$ like constants." We now integrate this function over the set $R_{(x,y)}$, resulting in a function of two variables $I(x, y) := \int_0^{4-y} f_{(x,y)}(z)dz = \int_0^{4-y} xyzdz$ we now have a continuous function $I : A \to \mathbb{R}$ of two variables and can perform the double integral $\int_A I(x, y)dxdy$, which we have been assured by our calculus text is equal to the original triple integral. Actually proving that this is the case is nontrivial. Note that, with our current notation, we are finding the measure of the region by integrating the measure of the "slice" of the region above each point $(x, y)$ when it is projected onto the $z$-axis. Note that the measure of $\pi_1^{-1}((x, y)$ is actually 0 since it is contained in a vertical line; that is why we project it to the one-dimensional space $\mathbb{R}$, where its measure is $4 - y$.

Proving that iterated Riemann integration works when the region is a rectangle is relatively straightforward; it is the passage from rectangles to more general regions that presents difficulties. The standard method involves covering a compact region in the plane by squares that meet only in their faces, the idea being to apply the theorem to each square and add the results. Even if one assumes that the function is defined on this larger region, one must still take a limit as the union of squares decreases to the original region–and as we have already observed Riemann integration fares poorly when it comes to limits. Instead, for Riemann integration one must proceed in a fairly *ad hoc* manner, extending the function as 0 on the larger region so that the integral over the larger region is equal to the integral over the original. Of course this extension need not be continuous, and one must pile on additional assumptions about the boundary of the region to ensure that the extended function is Riemann integrable. None of this is necessary with Lebesgue integration, for which adequate limiting theorems are available.

The proof of Fubini's theorem is somewhat involved, but general strategy is widely applicable in analysis: prove the theorem for characteristic functions, then for simple functions, and then use Proposition 292 to finish it off. This will not be the last time we follow these steps.

Our starting point should also be familiar from elementary calculus; we will show that the measure of a compact set (which is the integral of its characteristic function) in $\mathbb{R}^n \times \mathbb{R}^m$ can be computed using "integration by slices".

**Definition 294** *Let $E \subset \mathbb{R}^n \times \mathbb{R}^m$ and let $\pi_1 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\pi_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be the projections. Define $E_1 := \pi_1(E) \subset \mathbb{R}^n$ and for any $\mathbf{x} \in E_1$ define*

$$E_{\mathbf{x}} := \{\mathbf{y} \in \mathbb{R}^m : (\mathbf{x}, \mathbf{y}) \in E\} = \pi_2(\pi_1^{-1}(\mathbf{x}) \cap E).$$

*The set $E_{\mathbf{x}}$ is called the slice of $E$ at $\mathbf{x}$.*

Note, the slice $E_{\mathbf{x}}$ is, strictly speaking, a subset of $\mathbb{R}^m$. Geometrically it is useful to think of $E_{\mathbf{x}}$ as being the set $\pi_1^{-1}(\mathbf{x}) \cap E$, which really is a "slice" of the set $E$ sitting above the point $\mathbf{x}$ in $\mathbb{R}^m$, but of course in $\mathbb{R}^{n+m}$ this set has measure 0! That is why we must project to $\mathbb{R}^m$ if we are to correctly determine its measure.

**Exercise 185** *Verify the equality* $\{\mathbf{y} \in \mathbb{R}^m : (\mathbf{x}, \mathbf{y}) \in E\} = \pi_2(\pi_1^{-1}(\mathbf{x}) \cap E)$ *in Definition 294.*

**Lemma 295** *If* $E \subset \mathbb{R}^n \times \mathbb{R}^m$ *is compact then for all* $\mathbf{x} \in E_1$, $E_{\mathbf{x}}$ *is compact.*

**Proof.** Since $\{\mathbf{x}\}$ is closed, $\pi_1^{-1}(\mathbf{x})$ is closed and $E$ is compact, hence closed and bounded. Therefore $\pi_1^{-1}(\mathbf{x}) \cap E$ is closed and bounded, hence compact. Finally, $E_{\mathbf{x}} = \pi_2(\pi_1^{-1}(\mathbf{x}) \cap E)$ is compact. ■

**Theorem 296** *(Measure by Slices) Let* $E \subset \mathbb{R}^n \times \mathbb{R}^m$ *be a bounded* $F_\sigma$ *set and define* $\eta : \mathbb{R}^n \to \mathbb{R}$ *by* $\eta(\mathbf{x}) := \mu(E_{\mathbf{x}})$*. Then* $E_1$ *is a Borel set,* $\eta$ *is a Borel function, and* $\int_{E_1} \eta = \mu(E)$*.*

**Proof.** Note that if $\mathbf{x} \notin E_1$ then $E_{\mathbf{x}} = \varnothing$ and $\eta(\mathbf{x}) = 0$. Therefore, in each case below when we have proved $\eta$ is Borel we will know that for any Borel set $A$

$$\text{if } E_1 \subset A \text{ then } \int_A \eta = \int_{E_1} \eta. \tag{4.2}$$

Suppose first that $E$ is a cube of the form $Q_1 \times Q_2$, where $Q_1 = E_1$ and $Q_2$ are cubes of side length $r$ in $\mathbb{R}^n$, $\mathbb{R}^m$, respectively. Then $E_{\mathbf{x}} = Q_2$ for all $\mathbf{x} \in E_1$ and therefore $\eta(\mathbf{x}) = \mu(Q_2) = r^m$, which is constant, hence Borel. In addition,

$$\int_{E_1} \eta = \int_{Q_1} r^m = r^m \mu(Q_2) = r^{n+m} = \mu(E).$$

Now suppose that $E$ is a semicube $(Q_1 \times Q_2) \setminus A$. A somewhat tedious argument shows that $E_1$ is a semicube $Q_1 \setminus A_1$ and $E_{\mathbf{x}}$ is a semicube $Q_2 \setminus A_{\mathbf{x}}$ for all $\mathbf{x}$. (For example, one needs to show by writing out the definitions of the sets involved that the boundary of $Q_1$ is the projection of a compact part of the boundary of $Q_1 \times Q_2$. This is easy to see in low dimensions.) Once again we have $\eta(\mathbf{x}) = r^m$, $\mu(E) = r^{n+m}$ and $\int_{E_1} \eta = \int_{Q_1} r^m = r^{n+m}$.

Next suppose that $E$ is a finite union of pairwise disjoint semicubes in $\mathbb{R}^n \times \mathbb{R}^m$; $E := K_1 \cup \cdots \cup K_N$. Now $\pi_1^{-1}(\mathbf{x}) \cap E$ is the pairwise disjoint union of the sets $\pi_1^{-1}(\mathbf{x}) \cap K_i$ and since $\pi_2$ is one-to-one on $\pi_1^{-1}(\mathbf{x})$, each $E_{\mathbf{x}}$ is the pairwise disjoint union of the sets $(K_i)_{\mathbf{x}}$. We have $\eta(\mathbf{x}) = \mu(E_{\mathbf{x}}) = \sum_{i=1}^N \mu((K_i)_{\mathbf{x}})$. That is, $\eta$ is a sum of Borel functions and hence is Borel. Since $(K_i)_1 \subset E_1$, we have from (4.2) that

$$\int_{E_1} \eta = \int_{E_1} \mu(E_{\mathbf{x}}) = \int_{E_1} \sum_{i=1}^n \mu((K_i)_{\mathbf{x}}) = \sum_{i=1}^n \int_{E_1} \mu((K_i)_{\mathbf{x}})$$

$$= \sum_{i=1}^n \int_{(K_i)_1} \mu((K_i)_{\mathbf{x}}) = \sum_{i=1}^n \mu(K_i) = \mu(E).$$

Now let $E$ be an arbitrary compact set. According to Theorem 274 there exists a collection $\{F_i\}$ of sets such that $E$ is the decreasing intersection of the collection and each $F_i$ is a finite union of pairwise disjoint semicubes. It is easy to check that $E_{\mathbf{x}}$ is the decreasing intersection of the collection $\{(F_i)_{\mathbf{x}}\}$ and therefore $\mu((F_i)_{\mathbf{x}}) \searrow \mu(E_{\mathbf{x}})$. By definition this means that if $\eta_i(\mathbf{x}) := \mu((F_i)_{\mathbf{x}})$ then $\eta_i \searrow \eta$. Since $\eta_i$ is Borel by the case we just proved, $\eta$ is Borel. In addition, $\int_{(F_1)_1} \eta_1 = \mu(F_1) < \infty$, and since the sequence is decreasing $\eta_i \leq \eta_1$ for all $i$. We can apply the Lebesgue Dominated Convergence Theorem. Applying Formula (4.2) twice we have

$$\int_{E_1} \eta = \int_{(F_1)_1} \eta = \lim \int_{(F_1)_1} \eta_i = \lim \int_{(F_i)_1} \eta_i = \lim \mu(F_i) = \mu(E).$$

The general case, when $E$ is an $F_\sigma$, is an exercise. ∎

**Exercise 186** *Finish the proof of Measure by Slices. Be sure to explicitly use the fact that E is bounded. Hint: Use the fact that E is the increasing union of compact sets* $\{E_i\}$.

**Example 297** *Let $E$ be the union of the $y$-axis and the segment of the $x$-axis from the point $(-1, 0)$ to the point $(1, 0)$. According to Exercise 171 $\mu(E) = 0$, but of course $E$ is not compact. Now let's see what happens if we try to use Theorem 296 to compute the measure of $E$. Then $E_1$ is the interval $[-1, 1]$ and $E_x = \{0\}$ for $x \neq 0$, but $E_0 = \mathbb{R}$, and $\mu(E_0) = \infty$. This means that the function $\eta$ has an infinite value at $0$ and hence is not a real valued function. Therefore we cannot apply our integration theory. Resolving this problem (and proving the general version of Fubini's Theorem) requires an integration theory for functions into the extended reals–or at least for functions that have infinite value only on a set of measure $0$. The infinite values then do not affect the integral. In the present example the function $\eta$ takes an infinite value at a single point, which has measure $0$, and is $0$ elsewhere. Thus in a more general theory the integral of $\eta$ would be $0$, as expected.*

Recall from elementary calculus that the integral of a non-negative continuous function real function defined on an interval is interpreted to be the area of the region bounded by the graph. We will make this idea precise and more general by showing that the integral of a bounded function defined on a compact set is the measure of the set bounded by the graph.

**Definition 298** *Let $f : X \to \mathbb{R}$ be a nonnegative function, where $X$ is a set. Define*

$$A(f) := \{(x, y) \in X \times \mathbb{R} : 0 \leq y \leq f(x)\}.$$

**Exercise 187** *Show that if $f : E \to \mathbb{R}$ is a nonnegative continuous function, where $E$ is a compact subset of $\mathbb{R}^n$ then $A(f)$ is compact. Hint: Use the Heine-Borel Theorem.*

**Theorem 299** *If $f : E \to \mathbb{R}$ is a nonnegative continuous function, where $E$ is a compact subset of $\mathbb{R}^n$ then*

$$\mu(A(f)) = \int_E f.$$

**Proof.** We will compute the measure of $A(f) \subset \mathbb{R}^n \times \mathbb{R}$ by slices. First note that $(A(f))_1 = E$ and for any $\mathbf{x} \in E$, $_\mathbf{x}$ $(A(f))_\mathbf{x}$is the interval $[0, f(\mathbf{x})]$ in $\mathbb{R}$, which has measure $f(\mathbf{x})$. Therefore

$$\mu(A(f)) = \int_E \mu((A(f))_\mathbf{x}) = \int_E f.$$

■

**Exercise 188** *Prove that if $f : E \to \mathbb{R}$ is a nonnegative continuous function, where $E$ is a compact subset of $\mathbb{R}^n$ then the graph of $f$ has measure $0$ in $\mathbb{R}^{n+1}$.*

**Definition 300** *Let $E \subset \mathbb{R}^n \times \mathbb{R}^m$, $\pi_1 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\pi_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be the projections, and $f : E \to \mathbb{R}$ be a function. Define $f_\mathbf{x} : E_\mathbf{x} \to \mathbb{R}$ by $f_\mathbf{x}(\mathbf{y}) := f(\mathbf{x}, \mathbf{y})$.*

Note that if $f$ is continuous then the function $g_\mathbf{x}(\mathbf{y}) := f(\mathbf{x}, \mathbf{y})$ is continuous and since $f_\mathbf{x}(\mathbf{y}) = f(g_\mathbf{x}(\mathbf{y}))$, $f_\mathbf{x}$ is continuous for any $\mathbf{x}$.

**Exercise 189** *Let $E \subset \mathbb{R}^n \times \mathbb{R}^m$ let $f = \sum_{i=1}^k c_i g_i$ where each $g_i : E \to \mathbb{R}$ is a function. Prove that for any $\mathbf{x} \in E_1$, $f_\mathbf{x} = \sum_{i=1}^k c_i \left(g_i\right)_\mathbf{x}$.*

**Exercise 190** *Let $f_i \nearrow f$ on $E \subset \mathbb{R}^n \times \mathbb{R}^m$. Prove that for any $\mathbf{x} \in E_1$, $(f_i)_\mathbf{x} \nearrow f_\mathbf{x}$. A similar statement holds for decreasing sequences.*

**Lemma 301** *Let $E \subset \mathbb{R}^n \times \mathbb{R}^m$. Then for any $\mathbf{x} \in \mathbb{R}^n$, $(\chi_E)_\mathbf{x} = \chi_{E_\mathbf{x}}$.*

**Proof.** For any $\mathbf{y} \in \mathbb{R}^m$, we have $(\chi_E)_\mathbf{x}(\mathbf{y}) = \chi_E(\mathbf{x}, \mathbf{y}) = 1$ if and only if $(\mathbf{x}, \mathbf{y}) \in E$ and $0$ otherwise. But $\chi_{E_\mathbf{x}}(\mathbf{y}) = 1$ if and only if $\mathbf{y} \in E_\mathbf{x}$, which is equivalent to $(\mathbf{x}, \mathbf{y}) \in E$, and $0$ otherwise. ■

Note that in Fubini's Theorem below, if we take $f_x = 1$ then the theorem reduces to Measure by Slices; hence the latter theorem is a special case of Fubini's Theorem.

**Theorem 302** *(Fubini's Theorem) If $E \subset \mathbb{R}^n \times \mathbb{R}^m$ is compact and $f : E \to \mathbb{R}$ is a nonnegative continuous function then $I(\mathbf{x}) := \int_{E_\mathbf{x}} f_\mathbf{x}$ is a Borel function and*

$$\int_E f = \int_{E_1} \left( \int_{E_\mathbf{x}} f_\mathbf{x} \right).$$

**Proof.** First let $g=\sum_{i=1}^{k}c_i\chi_{F_i}$ where each $F_i\subset E$ is a bounded $F_\sigma$. Then for any $\mathbf{x}\in E_1$, by Exercise 189 and Lemma 301,

$$\int_{E_{\mathbf{x}}}g_{\mathbf{x}}=\int_{E_{\mathbf{x}}}\sum_{i=1}^{k}c_i\left(\chi_{F_i}\right)_{\mathbf{x}}=\sum_{i=1}^{k}c_i\int_{E_{\mathbf{x}}}\chi_{(F_i)_{\mathbf{x}}}=\sum_{i=1}^{k}c_i\mu((F_i)_{\mathbf{x}})$$

and by Theorem 296 the function $h(\mathbf{x}):=\int_{E_{\mathbf{x}}}g_{\mathbf{x}}$ is a Borel function and

$$\int_{E_1}\left(\int_{F_{\mathbf{x}}}g_{\mathbf{x}}\right)=\int_{E_1}\left(\sum_{i=1}^{k}c_i\mu((F_i)_{\mathbf{x}})\right)=\sum_{i=1}^{k}c_i\int_{E_1}\mu((F_i)_{\mathbf{x}})$$

$$=\sum_{i=1}^{k}c_i\int_{F_i}\mu((F_i)_{\mathbf{x}})=\sum_{i=1}^{k}c_i\mu(F_i)=\int_E g.$$

According to Proposition 292 there exist simple functions $g_i$ of the form considered above such that $g_i\nearrow f$ (the $F_\sigma$ sets are bounded because they are contained in the bounded set $E$). By Exercise 190, for any $\mathbf{x}\in E_1$, $(g_i)_{\mathbf{x}}\nearrow f_{\mathbf{x}}$. By the Monotone Convergence Theorem $\int_{E_{\mathbf{x}}}(g_i)_{\mathbf{x}}\nearrow\int_{E_{\mathbf{x}}}f_{\mathbf{x}}$ for all $\mathbf{x}$. Since each function $I_i(\mathbf{x}):=\int_{E_{\mathbf{x}}}(g_i)_{\mathbf{x}}$ is a Borel function and $I_i\nearrow I$, $I$ is also a Borel function. We can now apply the Monotone Convergence Theorem to conclude that

$$\int_{E_1}I=\lim\int_{E_1}I_i=\lim\int_{E_1}\int_{E_{\mathbf{x}}}(g_i)_{\mathbf{x}}=\lim\int_{E_1}g_i=\int_{E_1}f.$$

■

Note that we could carry out the above proof as well by letting $E_2$ be the projection of $E$ onto $\mathbb{R}^m$ and for any $\mathbf{y}\in E_2$ defining $f_{\mathbf{y}}$ in an analogous fashion to obtain:

$$\int_{E_2}\left(\int_{E_{\mathbf{y}}}f_{\mathbf{y}}\right)=\int_E f=\int_{E_1}\left(\int_{E_{\mathbf{x}}}f_{\mathbf{x}}\right).$$

In other words we can "reverse the order of integration."

**Exercise 191** *Let $f:A\to\mathbb{R}$ and $g:B\to\mathbb{R}$ be continuous nonnegative real functions and define $h:A\times B\to\mathbb{R}$ by $h(x,y)=f(x)g(y)$, where $A$ and $B$ are compact subsets of $\mathbb{R}$. Prove that $\int_{A\times B}h=\left(\int_A f\right)\left(\int_B g\right)$.*

## 4.8 Integration of Arbitrary Borel Functions

Of course one needs to consider integration of functions that are not necessarily nonnegative. Nonnegative functions are a bit easier to deal with, especially when we are dealing with integrals that may be infinite. We treat arbitrary Borel functions by writing them as a difference of nonnegative functions, and then applying the theory we have already developed to each of those nonnegative functions. This allows us to quickly extend our theory to more general functions.

**Definition 303** *Let $f : E \to \mathbb{R}$ be a function defined on a set $E$. Define $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) := -\min\{f(x), 0\}$.*

**Exercise 192** *Let $f : E \to \mathbb{R}$ be a Borel function defined on a Borel set $E \subset \mathbb{R}^n$. Prove that*

1. *$f^+$ and $f^-$ are nonnegative Borel functions on $E$,*
2. *$f = f^+ - f^-$, and*
3. *$|f| = f^+ + f^-$.*

**Definition 304** *Let $f : E \to \mathbb{R}$ be a Borel function defined on a set $E \subset \mathbb{R}^n$. If at least one of $\int_E f^+$ or $\int_E f^-$ is finite we say that $\int_E f$ exists and define $\int_E f := \int_E f^+ - \int_E f^-$. If $\int_E f$ exists and is finite then we say that $f$ is integrable.*

Note that $f$ is integrable if and only if both $f^+$ and $f^-$ are integrable. There are analogs for arbitrary Borel functions of most of the main theorems that we have proved about nonnegative functions, plus a few, like the next two lemmas, that are trivial for nonnegative functions. We must take care to assume (or assert) the existence of the integral where needed, and some care must be taken to avoid trying to add $+\infty$ and $-\infty$. Because of such problems many of the theorems below will be stated only for integrable functions, although some more general statements (with appropriate caveats) are also true.

**Lemma 305** *For any $f : E \to \mathbb{R}$, if $\int_E f$ exists then, $\left|\int_E f\right| \leq \int_E |f|$. Moreover, $f$ is integrable on $E$ if and only if $|f|$ is integrable.*

**Proof.** We compute

$$\left|\int_E f\right| = \left|\int_E f^+ - \int_E f^-\right| \leq \int_E f^+ + \int_E f^- = \int_E (f^+ + f^-) = \int_E |f|\,.$$

Now $f$ is integrable if and only if both $\int_E f^+$ and $\int_E f^-$ are finite, and since $|f| = f^+ + f^-$ this is equivalent to $\int_E |f|$ being finite. ∎

A few comments about the above lemma are in order. First of all, one should contrast what this lemma tells us with what we already know about series–that a series may converge but not converge absolutely. Lemma 305 is true because we separate out the positive and negative parts and only proceed if the integral of one of these parts is finite. In a series that converges, but not absolutely, the series of negative terms and the series of positive terms must both, by themselves, diverge. There is a similar distinction between Lebesgue integration and improper Riemann integration. While for nonnegative continuous functions we have already seen that Lebesgue integration and improper Riemann integration agree, the same is not true for arbitrary continuous functions, as is shown in the following exercise.

**Exercise 193** *Show that of the improper Riemann integrals* $\int_0^\infty \frac{\sin x}{x} dx$ *and* $\int_0^\infty \left|\frac{\sin x}{x}\right| dx$*, the first is finite but the second is infinite, using the steps below. The latter integral must equal the Lebesgue integral* $\int_{(0,\infty)} \left|\frac{\sin x}{x}\right|$*, and so by Lemma 305,* $\frac{\sin x}{x}$ *is not Lebesgue integrable on* $(0,\infty)$*. Therefore the improper Riemann integral* $\int_0^\infty \frac{\sin x}{x} dx$ *is not equal to the Lebesgue integral* $\int_{(0,\infty)} \frac{\sin x}{x}$*. You may use theorems from elementary calculus to do the calculations.*

1. *Show that* $\frac{\sin x}{x}$ *is the restriction of a continuous function defined on* $[0,\infty)$ *and therefore* $\int_0^\infty \frac{\sin x}{x} dx$ *is finite if and only if* $\int_1^\infty \frac{\sin x}{x} dx$ *is finite.*

2. *Use integration by parts and the fact that* $\int_1^\infty \frac{dx}{x^2} < \infty$ *to finish showing* $\int_0^\infty \frac{\sin x}{x} dx < \infty$*.*

3. *Prove that* $\int_{[0,n]} \frac{|\sin \pi x|}{x} \geq \frac{2}{\pi} \sum_{i=1}^n \frac{1}{i}$ *and use this to show that* $\int_0^\infty \left|\frac{\sin x}{x}\right| dx = \infty$*. Hint: Consider the intervals* $[i-1, i]$*.*

The (improper) Riemann integrability of $|f|$ is sometimes referred to as "absolute" integrability of $f$ in analogy with series. The above exercise therefore exhibits a function that is Riemann integrable but not absolutely Riemann integrable. We know already that there is no such distinction for Lebesgue integrals, although if the Lebesgue integral of a Borel function does not exist one can sometimes still consider improper Lebsegue integrals using limits in analogy with improper Riemann integrals.

**Lemma 306** *Suppose* $f = f_1 - f_2$*, where* $f_1$ *and* $f_2$ *are nonnegative integrable functions defined on* $E \subset \mathbb{R}^n$*. Then* $f$ *is integrable and* $\int_E f = \int_E f_1 - \int_E f_2$*.*

**Proof.** First, $|f| \leq |f_1| + |f_2|$, and since $f_1$ and $f_2$ are integrable, so is $f$ by Lemma 305. Since $f^+ - f^- = f_1 - f_2$, $f^+ + f_2 = f^- + f_1$ and therefore

$$\int_E f^+ + \int_E f_2 = \int_E f^- + \int_E f_1.$$

Rearranging the terms (which are all finite) finishes the proof. ∎

**Lemma 307** *If* $f : E \to \mathbb{R}$ *is an integrable function, where* $E \subset \mathbb{R}^n$ *and* $A \subset E$ *are Borel sets, then* $f$ *is integrable on* $A$*.*

**Proof.** If $f$ is integrable on $E$ then each of the nonnegative functions $f^+$ and $f^-$ has finite integral on $E$ and hence on $A$. Therefore $f$ is integrable on $A$. ∎

We now verify the basic properties of the integral. We will list them here and verify them below. Of course positivity no longer holds and normalization only concerns a nonnegative function, which leaves three of the original five.

1. (Linearity) If $f$ and $g$ are integrable on $A \subset \mathbb{R}^n$ then for any $c \in \mathbb{R}$, $cf + g$ is integrable and $\int_A (cf + g) = c\int_A f + \int_A g$.

2. (Countable Set Additivity) If $E = \bigcup_{i=1}^{\infty} E_i$ where $\{E_i\}_{i=1}^{\infty}$ is a pairwise disjoint collection of Borel sets and $f$ is integrable on $E$ then $\int_E f = \sum_{i=1}^{\infty} \int_{E_i} f$.

3. (Translation Invariance) If $f$ is integrable on $A \subset \mathbb{R}^n$ and $h(\mathbf{x}) = f(\mathbf{x} - \mathbf{v})$, which is defined on $\mathbf{v} + A$, then $h$ is integrable and $\int_A f = \int_{\mathbf{v}+A} h$.

**Proof.** We will prove linearity as two separate parts, proving $\int_A cf = c\int_A f$ first. If $c = 0$ the proof is trivial. If $c > 0$ then $(cf)^+ = \max\{0, cf\} = c\max\{0, f\} = cf^+$ and likewise $(cf)^- = cf^-$. We now compute

$$\int_A cf = \int_A (cf)^+ - (cf)^- = \int_A cf^+ - \int_A cf^- = c\left(\int_A f^+ - \int_A f^-\right) = c\int_A f.$$

The proof for $c < 0$ is an exercise. Since

$$f + g = (f^+ + g^+) - (f^- + g^-)$$

Lemma 306 implies that $f + g$ is integrable and

$$\int_A f + g = \int_A (f^+ + g^+) - \int_A (f^- + g^-)$$

$$= \int_A f^+ - \int_A f^- + \int_A g^+ - \int_A g^- = \int_A f + \int_A g.$$

The rest of the the proof is an exercise. ■

**Exercise 194** *Finish the above proof.*

**Lemma 308** *If $f$ and $g$ are Borel functions defined on a set $E \subset \mathbb{R}^n$ such that $f(x) \le g(x)$ for all $x \in E$ and $\int_E f$ and $\int_E g$ exist then $\int_E f \le \int_E g$.*

**Lemma 309** *If $f$ is a Borel function defined on a Borel set $E \subset \mathbb{R}^n$ of measure 0 then $\int_E f$ exists and is equal to 0.*

**Exercise 195** *Prove the above two lemmas.*

**Exercise 196** *Suppose that $f$ and $g$ are integrable Borel functions such that $f \ge g$ a.e. on a Borel set $E$.*

1. *Prove that if $\int_E f = \int_E g$ then $f = g$ a.e. on $E$.*

2. *Show by example that the statement is false if one removes the requirement that $f \ge g$ a.e.*

**Theorem 310** *(General Lebesgue Dominated Convergence Theorem) Let $f_i : E \to \mathbb{R}$ be a sequence of Borel functions such that $f_i \xrightarrow{p} f$ for some $f : E \to \mathbb{R}$, where $E$ is a Borel subset of $\mathbb{R}^n$. If there exists a nonnegative integrable function $g : E \to \mathbb{R}$ such that $|f_i(x)| \le g(x)$ for all $i$ and $x \in E$, then*

$$\int_E f = \lim \int_E f_i.$$

**Proof.** Note that

$$0 \leq g - f_i \leq g - (-g) = 2g$$

and $(g - f_i) \xrightarrow{p} g - f$. The Lebesgue Dominated Convergence Theorem implies

$$\int_E g - \lim \int_E f_i = \lim \int_E (g - f_i) = \int_E g - \int_E f$$

and the proof is finished since all the integrals are finite for large $i$. ■

**Theorem 311** *(Fubini's Theorem) If $E \subset \mathbb{R}^n \times \mathbb{R}^m$ is compact and $f : E \to \mathbb{R}$ is a continuous function then $I(x) := \int_{E_{\mathbf{x}}} f_{\mathbf{x}}$ is a Borel function and*

$$\int_E f = \int_{E_1} \left( \int_{E_{\mathbf{x}}} f_{\mathbf{x}} \right).$$

**Proof.** Since $f$ is continuous on a compact set $f$ is bounded below by some real number $M$ and $g := f - M$ is nonnegative and continuous. One can easily check that $f_x = g_x + M$ and so $f_x$ is Borel and integrable. We can apply the previous version of Fubini's Theorem to $g$:

$$\int_E g = \int_{E_1} \left( \int_{E_{\mathbf{x}}} g_{\mathbf{x}} \right) = \int_{E_1} \left( \int_{E_{\mathbf{x}}} f_x - M \right)$$

$$= \int_{E_1} \left( \int_{E_{\mathbf{x}}} f_{\mathbf{x}} \right) - M \int_{E_1} \left( \int_{E_{\mathbf{x}}} 1 \right) = \int_{E_1} \left( \int_{E_{\mathbf{x}}} f_{\mathbf{x}} \right) - M\mu(E).$$

On the other hand,

$$\int_E g = \int_E (f - M) = \int_E f - M\mu(E).$$

■

**Proposition 312** *Let $f : U \to \mathbb{R}$ be an integrable Borel function defined on an open set $U \subset \mathbb{R}^n$, and suppose that $C \subset U$ is compact. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that $N(C, \delta) \subset U$ and $\left| \int_{N(C,\delta)} f - \int_C f \right| < \varepsilon$.*

**Exercise 197** *Prove Proposition 312.*

At this point we would like to describe the most direct method for showing the existence of the Lebesgue integral. One can start with the conclusion of Theorem 274, defining the measure of an arbitrary set in $\mathbb{R}^n$ to be the infimum of sums $\sum_{i=1}^k \mu(Q_i)$ such that each $Q_i$ is a cube and $C \subset \bigcup_{i=1}^k Q_i$. An immediate problem is that this measure is not countably additive for arbitrary subsets of $\mathbb{R}^n$; to get countable additivity one must restrict consideration to a collection of sets for which countably additivity is valid, such as Borel sets or measurable sets that were defined earlier. This measure allows one to define the integral of simple functions. One then defines the integral of an arbitrary nonnegative Borel function $f$ (or more generally *measurable* function, that is, inverse images of open sets are measurable) to be the supremum of the integrals of all simple functions less than or equal to $f$ (see Corollary 293). The extension to arbitrary Borel or measurable functions is then exactly the same as what we did this section. While this whole procedure seems straightforward, the details of the proofs of the various properties are somewhat involved. In some graduate courses this presentation is avoided and the Lebesgue integral appears somewhat mysteriously as an application of a theorem known as the Riesz Representation Theorem to the Riemann integral; one purpose of this chapter has been to facilitate the transition to that more sophisticated perspective.

We conclude this section with a brief discussion of integrals of complex valued functions. If $E \subset \mathbb{C}$ is a Borel set and $f : E \to \mathbb{C}$ is a function, we say that $f$ is integrable if and only if the real and imaginary parts of $f$ are integrable. More specifically, letting $f(z) = g(z) + ih(z)$ where $g$ and $h$ are real function, $f$ is said to be integrable if and only if $g$ and $h$ are integrable (as real valued functions defined on $\mathbb{C} = \mathbb{R}^2$) and we define $\int_E f$ to be the complex number $\int_E g + i \int_E h$. Using methods similar to what we used in this chapter one can prove linearity, countable set additivity, and translation invariance, plus a version of the Lebesgue Dominated Convergence Theorem for complex integrals.

## 4.9 L$^p$ Spaces

The purpose of this section is to introduce the most important and well-known infinite dimensional normed vector spaces, the $L^p$ spaces. We saw in Theorem 139 that all norms on $\mathbb{R}^n$ are bilipschitz equivalent and hence, since the standard norm on $\mathbb{R}^n$ is complete, it follows that all norms on $\mathbb{R}^n$ are complete. The same cannot be said for norms on infinite dimensional spaces, although we will not discuss specific examples here. Completeness is important enough a property that complete normed vector spaces have their own name.

**Definition 313** *A Banach space is a vector space $V$ with a norm such that the metric induced by the norm is complete.*

A theorem from linear algebra tells us that every finite dimensional (real) vector space is isomorphic to $\mathbb{R}^n$ for some $n$ and hence we may conclude that all normed vector spaces of dimension $n$ are Banach spaces that are topologically equivalent and isometric to $\mathbb{R}^n$ with some norm. The most basic infinite dimensional Banach spaces are defined as follows:

**Definition 314** *Let $E$ be a nonempty Borel subset of $\mathbb{R}^n$ and $1 \leq p < \infty$. We define $L^p(E)$ to be the set of all functions $f : E \to \mathbb{R}$ such that $\int_E |f|^p < \infty$. None of the results in this section depends on the particular properties of $E$, and hence we will simply denote the space by $L^p$, as is customary.*

**Exercise 198** *Prove using theorems of elementary calculus that $x^{-\frac{1}{q}} \in L^p((0,1))$ if $p > q$, but $x^{-\frac{1}{q}} \notin L^q((0,1))$.*

Note that $L^p$ is a vector space. In fact, the set of all real valued functions on $E$ is easily checked to be a vector space with respect to the usual adding and scalar multiplication of vectors. Therefore we need only check closure under scalar multiplication and addition. Certainly if $f \in L^p$ then $cf \in L^p$ for any $c \in \mathbb{R}$. If $f, g \in L^p$ we have for any $x \in E$,

$$\begin{aligned} |f(x)+g(x)|^p &\leq (|f(x)|+|g(x)|)^p \leq (2\max\{|f(x)|,|g(x)|\})^p \\ &= 2^p \max\{|f(x)|^p,|g(x)|^p\} \leq 2^p(|f(x)|^p+|g(x)|^p) \end{aligned}$$

and therefore

$$\int_E |f+g|^p \leq 2^p\left(\int_E |f|^p + \int_E |g|^p\right) < \infty.$$

We define the *norm* of $f \in L^p$ by

$$\|f\|_p := \left(\int_E |f|^p\right)^{\frac{1}{p}}. \tag{4.3}$$

We will verify that this is a norm–almost! It turns out that this "norm" is not positive definite, and we will have to deal with this later. For now, observe that certainly for any constant $c$, $\|cf\|_p = |c|\,\|f\|_p$. The triangle inequality (also known, in this context, as the Minkowski Inequality), will follow from some other inequalities that are important in their own right.

**Lemma 315** *(Young's Inequality) Let $f : [0,\infty) \to [0,\infty)$ be a continuous, one-to-one function such that $f(0) = 0$. Then for any $a, b > 0$ such that $b$ lies in the range of $f$,*

$$ab \leq \int_0^a f + \int_0^b f^{-1}$$

*and equality holds if and only if $b = f(a)$.*

**Proof.** Let $A = A(f\mid_{[0,a]})$ and $B = A(f^{-1}\mid_{[0,b]})$ (see Definition 298). Note that $A \cap B$ is the graph of $f$, which by Exercise 188 has measure 0. Therefore

$$\mu(A \cup B) = \mu(A) + \mu(B) = \int_0^a f + \int_0^b f^{-1}.$$

Since $f(0) = 0$ and $f$ is nonnegative, $f$ must be strictly increasing by Corollary 272. Let $R := [0,a] \times [0,b]$, which has measure $ab$. If $(x,y) \in R$, then $y \leq f(x)$ or $y \geq f(x)$. If $y \leq f(x)$ then $(x,y) \in A$. If $y \geq f(x)$ then $f^{-1}(y) \geq x$ and $(x,y) \in B$. In other words, $R \subset A \cup B$ and we have $ab \leq \int_0^a f + \int_0^b f^{-1}$. Suppose $a = f(b)$. Since $f$ is increasing, if $(x,y) \in A$ then $y = f(x) \leq f(a) = b$ and so $(x,y) \in R$. Likewise $B \subset R$ and therefore $A \cup B \subset R$ and the opposite inequality holds. If $b > f(a)$ then the region $A(f\mid_{(a,f^{-1}(b)]})$ is contained in $R$, but not in $A$. This region has positive measure since $f$ is continuous and positive. That is, $ab > \int_0^a f + \int_0^b f^{-1}$. By Lemma 105, $f^{-1}$ is continuous and a similar argument shows the opposite strict inequality if $b < f(a)$. ■

**Exercise 199** *Illustrate with a graph the various cases in the proof of Young's Inequality; this "proof by picture" is quite simple and clear if the integral is interpreted as area.*

**Corollary 316** *Let* $a, b \geq 0$ *and* $1 < p, p' < \infty$ *be such that* $\frac{1}{p} + \frac{1}{p'} = 1$. *Then*

$$ab \leq \frac{a^p}{p} + \frac{b^{p'}}{p'}.$$

**Exercise 200** *Prove Corollary 316. Hint: Use* $f(x) = x^r$ *for* $r > 0$ *in Young's Inequality and then let* $r = p - 1$. *You may use theorems from elementary calculus to do the necessary integrals.*

**Theorem 317** *(Hölder Inequality) Suppose that* $1 < p, p' < \infty$ *are such that* $\frac{1}{p} + \frac{1}{p'} = 1$ *and let* $f, g$ *be Borel functions defined on a Borel set* $E \subset \mathbb{R}^n$ *such that* $|f|^p$ *and* $|g|^{p'}$ *are integrable. Then*

$$\int_E |fg| \leq \|f\|_p \|g\|_{p'} .$$

**Proof.** If $\|f\|_p = 0$ then by Lemma 273 $|f|^p = 0$ a.e. and hence $|f| = 0$ a.e. This implies that $|fg| = 0$ a.e and hence $\int_E |fg| = 0$ and the inequality must be true. A similar proof covers the case when $\|g\|_{p'} = 0$, so we may assume both are nonzero. $h := \frac{f}{\|f\|_p}$ and $k := \frac{g}{\|g\|_{p'}}$. Now

$$\int_E |fg| = \left( \|f\|_p \|g\|_{p'} \right) \int_E |hk| .$$

Corollary 316 implies that

$$\int_E |hk| = \int_E |h| |k| \leq \int_E \left( \frac{|h|^p}{p} + \frac{|k|^{p'}}{p'} \right)$$

$$= \frac{1}{\|f\|_p^p} \int_E \frac{|f|^p}{p} + \frac{1}{\|g\|_{p'}^{p'}} \int_E \frac{|g|^{p'}}{p'} = \frac{1}{p} + \frac{1}{p'} = 1.$$

■

In the special case $p = p' = 2$ we have:

**Corollary 318** *(Cauchy-Schwarz Inequality) Let* $f, g$ *be Borel functions defined on a Borel set* $E \subset \mathbb{R}$ *such that* $f^2$ *and* $g^2$ *are integrable. Then*

$$\int_E |fg| \leq \|f\|_2 \|g\|_2 .$$

**Exercise 201** *Prove the Minkowski Inequality for* $p = 1$.

For $p > 1$,

$$\|f+g\|_p^p = \int_E |f+g|^p = \int_E |f+g|\,|f+g|^{p-1} \le \int_E |f|\,|f+g|^{p-1} + \int_E |g|\,|f+g|^{p-1}$$

Applying the Hölder Inequality with $p' = \frac{p}{p-1}$ gives us

$$\int_E |f|\,|f+g|^{p-1} \le \left(\int_E |f|^p\right)^{\frac{1}{p}} \left(\int_E |f+g|^p\right)^{\frac{p-1}{p}} = \|f\|_p \, \|f+g\|_p^{p-1}.$$

Doing a similar calculation with $\int_E |g|\,|f+g|^{p-1}$ and combining the results yields

$$\|f+g\|_p^p \le \left(\|f\|_p + \|g\|_p\right) \|f+g\|_p^{p-1}$$

and the triangle inequality follows by dividing each side by $\|f+g\|_p^{p-1}$.

Do we now have a norm? Unfortunately there is a little problem. Recall that a norm must be postive definite, but, for example, if $f_\delta$ is the Dirichlet function then by Corollary 270 we know that $\|f_\delta\|_1 = 0$, but $f_\delta$ is certainly not the 0 function. In fact, if $f$ any function that is 0 a.e. then $|f|^p$ has the same property and hence $\|f\|_p = 0$ for all $p$. Not to be deterred, we define two functions $f$ and $g$ to be equivalent if $f = g$ a.e. Rather than considering functions, we consider equivalence classes of functions, and these equivalence classes form a vector space, also referred to as $L^p$, with a norm naturally induced by $\|\cdot\|_p$. For example, if $[f]$ denotes the equivalence class of $f$ in $L^p$, we define scalar multiplication by $c\,[f] = [cf]$. We need to check that this scalar multiplication is well-defined. Suppose that $[f] = [g]$. Then $f(x) = g(x)$ for almost every $x$ and therefore $cf(x) = cg(x)$ for almost every $x$. That is, $cf$ is equivalent to $cg$, i.e., $[cf] = [cg]$. The remaining details are an exercise. Although it is necessary to use equivalence classes in order for $L^p$ to be formally a normed vector space, in practice no notation for equivalence classes is used and customary to consider $L^p$ as a normed vector space but still treat its elements as functions. On the other hand, since we really are dealing with equivalence classes we may in any proof concerning $L^p$ spaces replace a given function by a (presumably nicer or more convenient) function that differs from the original on a set of measure 0.

**Exercise 202** *Define the sum and norm of equivalence classes of $L^p$ and show that each of these is well-defined. Check the distributive law; the proofs of the other axioms of a vector space are similar.*

**Definition 319** *If $E$ is a Borel subset of $\mathbb{R}^n$ and $f_i, f : E \to \mathbb{R}$ are Borel functions, we say $f_i \xrightarrow{p} f$ a.e. if there is some $F \subset E$ such that $\mu(F) = 0$ and $f_i \xrightarrow{p} f$ on $E \backslash F$. We will write $f_i \to f$ in $L^p$ to mean that $f_i$ converges to $f$ in the metric space $L^p$.*

**Exercise 203** *Show that if $f_i \xrightarrow{p} f$ a.e., where the functions $f_i$ and $f$ are all in $L^p$, then there exist functions $f_i', f' \in L^p$ such that*

1. $f_i' \xrightarrow{p} f'$ *and*
2. $f_i \to f$ *in* $L^p$ *if and only if* $f_i' \to f'$ *in* $L^p$.

**Exercise 204** *Let* $f_i \to f$ *in* $L^p$ *and* $\varepsilon > 0$, *and define* $E_i := \{\mathbf{x} \in E : |f_i(\mathbf{x}) - f(\mathbf{x})| > \varepsilon\}$. *Show that* $\lim_{i\to\infty} \mu(E_i) = 0$.

**Exercise 205** *Let* $1 \le p < q < \infty$. *Show that if* $\mu(E) < \infty$ *then* $L^q(E) \subset L^p(E)$. *Hint: Consider* $F := \{\mathbf{x} : f(x) \ge 1\}$ *and* $E \backslash F$.

**Lemma 320** *If* $(f_i)$ *is a Cauchy sequence in* $L^p$ *and* $f_i \xrightarrow{p} f$ *a.e. then* $f \in L^p$ *and* $f_i \to f$ *in* $L^p$.

**Proof.** By Exercise 203, without loss of generality we may assume $f_i \xrightarrow{p} f$. Let $\varepsilon > 0$. There is some $j$ such that if $i, k \ge j$,

$$\|f_k - f_i\|_p = \left(\int_E |f_k - f_i|^p\right)^{\frac{1}{p}} < \varepsilon.$$

Fixing $k = j$, an application of Fatou's Lemma gives

$$\int_E |f_j - f|^p \le \liminf_{i\to\infty} \int_E |f_k - f_i|^p \le \varepsilon^p.$$

Then $f_j - f \in L^p$ and therefore $f = f_j - (f_j - f) \in L^p$. Moreover,

$$\|f_j - f\|_p = \left(\int_E |f_j - f|^p\right)^{\frac{1}{p}} \to 0$$

and therefore $f_j \to f$ in $L^p$. ■

**Theorem 321** *For all* $1 \le p < \infty$, $L^p$ *is a Banach space.*

**Proof.** If $(f_i)$ be is a Cauchy sequence in $L^p$ then using an iterative construction we can find $n_1 < n_2 < \cdots$ such that for every $j$ $\left\|f_{n_j} - f_{n_j+1}\right\|_p \le 2^{-j}$. Define $h_k := \sum_{j=1}^{k} \left|f_{n_j} - f_{n_{j+1}}\right|$; so $(h_k)$ is monotone increasing and nonnegative. Moreover,

$$\|h_k\|_p = \left\|\sum_{j=1}^{k} \left|f_{n_j} - f_{n_{j+1}}\right|\right\|_p \le \sum_{j=1}^{k} \left\|f_{n_j} - f_{n_j+1}\right\|_p < 1$$

for any $k$; that is, $\int_E h_k^p < 1$ for all $k$. According to Exercise 176, $(h_k)$ is pointwise convergent a.e. For any $\mathbf{x} \in E$ and $k$ we have

$$\left|\sum_{j=1}^{k} \left(f_{n_j}(\mathbf{x}) - f_{n_{j+1}}(\mathbf{x})\right)\right| \le \sum_{j=1}^{k} \left|f_{n_j}(\mathbf{x}) - f_{n_{j+1}}(\mathbf{x})\right| = h_k(\mathbf{x}).$$

That is, the telescoping series $\sum_{j=1}^{\infty}\left(f_{n_j}(\mathbf{x})-f_{n_{j+1}}(\mathbf{x})\right)$ is convergent for almost every $\mathbf{x}$. According to Exercise 130, this means that $(f_{n_j}(\mathbf{x}))$ is convergent for almost every $\mathbf{x}$. That is, $(f_{n_j})$ is pointwise convergent a.e. to some function $f$. Since $(f_{n_j})$ is a subsequence of a Cauchy sequence, hence Cauchy, Lemma 320 implies that $f \in L^p$ and $f_{n_j} \to f$ in $L^p$. The proof is now finished by Lemma 156. ■

Let's turn to the special case $p = 2$. What is special about $p = 2$ is that we may take $p' = 2$ and the (most recent) Cauchy-Schwarz inequality implies that if we define

$$\langle f, g\rangle := \int_E fg$$

then $\langle f, g\rangle$ is finite. It is an exercise below that $\langle f, g\rangle$ has the same basic properties as the Euclidean dot product. The Cauchy-Schwarz Inequality in this setting can now be restated in a form nearly identical to the inequality for Euclidean spaces:

$$|\langle f, g\rangle| \leq \|f\|_2 \, \|g\|_2 \, .$$

The space $L^2$ is referred to as separable Hilbert space. "Separable" means that there is a sequence having every element of $L^2$ as a cluster point. $\mathbb{R}^n$ is separable; simply take a sequence having as its image the (countable) collection of all points having rational coordinates. A Hilbert space is a vector space (often assumed to be infinite dimensional) with a dot product (also called an *inner product* or *positive definite, symmetric bilinear form*) that is complete with respect to the norm defined by the dot product. It turns out that any two separable Hilbert spaces are isomorphically isometric, but we will not prove this here (nor the fact that $L^2$ is separable). Since the geometry of Euclidean space is entirely determined by the dot product, any Hilbert space is geometrically essentially the same as the Euclidean spaces, except that it is infinite dimensional.

**Exercise 206** *Show that on $L^2$, $\langle\cdot,\cdot\rangle$ has the same properties of the dot product in Euclidean space, namely it is symmetric, bilinear, and positive definite–and by definition $\|f\|_2^2 = \int_E f^2 = \langle f, f\rangle$.*

**Exercise 207** *Let $f_1 := \chi_{[0,1]}$, $f_2 := (-1)\chi_{[0,\frac{1}{2})} + \chi_{[\frac{1}{2},1]}$.*

1. *Show that $f_1$ and $f_2$ are orthonormal in $L^2([0,1])$; that is, $\langle f_i, f_j\rangle = 0$ if $i \neq j$ and $\langle f_i, f_j\rangle = 1$ if $i = j$.*

2. *Show how to continue this process to construct a sequence $(f_i)$ of orthonormal functions. (Drawing the graphs of the next couple of functions will suffice.)*

3. *Show that for any $i \neq j$, $\|f_i - f_j\|_2 = \sqrt{2}$.*

4. *Show that the Heine-Borel Theorem fails for $L^2$.*

**Exercise 208** *The $L^p$ spaces have finite dimensional analogs: For $(v_1, ..., v_n) \in \mathbb{R}^n$ and $1 < p < \infty$, define*

$$\|(v_1, ..., v_n)\|_p := \left( \sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}}.$$

*Verify a version of the Hölder Inequality for this norm. The proof of the triangle inequality, also called the Minkowski Inequality in this case, is very similar to the proof for $L^p$; you do not need to write the details.*

# Chapter 5

# Differentiation

## 5.1 A Little Linear Algebra

Differentiation is the process of approximating certain reasonable functions between Euclidean spaces using linear functions. We will first establish our notation for basic linear algebra. We denote the *standard basis vectors* by

$$\mathbf{e}_i := (0, ..., 0, 1, 0, ..., 0)$$

where the "1" is the $i^{th}$ coordinate. A vector $\mathbf{v} = (v_1, ..., v_n)$ is uniquely expressed as $\mathbf{v} = \sum_{i=1}^{n} v_i e_i$, where the numbers $v_i$ are called the components of $\mathbf{v}$. We will denote (real) matrices with bold capitals (e.g. $\mathbf{A}$) sometimes reserving for row or column vectors some of the letters near the end of the alphabet (e.g. $\mathbf{X}$). The entries of a matrix will be denoted by smaller case (not bold) characters, since they are real numbers: the entry of $\mathbf{A}$ that is in the $i^{th}$ row and $j^{th}$ column will be denoted by $a_{ij}$ or $(\mathbf{A})_{ij}$. For example the product of matrices may be expressed as

$$(\mathbf{AB})_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

If $\mathbf{A}$ is and $n \times m$ matrix and $\mathbf{X} := (x_1, ..., x_m) \in \mathbb{R}^m$ then the expression $\mathbf{Y} = \mathbf{AX}$ describes a function from $\mathbb{R}^m$ to $\mathbb{R}^n$ given by matrix multiplication. Explicitly,

$$\mathbf{A}(x_1, ..., x_m) = \sum_{i=1}^{n} \sum_{k=1}^{m} a_{ik} x_k \mathbf{e}_i$$

In particular,

$$\mathbf{A}\mathbf{e}_j = \sum_{i=1}^{n} a_{ij} \mathbf{e}_i \text{ or } (\mathbf{A}\mathbf{e}_j)_i = a_{ij}.$$

This function has continuous components and hence is continuous. Moreover this function is linear in the sense that for every $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $t \in \mathbb{R}$ we have $\mathbf{A}(\mathbf{v} + t\mathbf{w}) = \mathbf{A}(\mathbf{v}) + t\mathbf{A}(\mathbf{w})$. Conversely, if $f : \mathbb{R}^n \to \mathbb{R}^m$ is linear then there is some matrix $\mathbf{A}$ such that $f(\mathbf{v}) = \mathbf{A}\mathbf{v}$ for all $\mathbf{v}$. In fact, the $j^{th}$ column of $\mathbf{A}$ is given by $f(\mathbf{e}_j)$. For our purposes it is convenient to simply use matrices and linear functions interchangeably, and so we will use the same kind of notation for linear functions that we do for matrices.

**Exercise 209** *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a function. Show that $f$ is linear if and only if each component of $f$ is linear. Hint: Using matrices is not the easiest way to do this.*

Note that if $\mathbf{M}$ is an $m \times n$ matrix and $\mathbf{N}$ is an $n \times k$ matrix then the product matrix $\mathbf{MN}$ represents the composition of the linear functions corresponding to $\mathbf{M}$ and $\mathbf{N}$. We will denote the determinant of a square matrix $\mathbf{M}$ by $\det \mathbf{M}$. An identity matrix $\mathbf{I}$ is a matrix having a 1 in each diagonal entry and 0 in each other entry; the $n \times n$ identity matrix represents the identity function on $\mathbb{R}^n$. We will need the facts that $\det \mathbf{MN} = \det \mathbf{M} \det \mathbf{N}$, and $\det \mathbf{M} = 0$ if and only if $\mathbf{M}$ represents an linear isomorphism of $\mathbb{R}^n$, i.e. a bijective linear function. In this case $\mathbf{M}$ (and the corresponding linear function) are called non-singular. In general, a linear function $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ is injective if and only if $\mathbf{L}^{-1}(\{\mathbf{0}\}) = \{\mathbf{0}\}$. If $n = m$ then the following are equivalent: $\mathbf{L}$ is injective; $\mathbf{L}$ is surjective; $\mathbf{L}$ is an isomorphism.

The next proposition is, in a sense, a transition point between linear algebra and analysis.

**Proposition 322** *If $\mathbf{L} : \mathbb{R}^m \to \mathbb{R}^n$ is linear then $\mathbf{L}$ is Lipschitz.*

**Proof.** Since $\mathbf{L}$ is continuous and $\mathbb{S}^{m-1}$ is compact (Exercise 97) there exists some $M > 0$ such that $\mathbf{L}(\mathbf{y}) \leq M$ for all $\mathbf{y} \in \mathbb{S}^{m-1}$. Now for any $\mathbf{x} \neq \mathbf{y}$ in $\mathbb{R}^n$, $\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|} \in \mathbb{S}^{m-1}$ and

$$\|\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{y})\| = \|\mathbf{L}(\mathbf{x} - \mathbf{y})\|$$

$$= \|\mathbf{x} - \mathbf{y}\| \left\| \mathbf{L}\left( \frac{\mathbf{x} - \mathbf{y}}{\|\mathbf{x} - \mathbf{y}\|} \right) \right\| \leq \|\mathbf{x} - \mathbf{y}\| M.$$

■

**Exercise 210** *Suppose that $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. For any $f \in L^q$, define $\phi : L^p \to \mathbb{R}$ by $\phi(g) := \int_E fg$.*

1. *Prove that $\phi$ is linear.*
2. *Prove that $\phi$ is Lipschitz. Hint: You cannot mimic the proof of Proposition 322 (cf. Exercise 207).*

The converse of the above exercise, i.e. that every "bounded linear functional" on $L^p$ is of the form $\phi(g) = \int_E fg$ for some $f \in L^q$, is an important theorem known as the Riesz Representation Theorem. The proof of this theorem requires topics that were not covered in the previous chapter.

## 5.2 Derivatives

In elementary calculus one learns that the derivative of a real function $f$ at the point $x_0$ is given by the formula

$$f'(x_0) := \lim_{h \to 0} \frac{f(x_0 + h) - f(x)}{h}$$

provided this limit exists. The function is assumed to be defined in an open interval about $x_0$ so that $x_0$ may be "approached from both sides." The limit is, of course, a real number; this formula therefore defines a new real function called the derivative, denoted by $f'$, defined wherever the limit exists. There are various problems with directly generalizing this definition to the case when more than one variable is involved, including the fact that one cannot divide by vectors, so the quotient makes no sense when the denominator is a vector $\mathbf{h}$. To generalize the notion of differentiability we must therefore look at an equivalent notion–also learned in elementary calculus–that of the "linear approximation" of a differentiable function. Recall that if $f$ is differentiable at $x_0$ then the linear approximation is given by $L(x) := f'(x_0)(x - x_0) + f(x_0)$, the graph of which is the tangent line to the graph of $f$ at $x_0$. The linear approximation provides a "first order" approximation to $f$ at $x_0$ in the following sense:

$$\lim_{x \to x_0} \frac{f(x) - L(x)}{x - x_0} = \lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} \tag{5.1}$$

$$= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) = 0.$$

More generally, a function $f$ is an $n^{th}$*-order approximation* of a function $g$ at $x_0$ if $\lim_{x \to x_0} \frac{f(x)-g(x)}{(x-x_0)^n} = 0$, where $n$ is a nonnegative integer. The idea is that as $x \to x_0$, $(x - x_0)^n \to 0$ for any $n \in \mathbb{N}$, and for larger $n$ the convergence is more rapid. In order for $\frac{f(x)-g(x)}{(x-x_0)^n}$ to converge to 0, the numerator must converge "more rapidly" than the denominator, so the order of the approximation is a measure of how well $g$ approximates $f$ near $x_0$, with larger $n$ implying a better approximation. If $f$ and $g$ are continuous then $g$ is a $0^{th}$-order approximation of $f$ if and only if $f(x_0) = g(x_0)$.

**Exercise 211** *To what best possible order do the following functions approximate $f(x) = x^3$ at $x = 0$? (You may use any theorems from elementary calculus.)*

1. $g(x) = 0$

2. $g(x) = x^2$

3. $g(x) = \cos x - 1$

In order to extend this idea to functions between higher dimensional Euclidean spaces one needs to replace the $x - x_0$ in the denominator by $\|\mathbf{x} - \mathbf{x}_0\|$ to obtain the formula

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{f(\mathbf{x}) - g(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}_0\|^n} = 0$$

for an $n^{th}$ order approximation at $\mathbf{x}_0$. We will say that a function between higher dimensional spaces is differentiable at a point if it has a linear first-order approximation at the point. Note that the function $L$ from elementary calculus described above is not actually a linear function (due to the constant term). Therefore the following definition is slightly reformulated from Formula (**??**):

**Definition 323** *Let $f : U \to \mathbb{R}^n$ be a function, where $U \subset \mathbb{R}^m$ is open. We say that $f$ is differentiable at $\mathbf{x}_0 \in U$ if there exists a linear function $\mathbf{L} : \mathbb{R}^m \to \mathbb{R}^n$ such that*

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{f(\mathbf{x}) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0) - f(\mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \mathbf{0}. \tag{5.2}$$

*If $A$ is any subset of $U$ we say that $f$ is differentiable on $A$ if $f$ is differentiable at every $\mathbf{x}_0 \in A$.*

We require that $U$ be open so that the above limit involves all points in a small ball about $\mathbf{x}_0$. Roughly speaking we want to be able to approach $\mathbf{x}_0$ "from all directions." Since $\mathbf{L}(\mathbf{0}) = \mathbf{0}$, we subtract the term $f(\mathbf{x}_0)$ so that $\mathbf{L}(\mathbf{x} - \mathbf{x}_0) - f(\mathbf{x}_0) = f(\mathbf{x}_0)$ when $\mathbf{x} = \mathbf{x}_0$.

Note that Formula (5.2) is equivalent to the real valued limit

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0) - f(\mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

It is an exercise to show that this limit is also equivalent to the following limits, which we will use frequently:

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{x}_0+\mathbf{h}) - \mathbf{L}\mathbf{h} - f(\mathbf{x}_0)}{\|\mathbf{h}\|} = \mathbf{0} \text{ and } \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|f(\mathbf{x}_0+\mathbf{h}) - \mathbf{L}\mathbf{h} - f(\mathbf{x}_0)\|}{\|\mathbf{h}\|} = 0. \tag{5.3}$$

**Exercise 212** *Prove that $f$ is differentiable at $\mathbf{x}_0$ if and only if either (hence both) of the equations in (5.3) is true. Hint: vague references to "substitution" won't do.*

We will now show that there is a good way to compute the linear approximation $\mathbf{L}$ (provided it exists) and that it is unique.

**Definition 324** *Let $f : U \to \mathbb{R}$ be a function, where $U \subset \mathbb{R}^m$ is open and $\mathbf{x}_0 = (x_1, ..., x_m) \in U$. For any $1 \le j \le m$, define the $j^{th}$ partial derivative of $f$ at $\mathbf{x}_0$ by*

$$\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0)}{t}$$

*provided the limit exists. In case $m = 1$, i.e. if $f$ is a real function, we will use the notations $f'(x_0)$ and $\frac{df}{dx}(x_0)$ and simply refer to the (only) partial derivative at $\mathbf{x}_0$ as "the" derivative. When $m = 2$ the partial derivatives are sometimes denoted by $\frac{\partial f}{\partial x} = f_x$ and $\frac{\partial f}{\partial y} = f_y$, with similar notation using $x, y, z$ when $m = 3$.*

**Exercise 213** *Let $f : U \to \mathbb{R}$ be a function, where $U \subset \mathbb{R}^m$ is open and $\mathbf{x}_0 = (x_1, ..., x_m) \in U$. For any $1 \leq j \leq m$, define*

$$h_j(t) := f(x_1, ..., x_{j-1}, x_j + t, x_{j+1}, ..., x_m)$$

*Prove that $\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = h_j'(0)$.*

In the next proof, and frequently in future proofs, we will replace phrases of the sort "there exists a $\delta > 0$ such that if $a < \delta$ then" by "for all sufficiently small $a$" or "for all $a$ sufficiently close to 0." This is completely analogous to our use of "for all large $n$" to replace "there exists and $N$ such that for all $n \geq N$" and is done for similar reasons, namely to shorten statements and to avoid having to choose several $\delta$'s and then choose the minimum of all those $\delta$'s. As an example, note that the definition of $\lim_{x \to x_0} f(x) = c$ for a real function $f$ may be stated as: "for all $\varepsilon > 0$, $|f(x) - c| < \varepsilon$ whenever $|x - x_0|$ is sufficiently small".

**Theorem 325** *Let $f : U \to \mathbb{R}^n$ be a function, where $U \subset \mathbb{R}^m$ is open and let $f_i : U \to \mathbb{R}$ be the components of $f$. If $f$ is differentiable at $\mathbf{x}_0 \in U$ and $\mathbf{L}$ is a linear function such that*

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(\mathbf{x}_0 + \mathbf{h}) - \mathbf{L}\mathbf{h} - f(\mathbf{x}_0)}{\|\mathbf{h}\|} = \mathbf{0}$$

*then for all $i$ and $j$ the partial derivative*

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)$$

*exists and is equal to $l_{ij}$. In particular, the linear function $\mathbf{L}$ is unique.*

**Proof.** If the above limit exists then the corresponding limits for the components exists, i.e., for every $\delta > 0$ and $i$, if $\mathbf{h}$ is sufficiently close to $\mathbf{0}$,

$$\frac{|f_i(\mathbf{x}_0 + \mathbf{h}) - (\mathbf{L}\mathbf{h})_i - f_i(\mathbf{x}_0)|}{\|\mathbf{h}\|} < \delta.$$

In particular, if for any $j$ we choose $\mathbf{h} := t\mathbf{e}_j$ then $\|\mathbf{h}\| = \|t\mathbf{e}_j\| = |t|$ and

$$(\mathbf{L}\mathbf{h})_i = (\mathbf{L}(t\mathbf{e}_j))_i = tl_{ij}.$$

Therefore if $t$ is sufficiently close to 0,

$$\left| \frac{f_i(\mathbf{x}_0 + t\mathbf{e}_j) - f_i(\mathbf{x}_0)}{t} - l_{ij} \right| = \left| \frac{f_i(\mathbf{x}_0 + t\mathbf{e}_j) - f_i(\mathbf{x}_0) - tl_{ij}}{t} \right|$$

$$= \frac{|f_i(\mathbf{x}_0 + \mathbf{h}) - \mathbf{L}(\mathbf{h})_i - f_i(\mathbf{x}_0)|}{\|\mathbf{h}\|} < \delta.$$

■

The matrix having components

$$l_{ij} = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)\right)$$

(or the linear function corresponding to it) is variously referred to as the *derivative* or *total derivative* or *Jacobian matrix* of $f$ at $\mathbf{x}_0$, and is written as $D(f)(\mathbf{x}_0)$ or $Df(\mathbf{x}_0)$ or $f'(\mathbf{x}_0)$. We will use the notations $Df$ to try to avoid potential confusion arising from the fact that the derivative is not a function from a subset of $\mathbb{R}^m$ into $\mathbb{R}^n$ (except when $m = 1$, in which case $Df$ is a column vector, hence an element of $\mathbb{R}^n$). In fact, $Df$ is a function that assigns a matrix (or linear function) to those points in $\mathbb{R}^m$ where $f$ is differentiable. That is, $Df$ is a function from a subset of $\mathbb{R}^m$ into a set of matrices, which can itself be regarded as a subset of $\mathbb{R}^{mn}$. However, studying the derivative from this standpoint is beyond the scope of this text. Since it is unique, one can occasionally prove differentiability and identify the derivative simply by checking that the "candidate" linear function satisfies the role of $\mathbf{L}$ in the definition. For example:

**Proposition 326** *If $\mathbf{L} : \mathbb{R}^m \to \mathbb{R}^n$ is linear then $\mathbf{L}$ is differentiable at every point $\mathbf{x}_0 \in \mathbb{R}^m$ and $D\mathbf{L}(\mathbf{x}_0) = \mathbf{L}$.*

**Proof.** For any $\mathbf{x}_0$ we have

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0) - \mathbf{L}(\mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_0) - \mathbf{L}(\mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \mathbf{0}.$$

■

**Exercise 214** *Prove that if $f : \mathbb{R}^m \to \mathbb{R}^n$ is constant then $f$ is differentiable and $Df(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^m$.*

We have seen lemmas like the next one already in the context of limits and continuity; it will be useful on a number of occasions.

**Lemma 327** *Let $f : U \to \mathbb{R}^n$ be a function, where $U \subset \mathbb{R}^m$ is open. Then $f$ is differentiable at $\mathbf{x}_0 \in U$ if and only if each component $f_i : U \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$. Moreover, $(Df)(\mathbf{x}_0)_i = D(f_i)(\mathbf{x}_0)$; that is, the components of the derivative of $f$ are the derivatives of the components of $f$.*

**Proof.** According to Propositions 130 and 65, the limit

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0) - f(\mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

is equivalent to

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|f_i(\mathbf{x}) - (\mathbf{L}(\mathbf{x} - \mathbf{x}_0))_i - f_i(\mathbf{x}_0)|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

for all $i$. Since $(\mathbf{L}(\mathbf{x}-\mathbf{x}_0))_i = \mathbf{L}_i(\mathbf{x}-\mathbf{x}_0)$, the proof follows from uniqueness. ■

Formula (5.1) shows that a real valued function satisfies Definition 323 if and only if the derivative of the function exists at each point. Combining this with Lemma 327 gives us:

**Corollary 328** *A function $f : U \to \mathbb{R}^n$, where $U \subset \mathbb{R}$ is open, is differentiable at $x_0 \in U$ if and only if for all $i$, $\frac{df_i}{dx}(x_0)$ exists.*

**Exercise 215** *Prove that if $f(x) = \sqrt{x}$ then for $x > 0$, $f'(x) = \frac{1}{2\sqrt{x}}$.*

The next example shows that for functions of more than one variable, mere existence of partial derivatives does not imply that a function is differentiable.

**Example 329** *Let*

$$f(x,y) = \begin{cases} 0 & \text{if } (x,y) = (0,0) \\ \frac{x^3}{x^2+y^2} & \text{otherwise} \end{cases}.$$

*First note that $f$ is continuous. The only possible discontinuity of $f$ is at the point $\mathbf{0}$. But when $0 < \|(x,y)\| < \varepsilon$ we have*

$$\left|\frac{x^3}{x^2+y^2}\right| = \frac{|x|\,x^2}{x^2+y^2} \leq \frac{|x|\left(x^2+y^2\right)}{x^2+y^2} = |x| \leq \sqrt{x^2+y^2} < \varepsilon,$$

*which proves continuity at $\mathbf{0}$.*

*When $(x,y) \neq \mathbf{0}$ the partial derivatives can be computed using elementary calculus. At $\mathbf{0}$ we must compute directly. Then*

$$f_x(0,0) = \lim_{t\to 0}\frac{t^3}{t^3} = 1,\ f_y(0,0) = \lim_{t\to 0}\frac{0}{t^2} = 0.$$

*Therefore all partial derivatives of $f$ exist at all points of $\mathbb{R}^2$. On the other hand, we can check differentiability directly at $(0,0)$, knowing that the linear function $\mathbf{L}$ in question would have to be represented by the matrix $\begin{pmatrix} 1 & 0 \end{pmatrix}$. We have*

$$\lim_{(x,y)\to\mathbf{0}} \frac{f(x,y) - \begin{pmatrix} 1 & 0 \end{pmatrix}(x,y) - f(0,0)}{\sqrt{x^2+y^2}} = \lim_{(x,y)\to\mathbf{0}}\left(\frac{x^3 - x(x^2+y^2)}{(x^2+y^2)^{\frac{3}{2}}}\right)$$

$$= \lim_{(x,y)\to\mathbf{0}}\left(\frac{x^3 - x(x^2+y^2)}{(x^2+y^2)^{\frac{3}{2}}}\right) = \lim_{(x,y)\to\mathbf{0}}\left(\frac{-xy^2}{(x^2+y^2)^{\frac{3}{2}}}\right)$$

*But the latter limit does not exist. In fact if one considers any sequence $(x_i, x_i) \to \mathbf{0}$,*

$$\lim\left(\frac{-x_i^3}{(x_i^2+x_i^2)^{\frac{3}{2}}}\right) = -2^{-\frac{3}{2}}$$

*but for any sequence $(x_i, 0) \to \mathbf{0}$,*

$$\lim\left(\frac{-x_i \cdot 0}{(x_i^2+0^2)^{\frac{3}{2}}} - x_i\right) = 0.$$

**Exercise 216** *Define* $f(x,y) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{x^2 y}{x^4+y^2} & \text{otherwise} \end{cases}$. *Prove that the partial derivatives of* $f$ *exist everywhere, but that* $f$ *is not continuous. Hint: Consider a sequence that lies on the graph of* $y = x^2$.

Example 329 leaves us with no convenient way to actually check whether or not a function with more than one variable is differentiable. We will remedy this situation later when we consider continuously differentiable functions.

We conclude by recalling two special cases of derivatives. A differentiable function $\alpha : (a,b) \to \mathbb{R}^n$, $n > 1$, is called a *differentiable curve* . The derivative of $\alpha$ is denoted by $\alpha' : (a,b) \to \mathbb{R}^n$, and is called the *velocity* of the curve. The scalar $\|\alpha'\|$ is called the *speed* and the second derivative $\alpha''$ is called the *acceleration* of the curve. The third derivative is referred to by physicists as the *jerk* (think about it!).

If $f$ is a real-valued differentiable function defined on $U \subset \mathbb{R}^n$ then (at least when $n > 1$) the derivative of $f$ is called the *gradient of* $f$, denoted by $\nabla f = (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n})$.

## 5.3 Basic Differentiation Theorems

**Proposition 330** *(Linearity) Let* $f, g : U \to \mathbb{R}^n$ *be functions, where* $U \subset \mathbb{R}^m$ *is open and suppose that* $f$ *and* $g$ *are differentiable at* $\mathbf{x}_0 \in U$. *Then for any* $c \in \mathbb{R}$, $D(cf+g)(\mathbf{x}_0) = cD(f)(\mathbf{x}_0) + D(g)(\mathbf{x}_0)$.

**Exercise 217** *Use uniqueness to prove the above proposition.*

**Lemma 331** *Let* $f : U \to \mathbb{R}^n$ *be a function, where* $U \subset \mathbb{R}^m$ *is open and suppose that* $f$ *is differentiable at* $\mathbf{x}_0 \in U$. *Then for some* $r, M > 0$,

$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\| \leq M \|\mathbf{x} - \mathbf{x}_0\|$$

*for all* $x \in B(x_0, r)$.

**Proof.** We have for $\mathbf{L} := D(f)(\mathbf{x}_0)$,

$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\| \leq \|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0)\| + \|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\|.$$

For some $r > 0$, if $0 < \|\mathbf{x} - \mathbf{x}_0\| < r$ then

$$\frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} \leq 1,$$

and hence (even for $\mathbf{x} = \mathbf{x}_0$),

$$\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0)\| \leq \|\mathbf{x} - \mathbf{x}_0\|.$$

Also,

$$\|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\| \leq N \|\mathbf{x} - \mathbf{x}_0\|$$

for some $N$ and all $\mathbf{x}$ by Proposition 322. Let $M := N + 1$. ■

**Corollary 332** *Let $f : U \to \mathbb{R}^n$ be a function, where $U \subset \mathbb{R}^m$ is open. If $f$ is differentiable at $\mathbf{x}_0 \in U$ then $f$ is continuous at $\mathbf{x}_0$.*

**Proof.** Let $\varepsilon > 0$ and let $r, M > 0$ be such that $\|f(\mathbf{x}) - f(\mathbf{x}_0)\| \leq M\|\mathbf{x} - \mathbf{x}_0\|$ when $\|\mathbf{x} - \mathbf{x}_0\| < r$. If $\|\mathbf{x} - \mathbf{x}_0\| < \min\left\{\frac{\varepsilon}{M}, r\right\}$ then $\|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon$. ■

One of the most important applications of the derivative involves identification of local maxima and minima.

**Definition 333** *Let $f : A \to \mathbb{R}$ be a function defined on $A \subset \mathbb{R}^m$. The function $f$ is said to have a local minimum (resp. maximum) at a point $\mathbf{x}_0 \in A$ if there exists an open set $U$ such that $\mathbf{x}_0 \in U \subset A$ and for every $\mathbf{x} \in U$, $f(\mathbf{x}_0) \leq f(\mathbf{x})$ (resp. $f(\mathbf{x}_0) \geq f(\mathbf{x})$). If $f$ has a local minimum or local maximum at $\mathbf{x}_0$ then we say that $f$ has a local extremum at $\mathbf{x}_0$.*

Note that by definition a local extremum must be at an interior point of $A$. If a real valued function $f$ has a maximum (resp. minimum) at $\mathbf{x}_0$ in an open set $U \subset \mathbb{R}^m$ then by definition $f$ has a local maximum (resp. minimum) at $\mathbf{x}_0$.

**Theorem 334** *(First Derivative Test) If $f : A \to \mathbb{R}$ is a function defined on $A \subset \mathbb{R}^m$ and $\mathbf{x}_0$ is a local extremum of $f$ then for any $j$, if $\frac{\partial f}{\partial x_j}(\mathbf{x}_0)$ exists then $\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = 0$.*

**Proof.** Suppose that $f$ has a local maximum at $\mathbf{x}_0$; the proof for a local minimum is similar. Let $t_i < 0$ satisfy $t_i \to 0$. By assumption, for any $j$, $\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = \lim \frac{f(\mathbf{x}_0 + t_i\mathbf{e}_j) - f(\mathbf{x}_0)}{t_i}$ exists. Because $f$ has a local max at $\mathbf{x}_0$, $f(\mathbf{x}_0) \geq f(\mathbf{x}_0 + t_i\mathbf{e}_j)$ for all large $i$ and therefore (using $t_i < 0$)

$$\frac{f(\mathbf{x}_0 + \mathbf{e}_j t_i) - f(\mathbf{x}_0)}{t_i} \geq 0,$$

which implies $\frac{\partial f}{\partial x_j}(\mathbf{x}_0) \geq 0$. A similar argument using a sequence with $t_i > 0$ implies the opposite inequality. ■

**Definition 335** *Let $f : A \to \mathbb{R}$ be a function defined on $A \subset \mathbb{R}^m$. A point $\mathbf{x}_0$ in the interior of $A$ is called a critical point of $f$ if for every $j$, $\frac{\partial f}{\partial x_j}(\mathbf{x}_0)$ does not exist or is 0.*

**Corollary 336** *If $f : U \to \mathbb{R}$ is a function, where $U \subset \mathbb{R}^n$ is open, and $f$ is differentiable on $U$, then every local extremum of $f$ is a critical point.*

In elementary calculus the Chain Rule is introduced as a mechanical computation method for finding derivatives of compositions of differentiable functions. The rule is named for this mechanical process in which one starts with the "outer function", differentiates it, then moves to the "next inner most" function, multiplies by the derivative of this function, and so on–thus creating a "chain" of multiplied derivatives. More complicated rules are learned for functions of more than one variable. In fact this computational method comes from a very elegant and natural theorem, which, through force of habit, we will still refer to as the Chain Rule–although it would be more aptly named the "Composition Rule". The connection between this theorem and the computation method learned in elementary calculus is addressed in Exercise 218 below.

**Theorem 337** *(Chain Rule) Let $g : U \to \mathbb{R}^m$ and $f : V \to \mathbb{R}^n$ be functions such that $U \subset \mathbb{R}^k$ and $V \subset \mathbb{R}^m$ are open and $g(U) \subset V$. If $g$ is differentiable at $\mathbf{x}_0 \in U$ and $f$ is differentiable at $g(\mathbf{x}_0)$ then $f \circ g$ is differentiable at $\mathbf{x}_0$ and*

$$D(f \circ g)(\mathbf{x}_0) = D(f)(g(\mathbf{x}_0)) \circ D(g)(\mathbf{x}_0)$$

*In other words the derivative of the composition is the composition of the derivatives.*

**Proof.** It is sufficient to show that for every $\varepsilon > 0$,

$$\|f(g(\mathbf{x})) - [D(f)(g(\mathbf{x}_0))]\,[D(g)(\mathbf{x}_0)]\,(\mathbf{x} - \mathbf{x}_0) - f(g(\mathbf{x}_0))\| < \varepsilon \|\mathbf{x} - \mathbf{x}_0\|$$

when $\|\mathbf{x} - \mathbf{x}_0\|$ is sufficiently small. Applying Lemma 331, let $M$ be such that $\|g(\mathbf{x}) - g(\mathbf{x}_0)\| \le M \|\mathbf{x} - \mathbf{x}_0\|$ when $\|\mathbf{x} - \mathbf{x}_0\|$ is small. Since $f$ is differentiable at $g(\mathbf{x}_0)$, we already know that

$$\|f(g(\mathbf{x})) - D(f)(g(\mathbf{x}_0))(g(\mathbf{x}) - g(\mathbf{x}_0)) - f(g(\mathbf{x}_0))\| <$$
$$\frac{\varepsilon}{2M} \|g(\mathbf{x}) - g(\mathbf{x}_0)\| \le \frac{\varepsilon}{2} \|\mathbf{x} - \mathbf{x}_0\|$$

when $\|g(\mathbf{x}) - g(\mathbf{x}_0)\|$ is sufficiently small, and hence (since $g$ is continuous at $\mathbf{x}_0$) when $\|\mathbf{x} - \mathbf{x}_0\|$ is sufficiently small. By the triangle inequality it is now sufficient to show that

$$\|[D(f)(g(\mathbf{x}_0))]\,[D(g)(\mathbf{x}_0)]\,(\mathbf{x} - \mathbf{x}_0) - D(f)(g(\mathbf{x}_0))(g(\mathbf{x}) - g(\mathbf{x}_0))\| < \frac{\varepsilon}{2} \|\mathbf{x} - \mathbf{x}_0\|$$

when $\|\mathbf{x} - \mathbf{x}_0\|$ is sufficiently small. The latter inequality is equivalent (for positive $\|\mathbf{x} - \mathbf{x}_0\|$) to

$$\left\| [D(f)(g(\mathbf{x}_0))] \left[ \frac{[D(g)(\mathbf{x}_0)]\,(\mathbf{x} - \mathbf{x}_0) - (g(\mathbf{x}) - g(\mathbf{x}_0))}{\|\mathbf{x} - \mathbf{x}_0\|} \right] \right\| < \frac{\varepsilon}{2}.$$

However, if $\|\mathbf{x} - \mathbf{x}_0\| > 0$ is sufficiently small,

$$\frac{[D(g)(\mathbf{x}_0)]\,(\mathbf{x} - \mathbf{x}_0) - (g(\mathbf{x}) - g(\mathbf{x}_0))}{\|\mathbf{x} - \mathbf{x}_0\|}$$

is close to $\mathbf{0}$ and we are finished by the continuity of the linear function $D(f)(g(\mathbf{x}_0))$. ■

Since composition of linear functions corresponds to the product of the corresponding functions, the Chain Rule may be equivalently stated in terms of matrices using matrix multiplication; that is

$$D(f \circ g)(\mathbf{x}_0) = [D(f)(g(\mathbf{x}_0))]\,[D(g)(\mathbf{x}_0)]\,.$$

For real functions, matrix multiplication is simply scalar multiplication and the Chain Rule becomes the familiar

$$(f \circ g)'(x_0) = f'(g(x_0))g'(x_0).$$

**Exercise 218** *In elementary calculus the chain rule is not stated in terms of matrix multiplication, but in terms of specific rules for specific kinds of functions. For example, one might learn that if* $s$ *and* $t$ *are functions of* $u$ *and* $v$ *and* $f$ *is a function of* $s$ *and* $t$*, then* $\frac{\partial f}{\partial u} = \frac{\partial f}{\partial s}\frac{\partial s}{\partial u} + \frac{\partial f}{\partial t}\frac{\partial t}{\partial u}$*. Show that this particular formula represents a single entry in a matrix that arises from the Chain Rule applied to* $f$ *composed with a function from* $\mathbb{R}^2$ *to* $\mathbb{R}^2$*.*

**Example 338** *We will find the extrema of the function* $f(x, y) = xy^2$ *on the closed unit ball* $C(0, 1) \subset \mathbb{R}^2$*. Of course the continuous function* $f$ *does have a max and a min on this compact set. First we check in the open ball for local extrema. We have* $\frac{\partial f}{\partial x} = y^2$ *and* $\frac{\partial f}{\partial y} = 2xy$*. These are simultaneously* $0$ *precisely when* $y = 0$*, and the functional values at these points are all* $0$*. To check for extrema on the boundary we observe that the function* $g(t) = (\cos t, \sin t)$ *is a differentiable function and its restriction to* $[0, 2\pi]$ *is onto the unit circle. Therefore the maximum and minimum values of* $f$ *on the unit circle will be the maximum and minimum values of the real function* $f \circ g$ *on* $[0, 2\pi]$*. By the chain rule, since* $x = \cos t$ *and* $y = \sin t$*,*

$$(f \circ g)'(t) = \left[\sin^2 t, 2\cos t \sin t\right] \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix} = -\sin^3 t + 2\cos^2 t \sin t$$

$$= \sin t(2\cos^2 t - \sin^2 t) = \sin t(3\cos^2 t - 1).$$

*Hence the derivative is* $0$ *precisely when* $\sin t = 0$ *or* $\cos^2 t = \frac{1}{3}$*, i.e. when* $y = 0$ *or* $x = \pm\frac{1}{\sqrt{3}}$*. We have already treated the case* $y = 0$*; if* $x = \pm\frac{1}{\sqrt{3}}$ *on the unit circle then* $y = \pm\sqrt{\frac{2}{3}}$*. Among these four points* $f$ *takes on a maximum of* $\frac{2}{3\sqrt{3}}$ *and a minimum of* $-\frac{2}{3\sqrt{3}}$*, which must be the maximum and minimum of the function. What we have done here is to first check for local extrema (which might be maxima or minima) in the interior of the region and then "parameterize" the boundary to be able to apply the First Derivative Test here as well. This allows us to avoid the problem that the boundary has no interior and therefore the First Derivative Test cannot be used directly. We will discuss parameterizations further in a later section. Note that in the above computation one could have written* $h(t) = f(g(t)) = \cos t \sin^2 t$ *and use the Chain Rule as described in elementary calculus to get the same answer.*

**Exercise 219** *Find the extrema of the function* $f(x, y) = x^2 + y^2$ *on the closed, bounded region in* $\mathbb{R}^2$ *bounded by the graph of* $y = x^2$ *and* $y = x + 1$*. You may use what you learned in elementary calculus about how to compute derivatives, except that you must use the matrix form of the chain rule as was done in Example 338.*

**Definition 339** *If* $\mathbf{u} \in \mathbb{R}^n$ *is a unit vector then* $D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u}$ *is called the directional derivative of* $f$ *in the direction of* $\mathbf{u}$.

**Exercise 220** *Let* $f : U \to \mathbb{R}$ *be differentiable,* $U \subset \mathbb{R}^n$.

1. *Let* $\mathbf{u} \in \mathbb{R}^n$ *be a unit vector and* $\mathbf{x} \in U$. *Define* $g(t) := f(\mathbf{x} + t\mathbf{u})$. *Show that* $D_{\mathbf{u}} f(\mathbf{x}) = g'(0)$. *This justifies the term "directional derivative" by showing that* $D_{\mathbf{u}}(f)$ *measures the rate of change of* $f$ *along the line through* $\mathbf{x}$ *parallel to* $\mathbf{u}$.

2. *Show that the derivative in the direction of* $\frac{\nabla f}{\|\nabla f\|}$ *is the maximum directional derivative, with value* $\|\nabla f\|$. *Note: you may not use the formula* $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \alpha$, *which we have not discussed. In fact this formula is used to define the angle between two vectors, and verifying that this definition is legitimate is no easier than (and uses the same theorem as) a direct solution to this exercise. Moreover, some linear algebra is required to show that, for* $n > 2$, *this notion of angle has the familiar geometric properties of the angle in the plane.*

## 5.4 The Mean Value Theorem and Applications

The Mean Value Theorem has wide ranging applications, from L'Hospital's Rule to optimization problems to the Fundamental Theorem of Calculus. The reader should pay close attention to when and how it is used because she will likely have reason to use it in the future.

**Theorem 340** *(Mean Value Theorem) Let* $f, g : [a, b] \to \mathbb{R}$ *be continuous on* $[a, b]$ *and differentiable on* $(a, b)$, *where* $a < b$. *Then there exists some* $x_0 \in (a, b)$ *such that*

$$[f(b) - f(a)]\, g'(x_0) = [g(b) - g(a)]\, f'(x_0).$$

**Proof.** Define $\phi : [a, b] \to \mathbb{R}$ by

$$\phi(t) := [g(b) - g(a)]f(t) - [f(b) - f(a)]g(t).$$

The function $\phi$ is continuous on $[a, b]$ and differentiable on $(a, b)$, $\phi(a) = \phi(b)$ and

$$\phi'(t) = [g(b) - g(a)]f'(t) - [f(b) - f(a)]g'(t).$$

The proof will therefore be finished if we can show that $\phi'(x_0) = 0$ for some $x_0 \in (a, b)$. If $\phi$ is constant then the proof is finished by Exercise 214. If $\phi$ is not constant then there exists some $c \in [a, b]$ such that $\phi(c) < \phi(a) = \phi(b)$ or $\phi(c) > \phi(a) = \phi(b)$. We will consider only the first case; the second is similar. Then the minimum $m$ of the continuous function $\phi$ satisfies $m \leq \phi(c) < \phi(a) = \phi(b)$ and therefore occurs at some $x_0 \in (a, b)$. But the First Derivative Test tells us that $\phi'(x_0) = 0$. ■

The Mean Value Theorem is a powerful workhorse in real analysis because it relates difference quotients to quotients of derivatives; this can be seen explicitly if $g(b) \neq g(a)$ and $g'(x_0) \neq 0$, in which case the statement can be written

$$\frac{[f(b) - f(a)]}{[g(b) - g(a)]} = \frac{f'(x_0)}{g'(x_0)}.$$

Letting $g(x) = x$ in Theorem 340 leads to the following theorem, which is also referred to as the Mean Value Theorem (sometimes Theorem 340 is called the Generalized Mean Value Theorem).

**Corollary 341** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. Then there exists some $x_0 \in (a, b)$ such that $f'(x_0) = \frac{f(b)-f(a)}{b-a}$.*

The next corollary is often referred to as Rolle's Theorem:

**Corollary 342** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$, and suppose that $f(a) = f(b)$. Then $f$ has a critical point in $(a, b)$.*

**Corollary 343** *If $f$ is a real function differentiable on $(a, b)$ and $f'(x) \geq 0$ (resp. $f'(x) \leq 0$, $f'(x) > 0$, $f'(x) < 0$) for all $x \in (a, b)$ then $f$ is increasing (resp. decreasing, strictly increasing, strictly decreasing) on $(a, b)$.*

**Proof.** We prove only the case for $f'(x) \geq 0$. For any $x_1 < x_2 \in (a, b)$ by the Mean Value Theorem we have for some $c$ such that $x_1 < c < x_2$ that

$$0 \leq f'(c)\,(x_2 - x_1) = f(x_2) - f(x_1).$$

■

**Exercise 221** *Let $f$ be a real function differentiable on $(a, b)$ such that $f'$ is bounded on $(a, b)$.*

1. *Prove that $f$ is Lipschitz on $(a, b)$.*
2. *Prove that $\lim_{x\to b} f(x)$ exists.*
3. *A similar statement holds for $a$; show that $f$ can be extended to a continuous function on $[a, b]$.*

**Exercise 222** *The purpose of this exercise is to prove one of the several versions of L'Hospital's rule. Let $f$ and $g$ be real functions differentiable on $(a, b)$, $g'(x) \neq 0$ for all $x \in (a, b)$ where $b < \infty$. Suppose that*

$$\lim_{x\to b} f(x) = \lim_{x\to b} g(x) = 0 \text{ and } \lim_{x\to b} \frac{f'(x)}{g'(x)} = L$$

*for some real number $L$. Show that $\lim_{x\to b} \frac{f(x)}{g(x)} = L$ by justifying the following steps: Let $\varepsilon > 0$ and $x_i \to b$ in $(a, b)$. Then*

1. *For all large* $i$, $\frac{f'(x_i)}{g'(x_i)} < L + \varepsilon$.
2. *For any large* $i$ *and* $j$,$g(x_i) - g(x_j) \neq 0$ *and* $\frac{f(x_i)-f(x_j)}{g(x_i)-g(x_j)} = \frac{f'(c)}{g'(c)} < L + \varepsilon$ *for some* $c$.
3. *For all large* $i$, $\frac{f(x_i)}{g(x_i)} \leq L + \varepsilon$ *and therefore* $\limsup \frac{f(x_i)}{g(x_i)} \leq L + \varepsilon$ *and therefore* $\limsup \frac{f(x_i)}{g(x_i)} \leq L$.

*A similar proof shows the opposite inequality for the* $\liminf$ *(you don't need to fill in the details).*

## 5.5 $C^1$ Functions

**Definition 344** *Let* $f : U \to \mathbb{R}^n$ *be a differentiable function, where* $U \subset \mathbb{R}^m$ *is open. If all the partial derivatives* $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ *exist for all* $\mathbf{x} \in U$ *and the resulting functions* $\frac{\partial f_i}{\partial x_j} : U \to \mathbb{R}$ *are continuous then* $f$ *is called* $C^1$.

One may analogously define $C^n$ for $n > 1$ by requiring that all $n^{th}$ partial derivatives be continuous, and $C^\infty$ to mean partial derivatives of all orders exist (and hence are continuous). The usual convention is also that $C^0$ simply means continuous.

**Exercise 223** *Show that the function* $f(x) = x^2 \sin(\frac{1}{x})$ *can be extended to a function defined on all of* $\mathbb{R}$ *that is differentiable but not* $C^1$ *(you may use differentiation theorems from elementary calculus).*

Verifying that a function is $C^1$ is often much simpler than verifying that a function is differentiable, at least in a concrete situation when one can actually compute the partial derivatives. Therefore the following theorem is very convenient. Not only does it have as a corollary that $C^1$ functions are differentiable, it shows that the differential approximates a $C^1$ function *uniformly*, in a sense, on any compact set.

**Theorem 345** *Let* $f : U \to \mathbb{R}^n$ *be a* $C^1$ *function, where* $U \subset \mathbb{R}^m$ *is open, and let* $A \subset U$ *be compact. For every* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that if* $\|\mathbf{x} - \mathbf{y}\| < \delta$ *with* $\mathbf{x} \in A$ *then*

$$\|f(\mathbf{y}) - f(\mathbf{x}) - D(f)(\mathbf{x})(\mathbf{y} - \mathbf{x})\| < \varepsilon \|\mathbf{x} - \mathbf{y}\| .$$

**Proof.** First suppose that $n = 1$. According to Proposition 90 for some $\delta_1 > 0$, $\overline{N(A, 2\delta_1)} \subset U$. It follows from the triangle inequality that, since $A$ is bounded, $\overline{N(A, \delta_1)} \subset U$ is bounded, hence compact, and the partial derivatives of $f$ are uniformly continuous on $V := N(A, \delta_1)$. Since there are finitely many partial derivatives, this means there exists some positive $\delta < \delta_1$ such that if $\mathbf{x} \in A$ and $d(\mathbf{x}, \mathbf{y}) < \delta$ then $\mathbf{y} \in V$, and for all $i$, $\frac{\partial f}{\partial x_i}(\mathbf{y})$ is defined and

$$\left| \frac{\partial f}{\partial x_i}(\mathbf{y}) - \frac{\partial f}{\partial x_i}(\mathbf{x}) \right| < \frac{\varepsilon}{m}. \tag{5.4}$$

For any $\mathbf{h}$ such that $\|\mathbf{h}\| < \delta$, write $\mathbf{h} = \sum_{i=1}^{m} h_i \mathbf{e}_i$ and set $\mathbf{w}_k := \sum_{i=1}^{k} h_i \mathbf{e}_i$ for $0 < k \leq m$ (so $\mathbf{w}_m = \mathbf{h}$) and $\mathbf{w}_0 := \mathbf{0}$. If $t \in [0,1]$ then for any $k$

$$\|\mathbf{w}_k + th_{k+1}\mathbf{e}_{k+1}\| = \sqrt{\sum_{i=1}^{k}(h_i)^2 + t^2 h_{k+1}^2} \leq \|\mathbf{h}\| < \delta.$$

In particular expressions such as $f(\mathbf{x} + \mathbf{w}_k + th_{k+1}\mathbf{e}_{k+1})$ when $\mathbf{x} \in A$ are defined when $t \in [0,1]$. We have a telescoping sum

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) := \sum_{i=1}^{m} \left[f(\mathbf{x}+\mathbf{w}_i) - f(\mathbf{x}+\mathbf{w}_{i-1})\right]. \tag{5.5}$$

For any $i$, let

$$\phi_i(t) = f(\mathbf{x} + \mathbf{w}_{i-1} + th_i\mathbf{e}_i) \text{ for } 0 \leq t \leq 1.$$

Then

$$f(\mathbf{x}+\mathbf{w}_i) - f(\mathbf{x}+\mathbf{w}_{i-1}) = \phi(1) - \phi(0)$$

and by the Mean Value Theorem and the chain rule, for some $t_i \in (0,1)$,

$$f(\mathbf{x}+\mathbf{w}_i) - f(\mathbf{x}+\mathbf{w}_{i-1}) = \phi'(t_i) = h_i \frac{\partial f}{\partial x_i}(\mathbf{y}_i)$$

where $\mathbf{y}_i = \mathbf{x} + \mathbf{w}_{i-1} + t_i h_i \mathbf{e}_i$. Combining this with Formula 5.5 we obtain

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \sum_{i=1}^{m} h_i \frac{\partial f}{\partial x_i}(\mathbf{y}_i)$$

and Formula (5.4) implies

$$|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - D(f)(\mathbf{x})\mathbf{h}| = \left|\sum_{i=1}^{m} h_i \frac{\partial f}{\partial x_i}(\mathbf{y}_i) - \sum_{i=1}^{m} h_i \frac{\partial f}{\partial x_i}(\mathbf{x})\right|$$

$$\leq \sum_{i=1}^{m} |h_i| \left|\frac{\partial f}{\partial x_i}(\mathbf{y}_i) - \frac{\partial f}{\partial x_i}(\mathbf{x})\right| \leq \sum_{i=1}^{m} |h_i| \frac{\varepsilon}{m} \leq \frac{\varepsilon}{m} \sum_{i=1}^{m} |h_i| \leq \varepsilon \max\{|h_i|\} \leq \varepsilon \|\mathbf{h}\|.$$

Letting $\mathbf{y} := \mathbf{x} + \mathbf{h}$ finishes the proof when $n = 1$. The proof for $n > 1$ is an exercise. ■

Taking $C$ to be the singleton set $\{\mathbf{x}\}$ we have the following corollary:

**Corollary 346** *If $f : U \to \mathbb{R}^n$ is a $C^1$ function, where $U \subset \mathbb{R}^m$ is open, then $f$ is differentiable on $U$.*

**Exercise 224** *Finish the proof of Theorem 345.*

**Definition 347** *Let $f : U \to \mathbb{R}^n$ be a differentiable function, where $U \subset \mathbb{R}^m$ is open. A point $\mathbf{x} \in U$ is called regular if the linear function $Df(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^n$is*

1. *surjective in the case when* $m \geq n$

2. *injective in the case when* $m < n$.

*A point that is not regular is called a critical point. If* $f$ *is* $C^1$ *and every point in* $U$ *is regular,* $f$ *is simply called regular (on* $U$*).*

**Remark 348** *The above definition unifies two standard uses of the word "regular". The first condition (surjectivity) is frequently defined for differentiable functions in general, without regard to dimension. However, the student familiar with linear algebra knows that if* $\mathbf{L} : \mathbb{R}^m \to \mathbb{R}^n$ *is linear then* $m$ *is the sum of the rank of* $\mathbf{L}$ *and the nullity of* $\mathbf{L}$ *(i.e. the dimensions of the image and kernel of* $\mathbf{L}$*, respectively). The former must be* $\leq n$ *and the latter is* $\geq 0$*. In other words, if* $m < n$ *it is impossible for the differential to be surjective and hence for any point to be regular; this definition is of no value in this case. The word "regular" is also used in reference to regular curves and surfaces, and in this case* $m < n$ *and the differential is assumed to be injective. Note that in the case* $m = n$*, regularity is actually equivalent to the differential being bijective. A little linear algebra shows that the two conditions in the above statement may actually be compressed to the simple statement:* $\dim(\ker Df(\mathbf{x})) = \max\{0, m - n\}$.

**Exercise 225** *This exercise uses some of the facts about linear algebra discussed at the beginning of this chapter. Let* $f : U \to \mathbb{R}^n$ *be a differentiable function, where* $U \subset \mathbb{R}^m$ *is open. Show the following:*

1. *If* $m = n$*,* $\mathbf{x} \in U$ *is regular if and only if* $\det Df(\mathbf{x}) \neq 0$.

2. *If* $m = 1$ *(i.e.* $f$ *is a curve) then* $x \in U$ *is regular if and only if* $f'(x) \neq \mathbf{0}$.

3. *If* $n = 1$ *(i.e.* $f$ *is real valued) then* $\mathbf{x} \in U$ *is regular if and only if* $\frac{\partial f}{\partial x_i}(\mathbf{x}) \neq 0$ *for some* $i$*. In particular, Definition 335 is consistent with the current more general one.*

4. *If* $f$ *is real valued and* $C^1$*, and* $\mathbf{x} \in U$ *is a regular point then for some* $\varepsilon > 0$*,* $f$ *is regular on* $B(\mathbf{x}, \varepsilon)$.

5. *If* $m = n + k$ *for* $k > 1$ *then* $\mathbf{x} \in U$ *is regular if and only if for some* $1 \leq m_1 \leq \cdots \leq m_n \leq n + k$ *the square matrix* $\mathbf{D}$ *with* $(\mathbf{D})_{ij} := \frac{\partial f_i}{\partial x_{m_j}}(\mathbf{x})$ *has nonzero determinant. This part requires some additional knowledge of linear algebra. Hint: For one direction, apply* $Df(\mathbf{x})$ *to the standard basis and pick out a maximal linearly independent subset of the resulting vectors.*

**Exercise 226** *Let* $g(\mathbf{x}) := \|\mathbf{x}\|$.

1. *Show that for* $\mathbf{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$*,* $\frac{\partial g}{\partial x_i}(\mathbf{x}) = \frac{x_i}{\|\mathbf{x}\|}$.

2. *Show that* $g$ *is regular on* $\mathbb{R}^n \backslash \{\mathbf{0}\}$.

By definition, a function is $C^1$ if and only if its partial derivatives are continuous. However, in many situations involving $C^1$ functions it is more useful to know that, in some sense, the differential itself is a continuous function. There are a couple of ways to interpret this, one of which is in the following exercise.

**Exercise 227** *We may identify the set $\mathcal{M}$ of real $m \times n$ matrices with $\mathbb{R}^{mn}$ using any scheme that sets up a bijection between the entries of a matrix and the components of a vector. Given any matrices $(a_{ij})$ and $(b_{ij})$ the number $\max|a_{ij} - b_{ij}|$ corresponds to the max metric on $\mathbb{R}^{mn}$ and therefore defines a metric on $\mathcal{M}$. Define a function $Jf : U \to \mathcal{M}$ that assigns to each $\mathbf{x} \in U$ the Jacobian matrix $Df(\mathbf{x})$. Show that $Jf$ is continuous if $f$ is $C^1$.*

Another method of considering continuity of the differential, which perhaps better captures the idea that the differentials at close points are close *as functions*, is the following:

**Proposition 349** *Suppose that $f : U \to \mathbb{R}^n$ is a $C^1$ function, where $U \subset \mathbb{R}^m$ is open. Then the function $Df : U \times \mathbb{R}^m \to \mathbb{R}^n$ defined by*

$$Df(\mathbf{x}, \mathbf{v}) = D(f)(\mathbf{x})(\mathbf{v})$$

*is continuous.*

**Proof.** By definition the functions $\frac{\partial f_j}{\partial x_i}$ are continuous. It suffices to consider any component

$$Df(\mathbf{x}, \mathbf{v})_i = \sum_{i=1}^{m} \frac{\partial f_i(\mathbf{x})}{\partial x_j} \mathbf{v}_j = \sum_{j=1}^{m} \frac{\partial f_i(\pi_1(\mathbf{x}, \mathbf{v}))}{\partial x_j} \cdot \pi_j(\pi_2(\mathbf{x}, \mathbf{v})),$$

where $\pi_k$ is the projection onto the $k^{th}$ factor, which is continuous. But we have represented this component as a combination of continuous functions using compositions and algebraic operations, and hence each component is continuous. ■

**Proposition 350** *Let $f_{ij} : X \to \mathbb{R}$ be continuous functions, $1 \leq i, j \leq m$, where $X$ is a metric space and let $\mathbf{F}(\mathbf{v})$ be the matrix such that $(\mathbf{F}(\mathbf{v}))_{ij} = f_{ij}(\mathbf{v})$. Then*

1. $\det \mathbf{F} : X \to \mathbb{R}$ *is continuous and*
2. *If $\mathbf{F}(\mathbf{v})$ is nonsingular for all $\mathbf{v}$ then the function $\left(\mathbf{F}^{-1}\right)_{ij} : X \to \mathbb{R}$ is continuous for all $i, j$.*

**Proof.** If $\mathbf{F}$ is $1 \times 1$ there is nothing to prove. If $\mathbf{F}$ is $2 \times 2$ then $\det \mathbf{F} = f_{11}f_{22} - f_{12}f_{21}$, which is continuous. The first part of the proposition may now be proved by induction, since the determinant of a matrix is an algebraic combination of determinants of lower demensional matrices; details are left to the reader. Similarly, the entries of the inverse of a matrix are given by Cramer's Rule, which expresses them using only algebraic combinations of the entries of the original matrix, including dividing by the (non-zero) determinant, which we already know is continuous. Again, details are left to the reader. ■

**Corollary 351** *If $f : U \to \mathbb{R}^n$ is $C^1$ where $U$ is an open subset of $\mathbb{R}^n$, then $\det Df : U \to \mathbb{R}$ is continuous. In particular if $\det Df(\mathbf{x}) \neq 0$ for some $\mathbf{x} \in U$ then for some $\varepsilon > 0$, $f$ is regular on $B(\mathbf{x}, \varepsilon)$.*

**Proposition 352** *Suppose that $A \subset U \subset \mathbb{R}^m$, where $A$ is compact and $U$ is open. If $f : U \to \mathbb{R}^n$ is $C^1$ then there exist $k_1, k_2 \in \mathbb{R}$ such that for all $\mathbf{x} \in A$ and $\mathbf{v} \in \mathbb{R}^n$,*

$$k_1 \|\mathbf{v}\| \leq \|Df(\mathbf{x})(\mathbf{v})\| \leq k_2 \|\mathbf{v}\|$$

*If $f$ is regular and $m < n$ then $k_1, k_2 > 0$. If $f$ is regular and $m = n$ then*

$$\frac{1}{k_2} \|\mathbf{v}\| \leq \left\| (Df)^{-1} (\mathbf{x})(\mathbf{v}) \right\| \leq \frac{1}{k_1} \|\mathbf{v}\| .$$

**Proof.** Consider the function $g(\mathbf{x}, \mathbf{v}) = \frac{\|Df(\mathbf{x})(\mathbf{v})\|}{\|\mathbf{v}\|} : A \times \mathbb{S}^{m-1} \to \mathbb{R}^n$, which is continuous (note that $\mathbf{v} \in \mathbb{S}^{m-1}$ implies $\mathbf{v} \neq \mathbf{0}$) on the compact set $A \times \mathbb{S}^{m-1}$. Therefore $g$ has a minimum $k_1$. If $f$ is regular and $m < n$ then $Df(\mathbf{x})(\mathbf{v}) \neq \mathbf{0}$ when $\mathbf{v} \neq \mathbf{0}$ and so $Df$ is never $\mathbf{0}$ on $A \times \mathbb{S}^{m-1}$; hence $k_1 > 0$. By definition of minimum, the left inequality holds for $\mathbf{v} \in \mathbb{S}^{m-1}$. Certainly the inequality is also true for $\mathbf{v} = \mathbf{0}$. For $\mathbf{v} \neq \mathbf{0}$ we let $\mathbf{u} := \frac{\mathbf{v}}{\|\mathbf{v}\|} \in \mathbb{S}^{m-1}$, and we have

$$\|Df(\mathbf{x})(\mathbf{v})\| = \|\mathbf{v}\| \|Df(\mathbf{x})(\mathbf{u})\| \geq k_1 \|\mathbf{v}\| \|\mathbf{u}\| = k_1 \|\mathbf{v}\| .$$

The second inequality is similar and the inequalities concerning the case $m = n$ are an exercise. ■

**Exercise 228** *Finish the proof of Proposition 352.*

## 5.6 The Inverse and Implicit Function Theorems

The next theorem is used later in this text, but more generally is useful in analysis, differential equations, differential geometry, and other areas of mathematics.

**Theorem 353** *(Inverse Function Theorem) Let $f : U \to \mathbb{R}^n$ be a $C^1$ function on an open set $U \subset \mathbb{R}^n$ and $\mathbf{x}_0 \in U$ be regular. Then there exists an open set $W \subset U$ containing $\mathbf{x}_0$ such that*

1. *$V = f(W)$ is open,*
2. *the function $g := f|_W$ is injective,*
3. *$g^{-1}$ is $C^1$ on $V$ and*
4. *$D(g^{-1})(g(\mathbf{x})) = (D(g)(\mathbf{x}))^{-1}$ for all $\mathbf{x} \in V$.*

**Proof.** By Corollary 351, $f$ is regular on some $B(\mathbf{x}_0, \varepsilon)$. By Proposition 352 there exists some $k > 0$ such that for all $\mathbf{x} \in C := C(\mathbf{x}_0, \frac{\varepsilon}{2})$ and $\mathbf{v} \in \mathbb{R}^n$,

$$k \|\mathbf{v}\| \leq \|Df(\mathbf{x})(\mathbf{v})\| . \tag{5.6}$$

According to Theorem 345 there is some $\delta > 0$ such that if $\|\mathbf{x}-\mathbf{y}\| < \delta$ with $\mathbf{x} \in C$ then

$$\|f(\mathbf{y}) - f(\mathbf{x}) - D(f)(\mathbf{x})(\mathbf{y}-\mathbf{x})\| < \frac{k}{2}\|\mathbf{x}-\mathbf{y}\|.$$

Let $0 < r < \frac{\min\{\varepsilon,\delta\}}{2}$. If $\mathbf{x}, \mathbf{y} \in K := C(\mathbf{x}_0, r)$ then $\mathbf{x} \in C$ and $\|\mathbf{x}-\mathbf{y}\| < \delta$. By the triangle inequality,

$$\begin{aligned}\|f(\mathbf{y}) - f(\mathbf{x})\| &\geq \|Df(\mathbf{x})(\mathbf{y}-\mathbf{x})\| - \|Df(\mathbf{x})(\mathbf{y}-\mathbf{x}) + f(\mathbf{x}) - f(\mathbf{y})\| \\ &\geq k\|\mathbf{y}-\mathbf{x}\| - \frac{k}{2}\|\mathbf{y}-\mathbf{x}\| = \frac{k}{2}\|\mathbf{y}-\mathbf{x}\| > 0. \qquad (5.7)\end{aligned}$$

It follows that $f$ is injective on $K$. Moreover, by Lemma 105, $g^{-1}$ is continuous on $f(K)$.

We will now show that $B(f(\mathbf{x}_0), \frac{\mathbf{kr}}{\mathbf{4}}) \subset f(K)$. Suppose not, i.e. there is some $\mathbf{y} \in B(f(\mathbf{x}_0), \frac{\mathbf{kr}}{\mathbf{4}}) \backslash f(K)$ and define $g(\mathbf{x}) := \|f(\mathbf{x}) - \mathbf{y}\|$. Since $f^{-1}(\{\mathbf{y}\})$ is closed, Exercise 226 implies $g$ is regular on an open set containing $K$. Since $K$ is compact, $g$ has a minimum at some $\mathbf{x} \in K$. But $g$ is regular and hence has no critical points, so $\|\mathbf{x}-\mathbf{x}_0\| = r$. But

$$\begin{aligned}g(\mathbf{x}) &= \|f(\mathbf{x}) - \mathbf{y}\| \geq \|f(\mathbf{x}) - f(\mathbf{x}_0)\| - \|f(\mathbf{x}_0) - \mathbf{y}\| \\ &\geq \frac{kr}{2} - \|f(\mathbf{x}_0) - \mathbf{y}\| > \frac{kr}{4} > \|f(\mathbf{x}_0) - \mathbf{y}\| = g(\mathbf{x}_0)\end{aligned}$$

a contradiction. Let $V := B(f(\mathbf{x}_0), \frac{\mathbf{kr}}{\mathbf{4}})$ and $W := f^{-1}(V) \cap B(\mathbf{x}_0, r)$. It is an exercise to show $f(W) = V$, and the first two parts are finished.

We next show that $g^{-1}$ is differentiable, with $D(g^{-1}) = (Dg)^{-1}$ on $V$. Applying Inequalities (5.6) and (5.7) and the fact that $g = f$ on $W$ we have for $\mathbf{x}, \mathbf{y} \in W$

$$\begin{aligned}&\frac{\left\|(Dg(\mathbf{x})^{-1}(g(\mathbf{x}) - g(\mathbf{y})) - g^{-1}(g(\mathbf{x})) + g^{-1}(g(\mathbf{y}))\right\|}{\|g(\mathbf{x}) - g(\mathbf{y})\|} \\ &\leq \frac{\left\|D(g(\mathbf{x}))\left[(Dg(\mathbf{x})^{-1}(g(\mathbf{x}) - g(\mathbf{y})) - \mathbf{x} + \mathbf{y})\right]\right\|}{k\|g(\mathbf{x}) - g(\mathbf{y})\|} \\ &\leq \frac{2}{k^2}\frac{\|g(\mathbf{x}) - g(\mathbf{y}) - Dg(\mathbf{x})(\mathbf{x}-\mathbf{y}))\|}{\|\mathbf{x}-\mathbf{y}\|}.\end{aligned}$$

The latter quantity is arbitrarily small when $\|\mathbf{x}-\mathbf{y}\|$ is small, and since $g^{-1}$ is continuous it is small when $\|g(\mathbf{x}) - g(\mathbf{y})\|$ is small, and the proof of the fourth part is finished. The final detail, that $g^{-1}$ is $C^1$ on $V$, follows from Proposition 350. ■

**Exercise 229** *Finish the proof of the Inverse Function Theorem.*

**Corollary 354** *If $f : U \to \mathbb{R}^n$ is a regular function on an open set $U \subset \mathbb{R}^n$ then $f$ is open.*

**Exercise 230** *Prove Corollary 354.*

**Exercise 231** *Consider the function* $f(x) = x + 2x^2 \sin(\frac{1}{x})M$.

1. *Show that* $f$ *may be extended to a differentiable function on all of* $\mathbb{R}$ *such that* $f'(0) \neq 0$.
2. *Show that the resulting function is not one-to-one on any open interval containing* $0$*, hence necessity of the requirement in the Inverse Function Theorem that the function be* $C^1$.

To motivate the Implicit Function Theorem, recall the process of "implicit differentiation" learned in elementary calculus. One is given an equation like $x^2y - 1 = 0$ and one finds $\frac{dx}{dy}$ (or $\frac{dy}{dx}$) without actually solving for the dependent variable–which can be difficult or impossible to do. Formally, we simply apply $\frac{d}{dy}$ to each side of the equation, treating $x$ as a function of $y$ and applying the chain rule (and product rule). The result is $2yx\frac{dx}{dy} + x^2 = 0$, and we may then solve for $\frac{dx}{dy}$. If the reader checks his or her calculus book, he/she will find buried somewhere in the discussion of implicit differentiation some fine print to the effect that, in carrying out implicit differentiation it is assumed that the variable in question is defined implicitly as a function of the other variable(s) by the given equation. What does this caveat mean? Letting $f(x, y) = x^2y - 1$, the above equation becomes $f(x, y) = 0$. The assumption is that we can write $x = g(y)$ where $g$ is differentiable function such that $f(g(y), y) = 0$. If such exists then implicit differentiation is clearly just the chain rule: $0 = \frac{d}{dy}(f(g(y), y) = \frac{\partial f}{\partial x}g'(y) + \frac{\partial f}{\partial y} \cdot 1$, then solve for $g'(y)$. There may not be a single function that works for all $y$. In the present case, we can solve for $x = \pm\frac{1}{\sqrt{y}}$. We then have two "inverse functions", one of which is defined for $x > 0$ and the other for $x < 0$. The Implicit Function Theorem justifies the assumption about the existence of $g$.

The Implicit Function Theorem concerns functions from open subsets of $\mathbb{R}^{n+k}$ into $\mathbb{R}^n$, with $k \geq 1$. The basic assumption is similar to regularity, except that in this case the Jacobian matrix is not square and has no determinant. In fact, the differential is has continuous entries $\frac{\partial f_i}{\partial x_j}$, where $i = 1, ..., n$ and $j = 1, ..., n+k$ and the assumption will be that the square matrix with $j = 1, ..., n$ has non-zero determinant. In the statement of the theorem we will identify $\mathbb{R}^{n+k}$ with $\mathbb{R}^n \times \mathbb{R}^k$ the elements of which are pairs of vectors $(\mathbf{x}, \mathbf{t})$.

**Theorem 355** *(Implicit Function Theorem) Let* $f : U \to \mathbb{R}^n$ *be a* $C^1$ *function on an open set* $U \subset \mathbb{R}^{n+k}$ *with* $k \geq 1$*. Suppose that for some* $(\mathbf{x}_0, \mathbf{t}_0) \in U$,

1. $f(\mathbf{x}_0, \mathbf{t}_0) = \mathbf{0}$ *and*
2. $\det \mathbf{D} \neq 0$*, where* $(\mathbf{D})_{ij} := \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0, \mathbf{t}_0)$ *with* $1 \leq i, j \leq n$.

*Then there exist an open subset* $U' \subset \mathbb{R}^k$ *containing* $\mathbf{t}_0$ *and a unique* $C^1$ *function* $g : U' \to \mathbb{R}^n$ *such that* $g(\mathbf{t}_0) = \mathbf{x}_0$ *and* $f(g(\mathbf{t}), \mathbf{t})) = \mathbf{0}$ *for all* $\mathbf{t} \in U'$.

**Proof.** Define a new function $F : U \to \mathbb{R}^{n+k}$ by $F(\mathbf{x}, \mathbf{t}) = (f(\mathbf{x}, \mathbf{t}), \mathbf{t})$. The Jacobian matrix for $F$ consists of a block matrix

$$\begin{bmatrix} \mathbf{D} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

where $\mathbf{D}$ is the matrix defined above and $(\mathbf{A})_{ij} = \frac{\partial f_i}{\partial t_j}$. It can be checked that the determinant of this matrix is $(\det \mathbf{D})(\det \mathbf{I}) = \det \mathbf{D}$, which is nonzero at $(\mathbf{x}_0, \mathbf{t}_0)$. Moreover this matrix has continuous entries, so the Inverse Function Theorem applies and there exist open sets $V$ and $W$ and a $C^1$ function $G : V \to W$ such that $G = F^{-1}$ when $F$ is restricted to $W$. Define $G_1 := \pi_1 \circ G$, where $\pi_1 : \mathbb{R}^{n+k} \to \mathbb{R}^k$ is the projection. Let $V' \subset \mathbb{R}^n$ and $U' \subset \mathbb{R}^k$ be open such that $(\mathbf{0}, \mathbf{t}_0) = F(\mathbf{x}_0, \mathbf{t}_0) \in V' \times U' \subset V$ (see Exercise 82) and define $g(\mathbf{t}) := G_1(\mathbf{0}, \mathbf{t})$ for all $\mathbf{t} \in U'$. Certainly $g$ is $C^1$, and

$$g(\mathbf{t}_0) = G_1(\mathbf{0}, \mathbf{t}_0) = G_1(F(\mathbf{x}_0, \mathbf{t}_0)) = \mathbf{x}_0.$$

Since $F$ is onto $V$, if $\mathbf{t} \in U'$ then there is some $(\mathbf{x}, \mathbf{t}) \in W$ such that $f(\mathbf{x}, \mathbf{t}), \mathbf{t}) = F(\mathbf{x}, \mathbf{t}) = (\mathbf{0}, \mathbf{t})$. Now

$$(f(G_1(\mathbf{0}, \mathbf{t}), \mathbf{t}), \mathbf{t}) = F(G_1(\mathbf{0}, \mathbf{t}), \mathbf{t})$$

$$= F(\pi_1(F^{-1}(F(\mathbf{x}, \mathbf{t})), \mathbf{t})) = F(\mathbf{x}, \mathbf{t}) = (\mathbf{0}, \mathbf{t}).$$

The first terms of these ordered pairs must be equal, so

$$f(g(\mathbf{t}), \mathbf{t})) = f(G_1(\mathbf{0}, \mathbf{t}), \mathbf{t})) = \mathbf{0}.$$

Now suppose $h$ is another such function. Then $f(g(\mathbf{t}), \mathbf{t}) = f(h(\mathbf{t}), \mathbf{t}) = \mathbf{0}$ and

$$F(g(\mathbf{t}), \mathbf{t}) = (f(g(\mathbf{t}), \mathbf{t}), \mathbf{t}) = (f(h(\mathbf{t}), \mathbf{t}), \mathbf{t}) = F(h(\mathbf{t}), \mathbf{t}).$$

Since $F$ is one-to-one on $V$, $g(\mathbf{t}) = h(\mathbf{t})$. ■

Note that in some cases, reordering the coordinates may allow the Implicit Function Theorem to apply. That is, the it may be possible to "collect" $n$ coordinates as the first $n$ coordinates, so that the determinant of the resulting matrix $\mathbf{D}$ is nonzero. This may be more precisely formulated using the notion of rank of a linear transformation, but we will not give the details here.

**Exercise 232** *Consider the function* $f(x, y) = x^2 y - 1$.

1. *Verify that the assumptions for the Implicit Function Theorem are valid whenever* $x \neq 0$.

2. *What is the largest possible set* $U'$ *whose existence the theorem guarantees–in particular, does* $U'$ *really depend on* $(x_0, y_0)$ *in this case?*

**Exercise 233** *Use the Implicit Function Theorem to prove the Inverse Function Theorem, using the following steps:*

1. *Given a function $f$ satisfying the conditions of the Inverse Function Theorem, consider the function $g(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}$.*
2. *Show that the resulting one-sided inverse $h : U' \to \mathbb{R}^n$ also satisfies the conditions of the Inverse Function Theorem and is 1-1.*
3. *Apply the same two steps to the function $h$ and show that the resulting function $k$ must be the restriction of $f$ to some open set.*

**Exercise 234** *Consider the function $f : \mathbb{C} \to \mathbb{C}$ defined by $f(z) = z^2$.*

1. *Rewrite $f$ as a function of two real variables from $\mathbb{R}^2$ to $\mathbb{R}^2$.*
2. *Show that $f$ is regular on $\mathbb{R}^2 \backslash (0, 0)$ but that it is not one-to-one.*

**Definition 356** *Let $h : U \to \mathbb{R}$ be a $C^1$ function, where $U \subset \mathbb{R}^n$. A regular value of $h$ is $c \in \mathbb{R}$ such that if $h(\mathbf{x}) = c$ then $\mathbf{x}$ is not a critical point of $h$. A regular surface is a non-empty set of the form $h^{-1}(c)$, where $c$ is a regular value of $h$.*

**Exercise 235** *Show that the graph of any $C^1$ function $f : U \to \mathbb{R}$, where $U \subset \mathbb{R}^{n-1}$ is open, is a regular surface.*

**Exercise 236** *Show that the sphere $S$ of radius $c^2 > 0$ in $\mathbb{R}^{n+1}$ centered at $\mathbf{0}$ is a regular surface.*

**Exercise 237** *The purpose of this exercise is to justify the following statement: every regular surface is "locally the graph of a function".*

1. *Let $h : U \to \mathbb{R}$ be a $C^1$ function, where $U \subset \mathbb{R}^n$ and $\mathbf{z} \in U$ is not a critical point of $h$. Define $S = h^{-1}(h(\mathbf{z}))$. Show that for some $1 \leq i \leq n$ the following holds: Let $\mathbb{R}^n := \mathbb{R} \times \mathbb{R}^{n-1}$, writing $\mathbf{x} = (x_i, \mathbf{t})$, with $\mathbf{t} := (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$. Then there exists an open set $V \subset \mathbb{R}^{n-1}$ and a $C^1$ function $g : V \to \mathbb{R}$ such that the graph of $g$ is the set of all $\mathbf{y} \in S$ such that $(y_1, ..., y_{i-1}, y_{i+1}, ...y_n) \in V$. Hint: Show that for some $i$, $\frac{\partial h}{\partial x_i}(\mathbf{z}) \neq 0$ and consider the function $f(\mathbf{x}) := h(\mathbf{x}) - h(\mathbf{z})$.*
2. *Illustrate the above statement by showing that the unit sphere in $\mathbb{R}^3$ is the union of six open hemispheres that are the graphs of six functions. Not a lot of details needed–just state what the functions are and draw a picture.*

## 5.7 Real Functions

If $f : A \to \mathbb{R}$ is differentiable on $A \subset \mathbb{R}$ then the derivative at each point defines a new real function $f' : A \to \mathbb{R}$. Recapping our theorems in the real case case we have for linearity $(cf + g)'(x_0) = cf'(x_0) + g'(x_0)$ and for the Chain Rule $(f \circ g)'(x_0) = f'(g(x_0))g'(x_0)$. Note that every partial derivative is also a derivative (according to Exercise 213) and therefore the theorems in this section apply to partial derivatives, and hence to the components of any total derivative.

**Exercise 238** *Let $f$ and $g$ be real functions differentiable at $x$, and $c \in \mathbb{R}$.*

1. *Prove the Product Rule: $(f \cdot g)'(x) = f'(x)g(x) + g'(x)f(x)$.*
2. *Use that fact that $f(x)\left(\frac{1}{f(x)}\right) = 1$ to prove that $\left(\frac{1}{f(x)}\right)' = -\frac{f'(x)}{f(x)^2}$ for all $x$ such that $f(x) \neq 0$.*
3. *Prove the Quotient Rule: $\left(\frac{g(x)}{f(x)}\right)' = \frac{f(x)g'(x)-g(x)f'(x)}{f(x)^2}$ for all $x$ such that $f(x) \neq 0$.*
4. *Prove the Powers Rule for natural numbers: For every natural number $n$, if $k(x) = x^n$ then $k'(x) = nx^{n-1}$.*
5. *Prove that if $f$ has an inverse $f^{-1}$ in some open interval containing $x$ and $f'(x)$ exists and is nonzero, then $\frac{df^{-1}}{dx}(f(x)) = \frac{1}{f'(x)}$. Hint: Use $f(f^{-1}(x)) = x$. Note that if $f$ is $C^1$ then one need only assume that $f'(x) \neq 0$ in the interval and apply the Inverse Function Theorem.*

We will not go farther in discussing the basic computational theorems in calculus beyond a few comments now and additional exercises about the exponential function later. The derivatives of trigonometric functions can be determined using geometric arguments and the basic theorems in the previous exercise. Alternatively one can define the trigonometric functions using their power series and then derive their geometric properties.

Two of the most important applications of the Mean Value Theorem are the two theorems together known as the Fundamental Theorem of Calculus. Before proving these theorems we establish a little notation. From now on we will denote the Lebesgue integral of an integrable Borel function $f$ on an interval $[a, b]$ by $\int_a^b f$ or $\int_a^b f(x)$ or $\int_a^b f(x)dx$. If $a < b < c$ then since the intervals $[a, b]$ and $[b, c]$ intersect in the set $\{b\}$, which has measure 0, we immediately have $\int_a^c f = \int_a^b f + \int_b^c f$. When $a < b$ we will define $\int_b^a f = -\int_a^b f$. This definition is made purely for computational convenience; see, for example, the next exercise. Also, since $[a, a]$ is the single point set $\{a\}$, $\int_a^a f = 0$ for any $f$.

**Exercise 239** *Show that the formula $\int_a^c f = \int_a^b f + \int_b^c f$ is true without any assumptions about the relative orders of the points $a, b, c$. Hint: Don't be exhaustive–do a couple of cases including one when two of the points are equal; the rest are similar.*

**Theorem 357** *(Fundamental Theorem of Calculus I): Suppose that $f$ is a real function differentiable on an open set containing $[a, b]$. Then $f'$ is a Borel function on $[a, b]$ and if $f'$ is bounded on $[a, b]$ then $\int_a^b f' = f(b) - f(a)$.*

**Proof.** Let $t_i \to 0$, with $t_i > 0$ for all $i$. Since $f$ is defined on an open set containing $[a, b]$, letting $h_i(x) := \frac{f(x+t_i)-f(x)}{t_i}$ we see that $h_i$ is defined on $[a, b]$

for all large $i$. By definition $f'$ is the pointwise limit of $(h_i)_{i=1}^{\infty}$ and hence is Borel. Moreover by the Mean Value Theorem for all large $i$ and all $x$ we have

$$|h_i(x)| = \left|\frac{f(x+t_i) - f(x)}{t_i}\right| = |f'(c)| \leq M,$$

where $M$ is a bound for $f'$ and $c$ is some point between $x$ and $x + t_i$. The Lebesgue Dominated Convergence Theorem applies and we have

$$\int_a^b f' = \lim \int_a^b h_i = \lim \int_a^b \frac{f(x+t_i) - f(x)}{t_i} = \lim \frac{1}{t_i}\left(\int_a^b f(x+t_i) - \int_a^b f(x)\right).$$

Invoking translation invariance yields

$$\int_a^b f' = \lim \frac{1}{t_i}\left(\int_{a+t_i}^{b+t_i} f - \int_a^b f\right) = \lim \frac{1}{t_i}\left(\int_b^{b+t_i} f - \int_a^{a+t_i} f\right)$$

$$\leq \lim \frac{1}{t_i}\left(t_i \cdot \max_{[b,b+t_i]} f - t_i \cdot \min_{[a,a+t_i]} f\right) = \lim \left(\max_{[b,b+t_i]} f - \min_{[a,a+t_i]} f\right) = f(b) - f(a)$$

since $f$ is continuous. Changing the max to a min and vice versa reverses the inequality in the last line, proving equality. ■

**Corollary 358** *If $f$ and $g$ are differentiable on some open set containing $[a, b]$ and $f'(x) = g'(x)$ for all $x \in (a, b)$ then for all $x$, $f(x) - g(x) = c$ for some $c \in \mathbb{R}$. In particular if $f'(x) = 0$ for all $x$ then $f$ is constant.*

**Proof.** Let $h(x) = f(x) - g(x)$. Then $h'(x) = 0$ and $h(x) - h(a) = \int_a^x h'(x) = 0$. Letting $c := h(a)$ finishes the proof. ■

Note that a $C^1$ function on an open set containing a closed, bounded interval always has bounded derivative on that interval.

**Exercise 240** *For this exercise you may use only results from this or previous sections and not your other knowledge about the function $e^x$. Suppose that $f : \mathbb{R} \to \mathbb{R}$ satisfies the equations $f'(x) = f(x)$ for all $x$ and $f(0) = 1$.*

1. *Prove that the function $h(x) := f(x)f(-x)$ is the constant function 1; conclude that $f$ is never 0 (hence positive), and that $f(-x) = \frac{1}{f(x)}$ for all $x$.*

2. *Prove that there is only one such function $f$. Hint: Consider the function $\frac{g}{f}$ where $f$ and $g$ are any two such functions.*

3. *Show that $f(a + x) = f(a)f(x)$ for any $a$ and $x$. Hint: Fix $a$ and apply uniqueness to the function $k(x) = \frac{f(a+x)}{f(a)}$.*

**Theorem 359** *(Fundamental Theorem of Calculus II) If $f : [a, b] \to \mathbb{R}$ is continuous then for any $c \in (a, b)$ the function $F(x) := \int_c^x f$ is differentiable on $(a, b)$ and $F'(x) = f(x)$.*

**Proof.** Let $\varepsilon > 0$. If $h > 0$ is small enough that $x + h \in (a, b)$ we have $F(x+h) - F(x) = \int_a^{x+h} f - \int_a^x f = \int_x^{x+h} f$ and

$$h \min_{[x,x+h]} f \leq \int_x^{x+h} f \leq h \max_{[x,x+h]} f,$$

which implies

$$\min_{[x,x+h]} f \leq \frac{F(x+h) - F(x)}{h} \leq \max_{[x,x+h]} f.$$

Since $f$ is continuous, when $h$ is small we have

$$f(x) - \varepsilon \leq \frac{F(x+h) - F(x)}{h} \leq f(x) + \varepsilon.$$

A similar argument shows the same inequality when $h$ is negative, and the proof is finished. ■

The function $F(x) := \int_c^x f$ is called an *antiderivative* of $f$, because $F'(x) = f(x)$ for all $x$. Given any other antiderivative $g$, we see that $\frac{d}{dx}(F(x) - g(x)) = f(x) - f(x) = 0$ for every $x$ and hence $F(x) - g(x)$ is constant. That is, any two antiderivatives of $f$ differ by a constant; conversely, for any constant $q$, $g(x) + q$ is also an antiderivative of $f$.

We next consider differentiation in connection with sequences of functions. The Fundamental Theorem of Calculus allows us to derive a useful differentiation theorem from our theorems about limits of integrals.

**Theorem 360** *Let $f_i : (a, b) \to \mathbb{R}$ be a sequence of differentiable functions such that $(f_i')$ converges uniformly to a function $g : (a, b) \to \mathbb{R}$ and for some $x_0 \in (a, b)$, $f_i(x_0)$ is convergent. Then $(f_i)$ converges uniformly to a differentiable function $f : (a, b) \to \mathbb{R}$ and for every $x \in (a, b)$,*

$$f'(x) = g(x) = \lim f_i'(x).$$

**Proof.** Let $\delta > 0$. Given any $x, x + h \in (a, b)$, applying the Mean Value Theorem and the fact that $(f_i')$ is Cauchy, we have for all large $i$ and $j$ there is some $c_{ij} \in (a, b)$ such that

$$\begin{aligned} &|f_i(x+h) - f_i(x) - f_j(x+h) + f_j(x)| \\ &= |(f_i - f_j)(x+h) - (f_i - f_j)(x)| \\ &= \left|h (f_i - f_j)' (c_{ij})\right| \leq \delta |h| . \end{aligned} \tag{5.8}$$

Now given $\varepsilon > 0$ we may take $\delta := \frac{\varepsilon}{2(b-a)}$ in the above inequality and note that for large $i$ and $j$, $|f_i(x_0) - f_j(x_0)| \leq \frac{\varepsilon}{2}$.

Letting $h := x_0 - x$ we obtain

$$|f_i(x) - f_j(x)| \leq |f_i(x) - f_j(x) - f_i(x_0) + f_j(x_0)| + |f_i(x_0) - f_j(x_0)|$$

$$\leq \frac{\varepsilon\,|x_0 - x|}{2(b-a)} + \frac{\varepsilon}{2} \leq \varepsilon.$$

This shows $(f_i)$ is Cauchy, and Proposition 220 implies that $f_i \to f$ for some $f$. Now for any $x, x+h \in (a,b)$ with $h \neq 0$, define $k_i(h) := \frac{f_i(x+h)-f_i(x)}{h}$. Formula (5.8) implies that $(k_i)$ converges uniformly to $k(h) := \frac{f(x+h)-f(x)}{h}$. Theorem 217 now implies that the following limits exist:

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

$$= \lim_{i\to\infty}\lim_{h\to 0} \frac{f_i(x+h) - f_i(x)}{h} = \lim_{i\to\infty} f_i'(x) = g(x).$$

■

We now turn our attention to Taylor's Theorem and power series for real functions. If $f$ is differentiable on some interval, then we can ask whether the function $f'$ is differentiable. If it is, its derivative will be denoted by $f''$. Continuing this process (as long as the derivatives in question exist), the *higher derivatives* of $f$ will be denoted by $f'', f''', f^{(4)}, \ldots$ or $\frac{d^2 f}{dx^2}, \frac{d^3 f}{dx^3}, \ldots$. In order to simplify the statement of Taylor's Theorem we will also denote $f$ by $f^{(0)}$. In the next theorem we will say that $c$ is *between* $x$ and $x_0$ if $x < c < x_0$ or $x_0 < c < x$.

**Theorem 361** *(Taylor's Theorem) Let $f$ be a real $C^n$ function defined on an open set $U$ and suppose that $f^{(n+1)}$ exists for every $x \in [a,b] \subset U$. For any $x_0, x \in [a,b]$ there exists some point $c$ between $x_0$ and $x$ such that*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \frac{f^{(n+1)}(c)(x-x_0)^{n+1}}{(n+1)!}.$$

**Proof.** If we define

$$M := \frac{f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k}{(x-x_0)^{n+1}}$$

then we need to show that for some $c$ between $x$ and $x_0$, $f^{(n+1)}(c) = (n+1)!M$. Letting

$$g(t) := f(t) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(t-x_0)^k - M(t-x_0)^{n+1} \tag{5.9}$$

it is an exercise to prove that $g(x_0) = g(x) = 0$ and $g^{(j)}(x_0) = 0$ for all $j \leq n$. We will show by induction that for all $1 \leq j \leq n+1$, $g^{(j)}(c_j) = 0$ for some $c_j$ between $x$ and $x_0$. The case $j = 1$ follows from Rolle's Theorem and the fact that $g(x_0) = g(x) = 0$. Suppose that we have proved that for some $k \leq n$, there exists some $c_k$ between $x_0$ and $x$ such that $g^{(k)}(c_k) = 0$. Since $g^{(k)}(x_0) = 0$ the Mean Value Theorem implies there exists some $c_{k+1}$ between $x_0$ and $c_k$, and hence between $x_0$ and $x$, such that $g^{(k+1)}(c_{k+1}) = 0$. Now since $\sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(t-x_0)^k$ is an $n^{th}$ degree polynomial, differentiating Formula (5.9) $(n+1)$ times yields

$$0 = g^{(n+1)}(c_{n+1}) = f^{(n+1)}(c_{n+1}) - 0 - (n+1)!M.$$

■

**Exercise 241** *Finish the proof of Taylor's Theorem.*

**Exercise 242** *Let $f : (0,\infty) \to \mathbb{R}$ be a function such that $f''(x)$ exists for all $x$ and $f''$ is bounded on $(0,\infty)$. Show that if $\lim_{x\to\infty} f(x) = 0$ then $\lim_{x\to\infty} f'(x) = 0$. Hint: Solve for the derivative in Taylor's formula and figure out how to make it small by making $x - x_0$ small and $x$ large.*

We are now in a position to discuss integration and differentiation of real power series. Although we previously studied power series for complex numbers, if the coefficients of a power series are real then the restriction of the power series to those real numbers for which it converges defines a real function. The domain of this function is the intersection of the domain of the complex series with $\mathbb{R}$; that is, if $R > 0$ is the radius of convergence of the complex power series (which depends only on the coefficients) then the real power series converges absolutely and uniformly on any interval $[-r, r]$ where $0 < r < R$, and converges pointwise on $(-R, R)$. A power series $\sum_{n=0}^{\infty} c_n x^n$, where $c_n$ and $x$ are real is called a real power series. Given such a power series one can "formally" differentiate and antidifferentiate it (i.e. compute now, prove later) as one would a polynomial, to obtain the real power series

$$\sum_{n=0}^{\infty} n c_n x^{n-1} \text{ and } \sum_{n=0}^{\infty} \frac{1}{n+1} c_n x^{n+1} \tag{5.10}$$

respectively. Since $\lim \sqrt[n]{n} = \lim \sqrt[n]{n+1} = 1$, these two series have the same radius as convergence $R$ as the original. Supposing that $R > 0$ we know that $\sum_{n=0}^{k} c_n x^n$ converges uniformly to the continuous function $\sum_{n=0}^{\infty} c_n x^n$ on any $[-r, r]$ where $0 < r < R$. From Theorem 360 we obtain:

**Proposition 362** *Let $f(x) = \sum_{n=0}^{\infty} c_n x^n$ be a real power series with radius of convergence $R > 0$. The the derivative and an antiderivative of $f$ are given by Formula (5.10), and these series also have radius of convergence $R$.*

Differentiating a power series $n$ times proves the next corollary, which essentially states that any power series with positive radius of convergence is equal to its own (infinite) Taylor series with $x_0 = 0$ (such series are often called Maclaurin series). A function having a Taylor series with nontrivial interval of convergence is called analytic in that interval.

**Corollary 363** *If $f(x) = \sum_{n=0}^{\infty} c_n x^n$ has radius of convergence $R > 0$ then $c_n = \frac{f^{(n)}(0)x^n}{n!}$ for all $n$.*

**Exercise 243** *Let $f$ be the function from Exercise 240, which is the unique function such that $f'(x) = f(x)$ for all $x \in \mathbb{R}$ and $f(0) = 1$.*

1. *Prove that $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ is the Maclaurin series for $f$ and hence that $f(x) = e^x$.*

2. *Show that $\frac{d}{dx}(\ln x) = \frac{1}{x}$.*

**Exercise 244** *Write out the first four terms of the Taylor series of $f(x) = \frac{1}{x}$ at $x_0 = 1$ and use this to find the first five terms of the Taylor series of $g(x) = x \ln x$ at $x_0 = 1$. Hint: For the last part multiply by $(x-1)+1$ rather than $x$.*

## 5.8 Linear Functions and Integration

We will review a little more linear algebra. A nonsingular linear function can be written as a composition of *elementary linear functions* of the following three types:

1. switch coordinates: $(x_1, .., x_i, ..., x_j, ..., x_n) \mapsto (x_1, .., x_j, ..., x_i, ..., x_n)$, $i < j$

2. scale coordinate: $(x_1, .., x_i, ..., x_n) \mapsto (x_1, .., ax_i, ..., x_n)$, $a \neq 0$

3. tilt coordinate: $(x_1, .., x_i, ..., x_n) \mapsto (x_1, .., x_i + ax_j, ..., x_n)$, $a \neq 0$, $i \neq j$

Note that the inverse of any elementary linear transformation is an elementary linear transformation of the same type. We will need to know the effect on a unit cube of these three types of elementary functions. The first one takes the unit cube back to itself. The second one takes the unit cube to a rectangular box of measure $|a|$, and the last one takes the unit cube to a parallelepiped of measure 1 (see also Exercise 247 below). We will only provide details of the last, computing the measure by slices. Let $\mathbf{E}$ be a linear function of the third type. Since the unit cube $Q^n$ is compact and $\mathbf{E}$ is continuous, $\mathbf{E}(Q^n)$ is compact, and has the form $\mathbf{E}(Q^n) = \{(x_1, .., x_i + ax_j, ..., x_n) : 0 \leq x_k \leq 1 \text{ for all } k\}$. Using $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}^1$, where $\mathbb{R}^1$ is the $i^{th}$ factor, we see that the projection $E_1$ of $\mathbf{E}(Q^n)$ onto $\mathbb{R}^{n-1}$ is simply $Q^{n-1}$. Fixing any $\mathbf{x} = (x_1, ..., x_{n-1}) \in E_1$, we have $ax_l \leq x_i + ax_l \leq ax_l + 1$ where $l = j$ if $j < i$ and $l = j-1$ if $j > i$. This means $E_{\mathbf{x}} = [ax_j, ax_j + 1]$, which has measure 1. Applying Theorem 296 shows that

$$\mu\left(\mathbf{E}(Q^n)\right) = \int_{Q^{n-1}} 1 = \mu(Q^{n-1}) = 1.$$

From elementary linear algebra we know that the determinant of an elementary linear transformation of the first or third type is 1, and the determinant of one of the second type is $a$. In other words, for any elementary linear transformation $\mathbf{E}$,

$$\mu(\mathbf{E}(Q^n)) = |\det \mathbf{E}| \,. \tag{5.11}$$

Before we continue we need to show that the image of a Borel set with respect to a homeomorphism is a Borel set. While this fact may seem trivial since homeomorphisms preserve all topological properties of a metric space, the definition of Borel sets as the smallest $\sigma$-algebra of sets with certain properties makes the argument a little tricky.

**Lemma 364** *Let $f : X \to Y$ be a homeomorphism between metric spaces. If $E \subset X$ is a Borel set then $f(E)$ is a Borel set.*

**Proof.** Let $\mathcal{B}(X)$ be the collection of all Borel sets in $X$ and $\mathcal{B}'(X) := \{f^{-1}(A) : A \subset X \text{ is a Borel set}\}$. It is easy to check that $\mathcal{B}'(X)$ is a $\sigma$-algebra and contains all open sets. Since $\mathcal{B}(X)$ is the smallest collection with these properties, $\mathcal{B}(X) \subset \mathcal{B}'(X)$. Now suppose that $A = f(E)$ where $E$ is a Borel set in $X$. Then $f^{-1}(E) = A \in \mathcal{B}(X) \subset \mathcal{B}'(X)$ and therefore $E = f^{-1}(C)$ for some Borel set $C$ in $Y$. But $f$ is one-to-one and $E = f^{-1}(A)$, so $f(E) = C$ and $f(E)$ is therefore a Borel set. This shows $\mathcal{B}'(X) \subset \mathcal{B}(X)$. ■

**Corollary 365** *If $f : Y \to Z$ is a Borel function and $h : X \to Y$ is a homeomorphism, where $X, Y, Z$ are metric spaces then $f \circ h$ is a Borel function.*

**Proof.** For any open set $U$ in $Z$, $(f \circ h)^{-1}(U) = h^{-1}(f^{-1}(U))$. Since $f^{-1}(U)$ is Borel and $h^{-1}$ is a homeomorphism, the proof is complete. ■

Note that a nonsingular linear function $\mathbf{L}$ has a linear, hence continuous, inverse, and hence is a homeomorphism.

**Theorem 366** *Let $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^n$ be nonsingular. For any nonnegative Borel function $f : E \to \mathbb{R}$, where $E \subset \mathbb{R}^n$ is a Borel set, $\int_E f = |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} (f \circ \mathbf{L})$. In particular for any Borel set $E \subset \mathbb{R}^n$ we have $\mu(\mathbf{L}(E)) = |\det \mathbf{L}|\, \mu(E)$.*

**Proof.** Since $\mathbf{L}$ is a homeomorphism, Lemma 364 and Corollary 365 imply that the set $\mathbf{L}^{-1}(E)$ and the function $f \circ \mathbf{L}$ are both Borel. Therefore the integral in the right side of the equation is defined. We will show that the function that assigns to $E$ and $f$ the function $I_E(f) := |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} (f \circ \mathbf{L})$ satisfies the properties of the Lebesgue integral, which, according to Theorem 247, is uniquely determined by these properties. Positivity is clear. For any constant $c > 0$ and nonnegative functions $f, g : E \to \mathbb{R}$ we have

$$I_E(cf + g) = |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} ((cf + g) \circ \mathbf{L}) = |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} (cf \circ \mathbf{L} + g \circ \mathbf{L})$$

$$= c\,|\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} f \circ \mathbf{L} + |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} g \circ \mathbf{L} = cI_E(f) + I_E(g)$$

$$= cI_E(f) + I_E(g).$$

If $c = 0$ the proof follows from Proposition 262.

Countable set additivity is an exercise.

Now let $\mathbf{v} \in \mathbb{R}^n$; it is easy to check that $\mathbf{L}^{-1}(E+\mathbf{v}) = \mathbf{L}^{-1}(E) + \mathbf{L}^{-1}(\mathbf{v})$. If $h(\mathbf{x}) = f(\mathbf{x} - \mathbf{v})$ then

$$h(\mathbf{L}(\mathbf{x})) = f(\mathbf{L}(\mathbf{x}) - \mathbf{v}) = f(\mathbf{L}(\mathbf{x} - \mathbf{L}^{-1}(\mathbf{v}))$$

and applying translation invariance to the functions $h \circ \mathbf{L}$ and $f \circ \mathbf{L}$ we have

$$I_{E+\mathbf{v}}(h) = |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E+\mathbf{v})} (h \circ \mathbf{L}) = |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)+\mathbf{L}^{-1}(\mathbf{v})} (h \circ \mathbf{L})$$
$$= |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} (f \circ \mathbf{L}) = I_E(f).$$

To show normalization, suppose first that $\mathbf{L}$ is elementary. According to Formula (5.11) we have

$$I_{Q^n}(1) = |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(Q^n)} (1) = |\det \mathbf{L}| \, \mu(\mathbf{L}^{-1}(Q^n)) = |\det \mathbf{L}| \frac{1}{|\det \mathbf{L}|} = 1. \tag{5.12}$$

This completes the proof of the theorem for any elementary transformation $\mathbf{L}$, and in particular we have that for any Borel set $E$,

$$\mu(\mathbf{L}(E)) = |\det \mathbf{L}| \, \mu(E). \tag{5.13}$$

But any linear transformation is a composition of elementary linear transformations, and since the determinant of the composition is the product of the determinants, Formula (5.13) implies that Formula (5.11) holds for an arbitrary linear transformation. Now Equation (5.12) holds, and finishes the proof. ■

**Corollary 367** *Let $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^n$ be nonsingular. For any integrable Borel function $f : E \to \mathbb{R}$, where $E \subset \mathbb{R}^n$ is a Borel set, $\int_E f = |\det \mathbf{L}| \int_{\mathbf{L}^{-1}(E)} (f \circ \mathbf{L})$.*

**Exercise 245** *Finish the proof of the above theorem.*

**Exercise 246** *Prove the above corollary.*

Recall from linear algebra that a *rotation* is an orthogonal linear function from $\mathbb{R}^n$ to itself that preserves orientation. An orthogonal linear function, in turn, is one that preserves the standard dot product, and such functions have determinant $\pm 1$. Therefore:

**Corollary 368** *Any orthogonal linear function, and in particular any rotation, preserves measure, i.e., if $f : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal and $A \subset \mathbb{R}^n$ is Borel then $\mu(A) = \mu(f(A))$.*

Another fact from linear algebra is that any isometry of $\mathbb{R}^n$ is the composition of an orthogonal linear function and a translation. We already know that translation preserves measure, and hence we have:

**Corollary 369** *Any isometry of* $\mathbb{R}^n$ *preserves measure.*

**Exercise 247** *The parallelepiped spanning* $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *is*

$$P(\mathbf{v}_1, ..., \mathbf{v}_n) := \left\{ \sum_{i=1}^{n} t_i \mathbf{v}_i : 0 \leq t_i \leq 1 \right\}.$$

*Show that* $\mu(P(\mathbf{v}_1, ..., \mathbf{v}_n)) = |\det \mathbf{M}|$, *where* $\mathbf{M}$ *is the matrix having* $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *as rows. Hint: Find a linear function that takes the unit cube to* $P(\mathbf{v}_1, ..., \mathbf{v}_n)$.

Another important type of linear function defined on $\mathbb{R}^n$ is *rescaling*. A rescaling of $\mathbb{R}^n$ with factor $s > 0$ is the function $\rho_s : \mathbb{R}^n \to \mathbb{R}^n$ given by $\rho_s(\mathbf{v}) = s\mathbf{v}$. The matrix for this linear function consists of $s\mathbf{I}$, where $\mathbf{I}$ is the identity matrix; that is, $s\mathbf{I}$ has $s$ for the diagonal entries and 0's elsewhere. The determinant of this function is the product of the diagonal elements, or $s^n$. We have proved:

**Corollary 370** *For any* $s > 0$ *and Borel set* $A \subset \mathbb{R}^n$, $\mu(\sigma_s(A)) = s^n \mu(A)$.

## 5.9 Change of Variables

In this section we will investigate the following important idea: if a $f : U \to \mathbb{R}^n$, $U \subset \mathbb{R}^n$ is $C^1$ then "locally" (that is, in some possibly small open set containing that point), the function behaves approximately like its derivative in terms of measure. While there are related results concerning points where the linear function $D(f)$ is singular, we will only consider the case when $\det D(f) \neq 0$ and the Inverse Function Theorem implies that $f$ is open and locally a homeomorphism. We start with the following proposition, which extends Theorem 345 and is used in the proof of the main theorem.

**Proposition 371** *Let* $f : U \to \mathbb{R}^n$ *be a regular function, where* $U \subset \mathbb{R}^n$ *is open, and let* $A \subset U$ *be compact. Given* $\mathbf{z} \in U$, *define* $g_{\mathbf{z}} := Df(\mathbf{z})^{-1} \circ f : U \to \mathbb{R}^n$. *For every* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that if* $\mathbf{z}, \mathbf{x} \in A$ *and* $\mathbf{y} \in U$ *such that* $\|\mathbf{x} - \mathbf{y}\| < \delta$, *then*

$$\|g_{\mathbf{z}}(\mathbf{y}) - g_{\mathbf{z}}(\mathbf{x}) - Dg_{\mathbf{z}}(\mathbf{x})(\mathbf{y} - \mathbf{x})\| < \varepsilon \|\mathbf{x} - \mathbf{y}\|.$$

**Proof.** By Proposition 352 there is some $k > 0$ such that for all $\mathbf{z} \in A$ and $\mathbf{v} \in \mathbb{R}^n$,

$$\left\| Df(\mathbf{z})^{-1}(\mathbf{v}) \right\| \leq k \|\mathbf{v}\|.$$

According to Theorem 345 there is some $\delta > 0$ such that if $\|\mathbf{x} - \mathbf{y}\| < \delta$ with $\mathbf{x} \in A$ then

$$\|f(\mathbf{y}) - f(\mathbf{x}) - Df(\mathbf{x})(\mathbf{y} - \mathbf{x})\| < \frac{\varepsilon \|\mathbf{x} - \mathbf{y}\|}{k}.$$

For all such $\mathbf{x}, \mathbf{y}$ and $\mathbf{z} \in A$ we then have $Dg_{\mathbf{z}}(\mathbf{x}) = Df(\mathbf{z})^{-1} \circ Df(\mathbf{x})$ and

$$\|g_{\mathbf{z}}(\mathbf{y}) - g_{\mathbf{z}}(\mathbf{x}) - D(g_{\mathbf{z}})(\mathbf{x})(\mathbf{y} - \mathbf{x})\|$$

$$= \left\| Df(\mathbf{z})^{-1} \left( f(\mathbf{y}) - f(\mathbf{x}) - Df(\mathbf{x})(\mathbf{y}-\mathbf{x}) \right) \right\|$$
$$\leq k \left\| f(\mathbf{y}) - f(\mathbf{x}) - Df(\mathbf{x})(\mathbf{y}-\mathbf{x}) \right\| < \varepsilon \left\| \mathbf{x}-\mathbf{y} \right\|.$$

■

**Theorem 372** *Let $f : U \to \mathbb{R}^n$ be a $C^1$ function on an open set $U \subset \mathbb{R}^n$ such that $D(\mathbf{x}) := |\det D(f)(\mathbf{x})| \neq 0$ for all $\mathbf{x} \in U$. For any compact set $A \subset U$ and $\rho > 0$ there exists a $\delta > 0$ such that if $Q$ is a semicube of radius $r < \delta$ centered at $\mathbf{x} \in A$,*

$$\left| \frac{\mu\left(f(Q)\right)}{(2r)^n} - D(\mathbf{x}) \right| < \rho.$$

**Proof.** Let $\lambda > 0$ be small enough that

$$1 - \rho < (1-\lambda)^n < (1+\lambda)^n < 1 + \rho.$$

According to Proposition 371 there exists a $\delta > 0$ such that if $\mathbf{z}, \mathbf{x} \in A$ and $\mathbf{y} \in U$ such that $\|\mathbf{x}-\mathbf{y}\| < \delta$, then

$$\left\| g_{\mathbf{z}}(\mathbf{y}) - g_{\mathbf{z}}(\mathbf{x}) - Dg_{\mathbf{z}}(\mathbf{x})(\mathbf{y}-\mathbf{x}) \right\| < \frac{\lambda}{\sqrt{n}} \left\| \mathbf{x}-\mathbf{y} \right\|.$$

where $g_{\mathbf{z}} := Df(\mathbf{z})^{-1} \circ f$. Applying Proposition 115, we have that if $\|\mathbf{x}-\mathbf{y}\|_{\max} < \delta$, then $\|\mathbf{x}-\mathbf{y}\| < \delta$ and

$$\left\| g_{\mathbf{z}}(\mathbf{y}) - g_{\mathbf{z}}(\mathbf{x}) - Dg_{\mathbf{z}}(\mathbf{x})(\mathbf{y}-\mathbf{x}) \right\|_{\max} < \frac{\lambda}{\sqrt{n}} \left\| \mathbf{x}-\mathbf{y} \right\| \leq \lambda \left\| \mathbf{x}-\mathbf{y} \right\|_{\max}. \quad (5.14)$$

Now fix any $\mathbf{x} \in A$. Since Lebesgue measure and distance are both translation invariant, by composing $f$ with translations we need only consider the case $\mathbf{x} = \mathbf{0}$ and $f(\mathbf{0}) = \mathbf{0}$. Taking $\mathbf{z} = \mathbf{x}$, we have $Dg_{\mathbf{0}}(\mathbf{0})) = Df(\mathbf{0})^{-1} Df(\mathbf{0}) = \mathbf{I}$ and Formula (5.14) simplifies to

$$\left\| g_{\mathbf{0}}(\mathbf{y}) - \mathbf{y} \right\|_{\max} < \lambda \left\| \mathbf{y} \right\|_{\max} \quad (5.15)$$

when $\|\mathbf{y}\|_{\max} < \delta$. If $r < \delta$ and $\mathbf{y} \in C_{\max}(\mathbf{0}, r)$ then

$$\left\| g_{\mathbf{0}}(\mathbf{y}) \right\|_{\max} \leq \left\| g_{\mathbf{0}}(\mathbf{y}) - \mathbf{y} \right\|_{\max} + \left\| \mathbf{y} \right\|_{\max} < (1+\lambda)\, r$$

which proves

$$g_{\mathbf{0}}(C_{\max}(\mathbf{0}, r)) \subset C_{\max}\left(\mathbf{0}, (1+\lambda)\, r\right).$$

From this we conclude that

$$\mu\left(g_{\mathbf{0}}(C_{\max}(\mathbf{0}, r)) \leq \left(2\left(1+\lambda\right) r\right)^n = (1+\lambda)^n (2r)^n < (1+\rho)(2r)^n. \quad (5.16)$$

Since $g_{\mathbf{0}} = Df(\mathbf{0})^{-1} \circ f$, we obtain from Theorem 366 that

$$\mu(f(C_{\max}(\mathbf{0}, r)) \leq (1+\rho) D(\mathbf{0}) (2r)^n.$$

We next will prove that

$$C := C_{\max}(\mathbf{0}, (1-\lambda)\, r) \subset g_{\mathbf{0}}(B_{\max}(\mathbf{0}, r)). \tag{5.17}$$

Calculating as in Formula (5.16) we then obtain

$$\mu\left(f(B_{\max}(\mathbf{0}, r)\right) \geq (1-\rho) D(\mathbf{0})(2r)^n \tag{5.18}$$

to finish the proof of the theorem.

To prove Formula (5.17)9, suppose first that $\|\mathbf{y}\|_{\max} = r$, so

$$\|g_{\mathbf{0}}(\mathbf{y})\|_{\max} \geq \|\mathbf{y}\|_{\max} - \|g_{\mathbf{0}}(\mathbf{y}) - \mathbf{y}\|_{\max} > r - \frac{\lambda}{2} r = (1 - \frac{\lambda}{2}) r.$$

That is, if $\|\mathbf{y}\|_{\max} = r$ then $g_{\mathbf{0}}(\mathbf{y}) \notin C$ and

$$A := g_{\mathbf{0}}(B_{\max}(\mathbf{0}, r)) \cap C = g_{\mathbf{0}}(C_{\max}(\mathbf{0}, r)) \cap C.$$

Now according to Corollary 354, $g_{\mathbf{0}}$ is open, so $A$ is open in $C$. On the other hand, $g_{\mathbf{0}}(C_{\max}(\mathbf{0}, r))$ is compact, hence closed, and so $A$ is also closed in $C$. Since $g_{\mathbf{0}}(\mathbf{0}) = \mathbf{0}$, $A$ is also non-empty, and since $C$ is connected, $C = A \subset g_{\mathbf{0}}(B_{\max}(\mathbf{0}, r))$. ■

**Proposition 373** *Let $\phi : U \to \mathbb{R}^n$ be a one-to-one $C^1$ function, where $U \subset \mathbb{R}^n$ is open and for all $\mathbf{x} \in U$, $D(\mathbf{x}) := |\det D\phi(\mathbf{x})| \neq 0$. Let $E \subset U$ be an $F_\sigma$ set and $F := \phi(E)$. Then*

$$\mu(F) = \int_E D.$$

**Proof.** We will prove the proposition through a sequence of weakening assumptions on $E$: starting with $E$ being a semicube. In this case consider the sequence $\{\mathcal{K}_m\}_{m=1}^{\infty}$ from Theorem 274, where $\mathcal{K}_m$ is a collection $\{Q_{mi}\}_{i=1}^{n_m}$ of disjoint semicubes that subdivide the semicube $E$. Define a simple function $s_m := \sum D(\mathbf{x}_{mi}) \chi_{Q_{mi}}$, where $\mathbf{x}_{mi}$ is the center of $Q_{mi}$. Since $D$ is continuous, for any $\mathbf{x} \in E$ and $\varepsilon > 0$, for large enough $m$, $|D(\mathbf{x}) - D(\mathbf{x}_{mi})| < \varepsilon$, where $\mathbf{x} \in Q_{mi}$. In other words, $s_m \xrightarrow{p} D$. According to Theorem 372, given $\varepsilon > 0$, for large enough $m$ we have that

$$\left| \frac{\mu(\phi(Q_{mi}))}{\mu(Q_{mi})} - D(\mathbf{x}_{mi}) \right| \leq \varepsilon$$

for all $i$. Now $F$ is the disjoint union of the sets $\phi(Q_{mi})$ for every $m$ and we have

$$\mu(F) = \sum_i \mu(\phi(Q_{mi})) \leq (1+\varepsilon) \sum_i D(\mathbf{x}_{mi}) \mu(Q_{mi}) = (1+\varepsilon) \int_E s_m.$$

Applying the Lebesgue Dominated Convergence Theorem (each $s_m$ is bounded above by the maximum of $D$ on the compact set $\overline{E}$), we have $\mu(F) \leq (1+\varepsilon) \int_E D$. A similar argument shows that $\mu(F) \geq (1-\varepsilon) \int_E D$ and this first step is finished.

Now suppose $E$ is an arbitrary compact set. Let $\varepsilon > 0$. Using Theorem 274 we can cover $E$ by disjoint semicubes $Q_i$ such that if $A := \bigcup_i Q_i$ then $\mu(A \backslash E) < \frac{\varepsilon}{M}$ and therefore $\int_{A\backslash E} D \le \varepsilon$. Now

$$\mu(F) \le \sum_i \mu(\phi(Q_i)) = \sum_i \int_{Q_i} D = \int_A D = \int_E D + \int_{A\backslash E} D \le \int_E D + \varepsilon$$

and since $\varepsilon$ was arbitrary we obtain

$$\mu(F) \le \int_E D.$$

Applying this to the set $A\backslash E$ we have $\mu(\phi(A\backslash E)) \le \int_{A\backslash E} D \le \varepsilon$. Since $\phi$ is one-to-one we have that $\phi(A)$ is the disjoint union of $F$ and $\phi(A\backslash E)$ and therefore

$$\mu(F) = \mu(\phi(A)) - \mu(\phi(A\backslash E)) \ge \sum_i \mu(\phi(Q_i)) - \varepsilon = \int_A D - \varepsilon \ge \int_E D - \varepsilon.$$

Finally, if $E$ is an $F_\sigma$ then according to Lemma 291 $E$ is the increasing union of compact sets $E_i$. Since $\phi$ is injective, $F$ is the increasing union of compact sets $F_i := \phi(E_i)$. We have

$$\mu(F) = \lim \mu(F_i) = \lim \int_{E_i} D = \int_E D.$$

■

**Theorem 374** *(Change of Variables) Let $\phi : U \to \mathbb{R}^n$ be a one-to-one $C^1$ function, where $U \subset \mathbb{R}^n$ is open and for all $\mathbf{x} \in U$, $D(\mathbf{x}) := |\det D\phi(\mathbf{x})| \ne 0$. Let $E \subset U$ be closed, $F := \phi(E)$, and $f : F \to \mathbb{R}$ be a Borel function. Then*

$$\int_F f = \int_E [(f \circ \phi)\, D]$$

*provided either integral exists.*

**Proof.** Suppose first that $f = \chi_B$ where $B \subset F$ is an $F_\sigma$–so also $A = \phi^{-1}(B)$ is an $F_\sigma$. By Proposition 373,

$$\int_F f = \mu(B) = \int_A D = \int_E \chi_{\phi^{-1}(A)} D = \int_E (f \circ \phi)\, D.$$

If $f = \sum_{i=1}^m c_i \chi_{F_i}$ is a simple function with each $F_i$ an $F_\sigma$ then the proof follows from linearity. If $f$ is a nonnegative Borel function then by Proposition 292 there exists a sequence $(f_i)$ of functions of the type just considered such that $f_i \nearrow f$ on $E$. Then since $D \ge 0$, $(f_i \circ \phi)\, D \nearrow (f \circ \phi)\, D$ and

$$\int_F f = \lim \int_F f_i = \lim \int_E (f_i \circ \phi)\, D = \int_F (f \circ \phi)\, D.$$

The rest of the proof is an exercise. ■

**Exercise 248** *Finish the proof of the Change of Variables Theorem.*

**Exercise 249** *Evaluate* $\int_E e^{xy}(x^4 - y^4)$ *where* $E$ *is the region bounded by the graphs of the functions* $xy = 1$, $xy = 6$, $x^2 - y^2 = 2$, $x^2 - y^2 = 4$.

# Bibliography


[1] J. R. Munkres, *Topology, second edition,* Prentice Hall, Upper Saddle River, NJ, 2000.

[2] W. Rudin, *Principles of Mathematical Analysis,* second edition, McGraw-Hill, New York, NY, 1964.

[3] R. Wheeden and A. Zygmund, *Measure and Integral: An Introduction to Real Analysis*, Marcel Dekker, New York, NY, 1977.

# Index