Introduction to Real Analysis
M361K

Last Updated: July 2011

Preface

These notes are for the basic real analysis class, M361K (The more advanced class is M365C.) They were written, used, revised and revised again and again over the past decade. Contributors to the text include both TA's and instructors: Grant Lakeland, Cody Patterson, Alistair Windsor, Tim Blass, David Paige, Louiza Fouli, Cristina Caputo and Ted Odell.

The subject is calculus on the real line, done rigorously. The main topics are sequences, limits, continuity, the derivative and the Riemann integral. It is a challenge to choose the proper amount of preliminary material before starting with the main topics. In early editions we had too much and decided to move some things into an appendix to Chapter 2 (at the end of the notes) and to let the instructor choose what to cover. We also removed much of the topology on $\mathbb{R}$ material from Chapter 3 and put it in an appendix. In a one semester course we are usually able to do the majority of problems from Chapter 3–6 and a small selection of certain preliminary problems from Chapter 2 and the two appendices.

July 2011

Contents

CHAPTER 1

Introduction

1. Goals

The purpose of this course is three-fold:

(1) to provide an introduction to the basic definitions and theorems of calculus and real analysis.
(2) to provide an introduction to writing and discovering proofs of mathematical theorems. These proofs will go beyond the mechanical proofs found in your Discrete Mathematics course.
(3) and most importantly to let you experience the joy of mathematics: the joy of personal discovery.

2. Proofs

Hopefully all of you have seen some proofs before. A proof is the name that mathematicians give to an explanation that leaves no doubt. The level of detail in this explanation depends on the audience for the proof. Mathematicians often skip steps in proofs and rely on the reader to fill in the missing steps. This can have the advantage of focusing the reader on the new or crucial ideas in the proof but can easily lead to frustration if the reader is unable to fill in the missing steps. More seriously these missing steps can easily conceal mistakes: many mistakes in proofs, particularly at the undergraduate level, begin with "it is obvious that".

In this course we will try to avoid missing any steps in our proofs. Each statement should follow from a previous one by a simple property of arithmetic, by a definition, or by a previous theorem, and this justification should be clearly stated in plain language. Writing clear proofs is a skill in itself. Often the shortest proof is not the clearest.

There is no mechanical process to produce a proof but there are some basic guidelines you should follow. The most basic is that every object that appears should be defined; when a variable, function, or set appears we should be able to look back and find a statement defining that object:

(1) Let $\epsilon > 0$ be arbitrary.

(2) Let $f(x) = 2x + 1$.
(3) Let $A = \{x \in \mathbb{R} : x^{13} - 27x^{12} + 16x^2 - 4 = 0\}$.
(4) By the definition of continuity there exists a $\delta > 0$ such that...

Always watch out for hidden assumptions. In a proof, you may want to say "Let $x \in A$ be arbitrary," but this does not work if $A = \varnothing$ (where $\varnothing$ denotes the empty set). A common error in real analysis is to write $\lim_{n\to\infty} a_n$ or $\lim_{x\to a} f(x)$ without first checking whether the limit exists (often the hardest part).

A key factor in determining how a proof should be written is the intended audience. For this course, your intended audience is another student in the class who is clueless as to how to prove the theorem, but who knows all the definition and the results of the course covered up to that point.

3. Logic

For the most part, we will avoid using overly technical logical notation in our definitions and statements. Instead we will use their plain English equivalents and we suggest you do the same in your proofs. Beyond switching to the contrapositive and negating a definition, formal logical manipulation is rarely helpful in proving statements in real analysis.

On the other hand, you should be familiar with the basic logical operators and we will give a short review here. If P and Q are **propositions**, i.e., statements that are either true or false, then you should understand what is meant by

(1) not P
(2) P or Q (the mathematical use of "or" is not exclusive so that "P or Q" is considered true if both P and Q are true).
(3) P and Q
(4) if P then Q (or "P implies Q")
(5) P if and only if Q (sometimes written P is equivalent to Q)

Similarly if $P(x)$ is a **predicate**, that is a statement that becomes a proposition when an object such as a real number is inserted for x, then you should understand

(1) for all x, $P(x)$ is true
(2) there exists an x such that $P(x)$ is true

Simple examples of such a $P(x)$ are "$x > 0$" or "x^2 is an integer." These statements are true for some values of x and not for others.

You should also be familiar with the formulae for negating the various operators and quantifiers.

Most of our theorems will have the form of implications: "if P then Q." P is called the **hypothesis** and Q the **conclusion**.

Definition. The **contrapositive** of the implication "if P then Q" is the implication "if not Q then not P."

The contrapositive is logically equivalent to the original implication. This means that one is valid (true) if and only if the other is valid. Sometimes it is easier or better to pass to the contrapositive formulation when proving a theorem.

Definition. The **converse** of the implication if P then Q is the implication if Q then P.

The converse is *not* logically equivalent to the original implication and this fact is the underlying source of error in many undergraduate proofs.

Definition. A statement that is always true is called a **tautology**. A statement that is always false is called a **contradiction**.

To show that an argument is not valid it suffices to find one situation in which the hypotheses are true but the conclusion is false. This type of situation is called a **counterexample**.

One technique of proof is by contradiction. To prove "P implies Q" we might assume that P is true and Q is false and obtain a contradiction. Whenever you use contradiction, it is usually a good idea to see if you can rephrase your proof in a way that does not use contradiction. Often times, contradiction is not necessary and avoiding its use can lead to cleaner (and more understandable) proofs.

You will notice that we mostly spoke in great generality in the preceding review. If at any point you found yourself confused in regards to the meaning of a definition or the purpose of a concept, you should create some examples of the situation being described. Be as concrete as you can. In fact, as we will emphasize later, you should use this strategy any time you are confused when reading these notes (or any mathematics textbook).

CHAPTER 2

Preliminaries: Numbers and Functions

What exactly is a number?

If you think about it, to give a precise answer to this question is surprisingly difficult. As is often the case, the word 'number' reflects a concept of which we have some intuitive understanding, but no concrete definition. In this introduction, we will attempt to pin down what we mean by a number by describing exactly what we should expect from a number system. In fact, though we will not prove it, the only collection that satisfies are of our demands is $\mathbb{R}$, the collection of **real numbers**. Thus we conclude that a number is an element of the set $\mathbb{R}$. Just as with numbers, most of us have probably heard the term 'real numbers,' but may not be exactly sure what they are. Studying real numbers will be one of the important purposes of this course.

As mentioned above, we all know the things that we should expect from a number system. Think back to when you first met the idea of a number. Probably the very first purpose of numbers in your life is that they allowed you to count things: 50 states, 32 professional football teams, 7 continents, 5 golden rings, etc. Needing to count things leads us to the invention (or discovery depending on your point of view) of the natural numbers (the numbers $1, 2, 3, 4, 5, \cdots$). Mathematicians typically denote the collection of natural numbers by the symbol '$\mathbb{N}$.' Though this collection can be constructed quite rigorously from the standard axioms of mathematics, we will assume that we are all familiar with the natural numbers and their basic properties (such as the concept of mathematical induction; see the appendices). The natural numbers fulfill quite successfully our goal of being able to count.

The next thing that we expect of our number system is that it should be able to answer questions like the following: "If the Big Twelve has 10 football teams and the Big Ten has 12 (shockingly it's true), how many teams do the conferences have between them?" In other words we will need to add. We will also multiply. The natural numbers are already well-suited for these tasks. Really this should not come as a surprise. After all, adding natural numbers is really just a different way of looking at counting (i.e., adding three and five is the same as taking three dogs and five cats and counting the total number of animals). As we all know, multiplication of natural numbers is really just repeated addition.

Having addition naturally leads us to subtraction. This is the first place the natural numbers will fail us. Subtracting 7 from 2 is an operation that cannot be performed within $\mathbb{N}$. The need for subtraction, therefore, is one of the reasons that $\mathbb{N}$ will not work as our entire number system. Thus we need to expand the set of natural numbers to the integers. As we all probably know, the integers are comprised of the natural numbers, the number zero, and the negatives of the natural numbers (at this point, you might protest and say that zero should be included as a natural number as it allows us to count collections which contain no objects; in fact many mathematicians do include zero in $\mathbb{N}$, but the distinction is of little importance). The collection of integers is denoted by $\mathbb{Z}$. Again we will assume we know all the basic properties of $\mathbb{Z}$.

The integers are a very good number system for most purposes, but they still have an obvious defect: we cannot divide. Surely any reasonable number system allows division: if you and I have a sandwich and we each want an equal share, a number should describe the portion we each get. Needing division, we throw in fractions: symbols which are comprised of two integers, one in the numerator and one in the denominator (of course the denominator is not allowed to be zero). A fraction will represent the number which results when the numerator is divided by the denominator.

Combining all the numbers we have so far gives $\mathbb{Q}$, the collection of rational numbers. Again, we will assume that we are familiar with all its basic properties. Before we go on to justify our assertion that $\mathbb{Q}$ is not a sufficient number system, we have another property to point out. Notice that most of our properties so far involve **operations** among our numbers: namely addition, subtraction, multiplication, and division. We call these types of properties **algebraic** (in mathematics, the word **algebra** describes the study of operations). The property we are going to discuss next is not algebraic.

Suppose then that I pick a rational number and you pick another. We can easily decide which is bigger: Namely $\frac{a}{b}$ is bigger than $\frac{c}{d}$ (where a, b, c, and d are integers) if ad is bigger than bc (assuming b and d both positive; we can easily assure both denominators are positive by moving any negative into the numerator). Since ad and bc are integers, we know how to compare them (because we know how to compare natural numbers and how to take negatives into account). Since we can always compare any two rational numbers in this way, we say that $\mathbb{Q}$ is totally **ordered**.

In retrospect, we should have demanded this property of our number system from the beginning. Numbers should come with some notion of size. Fortunately, we got it for free. Moreover, it is interesting to notice that our expectation that a number system should include the natural numbers and that it should have certain algebraic properties is enough to lead us to include all of $\mathbb{Q}$. We did not need to insist that our system be ordered to find $\mathbb{Q}$. The order properties turn out to be more important in telling us which potential numbers we should *not* include (such as the imaginary number i).

$\mathbb{Q}$ comes very close to satisfying everything we want in a number system. Unfortunately it is still lacking. Suppose we draw a circle whose diameter is 1. The area of a circle with radius one (usually called the **unit circle**) should certainly be a number. If, however, we restrict ourselves to the rational numbers, this area will not be a number (the number is of course usually denoted π and it is not a rational number). The same could be said of the length of one of the sides of a square whose area is 2 (this number is usually denoted $\sqrt{2}$).

These two examples merely comprise our attempt to give a (geometric) demonstration that $\mathbb{Q}$ is lacking as a number system. The real (more general) property that we seek, called 'completeness', is actually quite subtle and has to do with the presence of something like 'gaps' in $\mathbb{Q}$ (the absence of the number $\sqrt{2}$ or of the number π is an example of such a gap). These gaps have to do with something called a 'monotone sequences' which we will study in detail in this course. One consequence of filling in these gaps is that we are able to perform calculus (showing this might be viewed as the main mathematical purpose of this course). This, in turn, allows us to express all the lengths, areas, volumes, etc. of geometric objects like the examples above as real numbers.

In that we have been a little bit vague in the preceding discussion, we formulate our demands precisely in the appendices (with the exception of the completeness axiom as it is a major object of study in this course). Once again, one of the fundamental results in mathematics is that the collection of real numbers is the *only* system of numbers which satisfies all of our demands. We thus conclude that the real numbers comprise the only possible choice of a number system (at least in the sense we have given; there are a surprising number of close competitors if we relax some of our demands).

To give an exact definition of the real number is surprisingly complicated. In fact, the first rigorous construction of the real numbers was given by Georg Cantor as late as 1873 (by comparison, the rational numbers where constructed in ancient times). For our purposes, we will first take it on faith that the real numbers exist as a number system and that they satisfy the demands we have described. Later in the book (towards the ends of Chapter 3), we will describe a way to define the real numbers rigorously using decimal expansions (there are actually several well-known ways and the way we choose, though perhaps the most famous, was not the first).

Finally, it is important to realize that the properties given in the appendix (which we will call **axioms**), together with the completeness axiom, are the *only* properties that we assume about $\mathbb{R}$. Strictly speaking, any other statement we want to make must be proven from either from our axioms or from properties we have already assumed about $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ (or, of course, some combination of the two).

In general, however, this can get to be a little bit tedious. Hence, we will allow you to assume all of the 'basic' or 'obvious' properties of the real numbers. Unfortunately, deciding which properties are obvious is a subjective process. Therefore, if there is any doubt about whether a statement is obvious, you should prove it rigorously from the axioms (or at least describe how to prove it rigorously). Actually, the ability to decide when statements are obvious or 'trivial' is an important skill in mathematics. Possessing this ability can often be a reflection of great mathematical maturity and insight. In general, make sure you are prepared to back up all your assertions to your fellow students and to your instructor.

In the appendix to this chapter, we will also derive some properties of $\mathbb{R}$ that follow from our axioms. We may work on some of these in class, but thereafter you may consider them "known." The appendix also contains a discussion of basic set theory, induction, and some supplementary material on cardinality.

1. Functions

Although most people may not realize it, the concept of a function is far more basic to mathematics than is the concept of a number. Roughly speaking, a function f from a set A to a set B is a rule that assigns to each element of A an element of B. In this case we write $f : A \to B$. What do we mean by 'rule'? Let's try to be more precise.

Definition. A **function** f from A to B, denoted by $f : A \to B$, is a subset f of the Cartesian product $A \times B = \{(a, b) : a \in A, b \in B\}$ satisfying

(1) for each $a \in A$ there exists $b \in B$ such that $(a, b) \in f$
(2) for all $a \in A$ and for all $b, b' \in B$ if $(a, b) \in f$ and $(a, b') \in f$ then $b = b'$.

The set A is called the **domain** of f and B is called its **codomain**.

At this point, it is probably a good time for some general advice regarding definitions. When taking a rigorous proof-based courses, many students merely skim over (or ignore) the definitions and go straight to the problems (specifically the ones they have been assigned). It is our recommendation that you avoid this behavior: the more deep thinking that you do about the concepts of a course, the easier time you will have succeeding. This is not to say that thinking deeply is easy. It is typically a very difficult time consuming process, but the results can be very rewarding.

Thus rather than simply read the definitions you come across, you should attempt to think deeply about them, particularly the ones that are confusing or long. At times, when we feel that a definition is particularly confusing, we will try to help you through this process, but you should be doing it all the time.

We will give some general advice now in how to think about definitions. We will also give some more advice later in this chapter (after we have more examples of definition that we may use to illustrate our points). Throughout the course, we will emphasize the fact that most (or perhaps all) definitions have two sides to them: the precise definition in mathematical language and the intuitive notion that the definition is an attempt to express.

For example, in the present case, we, speaking intuitively, stated that a function is a rule. We then proceeded to give the precise mathematical definition. When you have both sides of the definition before you, you should ask yourself how the precise definition captures the intuitive notion. Does it capture it fully? Is anything missing? Is the precise definition more broad than the intuitive notion? Is it more narrow? Ask yourselves these questions about the definition above. We will often give a precise definition without giving an intuitive description. In those cases, you should describe for yourself the intuitive idea that is attempting to be captured.

In general, both sides of the definition are important. When giving a proof of a statement which involves a term that we have defined, you will likely need to use the precise definition in your proof. Nevertheless, it is oftentimes the intuitive notion that *leads* your to the proof. Probably more often than not, mathematicians think about a result intuitively and then write down a rigorous proof to back up their intuition.

Technically, then, a function from A to B is just a special subset of $A \times B$. Mathematicians, however, rarely think of functions in this way. Rather we think of their more intuitive notion: a rule. For this reason, we typically use a different notation when discussing functions. Explicitly, instead of saying that the point (a, b) is an element of our function, we write $f(a) = b$. Translating the definition of a function into this notation gives the following definition for function:

> for each $a \in A$ there exists a unique $b \in B$ such that $f(a) = b$

You should see for yourself why this statement is the same as the above definition.

Reflecting the intuitive notion that they capture, a function is sometimes called a **mapping** or a **transformation**. Correspondingly, if $f(a) = b$, one might say "f maps a to b" or "f sends a to b." We should point out, however, that a function is more than just a rule. It actually has three ingredients: the domain, the codomain, and a rule which sends elements of A to elements of B. For example the function $f : [0, 1] \to \mathbb{R}$ given by $f(x) = x^2$ and the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x^2$ are not the same.

As our next bit of general advice regarding definitions, we point out that some definitions, like the previous one, define terms, such as 'function,' that we have all heard before. If you come across a definition that you have already learned, you should compare the definition given with idea in your own head on the other. Are the two notions the same? Is the definition given more general than the one in your head? Is it more specific?

Along these lines, many beginning students believe that a function is the same things a formula. In other words, many students think that in order to specify a function, they need to find a formula using variables. This is not the case. It is perfectly reasonable to define a function by saying something like:

> Define a function from the set of real numbers to the set $\{0, 1\}$ by assigning the value 1 to all rational numbers and the value 0 to all irrational numbers.

Since every number has been given a value and no number has been given more then one, our rule gives a function.

This function $f : \mathbb{R} \to \{0,1\}$ would probably be more commonly described by saying that for $x \in \mathbb{R}$,

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ \\ 0 & \text{if } x \in \mathbb{R}\backslash\mathbb{Q} \end{cases},$$

but either description would work. Notice that it would be essentially impossible to find what most people would call a 'formula' to describe this function.

You should come up with some examples of this kind on your own. In other words, give some examples of functions that don't have a formula in this sense.

We now give define some more terms related to general functions.

Suppose that $f : A \to B$ is a function and suppose that S is a subset of A. We can define a new function $\widetilde{f} : S \to B$ by using same rule as for f but by restricting ourselves to points in S. That is, $\widetilde{f}$ is defined by $\widetilde{f}(x) = f(x)$ for $x \in S$. $\widetilde{f}$ is called the **restriction** of f to S and is usually denoted $f|_S$.

Definition. If $f : A \to B$ is a function, the **range** of f, denoted by $f(A)$, is

$$f(A) = \{f(a) : a \in A\}.$$

Some functions have certain important properties that we shall name.

Definition. Let $f : A \to B$.

(1) f is **surjective** (or **onto**) if $f(A) = B$. That is, f is onto if, for every $b \in B$, there exists some $a \in A$ such that $f(a) = b$.
(2) f is **injective** (**one-to-one** or **1–1**) if for all $a_1, a_2 \in A$, $f(a_1) = f(a_2)$ implies $a_1 = a_2$. That is, f is 1–1 if, for all $a_1, a_2 \in A$ such that $a_1 \neq a_2$, we have $f(a_1) \neq f(a_2)$.
(3) f is a **bijection** (or a **1–1 correspondence**) if it is 1–1 and onto. This is equivalent to: for all $b \in B$ there exists a unique $a \in A$ with $f(a) = b$. (Note: it is usually simpler to show that a function is a bijection by showing it is 1–1 and onto separately.)

We point out that these notions depend on the domain and codomain of the function just as much as the depend on the rule.

2.1. Give an example of a function, $f : \mathbb{R} \to \mathbb{R}$, that is not injective, but becomes injective when restricted to a smaller set.

If a function is a bijection then you can 'reverse it' to obtain a function going the other way. The following theorem makes this precise.

2.2. Let $f : A \to B$ be a bijection. Then there exists a bijection $g : B \to A$ satisfying

(1) for all $a \in A$, $g(f(a)) = a$.
(2) for all $b \in B$, $f(g(b)) = b$.

Furthermore (and this is still part of the problem), this function g is unique; if g_1 and g_2 are bijections satisfying (1) and (2) then $g_1 = g_2$. (Would it be enough to assume g_1 and g_2 both satisfy (1)?)

The bijection g is called the **inverse function** of f and is usually denoted by f^{-1}. Do not confuse this with "$1/f$" (which would mean what?)

Definition. Let $f : A \to B$. Let $D \subseteq A$, and $C \subseteq B$.

(1) The **image** (or **direct image**) of D under f, denoted $f(D)$, is
$$f(D) = \{f(x) : x \in D\}.$$
(2) The **pre-image**, of C under f, denoted $f^{-1}(C)$, is
$$f^{-1}(C) = \{a \in A : f(a) \in C\}.$$

The definition above might lead to some temporary confusion in that we are using the symbol f^{-1} in two different ways. Explicitly, if the function f^{-1} exists, which is not always the case, then there are two different ways of reading $f^{-1}(C)$: it can be read as the direct image of the set C under the function f^{-1} or it can be read as the inverse image of C under the function f. Check for yourself that these two interpretations give the same set and so there is no ambiguity. We point out that $f^{-1}(C)$ always exists even if the function f^{-1} does not.

2.3. Let P be the collection of nonnegative real numbers. Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ that meets each of the following criteria. You can (and will have to) use different functions for different examples.

(1) $f^{-1}(P) = \varnothing$,
(2) $f(P) = \{-10, 10\}$, and

(3) there is some set $D \subset \mathbb{R}$ so that $f(f^{-1}(D)) \neq D$ (and you should specify the set D as well).

We next given an important way to combine two functions.

Definition. Let $f : A \to B$ and $g : B \to C$. The **composition** $g \circ f : A \to C$ is defined by $(g \circ f)(a) = g(f(a))$.

2.4. Let $f : A \to B$ and $g : B \to C$ be two bijections. Then $g \circ f : A \to C$ is also a bijection.

Caution: If f and g are functions, it is *not* true in general that $f \circ g = g \circ f$. In fact, these two compositions may have completely different domains and codomains!

2.5. Give an example of two function $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ such that $f \circ g \neq g \circ f$.

If $f : A \to B$ is a bijection then $f^{-1} \circ f : A \to A$ is the identity map on A and $f \circ f^{-1} : B \to B$ is the identity map on B (the identity map on a set S is the map $id : S \to S$ defined by $id(x) = x$ for all $x \in S$.) f^{-1} is the only function with these properties.

2. The Absolute Value

Now that we have given the general framework for functions, we move on to consider real numbers. In this section, we define an extremely important function on the real numbers.

Definition. Given a real number $a \in \mathbb{R}$, we define the absolute value of a, denoted $|a|$ to be a if a is nonnegative and $-a$ if a is negative.

Our definition might appear a bit strange at first. Surely we all know that the absolute value of a number should never be negative. Yet we said that the absolute value of a, at least for some values of a, is $-a$, which appears to be negative. Try some examples to figure out what is going on and explain this apparent contradiction to yourself. This sort of thinking will help you in the proof of the next result.

2.6. For $a \in \mathbb{R}$:

(1) $|a| \geq 0$,
(2) $|a| = 0$ if and only if $a = 0$,

(3) $|a| \geq a$, and
(4) $|-a| = |a|$.

Thus the absolute value gives us a function whose domain is $\mathbb{R}$ and whose codomain is the collection of non-negative real numbers.

The following technical observations will be of assistance in some arguments involving the absolute value.

2.7. For all $a \in \mathbb{R}$, $a^2 = |a|^2$.

2.8. For all $a, b \in \mathbb{R}$ with $a, b \geq 0$ we have $a^2 \leq b^2$ if and only if $a \leq b$. Likewise, $a < b$ if and only if $a^2 < b^2$.

The next statement gives two fundamental properties of the absolute value.

2.9. Let $a, b \in \mathbb{R}$, then

(1) $|ab| = |a||b|$ and
(2) $|a + b| \leq |a| + |b|$

Hint: O

ne can prove these by laboriously checking all the cases (e.g., $a > 0$, $b \leq 0$) but in each case an elegant proof is obtained by using our previous observations to eliminate the absolute value and then proceeding using the properties of arithmetic.

The second inequality above is perhaps the most important inequality in all of analysis. It is called the **triangle inequality**.

The remaining results in this section are important consequences of the triangle inequality.

2.10. Let $a, b, c \in \mathbb{R}$. Then we have

(1) $|a - b| \geq \big||a| - |b|\big|$ and
(2) $|a - c| \leq |a - b| + |b - c|$.

The major importance of the absolute value is that it will give us some notion of 'distance' or 'length.' Indeed, you have probably measured the length of something using a yardstick. Needless to say, you typically line up one end of the object with zero and read the length by looking to see where the other end hits the ruler. In a tight place, however, you might line up one end at $7''$ and the other at $13''$. What is the length of the object in this case? Of course it is $13'' - 7'' = 6''$.

A bit more abstractly, if I told you one end was at x and the other was at y, what would be the length? Well, $y - x$ if $y > x$ and $x - y$ if $x > y$. In other words, it would be $|x - y|$. So, we can regard $|x - y|$ as the **distance** between the numbers x and y. Note this works for all real numbers, even if one or both is negative. It also explains why we call the inequality in Problem 2.9(2) the triangle inequality (why? See Problem 2.10(2)).

The statement of Problem 2.10(1) is often called the **reverse triangle inequality**.

2.11. Let $x, \epsilon \in \mathbb{R}$ with $\epsilon > 0$. Then:

(1) $|x| \leq \epsilon$ if and only if $-\epsilon \leq x \leq \epsilon$, where the double inequality $-\epsilon \leq x \leq \epsilon$ means $-\epsilon \leq x$ and $x \leq \epsilon$.
(2) If $a \in \mathbb{R}$, $|x - a| \leq \epsilon$ if and only if $a - \epsilon \leq x \leq a + \epsilon$.

We point out that, by a similar proof, the same properties hold with $\leq$ replaced by $<$.

3. Intervals

Intervals are a very important type of subset of $\mathbb{R}$. Loosely speaking they are sets which consist of all the numbers between two fixed numbers, called the endpoints. We also (informally) allow the endpoints to be $\pm\infty$. Depending on whether the endpoints are finite and whether we include them in our sets, we arrive at 9 different types of intervals in $\mathbb{R}$.

Definition. An interval is a set which falls into one of the following 9 categories (assume $a, b \in \mathbb{R}$ with $a < b$). We apply the word 'bounded' if both the endpoints, a and b, are finite. Otherwise we use the word 'unbounded'.

(1) Bounded open intervals are sets of the form
$$(a, b) := \{x \in \mathbb{R} : a < x < b\}.$$
(2) Bounded closed interval are sets of the form
$$[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}.$$
(3) There are two type of half-open bounded intervals. One type is sets of the form
$$[a, b) := \{x \in \mathbb{R} : a \leq x < b\}.$$

(4) The other is sets of the form
$$(a, b] := \{x \in \mathbb{R} : a < x \leq b\}.$$
(5) There are also two types of unbounded open intervals not equal to $\mathbb{R}$. One type is sets of the form
$$(a, +\infty) := \{x \in \mathbb{R} : a < x\}.$$
(6) The other is sets of the form
$$(-\infty, b) := \{x \in \mathbb{R} : x < b\}.$$
(7) There are two types of unbounded closed intervals not equal to $\mathbb{R}$. One type is sets of the form
$$[a, +\infty) := \{x \in \mathbb{R} : a \leq x\}.$$
(8) The other is sets of the form
$$(-\infty, b] := \{x \in \mathbb{R} : x \leq b\}.$$
(9) The whole real line $\mathbb{R} = (-\infty, \infty)$ is an interval. We count $\mathbb{R}$ as being open, closed, and unbounded.

Some mathematicians include the empty set, $\varnothing$, and single points, $\{a\}$ for some $a \in \mathbb{R}$, as intervals. To distinguish these special sets, people often call them 'degenerate intervals' whereas sets of the above would be 'non-degenerate intervals.' We will reserve the word 'interval' for the non-degenerate case. That is, in our language, an interval is not allowed to be $\varnothing$ or $\{a\}$.

We now give some more general advice regarding definitions. One of the first things that you should do when you come across a definition is to come up with some examples of things that fit into the definition (at least in your own head, but it might help to write down your examples). In the present case, we will get you started: an example of an interval is all the numbers between 2 and 3, not including 2 or 3. More specifically, this is an example of a bounded open interval.

You should also attempt to make your examples as interesting or weird as possible so that you can test the outer reaches of what a definitions entails. For example, if we wanted to give an example of a number, we could certainly say the number 1 or the number 2. However, more exotic examples include numbers like $\frac{1}{3}$, $-\frac{7}{3}$, π or $\pi^2 + 7$ (at least once we prove they exist). The benefit of giving weird examples is that we can catch ourselves thinking too narrowly about a concept. If we say the word "number," many people think of $1, 2, 3, 4, 5 \ldots$, but in act there are many more type numbers and forgetting this fact can sometimes lead to trouble.

Another way of looking at this bit of advice is that you should come up with examples that are not basically the same. In other words, if you want to create three examples of numbers, it would suffice to say 1, 2, and 3, but it might be more informative of the nature of numbers to say 0, $-\frac{21}{8}$ and $-e^2+\frac{1}{2}+\pi$. In fact, it is important to generate both ‘easy’ or normal examples and ‘weird’ elaborate ones.

Perhaps equally important is to think of examples that do NOT fall into the definition. An easy way to do this is to name something that doesn’t have anything to do with the definition. For example, the University of Texas football team is not an interval and neither is your roommate. However, it might be a better idea to come up with some examples that are close to the definition, but don’t fall into it. For example, can you think of some subsets of the real numbers that are not intervals? Try to come up with some cheap examples and some clever ones. In general, try to get as close as you can to the definition without satisfying it. If a definition has two parts, try to come up with an example which fits one part but not the other.

Another question you should ask yourself: “why does this definition exist?” Why is the concept so important that generations of mathematicians have agree it should have a name? This question is not always to easy to answer, especially if you haven’t seen the definition used a few times. Nevertheless, you should keep the question in mind as you go through the exercises and results surrounding the definition. In general the answer to this question can be closely related to the intuitive notion behind the definition.

You will notice that we actually gave this type of explanation for the absolute value: we said that it is important as way to measure length or size. We won’t spell out exactly why intervals were given a name (you should try to come up with some reasons on your own), but we will say a word about why the notation exists. If we want to specify a set which includes exactly “all the numbers between 3 and π,” we notice that the English phrase necessary is a bit long. It’s also ambiguous: do we want to conclude 3 and π or leave them out? Thus to be clear, we really have to say something like “all the numbers between 3 and π, including 3 but not including π.”

This terminology is definitely getting very cumbersome and so mathematicians have found it convenient to replace all these words with the symbols $[3, \pi)$. This is definitely much shorter, but it comes at a price: the meaning of the symbols might not be obvious to somebody who was already familiar with them and, more importantly, the notation might disguise some subtlety in the definition. This balance is one that any mathematician has to strike for himself or herself.

2.12. Let $a \in \mathbb{R}$ and $\epsilon > 0$. Write the set

$$\{x \in \mathbb{R} : \ |x - a| \leq \epsilon\}$$

as an interval. Write

$$\{x \in \mathbb{R} : \ |x - a| < \epsilon\}$$

as an interval.

Definition. The closure of an interval I, denoted $\overline{I}$, is the union of I and its finite endpoints.

Thus, for $a < b$,

$$\begin{gathered}
\overline{(a,b)} = \overline{[a,b]} = \overline{[a,b)} = \overline{(a,b]} = [a,b] \\
\overline{(a,+\infty)} = \overline{[a,+\infty)} = [a,+\infty) \\
\overline{(-\infty,b)} = \overline{(-\infty,b]} = (-\infty,b] \\
\overline{\mathbb{R}} = \mathbb{R}
\end{gathered}$$

Definition. The interior of an interval I, denoted I°, is I minus its endpoints.

Thus

$$\begin{gathered}
(a,b)^\circ = [a,b]^\circ = [a,b)^\circ = (a,b]^\circ = (a,b) \\
(a,+\infty)^\circ = [a,+\infty)^\circ = (a,+\infty) \\
(-\infty,b)^\circ = (-\infty,b]^\circ = (-\infty,b) \\
\mathbb{R}^\circ = \mathbb{R}
\end{gathered}$$

Now that we have given all the basic terminology, we will begin our study of some of the deeper properties of numbers.

CHAPTER 3

Sequences

1. Limits and the Archimedean Property

Our first basic object for investigating real numbers is the sequence. Before we give the precise definition of a sequence, we will give the intuitive description. To begin, a **finite sequence** is just a finite ordered list of real numbers. For example, $(1, 2, 3, 4, 5)$ is a sequence with five terms and $(\pi, e, 3, 3)$ is a sequence with four.

We added the adjective 'ordered' above to reflect the fact that the order of the terms matters. For example, the 2 term sequence $(\pi, 7)$ is not the same as the 2 term sequence $(7, \pi)$. Notice that a sequence is not the same as a set for which order (and also repetition) do not matter. The sets $\{7, \pi\}$ and $\{\pi, 7\}$ are the same. Likewise the sets $\{3\}$ and $\{3, 3\}$ are the same, whereas the sequences (3) and $(3, 3)$ are not.

So far we have been discussing finite sequences, but for this course, these objects will not be the focus of our study. For us, a sequence will always be an infinite sequence. As you have probably guessed, an infinite sequence is essentially an infinite ordered list of real numbers. We now give the precise definition.

Definition. A **sequence** is a function $f : \mathbb{N} \to \mathbb{R}$.

At this point you decide for yourself whether the definition we have given captures the intuitive idea we were seeking. Giving a precise definition for an intuitive idea is a very important skill in studying mathematics.

For example, the functions defined by $f(n) = n^2$ and $g(n) = \frac{1}{n}$ are both sequences.

Although a sequence is technically a function, we typically do not use functional notation to discuss sequences as this notation does not reflect the intuitive notion we are attempting to describe. Instead of writing something like $f(n) = n^2$ or $g(n) = \frac{1}{n}$ to denote a sequence, we write $(n^2)_{n=1}^{\infty}$ or $(\frac{1}{n})_{n=1}^{\infty}$. If we say "consider the sequence $(a_n)_{n=1}^{\infty}$," we are referring to the sequence whose value at n is the real number a_n (in other words the sequence whose nth term is a_n).

Like any function, a sequence does not have to be defined via an elementary formula: any random list of numbers will work. For example, a sequence could be defined by saying the nth term is the nth decimal of π (we have not actually defined decimal expansions yet, but this infinite list of numbers certainly does not follow an elementary formula).

One good way to define a sequence is **recursively**: one states the first term (or the first several terms) and then gives a rule for getting each term from the previous ones.

3.1. A famous example of a recursively defined sequence is the **Fibonacci sequence**. It is defined by

$$a_1 = 1, \qquad a_2 = 1, \qquad a_{n+2} = a_{n+1} + a_n \text{ for } n \geq 1.$$

Find the third, fourth, and fifth terms for the Fibonacci sequence.

Although is is usually defined recursively, it is actual possible to give a formula for the Fibonacci sequence. This need not be the case for a recursively defined sequence.

A **constant sequence** is a sequence whose every term is the same number. For example, $(1, 1, 1, \dots,)$ is the constant sequence of value one. Once again we emphasize that a sequence is not to be confused with a set. For example, the set $\{1, 1, 1, \dots\}$ is really just the set $\{1\}$ but $(1, 1, 1, \dots)$ denotes the function $f : \mathbb{N} \to \mathbb{R}$ with $f(n) = 1$ for all n. It is true that every sequence gives a set, namely the set of values that it takes (in other words the range of the corresponding function), but different sequences can give the same set (and you should give some examples of this occurrence).

Before we continue, we should point a common convention of the notation here. Sometimes we will write "consider the sequence $(a_n)_{n=4}^{\infty}$." To be precise, this means our sequence is given by the function $f(n) = a_{3+n}$ for $n \in \mathbb{N}$. We do this because it is often notationally convenient to do something like "starting the sequence at $n = 4$."

We now proceed to discuss the notion of convergence. We cannot overstate the importance of this concept. In fact, it is easily the most importance concept of this chapter and in understanding the structure of the real numbers (beyond the algebraic and ordering properties we described in the appendices). In some sense the entire purpose of introducing sequences is to give a framework under which we can study convergence in $\mathbb{R}$.

All that being said, many students find the definition of convergence a bit confusing (at least at first) and so you should certainly attempt to think carefully about it and from several different points of view (and we will attempt to help you do so). As with sequences, we will begin by giving an intuitive description.

Intuitively, a sequence converges to a number L if the terms of the sequence are 'heading towards L' as we go down the list. In one attempt to formulate a definition, one might say that a sequence $(a_n)_{n=1}^{\infty}$ converges to a limit L if "the terms of (a_n) get closer and closer to L as n gets larger and larger." For example the sequence $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots)$ is clearly getting closer and closer to zero as we proceed through the list.

However, though they shed some light on the concept, the words we have said above do not quite capture the idea fully. For instance, we might also consider the sequence $(1, 0, \frac{1}{2}, 0, \frac{1}{3}, 0, \ldots)$. This sequence is certainly also heading to zero and yet the terms are not always getting closer to zero as we go along. For example, the first occurrence of 0 is certainly closer to 0 than is the later term $\frac{1}{2}$. Furthermore, returning to our previous example of $(1, \frac{1}{2}, \frac{1}{3}, \ldots)$, we see that terms of the sequence are also in fact getting closer and closer to -1 (or any other number less than zero).

Thus is perhaps more appropriate to say that $(a_n)_{n=1}^{\infty}$ converges to L if "the terms of (a_n) get arbitrarily close to L as n gets larger." This definition solves the problems to which we have already alluded, but it introduces some other linguistic difficulties. Specifically, it may not be clear exactly what we mean by the phrase 'arbitrarily close'. To be a little clearer, we might say that no matter how close we want to be to the limit L, there is a point in the sequence past which we are always at least that close to L. The language here is more specific, but it also more convoluted. Expressing precise mathematical ideas in plain language is often quite difficult, but the attempt can be a very beneficial exercise.

We now give the precise definition.

Definition. A sequence $(a_n)_{n=1}^{\infty}$ is said to **converge** to $L \in \mathbb{R}$ if for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ so that for all $n \geq N$, $|a_n - L| < \epsilon$.

Again this definition might appear a bit confusing and so you should think about it carefully. Does this match the intuitive idea we had in mind? How are we measuring closeness to L? What is the purpose of the number N? Can you choose different values of N for two different values of ϵ? What is the relationship between ϵ and N? Come up with some other open-ended questions to ask yourself.

Although the intuitive idea of convergence has been around for quite some time, this precise definition seems to have been first published by Bernard Bolzano, a Czech mathematician, in 1816. As you might have gleaned from your calculus courses, it is the notion of convergence or of limits that distinguishes analysis/calculus from, say, algebra. Nevertheless, the precise formulation came about 150 years after the creation of calculus (due independently to Newton and Leibniz). The fact that it will probably take you some time to understand and become comfortable with it is therefore no surprise: it took even the world's most brilliant mathematicians more than a century to nail it down precisely (of course they were trying to accomplish the task without the aid of textbooks and instructors).

Definition. If the sequence $(a_n)_{n=1}^{\infty}$ converges to L, L is called a **limit** of the sequence. If there exists any $L \in \mathbb{R}$ such that $(a_n)_{n=1}^{\infty}$ converges to L then we say $(a_n)_{n=1}^{\infty}$ **converges** or that $(a_n)_{n=1}^{\infty}$ is a **convergent** sequence.

So far we have not actually shown that a number sequence cannot have more than one limit. If we want our intuitive understanding of limit to be satisfied, we will certainly want this to be the case. You will now show that it is, beginning with the following helpful lemma.

3.2. Let $a \geq 0$ be a real number. Prove that if for every $\epsilon > 0$ we have $a < \epsilon$ then $a = 0$.

3.3. Prove that if $(a_n)_{n=1}^{\infty}$ converges to $L \in \mathbb{R}$ and $(a_n)_{n=1}^{\infty}$ converges to $M \in \mathbb{R}$ then $L = M$.

Thus it makes sense to talk about *the* limit of a sequence (rather than *a* limit of a sequence). If the limit of $(a_n)_{n=1}^{\infty}$ is L we often write $\lim\limits_{n\to\infty} a_n = L$. In other words, $\lim\limits_{n\to\infty} a_n$ is the limit of the sequence $(a_n)_{n=1}^{\infty}$. Similarly, we often write $a_n \to L$ if $\lim_{n\to\infty} a_n = L$.

Caution: Before we write $\lim\limits_{n\to\infty} a_n$ we must know that a_n has a limit: we will see below that many sequences do not have limits.

We have emphasized repeatedly that one of the keys to understanding a definition is creating and understanding examples. At this point you should come up with some sequences for which you can figure out the limit. We have already mentioned that $(1/n)_{n=1}^{\infty}$ should converge to 0. Can you give some other examples? The intuition that you have developed in you calculus courses should be helpful.

Of course there is a difference between knowing intuitively that something is true (or being told it is true by an authority figure like a teacher) and having a mathematical proof (and thus a rigorous understanding) of the fact. We now need to move from the former to the latter, beginning with the simplest sequences.

3.4. Show using the definition of convergence that the constant sequence of value $a \in \mathbb{R}$ converges to a.

3.5. Prove using the definition of a limit that $\lim\limits_{n\to\infty} \frac{1}{n} = 0$.

Let's examine the proof here. Your proof should look something like this.

Proof. We will show $\lim\limits_{n\to\infty} \frac{1}{n} = 0$. Let $\epsilon > 0$ be arbitrary but fixed. We must find $N \in \mathbb{N}$ so that if $n \geq N$ then $|\frac{1}{n} - 0| < \epsilon$ which is the same as $\frac{1}{n} < \epsilon$. Choose $N \in \mathbb{N}$ so that $\frac{1}{N} < \epsilon$. Then if $n \geq N$, $\frac{1}{n} \leq \frac{1}{N} < \epsilon$. $\square$

The proof above relies on two things. Firstly, we used the basic properties of order which we have assumed are known. Secondly, we have used the fact that, given an $\epsilon > 0$, there exists $N \in \mathbb{N}$ with $\frac{1}{N} < \epsilon$ or in other words with $\frac{1}{\epsilon} < N$. This is actually a fact that needs to be proven, but we will temporarily take it as known.

Explicitly, we assume that $\mathbb{N} \subset \mathbb{R}$ has the **Archimedean Property**, which says that for every $x \in \mathbb{R}$ there exists an $n \in \mathbb{N}$ with $x < n$. The proof of this seemingly obvious fact is surprisingly delicate. In fact, it necessarily relies on the completeness axiom which we have yet to formulate. This is an example of a situation where a seemingly obvious fact is not so obvious when we attempt to prove it.

3.6. Prove the following consequence to the Archimedean property: For every $\epsilon > 0$ there exists an $n \in \mathbb{N}$ such that

$$0 < \frac{1}{n} < \epsilon.$$

Armed with the Archimedean property and its consequences, we can now study some more concrete examples of limits precisely.

3.7. Prove using the definition of a limit that

$$\lim_{n\to\infty} 1 - \frac{1}{n^2+1} = 1 \text{ and } \lim_{n\to\infty} \frac{n^2-1}{2n^2+3} = \frac{1}{2}.$$

3.8. Let $(a_n)_{n=1}^{\infty}$ be sequence defined by saying that $a_n = \frac{1}{n}$ for n odd and $a_n = 0$ for n even. Show that $\lim\limits_{n\to\infty} a_n = 0$.

We will see in this chapter that convergent sequences have many special and important properties. We give the first now.

Definition. A sequence $(a_n)_{n=1}^{\infty}$ is called **bounded** if the associated set $\{a_n : n \in \mathbb{N}\}$ is contained in a bounded interval.

Notice that (a_n) is bounded if and only if there exists $K \geq 0$ with $|a_n| \leq K$ for all $n \in \mathbb{N}$. (why?)

3.9. Prove that if $(a_n)_{n=1}^{\infty}$ is a convergent sequence with $\lim\limits_{n\to\infty} a_n = L$ then $(a_n)_{n=1}^{\infty}$ is bounded.

Now that we have a considered a few affirmative examples, we need to consider some negatives ones. In other words, we need to think about what it means for a sequence *not* to converge.

3.10. Negate the definition of $\lim\limits_{n\to\infty} a_n = L$ to give an explicit definition of "$(a_n)_{n=1}^{\infty}$ does not converge to L."

We can write $(a_n)_{n=1}^{\infty}$ does not converge to L as $(a_n)_{n=1}^{\infty} \not\to L$.

Caution: We should not write $\lim\limits_{n\to\infty} a_n \neq L$ to mean $(a_n)_{n=1}^{\infty}$ does not converge to L unless we know that $\lim\limits_{n\to\infty} a_n$ exists.

Definition. A sequence (a_n) is said to **diverge** or be **divergent** if it does not converge to L for any $L \in \mathbb{R}$.

3.11. Without making a reference to the definition of convergence, formulate in precise logical language (as in the definition of convergence) a definition for '$(a_n)_{n=1}^{\infty}$ is divergent.' Avoid using a phrase like 'there exists no L.' Compare your definition to the one you gave for '$(a_n)_{n=1}^{\infty}$ does not converge to L'.

Definition. We will distinguish two special types of divergence:

(1) A sequence (a_n) is said to **diverge to** $+\infty$ if for all $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $a_n \geq M$. In an abuse of notation we often write $\lim_{n\to\infty} a_n = +\infty$ or $a_n \to +\infty$.
(2) A sequence (a_n) is said to **diverge to** $-\infty$ if for all $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $a_n \leq M$. Again in an abuse of notation we often write $\lim_{n\to\infty} a_n = -\infty$ or $a_n \to -\infty$.

In fact, we need to justify this terminology.

3.12. Show that if (a_n) diverges to ∞ then a_n diverges. Likewise, show that if (a_n) diverges to $-\infty$, it diverges.

Thus one way for a sequence to diverge is for it to 'head off to $\pm\infty$' (i.e., to ∞ or $-\infty$). This is, however, not the only way.

3.13. We have seen that every convergent sequence is bounded. Give an example of a sequence which is bounded and yet divergent. Show that it does not diverge to $\pm\infty$.

3.14. Which, if any, of the following conditions are equivalent to $(a_n)_{n=1}^{\infty}$ converges to L. If a condition is equivalent, prove it. If not, give a counter-example.

(1) There exists an $\epsilon > 0$ such that for all $N \in \mathbb{N}$ there is an $n \geq N$ with $|a_n - L| < \epsilon$.
(2) For all $\epsilon > 0$ and for all $N \in \mathbb{N}$, there is an $n \geq N$ with $|a_n - L| < \epsilon$.
(3) For all $N \in \mathbb{N}$, there exists an $\epsilon > 0$ such that for all $n \geq N$, $|a_n - L| < \epsilon$.
(4) For all $N \in \mathbb{N}$ and $n \geq N$, there is an $\epsilon > 0$ with $|a_n - L| < \epsilon$.

2. Properties of Convergence

Now that we have established the basic terminology of convergence (and of divergence), we need to study all the basic properties. These properties will both help us to understand limits and help us to prove things about limits.

3.15. Prove that a sequence $(a_n)_{n=1}^{\infty}$ converges to $L \in \mathbb{R}$ if and only if for all $\epsilon > 0$ the set

$$\{n \in \mathbb{N} : |a_n - L| \geq \epsilon\}$$

is finite.

Thus a sequence converges to L if and only if, for each $\epsilon > 0$, the number of terms which are more than ϵ distance from L is finite. This observation has a couple of interesting corollaries.

3.16. Prove:

(1) If we change finitely many terms of a sequence, we do not alter its limiting behavior: if the sequence originally converged to L then the altered sequence still converges to L, and if the original sequence diverged to $\pm\infty$ or diverged in general then so does the altered sequence.
(2) If we remove a finite number of terms from a sequence then we do not alter its limiting behavior.

The previous results confirm a fact which is perhaps intuitively clear: if we change or remove some terms at the beginning of a sequence, we do not change where it is headed.

3.17. Let $S \subset \mathbb{R}$ be a set. Assume that for all $\epsilon > 0$ there is an $a \in S$ with $|a| < \epsilon$. Prove there there is a convergent sequence $(a_n)_{n=1}^{\infty}$, with $a_n \in S$ for all n, and $\lim_{n\to\infty} a_n = 0$.

3.18. Prove or disprove:

(1) If $\lim_{n\to\infty} a_n = L$ then $\lim_{n\to\infty} |a_n| = |L|$.
(2) If $\lim_{n\to\infty} |a_n| = |L|$ then $\lim_{n\to\infty} a_n = L$.
(3) If $\lim_{n\to\infty} |a_n| = 0$ then $\lim_{n\to\infty} a_n = 0$.
(4) If $\lim_{n\to\infty} a_n = 0$ then $\lim_{n\to\infty} |a_n| = 0$.

This next result is often called the **Squeeze Theorem** for sequences.

3.19. If $a_n \leq c_n \leq b_n$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = L$ then $\lim_{n\to\infty} c_n = L$.

We next prove a collection of results known as the **Limit Laws** for sequences. In fact, this name is a misleading as essentially all the results of this chapter could equally be given the same name (as they are all facts about limits). For historical reasons, the results given this name are those that describe the relation between limits and the algebraic properties of $\mathbb{R}$.

The following result is not a limit law, but it will be very useful in proving them (specifically the one related to division).

3.20. Let $(b_n)_{n=1}^{\infty}$ be a convergent sequence whose limit M is nonzero. Prove that there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ we have $|b_n| > \frac{|M|}{2}$.

Suppose now that $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are sequences and $c \in \mathbb{R}$ is a real number. We can define new sequences $(c \cdot a_n)_{n=1}^{\infty}$, $(a_n + b_n)_{n=1}^{\infty}$, and $(a_n \cdot b_n)_{n=1}^{\infty}$. If $b_n \neq 0$ for all $n \in \mathbb{N}$ then we can define

$$\left(\frac{a_n}{b_n}\right)_{n=1}^{\infty}.$$

3.21. Let $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ be sequences and let $c \in \mathbb{R}$. Prove the Limit Laws:

(1) If $(a_n)_{n=1}^{\infty}$ is a convergent sequence and $c \in \mathbb{R}$ then $(c \cdot a_n)_{n=1}^{\infty}$ is a convergent sequence and

$$\lim_{n\to\infty}(c \cdot a_n) = c \cdot \lim_{n\to\infty} a_n.$$

(2) If $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are convergent sequences then $(a_n + b_n)_{n=1}^{\infty}$ is a convergent sequence and

$$\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n.$$

(3) If $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are convergent sequences then $(a_n \cdot b_n)_{n=1}^{\infty}$ is a convergent sequence and

$$\lim_{n\to\infty}(a_n \cdot b_n) = \big(\lim_{n\to\infty} a_n\big) \cdot \big(\lim_{n\to\infty} b_n\big).$$

(4) If $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are convergent sequences with $b_n \neq 0$ for all $n \in \mathbb{N}$ and $\lim\limits_{n\to\infty} b_n \neq 0$ then

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

Hint: For the third part, the problem is that we have two quantities changing simultaneously. To deal with this we use a very common trick in analysis: we add and subtract additional terms, which does not affect the value, and then group terms so that each term is a product of things we can control. Let $L = \lim_{n\to\infty} a_n$ and $M = \lim_{n\to\infty} b_n$. We can write

$$\begin{aligned} a_n \cdot b_n - L \cdot M &= a_n \cdot b_n - L \cdot M + L \cdot b_n - L \cdot b_n \\ &= (a_n - L) \cdot b_n + L \cdot (b_n - M). \end{aligned}$$

Hint: For the fourth part, given the third part, it suffices to prove (explain why) that

$$\lim_{n\to\infty} \frac{1}{b_n} = \frac{1}{\lim\limits_{n\to\infty} b_n}.$$

Let $M = \lim_{n\to\infty} b_n$ and notice that

$$\left|\frac{1}{b_n} - \frac{1}{M}\right| = \frac{|M - b_n|}{|Mb_n|} < \frac{|M - b_n|}{(M^2/2)}$$

if $|b_n| > \frac{|M|}{2}$.

3.22. Give an example of two sequences (a_n) and (b_n) such that (a_n) and (b_n) diverge and yet $(a_n + b_n)$ converges. Give an example where (a_n) and (b_n) diverge and yet $(a_n b_n)$ converges.

This next result is also sometimes including among the Limit Laws. Needless to say, it gives the interaction between the notion of a limit and the ordering properties of $\mathbb{R}$.

3.23. Suppose $a \le a_n \le b$ for all $n \in \mathbb{N}$. Prove that if $\lim_{n\to\infty} a_n = L$, then $L \in [a, b]$. Prove that the conclusion is still true if a_n is outside $[a, b]$ for only finitely many n.

3.24. Let $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ be convergent sequences with $a = \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. Prove that if $a_n - b_n = \frac{1}{n}$ for all $n \in \mathbb{N}$ then $a = b$. Would this theorem still be true if, instead of equality, you had $a_n - b_n < \frac{1}{n}$? What if $a_n - b_n = \frac{1}{2^n}$?

3. Monotone Sequences

In this section, we will finally formulate the completeness axiom. To do so, we introduce another class of sequences.

Definition. Let (a_n) be a sequence. We say that (a_n) is **increasing** if for all $n \in \mathbb{N}$, $a_n \leq a_{n+1}$. Likewise, we say that (a_n) is **decreasing** if for all $n \in \mathbb{N}$, $a_n \geq a_{n+1}$. A sequence is called **monotone** if it is either increasing or decreasing.

3.25. Prove the following:

(1) If (a_n) is increasing and unbounded then (a_n) diverges to $+\infty$.
(2) If (a_n) is decreasing and unbounded then (a_n) diverges to $-\infty$

Thus we see that an increasing sequence can diverge if it "escapes to $+\infty$." Suppose though that we decided ahead of time that this was not allowed. In other words consider an increasing sequence which is also bounded above. For the sake of intuition, assume that the sequence is **strictly increasing** meaning that each term is strictly larger than the last (i.e., not equal to the last).

In some sense we should expect that this sequence should be 'trapped.' On the one hand, the terms of the sequence are getting larger and larger. On the other, we have assumed that there is a ceiling that they cannot break. Of course, there is no guarantee that they will every get very close to our ceiling. But that just means we could pick a smaller ceiling. If they don't get close to that one either, we can just pick another and so on. In this way, we should be able to trap the sequence into smaller and smaller spaces (as the terms gets larger).

Thus it seems reasonable that these terms should be 'headed' somewhere specific. However, if you try to make a rigorous proof out of our intuitive ideas, it will always fail. Thus to make our idea work, we have to add an assumption to our axioms for the real numbers. This is the essence of the *completeness axiom.* The **completeness axiom** states that every bounded increasing sequence of real numbers converges.

In the introduction to chapter 2, we said that the completeness axiom forbids the existence of 'gaps' in the real line. In some sense, our formulation of the completeness axiom captures this idea because it asserts that there are no gaps towards which the 'trapped' sequence can head: it has to be heading towards a specific number. Our use of the word 'gap' here is an attempt to express an abstract mathematical ideal in plain language (a theme we have emphasized throughout the course). Nevertheless, we will see below that it is not a perfect choice of words. Perhaps after you complete and understand the material of this chapter, you can give a better explanation (and you should certainly try).

We can however finally prove the Archimedean property.

3.26. Suppose that $x \in \mathbb{R}$. Then there is some $n \in \mathbb{N}$ with $n > x$.

Hint: Consider the sequence $(n)_{n=1}^{\infty}$. If the Archimedean property is false then it is bounded. Since it is increasing, it must have limit L. Consider precisely the implications of assuming that all the natural numbers head towards a fixed number L.

You will notice that you used the completeness axiom in your proof. This is not an accident: it is impossible to prove the Archimedean property without assuming the completeness axiom . More precisely, the Archimedean property cannot be proven using the other axioms alone. We can view this fact as yet more motivation for assuming the completeness axiom (as it seems like a strange thing to assume at first). Certainly the set $\mathbb{N}$ should not be bounded by a real number.

3.27. Suppose that (a_n) is decreasing and bounded. Than (a_n) converges.

We will now give several results which demonstrate the power of the completeness axiom. We begin by studying an important class of sequences known as **geometric sequences**.

3.28. Let $r \in \mathbb{R}$.

(1) Prove that if $0 \le r < 1$ then $(r^n)_{n=1}^{\infty}$ converges to 0.
(2) Prove that if $r > 1$ then $(r^n)_{n=1}^{\infty}$ diverges to $+\infty$.
(3) Prove that the sequence (r^n) converges if and only if $-1 < r \le 1$. If $r = 1$ prove that $\lim_{n\to\infty} r^n = 1$. If $-1 < r < 1$, prove that $\lim_{n\to\infty} r^n = 0$.

Hint: For the first part, show the sequence is decreasing and bounded below. Let L be the limit and proceed by contradiction: suppose $L > 0$, find a_n with $L \le a_n < r^{-1}L$, and hence produce a contradiction.

3.29. Let $a_1 = 2$ and let (a_n) be generated by the recursive formula

$$a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right)$$

for $n \ge 1$.

(1) Prove that a_n is well-defined and positive.
(2) Prove that $2 < a_n^2$ for all $n \in \mathbb{N}$.
(3) Prove that a_n is decreasing and hence converges.

(4) Now take limits on both sides of the recursive formula and prove and use the fact that

$$\lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} a_n$$

to show that $a^2 = 2$ if $a = \lim_{n\to\infty} a_n$. Show that $a > 0$.

Hint: For the second part, assuming $a_n > 0$ turn

$$\frac{1}{2}\left(a_n + \frac{2}{a_n}\right) > \sqrt{2}$$

into an equivalent condition on a quadratic polynomial. Proceed by induction.

Needless to say the real number a considered in the previous result is usually denoted by $\sqrt{2}$. Thus the completeness axiom also implies that $\sqrt{2}$ is a well-defined real number.

3.30. $\sqrt{2}$ is not a rational number.

Of course we call a real number which is not rational an **irrational number**. Thus we have proven the ('obvious') fact that there exist irrational numbers. In other words, the collection of real numbers is larger than the collection of rational numbers. We will, however, see now that it is not too much larger (in an appropriate sense).

3.31. Prove the following using the Archimedean property:

(1) for every $x \in \mathbb{R}$ there exists an $m \in \mathbb{Z}$ such that

$$m \leq x < m + 1,$$

(2) for every $x \in \mathbb{R}$ and $n \in \mathbb{N}$ there exists an $m \in \mathbb{Z}$ such that

$$\frac{m}{n} \leq x < \frac{m+1}{n},$$

Using these we can prove that every real number can be approximated arbitrarily well by a rational number. The following result is referred to by saying that $\mathbb{Q}$ is **dense** in $\mathbb{R}$.

3.32. Prove:

(1) For every $x \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\frac{m}{n} \in \mathbb{Q}$ such that

$$0 \leq x - \frac{m}{n} < \epsilon.$$

(2) For every $x \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\frac{m}{n} \in \mathbb{Q}$ such that

$$-\epsilon < x - \frac{m}{n} \leq 0$$

(3) For every $x \in \mathbb{R}$ and $\epsilon > 0$ there exists an $a \in \mathbb{Q}$ with $|x - a| < \epsilon$.

(4) For every $x \in \mathbb{R}$ there is a sequence (x_n) of rational numbers with $\lim_{n\to\infty} x_n = x$.

The previous result shows a weakness of the language we used in our attempt to describe the completeness axiom. We stated that the truth of the completeness axiom was tantamount to saying that $\mathbb{R}$ contains no gaps. But this result implies that if we draw a 'gap' (i.e., an interval) on the real line, no matter how small, there is always a rational number inside of it. Thus there are no gaps of this kind in $\mathbb{Q}$ either, despite the fact that $\mathbb{Q}$ is not complete (i.e., does not satisfy the completeness axiom: there are bounded increasing sequence in $\mathbb{Q}$ which do not converge to a limit in $\mathbb{Q}$). It might be better to say that the completeness axiom states that $\mathbb{R}$ has no 'infinitely-small gaps.' But perhaps at this point our language starts to lose meaning.

The irrational numbers are also dense in $\mathbb{R}$.

3.33. Prove:

(1) For all $\frac{m}{n} \in \mathbb{Q} \setminus \{0\}$ the number $\sqrt{2}\,\frac{m}{n}$ is irrational.

(2) For every $x \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\frac{m}{n} \in \mathbb{Q}$ such that

$$0 \leq x - \sqrt{2}\frac{m}{n} < \epsilon.$$

(3) For every $x \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\frac{m}{n} \in \mathbb{Q}$ such that

$$-\epsilon < x - \sqrt{2}\frac{m}{n} \leq 0$$

3.34. Let $x \in \mathbb{R}^+$. What can you say about the sequence given by $a_1 = 1$ and $a_{n+1} = \frac{1}{2}(a_n + \frac{x}{a_n})$ for $n \geq 1$?

3.35. For each n, let $I_n = [a_n, b_n]$ be a bounded closed interval. Suppose that $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$.

(1) Prove that there exists a $p \in \mathbb{R}$ such that $p \in I_n$ for all $n \in \mathbb{N}$.

(2) Suppose in addition that $\lim\limits_{n\to\infty}(b_n - a_n) = 0$. Prove that if $q \in I_n$ for all $n \in \mathbb{N}$ then $q = p$.

4. Subsequences

Intuitively, a subsequence of the sequence (a_n) is a (new) sequence formed by skipping (possibly infinitely many) terms in a_n.

Definition. A sequence $(b_k)_{k=1}^{\infty}$ is a **subsequence** of $(a_n)_{n=1}^{\infty}$ if there exists a strictly increasing sequence of natural numbers $n_1 < n_2 < \cdots$ so that for all $k \in \mathbb{N}$, $b_k = a_{n_k}$.

Again you should see for yourself that this definition captures the intuitive idea above. As an example, $(1, 1, 1, \ldots)$ is a subsequence of $(1, -1, 1, -1, \ldots)$. In fact, any sequence of ± 1's is a subsequence of $(1, -1, 1, -1, \ldots)$.

In particular, we see that a subsequence of a divergent sequence may be convergent.

3.36. Give an example of a sequence $(a_n)_{n=1}^{\infty}$ of natural numbers (i.e., with $a_n \in \mathbb{N}$ for each n) so that every sequence of natural numbers is a subsequence of (a_n). Can you do the same if $\mathbb{N}$ is replaced with $\mathbb{Z}$?

$(\frac{1}{n^2})_{n=1}^{\infty}$ is a subsequence of $(\frac{1}{n})_{n=1}^{\infty}$, and both $\left(\frac{1}{n}\right)_{n=1}^{\infty}$ and $\left(\frac{1}{n^2}\right)_{n=1}^{\infty}$ converge to 0. The next problem asks if this example can be generalized.

3.37. Prove or disprove: If (b_n) is a subsequence of (a_n) and $\lim\limits_{n\to\infty} a_n = L$, then $\lim\limits_{n\to\infty} b_n = L$.

Subsequences are useful because, if a specified sequence does not have a certain desirable property, we can often find a subsequence which does.

3.38. Prove that every sequence of real numbers has a monotone subsequence.

Hint: Consider two cases. The first is the case in which every subsequence has a minimum element.

Thus we get the following very useful result.

3.39. Prove that every bounded sequence of real numbers has a convergent subsequence.

5. Cauchy Sequences

One of the reasons that the completeness axiom is so strong is that it (by definition) allows us to conclude that certain sequences converge without knowing their limit beforehand (for example we were able to show that $\sqrt{2}$ exists as a real number without assuming it beforehand).

But bounded monotone sequences are certainly not the only type of sequence which converges. In this section, we give a condition which is equivalent to convergence, but makes no reference to knowing the limit of the sequence.

Definition. A sequence (a_n) is called **Cauchy** if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $m, n \geq N$ we have $|a_n - a_m| < \epsilon$.

3.40. Negate the definition of Cauchy sequence to give an explicit definition of "$(a_n)_{n=1}^{\infty}$ is not a Cauchy sequence."

3.41. Which of the following conditions, if any, are equivalent to "$(a_n)_{n=1}^{\infty}$ is Cauchy" or "(a_n) is not Cauchy?"

(1) For all $\epsilon > 0$ and $N \in \mathbb{N}$ there are $n, m > N$ with $n \neq m$ and $|a_n - a_m| < \epsilon$.
(2) For all $n \in \mathbb{N}$ there is an $N \in \mathbb{N}$ such that for all $p, q \geq N$, $|a_p - a_q| < 1/n$.
(3) There exists an $\epsilon > 0$ and $n \neq m$ in $\mathbb{N}$ with $|a_n - a_m| < \epsilon$.
(4) There exists $N \in \mathbb{N}$ such that for all $\epsilon > 0$ there are $n \neq m$ with $n, m > N$ and $|a_n - a_m| < \epsilon$.

3.42. Prove that every convergent sequence is Cauchy.

This is particularly useful in the contrapositive form: if $(a_n)_{n=1}^{\infty}$ is not Cauchy then $(a_n)_{n=1}^{\infty}$ diverges.

3.43. Prove that every Cauchy sequence is bounded. Is the converse true?

3.44. Let (a_n) be a Cauchy sequence and let (a_{n_k}) be a convergent subsequence. Prove that (a_n) is convergent and $\lim_{n\to\infty} a_n = \lim_{k\to\infty} a_{n_k}$.

We are thus led to the following extremely important result.

3.45. Show that a sequence is convergent if and only if it is Cauchy.

Hence, a Cauchy sequence in $\mathbb{R}$ is the same as a convergent sequence in $\mathbb{R}$. Again the advantage is that the definition of a Cauchy sequence makes no reference to the limit: it is an intrinsic property of the sequence.

3.46. Is every Cauchy sequence in $\mathbb{Q}$ convergent to some point in $\mathbb{Q}$?

We remark that the equivalence of Cauchy sequence and convergent sequence is itself actually equivalent to the completeness axiom. In other words, one can assume that Cauchy sequences converge and prove that bounded monotone sequences converge. Many books begin with the Cauchy perspective and call the above equivalence the completeness axiom (the fact that bounded monotone sequence converge is then called the 'monotone convergence theorem').

3.47. Prove or give a counterexample: if a sequence of real numbers $(x_n)_{n=1}^{\infty}$ has the property that for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $|x_{n+1} - x_n| < \epsilon$, then (x_n) is a convergent sequence. How is this different from the definition of a Cauchy sequence?

6. Decimals

So far we have not had a systematic way of describing real numbers. In other words, unless we are dealing with a special number like a rational number or $\sqrt{2}$, we have no systematic way to specify a number. We will attempt to solve this problem by using decimal expansions.

Of course this method will fall somewhat short as many decimal expansions cannot be written down or even specified easily. In some sense, the only decimal expansion that we can easily write down either terminate

$$\frac{1}{8} = 0.125 \qquad \frac{27}{50} = 0.54$$

or become periodic

$$\frac{1}{3} = 0.3333\ldots = 0.\overline{3} \qquad \frac{1}{7} = 0.142857142857\ldots = 0.\overline{142857}$$

and we will see that such expansions do not encompass all numbers.

3.48. Write down a decimal expansion that is not periodic in such a way that the pattern is clear.

It is an interesting fact that a decimal expansion which either terminates or becomes periodic always represents a rational number (and we will show this fact below).

Another 'problem' with decimal expansions is that a single number can have two different expansions. For example, we will see that

$$\frac{1}{8} = 0.124\overline{9} = 0.125.$$

Fortunately, we will also see that all the numbers that have multiple decimal expansions are in fact rational. However, not all rational numbers have multiple expansions.

Shortly we will prove that every real number has a decimal expansion. For example the real number π has a decimal expansion even though not all the digits are known. As of this writing, the first 2,699,999,990,000 decimal digits have been calculated. Ironically a Frenchman named Fabrice Bellard was able to complete this calculation on his home computer using a new algorithm of his own design. It took Bellard's computer 131 days to complete the program. This accomplish is particulary impressice since the previous record was calculated on a Japanese super computer known as the T2K Open Supercomputer.

Chao Lu of China holds the Guinness Book of Records record for reciting digits of π. In just over twenty-four hours, he recited the first $67,890$ digits of π. He was given no breaks of any kind as the rules stated that he had to recite each digit within 15 seconds of the previous one.

We now define and construct decimal expansions precisely. Consider the interval $[0,1]$. Firstly, we divide the interval $[0,1]$ into 10 equal closed subintervals (of length $\frac{1}{10}$):

$$\left[0,\tfrac{1}{10}\right] \cup \left[\tfrac{1}{10},\tfrac{2}{10}\right] \cup \left[\tfrac{2}{10},\tfrac{3}{10}\right] \cup \left[\tfrac{3}{10},\tfrac{4}{10}\right] \cup \left[\tfrac{4}{10},\tfrac{5}{10}\right] \cup \left[\tfrac{5}{10},\tfrac{6}{10}\right] \cup \cdots \cup \left[\tfrac{9}{10},1\right].$$

Label these intervals $I_0, I_1, \ldots I_9$, respectively.

Now let k_1 be an integer between 0 and 9. Then I_{k_1} is the interval $I_{k_1} = [\frac{k_1}{10}, \frac{k_1+1}{10}]$. We may further divided this interval into ten equal pieces (of length $\frac{1}{100}$):

$$\left[\frac{k_1}{10}, \frac{k_1}{10} + \frac{1}{100}\right] \cup \left[\frac{k_1}{10} + \frac{1}{100}, \frac{k_1}{10} + \frac{2}{100}\right] \cup \cdots \cup \left[\frac{k_1}{10} + \frac{9}{100}, \frac{k_1+1}{10}\right].$$

Label these intervals $I_{k_1,0}, I_{k_1,2}, \ldots, I_{k_1,9}$. By dividing further and further, we may define the interval $I_{k_1,k_2,\ldots,k_n}$ where $(k_1,\ldots,k_n)$ is any finite sequence with values in $0,\ldots,9$.

3.49. What are the endpoints of the interval $I_{k_1,\ldots,k_n}$?

Definition. Suppose that $a \in [0, 1]$. A **decimal expansion** for x is an infinite sequence, $(k_i)_{i=1}^{\infty}$, such that each $0 \le k_i \le 9$ is an integer and so that $a \in I_{k_1,\ldots,k_n}$ for each n.

This definition might seem a little bit strange at first. How does it compare to the picture you have in mind for a decimal expansion?

3.50. Explain how a number in $[0, 1]$ can have more than 1 decimal expansion.

3.51. Suppose $a \in [0, 1]$. First prove that we may find a decimal expansion for a. Next assume that $(k_i)_{i=1}^{\infty}$ is a decimal expansion for $a \in [0, 1]$. Define a sequence $(a_n)_{n=1}^{\infty}$ by saying that

$$a_n = \frac{k_1}{10} + \frac{k_2}{100} + \cdots + \frac{k_n}{10^n} = \sum_{i=1}^{n} \frac{k_i}{10^i}.$$

Prove that (a_n) is convergent and $\lim_{n\to\infty} a_n = a$.

You may recall from your calculus course that we denote the limit of of the above sequence (a_n) by

$$\sum_{i=1}^{\infty} \frac{k_i}{10^i}.$$

There is a converse to the last result which perhaps demonstrates that our definition of decimal expansion is a reasonable one.

3.52. Suppose that $(k_i)_{i=1}^{\infty}$ is such that k_i is an integer with $0 \le k_i \le 9$. Show that

$$\sum_{i=1}^{\infty} \frac{k_i}{10^i}$$

always exists. Furthermore show that if

$$a = \sum_{i=1}^{\infty} \frac{k_i}{10^i}$$

then (k_i) is a decimal expansion for a.

It perhaps goes without saying that if (k_i) is a decimal expansion for a, we often write $a = 0.k_1k_2k_3\ldots$. Of course this notation does not reflect any kind of multiplication of the k_i. If we write a terminating decimal sequence like $0.k_1k_2\cdots k_n$, we mean the first n terms are $k_1, \ldots, k_n$ and all the other terms are zero.

So far we have only considered decimal expansions for numbers in $[0, 1]$. Of course we should define decimal expansions for all real numbers.

3.53. Extend the definition of decimal expansion to all real numbers. Explicitly state what it means for $(k_i)_{i=0}^{\infty}$ to be a decimal expansion for $a \in \mathbb{R}$. Show that the previous two results work for all real numbers and not just numbers in $[0, 1]$.

Again we usually denote the decimal expansion $(k_i)_{i=0}^{\infty}$ by $k_0.k_1k_2k_3\ldots$. Now that we have defined decimal expansions, we prove that all the basic results that we expect are true.

3.54. Suppose $x, y \in \mathbb{R}$ are not equal and suppose (k_i) and (ℓ_i) are decimal expansions for x and y respectively. In addition, assume without loss of generality that neither (k_i) nor (ℓ_i) ends with a constant sequence of 9's. Show that $x < y$ if and only if there exists a $r \in \mathbb{N}$ with $k_0.k_1k_2\ldots k_r < \ell_0.\ell_1\ell_2\ldots\ell_r$.

3.55. Suppose $x, y \in \mathbb{R}$ and suppose $\{k_i\}$ and $\{\ell_i\}$ are decimal expansions for x and y respectively. Describe (and prove) how to find a decimal expansion for $x + y$, for $-y$ and for $x - y$.

3.56. Suppose $x, y \in \mathbb{R}$ and suppose $\{k_i\}$ and $\{\ell_i\}$ are decimal expansions for x and y respectively. Describe (and prove) how to find a decimal expansion for xy, for $1/y$ and for x/y.

3.57. Prove that a decimal expansion is eventually periodic if and only if it comes from a rational number.

We conclude our study of decimal expansions by remarking that the ideas of this section can be used to construct the real numbers rigorously (rather than taking their existence on faith). One defines the real numbers to be the set of integer sequences $(k_i)_{i=0}^{\infty}$ such that $0 \leq k_i \leq 9$ for $i \geq 1$. We have to a little bit careful in that we need to specify that decimals that should give the same number are equal (this is accomplished by putting a equivalence relation on the set).

We then define addition, subtraction, multiplication, and division in the way that we did in the previous exercises. Likewise we define an order on the collection. We then verify that all the axioms (including the completeness axiom) hold. If you feel particularly ambitious, you could attempt to carry out this construction precisely and verify all the axioms. You certainly have the intellectual tools at your disposal, but the construction and verification might be a bit long and it will certainly take persistence.

7. Supremums and Completeness

In this section, we give yet another equivalent formulation of the completeness axiom. This version is a little bit different than the others in that it makes no mention of sequences and instead deals with sets. We begin with a few definitions.

Definition. Let $A \subseteq \mathbb{R}$ and $x \in \mathbb{R}$.

(1) x is called an **upper bound for** A if, for all $a \in A$, we have $a \leq x$.
(2) x is called an **lower bound for** A if, for all $a \in A$, we have $x \leq a$.
(3) The set A is called **bounded above** if there exists an upper bound for A. The set A is called **bounded below** if there exists a lower bound for A. The set A is called **bounded** if it is both bounded above and bounded below.
(4) x is called a **maximum of** A if $x \in A$ and x is an upper bound for A. We will often write $x = \max A$.
(5) x is called a **minimum of** A if $x \in A$ and x is a lower bound for A. Again we will often write $x = \min A$.

Previously we said a sequence, (a_n), was bounded if the associated set, $\{a_n : n \in \mathbb{N}\}$ was contained in a bounded interval. How do these two uses of the same word relate to each other?

Note that if a set has a maximum or a minimum, it only has one of each (why?)

3.58. For each of the following subsets of $\mathbb{R}$:

(a) $A = \varnothing$,
(b) $A = [0, 1]$,
(c) $A = (0, 1) \cap \mathbb{Q}$.
(d) $A = [0, \infty)$,

(1) Find all lower bounds for A and all upper bounds for A,

(2) Discuss whether A is bounded.
(3) Discuss whether A has a maximum and if so find it.
(4) Discuss whether A has a minimum and if so find it.

In particular, we see that not every set has a maximum and so the usefulness of the concept is a bit limited. Thus we introduce a more useful term which captures a similar idea.

Definition. Let $A \subseteq \mathbb{R}$ and $x \in \mathbb{R}$.

(1) x is called a **least upper bound of** A or **supremum of** A if
  (a) x is an upper bound for A,
  (b) for all y, if y is an upper bound for A then $x \leq y$.
(2) x is called a **greatest lower bound of** A or **infimum of** A if
  (a) x is a lower bound for A,
  (b) for all y, if y is a lower bound for A then $y \leq x$.

3.59. Let $A \subseteq \mathbb{R}$. Prove that if x is a supremum of A and y is a supremum of A then $x = y$.

Hence, if the set A has a supremum then that supremum is unique and we can speak of *the* supremum of A and write $\sup A$. Similarly, if the set A has an infimum then that infimum is unique and we can speak of *the* infimum of A and write $\inf A$. The supremum is a generalization of the idea of a maximum and the infimum is a generalization of the idea of a minimum.

3.60. Suppose that $A \subset \mathbb{R}$ is a set and that $\max A$ exists. Show that $\sup A$ exists and $\max A = \sup A$. Similarly if $\min A$ exists show that $\inf A$ exists and $\min A = \inf A$.

To better understand the infimum and the supremum, we should consider some examples.

3.61. Find $\inf A$ and $\sup A$, if they exist, for each of the following subsets of $\mathbb{R}$:

(1) $A = \varnothing$,
(2) $A = [0, 1]$,
(3) $A = (0, 1) \cap (\mathbb{R} \backslash \mathbb{Q})$
(4) $A = [0, \infty)$.

Thus we see that, like the maximum, the supremum also does not always exist. On the other hand, let us consider the reasons why the supremum failed to exist in some of the previous examples. Firstly, if the set in question is the empty set it cannot have a supremum. Hence assume that it is nonempty. We saw in the previous exercise a nonempty set without an supremum. What was it about that set which forced it not to have a supremum?

In fact this is the only other reason a supremum might not exist.

3.62. Suppose that A is nonempty and bounded above. Then A has a supremum.

Hint: Show that there is a least element $k_1 \in \mathbb{Z}$ such that k_1 is an upper bound for A. If k_1 is not a least upper bound for A, show there is a least $k_2 \in \mathbb{N}$ such that $k_1 - \frac{1}{2^{k_2}}$ is an upper bound for A. Proceed in this way to find the supremum.

Again we needed to use the completeness axiom in the proof of the previous result. Like the convergence of Cauchy sequences, the truth of the last result is equivalent to the completeness axiom. In fact, most authors take the last result as their formulation of the completeness axiom.

We have a similar existence result for infimums.

3.63. Suppose that A is nonempty and bounded below. Then $\inf A$ exists.

3.64. Suppose that $A \neq \varnothing$ is bounded. Prove that $\inf A \leq \sup A$.

3.65. Give examples, if possible, of the following.

(1) A set A with a supremum but no maximum.
(2) A decreasing sequence $(a_n)_{n=1}^{\infty}$ so that

$$\inf\{a_n | n \in \mathbb{N}\}$$

does not exist.
(3) An increasing sequence $(a_n)_{n=1}^{\infty}$ so that

$$\inf\{a_n | n \in \mathbb{N}\}$$

does not exist.

Now that we have proven the existence of supremums and infimums for nonempty bounded sets, we will study their properties. For example, though $\sup A$ need not be in the set A there are elements of A arbitrarily close to $\sup A$.

3.66. Let $A \subseteq \mathbb{R}$ be non-empty and bounded above. Prove that if $s = \sup A$ then for every $\epsilon > 0$ there exists an $a \in A$ with $s-\epsilon < a \leq s$.

3.67. What result regarding the infimum would correspond to the previous result? State and prove it.

3.68. Let $A \subseteq \mathbb{R}$ be bounded above and non-empty. Let $s = \sup A$. Prove that there exists a sequence $(a_n) \subseteq A$ with $\lim_{n\to\infty} a_n = s$. Prove that, in addition, the sequence (a_n) can be chosen to be increasing.

3.69. Again state and prove an analogous result for infimums.

3.70. Prove or disprove:

(1) If $A, B \subseteq \mathbb{R}$ are nonempty sets such that for every $a \in A$ and for every $b \in B$ we have $a \leq b$ then $\sup A$ and $\inf B$ exist and $\sup A \leq \inf B$.
(2) If $A, B \subseteq \mathbb{R}$ are nonempty sets such that for every $a \in A$ and for every $b \in B$ we have $a < b$, then $\sup A$ and $\inf B$ exist and $\sup A < \inf B$.

3.71. Let $A \subset \mathbb{R}$ be nonempty and suppose that $A \subset [a,b]$. Show that $\sup A$ and $\inf A$ exist and that

$$a \leq \inf A \leq \sup A \leq b.$$

3.72. Let $A, B \subset \mathbb{R}$ be non-empty, bounded sets. We define

$$A + B = \{a + b : a \in A \text{ and } b \in B\}.$$

Prove that $\sup(A+B) = \sup(A) + \sup(B)$. State and prove a similar result regarding infimums.

3.73. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ and $A \subseteq \mathbb{R}$, with $A \neq \varnothing$. Assume that $f(A)$ and $g(A)$ are bounded. Prove that

$$\sup(f+g)(A) \leq \sup f(A) + \sup g(A) \ .$$

Give an example where one has equality in the inequality and an example where one has strict inequality.

3.74. Let $A \subset \mathbb{R}$ be a non-empty, bounded set. Let $\alpha = \sup(A)$ and $\beta = \inf(A)$, and let $(a_n)_{n=1}^{\infty} \subset A$ be a convergent sequence, with $a = \lim_{n\to\infty} a_n$. Prove that $\beta \leq a \leq \alpha$.

3.75. In the notation of the previous result, give an example where the sequence (a_n) is strictly increasing and yet $a \neq \alpha$. Give an example where the sequence is strictly decreasing and yet $a \neq \beta$.

3.76. If $A \subset \mathbb{R}$ is a non-empty, bounded set and $B \subset A$ is nonempty, prove

$$\inf(A) \leq \inf(B) \leq \sup(B) \leq \sup(A)\ .$$

3.77. If $A, B \subset \mathbb{R}$ are both non-empty, bounded sets, prove

$$\sup(A \cup B) = \max\{\sup(A), \sup(B)\}\ .$$

8. Real and Rational Exponents

In calculus, we often consider functions $f : \mathbb{R} \to \mathbb{R}$ of the form $f(x) = 2^x$. To understand what the expression 2^x means we need to define a notion for exponentiating by an arbitrary real number, rather than just by integers. We will do so in this section, beginning with exponentiating by rational numbers.

3.78. Suppose that $a < b$ are positive real numbers. Show that, if k is any natural number (not including zero), $a^k < b^k$.

Definition. Suppose that $a \geq 0$ and $k \in \mathbb{N}$. We say that x is a k**th root** of a if $x \geq 0$ and $x^k = a$. The previous result shows that that kth roots are unique. (why?) We denote the kth root of a by $\sqrt[k]{a}$ or $a^{1/k}$.

3.79. Suppose that $x, a \in \mathbb{R}$ satisfy $x^k < a$. Show that there is a number $y \in \mathbb{R}$ so that $x < y$ and yet $y^k < a$. If $x^k > a$ show that there is a number $y \in \mathbb{R}$ such that $x > y$ and yet $y^k > a$.

Hint: For the first part, consider the sequence $(x + \frac{1}{n})^k$. To what real number does it converge? What does that tell you?

The next result demonstrates a powerful application of supremums.

3.80. Let $a \geq 0$ and $k \in \mathbb{N}$. Show that $\sqrt[k]{a}$ exists in $\mathbb{R}$.

We point out that the roots are a purely algebraic concept and yet the completeness of $\mathbb{R}$ implies their existence. This is a very special property of $\mathbb{R}$. We have seen for example that it does not hold for $\mathbb{Q}$.

3.81. Suppose that $r \in \mathbb{Q}$ and $a \geq 0$. Give a definition for a^r. Show that it agrees with the special cases when $r \in \mathbb{N}$ or $r = 1/k$ for some $k \in \mathbb{N}$. Prove the usual rules of exponents:

(1) $a^{sr} = (a^r)^s$ for $r, s \in \mathbb{Q}$ and $a \geq 0$,
(2) $a^{s+r} = a^r a^s$ for $r, s \in \mathbb{Q}$ and $a \geq 0$, and
(3) $a^{-r} = \dfrac{1}{a^r}$ for $r \in \mathbb{Q}$ and $a \geq 0$.

Of course we can also consider roots for negative numbers.

3.82. Suppose that $a \in \mathbb{R}$ (a not necessarily positive), for which k is there a real number x with $x^k = a$. Give the values for all such x in terms of $\sqrt[k]{|a|}$, which we have already defined.

Hint: Of course we know that if k is even and $a < 0$, there is no $x \in \mathbb{R}$ with $x^k = a$. Proving it, however, is not completely trivial. Derive a contradiction by deciding whether such an x would have to be positive or negative.

We now lay the ground work for exponentiating by real numbers.

3.83. Suppose that $a > 0$ and show that $(\sqrt[n]{a})_{n=1}^{\infty}$ converges to one.

Hint: Begin with the case $a > 1$.

3.84. Suppose that (r_n) is a Cauchy sequence of rational numbers. Show that if $a \geq 0$, (a^{r_n}) is a also Cauchy.

3.85. Suppose that (r_n) and (s_n) are two sequences of rational numbers which converge to the same real number r. Show that $\lim\limits_{n\to\infty} a^{r_n} = \lim\limits_{n\to\infty} a^{s_n}$.

Definition. Suppose that $r \in \mathbb{R}$ and $a \geq 0$. Pick a sequence (r_n) of rational numbers that converges to r. Then we define a^r to be $\lim_{n\to\infty} a^{r_n}$.

3.86. Show that the normal rules of exponents hold for real exponents as well as rational ones. Show that if $a > 0$, the map $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = a^x$ is injective.

CHAPTER 4

Limits of Functions and Continuity

1. Limits of Functions

In the previous chapter, we used the notion of convergence to refine our understanding of the real numbers. In this chapter, we use many of these same intuitive notions to study *functions*. First a word about conventions. In the section in which we introduced functions, we were very careful to specify that a function has three ingredients: a domain, a codomain, and a rule (which is technically a certain subset of the cartesian product of the domain and codomain).

For the remainder of the book, however, (unless we specify otherwise) the codomain of our functions will always be $\mathbb{R}$ itself. Likewise the domain will always be some subset of $\mathbb{R}$. Thus when we say f is a *function*, we mean that it is a function whose codomain is $\mathbb{R}$ and whose domain is some (potentially unspecified) subset of $\mathbb{R}$.

Furthermore, though it is also technically incorrect, we will often specify functions by giving only a formula or rule (as calculus textbooks often do). For example, we might say "consider the function $f(x) = 1/x$." This is technically not enough information. In this case, there are a large variety of subsets of $\mathbb{R}$ which could serve as the domain. Nevertheless when we say something like the above, we really mean that the domain should be every real number for which the formula makes sense. In our previous case then, the domain would be $\mathbb{R}\backslash\{0\}$, that is, all real numbers except for zero.

At times, this sort of vagueness can get one into trouble, but in all the cases we will consider, the intended domain will be clear. In fact, it is a (perhaps unfortunate) common practice in mathematics for an author, when defining an object, to specify only the noteworthy piece of data and leave it up to the reader to surmise the other pieces. In the case of functions, this is the practice taken by most calculus books (though perhaps without warning) and it will be our practice as well.

Now that we have made our conventions clear, we proceed to discuss limits of functions. Intuitively, we saw that a sequence (a_n) has limit L if it gets arbitrarily close to L as n gets larger. The intuitive idea of a function is similar, but whereas sequences themselves (possibly) have limits, functions (possibly) have limits *at points* in $\mathbb{R}$. In other words, if we want to consider the limit of a function, we need to specify the point at which we are focusing our attention.

Intuitively then, a function f has limit L at the point $p \in \mathbb{R}$ if the value, $f(x)$, of the function gets arbitrarily close to L as x gets very close to p. In particular, the limit does not depend on the value of f at p but only on the value of f at points x *near* p. Indeed, for a limit to exist at p it is not even necessary that f be defined at p. It is, however, necessary that f be defined at points x *near* p, in a sense made precise below.

Definition. Let f be a function and $L, p \in \mathbb{R}$. We say that L is a **limit of f as x approaches p** if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x \in \mathbb{R}$ with $0 < |x - p| < \delta$, x is in the domain of f and $|f(x) - L| < \epsilon$.

Again you should think careful about this definition. How is closeness to L measured? How is closeness to p measured? Where does $f(x)$ need to be defined? How does δ relate to ϵ? How does this definition compare to the notion of a limit that we have described above? How does it compare to the notion of a limit that you developed in your calculus courses? You should also try to draw some pictures of the situation. Where should ϵ go? What about δ?

Notice that when proving that L is a limit of f as x approaches p, we are given an arbitrary $\epsilon > 0$ and have to find a $\delta > 0$ exactly as we had to find a $N \in \mathbb{N}$ when proving that L was the limit of a sequence. In your proofs, you will need to choose δ judiciously, depending upon what f, p, and ϵ are.

Exactly as for sequences, we have to show that limits of functions are in fact unique.

4.1. Let f be a function and $p \in \mathbb{R}$. Suppose L and M are both limits of f as x approaches p. Show that $L = M$.

This shows that if a limit of f as x approaches p exists then it is unique. As with sequences, we can talk about *the* limit of f as x approaches p and write $\lim\limits_{x \to p} f(x) = L$.

Similar to sequences, we should not write $\lim\limits_{x \to p} f(x)$ unless we know that limit exists. In other words, as with sequences, when we write

$\lim_{x\to p} f(x) = L$ we are making two assertions: the limit of f as x approaches p exists, and its value is L.

4.2. Let f be a function and $p, L \in \mathbb{R}$. Give the negation of the definition of "$\lim_{x\to p} f(x) = L$".

Again the negation of "$\lim_{x\to p} f(x) = L$" is not "$\lim_{x\to p} f(x) \neq L$" since the latter implies implies the existence of the limit. The negation could read "$f(x)$ does not approach L as x approaches p." For this there are two possibilities: $\lim_{x\to p} f(x)$ exists but does not equal L, or f has no limit as x approaches p, but try to give the negation in terms of ϵ and δ as in the definition.

4.3. Let f be a function and $p, L \in \mathbb{R}$. Assume there exists $\epsilon_0 > 0$ so that x is in the domain of f if $0 < |x - p| < \epsilon_0$. Determine which, if any, of the following are equivalent to $\lim_{x\to p} f(x) = L$.

(1) For all $\epsilon > 0$, there is a $\delta > 0$ and an x with $0 < |x - p| < \delta$ and $|f(x) - L| < \epsilon$.
(2) For all $n \in \mathbb{N}$, there exists a $\delta > 0$ such that for all $x \in \mathbb{R}$ with $0 < |x - p| < \delta$, we have $|f(x) - L| < 1/n$.
(3) For all $n \in \mathbb{N}$, there exists a $m \in \mathbb{N}$ such that for all $x \in \mathbb{R}$ with $0 < |x - p| < 1/m$, we have $|f(x) - L| < 1/n$.
(4) For all $\delta > 0$ there is an x in the domain of f such that $0 < |x - p| < \delta$ and $f(x) = L$.
(5) There exists a $\delta > 0$ such that for all $\epsilon > 0$, if x is in the domain of f and $0 < |x - p| < \delta$ then $|f(x) - L| < \epsilon$.

As usual, an important first step in understanding a new concept (or a least a concept that we are meeting rigorously for the first time), is considering examples. We begin with the simplest functions: constant functions.

4.4. Let $a \in \mathbb{R}$ and let f be the function given by $f(x) = a$ for all $x \in \mathbb{R}$. Let $p \in \mathbb{R}$ be arbitrary. Prove that $\lim_{x\to p} f(x) = a$.

Notice that, in the case of a constant function, you can choose a δ that does not depend on ϵ. This behavior is extremely rare for a function.

We now proceed with a few more examples.

4.5. Let f be the function given by $f(x) = x$ and let $p \in \mathbb{R}$. Prove $\lim_{x \to p} f(x) = p$.

Now we see that δ depends on ϵ and that δ goes to 0 as ϵ goes to 0. However the choice of δ still does not depend on the choice of p.

4.6. Let f be the function given by $f(x) = 3x - 5$ and let $p \in \mathbb{R}$. Prove $\lim_{x \to p} f(x) = 3p - 5$.

4.7. Let f be the function given by $f(x) = x/x$. Show that $\lim_{x \to 0} f(x) = 1$, despite the fact that f is not defined at zero.

The previous example shows that indeed a function need not be defined at a point to have a limit there. In fact, this is even more obvious than we are making it seem: we can always take a function that has a limit at some point and then create a new function by removing this point from the domain. This new function will have the behavior to which we are referring. The previous example is, however, interesting since it is an example of a function of this kind defined by a simple formula.

4.8. Let f be the function given by $f(x) = x^2$ and let $p \in \mathbb{R}$. Prove $\lim_{x \to p} f(x) = p^2$.

Notice that regardless of how we produce our choice of δ for the previous proof, it always depends on the point p. For a fixed ϵ we see that the δ required gets smaller and smaller as $|p|$ gets bigger and bigger. How could you tell this from looking at the graph of f?

4.9. Prove that $\lim_{x \to 2}(2x^2 - x + 1) = 7$.

4.10. Let f be given by

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

Prove that $\lim_{x \to p} f(x)$ does not exist for any $p \in \mathbb{R}$.

4.11. Let f be given by

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

Prove that $\lim_{x \to p} f(x)$ exists for only one value of p.

We now add a brief discussion regarding the domain of a function for which we want to consider a limit at $p \in \mathbb{R}$. Indeed, one of the requirements of the definition of a limit says that we must find a $\delta > 0$ such that for all $x \in \mathbb{R}$ with $0 < |x - p| < \delta$, x is in the domain of f. In other words, the set $(p - \delta, p + \delta) \setminus \{p\}$ must be contained in the domain of f. Hence we see that f is defined on $I \setminus \{p\}$, where I is an open interval containing p. It is thus convenient to make the following definition.

Definition. Let $p \in \mathbb{R}$. We say a function f is **defined near** p if the domain of f contains a set of the form $I \setminus \{p\}$ where I is an open interval containing p.

In other words f is defined near p if there is some $\delta > 0$ such that for $x \in \mathbb{R}$ with $0 < |x - p| < \delta$, x is in the domain in f. Essentially that means that we can find a small sliver of the real line containing p on which f is defined (except for possibly at p). Do you think that this is a good meaning for the phrase 'near p'? Why or why not?

4.12. Suppose $p, L \in \mathbb{R}$ and that f is a function. Show that $\lim_{x \to p} f(x) = L$ if and only if f is defined near p and for every $\epsilon > 0$ there is a $\delta > 0$ such that for $x \in \mathbb{R}$ with $0 < |x - p| < \delta$, $|f(x) - L| < \epsilon$ whenever $f(x)$ is defined.

The preceding discussion and result had very little content to it and was essentially semantics or technicalities. Nevertheless, though we would all prefer to avoid them, technicalities play an essential role in the study of mathematics. In particular, the discussion above is important because of certain complications that might be introduced if were less careful.

For example, suppose that we had defined "$f(x)$ converges to L as x approaches p" to mean "for every $\epsilon > 0$ there is a $\delta > 0$ such that for

$x \in \mathbb{R}$ with $0 < |x - p| < \delta$, $|f(x) - L| < \epsilon$ whenever $f(x)$ is defined." We just saw that this definition is equivalent to ours, provided that f is defined near p. If, however, the domain of f contained no points in an open interval containing p, then *any* real number would satisfy this definition of limit at p! Clearly this means that the definition above is not a good one. We conclude that our definition needs to force f to be defined near p (or at least at some points near p: some authors use a slightly more general definition, but we have used ours to avoid even more technicalities).

More generally, we say a property of functions is true **near** p if it is true on a set of the form $I - \{p\}$ with I an open interval containing p. For example we might say $f(x) \leq g(x)$ is true near p to say that $f(x) \leq g(x)$ is true for all x on some set of the form $I \backslash \{p\}$ (in particular we are asserting that both f and g are defined near p).

Now that we have considered a few examples and thought a little bit about the above definition, the next logical step is to study its basic properties. Before we do so however, we introduce another important concept which will help us to better understand limits. It is the notion of a 'one-sided limit'. Intuitively a one-sided limit describes the situation when we approach the number p from only one side rather then from either side.

Definition. Let f be a function and $p, L \in \mathbb{R}$. We say that L is a **right-hand limit of f as x approaches p** if for every $\epsilon > 0$, there exists a $\delta > 0$ such that for $x \in \mathbb{R}$ with $p < x < p + \delta$, x is in the domain of f and $|f(x) - L| < \epsilon$. Similarly, we say that L is a **left-hand limit of f as x approaches p** if for every $\epsilon > 0$ there exists a $\delta > 0$, such that for $x \in \mathbb{R}$ with $p - \delta < x < p$, x is in the domain of f and $|f(x) - L| < \epsilon$.

4.13. Show that left-hand and right-hand limits are unique.

The left-hand limit is denoted by $\lim\limits_{x \to p^-} f(x) = L$ and the right-hand limit is denoted by $\lim\limits_{x \to p^+} f(x) = L$.

Left-hand and right-hand limits also require something about the domain of the function f. We might say that a right-hand limit requires f to be defined "near p on the right" and likewise for left-hand limits.

4.14. Give an example of a function f and a point $p \in \mathbb{R}$ such that $\lim\limits_{x \to p^+} f(x)$ exists but $\lim\limits_{x \to p^-} f(x)$ does not. Give an example of a function and a point where the left-hand limit and the right-hand limit both exist but they are not equal.

4.15. Let f be a function and $p \in \mathbb{R}$. Prove that $\lim_{x\to p} f(x) = L$ if and only if $\lim_{x\to p^-} f(x) = \lim_{x\to p^+} f(x) = L$.

Thus in order for the limit at p to exist it is necessary and sufficient for the right-hand and left-hand limits to exist and be equal. Now that we have considered one-sided limits, we proceed to study the basic properties of limits of functions. Our first task is give a relationship between limits of functions and limits of sequences. Indeed, the next result is known as the **sequential characterization of limits** (though it might be more precise to say that it is the "sequential characterization of limits of functions.")

4.16. Let $p, L \in \mathbb{R}$ and suppose that f is a function defined near p. Let D be the domain of f. Prove that $\lim_{x\to p} f(x) = L$ if and only if for every sequence $(x_n) \subseteq D \setminus \{p\}$, with $x_n \to p$, we have $\lim_{n\to\infty} f(x_n) = L$.

Hint: To prove "if for every sequence $(x_n) \subset D \setminus \{a\}$ with $\lim_{n\to\infty} x_n = p$ we have $\lim_{n\to\infty} f(x_n) = L$ implies that $\lim_{x\to p} f(x) = L$" you should switch to the contrapositive.

Why did we have to assume that f was defined near p in the previous result?

4.17. a) State the contrapositive of the sequential characterization of limits (i.e., get a new if and only if statement by negating both sides).

b) Let $f : D \to \mathbb{R}$ be defined near a point $p \in \mathbb{R}$. Prove that the limit of f as x approaches p does not exist if and only if there exists a sequence $(x_n) \subseteq D \setminus \{p\}$ with $x_n \to p$ so that $(f(x_n))_{n=1}^{\infty}$ diverges.

This statement is quite useful for proving that a function has no limit as x approaches p.

4.18. Which of these limits (if any) exist? Prove your answer.

(1) $\lim_{x\to 0} \sin\left(\frac{1}{x}\right)$.
(2) $\lim_{x\to 0} x \sin\left(\frac{1}{x}\right)$.

Hint: By definition $\sin(\theta)$ is the y-coordinate of the point on the unit circle at angle θ from the positive x-axis. You will be able to complete this exercise using only this definition (and the results of the course thus far).

We now prove the Limit Laws for functions (which are of course analogous to the limit laws for sequences). As in the sequential cause, we need to begin with a preliminary result which will help us.

4.19. Let f be a function let $p \in \mathbb{R}$. Assume $\lim\limits_{x \to p} = L$ and $L > 0$. Prove $f(x) > L/2$ near p.

Now we prove the Limit Laws. If f and g are functions, we remark the the formulas $f(x) + g(x)$ and $f(x)g(x)$ make sense as long as x is in both the domain of f and the domain of g and so the domain of the functions given by these formulas is defined to be the intersection of the domain of f with that of g. Likewise, the domain of the function given by $f(x)/g(x)$ is the intersection of the domain of f and the set on which g is nonzero.

4.20. Let $p \in \mathbb{R}$ and let f and g be functions satisfying

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

Let $c \in \mathbb{R}$. Prove that

(1) $\lim\limits_{x \to p} c \cdot f(x) = c \cdot L$.
(2) $\lim\limits_{x \to p} \big(f(x) + g(x)\big) = L + M$.
(3) $\lim\limits_{x \to p} \big(f(x) \cdot g(x)\big) = L \cdot M$.
(4) If g is nonzero near p and $M \neq 0$ then $\lim\limits_{x \to p} \frac{f(x)}{g(x)} = \frac{L}{M}$.

Hint: There are two ways to prove these statements; one is to use the definition of limit directly as with sequences, and the other is to use the sequential characterization of limits and the Limit Laws for sequences.

As with sequences, we have a result discussing the interplay between the order on $\mathbb{R}$ and limits of functions.

4.21. Let f be a function and $p \in \mathbb{R}$. Assume that $a \leq f(x) \leq b$ near p. Prove that if $L = \lim\limits_{x \to p} f(x)$, then $L \in [a, b]$.

We also have the Squeeze Theorem for functions.

4.22. Let f, g, and h be functions and let $p \in \mathbb{R}$. Suppose that $g(x) \leq f(x) \leq h(x)$ near p. Prove that if $\lim_{x \to p} g(x) = \lim_{x \to p} h(x) = L \in \mathbb{R}$, then $\lim_{x \to p} f(x) = L$.

2. Continuous Functions

Hopefully your calculus courses convinced you that continuity is an important property. At some point, you probably heard your instructor say something to the effect of "a function is continuous if you can draw its graph without picking up your pencil." In some sense this is the intuitive idea behind continuity, but in this case we will find the precise definition leads us a little further from the intuitive notion than in the previous cases we have considered.

With a few technicalities to be considered, to say that a function f is continuous at $p \in \mathbb{R}$ is to say that p is in the domain of f (which is not a requirement to consider the limit of the function at p) and the value of $f(x)$ gets arbitrarily close to $f(p)$ as x gets close to p.

Definition. Let f be a function and let p be in the domain of f. We say that **f is continuous at p** if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for $x \in \mathbb{R}$ with $|x - p| < \delta$, we have $|f(x) - f(p)| < \epsilon$ whenever x is in the domain of f.

4.23. Let f be a function and let p be in the domain of f. Negate the definition of "f is continuous at p".

As always, we begin by considering examples

4.24. Show that the function $f(x) = x$ is continuous for every $p \in \mathbb{R}$.

4.25. Let f be given by

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

Find all points $p \in \mathbb{R}$ at which f is continuous. Justify your answer.

4.26. Let f be given by

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

Find all points $p \in \mathbb{R}$ at which f is continuous. Justify your answer.

4.27. Let f be given by

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

Find all points $p \in \mathbb{R}$ at which f is continuous. Justify your answer.

The next result can help add to our intuition of continuity.

4.28. Suppose that f is function and p is in its domain. Show that f is continuous at p if and only if for each open interval I containing $f(p)$, there is an open interval, J, containing p such that $f(x) \in I$ whenever x is in the domain of f and $x \in J$.

Thus we might say that f is continuous at p if, for x near p, $f(x)$ is always near (or equal to) $f(p)$ (when defined).

We will see momentarily that continuity and limits of functions are closely related (as we might expect by comparing their definitions). However there are some notable differences.

4.29. Define a function $f : \{0\} \to \mathbb{R}$ by putting $f(0) = 1$. Show that f is continuous at 0. Is the same true if $f : \{0\} \cup [1, 2] \to \mathbb{R}$ is given by $f(0) = 1$ and $f(x) = x$ for $x \in [1, 2]$.

The function above is certainly not defined near zero and yet it is continuous at zero. A difference then between our definition of continuity at a point and our definition of the limit at a point then is that the former does not require the function to be defined near the point. There are certain important reasons that we want to do it this way (mostly to line up with notions from higher mathematics courses such as ‘topology’), but other authors may use a different convention.

We now make the connection between limits and continuity explicit.

4.30. Let f be a function and let p be in the domain of f. Assume that f is defined near p. Prove that f is continuous at p if and only if $\lim_{x \to p} f(x) = f(p)$.

There is also a connection between limits of sequences and continuity. Not surprisingly, it is called the **sequential characterization of continuity** and it is the next result. In proving it, you should be careful to consider what happens if f is not defined near p.

4.31. Let f be a function, let D be the domain of f and let $p \in D$. Prove that f is continuous at p if and only if for every sequence $(x_n) \subseteq D$ with $\lim_{n \to \infty} x_n = p$, we have $\lim_{n \to \infty} f(x_n) = f(p)$.

4.32. Give the contrapositive to the sequential characterization of continuity.

Not surprisingly, we have results for continuity which are analogous to the Limit Laws.

4.33. Let f and g be functions and let p be a real number in the domain of each. Assume that f and g are continuous at p. Let $c \in \mathbb{R}$. Prove:

(1) $f + g$ is continuous at p.
(2) $c \cdot f$ is continuous at p.
(3) $f \cdot g$ is continuous at p.
(4) If $g(p) \neq 0$ then $\frac{f(x)}{g(x)}$ is continuous at p.

In the above problem you should consider the domains of the various functions. For example the domain of $f + g$ is $\{x : x \in \mathrm{dom}(f) \cap \mathrm{dom}(g)\}$. For the next problem, what is the domain of $g \circ f$?

4.34. Let f and g be functions and let p be a point in the domain of f such that $f(p)$ is in the domain of g. Prove that if f is continuous at p and g is continuous at $f(p)$ then $g \circ f$ is continuous at p.

This can be proved either directly from the definition or by repeated application of Problem 4.31.

So far, we have been discussing continuity only at points, but it is probably far more important to consider continuity on sets.

Definition. Let f be a function and S a subset of $\mathbb{R}$. We say that f is **continuous on S** if S is contained in the domain of f and if for each $p \in S$ and $\epsilon > 0$, there is a $\delta > 0$ such that for all $x \in \mathbb{R}$ with $|x - p| < \delta$, we have $|f(x) - f(p)| < \epsilon$ whenever $x \in S$. We say a function is **continuous** if it is continuous on its domain.

We remark that the previous definition is used most frequently in the case that S is an interval (although not always). It is also a bit subtle. For example there is a difference between being continuous on S and being continuous at every point of S.

4.35. Define $f : [0, \infty) \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & x \in [0,1] \\ 1 & x \in (1, \infty). \end{cases}$$

Show that f is continuous on $[0,1]$ and yet it is not continuous at every point of $[0,1]$. What part of the definition of 'continuous on S' allows this to be the case? What is

$$\{x \in [0, \infty) : f \text{ is continuous at } x\}?$$

Assuming we rid ourselves of some technicalities, however, the two notions do indeed coincide.

4.36. Suppose that S is a subset of $\mathbb{R}$ and f is a function. Assume in addition that S is the domain of f. Show that f is continuous on S if and only if it is continuous at every point of S.

In other words, if f is a function whose domain contains a set S, f is continuous on S if and only if $f|_S$ is continuous.

4.37. Let f and g be functions and let S be a subset of $\mathbb{R}$. Suppose that f and g are continuous on S. Let $c \in \mathbb{R}$. Prove:

(1) $f + g$ is continuous on S.
(2) $c \cdot f$ is continuous on S.
(3) $f \cdot g$ is continuous on S.
(4) $\frac{f(x)}{g(x)}$ is continuous on S, provided that g is never zero on S.

The terminology here is perhaps also sightly different than the terminology one would see in a calculus course.

4.38. Show that function $f(x) = 1/x$ is continuous.

Of course a calculus student would never say $f(x) = 1/x$ is continuous. What we are really saying is that it is continuous on its domain, that is, it is continuous on the set $(-\infty, 0) \cup (0, \infty)$. Calculus students would agree that this is the appropriate result. It is indeed false that the function is continuous on $\mathbb{R}$ because it is not even defined on all of $\mathbb{R}$.

Definition. We recall that a **polynomial** is a function on $f : \mathbb{R} \to \mathbb{R}$ of the form

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

for an integer $n \geq 0$ and $a_i \in \mathbb{R}$. If f is not the zero function (that is, the function whose every value is zero), we may assume $a_n \neq 0$ and we call n the **degree** of f. A **linear function** is a function which is either a polynomial of degree zero or one or a function which is identically zero. Polynomials of degrees 2, 3, 4, and 5 are called **quadratic functions**, **cubic functions**, **quartic functions**, and **quintic functions**, respectively.

Definition. A **rational function** is a function of the form $f(x) = g(x)/h(x)$ where g and h are polynomials (so that the domain of f is the set where h is nonzero).

4.39. Prove that every polynomial is continuous on $\mathbb{R}$ and that every rational function is continuous.

4.40. Let $a \geq 0$. Show that the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = a^x$ is continuous.

3. Theorems About Continuous Functions

One of the fundamental theorems regarding continuous functions, more specifically functions which are continuous on an closed intervals, is the **Intermediate Value Theorem**.

We give the explicit formulation next, but first we discuss the intuitive statement. Suppose that f is a function and we know that f is continuous on $[a, b]$. Intuitively, we can draw the graph of f (at least between a and b) without picking up our pencil. Hence if y is some value between the beginning value, $f(a)$, of f and the ending value, $f(b)$, of f, we should expect that at some point the graph of f should have to 'cross' the value y. In other words, there should be a $c \in [a, b]$ with $f(c) = y$.

4.41. Suppose $a < b$ and f is a function which is continuous on $[a, b]$. If $f(a) \leq y \leq f(b)$ or $f(a) \geq y \geq f(b)$ then there exists $c \in (a, b)$ with $f(c) = y$.

Hint: Suppose $f(a) < y < f(b)$ and let $E = \{x \in [a, b] : f(x) < y\}$. Let $p = \sup E$. Show the point p can be written as the limit of a sequence $x_n \in E$ and as the limit of a sequence $x_n' \in [a, b] \setminus E$. Then prove that $f(p) = y$.

Notice that in proving the previous result, we needed to use the completeness axiom. This is not a coincidence as should perhaps be intuitively clear. Indeed, if there were 'gaps' in the real line than those gaps would allow a function to 'jump over' the value y without actually hitting it.

4.42. Suppose that f is a polynomial of odd degree. Show that f has a zero. That is, show that there is some $a \in \mathbb{R}$ with $f(a) = 0$.

4.43. Suppose that $a \geq 0$ and suppose $k \in \mathbb{N}$. Use the intermediate value theorem to give another proof that $\sqrt[k]{a}$ exists.

4.44. Suppose that $a, b > 0$. Show that there is a unique number $c \in \mathbb{R}$ so that $a^c = b$.

Definition. If $a, b \geq 0$, the unique number c with $a^c = b$ is called the **logarithm base *a* of** b and denoted $\log_a(b)$.

4.45. Show the usual rules of logarithms hold.

We next study the images of intervals under continuous functions.

4.46. Let $I \subseteq \mathbb{R}$ be any interval and suppose that f is a function which is continuous on I. Furthermore, assume that f is nonconstant on I (meaning that f takes on more than one value on I). Prove that $f(I)$ is an interval.

Hint: First prove that it suffices to show that given any two points $c, d \in f(I)$ the entire interval between them is contained in $f(I)$.

In general we cannot say any more about the interval $f(I)$. In other words, it might be bounded or unbounded and it might be open or closed or neither.

Definition. Let S be a subset of R and let f be a function. Then we say that f is **bounded** on S if S is contained in the domain of f and the set $f(S)$ is bounded. Thus if f is defined on S, it is bounded on S if and only if there exists $K \in \mathbb{R}$ such that $|f(x)| \leq K$ for all $x \in S$ (why?).

4.47. Give an example of a function which is continuous on $(0, 1)$ but not bounded on $(0, 1)$. Given an example of a function which is continuous on $(0, 1]$ but bounded neither above nor below on $(0, 1]$.

Thus the continuous image of a bounded interval may be unbounded.

4.48. Give an example of a continuous function $f : (0, 1) \to \mathbb{R}$ such that $f\big((0, 1)\big)$ is a closed and bounded interval.

Thus the continuous image of an open interval may be a closed interval.

However, in the special case of a continuous function on a closed and bounded interval we can say a lot more.

4.49. Let I be a closed bounded interval and suppose that f is a function which is continuous on I. Show that f is bounded on I.

Hint: Put $I = [a, b]$ and proceed by contradiction. Suppose that f is not bounded above on I and construct a sequence $(x_n)_{n=1}^{\infty}$ with $x_n \in [a, b]$ for all $n \in \mathbb{N}$ such that

$$\lim_{n\to\infty} f(x_n) = +\infty$$

Now apply the sequential characterization of continuity, Problem 4.31, to obtain a contradiction.

The next result is called the **Extreme Value Theorem**.

4.50. Let I be a closed and bounded interval and suppose that f is a function which is continuous on I. Show that there exist $x_m, x_M \in I$ such that $f(x_m) \leq f(x) \leq f(x_M)$ for all $x \in [a, b]$. In other words, f attains a **maximum value** and a **minimum value** on I.

Hint: Construct a sequence $(x_n)_{n=1}^{\infty}$ with $x_n \in [a,b]$ for all $n \in \mathbb{N}$ such that

$$\lim_{n\to\infty} f(x_n) = \sup\{f(x) : x \in [a,b]\}.$$

4.51. Give an example of a function which is bounded and continuous on $(0,1)$ but has neither a maximum nor a minimum on $(0,1)$. Can you do the same for $(0,1]$?

4. Uniform Continuity

In this section we discuss an important property for functions which is actually stronger than continuity. Recall that the function f is continuous on the set S if S is in the domain of f and if for all $p \in S$ and for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x \in S$ with $|x-p| < \delta$ we have $|f(x) - f(p)| < \epsilon$.

For a continuous function, the δ generally depends upon both ϵ and the point p as previous exercises have illustrated. If we remove the dependence on p, we get our new concept.

Definition. Let f be a function and S a subset of $\mathbb{R}$ contained in the domain of f. We say that f is **uniformly continuous** on S if for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x, y \in S$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \epsilon$.

4.52. Suppose $S \subseteq \mathbb{R}$ and let f be a function. Prove that if f is uniformly continuous on S than it is continuous on S.

4.53. Let f be a **linear function**. That is, let $f(x) = mx + b$ for some $m, b \in \mathbb{R}$. Prove that $f(x)$ is uniformly continuous on $\mathbb{R}$.

4.54. Negate the definition of uniform continuity.

4.55. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Prove that f is not uniformly continuous on $\mathbb{R}$.

Hint: Fix an $\epsilon > 0$ and show that no matter what $\delta > 0$ is chosen you can always choose $x, y \in \mathbb{R}$ such that $|x - y| < \delta$ and $|x^2 - y^2| \geq \epsilon$.

4.56. Prove that if I is a bounded interval and f is a function which is uniformly continuous on I then f is bounded on I.

This together with Problem 4.47 shows that we can have continuous functions on $(0,1)$ that are not uniformly continuous (why?). As in the previous section the case of functions on closed and bounded intervals is very different.

4.57. Let I be a closed and bounded interval. Show that a function is continuous on I if and only it is uniformly continuous on I.

Hint: Suppose that f is not uniformly continuous. Then there exists an $\epsilon > 0$ such that for every $\delta > 0$ there exists $x, y \in [a,b]$ such that $|x-y| < \delta$ but $|f(x)-f(y)| \geq \epsilon$. Show that there exists two sequences $(x_n), (y_n) \subset [a,b]$ which both converge to the same point $p \in [a,b]$ but such that $|f(x_n) - f(y_n)| \geq \epsilon$. Show that this implies that f is not continuous on I.

CHAPTER 5

Differentiation

1. Derivatives

Towards the beginning of Chapter 3, we stated that it is convergence and limits that separate calculus from algebra or other subjects. In fact, this is perhaps a bit misleading as convergence and limits really belong exclusively to a branch of mathematics known as topology (parts of which of course appear in calculus). Calculus is characterized by the derivative and the integral. In this chapter, we study the former.

Definition. Let f be a function and p a point in the domain of f. Then f is said to be **differentiable at p** if

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

exists. If f is differentiable at p then the above limit is called the **derivative of** f **at** p, and is usually denoted by $f'(p)$. If S is a subset of $\mathbb{R}$, f is said to be **differentiable on** S if f is differentiable at p for all $p \in S$ (which in particular implies that f is defined on S). f is called **differentiable** if it is differentiable on its domain.

Before we go on to give some examples, we will give a brief discussion the relationship between derivatives and the domain. We know that in order for the limit above to exist, the function

$$\frac{f(x) - f(p)}{x - p}$$

must be defined near p. We see that this will happen if and only of f is defined near p (why?). Thus an implicit requirement for f to be differentiable at p is that it has to be defined near p (and it is explicitly assumed to be defined *at* p)

Also, rather than always speaking of the derivative at individual points, we may, given a function f, define its **derivative** as a function. The domain of derivative is the set of points at which f is differentiable and for such a point p, the value of the derivative is $f'(p)$. For obvious reasons, the derivative is typically denoted by f'.

5.1. We give some examples of derivatives.

(1) Let $c \in \mathbb{R}$ be arbitrary and let f be defined by defined by $f(x) = c$. Prove that f is differentiable on $\mathbb{R}$ and that $f'(x) = 0$ for all $x \in \mathbb{R}$.
(2) Let f be defined by $f(x) = x$. Prove that f is differentiable on $\mathbb{R}$ and that $f'(x) = 1$ for all $x \in \mathbb{R}$.
(3) Let f be defined by $f(x) = 3x-7$. Prove that f is differentiable on $\mathbb{R}$ and that $f'(x) = 3$ for all $x \in \mathbb{R}$.
(4) Let f be defined by $f(x) = x^2 - x + 1$. Prove that f is differentiable on $\mathbb{R}$ and find $f'(x)$ for $x \in \mathbb{R}$.
(5) Let f be defined by $f(x) = |x|$. Prove that f is differentiable on $\mathbb{R} - \{0\}$, but not at zero. Find f'.

We can also relate the notion of f being differentiable at p with our earlier notion of continuity.

5.2. Let f be a function and let p be in the domain of f. Prove that f is differentiable at p if and only if there exists a function ϕ, defined near and at p, such that ϕ is continuous at p and

$$f(x) = f(p) + (x-p)\phi(x)$$

near p. Moreover, if such a ϕ exists then $f'(p) = \phi(p)$.

Now many of our results about differentiability of f will follow from our rules for continuous functions applied to ϕ.

5.3. Let f be a function and p a point in the domain of f. Prove that if f is differentiable at p then f is continuous at p.

In particular, many of the usual rules of differentiation, such as the Product Rule and the Quotient Rule come from Problem 4.37 after some simple algebra.

5.4. Let f and g be functions, $c \in \mathbb{R}$, and let p be a point in the domain of f and the domain of g. If f is differentiable at p and g is differentiable at p then

(1) $c \cdot f$ is differentiable at p and

$$(c \cdot f)'(p) = c \cdot f'(p).$$

(2) $f + g$ is differentiable at p and

$$(f+g)'(p) = f'(p) + g'(p).$$

(3) $f \cdot g$ is differentiable at p and

$$(f \cdot g)'(p) = f(p) \cdot g'(p) + f'(p) \cdot g(p).$$

(4) if $g(p) \neq 0$ then $\frac{f}{g}$ is differentiable at p and

$$\left(\frac{f}{g}\right)'(p) = \frac{g(p) \cdot f'(p) - f(p) \cdot g'(p)}{\big(g(p)\big)^2}.$$

Hint: You can prove these either directly from the definition or by writing $f(x) = f(p) + (x - p) \cdot \phi(x)$ and $g(x) = g(p) + (x - p) \cdot \psi(x)$ and using Problem 5.2. If you really want to help yourself understand, you should try both ways and decide which is easier or clearer.

We can also prove the power rule.

5.5. Suppose that n is a natural number. Define a function by $f(x) = x^n$. Show that f is differentiable for all points $p \in \mathbb{R}$ and $f'(p) = np^{n-1}$.

5.6. Suppose that n is a natural number. Define a function by $f(x) = x^{1/n}$. State the domain of f. Find f' (with proof) and state its domain.

Hint: For two real numbers a and b, check that

$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}).$$

How does this help?

5.7. Let $P(x) = a_n x^n + \cdots + a_1 x + a_0$ be a polynomial function. Show that P is differentiable and find (with proof) its derivative.

5.8. Let

$$R(x) = \frac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0}$$

be a rational function. Show that P is differentiable and find (with proof) its derivative.

There is one more standard differentiation rule: the Chain Rule.

5.9. Let f and g be functions. Let $p \in I$. Prove that if f is differentiable at p and g is differentiable at $f(p)$ then $g \circ f$ is differentiable at p and

$$(g \circ f)'(p) = g'\big(f(p)\big) \cdot f'(p).$$

Hint: Since g is differentiable at $f(p)$ there exists an appropriate function ψ such that

$$g(x) = g\big(f(p)\big) + \big(x - f(p)\big) \cdot \psi(x).$$

Replace x by $f(x)$ and then use the fact that $f(x) - f(p) = (x - p) \cdot \phi(x)$ for an appropriate ϕ.

5.10. Suppose that $r \in \mathbb{Q}$. Find the domain of $f(x) = x^r$. Find f' and its domain (with proof).

2. Theorems About Differentiable Functions

Perhaps the most important application of differentiation is optimization. Optimization is a word given to the process of finding the maximum or minimum value of a given function (often subject to one, or more, constraints).

Definition. Let f be a function. We say a point p in the domain of f is a **local maximum** for f if $f(x) \le f(p)$ for x near p. Explicitly this means that we can find a $\delta > 0$ such that $f(x) \le f(p)$ for all $x \in \mathbb{R}$ with $|x - p| < \delta$. Likewise, we say p is a **local minimum** if $f(x) \ge f(p)$ for x near p.

We begin with a result that relates local maxima and minima with differentiation.

5.11. Suppose that f is a function and p is either a local maximum or a local minimum of f. If f is differentiable at p, show that $f'(p) = 0$.

Hint: Assume p is a local maximum and consider the sign of

$$\lim_{x \to p^+} \frac{f(x) - f(p)}{x - p} \text{ and } \lim_{x \to p^-} \frac{f(x) - f(p)}{x - p}.$$

In particular, we can derive the theory behind a technique for solving many routine calculus problems.

Definition. Suppose that f is a function. $x \in \mathbb{R}$ is called a *critical point* of f if either f is not differentiable at x or if $f'(x) = 0$.

5.12. Suppose that f is a function which is continuous on the interval $[a, b]$. Then f attains a maximum and a minimum on $[a, b]$ and each occurs either at a critical point of f or at an endpoint of the interval $[a, b]$ (that is at a or b).

The previous observations also allow us to an important result known as Rolle's Theorem. A version of the theorem was first stated by Indian astronomer Bhaskara in the 12th century however. The first proof, however, seems to be due to Michel Rolle in 1691.

5.13. Let $a < b$. Suppose that f is a function which is continuous on $[a, b]$ and differentiable on (a, b) with $f(a) = f(b)$. Then there exists a point $c \in (a, b)$ with $f'(c) = 0$.

Hint: Show that f has either a maximum value or a minimum value at some $c \in (a, b)$.

An immediate consequence of Rolle's Theorem is the extremely important Mean Value Theorem. The Mean Value Theorem is used extensively in estimating the values of functions and for many other purposes (some of which we demonstrate below).

5.14. Let $a < b$. Suppose that f is a function which is continuous on $[a, b]$ and differentiable on (a, b) then there exists $c \in (a, b)$ with

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

Hint: Construct a linear function $l(x)$ with $l(a) = f(a)$, $l(b) = f(b)$ and consider $g(x) = f(x) - l(x)$.

We now give some consequences of the Mean Value Theorem.

5.15. Let $a < b$. Suppose f is a function which is continuous on $[a, b]$ and differentiable on (a, b). Assume the derivative of f is zero on (a, b). Then f is a constant on $[a, b]$. Conclude that the only functions of the form $f : \mathbb{R} \to \mathbb{R}$ whose derivatives are identically zero are constant.

5.16. Suppose that f and g are two functions which are differentiable on an open interval I. Suppose that that $f' = g'$ on I. Show that f and g differ by a constant. In other words, show that there is a $c \in \mathbb{R}$ with $g = f + c$.

5.17. Let $a < b$. Suppose that f is a function which is continuous on $[a, b]$ and differentiable on (a, b). Prove that

(1) if $f'(x) \geq 0$ for all $x \in (a, b)$ then f is increasing on $[a, b]$, i.e. if $a \leq x < y \leq b$, then $f(x) \leq f(y)$.
(2) if $f'(x) \leq 0$ for all $x \in (a, b)$ then f is decreasing on $[a, b]$, i.e. if $a \leq x < y \leq b$, then $f(x) \geq f(y)$.

5.18. Let $f(x) = ax^2 + bx + c$ be a quadratic function (so that in particular $a \neq 0$). Prove that if $a > 0$ then $f(x)$ has an unique absolute minimum. In other words, there is a $p \in \mathbb{R}$ such that $f(x) > f(p)$ for all $x \neq p$. Find p and $f(p)$. Likewise if $a < 0$, show that f has a unique absolute maximum and find it coordinates. Show that f has a root if and only if $b^2 - 4ac \geq 0$.

5.19. Suppose that $f(x) = ax^3 + bx^2 + cx + d$ is a cubic function. Suppose that $b^2 - 3ac < 0$. Show that f has exactly one zero.

Hint: We have already shown that f has a root. To show that it has at most one, argue by contradiction using the Mean Value Theorem (or more precisely Rolle's Theorem).

CHAPTER 6

Integration

1. The Definition

Our final chapter is the other half of calculus: the integral. The intuitive notion behind the integral is very simple. Specifically suppose that f is a function which is continuous on $[a, b]$ (we will see that it is not necessary to assume that f is continuous, but we will make this assumption for the sake of our intuition). Also assume for intuitional simplicity that f is positive on $[a, b]$. Finally assume that f is simple enough that we may easily draw its graph.

If, on the Cartesian plane, we draw the vertical lines $x = a$ and $x = b$, we see that there is a shape defined by the region surrounded by $x = a$, $x = b$, the graph of f and the x-axis. You should draw some examples of this situation. We have given one example in Figure 1. The point of integration is to find the area of this shape. Unless f is extremely simple (like a straight line or something), geometry does not give us a formula for this area. In fact, it's not even clear that we have precise definition for word 'area.'

Thus we need a new approach to finding (or even defining) this area. We begin by thinking of some objects of which we do know the area. The simplest is probably the rectangle: the area of a recangle should certainly be the product of its length and width. Our approach to finding more complicated areas will actually be a very clever use of this simple observation.

Our strategy begins by estimating the area in question. Indeed, one naive way of estimating the area with rectangles is with a single rectangle. To do this, as shown in Figure 2, we place a rectangle with its bottom side on the x-axis, and its left and right sides on the line $x = a$ and $x = b$. We place the top of the rectangle at *some* value of the function on $[a, b]$. In the example shown in the figure, we chose the lowest value of f, but any choice would give an estimate.

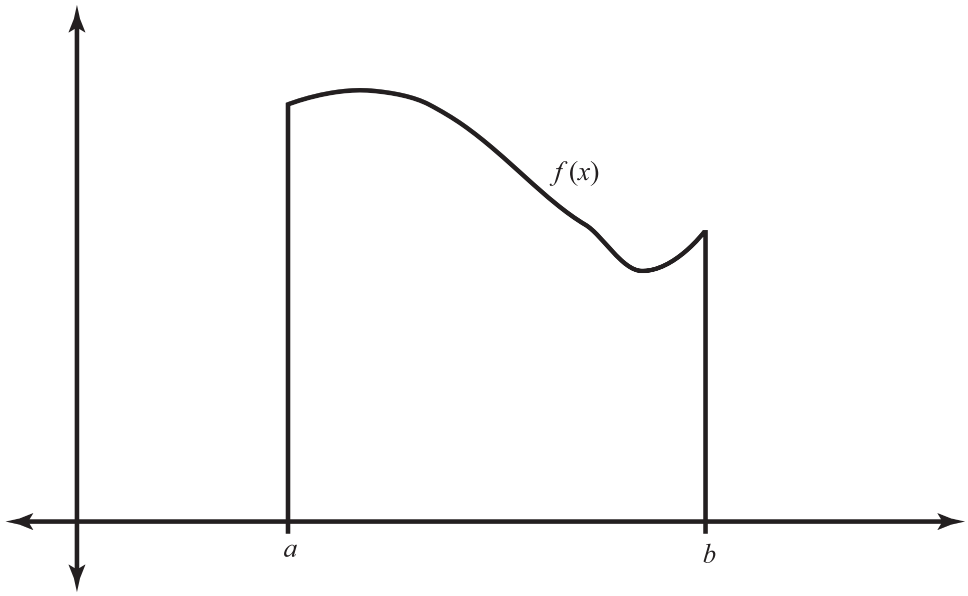

FIGURE 1. This is the shape of which we want to find the area.

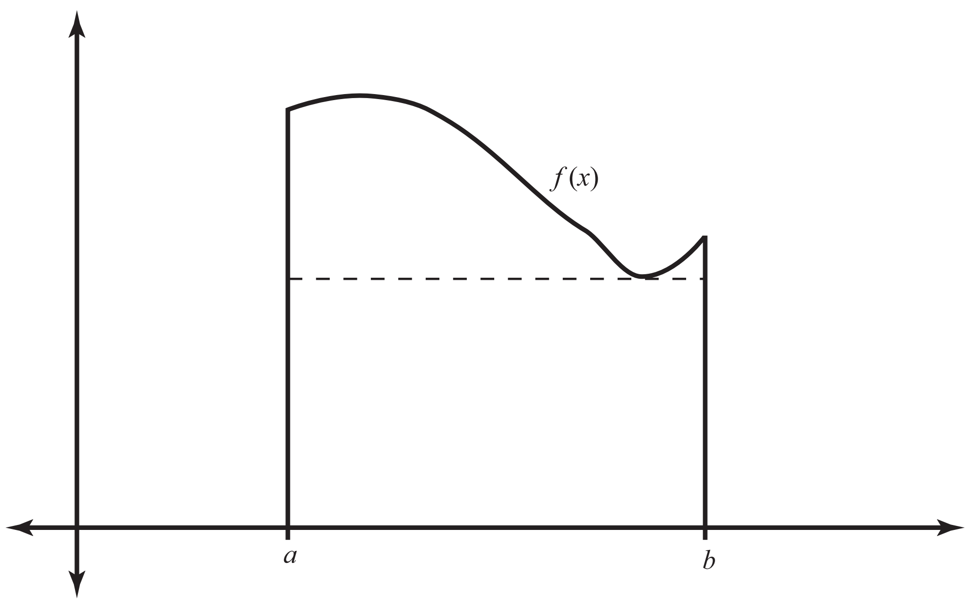

FIGURE 2. A naive way to estimate the area is by drawing a rectangle of similar size.

As you can tell from the figure, this estimate may not be all that precise: large parts of the shape lie outside of the rectangle and part of the rectangle lies outside of the shape. Thus we need to refine our method of estimation: instead of using one rectangle, we will use two, as the top picture of Figure 3.

We split the interval $[a, b]$ into two pieces by choosing some point, x_1, between a and b. We draw two rectangles: both of them have bottom side at the x-axis. One has left and right sides $x = a$ and $x = x_1$. Its top is at some value of f on $[a, x_1]$ (in this case the lowest). The other rectangle is similarly constructed between x_1 and b. To estimate the area of the shape in question, we simply add the areas of the two rectangles.

As we can see in the pictures, estimating in this way tends to work better than the first way. This should not be too surprising: we are using more information from the function than with the first estimation (i.e. two values rather than one). In the bottom picture of Figure 3, we use four rectangles: for this we have to choose 3 points between a and b and apply a similar construction. We expect that this estimation of the area will be even better than the second.

We now define the notion of integral precisely. An important step is our process was splitting the interval $[a, b]$ into multiple pieces so that we can use multiple rectangles. This leads to the first definition of this section.

Definition. A **partition** of an interval $[a, b]$ is a finite sequence $(x_1, \ldots, x_n)$ such that

$$a = x_0 < x_1 < \cdots < x_n = b.$$

If P and Q are two partitions of $[a, b]$ we say that Q **refines** P if every point in $[a, b]$ which occurs in P also occurs in Q.

What is the intuitive purpose of the notion of a partition? What about of the notion of refinement?

As above, a partition leads to an estimate of the area in question. Indeed, a partition with n elements leads us to n rectangles: the bottom, left, and right sides of the rectangles are given and we need only choose the tops. Of course the top of the rectangle which lies between x_i and x_{i+1} should be a value of f between x_i and x_{i+1}. There are basically two systematic ways of choosing this value. We can use the highest value of f or we can choose the lowest value. Of course, not every function has a highest or lowest value on an interval, but we can use the next best thing.

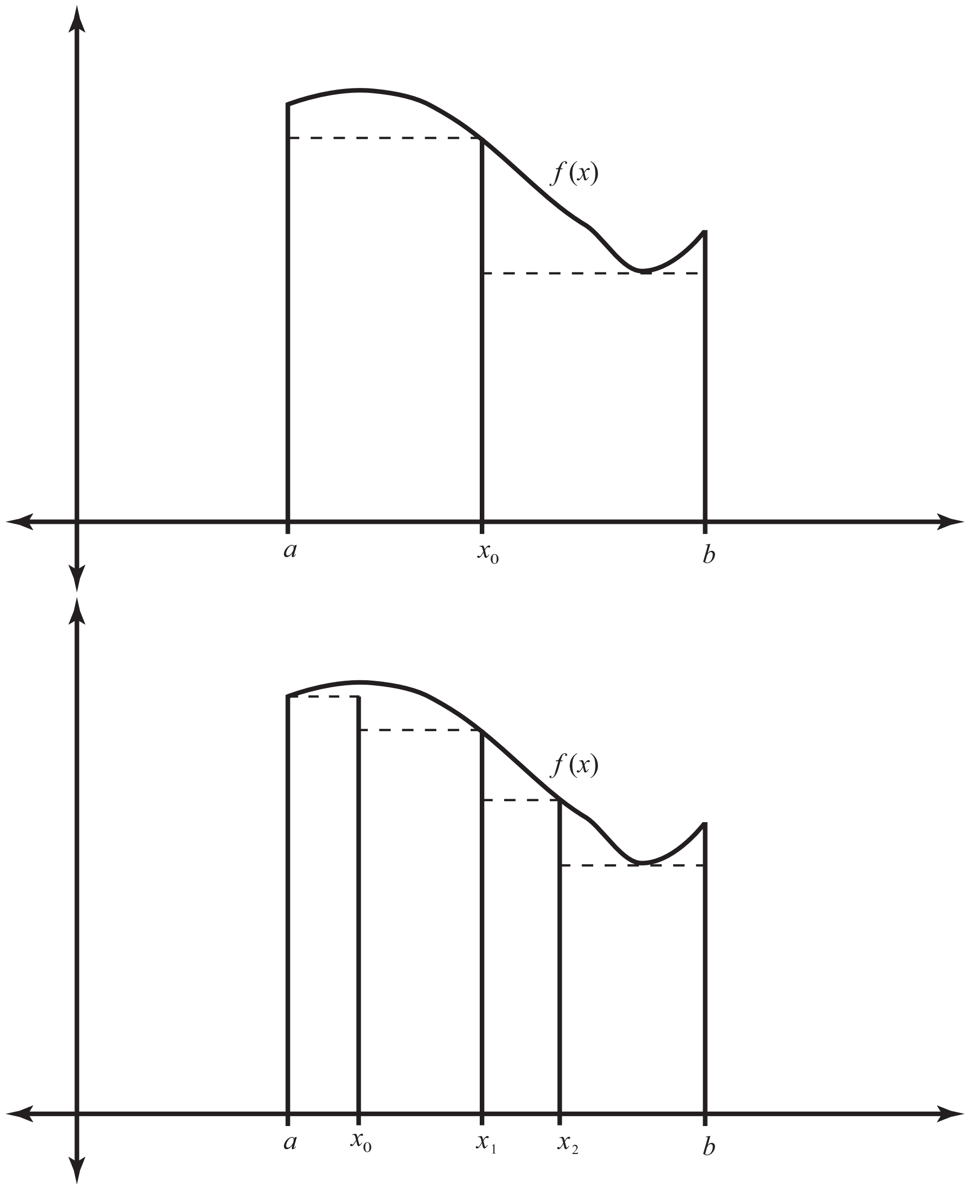

FIGURE 3. A somewhat better estimate is achieved by drawing multiple rectangles

Definition. Let $P = (x_0, x_1, \ldots, x_n)$ be a partition of $[a, b]$ and let f be a function whose domain contains $[a, b]$ which is bounded on $[a, b]$.

For $1 \leq i \leq n$ we set

$$\begin{aligned} M_i(f,P) &= \sup\{f(x) : x_{i-1} \leq x \leq x_i\} \\ &= \sup f\big([x_{i-1}, x_i]\big) \\ m_i(f,P) &= \inf\{f(x) : x_{i-1} \leq x \leq x_i\} \\ &= \inf f\big([x_{i-1}, x_i]\big) \\ \Delta_i &= x_i - x_{i-1}. \end{aligned}$$

Thus $m_i(f,P)$ is the 'lowest' value of f on the interval $[x_{i-1}, x_i]$ and $M_i(f,P)$ is the 'highest.' Notice that $m_i(f,P)$ and $M_i(f,P)$ exist because f is bounded on $[a,b]$. Δ_i is the length of the interval (and so the width of the resulting rectangle). These notions thus give us to systematic ways of estimating the area, given a partition.

Definition. Let P be a partition of the interval $[a,b]$ and let f be a function which is bounded on $[a,b]$. We define the **upper Riemann sum** of f with respect to P, denoted $U(f,P)$, by

$$U(f,P) = \sum_{i=1}^{n} M_i(f,P)\,\Delta_i$$

and we define the **lower Riemann sum** of f with respect to P, denoted $L(f,P)$, by

$$L(f,P) = \sum_{i=1}^{n} m_i(f,P)\,\Delta_i.$$

Check for yourself that both of these definitions provide estimates as we have described. Check that (given P) the upper Riemann sum is the highest estimate of the type we have described and the lower Riemann sum is the lowest. Given that our definitions have been a bit involved, we should also try some examples.

6.1. Let $f(x) = x$ and $g(x) = x^2$. Let $n \in \mathbb{N}$ and let P_n be the partition of $[0,1]$ given by $(0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1)$ be a partition of $[0,1]$. Draw a picture illustrating $L(f,P_4)$, $U(f,P_4)$, $L(g,P_4)$, and $U(g,P_4)$. Find expressions for $L(f,P_n)$, $U(f,P_n)$, $L(g,P_n)$, and $U(g,P_n)$.

6.2. Let $f : [a,b] \to \mathbb{R}$ be bounded and let P be a partition of $[a,b]$. Let $m = \inf f([a,b])$ and $M = \sup f([a,b])$.

(1) Prove

$$m\,(b-a) \le L(f,P) \le U(f,P) \le M\,(b-a).$$

(2) Prove that if Q refines P then

$$L(f,P) \le L(f,Q) \le U(f,Q) \le U(f,P).$$

Hint: For the second part, first prove that it suffices to show this when Q has one more element that P.

Thus we see that as our estimates get more and more refined, the lower sums go up. Thus we might expect that the best estimate that a lower sum will give us is the highest one. Likewise, we expect that the best estimates that the upper sums can give us is the lowest one. This leads us to the following definition (once again the highest lower sum and the lowest upper sum may not exist).

Definition. Let f be a function which is bounded on the interval $[a,b]$. We define the **upper Riemann integral** of f, denoted $U(f,[a,b])$, by

$$U(f,[a,b]) = \inf\big\{U(f,P) : P \text{ is a partition of } [a,b]\big\}.$$

We define the **lower Riemann integral** of f, denoted $L(f,[a,b])$, by

$$L(f,[a,b]) = \sup\big\{L(f,P) : P \text{ is a partition of } [a,b]\big\}.$$

(Why do these exist?) We say f is **Riemann integrable** on $[a,b]$ if $L(f,[a,b]) = U(f,[a,b])$. In this case we call the common value of $U(f,[a,b])$ and $L(f,[a,b])$ the **(definite) Riemann integral of** f over the interval $[a,b]$ which we denote by $\int_a^b f$ or $\int_a^b f(x)dx$.

6.3. Show that for f and g as in Problem 6.1 for all $n \in \mathbb{N}$, $U(f,P_n) \neq U(f,[0,1])$, $L(f,P_n) \neq L(f,[0,1])$ and similarly for g.

We begin by giving a non-example: that is, we show that not every bounded function is integrable.

6.4. Let f be given by

$$f(x) = \begin{cases} 0 & x \in [0,1] \setminus \mathbb{Q} \\ 1 & x \in [0,1] \cap \mathbb{Q}. \end{cases}$$

Show that f is not integrable.

2. Integrable Functions

Of course we hope that many functions are integrable as this will allow us to compute a lot of different areas. In this section, we will show that all continuous functions are in fact integrable, because the definition of integrability is a bit elaborate, it is necessary to do some preliminary work.

6.5. Let f be a function which is bounded on $[a, b]$ and let P and Q be partitions of $[a, b]$. Prove that $L(f, P) \leq U(f, Q)$.

Hint: Consider the partition $R = P \cup Q$.

6.6. Let f be a function which is bounded on $[a, b]$. Show that $L(f) \leq U(f)$.

6.7. Let $f(x) = x$. Let P_n be the partition of $[0, 1]$ given by $P_n = (0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1)$.

(1) Find $L(f, P_n)$ and $U(f, P_n)$.
(2) Find $U(f, P_n) - L(f, P_n)$.
(3) Show f is integrable on $[0, 1]$ and find $\int_0^1 f$.

The next result gives a very useful characterization of integrability.

6.8. Let f be a function which is bounded on $[a, b]$. Show that f is integrable on $[a, b]$ if and only if for all $\epsilon > 0$ there exists a partition $P = (x_0, \ldots, x_n)$ of the interval $[a, b]$ such that

$$U(f, P) - L(f, P) = \sum_{i=1}^{n} \big(M_i(f, P) - m_i(f, P)\big)\Delta_i < \epsilon.$$

6.9. Let f be a function which is increasing and bounded on $[a, b]$. Prove that f is integrable on $[a, b]$.

Hint: For a monotone function we know explicitly what $M_i(f, P)$ and $m_i(f, P)$ are.

6.10. Let f be a function which is continuous on $[a, b]$. Prove that f is integrable on $[a, b]$.

Hint: Use Problem 4.57 to conclude that f is uniformly continuous on $[a, b]$. Let $\epsilon > 0$ be arbitrary and choose $\delta > 0$ so that if $x, y \in [a, b]$

with $|x-y|<\delta$ then $|f(x)-f(y)|<\frac{\epsilon}{b-a}$. Let P be any partition with each $\Delta_i x<\delta$ and use Problem 6.8.

3. Properties of Integrals

For every concept we have introduced, we have seen that there is something like 'Limit Laws,' which are useful in computing examples. As you probably already know, integrals are no exception.

6.11. Let f and g be functions which are integrable on $[a,b]$. Let $c \in \mathbb{R}$ be an arbitrary constant. Prove that

(1) $c \cdot f$ is integrable on $[a,b]$ and
$$\int_a^b c\cdot f = c\int_a^b f.$$
(2) $f+g$ is integrable on $[a,b]$ and
$$\int_a^b (f+g) = \int_a^b f + \int_a^b g.$$

Hint: Show
$$M_i(f+g,P) \le M_i(f,P) + M_i(g,P)$$
and hence conclude that $U(f+g,P) \le U(f,P)+U(g,P)$. Similarly, show that $L(f+g,P) \ge L(f,P)+L(g,P)$. Use Problem 6.8 to conclude integrability. Then prove the equation.

6.12. Let f and g be functions which are integrable on $[a,b]$ with $f(x) \le g(x)$ for all $x \in [a,b]$. Prove that $\int_a^b f \le \int_a^b g$.

6.13. Let $a<c<b$ and let f be a function.

(1) Assume f is integrable on $[a,b]$. Prove that f is integrable on $[a,c]$ and $[c,b]$.
(2) Assume that f is integrable on $[a,c]$ and on $[c,b]$. Prove that f is integrable on $[a,b]$ and
$$\int_a^b f = \int_a^c f + \int_c^b f\ .$$

It is actually convenient for notational reasons for us to consider something like backwards integrals. In other words, for $a<b$, the concept we have been studying so far might be called the "integral from a to b." The "integral from b to a" is just the opposite.

Definition. If f is a function which is integrable on $[a,b]$ we define $\int_b^a f = -\int_a^b f$. We also define $\int_a^a f = 0$.

6.14. Let f be integrable on an interval containing a, b and c. Prove that, no matter the order of a, b, and c, we have

$$\int_a^b f = \int_a^c f + \int_c^b f \ .$$

6.15. Let f be a function which is integrable on $[a,b]$. Prove that

(1) $|f|$ is integrable on $[a,b]$
(2) $\left|\int_a^b f\right| \leq \int_a^b |f|$.

Hint: For the first part, prove that $M_i(|f|,P) - m_i(|f|,P) \leq M_i(f,P) - m_i(f,P)$.

6.16. Prove that if f and g are integrable on $[a,b]$ then $f \cdot g$ is integrable on $[a,b]$.

Hint: Since f and g are integrable we may define

$$M_f = \sup\{|f(x)| : a \leq x \leq b\},$$
$$M_g = \sup\{|g(x)| : a \leq x \leq b\}.$$

Use the fact that

$$f(x) \cdot g(x) - f(y) \cdot g(y) = f(x) \cdot \big(g(x) - g(y)\big) + \big(f(x) - f(y)\big) \cdot g(y)$$

to conclude

$$\begin{aligned} &M_i(f \cdot g, P) - m_i(f \cdot g, P) \\ &\qquad \leq L_f \cdot \big(M_i(g,P) - m_i(g,P)\big) + \big(M_i(f,P) - m_i(f,P)\big) \cdot M_g. \end{aligned}$$

Now use Problem 6.8

Thus the product of integral functions is integrable. Integrals, however, do not behave so well with respect to multiplication. In other words, we have seen the the product of sequences converges to the product of the limits of the two sequences and likewise for limits of functions. The analogous statement, however is not true for integrals.

6.17. Give an example of an interval $[a, b]$ and two functions f and g which are integrable on $[a, b]$ such that

$$\left[\int_a^b f(x)dx\right]\left[\int_a^b g(x)dx\right] \neq \int_a^b f(x)g(x)dx.$$

Given the complication resulting from the previous behavior, you are probably aware of some techniques people have developed to deal with integrals of products. We will state and prove them in a later section.

6.18. Given an example of a function f which is integrable on $[0, 1]$, but not continuous.

6.19. Show that the function $f(x) = \sqrt{1 - x^2}$ is integrable on $[0, 1]$.

Thus

$$\int_0^1 \sqrt{1 - x^2}dx$$

exists as a real number. But we also see that it is one fourth the area of the unit circle, that is it equals $\pi/4$. Multiplying by 4, we conclude that π exists as a real number.

4. Fundamental Theorems of Calculus

We saw that even computing $\int_0^1 xdx$ from the definition of integral took some work. Of course nobody actually computes integrals this way in calculus. Instead they use the Fundamental Theorem of Calculus. Loosely speaking, this theorem says that the derivative and the integral are opposites (or probably more appropriately inverses). As its name suggests, it is considered the crucial result in understanding calculus. It is commonly broken into two different theorems and we will also prove it this way.

6.20. Prove the first Fundamental Theorem of Calculus:

Let f be a function which is integrable on $[a, b]$. Define a function $F : [a, b] \to \mathbb{R}$ by

$$F(x) = \int_a^x f = \int_a^x f(t)dt$$

F is uniformly continuous on $[a, b]$, and if f is continuous at $c \in (a, b)$ then $F'(c) = f(c)$.

Hint: For uniform continuity show that if $a \le x \le y \le b$ then

$$F(y) - F(x) = \int_x^y f.$$

Now estimate $\int_x^y f$ from above and below in terms of $y-x$. For $F'(c) = f(c)$ show that

$$\frac{F(x) - F(c)}{x - c} - f(c) = \frac{1}{x - c} \int_c^x [f - f(c)] \, .$$

Use the fact that f is continuous at c to conclude that the right hand side can be made arbitrarily small by taking x sufficiently close to c.

The second part of the the Fundamental Theorem is the one that it often more convenient to use.

6.21. Prove the second Fundamental Theorem of Calculus:

If f is a function which is continuous on $[a, b]$ and differentiable on (a, b), then $\int_a^b f'(x)dx = f(b) - f(a)$.

Hint: If $P = (x_0, x_1, \ldots, x_n)$ is any partition of $[a, b]$ then, by Problem 5.14, for each $1 \le i \le n$ we can find $t_i \in (x_{i-1}, x_i)$ so that

$$f(b) - f(a) = \sum_{i=1}^{n} \big(f(x_i) - f(x_{i-1})\big) = \sum_{i=1}^{n} f'(t_i)\Delta_i \, .$$

Show $L(f', P) \le f(b) - f(a) \le U(f', P)$.

6.22. Suppose that $k \in \mathbb{N}$. Compute

$$\lim_{n \to \infty} \frac{1}{n^{k+1}} \sum_{i=1}^{n} i^k.$$

5. Integration Rules

Once we have the fundamental theorems of calculus we can prove two of the important rules of integration; integration by parts, and integration by substitution (or u-substitution). These turn out to be reinterpretations of the product rule and of the chain rule, respectively.

6.23. Prove that if f and g are functions that are differentiable on $[a, b]$ and both f' and g' are integrable on $[a, b]$ then

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

6.24. Suppose that u is differentiable on $[c, d]$ and u' is continuous on $[c, d]$. Let $a = u(c)$ and $b = u(d)$. Suppose that f is continuous on $u\big([c, d]\big)$. Prove that

$$\int_a^b f = \int_c^d (f \circ u) \cdot u'.$$

Hint: Let $F(x) = \int_a^x f$. Use Problem 6.20 to show that F is differentiable and use Problem 5.9 to show that $F \circ u$ is differentiable. Now apply Problem 6.21.

Notice we don't need to assume that $u\big([c, d]\big) = [a, b]$.

APPENDIX A

Prerequisite Knowledge

Given that we are making every effort to make our course rigorous, the truth of all of our proofs will ultimately rely on the axioms of mathematics (which are typically formulated in terms of the theory of sets). In this appendix, we will review all of our basic assumptions. They are of two varieties: the general theory of sets and the assumptions we will place specifically on the set of real numbers and on the set of natural numbers. The latter can be proved from the former, but doing so would take us outside of the scope of the text.

1. Set Theory

We will not try and define what we mean by a **set**. Surprisingly this point is quite complicated and philosphical and many mathematics have devoted their lives to these considerations. **Set theory** is the study of the basic definitions and properties surrounding the notion of a set.

Definition. You are probably already familiar with the most basic concepts of set theory:

(1) $\varnothing$ is the **empty** set, i.e. the set with no elements.
(2) $A \subseteq B$ means that every element of A is an element of B, or for all $x \in A$, $x \in B$. It is read "A is a **subset** of B."
(3) $A = B$ means that A and B have the same elements. Another way of saying this is $x \in A$ if and only if $x \in B$.
(4) $A \cap B = \{x : x \in A \text{ and } x \in B\}$. $A \cap B$ is read as "A intersect B" and is called the **intersection** of A and B.
(5) $A \cup B = \{x : x \in A \text{ or } x \in B\}$. $A \cup B$ is read as "A union B" and is called the **union** of A and B. If $x \in A \cup B$ then $x \in A$ or $x \in B$. It could be in both.
(6) A and B are called disjoint sets if $A \cap B = \varnothing$. A and B are disjoint if they have no elements in common.

Logically speaking, we are taken for granted the fact that given two sets, there is another set that is their union (and likewise with intersections and so forth). In the mundane cases with which we will be primarily concerned, this fact is of course trivial, but again one must be quite careful when formulating these notions in general. It turns out that the operations of union and intersection satisfy "distributive laws" (similar to those that hold for numbers: see the next section).

A.1. Suppose that A, B, and C are sets. Then the following identities hold:

(1) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
(2) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Definition. (1) $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$. $A \setminus B$ is read as "A minus B" and is called the "set theoretic difference of A and B."
(2) $A \triangle B = (A \setminus B) \cup (B \setminus A)$ and is called the **symmetric difference** of A and B.
(3) $A^c = \{x : x \notin A\}$ is called the **complement** of A.

Caution: Actually, A^c is not really a set. When we use A^c we must have a ambient set U in mind. This set is often unspecified and is simply inferred from the context. To be explicit, we can write $A^c = U \setminus A$ which is unambiguous.

A.2. Let A and B be sets. Then the following are true:

(1) $(A^c)^c = A$.
(2) $(A \cap B)^c = A^c \cup B^c$.
(3) $(A \cup B)^c = A^c \cap B^c$.
(4) $A \triangle B = (A \cup B) \setminus (A \cap B)$.

Statements (2) and (3) are called **De Morgan's Laws**.

Definition. If A and B are sets then the **Cartesian product** of A and B is defined to be

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

where (a, b) denotes the ordered pair.

$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is the Cartesian plane.

Definition. We can define unions and intersections of large collections of sets. If I is a set, called the **index set**, and for all $i \in I$, A_i is a set then we define

$$\bigcup_{i\in I} A_i = \{x : \text{ there exists } i \in I \text{ such that } x \in A_i\}$$
$$\bigcap_{i\in I} A_i = \{x : \text{ for all } i \in I,\, x \in A_i\}.$$

A.3. If I is a set and for all $i \in I$, A_i is a set, then

(1) $(\bigcup_{i\in I} A_i)^c = \bigcap_{i\in I} A_i^c$.
(2) $(\bigcap_{i\in I} A_i)^c = \bigcup_{i\in I} A_i^c$.

These are the De Morgan's laws for unions and intersections over arbitrary index sets.

2. The Field Properties of the Real Numbers

In this and the following two section, we formulate precisely the properties we are assuming regarding the set of real numbers with the exception of the completeness axiom which is discussed in Chapter 3. The completeness axiom is a major object of study in this course.

Mathematicians study many different types of mathematical objects. You may have heard of groups, rings, topological spaces, smooth manifolds, vector spaces, Banach spaces, affine varieties, elliptic curves, etc. One of the objects which mathematicians study is called a *field*. In the introduction to the chapter, we mentioned several algebraic properties of $\mathbb{R}$. The crucial algebraic properties of $\mathbb{R}$ can be summarized by saying that $\mathbb{R}$ is a field. Notice that all the field properties (listed below) would certainly be demanded of any number system.

As we mentioned above, we will take all the properties on faith. Hence will call them axioms (in mathematics an axiom is a basic statement which is accepted without proof, for example the statement which intuitively says "there exists a set" is a basic axiom of mathematics).

Axiom 1: There exists a set $\mathbb{R}$, which contains $\mathbb{Q}$. We may define two operations on $\mathbb{R}$ called addition and multiplication, which extend normal addition and multiplication of rational numbers.

When we say that addition on $\mathbb{R}$ extends addition on $\mathbb{Q}$, we mean that adding two real numbers which happen to be rational would be the same as the normal addition of rational numbers (and likewise for multiplication).

We will use all the standard notations regarding operations among numbers. For example $a+b$ is the sum of $a, b \in \mathbb{R}$. As always, we write the symbol '$=$' between two real numbers which are the same and the symbol '$\neq$' between two which are not.

AXIOM 2: Addition of real numbers is commutative: For every $a, b \in \mathbb{R}$, $a + b = b + a$.

AXIOM 3: Addition of real numbers is associative: For every $a, b, c \in \mathbb{R}$, $a + (b + c) = (a + b) + c$.

AXIOM 4: The real number zero is an additive identity: For each $a \in \mathbb{R}$, $0 + a = a$.

AXIOM 5: Every real number has an additive inverse: For every $a \in \mathbb{R}$, there is a number $b \in \mathbb{R}$ such that $a + b = 0$.

We mentioned above that there are many obvious facts about the real numbers that, strictly speaking, must be proven from the axioms. The following is an example (as are most of the exercises in this section).

A.4. For every $a \in \mathbb{R}$, the additive inverse of a is unique. That is, if b and c are real numbers which satisfy $a + b = 0$ and $a + c = 0$, we may conclude that $b = c$.

The previous problem justifies us saying *the* additive inverse of $a \in \mathbb{R}$ (rather then *an* additive inverse). As usual, we will use the symbol $-a$ for the additive inverse of a. Notice that, strictly speaking, $-a$ is not the same symbol as $(-1) \cdot a$ (that is the number negative 1 times the number a). That the two symbols represent the same number will be one the obvious facts we prove below.

We can also now define subtraction: If a and b are natural numbers, then $a - b$ is defined to be $a + (-b)$ (in words $a - b$ is the the sum of a and the additive inverse of b).

AXIOM 6: Multiplication of real numbers is commutative: For every $a, b \in \mathbb{R}$, $ab = ba$.

AXIOM 7: Multiplication of real numbers is associative: For every $a, b, c \in \mathbb{R}$, $a(bc) = (ab)c$.

AXIOM 8: The number one is a multiplicative identity: For every $a \in \mathbb{R}$, $a \cdot 1 = a$.

AXIOM 9: Every real number besides zero has a multiplicative inverse: For every $a \in \mathbb{R}$ with $a \neq 0$, there is a number b such that $ab = 1$.

We have a result about multiplicative inverses analogous to the one we had for additive inverse.

A.5. For every $a \in \mathbb{R}$ (other than zero), the multiplicative inverse of a is unique. That is, if b and c are real numbers which satisfy $ab = 1$ and $ac = 1$, we may conclude that $b = c$.

Again we are now justified in referring to *the* multiplicative inverse of a, which we will denote by a^{-1}. We define division in a similar manner as subtraction: a/b is defined to be ab^{-1} (that is, a/b is defined to be the product of a and the multiplicative inverse of b).

AXIOM 10: Multiplication and addition satisfy the distributive property: For every $a, b, c \in \mathbb{R}$, $a(b + c) = ab + ac$.

These algebraic properties of the real numbers are very important, but they are not unique to $\mathbb{R}$. $\mathbb{Q}$ would satisfy all of these axioms and so is also a field. In general, there exist many different fields. The collection of complex numbers, $\mathbb{C}$, (with usual notions of addition and subtraction) is a field. For a prime number p, you may be familiar with the collection of numbers modulo p, often denoted $\mathbb{Z}_p$. It too is a field (and in contrast to $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ has finitely many elements). We will thus need more properties of $\mathbb{R}$ to describe it uniquely.

The following (relatively simple) question might help you to better understand the axioms:

A.6. Which axioms would still be satisfied if $\mathbb{R}$ were replaced with $\mathbb{Q}$? with $\mathbb{Z}$? with $\mathbb{N}$?

We will now give some more basic properties about the real numbers which follow from these axioms.

For our first result, we will see that the multiplication operation on $\mathbb{R}$ still boils down to repeated addition (as long as one of the numbers is a natural number).

A.7. Multiplication of real numbers by natural numbers is just repeated addition. That is, if $a \in \mathbb{R}$ and $n \in \mathbb{N}$, na is the same as the number which results when a is added to itself n times.

Hint: Use induction.

As promised, we will show that the additive inverse of a real number is just that number, multiplied by -1.

A.8. For every $a \in \mathbb{R}$, the product of a and -1 is the additive inverse of a. That is, $-a = (-1)a$.

We can also define integral powers of real numbers (that is, raising a real number to an integral) in the usual way.

Definition. Let $a \in \mathbb{R}$. If n is a natural number, we define a^n to be the product of a with itself n times. Likewise a^{-n} is defined to be the product of a^{-1} with itself n times. We also define a^0 to be 1.

We have the usual basic properties of powers:

A.9. If $a, b \in \mathbb{R}$ and $m, n \in \mathbb{Z}$, then

(1) $(a^m)^n = a^{mn} = (a^n)^m$,
(2) $a^{m+n} = a^m a^n$, and
(3) $(ab)^m = a^m b^m$.

Hint: These properties are by no means automatic. They must be proven, by careful reasoning, from the axioms.

3. The Order Properties of the Real Numbers

In the previous section we saw the algebraic (or field) properties of $\mathbb{R}$. In this one we will study the order properties. As mentioned in the introduction, a set is ordered if we have a rule which tells us, given two elements of the set, which is bigger.

Axiom 11: The real numbers come equipped with an order which extends the order on $\mathbb{Q}$.

Again by 'extends,' we mean that if a and b are rational numbers, then a is less than b according to the order on $\mathbb{Q}$ if and only if a is less than b according to the order on $\mathbb{R}$.

As usual, we denote the order by $\leq$.

Axiom 12: The order is reflexive: For every $a \in \mathbb{R}$, $a \leq a$.

Axiom 13: The order is transitive: For every $a, b, c \in \mathbb{R}$ such that $a \leq b$ and $b \leq c$, we have $a \leq c$.

Axiom 14: The order is antisymmetric: For every $a, b \in \mathbb{R}$ such that $a \leq b$ and $b \leq a$, we have $a = b$.

Axiom 15: The order is a total order: For every $a, b \in \mathbb{R}$, either $a \leq b$ or $b \leq a$.

$\mathbb{R}$ is by no means the only set that comes with an order. In fact, an order can be defined on any set (and many sets, like for example the set consisting of all the months in the year, have an obvious order). Actually, there are many different ways to define an order on $\mathbb{R}$, but there is only one order that will satisfy all the axioms we will list (and have listed).

We will also use the symbols $<$, $>$, and $\geq$ with their usual meanings (i.e., $a < b$ means $a \leq b$ and $a \neq b$). To make our words precise, we will pronounce $a \leq b$ as "a is less than b" and $a < b$ as "a is strictly less than b" (with similar phrasing for $\geq$ and $>$). Note then that "a is less than b" includes the possibility that $a = b$. This is only a convention, but it is one used by many mathematicians.

Again we have many basic and obvious properties.

A.10. $\mathbb{R}$ satisfies the trichotomy property: if $a, b \in \mathbb{R}$, then exactly one of the following holds:

(1) $a < b$,
(2) $a > b$, or
(3) $a = b$.

As expected, a number which is strictly greater than zero is called **positive**, whereas a number which is either positive or zero (in other words a number that is greater then zero) is called **nonnegative**. We use the terms **negative** and **nonpositive** similarly (though nonpositive is typically used with less frequency).

4. The Ordered Field Properties of the Real Numbers

In this section, we will discuss how the algebraic (field) properties of $\mathbb{R}$ interact with the order properties (again in ways that, if you think about them, should work in any system of numbers).

Axiom 16: The order is preserved under addition by a fixed number: If $a, b, c \in \mathbb{R}$ and $a \leq b$ then $a + c \leq b + c$.

Axiom 17: The product of two nonnegative numbers is again nonnegative: If $a, b \in \mathbb{R}$, $0 \leq a$, and $0 \leq b$ then $0 \leq ab$.

To say that $\mathbb{R}$ satisfies these additional properties is to say that it is an **ordered field**. Notice that being an ordered field is much more restrictive than being a field and having an order. The field properties and the order properties must also interact in the right way (as described by the previous two axioms). For example, although there many orders on the set of complex numbers, $\mathbb{C}$, there is no order which makes it into an ordered field (this is not too difficult to prove and we will do so below). Demanding that our numbers form an ordered field tells us that we cannot include imaginary numbers (or complex numbers) in our number system. It also turns out that there is no order on $\mathbb{Z}_p$ which makes it into an ordered field.

Nevertheless, $\mathbb{R}$ is not the only ordered field. $\mathbb{Q}$ is an ordered field and there are many others. We will need one additional property, called the completeness axiom, to uniquely define $\mathbb{R}$. As we mentioned in the introduction to Chapter 2, the completeness axiom is significantly deeper than the others and we will need to develop several new concepts in the Chapter 3 before we can describe it.

Again, we have many basic properties that follow from the axioms. As always, be careful not to use any facts other than the axioms (and other facts we have proven).

The next result shows that we may multiply inequalities by -1 as long as we are willing to reverse the sign.

A.11. Let $a, b \in \mathbb{R}$. If $a \leq b$ then $-b \leq -a$.

More generally, we may multiply an inequality by a real number, but, as expected, we must reverse the sign if the number is negative.

A.12. Suppose $a, b \in \mathbb{R}$ and $a \leq b$. If $c \in \mathbb{R}$ is nonnegative, then $ac \leq bc$. If c is nonpositive then $bc \leq ac$.

Of course the same result holds for strict inequalities unless $c = 0$ (by a similar proof).

A.13. Given any number $a \in \mathbb{R}$, there is a number which is strictly larger.

Hint: Finding a number is not difficult, but prove rigorously that it is larger.

A.14. If $a \in \mathbb{R}$, $a^2 \geq 0$ with $a^2 = 0$ if and only if $a = 0$.

A.15. Suppose we have a field F (that is, F satisfies all the properties which we gave for $\mathbb{R}$ in the section on the field properties). In addition suppose there is an element $i \in F$ which satisfies $i^2 = -1$. Then there is no order on F which makes it into an order field (that is, no order which will satisfy all the properties given for $\mathbb{R}$ in this chapter).

Hint: Suppose F does indeed satisfy all the properties we have given so far for $\mathbb{R}$. How does i compare to zero?

We will not have occasion to use the next result, but they are of general interest and they provide insights into ordered fields.

A.16. There is no order on $\mathbb{C}$ which makes it an ordered field.

This last result is important because $\mathbb{C}$ does satisfy the completeness axiom (or at least an appropriate formulation of it). Thus there are fields other then $\mathbb{R}$ which satisfy the completeness axiom, but no other ordered fields which satisfy it. Notice that $\mathbb{Q}$ is an ordered field which is not complete and $\mathbb{C}$ is a completed field which is not ordered.

All of the assumed properties of $\mathbb{R}$ are now in place except completeness. The remaining statements must therefore be proven from our axioms.

5. Mathematical Induction

We mentioned in the introduction to Chapter 2 that we will assuming all the basic properties of $\mathbb{N}$. This includes the following fact.

Axiom 18: $\mathbb{N}$ is **well-ordered** under the usual ordering. In other words, every nonempty set of $\mathbb{N}$ contains an element which is smaller than the others.

The $\mathbb{N}$ is well-ordered axiom be proven from the more basic axioms of mathematics, but we will not attempt to carry out this proof. We use the previous axiom to prove the validity of an extremely important technique of proof. Hopefully you are already familiar with this method.

THE THEOREM OF MATHEMATICAL INDUCTION: Let $P(1), P(2), P(3), \ldots$ be a list of statements, each of which is either true or false. Suppose that

i) $P(1)$ is true
ii) For all $n \in \mathbb{N}$, if $P(n)$ is true then $P(n+1)$ is true.

Then for all $n \in \mathbb{N}$, $P(n)$ is true.

A.17. Prove the above theorem.

Hint: Suppose it were not true. Use the fact that $\mathbb{N}$ is well-ordered to find an n_0 which is the smallest element of $\mathbb{N}$ so that $P(n_0)$ is false.

A.18. Use mathematical induction to establish the following. Make sure in your proof to precisely state what you are taking "$P(n)$" to be.

a) For all $n \in \mathbb{N}$, $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$
b) For all $n \in \mathbb{N}$, $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$
c) For all $n \in \mathbb{N}$, if $n \geq 4$ then $2^n < n!$.

Note: $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot (n-1) \cdot n$. This is called "$n$ factorial."

APPENDIX B

Appendix to Chapter 3

The study of mathematics goes well beyond a knowledge of the real numbers. In this chapter, we discuss some topics which are tangentially related to the material of this course, but which can take a willing student a bit further into more advanced mathematics.

1. Cardinality

Intuitively, cardinality is the major of the 'size' of a set. For example, we should predict that a set containing two elements (regardless of what they are) should be 'larger' in some essential way than a set containing only one. In fact, mathematicians were surprised to discover that an infinite set can actually be essentially larger than another, a fact we will prove.

Definition. If A and B are sets, we write $|A| = |B|$ if there exists a bijection $f : A \to B$. We read $|A| = |B|$ as "the cardinality of A equals the cardinality of B.
Note. By Theorem 2.7 in Chapter 2, $|A| = |B| \Rightarrow |B| = |A|$.

Again $|A| = |B|$ intuitively means that both sets have the "same number of elements". This is not startling for finite sets. It is no surprise that $|\{a, b, c\}| = |\{1, 2, 3\}|$. However this definition can lead to non-intuitive results. We can have $A \subseteq B$, $A \neq B$ yet $|A| = |B|$ (how?).

Definition. A is *finite* if $A = \emptyset$ or if there exists $n \in \mathbb{N}$ with $|A| = |\{1, 2, \ldots, n\}|$. (We then say $|A| = 0$ or $|A| = n$ accordingly.) A is *infinite* if A is not finite. A is *countably infinite* if $|A| = |\mathbb{N}|$. A is *countable* if A is finite or countably infinite.

Are all infinite sets also countably infinite?

B.1. Prove that a set A is

(1) countably infinite if and only if we can write $A = \{a_1, a_2, \dots\}$ where $a_i \neq a_j$ if $i \neq j$.
(2) countably infinite if and only if A is infinite and we can write $A = \{a_1, a_2, \dots\}$.
(3) countable if and only if $A = \emptyset$ or we can write $A = \{a_1, a_2, \dots\}$.

Deduce that if $B \subseteq A$ and A is countable, then B is countable.

B.2. Let $|A| = |B|$ and $|B| = |C|$. Prove that $|A| = |C|$.

B.3. Prove that

(1) $|\mathbb{N}| = |\{2, 4, 6, 8, \dots\}|$
(2) $|\mathbb{N}| = |\mathbb{Z}|$
(3) $|\mathbb{N}| = |\{x \in \mathbb{Q} : x > 0\}|$
(4) $|\mathbb{N}| = |\mathbb{Q}|$.

Hint: For the third part, assuming $a_n > 0$ turn

$$\frac{1}{2}\left(a_n + \frac{2}{a_n}\right) > \sqrt{2}$$

into an equivalent condition on a quadratic polynomial. Proceed by induction.

B.4. If A is countable and B is countable prove that $A \times B$ is countable.

Hint: You want to construct a list of all elements in $A \times B$ (see 1.17). Can you make an infinite matrix of these elements starting with

$$\begin{array}{c|cccc} & a_1 & a_2 & a_3 & \dots \\ \hline b_1 & & & & \\ b_2 & & & & \\ b_3 & & & & \\ \vdots & & & & \end{array}$$

Can you take this matrix and make a list as in 1.17c)?

Our next problem is due to G. Cantor. It is a famous result which shook the mathematical world and has found its way into numerous "popular" math/science books. Cantor went insane. The problem's solution relies on the decimal representation of a real number. In turn this actually involves the notion of convergence of a sequence of reals which we address in chapter 3. But you can use it here. ($1/3 = .333...$ means that $1/3 = \lim_{n\to\infty} x_n$ where $x_n = .33...3$ (n entries). Beware of this fact: Some numbers have 2 decimal representations, e.g., $1 = 1.000... = .999...$. This can only happen to numbers which can be represented as decimals with 9 repeating forever from some point.)

B.5. In this problem, we consider the cardinality of $\mathbb{R}$.

(1) Prove that $(0,1)$ is not countable.
(2) Show that $|(0,1]| = |[0,1]|$.
(3) If $a < b$ show that $|(0,1)| = |(a,b)| = |[0,1]| = |[a,b]|$.

Hint: For the first part, assume that $(0,1)$ is countable. Then we can list $(0,1) = \{a_1, a_2, a_3, \ldots\}$. Write each a_i as a decimal to get an infinite matrix as the following example illustrates.

$$\begin{aligned} a_1 &= 0.13974\cdots \\ a_2 &= 0.000002\cdots \\ a_3 &= 0.55556\cdots \\ a_4 &= 0.345587\cdots \\ a_5 &= 0.9871236\cdots \\ &\vdots \end{aligned}$$

Can you find a decimal in $(0,1)$ that is not on this list? Can you describe an algorithm for producing such a number?

We prove in the main body of the text that irrationals exist, but here we can prove much more.

B.6. Prove the following.

(1) If A and B are countable then $A \cup B$ is countable.
(2) $\mathbb{R} \setminus \mathbb{Q}$ is *uncountable* (i.e., not countable).
(3) If $a < b$ then $(a,b) \cap (\mathbb{R} \setminus \mathbb{Q}) \neq \emptyset$.
(4) Prove that if I is countable and for all $i \in I$, A_i is a countable set then $\bigcup_{i \in I} A_i$ is countable.

So irrationals do exist. Does this proof give you any explicit number in $\mathbb{R} \setminus \mathbb{Q}$?

We have not defined $|A| \leq |B|$ yet.

B.7. Give a definition for $|A| \leq |B|$. Your definition should satisfy

(1) $|A| \leq |A|$
(2) $|A| \leq |B|$ and $|B| \leq |C|$ implies that $|A| \leq |C|$.

B.8. Suppose A and B are sets with $|A| \leq |B|$ and $|B| \leq |A|$. Then $|A| = |B|$.

Hint: This seemingly trivial statement is actually quite challenging to prove.

Definition. $|A| < |B|$ if $|A| \leq |B|$ and $|A| \neq |B|$.

Definition. If A is a set, the **power set of** A, denoted $\mathcal{P}(A)$, is the set of all subsets of A.

Thus, for example,

$$\mathcal{P}(\{1,2\}) = \{\phi, \{1\}, \{2\}, \{3\}\}.$$

B.9. Prove that for all sets A, $|A| < |\mathcal{P}(A)|$.

Hint: Show there does not exist a function $f : A \to \mathcal{P}(A)$ which is onto by assuming such an f exists and considering $B \in \mathcal{P}(A)$ where $B = \{a \in A : a \notin f(a)\}$.

The previous problem demonstrates that there is not largest cardinal.

2. Open and Closed Sets

We have remarked already that the material of the previous section belongs to the branch of mathematics known as set theory. The material of this section, belongs to a branch known as topology. Roughly speaking, topology might be considered the study of what it means for two elements of a set to be "close to each other." We will restrict ourselves exclusively to the topological properties of $\mathbb{R}$, but topology is a very rich subject whose objects goes well beyond merely the real numbers.

Definition. Let $\epsilon > 0$. The interval $(a - \epsilon, a + \epsilon)$ is said to be an *open interval centered at a of radius ϵ.*

B.10. Let $a < b$. Show that (a, b) is an open interval of radius $\epsilon$for some $\epsilon > 0$. What is the center? What is ϵ?

Definition. Let $S \subseteq \mathbb{R}$.

(1) S is *open* if for all $a \in S$ there exists $\epsilon > 0$ with $(a-\epsilon, a+\epsilon) \subseteq S$
(2) S is *closed* if $C(S) = \mathbb{R} \setminus S$ is open.

Is every $S \subseteq \mathbb{R}$ either open or closed? Can you justify your answer?

B.11. Prove that every open interval is an open set and every closed interval is a closed set.

B.12. Classify as open, closed, both or neither

(1) $\emptyset$
(2) $[0, 1]$
(3) $\mathbb{Q}$
(4) $\mathbb{R} \setminus \mathbb{Q}$
(5) $\mathbb{R}$
(6) $[0, 1] \cup [2, 3]$
(7) $\{\frac{1}{n} : n \in \mathbb{N}\}$

Definition. Let $S \subset \mathbb{R}$.

(1) $x \in \operatorname{int}(S)$ if there exists $\epsilon > 0$ with $(x - \epsilon, x + \epsilon) \subseteq S$.
(2) $x \in \operatorname{bd}(S)$ if for all $\epsilon > 0$, $(x-\epsilon, x+\epsilon) \cap S \neq \emptyset$ and $(x-\epsilon, x+\epsilon) \cap C(S) \neq \emptyset$.

We point out that "int" is short for *interior* and "bd" is short for *boundary.*

B.13. For each S find $\operatorname{int}(S)$ and $\operatorname{bd}(S)$

(1) $[0, 1]$
(2) $(0, 1)$
(3) $\mathbb{Q}$
(4) $\mathbb{R}$
(5) $\{1, 2, 3\}$
(6) $\{\frac{1}{n} : n \in \mathbb{N}\}$

B.14. Suppose that $S \subseteq \mathbb{R}$. Prove the following

(1) $\operatorname{int}(S) \subseteq S$ and $\operatorname{int}(S)$ is an open set.
(2) S is open $\Leftrightarrow S = \operatorname{int}(S)$.
(3) S is open $\Leftrightarrow S \cap \operatorname{bd}(S) = \emptyset$.
(4) S is closed $\Leftrightarrow S \supseteq \operatorname{bd}(S)$.

B.15. Prove the following

(1) If I is a set and for all $i \in I$, A_i is an open set, then $\bigcup_{i\in I} A_i$ is open.
(2) If I is any set and for all $i \in I$, F_i is a closed set then $\bigcap_{i\in I} F_i$ is closed.
(3) If $n \in \mathbb{N}$ and A_i is an open set for each $i \leq n$ then $\bigcap_{i=1}^{n} A_i$ is open.
(4) If $n \in \mathbb{N}$ and A_i is a closed set for each $i \leq n$ then $\bigcup_{i=1}^{n} A_i$ is closed.

B.16. Show by example that the last two parts of the previous problem cannot be extended to infinite intersections or unions.

Definition. Let $S \subseteq \mathbb{R}$, $x \in \mathbb{R}$.

(1) x is an *accumulation point* of S if for all $\epsilon > 0$, $\{y \in \mathbb{R} : 0 < |x - y| < \epsilon\} \cap S \neq \emptyset$.
(2) $S' = \{x : x \text{ is an accumulation point of } S\}$.
(3) x is an *isolated point* of S if $x \in S \setminus S'$.
(4) $\bar{S} = S \cup S'$.

$\bar{S}$ is called the *closure* of S.

B.17. Let $S \subseteq \mathbb{R}$. Prove the following.

(1) $x \in S$ is an isolated point of S if and only if there exists $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \cap S = \{x\}$. Let $S \subseteq \mathbb{R}$.
(2) Let $x \in \mathbb{R}$. Prove that $x \in S'$ if and only if for all $\epsilon > 0$, $(x - \epsilon, x + \epsilon) \cap S$ is infinite.

B.18. For each set S below find S', $\bar{S}$ and all isolated points of S.

(1) $\mathbb{R}$
(2) $\emptyset$
(3) $\mathbb{Q}$
(4) $(0, 1]$
(5) $\mathbb{Q} \cap (0, 1)$
(6) $(\mathbb{R} \setminus \mathbb{Q}) \cap (0, 1)$
(7) $\{\frac{1}{n} : n \in \mathbb{N}\}$

B.19. Prove the following. $S \subseteq \mathbb{R}$.

(1) S is closed if and only if $S \supseteq S'$.
(2) $\bar{S}$ is closed.

(3) S is closed if and only if $S = \bar{S}$.
(4) If $F \supseteq S$ and F is closed then $F \supseteq \bar{S}$.

3. Compactness

Our next topic in topology is compactness. The definition is quite abstract and will take effort to absorb. We will later prove that a continuous function on a compact domain achieves both a maximum and a minimum value — quite a useful thing in applications.

Definition. Let $S \subseteq \mathbb{R}$.

(1) Let $\{A_i\}_{i\in I}$ be a family of open sets. $\{A_i\}_{i\in I}$ is an *open cover* for S if $S \subseteq \bigcup_{i\in I} A_i$.
(2) Let $\{A_i\}_{i\in I}$ be an open cover for S. A *subcover* of this open cover is any collection $\{A_i\}_{i\in I_0}$ where $I_0 \subseteq I$ such that $\bigcup_{i\in I_0} A_i \supseteq S$.
(3) S is *compact* if every open cover of S admits a finite subcover, i.e., whenever $\{A_i\}_{i\in I}$ is a family of open sets such that $S \subseteq \bigcup_{i\in I} A_i$ then there exists a finite set $F \subseteq I$ so that $S \subseteq \bigcup_{i\in F} A_i$.

For example, $\{(n-1, n+1) : n \in \mathbb{Z}\}$ is an open cover of $\mathbb{R}$. For all $x \in \mathbb{Q}$, let ϵ_x be a positive number (which might be different for different x. Is $\{(x-\epsilon_x, x+\epsilon_x) : x \in \mathbb{Q}\}$ necessarily an open cover of $\mathbb{R}$?

This definition above is very abstract and may require study and time to absorb. Note that the definition requires that *every* open cover of S admits a finite subcover. To show S is not compact you only need construct *one* open cover without a finite subcover. Compactness plays a key role in analysis (and topology).

B.20. Which of the following sets are compact?

(1) $\{1, 2, 3\}$
(2) $\emptyset$
(3) $(0, 1)$
(4) $[0, 1)$
(5) $\mathbb{R}$

B.21. Let $S \subseteq \mathbb{R}$ be compact. Prove that

(1) S is bounded.
(2) S is closed.

Hint: Assume not in each case and produce an open cover without a finite subcover.

B.22. Prove that $[0,1]$ is compact.

Hint: Let $\{A_i\}_{i\in I}$ be any open cover of $[0,1]$. Let

$B = \{x \in [0,1] : [0,x]$ can be covered by a finite subcover of $\{A_i\}_{i\in I}\}$.

Then $0 \in B$ so $B \neq \emptyset$. Let $x = \sup(B)$. Show $x \in B$. Show $x = 1$.

B.23. Let $K \subseteq \mathbb{R}$ be compact and let $F \subseteq K$ be closed. Prove that F is compact.

Hint: If $\{A_i\}_{i\in I}$ covers F then $\{A_i\}_{i\in I} \cup \{C(F)\}$ covers K.

B.24. Let $K \subseteq \mathbb{R}$ be closed and bounded.

(1) Prove $\min(K)$ and $\max(K)$ both exist if $K \neq \emptyset$.
(2) Prove that K is compact.

We see that $K \subseteq \mathbb{R}$ is compact $\Leftrightarrow$ K is closed and bounded.

B.25. Let $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$ be a nested sequence of closed, bounded and nonempty sets in $\mathbb{R}$. Then

$$\bigcap_{n=1}^{\infty} I_n \neq \emptyset.$$

Hint: Assume it is empty. Then

$$\mathbb{R} = C\left(\bigcap_{n=1}^{\infty} I_n\right) = \bigcup_{n=1}^{\infty} C(I_n) \supseteq I_1\ .$$

B.26. Let $K \subseteq \mathbb{R}$ be compact and infinite. Prove that $K' \neq \emptyset$.

Hint: Assume $K' = \emptyset$.

B.27. Let $A \subseteq \mathbb{R}$ be bounded and infinite. Prove that $A' \neq \emptyset$.

4. Sequential Limits and Closed Sets

Definition. Let $A \subseteq \mathbb{R}$. A is *sequentially* closed if whenever $(a_n)_{n=1}^{\infty}$ is a sequence in A converging to a limit a, then $a \in A$.

B.28. If $A \subseteq \mathbb{R}$ is closed then it is sequentially closed.

B.29. If $A \subseteq \mathbb{R}$ is sequentially closed then it is closed.

B.30. If $A \subseteq \mathbb{R}$ then A is closed if and only if it is sequentially closed.