Chapter 1

Sets and Functions

Sets
Mathematicians try very hard to precisely define new concepts using only previously defined concepts. There is, at the beginning of this process, a concept that is not defined with previous concepts and this is the concept of a set. Sets are only defined intuitively as collections of objects. There is a language that allows us to communicate exactly what is in a particular set. What we do in the case of very small sets is write the elements of the set between braces, separating the elements with commas. Thus, we would write

\{1, 5, 6\}

to indicate the set with three elements, the elements being the numbers 1, 5, and 6. Sometimes we can write infinite sets this way if their elements form a pattern. This is done by explicitly indicating the first few elements until the pattern is discernible, then writing three dots to indicate all the rest are included. Using this method

\{1, 2, 3, \ldots\}

is intended to indicate all the natural numbers, so it includes 4, 5, 6, and on forevermore. Using dots at both the left and right let us indicate the integers

\{\ldots, -2, -1, 0, 1, 2, \ldots\},

which include all natural numbers and all of their negatives. You can see that zero is in this set too.

Often the elements of a set can be determined by a rule. When this is the case, we may write

\{elements \mid \text{the rule}\}

to indicate the set. Thus the set of prime numbers may be indicated by writing

\{n \mid n \text{ is a prime number}\}. 
Problem 1.1 Use the language of set notation to indicate the set of even integers bigger than 9. Give two different answers.

The set of rational numbers is

\[ \{ \frac{m}{n} \mid m \text{ and } n \text{ are integers, } n \neq 0 \}. \]

By looking at the definition of the set of rational numbers just given, it is hoped we have communicated that a number is rational exactly when it is a ratio of integers.

The common method of handling sets is to first define the set using braces and, at the same time, introduce a symbol to represent the set. An illustration of this would be to write

\[ \mathbb{Q} = \{ \frac{m}{n} \mid m \text{ and } n \text{ are integers, } n \neq 0 \}. \]

The symbol \( \mathbb{Q} \) now represents the set of rational numbers. We then write \( x \in \mathbb{Q} \) to indicate that \( x \) is a rational number. In general, \( y \in S \) means “\( y \) is an element of the set \( S \),” and \( y \notin S \) means “\( y \) is not an element of the set \( S \).” The sets that we use frequently all get their own symbols.

**Index of Sets**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \emptyset )</td>
<td>the empty set</td>
<td>{ }</td>
</tr>
<tr>
<td>( \mathbb{N} )</td>
<td>natural numbers</td>
<td>{1, 2, 3, \ldots }</td>
</tr>
<tr>
<td>( \mathbb{Z} )</td>
<td>integers</td>
<td>{ \ldots , -2, -1, 0, 1, 2, \ldots }</td>
</tr>
<tr>
<td>( \mathbb{Q} )</td>
<td>rational numbers</td>
<td>{ \frac{m}{n} \mid m, n \in \mathbb{Z} \text{ and } n \neq 0 }</td>
</tr>
<tr>
<td>( \mathbb{R} )</td>
<td>real numbers</td>
<td>rationals and irrationals</td>
</tr>
<tr>
<td>( \mathbb{R}^2 )</td>
<td>Cartesian plane</td>
<td>{ (x, y) \mid x, y \in \mathbb{R} }</td>
</tr>
<tr>
<td>( \mathbb{R}^3 )</td>
<td>Euclidean 3-space</td>
<td>{ (x, y, z) \mid x, y, z \in \mathbb{R} }</td>
</tr>
<tr>
<td>( \mathbb{F}^+ )</td>
<td>positive elements of ( \mathbb{F} )</td>
<td>{ x \mid x \in \mathbb{F} \text{ and } 0 &lt; x }</td>
</tr>
<tr>
<td>( \mathbb{C} )</td>
<td>complex numbers</td>
<td>{ x + iy \mid x, y \in \mathbb{R} }</td>
</tr>
</tbody>
</table>

In a truly formal mathematical development of a subject all of the definitions would be given in terms of sets, so that only a single concept (the set) remains without a formal definition. While this is the spirit of pure mathematics, this is usually not a practical way to become acquainted with a new mathematical idea since the intuition that lies behind the mathematical idea is often lost when a set theoretical model is given to describe that idea. For example, we indicated above that the number 1 is an element of the set \( \mathbb{N} \), or more briefly \( 1 \in \mathbb{N} \).
You probably did not notice that we failed to give a formal definition of what
the number 1 is! Such a definition would be given if we intended to build a
tory that enabled us to prove facts involving the number 1, facts that are
very familiar to us intuitively. This definition is part of what you will find in a
book on set theory, but to understand the set theoretical model of the number
1 it is essential that the intuition of numbers be firmly kept in mind. It is
very unlikely that a set theoretical model of numbers would convey any of the
intuition of numbers to a young student, which is why axiomatic set theory is
so unpopular in kindergartens. Our description of $\mathbb{Q}$ also diverges from a truly
formal development. For on thing, the definition is too vague since we never
mention how we are supposed to think of $4/2$ and $2/1$ as the same element. Our
description of $\mathbb{R}$ is even worse since we refer to irrational numbers without giving
you the slightest idea of what they are. Rest assured that formal set theoretical
definitions of $\mathbb{Q}$ and $\mathbb{R}$ are awaiting you in the mathematical literature (they can
be found in books on abstract algebra and real analysis). In the next chapter you
will see how we will manage to avoid these formal definitions and still adhere
to the spirit of pure mathematics. For the time being you need only have an
intuitive understanding of the sets $\mathbb{Q}$ and $\mathbb{R}$.

A way to obtain a new set from a given set is to extract part of it; we will
write $E \subseteq F$ to indicate that every element of $E$ also belongs to $F$, and we say
that $E$ is a subset of $F$. For example, if we define

$$E = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \},$$

then $E \subseteq \mathbb{R}^2$. Mathematicians picture $\mathbb{R}^2$ as a set of points that constitute an
infinite plane and they picture $E$ as a circle of radius one inside of this plane
(see Figure 1.1).

![Figure 1.1: The points on the circle constitute the subset $E$ of $\mathbb{R}^2$](image)

**Problem 1.2** Use set notation to describe the following subsets of $\mathbb{R}^2$;

1. The set of points on the $x$-axis.
2. The set of points on and above the \( x \)-axis.

3. The set of points on a parabola (any parabola will do).

Given two sets \( A \) and \( B \) you can form a union which is denoted \( A \cup B \) and defined by
\[
A \cup B = \{ x \mid x \in A \text{ or } x \in B \},
\]
and you can also form an intersection which is denoted \( A \cap B \) and defined by
\[
A \cap B = \{ x \mid x \in A \text{ and } x \in B \}.
\]
If we are dealing with a set that has an order defined, like \( \leq \) in \( \mathbb{R} \) and \( \mathbb{Q} \), then we can define interval subsets;
\[
[x, y] = \{ z \mid x \leq z \leq y \}
\]
\[
[x, y) = \{ z \mid x \leq z < y \}
\]
\[
(x, y] = \{ z \mid x < z \leq y \}
\]
\[
(x, y) = \{ z \mid x < z < y \}
\]

It is important to realize that we have just introduced conflicting notation; writing \((1, 2)\) can mean one of two things. It can either mean a point in the Cartesian plane, as drawn in Figure 1.1, or it can mean a subset of \( \mathbb{R} \), those numbers that lie between 1 and 2. Just as certain words can have multiple meanings so can mathematical definitions. You must rely on the context of usage to know which meaning applies.

**Problem 1.3** Is it true that \((5, 6) \notin (5, 6)\)? Explain the meaning of \((5, 6)\) in each usage.

**Functions**

In the spirit of pure mathematics it is possible to define a function as a particular type of set (just as the number 1 can be). You have probably graphed functions before; the definition of a function as a set is obtained by using the graph (which is a set) as the definition of the function. In the text that follows we will frequently define functions by drawing a graph. In function graphing exercises, the student is first given a formula, such as \( f(x) = x^3 \), and then the student is asked to draw the graph, i.e. to draw the set
\[
\{(x, f(x)) \mid x \in \mathbb{R} \}.
\]

The point of view taken here is that the set is the function, not the formula. In fact, it is hoped that you will come to appreciate the fact that most functions do not arise from formulas!

In order to build your intuition of what mathematicians think of when they say “function”, we will start with an intuitive definition, a definition that uses words that have not been previously given a precise mathematical meaning.
Definition 1.1 A function $f$ from a set $S$ to another set $T$ is a rule that assigns to each element of $S$ a unique element of $T$. We will refer to the elements of $S$ as inputs, and the function provides to each input a corresponding unique output which lies in the set $T$.

The way we indicate briefly that we have such a function is to write

$$f : S \rightarrow T,$$

which labels the function as $f$, the set of allowable inputs as $S$, and the set where the outputs reside as $T$. It is standard practice to call $S$ the domain of $f$, but authors do not all agree on a term for $T$. This is a good opportunity to point out that mathematical terminology is frequently not standardized. It is important to realize that the same term might mean different things in two books. The meaning of a mathematical term used in a book must be understood as it is defined in that same book. We will not need a special term for $T$ in this book so we will not give one.

Students frequently carry the misconception that functions are given only by formulas. As we mentioned earlier, the truth is that most functions can not be given by formulas. The ones that can are much easier to understand mathematically, which is why they are dominant in math textbooks. To fully understand the function concept it is important to drop this impression that functions arise only from formulas and begin looking at functions without formulas. This is why we prefer the view that graphs define functions.

The functions we study in this book take numbers as input and return numbers as output. To define such a function graphically, we will draw a picture as in Figure 1.2. The numbers to input into the function are on the horizontal

![Figure 1.2](image_url)

Figure 1.2:

line, and the numbers the function outputs are on the vertical line. To decide what the function outputs when 2 is the input, one looks for the point on the curve with 2 as its $x$-coordinate. The $y$-coordinate of that point is the output, which for the function defined by Figure 1.2, looks to be about 1. This method only works if there is a unique point on the curve with 2 as the $x$-coordinate. The curve in Figure 1.3 does not define a function since it cannot be decided
what the output is when 2 is input. In general, a curve defines a function if every vertical line intersects the curve in at most one point.

Notice that not every number can be input into the function defined by Figure 1.2. A number on the x-axis is in the domain of the function if the vertical line through that number touches the curve in exactly one place. You can see that 6 is not in the domain while 2 and 4 are. If a circle is drawn on the curve, this is our way of indicating that point is not on the curve. Thus (1, 1) is not on the curve. The black dot at the point (1, 2) is meant to indicate that 2 is the output when 1 is input. Now is a good time to abandon the word “curve” since you probably do not consider (1, 2) to be a part of the curve drawn in Figure 1.2, but we do intend (1, 2) to be a point on the graph of the indicated function. Thus anytime we have a subset of the Cartesian plane with the property that every vertical line intersects the subset in at most one point, then that subset defines a function with the points of that subset constituting the graph of that function.

Problem 1.4 Use interval notation, sets, and unions to describe the domain of the function whose graph is pictured in Figure 1.4.

Problem 1.5 Use set notation and some ingenuity to indicate the graph of a
function that is impossible to draw.

There is notation that we use to represent functions, their input, and their output. We use letters like \( f \), \( g \), or \( h \) to represent functions, letters like \( x \), \( y \), and \( z \) to represent inputs, and we write \( f(x) \) to represent the output corresponding to the input \( x \) by the function \( f \). Thus, if we use \( f \) to represent the function in Figure 1.2, then \( f(1) = 2 \) expresses that 2 is the output when 1 is the input (as we discussed in the previous paragraph). We might say that \( f(6) \) is not defined since 6 is not in the domain of \( f \). Similarly, \( f(0) \) is undefined since a circle appears on the curve and there is no black dot above or below 0.

**Problem 1.6** Draw a picture that defines a function \( f \) with \( f(1) = 1 \), \( f(2) = 3 \), \( f(3) \) undefined, and \( f(4) = 1 \).

**Problem 1.7** Draw pictures that define two functions \( f \) and \( g \) with \( f(1) = g(1) \), \( f(2) = 3 \), \( g(f(2)) = 2 \), and such that the domain of both functions is \([0, 4]\).
Chapter 2

Axioms and Definitions

Axiomatics

Mathematicians are in the business of making logical deductions from assumed truths. If you assume that “two sets are equal when they contain exactly the same elements”, then this assumption will let you deduce that \( \{1, 2\} \neq \{1, 3\} \). Mathematicians try to assume as little as possible and deduce as much as possible. You may be surprised to learn that all of calculus may be derived from only seven basic assumptions! No one hesitates to assume the truth of these basic assumptions because our intuition screams that they must be true; the assumptions seem obvious. Does the statement “two sets are equal when they contain exactly the same elements” seem like it has to be true? This is in fact one of the seven basic assumptions out of which calculus flows. The statements that we assume true (in order to form a starting point for logical deductions) are called axioms. My dictionary \(^1\) defines an axiom to be a “proposition regarded as a self-evident truth”, and this is precisely the mathematical meaning of the word axiom when used to describe these seven assumptions.

Mathematicians will also use the word axiom in another context that has a slightly different meaning than the one quoted from the dictionary. Mathematicians often define mathematical objects in terms of sets and a list of logical statements. For example, the real numbers will be defined as a set with an addition, a multiplication, and an order that satisfy 14 logical statements. Mathematicians will also use the word “axiom” to refer to the logical statements in a definition. The two uses of the word have one thing in common; in both instances the axioms serve as a starting point for deduction. However, this second usage of the word applies to statements that are not really self-evident truths. In the case of the real numbers the logical statements are called the axioms for an ordered field, but the existence of such a mathematical object is far from self-evident.

If we had a lot of time on our hands and really wanted to do things right, we would start with the seven basic axioms of set theory referred to in the first

---

\(^1\)Webster’s Seventh New Collegiate Dictionary
paragraph. From only these seven self evident truths we would deduce the existence of the real numbers and prove that this set satisfies all the properties in our list of 14 logical statements. From these 14 logical statements it is then possible to prove all the theorems you encounter in calculus. The problem with this approach is it would take more time than we have. An even worse problem is that the technicalities and abstraction needed to build the theory from scratch would make it nearly impossible to see what is happening intuitively! Just as kindergartners need to first develop the intuition of numbers, learning the fundamental theorems of calculus will also involve building intuition. Our strategy will be to simply lay down the 14 logical statements that characterize the real numbers and deduce away from these statements.

We would like to present an analogy to the process of laying down an assumption and proceeding to build a theory based on that assumption, without questioning that assumption. You have been told that the area of the rectangle in Figure 2.1 is \(xy\). If we allow ourselves to assume the area to be \(xy\), then we will be able to prove some very useful facts. This is like adopting the 14 statements that describe the real numbers as true, then deducing the theorems of calculus from them. However, someday we should come back to this assumption that the area is \(xy\) and convince ourselves that it is indeed a consequence of self evident truths. This is like starting from the seven axioms of set theory and using them to deduce the 14 statements that describe the real numbers. In order to verify that the area of our rectangle is \(xy\) we need to state assumptions that capture our intuition of what we mean by area.

**Assumption 2.1** *The area of the whole is the sum of the areas of parts.*

If we all agree that area should have this property then Assumption 2.1 becomes an axiom. We will assume that area satisfies this property and make deductions based on this assumption. If we mathematize this discussion we will discover that the formula for the area of a rectangle may be deduced from Assumption 2.1 (plus some other assumptions that we will discover along the way). Let us write \(A(x, y)\) to represent the area of the rectangle in Figure 2.1. If we divide the rectangle as in Figure 2.2 we can write an equation that describes
Assumption 2.1:

\[ A(x_1, y) + A(x_2, y) = A(x_1 + x_2, y). \quad (2.1) \]

If we take both \( x_1 \) and \( x_2 \) equal to half of \( x \) we discover that

\[ A\left(\frac{1}{2}x, y\right) = \frac{1}{2}A(x, y). \]

Dividing \( x \) into \( n \) equal parts reveals that

\[ A\left(\frac{1}{n}x, y\right) = \frac{1}{n}A(x, y). \]

What we just written is very typical of mathematical exposition; we have recorded Assumption 2.1 in Equation 2.1 and then we deduced the two equations above from Equation 2.1. An experienced reader of mathematics always has a pad and pencil nearby to help them follow such deductions. It is a very bad idea to simply read along a chain of deductions without actually performing the deductions yourself. You should strive to perform the deductions yourself, using what is written in the text as a guide.

**Problem 2.1** Prove that \( mA(x, y) = A(mx, y) \) for any natural number \( m \).

Now we see that

\[ A\left(\frac{m}{n}x, y\right) = \frac{m}{n}A(x, y) \]

for every positive rational number \( m/n \). A particular case of this can be written as

\[ A(x, y) = xA(1, y) \]

for all \( x \in \mathbb{Q}^+ \) (see page 2). If we repeat the reasoning above we will soon convince ourselves that

\[ A(x, y) = yA(x, 1) \]

for all \( y \in \mathbb{Q}^+ \). Combining these two results leads us to

\[ A(x, y) = xyA(1, 1) \quad (2.2) \]
for all \( x, y \in \mathbb{Q}^+ \). Recall that we wanted 
\[
A(x, y) = xy,
\]
but this only happens if \( A(1, 1) = 1 \) in Equation 2.2. There is no hope of proving \( A(1, 1) = 1 \) based only on Assumption 2.1, as the next problem is intended to convince you.

**Problem 2.2** Forget that anyone ever told you what the area of a rectangle is. Assume you walk into school and your teacher announces that the area of the rectangle in Figure 2.1 is \( 9xy \) (not \( xy \))! Divide the rectangle into two rectangles and prove that the area of the whole is the sum of the areas of the parts, using this new definition of area.

The point of the previous problem is that there are many (infinitely many in fact) definitions for an area of a rectangle that all satisfy Assumption 2.1. The definition of area that the teacher produced in Problem 2.2 is not really so strange. If \( x \) and \( y \) are measured in yards then the area of the rectangle is \( 9xy \) square feet, so the fictional teacher was right after all. The following assumption is very much like a choice of units (like square yards).

![Figure 2.3: If the small squares are assigned an area of 9, then the big rectangle has area 216, which equals 9xy.](image)

**Assumption 2.2** \( A(1,1) = 1 \)

Our two assumptions now let us deduce that the area of our rectangle is \( xy \), as long as \( x \) and \( y \) are positive rationals. What happens if one of the sides is an irrational number? Once again we need another assumption to proceed.

**Assumption 2.3** Area is continuous.

Exactly what is meant by continuity is an issue that we will — indexcontinuity devote an entire chapter of this book to, and proving that the area of a rectangle is \( xy \) for all real numbers \( x \) and \( y \) requires some sophisticated mathematics! Intuitively the assumption merely asserts that rectangles with nearly the same dimensions should have nearly the same area. The sophistication involves putting this intuition on a mathematical foundation.

**Problem 2.3** In the previous discussion of the area of a rectangle there were some assumptions made on the sly in addition to the three stated so forcefully. Find one.
Let us now see what we can prove based on the assumption that the area of a rectangle is $xy$.

**Problem 2.4** Prove that the area of a right triangle is half the base times the height (see Figure 2.4).

![Figure 2.4:](image)

**Problem 2.5** Prove the Pythagorean Theorem (see Figure 2.5).

![Figure 2.5:](image)

The little mathematics we have done thus far contains a flavor of what mathematics is all about. We all carry an intuitive idea about various mathematical concepts, such as area. A mathematician will try to formulate a rigid, logical statement that captures this intuition. Part of the intuition we have for area was recorded in Assumption 2.1. If someone produces a definition of area for a rectangle that does not satisfy Assumption 2.1, then it is a bad definition!
Problem 2.6 Show that the following is a bad definition for the area of a rectangle: if a rectangle has sides of length $x$ and $y$ then its area is defined to be $x^2y^2$.

On the other hand, in the process of making a connection between our intuition and the logical definition, we discover that there are several (infinitely many in fact) legitimate definitions for area. This is a recurring theme throughout modern mathematics; with a concrete mathematical concept in hand, mathematicians go out and search for a list of logical statements that describe that concept. All facts about this concept are then drawn as logical deductions from the list (and the deductions are given names like “theorem”, “proposition”, or “lemma”). What happens almost every time is that we discover that there are myriad other mathematical entities that behave just like the concept we started with, and are also described by our list of logical statements. We are left with a general mathematical theory that applies to much more than we had foreseen. The act of creating the list of statements is called abstraction, and the resulting theory produced from the list is an instance of generalization. If you are curious what the final form of our discussion of area might look like, peruse a book that covers the subject of measure theory.

Fields

We will now give the definition of a field, which is a mathematical object that one studies extensively in the subject of abstract algebra. The definition of a field will use up nine of the fourteen logical statements that describe the set of real numbers, and these nine will be referred to as field axioms. The next four statements will appear later in this chapter, and the last statement will have to wait until we have introduced enough language to express it.

When we say that we have a set $F$ on which an addition is defined we mean there exists a rule (a function) that associates an element $a + b \in F$ to each pair of elements $a, b \in F$. For example, there is an addition defined on $\mathbb{N}$ that assigns the number $5 \in \mathbb{N}$ to the pair $2, 3 \in \mathbb{N}$. We mean the same thing when we say a multiplication is defined on $F$, except $a \cdot b$ denotes the output of multiplication. There are many concrete instances of sets with addition and multiplication defined on them, such as $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$. The last three of these satisfy all the statements below, and hence are fields, while the first two are not fields. The letter $F$ is being used like a variable, representing any set that is a field.

Definition 2.1 A field $F$ is a set on which an addition $+$ and multiplication $\cdot$ are defined such that

(Field Axioms)

1. $a + (b + c) = (a + b) + c$ for all $a, b, c \in F$.

2. $a + b = b + a$ for all $a, b \in F$.
3. There exists an additive identity; that is, there exists an element denoted 0 in F with the property that
\[ a + 0 = 0 + a = a \text{ for all } a \in F. \]

4. Each element \( a \in F \) has an additive inverse in F denoted \(-a\) and satisfying
\[ a + (-a) = (-a) + a = 0 \text{ for all } a. \]

5. \((a \cdot b) \cdot c = a \cdot (b \cdot c)\) for all \(a, b, c \in F\).

6. \(a \cdot b = b \cdot a\) for all \(a, b \in F\).

7. There exists a multiplicative identity element denoted 1 in F with the property that
\[ a \cdot 1 = 1 \cdot a = a \text{ for all } a \in F. \]

8. Each element \(a \in F\), except 0, has a multiplicative inverse in F denoted \(a^{-1}\) and satisfying
\[ a \cdot a^{-1} = a^{-1} \cdot a = 1 \text{ for all } a. \]

9. \((a + b) \cdot c = a \cdot c + b \cdot c\) for all \(a, b, c \in F\).

Although subtraction and division are not explicitly mentioned in the list of axioms, they are implicit in the axioms declaring the existence of inverses. Thus \(3 - 2\) refers to the sum of 3 with the additive inverse of 2. In general, \(x - y\) is defined to be \(x + (-y)\), i.e. the sum of \(x\) with the additive inverse of \(y\). Similarly, \(x/y\) is defined to be the product of \(x\) with the multiplicative inverse of \(y\), i.e. \(x/y\) is defined to be equal to \(x \cdot y^{-1}\). It is customary to omit the dot that indicates the operation of multiplication and we will follow this custom. Thus \(xy\) is written to indicate \(x \cdot y\). When confusion may arise we will use parenthesis liberally; so instead of writing 35 we will write \((3)(5)\) to indicate the product of 3 and 5. The set of rational numbers is a field, and one might imagine someone dreaming up the list of properties that define a field using the rational numbers as a model. In so doing, the dreamer would avoid listing properties that can be derived from other properties in the hope of obtaining a minimal list from which all the familiar algebraic facts about rational numbers can be deduced. Our list is not minimal in the sense that some statements in the list can be proven using other statements in the list. A second definition of a field is equivalent to ours if the set of statements provable from each definition is the same. As a consequence, if you have a definition of a field containing a statement that is provable from the other statements, then it can be removed from the definition resulting in a smaller equivalent definition.
Problem 2.7 Find a statement in the list of field axioms that, when omitted, results in an equivalent definition.

When we are taught what multiplication is, it usually sounds like “three times five is five added to itself three times.” Notice that this is Field Axiom (9) since

\[ 3 = 1 + 1 + 1 \text{ and } (3)(5) = (1 + 1 + 1)(5) = 5 + 5 + 5. \]

If we are careful, we can also convince the young student that

\[ (3)(-5) = -15 \]

by saying you should add \(-5\) to itself three times. Notice that this is Axiom (9) again. But how do we convince a third grader that \((-3)(-5) = 15\)? I have seen some attempts to do so but I sympathize with the student who feels like they are being brainwashed. If we are willing to accept each of the nine properties, then we can use these to prove \((-3)(-5) = 15\). Here’s an outline of how a proof could go (feel free to fill in the details!). As stated in Axiom (4), each \(a \in F\) has an additive inverse; maybe it has many! The first step is to prove it only has one inverse, i.e. prove that \(a + b = 0\) and \(a + c = 0\) implies \(b = c\). Once this has been established one can prove \((-5)(-3) = 15\) by showing that \((-5)(-3)\) is an inverse of \(-15\) (since there is only one inverse and that one is 15). Thus we are left with trying to prove \((-5)(-3) + (-15) = 0\). If you bought that \(-15 = (-5)(3)\), then use this and Axiom (9) to finish the proof off.

Ordered Fields

Although the rational numbers may well have been a model for the list of field axioms, it turns out that there are many fields that are very different than \(\mathbb{Q}\). There is even a field with only two elements! The set of rational numbers has more structure that is not captured in the list of field axioms, structure that is not algebraic but involves the concept of order. The following axioms define a structure called an ordered field: Axioms (10) and (11) involve only the idea of order, while Axioms (12) and (13) connect the field structure with the order. When we say that there is an order \(<\) defined on the set \(F\) we mean that there is a distinguished subset \(P\) of \(F\) consisting of the positive elements, and we write \(x < y\) to indicate that \(y - x \in P\). For example, \(\mathbb{Z}\) has an order \(<\) defined by taking the set of positive elements to be the natural numbers, i.e. \(P = \mathbb{N}\). In this context, writing \(10 < 65\) simply says that \(65 - 10 = 55 \in \mathbb{N}\), while \(65 \not< 10\) since \(10 - 65 = -55\) is not in \(\mathbb{N}\).

Definition 2.2 An ordered field is a field \(F\) upon which an order \(<\) is defined that satisfies the following
(Order Axioms)

10. If \(x \neq y\) then either \(x < y\) or \(y < x\).
11. If \( x < y \) and \( y < z \) then \( x < z \).
12. If \( x < y \) then \( x + z < y + z \).
13. If \( 0 < x \) and \( 0 < y \) then \( 0 < xy \).

You can find a list of statements that follow from Axioms (1) through (14) at the end of the chapter. In the text we will develop only what is needed to define absolute value, define distance, and then prove the most important fact in this chapter; the triangle inequality. The triangle inequality gets its name from the geometric picture it represents in the Cartesian plane (Figure 2.6). The inequality is simply a formula that captures the fact that the length of one side of a triangle is less than the sum of the lengths of the other two sides. The formula is obtained by introducing symbolism that represents the lengths on the sides of the triangle: if \( x \) and \( y \) are any two points in the plane, define \( d(x, y) \) to be the distance from \( x \) to \( y \). Now if \( x, y \) and \( z \) denote the vertices of a triangle, then the triangle inequality may be expressed as

\[
d(x, z) < d(x, y) + d(y, z).
\]

If we want a formula that is valid for all points \( x, y \) and \( z \) (not requiring them to be vertices of a triangle) we could write

\[
d(x, z) \leq d(x, y) + d(y, z),
\]

since if \( x, y \) and \( z \) are not vertices of a triangle then they lie on the same line and it is possible that equality will hold in Equation 2.3 (writing \( \leq \) is shorthand for either < or =).
It is important to realize that we have not proved the triangle inequality is valid in the Cartesian plane, or in any other space for that matter. What we have done is appealed to our intuition to write down an inequality that should be provable, provided that the axioms that describe the structure of the plane have been laid down correctly. As we will see later in the book, the proof that the triangle inequality holds in the plane is a consequence of the fact that the triangle inequality it true in \( \mathbb{R} \). Part of the definition of \( \mathbb{R} \) is that it is an ordered field; we will now set out to prove that the triangle inequality is true in every ordered field.

What we are able to do in any ordered fields is construct a distance function \( d \) that satisfies Equation 2.3. The distance function is defined in terms of the absolute value function, which will define now. Axiom (10) states if \( 0 \neq a \), then either \( 0 < a \) or \( a < 0 \). In particular, given an arbitrary \( a \in F \), exactly one of the following three cases must occur;

\[
a = 0 \quad \text{or} \quad 0 < a \quad \text{or} \quad a < 0.
\]

The absolute value of \( a \), denoted \( |a| \), is defined by these three cases;

\[
|a| = \begin{cases} 
0 & \text{if } a = 0 \\
\quad a & \text{if } 0 < a \\
\quad -a & \text{if } a < 0
\end{cases}
\]

Finally, our distance function is defined by \( d(x, y) = |x - y| \).

**Problem 2.8** Show that \( 0 \leq |a| \) for all \( a \in F \).

**Problem 2.9** Show that \( |a| = |-a| \) for all \( a \in F \).

**Problem 2.10** Prove that \( 0 \leq d(x, y) \) and \( d(x, y) = d(y, x) \).

The triangle inequality can be established by first proving \( |a+b| \leq |a|+|b| \), which is itself proved by considering all cases of \( a \) and \( b \) being positive or negative.

**Problem 2.11** If \( 0 \leq a \) and \( 0 \leq b \), prove that \( |a+b| \leq |a|+|b| \).

**Problem 2.12** If \( a < 0 \) and \( 0 \leq b \), prove that \( |a+b| \leq |a|+|b| \).

**Problem 2.13** Prove that \( |a+b| \leq |a|+|b| \) for all \( a, b \in F \).

**Problem 2.14** Prove that \( d(x, y) \leq d(x, z) + d(z, y) \) for all \( x, y, z \in F \).

We mentioned earlier that there is a field with two elements. The two elements are forced by the field axioms to be the additive identity and the multiplicative identity.

**Problem 2.15** In the two element field, prove that \( 1 + 1 = 0 \).
As we add more and more axioms it becomes harder for structures to satisfy all the statements. You can think of all the fields gathered together in one room when a proclamation is made that only the ordered fields can remain in the room and all other fields should leave immediately. While $\mathbb{Q}$ will be allowed to remain in the room, the next problem says that the two element field is one of those who must leave. We only have one more axiom to add to the list, and we might imagine that all ordered fields that do not satisfy this axiom will again be asked to leave the room. It is a fact that after all the deficient ordered fields have left, $\mathbb{R}$ will be the only field left in the room!

**Problem 2.16** *Show that there does not exist a two element ordered field.*

**Field Exercises**

1. Prove that additive inverses are unique.
2. Prove that multiplicative inverses are unique.
3. Prove that $-(-x) = x$ and that $(x^{-1})^{-1} = x$ when $x \neq 0$.
4. Prove that the additive identity and the multiplicative identity are unique.
5. Prove that $x \cdot 0 = 0$ for all $x \in F$.
6. Prove that if $xy = 0$ then either $x$ or $y$ is 0.
7. Prove that the additive inverse of $xy$ equals $-x)$.y.
8. Prove that $(-1)(-1) = 1$.

**Ordered Field Exercises**

1. Prove that $0 < x$ and $y < z$ implies $xy < xz$.
2. Prove that $x < y$ implies $-y < -x$.
3. Prove that $0 < x^2$ for all $x \neq 0$ in $F$.
4. Prove that $0 < 1$.
5. Prove that $0 < x < y$ implies $0 < y^{-1} < x^{-1}$. 
Chapter 3

Intuitive Calculus

There are two major branches of calculus, differential and integral. The motivation for developing differential and integral calculus comes from solving problems that appear to be unrelated, but whose solutions both employ a common ingredient, the notion of a limit. Differential calculus arises from the problem of understanding rates of change, while integral calculus is first introduced as a tool for finding areas of general regions. The most important and powerful theorem of calculus is called the fundamental theorem of calculus, and it is this theorem that provides a surprising and intimate connection between differential and integral calculus.

The purpose of this chapter is to familiarize you with the principle objects of differential and integral calculus. The goal for now is to know what these objects represent graphically. We will provide logical statements that define these objects in later chapters, after which we will be in a position to prove things about them.

Integral Calculus

The process of translating a problem into a mathematical problem is one in which mathematical objects, in particular sets and functions, are introduced to represent the objects of the problem. If you are given the problem of finding the
area of the region in Figure 3.1 you might mathematize the problem by thinking of the boundary as graphs of functions. The function $f$ is defined for numbers $x$ in the interval $[a, b]$ and its graph is the uppermost boundary of the region. Similarly, $g$ is defined for $x \in [a, b]$ and its graph is the lowermost boundary of the region. The problem has now been put in terms of mathematical objects and it is solved by defining a process, called integration, which produces the area between the graph of a function and the $x$-axis. The symbol $\int_a^b f$ denotes the shaded area in Figure 3.3. You now have a mathematical method to find the area of the region in Figure 3.1, assuming that $\int_a^b f$ and $\int_a^b g$ are computable. The area is then $\int_a^b f - \int_a^b g$ (since $\int_a^b g$ represents the shaded area in Figure 3.4).

The symbol $\int_a^b f$ is called the integral of $f$ on the interval $[a, b]$. The integral is itself a function, but it is perhaps unlike functions you have thought about in the past because its domain is not a set of numbers. What one inputs into the integral is a function and an interval of numbers and the integral then outputs a number that represents an area. Actually, we have only given you an idea of what $\int_a^b f$ is for functions whose graph is above the $x$-axis. When the graph of the function is below the $x$-axis one defines $\int_a^b f$ to be the negative area between the graph and the $x$-axis. The reason for doing this is to ensure that the integral
Figure 3.4:

becomes what is called a linear function. Linear functions are so important that there is a whole subject of mathematics devoted to them called linear algebra. With this extended definition the area of the shaded region in Figure 3.5 is

Figure 3.5:

\[ - \int_a^b f. \]

The integral is also defined in a way that respects our demand that the area of a whole be the sum of areas of parts. One instance of this is captured in the equation

\[ \int_a^c f = \int_a^b f + \int_b^c f, \]

(3.1)

for \( a < b < c \).

**Problem 3.1** Draw a picture that illustrates the meaning of Equation 3.1 for a function whose graph is above the x-axis.

**Problem 3.2** Assume \( f \) is the function whose graph appears in Figure 3.6. Find the following: (a) \( f(0) \)  (b) \( \int_0^1 f \)  (c) \( \int_2^5 f \)  (d) \( \int_0^2 f \)

**Differential Calculus**

The integral is the principle object of integral calculus and the derivative is the principle object of differential calculus. To acquaint you with the idea of a
derivative imagine that a car is driving along a straight road. A mathematization of this scenario might be to think of the road as a number line, with the origin at some specified place (see Figure 3.7). One might then let $f$ represent the function that outputs the number on the line where the car is when the time is input. Perhaps at time 3 (units are not particularly relevant) the car is at 4.8, which can be expressed briefly by writing $f(3) = 4.8$. Of course there are other functions around; we could let $g$ represent the function that returns the car’s velocity when the time is input. The idea of the derivative is that $g$ can be obtained from $f$. In fact, if you look at the graph of $f$ it is possible to read off (at least approximately) what $g(t)$ is at any time $t$. The illustration in Figure 3.8 shows you how this can be done. Here is why this works; if you

![Figure 3.6:](image1)  

![Figure 3.7: Snapshot at time 3](image2)  

![Figure 3.8: The graph of $f$](image3)
were asked to find the average velocity of the car in the interval of time between $t = 1$ and $t = 3$ you would divide the change of position by the elapsed time. In terms of our symbols this is exactly

\[
\frac{f(3) - f(1)}{3 - 1}.
\]

In terms of the graph of $f$ this number is the slope of the secant line joining two points on the graph (see Figure 3.9). If you want to know the velocity at the instant when $t = 1$ you could approximate it by taking averages over shorter and shorter time intervals (Figure 3.10). The slopes of the secant lines then approach the slope of the line tangent to the graph of $f$ at $(1, f(1))$, and the slope of this tangent line is $g(1)$. If you are given a function $f$, the derivative of
f is denoted \( f' \) and it is another function. When \( x \) is input into the derivative the output is the slope of the tangent line to the graph of \( f \) at the point \((x, f(x))\).

In the preceding discussion the function \( g \) was the derivative of \( f \), which may be expressed symbolically by writing \( g = f' \). The word “tangent” can be misleading in certain instances, such as when the graph of the function is a straight line. A better interpretation of what the derivative function outputs is obtained as follows; if you are interested in \( f'(3) \), look at the graph of \( f \) around the point \((3, f(3))\) and zoom in. If upon repeated magnification the graph approaches a line, as in Figure 3.11, then \( f'(3) \) is the slope of that line. A particular instance is when the graph of \( f \) is itself a line, in which case \( f'(3) \) is the slope of that line. For example, the function that is defined by the equation \( f(x) = 5x \) has a derivative function that outputs the number 5, no matter what number is input. If upon magnification the graph does not approach a line (see Figure 3.11 again), then we say that \( f'(3) \) does not exist. This viewpoint of a derivative is a very good one; if \( f'(x) \) exists, we interpret this by thinking of \( f \) as behaving like a linear function (a function whose graph is a line) for numbers close to \( x \).

![Figure 3.11: The derivative at 3 exists, but the derivative does not exist at 4](image)

**Problem 3.3** If \( f \) is the function whose graph appears in Figure 3.12 then find

(a) \( f(2) \)  \hspace{1cm} (b) \( \int_0^1 f \)  \hspace{1cm} (c) \( f'(3) \)  \hspace{1cm} (d) \( f'(1) \)  \hspace{1cm} (e) \( \frac{f(3) - f(1)}{3 - 1} \)  \hspace{1cm} (f) \( \frac{f(4) - f(3)}{4 - 3} \)

**The Fundamental Theorem**

The two branches of calculus described above appear to have nothing to do with each other, but actually they are intimately related. The relation between the integral and differential calculus is the content of the fundamental theorem of calculus, which provides a careful mathematical formulation of the relationship in the statement of the theorem, together with the formal mathematical proof that the relationship is true. We have not yet defined the concepts needed to give a formal statement of the theorem, but we are in a position to provide an
intuitive description of the statement. The goal of the next several chapters is to build the mathematical tools needed to give a precise statement and proof of the fundamental theorem.

If you have a function $f$ like the one pictured in Figure 3.11, then it is possible to use the theory of integration to define a related function $g$ that outputs the area bounded by the graph of $f$ on the interval $[1, x]$, as illustrated in Figure 3.13. Notice that $g$ is a function of the interval of integration; imagine $x$ changing, so that the interval of numbers $[1, x]$ changes, and visualize the corresponding shaded area changing. When $x = 1$ there is absolutely no area, so $g(1) = 0$. As $x$ moves to the right, $g(x)$ gets larger. The fundamental theorem states that when one associates such a function $g$ to a function $f$ in this manner, it frequently happens that the derivative of $g$ is $f$. We use the word “frequently” because there are functions $f$ where this constructions yields a function $g$ with $g' = f$, and then there are functions $f$ where this construction fails altogether. The fundamental theorem gives a precise condition on $f$ that guarantees the resulting function $g$ satisfies $g' = f$. 
The practical importance of the fundamental theorem is that it provides a means by which it becomes possible to calculate integrals, and this is where the name of the subject calculus derives from. If we desperately needed to know what $\int_1^5 f$ is, then the definition of $g$ tells us that the answer is $g(5)$, and the fundamental theorem tells us that $g$ has $f$ as its derivative. It turns out that if we find any function $h$ that has $f$ as its derivative, this function $h$ differs from $g$ by a constant, that is $g(x) - h(x)$ is the same value no matter what $x$ is input (this is a fact that will be established when we develop differential calculus). If we put this information together it gives a recipe for finding $\int_1^5 f$. The first step is to find any function $h$ that satisfies $h' = f$. The fact that $g(5) - h(5) = g(1) - h(1)$ then leads immediately to the answer

$$h(5) - h(1) = g(5) - g(1) = g(5) = \int_1^5 f.$$ 

This reduces the problem of finding areas to the computational task of finding antiderivatives.

**Problem 3.4** Assume $f$ is the function whose graph is illustrated in Figure 3.13 and $g$ is the corresponding function defined by $g(x) = \int_1^x f$. 

*Find approximate values for $f(2)$, $g(2)$, $g'(3)$, and $g'(4)$.*

![Figure 3.14: The graph of $f$.](image)

**Problem 3.5** Assume that $f$ is the derivative of the function $g$ that returns an output of $\frac{x^4}{4} - 3x^3 + 13x^2 - 21x$ when $x$ is input, i.e. $g(x) = \frac{x^4}{4} - 3x^3 + 13x^2 - 21x$ for every $x$. The graph of $f$ is pictured in Figure 3.14; find the area of the region shaded in Figure 3.14.
Chapter 4

If–Then Statements

Mathematical proofs consist of setting down a list of axioms that reflect our intuition of a mathematical object, then drawing logical deductions from these axioms. The axioms are assumed to be true statements, and from them one seeks true if–then statements. Much of the art of creating mathematics involves finding appropriate axioms. In the previous chapter you learned the intuition underlying the basic objects of calculus and soon we will confront the task of writing logical statements that captures this intuition.

The English meaning of if–then is not quite the same as the logical meaning of if–then, which can be an impediment to learning mathematics. The difference is largely a degree of precision. As an illustration, imagine that you are planning a tennis date with a friend and you have the option of playing indoors or outdoors. When you ask your tennis partner where to play, you would be satisfied if she replied, “If it’s raining then let’s play indoors.” Most people would expect this response to mean you will play outdoors if it is sunny. The point is that this loose understanding of the reply is more than what was said. It was never explicitly said that, “If it’s not raining then we play outdoors.” In mathematics and logic, one must never infer that the converse of an if–then statement is true just from knowing that the if–then statement is true, as was done in the illustration above.

To make matters even more complicated, the common usage of if–then statements in mathematical writing is slightly different than the formal logical meaning of an if–then statement. Logicians are very careful to quantify every variable used in an if–then sentence, whereas mathematicians tend to forget quantifying variables occasionally. An example is the true statement

if \( x + 5 = 10 \) then \( x = 5 \).

A mathematician would not bother to mention that this is a true if–then statement for all \( x \), whereas a logician would have. In this book we will study the meaning of an if-then statement as it is used in mathematical writing, but we will keep the discussion slightly naive and avoid the formalism that the interested student will find in a deeper study of logic.
We will now define the meaning of if–then as it will be used in this book and it is important that you adopt this meaning when you encounter if–then sentences in mathematics. An if–then statement has two components, a hypothesis and a conclusion, as illustrated:

\[
\text{if } \text{hypothesis} \text{ then } \text{conclusion}.
\]

The hypothesis and conclusion are both logical statements that usually contain variables. For example, the statement “\(x + 5 = 10\)”, which contains the variable \(x\), is the hypothesis of the statement “\(x + 5 = 10 \text{ then } x = 5\)”. This hypothesis is true when \(x = 5\) and the hypothesis is false when \(x \neq 5\). Similarly, the conclusion is true exactly when \(x = 5\).

**Definition 4.1** We define an if–then statement to be a true statement provided the conclusion is true whenever the hypothesis is true.

Thus “\(x + 5 = 10 \text{ then } x = 5\)” is a true if–then statement, and so is “\(x + 5 = 10 \text{ then } x \in \mathbb{N}\)”. When the hypothesis and conclusion are interchanged one obtains what is called the converse if–then statement. The converse of “\(x + 5 = 10 \text{ then } x = 5\)” is “\(x = 5 \text{ then } x + 5 = 10\)”, which also happens to be a true if–then statement. When both the if–then statement and its converse are true we say that the hypothesis and conclusion are equivalent statements and this is often expressed by writing

\[
\text{hypothesis if and only if conclusion}
\]

Thus we have “\(x + 5 = 10 \text{ if and only if } x = 5\)”. However, it is not true that “\(x + 5 = 10 \text{ if and only if } x \in \mathbb{N}\)”, since the if–then statement “\(x \in \mathbb{N} \text{ then } x + 5 = 10\)” is false. To prove that an if–then statement is false, one provides what is called a counterexample; this is an example of a value for the variable that yields a true hypothesis but a false conclusion. In the false if–then statement “\(x \in \mathbb{N} \text{ then } x + 5 = 10\)”, letting \(x = 1\) gives a counterexample because the hypothesis “\(x \in \mathbb{N}\)” is true, whereas the conclusion “\(x + 5 = 10\)” is false. This counterexample proves the statement false because the condition in Definition 4.1 that defines a true if–then statement fails; it is not true that the conclusion is true whenever the hypothesis is true, and \(x = 1\) gives an instance of such a failure.

We occasionally encounter an if–then statement where the conclusion is always true, or we may find one where the hypothesis is always false. Both of these if–then statements are automatically true since the conclusion is true whenever the hypothesis is true. For example the if–then statement “\(\text{if } x \in \emptyset \text{ then } x \text{ is a tree}\)” is a true statement; since the hypothesis is never true you will never find a counterexample for this statement. (The empty set has no elements so it is never true that \(x \in \emptyset\).)

As the variables in the if–then statement vary the hypothesis will be true at times and false at times. To check the truth of the if–then statement itself one need only concern themselves with the instances when the hypothesis it true and then make sure that the conclusion is true. This reveals how one goes about
proving that an if–then statement is true; you assume an arbitrary instance of truth in the hypothesis and deduce from this the truth of the conclusion. To prove that the statement “if \( x + 5 = 10 \) then \( x = 5 \)” is true you begin by assuming the hypothesis; assume it true that \( x + 5 = 10 \). You can then deduce from this assumption, together with the axioms in chapter 2, that

\[
x = x + 0 = x + (5 + (-5)) = (x + 5) + (-5) = 10 + (-5) = 5.
\]

Problem 4.1 Determine whether the following if–then statements and their converses are true or false. If a statement is false, prove it by giving a counterexample.

1. if \( x^2 = 1 \) then \( x = -1 \).
2. if \( 0 \neq 1 \) then \( 0 = 1 \).
3. if \( x \in \mathbb{N} \) then \( x \in \mathbb{Q} \).
4. if \( x \in \emptyset \) then \( 10 = 12 \).

Saying that the conclusion is true whenever the hypothesis is true is exactly the same as saying the hypothesis is false whenever the conclusion is false. Convince yourself of this! When you believe this you will understand that it is possible to prove an if–then statement by assuming the conclusion is false and deducing from this that the hypothesis is false. This is called proving by contrapositive. If you have the if–then statement

\[
\text{if } \boxed{\text{statement 1}} \text{ then } \boxed{\text{statement 2}},
\]

you can construct an equivalent if–then statement

\[
\text{if } \boxed{\text{not statement 2}} \text{ then } \boxed{\text{not statement 1}}
\]

called the contrapositive statement. If you can prove the contrapositive statement is true you have proven that the original if–then statement is true.

Problem 4.2 Write the contrapositive of the if–then statement “if \( x \in \mathbb{N} \) and \( x \geq 4 \) then \( x \) may be written as a sum of two prime numbers”. Find a counterexample to the if–then statement and a counterexample to the contrapositive statement. Is the converse statement true or false?

All you need to know about prime numbers in order to do the previous problem is that they are natural numbers and the first five of them are 2, 3, 5, 7 and 11.

We would like to mention an important comment concerning mathematical definitions which you will find throughout the mathematical literature, and in particular in this book. Mathematicians will frequently use an if–then statement to define a term, such as “if all sides of a rectangle have equal lengths,
then that rectangle is a *square*, a statement that defines a square. Definitions are always intended to be if and only if statements, however! When you read a definition that is given in the form of an if-then statement, you must understand that the statement is true exactly when its converse is true. Thus “if a rectangle is a square then all sides of that rectangle have equal lengths” is true when the converse is true! It is curious that even in the technical writings of mathematicians you will find occasions when they slip into colloquial usages of if-then statements.
A powerful tool that we will use to prove calculus theorems is the concept of a sequence. Assume that $F$ is an ordered field. (What we are going to do with $F$ applies when $F$ is the set of rationals or when it is the set of real numbers. Think of $F$ as a variable that can be any object that satisfies the thirteen listed axioms given in Chapter 2.) A sequence in $F$ is thought of intuitively as an infinite list of elements of $F$. For example, we might write $1, \frac{1}{2}, \frac{1}{3}, \ldots$ to indicate a sequence in $\mathbb{Q}$. The three dots mean the list continues forever to the right; the next (unwritten) term of the sequence is $1/4$.

Calculus constructions all have one common ingredient; they are all built using the idea of a limit. Maybe we can not compute the area of the region drawn in Figure 5.1 because the boundary is curved. We could, however, approximate

![Figure 5.1:](image)

the area by inscribing rectangles and triangles, as in Figure 5.2. With patience we could get very close to the exact area of the region. One construct you meet in calculus (called an integral) is built to obtain areas of regions in just this way; the area is obtained indirectly as a limit of approximations.

We have not said yet exactly what a limit is, but you should think of it intuitively as the number at the end of all the approximations. You can think
of the terms in a sequence as the approximations which get closer to the answer as you move further out in the sequence. With only this knowledge you can probably do most of the following problem.

**Problem 5.1** Guess the limits of the following sequences.

1. $1, \frac{1}{2}, \frac{1}{3}, \ldots$
2. $3, 3.1, 3.14, 3.141, 3.1415, 3.14159, \ldots$
3. $\frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \frac{6}{5}, \ldots$
4. $1, -1, 1, -1, \ldots$

To make the intuition of a sequence mathematically precise we must associate a mathematical object with it that captures this idea of an infinite list. The first impulse might be to associate the set of terms of the sequence, i.e. think of the sequence $1, \frac{1}{2}, \frac{1}{3}, \ldots$ as being the set $\{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$. This approach has a serious flaw; part of the important intuition we wish to capture about a sequence is “where it is going” and this may be lost if one only looks at the set of its terms. For example, the sequence $-1, 1, 1, 1, \ldots$ (it says at 1 forevermore) is going to 1; in fact, it got there and stayed there! However, the sequence $-1, 1, -1, 1, -1, 1, \ldots$ (it alternates forevermore) is going absolutely nowhere. These are very different sequences even though they both have $\{1, -1\}$ as the set of their terms.

The information we wish to capture is knowing what the $10^{th}$ term is, what the $100^{th}$ term is, and what the $n^{th}$ term is for any $n \in \mathbb{N}$. The mathematical object that expresses this is a function whose domain is $\mathbb{N}$ and whose output consists of numbers in $F$.

**Definition 5.1** A sequence in $F$ is a function $s : \mathbb{N} \rightarrow F$.

We will indicate a sequence in $F$ by writing $(s_1, s_2, s_3, \ldots)$ or more briefly by writing $(s_n)$. With this notation it is to be understood that the function outputs $s_n$ when $n$ is input.
Before we define what a limit of a sequence is, we need to recall some terminology. If \( x, y \in F \) recall that
\[
\begin{align*}
[x,y] & = \{ z \mid x \leq z \leq y \} \\
[x,y) & = \{ z \mid x \leq z < y \} \\
(x,y] & = \{ z \mid x < z \leq y \} \\
(x,y) & = \{ z \mid x < z < y \}
\end{align*}
\]
The last one is called an open interval, the first a closed interval, and the middle two are called half open intervals. An open interval does not contain its endpoints, so if \( z \) is in an open interval then every point sufficiently close to \( z \) is also in that open interval. This key property of open intervals enables us to use open intervals to build a formal definition of a limit.

Here is how a mathematician might take the intuitive idea of a limit and formulate a logical statement that captures that idea. The mathematician has a concrete example in mind of a sequence approaching a particular number; perhaps it is the sequence \( \left( \frac{1}{n} \right) \) approaching 0. The goal is to find a logical statement that captures the idea of a sequence \( (s_n) \) converging to the limit \( s \), and the search begins by looking carefully at the concrete example. The mathematician will look for an if–then statement that has variables \( s_n \) and \( s \) and has the property that, when applied to the sequence with \( s_n = \frac{1}{n} \), is a true if–then statement only when \( s = 0 \). There is some danger that the resulting statement relies too heavily on particular properties of the concrete example (such as the fact that the terms of the sequence are getting successively smaller). It would later be necessary to test the statement on a variety of sequences to see whether the statement is true exactly when we consider the sequence \( s_n \) to be converging to \( s \). In the end the statement must stand a test of time; future generations will scrutinize the definition to ensure that it does in fact capture the intuition accurately. Sometimes definitions that have been in use for extended periods are discovered to be inadequate or incomplete and then they undergo a modification. This in fact happened when mathematicians were trying to agree on what was meant by a limit of a sequence of functions, which you will meet later in the book. We will now describe the insight behind the definition of a limit of a sequence of numbers, a definition that has passed the test of time.

Think about any open interval of rational numbers containing 0. That interval is determined by two numbers \( x, y \in \mathbb{Q} \) and is defined to be the set \( \{ z \mid x < z < y \} \). Since 0 is an element of this set we know that \( y \) is a positive rational number, so it is of the form \( y = \frac{p}{q} \) with \( p, q \in \mathbb{N} \). It is now possible to use the axioms for an ordered field to prove that
\[
0 < \frac{1}{2q} < \frac{p}{q},
\]
which implies that every term of the sequence \( \left( \frac{1}{n} \right) \) is in the open interval, except possibly for the finitely many terms whose denominators are smaller than \( 2q \). This leads to the realization that when \( s_n = \frac{1}{n} \) and \( s = 0 \), then the following
is a true if–then statement: if \( G \) is an open interval containing \( s \), then \( G \) also contains all of the numbers \( s_n \), except for finitely many terms of the sequence. Furthermore, 0 is the only value for \( s \) that makes this statement true, since if \( s \neq 0 \) we could find an open interval \( G \) that provides a counterexample to the statement. It will be convenient to introduce terminology that precisely describes the phrase “all but finitely many terms of the sequence”.

**Definition 5.2** We will say that the sequence \((s_n)\) is eventually in the set \( G \) in case the set of indices \( \{ n \mid s_n \notin G \} \) is a finite set.

**Definition 5.3** The sequence \((s_n)\) converges to \( s \) if the following is a true if–then statement:

\[
\text{if } G \text{ is an open interval containing } s \text{ then } (s_n) \text{ is eventually in } G.
\]

**Problem 5.2** Assume that \( s_n = \frac{1}{n} \) and \( G = (.1, 1) \). Explicitly list all the elements of the set \( \{ n \mid s_n \notin G \} \). Find a sequence \((t_n)\) that is eventually in \( G \).

To prove that a sequence \((s_n)\) converges to \( s \) you show that the if–then statement in Definition 5.3 is true; i.e. you assume you have an arbitrary open interval \( G = (x, y) \) with \( x < s < y \) and you prove that \( \{ n \mid s_n \notin G \} \) is finite. To prove that a sequence \((s_n)\) does not converge to \( s \) you exhibit counterexample to the if–then statement in Definition 5.3.

**Problem 5.3** Prove that \((\frac{1}{n})\) does not converge to 1.

A constant sequence is one where \( s_n = s_m \) for all \( n, m \in \mathbb{N} \). We say that a sequence \((s_n)\) converges when there exists \( s \) that makes the if–then statement in Definition 5.3 true.

**Problem 5.4** Prove that a constant sequence converges.

**Problem 5.5** Prove that in \( \mathbb{Q} \) the sequence \((\frac{1}{n})\) converges to 0.

A useful result about limits is one that says if you are squeezed between two things going to the same place, then you have to go there too; we refer to this as the squeeze theorem. A precise formulation of the squeeze theorem for sequences appears in the following problem.

**Problem 5.6** Assume that you have three sequences \((r_n)\), \((s_n)\), and \((t_n)\) such that \( r_n \leq s_n \leq t_n \) for all \( n \in \mathbb{N} \), and such that both \((r_n)\) and \((t_n)\) converge to the same number \( s \). Prove that \((s_n)\) converges to \( s \).

When two people are looking for a logical statement that describes the intuition of a converging sequence it is possible that they both arrive at different statements, yet they are both right. Assume that the first person arrives at a statement that we will call “statement 1” and the second person arrives at “statement 2”. If it happens that statement 1 is true if and only if statement 2
is true, then the two statements are logically equivalent and consequently both statements encode exactly the same intuition. The definition of convergence of a sequence that is found in most calculus books is the following statement;

$$\text{if } 0 < \epsilon \text{ then there exists } N \in \mathbb{N} \text{ such that } |s_n - s| < \epsilon \text{ for all } n \geq N.$$ 

**Problem 5.7** Prove that the logical statement above is equivalent to definition 5.3.

After the equivalence of the two statements defining convergence is established you can use either statement to prove that a sequence converges or fails to converge. It is often the case that a proof will follow from one of the statements more easily than the other. The nice thing is, if you have trouble proving a sequence converges using the definition, you can try proving it using the equivalent statement. Do not give up until you have tried both ways.

Mathematicians give a special name to true if–then statements that are particularly important; they call them theorems. The next three theorems establish very important facts about addition, multiplication and division.

**Definition 5.4** To abbreviate the exposition we introduce the symbolism

$$(s_n) \rightarrow s$$

**Theorem 5.1** If $(s_n) \rightarrow s$ and $(r_n) \rightarrow r$ then $(s_n + r_n) \rightarrow s + r$.

**Proof.** It is enough to prove that the following is a true if–then statement; if $0 < \epsilon$ then there exists $N \in \mathbb{N}$ such that

$$|(s_n + r_n) - (s + r)| < \epsilon \text{ for all } n \geq N.$$ 

To do this assume that the hypothesis is true, i.e. that $0 < \epsilon$. We are also assuming that $(s_n) \rightarrow s$ and $(r_n) \rightarrow r$ which says that

$$|s_n - s| < \frac{\epsilon}{2} \text{ and } |r_n - r| < \frac{\epsilon}{2} \quad (5.1)$$

when $n$ is sufficiently large. If $N \in \mathbb{N}$ is chosen so that equation 5.1 holds for all $n \geq N$ then

$$|(s_n + r_n) - (s + r)| = |(s_n - s) + (r_n - r)| \leq |(s_n - s)| + |(r_n - r)| < \epsilon.$$ 

$\triangle$
To prove the next theorem we need to know that a convergent sequence is bounded. A *convergent sequence* \((s_n)\) is one for which a number \(s\) exists so that \((s_n) \to s\). We say a sequence \((s_n)\) is bounded if the sequence is contained in some interval; i.e. if there exist \(x, y \in F\) such that
\[
x < s_n < y
\]
for all \(n \in \mathbb{N}\). An equivalent definition is that there exists a positive number \(z\) such that \(z > |s_n|\) for all \(n \in \mathbb{N}\). (You are encouraged to prove this equivalence.) All the sequences in Problem 5.1 are bounded.

**Problem 5.8** Prove if \((s_n)\) converges then \((s_n)\) is bounded.

**Problem 5.9** Prove that the converse of Problem 5.8 is false.

**Theorem 5.2** If \((s_n) \to s\) and \((r_n) \to r\) then \((s_n r_n) \to sr\).

**Proof.** As in the previous theorem we begin by assuming that \(0 < \epsilon\). We need to prove that eventually
\[
|s_n r_n - sr| < \epsilon.
\]
The insight needed now is to see that
\[
|s_n r_n - sr| = |s_n(r_n - r) + (s_n - s)r| \leq |s_n| |r_n - r| + |s_n - s||r|.
\]
Both \(|s_n - s|\) and \(|r_n - r|\) can be made as small as desired by the assumption that \((s_n) \to s\) and \((r_n) \to r\). Let \(x\) be a positive number such that \(x > |s_n|\) for all \(n \in \mathbb{N}\) and choose \(N \in \mathbb{N}\) so that
\[
|s_n - s| < \frac{\epsilon}{2|x|} \quad \text{and} \quad |r_n - r| < \frac{\epsilon}{2x}
\]
for all \(n \geq N\). Now when \(n \geq N\) we have
\[
|s_n r_n - sr| \leq |s_n||r_n - r| + |s_n - s||r| < \epsilon.
\]
\(\square\)

**Theorem 5.3** If \((s_n) \to s\), \(s \neq 0\), and \(s_n \neq 0\) for all \(n \in \mathbb{N}\), then \((\frac{1}{s_n}) \to \frac{1}{s}\).

**Proof.** Once again assume that \(0 < \epsilon\), so we need to prove that eventually
\[
|\frac{1}{s_n} - \frac{1}{s}| < \epsilon.
\]
To figure out how this quantity can be made small we rewrite it by getting a common denominator and see that
\[
|\frac{1}{s_n} - \frac{1}{s}| = \left|\frac{s - s_n}{ss_n}\right| = |s - s_n| \frac{1}{|s||s_n|}.
\]
(5.2)
Since \((s_n) \rightarrow s\) and \(s \neq 0\) we must have that eventually \(|s_n| > |s|/2\); choose \(N \in \mathbb{N}\) so that this inequality holds for all \(n \geq N\), and so that
\[
|s - s_n| < \frac{\epsilon |s|^2}{2}
\]
for \(n \geq N\). Substituting into Equation 5.2 then gives us
\[
\left| \frac{1}{s_n} - \frac{1}{s} \right| < \epsilon
\]
when \(n \geq N\).

\(\triangle\)

Reading proofs of theorems is very much like doing problems or exercises. One usually has to justify to oneself why certain steps are valid. The proofs of the three theorems just given each becomes progressively sketchier, requiring progressively more work from the reader, but they are all quite detailed in comparison to advanced texts. Most mathematicians believe that a reader better understands a sketchy proof whose logical gaps were pondered and filled than a detailed proof that was read passively, and this belief is reflected in their writing.

In Problem 5.5 the reader was asked to prove that \(\left(\frac{1}{n}\right)\) converges to 0, which can be done using properties of the rational numbers. It might come as a shock to learn that it is possible to construct an example of an ordered field for which \(\left(\frac{1}{n}\right)\) does not converge to 0. The point of mentioning this unhappy example is to inform you that it is not possible to prove that \(\left(\frac{1}{n}\right)\) converges to 0 using only the axioms for an ordered field (the unhappy example is a counterexample).

The final forthcoming axiom that is needed to establish the real number system will not only allow us to prove that \(\left(\frac{1}{n}\right)\) converges to 0, but it will be an essential assumption buried in all of the proofs of the important theorems of calculus. Let us introduce one more definition after which we will give the axiom for a complete ordered field.

We say that a sequence \((s_n)\) in \(F\) is increasing if
\[
s_1 \leq s_2 \leq s_3 \leq \ldots \leq s_n \leq \ldots ,
\]
i.e. in case \(s_n \leq s_{n+1}\) for all \(n \in \mathbb{N}\).

**Definition 5.5** A complete ordered field is an ordered field \(F\) that satisfies the additional axiom that
\[
(14) \text{ Every bounded increasing sequence converges.}
\]

We mentioned earlier that it is possible to prove that any two complete ordered fields are essentially the same. For this reason we use one symbol and one name to represent a complete ordered field; the name is the field of real numbers, and the symbol is \(\mathbb{R}\).
When we were discussing convergence of a sequence we pointed out that there is more than one logical statement that would serve as the definition of a sequence converging to a limit. The same thing happens when we define an object with a list of axioms, such as a complete ordered field whose definition involves fourteen axioms. There are several statements that would work in the place of Axiom (14). One such statement is “every bounded decreasing sequence converges.”

**Problem 5.10** Give a formal definition of a decreasing sequence.

**Problem 5.11** Deduce from the axioms for a complete ordered field that every bounded, decreasing sequence converges.

To get started on Problem 5.11 you should assume you have a bounded, decreasing sequence \((s_n)\). The next step is to use an algebraic trick to convert this to a bounded, increasing sequence. Use Axiom (14) and convert again to finish the problem.

We are nearly in a position to prove that \(\frac{1}{n} \to 0\) in \(\mathbb{R}\), but there is still a subtle point. You should be able to prove right now that \(\frac{1}{n}\) converges in \(\mathbb{R}\), but proving that 0 is the limit needs some work. The next problem is given to supply some of the ammunition you may use.

**Problem 5.12** Prove that limits of a sequence are unique; i.e. if \((s_n) \to s\) and \((s_n) \to r\) then \(s = r\).

One way to prove \(s = r\) is to prove that \(|s - r| = 0\). Since by its definition \(|s - r|\) is either positive or it is 0, you can prove \(|s - r| = 0\) by showing it is not positive. Let \(\epsilon\) be an arbitrary positive number. It is possible to prove \(|s - r| < \epsilon\) by using the triangle inequality

\[|s - r| \leq |s - s_n| + |s_n - r|\]

and the hypothesis that both \((s_n) \to s\) and \((s_n) \to r\). This strategy is tailored to the characterization of convergence obtained in problem 5.7.

A second way to prove that \(s = r\) is to prove the following if–then statement; if \(s \neq r\) then there exists an open interval \(G\) that contains \(s\) but not \(r\). If you use the information in the hypothesis of problem 5.12 and the contrapositive of the if–then statement above, the problem will be done. This strategy is most easily executed using the definition of convergence of a sequence. Choose your strategy; you are learning that there are many correct answers in mathematics.

To prove that \(\frac{1}{n} \to 0\) in \(\mathbb{R}\) one may use a technique that amounts to a variant of a proof by contrapositive, and that is a proof by contradiction. The axioms that we use to develop mathematics must have the property that it is not possible to prove both a statement and that statement’s negation; such an axiom system is referred to as consistent. A proof by contradiction starts by assuming the negation of what you intend to prove, and it ends by arriving at the negation of a statement that is already known to be true. Thus if you want to prove that
\((\frac{1}{n}) \to 0\), in a proof by contradiction you would assume that \((\frac{1}{n}) \to s\) and \(s \neq 0\). From this assumption it is possible to find a sequence that converges to two distinctly different numbers, which problem 5.12 says is impossible. What this establishes is the truth of the following statement; if \((\frac{1}{n}) \to s\) and \(s \neq 0\), then problem 5.12 is false. The contrapositive of this statement is our solution that shows \((\frac{1}{n}) \to 0\), since we know that problem 5.12 is true.

In the next problem you can assume \((\frac{1}{n}) \to s\) with \(s \neq 0\), use Theorem 5.3 to infer something about the sequence \((1, 2, 3, \ldots)\), and then use Theorem 5.1 to infer something about the sequence \((2, 3, 4, \ldots)\). You should see a contradiction to problem 5.12.

**Problem 5.13** Prove that \((\frac{1}{n}) \to 0\) in \(\mathbb{R}\).

**Problem 5.14** Prove the following is a true if–then statement: if \(0 < x\) and \(y \in \mathbb{R}\) then \(y < nx\) for some \(n \in \mathbb{N}\).

The content of Problem 5.14 is usually referred to as the Archimedean property of \(\mathbb{R}\). You can prove it using Problem 5.13.

An element \(x\) of an ordered field \(F\) is called an upper bound of a subset \(E \subseteq F\) if \(y \leq x\) for every \(y \in E\). Thus 3 is an upper bound of \(E = \{-1, 0, 1\}\) while 1/2 is not. We call \(x\) the least upper bound of \(E\) if \(x\) is an upper bound of \(E\) and every other upper bound of \(E\) is greater than \(x\). Thus if you were to prove \(x\) is the least upper bound of \(E\) you must prove the following are true if–then statements:

1. if \(y \in E\) then \(y \leq x\)
2. if \(z\) is an upper bound of \(E\) then \(x \leq z\).

In the previous example where \(E = \{-1, 0, 1\}\) the least upper bound is 1. If \(E = (0, 1)\) the least upper bound is again 1. These two examples show that a least upper bound of \(E\) may or may not be in the set \(E\). An ordered field \(F\) has the least upper bound property if every non-empty subset \(E\) that has an upper bound in \(F\) actually has a least upper bound in \(F\).

**Problem 5.15** Show that \(\mathbb{Q}\) does not have the least upper bound property.

**Problem 5.16** Assume that \(F\) is an ordered field that has the least upper bound property. Prove that \(F\) is a complete ordered field.

We could have taken Axiom (14) to be the assertion that “\(F\) has the least upper bound property” and we would have an equivalent definition of a complete ordered field. Of course, there’s a greatest lower bound property that would also work in place of Axiom (14).

**Problem 5.17** Formulate the definition of a lower bound of \(E\), a greatest lower bound of \(E\), and what it means to say that an ordered field has the greatest lower bound property.
Chapter 6

Limits

The purpose in the preceding chapter of constructing a theory that captures the essential properties of sequences is to use this theory as a foundation upon which the fundamental theorems of calculus can be built. Our plan is to extend the concept of a sequential limit to the idea of a limit of a function. Once we have established the basic properties of limits for functions we will be in a position to define derivatives and integrals and prove the deepest theorems of calculus.

Intuition

Let us begin by describing the intuition that underlies a functional limit. Suppose that $f$ is the function whose graph appears in Figure 6.1, so the domain of $f$ is

$$ (0, 1) \cup (1, 3]. $$

Even though 1 is not in the domain of $f$ we can see from the graph that as $x$ approaches 1 on the horizontal axis, $f(x)$ approaches 1 on the vertical axis. We
express this information by writing
\[ \lim_{x \to 1} f(x) = 1, \]
or other times we may write \( f(x) \to 1 \) as \( x \to 1 \). Notice that as \( z \) approaches 2 on the horizontal axis, \( f(z) \) approaches 3 on the vertical axis. This is expressed by writing \( \lim_{z \to 2} f(z) = 3 \). What we call the variable is irrelevant; we mean the same thing by writing \( \lim_{t \to 2} f(t) = 3 \) or \( \lim_{y \to 2} f(y) = 3 \). These both mean “as the input gets close to 2 the output of \( f \) approaches 3.” The limit of \( f \) as you approach 2 has absolutely nothing to do with what \( f \) does with the input 2. The limit is completely determined by the set of numbers close to, but not equal to 2. In fact, in this example \( f(2) = 1 \) while \( \lim_{x \to 2} f(x) = 3 \). The function might not even be defined at a point where the limit is taken; for example \( f(1) \) is undefined but \( \lim_{x \to 1} f(x) = 1 \).

The fact that \( f(2) \neq \lim_{x \to 2} f(x) \) expresses that the graph of \( f \) breaks at \( x = 2 \), and there is mathematical terminology to describe this situation. We will say that \( f \) is \textit{continuous} at \( x = a \) if
\[ f(a) = \lim_{x \to a} f(x), \]
and we will say \( f \) is discontinuous at \( x = a \) otherwise. With this terminology \( f \) is discontinuous at \( x = 2 \) but \( f \) is continuous at \( x = 1/2 \). We say that \( f \) is continuous on a set if \( f \) is continuous at every point of the set. The function defined in Figure 6.1 is continuous on the set \( (1, 2) \) but not on the set \( (1, 2] \). Implicit in the definition of continuity is that a point of continuity must be in the domain of a function. Thus \( f \) is discontinuous at \( x = 1 \) since this point is not in the domain.

Let us now move on to the function \( f \) whose graph appears in Figure 6.2. The arrows are meant to indicate that the curve continues forever with the same general shape. This ambiguity makes it impossible to tell what the behavior of \( f \) is outside of the interval \([-1, 5]\), but the arrows do tell us that the domain of \( f \) includes the sets \( \{ x \in \mathbb{R} \mid x < -1 \} \) and \( \{ x \in \mathbb{R} \mid x > 5 \} \). As drawn, the domain of \( f \) is exactly the set \( \{ x \in \mathbb{R} \mid x \neq 3 \} \). Even though \( f \) is not defined at 3 you
can see from the graph that \( \lim_{x \to 3} f(x) = 2 \). The function \( f \) is discontinuous at \( x = 3 \) since \( f(3) \) is undefined. We emphasize again that it still makes sense to talk about the limit of a function at a point not in the function’s domain. It will frequently be the case that we will be interested in the behavior of a function arbitrarily close to a point where a function is not defined.

If \( x \) approaches 2 from the right then \( f(x) \) approaches 1. We express this situation by writing

\[
\lim_{x \to 2^+} f(x) = 1.
\]

If \( y \) approaches 2 from the left then \( f(y) \) approaches 2, and the notation for this is

\[
\lim_{y \to 2^-} f(y) = 2.
\]

Once again, the choice of a letter to use as the variable is not important;

\[
\lim_{y \to 2^-} f(y) = \lim_{x \to 2^-} f(x) = \lim_{t \to 2^-} f(t) = 2.
\]

In this situation we will say \( \lim_{x \to 2^-} f(x) \) does not exist because the left and right limits are not equal. In general, \( \lim_{x \to a} f(x) \) exists if and only if both the left and right hand limits exist and

\[
\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x).
\]

Thus, using this new terminology and symbolism, we agree that

\[
\lim_{x \to 1^-} f(x) = 2 \quad \text{and} \quad \lim_{x \to 1^+} f(x) = 1,
\]

so \( \lim_{x \to 1^-} f(x) \) does not exist. We would like to point out that \( \lim_{x \to 2^-} f(x) \) not existing also implies \( f \) is discontinuous at \( x = 2 \). In order for a function to be continuous at a point the limit must exist and be equal to the function’s output at that point. This is what the symbolism \( \lim_{x \to a} f(x) = f(a) \) expresses very succinctly. For the function under discussion it is true that

\[
\lim_{x \to 2^-} f(x) = 2 = f(2).
\]

In this situation it is said that \( f \) is continuous from the left at \( x = 2 \). In contrast this same function is discontinuous from both the left and the right at \( x = 1 \).

**Problem 6.1** Define what it must mean for a function to be continuous from the right at \( x = a \) and what it means to be discontinuous from the right at \( x = a \).
We will have frequent occasion to say that “f is continuous on \([a, b]\)” (see Figure 6.3). What we mean by this is that f is continuous at every number \(t \in [a, b]\), f is continuous from the right at \(x = a\), and f is continuous from the left at \(x = b\). Each function in Figure 6.4 fails to be continuous on \([a, b]\).

**Problem 6.2** Using Figure 6.5, find:

\[
\begin{align*}
(a) & \ f(0) & (b) & \ f(\frac{5}{2}) & (c) & \ f(1) & (d) & \ \lim_{x \to 1} f(x) \\
(e) & \ \lim_{x \to 2} f(x) & (f) & \ \lim_{x \to 2^+} f(x) & (g) & \ \lim_{x \to 3^-} f(x) & (h) & \ \lim_{x \to 3^+} f(x)
\end{align*}
\]
Problem 6.3 Draw the graph of a single function $f$ with all of the following properties:

(a) $\int_0^1 f = 1$  
(b) $f$ is continuous on $[0, 1]$  
(c) $f(3) = 0$  
(d) $\lim_{x \to 2^+} f(x) = 3$  
(e) $\lim_{x \to 2^-} f(x) = 2$  
(f) $f(2) = 1$  
(g) $\lim_{x \to 3^-} f(x) = 2$  
(h) The domain of $f$ is $[0, 4]$  
(i) $f'(1) = 0$

Defining the derivative of $f$ involves associating a new function with $f$ that returns slopes of secant lines. For example, if $f$ is the squaring function and we want to compute $f'(2)$, then the first thing we do is construct the function

$$g(x) = \frac{f(x) - f(2)}{x - 2} = \frac{x^2 - 4}{x - 2},$$

which returns slopes of secant lines. If $f$ is not the squaring function we can still construct $g$ in the same way, as illustrated in Figure 6.6. It is important to realize that $g$ depends on $f$ and a number (the number 2 in both examples encountered so far). If you change $f$, then $g$ will change, and if you change the number, then $g$ changes again.

![Figure 6.6](image)

Stare at Figure 6.6 and imagine what happens to this secant line as $x$ approaches 2; as $x$ moves towards 2 the secant line pivots on the point $(2, f(2))$ and approaches the tangent line to the curve at $(2, f(2))$. This is how one is led to define the derivative at 2 as $\lim_{x \to 2} g(x)$, i.e. as

$$\lim_{x \to 2} \frac{f(x) - f(2)}{x - 2}.$$

Problem 6.4 Graph the function $g$ obtained from the squaring function and the number 2, i.e.

$$g(x) = \frac{f(x) - f(2)}{x - 2}$$

with $f(x) = x^2$. 

Problem 6.5 Graph the function \( g \) obtained from the squaring function and the number 0, i.e.
\[
g(x) = \frac{f(x) - f(0)}{x - 0}
\]
with \( f(x) = x^2 \).

Problem 6.6 Graph the function \( g \) obtained from the number 1 and the function \( f \) whose graph is given in Figure 6.6. (Give reasonable estimations.)

Problem 6.7 Estimate \( \lim_{x \to 2} g(x) \).

Problem 6.8 Let \( f \) be the cubing function, and let \( g \) be the function that returns the slope of the line joining the points \((2, f(2))\) and \((x, f(x))\). Graph \( g \) and find \( \lim_{x \to 2} g(x) \).

If you think of \( x \) as being a little bit added onto 2, i.e. \( x = 2 + h \) (and \( h \) would be approximately \(-1.5\) to obtain the \( x \) pictured in Figure 6.6), then the formula that gives the slopes of the secant lines takes the form
\[
\frac{f(2 + h) - f(2)}{h}.
\]
Thus an alternative definition of the derivative at 2 is given by
\[
\lim_{h \to 0} \frac{f(2 + h) - f(2)}{h}.
\]

Problem 6.9 Using Figure 6.7, find:

(a) \( \lim_{x \to 1} f(x) \)  
(b) \( \lim_{x \to 0^+} f(x) \)  
(c) \( \lim_{x \to 0} f(x) \)  
(d) \( \lim_{x \to 0^-} f(x) \)  
(e) \( \lim_{h \to 0} \frac{f(1.5 + h) - f(1.5)}{h} \)  
(f) \( \lim_{h \to 0} \frac{f(2 + h) - f(2)}{h} \)  
(g) \( \int_1^2 f \)  
(h) \( \lim_{h \to 0^-} \frac{f(2 + h) - f(2)}{h} \)  
(i) \( \frac{f(1 + h) - f(1)}{h} \) when \( h = -\frac{1}{2} \)

Figure 6.7: Graph of \( f \)
Rigor

The intuition of a functional limit must now be captured in a rigorous, logical statement, just as was done for sequential limits. There is not a single correct way to do this; there are many equivalent logical statements that will serve the purpose. The definition of a limit given in this book might appear to be different than the definition found in standard calculus texts, but we will prove that our definition is logically equivalent; we prove that \( \lim_{x \to a} f(x) = L \) is true according to our definition if and only if \( \lim_{x \to a} f(x) = L \) is true according to the definition found in standard calculus texts. The advantage of the definition we use is that it lets us draw on the theory of sequences that we have already established.

Recall that \( (x_i) \rightarrow a \) means the sequence \( (x_i) \) converges to \( a \). In such a situation it is possible that some of the terms of the sequence are equal to \( a \). Indeed, if \( x_i = a \) for all \( i \) then \( (x_i) \rightarrow a \). In the definition of a limit we want to approach \( a \) with numbers and see how the function behaves near \( a \), but we are not at all concerned with how the function behaves at \( a \) itself. In fact, we would like \( \lim_{x \to a} f(x) \) to be defined even if \( f \) is not defined at \( x = a \). For this reason we would like to introduce the following notation: we will write \( (x_i) \leftrightarrow a \) to indicate that \( (x_i) \rightarrow a \) and \( x_i \neq a \) for all \( i \in \mathbb{N} \). For the discussion that follows we will also require that all sequences reside in the domain of \( f \), so that given such a sequence \( (x_i) \) of inputs, the sequence \( (f(x_i)) \) of outputs is defined. Finally, in order to talk about the limit of a function \( f \) at \( x = a \) it is necessary to assume that there exists a sequence in the domain of \( f \) such that \( (x_i) \leftrightarrow a \), so we assume this also.

**Definition 6.1** We write \( \lim_{x \to a} f(x) = L \) when the following is true:

\[
\text{if } (x_i) \leftrightarrow a \text{ then } (f(x_i)) \rightarrow L.
\]

What our definition is saying is that all sequences in the domain that converge to the number \( a \) must be taken by \( f \) to sequences that converge to \( L \).

Figure 6.8 illustrates one sequence that converges to \( a \) (on the horizontal axis) and you can see by the graph that this sequence is mapped to a new sequence (on the vertical axis) that converges to \( L \). To prove that \( \lim_{x \to a} f(x) = L \) one begins by assuming they have an arbitrary sequence \( (x_i) \) for which \( (x_i) \leftrightarrow a \), and then prove that \( (f(x_i)) \rightarrow L \). To prove that \( \lim_{x \to a} f(x) \neq L \) one needs to give a counterexample to the if–then statement in Definition 6.1; exhibit a specific sequence for which \( (x_i) \leftrightarrow a \), and then show that, for this specific example, \( (f(x_i)) \) does not converge to \( L \).

**Problem 6.10** Prove that \( \lim_{x \to -2} f(x) \neq 3 \) if \( f \) is the function whose graph appears in Figure 6.9.

We now have a logical statement that defines what it means for \( f(x) \rightarrow L \) as \( x \rightarrow a \), and this definition has withstood a test of time; mathematicians are comfortable that this statement accurately reflects our intuition of a functional limit. From now on, when we say that the limit of \( f \) exists as \( x \) approaches \( a \), or if
we say \( \lim_{x \to a} f(x) \) exists, we mean that the if–then statement in Definition 6.1 is true for some number \( L \). If there does not exist a number \( L \) that makes the if–then statement in Definition 6.1 true we say that \( \lim_{x \to a} f(x) \) does not exist.

It happens in Figure 6.9 that \( \lim_{x \to -2} f(x) \) does not exist, since the left and right limits are not equal. This is not a proof that \( \lim_{x \to -2} f(x) \) does not exist, but it is a statement that we should be able to establish with a sequence of logical deductions beginning with Definition 6.1. The first thing to do is to provide a logical definition that captures the intuition of one-sided limits. This is done by slightly modifying Definition 6.1.

Problem 6.11 Give a logical definition of

\[
\lim_{x \to a^+} f(x) = L \quad \text{and} \quad \lim_{x \to a^-} f(x) = L.
\]
Problem 6.12 Prove that \( \lim_{x \to a} f(x) = L \) if and only if
\[
\lim_{x \to a^+} f(x) = L \quad \text{and} \quad \lim_{x \to a^-} f(x) = L.
\]

Recall that an “if and only if” statement is really two if–then statements, and both need to be proved. As it often happens, one of the if–then statements is much easier to prove than the other; the easier one in Problem 6.12 is “if \( \lim_{x \to a} f(x) = L \) then \( \lim_{x \to a^+} f(x) = L \) and \( \lim_{x \to a^-} f(x) = L \).”

We now wish to compare our definition of a functional limit with the definition found in most calculus books. The intent is to prove that the two definitions are equivalent, but until the equivalence is established we will use a different symbolism for the limit defined below in order to keep the two definitions separate.

Definition 6.2 We write \( \lim_{x \to a} f(x) = L \) when the following is true; for every \( \varepsilon > 0 \) there exists a \( \delta > 0 \) such that if \( 0 < |x - a| < \delta \) then \( |f(x) - L| < \varepsilon \).

This definition, presented as it is found in most Calculus textbooks, is a good illustration of how mathematicians will sometimes explicitly quantify variables (for every quantifies \( \varepsilon \) and there exists quantifies \( \delta \)) and then they will slip an unquantified variable into an if–then statement (we really mean for all \( x \in \mathbb{R} \), if \( 0 < |x - a| < \delta \) then \( |f(x) - L| < \varepsilon \)). It is possible to present the statement as a nest of if–then statements, which exaggerates the failure to quantify the for all variables, but also may clarify how one should prove or disprove the statement. Such a presentation goes as follows: “if \( \varepsilon > 0 \) then there exists a \( \delta > 0 \) that makes the following a true if–then statement: if \( 0 < |x - a| < \delta \) then \( |f(x) - L| < \varepsilon \).” When written this way you see how to prove that \( \lim_{x \to a} f(x) \neq L \); you exhibit a counterexample! You should find an \( \varepsilon \) that is positive but for which the conclusion is false. Saying the conclusion is false for a specific \( \varepsilon \) means for every \( \delta \) there is a counterexample to the statement “if \( 0 < |x - a| < \delta \) then \( |f(x) - L| < \varepsilon \).” This is illustrated in the proof of the forthcoming Proposition 6.1.

You can find lengthy expositions on the geometric meaning of this logical statement (in terms of the graph of the function \( f \)) in virtually every standard calculus text, so we will not give another one here. The reader is strongly encouraged to look up one or more of these expositions and try to reconcile for themselves why this logical statement accurately captures the intuition of a limit.

We have a logical statement given in Definition 6.1 that, when true, defines the meaning of \( \lim_{x \to a} f(x) = L \). We have a second logical statement given in Definition 6.2 that, when true, defines the meaning of \( \lim_{x \to a} f(x) = L \). What can be proved is that the logical statements in Definition 6.1 and 6.2 are true at exactly the same time; that is \( \lim_{x \to a} f(x) = L \) if and only if \( \lim_{x \to a} f(x) = L \). The consequence of this is any statement that can be proved true using a bunch of axioms and Definition 6.1 can also be proved true using the same bunch of axioms and Definition 6.2, and vice versa! We say that the two definitions are
equivalent in this situation, since the same collection of provable statements flows out of either definition.

The task of proving the equivalence of these two definitions boils down to proving two if–then statements true. Here is a sample of how one of the if–then statements is proved.

**Proposition 6.1** If \( \lim_{x \to a} f(x) = L \) then \( \lim_{x \to a} f(x) = L \).

**Proof.** We will prove the contrapositive of this if–then statement; i.e. we intend to prove “if \( \lim_{x \to a} f(x) \neq L \) then \( \lim_{x \to a} f(x) \neq L \)”. Assume the hypothesis is true, i.e. that \( \lim_{x \to a} f(x) \neq L \). By looking up the definition given above we find that we are assuming it is not true that

if \( \varepsilon > 0 \) then there exists a \( \delta > 0 \) that makes a true if–then statement:

if \( 0 < |x - a| < \delta \) then \( |f(x) - L| < \varepsilon \).

Thus there is a counterexample; and example of an \( \varepsilon_0 > 0 \) for which the conclusion is false. Saying the conclusion is false is the assertion that “there is no \( \delta > 0 \) that makes a true if–then statement” or more to the point, “for every \( \delta > 0 \) there is a counterexample to the statement: if \( 0 < |x - a| < \delta \) then \( |f(x) - L| < \varepsilon_0 \). In particular there is a counterexample when \( \delta = 1 \); i.e. there exists \( x_1 \) such that \( 0 < |x_1 - a| < 1 \) but \( |f(x_1) - L| \geq \varepsilon_0 \). There is also a counterexample when \( \delta = \frac{1}{2} \) and when \( \delta = \frac{1}{3} \) giving us \( x_2 \) and \( x_3 \). In general, if \( \delta = \frac{1}{n} \) then there is a number \( x_n \) such that \( 0 < |x_n - a| < \frac{1}{n} \) but \( |f(x_n) - L| \geq \varepsilon_0 \). Since \( 0 < |x_n - a| < \frac{1}{n} \) it must follow that \( x_n \to a \) (by the squeeze theorem), but \( f(x_n) \) does not converge to \( L \) since the interval \((L - \varepsilon_0, L + \varepsilon_0)\) contains absolutely none of the numbers \( f(x_n) \). Thus we have a counterexample to the if–then statement

if \( x_n \to a \) then \( f(x_n) \to L \)

so that \( \lim_{x \to a} f(x) \neq L \).

\( \triangle \)

**Problem 6.13** Prove the following is true:

if \( \lim_{x \to a} f(x) = L \) then \( \lim_{x \to a} f(x) = L \).

Now that we know \( \lim_{x \to a} f(x) = L \) is equivalent to \( \lim_{x \to a} f(x) = L \) we can stop writing \( \lim_{x \to a} f(x) = L \) and always write \( \lim_{x \to a} f(x) = L \). Notice that we now have two ways to prove or disprove \( \lim_{x \to a} f(x) = L \); we may either use the logical statement given in Definition 6.1 or the one given in Definition 6.2. Most of the proofs are easier using Definition 6.1 since this statement allows us to use the information accumulated about sequences.

**Problem 6.14** Let \( f \) be the function that outputs 0 for all numbers except numbers of the form \( \frac{1}{n} \) with \( n \in \mathbb{N} \). For these numbers assume that \( f \) outputs 1; i.e. \( f(\frac{1}{n}) = 1 \) for all \( n \in \mathbb{N} \). Draw the graph of \( f \) and prove that \( \lim_{x \to 0} f(x) \neq 0 \).
The next problem says that functional limits are unique. This problem can be solved using a related problem about sequences that said sequential limits are unique.

**Problem 6.15** If \( \lim_{x \to a} f(x) = L \) and \( \lim_{x \to a} f(x) = M \) then \( L = M \).

Recall that a function \( f \) is continuous at \( a \) when \( \lim_{x \to a} f(x) = f(a) \). This definition has now been made logically precise since it is phrased in terms of a limit and we have just finished giving two equivalent logical statements that define what it means to say \( \lim_{x \to a} f(x) = L \). One way to prove that \( f \) is continuous at \( a \) is to assume that \( (x_i) \to a \) and use this assumption (and the definitions) to prove that \( f(x_i) \to f(a) \). To prove that \( f \) is not continuous at \( a \) (i.e. that \( f \) is discontinuous at \( a \)) you would look for a counterexample. You should find an example of a sequence such that \( (x_i) \to a \) but \( f(x_i) \) fails to converge to \( f(a) \).

Of all the functions that abound it is a fact that the continuous functions form a small minority. On the other hand, the continuous functions are somewhat mathematically tractable and much of calculus involves what can be said when continuity appears in a hypothesis. Two results about sequences that we proved were really continuity assertions. The first says if \( (s_n) \to s \) and \( (r_n) \to r \) then \( (s_n + r_n) \to s + r \), and this may be expressed briefly by saying “addition is continuous.” The second assertion says “multiplication is continuous”. When these two facts are combined it is possible to show that all functions that operate by adding and multiplying are continuous; these functions constitute the polynomials. Proving the continuity of polynomials is what we are about to do now. The simplest of all polynomials are the polynomials of degree 0; these are the constant functions.

**Problem 6.16** Prove that polynomials of degree 0 are continuous.

Functions whose graphs are straight lines are particularly nice because they can easily be expressed with a formula. Recall that if the line has a slope \( m \) and intersects the vertical axis at the number \( b \), then the function \( f \) that has this line as its graph will output

\[
f(x) = mx + b
\]

when \( x \) is input. When \( m \) is not zero these functions (we get a different one for every choice of \( m \) and \( b \)) are the polynomials of degree 1.

**Problem 6.17** Prove that polynomials of degree 1 are continuous.

There is really no reason for writing \( mx + b \) instead of \( a_1x + a_0 \), as long as everyone understands that the variable is \( x \) and the \( a \)'s are constants. The next family of functions that we will deal with are the polynomials of degree 2, which the reader has probably seen expressed by the formula

\[
ax^2 + bx + c,
\]
for which there is a quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

that gives the roots of the polynomial. If we pick specific values for $a$, $b$, and $c$, with $a \neq 0$, and if $f$ is the function that outputs $ax^2 + bx + c$ when $x$ is input, then $f$ is a polynomial of degree 2. Remember that the graphs of such functions look like parabolas. The quadratic formula above tells you what inputs give an output of zero, and those inputs are called the roots of the polynomial.

**Problem 6.18** Prove that polynomials of degree 2 are continuous.

Once again there is no reason to write $ax^2 + bx + c$ instead of $a_2x^2 + a_1x + a_0$, but there is a good reason to write the latter instead of the former; the latter extends naturally to a formula the describes a polynomial of arbitrary degree. A polynomial of degree $n$ is a function $f$ that outputs a number of the form

$$a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0$$

when $x$ is input, and with $a_n \neq 0$. The next problem would likely ask you to prove polynomials of degree 3 are continuous. Instead of filling the rest of this book with the sequels of Problem 6.18, we will now illustrate how to prove that polynomials of degree $n$ are continuous for all $n \in \mathbb{N}$ using a proof technique called mathematical induction.

Assume $s(n)$ denotes the statement “all polynomials of degree $n$ are continuous.” You proved $s(1)$ is true in Problem 6.17 and $s(2)$ was proved true in Problem 6.18. A proof by mathematical induction involves two steps:

1. Prove $s(1)$ is true (this is called the base case).

2. Prove the following if–then statement is true for all $n \geq 2$; if $s(n - 1)$ is true then $s(n)$ is true (this is called the induction step).

Note that the second step does not state that $s(n)$ is true for all $n \in \mathbb{N}$, it says $s(n)$ is true when its predecessor $s(n - 1)$ is true. Only when both step 1 and step 2 are established is it known that $s(n)$ is true for all $n \in \mathbb{N}$. This is because step 1 tells us $s(1)$ is true. Then step 2 applies with $n = 2$ to tell us $s(2)$ is true. Then step 2 applies again with $n = 3$ to tell us $s(3)$ is true. Then step 2 applies again, and again, and like dominoes falling, we have $s(n)$ true for all $n \in \mathbb{N}$.

**Problem 6.19** Use induction to prove that polynomials of degree $n$ are continuous for all $n \in \mathbb{N}$.

If $s(n)$ is any collection of logical statements, one statement for each $n \in \mathbb{N}$, mathematical induction will work to prove $s(n)$ is true for all $n \in \mathbb{N}$, as long as you can prove both the base case and the induction step.

**Problem 6.20** Prove that the sum of the first $n$ natural numbers equals $\frac{n(n+1)}{2}$, i.e. prove $1 + \ldots + n = \frac{n(n+1)}{2}$ for all $n \in \mathbb{N}$. 

What is intended in the equation above is the sequence of equations
\[
\begin{align*}
1 &= \frac{1(1+1)}{2} \quad (n = 1) \\
1 + 2 &= \frac{2(2+1)}{2} \quad (n = 2) \\
1 + 2 + 3 &= \frac{3(3+1)}{2} \quad (n = 3) \\
\vdots & \quad \vdots
\end{align*}
\]

We bring this seemingly extraneous problem into the discussion to illustrate how an induction proof works. This problem turns out to be not so extraneous after all, however, as we will see when we put it to use in the chapter on integration.

Given two (or more) functions \( f \) and \( g \), there are various ways of combining them to form new functions. If both \( f \) and \( g \) output real numbers then \( f + g \) and \( fg \) may be defined \textit{pointwise}; \( f + g \) returns an output of \( f(x) + g(x) \) when \( x \) is input, and \( fg \) returns \( f(x)g(x) \). For example, if \( f \) is the squaring function and \( g \) is the cubing function, then \( f + g \) is the polynomial that outputs \( x^2 + x^3 \) when \( x \) is input. The function \( fg \) returns \( x^5 \) when \( x \) is input.

\textbf{Problem 6.21} Assume \( f_0, f_1, f_2, \ldots, f_n \) are continuous functions. Use induction to prove that \( f_0 + f_1 + f_2 + \ldots + f_n \) and \( f_0f_1f_2\ldots f_n \) are continuous.

There is another important method of combining functions called composition. The idea of composition is to use the output of one function as the input of the next function. If \( f \) and \( g \) are functions then \( f \circ g \) denotes the function that outputs \( f(g(x)) \) when \( x \) is input (see Figure 6.10). Thus if \( f \circ g \) is thought

of as an input–output machine, upon inputting a number \( x \) the machine would first input \( x \) into \( g \), then the machine would feed the output of \( g \) into the input of \( f \), and \( f \) would return \( f(g(x)) \). If \( f \) is the squaring function and \( g \) is the polynomial that returns \( x^2 + 5 \) when \( x \) is input, then \( f \circ g \) would return \((x^2 + 5)^2\) when \( x \) is input, while \( g \circ f \) would return \( x^4 + 5 \) when \( x \) is input. You see that \( f \circ g \) is a different function than \( g \circ f \), i.e. \( f \circ g \neq g \circ f \), unlike multiplication of numbers in a field. There is a term for this; we describe the inequality by saying function composition is not commutative.

In order for \( f \circ g \) to be defined it is necessary for \( g(x) \) to be in the domain of \( f \) for every \( x \) in the domain of \( g \). We assume this to be the case in the next problems.
Problem 6.22 Prove the following is true; if $f$ and $g$ are continuous then $f \circ g$ is continuous.

Problem 6.23 Prove the following is true; if $f_0$, $f_1$, $\ldots$, $f_n$ are all continuous then $f_0 \circ f_1 \circ \ldots \circ f_n$ is continuous.

Notice that Problem 6.22 is the base case for a proof by induction of Problem 6.23.
Chapter 7

Theorems About Continuous Functions.

There are two fundamental theorems of continuous functions that lie at the heart of calculus and out of which flow the deepest theorems of differentiation and integration: the extreme value theorem and the intermediate value theorem. The title of this chapter leads you to think these theorems are about continuous functions, but the theorems also involve the type of domain the functions have in a very critical way. Indeed, one of the facts we would like to stress is that both of the theorems are false if the domain is a closed interval of rational numbers. There are very special properties about closed intervals of real numbers that get used in the proof of these theorems. We begin with the extreme value theorem.

Theorem 7.1 (Extreme Value Theorem.) If \( f : [a, b] \rightarrow \mathbb{R} \) is a continuous function then there exists \( x_{\text{max}} \in [a, b] \) such that \( f(x) \leq f(x_{\text{max}}) \) for all \( x \in [a, b] \).

To prove the extreme value theorem it is necessary to develop enough mathematical machinery to reveal a very subtle property of closed intervals in \( \mathbb{R} \). We present the following problems before developing this machinery in the hope of providing more familiarity with the extreme value theorem before plunging into its proof.

Problem 7.1 Prove that the following is false;

\[
\text{if } f : \mathbb{R} \rightarrow \mathbb{R} \text{ is continuous then there exists } x_{\text{max}} \in \mathbb{R} \text{ such that } f(x) \leq f(x_{\text{max}}) \text{ for all } x \in \mathbb{R}.
\]

Problem 7.2 Prove that the following is false;

\[
\text{if } f : (a, b) \rightarrow \mathbb{R} \text{ is continuous then there exists } x_{\text{max}} \in (a, b) \text{ such that } f(x) \leq f(x_{\text{max}}) \text{ for all } x \in (a, b).
\]

Problem 7.3 Prove that the following is false;
if \( f : [a, b] \rightarrow \mathbb{R} \) is a function then there exists \( x_{\text{max}} \in [a, b] \) such that \( f(x) \leq f(x_{\text{max}}) \) for all \( x \in [a, b] \).

All of the problems above involve making slight changes in the hypothesis of the extreme value theorem and seeing that the resulting if-then statement is no longer true. Most of the changes we made were to replace the closed interval domain with some other domain, like an open interval or all of \( \mathbb{R} \). Just because open intervals don’t work in the hypothesis does not mean that closed intervals are the only type of set that will work in the hypothesis. Use the extreme value theorem (even though we have not proved it yet) to do the following problem.

**Problem 7.4** Prove that the following is true:

if \( f : [a, b] \cup [c, d] \rightarrow \mathbb{R} \) is a continuous function then there exists \( x_{\text{max}} \in [a, b] \cup [c, d] \) such that \( f(x) \leq f(x_{\text{max}}) \) for all \( x \in [a, b] \cup [c, d] \).

We now intend to build the mathematical machinery needed in the proof of the extreme value theorem, and we will use sequences to build this machinery. If you are given a sequence \((s_1, s_2, \ldots)\) then you can use it to construct new sequences by selecting some of the terms of the given sequence. For example, you might decide to select the even terms, which results in the sequence \((s_2, s_4, \ldots)\), or you might select the terms that are multiples of one hundred, resulting in the sequence \((s_{100}, s_{200}, \ldots)\). These two sequences are called *subsequences* of the sequence \((s_1, s_2, \ldots)\). As a concrete example, note that both \((\frac{1}{2}, \frac{1}{4}, \ldots)\) and \((\frac{1}{100}, \frac{1}{200}, \ldots)\) are subsequences of the sequence \((1, \frac{1}{2}, \frac{1}{3}, \ldots)\). You can see that the method of constructing a subsequence amounts to selecting the indices (the subscripts); in the first instance we are selecting the indices \(2 < 4 < 6 < \cdots\) and in the second instance we select the indices \(100 < 200 < 300 < \cdots\). This observation leads us to the *formal* definition of a subsequence.

**Definition 7.1** A subsequence of the sequence \((s_n)\) is a sequence of the form \((s_{n_1}, s_{n_2}, s_{n_3}, \ldots)\) where \(n_1 < n_2 < n_3 < \cdots\) are the selected indices.

If you start with a sequence that does not converge it is quite possible that some subsequence does converge. For example, the sequence

\[(s_n) = (1, -1, 1, -1, \ldots)\]

does not converge but the subsequence \((s_2, s_4, \ldots)\) converges to \(-1\). On the other hand, there are examples of sequences that fail to converge and every one of their subsequences also fails to converge.

**Problem 7.5** Show that every subsequence of \((1, 2, 3, \ldots)\) fails to converge.

We are going to prove some facts about sequences and their subsequences that will be used to prove the extreme value theorem. When an if-then statement is true and it is used as a building block for a bigger proof, mathematicians will often call the if-then statement a *lemma*. The next problem is really a lemma used in the proof of Lemma 7.1!
Problem 7.6 Assume \((s_n)\) is a sequence with the property that to every index \(k\) corresponds an index \(n\) such that \(n > k\) and \(s_n \geq s_k\). Prove that \((s_n)\) has a subsequence that is increasing.

Lemma 7.1 If \((s_n)\) is a sequence in \(\mathbb{R}\) then \((s_n)\) either has an increasing subsequence or a decreasing subsequence.

Proof. There are two mutually exclusive cases that we can consider; either \((s_n)\) has an increasing subsequence (in which case the proof is done) or it doesn’t have an increasing subsequence, in which case we need to prove there is a decreasing subsequence. Thus the second case requires our attention: assume there is no increasing subsequence of \((s_n)\). We need to select indices \(n_1 < n_2 < n_3 < \cdots\) for which \(s_{n_1} \geq s_{n_2} \geq s_{n_3} \geq \ldots\). There must exist an index \(n_1\) for which \(s_n < s_{n_1}\) for all \(n \geq n_1\) (this follows from Problem 7.6). The same reasoning implies that there must exist a second index \(n_2\) such that \(n_1 < n_2\) and \(s_n < s_{n_2}\) for all \(n \geq n_2\). Continuing in this way we obtain a subsequence such that \(s_{n_1} \geq s_{n_2} \geq s_{n_3} \geq \ldots\), which completes the proof.

\[\triangle\]

The subtle property of closed intervals that is at the heart of the extreme value theorem is the content of the next lemma.

Lemma 7.2 If \((s_n)\) is a sequence in a closed interval \([a, b]\) then there exists a subsequence of \((s_n)\) that converges to a number \(s \in [a, b]\).

Proof. By Lemma 7.1 \((s_n)\) either has an increasing subsequence or a decreasing subsequence. If it has an increasing subsequence then this subsequence converges to a limit \(s\) by Axiom (14), since every bounded increasing sequence converges. Since \(a \leq s_n \leq b\) for all \(n \in \mathbb{N}\) we must have \(s \in [a, b]\). If the sequence has a decreasing subsequence then we get the existence of a limit \(s\) from Problem 5.11.

\[\triangle\]

Lemma 7.3 If \(f\) is a continuous function on \([a, b]\) then the set

\[
\{ f(x) \mid x \in [a, b] \}
\]

has an upper bound.

Proof. We will prove the contrapositive of the statement; i.e. we will prove that if \(\{ f(x) \mid x \in [a, b] \}\) has no upper bound then \(f\) is not continuous on \([a, b]\). Assume the set \(\{ f(x) \mid x \in [a, b] \}\) has no upper bound. Since \(1\) is not an upper bound there is some element \(f(s_1)\) of the set bigger, and since \(2\) is not an upper bound there’s another element \(f(s_2)\) of the set with \(f(s_2) > 2\). Continuing this process we obtain a sequence \((s_n)\) in \([a, b]\) for which \(f(s_n) > n\).
By Lemma 7.2 there is a subsequence \((s_{n_1}, s_{n_2}, s_{n_3}, \ldots)\) of \((s_n)\) that converges to a number \(s \in [a, b]\). But, just as in Problem 7.5, there is no convergent subsequence of \((f(s_1), f(s_2), f(s_3), \ldots)\) so we have found a convergent sequence \((s_{n_1}, s_{n_2}, s_{n_3}, \ldots) \to s\) for which \((f(s_{n_1}), f(s_{n_2}), f(s_{n_3}), \ldots)\) does not converge. We conclude that \(f\) is not continuous at \(s\), and hence \(f\) is not continuous on \([a, b]\).

\[\triangle\]

We will now assemble the lemmas into a proof of the extreme value theorem!

**Proof of Extreme Value Theorem.** The set \(\{f(x) \mid x \in [a, b]\}\) has an upper bound by Lemma 7.3, and it is certainly not the empty set. Thus this set has a least upper bound, which we will denote \(\alpha\). Since \(\alpha\) is the least upper bound there must exist elements \(f(s_n)\) of the set such that \(\alpha - \frac{1}{n} < f(s_n)\). Thus the sequence \(f(s_n)\) converges to \(\alpha\) by the squeeze theorem, and moreover every subsequence of \(f(s_n)\) converges to \(\alpha\). By Lemma 7.2 there is a subsequence \((s_{n_1}, s_{n_2}, s_{n_3}, \ldots)\) of \((s_n)\) that converges to a number \(s \in [a, b]\). Since \(f\) is continuous at \(s\) we have that \((f(s_{n_1}), f(s_{n_2}), f(s_{n_3}), \ldots)\) converges to \(f(s)\). Since this sequence also converges to \(\alpha\) we conclude from Problem 5.12 that \(f(s) = \alpha\). If \(x_{\text{max}} = s\) we have that \(f(x) \leq f(x_{\text{max}}) = \alpha\) for all \(x \in [a, b]\), since \(\alpha\) is the least upper bound of \(\{f(x) \mid x \in [a, b]\}\).

\[\triangle\]

There are many variants of the extreme value theorem. One of the variants asserts that minimum values are attained. Knowing how to convert a function \(f\) into a function \(g\) so that \(f(x) \leq f(y)\) if and only if \(g(y) \leq g(x)\) gives an easy way of proving one form of the extreme value theorem from the other.

**Problem 7.7** Use the extreme value theorem to prove the following is true;

if \(f : [a, b] \to \mathbb{R}\) is continuous then there exists \(x_{\text{min}} \in [a, b]\) such that \(f(x_{\text{min}}) \leq f(x)\) for all \(x \in [a, b]\).

**Problem 7.8** Draw the graph of a function \(f : [0, 1] \to \mathbb{R}\) that satisfies the conclusion of the extreme value theorem but does not satisfy the hypothesis.

**Problem 7.9** Prove that the converse of the extreme value theorem is false.

**Problem 7.10** Is it possible to draw the graph of a function that satisfies the hypothesis of the extreme value theorem but not the conclusion? If it is, do it. Otherwise, say why it can not be done.

We will now look at the second important theorem about continuous functions, the intermediate value theorem. It has virtually the same hypothesis as the extreme value theorem; a continuous function with a closed interval of real numbers as a domain. It is interesting that very different properties of a closed interval \([a, b]\) make each theorem true. The property of \([a, b]\) that makes the
extreme value theorem true is called compactness and the property of \([a, b]\) used to prove the intermediate value theorem is called connectedness. These concepts are thoroughly covered in a course in real analysis or topology.

**Theorem 7.2 (Intermediate Value Theorem.)** If \(f : [a, b] \to \mathbb{R}\) is continuous and \(f(a) < 0 < f(b)\) then there exists \(x \in (a, b)\) such that \(f(x) = 0\).

**Problem 7.11** Draw the graph of a function that illustrates the meaning of the intermediate value theorem, with \(a, b,\) and \(x\) labeled.

**Problem 7.12** Draw the graph of a function that satisfies the conclusion of the intermediate value theorem but not the hypothesis.

There is a more general statement that deserves to be called the intermediate value theorem. It is often the case that one name refers to a family of theorems that are all so closely related that each theorem follows easily from any of the others.

**Problem 7.13** Use the intermediate value theorem to prove the following true:

\(\text{if } f : [a, b] \to \mathbb{R} \text{ is a continuous function and } y \text{ is between } f(a) \text{ and } f(b) \text{ then there exists } x \in [a, b] \text{ such that } f(x) = y.\)

Problem 7.13 is more general than the intermediate value theorem because the intermediate value theorem is a special case; let \(y = 0\). To prove Problem 7.13 from the intermediate value theorem you have to consider two cases; do a proof when you assume \(f(a) < y < f(b)\), then do another proof for \(f(b) < y < f(a)\). To use the extreme value theorem you have to construct a function to apply its hypothesis to; try \(g(x) = f(x) - y\) for one of the cases.

In Problem 7 we saw that the extreme value theorem is still true when the domain \([a, b]\) is replaced with a union \([a, b] \cup [c, d]\). The property of \([a, b]\) needed to prove the intermediate value theorem is no longer available with a union, and consequently the analogue of Problem 7 for the intermediate value theorem is obtained with a counterexample. Assume that \(a < b < c < d\) for the following problem

**Problem 7.14** Prove the following is false: If \(f : [a, b] \cup [c, d] \to \mathbb{R}\) is continuous and \(f(a) < 0 < f(d)\), then there exists \(x \in [a, b] \cup [c, d]\) such that \(f(x) = 0\).

The intermediate value theorem can be viewed as a generalization of the assertion that every positive real number has a square root. This assertion is true primarily because of the completeness Axiom (14), and it is simply false if the real numbers is replaced with an arbitrary ordered field. A glaring counterexample is obtained when the ordered field \(\mathbb{Q}\) replaces \(\mathbb{R}\). The fact is that 2 has no square root in \(\mathbb{Q}\)! The proof of this fact is an argument by contradiction. Either 2 does or does not have a square root in \(\mathbb{Q}\); if you assume 2 does have a square root in \(\mathbb{Q}\), and from this assumption you are able to deduce a contradiction, then the assumption that 2 has a square root must have been false! You then conclude that 2 can not have a square root in \(\mathbb{Q}\). Let’s try it;
assume 2 has a square root in Q. This means there is a rational number \( \frac{m}{n} \in \mathbb{Q} \) whose square is 2, and we may assume that either \( m \) or \( n \) is odd (by canceling as many twos as possible). Since

\[
\frac{m^2}{n^2} = 2
\]

we have \( m^2 = 2n^2 \). It follows that \( m^2 \) is even (since it is two times an integer).

**Problem 7.15** Prove the following is true:

if \( m \in \mathbb{Z} \) and \( m^2 \) is even then \( m \) is even.

Thus \( m \) is even. It follows that \( m^2 \) is divisible by 4, hence \( 2n^2 \) is divisible by 4 and \( n^2 \) is even. Thus \( n \) is even by Problem 7.15. Here’s our contradiction; we started with either \( m \) or \( n \) odd and we have just finished proving them both even. Thus our original assumption, that 2 has a square root in \( \mathbb{Q} \), must be wrong. We thus conclude that 2 has no square root in \( \mathbb{Q} \).

**Problem 7.16** Use the intermediate value theorem to prove the following:

there exists \( x \in [1, 3] \) such that \( x^2 = 2 \).

Knowing that 2 has no square root in \( \mathbb{Q} \) gives you a way to do the following problem (which involves constructing a counterexample). Assume for the next two problems that \( [a, b] \) denotes a closed interval of rational numbers.

**Problem 7.17** Prove that the following is false:

if \( f : [a, b] \rightarrow \mathbb{Q} \) is continuous and \( f(a) < 0 < f(b) \) then there exists \( x \in [a, b] \) such that \( f(x) = 0 \).

**Problem 7.18** Prove that the following is false:

if \( f : [a, b] \rightarrow \mathbb{Q} \) is continuous then there exists \( x \in [a, b] \) such that \( f(w) \leq f(x) \) for all \( w \in [a, b] \).

When we first introduced the intermediate value theorem we discussed what is meant by a more general statement. Much of what mathematicians do involves asking questions which arise as generalizations of statements that are known to be true. The mathematician will try to prove the new generalized statement, and if no success is met the mathematician will soon start to look for a counterexample.

**Problem 7.19** Generalize the following statement as much as you can;

2 has a square root in \( \mathbb{R} \).

**Problem 7.20** Prove or disprove your generalization given in the previous problem.
Chapter 8

Differentiation

Recall the intuitive idea of a derivative. If you have a function \( f \) whose graph at the point \((a, f(a))\) has the property that upon magnification it approaches a line, then we say \( f \) is differentiable at \( a \) and the derivative at \( a \), denoted \( f'(a) \), is the slope of that line. The value of \( f'(a) \) may be obtained as a limit of slopes of secant lines, and this then gives us a way of capturing the intuitive idea with a rigorous logical statement.

**Definition 8.1** The derivative of \( f \) at \( a \) is defined by

\[
f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.
\]

If the limit defining \( f'(a) \) exists we say that \( f \) is differentiable at \( a \), otherwise we say that the derivative doesn’t exist at \( a \).
We say that \( f \) is differentiable on a set \( E \) if \( f \) is differentiable at every \( x \in E \).

**Problem 8.1** Give a rigorous definition that captures the intuition of a left derivative and a right derivative.

**Problem 8.2** Draw the graph of a function whose left and right derivatives exist at \( a = 1 \), but are not equal.

If you have a function \( f \) that can be defined using a formula, such as a polynomial, then it may be possible to use the definition of \( f' \) to find a formula for \( f' \). Note that \( f' \) is not a number, it is a function. Only after receiving an input will \( f' \) return a number; \( f'(a) \) denotes the output obtained when \( a \) is input. A formula for \( f' \) is a mathematical expression that tells you what the output is in terms of the input. For example, suppose \( f \) is the squaring function; thus \( f \) can be described by the formula \( f(x) = x^2 \), which means \( f \) returns \( x^2 \) when \( x \) is input. Thus \( f \) returns \( a^2 \) when \( a \) is input, which we write symbolically as \( f(a) = a^2 \). You can substitute these expressions into the definition of \( f' \) to get

\[
f'(a) = \lim_{x \to a} \frac{x^2 - a^2}{x - a}.
\]

If you want to know what \( f'(2) \) is you would have to evaluate the limit

\[
f'(2) = \lim_{x \to 2} \frac{x^2 - 4}{x - 2}.
\]

You probably remember that \( x^2 - 4 = (x - 2)(x + 2) \). If you substitute this into the numerator and then cancel the \( x - 2 \) terms you are left with a limit that can be calculated (since polynomials are continuous). If you followed this discourse you will already know that \( f'(2) = 4 \). You can use this reasoning to obtain the following generalization.

**Problem 8.3** Let \( f \) be the squaring function. Find a formula that gives \( f'(a) \) for any \( a \in \mathbb{R} \) and give a proof that your formula is correct.

We defined \( f'(a) \) to be the limit

\[
\lim_{x \to a} \frac{f(x) - f(a)}{x - a}.
\]

If instead of using \( x \) to label the number getting close to \( a \) we had instead labeled it \( a + h \) with \( h \) close to zero, the definition of \( f'(a) \) takes the form

\[
f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}.
\]

**Problem 8.4** Find a formula for

\[
\lim_{h \to 0} \frac{f(a + h) - f(a)}{h}
\]

when \( f \) is the constant function defined by the formula \( f(x) = 2 \) for all \( x \in \mathbb{R} \). Prove that your formula is correct.
Teaching experience has shown that students often mistake what \( f(a + h) \) denotes, particularly when \( f \) is a constant function. For example, in Problem 8.4 \( f(a + h) \neq 2 + h \).

**Problem 8.5** Prove that the derivative of any constant function is the constant zero function.

**Problem 8.6** Prove that the derivative of the function defined by the formula \( f(x) = x \) is the constant 1 function.

If \( f \) is a function defined by the formula \( f(x) = x^n \), then we have discovered what \( f' \) is as long as \( n \in \{0, 1, 2\} \). A generalization of what we have proved thus far would be a formula for \( f' \) that works for any \( n \in \mathbb{N} \). It turns out that

\[
f'(x) = nx^{n-1}
\]

for all \( n \in \mathbb{N} \), and a proof of this can be accomplished using mathematical induction. The idea is to write \( f \) as

\[
f(x) = xx^{n-1}
\]

and then use the induction hypothesis together with something called the product rule. The next problem is the first thing you need in order to prove the product rule. To do the problem you can write

\[
f(x) - f(a) = \frac{f(x) - f(a)}{x-a} (x-a)
\]

and then take the limit as \( x \) approaches \( a \).

**Problem 8.7** Prove the following is true; if \( f \) is differentiable at \( a \) then \( f \) is continuous at \( a \).

If you have two functions \( f \) and \( g \) then you can form the product function \( fg \) as described on Page 55. If we know the derivatives of both \( f \) and \( g \), then the product rule allows us to find the derivative of the product. There is a trick involved in proving the product rule and that is to write

\[
\frac{f(x)g(x) - f(a)g(a)}{x-a} = \frac{f(x)g(x) - f(a)g(x) + f(a)g(x) - f(a)g(a)}{x-a}
\]

Once this is done you can take the limit as \( x \) approaches \( a \) and, after a slight algebraic manipulation, the product rule emerges.

**Problem 8.8** Prove the following: if \( f'(a) \) and \( g'(a) \) exist, then

\[
(fg)'(a) = f'(a)g(a) + f(a)g'(a).
\]

**Problem 8.9** Prove that \( f'(a) = na^{n-1} \) for all \( n \in \mathbb{N} \) when \( f \) is the function defined by \( f(x) = x^n \).
Differentiation theorems can be divided into two types. There are the computational theorems that tell us how to compute derivatives, and there are the theoretical theorems that tell us how to deduce information about a function from properties of its derivative. The computational theorems are often referred to as rules. In the problems above you proved the product rule and the power rule. There is also a quotient rule and, the most important rule of all, the chain rule. An example of a theoretical theorem is the true if–then statement; “if $f'$ is the constant zero function then $f$ is a constant function.” The converse of this statement also happens to be true, as you proved in Problem 8.5. The difference between this statement and its converse is that the converse is a computational statement (it tells you what the derivative of a constant function is) while the statement itself is in the theoretical camp; it gives you information about the function knowing only properties of its derivative. Thus the statement and its converse fall into different branches of our division. The statement is also surprisingly difficult to prove (although you probably found proving the converse very easy).

There is a single theorem, called the mean value theorem, that is central to all of the theorems in the theoretical branch. Once this theorem is proved the other theorems in the theoretical branch may be deduced as consequences.

**Theorem 8.1** If $f$ is continuous on $[a, b]$ and differentiable on $(a, b)$ then there exists $x \in (a, b)$ where

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

The conclusion of the mean value theorem says there must be a place on the graph of $f$ where the tangent line is parallel to the secant line joining the endpoints. For the function whose graph appears in Figure 8.2 there are several values of $x$ where $f'(x)$ equals the slope of the line joining $(a, f(a))$ and $(b, f(b))$. 

![Figure 8.2:](image-url)
We will eventually have a proof of the mean value theorem, but not until the last problem of this chapter will that proof be complete. Before we prove the mean value theorem we will experiment with the meaning of its statement to acquire familiarity.

**Problem 8.10** Draw the graph of a function that satisfies the conclusion of the mean value theorem but not the hypothesis.

**Problem 8.11** Draw the graph of a function that satisfies neither the hypothesis nor the conclusion of the mean value theorem.

In the following problem you are asked to prove that a function $f$ is a constant function. One way to prove that $f$ is a constant function is to prove that $f(a) = f(b)$ for all $a, b \in \mathbb{R}$. To do the next problem you can let $a$ and $b$ denote arbitrary numbers with $a < b$ and then apply the mean value theorem on the interval $[a, b]$.

**Problem 8.12** Use the mean value theorem to prove the following; if $f'(x) = 0$ for all $x \in \mathbb{R}$ then $f$ is a constant function.

**Problem 8.13** If $F$ and $G$ are two differentiable functions with the same derivative then $F$ and $G$ differ by a constant.

A function $f$ is *increasing* if the following is a true if–then statement: if $a < b$ then $f(a) \leq f(b)$. In terms of the graph of $f$ this logical statement captures the picture of a graph that is constant or rising as you look from left to right. Similarly, a function is decreasing when the following is true; if $a < b$ then $f(b) \leq f(a)$. You should be able to use these definitions to prove that a function is both increasing and decreasing if and only if the function is constant. To prove the statement in the next problem you can mimic the solution of Problem 8.12.

**Problem 8.14** Use the mean value theorem to prove the following; if $f'(x) \geq 0$ for all $x \in \mathbb{R}$ then $f$ is increasing.
A special case of the mean value theorem is a statement that goes by the name of Rolle’s theorem. (Notice how being a special case is the opposite of being a generalization. We could just as well have said that the mean value theorem is a generalization of Rolle’s theorem.)

**Theorem 8.2** If \( g \) is continuous on \([a, b]\) and differentiable on \((a, b)\) and if \( g(a) = g(b) \) then there exists \( x \in (a, b) \) such that \( g'(x) = 0 \).

**Problem 8.15** Use the mean value theorem to prove Rolle’s theorem.

You should beware that we have not proved Rolle’s theorem yet. We also have not proved the if–then statements that appeared in Problems 8.12 and 8.14. What we have established is that these if–then statements are all logical consequences of the mean value theorem; so as soon as the mean value theorem has been proved so have these if–then statements.

**Problem 8.16** Draw the graph of a function that satisfies the hypothesis of the mean value theorem but not the hypothesis of Rolle’s theorem, and also satisfies the conclusion of both.

You have already seen several instances when we have a family of statements that are so closely related that any one can be proved easily after assuming any of the others. We are in that situation again because Rolle’s theorem is so closely related to the mean value theorem that it is possible to construct an easy proof of the mean value theorem if you assume that Rolle’s theorem is true. To do this proof you would first assume that Rolle’s theorem is true. Then, in order to prove the mean value theorem, you would assume its hypothesis was true; i.e. you are assuming you have a function \( f \) that is continuous on \([a, b]\) and differentiable on \((a, b)\). You can not use Rolle’s theorem yet because you are not assuming \( f(a) = f(b) \). The trick is to construct a function related to \( f \) that satisfies the hypothesis of Rolle’s theorem.

**Problem 8.17** Let \( g \) be the function described in Figure 8.4. Show that \( g \) satisfies the hypothesis of Rolle’s theorem, then apply the conclusion of Rolle’s theorem to \( g \) and see what that says about \( f \).

It is very important that you realize we have neither a proof of Rolle’s theorem or the mean value theorem yet. What we have established is that we need only prove one of the theorems and the other will then follow as a consequence. We will now set out to give a direct proof of Rolle’s theorem. Our proof of Rolle’s theorem requires considering two cases, one of which is very easy. The easy case is when the function is constant.

**Problem 8.18** Prove Rolle’s theorem in the case when \( g \) is a constant function.

The other case is when \( g \) is not constant, which itself breaks up into two cases. There’s the possibility that there exists \( w \in (a, b) \) such that \( g(a) < g(w) \) and then there’s the possibility that \( g(w) < g(a) \) for some \( w \in (a, b) \).
possible that both cases happen at the same time, but what is important for this proof is that one of the two cases must occur.) In each case the crucial step is the application of the extreme value theorem to find an extreme value that the function takes on \([a, b]\). Thus if you are in the first case you would say there exists \(x \in [a, b]\) where \(g(z) \leq g(x)\) for all \(z \in [a, b]\). The assumption in this case lets you deduce that \(x \in (a, b)\), and from this you can prove \(g'(x) = 0\).

The way this problem is usually done is to consider

\[
\lim_{z \to x^-} \frac{g(z) - g(x)}{z - x} \quad \text{and} \quad \lim_{z \to x^+} \frac{g(z) - g(x)}{z - x}.
\]

Since we are assuming \(g'(x)\) exists then both the one sided limits are equal and both equal \(g'(x)\). Looking at the one sided limits carefully you will see that one is less than or equal to 0 and the other is greater than or equal to 0, so \(g'(x)\) is both \(\leq 0\) and \(\geq 0\).
You are now ready to prove Rolle’s theorem. If you feel lost, you should reread from Problem 8.18 to here.

**Problem 8.20** *Prove Rolle’s theorem.*
Chapter 9

Integration

The intuitive idea behind integration is not too difficult: to find the area under the graph of a function \( f \), as depicted in Figure 9.1, you can inscribe rectangles and obtain an approximation. The finer the rectangles fit, the better the approximation. The area should then be the limit of these approximations. What is difficult is formalizing this process; that is, finding appropriate definitions, notation, and logical statements that capture this intuition.

Let’s begin by introducing notation. We are after a mathematical expression that represents the sum of the areas of rectangles in Figure 9.1. The first step in this process is to label the edges of the rectangles on the x-axis, as depicted in Figure 9.2. For notational consistency we would like to have \( a = x_0 \) and \( b = x_6 \), so that we can indicate the width of any rectangle by \( x_i - x_{i-1} \), for some \( i \in \mathbb{N} \). In particular, the width of the shaded rectangle is \( x_i - x_{i-1} \) with \( i = 4 \). The height of each rectangle is of the form \( f(x_{i-1}) \) for some \( i \in \mathbb{N} \). In particular, the height of the shaded rectangle is \( f(x_{i-1}) \) and its area is \( f(x_{i-1})(x_i - x_{i-1}) \) with \( i = 4 \). As \( i \) varies, the same expression represents the area of the various rectangles,
Figure 9.2: 

Figure 9.3: Area of $i$th rectangle is $f(x_{i-1})(x_i - x_{i-1})$.

as depicted in Figure 9.3. When $7 \leq i$, the expression $f(x_{i-1})(x_i - x_{i-1})$ is meaningless since $x_7$, $x_8$, $x_9$, \ldots are not defined. Mathematicians use the symbol $\sum$ to indicate a sum. To express the sum of the areas of rectangles we write

$$\sum_{i=1}^{6} f(x_{i-1})(x_i - x_{i-1}),$$

which is read, “the sum from $i$ equal 1 to 6 of $f(x_{i-1})(x_i - x_{i-1})$.” The sum of the first two rectangles can be indicated by writing

$$\sum_{i=1}^{2} f(x_{i-1})(x_i - x_{i-1})$$
and the sum of the last two rectangles is
\[ \sum_{i=5}^{6} f(x_{i-1})(x_i - x_{i-1}). \]

**Problem 9.1** Draw the rectangles whose areas are summed in the expression
\[ \sum_{i=1}^{4} f(x_{i-1})(x_i - x_{i-1}), \]
where \( f \) and \( x_0, x_1, x_2, x_3, x_4 \) are given in Figure 9.4.

![Figure 9.4:](image)

**Problem 9.2** Using Figure 9.4, draw the rectangles whose areas are summed in the expression
\[ \sum_{i=2}^{4} f\left(\frac{x_i + x_{i-1}}{2}\right)(x_i - x_{i-1}). \]

**Problem 9.3** Assume \( f \) is the squaring function, \( x_0 = 0, x_1 = 1, x_2 = 2 \) and \( x_3 = 3 \). Draw the graph of \( f \) and the rectangles whose areas are summed in the expression
\[ \sum_{i=1}^{3} f(x_i)(x_i - x_{i-1}). \]

Find the exact numerical value of this expression.

The next step is to make definitions that give names to the expressions we just introduced. A partition \( P \) of an interval \([a, b]\) is a finite subset of \([a, b]\) that contains \( a \) and \( b \). To indicate a partition we will always write
\[ P = \{x_0, x_1, \ldots, x_n\} \]
with \( a = x_0 < x_1 < \cdots < x_n = b \).
In the previous exercises we were not very picky what we took for the height of our rectangles. The one common method used for obtaining the heights was to use the value \( f(x_i^*) \) for some number \( x_i^* \) in the interval \([x_{i-1}, x_i] \). In Problem 9.1 we took \( x_i^* \) to be the left hand endpoint of the interval, in Problem 9.2 \( x_i^* \) was the middle of the interval, and in Problem 9.3 \( x_i^* \) was the right endpoint of the interval. If \( P = \{x_0, x_1, \ldots, x_n \} \) is a partition and \( x_i^* \in [x_{i-1}, x_i] \) for each \( i \), we will call the expression

\[
\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1})
\]

a Riemann sum for \( f \) and the partition \( P \). The intuitive definition of a Riemann sum to keep in mind is that it represents a sum of areas of rectangles, but rectangles for which we are not fussy about where their tops touch the curve, as illustrated in Figure 9.5.

![Figure 9.5:](image)

There are two extreme methods of selecting the numbers \( x_i^* \) from the intervals \([x_{i-1}, x_i] \), one resulting with rectangles inscribed under the graph of \( f \), and the other resulting with superscribed rectangles (see Figure refInOut). If \( x_i^* \) is taken where \( f \) attains a minimum value on the interval \([x_{i-1}, x_i] \) then the resulting rectangles will be inscribed, and if \( x_i^* \) is taken where \( f \) attains a maximum value on the interval \([x_{i-1}, x_i] \) then the resulting rectangles will be superscribed. The Riemann sum associated with the superscribed rectangles is called the upper Riemann sum and the sum associated with the inscribed rectangles is called the lower Riemann sum.

**Problem 9.4** Draw the rectangles that correspond to the upper and lower Riemann sums for the two functions and the partitions illustrated in Figure 9.7.

The point of the preceding problem is to show you that our definition of upper and lower Riemann sum has a problem, but if you drew the same rectangles in both graphs you may have (wittingly or unwittingly) stumbled onto the
remedy. The problem is that functions do not always attain maximum and minimum values, so if we need the concept of upper and lower sum (which we will!) then we either have to restrict the definition to functions that attain extreme values, or we have to modify the definition. We wish to modify the definition so we can integrate functions like the ones in Figure 9.7.

If we have a partition $P = \{x_0, x_1, \ldots, x_n\}$ then it might happen that $f$ attains no maximum value on one of the subintervals $[x_{i-1}, x_i]$, even though $\{f(x) \mid x \in [x_{i-1}, x_i]\}$ has an upper bound. This is exactly the situation for the discontinuous function illustrated in Figure 9.7. The function is bounded so that $\{f(x) \mid x \in [x_2, x_3]\}$ has an upper bound, but $f$ does not attain a maximum value on the interval $[x_2, x_3]$. A reasonable number to use for the height of our rectangle in this case is the least upper bound of the set $\{f(x) \mid x \in [x_2, x_3]\}$, and that is exactly what you did if you drew the same rectangles in both graphs pictured in Figure 9.7. Thus to each subinterval $[x_{i-1}, x_i]$ we would like to associate the least upper bound of the set $\{f(x) \mid x \in [x_{i-1}, x_i]\}$. Let us denote the least upper bound of $\{f(x) \mid x \in [x_{i-1}, x_i]\}$ by $M_i$, so we get a number $M_i$
for each value of \( i \in \{1, \ldots, n\} \). We now define the upper Riemann sum to be

\[
\sum_{i=1}^{n} M_i(x_i - x_{i-1}).
\]

To obtain the new definition of a lower Riemann sum, let \( m_i \) denote the greatest lower bound of the set \( \{ f(x) \mid x \in [x_{i-1}, x_i] \} \) and define the lower Riemann sum to be

\[
\sum_{i=1}^{n} m_i(x_i - x_{i-1}).
\]

Although we have remedied the problem of defining an upper and lower Riemann sum for the discontinuous function in Figure 9.7, our new definitions still require that we place some restriction on \( f \); we must assume that \( f \) is bounded so that the least upper bounds and greatest lower bounds are sure to exist.

**Problem 9.5** Prove that

\[
\sum_{i=1}^{n} m_i(x_i - x_{i-1}) \leq \sum_{i=1}^{n} M_i(x_i - x_{i-1}).
\]

Give an example where the lower sum equals the upper sum and give an example where the lower sum is strictly less than the upper sum.

The intuition that we hope to capture with our formal statements is that all of the upper sums give us an overestimated approximation of the area and all of the lower sums give an underestimated approximation. Intuition tells us that if there is an answer to the problem, if there is a number that will represent the area under the graph of the function, then this number is the unique value sandwiched between all of the upper and lower Riemann sums. This is the intuition that motivates the following definition.

**Definition 9.1** A function \( f \) is integrable on the interval \([a, b]\) if there exists a unique number \( s \) with the property that, for every partition,

\[
\sum_{i=1}^{n} m_i(x_i - x_{i-1}) \leq s \leq \sum_{i=1}^{n} M_i(x_i - x_{i-1}).
\]

When \( f \) is integrable we denote this unique number \( s \) by the symbol \( \int_{a}^{b} f \).

We now have a formal meaning for the symbol \( \int_{a}^{b} f \). To prove statements involving \( \int_{a}^{b} f \) you must use this formal definition and its logical consequences. You may let your intuition be your guide, but intuition alone does not constitute a proof. For example, if you are to prove that a function \( f \) is integrable on \([a, b]\) then you need to prove that two if–then statements are true;
1. (Existence of $s$) If $P$ is a partition with corresponding upper sum $\sum_{i=1}^{n} M_i(x_i - x_{i-1})$ and lower sum $\sum_{i=1}^{n} m_i(x_i - x_{i-1})$, then

$$\sum_{i=1}^{n} m_i(x_i - x_{i-1}) \leq s \leq \sum_{i=1}^{n} M_i(x_i - x_{i-1}).$$

2. (Uniqueness of $s$) If $r$ is a number which also satisfies

$$\sum_{i=1}^{n} m_i(x_i - x_{i-1}) \leq r \leq \sum_{i=1}^{n} M_i(x_i - x_{i-1})$$

for every partition, then $r = s$.

If you are assuming that $f$ is integrable on $[a, b]$ and you wish to prove that $\int_a^b f$ equals a certain value $r$, then all you have to do is prove that

$$\sum_{i=1}^{n} m_i(x_i - x_{i-1}) \leq r \leq \sum_{i=1}^{n} M_i(x_i - x_{i-1})$$

for every partition. This amounts to verifying the hypothesis of the second if-then statement, which then will tell you that $\int_a^b f = r$.

**Problem 9.6** Assume that $f$ is the constant 3 function; i.e. $f(x) = 3$ for all $x$. Prove that $\int_0^2 f = 6$.

**Problem 9.7** Assume that $f$ is the constant $c$ function; i.e. $f(x) = c$ for all $x$. Prove that $\int_a^b f = c(b - a)$

After doing the previous problems you probably feel that we have traded an easy way to compute the area of rectangles for an extremely hard way, and you are right! The goal is not to make life difficult but to come up with a definition of area that applies to very general objects, not just triangles, rectangles, or circles. The definition we have at the moment is not much good because we do not know enough about it to be able to compute with it. The story has a wonderful ending, however. With a bit of effort we will be able to establish a theorem which tells us that finding $\int_a^b f$ can be as easy as plugging numbers into a function! We now begin to assemble the proof of this theorem with the following problems.

**Problem 9.8** Use induction to prove that

$$\sum_{i=1}^{n} (w_i - w_{i-1}) = w_n - w_0$$

for any collection of numbers $w_0, w_1, \ldots, w_n$. 
CHAPTER 9. INTEGRATION

Problem 9.9 Assume that \( P = \{x_0, x_1, \ldots, x_n\} \) is a partition of \([a, b]\) and that \( x^*_i \in [x_{i-1}, x_i] \) for each \( i = 1, \ldots, n \). Prove that

\[
\sum_{i=1}^{n} m_i(x_i - x_{i-1}) \leq \sum_{i=1}^{n} f(x^*_i)(x_i - x_{i-1}) \leq \sum_{i=1}^{n} M_i(x_i - x_{i-1}).
\]

Problem 9.10 Assume that \( F \) is a differentiable function with \( F' = f \), and assume that \( P = \{x_0, x_1, \ldots, x_n\} \) is a partition of \([a, b]\). Apply the mean value theorem on each interval \([x_{i-1}, x_i]\) to conclude that there exist numbers \( x^*_i \in [x_{i-1}, x_i] \) such that

\[
F(b) - F(a) = \sum_{i=1}^{n} f(x^*_i)(x_i - x_{i-1}).
\]

To do the next problem you can use the three previous problems and the definition of “integrable”. If you get stuck, reread the paragraph after Definition 9.1.

Problem 9.11 Assume that \( F \) is a differentiable function with \( F' = f \), and assume that \( f \) is integrable. Prove that

\[
F(b) - F(a) = \int_{a}^{b} f.
\]

As things stand now, if you know that a function \( f \) is integrable, and if you know an antiderivative of that function, i.e. if you know a function \( F \) that has \( f \) as its derivative, then you can calculate the area under the graph of \( f \) in a most painless way; you simply plug two numbers into \( F \) and subtract! That this should work is utterly amazing!

Problem 9.12 Assume for the moment that the squaring function is integrable. Use this assumption and Problem 9.11 to find the area illustrated in Figure 9.8.

The task ahead involves determining which functions are integrable. There are some very uncivilized functions for which integration, as we have defined it, makes absolutely no sense.

Problem 9.13 Let \( f \) be the function that returns 1 when a rational number is input and returns 0 when an irrational number is input. Compute the upper and lower sums for \( f \) and the partition \( P = \{x_0, x_1, \ldots, x_n\} \). Prove that \( f \) is not integrable on \([0, 1]\).

The flaw with the function in Problem 9.13 is that it is very discontinuous; in fact, the function is discontinuous at every point! On the other extreme are the continuous functions, which we intend to prove are integrable. There are functions that are not continuous which are still integrable, they just can not have too many points of discontinuity. Increasing and decreasing functions
might be discontinuous at infinitely many points, but they still turn out to be integrable. There is a very nice theorem that characterizes integrability in terms of how large the set of discontinuities is, but the proof is beyond the scope of this book. Our immediate need is to build the machinery needed to prove integrability, and the first step in this direction is to see what happens to the upper and lower sums if we add elements to a partition.

When we defined the upper and lower Riemann sums we were intentionally a little sloppy. We said that an upper sum was defined by

$$\sum_{i=1}^{n} M_i(x_i - x_{i-1}),$$

but we did not indicate carefully that the numbers $M_i$ depend not only on what $i$ is, but also on what the partition is. We could indicate this by writing $M_{i,P}$ instead of $M_i$. Thus we are using a slightly different symbolism to represent the same quantities, so for each $i$ it is true that $M_{i,P} = M_i$ but the extra subscript is meant to remind us that $M_i$ will change if $P$ changes. If we wanted to go a step further we would point out these numbers also depend on the function $f$, so we really should be writing $M_{i,P,f}$. You probably see why we avoided this issue; if there is only one function and one partition in the immediate discussion then the symbol $M_i$ is unambiguous and much less intimidating than $M_{i,P,f}$. However, if we wish to compare two upper sums for $f$ corresponding to the two different partitions $P = \{x_0, \ldots, x_n\}$ and $Q = \{w_0, \ldots, w_m\}$, then we need to distinguish the least upper bound of the set \{ $f(x) \mid x \in [x_{i-1}, x_i]$ \} from the least upper bound of \{ $f(w) \mid w \in [w_{i-1}, w_i]$ \}. We could distinguish these two numbers with the symbols $M_{i,P,f}$ and $M_{i,Q,f}$, but since there will only be one function involved in the following discussion let us use the symbols $M_{i,P}$ and $M_{i,Q}$. Similarly we use the symbols $m_{i,P}$ and $m_{i,Q}$ to denote the greatest
lower bounds of the sets above. If you are confused when you attempt the next
problem it would be helpful to draw pictures that represent the upper and lower
sums that appear in the problem.

**Problem 9.14** Of all partitions of \([a, b]\), the simplest is \(P = \{x_0, x_1\}\) and the
second simplest is \(Q = \{w_0, w_1, w_2\}\) (since these are both partitions of \([a, b]\) you
automatically know what \(x_0, x_1, w_0, \) and \(w_2\) are). Prove that

\[
\sum_{i=1}^{3} M_{i,Q} (w_i - w_{i-1}) \leq \sum_{i=1}^{2} M_{i,P} (x_i - x_{i-1})
\]

and

\[
\sum_{i=1}^{2} m_{i,P} (w_i - w_{i-1}) \leq \sum_{i=1}^{3} m_{i,Q} (x_i - x_{i-1}).
\]

Give an example where equality holds in both equations above and give a second
example where strict inequalities hold.

If you have any finite set \(P\) of real numbers with two or more elements,
then this is a partition of some interval. If the smallest element of the finite
set \(P\) is the number \(r\) and the largest element is \(s\), then \(P\) is a partition of
\([r, s]\). If you write your partition \(P = \{x_0, x_1, \ldots, x_n\}\) following the custom
that \(x_0 < x_1 < \cdots < x_n\), and if you isolate some numbers

\[x_0 < x_k < x_{k+1} < \cdots < x_{k+l} < x_n,\]

then \(P_1 = \{x_0, x_1, \ldots, x_k\}\) is a partition of \([x_0, x_k]\), \(P_2 = \{x_k, x_{k+1}, \ldots, x_{k+l}\}\)
is a partition of \([x_k, x_{k+l}]\), and \(P_3 = \{x_{k+l}, \ldots, x_n\}\) is a partition of \([x_{k+l}, x_n]\).
You should convince yourself that

\[
\sum_{i=1}^{k} M_{i,P_1} (x_i - x_{i-1}) + \sum_{i=k+1}^{k+l} M_{i,P_2} (x_i - x_{i-1}) + \sum_{i=k+l+1}^{n} M_{i,P_3} (x_i - x_{i-1})
\]
is exactly the same as \(\sum_{i=1}^{n} M_{i,P} (x_i - x_{i-1})\). Let us now figure out what happens
to the upper and lower sums when one more element is added to a partition.
Assume that \(P = \{x_0, x_1, \ldots, x_n\}\) is a given partition and \(x_k < v < x_{k+1}\).
Form a new partition \(Q\) by adding the number \(v\) to \(P\), so \(Q = P \cup \{v\}\). Now
you can isolate the two numbers \(x_k < x_{k+1}\) from \(P\) to get three partitions \(P_1, P_2\) and \(P_3\) as in the preceding discussion. You can also isolate the three numbers
\(x_k < v < x_{k+1}\) from \(Q\) to get partitions \(Q_1, Q_2\) and \(Q_3\). Problem 9.14 tells you
how the upper and lower sums corresponding to \(P_1, P_2\) and \(P_3\) relate, and if you
are not lost in the notational technicalities you should see how the remaining
upper sums relate.

**Problem 9.15** Assume that \(P\) and \(Q\) are partitions of \([a, b]\), \(P \subset Q\), and \(Q\)
has just one more element than \(P\). Prove that the upper sum corresponding to
\(Q\) is less than or equal to the upper sum corresponding to \(P\) and the lower sum
sum corresponding to \(P\) is less than or equal to the lower sum corresponding to \(Q\).
In the previous problem we could have labeled the elements of $P$ and $Q$ as $P = \{x_0, x_1, \ldots, x_n\}$ and $Q = \{w_0, w_1, \ldots, w_m\}$ and then asked you to prove
\[
\sum_{i=1}^{n} M_{i,Q} (w_i - w_{i-1}) \leq \sum_{i=1}^{n} M_{i,P} (x_i - x_{i-1}).
\]
Doing so would have been asking you to prove the exact same thing about the upper sums, but the choice of notation would make the problem difficult because it hides the crucial hypothesis that $P$ is a subset of $Q$ with exactly one less element. To do Problem 9.15 you should label the elements of $P$ and $Q$, but following the notation in the paragraph preceding Problem 9.15 will lead you to a cleaner argument. We could have expressed ourselves more succinctly if, prior to Problem 9.15, we had introduced symbols to represent upper and lower Riemann sums. Thus if we agree to let $U(f, P)$ represent the upper Riemann sum and $L(f, P)$ represent the lower sum of $f$ corresponding to a partition $P$, then in Problem 9.15 you are asked to prove that $U(f, Q) \leq U(f, P)$ and $L(f, P) \leq L(f, Q)$.

**Problem 9.16** Assume that $P$ and $Q$ are partitions of $[a, b]$ and $P \subset Q$. Use induction on the number of elements in $Q$ that are not in $P$ to prove that $U(f, Q) \leq U(f, P)$ and $L(f, P) \leq L(f, Q)$.

If you are given two arbitrary partitions $P$ and $Q$ of $[a, b]$ then it is quite possible that neither one is a subset of the other. For example, both $P = \{1, 2, 3\}$ and $Q = \{1, 1.5, 3\}$ are partitions of $[1, 3]$ but $P$ is not a subset of $Q$ and $Q$ is not a subset of $P$. However, both $P$ and $Q$ are subsets of $P \cup Q = \{1, 1.5, 2, 3\}$. To do the next problem you may assume that $P$ and $Q$ are arbitrary partitions of $[a, b]$ and apply the result of Problem 9.16 to the related partitions $P$ and $P \cup Q$. Problem 9.16 also applies to the related partitions $Q$ and $P \cup Q$. Finally, you can apply Problem 9.5 to the partition $P \cup Q$ and, when you combine this information you should see how to conclude that $L(f, Q) \leq U(f, P)$.

**Problem 9.17** Assume that $P$ and $Q$ are arbitrary partitions of $[a, b]$. Prove that $L(f, Q) \leq U(f, P)$.

The next problem will follow from the definition of least upper bound and the definition of greatest lower bound. If you can not recall a precise definition you should definitely look it up. To point you in the right direction, observe that saying $L(f, Q) \leq U(f, P)$ for every partition $P$ means that $L(f, Q)$ is a lower bound of the set $\{ U(f, P) \mid P \text{ a partition of } [a, b] \}$.

**Problem 9.18** Let $s_u$ be the greatest lower bound of the set
\[
\{ U(f, P) \mid P \text{ a partition of } [a, b] \},
\]
and let $s_l$ be the least upper bound of
\[
\{ L(f, P) \mid P \text{ a partition of } [a, b] \}.
\]
Prove that
\[ \sum_{i=1}^{n} m_{i,P} (x_i - x_{i-1}) \leq s_l \leq s_u \leq \sum_{i=1}^{n} M_{i,P} (x_i - x_{i-1}) \]
for any partition \( P \).

**Problem 9.19** With \( s_l \) and \( s_u \) defined as in Problem 9.18 prove that \( f \) is integrable if and only if \( s_l = s_u \).

One of the consequences of Problem 9.18 is that there always exists a number that is between all the upper and lower Riemann sums. Thus a function is integrable if and only if there is *only one* number between all upper and lower Riemann sums. In particular, if you ever wish to use Definition 9.1 to prove a function is integrable, you need not verify the existence of a number between all the upper and lower sums, since we have just proved such a number always exists. However, you still need to prove the uniqueness of such a number.

We now have two logical statements that are equivalent to the integrability of \( f \), the statement that appears in Definition 9.1 and the assertion that \( s_l = s_u \). There are several other logical statements that we wish to prove are equivalent to the integrability of \( f \). There is an effort saving device that mathematicians use to prove the equivalence of three or more statements, a device that we will now illustrate with three statements. Suppose that \( S_1, S_2 \), and \( S_3 \) are three statements that we would like to prove equivalent. We could establish the equivalence of each pair of statements by proving two if–then statements per pair, which amounts to proving six if–then statements (since there are three pairs). Alternatively, suppose we could prove true all three if–then statements “if \( S_1 \) then \( S_2 \)”, “if \( S_2 \) then \( S_3 \)”, and “if \( S_3 \) then \( S_1 \)”. In this situation it is impossible for one of the statements to be true while another statement is false, since the truth of any one statement will force all the other statements true (see Figure 9.9). This means that all three statements must then be equivalent!

Figure 9.9: Follow the arrows around.

**Theorem 9.1** The following statements are equivalent;

1. \( f \) is integrable on \([a, b]\).
2. if \( \epsilon > 0 \) then there exists a partition \( P \) of \([a, b]\) such that
   \[ U(f, P) - L(f, P) < \epsilon. \]
3. There exists a sequence $P_n$ of partitions of $[a, b]$ such that

$$U(f, P_n) - L(f, P_n) \to 0.$$ 

**Proof.** We begin by proving that the first statement implies the second, so assume the first statement is true and assume the hypothesis of the second statement; i.e. assume that $\epsilon > 0$. Since $f$ is assumed to be integrable, Problem 9.19 tells us that $s_i = s_u$. Now $s_u$ is the greatest lower bound of the set of upper sums, so any larger number fails to be a lower bound: in particular $s_u + \frac{\epsilon}{2}$ is not a lower bound so there is some upper sum $U(f, P_1)$ such that $U(f, P_1) < s_u + \frac{\epsilon}{2}$. Similarly, there is an element $L(f, P_2)$ of the set of lower sums such that $L(f, P_2) > s_l - \frac{\epsilon}{2}$. If $P = P_1 \cup P_2$ then by Problem 9.16 we have $U(f, P) \leq U(f, P_1)$ and $L(f, P_2) \leq L(f, P)$, so that

$$U(f, P) - L(f, P) \leq U(f, P_1) - L(f, P_2) < (s_u + \frac{\epsilon}{2}) - (s_l - \frac{\epsilon}{2}) = \epsilon.$$ 

Let us now prove that the second statement implies the third statement, so assume the second statement is true. If we apply the second statement to the number $\epsilon = 1$ then we get a partition $P_1$ such that $U(f, P_1) - L(f, P_1) < 1$. Apply the second statement again with $\epsilon = \frac{1}{2}$ to get a partition $P_2$ such that $U(f, P_2) - L(f, P_2) < \frac{1}{2}$. Continuing in this way we obtain with $\epsilon = \frac{1}{n}$ a partition $P_n$ such that $U(f, P_n) - L(f, P_n) < \frac{1}{n}$. Thus $0 \leq U(f, P_n) - L(f, P_n) < \frac{1}{n}$ implies that $U(f, P_n) - L(f, P_n) \to 0$ by the squeeze theorem.

Finally, let us prove that the third statement implies the first, so assume $P_n$ is a sequence of partitions such that $U(f, P_n) - L(f, P_n) \to 0$. As we mentioned immediately after Problem 9.19 we need only prove that there is only one number between all upper and lower Riemann sums. This amounts to proving that the following if-then statement is true: if both $r$ and $s$ are between all upper and lower Riemann sums then $r = s$. Assume that $r$ and $s$ are between every upper and lower sum. In particular we have $L(f, P_n) \leq r \leq U(f, P_n)$ and $L(f, P_n) \leq s \leq U(f, P_n)$ for all $n \in \mathbb{N}$. Subtracting $L(f, P_n)$ from each term in the inequality for $r$ we see that

$$0 \leq r - L(f, P_n) \leq U(f, P_n) - L(f, P_n)$$

so $r - L(f, P_n) \to 0$ by the squeeze theorem, i.e. $L(f, P_n) \to r$. Using the inequality for $s$ and the same reasoning we conclude that $L(f, P_n) \to s$. It follows from Problem 5.12 (limits are unique) that $r = s$.

$\triangle$

It is worth mentioning a fact that is implicit in the proof just given. If we can ever find a sequence of partitions $P_n$ such that $U(f, P_n) - L(f, P_n) \to 0$ then $U(f, P_n) \to \int_a^b f$ and $L(f, P_n) \to \int_a^b f$. In Problem 9.12 you were asked to find the value of $\int_1^3 f$ when $f$ is the squaring function. To do this you
assumed something about the squaring function that has not yet been proved; you assumed that the squaring function was integrable. The culmination of the following set of problems is a proof that the squaring function is integrable by exhibiting a sequence of partitions \( P_n \) for which \( U(f, P_n) - L(f, P_n) \to 0 \).

**Problem 9.20** Let \( f \) be the function that always outputs 1 except when 0 is input, in which case \( f \) outputs 2. Draw the graph of \( f \) and prove that \( f \) is integrable on \([-1, 1]\).

A particularly civilized partition of an interval is obtained when you divide the interval up into \( n \) subintervals of equal length. For example, if you have an interval \([a, b]\) you can obtain a partition \( P_2 \) by dividing the interval in half resulting with

\[
P_2 = \{a, a + \frac{b-a}{2}, b\}.
\]

The way the middle point is obtained is by adding half the length of the interval \([a, b]\) to \( a \), the length of \([a, b]\) being \( b-a \). If you divide the interval into three subintervals of equal length you obtain a partition

\[
P_3 = \{a, a + \frac{b-a}{3}, a + 2\frac{b-a}{3}, b\}.
\]

Once again, the middle points are obtained by adding one third the length of the interval to \( a \) and adding two thirds the length to \( a \). You probably see the general pattern; if you divide the interval \([a, b]\) into \( n \) subintervals of equal length you get the partition

\[
P_n = \{a, a + \frac{b-a}{n}, a + 2\frac{b-a}{n}, \ldots, a + n\frac{b-a}{n}\},
\]

and notice that the last element written is \( b \) (as it should be). Thus \( P_n \) is the partition \( \{x_0, \ldots, x_n\} \) with

\[
x_i = a + i\frac{b-a}{n}.
\]

Since the length of each subinterval is the length of \([a, b]\) divided by \( n \) we have

\[
x_i - x_{i-1} = \frac{b-a}{n}
\]

for each \( i \in \{1, \ldots, n\} \), and as a consequence

\[
U(f, P_n) = \frac{b-a}{n} \sum_{i=1}^{n} M_i, P_n \quad \text{and} \quad L(f, P_n) = \frac{b-a}{n} \sum_{i=1}^{n} m_i, P_n.
\]

The situation is simplified even further if you have an increasing function (if you have forgotten the definition of an increasing function you should reread
the paragraph after Problem 8.12). In this case you have \( M_i, P_n = f(x_i) \) and \( m_i, P_n = f(x_{i-1}) \), so that

\[
U(f, P_n) = \frac{b-a}{n} \sum_{i=1}^{n} f(x_i) \quad \text{and} \quad L(f, P_n) = \frac{b-a}{n} \sum_{i=1}^{n} f(x_{i-1}).
\]

You may use these comments and Problem 9.8 to do the following.

**Problem 9.21** If \( f \) is an increasing function on \([a, b]\) then \( U(f, P_n) - L(f, P_n) \rightarrow 0 \).

**Problem 9.22** Prove that \( f_1^3 f = \frac{26}{3} \) when \( f \) is the squaring function.

**Problem 9.23** Prove that the following is true; if \( f \) is a decreasing function on \([a, b]\) then \( f \) is integrable on \([a, b]\).

If \( f \) is neither increasing nor decreasing then the best one can say is that

\[
U(f, P_n) - L(f, P_n) = \frac{b-a}{n} \sum_{i=1}^{n} (M_i, P_n - m_i, P_n).
\]

Writing the difference of the upper and lower sum this way does reveal a strategy that may be used to prove \( f \) is integrable. Imagine what happens if every one of the numbers \( M_i, P_n - m_i, P_n \) is less than a number \( \epsilon \); you could then substitute the larger number to obtain the inequality

\[
U(f, P_n) - L(f, P_n) \leq \frac{b-a}{n} (\epsilon + \ldots + \epsilon) = \epsilon (b-a).
\]

This is the insight that leads one to stating and proving the following lemma. Some of the technicalities are simplified if we first introduce still more terminology. Let \( I \) denote a closed interval of numbers contained in the domain of \( f \) and let \( M_I \) and \( m_I \) denote the least upper bound and the greatest lower bound of the set \( \{ f(x) \mid x \in I \} \). We say \( f \) is uniformly continuous on its domain when the following is true; if \( \epsilon > 0 \) then there exists \( \delta > 0 \) such that \( M_I - m_I < \epsilon \) for every interval of length less than \( \delta \).

**Problem 9.24** Prove the following is true; if \( f \) is uniformly continuous on \([a, b]\) then \( f \) is continuous on \([a, b]\).

The converse of Problem 9.24 is also true, but it is much harder to prove. A key ingredient in the proof of the converse is the hypothesis that the domain of \( f \) be a closed interval. Without this hypothesis the converse is false; a counterexample is the squaring function viewed with the domain \( \mathbb{R} \).

**Lemma 9.1** If \( f \) is a continuous function on \([a, b]\) then \( f \) is uniformly continuous on \([a, b]\).
Proof. We will prove the contrapositive statement; if \( f \) is not uniformly continuous on \([a, b]\) then \( f \) is not continuous on \([a, b]\). Assume that \( f \) is not uniformly continuous, so the if–then statement that defines uniform continuity is false. Thus there is a positive number \( \epsilon \) for which the conclusion is false. This means that for every \( \delta > 0 \) we must have \( M_I - m_I \geq \epsilon \) for some interval of length less than \( \delta \). In particular, for \( \delta = 1 \) there is an interval \( I = [a_1, b_1] \) with \( b_1 - a_1 < 1 \) but \( M_I - m_I \geq \epsilon \), which then allows us to find \( x_1, y_1 \in (a_1, b_1) \) with

\[
f(x_1) - f(y_1) \geq \frac{\epsilon}{2}.
\]

Repeat the process with \( \delta = 1/2 \) to get an interval \([a_2, b_2]\) and numbers \( x_2, y_2 \in [a_2, b_2] \) for which

\[
b_2 - a_2 < \frac{1}{2} \quad \text{and} \quad f(x_2) - f(y_2) \geq \frac{\epsilon}{2}.
\]

Continuing inductively we obtain, with \( \delta = 1/n \), an interval \([a_n, b_n]\) and numbers \( x_n, y_n \in [a_n, b_n] \) for which

\[
b_n - a_n < \frac{1}{n} \quad \text{and} \quad f(x_n) - f(y_n) \geq \frac{\epsilon}{2}.
\]

Since \( x_n, y_n \in [a_n, b_n] \) we have

\[
0 \leq |x_n - y_n| \leq b_n - a_n < \frac{1}{n}
\]

so \( x_n - y_n \to 0 \) by the squeeze theorem. By Lemma 7.2 there exists a subsequence \((x_{n_1}, x_{n_2}, \ldots)\) of \((x_n)\) that converges to a number \( x \in [a, b] \). Since \( x_n - y_n \to 0 \) we conclude that the subsequence \((y_{n_1}, y_{n_2}, \ldots)\) also converges to \( x \). If \( f \) were continuous at \( x \) we would have both \((f(x_{n_1}), f(x_{n_2}), \ldots)\) and \((f(y_{n_1}), f(y_{n_2}), \ldots)\) converging to \( f(x) \), which is impossible since

\[
f(x_{n_k}) - f(y_{n_k}) \geq \frac{\epsilon}{2}
\]

for all \( n \in \mathbb{N} \). Thus \( f \) is not continuous at \( x \) and \( f \) is not continuous on \([a, b]\).

\( \triangle \)

Problem 9.25 Prove the following is true; if \( f \) is continuous on \([a, b]\) then \( f \) is integrable on \([a, b]\).
Chapter 10

Fundamental Theorems of Calculus

When I look up the word “calculus” in my dictionary \(^1\) I find that the word derived from Latin. The Latin meaning is “pepple, stone in the bladder or kidney, stone used in reckoning”. The modern definition I found that seems pertinent to this book is “a method of computation or calculation in a special symbolic notation”. There is a single theorem in the subject of calculus that, more than any other, earns the subject its name. This is the theorem that makes the complicated process of integration computable.

**Theorem 10.1 (Fundamental Theorem of Calculus.)** If \(f\) is continuous on \([a,b]\) and \(F\) is an antiderivative of \(f\), then \(\int_a^b f = F(b) - F(a)\).

**Problem 10.1** Prove Theorem 10.1.

Once again, there are several theorems that are all called the fundamental theorem of calculus. The version stated above is a computational theorem; it tells you how you may compute the value of an integral. The second version that we will encounter in this chapter is a much more powerful theorem. It is quite easy to prove Theorem 10.1 if you assume the second version true, but assuming that Theorem 10.1 is true provides no help at all in the proof of the second version.

**Theorem 10.2 (Fundamental Theorem of Calculus 2.)** If \(f\) is continuous on \([a,b]\) and \(G\) is defined by

\[
G(x) = \int_a^x f,
\]

then \(G\) is an antiderivative of \(f\).

\(^1\)Webster’s Seventh New Collegiate Dictionary
The primary goal of this chapter is to give a proof of this theorem. Before we start building the machinery that will go into this proof we will attempt to illuminate the meaning of this theorem.

If you have a continuous function \( f \) that is already known to have an antiderivative \( F \), then Theorem 10.1 tells us that \( \int_a^x f = F(x) - F(a) \). Thus if \( G(x) = \int_a^x f \) then \( F \) and \( G \) differ by the constant \( F(a) \) and consequently have the same derivative; i.e. \( G \) is an antiderivative of \( f \). The point is that Theorem 10.2 is very easy to prove if one assumes Theorem 10.1 is true and if one also assumes that \( f \) has an antiderivative. The really new content of Theorem 10.2 is that every continuous function has an antiderivative.

**Problem 10.2** Let \( f \) be the squaring function and define \( G \) by \( G(x) = \int_1^x f \). Prove that \( G \) is a polynomial of degree 3 and give a formula for this polynomial. Verify directly that the derivative of \( G \) is \( f \).

**Problem 10.3** Let \( f \) be defined by the formula

\[
f(x) = \begin{cases} 
0 & \text{if } 0 \leq x \leq 1 \\
1 & \text{if } 1 < x \leq 2
\end{cases}
\]

and let \( G(x) = \int_0^x f \). Draw the graph of both \( f \) and \( G \). Is \( G \) an antiderivative of \( f \)?

**Problem 10.4** Let \( f \) be defined by the formula

\[
f(x) = \begin{cases} 
2x & \text{if } 0 \leq x \leq 1 \\
4 - 2x & \text{if } 1 < x \leq 2
\end{cases}
\]

and let \( G(x) = \int_0^x f \). Draw the graph of both \( f \) and \( G \). Is \( G \) an antiderivative of \( f \)?

**Problem 10.5** Let \( f \) be defined by the formula

\[
f(x) = \begin{cases} 
2x & \text{if } 0 \leq x \leq 1 \\
-2x & \text{if } 1 < x \leq 2
\end{cases}
\]

and let \( G(x) = \int_0^x f \). Draw the graph of both \( f \) and \( G \). Is \( G \) an antiderivative of \( f \)?

The first thing you will stare at when you set out to prove Theorem 10.2 is

\[
\frac{G(x + h) - G(x)}{h},
\]

since \( G'(x) \) is the limit of this expression as \( h \) approaches 0. Assume for awhile that \( h \) is positive. If you look at a picture of the areas involved with the expression \( G(x + h) - G(x) \) (see Figure 10.1) you will see intuitively that

\[
G(x + h) - G(x) = \int_x^{x+h} f,
\]
and if our definitions accurately capture this intuition we should be able to prove this equality. That is the direction we are headed now; specifically, we intend to prove that

\[
\int_x^y f + \int_y^z f = \int_x^z f \text{ whenever } x < y < z.
\]

Recall from Problem 9.18 that we let \( s_u \) denote the greatest lower bound of the set

\[
\{ U(f, P) \mid P \text{ a partition of } [a, b] \}
\]

and we let \( s_l \) denote the least upper bound of

\[
\{ L(f, P) \mid P \text{ a partition of } [a, b] \}.
\]

This notation was fine at that time, when there was only one function and one interval under discussion. If you think about the definition of \( s_u \) (and \( s_l \)) you will realize the value depends on the interval \([a, b]\) and on the function \( f \), so a more descriptive, and yet more cumbersome, symbolism would indicate the interval and the function. This is completely analogous to the choice of symbolism used in Definition 9.1 to describe what \( f^b_a \) is. We said that \( f \) is integrable on an interval \([a, b]\) if there is a unique number \( s \) that is between all the upper and lower Riemann sums. If such a number exists we said that we would write it as \( f^b_a \) to reflect the fact that this number \( s = f^b_a \) depends on both the interval and the function. Since the discussion ahead of us involves integration over several different intervals we need to introduce a symbolism that reflects this; from now on let us write \( f^b_a \) instead of \( s_u \) and \( f^b_a \) instead of \( s_l \). With this notation we can express Problem 9.19 by saying \( f \) is integrable if and only if \( f^b_a = f^b_a \), in which case \( f^b_a \) is defined to equal this common value.
Lemma 10.1 If $f$ is any bounded function on $[x, z]$ and $x < y < z$ then

$$\int_x^y f + \int_y^z f = \int_x^z f.$$ 

Proof. By its definition the symbol $\int_x^z f$ denotes the greatest lower bound of the set

$$\{ U(f, P) \mid P \text{ a partition of } [x, z] \}.$$ 

We will now prove that $\int_x^y f + \int_y^z f$ is also a lower bound of this set, which will establish the inequality

$$\int_x^y f + \int_y^z f \leq \int_x^z f.$$ 

Assume that $P$ is an arbitrary partition of $[x, z]$. Let

$$Q_0 = P \cup \{y\}, Q_1 = \{ t \in P \mid t \leq y \} \cup \{y\}, \text{ and } Q_2 = \{ t \in P \mid t \geq y \} \cup \{y\},$$

so $Q_0$ is a partition of $[x, z]$, $Q_1$ is a partition of $[x, y]$, and $Q_2$ is a partition of $[y, z]$. You should convince yourself that

$$U(f, Q_1) + U(f, Q_2) = U(f, Q_0) \leq U(f, P)$$

(the rightmost inequality follows from Problem 9.16). Since $\int_x^y f$ is the greatest lower bound of the set

$$\{ U(f, Q) \mid Q \text{ a partition of } [x, y] \}$$

and $U(f, Q_1)$ is an element of this set we have $\int_x^y f \leq U(f, Q_1)$. Similarly we have $\int_y^z f \leq U(f, Q_2)$ so that

$$\int_x^y f + \int_y^z f \leq U(f, Q_1) + U(f, Q_2) = U(f, Q_0) \leq U(f, P).$$

It follows that $\int_x^y f + \int_y^z f$ is a lower bound of the set

$$\{ U(f, P) \mid P \text{ a partition of } [x, z] \}$$

so $\int_x^y f + \int_y^z f \leq \int_x^z f$.

We now intend to prove that

$$\int_x^y f \leq \int_x^y f + \int_y^z f.$$ 

To do this, assume that $Q_1$ is an arbitrary partition of $[x, y]$ and $Q_2$ is an arbitrary partition of $[y, z]$. If $P = Q_1 \cup Q_2$ then $P$ is a partition of $[x, z]$ so

$$\int_x^y f \leq U(f, P) = U(f, Q_1) + U(f, Q_2).$$
It follows that \( \int_x^z f - U(f, Q_1) \leq U(f, Q_2) \) so \( \int_x^z f - U(f, Q_1) \) is a lower bound of the set \{ U(f, Q) \mid Q \text{ a partition of } [y,z] \}.

Since \( \int_y^z f \) is the greatest lower bound we conclude that \( \int_x^z f - U(f, Q_1) \leq \int_y^z f \).

Know we have \( \int_x^z f - \int_y^z f \leq U(f, Q_1) \) which reveals that \( \int_x^z f - \int_y^z f \) is a lower bound of \{ U(f, Q) \mid Q \text{ a partition of } [x,y] \},

so \( \int_x^z f - \int_y^z f \leq \int_y^z f \). Combining this inequality with the one obtained in the first paragraph gives us

\[
\int_x^z f = \int_x^y f + \int_y^z f.
\]

\( \triangle \)

**Problem 10.6** Prove that the following is true; if \( f \) is any bounded function on \([x,z]\) and \( x < y < z \) then

\[
\int_x^y f + \int_y^z f = \int_x^z f.
\]

**Problem 10.7** Assume that \( f \) is a continuous function on \([a,b]\), \( x \in (a,b) \) and \( h \) is a positive number such that \( a < x < x+h \leq b \). If \( G \) is the function defined in the statement of Theorem 10.2, prove that \( G(x+h) - G(x) = \int_x^{x+h} f \).

Let us first determine what

\[
l \lim_{h \to 0^+} \frac{G(x+h) - G(x)}{h}
\]

is. To do this we will assume that \( (h_n) \) is a sequence of positive numbers such that \( x + h_n \in [a,b] \) and \( h_n \to 0 \). The goal is to prove that

\[
\frac{G(x+h_n) - G(x)}{h_n} \to f(x).
\]

The insight that will lead us to this proof is obtained by using Problem 10.7 to realize the numerator as a particular area under the graph of \( f \), and then drawing an illuminating picture (see Figure 10.2). For any positive number \( h \) you can restrict your attention to the function \( f \) on the interval \([x,x+h]\), so the area under the graph of \( f \) is \( \int_x^{x+h} f \). The extreme value theorem gives us numbers \( v \) and \( w \) in the interval \([x,x+h]\) where \( f \) attains a maximum and minimum value (respectively). The area of the large rectangle is actually an upper Riemann sum corresponding to the partition \{\( x, x+h \)\}, and the area of the small rectangle is the lower Riemann sum for this partition, so by Definition 9.1 you get \( f(v)h \leq \int_x^{x+h} f \leq f(v)h \). If you apply this reasoning to each term of the sequence \( (h_n) \) you obtain a sequence \( (v_n) \) and a sequence \( (w_n) \) of numbers where \( f \) attains a maximum and minimum value on the interval \([x,x+h_n]\) (respectively). The following problem is amounts to an application of the squeeze theorem and the definition of continuity.
Problem 10.8 With the sequences \( (v_n) \) and \( (w_n) \) defined as in the previous paragraph, prove that \( f(v_n) \to f(x) \) and \( f(w_n) \to f(x) \).

Problem 10.9 Prove that

\[
f(x) = \lim_{h \to 0^+} \frac{G(x+h) - G(x)}{h}.
\]

Problem 10.10 Prove that

\[
f(x) = \lim_{h \to 0^-} \frac{G(x+h) - G(x)}{h}.
\]

Problem 10.11 Prove Theorem 10.2.

You will find Problem 8.13 helpful to do the following.

Problem 10.12 Use Theorem 10.2 to give another proof of Theorem 10.1.
Chapter 11

Functions with Formulas

Among the functions that map real numbers to real numbers are a few that may be described with a formula. These are the most important ones mathematically because they are the ones that are susceptible to deep mathematical investigation. Functions that arise in nature are often not given by any formula, but it is frequently possible to approximate them with functions that are given by formulas. Thus the machinery that is used to study functions with formulas allows us to get information about the functions that may be approximated with formulas, and this turns out to be a pretty large collection of functions.

11.1 Polynomials

Using the operations of multiplication, addition, and subtraction, one generates the class of polynomial functions. These are the functions $f$ that admit a formula of the form

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

**Problem 11.1** For each graph in Figure 11.1, find $a_0, a_1, \text{ and } a_2$ so that

$$f(x) = a_0 + a_1 x + a_2 x^2.$$

![Graphs of three $f$'s](image)

Figure 11.1: Graphs of three $f$’s
The coefficients $a_i$ that determine the polynomial $f$ are intimately related to the behaviour of $f$ at the number 0. Indeed, the number $a_0$ is exactly $f(0)$ and $a_1$ equals $f'(0)$. Any polynomial can be represented with a formula that relates the function to its values at any other number $c$ by writing

$$f(x) = b_n(x - c)^n + b_{n-1}(x - c)^{n-1} + \cdots + b_1(x - c) + b_0.$$  

When a polynomial is written this way we say it is written centered at $c$.

**Problem 11.2** Assume the polynomial $f$ is written centered at $c$, as in the equation above. Find $b_2$ in terms of $f''$. (The answer is not equal to $f''(c)$, although it is close to this!)

**Problem 11.3** For each graph in Figure 11.1 find $b_0, b_1,$ and $b_2$ so that

$$f(x) = b_0 + b_1(x - 1) + b_2(x - 1)^2.$$  

You can use the results of the two previous problems to do this easily.

**Problem 11.4** Assume $f(x) = 4 + 2x - 3x^2 + x^3$. Find $b_0, b_1, b_2,$ and $b_3$ so that this $f$ is expressed by the formula

$$f(x) = b_0 + b_1(x - 1) + b_2(x - 1)^2 + b_3(x - 1)^3.$$  

**Problem 11.5** Find the equation of the line tangent to the graph of $f$ in Problem 11.4 at $(1,4)$. What does the solution have to do with the $b$‘s you found in the Problem 11.4?

**Problem 11.6** Without doing any computation, give the equation of the line tangent to the graph of $f(x) = 2 + 5(x - 2) + 9(x - 2)^2$ at $(2,2)$. Explain why your solution is correct.

**11.1.1 Graphs of polynomials**

**Problem 11.7** Draw the graphs of the functions given by the following formulas.

(i) $f(x) = 2x^2 + 4x + 1$

(ii) $f(x) = \frac{5}{12}x^3 + 2x^2 - 4x + 1$

(iii) $f(x) = (x - 1)(x - 3)$

The calculus developed earlier lets us deduce general facts that relate to what graphs of polynomials look like. For example, we will now show that the number of roots of a non-constant polynomial cannot exceed the degree of a polynomial. Recall that a root of a polynomial $f$ is an input $x$ for which $f(x) = 0$. Equivalently, a root of $f$ is a number on the $x$-axis where the graph of $f$ intersects with the $x$-axis (see figure 11.2). The degree of a polynomial is the largest power that appears in the formula that defines the polynomial. For example, the polynomials $x^2, 2x^2 + 5x + 1,$ and $x + 20x^2$ have degree two, and the polynomial $3x^5 + 15x^2 + 1$ has degree five. Thus we are interested in proving that the
number of times the graph of a polynomial crosses the $x$-axis cannot exceed the
degree of the polynomial. We will use induction and the mean value theorem in
the proof. The base case of the induction proof is the solution to the following.

![Graph of a polynomial](image)

Figure 11.2: Not the graph of a polynomial of degree 2 or less

**Problem 11.8** Assume $f(x) = ax + b$ with $a \neq 0$. Show that $f$ has exactly one root.

**Problem 11.9** Assume a polynomial has two roots at $a$ and $b$ with $a < b$. Apply the mean value theorem on $[a, b]$; what is the conclusion?

A critical point of a function is an point $x$ in its domain where the derivative exists and equals zero. In other words, the number $x$ is a critical point of $f$ if and only if $f'(x) = 0$.

**Problem 11.10** Draw the graph of a function whose derivative exists everywhere in its domain and which has six roots. Point to five places where the derivative is zero. Explain how the mean value theorem proves that such a function must have at least five critical points.

**Problem 11.11** Assume $f$ is a polynomial and prove the following: If $f$ has at least $m$ roots, then $f'$ has at least $m - 1$ roots.

**Problem 11.12** Write the contrapositive of Problem 11.11.

You can now use Problem 11.12 and induction to prove that polynomials of
degree $n$ have at most $n$ roots. The following is the inductive step.

**Problem 11.13** Prove the following: if polynomials of degree $n - 1$ have at most $n - 1$ roots, then polynomials of degree $n$ have at most $n$ roots.
11.1.2 Quadratics

Quadratics refer to polynomials of degree two. You remember the quadratic formula which gives the roots of \( f(x) = ax^2 + bx + c \) as

\[
x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.
\]

Whether there are two, one, or no roots at all depends on the sign of the number \( b^2 - 4ac \). The graph of every quadratic constitutes what is called a parabola, and these all look like either a big smile or a big frown. A function is said to be \textit{concave up} (smiling) on an interval if its derivative is increasing on that interval, and it is called \textit{concave down} (frowning) if its derivative is decreasing.

**Problem 11.14** Give a condition on the numbers \( a, b, \) and \( c \) that determines whether the graph of \( f \) is concave up or down.

**Problem 11.15** Find a formula for the critical point of \( f \) that involves only \( a, b, \) and \( c \).

11.1.3 Two points determine a line

**Problem 11.16** Find the equation of the unique line that passes through the point \((1, 2)\) and \((2, 5)\).

**Problem 11.17** Find the equation of the unique line that passes through the point \((1, 2)\) and whose derivative is 5 at \( x = 1 \).

The previous problems are meant to convince the reader that polynomials of degree one are completely determined by two bits of data. This data may be given as two points in the plane, through which the graph of the polynomial must pass, or the data may be given as the two values the polynomial and its derivative return at a specified point. A nice generalization of this is the fact that polynomials of degree \( n \) are completely determined by \( n + 1 \) bits of data given in either of the two ways mentioned above. The following problem can be solved by writing \( p(x) = a_0 + a_1x + a_2x^2 \) and using the given data to solve for \( a_0, a_1, \) and \( a_2 \).

**Problem 11.18** Find the formula of the unique quadratic \( p \) (i.e. second degree polynomial) that satisfies \( p(0) = 1, \ p'(0) = 2, \) and \( p''(0) = 3 \).

In the next problem, try writing \( p(x) = a_2(x - 5)^2 + a_1(x - 5) + a_0 \) and use the data to find \( a_2, a_1, \) and \( a_0 \).

**Problem 11.19** Find the formula of the unique quadratic \( p \) that satisfies \( p(5) = 0, \ p'(5) = 1, \) and \( p''(5) = 10 \).

**Problem 11.20** Find the formula of the unique quintic \( p \) (degree 5 polynomial) that has

\[
p(1) = 1, \ p'(1) = 2, \ p''(1) = 6, \ p'''(1) = 12, \ p''''(1) = 48, \) and \( p'''''(1) = 5! \)
11.1. POLYNOMIALS

Hopefully you realized that $5!$ means the product $(5)(4)(3)(2)(1)$ and you probably see how factorials arise when taking higher derivatives. The fifth derivative of $p$ is traditionally denoted $p^{(5)}$ instead of $p^{(5)}$, so we will be able to refer to the $n^{th}$ derivative as $p^{(n)}$.

**Problem 11.21** Find a formula for the unique $n^{th}$ degree polynomial that satisfies $p(a) = y_0$, $p^{(1)}(a) = y_1$, $p^{(2)}(a) = y_2$, \ldots, $p^{(n)}(a) = y_n$.

The formula you obtain in the previous problem is a general expression in terms of the data $a, y_0, y_1, y_2, \ldots, y_n$. You obtain it by recognizing the pattern arising in the previous problems where concrete data was given for $a, y_0, y_1, y_2, \ldots, y_n$. If you didn’t see the pattern in those problems, you can change the data in Problem 11.20 and do the new problem, watching carefully for the pattern to emerge.

The main point of our discussion thus far is that there exists exactly one polynomial $p$ of degree $n$ that fits the $n+1$ data requirements of the form $p(a) = y_0$, $p^{(1)}(a) = y_1$, $p^{(2)}(a) = y_2$, \ldots, $p^{(n)}(a) = y_n$. To clearly understand this statement, think of it as the generalization of the "point-slope" formula for a line; you can write the formula for a line if you know a point on the line $(p(a) = y_0)$ and the slope of the line $(p'(a) = y_1)$. You can write the formula for a quadratic knowing only a point on the curve $(p(a) = y_0)$, the slope of the curve at that point $(p'(a) = y_1)$, and the second derivative at that point $(p''(a) = y_2)$.

Now imagine that you have a function $f$ (that might not even have a known formula) for which it is possible to determine all of the values of the higher derivatives at some input $a$. There is then a unique $n^{th}$ degree polynomial $p$ that satisfies

$$p(a) = f(a), \quad p^{(1)}(a) = f^{(1)}(a), \ldots, \quad p^{(n)}(a) = f^{(n)}(a).$$

In other words, for each $n$ there is exactly one polynomial of degree $n$ that returns the same value as $f$ at $a$, and whose higher derivatives also return the same values as $f$’s higher derivatives at $a$ (up to the $n^{th}$ derivative). This polynomial is called the Taylor polynomial of $f$ of degree $n$ at $a$, and we will denote it $p_n$ for now. Since there is a polynomial for each value of $n$ we actually have a sequence of polynomials. As $n$ increases, how does $p_n(x)$ compare with $f(x)$? Does the sequence of numbers $p_n(x)$ always converge, no matter what $x$ we select? Could it possibly converge to $f(x)$?

**Problem 11.22** Assume $f(x) = \frac{1}{1-x}$ and $a = 0$. Find a formula for $p_5(x)$.

**Problem 11.23** With $f(x) = \frac{1}{1-x}$ and $a = 0$, find a formula for $p_n(x)$.

**Problem 11.24** Show that $(1-x)(1+x+\ldots+x^n) = 1-x^{n+1}$ for all $n$.

**Problem 11.25** With $f(x) = \frac{1}{1-x}$ and $p_n(x)$ as determined in problem 11.23 show that $p_n(x) \to f(x)$ exactly when $|x| < 1$. 
The situation exhibited by the one example when \( f(x) = \frac{1}{1-x} \) turns out to be typical of all examples; if \( f \) has derivatives of all orders at 0, so the Taylor polynomials \( p_n(x) \) for \( f \) at 0 can be defined, then there will be a radius \( R \) (called the \textit{radius of convergence}) for which \( p_n(x) \to f(x) \) when \(|x| < R \) and \( p_n(x) \to f(x) \) when \( R < |x| \). Essentially the same thing happens when considering the Taylor polynomials of \( f \) at a number \( a \) (other than 0), except the \textit{interval of convergence} is centered at \( a \) instead of 0. This may be expressed by saying that \( p_n(x) \to f(x) \) when \(|x - a| < R \) and \( p_n(x) \) is the Taylor polynomial of \( f \) at \( a \). Note that the radius \( R \) depends on \( a \).

![Interval of convergence for generalized polynomials of \( \frac{1}{1-x} \) at 0 and \( \frac{1}{1-x} \) at \( \frac{1}{2} \)](image)

\[ \text{interval of convergence for} \]
\[ \text{generalized polynomials of} \quad \frac{1}{1-x} \quad \text{at} \quad 0 \]
\[ \text{interval of convergence for} \]
\[ \text{generalized polynomials of} \quad \frac{1}{1-x} \quad \text{at} \quad \frac{1}{2} \]

\text{Figure 11.3:}

**Generalized Polynomials**

There is a shorthand notation that is commonly used to express the limit of Taylor polynomials. In the special case when \( f(x) = \frac{1}{1-x} \) and \( p_n \) is the Taylor polynomial of \( f \) at 0, the shorthand notation for the limit of \( p_n(x) \) is

\[ 1 + x + x^2 + \ldots, \]

which is shortened even further by writing \( \sum_{n=0}^{\infty} x^n \). It must be emphasized that it makes no mathematical sense to add infinitely many numbers together, and the shorthand expressions don’t mean that! They are simply symbols that represent a limit of a sequence; in particular, \( \sum_{n=0}^{\infty} x^n \) is the symbol that represents the limit of the sequence

\[ 1, 1 + x, 1 + x + x^2, \ldots \]

With this notation we may express the results of the previous problems by writing

\[ 1 + x + x^2 + \ldots = \frac{1}{1-x} \]

when \(|x| < 1 \).

**Problem 11.26** Find \( 1 + \frac{1}{2} + \frac{1}{2^2} + \ldots \) and give the sequence this symbol represents the limit of.

**Problem 11.27** Find \( 1 + \frac{1}{3} + \frac{1}{3^2} + \ldots \) and give the sequence this symbol represents the limit of.
11.1. **POLYNOMIALS**

**Problem 11.28** Find \( 1 - \frac{1}{3} + \frac{1}{3^2} - \frac{1}{3^3} + \ldots \) and give the sequence this symbol represents the limit of.

**Problem 11.29** Find \( 1 + 2 + 2^2 + \ldots \) and give the sequence this symbol represents the limit of.

The advantage of the symbolism introduced for the limit of Taylor polynomials is that symbolic manipulations that one performs on polynomials also work on these limits. The symbolism suggests that the derivative of \( 1 + x + x^2 + \ldots \) is \( 1 + 2x + 3x^2 + \ldots \), and inside the radius of convergence this actually works. The general symbolism is obtained by taking the \( n \)th degree Taylor polynomial \( a_0 + a_1 x + \ldots + a_n x^n \) and adding three dots to suggest the remaining (infinitely many) terms that follow. The alternative symbolism is obtained using the sigma notation, where \( \sum_{i=0}^{\infty} a_i x^i \) is the same as

\[
a_0 + a_1 x + \ldots + a_n x^n + \ldots
\]

(which represents the limit of the Taylor polynomials at centered at 0).

**Problem 11.30** Introduce the symbolism that expresses the limit of the Taylor polynomials at \( a = 5 \) (see problem 11.19 for guidance).

By starting with one known expression such as

\[
\frac{1}{1-x} = 1 + x + x^2 + \ldots,
\]

several more expressions may be obtained by differentiating, integrating, adding, and multiplying. The expression on the right is called the Taylor series of \( \frac{1}{1-x} \) centered at 0.

**Problem 11.31** Differentiate to obtain a Taylor series for the function \( \frac{1}{(1-x)^2} \).

**Problem 11.32** Find \( 1 + 2(\frac{1}{2}) + 3(\frac{1}{2})^2 + \ldots \) and give the sequence this symbol represents the limit of.

**Problem 11.33** Integrate to obtain a Taylor series for the function \( f \) defined by

\[
f(x) = \int_0^x \frac{1}{1-t} \, dt
\]

If you start with a function \( f \) and then determine the coefficients \( a_0, a_1, a_2, \ldots \) so that

\[
f(x) = a_0 + a_1(x - a) + a_2(x - a)^2 + \ldots,
\]

then the object on the right of the equality is called the Taylor series of \( f \) centered at \( a \). If you don’t start with a function but are just interested in investigating objects of the form

\[
a_0 + a_1(x - a) + a_2(x - a)^2 + \ldots,
\]
these objects are traditionally called \emph{power series} centered at \(a\).

You are encouraged to think of a power series as a generalization of a polynomial. You are so strongly encouraged to adopt this point of view that we are going to call power series \emph{generalized polynomials}. Addition, multiplication, differentiation, and integration of generalized polynomials work exactly as they do for polynomials. In particular, just as two polynomials are equal exactly when they have the same coefficients, two generalized polynomials centered at \(a\) are equal exactly when their coefficients are equal. Symbolically, this is expressed by writing

\[a_0 + a_1(x - a) + a_2(x - a)^2 + \ldots = b_0 + b_1(x - a) + b_2(x - a)^2 + \ldots\]

if and only if \(a_i = b_i\) for all \(i\). While every polynomial is a generalized polynomial (they are the ones where there exists an \(n\) so that \(a_i = 0\) when \(i \geq n\)), not every generalized polynomial is a polynomial.

\textbf{Problem 11.34} Assume \(p\) is a polynomial. Prove that \(p^{(n)}\) is the zero function for some \(n\).

It is possible to discover conditions on the coefficients \(a_i\) that certain properties determine. For example, if you are looking to see if there is a generalized polynomial that is equal to its own derivative, by equating coefficients (as discussed above) you will force a relation between \(a_0\) and \(a_1\) (by equating coefficients of \(x^0\)). Looking at coefficients of \(x\), one sees a relation between \(a_1\) and \(a_2\), and in general one obtains a relation between \(a_n\) and \(a_{n+1}\) by equating coefficients of \(x^n\). Once a value is specified for \(a_0\), the relation between \(a_n\) and \(a_{n+1}\) then forces specific values for all of the coefficients.

\textbf{Problem 11.35} Determine the coefficients \(a_n\) for the Taylor polynomial

\[f(x) = a_0 + a_1x + a_2x^2 + \ldots\]

that has \(a_0 = 1\) and \(f = f'\).

It is now possible to see that every generalized polynomial that equals its derivative is a constant multiple of the one found in Problem 11.35. One does this by writing

\[h(x) = b_0 + b_1x + b_2x^2 + \ldots\]

and then proving that \(\frac{h}{b_0}\) is the function appearing in Problem 11.35.

\textbf{Problem 11.36} Let \(f\) be the Taylor polynomial determined in Problem 11.35 and assume

\[h(x) = b_0 + b_1x + \ldots\]

satisfies \(h = h'\). Show that \(h = b_0f\).

Let us continue to assume that \(f\) denotes the function determined in Problem 11.35. An \emph{inverse} function of \(f\) is a function \(g\) that satisfies

\[f(g(x)) = x\] and \(g(f(x)) = x\]
for all $x$ for which these equations are defined ($x$ must be in the domain of $g$ in the left equation and in the domain of $f$ in the equation on the right). For example, the cube root function is the inverse of the cubing function, logs invert exponentiation, and arcsin $x$ is the inverse of sin $x$. Not every function has an inverse (which we delve further into shortly), but if a function does have an inverse, and you know how to differentiate that function, then the formula that defines the inverse allows you to compute the inverse’s derivative. Assume for the moment that $f$ does have an inverse that we call $g$, and let us find the derivative of $g$.

**Problem 11.37** Take the derivative of both sides of the equation $f(g(x)) = x$ (using the formula for the chain rule on the left) and use the fact that $f = f'$ to obtain a simple formula for $g'$: (The derivative of $f(g(x))$ is obtained from the chain rule, which expresses the derivative in terms of a derivative of $f$ and a derivative of $g$. See any Calculus book and look up the chain rule in the index.)

**Problem 11.38** With $g$ given in Problem 11.37, show that $g(2)$ equals the area depicted in Figure 11.4.

![Figure 11.4: Graph of $\frac{1}{x}$](image)

Knowing what the derivative of $g$ is enables one to prove that

$$g(ax) = g(a) + g(x)$$

for all numbers in the domain of $g$. Use the following steps to establish this equality in Problem 11.39:

1. Take the derivative of $g(ax)$ (use the chain rule and treat $a$ as a constant and $x$ as the variable).

2. Use the result of Problem 11.37 to simplify the derivative of $g(ax)$.

3. Take the derivative of $g(a) + g(x)$ (again treating $a$ as a constant and $x$ as the variable).
4. Use the fact that two functions with the same derivative must differ by a constant (this is Problem 8.13).

5. The last step is to determine what the constant is that the two functions differ by; this can be done by setting $x = 1$ and figuring out what number $g(1)$ is.

**Problem 11.39** Prove that $g(ax) = g(a) + g(x)$.

**Problem 11.40** Use Problem 11.39 to show that
\[ f(a + x) = f(x)f(a) \]
for all numbers in the domain of $f$.

Both Problem 11.40 and the last step of Problem 11.39) are done by recalling the relationship that defines what it means to say that $f$ and $g$ are inverses of each other.

**Problem 11.41** Let $e$ denote the output of $f$ when 1 is input; i.e. define
\[ e = f(1). \]
Show that $e^2 = f(2), e^3 = f(3)$, and in general $e^n = f(n)$ for all $n \in \mathbb{N}$.

**Problem 11.42** Show that $e^{\frac{1}{n}} = f\left(\frac{1}{n}\right)$ for all $n \in \mathbb{N}$.

**Problem 11.43** Show that $e^x = f(x)$ for all $x \in \mathbb{Q}$ (Hint: Let $x = \frac{m}{n}$ and so $f(x) = f\left(\frac{1}{n} + \ldots + \frac{1}{n}\right)$ where $\frac{1}{n} + \ldots + \frac{1}{n}$ contains $m$ summands.)

You certainly know what someone means when they write $2^3, 2^\frac{1}{2}$, or even $2^\sqrt{2}$. But what does $2^{\sqrt{3}}$ mean? It makes no sense intuitively to multiply a number by itself an irrational number of times. What is needed is a reasonable definition of $2^{\sqrt{3}}$. The previous examples involve a base other than 2, but for that base it is pretty clear how irrational powers should be defined. If $x$ is any real number, define
\[ e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots. \]
Thus we now have a perfectly good definition of $e^x$, even though both the base and the exponent are irrational numbers!

**Inverse Functions**

**Problem 11.44** Draw the graph of the inverse of $f$ in Figure 11.5.

**Problem 11.45** Indicate an interval on which the function $f$ given in Figure 11.6 is invertible and graph the inverse function.
Figure 11.5: Graph of $f$

**Problem 11.46** Prove that the function $f$ given in Figure 11.6 is not one-to-one.

**Problem 11.47** Use the Mean Value Theorem to prove the following: if $f$ is differentiable and $f'(x) \neq 0$ for all $x$, then $f$ is one-to-one.
Index

addition, 14
  definition, 14
identity, 15, 18
inverse, 15
of functions, 55
of sequences, 37
antiderivative, 78, 87, 88
Archimedian property, 41
assumption, 9–11
axiom, 9, 10
  as logical statement, 9
  as used in proofs, 29
  definition, 9
bounded, 37
calculus, 21, 87
  differential, 23
  Fundamental Theorem of, 26
  integral, 21
Cartesian plane, 2, 17
  circle as subset, 3
  subset of, 6
chain rule, 66
circle, 3
Complete Ordered Field, 39
  axiom, 39
complex numbers, 2
composition, 55
conclusion, 30
constant function, 53
  derivative, 65
constant sequence, 36
continuity, 57, 88, 91
  definition, 44
  intuition, 12
  of composition, 55
  on an interval, 46
  one-sided, 45
uniform, 85
contradiction, 40
contrapositive, 31
convergence, 36, 37
  bounded, 37
  interval of, 97
  notation, 37
  radius of, 97
converse, 29, 30
counterexample, 30, 36
critical point, 96
deduction, 9–11
degree, 53, 95
derivative, 23, 25, 47, 63
  appearance of factorial, 96
  constant, 26, 65
  definition, 63
differentiable
  at a point, 63
  implies continuity, 65
  on a set, 64
does not exist, 26
  figure, 26
  limit of secant slope, 63
  of constant function, 65
  of functions differing by a constant, 28
  of polynomial, 64
  one-sided, 64
  output, 26
two functions with the same, 67
differential calculus, 23
  chain rule, 66
computational theorems, 66
derivative, 23, 25
power rule, 66
product rule, 65
quotient rule, 66
secant line, 25
tangent line, 25
theoretical theorems, 66
discontinuity, 44
division, 15
domain, 5, 6, 43
equivalent, 30, 37
proving statements to be, 82
Euclidean 3-space, 2
Extreme Value Theorem, 57
proof, 60
statement, 57
factorial, 96
Field, 14
addition, 14
definition, 14
identity, 15, 18
inverse, 15
subtraction as, 15
definition, 14
multiplication, 14
definition, 14
division as, 15
identity, 15, 18
inverse, 15
of two elements, 16, 18, 19
order on, 16
ordered, 16
representation, 14
Field Axioms, 14
function, 4–6
addition, 55
as a subset, 4
as defined by a formula, 4
as defined by a graph, 4, 5
composition, 55
constant, 53
continuous, 44, 57
composition, 55
on a set, 44
on an interval, 46
one-sided, 45
upper bound, 59
critical point, 96
decreasing, 67
definition, 5
degree, 53
domain, 5, 6
increasing, 67
input, 5
representation, 7
inverse, 100
inverse trigonometric, 94
limit, 43
intuition, 43
notation, 44
logarithm, 94
multiplication, 55
not defined, 5
not defined at a point, 7
output, 5
representation, 7
polynomial, 93
representation, 5, 7
root, 94
used to approximate, 93
with formula, 93
without formula, 5
approximation, 93
Fundamental Theorem of Calculus,
26, 87
generalization, 87
relating integration and differentiation, 27
statement, 87
generalized polynomial, 99
equality, 100
genrealized polynomial
notation, 97
radius of convergence, 97
graph, 4
as a subset, 4
defining a function, 4
not defining a function, 5
of a function, 4
greatest lower bound, 41, 76, 81, 90
least upper bound, 41, 60, 75, 81

INDEX

hypothesis, 30
if and only if, 30
if-then statement, 29, 35
as definition, 31
as theorem, 37
contrapositive, 31
converse, 30
as in definitions, 32
counterexample, 30
defined to be true, 30
hypothesis, 30
if and only if, 30
logical definition, 29
proving true, 31

increasing
function, 67
sequence, 39

induction, 54

integers, 2

integral, 22
addition, 89
calculus, 21
definition, 76
function, 22
negative, 22
sum of parts, 23
symbol, 22

integration, 22, 71, 76
antiderivative, 78
calculation, 78
definition, 76
inscribed rectangles, 74
intuition, 71
logarithm, 94
notation, 76
partition, 73, 79
sequence, 83
simple, 80
subset of, 80
with subintervals of equal length, 84

Reimann sum, 89
Riemann sum, 74
lower, 74
upper, 74
sum of areas of rectangles, 72
superscribed rectangles, 74

Intermediate Value Theorem, 61
intersection, 4
interval, 4
closed, 35
half open, 35
open, 35
inverse, 100
existence, 100
graph, 102
interval, 102

inverse trigonometric function, 94
irrational numbers, 2, 3

least upper bound property, 41, 60
least upper bound, 41, 60
lemma, 14, 58
limit, 21
of a function, 43
definition, 49

equiv
talen
t definition, 50
existence, 45, 50
one-sided, 45, 50
unique, 52
of a sequence, 33
unique, 40
of a Taylor polynomial, 98
of average velocities, 25

linear, 23
linear algebra, 23
logarithm, 94

lower bound, 41
greatest, 41

Mean Value Theorem, 66, 78, 102
Rolle's Theorem, 68

multiplication, 14
by a negative number, 16
definition, 14
identity, 15, 18
inverse, 15
of functions, 55

natural numbers, 1, 2

numbers
  complex, 2
  integers, 2
  irrational, 2, 3
  natural, 2
  order on, 4
  rational, 2, 15
  real, 2, 10, 18

one-to-one, 102

order, 4
  definition, 16
Order Axioms, 16

Ordered Field, 16
  axioms, 16
  complete, 39
  deficient, 19
  definition, 16
  sequence on, 33

parabola, 53, 96
partition, 73
plane, 2
point, 4
polynomial, 53, 93
  antiderivative, 94
  critical point, 96
  degree, 95
  derivative, 64
  determined by, 96
  generalized, 99
  quadratic, 95
  quintic, 96
  rational functions, 94
  root, 94
    quadratic formula, 95
  Taylor, 97
power rule, 66
power series, 99
  generalized polynomial, 99
product rule, 65
proof
  by contradiction, 40

by contrapositive, 31
by counterexample, 30
by induction, 54
by logical deduction, 29
  of converse, 30
propoosition, 14
Pythagorean Theorem, 13
quadratic, 95
  parabola as graph, 96
quadratic formula, 53, 95
quintic, 96
quotient rule, 66

rational numbers, 2, 61
  as a field, 15
  as an ordered field, 16
  definition, 2
  order on, 4
real numbers, 2
  Archimedean property, 41
  as an ordered field, 18
  as complete ordered field, 39
  characterizing axioms, 10
  closed interval, 57
  closed intervals, 59
  definition, 2
  function on, 93
  order on, 4
Riemann sum, 74
  lower, 74
  upper, 74
Rolle's Theorem, 68
root, 94
sequence, 33
  addition, 37
  constant, 36
  convergent, 36, 37
    bounded, 37
    notation, 37
  decreasing, 39
  definition, 34
  eventually in a set, 36
  increasing, 39
  intuition, 33
limit, 33
  intuition, 33
  unique, 40
  not reaching limit, 49
  notation, 34
  of partitions, 83
  of polynomials, 97
  subsequence, 58

set
  definition, 1
  elements, 1
  positive, 2
  representation, 2
  empty, 2
  infinite, 1
  intersection, 4
  notation, 1
  order on, 4
  representation, 2
  subset, 3
  interval, 4
  union, 4

space
  Euclidean, 2
  squeeze theorem, 36, 91
  subsequence, 58
  convergent, 58
  definition, 58
  existence lemma, 59

subtraction, 15

Taylor polynomial, 97
  as sequence, 97
  coefficient, 100
  interval of convergence, 97
  limit, 98
  sigma notation, 99
  power series, 99

Taylor series
  at a point, 99
  theorem, 14, 37
  triangle inequality, 17

union, 4

upper bound, 41, 59, 75
  least, 41

velocity, 24
  average, 25
  limit of average, 25