Analysis for the Beginning Mathematician

Michael Klipper

Preface

This has not yet been written.

Contents

List of Figures

Chapter 1

Introduction to Reasoning

1.1 The Language of Logic

Before we can really delve into mathematics, and calculus in particular, we need to make some important conventions about how we present our ideas to one another. Calculus has some very powerful ideas, but they are easy to abuse if we aren't careful about what steps we are allowed to do as we solve problems.

Why should we be formal?

The English language, or any language that people use, is full of ambiguities. Suppose I say, "Either I will go to the store, or I'll mow the lawn." If I go to the store but don't mow the lawn, then you'll probably believe I said something true. If I mow the lawn but don't go to the store, then you'll think I was true as well. But what if I do both? Was I lying?

Here's another puzzler. Suppose I say, "Everyone owes somebody money." There are two ways to interpret this statement. The first way says each person has their own debt to pay, but possibly to different people. (For instance, I might be paying off loans for a car, while you're paying off loans for school.) However, there is a second way: maybe one person in the world is so powerful and so influential that everyone owes THAT person money. Which of these two meanings did I intend? Normally, in speech, people use context to tell the difference, but in mathematics, the situation can be rather complicated and it may become harder to tell which interpretation "makes the most sense."

These examples illustrate the need for precision, and careful logic will help us be more precise. However, there is another important benefit to being more formal. In many cases, when we work more formally, the logic imposes some structure on the problem, giving us ideas of how to proceed. We'll see examples of this later in this chapter.

1.1.1 Statements and Connectives

The simplest building blocks of logic are *statements*. A statement is roughly an expression which is either true or false. Thus, sentences like "3=3" and "every prime number is odd" are statements, but "3" and "x+1" are not. Often we use capital letter variables like P and Q to represent statements.

We want to break down sophisticated statements into simple components to make our analysis easier. For example, the complicated statement "If I don't go to the store, then I'll be hungry" is made out of the simple pieces "I don't go to the store" and "I'll be hungry", joined with the word "if". The word "if" is an example of a *connective*. Connectives glue statements together; popular examples include the words "and", "or", and "if". Let's look at what different connectives do.

Negation

To get ourselves started, we'll look at the simplest connective, "not".

Definition 1.1. If P is a statement, then the *negation* of P, denoted $\neg P$ or "*not* P", is true when P is false, and it is false when P is true.

Frequently, to describe complex statements, we draw a *truth table* to describe the values of a statement in terms of its simple pieces (these are frequently variables). To illustrate what goes in a truth table, let's look at the table for $\neg P$:

P	$\neg P$
T	F
F	T

A truth table has columns for each of the simple pieces (here, there's only one simple piece, P) and columns for each of the more complex statements (the only one here is $\neg P$). The rows start with the possible combinations values that the simple pieces can have (each piece can be either true T or false F). The corresponding values for the other columns are filled in (so the first row should be read as "when P is true, $\neg P$ is false"). Negation only applies to one statement, so we say it is a *unary* connective.

And and Or

Negation is the only basic connective to apply to one statement. Most connectives work with two statements, so they are *binary* connectives. We introduce a couple more connectives.

Definition 1.2. The *conjunction* of P and Q, denoted $P \wedge Q$ or "P *and* Q", is true when P and Q are both true; otherwise it is false.

The truth table for $P \wedge Q$ has two simpler components, P and Q, each of which can take 2 values, so the table has $2 \cdot 2 = 4$ rows:

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

Definition 1.3. The *disjunction* of P and Q, denoted $P \vee Q$ or "P *or* Q", is false when P and Q are both false; otherwise it is true.

The truth table:

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

Note that this definition of "or" does consider "true or true" to be true. For example, the statement "3 is positive, or 3 is prime" is true. However, in common speech, this case is usually meant to be false, like when a waiter says, "You may have soup or salad for that price." When this meaning is intended, the connective is instead called *exclusive or*, or *xor* for short.

Example 1.4:

Let's make a truth table for the statement $P \wedge (\neg Q \vee P)$. (Note that parentheses are frequently necessary when dealing with layers of complexity in your statement.) The basic components are P and Q. After that, $\neg Q$ exhibits the next layer of complexity, followed by $\neg Q \vee P$ and lastly $P \wedge (\neg Q \vee P)$. We'll make a column in our truth table for each of these statements, so that we only need to keep one connective in mind at any time when building the table.

Our truth table is:

P	Q	$\neg Q$	$\neg Q \vee P$	$P \wedge (\neg Q \vee P)$
T	T	F	T	T
T	F	T	T	T
F	T	F	F	F
F	F	T	T	F

Note that $P \wedge (\neg Q \vee P)$ is true precisely when P is true. We say that $P \wedge (\neg Q \vee P)$ is *equivalent* to P, and the truth table proves this equivalence. Thus, we may replace anything of the form $P \wedge (\neg Q \vee P)$ with P in any statement, possibly making our work easier.

There are many other examples of equivalent statements, which allow us to do simplification. See the exercises for a number of useful equivalences. $\square$

If-Then

For our next connective, we would like to address statements like "if P, then Q." (Note that the words "is true" are frequently omitted after variables.) When should "if P, then Q" be true? When P is true, Q should be true in order for "if P, then Q" to be true. However, what do we do if P is false? Should "if there is a cow on the moon, then there is a purple cow on the moon" be true? What about "if there is a cow on the moon, then $0 = 0$"?

By convention, we decide that these statements are true, because there is no cow on the moon. This leads to the following definition:

Definition 1.5. The *implication* $P \to Q$, also called "P *implies* Q", "*if* P *then* Q", etc., is false only when P is true and Q is false; otherwise it is true. P is called the *hypothesis*, and Q is called the *conclusion*.

The truth table:

P	Q	$P \to Q$
T	T	T
T	F	F
F	T	T
F	F	T

This definition may seem uncomfortable at first, but it's absolutely necessary for mathematics. The goal of a proof in mathematics is to start from some hypotheses and to work toward conclusions, with each step along with the way being a true implication. If the hypotheses of a theorem are false in some situation, then that doesn't make the theorem wrong; the theorem is still true, but it doesn't have any use in that situation. Try looking at $P \to Q$ as a promise: if $P \to Q$ is true, then we promise to prove Q when provided a proof of P. If you fail to show me P is true, then we don't have to do anything, and we have not broken our promise.

The last connective we will mention here will be defined in terms of the previous connectives.

Definition 1.6. $P \leftrightarrow Q$, also written as $P \equiv Q$ or "P *if and only if* Q", is defined as $(P \to Q) \wedge (Q \to P)$.

Thus, $P \leftrightarrow Q$ is true if each of P and Q implies the other. Since $P \leftrightarrow Q$ is built out of $P \to Q$ and $Q \to P$, we'll include them in our truth table:

P	Q	$P \to Q$	$Q \to P$	$P \leftrightarrow Q$
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

Thus, $P \leftrightarrow Q$ is true when P and Q have the same value. Since the phrase "if and only if" is so common in mathematics, it is often abbreviated as iff!

Remark. In some other textbooks, double-thickness arrows, like $\Rightarrow$ and $\Leftrightarrow$, are used instead of $\to$ and $\leftrightarrow$.

1.1.2 Sets and Quantifiers

Now that we have basic statements, we would like to make statements that are more sophisticated. While our current connectives allow for a lot of expressibility, they don't provide us any simpler way of breaking down many important mathematical statements like "Every real number is smaller than some other real number." We need to discuss how to encode words like "every" and "some" into mathematical statements. Also, since mathematical statements rarely talk about EVERY object in the known universe, we need a device to be able to talk about all objects within some restricted group. Sets and quantifiers will provide the devices we need.

Sets

A *set* is a collection of objects. The objects in a set are called *elements*, or *members*, or any other common synonym. Frequently, sets are specified by giving some property of their elements. For example, "the real numbers" form a set, but there are also non-mathematical sets, like "the people I have met", and "the pizzas I sold yesterday". We often use capital letters like S or T to represent sets and lowercase letters like s or t to represent elements. We use the notation $x \in S$ to say "x is a member of S". To say x is not a member of S, we may write $x \notin S$ instead of $\neg(x \in S)$.

Certain common sets get special names to represent them. The set of real numbers is denoted by $\mathbb{R}$ (that style of writing R is called "blackboard bold", in case you're curious). The set of complex numbers is $\mathbb{C}$. The set of integers is $\mathbb{Z}$ (from the German "Zahlen", meaning "number"). The rational numbers, which are fractions of integers, form the set $\mathbb{Q}$ (for "Quotient"). The natural numbers are denoted by $\mathbb{N}$; however, not all mathematicians agree which numbers ought to be "natural". Specifically, it is quite controversial whether 0 ought to be natural. This book will define $\mathbb{N}$ to consist of all positive integers and 0, as 0 is quite natural (at least to the author) to use in many situations. If we wish to talk about just the positive integers, then we'll use the symbol $\mathbb{N}^*$.

In our discussion in this chapter, we will take for granted some familiarity with these basic sets and their properties. Many of these results will be proven later, especially in Chapter 2, where we study $\mathbb{R}$ more formally.

Free Variables

How can we incorporate sets into our statements? As a starting point, sentences like "$x \in \mathbb{R}$" seem to be statements. Unfortunately, there's one problem here: we don't know if that sentence is true or false unless we know what x is. However, as soon as a value of x is specified, then we obtain a statement: "$3 \in \mathbb{R}$", "$4 \in \mathbb{R}$", and "(the President of the U.S.A.) $\in \mathbb{R}$" are three statements (the last one is false).

In some sense, "$x \in \mathbb{R}$" is an abbreviation for a whole collection of statements, where each value of x produces a new statement. We call this kind of phrase a *proposition*. x is called a *free variable* of the proposition, and we write $P(x)$, read as "P of x", to denote that P may depend on x. (x can also be called a *parameter* of P.) We can have multiple free variables in a proposition: $P(x, y)$ (read as "P of x and y") denotes a proposition P depending on two values x and y, and so forth. Any free variable occuring in a proposition MUST be listed in the parenthesized list after the proposition name (though possibly some extra variables are listed too). Statements are considered the same as propositions with no free variables.

For example, if $P(x)$ denotes "$x = 3$", then $P(x)$ is a proposition with one free variable x. However, propositions with free variables occur outside mathematics too: if S is the set of students in your school, then "$x \in S$, $y \in S$, and x loves y" and "$x \in S$, $y \in S$, and x and y are in a class together" are propositions with two free variables x and y.

The connectives can also be used with propositions, yielding other propositions. For this reason, the distinction between proposition and statement is commonly blurred, and many people use the words proposition and statement interchangeably.

Quantifiers

Now that we have introduced free variables, we need ways to deal with them effectively. There are many times in mathematics where we want to talk about all possible values of a variable, like in the statement "Every even integer n greater than 2 is not prime." In this example, the variable n is supposed to range over all even integers. There are also many times where we want to talk about existence of some value, as in "There exists a complex number i with $i^2 = -1$." For these scenarios, we use new tools called *quantifiers*. We use two kinds of quantifiers.

Definition 1.7. If $P(x)$ is a proposition and S is a set, then $\forall x \in S\ P(x)$ is true precisely when $P(x)$ is true for every member x of S. It is read as "*for* *all x in S, $P(x)$*" (though this is not the only way to read $\forall x \in S\ P(x)$ in English). The symbol $\forall$ is the *universal quantifier*. (Think of an upside-down A standing for "All".)

Definition 1.8. If $P(x)$ is a proposition and S is a set, then $\exists x \in S\ P(x)$ is true precisely when $P(x)$ is true for at least one value of x in S. It is read as "*for some x in S, $P(x)$*" (this is also not the only way to read it). The symbol $\exists$ is the *existential quantifier*. (Think of a backwards E standing for "Exists".)

For example, "every integer n greater than 2 is not prime" can be alternately written as, "for every integer n, if $n > 2$ then n is not prime", which is "$\forall n \in \mathbb{Z}\ (n > 2 \to n$ is not prime)". This is a *quantified* statement. For another example, the statement "there exists a complex number i with $i^2 = -1$" can be written as "$\exists i \in \mathbb{C}\ (i^2 = -1)$".

These are the key ingredients to a quantified proposition:

- A quantifier symbol.
- A variable. This variable is said to be *bound* by the quantifier. The actual name of the bound variable is not important; "every integer n is odd" means the same thing as "every integer k is odd", since n and k are just placeholders to represent arbitrary integers.
- A set of possible values for the quantified variable, called the *domain* of the variable. This lets us restrict our attention to values that we actually want to study. Occasionally people will write quantifiers with no domain (like $\forall x\ x = x$), but then a domain is understood by context (sometimes that domain is called the *universal set of discourse*). However, we should adopt good habits and make our domains clear.
- A proposition. This forms the *scope* of the variable, or it is also called the *body* of the quantifier.

Example 1.9:

Consider the proposition $\forall x \in \mathbb{R}(y - x > 0)$. Its body is $y - x > 0$, which has x and y free, but once the quantifier is added, x is no longer free; x is bound by the quantifier. The variable y is still free, however. Call this quantified proposition $Q(y)$.

Why do we distinguish between free and bound variables? When looking at $Q(y)$, we need to know a value for y before we can start to decide if $Q(y)$ is true or false. Hence, y is "free for the reader to choose". However, x is assigned a meaning by the quantifier. We don't have free choice for a value of x: the quantifier forces us to consider ALL real numbers x to decide if $Q(y)$ is true. In essence, when it comes to x, our hands are tied.

Since the quantifier gives x its meaning, once we leave the body of the quantifier, the quantifier is done and x no longer makes sense. Consider the phrase "$(\forall x \in \mathbb{R}\ (x = x)) \wedge (x = 3)$": this is not a well-formed statement, as the scope of x is $x = x$, and the last occurrence of x in our formula falls outside that scope. To fix this, we should change the bound variable name of x to something different, such as y, and we get the legitimate proposition "$(\forall y \in \mathbb{R}\ (y = y)) \wedge (x = 3)$" with x free. $\square$

Example 1.10:

Quantifiers can be even be placed inside other quantified propositions! To keep things from being confusing, however, different variable names should be used for each quantifier. For example, consider the sentence "for each real number x, there is some real number y with $x + y = 0$", which says that each real number has a negation. This can be written as "$\forall x \in \mathbb{R}\ (\exists y \in \mathbb{R}\ (x + y = 0))$". When one quantifier immediately follows another, we often can omit parentheses around the scope of the first one, so we may write this also as "$\forall x \in \mathbb{R}\ \exists y \in \mathbb{R}\ (x + y = 0)$". $\square$

Example 1.11:

In Section 1.1, we discussed the sentence "Everyone owes somebody money." The first interpretation is "For each person, there exists a person to whom money is owed". How is that written in logic? It says that given an arbitrary person, there exists some other person DEPENDING ON OUR FIRST PERSON to whom the money is owed. If P is the set of people, then we write this as "$\forall x \in P\ \exists y \in P$ (x owes y money)". The idea here is that we first take an arbitrary person x, and then afterwards we look for some specific person y, depending on x, which is the one to whom x owes money. Thus, $\exists y$ comes after $\forall x$, as you can't pick y based on x if you haven't picked x!

The second interpretation, however, is "There is one person to whom everyone owes money". In this setting, we FIRST choose the person to whom everyone owes money, and then we talk about arbitrary people. Thus, this meaning is denoted by "$\exists y \in P\ \forall x \in P$ (x owes y money)". Here, y is chosen independently of x.

This subtlety about quantifier orders is quite important in mathematics too. The statement "$\forall x \in \mathbb{R}\ \exists y \in \mathbb{R}\ (x < y)$" is true (because once x is given, y can be chosen as $x + 1$), but "$\exists y \in \mathbb{R}\ \forall x \in \mathbb{R}\ (x < y)$" is false, as there is no largest real number! □

We now have the ability to express basically every mathematical statement we wish. However, there are many equivalent ways of expressing the same statement with quantifiers. The exercises will cover some important equivalences, and we'll point out another here as an example.

Example 1.12:

Consider the awkward-sounding statement "it is not true that every student passed the exam". A better way of writing this is "there is some student who did not pass the exam" (or even better, "some student failed the exam"). However, it is NOT correct to rewrite this as "everyone failed the exam"; only one student needs to fail the exam in order to spoil a 100% pass rate for the class. In symbols, if S is the set of students, then this shows that "$\neg(\forall p \in S\ (p \text{ passed the exam}))$" is equivalent to $\exists p \in S\ (\neg(p \text{ passed the exam}))$" and is NOT equivalent to "$\forall p \in S\ (\neg(p \text{ passed the exam}))$".[1]

In general, if $P(x)$ is any proposition and S is any set, then the two statements

$$\neg(\forall x \in S\ P(x)) \qquad \exists x \in S\ (\neg P(x))$$

are equivalent. It's also true that the two statements

$$\neg(\exists x \in S\ P(x)) \qquad \forall x \in S\ (\neg P(x))$$

are equivalent. These two equivalences are very helpful, because they allow us to rewrite statements with negations outside quantifiers as statements with negations inside quantifiers. Hence, these equivalences are often called "negation-pushing" rules, since the negation is pushed inside of a quantifier. Note that the quantifier "flips" from $\forall$ to $\exists$ or vice versa when performing a negation push.

By continually applying equivalences like these, we can make statements much more readable. Consider, for instance, the sentence "not everyone lives with someone". If P is the set of people, then this is "$\neg(\forall x \in P\ \exists y \in$

[1] When we say "not (passing the exam)" is equivalent to "failing the exam", we are assuming that every student in the class took the test! This may be a safe assumption, depending on how the class is run, but it's important to acknolwedge this assumption.

P (x lives with y)). Doing one push yields "For some person, there does not exist someone with whom they live", i.e. "$\exists x \in P\ \neg(\exists y \in P\ x$ lives with y)". Pushing once again yields the much better-sounding sentence "someone lives with nobody", i.e. "$\exists x \in P\ \forall y \in P$ (x does not live with y)".[2] $\square$

1.2 Exercises

1. Build truth tables for the following statements in terms of P and Q:

   (a) $P \to (Q \to P)$
   (b) $(P \to Q) \to P$
   (c) $P \lor \neg(P \land Q)$
   (d) $(P \land Q) \leftrightarrow \neg Q$

2. Build truth tables for the following statements in terms of P, Q, and R (these truth tables will need 8 rows since there are $2 \cdot 2 \cdot 2 = 8$ ways to choose the values of P, Q, and R):

   (a) $P \to (Q \land \neg R)$
   (b) $(R \land P) \lor (R \land Q)$

3. We mentioned the "xor" connective briefly after mentioning "or". While "P or Q" is true when at least one of the two statements P, Q is true, "$P \operatorname{xor} Q$" is true when exactly one of the two statements is true.
   (a) Build a truth table for "$P \operatorname{xor} Q$".
   (b) How can we write "$P \operatorname{xor} Q$" in terms of the connectives $\lor$, $\land$, and $\neg$ (i.e. "or", "and", and "not")?

4. Show the following equivalences are true for all statements P,Q,R:
   (a) *Commutativity:*
   $(P \land Q) \leftrightarrow (Q \land P)$ and $(P \lor Q) \leftrightarrow (Q \lor P)$
   (b) *Associativity:*
   $(P \land Q) \land R \leftrightarrow P \land (Q \land R)$ and $(P \lor Q) \lor R \leftrightarrow P \lor (Q \lor R)$
   (c) *Distributivity:*
   $P \lor (Q \land R) \leftrightarrow (P \lor Q) \land (P \lor R)$ and $P \land (Q \lor R) \leftrightarrow (P \land Q) \lor (P \lor R)$
   These are often described by saying "or" distributes over "and" (for the first one) and "and" distributes over "or" (for the second one.)

[2] As a convention, we believe that a person does not live with himself / herself. You are welcome to disagree, however.

(d) *Double Negation:* $\neg\neg P \leftrightarrow P$

(e) *DeMorgan's Laws:*
$\neg(P \vee Q) \leftrightarrow (\neg P \wedge \neg Q)$ and $\neg(P \wedge Q) \leftrightarrow (\neg P \vee \neg Q)$
These laws allow us to push negations inside of complicated statements.

(f) $\neg(P \to Q) \leftrightarrow (P \wedge \neg Q)$

5. If P is the set of people, $K(x, y)$ means "x knows y", and m is "myself" (so $m \in P$), then write the following English sentences using quantifiers:
   (a) I know everyone.
   (b) Nobody knows me.
   (c) Everyone I know also knows me.
   (d) I know someone who does not know himself/herself.

6. In the following propositions, identify the free variables, the bound variables, and the scopes of the bound variables:
   (a) $\forall x \in \mathbb{R}\ ((\exists y \in \mathbb{R}\ xy = 0) \vee (xz = 1))$
   (b) $(\exists z \in \mathbb{R}\ \forall x \in \mathbb{R}\ x + z = x) \wedge (\exists t \in \mathbb{R}\ y + t = 2)$

7. For each of these following statements, if it is true for all propositions $P(x)$, $Q(x)$, and all sets S, then explain why. Otherwise, find a counterexample (i.e. pick a $P(x)$, $Q(x)$, and S which makes the statement false.)
   (a) $(\forall x \in S\ (P(x) \wedge Q(x)) \leftrightarrow (\forall x \in S\ P(x)) \wedge (\forall x \in S\ Q(x))$
   (b) $(\forall x \in S\ (P(x) \vee Q(x)) \leftrightarrow (\forall x \in S\ P(x)) \vee (\forall x \in S\ Q(x))$
   (c) $(\exists x \in S\ (P(x) \wedge Q(x)) \leftrightarrow (\exists x \in S\ P(x)) \wedge (\exists x \in S\ Q(x))$
   (d) $(\exists x \in S\ (P(x) \vee Q(x)) \leftrightarrow (\exists x \in S\ P(x)) \vee (\exists x \in S\ Q(x))$

8. If $P(x)$ is a proposition and x is a set, then we know "$\exists x \in S\ P(x)$" means that $P(x)$ is true for at least one $x \in S$.
   (a) How would you use connectives and quantifiers to express "$P(x)$ is true for EXACTLY one $x \in S$"? In other words, come up with a way to write "there exists a unique $x \in S$ satisfying $P(x)$".
   (b) How would you express "$P(x)$ is true for exactly two members $x, y \in S$"? (Thus, x and y have to have distinct values.)

9. Using the negation-pushing rules in Example 1.12 and DeMorgan's Law from Exercise 1.2.4.(e), rewrite the following statements by pushing the negations as much as possible. For example, $\neg(\forall x \in S\ (P(x) \wedge Q(x))$ becomes $\exists x \in S\ \neg(P(x) \wedge Q(x))$ by negation pushing, which then becomes $\exists x \in S\ (\neg P(x) \vee \neg Q(x))$ by DeMorgan's Law.
   (a) $\neg(\forall x \in \mathbb{R}\ \forall y \in \mathbb{R}\ (x < y))$
   (b) $\neg(\forall x \in \mathbb{R}\ (x = 0 \vee \exists y \in \mathbb{R}\ (xy = 1))$
   (c) $\neg((\forall x \in \mathbb{Q}\ x^2 \geq 0) \wedge (\exists x \in \mathbb{Q}\ x^2 = -1))$

1.3 Operations with Sets

Sets are important in mathematics, so it's important to have convenient ways of specifying and building sets. Much like we build sophisticated statements out of simple ones using connectives, we operations for building more complicated sets out of simple ones.

Defining a Set from Scratch

Before we have any other sets to use, we have a couple ways to build sets. The first, and simplest, option is to list all the elements of a set in curly braces. Thus, $\{1, 2, 5\}$ is the set whose three members are 1, 2, and 5. One set with four countries in it is $\{\text{U.S.A.}, \text{Mexico}, \text{Canada}, \text{Belize}\}$. Our elements listed can be of different types, like $\{\text{Uganda}, 3\}$. In fact, the elements of our sets could be other sets! For example, $\{1, 2, \{1\}\}$ is a set with three elements: the number 1, the number 2, and a set whose sole member is 1 (if you think of sets as boxes holding items, then think of this example like a small box inside of a big box).

We can even abuse this notation slightly and use ellipses (i.e. $\ldots$) when we wish to convey a pattern that the elements follow, without actually writing every single element. For example, $\{1, 2, \ldots, 10\}$ really means the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, and $\{-3, -2, -1, 0, 1, \ldots\}$ really means the set of all integers that are at least as big as -3 (note that this last set is an infinite set, unlike the previous examples). If you plan on using ellipses in set notation, then it should be very clear what the pattern is.

One particularly special example is the following:

Definition 1.13. The *empty set*, written as $\emptyset$ or $\{\}$, is the set with no members.

When talking about a set, all that matters is which elements belong to the set and which do not. $\{1,3,2\}$ and $\{1,2,3\}$ are the same set, since the same three integers appear in each; the order of presentation does not matter for a set. Similarly, writing an element multiple times does not matter; the sets $\{1,1,2\}$ and $\{1,2\}$ are the same, since they both contain precisely the elements 1 and 2. The only way to show two sets are different is to find a member of one of them which is not in the other: $\emptyset \neq \{\emptyset\}$ since $\emptyset \in \{\emptyset\}$ and $\emptyset \notin \emptyset$ (an empty box is not the same as a box with another empty box inside!).

Defining a Set with a Property

The second basic option is to use some proposition to specify a set, like saying "the set of all real numbers less than 2". If $P(x)$ is a proposition (like "$x \in \mathbb{R} \wedge x < 2$"), then $\{x \mid P(x)\}$ or $\{x : P(x)\}$ is written to denote the set of all objects x so that $P(x)$ is true, and we read this as "the set of x such that $P(x)$ is true". Sometimes other synonyms are used, such as saying "$P(x)$ holds" instead of "$P(x)$ is true". Also, when used in this context, the proposition $P(x)$ is sometimes called a *property* of the elements, so $\{x \mid P(x)\}$ is said to be the set of x values with property P.

Here, x is a variable bound to the set, so it is no longer in scope once the set has been finished, similar to how bound variables in quantifiers work. For example, consider $\{n \mid n \in \mathbb{Z} \ \wedge \ n \geq -3\}$, which is the set of all integers greater than or equal to -3. After the $\}$ symbol, the letter n doesn't mean anything anymore. Just like with variables bound to quantifiers, the variable name can be changed without harm, so our set could also be called $\{k \mid k \in \mathbb{Z} \ \wedge \ k \geq -3\}$.

There are other notational shortcuts we use to make life easier. For instance, we can list several propositions together with commas separating them instead of using the $\wedge$ symbol, as in $\{r \mid r \in \mathbb{R}, r^2 < 2\}$, the set of real numbers whose squares are less than 2. Another very common practice is, just like we specify domains for quantifiers, we can also specify domains for sets made out of a property. For instance, $\{r \in \mathbb{R} \mid r^2 < 2\}$ is the same as the set $\{r \mid r \in \mathbb{R}, r^2 < 2\}$ but looks much better. It's like the difference between saying "the set of real numbers r with $r^2 < 2$" and saying "the set of items r so that r is a real number and $r^2 < 2$".

Frequently, we wish to convey that the elements of our set have a certain form. For example, the odd numbers are the numbers which can be written in the form $2n+1$ where $n \in \mathbb{Z}$. In other words, this is "the set of numbers k which can be written $k = 2n+1$ for some $n \in \mathbb{Z}$", i.e. $\{k \mid \exists n \in \mathbb{Z}\ (k = 2n+1)\}$. However, k is only used as a placeholder for the form $2n+1$. It is more convenient for us to just say "for each $n \in \mathbb{Z}$, put $2n+1$ into the set". To do this, we write $\{2n+1 \mid n \in \mathbb{Z}\}$. Technically, this set definition is just shorthand for the earlier definition, but this form is very handy.

Also, set definitions can use more than one variable to specify them. In the style of the last paragraph, we may write $\{2n+3k \mid n \in \mathbb{Z}, k \in \mathbb{Z}\}$ to denote all integers which can be written as a multiple of 2 plus a multiple of 3. (This definition looks much nicer than the formal $\{x \mid \exists n \in \mathbb{Z}\ \exists k \in \mathbb{Z}\ (x = 2n+3k)\}$.) This definition corresponds to saying "For each integer n and each integer k, put $2n+3k$ in the set".

Example 1.14:

If a and b are real numbers with $a < b$, then we can define the *closed interval* $[a, b]$ as $\{x \in \mathbb{R} \mid a \le x \le b\}$. The *open interval* (a, b) is $\{x \in \mathbb{R} \mid a < x < b\}$ (parentheses are used when an end does not belong to an interval, and square brackets are used when it does). There are also two *half-open intervals* $[a, b) = \{x \in \mathbb{R} \mid a \le x < b\}$ and $(a, b] = \{x \in \mathbb{R} \mid a < x \le b\}$.

In the special case where $a = b$, we have $[a, b] = \{a\} = \{b\}$ but the open and half-open intervals are empty. Also, if $a > b$, then $[a, b]$ is also considered empty.

Also, if a is a real number, then we write (a, ∞) to denote $\{x \in \mathbb{R} \mid x > a\}$, and similarly $[a, \infty)$ denotes $\{x \in \mathbb{R} \mid x \ge a\}$. Also, $(-\infty, a)$ and $(-\infty, a]$ are defined similarly. Note that ∞ is not a real number; the symbol ∞ is just a convenient shorthand to denote this interval. Lastly, $(-\infty, \infty)$ is another way to write $\mathbb{R}$. $\square$

Example 1.15 (Russell's Paradox):

Consider the following very curious set: $R = \{x \mid x \notin x\}$. In other words, R is the set of all sets which do not contain themselves. Many objects you know belong to R: the set of all people is not a person, the set of all hamburgers is not a hamburger, etc. (However, perhaps the set of all ideas really is an idea... this matter is better left to philosophers.)

Is $R \in R$ true? If $R \in R$ were true, then the definition of R would say $R \notin R$ (when x has the value R), which contradicts our assumption. However, if $R \in R$ were false, then $R \notin R$ would be true, and hence R would be a valid value for x in R's definition. Thus, $R \notin R$ implies $R \in R$ as well! It seems that $R \in R$ can neither be true nor false.

This is known as *Russell's Paradox*, named after the mathematician Bertrand Russell. It illustrates that if you allow ANY property to be used when building a set, then havoc can result. People studying formal logic and set theory avoid this paradox by placing restrictions on what kinds of properties may be used to build a set, i.e. they only allow $P(x)$ to take certain forms.

More frequently, though, to avoid Russell's Paradox, mathematicians often work within what is called a *universe of discourse*. For each argument we study, there is some set U, called the universe of discourse for that argument, such that all mathematical objects we consider in the argument come from U. For instance, U may contain all real numbers, all functions from real numbers to real numbers, and other similar objects. By restricting our attention to U, Russell's Paradox does not occur.[3]

We will not have to worry much about these technical details, as the paradoxes like Russell's Paradox don't show up when doing calculus. However, if this paradox intrigues you, I would highly suggest a more formal course in set theory. $\square$

Combining Sets

Now, we aim to build more sophisticated sets out of basic ones. We'll introduce four useful operations for combining sets, analogous to how connectives combine statements.

Definition 1.16. If A and B are two sets, then their *intersection* $A \cap B$ is the set $\{x \mid x \in A \;\wedge\; x \in B\}$, so $A \cap B$ is the set of objects belonging to both A and B. If $A \cap B = \emptyset$, then A and B are said to be *disjoint*.

As an example, consider $\{2n \mid n \in \mathbb{Z}\} \cap \{3n \mid n \in \mathbb{Z}\}$ (the uses of n in the different sets are unrelated to one another, since one scope ends before the other begins). This set consists of numbers which are multiples of 2 AND multiples of 3. With a little effort, you can see that these are precisely the multiples of 6, i.e. this intersection is $\{6n \mid n \in \mathbb{Z}\}$.

Definition 1.17. The *union* $A \cup B$ is the set $\{x \mid x \in A \;\vee\; x \in B\}$, so $A \cup B$ consists of elements in at least one of A or B.

[3]For a challenge, figure out why the set $\{x \in U \mid x \notin x\}$ doesn't cause any paradox!

For example, if P is the set of people, then the set $\{p \in P \mid \text{firstname}(p) = \text{“Michael”}\} \cup \{p \in P \mid \text{lastname}(p) = \text{“Klipper”}\}$ is the set of all people who share a name in common with the author. All people with the first name of “Michael” are in this set, as are the people with the last name of “Klipper”. In contrast, the intersection of these two sets would only have people with the exact same name.[4]

Just like a truth table is a convenient way to lay out all the information about statements, a *Venn diagram* is a convenient way to display information about sets. There is one circle drawn for each basic set, with overlaps between every pair of basic sets, making some regions. The regions that lie in our sophisticated set are shaded. See Figure 1.1, where the circle on the left represents A and the circle on the right represents B.

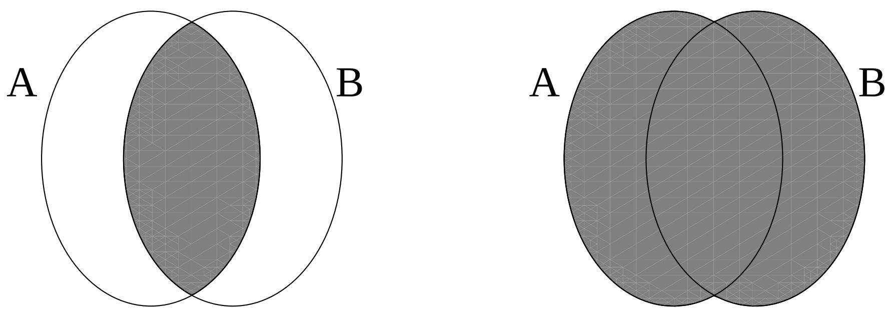

Figure 1.1: Venn diagrams for $A \cap B$ (left) and $A \cup B$ (right)

Definition 1.18. The *difference* $A - B$ (also written $A \setminus B$) is $\{x \mid x \in A, x \notin B\}$, so $A - B$ has the elements belonging to A which also do not belong to B.

For example, $\mathbb{R} - \mathbb{Q}$ is the set of irrational numbers. As another example, if $A = \{1, 2, 3\}$ and $B = \{2, 4\}$, then $A - B = \{1, 3\}$. In this example, although 4 doesn't occur in A, it doesn't matter that it occurs in B: elements of $A - B$ have to be in A as well as also being absent from B. In the exercises, you are asked to show that $A - B$ is always the same set as $A - (A \cap B)$.

Occasionally, people want to deal with elements to belong to A or B but not both. This can be written as $(A \cup B) - (A \cap B)$, and this is also the same

[4]However, the intersection could still have more than one member, because different people may have the same name!

set as $(A - B) \cup (B - A)$. This set is common enough to be given its own name: it is called the *symmetric difference* of A and B and is written $A \Delta B$. See Figure 1.2.

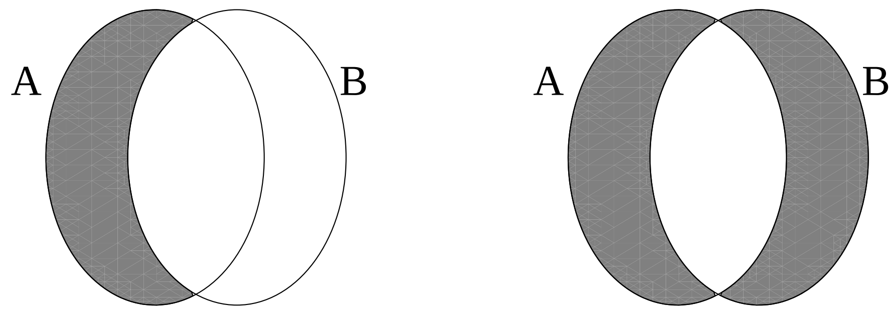

Figure 1.2: Venn diagrams for $A - B$ (left) and $A \Delta B$ (right)

Also, we frequently want to talk about when certain sets are contained in other sets. For example, every multiple of 4 is also a multiple of 2, so the set of multiples of 4 is contained in the set of multiples of 2. We define "containing" as follows:

Definition 1.19. If A and B are sets, then the statement $A \subseteq B$, read as "A is a *subset* of B", or as "A is *contained* in B" (or other synonyms), means "$\forall x \in A\ (x \in B)$".

Thus, we may rephrase our last example as saying $\{4n \mid n \in \mathbb{Z}\} \subseteq \{2n \mid n \in \mathbb{Z}\}$. Another very important example is

$$\mathbb{N}^* \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$$

(by chaining these together, analogously to how people chain together inequalities, we're saying that $\mathbb{N} \subseteq \mathbb{Z}$ and $\mathbb{Z} \subseteq \mathbb{Q}$, and so forth).

In fact, equality of sets is defined as

$$A = B \text{ iff } A \subseteq B \wedge B \subseteq A$$

If $A \subseteq B$ and $A \neq B$, then we say that A is a *proper subset* of B and write $A \subset B$ (it's like how $\leq$ means "less than or equal" but $<$ means "strictly less than"). Unfortunately, there are a couple different conventions for defining notation, and the other convention uses the symbol $\subset$ to denote what we have called $\subseteq$. Therefore, to make ourselves absolutely clear, we sometimes write $\subsetneq$ to mean "proper subset" (the subset symbol has part of the not-equals $\neq$ symbol under it).

Example 1.20:

The empty set is contained in every single set, i.e. for all sets A, $\emptyset \subseteq A$. This sounds quite plausible, since the empty set has no members, so it can be fit inside of any other set. However, writing out the definition of subset, we get "$\forall x \in \emptyset\ (x \in A)$". What exactly does "$\forall x \in \emptyset$" mean, since there are no x values in $\emptyset$?

Well, if "$\forall x \in \emptyset\ (x \in A)$" were false, i.e. "$\neg(\forall x \in \emptyset\ (x \in A))$" were true, then pushing the negation yields "$\exists x \in \emptyset\ (x \notin A)$". There can't exist any value x which belongs to $\emptyset$, let alone belonging to $\emptyset$ and also not belonging to A! By this reasoning, any statement of the form $\forall x \in \emptyset\ P(x)$ is automatically true. We say such statements are *vacuously true*.

As another example of a vacuously true statement, consider "every cow on the moon is purple." This is equivalent to "for all animals a, if a is a cow on the moon, then a is purple". However, the hypothesis of the if-then is always false, so the whole statement is true! $\square$

1.4 Building Proofs

At this point, you now have the major tools for building your own mathematical objects and writing propositions about them. However, having the tools and knowing how to use them are rather different things. After all, mathematics is not just about discovering results; it is also about convincing others (and yourself) that you are correct. We, as mathematicians, should hold ourselves to a higher standard and provide proofs for what we believe, to show others that our work is valid. Writing a proof of a mathematical statement which you believe to be true takes a lot of practice, and mathematicians get better at it as they do more proofs and see more examples of good reasoning.

Direct Proofs

To get ourselves started, let's look at the construction of a few simple proofs and see how they work.

Example 1.21:

Suppose A, B, and C are sets. It seems quite plausible that if A is contained in B, and B is contained in C, then A is also contained in C. For example,

$\mathbb{N} \subseteq \mathbb{Q}$ and $\mathbb{Q} \subseteq \mathbb{R}$ are both true (every natural number is rational, and every rational number is real), so these should imply every natural number is real, or $\mathbb{N} \subseteq \mathbb{R}$. Thus, let's aim to prove the following:

$$A \subseteq B \wedge B \subseteq C \to A \subseteq C$$

Strategy. Before making the proof, let's discuss a strategy for approaching the proof. This statement is an implication: we want to show that IF $A \subseteq B \wedge B \subseteq C$ is true, THEN $A \subseteq C$ is true. Hence, we might as well assume $A \subseteq B \wedge B \subseteq C$ is true, otherwise the hypothesis is false and we're already done (remember that a false statement implies anything!).

Next, with our new assumption, we'd like to show $A \subseteq C$. To do this, we should examine what $\subseteq$ means. By definition, $A \subseteq C$ means "$\forall x \in A\, x \in C$", so we want to show that every element of A is also an element of C. Thus, we should start with some arbitrary element x of A and show $x \in C$.

At this point, I have the assumptions $A \subseteq B \wedge B \subseteq C$ (thus I know $A \subseteq B$ and $B \subseteq C$ are both true) and $x \in A$. How do I put these assumptions together? Well, $A \subseteq B$ says that members of A are also members of B, and our particular value x is a member of A, so it is also a member of B! Similarly, since $x \in B$ and $B \subseteq C$, we conclude $x \in C$, as desired.

Now, we tidy up a little, being more precise about the quantifiers used in $A \subseteq B$ and $B \subseteq C$, and we get the actual proof.

Proof. Assume $A \subseteq B$ and $B \subseteq C$. Now let $x \in A$ be arbitrary. Because $A \subseteq B$, by definition $\forall y \in A\ (y \in B)$ is true[5], and $x \in A$ in particular, so we infer $x \in B$. Since $B \subseteq C$, or in other words $\forall y \in B\ (y \in C)$, from $x \in B$ we infer $x \in C$. Therefore, every $x \in A$ satisfies $x \in C$, so $\forall x \in A\ (x \in C)$ is true, i.e. $A \subseteq C$. ■

At the end of a proof, there is usually some marker to let you know the proof is complete. A box like ■ is one such marker. Another common marker is the phrase QED, abbreviating the Latin *Quod erat demonstrandum*, meaning "that which was to be demonstrated". □

The previous proof has some features which occur very frequently in mathematics:

[5] We had originally written the definition of subset using the variable name "x". However, since x has already been declared as arbitrary in this proof, we change the dummy variable to y.

- The statement to prove was an implication, so we assumed a hypothesis (e.g. $A \subseteq B \wedge B \subseteq C$) and proved a conclusion (e.g. $A \subseteq C$). This strategy for handling implications, called *direct proof*, helped us organize our work. (However, the assumption $A \subseteq B \wedge B \subseteq C$ should be viewed as temporary; we may make that assumption only to prove $A \subseteq C$. After we're done with that proof, we have to discard the assumption.)
- To show $A \subseteq C$, we used the definition of $\subseteq$. Many proofs need to use careful definitions of the terms involved.
- Since we wanted to prove a "for all" statement (e.g. $\forall x \in A(x \in C)$), we started with an arbitrary $x \in A$ and aimed to prove $x \in C$. This process of letting x be arbitrary is the main way to show a "for all x" statement (the phrase "Let $x \in A$ be given" is also frequently used to describe taking an arbitrary $x \in A$).

Why is the last point a valid proof strategy? Think of the proof of $\forall x \in A\ (x \in C)$ as a battle between you and a rival. You claim the result is true, but your rival is suspicious. He says, "Oh yeah? Is it true for this value?" and he gives you some value of x in A. You, undaunted, show him that $x \in C$. Flustered, he challenges you with a different value from A, but you still show him that value is in C. Your proof doesn't use any more information about x except that x is in A, so every time your rival challenges you, you successfully rise to his challenge. Eventually, after fighting off a few examples, you get annoyed and say, "Look, no matter which $x \in A$ you give me, I can show $x \in C$" and you present your argument using an arbitrary $x \in A$. Your rival acknowledges that he has no examples that can defeat you, and you win!

Let's look at another example which uses this idea.

Example 1.22:

Let's prove the following straightforward fact about integers:

$$\forall n \in \mathbb{Z}\ \forall k \in \mathbb{Z}\ (n \text{ is odd} \wedge k \text{ is odd} \rightarrow n + k \text{ is even})$$

Before we can start, we need to know precisely what "odd" and "even" mean. The *even numbers* are defined as the set $\{2z \mid z \in \mathbb{Z}\}$, i.e. the integers which are twice some other integer. The *odd numbers* are the set $\{2z+1 \mid z \in \mathbb{Z}\}$. Now that we have definitions, we can start talking strategy.

Strategy. Let's start with arbitrary odd numbers n and k (we're anticipating our rival will challenge us with these numbers). What do we know from them being odd? By definition, n has the form $2z + 1$ for some $z \in \mathbb{Z}$, i.e. $\exists z \in \mathbb{Z}\ (n = 2z + 1)$. We ought to choose a value for z out of all the possible values, just so that we have a value available for future use.

k also has a similar form, but it would be a mistake to say that $k = 2z+1$. Basically, the choice that worked for n above might not work for k, since n and k could be different! However, we can still make a choice for k, but we give it a different name; suppose we choose $y \in \mathbb{Z}$ so that $k = 2y + 1$.

Now that we have these forms for n and k, we can use them to rewrite $n + k$, and we want to show that sum is even. Hence, by definition of even, we want to end up with something looking like 2 times an integer. We're now ready for the proof.

Proof. Let $n \in \mathbb{Z}$ and $k \in \mathbb{Z}$ be given. (Sometimes, since the domain is the same, we shorten this to "Let $n, k \in \mathbb{Z}$ be given".) Assume n and k are both odd, so we may choose $z, y \in \mathbb{Z}$ so that $n = 2z+1$ and $k = 2y+1$. Therefore,

$$n + k = (2z + 1) + (2y + 1) = 2z + 2y + 2 = 2(z + y + 1)$$

is even because $z + y + 1 \in \mathbb{Z}$. ■

A natural question you might have at this point is "Didn't your strategizing already do the proof?" While the work was similar, the strategizing isn't the exact same thing as the proof. We develop strategy at the beginning as brainstorming, to generate ideas and to see where we think a proof is going to go. We allow ourselves to be more informal, we leave out some small details, and we make notes to ourselves about WHY we do certain steps in the proof.

However, in the actual proof, many of these motivational notes are gone (though a couple are sometimes left in, like saying what the current goal is, to help the reader keep track of the argument), and the steps are more rigorous. In essence, the final proof has all the steps, whereas the strategizing discusses the IDEAS behind the steps. These may look about the same right now, but that's mainly because these proofs are short and direct; for longer proofs, the strategizing becomes much more valuable. □

This previous proof had some similar logic components to the first example of the section: they both featured implications and universal quantifiers. However, this proof did something quite different: it used a statement with the form of "there exists", i.e. an existential quantifier. When we use statements like those (e.g. "there is some $z \in \mathbb{Z}$ so that $n = 2z + 1$"), our main action is to *choose* a value with the stated property (i.e. $z \in \mathbb{Z}$ is chosen so that $n = 2z + 1$).

Why is this a valid strategy for existential quantifiers? Think of the possible values of the variable z as being laid out on a shelf in the library; you see them there, but you don't own any of them. When you choose one, you finally have a value in your hand, and that value can be used in the rest of your proof. You didn't care which of the z's you grabbed; all you cared about was that it had the property you needed, which was that $n = 2z + 1$. Because this is a library, the librarian lets you take the value for free, but you must return the value when you're done with your proof. In our case, after proving $n + k$ is even, which is a statement that doesn't use z, we no longer need z and can return it to the library.

Our examples so far have focused a lot on how to work with quantifiers. Many mathematical statements have at least one quantifier in them, sometimes more, so developing good approaches for them is quite important. Let's do another example which mixes the two kinds of quantifiers.

Example 1.23:

Suppose that f is a function which takes real numbers as input and produces real numbers as output. (We'll talk more about functions in Section 1.6.) Let's show that

$$\forall c \in \mathbb{R}\ ((\exists x \in \mathbb{R}\ f(x) = 0) \leftrightarrow (\exists y \in \mathbb{R}\ f(y - c) = 0))$$

Strategy. After taking an arbitrary $c \in \mathbb{R}$, what does this statement mean in simpler terms? From experience with algebra, you probably have seen that $f(x-c)$ has a graph which is shifted c units to the right of the graph of $f(x)$. Also, saying $f(x) = 0$ is the same as saying x is an x-intercept of $f(x)$'s graph, or a *zero* of $f(x)$. Thus, this result is saying that $f(x)$ has a zero iff the shifted version $f(x - c)$ does.[6]

Now the result seems quite plausible. After all, a zero of $f(x)$ can be shifted by c units to get a zero of $f(x - c)$, and vice versa! To do the proof, we want to show two propositions are equivalent. Thus, we should show the

[6] Technically, a zero is not the same thing as a root; zeros belong to functions, and roots belong to equations. However, we will often ignore this distinction.

one on the left implies the one on the right, and we should also show the right one implies the left one.

To do left implies right, let's choose an $x \in \mathbb{R}$ with $f(x) = 0$. We want to find some number $y \in \mathbb{R}$ with $f(y - c) = 0$. (We will sometimes say y *witnesses the truth of* $\exists x \in \mathbb{R} f(x - c) = 0$.) One simple way to find y is to have $y - c = x$, or $y = x + c$. Thus, $x + c$ is our desired witness.

Showing right implies left is very similar, except that $y \in \mathbb{R}$ is chosen with $f(y - c) = 0$, so therefore $x = y - c$ is a witness to $\exists x \in \mathbb{R}\ f(x) = 0$. Now we build the proof.

Proof. Let $c \in \mathbb{R}$ be given. First assume $\exists x \in \mathbb{R}\ f(x) = 0$, and choose such a value for x. Then $f((x + c) - c) = f(x) = 0$, so $x + c$ witnesses $\exists y \in \mathbb{R}\ f(y - c) = 0$, finishing one direction of proof.

For the other direction, assume $\exists y \in \mathbb{R}\ f(y - c) = 0$, and choose such a y. Thus, $y - c$ witnesses $\exists x \in \mathbb{R}\ f(x) = 0$, and we are done. ■

This proof can be confusing at first glance. One reason for this is that the variable x is used differently in each direction of the argument. In one direction, we choose a value of x, and in the other direction, x is set to $y - c$. This is fine, since the choice in the first direction is only valid for that direction. Using the metaphor from earlier, the first argument "returns x to the library" before the other direction "checks out" a value for y and uses it to obtain the witness $x = y - c$. □

Remark. Some people write the previous theorem written in the form

$$\forall c \in \mathbb{R}((\exists x \in \mathbb{R}\ f(x) = 0) \leftrightarrow (\exists x \in \mathbb{R}\ f(x - c) = 0))$$

This is equivalent to the original statement, since only the name of a dummy variable has been changed. Also, this form uses fewer variable names, which reminds us that the different uses of x are related (though not the same). However, this form looks more confusing, since it raises the question "Are you saying that $x = x - c$? What if $c \neq 0$?"

The issue here is that the different occurrences of x have different meanings. There are two existential quantifiers, and their scopes don't overlap at all. Hence, the value of x doesn't have to be the same on each side of the $\leftrightarrow$.

As we proceed through these examples, we see more techniques for working with different forms of statements. Here's an example of a very common kind of proof in mathematics: showing that two sets are the same.

Example 1.24:

Let's show a result about inequalities:

$$\{x \in \mathbb{R} \mid 0 < x < 1\} = \left\{x \in \mathbb{R} \mid \frac{1}{2} < \frac{1}{1+x} < 1\right\}$$

Strategy. Suppose A is the set on the left and B is the set on the right. Recall that $A = B$ means that $A \subseteq B$ and $B \subseteq A$. Thus, we should prove both $A \subseteq B$ and $B \subseteq A$.

To show $A \subseteq B$, let $x \in A$ be given. We want to show $x \in B$. Right now, we know $0 < x < 1$, and we'd like to build $1/(1+x)$ out of x. Thus, we should add 1 to all parts of the inequality and then take reciprocals.

To show $B \subseteq A$, we let $x \in B$ be given, so $1/2 < 1/(1+x) < 1$, and we want to show $0 < x < 1$. The steps used are the reverse of the steps for showing $A \subseteq B$. We're now ready to write the proof.

Proof. Let A be the set on the left and B be the set on the right. First we show $A \subseteq B$. Let $x \in A$ be given, so $x \in \mathbb{R}$ and $0 < x < 1$. Adding 1 throughout, we get $1 < 1 + x < 2$. Taking reciprocals throughout (which is valid because all the numbers have the same sign) flips the inequalities, so we get $1/2 < 1/(1+x) < 1$, which means $x \in B$.

For the other direction, let $x \in B$ be given. Thus, $x \in \mathbb{R}$ and $1/2 < 1/(1+x) < 1$. Taking reciprocals gives $1 < 1+x < 2$. Subtracting 1 to both sides yields $0 < x < 1$, so $x \in A$. Therefore, $B \subseteq A$. ■

In our case, since the two subproofs just use the same implications but in opposite directions, we can combine them and prove the equivalent statement "$\forall x \in \mathbb{R}(x \in A \leftrightarrow x \in B)$". We can write our work in a more table-like format, with justifications provided on the side:

$$\begin{array}{lcll}
\forall x \in \mathbb{R} & & x \in A & \\
& \leftrightarrow & 0 < x < 1 & \text{definition} \\
& \leftrightarrow & 1 < 1 + x < 2 & \text{adding 1 throughout} \\
& \leftrightarrow & \dfrac{1}{2} < \dfrac{1}{1+x} < 1 & \text{reciprocals of positive numbers} \\
& \leftrightarrow & x \in B & \text{definition}
\end{array}$$

□

In the last example, we wrote our work conveniently as a sequence of statements following each other, rather than using full English sentences.

You've probably used a format like this before when solving equations. For example, if you wanted to find the roots of $x^3 + x^2 - x - 1 = 0$, then you'd probably have written

$$\begin{array}{rcl} x^3 + x^2 - x - 1 & = & 0 \\ x^2(x+1) - 1(x+1) & = & 0 \\ (x^2 - 1)(x+1) & = & 0 \\ (x-1)(x+1)(x+1) & = & 0 \\ x & = & 1, -1 \end{array}$$

on your paper. However, there are three small but important things that the proof in Example 1.24 has which this algebra doesn't have:

1. *A quantifier*: The proof in Example 1.24 says where x comes from and whether it is arbitrary or chosen.
2. *Justifications*: Each line has a reason for why it works.
3. *Iff ($\leftrightarrow$) symbols*: These say that the logic flows both ways. Without them, the algebra above only says "if $x^3 + x^2 - x - 1 = 0$, then $x = 1$ or $x = -1$", i.e. the logic only flows from top to bottom. This shows $x = 1$ or $x = -1$ are the only POSSIBLE solutions, but the top-to-bottom direction doesn't check whether they are ACTUALLY solutions. However, by writing in the $\leftrightarrow$ symbols, we see $x = 1$ and $x = -1$ are precisely the two solutions.

Out of those three points, the third is generally the most important, as forgetting to check two directions leads to many common mistakes by students. Note the following example.

Example 1.25:

Let's try and find the roots to

$$\sqrt{x-3} = x - 5$$

where $x \in \mathbb{R}$. Let $x \in \mathbb{R}$ be given, and let's try solving for the roots:

$$\begin{array}{rcll} \sqrt{x-3} & = & x - 5 & \\ x - 3 & = & (x-5)^2 & \text{squaring} \\ x - 3 & = & x^2 - 10x + 25 & \text{expanding} \\ 0 & = & x^2 - 11x + 28 & \text{collecting terms} \\ 0 & = & (x-4)(x-7) & \text{factoring} \\ x = 4 & \text{or} & x = 7 & \text{one factor must be zero} \end{array}$$

However, if we check our answer, then we see that $\sqrt{7-3} = 2 = 7-5$, but $\sqrt{4-3} = 1$ and $4-5 = -1$, so $\sqrt{4-3} \neq 4-5$! What went wrong with 4? The problem is in the first two rows. It is true that $\sqrt{x-3} = x-5$ IMPLIES $x-3 = (x-5)^2$ by squaring both sides, but the opposite direction doesn't work: taking the square root of both sides of $x-3 = (x-5)^2$ actually yields $\sqrt{x-3} = |x-5|$. This shows that by being careless and not writing $\leftrightarrow$ symbols in our proof, we accidentally obtained an extra solution to our equation.

However, we still do have a valid proof from top to bottom by inserting $\rightarrow$ symbols in our work. We just have to realize our work does not flow in both directions, and thus we need only check the possible solutions in the original equation. $\square$

Proof By Cases

Sometimes in a proof, we end up in a situation where we have a few different possibilities but don't know which possibility is the correct one. For example, maybe we've managed to find an integer n in our proof, but we aren't sure how to proceed unless we know whether n is odd or n is even. As another example, maybe we have two sets A and B, and we've found an element $z \in A \cup B$, so it follows that at least one of $z \in A$ or $z \in B$ is true. Which one is true? In these kinds of situations, *proof by cases* is useful. To do proof by cases, consider each possibility one at a time as a separate case, and try to draw the same conclusion from each case.

Let's go over a couple of examples to see proof by cases in action.

Example 1.26:

Let's prove an inequality with absolute values:

$$\forall x \in \mathbb{R}\ (|x| > 1 \rightarrow x^2 > 1)$$

Strategy. By now we know to start with an arbitrary $x \in \mathbb{R}$ and assume $|x| > 1$. However, $|x|$ is defined by two cases: it is x if $x \geq 0$, and it is $-x$ if $x < 0$. Since we don't know which of $x \geq 0$ and $x < 0$ is true, we should handle the cases separately.

In the first case, $|x| = x > 1$, so x is positive and we can safely multiply it through the inequality to get $x^2 > x$. In the second case, $|x| = -x > 1$, so $x < -1$ and x is negative, thus multiplying it through the inequality flips the inequality direction. In either case, though, we'll get $x^2 > 1$. We're now ready to do the proof.

Proof. Let $x \in \mathbb{R}$ be given, and assume $|x| > 1$. Either $x \geq 0$ or $x < 0$.

In the first case, $x \geq 0$. Thus, $|x| = x$, so $x > 1$. As x is positive, we may multiply it to both sides of the inequality and obtain $x^2 > x$. Since $x^2 > x$ and $x > 1$, we have $x^2 > 1$, finishing the first case.

In the second case, $x < 0$. Thus, $|x| = -x$, so $-x > 1$ and hence $x < -1$. As x is negative, multiplying it to both sides of the inequality flips the inequality, so $x^2 > -x$. As $x^2 > -x$ and $-x > 1$, we have $x^2 > 1$.

Thus, in either case $x^2 > 1$ is true. ■ □

Example 1.27:

Let's prove a nice result about integers:

$$\forall n \in \mathbb{Z}\ (n^2 - n \text{ is even})$$

Strategy. If $n \in \mathbb{Z}$, then how can we show $n^2 - n$ is even? Since this is a question about even versus odd, it seems useful to break into cases based on whether n is even or odd (we say we're considering the *parity* of n). After that, it should be routine calculation to see whether we can factor a 2 out of $n^2 - n$ (note as well that $n^2 - n = n(n-1)$, which gives us another way to try deciding the parity of $n^2 - n$).

Proof. Let $n \in \mathbb{Z}$ be given. Either n is even or n is odd.

If n is even, then we may choose some $k \in \mathbb{Z}$ with $n = 2k$. It follows $n^2 - n = 4k^2 - 2k = 2(2k^2 - k)$, which is even because $2k^2 - k \in \mathbb{Z}$.

On the other hand, if n is odd, then we choose $k \in \mathbb{Z}$ so that $n = 2k + 1$. It follows that $n^2 - n = (2k+1)^2 - (2k+1) = 4k^2 + 4k + 1 - 2k - 1 = 4k^2 + 2k = 2(2k^2 + k)$. Since $2k^2 + k \in \mathbb{Z}$, this shows $n^2 - n$ is even.

In either case, $n^2 - n$ is even. ■ □

On a related note, suppose that rather than ASSUMING one of several possibilities holds (like in the last two proofs), what if you want to SHOW that one of several possibilities holds, but you don't know which one? In other words, say we have a goal of the form $P \vee Q$, but we're not sure which of P or Q is going to be true. We have a nice tactic for these situations. Since $P \vee Q$ is equivalent to $(\neg P) \to Q$ (i.e. if at least one of P or Q is true, but P is not true, then Q must be the true one), we can try to show $P \vee Q$ by assuming P is false and then proving Q. Let's see this in action.

Example 1.28:

Let's prove a very important result about factoring:

$$\forall x, y \in \mathbb{R}\ (xy = 0 \to (x = 0) \vee (y = 0))$$

Strategy. This result helps us find zeroes of products and is used all the time when factoring polynomials. How will we approach the proof? Assume $xy = 0$, and $(x = 0) \vee (y = 0)$ is our goal. This goal has the form $P \vee Q$ where P is $x = 0$ and Q is $y = 0$ (note that by this point in the proof, x and y have been given values, so they're not really free variables and thus I don't have to write $P(x)$ or $Q(y)$). Thus, we'll treat this as $(\neg P) \to Q$, by assuming $x \neq 0$ and proving $y = 0$.

Proof. Let $x, y \in \mathbb{R}$ be given, and assume $xy = 0$. It suffices to show that $x \neq 0 \to y = 0$, so assume $x \neq 0$. Therefore, x may be divided from the equation to obtain $xy/x = 0/x$, i.e. $y = 0$. ■ □

Alternatives to Direct Proof

Every time so far that we've dealt with an implication of the form $P \to Q$, we've assumed the hypothesis P and aimed to get the conclusion Q. This is usually the easiest approach to read and use, but not always. Consider, for example, showing that for all $n \in \mathbb{Z}$, if n^2 is even then n is even (so P is "n^2 is even" and Q is "n is even"). Knowing that n^2 is even allows us to write $n^2 = 2k$ for some $k \in \mathbb{Z}$, but then it's not clear how one could show n is even from that. The problem is that knowing a factor of n^2 doesn't lead right away to a factor of n.

However, we do have a couple approaches for getting around this problem. One very handy approach involves an alternate way of writing an implication:

Definition 1.29. The *contrapositive* of an implication $P \to Q$ is the statement $(\neg Q) \to (\neg P)$.

There are a couple other related definitions: the *converse* of $P \to Q$ is $Q \to P$, and the *inverse* of $P \to Q$ is $(\neg P) \to (\neg Q)$. However, as can be verified from truth tables, the contrapositive and the original are equivalent! For example, the statements "if it rains, then I carry an umbrella" and "if I'm not carrying an umbrella, then it is not raining" are equivalent.

Going back to proving that "(n^2 is even) implies (n is even)", let's try handling the contrapositive instead. Hence, we want to show "(n is not even) implies (n^2 is not even)", or in other words, "n odd $\to n^2$ odd". Thus, assuming n is odd yields $n = 2k+1$ for some $k \in \mathbb{Z}$, and now we may write n^2 in terms of k. This yields the following proof:

Proof. Let $n \in \mathbb{Z}$ be given. We prove the contrapositive: if n is odd, then n^2 is odd. Assume n is odd, so choose some $k \in \mathbb{Z}$ so that $n = 2k+1$. Hence, $n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$ is odd ($2k^2 + 2k \in \mathbb{Z}$). ■

The other approach is called *proof by contradiction*, and it is useful in many situations. The idea is that if you want to prove something, then instead assume it is false and obtain a contradiction. This is a very common tactic for proving something whose negation is a very useful assumption. For example, the proof above could have been rewritten as a proof by contradiction as follows:

Proof. Let $n \in \mathbb{Z}$ be given, and assume n^2 is even. We want to prove n is even. Suppose for contradiction that n is odd, thus we may choose $k \in \mathbb{Z}$ so that $n = 2k+1$. Thus, $n^2 = 2(2k^2 + 2k) + 1$ is odd, contradicting the assumption that n^2 is even. Hence, n must be even. ■

The last example we'll provide in this section is an argument that really shows the power of proof by contradiction.

Example 1.30:

It's well-known that sometimes a rational number to a rational power is rational (e.g. 2^2) and sometimes it is irrational (e.g. $2^{1/2} = \sqrt{2}$). However, is an irrational number to an irrational power still irrational? Not necessarily: we'll prove

$$\exists x \in \mathbb{R} - \mathbb{Q}\ \exists y \in \mathbb{R} - \mathbb{Q}\ (x^y \in \mathbb{Q})$$

Strategy. A direct proof would figure out what the values of x and y are, though we probably don't have any good guesses for x or y right now, making that tactic difficult. Let's try a contradiction argument, so we assume $\neg(\exists x, y \in \mathbb{R} - \mathbb{Q}\ (x^y \in \mathbb{Q}))$. By doing a negation push, our assumption is $\forall x, y \in \mathbb{R} - \mathbb{Q}\ (x^y \notin \mathbb{Q})$. This is quite a useful assumption since it applies to all irrationals!

Since this assumption applies to all irrationals, let's try some simple choices for x and y to substitute into our assumption to hope to get a contradiction. One of the simplest irrational numbers to work with is $\sqrt{2}$, so if we put in $x = \sqrt{2}$ and $y = \sqrt{2}$, then we find $(\sqrt{2})^{\sqrt{2}}$ is irrational. This means that this new number can be substituted in our assumption! Hence, we try $x = (\sqrt{2})^{\sqrt{2}}$ and $y = \sqrt{2}$, and algebra properties of powers simplify x^y a lot. (On the other hand, if we chose $x = \sqrt{2}$ and $y = (\sqrt{2})^{\sqrt{2}}$, then x^y wouldn't simplify at all.) Now we write the proof.

Proof. Suppose for contradiction that $\exists x, y \in \mathbb{R} - \mathbb{Q}\ (x^y \in \mathbb{Q})$ is false, so thus $\forall x, y \in \mathbb{R} - \mathbb{Q}\ (x^y \notin \mathbb{Q})$ is true. Since $\sqrt{2}$ is irrational, plugging in $x = \sqrt{2}$ and $y = \sqrt{2}$ shows that $(\sqrt{2})^{\sqrt{2}}$ is irrational; for ease of notation, call this number c. Now, plugging in $x = c$ and $y = \sqrt{2}$ to our assumption yields $c^{\sqrt{2}} \notin \mathbb{Q}$, but

$$c^{\sqrt{2}} = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{2})^{(\sqrt{2}\cdot\sqrt{2})} = (\sqrt{2})^2 = 2$$

and $2 \in \mathbb{Q}$, contradicting $c^{\sqrt{2}} \notin \mathbb{Q}$. ■ □

Remark. Although proof by contradiction is a very powerful tactic, it can be difficult to use and read. The issue is: you don't know ahead of time what the contradiction will be! This makes it tricky to design the argument. In contrast, when using the contrapositive, you know what the goal is: to prove $P \to Q$, you assume $\neg Q$ and must show $\neg P$.

In fact, the method of proof of contradiction is risky for another important reason. If you make a small mistake in your argument, and you arrive at an impossible situation, then you will be led to conclude that you found your contradiction, finishing the proof. However, the error may not be in your hypothesis you aim to contradict; the error may be in your work! The method of proof by contrapositive doesn't have this problem as often, because it gives you a specific goal to achieve.

When proving $P \to Q$ by contradiction, the most effective arguments assume P and $\neg Q$ and USE BOTH. If the assumption of P is never used, then you've probably proven $\neg P$ directly from $\neg Q$, and hence you've built a contrapositive argument! For instance, consider the argument:

"To prove $(x > 0 \to -x < 0)$, assume $x > 0$, and also suppose for contradiction that $-x$ is not less than 0. Thus, $-x \geq 0$. By negating both sides of this, we obtain $x \leq 0$, which contradicts $x > 0$."

In this argument, the hypothesis of $x > 0$ is not used except to obtain the final contradiction! Essentially, this proof is really showing $\neg(-x < 0) \to$

$\neg(x > 0)$. We rewrite this as a cleaner contrapositive argument: "Assume $\neg(-x < 0)$, so $-x \geq 0$. Negate both sides to obtain $x \leq 0$. Therefore, $\neg(x > 0)$."

Now that you've seen a number of different examples of proofs and proof strategies, it's good to notice that they all share one important feature: when building these proofs, the logical structure of the goal sets up our proof. Consider, for instance, Example 1.23, which proved the following for any function f from real numbers to real numbers:

$$\forall c \in \mathbb{R}\ ((\exists x \in \mathbb{R}\ f(x) = 0) \leftrightarrow (\exists x \in \mathbb{R}\ f(x - c) = 0))$$

Immediately after seeing this result, we can already build an outline of the proof steps, with two gaps in it marked GAP:

1. Let $c \in \mathbb{R}$ be given.
2. Assume $\exists x \in \mathbb{R}\ f(x) = 0$.
3. Goal: $y \in \mathbb{R}$ with $f(y - c) = 0$.
4. Choose $x \in \mathbb{R}$ with $f(x) = 0$.
5. GAP
6. Thus, $\exists y \in \mathbb{R}\ f(y - c) = 0$.
7. Assume $\exists y \in \mathbb{R}\ f(y - c) = 0$.
8. Goal: $x \in \mathbb{R}$ with $f(x) = 0$.
9. Choose $y \in \mathbb{R}$ with $f(y - c) = 0$.
10. GAP
11. Thus, $\exists x \in \mathbb{R}\ f(x) = 0$.

Line 1 comes from "$\forall c \in \mathbb{R}$". Next, the $\leftrightarrow$ symbol tells us to do two directions of proof: the first is in lines 2 through 6, and the second is in lines 7 through 11. Each direction has its own assumption and its own goal. The goals, having "$\exists x \in \mathbb{R}$" or "$\exists y \in \mathbb{R}$" in them, tell us to look for witnesses, so we make a goal to find some value with the desired property (lines 3 and 8). Lastly, our assumptions with "$\exists x \in \mathbb{R}$" or "$\exists y \in \mathbb{R}$" in them tell us to choose a value as specified (lines 4 and 9).

After doing that outline, it becomes much clearer how to fill in the gaps. Line 5 should be "Define $y = x + c$, so $f(y - c) = f(x) = 0$". Line 10 should be "Define $x = y - c$, so $f(x) = f(y - c) = 0$".

By building the outline for ourselves, we made our remaining steps much clearer. This is the ultimate goal of logic: to make reasoning clearer, both for the reader and the writer. Unfortunately, many students, when first exposed to logic, are not shown this process of building an outline, so they view logic as an unnecessary technicality and hindrance. In fact, the logical process gives us guidance, clarity, and precision.

1.5 Exercises

1. Prove the following for all sets A, B, and C:
   (a) $A - B = A - (A \cap B)$
   (b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
   (c) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

   Suggestion for parts (b) and (c): use the distributivity laws in Exercise 1.2.4.(c).

   (d) $A - (B \cup C) = (A - B) \cap (A - C)$
   (e) $(A \cup B) \subseteq C \leftrightarrow ((A \subseteq C) \wedge (B \subseteq C))$
   (f) $A \subseteq (B \cap C) \leftrightarrow ((A \subseteq B) \wedge (A \subseteq C))$

2. For each of the following statements, either prove the statement if it is true for all sets A, B, and C, or provide a counterexample.
   (a) $A - (B \cap C) \subseteq (A - B) \cup (A - C)$
   (b) $(A - B) \cup (A - C) \subseteq A - (B \cap C)$
   (c) $(A - B) - C = A - (B \cup C)$
   (d) $A - (B - C) = A - (B \cap C)$

3. For each of the following statements, suppose that A,B,C,D are arbitrary sets satisfying $A \subseteq B$ and $C \subseteq D$. If the statement is true, then prove it. Otherwise, give a counterexample.

   (a) $A \cap C \subseteq B \cap D$
   (b) $A \cup C \subseteq B \cup D$
   (c) $A - C \subseteq B - D$
   (d) $A \Delta C \subseteq B \Delta D$

4. Prove that for all $n \in \mathbb{Z}$, if n is even then $(n + 1)(n + 3) - 1$ is even.

5. Prove that for all $x \in \mathbb{R}$, if $x \neq 4$ then there is a unique $y \in \mathbb{R}$ so that $xy = 4y + 5$.

6. Prove that for all $x \in \mathbb{R}$, $x \in [-1, 1]$ iff $3x^2 \in [0, 3]$. (Hint: Recall that $\sqrt{x^2}$ is $|x|$, not x.)

7. Prove by contrapositive that for all $a, b, c \in \mathbb{Z}$, if $a^2 + b^2 = c^2$, then at least one of a, b, or c is even. (You may assume that no integer is both even and odd.)

1.6 Functions

Functions are the basic tool we use in mathematics when we want to take input and produce output. For example, when we wrote "$f(x) = x^2$" in high school algebra, we were describing a function that takes a real number x as input and produces its square as output. However, functions can be significantly more general than just functions from the real numbers to the real numbers. For example, a machine that takes a person's name as input and prints out their birthday is computing a function from names to days of the year. If we call that function f, then, for instance, we have f("Michael Klipper") = "July 20", f("John Mackey") = "January 13", and so on.[7]

Since functions occur quite often in mathematics, it's good to have a precise definition for them. To do this, we consider *ordered pairs*. An ordered pair is an object of the form (x, y) for some objects x and y, where we say that x is the *first coordinate* of the pair and y is the *second coordinate*.

Unfortunately, the notation for the ordered pair (x, y) is the same as the notation for an open interval; hopefully in context it will be clear which of the two is meant when we see that symbol. Also, note that the ordered pair (x, y) is not the same as the set $\{x, y\}$. First, in (x, y), x and y are allowed to be the same, whereas the set $\{x, y\}$ simplifies to just $\{x\}$ if x and y are the same. Second, order matters in an ordered pair: (x, y) is not the same as (y, x) if $x \neq y$, but $\{x, y\} = \{y, x\}$.

We frequently want to consider sets of ordered pairs, such as the following:

Definition 1.31. If A and B are sets, then the set $A \times B$, called the *Cartesian product* of A and B, is the set $\{(x, y) \mid x \in A, y \in B\}$. In other words, it is the set of all ordered pairs whose first coordinates come from A and whose second coordinates come from B.

Example 1.32:

Here are some simple examples of Cartesian products:

- $\{0, 1\} \times \{0, 2\} = \{(0, 0), (0, 2), (1, 0), (1, 2)\}$.
- $\{0, 1, 2\} \times \{0\} = \{(0, 0), (1, 0), (2, 0)\}$.

[7] However, this is only a well-defined function if a person can be uniquely identified by their name! One alternative might be to have f take two inputs, such as a name and a birthday.

- If my set of ties is {red, yellow, blue} and my set of shoes is {black, brown}, then the Cartesian product of these sets gives the six tie-shoes combinations {(red, black),(red, brown),(yellow, black),(yellow, brown),(blue, black),(blue, brown)}.
- $\mathbb{R} \times \mathbb{R}$ is the set of points in the plane. This is often given the simpler name $\mathbb{R}^2$. In general, A^2 means $A \times A$ when A is a set.
- For any set B, $\emptyset \times B = \emptyset$. Why? Since there are no members of $\emptyset$, it's not possible to make any pairs (x, y) where $x \in \emptyset$ and $y \in B$. Similarly, $A \times \emptyset = \emptyset$ for any set A.

□

Similarly to ordered pairs, we can define *ordered triples* to be objects of the form (x, y, z) and make products of three sets $A \times B \times C = \{(x, y, z) \mid x \in A, y \in B, z \in C\}$. Thus, $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ is the set of points in three-dimensional space, though this set is often given the simpler name $\mathbb{R}^3$ (in general, A^3 means $A \times A \times A$ when A is a set). We can even make ordered quadruples (x, y, z, w), and in general for each positive integer n we can make *ordered n-tuples* with n coordinates. However, for our purposes, ordered pairs are usually enough.

Ordered pairs allow us to write a precise definition of functions:

Definition 1.33. A *function* f from a set A to a set B is a subset of $A \times B$ such that for each member x of A, there is exactly one $y \in B$ satisfying $(x, y) \in f$. This unique value of y is written as $f(x)$. A is called the *domain* of f (written as $\text{dom}(f) = A$), and B is called a *codomain* of f. We write "$f : A \to B$" to say "f is a function from A to B".

Many functions with which you are familiar from algebra are functions from $\mathbb{R}$ to $\mathbb{R}$. For instance, writing $f = \{(x, x^2) \mid x \in \mathbb{R}\}$ means $f(x) = x^2$ for each $x \in \mathbb{R}$. Technically speaking, the phrase "$f(x)$" does not represent a function; it represents the value of the function at some number x, so x has to be known for this to make sense. Thus, writing "$f(x) = x^2$" isn't really a proper definition of a function. Instead, we write "$f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ for all $x \in \mathbb{R}$", which includes both a specification of domain and codomain and also writes "for all $x \in \mathbb{R}$".

Example 1.34:

A definition of a function has to make sure that each domain value is assigned to only one codomain value. For example, consider the statement "when

$x \in [-1, 1]$ and $y \in \mathbb{R}$, $f(x) = y \leftrightarrow x^2 + y^2 = 1$", which means the same as saying $f = \{(x, y) \in [-1, 1] \times \mathbb{R} \mid x^2 + y^2 = 1\}$. This does not define a function, because when $x = 0$, both $0^2 + 1^2 = 1$ and $0^2 + (-1)^2 = 1$ are true, so $(0, 1)$ and $(0, -1)$ are two pairs in f with the same first coordinate. (When we graph functions from $\mathbb{R}$ to $\mathbb{R}$ in the usual way from algebra, we sometimes say that this equation for f *fails the Vertical Line Test*, because the vertical line $x = 0$ strikes the graph twice.) However, if we instead say "when $x \in [-1, 1]$, $f(x) = y \leftrightarrow (x^2 + y^2 = 1$ and $y \geq 0)$", then f is a function. We can manipulate algebraically to get the equivalent statement "$f(x) = \sqrt{1 - x^2}$ when $x \in [-1, 1]$".

However, there's no reason why a function from $\mathbb{R}$ to $\mathbb{R}$ has to be specified by one single equation. Any proposition of x and y that makes sure every x in the domain is paired with exactly one y in the codomain is good enough. For example, we could define a function $f : \{0, 1\} \to \{-1, 0, 1\}$ by just saying "$f(0) = 0$ and $f(1) = -1$". As another example, we can define a function by cases: the absolute-value function $f : \mathbb{R} \to \mathbb{R}$, defined by

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

is one of the most popular examples. Its definition can also be phrased as "for all $x, y \in \mathbb{R}$, $(x, y) \in f$ iff $(x \geq 0 \wedge y = x) \vee (x < 0 \wedge y = -x)$". $\square$

One-to-One and Onto Functions

In the definition of function, the domain is the set of all inputs to the function. Hence, a function can only have one possible domain. However, the codomain is not uniquely specified; a codomain has to contain all possible outputs, but it's not required to EQUAL the set of all outputs. Thus, every function can have plenty of possible codomains! For example, consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ for $x \in \mathbb{R}$. For any $x \in \mathbb{R}$, the statement $f(x) \in \mathbb{R}$ is true, but also $f(x) \in \mathbb{C}$ is true[8]; therefore, we may also write $f : \mathbb{R} \to \mathbb{C}$.

In general, if $f : A \to B$, then any set containing B is also a valid codomain for f, i.e. codomains can be made larger. This raises the question:

[8]The author adopts the view that $\mathbb{R} \subseteq \mathbb{C}$. Not all mathematicians agree with this, claiming that real numbers and complex numbers are very different types of objects. The details are technical and will not be covered here.

what is the smallest possible codomain for f? A set B is a codomain for f iff $f \subseteq A \times B$, i.e. whenever $(x, y) \in f$, $y \in B$. Thus, the set of all such y's is the smallest codomain. This set is called the *range* of f and can be written as $\{y \mid \exists x \in A\ (x, y) \in f\}$, as $\{f(x) \mid x \in A\}$, or as $\mathrm{ran}(f)$.

Example 1.35:

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

$$f(x) = \frac{1}{1+x^2} \quad \text{for all } x \in \mathbb{R}$$

We note this is well-defined because $1 + x^2$ is never zero when $x \in \mathbb{R}$ (on the other hand, that formula would not define a function from $\mathbb{C}$ to $\mathbb{C}$ because the formula wouldn't make sense when $x = i$). What is the range of f? If y is an arbitrary member of the codomain (i.e. $y \in \mathbb{R}$), then we find that

$$\begin{array}{lllcll}
 & & y & \in & \mathrm{ran}(f) & \\
\leftrightarrow & \exists x \in \mathrm{dom}(f) & y & = & f(x) & \text{def. of ran}(f) \\
\leftrightarrow & \exists x \in \mathbb{R} & y & = & \dfrac{1}{1+x^2} & \text{def. of } f \\
\leftrightarrow & \exists x \in \mathbb{R} & (1+x^2)y & = & 1 & \text{multiplying} \\
\leftrightarrow & \exists x \in \mathbb{R} & (y)x^2 + (y-1) & = & 0 & \text{distributing}
\end{array}$$

(Note that we really do have to keep writing "$\exists x \in \mathbb{R}$", because we're continually saying there is SOME corresponding domain member x, but we're never actually choosing x.)

Now, if y is 0, then this last line reduces to "$\exists x \in \mathbb{R}\ (-1 = 0)$", which is false, so $0 \notin \mathrm{ran}(f)$. If $y \neq 0$, then this last line is a quadratic equation. Thus, when $y \neq 0$, the last line is equivalent to

$$\exists x \in \mathbb{R}\ x = \frac{\pm\sqrt{-4y(y-1)}}{2y}$$

which is really just saying

$$\frac{\sqrt{-4y(y-1)}}{2y} \in \mathbb{R}$$

(since if the positive square root is real, then so is the negative square root).

Thus, we want to know: which nonzero real values of y satisfy $-4y(y-1) \geq 0$? (Remember that the square root of a negative number is purely imaginary.) The zeroes of $-4y(y-1)$ occur at $y = 0$ and $y = 1$, so on each of the three intervals $(-\infty, 0]$, $[0, 1]$, and $[1, \infty)$, the sign of $-4y(y-1)$ does not change (see Figure 1.3, where the sign on each interval is marked with $+$ for positive or $-$ for negative).

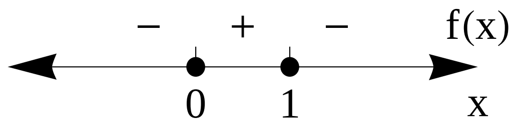

Figure 1.3: A number line displaying the signs of $-4y(y-1)$ on each interval

By testing some values, we find that $-4y(y-1) \geq 0$ iff $y \in [0, 1]$. However, we must also have $y \neq 0$, so we find that the set of possible y values is $\text{ran}(f) = (0, 1]$. $\square$

When we present a function with the "best possible" codomain, where the codomain contains all the outputs of the function and ONLY the outputs of the function, we use a special term to explain this situation:

Definition 1.36. If A and B are sets and $f : A \to B$, then f is *surjective onto* B (or more simply "f maps *onto* B") if $B = \text{ran}(f)$. We also say f is a *surjection* onto B[9]. Occasionally, people also say f *is onto* (using the preposition as an adjective).

One question you might have at this point is: why do we ever work with codomains that are larger than the range? In other words, why is surjectivity useful? There are a couple of reasons. First of all, sometimes finding the range of a function is hard or takes a lot of work, but finding an appropriate codomain is easier. For example, with the function f in Example 1.35, it's easy to see that f produces real numbers as output, but it took significant work to find $\text{ran}(f)$. Hence, for convenience, we will often just write $f : \mathbb{R} \to \mathbb{R}$ and be more specific later, if necessary, about which numbers are in the range.

Secondly, when working with several functions and combining them, it is more convenient to have all of the functions use the same codomain. For instance, we may want to add or multiply several functions from $\mathbb{R}$ to $\mathbb{R}$. By using the same codomain with each function, the reasoning is simpler.

[9]Technically, it's more correct to say "f is surjectively presented", since you need to know what the chosen codomain is in order to determine if the function is surjective. Most of the time, when we say a function is surjective, B is known in context.

Example 1.37:

The function defined by $f(x) = x^2$ for all $x \in \mathbb{R}$ does not map onto $\mathbb{R}$, because there is no $x \in \mathbb{R}$ satisfying $f(x) = -1$. However, if f is viewed as a function from $\mathbb{R}$ to $[0, \infty)$, then f is surjective.

As a nonmathematical example, the function which takes a person and maps them to their birthday is surjective (there is no calendar day of the year on which nobody has been born). ☐

Another way to phrase the definition of surjectivity is as follows: if $f : A \to B$, then f is surjective iff for all $y \in B$, there is AT LEAST one $x \in A$ satisfying $(x, y) \in f$. In other words, each possible output arises from at least one input.

What kind of property do we obtain if we replace "at least" with "at most" in that statement? If each possible output arises from at most one input, then every actual output (i.e. every member of the range) arose from a unique input. To put this in symbols, if $x, x' \in A$ produce the same output $f(x) = f(x')$, then they must be the same input, so $x = x'$. Equivalently, by using the contrapositive, if $x \neq x'$, then we must have $f(x) \neq f(x')$.

Definition 1.38. A function $f : A \to B$ is *injective* (or *one-to-one*) if for every $y \in B$, there is at most one $x \in A$ satisfying $(x, y) \in f$. Equivalently, f satisfies $\forall x, x' \in A\ (f(x) = f(x') \to x = x')$. We also say that f is an *injection*.

Example 1.39:

The function defined by $f(x) = x^2$ for all $x \in \mathbb{R}$ is not injective, because $f(1) = f(-1) = 1$. (When we graph the function f for all real x, sometimes we say that f *fails the Horizontal Line Test*, since the horizontal line $y = 1$ strikes the graph twice). However, consider the function $g = \{(x, y) \in f \mid x \in [0, \infty)\}$; we say that g is the *restriction* of f to $[0, \infty)$ and use the notation "$g = f \upharpoonright [0, \infty)$". g is injective: whenever $x, x' \geq 0$ and $x^2 = (x')^2$, $|x| = |x'|$ is true and hence $x = x'$.

As another example, consider the function $h : [0, \infty) \to \mathbb{R}$ defined by $h(x) = x^2 + x + 1$ for all $x \geq 0$. Let's show h is injective by using the contrapositive. Let $x, x' \geq 0$ be given and assume $x \neq x'$; without loss of generality, we may assume $x < x'$. (Frequently, "without loss of generality" is abbreviated as WLOG.) Therefore, $x^2 < (x')^2$ and thus $x^2 + x + 1 < (x')^2 + x' + 1$, showing that $h(x) < h(x')$. In particular, $h(x) \neq h(x')$.

The function which takes a person and produces their birthday is also not injective, since many people have the same birthday. In fact, it's impossible to avoid repetitions of birthdays unless there are at most 366 people in the domain (accounting for leap years). In contrast, the function that takes an American driver and returns their driver's license number should be injective (there's less confusion for the government and police that way). □

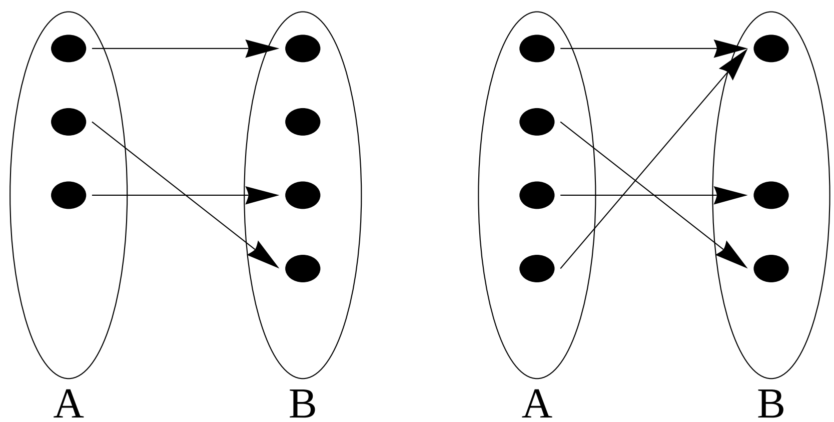

Figure 1.4: An injective but not surjective function (left) and a surjective but not injective function (right)

Take a look at Figure 1.4. In this picture, the domain and codomain are drawn as ovals, with solid dots inside standing for members, and an arrow pointing from a member $x \in A$ on the left to a member $y \in B$ on the right means (x, y) is in the function. The left picture has one right-side dot with no arrow pointing to it, so that function is not surjective. However, the left function is injective, since no two arrows point to the same right-side dot. The right picture has two left-side dots pointing to the same right-side dot, so that function is not injective. However, the right function has an arrow to each right-side dot, so it is surjective.

Definition 1.40. A function $f : A \to B$ is *bijective* (so it is a *bijection* or a *1-1 onto correspondence*) if f is both injective and surjective. In other words, for each $y \in B$, there is EXACTLY one $x \in A$ satisfying $(x, y) \in f$.

For example, the left function in Figure 1.4 becomes a bijection by removing the right-side dot with no arrow pointing to it (i.e. we remove one codomain element). Also, the right function becomes a bijection by removing the bottom left-side dot (i.e. we restrict the domain to be smaller).

Example 1.41:

If A is any set, then one basic example of a bijection is the function $f : A \to A$ defined by $f(x) = x$ for all $x \in A$. This function is called the *identity function on A*. We'll use the symbol id_A to represent this function.

As a harder example, consider the function $g : (0, \infty) \to (0, 1)$ defined by

$$g(x) = \frac{x}{x+1} \quad \text{for all } x \in (0, \infty)$$

Note that g is well-defined since $x + 1 \neq 0$ whenever $x > 0$. We'll show g is a bijection. This involves showing both that g is injective and that g is surjective.

To show injectivity, let $x, x' \in (0, \infty)$ be given, and assume $g(x) = g(x')$. The following proves that $x = x'$:

$$\begin{array}{rrcll}
 & \dfrac{x}{x+1} & = & \dfrac{x'}{x'+1} & \\
\to & x(x'+1) & = & x'(x+1) & \text{cross-multiplying} \\
\to & xx' + x & = & xx' + x' & \text{distributing} \\
\to & x & = & x' & \text{subtracting to both sides}
\end{array}$$

Now, to show surjectivity, let's show the range is $(0, 1)$. Let $y \in \mathbb{R}$ be given; we want to show $y \in \mathrm{ran}(g)$ iff $y \in (0, 1)$. We find

$$\begin{array}{llrcll}
 & & y & \in & \mathrm{ran}(g) & \\
\leftrightarrow & \exists x \in \mathrm{dom}(g) & y & = & g(x) & \text{def. of } \mathrm{ran}(g) \\
\leftrightarrow & \exists x \in (0, \infty) & y & = & \dfrac{x}{x+1} & \text{def. of } g \\
\leftrightarrow & \exists x \in (0, \infty) & y(x+1) & = & x & \text{multiplying} \\
\leftrightarrow & \exists x \in (0, \infty) & xy - x & = & -y & \text{distributing and rearranging} \\
\leftrightarrow & \exists x \in (0, \infty) & x(y-1) & = & -y & \text{factor out } x
\end{array}$$

This shows that in order for $y \in \mathbb{R}$ to be in the range, we must have $y \neq 1$ (since otherwise the last line simplifies to the false statement "$\exists x \in (0, \infty)\ 0 = -1$"). When $y \neq 1$, we may divide and obtain the equivalent statement

$$\exists x \in (0, \infty)\ x = \frac{-y}{y-1}$$

which is the same as saying

$$\frac{-y}{y-1} > 0$$

To solve this inequality for y, note that $-y/(y-1)$ is zero at $y = 0$ and undefined at $y = 1$, so we should check three intervals: $(-\infty, 0)$, $(0, 1)$, and $(1, \infty)$. (Here, we don't include the interval endpoints because we need $-y/(y-1)$ to be STRICTLY bigger than 0). If $y < 0$, then y and $y-1$ are both negative, so $-y/(y-1)$ is negative. If $0 < y < 1$, then y is positive but $y-1$ is negative, so $-y/(y-1)$ is positive. Lastly, if $y > 1$, then y and $y-1$ are positive, so $-y/(y-1)$ is negative.

This shows that $y \in \mathrm{ran}(g)$ iff $y \in (0, 1)$, so g is surjective. This finishes the proof that g is a bijection. $\square$

Composition of Functions

At this point, we've developed a definition for functions, and we've looked at some examples. We've also seen properties that a function could have, like not producing duplicate outputs (injectivity) and producing all the outputs in the codomain (surjectivity). However, why stop at just studying individual functions? We have ways of combining functions together.

When we're working with two functions $f, g : \mathbb{R} \to \mathbb{R}$, we have certain useful algebraic ways of combining them. One way is to add the functions: $f + g : \mathbb{R} \to \mathbb{R}$ is defined by $(f + g)(x) = f(x) + g(x)$ for all $x \in \mathbb{R}$. We can similarly define $f - g$ and fg to be respectively the difference and product of the functions. If $x \in \mathbb{R}$ and $g(x) \neq 0$, then we can also define f/g at x by $(f/g)(x) = f(x)/g(x)$. These operations are usually called *pointwise* addition, subtraction, multiplication, and division of functions.

However, for more general functions, we have a very important operation, which corresponds to using one function right after another:

Definition 1.42. Suppose A, B, and C are sets, and $f : A \to B$ and $g : B \to C$ are functions. The *composite function* $g \circ f : A \to C$ is defined by $(g \circ f)(x) = g(f(x))$ for all $x \in A$. The function $g \circ f$ is called the *composition* of f *followed by* g (since when evaluating the function, f is applied first, and then g is applied).

See Figure 1.5 for a picture of what composition does. It takes one function f and immediately follows it with another function g. It is important that the domain of g be a codomain for f, otherwise g won't know how to handle the outputs of f.

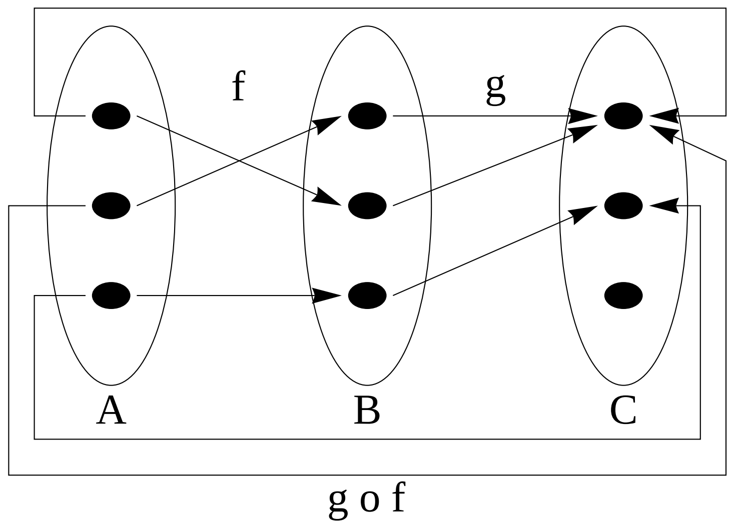

Figure 1.5: Two functions f and g and their composite $g \circ f$

Example 1.43:

When dealing with a graph of a function $f : \mathbb{R} \to \mathbb{R}$ in the plane, algebra courses teach that to stretch the picture vertically by a factor of c, you multiply the formula by c throughout. For example, the graph of $2x^2$ stretches twice as tall as the graph of x^2. We may express stretching in other terms by defining $s : \mathbb{R} \to \mathbb{R}$ by $s(x) = cx$ for all $x \in \mathbb{R}$, and then we say that the stretched version of f is $s \circ f$. For example, if $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(x) = x^2$ for all $x \in \mathbb{R}$, then $(s \circ f)(x) = s(x^2) = 2x^2$.

Multiple stretches and shifts can be done by multiple compositions. For example, replacing x^2 by $(x-1)^2$ shifts the graph one unit to the right. This is the same as replacing f with $f \circ t$, where $t(x) = x - 1$ for all $x \in \mathbb{R}$. (Note that the shift happens BEFORE f is applied!) By combining stretching and shifting, we can replace f with $s \circ f \circ t$, which has the formula $s(f(t(x))) = s(f(x-1)) = s((x-1)^2) = 2(x-1)^2$ for all $x \in \mathbb{R}$. Technically, we're only supposed to compose two functions at a time, but you can check that $(s \circ f) \circ t$ is the same as $s \circ (f \circ t)$ (they produce the same outputs for each input), so either of these expressions makes sense as a definition for $s \circ f \circ t$.

It is good to note that usually the order in which functions are composed matters. For example, when $x \in \mathbb{R}$, $(f \circ s)(x) = (x-1)^2$, but $(s \circ f)(x) = x^2 - 1$. Plugging in $x = 0$, we see that $(f \circ s)(0) = 1$, $(s \circ f)(0) = -1$, and

$1 \neq -1$, so the functions $f \circ s$ and $s \circ f$ are unequal. $\square$

A good question to consider here is: if $f : A \to B$ and $g : B \to C$ are functions, then which properties of f and g also hold for the composite $g \circ f$? We say that these properties are *preserved* by composition.

As an example, suppose that $A = B = C = \mathbb{R}$. Consider first when $f(x) = 1$ and $g(x) = 2$ for all $x \in \mathbb{R}$. f and g are constant functions, and $(g \circ f)(x) = g(f(x)) = g(1) = 2$ for all $x \in \mathbb{R}$, so $g \circ f$ is also a constant function. More generally, if A and B are any sets and $f : A \to B$ is a function, then we say f is a *constant function* if all outputs of f are the same, i.e. $\forall x, x' \in A\ f(x) = f(x')$. The property of being a constant function is preserved by composition: if A, B, C are any sets and $f : A \to B$ and $g : B \to C$ are any constant functions, then you can prove that $g \circ f$ is also constant.[10]

However, some properties are not preserved by composition. Consider when $f(x) = x - 1$ and $g(x) = x + 1$ for each $x \in \mathbb{R}$. Both f and g have the property that they never map x to itself; i.e. $\forall x \in \mathbb{R}\ f(x) \neq x$ and $\forall x \in \mathbb{R}\ g(x) \neq x$. However, $(g \circ f)(x) = g(x + 1) = (x - 1) + 1 = x$ maps every $x \in \mathbb{R}$ to itself. Thus, the property of never mapping a value to itself is not preserved by composition.

We are most interested in the properties we have defined in this chapter. Those properties are preserved by composition:

Theorem 1.44. *Let A, B, C be arbitrary sets, and suppose $f : A \to B$ and $g : B \to C$ are functions.*

(i) If f and g are injective, then so is $g \circ f$.

(ii) If f and g are surjective, then so is $g \circ f$.

(iii) If f and g are bijective, then so is $g \circ f$.

Strategy. Let's start with part (i), so we assume f and g are injective. This means that for all $x, x' \in A$, $f(x) = f(x') \to x = x'$, and similarly, for all $y, y' \in B$, $g(y) = g(y') \to y = y'$ (we'll use a different letter from x just to remind ourself these variables come from a different set). Now, we want to let $x, x' \in A$ be given and assume $(g \circ f)(x) = (g \circ f)(x')$, i.e. $g(f(x)) = g(f(x'))$, with the goal of showing $x = x'$.

Everything until now has just been using the definitions we previously introduced. At this point, we need to put our assumptions to good use. Since

[10] In fact, you only need g to be constant in order to ensure $g \circ f$ is constant!

we currently know that $g(f(x)) = g(f(x'))$, which features g as the outermost function, the assumption about g allows me to remove the application of g from each side to get $f(x) = f(x')$ (think of choosing $y = f(x)$ and $y' = f(x')$ in the definition above). Similarly, at this point, our assumption about f comes in handy to give $x = x'$.

The strategy for (ii) will also have to use definitions carefully. To show $g \circ f$ is surjective, we need to show that each $z \in C$ occurs as an output, i.e. there is $x \in A$ with $g(f(x)) = z$. This shows that we want z to be an output of g; here's where the surjectivity of g tells us there is some $y \in B$ with $g(y) = z$. Now, if $f(x)$ were y, then that would finish our argument, and fortunately we can find such an x because f is surjective.

In essence, the approaches for parts (i) and (ii) end up working with each function, one at a time, starting from the outside. We are "peeling off" the functions, first peeling off g and then peeling off f.

For (iii), if f and g are bijective, then they are both injective and surjective, so (i) and (ii) apply to them.

Proof. (i) Assume f and g are injective. Let $x, x' \in A$ be given with $(g \circ f)(x) = (g \circ f)(x')$, i.e. $g(f(x)) = g(f(x'))$. We want to show $x = x'$. Because g is injective, $g(f(x)) = g(f(x'))$ implies $f(x) = f(x')$. Because f is injective, $f(x) = f(x')$ implies $x = x'$.

(ii) Assume f and g are surjective. Let $z \in C$ be given; we wish to find $x \in A$ so that $(g \circ f)(x) = g(f(x)) = z$. Because g is surjective, we may choose $y \in B$ so that $g(y) = z$. Because f is surjective, we may choose $x \in A$ so that $f(x) = y$. Therefore, $g(f(x)) = g(y) = z$.

(iii) If f and g are both bijective, then they are both injective and surjective. Therefore, by parts (i) and (ii), $g \circ f$ is both injective and surjective, so it is bijective. ■

Example 1.45:

In Example 1.41, we saw that the function $g : (0, \infty) \to (0, 1)$, given by $g(x) = x/(x + 1)$ for all $x \in (0, \infty)$, is a bijection. A little work shows that $f : (0, \infty) \to (0, \infty)$, given by $f(x) = 2x$ for all $x \in (0, \infty)$, is also a bijection. Therefore, Theorem 1.44 tells us that the function $g \circ f : (0, \infty) \to (0, 1)$, defined by

$$(g \circ f)(x) = g(2x) = \frac{2x}{2x + 1} \quad \text{for all } x \in (0, \infty)$$

is a bijection.

As a nonmathematical example, suppose P is the set of all people who have ever lived, D is the set of all possible days in a calendar year (so D has 366 members because of leap years), and M is the set of all calendar months. Consider the function $b : P \to D$, which maps people to the calendar date of their birthday, and the function $m : D \to M$, which maps each day of the year to its month. (Note that capitalization is important. m and M are not the same!) Certainly b and m are surjective, as each day of the year has some person born on it, and each month has some day belonging to it. Thus, $m \circ b : P \to M$, which maps people to their birth month, is surjective. This makes sense, since every month has someone born in it. $\square$

Example 1.46:

Theorem 1.44 is very useful for studying properties of functions, but it is easy to accidentally believe the theorem says more than it really does. One of the statements it proves is "for all sets A,B,C, and for all functions $f : A \to B$ and $g : B \to C$, if f and g are injective then $g \circ f$ is injective". The innermost body of this statement is an implication but NOT an equivalence; let's show that "for all sets A,B,C, and for all functions $f : A \to B$ and $g : B \to C$, if $g \circ f$ is injective then f and g are injective" is false.

To show the converse is false, we find a *counterexample*, which is an example showing that the claim fails. (When you have a statement of the form $\forall x\ P(x)$, a counterexample is a value of x for which $P(x)$ fails). Thus, we'd like to exhibit sets A, B, C and functions $f : A \to B$ and $g : B \to C$ such that $g \circ f$ is injective but f and g are not both injective (this is how you negate "for all sets A, B, C and all functions $f : A \to B$ and $g : B \to C$, if $g \circ f$ is injective, then so are f and g"). Let's choose $A = C = \{1\}$ and $B = \{1, 2\}$, and then we define f and g by $f(1) = 1$ and $g(1) = g(2) = 1$. f and $g \circ f$ are both injective because their domain only has one element. However, since $g(1) = g(2)$ and $1 \neq 2$, g is not injective. See Figure 1.6.

What causes this counterexample to work? Look at the hollow dot in Figure 1.6. This dot represents a member of B which is not in $\text{ran}(f)$. When computing the composite, $g \circ f$ evaluates g at points of the form $f(x)$ for $x \in A$, i.e. g is only evaluated at points in $\text{ran}(f)$. Thus, the hollow dot is not used when constructing $g \circ f$. Although that dot makes g fail to be injective, $g \circ f$ never notices it!

Despite this example, which shows that the converse isn't true, there are

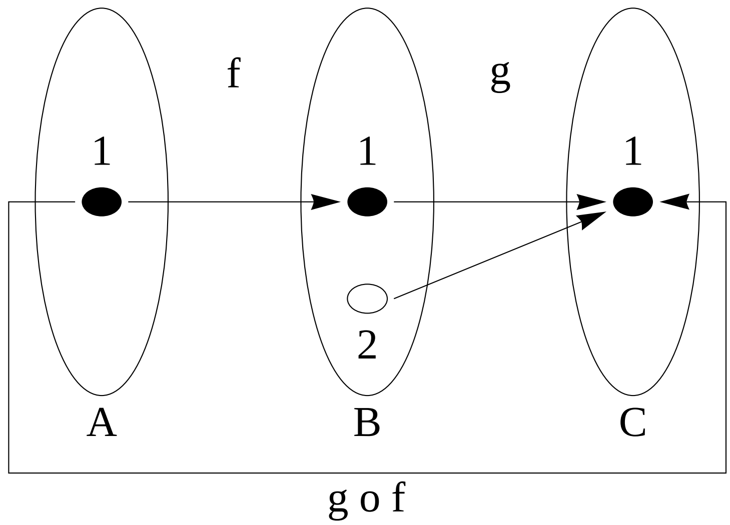

Figure 1.6: $g \circ f$ is injective but g is not

still some statements we can prove about how $g \circ f$ influences g and f. See the exercises. $\square$

Inverses

As we've mentioned, one way to think of a function is to think of it as something which converts inputs into outputs. This raises the question: can we also make a function to reverse this process and send outputs back to the inputs which made them? Essentially, this second function "cancels out" the first function. For example, the functions $f, g : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x + 1$ and $g(x) = x - 1$ for each $x \in \mathbb{R}$, cancel each other, since for any $x \in \mathbb{R}$, $g(f(x)) = g(x + 1) = (x + 1) - 1 = x$ (g sends $f(x)$ back to x) and $f(g(x)) = f(x - 1) = (x - 1) + 1 = x$ (f sends $g(x)$ back to x). We formalize this "canceling" notion with the definition of an *inverse function*.

Definition 1.47. Let A, B be sets and let $f : A \to B$ be a function. A function $g : B \to A$ is said to be an *inverse* for f iff

$$\forall x \in A\ g(f(x)) = x \quad \text{and} \quad \forall y \in B\ f(g(y)) = y$$

Equivalently, using the identity functions defined in Example 1.41,

$$g \circ f = \text{id}_A \quad \text{and} \quad f \circ g = \text{id}_B$$

If f has an inverse, then we say that f is *invertible*.

A couple things are worth noting about this definition. First of all, the domain of g must be precisely the codomain we used to describe f, and the codomain of g must be the domain of f. This is required so that the composites $g \circ f$ and $f \circ g$ both exist. Secondly, we want g and f to cancel each other no matter which direction we compose them; sometimes, to place emphasis on this, g is called a *two-sided inverse*. There are also definitions for one-sided inverses: see the exercises.

Example 1.48:

The functions $f, g : \mathbb{R} \to \mathbb{R}$ satisfying $f(x) = x + 1$ and $g(x) = x - 1$ for all $x \in \mathbb{R}$ are inverses for each other. In general, for any $c \in \mathbb{R}$, the functions defined by $f(x) = x + c$ and $g(x) = x - c$ for all $x \in \mathbb{R}$ are inverses for each other. Similarly, if $c \neq 0$, then the functions defined by $f(x) = cx$ and $g(x) = x/c$ for all $x \in \mathbb{R}$ are inverses for each other. (We often just say f and g are *inverses* instead of saying "inverses for each other".)

Some common functions are even introduced as inverses for other functions. For instance, it is known that the function $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x^3$ for each $x \in \mathbb{R}$, is a bijection. (We'll prove this formally in Chapter 3.) We'll soon show that this implies f has exactly one inverse. The cube-root function $g : \mathbb{R} \to \mathbb{R}$, satisfying $g(x) = x^{1/3}$ for each $x \in \mathbb{R}$, is DEFINED as the inverse for f.

However, if $f : \mathbb{R} \to \mathbb{R}$ and $g : [0, \infty) \to \mathbb{R}$ are defined by $f(x) = x^2$ and $g(x) = \sqrt{x}$ for all $x \in [0, \infty)$, then f and g are NOT inverses. First of all, the domain for g is not the codomain for f. Second, for any $x \in \mathbb{R}$, $g(f(x)) = \sqrt{x^2} = |x|$, since g only produces the nonnegative square root.

To fix these issues, we first restrict f to the smaller domain $[0, \infty)$. Define $h = f \restriction [0, \infty)$ (this means that $\mathrm{dom}(h) = [0, \infty)$ and $h(x) = f(x)$ for all $x \in [0, \infty)$). Second, we choose the codomain $[0, \infty)$ for g and h instead of choosing $\mathbb{R}$. After doing these steps, h and g ARE inverses. $\square$

The following lemma provides a convenient way to check whether two functions are inverses. (A *lemma* is a result which usually isn't so important on its own, but it serves as a helpful tool for other theorems.)

Lemma 1.49. *Let A,B be sets, and let $f : A \to B$ and $g : B \to A$ be functions. f and g are inverses iff*

$$\forall x \in A\ \forall y \in B\ (f(x) = y \leftrightarrow g(y) = x)$$

Strategy. For the forward direction, we assume that g is an inverse for f. Therefore, for all $x \in A$, $g(f(x)) = x$, so letting $y = f(x)$ gives us one part of the conclusion. Also, for all $y \in B$, $f(g(y)) = y$, so letting $x = g(y)$ gives the other part. The backward direction will be similar.

Proof. For the forward direction, assume that g is an inverse for f. Let $x \in A$ and $y \in B$ be given. If $f(x) = y$, then by the definition of inverse, $g(y) = g(f(x)) = x$. Conversely, if $g(y) = x$, then the definition of inverse says $f(x) = f(g(y)) = y$. Hence, $f(x) = y \leftrightarrow g(y) = x$.

For the backward direction, assume that for all $x \in A$ and all $y \in B$, $f(x) = y$ iff $g(y) = x$. We must show that for all $x \in A$, $g(f(x)) = x$, and that for all $y \in B$, $f(g(y)) = y$.

First, if $x \in A$ is given, then choose $y = f(x)$, so $y \in B$. Because $f(x) = y$ is true, by the assumption for this direction, $g(y) = x$ is true, so $g(f(x)) = g(y) = x$.

Second, if $y \in B$ is given, then choose $x = g(y)$, so $x \in A$. Because $g(y) = x$ is true, by the assumption for this direction again, $f(x) = y$ is true, so $f(g(y)) = f(x) = y$. ■

Up until now, we've had to use the phrase "an inverse" because we don't yet know how many inverses a function has. However, using Lemma 1.49, we will now show uniqueness of inverses as a corollary, so we may use the phrase "the inverse". (A *corollary* is a result which follows quickly from a previously proven result, much like a footnote to a paragraph.)

Corollary 1.50. *Let A, B be sets, $f : A \to B$ and $g, h : B \to A$ be functions, and suppose g and h are both inverses for f. Then $g = h$, i.e. for all $y \in B$, $g(y) = h(y)$. Thus, if f has an inverse, then it has a unique inverse, frequently denoted by f^{-1}.*

Strategy. Suppose that g, h are both inverses for f. For all $y \in B$, we want to show $g(y) = h(y)$. If we give $g(y)$ the name x, so $x \in A$, then Lemma 1.49 lets us manipulate the statement "$g(y) = x$" into other convenient forms.

Proof. Suppose that $g, h : B \to A$ are both inverses for $f : A \to B$. Let $y \in B$ be given; we wish to show $g(y) = h(y)$. Define $x = g(y)$, so that $x \in A$. Since g is an inverse for f, by Lemma 1.49, $f(x) = y$ iff $g(y) = x$, so $f(x) = y$. By Lemma 1.49 again, $f(x) = y$ iff $h(y) = x$ (because h is also an inverse for f), so $h(y) = x$. Thus, $g(y) = h(y)$. ■

Now we wish to answer the biggest question of this subsection: which functions are invertible?

Theorem 1.51. *Let A,B be sets, and let $f : A \to B$ be a function. f is invertible iff f is bijective.*

Strategy. For the forward direction, we want to show invertible functions are bijective. We have two parts to the definition of inverse. The part that says $f^{-1}(f(x)) = x$ for all $x \in A$ will allow us to show injectivity of f, because if $f(x_1) = f(x_2)$ for some $x_1, x_2 \in A$, then we can apply f^{-1} on each side to get $x_1 = x_2$. The other part of the definition, saying $f(f^{-1}(y)) = y$ for all $y \in B$, allows us to take an element y of B and "undo f" to find a choice of $x \in A$, namely $f^{-1}(y)$, that f maps to y. This shows every $y \in B$ is in $\mathrm{ran}(f)$, so f is surjective.

For the backward direction, if f is bijective, then each $y \in B$ has exactly one $x \in A$ with $(x, y) \in f$. Since we want an inverse for f to go "in the reverse direction" of f, if f sends x to y, then the inverse should send y to x! In other words, we define $g : B \to A$ by saying that for each $y \in B$, $g(y)$ should be THE $x \in A$ which satisfies $f(x) = y$. We'll show $g = f^{-1}$ by using Lemma 1.49.

Proof. For one direction of proof, assume f^{-1} exists. To show f is bijective, we show f is injective and f is surjective. To show f is injective, let $x, x' \in A$ be given with $f(x) = f(x')$, and we'd like to show $x = x'$. Applying f^{-1} to both sides yields $f^{-1}(f(x)) = f^{-1}(f(x'))$, i.e. $x = x'$. To show f is surjective, let $y \in B$ be given, and we'd like to find $x \in A$ with $f(x) = y$. Choose $x = f^{-1}(y)$, so that $f(x) = f(f^{-1}(y)) = y$. This proves f is bijective.

For the other direction, assume f is bijective. Thus, for every $y \in B$, there is exactly one $x \in A$ with $(x, y) \in f$. We define $g : B \to A$ as follows: for each $y \in B$, $g(y)$ is the UNIQUE $x \in A$ satisfying $f(x) = y$. (More precisely, g can be written as the set of ordered pairs $\{(y, x) \in B \times A \mid (x, y) \in f\}$.) This definition shows that for all $x \in A$ and all $y \in B$,

$$f(x) = y \leftrightarrow g(y) = x$$

By Lemma 1.49, g is an inverse for f. (Note that by Corollary 1.50, we may write $g = f^{-1}$.) ■

We conclude with a demonstration of how to use Lemma 1.49 to find inverses of functions. Looking back at Example 1.41, where we started with the function $g : (0, \infty) \to (0, 1)$, defined by

$$g(x) = \frac{x}{x+1} \quad \text{for all } x \in (0, \infty)$$

we showed that g is a bijection. During that proof, we showed that for all $y \in \mathbb{R}$ with $y \neq 1$,

$$y \in \operatorname{ran}(g) \leftrightarrow \exists x \in (0, \infty)\ x = \frac{-y}{y-1}$$

which is equivalent to

$$\forall x \in (0, \infty)\ \forall y \in \mathbb{R} - \{1\} \left(g(x) = y \leftrightarrow x = \frac{-y}{y-1} \right)$$

After that, the proof showed both sides of the $\leftrightarrow$ in the body are true when $y \in (0, 1)$. Hence if we choose $A = (0, \infty)$ and $B = (0, 1)$, we have

$$\forall x \in A\ \forall y \in B\ \left(g(x) = y \leftrightarrow \frac{-y}{y-1} = x \right)$$

Lemma 1.49 now tells us that

$$g^{-1}(y) = \frac{-y}{y-1} \quad \text{for all } y \in (0, 1)$$

This shows that a process like the one used in Example 1.41 to show a function is a bijection can also produce the inverse as a byproduct!

1.7 Exercises

1. Suppose $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2 - 4x + 3$ for all $x \in \mathbb{R}$. Find $\operatorname{ran}(f)$.

2. (a) Suppose $g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ is defined by $g(a, b) = 2a - 3b$ for all $a, b \in \mathbb{Z}$ (g takes PAIRS as input, so we think of g as taking two integer inputs.) Find $\operatorname{ran}(g)$.
   (b) Do the same as (a) but with $g(a, b) = 6a + 4b$ for all $a, b \in \mathbb{Z}$. (Suggestion: if you want to show the range is some set $B \subseteq \mathbb{Z}$, then first show $B \subseteq \operatorname{ran}(f)$ and then show $\operatorname{ran}(f) \subseteq B$ separately.)

3. Perhaps this is surprising, but $\emptyset$ is a function. Explain why $\emptyset$ is a function. What are its domain and range?

4. Find functions $f, g : \mathbb{R} \to \mathbb{R}$ so that f and g are both bijective but $f + g : \mathbb{R} \to \mathbb{R}$ is neither injective nor surjective. This shows that pointwise addition does not preserve injectivity or surjectivity.

5. Prove that the function $f : (\mathbb{R} - \{1\}) \to (\mathbb{R} - \{2\})$, defined by

$$f(x) = \frac{2x+5}{x-1} \quad \text{for all } x \in \mathbb{R} - \{1\}$$

is a bijection. (Suggestion: To make the algebra more convenient, it helps to note that $2x + 5 = 2(x - 1) + 7$.) What is a formula for f^{-1}?

6. For each of the parts of this exercise, suppose that $f : \mathbb{R} \to \mathbb{R}$.
   (a) We say that f is *strictly increasing* to mean

   $$\forall x_1, x_2 \in \mathbb{R}\ (x_1 < x_2 \to f(x_1) < f(x_2))$$

   (In other words, f sends smaller inputs to smaller outputs.) Similarly, we say f is *strictly decreasing* when

   $$\forall x_1, x_2 \in \mathbb{R}\ (x_1 < x_2 \to f(x_1) > f(x_2))$$

   Prove that if f is strictly increasing, then f is injective. (The same type of proof works to show that strictly decreasing functions are also injective.)
   (b) Suppose that f is strictly increasing, and also assume that f is surjective. Thus, by part (a) and Theorem 1.51, $f^{-1} : \mathbb{R} \to \mathbb{R}$ exists. Prove that f^{-1} is also strictly increasing. (Hint: Use a proof by contradiction.)
   (c) Prove that if a strictly decreasing function is invertible, then its inverse is also strictly decreasing. (Note: There is a simple way to reuse part (b) to do this.)
   (d) Give an example where f is injective but is neither strictly increasing nor strictly decreasing.

7. Suppose that $f : A \to B$ and $g : B \to C$ are bijections. Thus, $g \circ f$ is a bijection and hence invertible. Prove that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

8. For each of these statements, either prove the statement if it is true for all sets A,B,C and all functions $f : A \to B$ and $g : B \to C$, or provide a counterexample if the statement is false.
   (a) If $g \circ f$ is injective, then f is injective.
   (b) If $g \circ f$ is injective, then g is injective.
   (c) If $g \circ f$ is surjective, then f is surjective.
   (d) If $g \circ f$ is surjective, then g is surjective.

The next four exercises deal with one-sided inverses, defined as follows.

Definition 1.52. If $f : A \to B$, then a *left inverse* for f is a function $g : B \to A$ such that $g \circ f = \mathrm{id}_A$ (i.e. g cancels f on the left). If f has a left inverse, then we say f is *left-invertible*. Similarly, a *right inverse* for f is a function $g : B \to A$ such that $f \circ g = \mathrm{id}_B$, and f is *right-invertible* if a right inverse exists.

As examples, we note that any bijection's inverse is both a left inverse and a right inverse. Also, the function $f : \mathbb{R} \to [0, \infty)$, defined by $f(x) = x^2$ for all $x \in \mathbb{R}$, has a right inverse g defined by $g(x) = \sqrt{x}$ for all $x \in [0, \infty)$, but g is not a left inverse. (See Example 1.48.)

9. Give examples which show that a left-invertible function could have more than one left inverse, and also show similar examples for right inverses. (Hint: if $f : A \to B$ has a left inverse and is not surjective, then what could a left inverse do with inputs from $B - \mathrm{ran}(f)$? For right inverses, if f has a right inverse and is not injective, consider elements $x, x' \in A$ with $x \neq x'$ and $f(x) = f(x')$. Also, drawing pictures should give you guidance.)

10. Prove that if a function $f : A \to B$ has both a left inverse $g : B \to A$ and a right inverse $h : B \to A$, then $g = h$. (Hint: Consider $g \circ f \circ h$.)

11. Prove the following theorem:

    Theorem 1.53. *When $A, B \neq \emptyset$, a function $f : A \to B$ is left-invertible iff f is injective.*

12. Prove the following theorem:

Theorem 1.54. *A function $f : A \to B$ is right-invertible iff f is surjective.*

13. Let $f : \mathbb{C} \to \mathbb{C}$ be defined by $f(z) = z^2$ for all $z \in \mathbb{C}$. Prove f is surjective. (Hint: This is a long problem. Given $a, b \in \mathbb{R}$, we wish to find $x, y \in \mathbb{R}$ so that $f(x+yi) = a+bi$. Thus, $(x+yi)^2 = x^2+2xyi-y^2 = a+bi$. This means that $x^2 - y^2 = a$ and $2xy = b$. Consider cases when $b = 0$ and when $b \neq 0$. In the second case, try to solve for y in terms of x and then find solutions for x.)

1.8 Induction

There is another proof technique, called *induction* (formally, it's called *mathematical induction*), which is helpful for proving statements about natural numbers. Before explaining more formally how induction works, let's examine a situation where induction is useful.

Example 1.55:

Consider the four squares of dots in Figure 1.7.

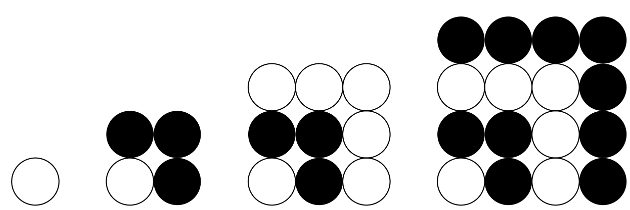

Figure 1.7: Four squares of dots with alternating colored layers

Each square consists of "L"-shaped layers, starting from the upper-right corner of the square, where the layers are alternately colored white and black. The first layer has 1 white dot, the next has 3 black dots, the next has 5 white dots, and so on. This suggests that the following pattern occurs:

$$\begin{aligned} 1 &= 1^2 \\ 1+3 &= 2^2 \\ 1+3+5 &= 3^2 \\ 1+3+5+7 &= 4^2 \end{aligned}$$

and in general, for any $n \in \mathbb{N}^*$, the sum of the first n odd numbers is n^2. If we use $P(n)$ to denote the proposition "$1 + 3 + \cdots + (2n-1) = n^2$" (so $P(n)$ is the n^{th} row of this pattern), we'd like to prove $\forall n \in \mathbb{N}^* P(n)$. However, if we start by taking an arbitrary $n \in \mathbb{N}^*$, then the tactics from the last section don't seem to help us prove $P(n)$.

Let's take a different approach. Rather than trying to prove that $P(n)$ is true for all $n \in \mathbb{N}^*$ at the same time (which is what taking an arbitrary $n \in \mathbb{N}^*$ tries to do), let's prove the statements $P(n)$ in order. Thus, we'll prove $P(1)$, then $P(2)$, then $P(3)$, and so forth. This is a useful approach because for any $n \in \mathbb{N}^*$, when we get to $P(n)$, we know we've already proven $P(1)$ through $P(n-1)$, so we may use them as hypotheses in our proof of $P(n)$ if we want.

Why do these extra hypotheses help? The main reason is because bigger squares are built from smaller squares by adding "layers" (for instance, Figure 1.7 portrays the fourth square as four layers of dots). More precisely, the $(n+1)^{\text{st}}$ square arises from the n^{th} square by adding one layer of $2n+1$ dots. Now, let's suppose we've managed to do n steps of proof, showing that $P(1)$ through $P(n)$ are all true. Since $P(n)$ is true, we know the n^{th} square has n^2 dots. Thus, the $(n+1)^{\text{st}}$ square has $2n+1$ more dots than that, i.e. it has

$$n^2 + (2n+1) = (n+1)^2$$

dots. This proves $P(n+1)$.

The previous paragraph proves that for all $n \in \mathbb{N}^*$, $P(n)$ implies $P(n+1)$. (In other words, we've shown how to "add a layer" in the proof.) In essence, we've proven the infinite chain of implications

$$P(1) \to P(2) \to P(3) \to \cdots \to P(n) \to P(n+1) \to \cdots$$

Now, we note that $P(1)$ says "$1 = 1^2$", which is true. Therefore, the chain above shows that $P(n)$ is true for every $n \in \mathbb{N}^*$, as desired. $\square$

The principle of induction is the proof strategy that uses the idea shown in the previous example. Suppose $P(n)$ is a proposition that we would like to prove for all $n \in \mathbb{N}^*$. We first start by showing $P(1)$ is true, which corresponds to starting the chain above. After that, for any $n \in \mathbb{N}^*$, we prove that $P(n)$ implies $P(n+1)$. Thus, $P(1)$ implies $P(2)$, which implies $P(3)$, which implies $P(4)$, and so on. As $P(1)$ is true, all the $P(n)$'s are true.

Definition 1.56. Let $P(n)$ be a proposition. A *proof of* $\forall n \in \mathbb{N}^* P(n)$ *by induction on* n consists of the following steps:

1. Establish the *base case (BC)*: Prove that $P(1)$ is true.
2. Let $n \in \mathbb{N}^*$ be given, and assume the *inductive hypothesis (IH)* that $P(n)$ is true.
3. Lastly, perform the *inductive step (IS)*: Prove that $P(n+1)$ is true. (The proof will generally use the IH.)

Another way to think of a proof by induction is to think of a row of dominoes, with one domino for each natural number. Think of domino n falling over as representing that $P(n)$ is true. At the beginning of the line is domino number 1, which we knock over by proving $P(1)$ as our base case. Next, we show that whenever the n^{th} domino falls for any $n \in \mathbb{N}^*$, it makes the $(n+1)^{\text{st}}$ domino fall, which corresponds to showing $P(n)$ implies $P(n+1)$ (the IH implies the IS). As a result, every single domino eventually falls down, so all the $P(n)$ statements are true.

Remark. The domino analogy makes induction more intuitively plausible, but it does not prove that mathematical induction is a legitimate proof tactic. In order to justify induction more formally, one needs to rigorously define $\mathbb{N}^*$, which is frequently done in a set theory course. (For more details, you should search for information about "well-ordering" $\mathbb{N}$.)

For our purposes, it will suffice to say that $\mathbb{N}^*$ has the property that every member of $\mathbb{N}^*$ can be obtained by repeatedly adding 1, starting from 1. In other words, for every $n \in \mathbb{N}^*$, a proof by induction eventually proves $P(n)$ after finitely many steps.

Here is an example of how to write a proof by induction:

Example 1.57:

Let's prove that for all $n \in \mathbb{N}^*$,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

In this case, $P(n)$ is "$1 + 2 + \cdots + n = n(n+1)/2$".

Proof. We prove $\forall n \in \mathbb{N}^* P(n)$ by induction on n.

Base Case: $P(1)$ says "$1 = 1(1+1)/2$", which is true.

Inductive Hypothesis: Let $n \in \mathbb{N}^*$ be given, and assume $P(n)$ is true, i.e. $1 + 2 + \cdots + n = n(n+1)/2$.

Inductive Step: We note that the sum $1 + 2 + \cdots + (n+1)$ contains the sum $1+2+\cdots+n$ as its first n terms, so we can apply our IH to those terms. This gives us

$$\begin{aligned}
1 + 2 + \cdots + (n+1) &= (1 + 2 + \cdots + n) + (n+1) \\
&= \left(\frac{n(n+1)}{2}\right) + (n+1) && \text{by the IH } P(n) \\
&= (n+1)\left(\frac{n}{2} + \frac{2}{2}\right) && \text{factoring} \\
&= \frac{(n+1)(n+2)}{2}
\end{aligned}$$

which proves that $P(n+1)$ is true. ■ □

Not all examples of proofs by induction have to use summations. Induction is a valid proof strategy when trying to prove any proposition $P(n)$, where n is supposed to be a natural number. Induction works well when $P(n+1)$ can be related to $P(n)$, so that the inductive step can "reuse" all the work that went into proving $P(n)$ to help prove $P(n+1)$. For instance, in the earlier examples, the summation on the left side of $P(n+1)$ contained the summation on the left side of $P(n)$. This makes induction ideal for reasoning about patterns, since the step from $P(n)$ to $P(n+1)$ essentially proves why "the pattern continues" or how we can "add a layer" to the problem.

For illustration, here is an induction example without summations:

Example 1.58:

Let's prove that for all $x \in \mathbb{R}$ and all $n \in \mathbb{N}^*$, if $x > 1$ then $x^n > 1$. Induction is a good strategy for this proof because x^{n+1} is xx^n, so a statement of the form $x^n > 1$ comes in handy when trying to prove $x^{n+1} > 1$.

Proof. Let $x \in \mathbb{R}$ be given, and assume $x > 1$. We will prove $\forall n \in \mathbb{N}^* x^n > 1$, by induction on n. (Frequently, if the choice of $P(n)$ is clear from context, we don't have to state $P(n)$ explicitly.)

Base case: "$x^1 > 1$" is true by assumption.

Inductive hypothesis: Let $n \in \mathbb{N}^*$ be given, and assume $x^n > 1$.

Inductive step: Because $x > 1$ and x^n is positive (by the IH), $xx^n > 1x^n$. Thus, $x^{n+1} > x^n$ and $x^n > 1$ by the IH, so $x^{n+1} > 1$. ■ □

Induction Variants and Strong Induction

There are several variants on the principle of induction. For example, an induction proof doesn't necessarily have to start with $P(1)$. Suppose you have a proposition $P(n)$, n_0 is some integer, and you'd like to prove $P(n)$ is true for all integers $n \geq n_0$. This basically means that you'd like to prove $P(n_0)$, then $P(n_0+1)$, then $P(n_0+2)$, etc. Hence, your base case should be to prove $P(n_0)$, and your inductive step proves that $P(n)$ implies $P(n+1)$ whenever $n \geq n_0$.

This proof tactic is often called "induction with a modified base case". We can also relate a proof of this style to a row of dominoes. Suppose that instead of pushing over domino #1, you start by pushing domino $\#n_0$. Every domino after that n_0^{th} domino will fall. However, the dominos that come before $\#n_0$ may or may not fall (for instance, some might have fallen if you were careless when pushing $\#n_0$ and accidentally knocked over some dominoes to the side).

The following example illustrates a proof with a modified base case:

Example 1.59:

Let's define the *factorial* $n!$ (read "n factorial") for $n \in \mathbb{N}$ as follows. We define $0! = 1$ and, whenever $n!$ is defined for any $n \in \mathbb{N}$, we define $(n+1)! = (n+1) \cdot n!$. Thus, $1! = 1 \cdot 0! = 1$ is defined in terms of $0!$, $2!$ is defined in terms of $1!$, $3!$ is defined in terms of $2!$, and so forth. (This is called an *inductive definition.*) It is also possible to write $n!$ as $1 \cdot 2 \cdot \dots \cdot n$.

We'd like to know which $n \in \mathbb{N}$ satisfy $n! > 2^n$. We try a few values of n:

n	$n!$	2^n
0	1	1
1	1	2
2	2	4
3	6	8
4	24	16
5	120	32

We see that $n! \leq 2^n$ when $n \leq 3$, but for $n \geq 4$, it seems that $n!$ is growing so quickly that it overtakes 2^n. Thus, if we write $P(n)$ to mean "$n! > 2^n$", we'd like to prove that $P(n)$ is true for all $n \geq 4$. We prove this by induction on n.

Base case: When $n = 4$, the table above shows $4! > 2^4$.

Inductive hypothesis: Let $n \geq 4$ be given, and assume $P(n)$.

Inductive step: We'd like to prove $P(n+1)$, i.e. $(n+1)! > 2^{n+1}$. By the definition of factorial, $(n+1)! = (n+1)n!$. Since $n! > 2^n$ by the IH, and $n+1$ is positive (in fact, $n+1 \geq 5$), $(n+1)n! > (n+1)2^n$. Also, since $n+1 \geq 5 > 2$, $(n+1)2^n > 2 \cdot 2^n = 2^{n+1}$. Therefore, $(n+1)! > 2^{n+1}$. $\square$

Another useful variant of induction is called strong induction. The main idea is that if we're trying to prove $P(n+1)$, and we've already proven $P(1)$ through $P(n)$, we should be able to use ALL of the statements $P(1)$, $P(2)$, and so forth up to $P(n)$ as hypotheses. (There is also a variant on strong induction allowing us to start at $P(n_0)$ for any integer n_0.) Thus, we have a stronger set of assumptions than just assuming $P(n)$.

Definition 1.60. Suppose $P(n)$ is a proposition. A *proof of* $\forall n \in \mathbb{N}^* P(n)$ *by strong induction* consists of the following steps:

1. Let $n \in \mathbb{N}^*$ be given, and assume the (Strong) Inductive Hypothesis: $P(1)$, $P(2)$, and so on up to $P(n)$ are all true. (If $n = 0$, then this is an empty set of assumptions.) In other words, assume that for all $k \in \mathbb{N}$ with $1 \leq k \leq n$, $P(k)$ is true.

2. Perform the Inductive Step: Prove that $P(n+1)$ is true.

Strong induction doesn't technically need a base case: $P(1)$ is proven in the inductive step as $P(n+1)$ where $n = 0$ (so the induction hypothesis is an empty statement). Here is an example showing a proof by strong induction.

Example 1.61:

The Fibonacci numbers f_n, where $n \in \mathbb{N}^*$, are a famous inductively-defined sequence of numbers. They are defined as follows: $f_1 = 1$, $f_2 = 1$, and for all $n \geq 3$, $f_n = f_{n-1} + f_{n-2}$. (Some people also define $f_0 = 0$, so that $f_2 = f_1 + f_0$ as well.) Hence, f_3 is defined in terms of f_2 and f_1, f_4 is defined in terms of f_3 and f_2, and so forth. The first few Fibonacci numbers are $1, 1, 2, 3, 5, 8, 13, \ldots$.

We'd like to prove that for all $n \in \mathbb{N}^*$, $f_n < 2^n$. How would the inductive step go for this proof? To prove $f_{n+1} < 2^{n+1}$, we need to know how to compute f_{n+1}. If $n+1 \geq 3$, then we may use the definition of the Fibonacci numbers to say that $f_{n+1} = f_n + f_{n-1}$, and then our IH will tell us information about f_n and f_{n-1}. Otherwise, if $n+1 \leq 2$, then either $n = 0$ or $n = 1$, and we can handle those cases separately using the definition $f_1 = f_2 = 1$.

Proof. Let $P(n)$ be "$f_n < 2^n$". We prove $P(n)$ for all $n \in \mathbb{N}^*$ by strong induction on n.

IH: Let $n \in \mathbb{N}$ be given, and assume that for all $k \in \mathbb{N}^*$, if $1 \le k \le n$, then $P(k)$ is true.

IS: Let's prove $f_{n+1} < 2^{n+1}$. In the first case, if $n = 0$, then $f_1 = 1 < 2^1$ by definition of f_1. Second, if $n = 1$, then $f_2 = 1 < 2^2$ by definition of f_2.

For the last case, suppose $n + 1 \ge 3$. Therefore, $f_{n+1} = f_n + f_{n-1}$. By our IH, $P(n)$ and $P(n-1)$ are true (since both $n-1$ and n are at least 1), so $f_n < 2^n$ and $f_{n-1} < 2^{n-1}$. We compute

$$\begin{array}{rcll} f_{n+1} & = & f_n + f_{n-1} & \text{definition of } f_{n+1} \\ & < & 2^n + 2^{n-1} & \text{IH with } P(n) \text{ and } P(n-1) \\ & = & 2^{n-1}(2+1) & \text{factoring} \\ & < & 2^{n-1}(2^2) & \\ & = & 2^{n+1} & \end{array}$$

which proves $P(n+1)$. ■ □

It turns out that any theorem which can be proven by strong induction can also be proven by ordinary induction, so the word "strong" is a slight misnomer. However, the process which turns a proof by strong induction into a proof by regular induction makes the proof look more cluttered and technical. Strong induction provides a convenient way to organize our work when we want to use multiple values in our induction hypothesis. We will not have much need for strong induction in this book, but it is a useful proof strategy in much of discrete mathematics.

1.9 Exercises

1. Prove the following summations for all $n \in \mathbb{N}^*$ by induction on n:
   (a) $1^2 + 2^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$
   (b) $1 - 2 + 3 - 4 + \cdots + (-1)^{n-1}n = (-1)^{n+1}\left(\dfrac{n(n+1)}{2}\right)$
   (c) $(1)1! + (2)2! + \cdots + (n)n! = (n+1)! - 1$ (see Example 1.59)
2. (a) Prove by induction on n that for any $x \neq 1$ and any $n \in \mathbb{N}^*$,
   $$1 + x + x^2 + \cdots + x^n = \frac{x^{n+1} - 1}{x - 1}$$

(b) What is the value of $1 + x + \cdots + x^n$ when $x = 1$?

3. Use induction to prove for all $n \in \mathbb{N}^*$ that $n^2 + 3n + 2$ is even. (Hint: If $f(n) = n^2 + 3n + 2$, then what is $f(n+1) - f(n)$?)

4. Prove by induction that for all $n \in \mathbb{N}^*$ and all $x \geq 0$, $(1+x)^n \geq 1+nx$.

5. In Example 1.61, we proved that $f_n < 2^n$. Suppose that we replace 2 with a positive number c everywhere in the proof, to try and prove that $f_n < c^n$. For which values of c does the proof still work?

6. For each $n \in \mathbb{N}$ with $n \geq 2$, let

$$a_n = \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\cdots\left(1 - \frac{1}{n}\right)$$

Guess a simpler formula for a_n, and prove your guess by induction.

7. The following proof is incorrect. Explain where the proof fails.

Proof. I prove that $1 + 2 + \cdots + n = \dfrac{n^2}{2} + \dfrac{n}{2} + 1$ by induction on $n \in \mathbb{N}^*$. Suppose as the IH that the statement is true for n. Then adding $n+1$ to each side yields

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= \frac{n^2}{2} + \frac{n}{2} + 1 + (n+1) \\ &= \frac{n^2}{2} + \frac{n}{2} + \frac{2n+2}{2} + 1 \\ &= \frac{n^2 + 2n + 1}{2} + \frac{n+1}{2} + 1 \\ &= \frac{(n+1)^2}{2} + \frac{n+1}{2} + 1 \end{aligned}$$

proving the statement is true for $n+1$. ■

8. Find the mistake in this proof by strong induction.

Proof. I prove that for all $x \neq 0$ and all $n \in \mathbb{N}^*$, $x^{n-1} = 1$. The proof is by strong induction on n.

Inductive hypothesis: For some $n \in \mathbb{N}$, assume $x^{1-1} = 1$, $x^{2-1} = 1$, and so on up to $x^{n-1} = 1$.

Inductive step: We wish to prove $x^n = 1$. If $n = 0$, then clearly $x^0 = 1$. Otherwise, we note

$$x^n = \frac{x^{n-1}x^{n-1}}{x^{n-2}}$$

and by the IH, $x^{n-1} = 1$ and $x^{n-2} = 1$. Thus, $x^n = 1$. ■

9. Find the mistake in this following argument, which tries to prove that all people have the same name.

   Proof. We prove the following statement by induction on $n \in \mathbb{N}^*$: "In any set of n people, all the people have the same name." For the base case, when $n = 1$, clearly any set with one person in it has all people named the same.

   Assume the statement is true for some $n \in \mathbb{N}^*$ as an inductive hypothesis, and let $\{p_1, p_2, \ldots, p_{n+1}\}$ be a set of $n+1$ people. By the IH applied to the set $\{p_1, p_2, \ldots, p_n\}$ with n people in it, the people p_1, p_2, and so on up to p_n all have the same name. Similarly, by the IH applied to the set $\{p_2, p_3, \ldots, p_{n+1}\}$ with n people, the people p_2, p_3, and so on up to p_{n+1} all have the same name. Thus, the people p_1, p_2, and so on up to p_{n+1} have the same name. ■

Chapter 2

The Real Numbers, Really

Much of our study of calculus will work with real numbers and with functions from real numbers to real numbers. The set of real numbers $\mathbb{R}$ has many useful properties, and we want to take advantage of those properties. Some of these properties include:

- The factoring property: For all $a, b \in \mathbb{R}$, if $ab = 0$, then $a = 0$ or $b = 0$. This was proven in the last chapter. This is a very useful tool for solving equations.

- The distributive property: For all $a, b, c \in \mathbb{R}$, $a(b + c) = ab + ac$. This helps us solve equations like "$x^2 + 5x = 0$" by noting $x^2 + 5x = x(x + 5)$, and then we use the factoring property above.

- Adding inequalities: For all $a_1, b_1, a_2, b_2 \in \mathbb{R}$, if $a_1 < b_1$ and $a_2 < b_2$, then $a_1 + a_2 < b_1 + b_2$. Inequalities will be extremely useful to us as a way to place restrictions on the values a variable can take, so we need tools for manipulating inequalities. For example, if we know $2 < x$ and $1 < y$ for some real numbers x, y, then we know $2 + 1 < x + y$, i.e. $3 < x + y$, which places a restriction on the sum $x + y$.

- Reciprocals: For all $a \in \mathbb{R}$, if $a \neq 0$ then $1/a$ exists. These reciprocals help us solve equations and inequalities. For example, say $2x = 3$ for some $x \in \mathbb{R}$. Because 2 has a reciprocal $1/2$, we can multiply $1/2$ to both sides and get $(1/2)(2x) = (1/2)3$, or $x = 3/2$.

- Density: For any two real numbers, there is always a real number in between them. More precisely, this means $\forall a, b \in \mathbb{R}$ $(a < b \rightarrow$

$\exists c \in \mathbb{R}\ a < c < b$). Why is this property meaningful? Consider the number line representing $\mathbb{R}$. If the density property were not true, say there were two numbers $a, b \in \mathbb{R}$ with no real numbers between them, then our line would have a "break" in between the points a and b. See Figure 2.1. Density tells us this behavior doesn't happen, so we draw our number line as a solid line without breaks. This relates as well to drawing nice smooth curves for graphs of functions; we'll return to that point in Chapter 3.

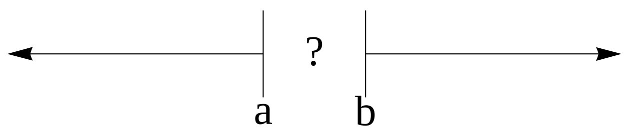

Figure 2.1: A break in the number line between two points

These properties aren't specific to the real numbers. The factoring and distributive properties, for instance, still work with the integers; i.e. "$\forall a, b \in \mathbb{Z}\ (ab = 0 \rightarrow a = 0 \vee b = 0)$" and "$\forall a, b, c \in \mathbb{Z}\ a(b + c) = ab + ac$" are both true. The natural numbers $\mathbb{N}$ satisfy those two properties as well. However, the natural numbers and the integers do not satisfy the last two properties of the list above.

The rational numbers $\mathbb{Q}$ satisfy every property in the list above. However, there are further properties of $\mathbb{R}$ which are not satisfied by $\mathbb{Q}$. For instance, the equation "$x^2 = 2$" has a solution with $x \in \mathbb{R}$, namely $x = \sqrt{2}$, but there is no solution with $x \in \mathbb{Q}$. We'll return to this point later.

Why are all these properties true for $\mathbb{R}$ in the first place? It is rather difficult to make a definition of the real numbers; many attempts at a definition lead to circular reasoning, like: "The real numbers are the rationals and the irrationals. The irrational numbers are the real numbers which are not rational." We will not aim to give a complete description of how $\mathbb{R}$ is constructed, although such constructions do exist: generally junior-level courses in real analysis provide a definition of $\mathbb{R}$.

For our purposes, we only need a basic system for reasoning about $\mathbb{R}$. To do this, we will introduce some fundamental properties of $\mathbb{R}$; these properties are not meant to be proven, but they will instead start as our beginning assumptions from which every other result will follow. (Assumptions with this purpose are frequently called *axioms* or *postulates*.) We will show that these axioms imply the major properties of the real numbers with which we are familiar.

We'll break up our collection of axioms into three groups: the algebraic axioms, the order axioms, and a last axiom called completeness. These different groups will help explain different facets of $\mathbb{R}$.[1]

2.1 The Algebraic Axioms

First, we will consider axioms which allow us to simplify and manipulate equations (inequalities follow a different set of rules and will be addressed in Section 2.3). After all, one of the major purposes of the laws of algebra is to enable us to take equations with variables in them and solve for the values of the variables. For this reason, these upcoming axioms will be called the *algebraic axioms*.

Traditionally, there are five major operations we first learn to perform with numbers: addition, subtraction, multiplication, division, and exponentiation (also known as the process of "taking powers"). These are generally called the *arithmetic operations*. The rules for adding and multiplying are simpler than the rules for the other operations. For instance, if a and b are any nonzero real numbers, then "$a + b = b + a$" and "$ab = ba$" are true, but "$a - b = b - a$", "$a/b = b/a$", and "$a^b = b^a$" are false except in certain special cases. Hence, we will only introduce axioms for addition and multiplication, and later we will define the other operations in terms of addition and multiplication.

We need to define some basic symbols. Addition and multiplication are functions from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$, i.e. $+, \cdot : (\mathbb{R} \times \mathbb{R}) \to \mathbb{R}$. Out of habit, for any $x, y \in \mathbb{R}$, instead of writing $+(x, y)$ to denote the *sum* of x and y, we write $x + y$; we say that we are using *infix notation* for $+$ (as opposed to the conventional *prefix notation* for functions). Similarly, we write $x \cdot y$, or simply xy, for the *product* instead of writing $\cdot(x, y)$. We also introduce the symbols $0, 1$ to stand for two distinct real numbers; i.e. $0 \in \mathbb{R}$, $1 \in \mathbb{R}$, and

[1] More precisely, any system which satisfies all of our axioms is called a *complete ordered field*. It is shown in junior-level analysis courses that there's essentially only one complete ordered field. More precisely, given any two complete ordered fields, there is a bijection between them that "preserves" all the operations and inequalities, so any statement which is true in one can be turned into a statement which is true in the other. This topic, however, is beyond the scope of this book.

$0 \neq 1$. Some major properties of 0 and 1 appear in the axioms.

With these symbols in place, we now present our six algebraic axioms. Instead of writing "Algebraic Axiom 1", "Algebraic Axiom 2", etc., we'll abbreviate with "AA1", "AA2", and so forth. Sometimes these axioms are also called the *field axioms*, because in abstract algebra, any set which satisfies these axioms is called a *field* (so algebraists simply say "$\mathbb{R}$ is a field with the usual definitions of $+$, $\cdot$, 0 and 1").

Axiom (AA1: *Commutativity*). For all $x, y \in \mathbb{R}$,

$$x + y = y + x \quad \text{and} \quad xy = yx$$

This is also described by saying addition and multiplication are *commutative*, or that they *commute*.

Axiom (AA2: *Associativity*). For all $x, y, z \in \mathbb{R}$,

$$(x + y) + z = x + (y + z) \quad \text{and} \quad (xy)z = x(yz)$$

This is also described by saying addition and multiplication are *associative*, or that they *associate*.

Remark. By AA2, when x, y, z are real numbers, there is no ambiguity when we write "$x+y+z$": the two possible interpretations as "$(x+y)+z$" or as "$x+(y+z)$" yield the same value. Similarly, we may write "xyz" without causing any confusion. In fact, whenever doing multiple additions or multiplications sequentially, we may omit parentheses; for example, for any real numbers a,b,c,d,e, the values "$(a + (b + c)) + (d + e)$", "$a + (((b + c) + d) + e)$", and "$(((a+b)+c)+d)+e$" are all the same, so we merely write "$a+b+c+d+e$" to denote the sum of all five numbers.

However, when mixing addition and multiplication, AA2 does not help. In this case, we commonly adopt an order of operations, which says that multiplications occur first and additions occur last, unless parentheses are used. Thus, for any real numbers a,b,c, "$ab + c$" means "$(ab) + c$" and not "$a(b + c)$".

Axiom (AA3: *Distributivity*). For all $x, y, z \in \mathbb{R}$,

$$x(y + z) = xy + xz$$

(hence $(y + z)x = yx + zx$ is true also by AA1). This is also called the *distributive property*, and we sometimes describe it by saying multiplication *distributes* over addition.

Axiom (AA4: *Identities*). For all $x \in \mathbb{R}$,

$$x + 0 = x \quad \text{and} \quad x \cdot 1 = x$$

(Hence, by AA1, $0+x = x$ and $1 \cdot x = x$ as well.) We say that 0 is an *additive identity* element, as the function $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x + 0$ for all $x \in \mathbb{R}$, is equal to the identity function $\mathrm{id}_{\mathbb{R}}$. Similarly, 1 is a *multiplicative identity* element, since the function $g : \mathbb{R} \to \mathbb{R}$, defined by $g(x) = x \cdot 1$ for all $x \in \mathbb{R}$, is $\mathrm{id}_{\mathbb{R}}$.

Axiom (AA5: *Negations*). For every $x \in \mathbb{R}$, there exists some $y \in \mathbb{R}$ so that

$$x + y = 0$$

(hence $y + x = 0$ as well by AA1). Note that this axiom does not say how many such values of y exist. We'll address this soon.

Axiom (AA6: *Reciprocals*). For every $x \in \mathbb{R}$, if $x \neq 0$ then there exists some $y \in \mathbb{R}$ so that

$$xy = 1$$

(hence $yx = 1$ as well by AA1). Note that this axiom does not say how many such values of y exist, nor does it say whether such a y exists when $x = 0$.

These six axioms will allow us to develop useful algebraic laws. However, before proceeding further, we want to address the question "Why should we find proofs for all these really obvious laws?" For $\mathbb{R}$, you are presumably rather familiar with these axioms and with many laws which follow from them, so you accept the laws as true with hardly a second glance.

However, there are other nontrivial examples of fields, some of them much less intuitive or commonplace than $\mathbb{R}$. For those fields, many of the laws we will discover will not be immediately evident. Once we discover proofs for our laws using only AA1 through AA6 as starting assumptions, though, these laws will apply to any field, even to the nontrivial fields. In other words, reasoning from the axioms allows us to be quite general, proving algebra laws for a wide variety of fields all at once.

To illustrate this point, let's demonstrate a couple other examples of fields. For the sake of these examples, we'll assume the usual algebraic properties of $\mathbb{R}$ (we'll prove many of them shortly).

Example 2.1:

The set $\mathbb{Q}[\sqrt{2}]$ is defined as $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. Note that $\mathbb{Q} \subseteq \mathbb{Q}[\sqrt{2}]$, because for all $x \in \mathbb{Q}$, x can be written as $x + 0\sqrt{2}$. In particular, 0 and 1 belong to $\mathbb{Q}[\sqrt{2}]$. Certainly $\mathbb{Q}[\sqrt{2}] \subseteq \mathbb{R}$, so we may use the usual addition and multiplication of $\mathbb{R}$ to perform calculations in $\mathbb{Q}[\sqrt{2}]$.

However, we need to check an important detail. Plus and times are supposed to be functions from $\mathbb{Q}[\sqrt{2}] \times \mathbb{Q}[\sqrt{2}]$ to $\mathbb{Q}[\sqrt{2}]$ (it's not enough to have them just be functions from $\mathbb{Q}[\sqrt{2}] \times \mathbb{Q}[\sqrt{2}]$ to $\mathbb{R}$, since we want to be able to perform repeated additions and multiplications). Thus, we need to show that for all $x, y \in \mathbb{Q}[\sqrt{2}]$, both $x+y$ and xy are in $\mathbb{Q}[\sqrt{2}]$ (sometimes we describe this by saying $\mathbb{Q}[\sqrt{2}]$ is *closed* under addition and multiplication).

Let's prove this closure property. Let $x, y \in \mathbb{Q}[\sqrt{2}]$ be given. Thus, there are $a, b, c, d \in \mathbb{Q}$ so that $x = a+b\sqrt{2}$ and $y = c+d\sqrt{2}$ (it turns out those values of a, b, c, d are uniquely determined, but we don't need that information for our proof... see Exercise 2.2.7). We compute

$$x + y = a + b\sqrt{2} + c + d\sqrt{2} = (a + c) + (b + d)\sqrt{2}$$

and

$$xy = ac + (ad + bc)\sqrt{2} + bd(\sqrt{2})^2 = (ac + 2bd) + (ad + bc)\sqrt{2}$$

Since $a + c$, $b + d$, $ac + 2bd$, and $ad + bc$ are all in $\mathbb{Q}$, $x + y$ and xy are in $\mathbb{Q}[\sqrt{2}]$, proving closure.

Since $\mathbb{Q}[\sqrt{2}]$ is closed under the usual plus and times of $\mathbb{R}$, properties AA1 through AA3 hold for $\mathbb{Q}[\sqrt{2}]$ (because they hold for $\mathbb{R}$). Property AA4 similarly holds since $0, 1 \in \mathbb{Q}[\sqrt{2}]$. To check AA5, if $x \in \mathbb{Q}[\sqrt{2}]$ is given, we need to find some y in $\mathbb{Q}[\sqrt{2}]$ (NOT just in $\mathbb{R}$) so that $x + y = 0$. Write $x = a + b\sqrt{2}$ for some $a, b \in \mathbb{Q}$; we can choose $y = (-a) + (-b)\sqrt{2}$.

Lastly, to verify AA6, suppose $a, b \in \mathbb{Q}[\sqrt{2}]$ is given with $a + b\sqrt{2} \neq 0$. Note that

$$(a + b\sqrt{2})(a - b\sqrt{2}) = a^2 + ab\sqrt{2} - ab\sqrt{2} - 2b^2 = a^2 - 2b^2$$

which leads us to

$$\frac{1}{a + b\sqrt{2}} = \frac{a - b\sqrt{2}}{a^2 - 2b^2} = \left(\frac{a}{a^2 - 2b^2}\right) + \left(\frac{-b}{a^2 - 2b^2}\right)\sqrt{2}$$

However, one detail of this should be checked: could $a^2 - 2b^2$ be zero? The answer is no, for if $a^2 - 2b^2$ were zero, then we would have $2 = a^2/b^2$ and hence $\sqrt{2} = \pm a/b$, contradicting the fact that $\sqrt{2}$ is irrational (see Exercise 2.2.6). This shows that $a + b\sqrt{2}$ has a multiplicative inverse in $\mathbb{Q}[\sqrt{2}]$, so AA6 holds. Thus, $\mathbb{Q}[\sqrt{2}]$ is a field. □

Example 2.2:

The last example showed that a particular subset of $\mathbb{R}$, namely $\mathbb{Q}[\sqrt{2}]$, is a field using the same addition and multiplication operations of $\mathbb{R}$ (but restricted to the subset under consideration). This example will use completely different operations!

Define $\mathbb{Z}_2 = \{0, 1\}$ (this is called the *integers modulo 2*, since it consists of all possible remainders when dividing integers by 2). For all $x, y \in \mathbb{Z}_2$, we define $x + y$ in $\mathbb{Z}_2$ to be the remainder of the usual value of $x + y$ in $\mathbb{R}$ when divided by 2, and xy in $\mathbb{Z}_2$ is the remainder of the usual value of xy in $\mathbb{R}$ when divided by 2. For example, in $\mathbb{R}$, $1 + 1 = 2$, which has remainder 0 when divided by 2, so in $\mathbb{Z}_2$, $1 + 1 = 0$. We have the following addition and multiplication tables:

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

$\mathbb{Z}_2$ satisfies AA1 because for all $x, y \in \mathbb{Z}_2$, in $\mathbb{R}$ we have $x + y = y + x$, so their remainders when dividing by 2 are the same (the same also holds for multiplication). We can similarly check that AA4 holds. However, AA2 and AA3 are more annoying to prove. Let $x, y, z \in \mathbb{Z}_2$ be given. If $x = 0$, then $(x + y) + z = (0 + y) + z = y + z = 0 + (y + z)$ and $(0y)z = 0z = 0 = 0(yz)$, showing AA2 holds in this case (similar calculations show that AA3 holds in this case too). Similar reasoning shows AA2 and AA3 hold if $y = 0$ or $z = 0$. The last case to consider is when $x = y = z = 1$, and you can check that AA2 and AA3 hold for that case as well.

Lastly, we need to address AA5 and AA6. For AA6, the only nonzero element is 1, and clearly its reciprocal is also 1. AA5 is proven by noting that $0+0 = 0$ and $1+1 = 0$ (so in $\mathbb{Z}_2$, we have $-1 = 1$!). Thus, $\mathbb{Z}_2$ is a field. In fact, since its only elements are 0 and 1, it is the smallest field.

This example can be generalized by considering remainders upon division by n for integers $n > 1$ (so this example is the case where $n = 2$). Some values of n produce fields, but some do not. See Exercise 2.2.3. □

Now that we've seen some other examples of fields, let's get started developing properties of fields, beginning with laws that will help us define subtraction and division. Note that since addition and multiplication are functions, we can always add or multiply a number to both sides of an equation: i.e. if a, b, c are real numbers and $b = c$, then $a+b = a+c$ and $ab = ac$. The next couple theorems address whether we can go the other way, starting from $a + b = a + c$ or $ab = ac$ and obtaining $b = c$.

Theorem 2.3 (Cancellation of Addition). *For all $a, x, y \in \mathbb{R}$, if $a+x = a + y$, then $x = y$.*

Strategy. Suppose we start with the assumption $a + x = a + y$. How do we cancel a from both sides? We are allowed to add numbers to both sides, so we should add a real number which "cancels out" a. We use AA5 to find such a number, use AA2 to group the additions properly so as to produce 0's, and use AA4 to remove the 0's from the equation.

Proof. Let $a, x, y \in \mathbb{R}$ be given, and assume $a + x = a + y$. By AA5, we may choose $b \in \mathbb{R}$ so that $a + b = 0$. By AA1, $b + a = 0$ as well.

Now, add b to both sides of $a+x = a+y$ to get $b+(a+x) = b+(a+y)$. By AA2, $(b+a)+x = (b+a)+y$. As $b+a = 0$, we have $0+x = 0+y$. By AA1, $x + 0 = y + 0$, and hence $x = y$ by AA4. ∎

The following similar theorem is left as an exercise:

Theorem 2.4 (Cancellation of Multiplication). *For all $a, x, y \in \mathbb{R}$, if $a \neq 0$ and $ax = ay$, then $x = y$.*

As a quick consequence of the cancellation theorems, we show that we can't have "two different 0's" or "two different 1's":

Corollary 2.5. *The numbers 0 and 1 are uniquely defined by the algebraic axioms. In other words, if y is any real number satisfying "$\forall x \in \mathbb{R}\ x + y = x$", then y must be 0. Similarly, if z is any nonzero real number satisfying "$\forall x \in \mathbb{R}\ xz = x$", then z must be 1.*

Strategy. For uniqueness of 0, if there were some $y \in \mathbb{R}$ with $x + y = x$ for all $x \in \mathbb{R}$, then we would have $x + y = x = x + 0$. Theorem 2.3 now applies. The proof is similar for uniqueness of 1.

Proof. First, for uniqueness of 0, let $y \in \mathbb{R}$ be given such that for all $x \in \mathbb{R}$, $x + y = x$. By AA4, for all $x \in \mathbb{R}$, $x = x + 0$, thus $x + y = x + 0$. By Theorem 2.3, $y = 0$.

You are asked to show uniqueness of 1 as an exercise. ■

There is another way to view Theorems 2.3 and 2.4. For any $a \in \mathbb{R}$, define $f_a, g_a : \mathbb{R} \to \mathbb{R}$ by $f_a(x) = a + x$ and $g_a(x) = ax$ for all $x \in \mathbb{R}$. Theorem 2.3 then says that f_a is an injection, and Theorem 2.4 says that g_a is an injection whenever $a \neq 0$. We'll now show that these functions are also surjections, and hence they are bijections.

Theorem 2.6. *For all $a, b \in \mathbb{R}$, there exists $x \in \mathbb{R}$ so that $a + x = b$. In other words, the function f_a from above is surjective.*

Strategy. We wish to find some x such that $a + x = b$. As in Theorem 2.3, we ought to try and cancel a from the left side by adding something appropriate to both sides. Thus, we use AA5 to get some number y with $a + y = 0$, and then $y + (a + x) = y + b$. This simplifies to $x = y + b$.

Proof. Let $a, b \in \mathbb{R}$ be given. By AA5, we may choose $y \in \mathbb{R}$ so that $a + y = 0$. Now, define $x = y + b$. We compute

$$\begin{array}{rcll} a + x & = & a + (y + b) & \text{def. of } x \\ & = & (a + y) + b & \text{AA2} \\ & = & 0 + b & \text{choice of } y \\ & = & b & \text{AA1 and AA4} \end{array}$$

which shows that $a + x = b$ as desired. ■

Similarly, you can prove the following theorem as an exercise:

Theorem 2.7. *For all $a, b \in \mathbb{R}$, if $a \neq 0$ then there exists $x \in \mathbb{R}$ so that $ax = b$. In other words, the function g_a from above is surjective.*

The last few theorems allow us to define subtraction and division.

Definition 2.8. Let $a, b \in \mathbb{R}$ be given. By Theorems 2.3 and 2.6, there exists exactly one $x \in \mathbb{R}$ so that $a + x = b$ (or equivalently, so that $x + a = b$ by AA1). This value of x is called the *difference* of b by a, written as $x = b - a$ and sometimes called "a *minus* b". The function $- : (\mathbb{R} \times \mathbb{R}) \to \mathbb{R}$ defined in this way is called *subtraction*.

As a special case, for all $b \in \mathbb{R}$, $0 - b$ is called the *negation*[2] of b and is simply written as "$-b$". Thus, in AA5, for any $x \in \mathbb{R}$, $-x$ is precisely the unique real number y with $x + y = 0$.

Similarly, for all $a, b \in \mathbb{R}$, Theorems 2.4 and 2.7 show that if $a \neq 0$, there exists a unique $x \in \mathbb{R}$ with $ax = b$ (or equivalently, with $xa = b$ by AA1). This value of x is called the *quotient* of b by a, written as $x = b/a$ or as $x = \frac{b}{a}$. x is also sometimes called "b *over* a", where b is the *numerator* and a is the *denominator*. The function $/ : (\mathbb{R} \times (\mathbb{R} - \{0\})) \to \mathbb{R}$ defined in this way is called *division*.

As a special case, for all $b \in \mathbb{R}$, if $b \neq 0$ then $1/b$ is called the *reciprocal* of b and is denoted b^{-1}. Thus, in AA6, for any $x \in \mathbb{R}$ with $x \neq 0$, x^{-1} is precisely the unique real number y with $xy = 1$.

From the definitions, we see that for all $x \in \mathbb{R}$, $x - 0 = x$, because $0 + x = x$ by AA4. This also shows that for all $x \in \mathbb{R}$, $x - x = 0$. In addition, for all $x \in \mathbb{R}$, $x/1 = x$ because $1x = x$ (which also shows $x/x = 1$ when $x \neq 0$).

Since we have convenient axioms for addition and multiplication, let's prove a useful theorem allowing us to write subtractions in terms of additions:

Theorem 2.9. *For all $a, b \in \mathbb{R}$, $b - a = b + (-a)$.*

Strategy. $b - a$ is, by definition, the unique number which is added to a to produce b (note that this is ALL we know about $b - a$ just from the definitions above). Thus, we need to show that adding a to $b + (-a)$ gives b as a result. Along the way, we'll use the definition of $-a$, which says that $-a + a = 0$.

Proof. Let $a, b \in \mathbb{R}$ be given. We compute

$$\begin{array}{rcll} a + (b + (-a)) & = & (b + (-a)) + a & \text{AA1} \\ & = & b + ((-a) + a) & \text{AA2} \\ & = & b + 0 & \text{def of } -a \\ & = & b & \text{AA4} \end{array}$$

so, by definition of subtraction, $b - a = b + (-a)$. ∎

Similarly, we also obtain

Theorem 2.10. *For all $a, b \in \mathbb{R}$, if $a \neq 0$ then $b/a = b(a^{-1})$.*

[2] Alternate terms include "the *opposite* of b" or "*minus* b".

Proof. Exercise. ■

The two theorems above allow us to rewrite subtractions and divisions, but they introduce negations and reciprocals into our work. Therefore, we should develop some results allowing us to work with negations and reciprocals. This next result, for instance, shows you never need to negate a number more than once, because double negations can be canceled.

Theorem 2.11. *For all $a \in \mathbb{R}$, $-(-a) = a$. In other words, the negation of a negation is the original number.*

Strategy. We need to use the definition of negations. We know $a + (-a) = 0$. Thus, $-a$ is what you add to a to make 0. Now, $-(-a)$ is what you add to $-a$ to make 0, but the equation above shows that a does the job.

Proof. Let $a \in \mathbb{R}$ be given. By definition of negations, $a + (-a) = 0$. This means, however, that adding a to $-a$ produces 0. Since $-(-a)$ is, by definition, the number which is added to $-a$ to produce 0, $-(-a) = a$. ■

Similarly, we never need to take reciprocals more than once:

Theorem 2.12. *For all $a \in \mathbb{R}$, if $a \neq 0$ then $(a^{-1})^{-1} = a$.*

Proof. Exercise. ■

This next theorem has a surprisingly tricky proof strategy:

Theorem 2.13. *For all $a \in \mathbb{R}$, $0a = 0$ (hence $a0 = 0$ too by AA1).*

Strategy. At this point, the most important property we have for 0 is AA4, which says adding 0 to a number does not change the number. To try and make this useful, we can replace a by $a + 0$ in the expression $0a$. Thus, $0a = 0(a + 0)$, but then AA3 says this equals $0a + 0 \cdot 0$. As $0a = 0a + 0 \cdot 0$, we may cancel $0a$ from each side by Thm 2.3.

However, this yields $0 \cdot 0 = 0$, which is not quite what we want. Looking back, we see we chose to replace a by $a + 0$. What if we replaced 0 by $0 + 0$ instead? Trying to do that, and using AA3 and cancellation, everything works.

Proof. Let $a \in \mathbb{R}$ be given. We compute

$$\begin{array}{rcll} 0a & = & (0+0)a & \text{AA4} \\ & = & 0a + 0a & \text{AA3} \end{array}$$

thus $0a + 0 = 0a + 0a$ by AA4 applied to the left side. By Theorem 2.3, $0 = 0a$. ∎

It also helps to have theorems about how negations interact with products and reciprocals.

Theorem 2.14. *For all $a, b \in \mathbb{R}$, $-(ab) = (-a)b$ (hence $-(ab) = -(ba) = (-b)a = a(-b)$ as well by a couple uses of AA1).*

Strategy. $-(ab)$ is the number which is added to ab to produce 0. Let's show $(-a)b$ also has that property. When computing the sum $ab + (-a)b$, we can factor out b by using AA3.

Proof. Let $a, b \in \mathbb{R}$ be given. We compute

$$\begin{array}{rcll} ab + (-a)b & = & (a + (-a))b & \text{AA3 (and AA1)} \\ & = & 0b & \text{def of } -a \\ & = & 0 & \text{by Thm 2.13} \end{array}$$

so, by definition, $-(ab) = (-a)b$. ∎

Corollary 2.15. *For all $b \in \mathbb{R}$, $-b = (-1)b$.*

Proof. Use Thm 2.14 with $a = 1$, since $1b = b$ by AA4. ∎

Theorem 2.16. *For all $a \in \mathbb{R}$, if $a \neq 0$ then $-a \neq 0$ and $(-a)^{-1} = -(a^{-1})$. In other words, the reciprocal of the negation is the negation of the reciprocal.*

Strategy. Suppose $a \neq 0$. If $-a = 0$, then $0 = a + (-a) = a + 0 = a$, which contradicts our assumption. Hence $-a \neq 0$. Next, $(-a)^{-1}$ is, by definition, the number which produces 1 when multiplied by $-a$. We just need to show $-(a^{-1})$ has this property. The previous theorem will be helpful to work with the negations that occur.

Proof. Let $a \in \mathbb{R}$ be given, and assume $a \neq 0$. Therefore, $a + 0 = a \neq 0$, so 0 cannot be the negation of a. This shows $-a \neq 0$, so $(-a)^{-1}$ exists.

We compute

$$\begin{array}{rcll} (-a)(-(a^{-1})) & = & -(a(-(a^{-1}))) & \text{Thm 2.14} \\ & = & -(-(a \cdot a^{-1})) & \text{Thm 2.14 with AA1} \\ & = & -(-(1)) & \text{def. of } a^{-1} \\ & = & 1 & \text{Thm 2.11} \end{array}$$

which shows, by definition of reciprocal, that $(-a)^{-1} = -(a^{-1})$. ■

At this point, we have discovered a number of useful algebra rules. There are some other useful properties we haven't yet explored, but those properties can mostly be derived by applying the results already shown in this section; see the exercises. In contrast, most of the proofs which have been presented in this section demonstrate important proof strategies with using the axioms, and they emphasize in particular the importance of referring to a definition in a proof.

2.2 Exercises

1. Prove the theorems listed as exercises in this section. Namely,

   (a) Prove Theorem 2.4.

   (b) Finish the proof of Corollary 2.5.

   (c) Prove Theorem 2.7.

   (d) Prove Theorem 2.10.

   (e) Prove Theorem 2.12.

2. Which of the algebraic axioms are satisfied by $\mathbb{Q}$, where the usual operations of $+$ and $\cdot$ are restricted to $\mathbb{Q} \times \mathbb{Q}$? Similarly, which algebraic axioms are satisfied by $\mathbb{Z}$? State your reasoning.

3. This problem will generalize Example 2.2. For any positive integer $n > 1$, we define the arithmetic of $\mathbb{Z}_n$, the *integers modulo* n, as follows. The members of $\mathbb{Z}_n$ are $\{0, 1, 2, \ldots, n-1\}$. If $a, b \in \mathbb{Z}_n$, then the sum $a + b$ in $\mathbb{Z}_n$ is defined to be the remainder of the usual sum $a + b$ in $\mathbb{R}$ when divided by n, and the product ab in $\mathbb{Z}_n$ is defined to be the remainder of the usual product ab in $\mathbb{R}$ when divided by n.

For example, consider when $n = 6$. In $\mathbb{R}$, $2+5 = 7$, which has remainder 1 when divided by 6. Thus, in $\mathbb{Z}_6$, $2+5 = 1$. Also, in $\mathbb{R}$, $2\cdot 5 = 10$, which has remainder 4 when divided by 6, so in $\mathbb{Z}_6$ we have $2 \cdot 5 = 4$. The full addition and multiplication tables for $\mathbb{Z}_6$ can be found in Figure 2.2.

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

·	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Figure 2.2: Addition and multiplication tables for $\mathbb{Z}_6$

For any $n > 1$, you may assume that $\mathbb{Z}_n$, with these definitions of plus and times, satisfies AA1 through AA3. (Showing this requires some number theory which is outside the scope of this book.)

(a) For all $n > 1$, prove that $\mathbb{Z}_n$ satisfies AA4 and AA5.

(b) If $n > 1$ is not prime, show that $\mathbb{Z}_n$ does not satisfy AA6, so $\mathbb{Z}_n$ is not a field. (In fact, when n is prime, $\mathbb{Z}_n$ is a field, but we will not prove this.)

4. Prove the following properties for all $a, b, c \in \mathbb{R}$, using any axioms and theorems desired from the previous section:

   (a) $a(b - c) = ab - ac$.

   (b) $(-a)(-b) = ab$.

   (c) $-(a + b) = (-a) + (-b)$.

   (d) $-(a - b) = b - a$.

   (e) 0 does not have a reciprocal (this can be casually stated as "1/0 does not exist").

5. Let $a, b, c, d \in \mathbb{R}$ be arbitrary, and assume $b \neq 0$ and $d \neq 0$. By the Factoring Theorem, $bd \neq 0$, so $(bd)^{-1}$ exists. Prove the following properties, using any axioms and theorems desired from the previous section or from the previous exercise:

(a) $(bd)^{-1} = b^{-1}d^{-1}$.

(b) $-\left(\frac{a}{b}\right) = \frac{-a}{b} = \frac{a}{-b}$.

(c) $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$.

(d) $\frac{a}{b} - \frac{c}{d} = \frac{ad-bc}{bd}$.

(e) $\left(\frac{a}{b}\right)\left(\frac{c}{d}\right) = \frac{ac}{bd}$.

(f) $\frac{a}{b} = 0$ iff $a = 0$.

(g) $c \neq 0 \to \left(\frac{a}{b}\right) / \left(\frac{c}{d}\right) = \frac{ad}{bc}$

6. In Example 2.1, we needed to use the fact that $\sqrt{2}$ was irrational in order to show $\mathbb{Q}[\sqrt{2}]$ satisfied AA6. Prove that $\sqrt{2}$ is irrational. (Hint: If it could be written as p/q where $p, q \in \mathbb{Z}$ and the fraction p/q is in lowest terms, then $2 = p^2/q^2$. Use this to show p is even, then use that information to show q is even, contradicting that p/q is in lowest terms.)

7. Using the result of Exercise 2.2.6, prove that for any $x \in \mathbb{Q}[\sqrt{2}]$, there is exactly one way to choose $a, b \in \mathbb{Q}$ so that $x = a + b\sqrt{2}$. In other words, prove that the function $f : \mathbb{Q}\times\mathbb{Q} \to \mathbb{Q}[\sqrt{2}]$, defined by $f(a,b) = a + b\sqrt{2}$ for all $a, b \in \mathbb{Q}$, is a bijection.

2.3 The Order Axioms

The algebraic axioms tell us a lot about the structure of $\mathbb{R}$. They also help us distinguish $\mathbb{R}$ from some other sets. For example, arithmetic in $\mathbb{N}$ behaves very differently from arithmetic in $\mathbb{R}$ because $\mathbb{N}$, with its usual plus and times, doesn't satisfy AA5 or AA6. In fact, the study of $\mathbb{N}$ and $\mathbb{Z}$ forms the basis of *number theory*, one of the oldest forms of mathematics.

The algebraic axioms, while useful, do not address inequalities, a very important aspect of $\mathbb{R}$. Inequalities are everywhere in day-to-day events: examples include comparing prices of goods, trying to find the shortest walking route to class, and driving below speed limits (there is no law that says you have to drive EXACTLY the speed limit; the speed limit law is an inequality). In mathematics, we want to find optimal solutions to problems, such as finding out which price to set to earn the most profit, or discovering how to minimize the cost of material needed to make a container with a fixed volume. All of these tasks require having a well-defined notion of comparing numbers, i.e. saying that some numbers are larger than others.

Thus, we should establish a few axioms allowing us to make statements about which numbers are larger than others, called the *order axioms*. We will abbreviate the order axioms with "OA". For convenience, it is common to state order axioms in terms of a distinguished subset $\mathbb{R}^+$ of $\mathbb{R}$, called the *positive numbers*. Just like we didn't define $+$ and $\cdot$ for $\mathbb{R}$ but we instead gave properties of them as axioms, the order axioms don't define $\mathbb{R}^+$ but merely state important properties of it. We can then define our usual comparisons in terms of $\mathbb{R}^+$:

Definition 2.17. The proposition $<(x, y)$, meaning "x is *less than* y", is defined as "$y - x \in \mathbb{R}^+$". We frequently write $<$ in the middle as an infix operator like we did with $+$ and $\cdot$. Thus, the definition can be rephrased as "For all $x, y \in \mathbb{R}$, $x < y$ iff $y - x \in \mathbb{R}^+$".

We also similarly define, for all $x, y \in \mathbb{R}$:

- "$x \leq y$" means "$x < y$ or $x = y$" (said "x is *less than or equal to* y")
- "$x > y$" means "$y < x$" (said "x is *greater* than y")
- "$x \geq y$" means "$y \leq x$" (said "x is *greater than or equal to* y")

Note that when $x, y \in \mathbb{R}$ and $x < y$, some mathematicians will choose to say "x is *strictly* less than y" to emphasize that $x \neq y$ (we shall prove that $x < y$ implies $x \neq y$ soon).

Axiom (OA1: Closure of $\mathbb{R}^+$ under Plus and Times). For all $x, y \in \mathbb{R}^+$, both $x + y$ and xy are in $\mathbb{R}^+$.

Axiom (OA2: Uniqueness of Signs). For all $x \in \mathbb{R}^+$, if $x \neq 0$ then exactly one of $x \in \mathbb{R}^+$ and $-x \in \mathbb{R}^+$ is true.

When $-x \in \mathbb{R}^+$ is true, we say that x is *negative*. (The *sign* of $x \neq 0$ is $+$ if $x \in \mathbb{R}^+$ and $-$ if $-x \in \mathbb{R}^+$). Similarly, if $x \geq 0$, then we say x is *nonnegative*, and if $x \leq 0$, then we say x is *nonpositive*.

Axiom (OA3: Zero Has No Sign). $0 \notin \mathbb{R}^+$

Before proceeding, a couple remarks are worthwhile concerning the sign of a number. For each $x \in \mathbb{R}$, we defined "x is positive" to mean "$x \in \mathbb{R}^+$". Since $x - 0 = x$, $x \in \mathbb{R}^+$ is equivalent to $0 < x$, so the positive numbers are the numbers strictly greater than 0. Similarly, $0-x=-x$, so x is negative iff $x<0$. (These results sound obvious, but they need to be checked because of the way $<$ was defined.) Thus, OA2 and OA3 imply that every nonzero real number is either positive or negative, but no real number is both positive and negative.

The first question we should address is: for any two real numbers x and y, is it always the case that either $x \le y$ or $y \le x$? In other words, is it possible to compare any two numbers? (If so, this makes the real numbers rather different from fruits, since as the old saying goes, you can't compare apples and oranges.) It turns out we can compare any two numbers:

Theorem 2.18 (Trichotomy). *For all $a, b \in \mathbb{R}$, exactly one of the three statements "$a<b$", "$a=b$", and "$a>b$" is true. Equivalently (by using the definitions of $\le$ and $\ge$), "$a \le b$" or "$a \ge b$" is true, with both being true iff $a=b$.*

Strategy. For all a, b, we defined $a<b$ to mean $b-a \in \mathbb{R}^+$. Thus, we want to consider the number $x=b-a$. If $x=0$, then $a=b$. Otherwise, either $x>0$ or $x<0$ but not both. This leads to either $a<b$ or $a>b$ but not both. We also need to use OA2 and OA3 carefully to show no two of "$a=b$", "$a<b$", and "$a>b$" can hold at the same time.

Proof. Let $a, b \in \mathbb{R}$ be given, and define $x=b-a$. In the first case, suppose $x=0$. It follows that $a=b$. By OA3, "$a<b$" is false. Also, since $a=b$, $a-b=0$ as well, so "$a>b$" is false by OA3.

In the second case, $x \ne 0$. Thus, by OA2, exactly one of "$x \in \mathbb{R}^+$" and "$-x \in \mathbb{R}^+$" is true. $x \in \mathbb{R}^+$ is equivalent to $a<b$ by definition of $<$, and $-x \in \mathbb{R}^+$ holds iff $-(b-a)=a-b \in \mathbb{R}^+$, which holds iff $a>b$ by definition of $>$. Thus, in all possibilities, exactly one comparison is true. ∎

Remark. Theorem 2.18 immediately implies that for all $a \in \mathbb{R}$, "$a<a$" is false, because "$a=a$" is true. Since $a=a$ is true for all $a \in \mathbb{R}$, we sometimes say that the $=$ operator is *reflexive*. In contrast, we just showed $<$ is *irreflexive*.

A crucial property of inequalities is that they can be "chained", as this next theorem states:

Theorem 2.19 (Transitivity of $<$). *For all $a, b, c \in \mathbb{R}$, if $a<b<c$ (this is shorthand for saying $a<b$ and $b<c$), then $a<c$.*

Strategy. $a < b$ means $b - a$ is positive, and $b < c$ means $c - b$ is positive. We want to show $c - a$ is positive. This is where OA1 helps, since $c - a = (b - a) + (c - b)$.

Proof. Let $a, b, c \in \mathbb{R}$ with $a < b$ and $b < c$ be given. Therefore, $b - a \in \mathbb{R}^+$ and $c - b \in \mathbb{R}^+$. Now, $(b-a)+(c-b) = b+(-a)+c+(-b)$ by Theorem 2.9, which equals $(-a) + c = c - a$. Thus, by OA1, $c - a \in \mathbb{R}^+$, and by definition of $<$, $a < c$. ■

Next, we show that adding the same number to both sides of an inequality is safe:

Theorem 2.20. *For all $a, b, c \in \mathbb{R}$, if $a < b$ then $a + c < b + c$.*

Strategy. We just need to compute $(b+c)-(a+c)$ in terms of $b-a$. It turns out they are exactly the same number.

Proof. Let $a, b, c \in \mathbb{R}$ be given with $a < b$. Therefore, $b - a \in \mathbb{R}^+$. It follows that $(b + c) - (a + c) = b + c + (-a) + (-c)$ (by Exercise 2.2.4.(c)), which equals $b + (-a) = b - a$. Thus, $(b + c) - (a + c) \in \mathbb{R}^+$, so $a + c < b + c$. ■

However, when multiplying the same number to both sides of an inequality, we have to pay attention to the sign of that number:

Theorem 2.21. *For all $a, b, c \in \mathbb{R}$, if $a < b$ and $c > 0$ then $ac < bc$.*

Strategy. Here, $bc - ac = (b - a)c$. If $b - a$ is positive, and c is positive, then so is the product.

Proof. Let $a, b, c \in \mathbb{R}$ be given with $a < b$ and $c > 0$. Therefore, $b - a \in \mathbb{R}^+$ and $c \in \mathbb{R}^+$, so $(b - a)c \in \mathbb{R}^+$ by OA1. However, $(b - a)c = bc - ac$ by AA1 and Exercise 2.2.4.(a), so $ac < bc$. ■

These last few results are the most important tools for working with inequalities. Many other properties can be obtained by applying these previous results; see the exercises. However, while we've discussed how to work with positive numbers, we have yet to come up with any numbers that we know are positive. Let's end this section by finding some positive numbers:

Theorem 2.22. *For all $a \in \mathbb{R}$, if $a \neq 0$ then $a^2 > 0$. In particular, when $a = 1$, we obtain $1 > 0$.*

Strategy. If $a \neq 0$, then a is either positive or negative. OA1 shows that if a is positive, then a^2 is positive. If a is negative, however, we can write it as $-b$ where $b = -a > 0$, and then squaring a "cancels the minus sign".

Proof. Let $a \in \mathbb{R}$ with $a \neq 0$ be given. If $a \in \mathbb{R}^+$, then $a^2 = a \cdot a \in \mathbb{R}^+$ by OA1. Otherwise, by OA2, $-a \in \mathbb{R}^+$, and thus $a^2 = a \cdot a = (-a)(-a)$ by Exercise 2.2.4.(b), and $(-a)(-a) \in \mathbb{R}^+$ by OA1. ■

2.4 Exercises

1. In the statement of OA2, it says that for nonzero $x \in \mathbb{R}$, exactly one of $x \in \mathbb{R}^+$ and $-x \in \mathbb{R}^+$ is true. Suppose we weaken this statement to "for all $x \in \mathbb{R}$ with $x \neq 0$, $x \in \mathbb{R}^+$ or $-x \in \mathbb{R}^+$"; call this weaker statement WOA2 (Weak Order Axiom 2). Thus, WOA2 doesn't exclude the possibility that both $x \in \mathbb{R}^+$ and $-x \in \mathbb{R}^+$ when $x \neq 0$.

   Prove that OA1, WOA2, and OA3 together imply OA2. Hence, when we want to show that a field F can be ordered (by finding a subset F^+ satisfying the order axioms), we only have to check that WOA2 is satisfied instead of showing that OA2 is satisfied.

2. Since our main properties above use the $<$ symbol, it's good to have some similar properties with $\leq$ in them. The following lemma helps when proving properties with $\leq$. Prove this lemma:

   Lemma 2.23. *For all $a, b \in \mathbb{R}$, if $a, b \geq 0$, then $a + b \geq 0$ and $ab \geq 0$.*

3. Prove the following properties of inequalities. You may use any results from the additive axioms (including the exercises), any results in the previous section, or any earlier parts of this problem:

   (a) The sum of any two negative numbers is negative.
   (b) $\forall a, b, c \in \mathbb{R}\ (a < b \wedge c < 0) \rightarrow ac > bc$. (Thus, multiplying by a negative number "flips" an inequality.)
   (c) $\forall a, b \in \mathbb{R}\ (a < b \leftrightarrow -a > -b)$. In particular, for all $a \in \mathbb{R}$, a is positive iff $-a$ is negative.
   (d) For all $a, b \in \mathbb{R}$, if $ab > 0$, then a and b are both nonzero and have the same sign.

(e) For all $a, b, c, d \in \mathbb{R}$, if $a < b$ and $c < d$ then $a + c < b + d$. (Thus, we can "add inequalities".)

4. Suppose that $a, b, c, d \in \mathbb{R}$ are given with $a < b$ and $c < d$.
   (a) Show that "$a - c < b - d$" might be false, i.e. we cannot "subtract inequalities".
   (b) Show that "$ac < bd$" might be false, i.e. we cannot "multiply inequalities". How can you add an extra hypothesis to make $ac < bd$ true?
   (c) Using your new hypothesis from part (b), as well as the old hypotheses $a < b$ and $c < d$, prove $ac < bd$.
   (d) Using part (c), prove that if $0 < a < b$, then $a^2 < b^2$.

5. (a) Prove that for all $a \in \mathbb{R}$, if $a \neq 0$ then a and a^{-1} have the same sign. (Hint: do a proof by contradiction.)
   (b) Prove that if $a, b \in \mathbb{R} - \{0\}$ satisfy $0 < a < b$, then $0 < b^{-1} < a^{-1}$. Also, show that "$a < b$ implies $b^{-1} < a^{-1}$" might be false.

6. Prove that for all $a, b, c \in \mathbb{R}$, if $a \leq b$, $b \leq c$, and $a = c$, then $a = b = c$.

7. (a) Prove that there is no largest real number, i.e. there does not exist $x \in \mathbb{R}$ such that for all $a \in \mathbb{R}$, $a \leq x$.
   (b) Prove that there is no smallest real number, i.e. there does not exist $x \in \mathbb{R}$ such that for all $a \in \mathbb{R}$, $x \leq a$. This and part (a) are sometimes summarized by saying that $\mathbb{R}$ *has no endpoints*.

8. Prove $\forall x \in \mathbb{R}\ (\forall h \in \mathbb{R}^+\ 0 \leq x < h) \rightarrow x = 0$. Thus, 0 is the only nonnegative number smaller than every positive number.

9. Prove that $\mathbb{C}$ cannot be ordered, i.e. it is impossible to pick a subset $\mathbb{C}^+ \subseteq \mathbb{C}$ of positive complex numbers satisfying OA1 through OA3. This shows an important way that $\mathbb{R}$ and $\mathbb{C}$ differ.

10. See Exercise 2.2.3 for a definition of $\mathbb{Z}_n$ where n is a positive integer greater than 1. Why can't $\mathbb{Z}_n$ be ordered for any $n > 1$?

11. (a) Show that for all $x, y \in \mathbb{R}$, $2xy \leq x^2 + y^2$. (Hint: first show $(x - y)^2 \geq 0$... this isn't quite the same as Theorem 2.22 because $\geq$ is used instead of $>$.)

(b) Suppose that $a, b \in \mathbb{R}$ satisfy $a, b \geq 0$. Using part (a), show that $\sqrt{ab} \leq (a+b)/2$, and that this is an equality iff $a = b$ (you may use the fact that the square root function is injective on the nonnegative real numbers). This result is known as the *Arithmetic-Geometric Mean Inequality*, or the AGM Inequality for short.

12. Prove that $\mathbb{R}$ is *dense*, i.e. for all $x, y \in \mathbb{R}$ with $x < y$, there exists $z \in \mathbb{R}$ with $x < z < y$.

2.5 The Completeness Axiom

The last few sections have explored what it means for $\mathbb{R}$ to be an *ordered field*: this means it satisfies axioms AA1 through AA6 (describing useful algebraic laws) and axioms OA1 through OA3 (describing important properties of inequalities and sign). You can show that $\mathbb{Q}$ is also an ordered field, using the same operations and ordering as $\mathbb{R}$ (but restricted to $\mathbb{Q}$). However, if both $\mathbb{R}$ and $\mathbb{Q}$ are ordered fields, why is $\mathbb{R}$ used so frequently in calculus (functions from $\mathbb{R}$ to $\mathbb{R}$ are especially common), whereas you rarely see a function in calculus from $\mathbb{Q}$ to $\mathbb{Q}$?

Example 2.24:

Consider the famous Pythagorean Theorem: $a^2 + b^2 = c^2$, where a and b are lengths of legs of a right triangle and c is the hypotenuse length. This equation can be used to find the third side of a triangle from knowledge of the other two sides. However, suppose $a = b = 1$, and we want to find the hypotenuse length c. Putting in our values of a and b, we obtain $c^2 = 2$. This equation has no solutions for c when c is rational, by Exercise 2.2.6, but intuitively this equation ought to have a positive solution (after all, the hypotenuse SHOULD have a length). In $\mathbb{R}$, unlike in $\mathbb{Q}$, there is a unique positive solution for c, namely $c = \sqrt{2} \approx 1.414214$.

Another way of looking at this situation is to consider the sets $A = \{q \in \mathbb{Q}^+ \mid q^2 < 2\}$ and $B = \{q \in \mathbb{Q}^+ \mid q^2 > 2\}$. It can be shown that for every $a \in A$ and every $b \in B$, $a < b$ holds (using exercise 2.4.4.(d)). Hence, a real number c satisfying $c^2 = 2$ would have to be between A and B. However, there is no such number in $\mathbb{Q}$, so there is a "gap" between the sets A and B

in $\mathbb{Q}$ with no number to fill the gap. Intuitively, this suggests that $\mathbb{R}$ may be more suitable for our calculus because $\mathbb{R}$ fills in $\mathbb{Q}$'s gaps.[3] $\square$

We haven't yet formally justified that 2 has a square root in $\mathbb{R}$; we've only explained why such a number ought to be in $\mathbb{R}$. In fact, it's impossible to prove that 2 has square roots using only the axioms for an ordered field (otherwise, $\mathbb{Q}$ would also contain $\sqrt{2}$, since $\mathbb{Q}$ is an ordered field). We need another axiom for $\mathbb{R}$ which will allow us to "fill the gap" between the sets A and B from the last example. To do this, we need some new terminology.

Definition 2.25. Let S be a subset of $\mathbb{R}$ and x be a member of $\mathbb{R}$. We say x is an *upper bound* of S if for all $a \in S$, $a \leq x$. If S has any upper bounds, we say S is *bounded above*, otherwise we say S is *unbounded above*. Similarly, x is a *lower bound* of S if for all $a \in S$, $x \leq a$. We say S is *bounded below* if it has any lower bounds, otherwise we say S is *unbounded below*. We say S is *bounded* if S is bounded both above and below, otherwise S is *unbounded*.

If x is both a member of S and an upper bound of S, we say that x is a *maximum* element of S, denoted $x = \max S$. Similarly, if x is both a member of S and a lower bound of S, we say that x is a *minimum* element of S, denoted $x = \min S$.

Remark. If S is a set with a maximum element, then the maximum element is uniquely determined. To prove this, suppose $x, y \in \mathbb{R}$ are both maximum elements of S. Since $x \in S$ (by the definition of maximum) and y is maximum, so y is an upper bound of S, we must have $x \leq y$. Also, since $y \in S$ and x is maximum, $y \leq x$. By trichotomy, $x = y$, so there can be at most one maximum element.

The same style of argument proves there is at most one minimum element. Thus, we may use the word "the" when describing maximum and minimum values.

Example 2.26:

Here are a few examples of subsets of $\mathbb{R}$ with information about their bounds:

- The set $\mathbb{R}^+$ is bounded below, because 0 is a lower bound. In addition, any number x with $x \leq 0$ is also a lower bound of $\mathbb{R}^+$. By Exercise 2.4.8, the only lower bound of $\mathbb{R}^+$ which is not smaller than 0 is 0 itself,

[3] In fact, a formal construction of $\mathbb{R}$ can be done by precisely defining gaps in $\mathbb{Q}$ and designing $\mathbb{R}$ to fill every gap. For more information, search for "Dedekind cuts".

so 0 is the greatest of the lower bounds of $\mathbb{R}^+$. Since $0 \notin \mathbb{R}^+$, $\mathbb{R}^+$ does not have a minimum element.

Exercise 2.4.7.(a) shows that $\mathbb{R}^+$ is not bounded above.

- The closed interval $[0, 1]$ has the minimum element 0 and the maximum element 1. In particular, $[0, 1]$ is bounded both above and below, so it is bounded.

- The open interval $(0, 1)$ has no minimum element. To prove this, let $x \in (0, 1)$ be given; we show that there is some $y \in (0, 1)$ with $x \not\leq y$ (i.e. $y < x$), which shows x is not a minimum element of $(0, 1)$. Since $2 = 1 + 1 > 1 > 0$, $1/2 > 0$ as well and $1/2 < 1$ by Exercise 2.4.5. Therefore, as x is positive, $x/2 < x < 1$ and $x/2 > 0$. This means we may choose $y = x/2$, and then $y \in (0, 1)$ and $y < x$, finishing the proof.

  A similar style of proof shows that $(0, 1)$ has no maximum element either. However, 0 is a lower bound of $(0, 1)$, and 1 is an upper bound, so $(0, 1)$ is bounded.

$\square$

When upper and lower bounds exist for sets, the bounds are certainly not unique; while 0 is a lower bound of $\mathbb{R}^+$, -1, -42, and $-1/6$ are also lower bounds of $\mathbb{R}^+$, to name a few. We feel 0 is the most useful of the lower bounds, because no greater number is a lower bound for $\mathbb{R}^+$. Similarly, 1 is the most useful upper bound of $[0, 1]$ and of $(0, 1)$, since it can be shown that no smaller number is an upper bound of either of those sets. These "most useful" upper and lower bounds receive special names:

Definition 2.27. Let $S \subseteq \mathbb{R}$ and $x \in \mathbb{R}$ be given. We say that x is a *least upper bound* of S, also called a *supremum* of S and written $x = \sup S$, if x is a minimum of the set of upper bounds for S. (Hence, since minimum elements are unique if they exist, a set can have at most one supremum.) Similarly, we say x is a *greatest lower bound* of S, also called an *infimum* of S and written $x = \inf S$, if x is a maximum of the set of lower bounds for S.

Note that not all sets of real numbers have to have suprema or infima. Furthermore, if $S \subseteq \mathbb{R}$ has a supremum, then $\sup S$ might not belong to S. It is possible to prove that $(\sup S) \in S$ iff S has a maximum element, and in that case $\max S = \sup S$. An analogous statement holds for the infimum and the minimum.

Thus, $\sup[0,1] = \sup(0,1) = 1$, $\inf[0,1] = \inf(0,1) = 0$, and $\inf \mathbb{R}^+ = 0$. However, $\sup \mathbb{R}+$ does not exist.

Looking back at Example 2.24, we created the sets $A = \{q \in \mathbb{Q}^+ \mid q^2 < 2\}$ and $B = \{q \in \mathbb{Q}^+ \mid q^2 > 2\}$. A does not have a maximum element, but A is bounded above, as every member of B is an upper bound of A. Intuitively, $\sqrt{2}$ should be an upper bound of A, but it should also be smaller than every member of B, so $\sqrt{2}$ should be a small upper bound of A. This gives us the idea to define $\sqrt{2}$ as the least upper bound of A, provided this number exists.

The last axiom for $\mathbb{R}$ allows us to fill the gap between A and B. It tells us, in essence, that $\mathbb{R}$ has no "holes" in it, so the axiom is commonly called the *completeness axiom*.

Axiom (C: *Completeness*). Every nonempty subset of $\mathbb{R}$ which is bounded above has a supremum.

You may ask: why is Axiom C stated in terms of suprema ("suprema" is the plural of "supremum") but says nothing in terms of infima? This is because Axiom C can be used to prove the following theorem:

Theorem 2.28. *Every nonempty subset of* $\mathbb{R}$ *which is bounded below has an infimum.*

Strategy. Suppose $S \subseteq \mathbb{R}$ is nonempty and bounded below, say $b \in \mathbb{R}$ is a lower bound. To find a greatest lower bound, we want to turn our problem into a problem we already know how to solve with Axiom C: the problem of finding least upper bounds.

One option is to "flip" our comparisons by using negation. This is because for all $x, y \in \mathbb{R}$, $x \le y$ is equivalent to $-x \ge -y$. Thus, we consider $-S$, the set of negations of members of S, which has the UPPER bound of $-b$. This means $-S$ has a supremum c, and we show that $-c$ is the infimum of S.

Another option is to consider the set of lower bounds of S. We want to show there is a maximum lower bound. Thus, we can use Axiom C to find a supremum for the lower bounds, and then we show it is actually a lower bound for S. You are asked in Exercise 2.7.1 to provide a proof of this theorem using this method.

Proof. Let $S \subseteq \mathbb{R}$ be given which is nonempty and bounded above. Hence, there is some $x \in S$ and some $b \in \mathbb{R}$ so that b is a lower bound of S.

Define $T = \{-y \mid y \in S\}$ (sometimes this set is written as $-S$). Since $x \in S$, $-x \in T$, so T is nonempty. Also, for all $y \in S$, $y \geq b$ (as b is a lower bound of S), so $-y \leq -b$, which shows $-b$ is an upper bound of T.

By Axiom C, T has a supremum, say $c = \sup T$. It remains to show $-c = \inf S$. First, to see why $-c$ is a lower bound of S, for all $y \in S$, $-y \leq c$ (because c is an upper bound of T and $-y \in T$), so $y \geq -c$. Second, if $d \in \mathbb{R}$ is any lower bound of S, then for all $y \in S$, $y \geq d$, so $-y \leq -d$, showing that $-d$ is an upper bound of T. Because c is the LEAST upper bound of T, $-d \geq c$, so $d \leq -c$, showing that $-c$ is the greatest lower bound. ■

Other important consequences of Axiom C will be discussed in the remainder of this chapter. Before proceeding, however, a couple remarks concerning $\sqrt{2}$ are in order. The discussion around Example 2.24 indicates how to find $\sqrt{2}$ as the supremum of a set, but we have not yet shown why that supremum c actually has the property that $c^2 = 2$. This theorem provides the necessary justification:

Theorem 2.29. *For every $x \geq 0$, there is a number $c \geq 0$ so that $c^2 = x$. c is called a nonnegative* square root *of x, and we write $c = \sqrt{x}$ or $c = x^{1/2}$. (Exercise 2.7.2 shows that there is exactly one nonnegative square root.)*

The main idea of the proof of Theorem 2.29 is to define $c = \sup\{y \in \mathbb{R} \mid y^2 < x\}$, once we have argued that this supremum is well-defined. We then show that we cannot have either $c^2 > x$ or $c^2 < x$. However, the proof is particularly difficult and unenlightening. (The proof shows that if $c^2 > x$, then c cannot be an upper bound, and if $c^2 < x$, then c cannot be a LEAST upper bound.) You may take this theorem on faith for now. In Chapter 3, when we study continuity, we will revisit the issue of defining roots and provide a simple proof to show that n^{th} roots of nonnegative numbers exist whenever $n \in \mathbb{N}^*$.

2.6 Calculations with Suprema

The completeness axiom, Axiom C, is useful for showing solutions exist for certain equations (for example, the previous section shows how to find a solution to the equation $x^2 = 2$), but the axiom doesn't tell you very much information about what the solution actually is. Hence, it will be useful to develop a few tools for doing calculations with suprema and infima. As one example, we have the following relationship between infima, suprema, and negatives:

Theorem 2.30. *If $S \subseteq \mathbb{R}$ is nonempty, then S is bounded below iff $-S = \{-x \mid x \in S\}$ is bounded above, in which case $-\inf S = \sup(-S)$.*

Most of the work of this theorem was done in the proof of Theorem 2.28, so we omit the proof.

The next result addresses what happens when the members of two sets are added together. If A and B are any subsets of $\mathbb{R}$, we define $A + B$ to be the set $\{a + b \mid a \in A, b \in B\}$. Note that in the sum $A + B$, every member of A gets added to every member of B. For example, if $A = \{1, 3, 4\}$ and $B = \{2, 5\}$, then $A+B = \{1+2, 3+2, 4+2, 1+5, 3+5, 4+5\} = \{1, 3, 5, 6, 8, 9\}$; in this sum of sets, the number 6 can be obtained in two ways, either as $4+2$ or as $1 + 5$.

Example 2.31:

Let's show that $\mathbb{Z} + \mathbb{Z} = \mathbb{Z}$. To prove $\mathbb{Z} \subseteq \mathbb{Z} + \mathbb{Z}$, if $n \in \mathbb{Z}$ is given, then n can be written as $n + 0$, and $0 \in \mathbb{Z}$, so $n \in \mathbb{Z} + \mathbb{Z}$. To prove $\mathbb{Z} + \mathbb{Z} \subseteq \mathbb{Z}$, if $a, b \in \mathbb{Z}$ are given, then their sum $a + b$ is also an integer, so $a + b \in \mathbb{Z}$.

One common mistake to make is to guess that $\mathbb{Z} + \mathbb{Z}$ consists only of even integers. After all, for any $a \in \mathbb{Z}$, the sum $a + a$ is even. However, the sum $\mathbb{Z} + \mathbb{Z}$ can add any integer to ANY other integer; certainly sums of the form $a + a$ appear in $\mathbb{Z} + \mathbb{Z}$, but other sums are included too. Exercise 2.7.4 explores this further.

As another example, consider $A = \{\sin x \mid x \in \mathbb{R}\}$ and $B = \{\cos x \mid x \in \mathbb{R}\}$. In other words, A and B are respectively the ranges of the sine and cosine functions, so $A = B = [-1, 1]$. In Exercise 2.7.3, you are asked to prove that $A + B = [-2, 2]$. Note that for this example, A and B both have the maximum element 1, and $A + B$ has the maximum element $1 + 1$, or 2. In other words, $\max(A + B) = \max A + \max B$. $\square$

Theorem 2.32. *Let $A, B \subseteq \mathbb{R}$ be given. If A and B both have suprema, then so does $A + B$, and $\sup(A + B) = \sup A + \sup B$. Similarly, if A and B both have infima, then so does $A + B$, and $\inf(A + B) = \inf A + \inf B$.*

Strategy. Let's just focus on the part with suprema; the other part of the theorem, dealing with infima, is left as Exercise 2.7.5. Suppose that $\sup A$ and $\sup B$ both exist. The main idea is that although $\sup A$ and $\sup B$ might not be members of A or of B, there are elements of A which are arbitrarily close to $\sup A$, and similarly for B. Therefore, the sums in $A+B$ should be arbitrarily close to $\sup A + \sup B$.

However, this needs to be more formal. To show $\sup(A+B) = \sup A + \sup B$, by the definition of supremum for the set $A+B$, we need to show that $\sup A + \sup B$ is an upper bound of the set $A+B$, and then we need to show that no other upper bound is smaller.

Let's consider how to do the second task. Suppose that $x < \sup A + \sup B$. We want to show that x is too small to be an upper bound of $A+B$ by finding elements $a \in A$ and $b \in B$ such that $a+b > x$. Ideally, a and b should be close to $\sup A$ and $\sup B$ respectively, but how close?

The trick we'll use is to examine distances. Let $h = \sup A + \sup B - x$, so $h > 0$ and h is the distance from x to $\sup A + \sup B$. If $h_1 = \sup A - a$ and $h_2 = \sup B - b$, how small do we want h_1 and h_2 to be? Note that $\sup A + \sup B - (a+b) = (\sup A - a) + (\sup B - b) = h_1 + h_2$. Since we want to have $a+b > x$, we'd like $\sup A + \sup B - (a+b) < \sup A + \sup B - x$, i.e. $h_1 + h_2 < h$. This suggests that we choose $h_1 = h_2 < h/2$.

Proof. We only prove the statement about suprema here; the statement about infima is Exercise 2.7.5. Assume that $\sup A$ and $\sup B$ exist. For all $a \in A$ and $b \in B$, $a \le \sup A$ and $b \le \sup B$, because the suprema are upper bounds. Thus, $a+b \le \sup A + \sup B$. This shows that $\sup A + \sup B$ is an upper bound of the set $A+B$.

Next, let $x \in \mathbb{R}$ be given with $x < \sup A + \sup B$; it remains to prove that x is not an upper bound of $A+B$. Let $h = \sup A + \sup B - x$, so $h > 0$. Therefore, $h/2 > 0$ as well. There must exist some $a \in A$ with $a > \sup A - h/2$, or else the number $\sup A - h/2$ would be a smaller upper bound of A than $\sup A$, contradicting the definition of supremum. Choose such an $a \in A$. Similarly, choose some $b \in B$ such that $b > \sup B - h/2$. Therefore,

$$a + b > (\sup A - h/2) + (\sup B - h/2) = (\sup A + \sup B) - h = x$$

finishing the proof. ∎

The previous proof has a couple of tactics which are very common when arguing with suprema and infima:

- The proof used the fact that $\sup A$ is the LEAST upper bound of A in order to guarantee that there is an element $a \in A$ with $a > \sup A - h/2$, where h could be any positive number. In other words, we could find elements of A arbitrarily close to $\sup A$. This makes the supremum of A very useful, more so than the other upper bounds.

- Note how the proof also used $h/2$ twice as a distance, so that in the end the two $h/2$'s would add together and form h. Frequently, when a proof wants to find a number within some specified distance (such as here, when we wanted to find a member of $A + B$ with distance less than h from $\sup A + \sup B$), that distance is often broken into a few smaller, but still positive, pieces. We'll see this tactic again when talking about limits in the next chapter.

For the last result in this section, recall the sets we used to explain the "gap" that $\sqrt{2}$ filled: they were $A = \{q \in \mathbb{Q}^+ \mid q^2 < 2\}$ and $B = \{q \in \mathbb{Q}^+ \mid q^2 > 2\}$. All of A's elements are smaller than all of B's elements, so intuitively A and B leave a number in the middle between them. In some sense, the members of A are underestimates for $\sqrt{2}$, whereas the members of B are overestimates for $\sqrt{2}$, and right in the middle is $\sqrt{2}$.

In general, a useful way to find a number is to create a set of overestimates and a set of underestimates, and then you argue that the number you want is in between the two sets. This will be especially useful in the definition of the integral in Chapter 5. Let's prove the main theorem that says there really is some number in the middle:

Theorem 2.33. *Let $A, B \subseteq \mathbb{R}$ be two nonempty sets such that for all $a \in A$ and all $b \in B$, $a \leq b$. Then* $\sup A$ *and* $\inf B$ *both exist, and* $\sup A \leq \inf B$.

Strategy. The set A has upper bounds, because any member of B is an upper bound of A. Thus, by Axiom C, $\sup A$ exists. However, $\sup A$ is the minimum of the upper bounds of A; since all the members of B are themselves upper bounds of A, $\sup A$ is smaller than or equal to all the members of B. This shows B has $\sup A$ as a lower bound, but $\inf B$ is the greatest lower bound of B.

Proof. Assume that $A, B \neq \emptyset$ and that for all $a \in A$ and all $b \in B$, $a \leq b$. Let $b \in B$ in arbitrary, so that we have $\forall a \in A\ a \leq b$. This shows that b is an upper bound of A. As A is nonempty as well, Axiom C says that $\sup A$ exists. Because $\sup A$ is the least upper bound of A, and b is an upper bound of A, $\sup A \leq b$.

Thus, $\forall b \in B\ \sup A \leq b$ is true, so $\sup A$ is a lower bound of B. Therefore, by Theorem 2.28, $\inf B$ exists. As $\inf B$ is the greatest lower bound of B, and $\sup A$ is a lower bound of B, we must have $\sup A \leq \inf B$. ■

2.7 Exercises

1. Give another proof of Theorem 2.28, using the second strategy outlined. Your proof should consider the set B of lower bounds of S, where S is a nonempty set which is bounded below, and it should prove $\sup B$ exists and equals $\max B$. (Hint: Once you've shown $\sup B$ exists, assume $\sup B$ is not in B and derive a contradiction.)

2. Let $a \geq 0$ be given. Thus, there exists at least one $c \in \mathbb{R}^+ \cup \{0\}$ with $c^2 = a$. Prove there is exactly one such c. (Hint: Solutions to the equation $x^2 = c^2$ for $x \in \mathbb{R}$ are solutions to the equation $x^2 - c^2 = 0$.)

3. Prove that $[-1, 1] + [-1, 1] = [-2, 2]$.

4. Let $f, g : [0, 1] \to [0, 1]$ be arbitrary. Assume that $\max(\text{ran} f)$ and $\max(\text{ran} g)$ both exist.
   (a) Prove that $\sup\{f(x)+g(x) \mid x \in [0, 1]\}$ exists and that $\sup\{f(x)+g(x) \mid x \in [0, 1]\} \leq \max(\text{ran} f) + \max(\text{ran} g)$.
   (b) Provide an example of functions $f, g : [0, 1] \to [0, 1]$ such that $\max(\text{ran} f)$ and $\max(\text{ran} g)$ both exist but $\sup\{f(x) + g(x) \mid x \in [0, 1]\} < \max(\text{ran} f) + \max(\text{ran} g)$.

   REMARK: This exercise shows that adding functions is not the same thing as adding sets! (Compare this to Theorem 2.32.)

5. Prove the statement in Theorem 2.32 pertaining to infima.

6. Let $A, B \subseteq \mathbb{R}$ be arbitrary nonempty sets. Prove that if $\sup(A + B)$ exists, then $\sup A$ and $\sup B$ also exist. (Hint: for any particular $a \in A$, if $b \in B$ is allowed to vary, how large can $a + b$ get? For which set does this information give you an upper bound?)

7. Let $A, B \subseteq \mathbb{R}$ be nonempty, and assume that for all $a \in A$ and $b \in B$, $a < b$.
   (a) Find an example of sets A, B as described where $\sup A < \inf B$, and find an example of sets A, B as described where $\sup A = \inf B$.
   (b) Let $C = \{b - a \mid b \in B, a \in A\}$. Prove that C is bounded below and that $\inf C = (\inf B) - (\sup A)$. This measures the "size of the gap" between A and B.

8. Let $A, B \subseteq \mathbb{R}$ be nonempty, and suppose that for all $a \in A$, there exists $b \in B$ with $a \leq b$. (This is not the same hypothesis as in Theorem 2.33.)
   (a) Find an example of such sets A, B where A and B are both unbounded above.
   (b) Prove that if $\sup B$ exists, then so does $\sup A$, and $\sup A \leq \sup B$.

2.8 Completeness and the Integers

Consider basic counting[4], which lists the elements of $\mathbb{N}^*$ in increasing order, starting with 1, then 2, then 3, and so forth. Certainly, this process doesn't stop, as every $n \in \mathbb{N}^*$ always has a successor $n+1$ which is greater. However, how large do the listed numbers really get? In other words, is there some upper bound of $\mathbb{N}^*$? How do you know there isn't some real number $*$ like in Figure 2.3?

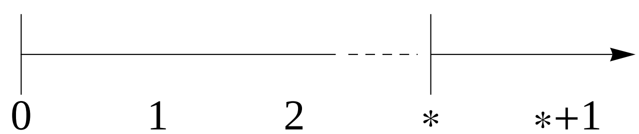

Figure 2.3: A number line with an upper bound $*$ of $\mathbb{N}^*$

It is tempting to say "there can't be an upper bound, because we can count forever". However, this reasoning is not correct. To see why, suppose that instead of listing the natural numbers, we instead list the elements of $S = \{1/2, 2/3, 3/4, \ldots\}$ in increasing order, so that the n^{th} number in our list is $n/(n+1)$ instead of n. Because S is an infinite[5] set, we can list its members forever. However, S does have an upper bound, namely 1. See also Exercise 2.9.2.

[4] In this book, we will assume familiarity with some basic principles of integers rather than proving them formally. For instance, we will take for granted that 1 is the smallest positive integer, and the sum or product of two integers is an integer. We will not cover a formal construction of $\mathbb{N}$ or $\mathbb{Z}$; much of this work is done in a junior-level real analysis course.

[5] Technical definitions of "finite" and "infinite" are often given in a beginning discrete mathematics course, but a casual understanding of these terms is sufficient for our purposes.

What, then, makes $\mathbb{N}^*$ different from the set S? One important difference is that any two successive members of $\mathbb{N}^*$ differ by exactly 1. In contrast, the successive members $n/(n+1)$ and $(n+1)/(n+2)$ differ by

$$\begin{aligned}\frac{n+1}{n+2} - \frac{n}{n+1} &= \frac{(n+1)^2 - n(n+2)}{(n+1)(n+2)} \\ &= \frac{n^2+2n+1-n^2-2n}{(n+1)(n+2)} = \frac{1}{(n+1)(n+2)}\end{aligned}$$

and upon plugging in larger and larger values of n, it seems this difference is becoming quite small. This says that, in essence, the natural numbers grow at a constant rate, whereas the members of S do not. However, it turns out even this property is not enough to guarantee that $\mathbb{N}^*$ is unbounded above. In formal logic courses, it is possible to create an ordered field which satisfies most of the properties of $\mathbb{R}$ (it satisfies all the axioms except for Axiom C), but in this field, the set $\{1, 1+1, 1+1+1, \ldots\}$ (where 1 is the number specified in AA4) IS bounded above!

It turns out that we can prove that $\mathbb{N}^*$ is unbounded above, but we need to use Axiom C in the proof.

Theorem 2.34. *$\mathbb{N}^*$ has no upper bound (in $\mathbb{R}$).*

Strategy. If there were an upper bound of $\mathbb{N}^*$, Axiom C would guarantee the existence of a least upper bound c. There exist natural numbers which are "pretty close" to c, i.e. there is some $n \in \mathbb{N}^*$ with $|c-n| < 1$. However, c has to be at least as big as both n and $n+1$, which is impossible.

Proof. Suppose for contradiction that $\mathbb{N}^*$ is bounded above. As $\mathbb{N}^*$ is clearly nonempty, Axiom C says that there exists a least upper bound $c = \sup \mathbb{N}^*$. Therefore, $c-1$, being smaller than c, is not an upper bound, so there is some $n \in \mathbb{N}^*$ with $c - 1 < n$.

Thus, $c < n+1$ by adding 1 to both sides. As $n+1 \in \mathbb{N}^*$, this contradicts the fact that c is an upper bound of $\mathbb{N}^*$. ∎

Corollary 2.35. *$\mathbb{Z}$ has no upper or lower bounds (in $\mathbb{R}$).*

Strategy. If b were a lower bound of $\mathbb{Z}$, it would also be a lower bound of $-\mathbb{N}^*$, but then $-b$ would be an upper bound of $\mathbb{N}^*$, which is impossible.

Proof. Because $\mathbb{N}^* \subseteq \mathbb{Z}$ and $\mathbb{N}^*$ has no upper bounds, neither does $\mathbb{Z}$. Next, suppose that there is some lower bound b of $\mathbb{Z}$ for a contradiction. For any $n \in \mathbb{N}^*$, because $-n \in \mathbb{Z}$, and b is a lower bound of $\mathbb{Z}$, we have $b \leq -n$, so therefore $-b \geq n$. This shows that $-b$ is an upper bound of $\mathbb{N}^*$, which is impossible, finishing the proof by contradiction. ∎

Corollary 2.36 (Archimedian Property). *For all $x > 0$ and all $y \in \mathbb{R}$, there exists an $n \in \mathbb{N}^*$ satisfying $nx > y$.*

Strategy. The idea is that this should be true for "large enough" values of n. How big should n be? We want to find n so that $nx > y$, so by dividing x to both sides, we get $n > y/x$.

Proof. Because $\mathbb{N}^*$ is unbounded above, y/x is not an upper bound of $\mathbb{N}^*$, so there is some $n \in \mathbb{N}^*$ with $n > y/x$. Because x is positive, multiplying x to both sides of the inequality yields $nx > y$. ∎

The Archimedian Property can be understood as follows: if you have a ruler which can measure lengths up to x (such as, for example, x is 12 inches, or x is 30 centimeters), and you wish to measure a total length of y, then you only need to place finitely many copies of your ruler, say n copies, end to end to measure the total length. Even though n could be quite large, it is still finite; there is no length y which is beyond your reach with a ruler. This prohibits "jumps to infinity" like those pictured in Figure 2.3 from earlier.

Thus, the completeness axiom tells us that the integers stretch out arbitrarily far to the left and right on a number line. As consecutive integers differ by 1, the integers are also evenly spaced along the line. Hence, it seems quite plausible that every real number of x lies between a unique pair of consecutive integers (where if x is already an integer, we make x the smaller of the two integers in the pair). This is surprisingly tricky to prove:

Theorem 2.37. *For all $x \in \mathbb{R}$, there is exactly one $n \in \mathbb{Z}$ with $n \leq x < n+1$. This value n is called the* greatest integer below x*, or the* floor *of x, and is written $\lfloor x \rfloor$.*

Strategy. Out of all the integers which are smaller than or equal to x, we want to find the "closest" integer to x. More precisely, we want the largest integer which does not exceed x. Let's say $S = \{k \in \mathbb{Z} \mid k \leq z\}$; thus, the floor of x, if it exists, should be $\max S$. Once we show that this max exists, it follows that the floor of x is uniquely determined, since sets with maximum elements have unique maximum elements.

We don't have any tools for showing that a set has a maximum element if we don't know what to guess for its maximum. Thus, let's consider a similar concept: $\sup S$. Thanks to Axiom C, we have a way to check whether $\sup S$ exists: we show that S is nonempty and bounded above. Neither of these details is very hard to check. Let's say $c = \sup S$.

At this point, though, we don't know whether c is an integer! To get around this, we know that there are members of S which are as close to c as we'd like. More precisely, there is an integer $n \in S$ such that $n > c - 1$. The next larger integer is $n + 1$, and $n + 1 > c$, so since c is an upper bound of S, it follows that $n + 1 \notin S$. This shows that no integer larger than n is in S, so $n = \max S$. (In fact, $n = c$, but we don't need this information.)

Proof. Let $x \in \mathbb{R}$ be given. Define $S = \{k \in \mathbb{Z} \mid k \leq x\}$. We claim that S has a (necessarily unique) maximum element. Once we have proven this claim, it follows that if $n = \max S$, then n is the floor of x. This is because $n + 1$ is an integer which is not in S (because it is larger than the max n), so by definition of S, we must have $n + 1 > x$ and hence $n \leq x < n + 1$.

First, we show that $\sup S$ exists. The set S is certainly bounded above by x. If S were empty, then for all $k \in \mathbb{Z}$ we'd have $k > x$, so x would be a lower bound of $\mathbb{Z}$, contradicting Corollary 2.35. Therefore, S is nonempty and bounded above, so Axiom C guarantees that $\sup S$ exists.

By the definition of supremum, $\sup S - 1$ is not an upper bound of S, so we may choose some $n \in S$ with $n > \sup S - 1$. It follows that $n + 1 > \sup S$. In fact, for any integer k satisfying $k > n$, we have $k \geq n + 1 > \sup S$, and it must follow that $k \notin S$ (because $\sup S$, as an upper bound of S, cannot be smaller than members of S). Thus, n is the largest member of S, so $n = \max S = \sup S$. ■

As a quick consequence of the previous theorem, we can say that any two numbers that are greater than 1 apart have an integer fitting in between them:

Corollary 2.38. *For all $x, y \in \mathbb{R}$, if $y - x > 1$, then there exists $n \in \mathbb{Z}$ with $x < n < y$.*

Strategy. Intuitively, we want to pick an integer greater than x which is as close to x as possible. Because $\lfloor x \rfloor$ is the closest integer to x which is smaller than x, $\lfloor x \rfloor + 1$ seems like a good choice for n.

Proof. Let $x, y \in \mathbb{R}$ be given with $y - x > 1$. Define $n = \lfloor x \rfloor + 1$, so $n \in \mathbb{Z}$ and $n - 1 \leq x < n$. Therefore, $n \leq x + 1 < y$, so $x < n < y$. ∎

The exercises have some other good properties of $\mathbb{Z}$, but there are also properties of the rationals $\mathbb{Q}$ and of the irrationals $\mathbb{R} - \mathbb{Q}$ which follow from the properties of $\mathbb{Z}$ in this section. Some of these properties will be used in later chapters.

2.9 Exercises

1. Prove that for all $x > 0$, there is some $n \in \mathbb{N}^*$ with $1/n < x$. (This is not the same as Exercise 2.4.8.)

2. If $S = \{1/2, 2/3, 3/4, \ldots\}$, prove that $\sup S = 1$. (Hint: Exercise 2.9.1 might be useful.)

3. Prove that for all $x \in \mathbb{R}$, there exists a unique $n \in \mathbb{Z}$ with $n-1 < x \leq n$. This n is called the *least integer above* x, or the *ceiling* of x, written $\lceil x \rceil$. (Hint: Note that $\lceil x \rceil = \min\{k \in \mathbb{Z} \mid k \geq x\}$. You can also use the floor function to solve this problem quickly!)

4. Prove that for all $x, y \in \mathbb{R}$ with $x < y$, there exists some $q \in \mathbb{Q}$ with $x < q < y$. We describe this by saying that the rationals are *dense* in $\mathbb{R}$. (Hint: Let's say $q = a/b$ where $b > 0$. In order to have $x < a/b < y$, we must have $bx < a < by$, so bx and by should have an integer between them. How large should b be in order to apply Corollary 2.38?)

5. The proof of Theorem 2.29 (which we did not cover) shows that $\sqrt{2} = \sup\{x \in \mathbb{R}^+ \mid x^2 < 2\}$. Using this fact and Exercise 2.9.4, prove that $\sqrt{2} = \sup\{x \in \mathbb{Q}^+ \mid x^2 < 2\}$.

6. Prove that for all $x, y \in \mathbb{R}$ with $x < y$, there exists $r \in \mathbb{R} - \mathbb{Q}$ with $x < r < y$. This shows that the irrationals are dense in $\mathbb{R}$. (Hint: This is similar to Exercise 2.9.4. However, instead of finding an element in between x and y of the form a/b, find an element of the form $a/(b\sqrt{2})$, where $a, b \in \mathbb{Z}$, $b > 0$, and $a \neq 0$.)

7. For all $a, b \in \mathbb{N}$ with $b > 0$, prove there exists exactly one pair $(q, r) \in \mathbb{N} \times \mathbb{N}$ with the properties that $a = qb + r$ and $0 \leq r < b$. The

process of obtaining q and r is called *Euclidean division*, where q is the *quotient* and r is the *remainder*. (Hint: You want to pick q so that $qb \le a < (q+1)b$.)

2.10 Using the Axioms in Real Mathematics

This chapter presented the important axioms of $\mathbb{R}$ for two main reasons. The first reason is to illustrate the process of building a series of results starting only from simple assumptions (the axioms) and some definitions, where the axioms summarize the most essential features of $\mathbb{R}$. The other reason is to show some of the important ways in which $\mathbb{R}$ differs from other sets of numbers like $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$, showing that $\mathbb{R}$ is a suitable set of numbers to use in the study of calculus.

However, most mathematicians do not bother with axioms of $\mathbb{R}$ very frequently. A proof from axioms is often long and mundane, but straightforward with enough practice. Furthermore, there are many statements like "$3 > 1$" that technically should receive formal proofs (such as "Because $1 > 0$ by Theorem 2.22, $3 = 2 + 1 > 2$, and $2 = 1 + 1 > 1$, proving $3 > 1$"), but including proofs for all of these "trivial" statements would clutter the book with technical details that distract you from the main points. Lastly, many of the properties of the real numbers that follow from the algebraic and order axioms are presumably quite familiar to you from years of mathematics education prior to now.

Thus, from this point on, we will be more casual when using properties of the real numbers which follow from the algebraic and order axioms. Since the completeness axiom is probably more unfamiliar, and it yields some very powerful consequences for calculus, we will still treat the completeness axiom formally. To be less verbose, though, in future proofs we will probably not cite Axiom C or Theorem 2.28 by name when saying that a supremum or infimum exists; we will merely check that the hypotheses of Axiom C or Theorem 2.28 apply.

Chapter 3

Limits and Continuity

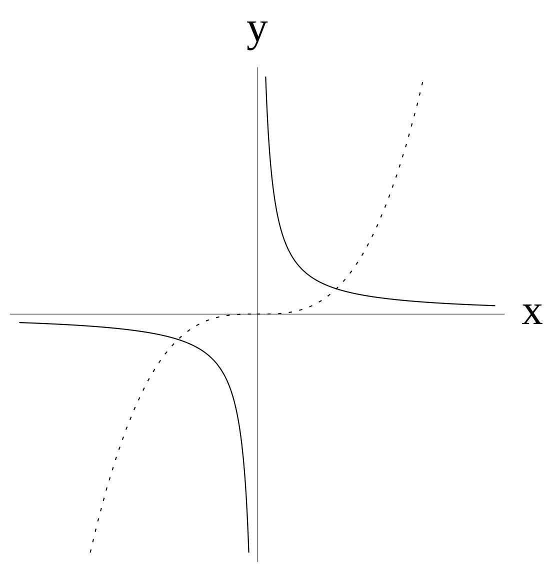

Figure 3.1: Graphs of $f(x) = x^3$ (dotted) and $g(x) = 1/x$ (solid)

In your previous mathematics education, many of the graphs you drew of functions from $\mathbb{R}$ to $\mathbb{R}$ were probably connected curves. Roughly, this means you could draw the entire graph without lifting your pencil (or pen, or other writing implement) off the paper (or touchscreen, or other drawing surface). There were sometimes certain examples, like the graph of $g(x) = 1/x$, where you had to lift your pencil off the paper at $x = 0$, but that was only because the function's graph went off to arbitrarily large values. Basically, wherever your function was actually defined, you could draw a connected curve. See Figure 3.1.

Let's consider the following example. Suppose that you are driving a car. If we measure time t in seconds after the start of your trip (so that $t = 0$ corresponds to when you start the car, $t = 60$ is a minute afterward, $t = 0.5$ is half a second after starting, and so on), we'll let $v(t)$ be the speed of the car after t seconds. (To be concrete, we'll measure $v(t)$ in miles per hour, or mph, but it is certainly possible to measure $v(t)$ in other units.) After 20 minutes (i.e. $t = 60 \cdot 20 = 1200$), you stop driving. Thus, v is a function from $[0, 1200]$ to $[0, \infty)$.

Note that your speed can't just "jump" from one value to another. Certainly there are times when instantaneous change of speed can be useful, like if you want to stop your car immediately to avoid an accident, but this is not possible: acceleration has to occur, and the car builds up or loses speed over a period of time. We describe this by saying that speed changes continuously with time, i.e. v is a continuous function. Now, the motion of your car doesn't have to be smooth: depending on how erratically you hit the gas pedal or the brakes, the car's acceleration can be quite jumpy. This means the acceleration doesn't have to be continuous; we'll explore this more when we talk about rates of change and derivatives in the next chapter.

The speed function v has some other useful properties. In only 20 minutes, you can't expect to make the car go infinitely fast (you'd need an infinite amount of time to do that), so your speed will be bounded over the course of the trip. In fact, there will be some point t^* on the trip where your car is going faster than at any other point. In other words, your car attains some maximum speed over the entire trip (which may not be the maximum speed that the car is physically capable of driving).

If we draw a graph of v as t goes from 0 to t^*, we'll draw a connected[1] curve (since the speed can't "jump" all of a sudden). Along that curve, we must hit every possible speed from $v(0)$ up to $v(t^*)$ somewhere along the way. For instance, if $v(t^*)$ is 60mph, and $v(0) = 0$ (corresponding to starting the car from a complete stop), then at some time between 0 and t^*, the car must have been going 30mph, because 30mph is between 0mph and 60mph.

We claim those properties of v follow just from the fact that v is continuous on $[0, 1200]$. However, in order to make such claims and prove them, we'll

[1]The word "connected" has a formal meaning, but we only intend to use the word casually.

need careful definitions of what exactly it means to say "v can be graphed with a connected curve". In general, the main idea will be roughly as follows: for any function f with codomain $\mathbb{R}$ and any $a \in \text{dom}(f)$, we say f is continuous at a to mean that for all $x \in \text{dom}(f)$, we can guarantee that $f(x)$ is as close as we like to $f(a)$ by taking x close enough to a.

Thus, intuitively, we see that even if we try to make the function g from Figure 3.1 defined at 0 by picking a value for $g(0)$, we can't get continuity at $a = 0$ because of the "jump" in the values of $g(x)$ from very negative values to very positive values that happens when x is close to 0. It seems that the values of $g(x)$ don't seem to approach any one particular value when x is near 0. We describe this by saying that $g(x)$ has no limit as x approaches 0.

However, along with the examples of "nice" functions in Figure 3.1, there are other, more complicated, examples which we will study. In order to properly analyze those functions, as well as to place our study on a firm foundation, we will need more formal definitions of a limit and of continuity, from which we'll be able to prove the properties we mentioned previously. Establishing formal definitions of limits and continuity, along with studying important consequences of the definitions, is the purpose of this chapter.

3.1 The Definition of a Limit

In this section, we'll aim to make a precise way to say things like "As x approaches 3, x^2 approaches 3^2." As this involves being able to say how far numbers are from one another, we should also come up with a way to measure distance between numbers.

Distance and the Absolute-Value Function

A useful tool for us will be the *absolute-value* function, whose value on any $x \in \mathbb{R}$ is written as $|x|$ and defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

(This function is well-defined because the Trichotomy Law says that exactly one of $x > 0$, $x = 0$, or $x < 0$ is true.) $|x|$ is called "the *absolute value of* x", "*absolute* x", or "the *magnitude* of x", among other phrases. Note that for all $x \in \mathbb{R}$, $|x| \geq 0$, and $|x| = 0$ iff $x = 0$.

How does this function relate to distance? If a and b are two real numbers, then the distance from a to b on a number line is the larger number minus the smaller one. In other words, the distance from a to b is $b - a$ if $b \geq a$ (i.e. $b - a \geq 0$), or is $a - b = -(b - a)$ if $b < a$ (i.e. $b - a < 0$). This means that the distance is the same as $|b - a|$. In particular, when $a = 0$, $|b|$ is the distance of b to the origin 0.

Unfortunately, the absolute-value function is difficult to use with algebraic laws, because for real a and b, $|a + b|$ might NOT be equal to $|a| + |b|$: $|-1 + 1| = 0$ but $|-1| + |1| = 2$. However, the following are true for every $a, b \in \mathbb{R}$:

$$|ab| = |a||b| \qquad \left|\frac{a}{b}\right| = \frac{|a|}{|b|} \text{ if } b \neq 0$$

We do frequently use absolute values in inequalities, however. For instance, let's consider the set of all points that are at a distance at most r from some central point c, i.e. all $x \in \mathbb{R}$ such that $|x - c| \leq r$. Looking at the number line in Figure 3.2, it seems that the smallest x can be is r less than c (i.e. $c - r$), and the largest x can be is r more than c (i.e. $c + r$). The following lemmas prove this more formally:

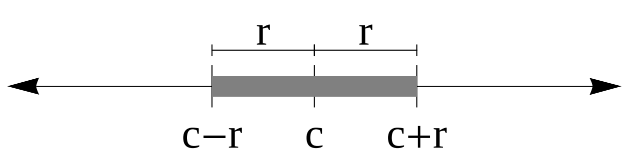

Figure 3.2: The points x with distance at most r from c

Lemma 3.1. *For all $x \in \mathbb{R}$, $-|x| \leq x \leq |x|$.*

Lemma 3.1 is easy to show by considering two cases, so we leave its proof to the reader. We use this result to establish the following:

Lemma 3.2. *For all $c, r, x \in \mathbb{R}$, $|x - c| \leq r$ iff $c - r \leq x \leq c + r$, thus $\{x \in \mathbb{R} \mid |x - c| \leq r\}$ is the set $[c - r, c + r]$. (This set is empty if $r < 0$.) In particular, when $c = 0$, $|x| \leq r$ iff $-r \leq x \leq r$.*

Strategy. Let's first concentrate on proving the special case when $c = 0$. If we can prove that case, then by applying it to the number $x - c$, we find that $|x - c| \leq r$ iff $-r \leq x - c \leq r$, which is equivalent to $c - r \leq x \leq c + r$

by adding c to all sides of the inequality. (We sometimes say describe this situation, where after proving one case, the rest of the cases follow easily, by saying that the lemma can be *reduced* to the case where $c = 0$.)

For one direction, if $-r \leq x \leq r$, then we consider cases based on whether $x \geq 0$ or $x < 0$. If $x \geq 0$, then $|x| = x$, so $x \leq r$ implies $|x| \leq r$. If $x < 0$, then $|x| = -x$, and $x \geq -r$ implies $-x \leq r$.

The other direction can be done in a similar manner, or we can use Lemma 3.1 to simplify our work: $-|x| \leq x \leq |x|$. Thus, if $|x| \leq r$, then $-|x| \geq -r$, and we can put all these inequalities together.

Proof. Let $c, r \in \mathbb{R}$ be given. It suffices to prove that for all $x \in \mathbb{R}$, $|x| \leq r$ is equivalent to $-r \leq x \leq r$, because then plugging in $x - c$ in place of x yields $|x - c| \leq r$ iff $-r \leq x - c \leq r$ iff $c - r \leq x \leq c + r$. Thus, we may assume $c = 0$ for the remainder of the proof.

For one direction of proof, assume $|x| \leq r$. By Lemma 3.1, we have $-|x| \leq x \leq |x|$. Thus,

$$-r \leq -|x| \leq x \leq |x| \leq r$$

or $-r \leq x \leq r$ in particular.

For the other direction, assume $-r \leq x \leq r$. If $x \geq 0$, then $x = |x|$ and $x \leq r$, so $|x| \leq r$. If $x < 0$, then $|x| = -x$ and $x \geq -r$, so $-x \leq r$ and $|x| \leq r$. Thus, $|x| \leq r$ in either case. ∎

Remark. A very similar proof can be used to prove a strict inequality version of the lemma: for all $c, r, x \in \mathbb{R}$, $|x - c| < r$ iff $c - r < x < c + r$, i.e. iff $x \in (c - r, c + r)$.

The following consequence of Lemma 3.2 is a very important tool in handling inequalities with absolute values:

Theorem 3.3 (Triangle Inequality). *For all $x, y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$.*

Before we prove this theorem, let's see why it is named the Triangle Inequality. When mathematicians work with vectors (which are essentially arrows from the origin with a length and a direction), they frequently use the notation $|x|$ to represent the length of a vector x. Our use of $|x|$ for real values of x is a special case of this definition, if you think of the real number x as a vector pointing from 0 to x on the number line. To add two vectors

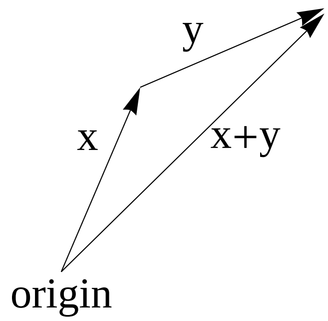

Figure 3.3: Adding two vectors in $\mathbb{R}^2$

x and y, you draw x starting from the origin, then you draw y starting from x's ending, and then $x + y$ is the arrow from the origin to y's final point. Figure 3.3 illustrates this in two dimensions.

The vectors x, y, and $x + y$ form a triangle, so the statement "$|x + y| \leq |x| + |y|$" says that the third side of the triangle cannot be longer than the sum of the other two side lengths. For another way to interpret this, when you walk from the origin to the final point on the vector $x + y$, it takes less time to walk straight along the vector $x + y$ than it takes to first walk along x and then walk along y. In other words, the straight-line path between two points is the shortest path.

Strategy. One way to prove this is to break the proof into four cases based on the signs of x and y, but that gets messy. It is simpler to use Lemma 3.2 and try to prove

$$-(|x| + |y|) \leq x + y \leq |x| + |y|$$

instead (where $r = |x| + |y|$ and $c = 0$ in the lemma). This is easy to show with two applications of Lemma 3.1.

Proof. Let $x, y \in \mathbb{R}$ be given. By Lemma 3.1, we know $-|x| \leq x \leq |x|$ and $-|y| \leq y \leq |y|$. Adding these inequalities yields

$$-(|x| + |y|) \leq x + y \leq |x| + |y|$$

so $|x + y| \leq |x| + |y|$ by Lemma 3.2. ■

The ϵ-δ Definition

Now that we have a way of conveniently expressing the distance between two real numbers, let's investigate a potential definition for limits. Namely, if f is a function with $\text{dom}(f), \text{ran}(f) \subseteq \mathbb{R}$ (for convenience, we'll say that such functions are *real functions*) and $L, a \in \mathbb{R}$, what should it mean to say that $f(x)$ approaches L as x approaches a?

First, we want $f(x)$ to be "as close as desired" to L, so let's specify how close we want $f(x)$ to be to L. Let's say we want the distance $|f(x)-L|$ to be less than some positive number, which is traditionally denoted by the Greek letter ϵ (pronounced "epsilon"). Hence, we want to make sure $|f(x) - L| < \epsilon$ when x is near enough to a, but what does "near enough" mean? Once again, we formalize this by using a positive number to denote how far apart x and a should be; traditionally, the Greek letter δ (pronounced "delta") is used for this purpose[2]. Thus, we want to pick $\delta > 0$ so that whenever $x \in \text{dom}(f)$ and $|x - a| < \delta$, it follows $|f(x) - L| < \epsilon$.

Putting this all together, we arrive at this potential definition: we want the property that for every $\epsilon > 0$ (representing a desired distance from the limit L), there exists some corresponding $\delta > 0$ (representing how close our input x should be to a) such that for every $x \in \text{dom}(f)$, if $|x - a| < \delta$, then $|f(x) - L| < \epsilon$.

Here's a casual example to illustrate what is going on here. Let's say you are holding a laser pointer, so that you can shine a spot on a screen far away from you. To control where the spot lands, you angle your hand: when your hand is at an angle of x radians away from being completely flat, the spot lands at height $f(x)$ on the screen, where f is some function of your hand's angle. Ideally, you'd like to have the pointer shine exactly at height L, which you could do if you could hold the laser pointer perfectly steady in your hand at an angle of a.

However, you're not that steady. Thus, the pointer wobbles a little, and this causes the spot to move up and down. Let's say your friend challenges you to keep that shined spot "pretty close" to L, say they give you a distance $\epsilon > 0$ and want the shined spot to be between $L - \epsilon$ and $L + \epsilon$. (In essence, your friend wants you to hit a "bullseye" centered at L with radius ϵ.) By holding yourself pretty steady, i.e. you never let your hand's angle deviate more than some distance $\delta > 0$ from a, your light shines within distance ϵ

[2]The use of ϵ and δ is due to the French mathematician Augustin-Louis Cauchy. The letter ϵ stands for "erreur", where we think of $f(x)$ as having "error less than ϵ" in measuring L. The letter δ stands for "différence", representing the difference between x and a.

of L. Your friend keeps challenging you with different values of ϵ, trying to make you more and more accurate, and each time, you find a corresponding control value of δ that dictates how steady your hand must stay. After many challenges, your friend acknowledges that while your aim is never perfect, at least it's clear you're able to make your spot approach the ideal height of L.

This is the idea of our attempted definition: when challenged with an accuracy of ϵ we must get for a function's outputs, we find out how much control δ needs to be placed on the inputs. The definition we have outlined above is almost the definition we want, but it has two issues.

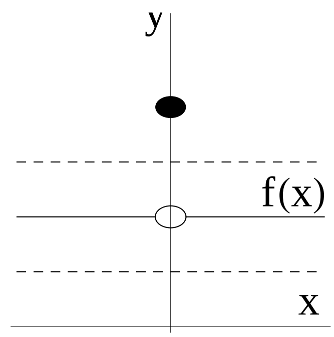

Figure 3.4: A function for which the limit as $x \to 0$ shouldn't equal $f(0)$

First, if we consider the function f in Figure 3.4, then the limit as x approaches 0 should be at the hollow circle. However, the actual value of $f(0)$ is at the solid circle, and that solid circle is not within distance ϵ of the hollow circle when ϵ is small enough (the region between the dashed lines represents the region from $L-\epsilon$ to $L+\epsilon$ on the y-axis). The issue is that the limit ought to depend on the values of $f(x)$ where x is CLOSE to 0 but NOT equal to 0. To fix this, our definition will say "if $0 < |x-a| < \delta$" instead of "if $|x-a| < \delta$". In other words, we will consider all points x near a except for a itself.

Second, if dom(f) doesn't have many points near a, then the statement "for all $x \in \text{dom}(f)$, if $|x-a| < \delta$ then $|f(x)-L| < \epsilon$" could be vacuously true for small choices of δ. To avoid this issue, we'll require that dom(f) contain a *deleted open interval around* a, which is a set of the form $(a_1, b_1) - \{a\}$ where $a_1 < a < b_1$. More simply, we'll say f needs to be defined *near* a. There are other kinds of requirements we could place on dom(f), but this requirement is simple to state and also seems quite plausible (as it requires that when x is close enough to a, $f(x)$ actually exists).

This leads us to the following definitions, including our official limit definition:

Definition 3.4. Let $a \in \mathbb{R}$ be given. A *deleted open interval around* a is a set of the form $(a_1, b_1) - \{a\}$, where $a_1, b_1 \in \mathbb{R}$ and $a_1 < a < b_1$. (The terminology is appropriate because we start with an interval containing a and then "delete" a from it.)

Let f be a real function. We say that f is defined *near* a if $\mathrm{dom}(f)$ contains a deleted open interval around a.

Definition 3.5. Let $a, L \in \mathbb{R}$ be given, and suppose f is a real function defined near a. We say that $f(x)$ *has limit* L *as* x *approaches* a, written in any of the following three ways

$$\lim_{x \to a} f(x) = L \qquad f(x) \to L \text{ as } x \to a \qquad f(x) \underset{x \to a}{\longrightarrow} L$$

if the following holds:

$$\forall \epsilon > 0 \; \exists \delta > 0 \; \forall x \in \mathrm{dom}(f) \; (0 < |x - a| < \delta \to |f(x) - L| < \epsilon)$$

If $f(x)$ does not approach L as x approaches a, we sometimes write "$f(x) \not\to L$ as $x \to a$".

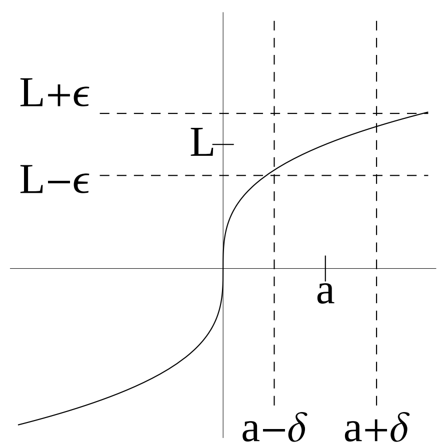

Figure 3.5: A picture of the limit definition

Another picture illustrating this definition is in Figure 3.5. The horizontal dashed lines above and below the limit L represent the points where $|f(x) -$

$L| < \epsilon$. The vertical dashed lines left and right of a represent the region where $|x - a| < \delta$. We note that the entire portion of the curve where $|x-a| < \delta$ lands within the two horizontal lines, i.e. for all these x, we have $|f(x) - L| < \epsilon$.

Remark. When you have a function f and a real number a, the definition of limit does NOT tell you what the value of L should be. In other words, the definition gives you a way to check if you guessed the correct limit, but it doesn't say how to obtain that guess in the first place. Sometimes we describe this by saying the definition is not constructive.

However, it is worth noting that a function can have at most one limit at any point. In other words, as $x \to a$, $f(x)$ cannot approach two distinct limits at the same time. See Exercise 3.2.9.

Example 3.6:

Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a constant function, i.e. there is some $c \in \mathbb{R}$ so that for all $x \in \mathbb{R}$, $f(x) = c$. For any $a \in \mathbb{R}$, it makes sense to guess that $f(x) \to c$ as $x \to a$. To prove this guess, let's use the definition of limit.

Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for any $x \in \mathbb{R}$, $0 < |x-a| < \delta$ implies $|f(x)-c| < \epsilon$. However, since $|f(x)-c| = |c-c| = 0$, ANY value of x works, so δ can be any positive number. (In other words, for any accuracy $\epsilon > 0$ we want, there is no need to control our inputs because the function values never change, so $\delta > 0$ can be anything.) $\square$

Example 3.7:

Suppose that $f : \mathbb{R} \to \mathbb{R}$ is the identity function, so for all $x \in \mathbb{R}$, $f(x) = x$. For any $a \in \mathbb{R}$, we prove that $f(x) \to a$ as $x \to a$.

Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for any $x \in \mathbb{R}$, $0 < |x - a| < \delta$ implies $|f(x) - a| < \epsilon$. However, since $|f(x) - a| = |x - a|$, it suffices to choose $\delta = \epsilon$. $\square$

In the last examples, we found that for any $a \in \mathbb{R}$, the limit of $f(x)$ as x approaches a is in fact $f(a)$. This is a useful property a function can have:

Definition 3.8. If f is a real function and $a \in \mathbb{R}$, we say that f is *continuous* at a if $a \in \text{dom}(f)$, $\lim\limits_{x\to a} f(x)$ exists, and $\lim\limits_{x\to a} f(x) = f(a)$. We also say that a is a *continuity point* of f. For any $S \subseteq \mathbb{R}$, if f is continuous at every $a \in S$, we say f is *continuous on S*.

Equivalently, for $a \in \mathbb{R}$, f is continuous at a if f is defined in an open interval containing a and the following holds:

$$\forall \epsilon > 0 \; \exists \delta > 0 \; \forall x \in \mathrm{dom}(f) \; (|x - a| < \delta \to |f(x) - f(a)| < \epsilon)$$

If f is defined near a but is not continuous at a, then we say f is *discontinuous at* a.[3]

Remark. In the definition of continuity, we may allow x to equal a (i.e. we don't have to write the $0 <$ in the definition of limit) since when $x = a$, we have $|f(x) - f(a)| = |f(a) - f(a)| = 0 < \epsilon$.

One-Sided Limits

In the definition of limit as x approaches a, we consider all values of x which are close enough to a, including both the values below a and the values above a. Thus, we sometimes say that the limit we defined is a *two-sided* limit. There are analogous definitions where we only consider values for x that are greater than a or only values that are smaller than a. These are called *one-sided limits.*

Definition 3.9. Let $a \in \mathbb{R}$ and a real function f be given. We say that f is defined *near* a *on the left* if $\mathrm{dom}(f)$ contains a set of the form (a_1, a) where $a_1 \in \mathbb{R}$ and $a_1 < a$. Similarly, we say that f is defined *near* a *on the right* if $\mathrm{dom}(f)$ contains a set of the form (a, b_1) where $b_1 \in \mathbb{R}$ and $a < b_1$.

Definition 3.10. Let $a, L \in \mathbb{R}$ be given, and let f be a real function defined near a on the left. We say $f(x)$ approaches L as x approaches a *from the left*, also written in the following ways

$$\lim_{x \to a^-} f(x) = L \qquad f(x) \to L \text{ as } x \to a^- \qquad f(x) \underset{x \to a^-}{\longrightarrow} L$$

if the following holds:

$$\forall \epsilon > 0 \; \exists \delta > 0 \; \forall x \in \mathrm{dom}(f) \; (0 < a - x < \delta \to |f(x) - L| < \epsilon)$$

If $f(a)$ is defined and $\lim_{x \to a^-} f(x)$ exists and equals $f(a)$, then we say f is *continuous from the left* at a.

[3]Some authors do not require f to be defined near a in order to have a discontinuity, and some place other requirements on f.

Similarly, if f is defined near a on the right, then we say $f(x)$ approaches L as x approaches a *from the right*, written in one of the following ways

$$\lim_{x\to a^+} f(x) = L \qquad f(x) \to L \text{ as } x \to a^+ \qquad f(x) \underset{x\to a^+}{\longrightarrow} L$$

if the following holds:

$$\forall \epsilon > 0\ \exists \delta > 0\ \forall x \in \mathrm{dom}(f)\ (0 < x - a < \delta \to |f(x) - L| < \epsilon)$$

If $f(a)$ is defined and $\lim_{x\to a^+} f(x)$ exists and equals $f(a)$, we say f is *continuous from the right* at a.

It turns out that a two-sided limit exists and equals L iff the corresponding one-sided limits also exist and equal L: see Exercise 3.2.7. However, the one-sided limits can sometimes provide us more information about how a function behaves. To illustrate this, we present a couple examples.

Example 3.11:

In Theorem 2.37, we established the existence of the floor function $f : \mathbb{R} \to \mathbb{R}$, where for any $x \in \mathbb{R}$, $f(x) = \lfloor x \rfloor$ is the unique integer satisfying $\lfloor x \rfloor \le x < \lfloor x \rfloor + 1$. In the usual drawing of the floor function in Figure 3.6, the function jumps up 1 unit at each integer, but otherwise the function looks smooth. Thus, we expect that the floor function is continuous at all points which are not integers. The next theorem makes this more precise:

Theorem 3.12. *For every $a \in \mathbb{R}$,*

$$\lim_{x\to a^+} \lfloor x \rfloor = \lfloor a \rfloor \qquad \textit{and} \qquad \lim_{x\to a^-} \lfloor x \rfloor = \begin{cases} \lfloor a \rfloor & \textit{if } a \notin \mathbb{Z} \\ a - 1 & \textit{if } a \in \mathbb{Z} \end{cases}$$

Thus, the floor function is continuous from the right everywhere but is not continuous from the left at integers.

Strategy. Let's say f is the floor function, and we'll consider the left and right-hand limits separately. First, from the right side, let's consider the interval $(a, \lfloor a \rfloor + 1)$. (This corresponds to choosing $\delta = \lfloor a \rfloor + 1 - a$, which is guaranteed to be positive.) On this interval, the function f is constant with value $\lfloor a \rfloor$. As constants are continuous, the right-hand limit should be $\lfloor a \rfloor$.

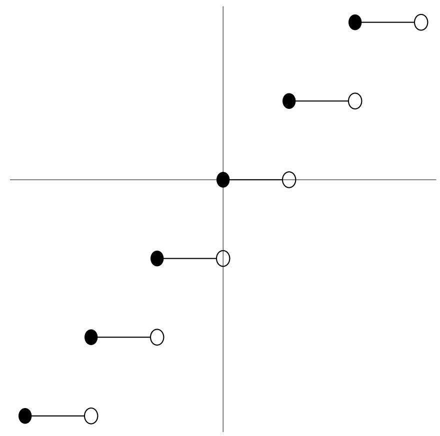

Figure 3.6: The function $\lfloor x \rfloor$ (solid dots are the function value, and hollow dots are "jumps")

However, for the left side, we need to consider whether $a \in \mathbb{Z}$ or not, since this determines whether $a - \lfloor a \rfloor$ is zero or not. If $a \notin \mathbb{Z}$, then on the interval $(\lfloor a \rfloor, a)$ (corresponding to $\delta = a - \lfloor a \rfloor$, which is positive), the function f takes the constant value of $\lfloor a \rfloor$. We already saw constant functions are continuous, so therefore the limit from the left should be $\lfloor a \rfloor$. On the other hand, if $a \in \mathbb{Z}$, then the interval $(\lfloor a \rfloor, a)$ is empty (it's the same as (a, a)), so we need to consider the interval $(a-1, a)$, on which the function f takes the constant value $a - 1$. This tells us the limit from the left should be $a - 1$.

Proof. Let $a \in \mathbb{R}$ be given. First, we prove that the floor function is continuous from the right at a. Let $\epsilon > 0$ be given. We choose $\delta = \lfloor a \rfloor + 1 - a$; note that $\delta > 0$ because $a < \lfloor a \rfloor + 1$. Now, suppose that $x \in \mathbb{R}$ is given with $0 < x - a < \delta$, i.e. $x \in (a, a + \delta)$. Therefore,

$$\lfloor a \rfloor \leq a < x < a + \delta = \lfloor a \rfloor + 1$$

In particular, $\lfloor a \rfloor \leq x < \lfloor a \rfloor + 1$, so the definition of the floor function says $\lfloor x \rfloor = \lfloor a \rfloor$ (since $\lfloor a \rfloor$ is an integer). Hence, $|\lfloor x \rfloor - \lfloor a \rfloor| = 0 < \epsilon$, proving the right-hand limit is $\lfloor a \rfloor$.

Next, let's consider the left-hand limit assuming that $a \notin \mathbb{Z}$. Let $\epsilon > 0$ be given. We choose $\delta = a - \lfloor a \rfloor$; because $a \notin \mathbb{Z}$, $a > \lfloor a \rfloor$, so δ is positive. Now, suppose that $x \in \mathbb{R}$ is given with $0 < a - x < \delta$, i.e. $x \in (a - \delta, a)$. Therefore,

$$\lfloor a \rfloor = a - \delta < x < a < \lfloor a \rfloor + 1$$

In particular, $\lfloor a \rfloor \leq x < \lfloor a \rfloor + 1$, so by definition, $\lfloor x \rfloor = \lfloor a \rfloor$. Hence, $|\lfloor x \rfloor - \lfloor a \rfloor| = 0 < \epsilon$, showing the left-hand limit is $\lfloor a \rfloor$.

Lastly, suppose $a \in \mathbb{Z}$. Let $\epsilon > 0$ be given, and we choose $\delta = 1$. Now, suppose $x \in \mathbb{R}$ is given with $0 < a - x < 1$, so $x \in (a-1, a)$. Because $a - 1 \in \mathbb{Z}$, and $a - 1 \leq x < (a-1) + 1$, we have $\lfloor x \rfloor = a - 1$. Hence $|\lfloor x \rfloor - (a-1)| = 0 < \epsilon$, proving that the left-hand limit is $a - 1$. ■ □

Example 3.13:

Let's consider a function f where some one-sided limit fails to exist. Let $f : (\mathbb{R} - \{0\}) \to \mathbb{R}$ be defined by $f(x) = 1/x$ for all $x \in \text{dom}(f)$. We claim that f has no limit as $x \to 0^+$. In other words, for all $L \in \mathbb{R}$, we'll show $f(x) \not\to L$ as $x \to 0^+$. This means we will prove

Theorem 3.14. *For all $L \in \mathbb{R}$,*

$$\exists \epsilon > 0 \ \forall \delta > 0 \ \exists x \in (0, \delta) \left| \frac{1}{x} - L \right| \geq \epsilon$$

which is precisely the negation of the statement "$1/x \to L$ as $x \to 0^+$".

Remark. Another way to think about this is that there is some accuracy ϵ which we fail to achieve: no matter which $\delta > 0$ we use to control our inputs, there is always some point x which is within distance δ to the right of 0 but for which $f(x)$ has distance at least ϵ from L.

Strategy. Which ϵ works? Intuitively, $1/x$ becomes arbitrarily large as x gets close to 0 from the right side, so any choice of ϵ should work. (In other words, the graph of $1/x$ eventually becomes arbitrarily far away from EVERY fixed value L.) To be definite, let's choose $\epsilon = 1$.

Now, let $\delta > 0$ be given. We wish to find a real number $x \in \text{dom}(f)$ so that $0 < x < \delta$ and $|1/x - L| \geq \epsilon = 1$. Using the remark following Lemma 3.2, this means we want some $x \in (0, \delta)$ satisfying either $1/x \geq L + 1$ or $1/x \leq L - 1$. Intuitively, since the values of $1/x$ are becoming very large, we expect that $1/x$ will eventually be greater than L, so we'd like to satisfy $1/x \geq L + 1$.

However, we can't take reciprocals of both sides of the inequality unless we know the sign of $L + 1$. This leads us to consider two cases. In the first case, if $L + 1 \leq 0$, then clearly any positive x satisfies $1/x \geq 0$. In the second case, if $L+1>0$, then we want x to be less than $1/(L+1)$. In each case, x can also be chosen to be smaller than δ. (After all, we need to show that we can find a suitable value of x no matter which control δ we try to place on our inputs.)

Proof. Let $L \in \mathbb{R}$ be given, and we choose $\epsilon = 1$. Let $\delta > 0$ be given. It suffices to find some $x \in (0, \delta)$ such that $1/x \geq L+\epsilon = L+1$. This is because for such an x, the statement

$$-1 < \frac{1}{x} - L < 1$$

is false, and thus by the remark after Lemma 3.2, "$|1/x - L| < 1$" is false, so "$|1/x - L| \geq 1$" is true.

We consider two cases. In the first case, $L+1 \leq 0$. We choose $x = \delta/2$. Note that $x \in (0, \delta)$, and since x is positive, $1/x > 0 \geq L+1$.

In the second case, $L+1 > 0$. Thus, for any $x \in (0,\delta)$, $1/x \geq L+1$ is equivalent to $x \leq 1/(L+1)$ by taking reciprocals of both sides. (This is possible because we know both sides are positive.) We choose

$$x = \frac{1}{2}\min\left\{\frac{1}{L+1}, \delta\right\}$$

(Essentially, we have two constraints we want x to satisfy: $x < 1/(L+1)$ and $x < \delta$. By using a min, we take the "stricter" of the two constraints.) Thus, $x < \delta$ and $x < 1/(L+1)$, so $x \in (0,\delta)$ and $1/x \geq L+1$ as desired. ■ □

Remark. Using Exercise 3.2.7, we now see that the two-sided limit $\lim_{x\to 0} \frac{1}{x}$ does not exist. This confirms our intuition in the beginning of this chapter. In fact, neither one-sided limit exists: see Exercise 3.2.6.

Remark. The previous two examples show that several different types of discontinuities can arise in a function. One type, called a *jump discontinuity*, occurs when both one-sided limits exist but are unequal. Thus, the floor function has jump discontinuities at each integer.

Another type, called a *removable discontinuity*, occurs when the two one-sided limits are equal, but the function value does not equal that limit. This is often the least worrisome type of discontinuity, because it can be "fixed" by redefining the function's value at the point! Figure 3.4 illustrates a removable discontinuity.

The discontinuity in $1/x$ at 0 is called an *infinite disconuity*, because the function values are unbounded near 0. More details can be found in the exercises of Section 3.4: see Definitions 3.32 and 3.33.

3.2 Exercises

1. Sometimes the Triangle Inequality is stated in terms of distances between three points as follows: for any $x, y, z \in \mathbb{R}$, the distance from x to z is at most the sum of the distances from x to y and then from y to z. In other words,

$$\forall x, y, z \in \mathbb{R}\ |z - x| \leq |y - x| + |z - y|$$

   Prove this version of the Triangle Inequality using Theorem 3.3.

2. Prove the following for all $x, y \in \mathbb{R}$:

   (a) $|-x| = |x|$
   (b) $\sqrt{x^2} = |x|$
   (c) $|x - y| \leq |x| + |y|$
   (d) $|x| - |y| \leq |x - y|$
   (e) $||x| - |y|| \leq |x - y|$

   (Hints: For (d), add $|y|$ to both sides. For (e), use Lemma 3.2.)

3. Let $a \in \mathbb{R}$ be given and let f be a real function. Prove that f is defined near a iff there exists some $\delta > 0$ so that $(a - \delta, a + \delta) - \{a\} \subseteq \mathrm{dom}(f)$. (This result simplifies some proofs where it is helpful to work with deleted open intervals around a which are symmetric about a.)

4. For any $a \in \mathbb{R}$, find the following limits and prove that your answer is correct:

   (a) $\lim_{x \to a} 2x + 5$
   (b) $\lim_{x \to a} 2 - x$
   (c) $\lim_{x \to a} mx + b$ for any $m, b \in \mathbb{R}$ where $m \neq 0$

5. The ceiling function is defined in Exercise 2.9.3. Find $\lim_{x \to a^-} \lceil x \rceil$ and $\lim_{x \to a^+} \lceil x \rceil$ for any $a \in \mathbb{R}$. Prove your answers are correct.

6. Prove that $\lim_{x \to 0^-} \frac{1}{x}$ does not exist.

7. Suppose that $a, L \in \mathbb{R}$ and f is a real function whose domain contains an open interval containing a (except for possibly a itself). Prove that

$$\lim_{x \to a} f(x) = L \leftrightarrow \left(\lim_{x \to a^+} f(x) = L\right) \wedge \left(\lim_{x \to a^-} f(x) = L\right)$$

   In other words, a two-sided limit exists iff the one-sided limits exist and are equal.

   (Hint: For one direction, we assume both one-sided limits exist. Thus, each one-sided limit has a corresponding choice of δ. If δ_+ is the choice for the right-hand limit, saying how close x should be to a from the right, and δ_- is the choice for the left-hand limit which says how close x should be to a from the left, then how close can x be to a from EITHER side?)

8. Find a function $f : \mathbb{R} \to \mathbb{R}$ and a number $a \in \mathbb{R}$ where $\lim_{x \to a} f(x) = a$ but $\lim_{x \to a} f(f(x)) \neq a$.

9. Suppose that $a, L, M \in \mathbb{R}$ and f is a real function defined near a. Prove that if $f(x) \to L$ and $f(x) \to M$ as $x \to a$, then $L = M$. This justifies us saying "the" limit of a function as opposed to saying "a" limit.

   (Hint: Suppose WLOG that $L < M$ for a contradiction. Imagine drawing a horizontal line right in between the lines $y = L$ and $y = M$ to serve as a "barrier" between L and M. Find a value of ϵ such that $f(x)$ is below this barrier when $|f(x) - L| < \epsilon$ and is above this barrier when $|f(x) - M| < \epsilon$.)

10. Prove that $\lim_{x \to 0} \frac{|x|}{x}$ does not exist. However, show that the one-sided limits $\lim_{x \to 0^+} \frac{|x|}{x}$ and $\lim_{x \to 0^-} \frac{|x|}{x}$ do exist by calculating these one-sided limits with proof.

11. Prove that $\lim_{x \to 0^+} \sqrt{x} = 0$.

12. Prove that $\lim_{x \to a} \sqrt{x} = \sqrt{a}$ for every $a \in \mathbb{R}^+$. (Hint: Note that $|\sqrt{x} - \sqrt{a}||\sqrt{x} + \sqrt{a}| = |x - a|$. How small can $\sqrt{x} + \sqrt{a}$ be?)

13. Suppose $f, g : \mathbb{R} \to \mathbb{R}$ and $p \in \mathbb{R}$ satisfy the property that for all $x \in \mathbb{R}$ with $x \neq p$, $f(x) = g(x)$. Suppose also that for some $a \in \mathbb{R}$ with $a \neq p$, $\lim_{x \to a} f(x)$ exists. Prove that $\lim_{x \to a} g(x) = \lim_{x \to a} f(x)$.[4]

    (Hint: How small should $|x - a|$ be in order to guarantee $x \neq p$?)

14. Consider the following function $f : \mathbb{R} \to \mathbb{R}$:

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

    Prove that for all $a \in \mathbb{R}$, $\lim_{x \to a} f(x)$ does not exist. Thus, f is discontinuous everywhere.

    (Hint: Use the facts that $\mathbb{Q}$ and $\mathbb{R} - \mathbb{Q}$ are dense in $\mathbb{R}$, as shown in Exercises 2.9.4 and 2.9.6.)

3.3 Arithmetic Limit Laws

The definition of a limit is rather cumbersome. It involves multiple quantifiers, and proving that the value of a limit is correct often involves careful work with inequalities or clever ways of writing formulas (for instance, see Exercise 3.2.12). Luckily, just like many functions are built from simpler functions by adding, multiplying, and so forth, many limits can also be computed from simpler limits. The theorems that help us compute limits in this way are frequently called *limit laws*.

We should be clear about how we build functions out of other functions. If f and g are two real functions, then the sum $f + g$ is also a real function, and it's defined by $(f + g)(x) = f(x) + g(x)$ whenever $f(x)$ and $g(x)$ are both defined. (We describe this style of definition by saying that $f + g$ is defined *pointwise*; roughly, "the value of the sum is the sum of the values".) Thus, $f + g$ has domain $\text{dom}(f) \cap \text{dom}(g)$. There is a subtle distinction here between the FUNCTION $f + g$ and the VALUE $f(x) + g(x)$ that the function takes when $x \in \text{dom}(f + g)$.

Similarly, the difference $f - g$ and product fg are defined pointwise: their domain is $\text{dom}(f) \cap \text{dom}(g)$, and at any $x \in \text{dom}(f) \cap \text{dom}(g)$, we have $(f -$

[4]This result is also true when $a = p$, but then the proof is much simpler: if $a = p$, then f and g have the same limit at a because the definition of limit doesn't consider the value of the function at a!

$g)(x) = f(x) - g(x)$ and $(fg)(x) = f(x)g(x)$. We also can do multiplication by constants: when $c \in \mathbb{R}$, cf is the function for which $(cf)(x) = cf(x)$ for all $x \in \text{dom}(f)$. Lastly, division f/g is defined pointwise by $(f/g)(x) = f(x)/g(x)$, but now the domain consists of values of x for which $f(x)$ and $g(x)$ are both defined AND $g(x) \neq 0$.

Before starting the limit laws, it will be worth our while to get a technical concern out of the way. Recall that in order for a function f to have a limit as the input approaches a, f needs to be defined near a. This next result is very helpful for showing that when you combine functions defined near a, the result is still defined near a.

Lemma 3.15. *For all $a \in \mathbb{R}$, the intersection of two open intervals containing a is another open interval containing a. It follows that the intersection of two deleted open intervals around a is also a deleted open interval around a.*

Strategy. Suppose that (a_1, b_1) and (a_2, b_2) are both open intervals containing a. Certainly their intersection contains a. Now we consider the question: what are bounds on the elements of $(a_1, b_1) \cap (a_2, b_2)$? If $x \in (a_1, b_1) \cap (a_2, b_2)$, then $x > a_1$ and $x > a_2$, i.e. a_1 and a_2 are lower bounds for the possible values of x. The bigger choice of lower bound yields the stricter condition (i.e. $x > \max\{a_1, a_2\}$ implies $x > \min\{a_1, a_2\}$). Similarly, we know $x < b_1$ and $x < b_2$, and the smaller of these upper bounds yields the stricter condition.

Proof. Let $a, a_1, a_2, b_1, b_2 \in \mathbb{R}$ be given so that $a \in (a_1, b_1) \cap (a_2, b_2)$. For all $x \in \mathbb{R}$, we have the following:

$$\begin{aligned} x \in (a_1, b_1) \cap (a_2, b_2) &\leftrightarrow (a_1 < x < b_1 \text{ and } a_2 < x < b_2) \\ &\leftrightarrow (\max\{a_1, a_2\} < x < \min\{b_1, b_2\}) \end{aligned}$$

This shows that $(a_1, b_1) \cap (a_2, b_2) = (\max\{a_1, a_2\}, \min\{b_1, b_2\})$, which is an open interval. Similarly, the intersection of deleted open intervals is

$$((a_1, b_1) - \{a\}) \cap ((a_2, b_2) - \{a\}) = (\max\{a_1, a_2\}, \min\{b_1, b_2\}) - \{a\}$$

which is a deleted open interval around a. ■

Constant Multiples

Let's start our analysis of limit laws by considering the simplest kind of example: multiplying a function by a constant.

Example 3.16:

It seems natural to guess that $\lim_{x\to 0} x^2 = 0^2 = 0$ and $\lim_{x\to 0}(3x^2) = 3 \cdot 0^2 = 0$. Indeed, we'll prove these limits, but as we proceed, we should try to see what effect the extra constant of 3 really has on the proof.

To start with the first limit, let $\epsilon > 0$ be given. We'd like to guarantee $|x^2 - 0| < \epsilon$ provided that $0 < |x - 0| < \delta$ for some $\delta > 0$. (We sometimes call δ the "control" value, since it controls how much choice we have over our input.) Note that $|x^2| < \epsilon$ iff $\sqrt{|x^2|} < \sqrt{\epsilon}$, and $\sqrt{|x^2|} = |x|$. Thus, we can take $\delta = \sqrt{\epsilon}$, and this finishes the first limit.

For the second limit, if $\epsilon > 0$ is given, we'd like a control δ that guarantees $|3x^2 - 0| < \epsilon$. Since $|3x^2 - 0| = 3|x^2|$, we have $3|x^2| < \epsilon$ iff $|x^2| < \epsilon/3$. As in the first proof, we could take square roots of both sides and thus we'd choose $\delta = \sqrt{\epsilon/3}$.

However, there's another way to do this. The first proof already showed that we could make $|x^2|$ as small as we'd like, provided we use an appropriate choice of δ. In the second proof, we'd like to guarantee $|x^2| < \epsilon/3$ (whereas the first proof guaranteed $|x^2| < \epsilon$). The steps of the first proof work perfectly for this second proof if you just replace the occurrences of ϵ with $\epsilon/3$ in the argument. Doing this gives you $\delta = \sqrt{\epsilon/3}$ instead of $\delta = \sqrt{\epsilon}$. $\square$

The key idea here is the tactic of "using a different ϵ". Since our first proof worked for arbitrary positive values of ϵ, we could use ANY positive number we want in place of ϵ. Basically, to guarantee an accuracy of ϵ for the second limit, we need to guarantee a "new accuracy" of $\epsilon/3$ in the first limit. The next theorem will use this tactic more generally.

Theorem 3.17. *Let $a, L \in \mathbb{R}$ be given, and let f be a real function satisfying $\lim_{x\to a} f(x) = L$. Then, for all $c \in \mathbb{R}$, $\lim_{x\to a}(cf(x)) = cL$. In other words,*

$$\lim_{x\to a}(cf(x)) = c \lim_{x\to a} f(x)$$

provided the limit on the right exists. This is sometimes described by saying "constant multiples can come out of limits".

Strategy. We want to guarantee that $|cf(x) - cL| < \epsilon$ for x close enough to a. Since $|cf(x) - cL| = |c||f(x) - L|$, we would like $|f(x) - L| < \epsilon/|c|$ if

$c \neq 0$ (the case of $c = 0$ is handled separately). Therefore, since $f(x) \to L$ as $x \to a$, there is some control $\delta > 0$ on x which guarantees $f(x)$ is within $\epsilon/|c|$ of L. (Our "new accuracy" is $\epsilon/|c|$ instead of ϵ.)

Proof. Let a, L, f be given as described in the theorem. Let $c \in \mathbb{R}$ be given. Since $\lim_{x\to a} f(x)$ exists, f is defined near a, so as $\mathrm{dom}(cf) = \mathrm{dom}(f)$, the function cf is also defined near a. If $c = 0$, then $cf(x) = 0$ for all $x \in \mathrm{dom}(f)$, and therefore

$$\lim_{x\to a}(cf(x)) = \lim_{x\to a} 0 = 0 = 0 \cdot L$$

Otherwise, if $c \neq 0$, then let $\epsilon > 0$ be given. Because $\epsilon/|c| > 0$ and $f(x) \to L$ as $x \to a$, we may choose some $\delta > 0$ so that any $x \in \mathrm{dom}(f)$ satisfying $0 < |x - a| < \delta$ also satisfies $|f(x) - L| < \epsilon/|c|$. Then, for all $x \in \mathrm{dom}(f)$ with $0 < |x - a| < \delta$, we have

$$|cf(x) - cL| = |c||f(x) - L| < |c|\left(\frac{\epsilon}{|c|}\right) = \epsilon$$

as desired. ■

Addition and Subtraction: Linear Combinations

Our next limit laws deal with adding and subtracting functions. Let's look at an example of a proof of a limit using a sum to gain an appreciation for the tactics the argument uses.

Example 3.18:

Let $f, g : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x$ and $g(x) = \lfloor x \rfloor$ for any $x \in \mathbb{R}$. We've previously seen that both f and g are continuous at $1/2$. Let's prove that $\lim_{x\to 1/2}(x + \lfloor x \rfloor) = 1/2 + \lfloor 1/2 \rfloor$.

For any $\epsilon > 0$, we want to find a control δ on x which guarantees

$$\left|(x + \lfloor x \rfloor) - \left(\frac{1}{2} + \left\lfloor \frac{1}{2} \right\rfloor\right)\right| < \epsilon$$

when $0 < |x - 1/2| < \delta$. As a first step, let's regroup things so that similar terms come together, so

$$\left|(x + \lfloor x \rfloor) - \left(\frac{1}{2} + \left\lfloor \frac{1}{2} \right\rfloor\right)\right| = \left|\left(x - \frac{1}{2}\right) + \left(\lfloor x \rfloor - \left\lfloor \frac{1}{2} \right\rfloor\right)\right|$$

This seems helpful, because $x \to 1/2$ and $\lfloor x \rfloor \to \lfloor 1/2 \rfloor$ as $x \to 1/2$, so the distances $|x - 1/2|$ and $|\lfloor x \rfloor - \lfloor 1/2 \rfloor|$ can be made as small as we'd like.

However, how do we relate $|(x - 1/2) + (\lfloor x \rfloor - \lfloor 1/2 \rfloor)|$ to those two distances? This is where the Triangle Inequality from Theorem 3.3 becomes useful, giving us:

$$\left|\left(x - \frac{1}{2}\right) + \left(\lfloor x \rfloor - \left\lfloor \frac{1}{2} \right\rfloor\right)\right| \le \left|x - \frac{1}{2}\right| + \left|\lfloor x \rfloor - \left\lfloor \frac{1}{2} \right\rfloor\right|$$

We'd like this sum of two distances on the right-hand side to be less than ϵ, and we know we can make each of those distances as small as we'd like. Thus, let's try to make each distance less than $\epsilon/2$, so that the total sum is less than ϵ. (Basically, we're using the "use a different ϵ" tactic twice.)

Since $0 < |x - 1/2| < \delta$, and we want $|x - 1/2| < \epsilon/2$, this suggests choosing $\delta = \epsilon/2$. However, we also want to have $|\lfloor x \rfloor - \lfloor 1/2 \rfloor| < \epsilon/2$. Using the ideas from the proof of Theorem 3.12, we consider choosing $\delta = 1/2$ (since $\lfloor x \rfloor$ is constantly 0 on $(1/2 - 1/2, 1/2 + 1/2) = (0, 1)$).

This suggests two values of δ: $\epsilon/2$ and $1/2$. How do we pick ONE value of δ that guarantees both $|x - 1/2| < \epsilon/2$ AND $|\lfloor x \rfloor - \lfloor 1/2 \rfloor| < \epsilon/2$? Basically, we want $0 < |x - 1/2| < \epsilon/2$ and $0 < |x - 1/2| < 1/2$ at the same time (corresponding to our two choices for δ). To make this happen, we use the stricter control on x, i.e. δ is the minimum of $\epsilon/2$ and $1/2$. Thus, when

$$0 < |x - 1/2| < \min\left\{\frac{\epsilon}{2}, \frac{1}{2}\right\}$$

we have both $|x - 1/2| < \epsilon/2$ and $|x - 1/2| < 1/2$, as we wanted. $\square$

Using the ideas from the example above, we prove a general addition law.

Theorem 3.19. *Let $a, L, M \in \mathbb{R}$ be given, and let f and g be real functions satisfying $\lim_{x \to a} f(x) = L$ and $\lim_{x \to a} g(x) = M$. Then $\lim_{x \to a}(f(x) + g(x)) = L + M$. In other words,*

$$\lim_{x \to a}(f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$$

provided the limits on the right exist. This is often described as saying "the limit of a sum is the sum of the limits".

Strategy. For any $\epsilon > 0$, we'd like to find values of x close enough to a so that $|(f(x) + g(x)) - (L + M)| < \epsilon$. As in the example, we'd like to group things in terms of $|f(x) - L|$ and $|g(x) - M|$, since we can control those distances. By regrouping terms and using the Triangle Inequality, we get

$$|(f(x)+g(x))-(L+M)| = |(f(x)-L)+(g(x)-M)| \leq |f(x)-L|+|g(x)-M|$$

To make the sum less than ϵ, we aim to make each term of the sum less than $\epsilon/2$. Since $f(x) \to L$ as $x \to a$, we can make $|f(x) - L| < \epsilon/2$ using some control $\delta_1 > 0$ on x. Similarly, we can make $|g(x) - M| < \epsilon/2$ using a control $\delta_2 > 0$ on x (we use subscripts because the two functions might not use the same value of δ). Since we want both $|f(x) - L| < \epsilon/2$ and $|g(x) - M| < \epsilon/2$ at the same time, we'll take δ to be the smaller of δ_1 and δ_2, i.e. $\delta = \min\{\delta_1, \delta_2\}$.

(This tactic of using the Triangle Inequality, followed by bounding two terms each by $\epsilon/2$ and choosing δ as a minimum of two possibilities, is common enough that sometimes it is called "*an $\epsilon/2$ argument*".)

Proof. Let a, L, M, f, g be given as described in the theorem. Since $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist, f and g are defined near a, so by Lemma 3.15, $\text{dom}(f) \cap \text{dom}(g)$ contains a deleted open interval around a. As $\text{dom}(f+g) = \text{dom}(f) \cap \text{dom}(g)$, $f + g$ is defined near a.

Now, let $\epsilon > 0$ be given. We wish to find some $\delta > 0$ so that for all $x \in \text{dom}(f + g)$,

$$0 < |x - a| < \delta \to |(f(x) + g(x)) - (L + M)| < \epsilon$$

Since $\epsilon/2 > 0$, by the definition of limit applied to f and g as $x \to a$, we may choose values $\delta_1, \delta_2 > 0$ so that

$$\forall x \in \text{dom}(f)\ (0 < |x - a| < \delta_1 \to |f(x) - L| < \epsilon/2)$$

and

$$\forall x \in \text{dom}(g)\ (0 < |x - a| < \delta_2 \to |g(x) - M| < \epsilon/2)$$

Now, let $\delta = \min\{\delta_1, \delta_2\}$, so $\delta > 0$. Let $x \in \text{dom}(f + g)$ be given with $0 < |x-a| < \delta$. Therefore, $x \in \text{dom}(f)$ and $x \in \text{dom}(g)$, and by the choice of δ, $0 < |x - a| < \delta_1$ and $0 < |x - a| < \delta_2$. Thus, using the Triangle Inequality,

$$\begin{aligned} |(f(x) + g(x)) - (L + M)| &= |(f(x) - L) + (g(x) - M)| \\ &\leq |f(x) - L| + |g(x) - M| \\ &< \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

as desired. ■

Corollary 3.20. *Let $a, L, M \in \mathbb{R}$ be given, and let f and g be real functions satisfying $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Then $\lim_{x\to a}(f(x) - g(x)) = L - M$. In other words,*

$$\lim_{x\to a}(f(x) - g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x)$$

provided the limits on the right exist. This is often described as saying "the limit of a difference is the difference of the limits".

Strategy. We'll omit the formal proof and just describe the strategy. One very quick way to prove this corollary is to reuse the laws we have already proven! First, we write $f(x) - g(x)$ as $f(x) + (-1)g(x)$. Next, we apply the constant-multiple law (Theorem 3.17) to $(-1)g(x)$. Lastly, we apply the addition law (Theorem 3.19) to $f(x) + (-1)g(x)$.

However, it's also worth noting that we can give a proof of the subtraction law in nearly the same way as for the addition law. The main change is that instead of the Triangle Inequality, we can use the result of Exercise 3.2.2.(c) to say

$$|(f(x)-g(x))-(L-M)| = |(f(x)-L)-(g(x)-M)| \le |f(x)-L|+|g(x)-M|$$

After this, we can use an $\epsilon/2$ argument with $|f(x) - L|$ and $|g(x) - M|$.

To summarize the results we have shown so far, we have shown that for any $a, c, d \in \mathbb{R}$ and any real functions f, g for which $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist, we have

$$\lim_{x\to a}(cf(x) + dg(x)) = c\lim_{x\to a} f(x) + d\lim_{x\to a} g(x)$$

(this follows just by applying the previous results). A sum of the form $cf(x)+dg(x)$, where we only use constant multiples (i.e. no product terms like $f(x)g(x)$) and sums or differences, is called a *linear combination* of $f(x)$ and $g(x)$. With this terminology, we say that a limit of a linear combination is a linear combination of limits, or more simply we say that taking limits is a *linear operation*. The important operations of calculus are all linear operations, as we will see in later chapters.

Products

Now we consider products. Once again, we will need a new tactic for handling these limits, which we illustrate with an example.

Example 3.21:

We showed earlier that $x^2 \to 0$ as $x \to 0$. Now, let's show $x^2 \to 1$ as $x \to 1$. Perhaps surprisingly, this will be a much harder task.

Let $\epsilon > 0$ be given, and we wish to find a control δ on x that guarantees $|x^2 - 1| < \epsilon$ when $0 < |x - 1| < \delta$. Since we can make the distance $|x - 1|$ small, we ought to try and express $|x^2 - 1|$ in terms of $|x - 1|$ if possible. To do this, we factor to get $|x^2 - 1| = |x - 1||x + 1|$.

One tempting idea is to say "if we want $|x - 1||x + 1| < \epsilon$, i.e. $|x - 1| < \epsilon/|x + 1|$, then why not pick $\delta = \epsilon/|x + 1|$?" This doesn't work, because in the limit definition, δ can only depend on ϵ, not on x as well. (After all, we choose x based on δ, since δ tells us how to control the x's, so having δ based on x as well would lead to circular reasoning!)

However, a modification of this idea does work: let's try to find a good bound for $|x + 1|$ which doesn't depend on δ, and then we'll use that bound to help us pick δ. The key tactic here is that since we're taking a limit as x goes to 1, we can make x as close to 1 as we'd like. This allows us to make bounds on x, which lead to bounds on $x + 1$.

Suppose we require that $|x - 1| < 1/2$ (i.e. we consider $1/2$ as a possible choice for δ). Then, $1/2 < x < 3/2$, so $3/2 < x + 1 < 5/2$. This means that

$$\frac{3}{2}|x - 1| < |x - 1||x + 1| < \frac{5}{2}|x - 1|$$

Thus, to make $|x - 1||x + 1| < \epsilon$, it's enough to guarantee $(5/2)|x - 1| < \epsilon$, i.e. $|x - 1| < 2\epsilon/5$, so we are led to consider $2\epsilon/5$ as another choice for δ.

As in the proof of the addition law, we now have two possibilities for δ, and we go with the stricter one. Thus, we take $\delta = \min\{1/2, 2\epsilon/5\}$. $\square$

Using ideas from the approach in the previous example, we come up with the following rule:

Theorem 3.22. *Let $a, L, M \in \mathbb{R}$ be given, and let f and g be real functions satisfying $\lim_{x \to a} f(x) = L$ and $\lim_{x \to a} g(x) = M$. Then $\lim_{x \to a}(f(x)g(x)) = LM$. In other words,*

$$\lim_{x \to a}(f(x)g(x)) = \left(\lim_{x \to a} f(x)\right)\left(\lim_{x \to a} g(x)\right)$$

provided the limits on the right exist. We sometimes describe this as "the limit of a product is the product of the limits".

Strategy. As usual, when $\epsilon > 0$, we wish to make $|f(x)g(x) - LM| < \epsilon$ for values of x close to a. However, unlike with the addition law, it's not clear how to write $|f(x)g(x) - LM|$ in terms of $|f(x) - L|$ or $|g(x) - M|$. The trick here is to add and subtract a term:

$$\begin{aligned} f(x)g(x) - LM &= f(x)g(x) - Lg(x) + Lg(x) - LM \\ &= g(x)(f(x) - L) + L(g(x) - M) \end{aligned}$$

Therefore, by the Triangle Inequality,

$$|f(x)g(x) - LM| \le |g(x)||f(x) - L| + |L||g(x) - M|$$

We'd like to make the whole sum less than ϵ, so as in the proof of the addition law, let's make each term in the sum less than $\epsilon/2$.

The second term of the sum is easy to estimate, because L does not depend on x. If we want $|L||g(x) - M|$ less than $\epsilon/2$, we should make $|g(x) - M|$ less than $\epsilon/(2|L|)$, but that's only allowed when $L \neq 0$. By making $|g(x) - M|$ even smaller, say less than $\epsilon/(2|L|+1)$, we'll have $|L||g(x) - M| < \epsilon/2$ even when $L = 0$. To ensure $|g(x) - M|$ is that small, we use some control δ_1 on x (as in the proof of the addition law, we'll need to consider several different possibilities for the control δ).

The first term of the sum is harder, because $|g(x)|$ depends on x. As in the previous example, let's find an upper bound on $|g(x)|$ which IS constant, and we'll use that constant to help us get another choice of δ.

Since $g(x) \to M$ as $x \to a$, we should expect that $g(x)$ is close to M for x near a. In fact, the definition of limit says we can guarantee the error $|g(x) - M|$ is as small as we'd like. To be definite, let's say we want $|g(x) - M| < 1$, so that $g(x)$ is between $M - 1$ and $M + 1$. In order to get this accuracy, we need some control δ_2 on x.

With this insight, we should be able to say that $|g(x)||f(x) - L| \le |M + 1||f(x) - L|$, which we want to be less than $\epsilon/2$. This is almost correct, but it has one error. If M is negative, then it is possible that $g(x) \le M + 1$ does NOT imply $|g(x)| \le |M + 1|$. However, we can fix this by noting that it is true that $|g(x)| \le |M| + 1$: writing $|g(x)|$ as $|M + (g(x) - M)|$, we find that

$$|g(x)| = |M + (g(x) - M)| \le |M| + |g(x) - M| \le |M| + 1$$

by the Triangle Inequality.

Hence, we have $|g(x)||f(x)-L| \leq (|M|+1)|f(x)-L|$. By picking a third control δ_3 on x, we can guarantee that $|f(x)-L| < \epsilon/(2(|M|+1))$, so that $(|M|+1)|f(x)-L| < \epsilon/2$. Finally, we have all the pieces we want, but we've introduced the need for three possible controls δ_1, δ_2, and δ_3 on x. We'll go with the strictest one, i.e. the minimum.

Proof. Let a, L, M, f, g be given as described in the theorem. Because f and g are defined near a (since they have limits as $x \to a$), Lemma 3.15 shows that $\text{dom}(f) \cap \text{dom}(g)$ contains a deleted open interval around a. Hence, fg is defined near a.

Now, let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for all $x \in \text{dom}(fg)$,

$$(0 < |x-a| < \delta) \to (|f(x)g(x) - LM| < \epsilon)$$

First, because $g(x) \to M$ as $x \to a$, and $\epsilon/(2|L|+1) > 0$, by the definition of limit, we may choose $\delta_1 > 0$ so that

$$\forall x \in \text{dom}(g) \ \left(0 < |x-a| < \delta_1 \to |g(x) - M| < \frac{\epsilon}{2|L|+1}\right)$$

Second, by the same reasoning, we may choose $\delta_2 > 0$ so that

$$\forall x \in \text{dom}(g) \ (0 < |x-a| < \delta_2 \to |g(x) - M| < 1)$$

Thus, using the Triangle Inequality, if $x \in \text{dom}(g)$ and $0 < |x-a| < \delta_2$, then

$$|g(x)| = |M + (g(x) - M)| \leq |M| + |g(x) - M| < |M| + 1$$

Third, because $f(x) \to L$ as $x \to a$, and $\epsilon/(2(|M|+1)) > 0$, by the definition of limit, we may choose $\delta_3 > 0$ so that

$$\forall x \in \text{dom}(f) \ \left(0 < |x-a| < \delta_3 \to |f(x) - L| < \frac{\epsilon}{2(|M|+1)}\right)$$

Choose $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Thus, for all $x \in \text{dom}(fg)$ with $0 < |x-a| < \delta$, we have $x \in \text{dom}(f)$, $x \in \text{dom}(g)$, $0 < |x-a| < \delta_1$, $0 < |x-a| < \delta_2$, and $0 < |x-a| < \delta_3$. By the choices of δ_1, δ_2, and δ_3, and the Triangle Inequality, we find

$$\begin{aligned}
|f(x)g(x) - LM| &= |f(x)g(x) - Lg(x) + Lg(x) - LM| \\
&= |g(x)(f(x) - L) + L(g(x) - M)| \\
&\leq |g(x)||f(x) - L| + |L||g(x) - M| \\
&< (|M| + 1)\left(\frac{\epsilon}{2(|M| + 1)}\right) + |L|\left(\frac{\epsilon}{2|L| + 1}\right) \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\end{aligned}$$

as desired. ■

Quotients

With the product law available, we'd like to handle quotients. To simplify our work, we'll prove a law allowing us to take limits of reciprocals easily. Afterward, if f and g are real functions defined near a, we'll write the fraction f/g as the product $f \cdot (1/g)$, so the product law and reciprocal law together tell us the limit of the fraction.

To see how handling limits of reciprocals works, let's try an example.

Example 3.23:

We've already seen that $1/x$ has no limit as $x \to 0$. However, we do have a limit as $x \to 1$: we'll prove $1/x \to 1$ in that case.

If $\epsilon > 0$ is given, we'd like to find out how close to 1 we want x to be in order to insure $|1/x - 1| < \epsilon$. A common tactic with fractions like these is to use a common denominator

$$\left|\frac{1}{x} - 1\right| = \left|\frac{1-x}{x}\right| = \frac{|x-1|}{|x|}$$

since $|1 - x| = |-(x - 1)| = |x - 1|$. We're taking a limit as $x \to 1$, so we can make $|x - 1|$ as small as we'd like. However, we need to find a bound on $1/|x|$ which doesn't depend on δ (like we did with our product law example).

The idea is to notice once again that x can be made as close to 1 as we like, so $|x|$ can be made close to 1. Say, for example, that we decide to make $|x - 1| < 1/2$ (i.e. we consider $1/2$ as a possible choice for δ), so that

$1/2 < x < 3/2$. It follows then that $1/2 < |x| < 3/2$, so $2/3 < 1/|x| < 2/1$ and

$$\frac{2}{3}|x-1| < \frac{|x-1|}{|x|} < 2|x-1|$$

Thus, we need only guarantee that $2|x-1| < \epsilon$, which suggests $\epsilon/2$ as a possibility for δ. With two possibilities available ($1/2$ and $\epsilon/2$), we make δ the minimum of those two. $\square$

Our example earlier used two important techniques. First, it made a common denominator out of a subtraction. Second, it estimated the denominator with bounds, much like the tactic used in the product law. We use those techniques more generally to make a reciprocal law:

Lemma 3.24. *Let $a, M \in \mathbb{R}$ be given, and let g be a real function for which $\lim_{x\to a} g(x) = M$. If $M \neq 0$, then g is defined and nonzero near a, and*

$$\lim_{x\to a} \frac{1}{g(x)} = \frac{1}{M}$$

Strategy. If $\epsilon > 0$, then we'd like to find out how close x should be to a to guarantee $|1/g(x) - 1/M| < \epsilon$. By collecting the terms over a common denominator, we find

$$\left|\frac{1}{g(x)} - \frac{1}{M}\right| = \left|\frac{M-g(x)}{Mg(x)}\right| = \frac{|g(x)-M|}{|Mg(x)|}$$

since $|M - g(x)| = |-(g(x)-M)| = |g(x)-M|$. Since $g(x) \to M$ as $x \to a$, the numerator $|g(x)-M|$ can be made as small as we'd like. We need to get bounds on the denominator $|g(x)|$ so that we can estimate the entire fraction.

Since $g(x) \to M$, the values of $g(x)$ can be made close to M. In particular, there is a control $\delta_1 > 0$ on x which guarantees that $|g(x) - M| < |M|/2$ (remember that M could be negative, and distances need to be nonnegative). It follows that $g(x)$ is between $M/2$ and $3M/2$, and thus g is nonzero near a.

Thus, we have

$$\frac{|M|^2}{2} < |M||g(x)| < \frac{3|M|^2}{2}$$

and therefore

$$\frac{2|g(x)-M|}{3|M|^2} < \frac{|g(x)-M|}{|Mg(x)|} < \frac{2|g(x)-M|}{|M|^2}$$

Therefore, we need $2|g(x)-M|/|M|^2 < \epsilon$, which suggests taking a control δ_2 to guarantee $|g(x)-M| < |M|^2\epsilon/2$. By taking a control stricter than δ_1 and δ_2, such as their minimum, we are done.

An alternate way to finish, which we'll use in the proof for the sake of variety, is to notice that our two controls are each trying to influence the same limit for g. Thus, we only need one control value δ corresponding to the tighest accuracy we need, i.e. one value of δ to guarantee that

$$|g(x)-M| < \min\left\{\frac{|M|}{2}, \frac{|M|^2\epsilon}{2}\right\}$$

Proof. Let a, g, M be given as described in the theorem. Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for all $x \in \text{dom}(1/g)$,

$$\left(0 < |x-a| < \delta \to \left|\frac{1}{g(x)} - \frac{1}{M}\right| < \epsilon\right)$$

Since $g(x) \to M$ as $x \to a$, we may choose $\delta > 0$ so that for all $x \in \text{dom}(g)$,

$$\left(0 < |x-a| < \delta \to |g(x)-M| < \min\left\{\frac{|M|}{2}, \frac{|M|^2\epsilon}{2}\right\}\right)$$

In particular, this implies that for all $x \in \text{dom}(g) \cap ((a-\delta, a+\delta) - \{a\})$, $|g(x)-M| < |M|/2$, so $g(x)$ is between $M/2$ and $3M/2$. In particular, $g(x) \neq 0$. As g is defined near a (it has a limit as $x \to a$), Lemma 3.15 shows that g is defined and nonzero in an open interval around a, except possibly at a.

Now, let $x \in \text{dom}(1/g) \subseteq \text{dom}(g)$ be given with $0 < |x-a| < \delta$. We have $|g(x)| > |M|/2$, and we also have $|g(x)-M| < |M|^2\epsilon/2$. Therefore,

$$\begin{aligned}
\left|\frac{1}{g(x)} - \frac{1}{M}\right| &= \left|\frac{M-g(x)}{Mg(x)}\right| \\
&= \frac{|g(x)-M|}{|Mg(x)|} \\
&< \frac{|M|^2\epsilon/2}{|M|^2/2} = \epsilon
\end{aligned}$$

as desired. ■

Corollary 3.25. *Let $a, L, M \in \mathbb{R}$ be given, and let f and g be real functions satisfying $\lim_{x \to a} f(x) = L$ and $\lim_{x \to a} g(x) = M$. Also assume that $M \neq 0$. Then $\lim_{x \to a}(f(x)/g(x)) = L/M$. In other words,*

$$\lim_{x \to a} \left(\frac{f(x)}{g(x)} \right) = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$$

provided the fraction on the right is defined. This is sometimes described as "the limit of a quotient is the quotient of the limits".

Strategy. The strategy was mentioned before Lemma 3.24. All we have to do is write f/g as $f \cdot (1/g)$ and use the reciprocal and product laws.

Proof. Let a, L, M, f, g be given as described. By Lemma 3.24, g is defined and nonzero in an open interval around a, except possibly at a. As f is also defined near a, Lemma 3.15 tells us that $\mathrm{dom}(f) \cap \mathrm{dom}(g)$ contains an open interval around a (except possibly a itself) within which g is nonzero. Thus, f/g is defined near a.

By using the product law (Theorem 3.22) and the previous lemma, we calculate

$$\begin{aligned} \lim_{x \to a} \left(\frac{f(x)}{g(x)} \right) &= \lim_{x \to a} \left(f(x) \cdot \frac{1}{g(x)} \right) \\ &= L \cdot \frac{1}{M} = \frac{L}{M} \end{aligned}$$

as desired. ∎

Basic Examples

Now, let's put the limit laws to good use by computing some limits. It is much easier to use the limit laws when possible than to use the definition of limit to prove these limits are correct.

In the last section, we saw that constant functions and the identity function $\mathrm{id}_{\mathbb{R}}$ (which maps x to x for each $x \in \mathbb{R}$) are continuous everywhere. It follows that the functions which map $x \in \mathbb{R}$ to x^2, x^3, etc. are all products of $\mathrm{id}_{\mathbb{R}}$ and are also continuous everywhere. For instance, for any $a \in \mathbb{R}$,

$$\lim_{x \to a}(x^2) = \lim_{x \to a}(x \cdot x) = \left(\lim_{x \to a} x \right) \left(\lim_{x \to a} x \right) = a \cdot a = a^2$$

Note that we wrote our calculations as if we went from left to right. However, technically, we need to know that $\lim_{x \to a} x$ exists BEFORE we are allowed to use the product law, so the justifications really go from right to left.

As a longer example, we do the following computation for any $a \in \mathbb{R}$ (the law used as justification is provided with each step):

$$\begin{array}{rcll} \lim_{x \to a}(3x^3 + 2x + 1) & = & \lim_{x \to a}(3x^3) + \lim_{x \to a}(2x) + \lim_{x \to a}(1) & \text{Sum} \\ & = & 3\lim_{x \to a}(x^3) + 2\lim_{x \to a} x + \lim_{x \to a} 1 & \text{Constant Multiple} \\ & = & 3\left(\lim_{x \to a} x\right)^3 + 2\lim_{x \to a} x + \lim_{x \to a} 1 & \text{Product} \\ & = & 3a^3 + 2a + 1 & \text{Known limits} \end{array}$$

(Note that when using the sum and product law above, we really used them twice, since we had sums and products of three terms.)

More generally, the following theorem is left as an induction exercise:

Theorem 3.26. *Every polynomial function, i.e. every function $f : \mathbb{R} \to \mathbb{R}$ of the form $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ for some $n \in \mathbb{N}^*$ and some $a_n, a_{n-1}, \ldots, a_0 \in \mathbb{R}$, is continuous everywhere.*

Similarly, we can easily compute limits of fractions of polynomials:

Corollary 3.27. *Every rational function, i.e. every function of the form f/g where f and g are polynomials, is continuous at every point in its domain.*

Example 3.28:

The limit laws can apply to more than just sums or products of two functions. By using these laws over and over, we can apply them to any finite sum or products of functions. To help state this result clearly, we should introduce some notation for writing sums and products more compactly.

When a, b are integers with $a \leq b$, and f is a function defined on the integers from a to b, we write the sum

$$f(a) + f(a+1) + f(a+2) + \cdots + f(b)$$

using the *summation operator* $\sum$ as

$$\sum_{i=a}^{b} f(i)$$

Here, the $\sum$ symbol is a capital sigma (remember that the "s" is for "sum"). The subscript $i = a$ tells us that we are using the variable i as an *index of summation* which starts counting at a. The superscript of b tells us that the value of i counts up to b. The whole expression is read as "the sum from $i = a$ to b of $f(i)$", and this amounts to plugging in a for i, then $a + 1$ for i, and so forth up to plugging in b for i, and finally adding up all the results together. (Also, if $b < a$, then we treat the whole sum as just 0, i.e. the "sum of nothing".)

For example, we have the following sums:

$$\sum_{i=1}^{5} i = 1 + 2 + 3 + 4 + 5 = 15 \qquad \sum_{i=2}^{4} i^2 = 2^2 + 3^2 + 4^2 = 29$$

(in the first sum, f is just the identity function, and in the second sum, f is the function which squares numbers). As another example, a polynomial of the form $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ can also be expressed as $\displaystyle\sum_{i=0}^{n} a_i x^i$.

The variable i is not the only valid choice of name for an index of summation. Any variable that isn't already being used for some other purpose can be used instead. For instance, the following sums are all equal:

$$\sum_{i=0}^{n} f(i) = \sum_{j=0}^{n} f(j) = \sum_{q=0}^{n} f(q)$$

(When the variable name can be changed freely like this, sometimes we say the variable is a *dummy variable*.) However, the index of summation is only in scope during the summation, and the variable has no meaning when the summation is done. Thus, an expression like $\displaystyle\left(\sum_{i=0}^{n} i\right) + i^2$ makes no sense.

With this notation, we can now generalize the sum law to the following. Suppose that $a \in \mathbb{R}$, $n \in \mathbb{N}$, and that for each i from 1 to n, f_i is a real function and L_i is a real number such that $f_i(x) \to L_i$ as $x \to a$. Then

$$\lim_{x \to a} \left(\sum_{i=1}^{n} f_i(x) \right) = \sum_{i=1}^{n} \left(\lim_{x \to a} f_i(x) \right) = \sum_{i=1}^{n} L_i$$

This result is proven by induction on n (using Theorem 3.19 to carry out the inductive step), so we omit the proof.

There is also similar notation for products using the *product operator* $\prod$ (which is a capital "pi"), so instead of

$$f(a) \cdot f(a+1) \cdot \dots \cdot f(b)$$

for integers a, b and a function f, we can write

$$\prod_{i=a}^{b} f(i)$$

For example, we have the following products:

$$\prod_{i=1}^{4} i = (1)(2)(3)(4) = 24 \qquad \prod_{i=-3}^{-1} i^2 = (-3)^2(-2)^2(-1)^2 = 36$$

We also have the convention is that when $b < a$, the product is 1.

You should formulate and prove a limit law for handling products. $\square$

Example 3.29:

The Quotient Law allows us to compute limits of quotients when the denominator does not approach 0. If the denominator does approach 0, that does NOT necessarily mean the limit fails to exist. Consider the example of

$$\lim_{x \to 0} \frac{x}{x}$$

Although x/x is not defined when $x = 0$, for all other values of x, $x/x = 1$. Therefore, since the limit is unaffected by whether the function is defined at 0 (only by whether it is defined NEAR 0), we have

$$\lim_{x \to 0} \frac{x}{x} = \lim_{x \to 0} 1 = 1$$

This is an example of what we sometimes call a *0/0 form* (read as "*0 over 0*"), where both the numerator and denominator approach 0. When you have a 0/0 form, the Quotient Law does not apply, so the formula should be rewritten in a form which is not 0/0. In some cases, the limit exists, and in some cases, it does not. For instance,

$$\lim_{x \to 0} \frac{x}{x^2}$$

is another $0/0$ form, but this limit does not exist (since $\lim_{x \to 0} \frac{1}{x}$ does not exist, as shown in the previous section.)

One common tactic with $0/0$ forms is to try and cancel common factors, as the next two calculations show:

$$\lim_{x \to 1} \frac{x^2 - x}{x^2 - 3x + 2} = \lim_{x \to 1} \frac{x(x-1)}{(x-2)(x-1)} = \lim_{x \to 1} \frac{x}{x-2} = \frac{1}{1-2} = -1$$

$$\lim_{x \to -1} \frac{x+1}{x^2 - 1} = \lim_{x \to -1} \frac{x+1}{(x-1)(x+1)} = \lim_{x \to -1} \frac{1}{x-1} = \frac{1}{-1-1} = \frac{-1}{2}$$

□

Variants of the Limit Definition: Asymptotes

Intuitively, when $a, L \in \mathbb{R}$ and f is a real function defined near a, the statement

$$\lim_{x \to a} f(x) = L$$

says that $f(x)$ can be made as close to L as desired by restricting x to be sufficiently close to a. In this definition, a and L are finite. However, there are also variations of the basic limit definition that consider "infinite" inputs or outputs. The main definitions are presented more formally in the exercises (see Definitions 3.30, 3.31, 3.32, and 3.33), but we will still mention the main ideas here.

When $L \in \mathbb{R}$, the statement

$$\lim_{x \to \infty} f(x) = L$$

informally means that $f(x)$ can be made as close to L as desired by taking x sufficiently large, i.e. "sufficiently close to infinity." There is a similar definition for a limit as $x \to -\infty$, i.e. when x is negative with a sufficiently large magnitude. If $f(x)$ has a finite limit L as either $x \to \infty$ or $x \to -\infty$, then we call the line $y = L$ a *horizontal asymptote* for f. (Thus, a function can have at most two horizontal asymptotes: one as $x \to \infty$ and one as $x \to -\infty$.)

On the other hand, we have limit definitions to address what happens when the values of f grow arbitrarily large or small. When $a \in \mathbb{R}$, the statement

$$\lim_{x \to a} f(x) = \infty$$

informally means that the values of $f(x)$ can be made arbitrarily large by restricting x to be close enough to a. There is a similar definition for saying $f(x) \to -\infty$ as $x \to a$. When $f(x)$ approaches ∞ or $-\infty$ as $x \to a$, we call the line $x = a$ a *vertical asymptote* of f.

There are some other variants on these definitions, such as one-sided vertical asymptotes or "double-infinity" limits like

$$\lim_{x\to\infty} f(x) = \infty$$

The reader is encouraged to try formulating definitions of those for him-self/herself. Furthermore, many of our limit laws, slightly modified, hold for these variants of the limit definition. To get a sense of how the proofs change, consult the exercises.

Remark. By convention, when we say a limit exists, we mean to say that the limit value L is finite. At a vertical asymptote, we do not say the limit "exists"; however, it fails to exist in a very special way.

3.4 Exercises

1. Find the following limits using the limit laws. Justify each step with an appropriate law.

   (a) $\displaystyle\lim_{x\to 0} \frac{x^2-1}{x+3}$

   (b) $\displaystyle\lim_{x\to 1} \frac{x}{1+\frac{x}{1+x}}$

2. The limit laws from the last section also work when the two-sided limits are replaced with one-sided limits (the proofs are quite similar). For instance, if $f(x) \to L$ and $g(x) \to M$ as $x \to a^+$, then $f(x) + g(x) \to L + M$ as $x \to a^+$. Using these laws (except in part (a)), find the following limits, justifying your steps:

   (a) $\displaystyle\lim_{x\to 0} |x|$

   (b) $\displaystyle\lim_{x\to 0^+} \frac{|x|}{x}$

   (c) $\displaystyle\lim_{x\to a^-} \frac{\lfloor x \rfloor}{x}$ provided that $a \in \mathbb{Z}$ and $a \neq 0$

   (d) $\displaystyle\lim_{x\to 0^-} \frac{\sqrt{x^2}}{x}$

   (Hint for (a): Consider cases based on the sign of x.)

3. (a) For which value(s) of $a \in \mathbb{R}$ does $\lim_{x \to a} (x^2 + 1) = 2a$?
   (b) If $a \in \mathbb{R}$ is given, and $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$\forall x \in \mathbb{R} \; f(x) = \begin{cases} ax^2 & \text{if } x \le 2 \\ bx + 5 & \text{if } x > 2 \end{cases}$$

   then for which value(s) of b (depending only on a) is f continuous at 2?

4. Find the following limits:

   (a) $\lim_{x \to 1} \dfrac{x^2 - 1}{x^2 - 2}$
   (b) $\lim_{x \to 1} \dfrac{x^2 - 1}{x^2 - x}$
   (c) $\lim_{x \to 2} \dfrac{x^4 - 16}{x^2 - x - 2}$
   (d) $\lim_{h \to 0} \dfrac{(x+h)^2 - x^2}{h}$
   (e) $\lim_{x \to 0} \dfrac{1}{x(x+1)} + \dfrac{1}{x(x-1)}$

   (Hints: For (d), h is the limit variable, so x should be treated as a constant. For (e), use a common denominator.)

5. Prove Theorem 3.26 by proving the following statement by induction on $n \in \mathbb{N}$: for all $a_0, a_1, \ldots, a_n \in \mathbb{R}$, the function which takes $x \in \mathbb{R}$ to $\sum_{i=0}^{n} a_i x^i$ is continuous everywhere. Also prove Corollary 3.27.

6. This exercise outlines a proof of the Limit Comparison Theorem (Theorem 3.34), which is stated more formally in the next section.
   (a) Suppose that f is a real function defined near a satisfying $f(x) \ge 0$ for all x near a (except possibly not at a). Prove that if $\lim_{x \to a} f(x)$ exists, then $\lim_{x \to a} f(x) \ge 0$. (Hint: Do a proof by contradiction.)
   (b) Now let f be a real function with the property that for some constant M, $f(x) \ge M$ for all x near a. Show that if $\lim_{x \to a} f(x)$ exists, then $\lim_{x \to a} f(x) \ge M$. (Hint: Use part (a) on a different function.)
   (c) Now suppose f and g are real functions defined near a satisfying $f(x) \ge g(x)$ for all x near a. Prove that if $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ both exist, then $\lim_{x \to a} f(x) \ge \lim_{x \to a} g(x)$.

For the next problems, we need some new definitions. Suppose that f is a real function whose domain contains some interval of the form (a, ∞) (let's say such functions are defined *near infinity*). Thus, $f(x)$ is defined for "very large" values of x. How do we express a statement saying that the values of $f(x)$ approach a limit L as x becomes "very large"? We want to say that $f(x)$ can be made as close to L as desired by taking x "large enough". The notion of "large enough" is formalized by saying there is some cutoff value M so that for all $x \geq M$, $f(x)$ is close enough to L. (See Figure 3.7).

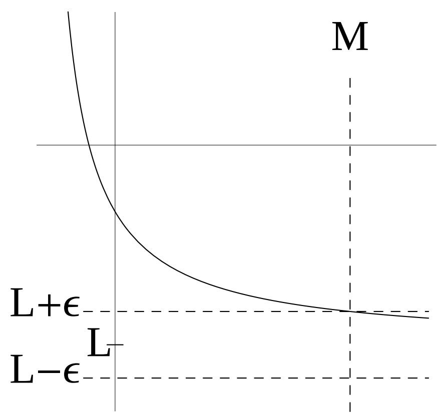

Figure 3.7: A limit L as $x \to \infty$

Definition 3.30. Let $L \in \mathbb{R}$ be given, and let f be a real function defined near infinity. We say that $f(x)$ *approaches* L *as* x *approaches infinity*, also written in any of the following ways

$$\lim_{x \to \infty} f(x) = L \qquad f(x) \to L \text{ as } x \to \infty \qquad f(x) \underset{x \to \infty}{\longrightarrow} L$$

if the following holds:

$$\forall \epsilon > 0 \; \exists M > 0 \; \forall x \in \mathrm{dom}(f) \; (x \geq M \to |f(x) - L| < \epsilon)$$

Similarly, we can define limits that go to $-\infty$ by talking about taking values of x which are "negative enough":

Definition 3.31. Let $L \in \mathbb{R}$ be given, and let f be a real function defined on an interval of the form $(-\infty, a)$ (i.e. f is defined *near negative infinity*).

We say that $f(x)$ *approaches* L *as* x *approaches negative infinity*, also written in any of the following ways

$$\lim_{x\to-\infty} f(x) = L \qquad f(x) \to L \text{ as } x \to -\infty \qquad f(x) \underset{x\to-\infty}{\longrightarrow} L$$

if the following holds:

$$\forall \epsilon > 0\ \exists M < 0\ \forall x \in \mathrm{dom}(f)\ (x \le M \to |f(x) - L| < \epsilon)$$

7. Use the definitions above to prove $\displaystyle\lim_{x\to\infty} \frac{1}{x} = 0$ and $\displaystyle\lim_{x\to-\infty} \frac{1}{x} = 0$.

8. Use the definitions above to prove that $\dfrac{1}{x} + \dfrac{2}{x^2} \to 0$ as $x \to \infty$. (Hint: You'll want to do a form of $\epsilon/2$ argument. If you handle $1/x$ and $2/x^2$ separately, each of them has their own "cutoff" M. How do you pick one cutoff that works for both?)

9. State analagous versions of the constant multiple, sum, difference, product, and quotient laws for limits as $x \to \infty$ or $x \to -\infty$. Describe informally what changes need to be made to the original proofs. (Hint: Instead of taking a min of "δ" values, the previous exercise suggests you should do something different with multiple "M" values.)

10. (a) Find $\displaystyle\lim_{x\to\infty} \frac{x^2}{x^2+1} = 1$ by multiplying the numerator and denominator by $1/x^2$. (Also, use Exercise 3.4.7 to help).

    (b) Using a similar tactic to the last part, compute $\displaystyle\lim_{x\to\infty} \frac{3x^2+2x}{4-x^2}$.

    (c) Compute $\displaystyle\lim_{x\to\infty} \frac{x-5}{x^2+7}$.

For these next exercises, we consider functions where limits fail to exist, but they fail to exist in a special way. The function whose value at any $x \neq 0$ is $1/x$, for example, does not have a limit as $x \to 0$, and this is because its values become arbitrarily large as x gets close to 0 from the right. To describe this situation, we say that $1/x \to \infty$ as $x \to 0^+$. The formal definition of this uses a notion of "cutoff point" similar to the definition as $x \to \infty$, but this time we use the cutoff with the function outputs instead of with the function inputs:

Definition 3.32. Let $a \in \mathbb{R}$ be given, and let f be a real function defined near a. We say that $f(x)$ *approaches infinity as* x *approaches* a, also written in any of the following ways

$$\lim_{x\to a} f(x) = \infty \qquad f(x) \to \infty \text{ as } x \to a \qquad f(x) \underset{x\to a}{\longrightarrow} \infty$$

if the following holds:

$$\forall M > 0\ \exists \delta > 0\ \forall x \in \operatorname{dom}(f)\ (0 < |x-a| < \delta \to f(x) \geq M)$$

Thus, the idea is that no matter what cutoff M you specify for values of $f(x)$, eventually by taking x close enough to a with a control δ, the values of $f(x)$ surpass that cutoff. There is a similar definition for functions whose values become very negative:

Definition 3.33. Let $a \in \mathbb{R}$ be given, and let f be a real function defined near a. We say that $f(x)$ *approaches negative infinity as* x *approaches* a, also written in any of the following ways

$$\lim_{x\to a} f(x) = -\infty \qquad f(x) \to -\infty \text{ as } x \to a \qquad f(x) \underset{x\to a}{\longrightarrow} -\infty$$

if the following holds:

$$\forall M < 0\ \exists \delta > 0\ \forall x \in \operatorname{dom}(f)\ (0 < |x-a| < \delta \to f(x) \leq M)$$

Remark. We note that if $f(x) \to \infty$ or $f(x) \to -\infty$ as $x \to a$, we do NOT say that $f(x)$ has a limit as $x \to a$. However, saying that $f(x) \to \infty$ or $f(x) \to -\infty$ as $x \to a$ does tell us information about SPECIFICALLY why there fails to be a finite limit for $f(x)$ as $x \to a$, and this can be useful information.

11. (a) Formulate similar definitions for what it means to write $f(x) \to \infty$ or $f(x) \to -\infty$ as $x \to a^+$ or $x \to a^-$. These are one-sided versions of the definitions above.
    (b) Use your definition to prove that $1/x \to \infty$ as $x \to 0^+$ and $1/x \to -\infty$ as $x \to 0^-$.

12. Prove that $1/x^2 \to \infty$ as $x \to 0$.

13. (a) Formulate a definition for what it means to write $f(x) \to \infty$ as $x \to \infty$. (We suggest you combine the ideas from all the definitions you've seen in these exercises.) There are similar definitions using $-\infty$ instead of ∞, but you don't need to write them all.
    (b) Use your definition to prove that $x^2 \to \infty$ as $x \to \infty$.
    (c) Prove that for any $n \in \mathbb{N}^*$, $x^n \to \infty$ as $x \to \infty$. (It turns out you do not need to use n^{th} roots to prove this, nor do you need to use induction!)
    (d) Prove that $\displaystyle\lim_{x\to\infty} \frac{x^2 - 1}{3x + 1} = \infty$ by using techniques like those in Exercise 3.4.10. (Hint: Divide x from the top and bottom, so the new denominator approaches 3 as $x \to \infty$. Thus, when x is large, you can find a constant upper bound for this new denominator. Use this to make a lower bound for the entire fraction, where the lower bound goes to ∞.)

14. Prove that $\displaystyle\lim_{x\to 0} \frac{1}{x^2} + x = \infty$. (Hint: With an appropriate choice of δ, you can make sure that $x > -1$. If you want $1/x^2 + x$ to be greater than a cutoff M, how large do you want to make $1/x^2$? Does this suggest a value of δ?)

15. (a) Prove that for any $a, L \in \mathbb{R}$ and any real functions f and g defined near a, if $f(x) \to \infty$ and $g(x) \to L$ as $x \to a$, then $f(x) + g(x) \to \infty$ as $x \to a$. (Thus, we have essentially "$\infty + L = \infty$".)
    (b) Prove that if $f(x) \to \infty$ and $g(x) \to \infty$ as $x \to a$, then $f(x) + g(x) \to \infty$ as $x \to a$. ("$\infty + \infty = \infty$".)
    (c) For any $r \in \mathbb{R}$, find examples of real functions f and g defined near 0 for which $f(x) \to \infty$, $g(x) \to -\infty$, and $f(x) + g(x) \to r$ as $x \to 0$. (This shows that sums of the form "$\infty - \infty$" don't have any clear outcome, similar to 0/0 forms.)

3.5 More Limit Theorems

We've now shown how limits interact with addition, multiplication, and so forth. However, we still need some more techniques for computing limits of many functions which occur frequently. First, we'll analyze how limits and inequalities are related. This will lead us to a useful theorem, called the Squeeze Theorem, which helps us find limits by analyzing limits of lower and upper bounds. After that, we'll look at how limits work with function composition; this is especially important because many complicated functions are composites of simple functions.

Inequalities with Limits

Suppose that we have two real functions, f and g, which are both defined near a point a. Also, suppose f is smaller near a, i.e. $f(x) \leq g(x)$ for all x in a deleted open interval around a. It makes sense that if f and g have limits at a, then f should have the smaller limit. After all, if $f(x)$ is never larger than $g(x)$ when x is near a, then there's no reason that the situation should suddenly switch in the limit!

More precisely, we have the following theorem, whose proof is outlined in Exercise 3.4.6:

Theorem 3.34 (Limit Comparison Theorem). *Let $a \in \mathbb{R}$ be given, and suppose f and g are real functions defined near a. Assume that $f(x) \leq g(x)$ for all x near a, and also assume that the limits of $f(x)$ and $g(x)$ exist as $x \to a$. Then*

$$\lim_{x \to a} f(x) \leq \lim_{x \to a} g(x)$$

This theorem is often summarized by saying "limits preserve inequalities" or by saying "you can take limits in an inequality, if the limits exist."

It is worth noting, though, that the Limit Comparison Theorem requires the limits of each side of the inequality to exist. Thus, it cannot be used to prove a limit exists. To illustrate this, we offer the following example.

Example 3.35:

Consider the following complicated function $f : \mathbb{R} \to \mathbb{R}$:

$$\forall x \in \mathbb{R}\ f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ -x & \text{if } x \notin \mathbb{Q} \end{cases}$$

Because there are infinitely many rationals and infinitely many irrationals inside every interval, the values of f keep jumping between the curves $y = x$ and $y = -x$. Thus, it's impossible to reliably graph this function!

Suppose we consider positive values of x. As the values of $f(x)$ jump between x and $-x$, the distance jumped, $2x$, gets very small as x goes to 0. This gives us the idea that the function starts "settling down" near $x = 0$.

Another way to describe the "jumpiness" of the function f is to use inequalities. In our case, for $x > 0$, we have the very convenient inequalities

$$-x \le f(x) \le x$$

Now, if we can take the limit of both sides of an inequality, then we get

$$\lim_{x\to 0} -x \le \lim_{x\to 0} f(x) \le \lim_{x\to 0} x$$

This seems to show that $0 \le \lim_{x\to 0} f(x) \le 0$, and hence $f(x) \to 0$ as $x \to 0$. Unfortunately, this work is not justified, because the Limit Comparison Theorem can only be used IF ALL THE LIMITS ARE KNOWN TO EXIST. Since we don't yet know whether $\lim_{x\to 0} f(x)$ exists ahead of time, we're not allowed to write limits into the inequality. $\square$

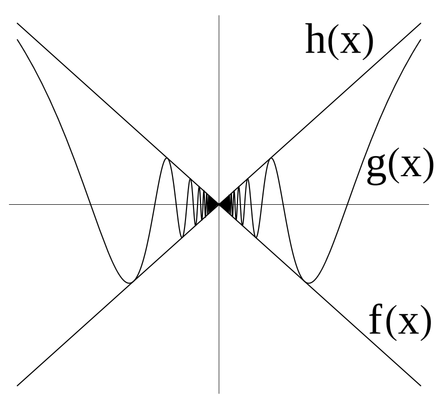

Figure 3.8: Squeezing $g(x)$ between $f(x)$ and $h(x)$ as $x \to a$

Despite this technical issue, however, the limit of $f(x)$ does exist as $x \to 0$, and that limit is 0. We obtain this by using a nice theorem called the Squeeze Theorem. It applies to functions that have upper and lower bounds with the same limit as $x \to a$ for some a. See Figure 3.8, where $g(x)$ is always between $f(x)$ and $h(x)$. As $x \to a$, both $f(x)$ and $h(x)$ approach the same value, so it looks like they "squeeze" $g(x)$ toward the same limit.

Theorem 3.36 (Squeeze Theorem). *Let $a \in \mathbb{R}$ be given, and let f, g, h be real functions all defined near a. Suppose that for all x near a, $f(x) \leq g(x) \leq h(x)$. If $\lim_{x \to a} f(x)$ and $\lim_{x \to a} h(x)$ both exist and equal the same number L, then $\lim_{x \to a} g(x) = L$ also.*

Strategy. First, let's perform a useful reduction. We subtract $f(x)$ from all sides of the given inequality to get

$$0 \leq g(x) - f(x) \leq h(x) - f(x)$$

By the limit laws and our assumptions on f and h, we know that $h(x) - f(x) \to 0$ as $x \to a$. We'd now like to show that $g(x) - f(x)$ goes to 0 as well as $x \to a$. (This is going to be easier than our original task, because we now have only two functions, $g(x) - f(x)$ and $h(x) - f(x)$, to consider instead of three.) Once we know $g(x) - f(x) \to 0$ as $x \to a$, we can add $f(x)$ again and use the limit laws to conclude $g(x) \to L$.

Thus, given $\epsilon > 0$, we'd like to find a control $\delta > 0$ on x which guarantees $|g(x) - f(x)| < \epsilon$. In this case, since $g(x) - f(x) \geq 0$, we just need to show $g(x) - f(x) < \epsilon$ (this is another handy benefit of the earlier reduction). Since $g(x) - f(x) \leq h(x) - f(x)$, it suffices to show $h(x) - f(x) < \epsilon$ if x is close enough to a. Since $h(x) - f(x) \to 0$, we just need to use the definition of limit for $h(x) - f(x)$, and we'll be done.

Proof. Let a, L, f, g, h be given as described in the theorem. We first prove that $g(x) - f(x) \to 0$ as $x \to a$. Since g and f are defined near a, so is $g - f$. Now, let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for all $x \in \mathrm{dom}(g - f)$,

$$0 < |x - a| < \delta \to |g(x) - f(x)| < \epsilon$$

Because $f(x)$ and $h(x)$ both approach L as $x \to a$, we have $h(x) - f(x) \to 0$ as $x \to a$ by the limit laws. Thus, we may choose some $\delta > 0$ such that

$$\forall x \in \mathrm{dom}(h - f)\ (0 < |x - a| < \delta \to |h(x) - f(x)| < \epsilon)$$

Furthermore, because $f(x) \leq g(x) \leq h(x)$ for all x near a, δ can be chosen small enough to also guarantee that for all $x \in \mathbb{R}$ satisfying $0 < |x - a| < \delta$, $f(x) \leq g(x) \leq h(x)$, so $h(x) - f(x) = |h(x) - f(x)| < \epsilon$.

Also, whenever $0 < |x - a| < \delta$, we have

$$|g(x) - f(x)| = g(x) - f(x) < h(x) - f(x) = |h(x) - f(x)| < \epsilon$$

This finishes the proof that $g(x) - f(x) \to 0$ as $x \to a$.

To end the proof, since $g(x) - f(x)$ and $f(x)$ both have limits as $x \to a$, we use the limit laws to conclude that

$$\lim_{x \to a} g(x) = \lim_{x \to a} (g(x) - f(x)) + \lim_{x \to a} f(x) = 0 + L = L$$

■

With the Squeeze Theorem, we can finally finish the proof that for our "jumpy" function f from earlier, we have $f(x) \to 0$ as $x \to 0$. Since $-x \le f(x) \le x$ for all $x \in \mathbb{R}$, and both $-x$ and x approach 0 as $x \to 0$, the Squeeze Theorem immediately tells us $f(x) \to 0$ as well.

Example 3.37:

The Squeeze Theorem is often very useful with trigonometric functions because sin and cos always produce values between -1 and 1, so they are convenient for inequalities. As an example, let's show

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

Note that the product law for limits does NOT apply here because we don't know whether $\sin(1/x)$ has a limit as $x \to 0$. (The issue is that $1/x$ has no limit as $x \to 0$. See Exercise 3.6.8 for more details.) Without the product law, our next best bet is to find bounds for $x \sin(1/x)$ and use the Squeeze Theorem. This seems promising, because for any $x \ne 0$, $-1 \le \sin(1/x) \le 1$.

When $x > 0$, we can multiply x throughout in the inequalities and get

$$-x \le x \sin(1/x) \le x \text{ when } x > 0$$

When $x < 0$, multiplying throughout flips the inequalities, giving

$$x \le x \sin(1/x) \le -x \text{ when } x < 0$$

Since $-x$ and x both approach 0 as $x \to 0$, one-sided modifications of the Squeeze Theorem show that $x \sin(1/x) \to 0$ both as $x \to 0^+$ and as $x \to 0^-$.

An alternate way to handle this, without having two cases for the sign of x, is to instead note that $|\sin(1/x)| \le 1$, and therefore

$$0 \le |x \sin(1/x)| \le |x|$$

Since the absolute-value function is continuous everywhere (the only case which isn't obvious is when $x = 0$, and that is handled in Exercise 3.4.2.(a)), we have $|x| \to 0$ as $x \to 0$. By the Squeeze Theorem, this means that $|x \sin(1/x)| \to 0$. From there, Exercise 3.6.1 shows that $x \sin(1/x) \to 0$. $\square$

Example 3.38:

The Squeeze Theorem is also very useful for establishing some basic properties of the sine and cosine functions. We'll use it to prove a very important limit in calculus:

$$\lim_{h \to 0} \frac{\sin h}{h} = 1$$

We will make use of this limit when we study derivatives in the next chapter.

To prove it, consider the picture in Figure 3.9. This picture shows a unit

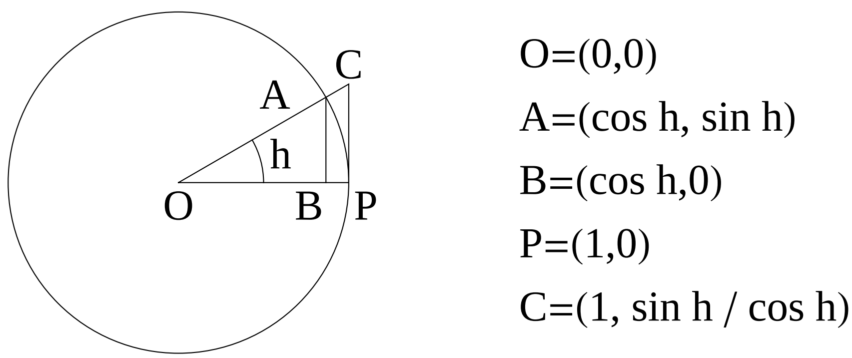

Figure 3.9: An angle of measure h in a unit circle

circle (a circle of radius 1) whose center is at the origin $(0, 0)$. We have also marked an acute angle h on the circle (measured counterclockwise from the segment OP), and the point A is placed on the circle so that the angle AOP has measure h. (Note that h is also the length of the arc from A to P when we measure h in radians.) Furthermore, the segment OA is extended with the same slope until it hits a point C with x-coordinate 1. Since the slope of the segment OA is $\sin h / \cos h$, the slope of OC is also $\sin h / \cos h$, which tells us the y-coordinate of C.

Remark. The notions of "area", "length of a curve", and "angle measure" are being used informally here. There are formal ways to define all these concepts (such as by using a system of axioms, as we did in Chapter 2 for the real numbers), but the work involved is long, complicated, and is not particularly interesting for the problem at hand. Similarly, we are not going to introduce a formal definition of sin and cos, although we may occasionally point out alternative ways to introduce the trigonometric functions. Upper-level real analysis courses cover more formal approaches to studying area and length.

Anyway, returning to the task at hand, the picture makes it clear that the length of AB is less than the arc length from A to P: hence $\sin h < h$ (when $h < \pi/2$, i.e. for acute angles). However, it seems that h is not too much larger than $\sin h$. In fact, if we decrease the angle h, the difference between the segment and the arc seems to diminish, so we suspect that $\sin h \approx h$ when h is small. This leads us to guess that $(\sin h)/h \to 1$ as $h \to 0^+$.

To verify this one-sided limit more formally, let's figure out some useful inequalities from the picture. We'll compare the areas of different figures. First, the triangle OAB is a right triangle with leg lengths $\sin h$ and $\cos h$, so its area is $(\sin h)(\cos h)/2$. Next, the circular sector OAP has area $h/2$ (this is because the whole circle has area π, and the whole circle is a sector with angle 2π, so the ratio of area to angle is $1/2$ for any sector). The picture shows that sector OAP has a bigger area than triangle OAB does. Lastly, the triangle OCP similarly is a right triangle with area $(\sin h)/(2\cos h)$, and this area is bigger than that of sector OAP. Thus,

$$\frac{(\sin h)(\cos h)}{2} < \frac{h}{2} < \frac{\sin h}{2\cos h}$$

By dividing through by $(\sin h)/2$ (which is positive), we get

$$\cos h < \frac{h}{\sin h} < \frac{1}{\cos h}$$

Lastly, by taking reciprocals, the inequalities flip (as all numbers involved are positive) and we get

$$\cos h < \frac{\sin h}{h} < \frac{1}{\cos h}$$

These inequalities are valid for any h with $0 < h < \pi/2$, so we may use them for the Squeeze Theorem. It turns out that cos is continuous everywhere (a proof of this is outlined in Exercise 3.6.10), so $\cos h \to \cos 0 = 1$ as $h \to 0^+$. Thus, $\cos h$ and $1/\cos h$ both approach 1 as $h \to 0^+$, and the Squeeze Theorem shows $(\sin h)/h \to 1$ as $h \to 0^+$.

How do we handle the case when $h \to 0^-$? Since sin is an odd function, we have $\sin(-h) = -\sin h$, so thus $(\sin(-h))/(-h) = (\sin h)/h$ for any $h \neq 0$. Also, cos is even, so $\cos(-h) = \cos h$. Thus, the same inequalities above work if $-\pi/2 < h < 0$, and the Squeeze Theorem shows that $(\sin h)/h \to 1$ as $h \to 0^-$ as well. $\square$

Composition

Recall that if f and g are functions, then their composite $g \circ f$ satisfies $\text{dom}(g \circ f) = \{x \in \text{dom}(f) \mid f(x) \in \text{dom}(g)\}$ and for any $x \in \text{dom}(g \circ f)$, $(g \circ f)(x) = g(f(x))$. Many functions can be expressed as composites of simpler functions, so it is important to have a way to deal with composition in limits.

We try the following approach: Suppose we'd like to find $\lim\limits_{x \to a} g(f(x))$. If f has a limit, say $f(x) \to L$ as $x \to a$, then intuitively $f(x)$ is close to L when x is close to a. Thus, if we let $y = f(x)$, we're interested to know what happens to $g(y)$ as $y \to L$, i.e. we'd like to find $\lim\limits_{y \to L} g(y)$. This leads us to the guess that

$$\lim_{x \to a} g(f(x)) = \lim_{y \to L} g(y) \qquad \text{where } L = \lim_{x \to a} f(x)$$

Let's look at an example of this in action.

Example 3.39:

Suppose that $g(x) = x^2$ and $f(x) = 2x + 1$ for all $x \in \mathbb{R}$. Thus, $g(f(x)) = (2x + 1)^2$. Since $g(f(x))$ also equals $4x^2 + 4x + 1$, which is a polynomial, we already know $g \circ f$ is continuous everywhere. However, in order to demonstrate how composition affects limits, let's try and use the definition of limit to prove that $\lim\limits_{x \to 0} (2x + 1)^2 = \lim\limits_{y \to 1} y^2$ (so in this case, $a = 0$ and $L = 1$). We've already found that $\lim\limits_{y \to 1} y^2 = 1$.

Let $\epsilon > 0$ be given. We'd like to find how close x should be to 0 in order to guarantee $|(2x + 1)^2 - 1| < \epsilon$. Let $y = 2x + 1$, so we'd like to guarantee $|y^2 - 1| < \epsilon$. (The idea here is we'd like to focus on one function at a time, so we consider $g(y)$ first, and we'll replace y with $f(x)$ later.) From the proof done in Example 3.21, we see that $|y^2 - 1| < \epsilon$ whenever $0 < |y - 1| < \delta_1$, where $\delta_1 = \min\{1/2, 2\epsilon/5\}$. In fact, because $y^2 = 1$ when $y = 1$, we actually have $|y^2 - 1| < \epsilon$ when $|y - 1| < \delta_1$ (i.e. we can allow y to be 1).

Now, we replace y with $f(x)$, so we see that $|(2x+1)^2 - 1| < \epsilon$ when $|(2x+1)-1| < \delta_1$. Now, what control δ on x will guarantee $|(2x+1)-1| < \delta_1$? Since $|(2x+1) - 1| = |2x| = 2|x|$, we can choose $\delta = \delta_1/2$. Hence, when $0 < |x| < \delta$, we have $|(2x+1) - 1| < \delta_1$ and thus $|(2x+1)^2 - 1| < \epsilon$, so we are done.

Note the key tactic here was the use of the number δ_1. This told us how much to control the input to the outer function g, but it also told us how accurate we needed the inner function f to be. Basically, one function's "ϵ" is another function's "δ". $\square$

The previous example shows our guess at a composition law seems quite plausible. However, unfortunately, our guess at a rule for limits of compositions doesn't work in all cases. Suppose $a = 0$, and consider the following two functions $f, g : \mathbb{R} \to \mathbb{R}$, whose values at any $x \in \mathbb{R}$ are

$$f(x) = 0 \qquad g(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

Here, for all $x \in \mathbb{R}$, $g(f(x)) = g(0) = 1$, so $g(f(x)) \to 1$ as $x \to 0$. However, $f(x) \to 0$ as $x \to 0$ (i.e. $L = 0$), and $g(y) \to 0$ as $y \to 0$. Thus, our guess is wrong for these functions. (Also, see Exercise 3.2.8.)

Why did our guess fail for these functions? The limit $\lim_{y \to 0} g(y)$ ignores the value of g at 0. However, as $f(x)$ approaches 0 when $x \to 0$, $f(x)$ IS the value 0, as opposed to just being near 0. Thus, only values of $y = 0$ are considered, never values of y that are in a deleted open interval around 0. This becomes a problem because $g(0)$ isn't the same as $\lim_{y \to 0} g(y)$, i.e. g is discontinuous at 0.

The next theorem we'll present, called the Composition Limit Theorem (abbreviated as the CLT), fixes our guess by requiring continuity of g at L:

Theorem 3.40 (Composition Limit Theorem). *Let $a, L \in \mathbb{R}$ be given, and let f and g be real functions where $\lim_{x \to a} f(x) = L$ and g is continuous at L. Then $\lim_{x \to a} g(f(x)) = g(L)$. In other words,*

$$\lim_{x \to a} g(f(x)) = g\left(\lim_{x \to a} f(x)\right)$$

provided g is continuous at $\lim_{x \to a} f(x)$. This is sometimes described by saying "you can move a limit inside of a continuous function".

Strategy. If $\epsilon > 0$ is given, we want to find how close x should be to a to guarantee that $|g(f(x)) - g(L)| < \epsilon$. We know that $g(y) \to g(L)$ as $y \to L$ by continuity, i.e. $g(y)$ gets close to $g(L)$ when the input, y, is close to L. More precisely, there is some control $\delta_1 > 0$ on y which makes $|g(y) - g(L)| < \epsilon$ when $|y - L| < \delta_1$ (recall the "$0 <$" in the limit definition is not needed for continuity).

In our situation, however, the input y to g is $f(x)$, so we'd like to get $|f(x) - L| < \delta_1$. This is where we use the assumption that $f(x) \to L$ as $x \to a$. This means that by picking a control $\delta_2 > 0$ on x, we can guarantee $f(x)$ is close enough to L, more precisely $|f(x) - L| < \delta_1$. Thus, δ_2 is the control we want to use for x to guarantee $|g(f(x)) - g(L)| < \epsilon$.

Proof. Let a, L, f, g be given as described in the theorem. For now, let's suppose that $g \circ f$ is defined near a without proof. (We'll address this non-trivial technical detail in the remarks following this theorem.) Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for all $x \in \text{dom}(g \circ f)$,

$$(0 < |x - a| < \delta \to |g(f(x)) - g(L)| < \epsilon)$$

By the continuity of g at L, we may choose $\delta_1 > 0$ so that

$$\forall y \in \text{dom}(g)\ (|y - L| < \delta_1 \to |g(y) - g(L)| < \epsilon)$$

Because $f(x) \to L$ as $x \to a$, we may choose $\delta > 0$ so that

$$\forall x \in \text{dom}(f)\ (0 < |x - a| < \delta \to |f(x) - L| < \delta_1)$$

Therefore, for any $x \in \text{dom}(g \circ f)$ with $0 < |x - a| < \delta$, we have $x \in \text{dom}(f)$, so $|f(x) - L| < \delta_1$. Letting $y = f(x)$, because $x \in \text{dom}(g \circ f)$, we have $f(x) \in \text{dom}(g)$, so from $|y - L| < \delta_1$, we obtain $|g(y) - g(L)| < \epsilon$. Thus, $|g(f(x)) - g(L)| < \epsilon$, as desired. ■

Remark. The CLT proof skips one step: it doesn't prove that $g \circ f$ is defined near a. This is fairly tricky to show since $\text{dom}(g \circ f)$ depends on $\text{dom}(g)$ and $\text{dom}(f)$ in a complicated way. (In contrast, $\text{dom}(f + g)$, $\text{dom}(fg)$, and $\text{dom}(f/g)$ were all pretty simple to express in terms of $\text{dom}(f)$ and $\text{dom}(g)$.) We merely give an outline here, since the proof that $g \circ f$ is defined near a is very similar to the CLT proof we just covered.

First, since g is continuous at L, there is some $\delta_1 > 0$ such that $|y - L| < \delta_1$ implies that $y \in \text{dom}(g)$. (See Exercise 3.2.3.) Next, as in the proof of the CLT above, we find some $\delta > 0$ so that $0 < |x-a| < \delta$ implies $|y-L| < \delta_1$. Putting these statements together, we find that $g \circ f$ is defined on $(a-\delta, a+\delta) - \{a\}$.

Many limits become much easier with the CLT. For instance, we compute

$$\lim_{x\to 1}(x^2+1)^6 = \left(\lim_{x\to 1}(x^2+1)\right)^6 = (1^2+1)^6 = 2^6 = 64$$

without having to expand out the entire sixth-degree polynomial or having to use the product law multiple times. In this calculation, we use the fact that the function which takes y to y^6 for all $y \in \mathbb{R}$ is continuous everywhere (so in particular, it is continuous at 2).

Another useful function for compositions is the square root function, which was shown to be continuous on its domain in Exercise 3.2.12 (except at 0, where it is only continuous from the right, but the CLT can be adapted to still work as long as the inner function is nonnegative). Thus, we can use this function in the CLT to compute limits like the following:

$$\begin{aligned}
\lim_{x\to 2}\sqrt{3x^2-1} &= \sqrt{\lim_{x\to 2}(3x^2-1)} && \text{CLT} \\
&= \sqrt{3(2)^2-1} && \text{Polynomials are continuous.} \\
&= \sqrt{11}
\end{aligned}$$

Example 3.41:

Another important technique for dealing with $0/0$ forms is called *rationalization*. Consider the following example:

$$\lim_{x\to 1}\frac{\sqrt{x}-1}{x-1}$$

This limit is in $0/0$ form, so we can't use the quotient law. However, when we multiply and divide by $\sqrt{x}+1$, we get

$$\lim_{x\to 1}\frac{(\sqrt{x}-1)(\sqrt{x}+1)}{(x-1)(\sqrt{x}+1)} = \lim_{x\to 1}\frac{(x-1)}{(x-1)(\sqrt{x}+1)} = \frac{1}{\sqrt{1}+1} = \frac{1}{2}$$

In general, when something of the form $\sqrt{f(x)} - g(x)$ appears in a limit as $x \to a$, and $\sqrt{f(x)} \neq g(x)$ when $x \neq a$, we can multiply and divide by $\sqrt{f(x)} + g(x)$ to get

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

This is called *rationalizing* the expression $\sqrt{f(x)} - g(x)$. Rationalization is frequently useful when the byproduct $f(x) - (g(x))^2$ cancels with something else. The value $\sqrt{f(x)} + g(x)$ which is multiplied and divided is sometimes called the *conjugate* of $\sqrt{f(x)} + g(x)$.

As another example, it's certainly possible to rationalize square roots in a denominator:

$$\begin{aligned}
\lim_{x \to 0} \frac{x^3}{\sqrt{x+4} - 2} &= \lim_{x \to 0} \frac{x^3(\sqrt{x+4} + 2)}{(\sqrt{x+4} - 2)(\sqrt{x+4} + 2)} \\
&= \lim_{x \to 0} \frac{x^3(\sqrt{x+4} + 2)}{(x+4) - (2^2)} = \lim_{x \to 0} \frac{x^3(\sqrt{x+4} + 2)}{x} \\
&= \lim_{x \to 0} x^2(\sqrt{x+4} + 2) = 0^2(\sqrt{0+4} + 2) = 0
\end{aligned}$$

□

3.6 Exercises

1. If $a \in \mathbb{R}$ is given, and f is a real function such that $\lim_{x \to a} |f(x)| = 0$, prove that $\lim_{x \to a} f(x) = 0$. (Hint: Use Lemma 3.2.)

2. Prove that for any $a, h \in \mathbb{R}$ and any real function f defined near $a + h$, $\lim_{x \to a+h} f(x)$ exists iff $\lim_{x \to a} f(x+h)$ exists. If both of these quantities exist, then show they are the same number.

3. Calculate the following limits, justifying all steps with an appropriate theorem or limit law:

   (a) $\displaystyle\lim_{x \to 1} \frac{\sqrt{x+1}}{2 - \sqrt{3x - 2}}$

   (b) $\displaystyle\lim_{x \to 0^+} \sqrt{x + \sqrt{x^4 + 1}}$

4. Find the following limits:

   (a) $\displaystyle\lim_{x \to 0} \frac{1 - \sqrt{1 - x^2}}{x^2}$

   (b) $\displaystyle\lim_{x \to 2} \frac{\sqrt{x+7} - 3}{x^2 - 3x + 2}$

   (c) $\displaystyle\lim_{x \to 0} \frac{\sqrt{x+1} - 1}{\sqrt{x+4} - 2}$

5. Let $a \in \mathbb{R}$ be given, and let f and g be real functions such that f is continuous at a and $\lim_{x\to\infty} g(x) = a$. (For the definition of a limit as $x \to \infty$, see the exercises in Section 3.4 near Definition 3.30.)
   (a) Prove that the composition $f \circ g$ is defined near infinity.
   (b) Prove that $\lim_{x\to\infty} f(g(x)) = f(a)$. (This is a variant of the CLT.)
   (c) Use the previous parts and the tactic from Exercise 3.4.10 to find $\lim_{x\to\infty} \sqrt{\frac{x^3+x}{4x^3+1}}$.
   (d) Find $\lim_{x\to\infty} \frac{x+2}{\sqrt{9x^2-1}}$. (Hint: Recall $1/x = \sqrt{1/x^2}$ when $x > 0$.)

6. Find the following limits. You may use Exercise 3.6.2 if necessary.
   (a) $\lim_{x\to 3} \frac{\sin(x-3)}{x-3}$
   (b) $\lim_{x\to 0} \frac{\sin(5x)}{\sin(7x)}$
   (c) $\lim_{x\to 1} \frac{\sin(x^2-1)}{(x-1)}$

   (Hint for (b) and (c): Multiply the top and bottom by something to get two simpler fractions.)

7. Prove that $\lim_{x\to 0} \frac{\cos x - 1}{x} = 0$ in each of the following two ways:
   (a) By using the half-angle identity $2\sin^2 x = 1 - \cos(2x)$.
   (b) By multiplying the top and bottom by $\cos x + 1$ and using an identity.

8. Let $f : \mathbb{R} - \{0\} \to \mathbb{R}$ be defined by $f(x) = \sin(1/x)$ for all $x \neq 0$. The graph of f oscillates, but unlike the graph of sin, the oscillations become steeper and steeper as x gets close to 0.

   Prove that $f(x)$ has no limit as $x \to 0$. This shows that although the values of f are always between -1 and 1, the function oscillates too wildly when x is near 0 (this is sometimes called a discontinuity by *oscillation*). (Hint: Show that no δ works for the accuracy $\epsilon = 1$ by analyzing the places where $f(x) = 1$ or where $f(x) = -1$.)

9. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined before the Squeeze Theorem:

$$\forall x \in \mathbb{R}\ f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ -x & \text{if } x \notin \mathbb{Q} \end{cases}$$

   We showed using the Squeeze Theorem that f is continuous at 0. Prove that for all $x \neq 0$, f is NOT continuous at x. Thus, f is a function with exactly one continuity point.

10. The following exercise outlines a method for proving that sin and cos are continuous everywhere. Recall that our proof that $(\sin h)/h \to 1$ as $h \to 0$ relied on the assumption that cos is continuous at 0.
    (a) The picture in Figure 3.9 showed that $\sin h < h$ for all $0 < h < \pi/2$. Use this to prove that $\sin h \to 0$ as $h \to 0$. (Thus, sin is continuous at 0.)
    (b) Next, use the identity $\sin^2 x + \cos^2 x = 1$, true for all $x \in \mathbb{R}$, to prove that $\cos h \to 1$ as $h \to 0$. (Thus, cos is continuous at 0.)
    (c) Lastly, use the sum identities

$$\sin(x + h) = (\sin x)(\cos h) + (\cos x)(\sin h)$$

    and

$$\cos(x + h) = (\cos x)(\cos h) - (\sin x)(\sin h)$$

    (valid for all $x, h \in \mathbb{R}$) to prove that sin and cos are continuous everywhere. (Exercise 3.6.2 might be useful.)

3.7 Some Unusual Functions

In this section, we'll look at some more unusual functions and how they behave when you take limits. We do this for a couple reasons. First, we consider these examples as good exercises of proof technique with limits. Second, and more importantly, these examples help us keep in mind a much larger variety of functions than those we frequently use. Many functions you've seen in practice are continuous everywhere they are defined, and they have only a few places where they don't exist (the polynomials, rational functions, and trigonometric functions fall into this category, except that tan and cot have a periodic set of discontinuities). Our examples will all be defined everywhere but will have more unusual sets of continuity points. In fact, it is largely due to examples like these that mathematicians need formal definitions of limit and continuity, so that results can still be proven even when studying very non-intuitive functions.

Examples Using Characteristic Functions

To help explain our first examples, we introduce the following:

Definition 3.42. If $A \subseteq \mathbb{R}$, then the *characteristic function* of A, denoted $\chi_A : \mathbb{R} \to \mathbb{R}$ (the Greek letter is "chi"), is defined by

$$\forall x \in \mathbb{R}\ \chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

In some sense, χ_A is like an on-off switch: it is "on" with value 1 when its input is in A, and it is "off" with value 0 when its input is not in A. One famous special case is when $A = [0, \infty)$, so that $\chi_{[0,\infty)}(x) = 1$ when $x \geq 0$ and $\chi_{[0,\infty)}(x) = 0$ when $x < 0$. $\chi_{[0,\infty)}$ is frequently called the *Heaviside function* and is denoted by H.

A more interesting example is the function $\chi_{\mathbb{Q}}$, because both $\mathbb{Q}$ and $\mathbb{R} - \mathbb{Q}$ are dense (see Exercises 2.9.4 and 2.9.6). This means, in particular, that every open interval contains rational numbers and also contains irrational numbers. This causes the graph of $\chi_{\mathbb{Q}}$ to continually keep "jumping" between $y = 0$ and $y = 1$. Using this idea, in Exercise 3.2.14, it was shown that $\chi_{\mathbb{Q}}$ has no points of continuity at all. In fact, for any $a \in \mathbb{R}$, as $x \to a$, $\chi_{\mathbb{Q}}(x)$ has no limit! (Also, see Exercise 3.8.2.)

We can also show that the "jumpy" function $f : \mathbb{R} \to \mathbb{R}$ from the last section, defined by

$$\forall x \in \mathbb{R}\ f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ -x & \text{if } x \notin \mathbb{Q} \end{cases}$$

can be defined in terms of $\chi_{\mathbb{Q}}$. Note that $f(x) + x$ is either $2x$ when $x \in \mathbb{Q}$, or it is 0 when $x \notin \mathbb{Q}$. It follows that $f(x) + x = 2x\chi_{\mathbb{Q}}(x)$ for all $x \in \mathbb{R}$. Let's define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x\chi_{\mathbb{Q}}(x)$ for all $x \in \mathbb{R}$, so we get

$$f(x) = 2g(x) - x \qquad \text{and} \qquad g(x) = \frac{f(x) + x}{2}$$

Since polynomials are continuous everywhere, the two equations above imply that f and g are continuous at exactly the same points. Combining this with the last section and Exercise 3.6.9, we see that g is continuous only at 0 and has no limits anywhere else. Note that in contrast to $\chi_{\mathbb{Q}}$, the graph of g keeps jumping between $y = x$ and $y = 0$, so when x approaches 0, an argument with the Squeeze Theorem shows that g "settles down" and becomes continuous at 0.

Example 3.43:

We have now seen an example of a function with no continuity points and an example of a function g with one continuity point. Let's make a function with two continuity points.

Let $h : \mathbb{R} \to \mathbb{R}$ be defined as follows:

$$\forall x \in \mathbb{R}\ h(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ x & \text{if } x \in \mathbb{Q} \text{ and } x < 1 \\ 2 - x & \text{otherwise} \end{cases}$$

When $x < 1$, $h(x) = g(x)$, i.e. h keeps jumping between $y = x$ and $y = 0$. The "jumpiness" settles down when $x = 0$, and so $x = 0$ is a point of continuity for h. However, for $x \geq 1$, h starts jumping between $y = 2 - x$ and $y = 0$, i.e. $h(x) = (2 - x)\chi_{\mathbb{Q}}(x)$. Here, the jumpiness settles down when $2 - x = 0$, i.e. when $x = 2$.

The previous paragraph is quite informal, but it can be formalized using the definition of limit and the Squeeze Theorem (see Exercise 3.8.3). In addition, you can show more generally in Exercise 3.8.4 that ANY finite set of real numbers is the set of continuity points for some function. (The function h is one such function corresponding to the set $\{0, 2\}$.) $\square$

The Ruler Function

This next function is also defined separately for rational inputs and for irrational inputs, but it exhibits much more complicated behavior.

The *ruler function* (there are plenty of other names for this function) is the function $r : \mathbb{R} \to \mathbb{R}$ defined as follows:

$$\forall x \in \mathbb{R}\ r(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ \dfrac{1}{q} & \text{if } x \in \mathbb{Q} \text{ and } x = p/q \text{ in lowest terms} \end{cases}$$

Here, "lowest terms" means that p and q have no factors in common and that q is positive. (Thus, $-12/10$ is $(-6)/5$ in lowest terms, 0 is $0/1$ in lowest terms, and so forth.)

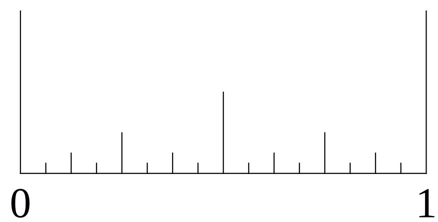

Figure 3.10: A partial picture of the ruler function on $[0, 1]$

A rough picture of r on rational inputs is given in Figure 3.10. This picture shows the values of $r(x)$ for $x = 0, 1/16, 2/16$, and so on up to 1 by drawing a line with height $r(x)$ from the axis. The markings resemble the lines on a ruler, hence the name "ruler function". (However, a more complete picture of the function would also show the values of $r(1/3)$, $r(2/3)$, and plenty more points, in addition to also showing that $r(x) = 0$ for all $x \notin \mathbb{Q}$.)

What makes the ruler function different from the previous examples? The previous examples consisted of functions whose values jump between two continuous functions ($\chi_{\mathbb{Q}}$ jumps between $y = 0$ and $y = 1$, and the function g from earlier jumps from $y = 0$ to $y = x$). While the ruler function does jump between two possibilities, depending on whether its input is rational or not, the larger possibility does NOT come from a continuous curve. For instance, if we consider the function which sends rational numbers p/q in lowest terms to $1/q$, then this function sends $1/2$ to $1/2$, but it sends all other rational numbers nearby to values which are at most $1/3$.

To see why this property makes such a difference, let's look at r only on the interval $[0, 1]$. There are two points where $r(x) = 1$: $x = 0$ and $x = 1$. After that, there is only one point with $r(x) = 1/2$: $x = 1/2$. There are two points with $r(x) = 1/3$: $x = 1/3$ and $x = 2/3$. In general, for any $n \in \mathbb{N}^*$, there are at most n points in $[0, 1]$ for which $r(x) = 1/n$. It follows that only finitely many values of x in $[0, 1]$ satisfying $r(x) \geq 1/n$. This suggests that, with "few" (i.e. finitely many) exceptions, the ruler function takes values close to 0, since $1/n$ gets as close to 0 as we'd like when n is large enough. This is proven formally in the following theorem:

Theorem 3.44. *For all $a \in \mathbb{R}$, the ruler function r satisfies $\lim_{x \to a} r(x) = 0$. Therefore, the ruler function is continuous precisely at the irrational numbers.*

Strategy. If $\epsilon > 0$, we want to find a control $\delta > 0$ on x which guarantees that $|r(x)| < \epsilon$. Hence, we ask the opposite question: which points $x \in \mathbb{R}$ do NOT satisfy $|r(x)| < \epsilon$? We need to take δ small enough so that if $|r(x)| \geq \epsilon$, then we do not have $|x - a| < \delta$.

Any such x is rational of the lowest-terms form p/q with $1/q \geq \epsilon$. Since only finitely many possible values of $q \in \mathbb{N}^*$ satisfy $1/q \geq \epsilon$, and there are only finitely many points near a of the form p/q for each q (there are infinitely many points overall, but only finitely many are within a short distance of a), there are only finitely many points x that we DON'T want to have within distance δ of a. Thus, for each x with $|r(x)| \geq \epsilon$, we want δ to be smaller than $|x - a|$. This suggests we take δ to be something like the minimum of $|x - a|$ for finitely many values of x. (This is valid because all finite sets have minimum elements.)

Proof. Let $a, \epsilon \in \mathbb{R}$ with $\epsilon > 0$ be given. We wish to find $\delta \in \mathbb{R}$ with $\delta > 0$ and so that

$$\forall x \in \mathbb{R}\ (0 < |x - a| < \delta \rightarrow |r(x)| < \epsilon)$$

Let's analyze the set S defined as

$$S = \{x \in \mathbb{R} \mid x \in ((a-1, a+1) - \{a\}) \text{ and } |r(x)| \geq \epsilon\}$$

Since $r(x) = 0$ for all $x \notin \mathbb{Q}$, every member of S is rational. If $x \in S$ and $x = p/q$ in lowest terms, then $1/q \geq \epsilon$. Therefore, $q \leq 1/\epsilon$.

For each $q \in \mathbb{N}^*$ with $q \leq 1/\epsilon$, there are at most $2q$ different $x \in S$ with $r(x) = 1/q$ (this is because any two such points have distance at least $1/q$, so at most $2q$ of them fit in the interval $(a-1, a+1)$ of length 2). Therefore, as every member x of S has the form p/q with $q \leq 1/\epsilon$, the size of S is at most

$$\sum_{q=1}^{\lfloor 1/\epsilon \rfloor} 2q \leq 2\left(\frac{1}{\epsilon}\right)^2$$

(note that the summation consists of at most $1/\epsilon$ terms, each at most $2(1/\epsilon)$). Hence, S is a finite set.

Now, choose δ as

$$\delta = \frac{1}{2}\min\{1, \min\{|x-a| \mid x \in S\}\}$$

(if S is empty, then we take $\delta = 1/2$). Because S is finite, the minima are all well-defined, and because $a \notin S$, we have $\delta > 0$.

Let $x \in \mathbb{R}$ be given with $0 < |x-a| < \delta$. As $\delta < 1$, this shows that $0 < |x-a| < 1$ and hence $x \in ((a-1, a+1) - \{a\})$. We must have $x \notin S$, or otherwise we would have $|x-a| < \delta < |x-a|$ (because $x \in S$), which is a contradiction. Therefore, as $x \in ((a-1, a+1) - \{a\})$ and $x \notin S$, we have $|r(x)| < \epsilon$, as desired. ■

We have now seen that the ruler function is continuous precisely at the irrational numbers. One natural question to ask is: is there a function defined on all of $\mathbb{R}$ whose continuity points are precisely the rational numbers? Interestingly, it is possible to prove that no such function exists, but the proof is beyond the scope of this book. (The proof is sometimes covered in upper-level analysis courses.)

Enumerating $\mathbb{Q}$

In order to prove the ruler function has limit 0 everywhere, we found ourselves arguing that a specific set S of rational numbers is finite. To do this, we found ourselves listing the possible members of S in a very methodical way, in order of increasing denominators. Since the members of S were at most distance 1 from a, and the denominators were bounded by $1/\epsilon$, we found that S had only finitely many members.

Now, $\mathbb{Q}$ is certainly not finite. However, we can still list out the elements of $\mathbb{Q}$ methodically. More formally, let's now show that $\mathbb{Q}$ can be *enumerated*, meaning that we can list the members of $\mathbb{Q}$ in an infinite sequence r_1, r_2, r_3, and so forth, with no value being repeated. (Equivalently, the function which maps $n \in \mathbb{N}^*$ to $r_n \in \mathbb{Q}$ is a bijection from $\mathbb{N}^*$ to $\mathbb{Q}$.)[5] We will need this enumeration for our last example in this section.

[5]In general, whenever there exists a bijection from $\mathbb{N}^*$ to a set S, we say that S is a *countably infinite* set, because it can be similarly enumerated. We will not have need to talk about countably infinite sets in general, but we will use the fact that $\mathbb{Q}$ is countably infinite several times in this book.

To start our sequence, we first list all the rational numbers which can be written in lowest terms p/q where $|p|, |q| \le 1$. There are only three such fractions:

$$\frac{-1}{1}, \frac{0}{1}, \frac{1}{1}$$

These will be the first three rational numbers in our sequence: $r_1 = -1$, $r_2 = 0$, and $r_3 = 1$. Next, we list the rational numbers with lowest-terms form p/q which haven't already been listed and which satisfy $|p|, |q| \le 2$:

$$\frac{-2}{1}, \frac{-1}{2}, \frac{1}{2}, \frac{2}{1}$$

This makes the next four numbers in the sequence, r_4 through r_7. After this, we list the rationals of the form p/q which haven't been previously listed and which satisfy $|p|, |q| \le 3$:

$$\frac{-3}{1}, \frac{-3}{2}, \frac{-2}{3}, \frac{-1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{3}{2}, \frac{3}{1}$$

This yields the next eight terms of the sequence, r_8 through r_{15}.

We continue this process, where for each $n \in \mathbb{N}^*$, in the n^{th} stage of the construction, we list all previously-unlisted lowest-terms fractions of the form p/q with $|p|, |q| \le n$. By design, we never list a rational number twice. Also, any rational number of the form p/q appears in our list by stage $\max\{|p|, |q|\}$). Thus, every member of $\mathbb{Q}$ appears exactly once in our final list.

Remark. In contrast, it is NOT possible to list all the irrational numbers in a sequence; any attempt at making a sequence of irrational numbers must leave some out. Equivalently, there is no surjection from $\mathbb{N}^*$ to $\mathbb{R} - \mathbb{Q}$. (In other words, $\mathbb{R} - \mathbb{Q}$ is not countably infinite.) For more details, search for the "diagonalization argument" by Cantor.

Definition of the Staircase Function

For our last major example of this section, we will build another function which has interesting behavior at rational inputs. However, this example will be built in a different manner from the earlier examples. We will build this function s by first defining a sequence of approximations g_1, g_2, g_3, and so forth. Informally, the graph of g_n will have "steps" in it located at the first

n rationals of our enumeration, r_1 through r_n. As n gets larger, the number of steps increases, and the graph of g_n becomes more jagged since it "jumps" at more rational numbers. For each $x \in \mathbb{R}$, we will take $s(x)$ to be the limit of $g_n(x)$ as $n \to \infty$.[6]

More precisely, we will build $s(x)$ as a supremum: $s(x)$ will be $\sup\{g_n(x) \mid n \in \mathbb{N}^*\}$. In order for this supremum to have the properties we want, we will guarantee two important things about the $g_n(x)$ sequence. The first guarantee is that g_n grows as n grows, i.e. for each $n \in \mathbb{N}$ and $x \in \mathbb{R}$, $g_{n+1}(x) \geq g_n(x)$. The second guarantee is that our "steps" will become smaller as n gets larger to ensure that $\{g_n(x) \mid n \in \mathbb{N}\}$ remains bounded. In essence, for each $x \in \mathbb{R}$, the values of $g_n(x)$ keep rising as n grows larger, until eventually they hit the limit value $s(x)$.

Now, we introduce a formal definition of what we want a "step" to be. For each $i \in \mathbb{N}^*$, we choose the i^{th} step to be the function taking $x \in \mathbb{R}$ to $(1/2)^i H(x - r_i)$, where H is the Heaviside function mentioned in Definition 3.42. This function equals 0 when $x - r_i < 0$, i.e. when $x < r_i$, and it equals $(1/2)^i$ when $x \geq r_i$. Essentially, this function "turns on" with value $(1/2)^i$ once x gets as large as r_i, i.e. its graph looks like it jumps upward by 2^{-i} at $x = r_i$.

Now that we have introduced what a step is, we define our n^{th} approximation $g_n : \mathbb{R} \to \mathbb{R}$ for $n \in \mathbb{N}^*$ as the sum of the first n steps:

$$\forall x \in \mathbb{R}\ g_n(x) = \sum_{i=1}^{n} \left(\frac{1}{2}\right)^i H(x - r_i)$$

(summation notation using $\sum$ was defined in Example 3.28). Thus, the n steps in g_n, at the first n rationals, have sizes ranging from $1/2$ to 2^{-n}. Note that $g_n(x)$ grows as n grows, since $g_{n+1}(x) - g_n(x) = (1/2)^{n+1} H(x - r_{n+1})$ (all other terms from the sums cancel), and this difference is nonnegative. In essence, as we go from n to $n+1$, we throw in one more step of size $2^{-(n+1)}$ at $x = r_{n+1}$.

Remark. This $(n+1)^{\text{st}}$ step doesn't have to appear to the right of all the previously-added steps! That would only happen if r_{n+1} were greater than all of r_1 through r_n. In general, for an arbitrary $n \in \mathbb{N}^*$, there's very little we can guarantee about where r_{n+1} lies compared to all the previous r_i values.

[6] Since n can only take values in $\mathbb{N}^*$, this notion of limit is not quite the same as the definition of a limit of $\lim\limits_{x\to\infty}$ from Definition 3.30 in the exercises of Section 3.4. We will revisit this type of limit more formally in Chapters 9 and 10.

It may be helpful to think of the graph of g_n as a stack of n blocks placed on top of each other, where the blocks have sizes from $1/2$ to 2^{-n}. The blocks are shifted so that the left edge of the i^{th} block lies at r_i, for each i from 1 to n, and blocks extend infinitely far to the right. When we add the $(n+1)^{\text{st}}$ step, we slide the $(n+1)^{\text{st}}$ block into place, and this may require us to lift other blocks upward to make room for the $(n+1)^{\text{st}}$ block.

At this point, we have fulfilled one of our guarantees: we made g_n grow when n grows. Now we address the second guarantee: how large can $g_n(x)$ get when $x \in \mathbb{R}$ and $n \in \mathbb{N}^*$? To answer this, we know that for each $x \in \mathbb{R}$ and each $i \in \mathbb{N}^*$, $H(x - r_i)$ is either 0 or 1. Therefore the i^{th} step is either 0 or 2^{-i}, and thus $g_n(x)$ is somewhere from 0 to $1/2 + 1/4 + \cdots + 2^{-n}$. We use Exercise 1.9.2 to bound this sum:

$$\begin{aligned}
\sum_{i=1}^{n} \left(\frac{1}{2}\right)^i &= \left(1 + \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^n\right) - 1 \\
&= \frac{\left(\frac{1}{2}\right)^{n+1} - 1}{\frac{1}{2} - 1} - 1 = \frac{1 - \left(\frac{1}{2}\right)^{n+1}}{1 - \frac{1}{2}} - 1 \\
&< \frac{1}{1 - \frac{1}{2}} - 1 = 1
\end{aligned}$$

Therefore, $g_n(x) < 1$.

Therefore, for all $x \in \mathbb{R}$, $\{g_n(x) \mid n \in \mathbb{N}^*\}$ is nonempty and bounded above by 1. We may now define the *staircase function* $s : \mathbb{R} \to \mathbb{R}$ by

$$\forall x \in \mathbb{R}\ s(x) = \sup\{g_n(x) \mid n \in \mathbb{N}^*\}$$

Our work above shows that $0 \le s(x) \le 1$ for all $x \in \mathbb{R}$. In fact, you can prove in Exercise 3.8.9 that $s(x)$ never equals 0 or 1, although the range does have infimum 0 and supremum 1. In other words, the range gets arbitrarily close to 0 and 1.

Some Properties Of The Staircase Function

The staircase function s has an important property that none of the earlier examples in this section have; the graph of s never falls as you move to the right. To be more precise, we introduce the following definitions: introduce the following definitions:

Definition 3.45. Let f be a real function, and let $I \subseteq \text{dom}(f)$ be given. We say that f is *increasing on I* if, for all $x, y \in I$ with $x \le y$, we have $f(x) \le f(y)$. If we have the stronger property that when $x, y \in I$ satisfy $x < y$, we also have $f(x) < f(y)$, we say f is *strictly increasing on I*.

Similarly, f is *decreasing on I* if $x \le y$ implies $f(x) \ge f(y)$ for all $x, y \in I$, and f is *strictly decreasing on I* if $x < y$ implies $f(x) > f(y)$ for all $x, y \in I$.

If f is increasing or decreasing on I, it is called *monotone on I*, and f is *strictly monotone on I* if f is either strictly increasing or strictly decreasing. In essence, a monotone function never "changes direction" on I.

In all of these definitions, we will often omit the phrase "on I" when I is $\text{dom}(f)$, e.g. we will just say "f is (strictly) increasing".

Example 3.46:

Many of the continuous functions with which you are familiar are monotone or can be broken into monotone pieces. For example, when $n \in \mathbb{N}^*$ is odd, the function mapping $x \in \mathbb{R}$ to x^n is strictly increasing. Also, if $n \in \mathbb{N}^*$ is even, the function mapping $x \in \mathbb{R}$ to x^n is strictly decreasing on the domain $(-\infty, 0]$ and strictly increasing on the domain $[0, \infty)$. (See Exercise 3.8.6.(a).) However, $\chi_{\mathbb{Q}}$ and the ruler function are not monotone on ANY interval, because their values keep jumping between 0 and some positive numbers infinitely often in any interval. □

We claim that the staircase function s is increasing. Informally, this result is proven in stages, corresponding to the process we used to create s. First, we will show that each step is an increasing function. Second, we will show that each g_n is increasing, since it is a sum of increasing steps. Lastly, Exercise 3.8.8 shows that a supremum of increasing functions is also increasing.

To help with the first two steps, we have the following intuitive lemma:

Lemma 3.47. *Suppose that f and g are increasing functions, and let $a, c \in \mathbb{R}$ be given with $a \ge 0$. Then the functions $f + g$ and af are increasing. Also, the function which takes $x \in \mathbb{R}$ to $f(x - c)$ (provided that $x - c \in \text{dom}(f)$) is also increasing. In other words, we may add increasing functions, multiply increasing functions by nonnegative constants, or shift increasing functions horizontally to yield increasing functions.*

Strategy. The proof of this lemma really just consists of using the definition of increasing. The only subtle point is that we need $a \ge 0$ to show that af is increasing, because multiplying by a nonnegative number does not flip an inequality.

Proof. Let a, c, f, g be given as described in the theorem. Let $x, y \in \mathbb{R}$ be given with $x \le y$.

First, if $x, y \in \text{dom}(f)$, then $f(x) \le f(y)$ because f is increasing. Similarly, if $x, y \in \text{dom}(g)$, then $g(x) \le g(y)$. Thus, if $x, y \in \text{dom}(f+g) = \text{dom}(f) \cap \text{dom}(g)$, then $(f+g)(x) = f(x)+g(x) \le f(y)+g(y) = (f+g)(y)$, showing that $f+g$ is increasing.

Second, if $x, y \in \text{dom}(f)$, then multiplying a to both sides of $f(x) \le f(y)$ does not flip the inequality because $a \ge 0$. We thus obtain $af(x) \le af(y)$, showing af is increasing.

Lastly, suppose $x-c, y-c \in \text{dom}(f)$. Then $x \le y$ implies $x-c \le y-c$, and therefore $f(x-c) \le f(y-c)$ because f is increasing. ∎

Remark. There is an analogous lemma which shows that adding decreasing functions, multiplying decreasing functions by nonnegative constants, and shifting decreasing functions still yield decreasing functions. The proof is almost exactly the same.

Now, to start our proof that s is increasing, it is easy to show that H is increasing (but not strictly increasing). Thus, by the lemma, $H(x-r_i)$ is also increasing for each i, and so is $2^{-i}H(x-r_i)$. This means each step is increasing. By the lemma again, g_n is a sum of n steps, so it is also increasing. For the last step, see Exercise 3.8.8.

In fact, although none of the g_n functions are strictly increasing (they are constant in between their steps), s IS strictly increasing! Intuitively, this happens because whenever $x, y \in \mathbb{R}$ are distinct, there is some rational number between x and y, so some step fits in between x and y. More formally, the proof that s is strictly increasing is outlined in Exercise 3.8.10.

Another interesting property of s that we'll mention here is that, like the ruler function, s is continuous at the irrational numbers and discontinuous at the rational numbers. Most of the work of this proof is left to Exercise 3.8.11, but we can still mention the basic motivation here. Each g_n is the sum of steps, where the i^{th} step is discontinuous at r_i. This suggests that g_n is discontinuous precisely at the first n rationals r_1 through r_n, so as n gets larger, g_n is discontinuous at more rationals. Eventually, in the limit as $n \to \infty$, s is discontinuous at all the rationals.

Monotonicity and One-Sided Limits

Montone functions are often quite well-behaved compared to other functions. Intuitively, a monotone function can't oscillate. In the worst case, a monotone function can only "jump" upwards or downwards. For instance, the floor function $\lfloor x \rfloor$ jumps up by 1 unit at each integer. On the other hand, the staircase function's graph has jumps at each rational number; see Exercise 3.8.11.

To make this idea more precise, we prove that all one-sided limits exist for a monotone function. The proof is a useful application of using suprema and infima to define limits. We will only state and prove the result for increasing functions, because the corresponding result for decreasing functions follows easily by regarding decreasing functions as the negations of increasing functions.

Theorem 3.48. *Let f be an increasing function on an open interval I. Then for all $a \in I$, $\lim_{x \to a^-} f(x)$ and $\lim_{y \to a^+} f(y)$ exist, and*

$$\lim_{x \to a^-} f(x) \leq f(a) \leq \lim_{y \to a^+} f(y)$$

The difference

$$\lim_{y \to a^+} f(y) - \lim_{x \to a^-} f(x)$$

is called the gap *of f at a. (With this terminology, we note that the two-sided limit $\lim_{x \to a} f(x)$ exists iff the gap at a is 0, in which case the limit must be $f(a)$.)*

Strategy. Let's suppose that $a \in I$. What should $\lim_{x \to a^-} f(x)$ be? Since f is increasing, for any $x < a$, we have $f(x) \leq f(a)$, and furthermore as x gets larger, the values of $f(x)$ grow as well. Thus, the one-sided limit $\lim_{x \to a^-} f(x)$ should be an upper bound of the $f(x)$ values. (The function is "climbing upwards" toward some peak, which represents the limit.) Furthermore, the limit should be as close to the $f(x)$ values as possible; after all, the $f(x)$ values should "approach" the limit. This suggests that the limit should be the strictest upper bound of the $f(x)$ values, i.e. the supremum. Note that since $f(a)$ is an upper bound, the supremum will be at most $f(a)$.

However, how will we actually show formally that this is the one-sided limit? Let's say L^- is the sup of the $f(x)$ values; for any $\epsilon > 0$, we want to find how close x should be to a to guarantee $|f(x) - L^-| < \epsilon$. Since L^-

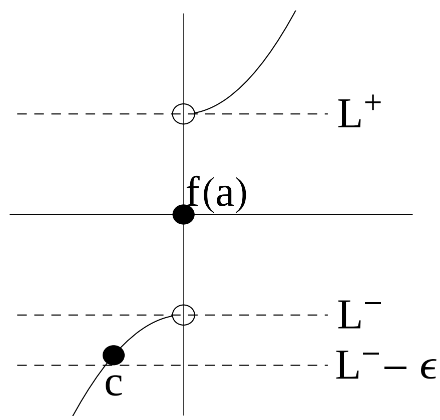

Figure 3.11: An increasing function f with one-sided limits to a

is an upper bound of the $f(x)$ values, we know $L^- \geq f(x)$, so we want to guarantee $L^- - f(x) < \epsilon$, i.e. $f(x) > L^- - \epsilon$. However, the value $L^- - \epsilon$ is smaller than the LEAST upper bound of the $f(x)$ values, so it is NOT an upper bound. Hence, there is some point $c < a$ where $f(c)$ is greater than the sup minus ϵ. As f is increasing, ALL values of x greater than c will have $f(x) \geq f(c) > L^- - \epsilon$. See Figure 3.11.

The one-sided limit from the right will use a similar idea. If we consider values y which are greater than a, then $f(a) \leq f(y)$, and the values of $f(y)$ decrease as y decreases. Thus, we want a lower bound instead, and the limit will be the greatest lower bound L^+ (the infimum). Since $f(a)$ is also a lower bound, and L^+ is the greatest lower bound, we have $L^+ \geq f(a)$.

Proof. Suppose f, I are given as described in the theorem, and let $a \in I$ be given. Therefore, as I is an open interval containing a, and f is defined on I, f is defined near a.

Let's define

$$S^- = \{f(x) \mid x \in I \text{ and } x < a\} \qquad S^+ = \{f(y) \mid y \in I \text{ and } y > a\}$$

Because I is open, it has no least or greatest member, so S^- and S^+ are both nonempty. Since f is increasing, whenever $x < a$, we have $f(x) \leq f(a)$, so S^- is bounded above by $f(a)$. Also, whenever $y > a$, we have $f(y) \geq f(a)$, so S^+ is bounded below by $f(a)$.

This shows that $\sup S^-$ and $\inf S^+$ both exist. Define $L^- = \sup S^-$ and $L^+ = \inf S^+$. Because $f(a)$ is an upper bound of S^- and a lower bound of

S^+, we have $L^- \leq f(a) \leq L^+$. We first prove that $\lim_{x \to a^-} f(x) = L^-$. Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for all $x \in I$,

$$0 < a - x < \delta \to |f(x) - L^-| < \epsilon$$

It is helpful to note that since L^- is an upper bound of S^-, and $f(x) \in S^-$ for each $x \in I$ less than a, we know $L^- \geq f(x)$, so $|f(x) - L^-| < \epsilon$ is equivalent to $L^- - f(x) < \epsilon$, which means the same as $f(x) > L^- - \epsilon$.

Because $L^- - \epsilon$ is NOT an upper bound of S^- (it is smaller than the least upper bound), there is some member of S^- larger than $L^- - \epsilon$, i.e. there is some $c \in I$ with $c < a$ and $f(c) > L^- - \epsilon$. Let $\delta = a - c$, so $\delta > 0$. Now suppose $x \in I$ is given with $0 < a - x < \delta$. Thus, $a - \delta < x < a$, so $c < x < a$ by the choice of δ. As f is increasing, $f(c) \leq f(x)$. Since $f(c) > L^- - \epsilon$ and $f(x) \geq f(c)$, we have $f(x) > L^- - \epsilon$, as desired. This proves the statement for the one-sided limit from the left.

For the other side, we claim that $\lim_{y \to a^+} f(y) = L^+$. Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that for all $y \in I$,

$$0 < y - a < \delta \to |f(y) - L^+| < \epsilon$$

As in the other case, since L^+ is a lower bound of S^+, and $f(y) \in S^+$ for all $y \in I$ greater than a, we have $L^+ \leq f(y)$, so $|f(y) - L^+| < \epsilon$ is equivalent to $f(y) - L^+ < \epsilon$, i.e. $f(y) < L^+ + \epsilon$.

Because $L^+ + \epsilon$ is not a lower bound of S^+, there is some $d \in I$ with $d > a$ and $f(d) < L^+ + \epsilon$. Let $\delta = d - a$, so $\delta > 0$. For all $y \in I$ with $0 < y - a < \delta$, we have $a < y < a + \delta$, so $a < y < d$ by the choice of δ. As f is increasing, $f(y) \leq f(d)$, and $f(d) < L^+ + \epsilon$. This shows $f(y) < L^+ + \epsilon$, as desired. ■

This theorem gives some insight as to why we call s the staircase function, instead of giving that name to the floor function (whose graph certainly looks like a staircase). First of all, the floor function already has a well-established name. Secondly, in Exercise 3.8.11, you can show that for each $n \in \mathbb{N}^*$, the staircase function has gap $1/2^n$ at r_n. Hence, unlike the floor function, s has a positive gap at each rational, so EVERY open interval contains gaps! However, the gap sizes approach 0 very quickly, so the staircase function remains bounded. In essence, the graph of s has more densely-packed tiny "stairs" than the graph of the floor function does. Unfortunately, this makes s very difficult to graph, although its approximations g_n can be graphed pretty easily.

3.8 Exercises

1. This exercise will introduce some simple calculations with characteristic functions. Let $A, B \subseteq \mathbb{R}$ be given.
   (a) Prove that $\chi_{\mathbb{R}-A} = 1 - \chi_A$.
   (b) Prove that if $A \subseteq B$, then $\chi_A(x) \leq \chi_B(x)$ for every $x \in \mathbb{R}$. Is the converse true? Either prove it is true, or give a counterexample.
   (c) Prove that $\chi_{A\cup B} = \max\{\chi_A, \chi_B\}$.
   (d) Prove that if A and B are disjoint, then $\chi_{A\cup B} = \chi_A + \chi_B$.
   (e) How can you write $\chi_{A\cap B}$ in terms of χ_A and χ_B? Prove that your answer is correct.
   (f) How can you write χ_{A-B} in terms of χ_A and χ_B? Prove that your answer is correct.

2. Let $A \subseteq \mathbb{R}$ be given.
   (a) Prove that for all $a \in \mathbb{R}$, $\lim_{x\to a} \chi_A(x) = 1$ iff A contains a deleted open interval around a.
   (b) Prove that for all $a \in \mathbb{R}$, $\lim_{x\to a} \chi_A(x) = 0$ iff $\mathbb{R}-A$ contains a deleted open interval around a. (Hint: The previous exercise might help.)
   (c) Prove that the previous two parts describe all possibilities for limits of χ_A. In other words, prove that if $a \in \mathbb{R}$ is given and $\lim_{x\to a} \chi_A(x)$ exists, then that limit must be 0 or 1.

3. Suppose that $q : \mathbb{R} \to \mathbb{R}$ is a function which is continuous everywhere. Prove that for all $a \in \mathbb{R}$, the function $q\chi_{\mathbb{Q}}$ is continuous at a iff $q(a) = 0$. As an example, the function in Example 3.43 is the special case where

$$q(x) = \begin{cases} x & \text{if } x < 1 \\ 2 - x & \text{if } x \geq 1 \end{cases}$$

   for all $x \in \mathbb{R}$.

   (Hint: If $q(a) = 0$, then the proof is pretty short and uses the Squeeze Theorem. If $q(a) \neq 0$, then WLOG we may suppose $q(a) > 0$, since the case where $q(a) < 0$ is handled by working with $-q$ instead of q. The continuity of q guarantees that for x close enough to a, $q(x)$ is between $q(a)/2$ and $3q(a)/2$. Use this idea to show that when $\epsilon = q(a)/2$, no choice of δ works for the function $q\chi_{\mathbb{Q}}$.)

4. For any finite $S \subseteq \mathbb{R}$, make a function $h_S : \mathbb{R} \to \mathbb{R}$ such that the set of continuity points of h_S is precisely S. Prove that your function h_S has this property. (Hint: Use Exercise 3.8.3.)

5. Consider a variant on the ruler function, $r_2 : \mathbb{R} \to \mathbb{R}$, defined as follows. If $x \in \mathbb{R}$ is irrational, then $r_2(x) = 0$. If $x \in \mathbb{Q}$ and has lowest-terms form p/q, then $r_2(x) = 1/q^2$.

   Prove that for all $a \in \mathbb{R}$, $\lim_{x \to a} r_2(x) = 0$. (Hint: It is possible to imitate the proof of Theorem 3.44, but there's a much quicker proof reusing what we know about the ruler function.)

6. (a) Prove that for any $n \in \mathbb{N}^*$, the n^{th} power function, which takes $x \in \mathbb{R}$ to x^n, is strictly increasing on $[0, \infty)$. (Hint: If $0 \le x < y$, then prove $x^n < y^n$ by induction on n. See Exercise 2.4.4.(d) as well.)
   (b) Prove by induction on $n \in \mathbb{N}^*$ that $n < 2^n$. This shows that the powers of 2 are unbounded above (because the natural numbers are unbounded above).

7. When defining the ruler function, we showed that for any $n \in \mathbb{N}^*$, $\sum_{i=1}^{n} \frac{1}{2^i} < 1$. This exercise establishes something stronger.
   (a) Prove that
   $$\sup \left\{ \sum_{i=1}^{n} \frac{1}{2^i} \mid n \in \mathbb{N}^* \right\} = 1$$
   (Hint: Exercise 1.9.2 gives the exact value of each sum in that set. How close do those sums get to 1? Exercise 3.8.6.(b) might help too.)
   (b) Use part (a) to prove that for each $\epsilon > 0$, there is some $N \in \mathbb{N}^*$ such that
   $$\sup \left\{ \sum_{i=N}^{n} \frac{1}{2^i} \mid n \in \mathbb{N}^* \right\} < \epsilon$$
   In essence, this is saying that if we throw out the first $N-1$ terms from the sums $\sum_{i=1}^{n} \frac{1}{2^i}$ for $n \in \mathbb{N}^*$, what remains can be made very small.

8. Suppose that S is a nonempty set of functions from $\mathbb{R}$ to $\mathbb{R}$, and every $g \in S$ is increasing.
   (a) If $\sup\{g(b) \mid g \in S\}$ exists for some $b \in \mathbb{R}$, then prove that $\sup\{g(a) \mid g \in S\}$ exists for every $a \in \mathbb{R}$ with $a \le b$. (Hint: If $a \le b$, then what is one number which serves as an upper bound for EVERY $g(a)$?)
   (b) Now assume that $\sup\{g(x) \mid g \in S\}$ exists for every $x \in \mathbb{R}$, and call it $f(x)$. Thus, f is a function from $\mathbb{R}$ to $\mathbb{R}$. Prove that f is increasing. (This implies that the staircase function s is increasing.)
   (c) Find an example of a set S of STRICTLY increasing functions such that $f(x) = \sup\{g(x) \mid g \in S\}$ exists for every $x \in \mathbb{R}$ but f is not strictly increasing.

9. (a) Prove that for all $x \in \mathbb{R}$, the staircase function s satisfies $0 < s(x) < 1$. (Hint: There must exist some rational numbers r_i, r_j with $r_i < x < r_j$. When n is large enough, can you come up with bounds on $g_n(x)$ based on i and j?)
   (b) Prove that for any $\epsilon > 0$, there exist $x_1, x_2 \in \mathbb{R}$ such that $s(x_1) < \epsilon$ and $s(x_2) > 1 - \epsilon$. Thus, $\inf(\operatorname{ran}(s)) = 0$ and $\sup(\operatorname{ran}(s)) = 1$. (Both parts of Exercise 3.8.7 should be useful.)

10. This exercise outlines a proof that the staircase function is strictly increasing.
    (a) Let $j \in \mathbb{N}^*$ be given, and suppose that $x \in \mathbb{R}$ is given with $x < r_j$. Prove that $g_n(x) + 1/2^j \le g_n(r_j)$ for all $n \in \mathbb{N}^*$ with $n \ge j$.
    (b) If j, x are given as in part (a), prove that $s(x) + 1/2^j \le s(r_j)$. (Hint: This proof uses a lot of the same ideas as Exercise 2.7.8.)
    (c) Use the previous parts to prove that whenever $x, y \in \mathbb{R}$ are given with $x < y$, then $s(x) < s(y)$.

11. The following exercise outlines a proof that the staircase function is continuous at each irrational number and has jump $1/2^j$ at r_j for each $j \in \mathbb{N}^*$.
    (a) Prove that for all $a \in \mathbb{R}$, then s is continuous from the right at a. (Hint: The main idea was used in the proof of Theorem 3.44, which chooses δ based on excluding a certain finite set of points. Use this together with Exercise 3.8.7.(b).)

(b) Next, prove that if $a \in \mathbb{R} - \mathbb{Q}$, then s is continuous from the left at a.

(c) Lastly, prove that for all $j \in \mathbb{N}^*$, $\lim_{x \to r_j^-} s(x) = s(r_j) - 1/2^j$. (Hint: If you try the proof of part (b) here, what makes the proof fail? Use that to adjust the proof of part (b) accordingly.)

3.9 Consequences of Continuity: Part 1

Up until now in this chapter, we have seen definitions of limits and continuity, we've proven some theorems about how to compute limits, and we've looked at a variety of examples. However, apart from drawing a few graphs and talking intuitively about "jumps" in a function and such, we haven't yet explained why continuity is a desirable property of a function. Therefore, the goal of the rest of this chapter is to present some important properties that continuous functions have. These properties are among the major reasons that the definition of continuity was created in the first place.

Our properties will be split into two sections. This section will introduce the Intermediate Value Theorem, which tells us roughly that the graph of a continuous function does not "skip over values". This theorem will also be applied to argue that certain continuous functions have inverses, and we will prove the continuity of those inverses as well. The second section will introduce the Extreme Value Theorem, which says that any continuous function attains maximum and minimum values when its domain is a closed interval. That section will also introduce a sharper version of continuity called uniform continuity, which will be especially useful when we study integrals.

Intermediate Value Theorem

The Intermediate Value Theorem (or the *IVT* for short) is essentially a formal way of stating "to draw a continuous function f, you never need to lift your pencil off the paper." Roughly, if you start drawing the graph of f from $x = a$ to $x = b$ (where $a, b \in \mathbb{R}$ with $a < b$), you won't skip over any value in between $f(a)$ and $f(b)$. For instance, take a look back at the speed function v from the introduction of this chapter, where $v(t)$ measures the speed of our car t seconds after the start of a trip. If $v(0) = 0$ and $v(10) = 60$ (roughly corresponding to accelerating up to highway speeds in 10 seconds), then at some time between 0 and 10 seconds, your speed must have been 20mph (which is, of course, the time when people behind you get impatient).

To prove the IVT, we'll first prove a special case of it called Bolzano's Theorem (which is an interesting theorem in its own right), and then we'll prove the IVT as a corollary. Bolzano's Theorem says that if a continuous function switches its sign as its input goes from a to b, then the function must cross the x-axis somewhere in between a and b.

Theorem 3.49 (Bolzano's Theorem). *Let f be a real function which is continuous on a closed interval $[a, b]$ (where $a < b$). If $f(a)$ and $f(b)$ have different signs, then there is some $c \in (a, b)$ with $f(c) = 0$.*

Remark. It turns out that we don't need two-sided continuity at the endpoints of the interval $[a, b]$: we only need continuity from the right at a and continuity from the left at b. When people say a function is *continuous on a closed interval* $[a, b]$, they frequently mean to say that the endpoints only need one-sided continuity. (In fact, f may only be defined on $[a, b]$, so a two-sided limit at a or at b wouldn't even make sense!)

Many proofs of results assuming continuity on the closed interval $[a, b]$ still work with minor modifications when we have only one-sided continuity at the endpoints. However, since the modifications of the proofs involve introducing special cases for a and b, and thus they are more difficult to read without being any more instructive, we will assume two-sided continuity at a and b to make life simpler.

Strategy. We know that $f(a)$ and $f(b)$ have different signs, so either $f(a) < 0 < f(b)$ or $f(b) < 0 < f(a)$. To make our work simpler, let's only handle the case when $f(a) < 0 < f(b)$. In the other case, we have $-f(a) < 0 < -f(b)$, and $-f$ is continuous on $[a, b]$, so we can use the argument for the first case with the function $-f$.

Let's imagine we're walking along the graph of f as we go from a to b. As $f(a) < 0$, we can think of this as starting our walk below the x-axis, and since $f(b) > 0$, eventually we end above the x-axis. Thus, we can't stay below the axis forever. A good informal question to ask is: how long can we walk before we have to stay above the axis? In other words, how large can x get before $f(x)$ has to be positive?

This leads us to consider something like "the largest x for which $f(x) < 0$" (i.e. x is the last point on the walk which is below the axis). Thus, let's let S be the set $\{x \in [a, b] \mid f(x) < 0\}$. However, S doesn't have to have a maximum element, even though S is bounded. Thus, we'll consider the next best thing: the supremum. This leads us to consider the guess that we can take c to be $\sup S$.

How can we show that $f(c) = 0$? We'll try a proof by contradiction, i.e. we'll show that $f(c) \neq 0$ is impossible. The main idea is we'd like to show that for all x sufficiently close to c, $f(x)$ and $f(c)$ have the same sign. Thus, if $f(c) < 0$, then $f(x) < 0$ for some x a little bigger than c, showing that there is a member of S bigger than c. This contradicts that c is an upper bound of S. On the other hand, if $f(c) > 0$, then $f(x) > 0$ for some x a little smaller than c, which suggests that x is a better upper bound for S than c is. This contradicts that c is a least upper bound for S.

(Note that the two cases each produce a contradiction: one contradicts that c is an upper bound, and one contradicts that c is "least". This occurs rather commonly in arguments with suprema.)

The last point to consider is: how can we show that $f(x)$ and $f(c)$ have the same sign if x is close to c? This is where we have to use the continuity of f. We can choose $\epsilon > 0$ to be small enough so that, for x close enough to c, $f(x)$ is at most ϵ away from $f(c)$, and this forces $f(x)$ and $f(c)$ to have the same sign. One such choice for ϵ is half the distance from $f(c)$ to 0, i.e. $|f(c)|/2$.

Proof. Let f, a, b be given as described. We may assume WLOG that $f(a) < 0 < f(b)$ (the other case follows by applying this argument to $-f$).

Let $S = \{x \in [a,b] \mid f(x) < 0\}$. Since $a \in S$ by assumption, and b is an upper bound of S (since $S \subseteq [a,b]$), S is nonempty and bounded above. Thus, S has a supremum. Define $c = \sup S$. Clearly $c \geq a$ since $a \in S$, and $c \leq b$ since b is an upper bound of S, so f is continuous at c. We'll show that $f(c) = 0$ by showing that $f(c) > 0$ and $f(c) < 0$ are each impossible.

For the first case, suppose $f(c) > 0$ for a contradiction. Let $\epsilon = |f(c)|/2$, so $\epsilon > 0$. By continuity of f at c, there is some $\delta > 0$ so that for all $x \in \mathrm{dom}(f)$, if $|x-c| < \delta$ then $|f(x)-f(c)| < |f(c)|/2$. It follows that for all $x \in (c-\delta, c+\delta) \cap \mathrm{dom}(f)$, $f(x)$ is between $f(c)/2$ and $3f(c)/2$, so $f(x) > 0$.

Therefore, no member of $(c-\delta, c]$ belongs to S. Since no value of x greater than c belongs to S (as c is an upper bound of S), this shows $c - \delta$ is an upper bound of S. However, this contradicts that c is the least upper bound. Thus, we cannot have $f(c) > 0$.

For the second case, suppose $f(c) < 0$ for a contradiction. By the same reasoning as the first case, there is some $\delta > 0$ so that for all $x \in \mathrm{dom}(f)$, if $x \in (c-\delta, c+\delta)$, then $|f(x) - f(c)| < |f(c)|/2$. Thus, for any $x \in (c-\delta, c+\delta) \cap \text{dom}(f)$, $f(x)$ is between $f(c)/2$ and $3f(c)/2$, so $f(x) < 0$.

In particular, when $x = c + (\delta/2)$, $f(x) < 0$ and hence $x \in S$. However, $x > c$, so this contradicts that c is an upper bound of S. Hence, we cannot have $f(c) < 0$ either.

Thus, since $f(c) > 0$ and $f(c) < 0$ are each impossible, we must have $f(c) = 0$. We already showed $c \in [a,b]$, and since $f(a) < 0 < f(b)$, we cannot have c equal to a or b, so $c \in (a,b)$. ■

From Bolzano's Theorem, the IVT follows quickly:

Corollary 3.50 (Intermediate Value Theorem). *Let f be a real function which is continuous on a closed interval $[a,b]$ (where $a < b$). For every value $k \in \mathbb{R}$ between $f(a)$ and $f(b)$, there is some $c \in (a,b)$ with $f(c) = k$.*

Strategy. Let's say f is continuous on $[a,b]$ and k is between $f(a)$ and $f(b)$. We'd like to find some value $c \in (a,b)$ with $f(c) = k$, or equivalenty, some $c \in (a,b)$ satisfying $f(c) - k = 0$. In other words, if g is the function satisfying $g(x) = f(x) - k$ for all $x \in \text{dom}(f)$, we're looking for a value $c \in (a,b)$ with $g(c) = 0$. Bolzano's Theorem is well-suited for this task, because g is a difference of continuous functions and is thus continuous.

Proof. Let f, a, b, k be given as described. We define a real function g with $\text{dom}(g) = \text{dom}(f)$ as follows: for all $x \in \text{dom}(f)$, $g(x) = f(x) - k$. Note that g is continuous on $[a,b]$ (as a difference of continuous functions) and that $g(a)$ and $g(b)$ have different signs because k is between $g(a)$ and $g(b)$. By Bolzano's Theorem applied to g, there is some $c \in (a,b)$ with $g(c) = 0$, i.e. $f(c) = k$. ■

Example 3.51:

The assumption of continuity in the ENTIRE interval $[a,b]$ is absolutely necessary for the IVT (or Bolzano's Theorem) to work. A single discontinuity, even at an endpoint of the interval, can make the theorem fail. As an example, consider the Heaviside function shifted down by $1/2$, i.e. we consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$\forall x \in \mathbb{R}\ f(x) = \begin{cases} 1/2 & \text{if } x \geq 0 \\ -1/2 & \text{if } x < 0 \end{cases}$$

When $a = -1$, $b = 1$, and $k = 0$, we have $f(-1) < 0$ and $f(1) > 0$, but there is no value $c \in (-1, 1)$ with $f(c) = 0$. Alternatively, if $a = -1$ and $b = 0$, then we still have $f(-1) < 0 < f(0)$ without having any value $c \in (-1, 0)$ satisfying $f(c) = 0$.

This shows that even a single discontinuity in $[a, b]$ (even at an endpoint) can make the IVT fail to apply. If you look at f on any interval which does not contain 0, then the IVT does apply. (However, this is trivial for f because f is constant on any interval not containing 0.) $\square$

Example 3.52:

Although Bolzano's Theorem only tells you that a zero of a function exists (it doesn't say how many there are, nor does it give more information about how to find them), repeated use of Bolzano's Theorem can help us find more information about zeroes. Let's demonstrate how we can approximate the zeroes of a cubic polynomial using Bolzano's Theorem.

Remark. A *zero* of a function f is an input x such that $f(x) = 0$. This is often called a root, but technically, a *root* is a solution to an equation. For example, we can either say "the polynomial $x^2 - 1$ has the two zeroes -1 and 1" or "the equation $x^2 - 1 = 0$ has the two roots -1 and 1". However, in common practice, the words "zero" and "root" are used interchangeably.

For example, consider the polynomial $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3 - 5x + 1$ for $x \in \mathbb{R}$. As a cubic polynomial, f has at most 3 zeroes. By taking a look at a graph, it seems that there should be zeroes in the intervals $(-3, -2)$, $(0, 1)$, and $(2, 3)$.

Bolzano's Theorem can confirm that these intervals do in fact contain zeroes as follows. Note that f is continuous everywhere. Also, $f(-3) = -11 < 0 < 3 = f(-2)$, so there is some $c_1 \in (-3, -2)$ with $f(c_1) = 0$. Next, $f(0) = 1 > 0 > -3 = f(1)$, so there is some $c_2 \in (0, 1)$ with $f(c_2) = 0$. Lastly, $f(2) = -1 < 0 < 13 = f(3)$, so there is some $c_3 \in (2, 3)$ with $f(c_3) = 0$.

Because the intervals we used in our applications of Bolzano's Theorem are each disjoint from the others, we know our three zeroes c_1, c_2, and c_3 are all distinct. As f has at most 3 zeroes, we now know the set of zeroes is precisely $\{c_1, c_2, c_3\}$. We sometimes say we've *isolated* the zeroes from one another (by finding a system of disjoint intervals, each containing a zero).

However, we can get better information about where the zeroes lie. Consider c_2 for instance: we currently only know that $c_2 \in (0, 1)$, which isn't a very accurate estimate for c_2. By looking at $x = 0.5$, the midpoint of $(0, 1)$, we see that $f(0.5) = -1.375$, so f changes sign on the interval $[0, 0.5]$ and must therefore have a zero in $(0, 0.5)$. Thus, as this zero must be c_2, we now have better information about c_2 than before. Trying this tactic again with the midpoint of $(0, 0.5)$, we find that $f(0.25) \approx -0.234$, so we find that actually $c_2 \in (0, 0.25)$.

This suggests the following useful algorithm for trying to find zeroes of continuous functions f:

1. Find initial values $a_0, b_0 \in \mathbb{R}$ where $f(a_0)$ and $f(b_0)$ have different signs (a graph usually helps for this). Also, we let $i = 0$ to start our algorithm (i will increase as the algorithm continues).
2. We now know there is a zero of f on the interval (a_i, b_i) (by the IVT). Let $c = (a_i + b_i)/2$, the midpoint of (a_i, b_i). If $f(c) = 0$, we have found our zero and may stop the algorithm.
3. Otherwise, we have two cases. If $f(c)$ has the same sign as $f(b_i)$, then let $a_{i+1} = a_i$ and $b_{i+1} = c$. Otherwise, if $f(c)$ has the same sign as $f(a_i)$, then let $a_{i+1} = c$ and $b_{i+1} = b_i$.
4. Increase i by one and go back to Step 2. We repeat this process until the distance $b_i - a_i$ is as small as we want it to be. (For instance, if you want to approximate the zero up to n decimal places of accuracy, you run this algorithm until $b_i - a_i < 10^{-n}$.)

As the distance $b_i - a_i$ gets cut in half each time Step 3 is performed, the distance $b_i - a_i$ gets close to 0 quite quickly. When we run this algorithm with the initial values of $a_0 = 0$ and $b_0 = 1$, it produces $a_1 = 0$ and $b_1 = 0.5$ in Step 3. The next time Step 3 is performed, it produces $a_2 = 0$ and $b_1 = 0.25$. The next time, we get $a_3 = 0.125$ and $b_3 = 0.25$. After more steps, we obtain $a_7 = 0.1953125$ and $b_7 = 0.203125$. Therefore, we now know that c_2, rounded to the first two decimal places, is 0.20. If we want more digits of accuracy, we need only run the algorithm longer.

This basic process of splitting our interval in half, finding out which half has our solution, and then continuing our search in a new interval of half the size, is a useful process called *binary search* or *bisection*. See Exercise 3.10.1 for more practice with isolating roots and using binary search. $\square$

n^{th} **Roots**

Back in Theorem 2.29, we claimed that every nonnegative number has a nonnegative square root, i.e. for each $a \in \mathbb{R}^+ \cup \{0\}$ there is some $x \in \mathbb{R}^+ \cup \{0\}$ with $x^2 = a$ (uniqueness of x was left as an exercise). Now that we have the Intermediate Value Theorem, we can make a rather simple proof of this theorem. In fact, we'll even generalize to n^{th} roots where $n \in \mathbb{N}^*$ as follows:

Theorem 3.53. *For all $a \geq 0$ and all $n \in \mathbb{N}^*$, there is a unique $x \geq 0$ with $x^n = a$. This value x is called the n^{th}* root of *a and is denoted either by $\sqrt[n]{a}$ or by $a^{1/n}$.*

Strategy. As is common with proving existence of unique objects, let's first show that an n^{th} root exists, and then we'll show uniqueness.

We'd like to find some $x \geq 0$ satisfying $x^n = a$. This is where the IVT becomes useful, because the n^{th} power function taking $x \in \mathbb{R}$ to x^n is continuous everywhere. This means we want to find values $a_1, b_1 \in \mathbb{R}$ such that $a_1^n < a < b_1^n$, and then it follows that for some $x \in (a_1, b_1)$, we have $x^n = a$.

To argue uniqueness, we note that the n^{th} power function is strictly increasing on $[0, \infty)$ (this was shown in Exercise 3.8.6.(a)). Thus, no two n^{th} powers of distinct numbers in $[0, \infty)$ are the same.

Proof. Let $a \geq 0$ and $n \in \mathbb{N}^*$ be given. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^n$ for all $x \in \mathbb{R}$. We have seen that f is continuous everywhere.

First, we prove existence of an n^{th} root. If $a = 0$, then we choose $x = 0$; since $0^n = 0$, we are done. Otherwise, we note that $f(0) = 0 < a$ and $f(1+a) = (1+a)^n \geq 1 + a > a$ (because $x^n \geq x$ whenever $x \geq 1$). Thus, by the IVT, there is some $x \in (0, 1+a)$ with $f(x) = a$, i.e. $x^n = a$ as desired.

Second, to prove uniqueness, note that f is strictly increasing on $[0, \infty)$. Thus, whenever $x, y \geq 0$ are distinct, say WLOG that $x < y$, then $x^n < y^n$. This means that no two n^{th} powers of nonnegative numbers are equal, so there can be at most one nonnegative n^{th} root of a. ■

The above theorem is really in some sense a definition and a theorem at the same time (a "definitheorem", if you will). This occurs commonly in mathematics, where a theorem asserts the existence and uniqueness of some object with a defining property.

How do we handle n^{th} roots of negative numbers? When n is even, we must have $x^n \geq 0$ for all $x \in \mathbb{R}$, so there is no such thing as an n^{th} root in $\mathbb{R}$ of a negative number. However, when n is odd, for every $a \in \mathbb{R}$ there is exactly one $x \in \mathbb{R}$ satisfying $x^n = a$. (When a is negative, that choice of x is precisely $-((-a)^{1/n})$.) That unique x is also called the n^{th} root of a, and the proof of its existence and uniqueness follows quickly from Theorem 3.53.

From n^{th} roots, you can make rational powers as follows. In Exercise 3.10.6, you will prove that when $a \geq 0$ and $m, n \in \mathbb{N}^*$, the quantities $(a^m)^{1/n}$ and $(a^{1/n})^m$ are the same (this is NOT the case when $a < 0$ for certain values of m and n). Furthermore, the quantity $(a^m)^{1/n}$ only depends on the ratio m/n, in the sense that if $p, q \in \mathbb{N}^*$ satisfy $m/n = p/q$, then $(a^m)^{1/n} = (a^p)^{1/q}$. Thus, we may define $a^{m/n}$ by $(a^m)^{1/n}$, and the usual algebraic laws of powers will hold. By also making the definitions $a^0 = 1$ and $a^{-y} = 1/(a^y)$ when $a, y > 0$, we obtain definitions of all rational powers (but this requires $a > 0$ as opposed to just $a \geq 0$, since negative powers involve reciprocals).

Note that we have NOT YET made any definition of what a^x means when x is irrational. We will have to wait until much later, when we cover exponential functions in Chapter 7, to have a proper treatment of the expression a^x for arbitrary $x \in \mathbb{R}$ and $a > 0$.

Continuity of Inverses

Although we've now defined n^{th} roots, we don't yet know whether the n^{th} root functions are continuous. None of our current limit laws apply for showing n^{th} roots are continuous, as we didn't build n^{th} roots out of the operations we've studied so far (adding, multiplying, composition, etc.). We need a new kind of proof if we want to show that the n^{th} roots are continuous.

Our main approach will be to look at the n^{th} root function as the inverse of the n^{th} power function (inverse functions were defined in Chapter 1). More precisely, when $n \in \mathbb{N}^*$, let's consider the n^{th} power function $f : [0, \infty) \to [0, \infty)$ defined by $f(x) = x^n$ for all $x \in \mathbb{R}$. We've seen that f is strictly increasing (because we are restricting ourselves to the domain $[0, \infty)$), hence it is injective. Furthermore, Theorem 3.53 proved that every $a \in [0, \infty)$ is in the range of f, namely $f(a^{1/n}) = (a^{1/n})^n = a$. Thus, f is also surjective, so f has an inverse function, the n^{th} root function taking $x \in [0, \infty)$ to $x^{1/n}$. (In fact, when n is odd, the n^{th} power function is also a strictly increasing bijection from $\mathbb{R}$ to $\mathbb{R}$.)

It turns out that reasoning like the proof of Theorem 3.53 can prove that any strictly increasing continuous function f on an interval $[a, b]$ is a bijection from $[a, b]$ to $[f(a), f(b)]$; see Exercise 3.10.8. Thus, there is an inverse function $g : [f(a), f(b)] \to [a, b]$. The graph of g is obtained from the graph of f by reflecting over the line $y = x$.

Because of this reflection property, and because we associate continuous functions with connected graphs, we suspect that g is also strictly increasing and continuous. Exercise 1.7.6 proved that g is strictly increasing. We now prove the continuity of g:

Theorem 3.54. *Let f be a real function which is continuous on a closed interval $[a, b]$. Assume that f is strictly increasing, so Exercise 3.10.8 ensures that f is a bijection from $[a, b]$ to $[f(a), f(b)]$. Let $g : [f(a), f(b)] \to [a, b]$ be the inverse of f, so g is also strictly increasing. Then g is continuous on $[f(a), f(b)]$.*

Strategy. We aim to take advantage of the symmetric relationship between f and g, as displayed in Figure 3.12; the graphs of f and g are reflections of one another over the dashed line. Let's say we're trying to show g is continuous at $d \in [f(a), f(b)]$. Since f and g are inverses, we know that we can write d as $f(c)$ for some unique $c \in [a, b]$. In fact, $c = g(d)$. Thus, on our figure, we draw two mirror-image points: one at (d, c) on the graph of g, and one at (c, d) on the graph of f.

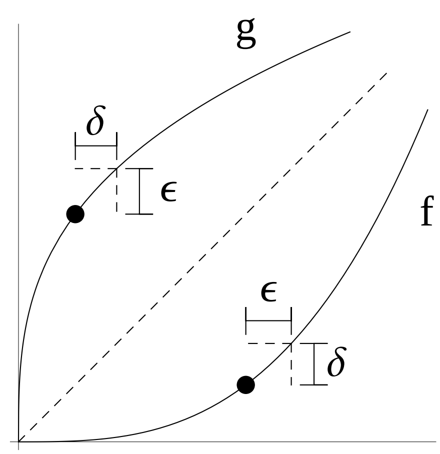

Figure 3.12: Two continuous inverses f and g with $f(c) = d$ and $g(d) = c$

Now, let's say $\epsilon > 0$ is given. For simplicity, we deal with the two one-sided limits of g separately. Let's look at the right-hand limit first. We want the graph of g to lie below the dashed horizontal line near (d, c), provided y stays close enough to d. (The dashed vertical line displays how close y should stay to d.) More precisely, we want to find some $\delta > 0$ so that $y < c + \delta$ implies $g(y) < g(c) + \epsilon$. We have labeled ϵ and δ on Figure 3.12.

Since f and g are mirror images, we can make mirror images of these dashed lines as well! When we do so, we find that ϵ now represents a HORIZONTAL distance for the graph of f, as labeled in the figure. Also, δ now represents a vertical distance for the graph of f. In fact, we can read off the value of δ from this picture: since the horizontal dashed line for f is at the height of $f(c+\epsilon)$, we see that δ should be $f(c+\epsilon) - d = f(c+\epsilon) - f(c)$.

Now, for any $x \in [a, b]$, we have $c < x < c + \epsilon$ iff $f(c) < f(x) < f(c+\epsilon)$ since f is strictly increasing. Note that $f(c+\epsilon) = f(c) + \delta$ by the choice of δ. Thus, $c < x < c+\epsilon$ iff $f(c) < f(x) < f(c) + \delta$. When we set $y = f(x)$ (so that $x = g(y)$) and $d = f(c)$ (so that $c = g(d)$), we find that $d < y < d+\delta$ iff $g(d) < g(y) < g(d) + \epsilon$, which is exactly what we wanted for the right-hand limit!

The left-hand limit works similarly, except that we choose $\delta = f(c) - f(c - \epsilon)$ instead (draw the appropriate dashed lines to see this).

Proof. Let f, a, b, g be given as described. Let $d \in [f(a), f(b)]$ be given. First, we prove that if $d < f(b)$, then g is continuous from the right at d, i.e. $g(y) \to g(d)$ as $y \to d^+$. Let $\epsilon > 0$ be given; we must find some $\delta > 0$ so that all $y \in (d, d+\delta)$ satisfy $|g(y) - g(d)| < \epsilon$. In fact, since g is strictly increasing, when $d < y$, we know $g(y) > g(d)$, so it will suffice to prove $g(y) < g(d) + \epsilon$.

Let's define $c = g(d)$; since $d < f(b)$, we have $c = g(d) < g(f(b)) = b$. WLOG, we may suppose ϵ is small enough to satisfy $c + \epsilon \le b$. Therefore, $f(c+\epsilon)$ is defined, and we choose

$$\delta = f(c+\epsilon) - f(c)$$

Note that $\delta > 0$ because f is strictly increasing.

To show that this choice of δ works for the right-hand limit, let $y \in \mathrm{dom}(g)$ be given satisfying $d < y < d + \delta$; we must show $g(d) < g(y) < g(d) + \epsilon$. Because $c = g(d)$ and f is the inverse of g, $f(c) = d$, our choice of δ implies

$$d + \delta = f(c) + (f(c+\epsilon) - f(c)) = f(c+\epsilon)$$

Therefore, $f(c) < y < f(c+\epsilon)$. Because g is strictly increasing, we may apply it to all the parts of this inequality and conclude that $g(f(c)) < g(y) < g(f(c+\epsilon))$. Because g is the inverse of f, this means $c < g(y) < c + \epsilon$, i.e. $g(d) < g(y) < g(d) + \epsilon$, as desired. This proves g is continuous from the right.

Next, we prove that if $d > f(a)$, then g is continuous from the left at d. Let $\epsilon > 0$ be given; we will find $\delta > 0$ so that $d - \delta < y < d$ implies

$g(d) - \epsilon < g(y) < g(d)$. To do this, we once again set $c = g(d)$; since $d > f(a)$, we have $c > g(f(a)) = a$. We may suppose WLOG that ϵ is small enough to satisfy $c - \epsilon \geq a$, and we choose

$$\delta = f(c) - f(c - \epsilon)$$

To show that this choice of δ works, let $y \in \mathrm{dom}(g)$ be given satisfying $d - \delta < y < d$. Our choice of δ implies that

$$d - \delta = f(c) - (f(c) - f(c - \epsilon)) = f(c - \epsilon)$$

Thus, $f(c - \epsilon) < y < f(c)$, and applying g to all sides yields $c - \epsilon < g(y) < c$. In other words, we have $g(d) - \epsilon < g(y) < g(d)$, as desired. ■

Remark. Theorem 3.54 was just proven for strictly increasing functions. It is also true that the inverse of a continuous strictly decreasing function is strictly decreasing and continuous. The proof of this can be reduced to the theorem just proven; see Exercise 3.10.9.

You might be wondering why we restrict our attention to strictly increasing or strictly decreasing functions. It turns out that a continuous bijection HAS to be strictly monotone; see Exercise 3.10.7.

Also, although the theorem was stated for functions defined on a closed bounded domain $[a, b]$, with appropriate modifications it also applies to functions defined on any kind of interval (even unbounded intervals like $[0, \infty)$). The details of the proof involve generalizing Theorem 3.48 as well; you should try to work out the details yourself.

Theorem 3.54 proves that the n^{th} root functions are continuous, because they are the inverses of the n^{th} power functions. It follows by the CLT that all rational power functions are continuous (since the function taking $x \in \mathbb{R}$ to $x^{m/n}$ is the composition of the m^{th} power and the n^{th} root functions).

Example 3.55:

Several other useful continuous functions are defined as inverses of strictly montone continuous functions. Some of the most famous examples are the inverse trigonometric functions, which we'll introduce now.

First, let's consider the sine function. sin is not strictly monotone on all of $\mathbb{R}$, but we can restrict our attention to an interval where it is strictly monotone, like $[-\pi/2, \pi/2]$. Thus, since sin is a continuous increasing bijection from $[-\pi/2, \pi/2]$ to $[-1, 1]$, it has a continuous increasing inverse from $[-1, 1]$ to $[-\pi/2, \pi/2]$. This inverse is written as $\sin^{-1}$ or as arcsin.

With the cosine function, cos is strictly decreasing on $[0, \pi]$. Thus, cos is a continuous strictly decreasing bijection from $[0, \pi]$ to $[-1, 1]$, so it has a continuous strictly decreasing inverse from $[-1, 1]$ to $[0, \pi]$. This is written as $\cos^{-1}$ or as arccos.

Lastly, the tangent function, being the ratio of sine and cosine, is continuous and strictly increasing from $(-\pi/2, \pi/2)$ to $(-\infty, \infty)$. Its inverse, denoted by $\tan^{-1}$ or by arctan, is continuous and strictly increasing from $(-\infty, \infty)$ to $(-\pi/2, \pi/2)$. Therefore, much like the staircase function, arctan is defined everywhere, is strictly increasing, and has bounded range, but unlike the staircase function, arctan is continuous everywhere. □

3.10 Exercises

1. In Example 3.52, we demonstrated how to take a continuous function, isolate its zeroes, and try and approximate one of its zeroes using the binary search algorithm. Perform these same steps on each of the following functions $f : \mathbb{R} \to \mathbb{R}$. For each, give your approximation of a zero accurate to two decimal places:

   (a) $f(x) = x^3 + 5x^2 + x - 2$

   (b) $f(x) = x^5 - 4x^3 + 3x + 1$

   (c) $f(x) = \dfrac{x}{2} - \sin x + 1$

2. Prove that for all $a, b \in \mathbb{R}$ with $a < b$ and all continuous functions $f : [a, b] \to [a, b]$, there is some $x \in [a, b]$ with $f(x) = x$. We say that this x is a *fixed point* of f. (Hint: Consider $f(x) - x$.)

3. Although $\tan(\pi/4) = 1$ and $\tan(3\pi/4) = -1$, prove that there is no $c \in (\pi/4, 3\pi/4)$ satisfying $\tan c = 0$. Why doesn't this contradict the IVT?

4. Let $n \in \mathbb{N}^*$ be given. If $f : \mathbb{R} \to \mathbb{R}$ is a polynomial function of degree n, then we say f is *monic* if it can be written with a leading coefficient of 1, i.e. there exist $a_0, a_1, \ldots, a_{n-1} \in \mathbb{R}$ so that for all $x \in \mathbb{R}$, $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Prove that every monic

polynomial of ODD degree has a real zero, and use that to show that all polynomials of odd degree have real zeroes.

(Hint: You want to show that when $|x|$ is large enough, x and $f(x)$ have the same sign. Let $C = \max\{|a_0|, |a_1|, \ldots, |a_{n-1}|\}$. Can you make a bound on $|a_{n-1}x^{n-1} + \cdots + a_1x + a_0|$ in terms of C, n, $|x|$?)

5. Define $f : \mathbb{R} - \{0\} \to \mathbb{R}$ by $f(x) = \dfrac{\sin x}{x}$ for all $x \neq 0$. Prove that there is some $c \in (0, 2)$ such that $f(c) = 0.95$. (Hint: f is not defined at 0, so it is not continuous at 0, but you can extend f to a new function F defined on all of $\mathbb{R}$ by specifying a value for $F(0)$. Which choice of value for $F(0)$ is the most helpful?)

6. This exercise addresses the formal details of verifying that rational powers $a^{m/n}$ are well-defined when $a \geq 0$ and $m, n \in \mathbb{N}^*$.
   (a) Let $a \geq 0$ and $m, n \in \mathbb{N}^*$ be given. Since a^m and a^n are both nonnegative, $(a^m)^{1/n}$ and $(a^{1/n})^m$ both exist. Prove that $(a^m)^{1/n} = (a^{1/n})^m$. (Hint: Be strict about using the definition of n^{th} root correctly.)
   (b) Find some $a < 0$ and some $m, n \in \mathbb{N}^*$ such that $(a^m)^{1/n}$ exists but $(a^{1/n})^m$ does not exist.
   (c) Let $a \geq 0$ and $m, n, p, q \in \mathbb{N}^*$ be given. Prove that if $m/n = p/q$, then $(a^m)^{1/n} = (a^p)^{1/q}$. (Hint: Raise both sides to an appropriate power.)
   (d) Find some $a < 0$ and some $m, n, p, q \in \mathbb{N}^*$ such that $(a^m)^{1/n}$ and $(a^p)^{1/q}$ both exist, and $m/n = p/q$, but $(a^m)^{1/n} \neq (a^p)^{1/q}$.

7. This exercise outlines a proof that a continuous injective function on a closed interval $[a, b]$ is strictly monotone on $[a, b]$. More specifically, we aim to prove

   Theorem 3.56. *Let $a, b \in \mathbb{R}$ be given with $a < b$, and let f be a real function which is continuous and injective on $[a, b]$. Also, suppose that $f(a) < f(b)$ (note that $f(a)$ cannot equal $f(b)$ when f is injective). Then f is strictly increasing on $[a, b]$.*

   Analogously, if $f(a) > f(b)$, then f is strictly decreasing on $[a, b]$; the proof of this claim is very similar to the proof of Theorem 3.56.

(a) First, prove the following lemma:

Lemma 3.57. *Let a, b, f be given as described in Theorem 3.56, and let $a_1, b_1, x_1 \in [a, b]$ be given satisfying $a_1 < x_1 < b_1$. If $f(a_1) < f(b_1)$, then $f(a_1) < f(x_1) < f(b_1)$.*

(Hint: If $f(x_1)$ is not in $(f(a_1), f(b_1))$, then there are two cases to consider. In each case, the IVT should yield a contradiction.)

(b) Now, let a, b, f be given as described in Theorem 3.56. Let $x, y \in [a, b]$ be given satisfying $x < y$. Use Lemma 3.57 twice (with possibly different values of a_1, b_1, and x_1 each time) to prove that $f(x) < f(y)$. This proves that f is strictly increasing on $[a, b]$.

8. Assume that f is a real function which is continuous on a closed interval $[a, b]$. Prove that if f is strictly increasing, then f is a bijection from $[a, b]$ to $[f(a), f(b)]$, and if f is strictly decreasing, then f is a bijection from $[a, b]$ to $[f(b), f(a)]$.

9. Prove an analogous version of Theorem 3.54 for strictly decreasing continuous functions. You may use the fact that the inverse of a strictly decreasing function is strictly decreasing, as shown in Exercise 1.7.6.

10. This exercise generalizes Theorem 3.54. Let $f : A \to B$ be a strictly increasing bijection, where $A, B \subseteq \mathbb{R}$. Let $g : B \to A$ be the inverse function of f. We know that g is strictly increasing.

    Let $a \in A$ be given, and let $b = f(a)$. Recall that in order for $\lim_{x \to a} f(x)$ to exist, f must be defined near a, i.e. A must contain an open interval around a. Similarly, in order for $\lim_{y \to b} g(y)$ to exist, g must be defined near b, i.e. B must contain an open interval around b.

    (a) Prove that if A contains an open interval around a and B contains an open interval around b, then g is continuous at b.

    (b) Prove that under the same hypotheses as part (a), f is continuous at a. (Hint: There is a quick way to do this by reusing part (a)!)

    Remark. The conditions in the exercise imply continuity of the inverse g at b, i.e. they are sufficient for continuity. Furthermore, they actually imply continuity of both the original function f at a AND the inverse g at b, as part (b) shows.

However, these conditions are not necessary for continuity of g at b. For example, consider when $g : B \to A$ is the staircase function s, where $B = \mathbb{R}$ and $A = \text{ran}(s) \subseteq (0,1)$. Let $f = g^{-1}$, so $f : A \to B$ and $g = f^{-1}$. Then f and g are both strictly increasing, and g is continuous at all irrational inputs. However, it can be shown that for all $a \in A$, A does not contain ANY open interval around a at all!

We describe this by saying that A is *nowhere dense*. In some sense, this means that the range of s is very "sparse" in the interval $(0,1)$. The properties of nowhere dense sets are studied more in upper-level real analysis courses.

3.11 Consequences of Continuity: Part 2

Extreme Value Theorem

The previous section presented the Intermediate Value Theorem, which says that as x goes from a to b, $f(x)$ goes from $f(a)$ to $f(b)$ without skipping anything, provided that f is continuous on $[a,b]$. This section shows that the range of f on $[a,b]$ is also bounded (so f doesn't have any "infinite jumps"). In fact, not only is f bounded on $[a,b]$, but f also attains maximum and minimum values, i.e. there exist $c, d \in [a,b]$ such that

$$f(c) = \max\{f(x) \mid x \in [a,b]\} \qquad f(d) = \min\{f(x) \mid x \in [a,b]\}$$

We say that the value $f(c)$ is the *absolute maximum* for f on $[a,b]$, and we say that c is a point where the maximum is attained (there might be more than one such value for c). Similarly, we say $f(d)$ is the *absolute minimum*, attained at d. The values $f(c)$ and $f(d)$ are called *absolute extreme values* (or *absolute extrema*) for f on $[a,b]$.

First, we'll prove that a continuous function f is bounded on a closed interval $[a,b]$. Note that this is not the same thing as saying "a continuous function has a bounded range"; in fact, not all continuous functions have bounded range. For example, the identity function $\text{id}_{\mathbb{R}}$ which maps any $x \in \mathbb{R}$ to itself is continuous and is unbounded on $\mathbb{R}$ (certainly $\mathbb{R}$ is not a bounded interval!). For another example, the function which takes a nonzero $x \in \mathbb{R}$ and produces its reciprocal $1/x$ is unbounded on the interval $(0,1)$. However, in each of these examples, the domain is not a closed bounded interval.

Lemma 3.58. *Let f be a real function which is continuous on a closed interval $[a, b]$. Then f is bounded on $[a, b]$, i.e. $\{f(x) \mid x \in [a, b]\}$ is a bounded set.*

Strategy. We start with a simple remark: if c is the midpoint of $[a, b]$, and f is bounded both on $[a, c]$ and on $[c, b]$, then f is also bounded on $[a, b]$. Why is this true? If f is bounded on $[a, c]$, then there is some $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in [a, c]$. (Hence, M is an upper bound, and $-M$ is a lower bound.) Similarly, if f is bounded on $[c, b]$, then there is some $N \in \mathbb{R}$ so that $|f(x)| \leq N$ for all $x \in [c, b]$. Therefore, for all $x \in [a, b]$, $|f(x)| \leq \max\{M, N\}$, showing that f is bounded on $[a, b]$.

However, our remark is especially useful when we look at its contrapositive: if f is unbounded on $[a, b]$, then it must be unbounded on at least one of $[a, c]$ or $[c, b]$. In other words, if f is unbounded on a closed interval, then f is unbounded on one of its halves. This gives us the idea to do a modified binary search tactic, which in this context is frequently called a *bisection argument*.

Thus, let's suppose f is unbounded on $[a, b]$, and we'll aim to get a contradiction. Let $a_0 = a$ and $b_0 = b$. Next, let $[a_1, b_1]$ be some half of $[a_0, b_0]$ on which f is unbounded. Then, let $[a_2, b_2]$ be some half of $[a_1, b_1]$ on which f is unbounded. Continuing this process, we keep subdividing intervals in half to get a sequence of intervals $[a_n, b_n]$ on which f is unbounded. Note also that since the lengths of these intervals keep getting divided in half, we have $b_n - a_n = (b - a)/2^n$, which gets close to 0 when n is large.

In essence, the intervals are "zooming in" toward a location where f is unbounded. To be more precise about where the zooming occurs, let's take a look at the sequence of a_n values. Because $a_{n+1} \in [a_n, b_n]$ for each n, $a_{n+1} \geq a_n$, so the sequence is increasing. The sequence is also bounded above by b. Thus, the values of the sequence have a supremum; let's say we denote $\sup\{a_n \mid n \in \mathbb{N}\}$ by α. (Basically, the $[a_n, b_n]$ intervals are zooming in on α. To develop this idea more fully, see Exercise 3.12.1.)

Now, since $\alpha \geq a_0 = a$, and $\alpha \leq b$, we know that f is continuous at α. Thus, when x is near α, $f(x)$ is near $f(\alpha)$, which tells us that f is bounded near α. To be more precise, if we take $\epsilon = 1$ (any positive number works here, so we'll use 1 for convenience), there is some $\delta > 0$ such that whenever $|x - \alpha| < \delta$, $|f(x) - f(\alpha)| < \epsilon = 1$. This implies that $|f(x)| < |f(\alpha)| + 1$ (via a calculation with the Triangle Inequality), so f is bounded on the interval $(\alpha - \delta, \alpha + \delta)$.

However, we'll show that for some n large, $[a_n, b_n]$ is contained in $(\alpha - \delta, \alpha + \delta)$, which gives our contradiction since f is bounded on $(\alpha - \delta, \alpha + \delta)$ and unbounded on $[a_n, b_n]$. An argument similar to the one in Theorem 3.48 shows that the values of the a_n's get arbitrarily close to α when n is large enough. Thus, for all n large enough, $a_n > \alpha - \delta$. Also, since the difference $b_n - a_n$ gets very close to 0 as n grows large, we can also guarantee $b_n < \alpha + \delta$ for n large. Thus, $[a_n, b_n] \subseteq (\alpha - \delta, \alpha + \delta)$.

Proof. Let f, a, b be given as described, and assume for contradiction that f is unbounded on $[a, b]$. It follows that if $c = (a + b)/2$ (so c is the midpoint of $[a, b]$), then f is either unbounded on $[a, c]$ or on $[c, b]$. Why is this? Assume otherwise, so there exist $M, N \in \mathbb{R}$ such that $|f(x)| \le M$ for all $x \in [a, c]$ and $|f(x)| \le N$ for all $x \in [c, b]$. It follows that $|f(x)| \le \max\{M, N\}$ for all $x \in [a, b]$, which contradicts the assumption that f is unbounded on $[a, b]$.

Now, we create a sequence $[a_n, b_n]$ of closed intervals on which f is unbounded as follows. We start by letting $a_0 = a$ and $b_0 = b$. Next, by the previous paragraph, f must be unbounded on at least one of the two halves $[a_0, (a_0 + b_0)/2]$ or $[(a_0 + b_0)/2, b_0]$ of the interval $[a_0, b_0]$. Choose $a_1, b_1 \in \mathbb{R}$ so that $[a_1, b_1]$ is a half of $[a_0, b_0]$ on which f is unbounded (if both halves satisfy that property, choose $[a_1, b_1]$ to be the left half).

In general, for any $n \in \mathbb{N}$, once $[a_n, b_n]$ has been defined such that f is unbounded on $[a_n, b_n]$, let $[a_{n+1}, b_{n+1}]$ be a half of $[a_n, b_n]$ on which f is unbounded (if both halves satisfy that property, then pick the left half). Note that $a_{n+1} \ge a_n$, so the sequence of a_n values is increasing. Also, because the length of $[a_{n+1}, b_{n+1}]$ is half the length of $[a_n, b_n]$, a routine induction argument shows that $b_n - a_n = (b - a)/2^n$.

Let $\alpha = \sup\{a_n \mid n \in \mathbb{N}\}$; this sup exists because $a_n \le b$ for every $n \in \mathbb{N}$. Thus, $\alpha \le b$ and $\alpha \ge a_0 = a$, so $\alpha \in [a, b]$ and f is continuous at α. This means there is some $\delta > 0$ such that for all $x \in [a, b]$, if $|x - \alpha| < \delta$ then $|f(x) - f(\alpha)| < 1$; in particular, by the Triangle Inequality,

$$|f(x)| = |(f(x) - f(\alpha)) + f(\alpha)| \le |f(x) - f(\alpha)| + |f(\alpha)| < 1 + |f(\alpha)|$$

This shows that f is bounded on $(\alpha - \delta, \alpha + \delta) \cap [a, b]$.

Next, as $\alpha - \delta$ is less than α, $\alpha - \delta$ is not an upper bound of $\{a_n \mid n \in \mathbb{N}\}$, so there is some $n_0 \in \mathbb{N}$ such that $a_{n_0} > \alpha - \delta$. Choose $n \ge n_0$ large enough so that $(b - a)/2^n < \delta$ (this is possible because Exercise 3.8.6.(b) ensures that the powers of 2 are unbounded). Therefore, as the a_n sequence is increasing with upper bound α, we know

$$\alpha - \delta < a_{n_0} \leq a_n \leq \alpha$$

and

$$b_n = a_n + \frac{b-a}{2^n} < a_n + \delta \leq \alpha + \delta$$

This implies that $[a_n, b_n] \subseteq (\alpha - \delta, \alpha + \delta)$. We also know $[a_n, b_n] \subseteq [a, b]$. Therefore, as f is bounded on $(\alpha - \delta, \alpha + \delta) \cap [a, b]$, f is bounded on $[a_n, b_n]$. However, this contradicts the choice of $[a_n, b_n]$ as an interval on which f is unbounded. ■

Remark. Where did the assumption of continuity actually get used in the previous proof? We note that we only used the continuity of f for one purpose: to show that f was bounded in some open interval containing α. However, the bound given there only works for inputs near α.

In general, whenever a real function f is continuous at some value x, then f is bounded in an open interval around x. However, the bound for f on that open interval won't necessarily be valid for inputs that are too far away from x. Thus, if you stay close to x, i.e. you stay in x's "neighborhood", f is bounded. This is sometimes summarized by saying that a continuous function f is *locally bounded*. (Sometimes, for contrast, when f is bounded on an entire interval, we say f is "globally bounded".)

Thus, the previous theorem says in essence that if a function f is locally bounded for every point in a closed interval $[a, b]$, then f is "globally" bounded, i.e. there is a bound on $|f(x)|$ that works for ALL $x \in [a, b]$. The bisection argument is one approach used to take a "local" property of a function and make it a "global" property of a function. We'll see another example of a bisection argument soon.

Lemma 3.58 tells us that if f is continuous on $[a, b]$, then $\sup\{f(x) \mid x \in [a, b]\}$ and $\inf\{f(x) \mid x \in [a, b]\}$ exist. Let's now prove that the sup and inf are actually max and min respectively. This famous result is called the *Extreme Value Theorem*, or the EVT for short.

Theorem 3.59 (Extreme Value Theorem). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$. Then f attains an absolute maximum and absolute minimum on $[a, b]$.*

Strategy. Let's first consider the case where f is a continuous function on $[a, b]$ with $\inf(\text{ran}(f)) = 0$. To show that f attains a minimum value, we want to show that 0 is actually $\min(\text{ran}(f))$. Assume $0 \notin \text{ran}(f)$ for a contradiction. This means that the values of f are all positive, and they get arbitrarily close to 0, but they never actually attain 0.

We can then apply the following nice trick: if a continuous function is always positive on $[a, b]$, then its reciprocal is continuous on $[a, b]$, and hence Lemma 3.58 tells us that the reciprocal is bounded on $[a, b]$. Thus, for our function f, $1/f$ should be bounded on $[a, b]$. However, since f takes values arbitrarily close to 0, $1/f$ should also be unbounded, which is a contradiction.

This proves that if a continuous function on $[a, b]$ has 0 as the inf of its range, then 0 has to actually be in the range. For all other values of the inf, those cases can be reduced to the case of the inf being 0 by doing vertical shifts. This shows that f attains an absolute minimum. To show that f attains an absolute maximum, we merely need to find an absolute minimum of $-f$.

Proof. Let f, a, b be given as described. We will show that f attains an absolute minimum, i.e. $\min(\text{ran}(f))$ exists. It will then follow that f has an absolute maximum because $\max(\text{ran}(f)) = \min(\text{ran}(-f))$, and $-f$ is continuous on $[a, b]$ as well.

By Lemma 3.58, we may declare $m = \inf(\text{ran}(f))$. Suppose for contradiction that $m \notin \text{ran}(f)$. Therefore, $f(x) > m$ for all $x \in [a, b]$, so $f(x) - m > 0$ for all $x \in [a, b]$. If $g : [a, b] \to \mathbb{R}$ is defined by $g(x) = f(x) - m$ for all $x \in [a, b]$, then g is continuous on $[a, b]$, $\inf(\text{ran}(g)) = 0$, and $g(x) > 0$ for all $x \in [a, b]$. Therefore, $1/g$ is continuous on $[a, b]$.

By Lemma 3.58, $1/g$ is bounded on $[a, b]$, so there exists some $M > 0$ such that for all $x \in [a, b]$, $1/g(x) \le M$. Thus, $g(x) \ge 1/M$ for all $x \in [a, b]$ (because M and $g(x)$ are positive). However, since $1/M$ is greater than $\inf(\text{ran}(g))$, i.e. $1/M$ is greater than the greatest lower bound of $\text{ran}(g)$, there exists some member of $\text{ran}(g)$ less than $1/M$. Thus, there is some $x \in [a, b]$ with $g(x) < 1/M$. This contradicts the statement that $g(x) \ge 1/M$ for all $x \in [a, b]$, and thus we must have $m \in \text{ran}(f)$, as claimed. ■

Example 3.60:

Consider the function $g : [\pi/2, 5\pi/2] \to \mathbb{R}$ defined by $g(x) = \sin x$ for all $x \in [\pi/2, 5\pi/2]$. (Thus, g is the restriction of sin to $[\pi/2, 5\pi/2]$.) We know

$|\sin(x)| \leq 1$ for all $x \in \mathbb{R}$, so this tells us $\inf(\text{ran}(g)) \geq -1$ and $\sup(\text{ran}(g)) \leq 1$. Furthermore, $g(\pi/2) = 1$ and $g(3\pi/2) = -1$, showing that the absolute maximum is 1 and the absolute minimum is -1. Note as well that $g(5\pi/2) = g(\pi/2) = 1$ because sin is periodic with period 2π. This shows that an absolute maximum can be attained at more than one point. □

Unfortunately, while the Extreme Value Theorem guarantees the existence of absolute extrema for continuous functions f on $[a, b]$, it does not provide a way to find the locations of these extrema. A binary search technique doesn't work for finding maximum values, because the EVT provides no way of knowing which half of $[a, b]$ contains a maximum for f (whereas in contrast, the IVT can determine which half of $[a, b]$ contains a zero of f if f changes signs). Nevertheless, the EVT is a useful theoretical tool, and we will need it when we study derivatives in the next chapter.

Extent: From Top To Bottom

Since the EVT shows us that continuous functions are bounded on closed intervals, we'd now like to see what we can say about the bounds. If the bounds are far apart, it says that the function rises or falls a lot. This suggests the following definition:

Definition 3.61. If a real function f is bounded on a set $S \subseteq \text{dom}(f)$, then we'll define the *extent*[7] of S for f, $\text{extent}(f, S)$, by the equation

$$\text{extent}(f, S) = \sup\{f(x) \mid x \in S\} - \inf\{f(x) \mid x \in S\}$$

$\text{extent}(f, S)$ measures the distance from the top of f's range on S to the bottom.

The notion of extent helps simplify some of our definitions. For instance, when f is continuous at a, $f(x)$ is close to $f(a)$ when x is close to a, so this suggests that the values of $f(x)$ aren't far away from one another when x comes from a sufficiently small open interval containing a. More formally, in Exercise 3.12.5.(a), you can prove that for all real functions f and all $a \in \text{dom}(f)$, saying that f is continuous at a is equivalent to saying

$$\forall \epsilon > 0 \; \exists \delta > 0 \; (\text{extent}(f, (a - \delta, a + \delta)) < \epsilon)$$

[7] This can also be called the *span* of f on S, but the word "span" frequently means something different in linear algebra. Sometimes, the phrase "the *oscillation* of f on S" is also used instead of "extent".

We will make use of this fact later.

As an illustration of the notion of extent, let's return to the hypothetical situation where you're driving a car. However, instead of doing a 20 minute trip, your trip is expected to take 8 hours. We'll use a function $f(t)$ to model the total amount of fuel used by the car (measured in gallons of gas) after t hours of driving, so $f : [0, 8] \to [0, \infty)$. Intuitively, f is an increasing continuous function. Therefore, for any subinterval $[a, b]$ of $[0, 8]$, $\text{extent}(f, [a, b]) = f(b) - f(a)$, meaning that as time progresses from $t = a$ to $t = b$, the car consumes $f(b) - f(a)$ gallons of gas.

Now, let's suppose that your car doesn't have a large enough gas tank to make the entire 8 hour trip with no breaks. Therefore, your trip has to be broken into segments to allow for refueling. (Let's assume that we can refuel the car without having to stop, which is not currently practical for most cars but will make our explanation simpler.) If you have a car with a 15-gallon tank of gas, then you want each segment of the trip to burn at most 15 gallons. In other words, the extent of each segment $[a, b]$ of the trip should be at most 15. After analyzing your trip, perhaps you find that you should break the trip into 5 segments, each lasting 1.6 hours.

However, if your car has a smaller tank that only holds 10 gallons of gas, then you need to break your trip into more segments, each having an extent of at most 10. In fact, no matter how small your tank is, you can always complete your 8-hour trip, provided you break the trip into enough segments. Clearly the number of segments needed is going to be related to how fast your car burns gas at different points in the trip: if the car is accelerating very quickly in one portion of the trip and thus going through fuel quickly, then that portion probably needs to be broken down into more segments in order for each segment to have a small extent of fuel consumption. Thus, we get detailed information about how quickly the car burns gas by examining the relationship between the number of segments of the trip and the extent of fuel consumption of each segment.

To put this example in more conventional terms, suppose we have a real function f which is continuous on a closed bounded interval $[a, b]$. We'd like to get a better picture of how much f rises or falls. To do this, we'll chop $[a, b]$ into equal-width subintervals, each having a small extent. Let's illustrate this idea with another example.

Example 3.62:

Consider $f : [1,3] \to \mathbb{R}$ given by $f(x) = x^2$ for all $x \in [1,3]$. Since f is strictly increasing on $[1,3]$, for all $a, b \in \mathbb{R}$ with $[a,b] \subseteq [1,3]$, we have $\sup\{x^2 \mid x \in [a,b]\} = b^2$ and $\inf\{x^2 \mid x \in [a,b]\} = a^2$. Thus, $\text{extent}(f,[a,b]) = b^2 - a^2$.

This shows that $\text{extent}(f,[1,3]) = 3^2 - 1^2 = 8$. If we cut up $[1,3]$ into the two equal-width subintervals $[1,2]$ and $[2,3]$, then $\text{extent}(f,[1,2]) = 3$ and $\text{extent}(f,[2,3]) = 5$. Here, the largest extent for a subinterval is at most 5. If we cut $[1,3]$ into four equal-width pieces $[1,1.5]$, $[1.5,2]$, $[2,2.5]$, and $[2.5,3]$, we find $\text{extent}(f,[1,1.5]) = 1.25$, $\text{extent}(f,[1.5,2]) = 1.75$, $\text{extent}(f,[2,2.5]) = 2.25$, and $\text{extent}(f,[2.5,3]) = 2.75$. Now, the largest extent we have is 2.75. The following table shows for different values of n how large the extents become if we cut $[1,3]$ into n equal-width pieces:

n	Largest extent when breaking into n equal-width subintervals
1	8
2	5
3	≈ 3.56
4	2.75
8	1.4375
10	1.16
12	≈ 0.972

In fact, it can readily be shown that for any n, when $[1,3]$ is broken into n equal-width pieces, the subinterval with the largest extent is $[3 - 2/n, 3]$, with

$$\text{extent}\left(f, \left[3 - \frac{2}{n}, 3\right]\right) = 9 - \left(3 - \frac{2}{n}\right)^2 = 9 - \left(9 - \frac{12}{n} + \frac{4}{n^2}\right) = \frac{12}{n} - \frac{4}{n^2}$$

and this extent can be made arbitrarily close to 0 by choosing n large enough. $\square$

In our example, we see that when we break $[1,3]$ into more pieces, the extents of the pieces decrease. In general, a smaller set has a smaller extent, as is shown in the following lemma:

Lemma 3.63. *Let f be a real function, and let $S, T \subseteq \text{dom}(f)$ be given. If $S \subseteq T$ and f is bounded on T (so it is also bounded on S), then* $\text{extent}(f,S) \leq \text{extent}(f,T)$.

The proof of this lemma is Exercise 3.12.6.

Going back to Example 3.62, we saw that for any $\epsilon > 0$, by dividing the domain $[1, 3]$ into enough pieces, the extents of the function f could all be made smaller than ϵ. Let's give this property a name:

Definition 3.64. Let f be a real function and let $\epsilon > 0$ be given. Let's say that a closed interval $[a, b]$ has the *ϵ-extent property* (for f) if $[a, b] \subseteq \mathrm{dom}(f)$ and there exists an $n \in \mathbb{N}^*$ such that when $[a, b]$ is broken into n equal-width subintervals

$$\left[a + \frac{b-a}{n}\right], \left[a + \frac{b-a}{n}, a + \frac{2(b-a)}{n}\right], \ldots, \left[a + \frac{(n-1)(b-a)}{n}, b\right]$$

the extent of f on each subinterval is less than ϵ.

We'd like to prove that, in general, for any continuous function, each closed interval $[a, b]$ has the ϵ-extent property for every $\epsilon > 0$. We'll use a bisection argument to do this, where we assume that the interval $[a, b]$ doesn't have the ϵ-extent property, and then we keep picking halves of intervals in order to "zoom in" on one point. The following lemma will help us do this:

Lemma 3.65. *If f is a real function defined on a closed bounded interval $[a, b]$, then let $c = (a + b)/2$ be the midpoint of $[a, b]$. For any $\epsilon > 0$, if $[a, c]$ and $[c, b]$ both have the ϵ-extent property, then $[a, b]$ has the ϵ-extent property.*

Strategy. To see this, suppose that $[a, c]$ can be split into n_1 equal-width pieces, each with extent less than ϵ, and $[c, b]$ can be split into n_2 equal-width pieces, each with extent less than ϵ. Upon doing this, $[a, b]$ is broken into n_1 pieces of width $(b - a)/(2n_1)$ and n_2 pieces of width $(b - a)/(2n_2)$. See Figure 3.13 for an example, which uses solid lines to break up the left half $[a, c]$ into 2 pieces and to break up the right half $[c, b]$ into 3 pieces.

However, we'd like to make all our subdivisions the same width. Thus, we further divide each piece of the left half in n_2 pieces, and we divide each piece of the right half in n_1 pieces. (Figure 3.13 uses dashed lines to represent these further subdivisions.) By Lemma 3.63, the extents of these new pieces have not grown any bigger, so they are all still less than ϵ. Now, each half has now been broken into $n_1 n_2$ pieces, all of the same width. Thus, we have broken up $[a, b]$ into $2n_1 n_2$ pieces, all with extent less than ϵ.

Proof. Let f, a, b, c be given as described. Let $\epsilon > 0$ be given, and assume that $[a, c]$ and $[c, b]$ have the ϵ-extent property. Thus, we may choose $n_1, n_2 \in \mathbb{N}^*$

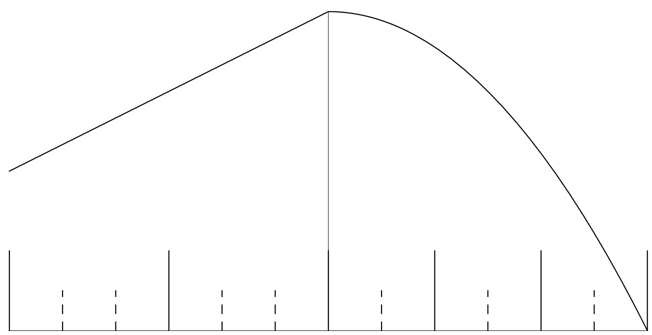

Figure 3.13: Splitting $[a, c]$ in $n_1 = 2$ pieces and $[c, b]$ in $n_2 = 3$ pieces

such that when $[a, c]$ is broken into n_1 equal-width subintervals, and $[c, b]$ is broken into n_2 equal-width subintervals, then all these subintervals have extent less than ϵ. To be more precise, when i is an integer from 1 to n_1, let S_i be the i^{th} subinterval of $[a, c]$ (counting from the left). S_i has width $(c-a)/n_1 = (b-a)/(2n_1)$. Similarly, if j is an integer from 1 to n_2, then let T_j be the j^{th} subinterval of $[c, b]$ (counting from the left). T_j has width $(b-c)/n_2 = (b-a)/(2n_2)$.

Now, for each i from 1 to n_1, by the choice of S_i, we have extent$(f, S_i) < \epsilon$. Divide S_i into n_2 equal-width subintervals, called $S_{i,1}$, $S_{i,2}$, and so on up to S_{i,n_2}. For each k from 1 to n_2, since $S_{i,k} \subseteq S_i$, Lemma 3.63 ensures that extent$(f, S_{i,k}) \le$ extent$(f, S_i) < \epsilon$. Also, the length of $S_{i,k}$ is the length of S_i divided by n_2, i.e. $(b-a)/(2n_1n_2)$.

Similarly, for each j from 1 to n_2, we may divide T_j into n_1 equal-width subintervals, called $T_{j,1}$, $T_{j,2}$, and so on up to T_{j,n_1}. We have extent$(f, T_{j,l}) \le$ extent$(f, T_j) < \epsilon$ for each l from 1 to n_1, and the length of $T_{j,l}$ is the length of T_j divided by n_1, i.e. $(b-a)/(2n_1n_2)$.

Hence, $[a, b]$ is the union of all the $S_{i,k}$ and $T_{j,l}$ intervals (where i and l are from 1 to n_1, and j and k are from 1 to n_2), each subinterval having length $(b-a)/(2n_1n_2)$ and extent less than ϵ. Thus, $[a, b]$ has the ϵ-extent property. ■

With the aid of Lemma 3.65, we can now prove the following:

Theorem 3.66. *Let f be a real function continuous on a closed interval $[a, b]$. Then for all $\epsilon > 0$, $[a, b]$ has the ϵ-extent property for f.*

Strategy. Let's assume that $[a, b]$ does not have the ϵ-extent property and try to get a contradiction. By Lemma 3.65, we know that if $[a, b]$ does not have the ϵ-extent property, then one of its halves does not either. Let's choose $[a_1, b_1]$ to be some half of $[a, b]$ which does not have the ϵ-extent property. After that, some half of $[a_1, b_1]$ doesn't have the ϵ-extent property: call that half $[a_2, b_2]$. In this manner, we generate smaller and smaller subintervals $[a_n, b_n]$ for each $n \in \mathbb{N}^*$ which don't have the ϵ-extent property.

As in the proof of Lemma 3.58, the intervals $[a_n, b_n]$ "zoom in" on the value $\alpha = \sup\{a_n \mid n \in \mathbb{N}^*\}$. Since f is continuous at α, we have seen that there is some $\delta > 0$ such that $\text{extent}(f, (\alpha - \delta, \alpha + \delta)) < \epsilon$. However, as in the proof of Lemma 3.58, when n is large, we'll actually have $[a_n, b_n] \subseteq (\alpha - \delta, \alpha + \delta)$. Hence, Lemma 3.63 implies that $\text{extent}(f, [a_n, b_n]) \leq \text{extent}(f, (\alpha - \delta, \alpha + \delta)) < \epsilon$, so $[a_n, b_n]$ has the ϵ-extent property (just break it into 1 subinterval!). This contradicts the choice of $[a_n, b_n]$.

Proof. Let f, a, b be given as described. Let $\epsilon > 0$ be given, and assume for contradiction that $[a, b]$ does not have the ϵ-extent property. We create a sequence $[a_n, b_n]$ of closed intervals which do not have the ϵ-extent property as follows. We start by letting $a_0 = a$ and $b_0 = b$. Next, by Lemma 3.65, at least one of the two halves $[a_0, (a_0 + b_0)/2]$ or $[(a_0 + b_0)/2, b_0]$ of the interval $[a_0, b_0]$ does not have the ϵ-extent property. Choose $[a_1, b_1]$ to be such a half (if both halves fail to have the ϵ-extent property, then choose $[a_1, b_1]$ to be the left half).

In general, for any $n \in \mathbb{N}$, once $[a_n, b_n]$ has been defined such that $[a_n, b_n]$ does not have the ϵ-extent property, let $[a_{n+1}, b_{n+1}]$ be a half of $[a_n, b_n]$ without the ϵ-extent property (if both halves fail to have the ϵ-extent property, then pick the left half). Note that $a_{n+1} \geq a_n$, so the sequence of a_n values in increasing. Also, because the length of $[a_{n+1}, b_{n+1}]$ is half the length of $[a_n, b_n]$, we see that $b_n - a_n = (b - a)/2^n$.

Let $\alpha = \sup\{a_n \mid n \in \mathbb{N}\}$; this sup exists because $a_n \leq b$ for every $n \in \mathbb{N}$. Thus, $\alpha \leq b$ and $\alpha \geq a_0 = a$, so $\alpha \in [a, b]$ and f is continuous at α. This means there is some $\delta > 0$ such that $\text{extent}(f, (\alpha - \delta, \alpha + \delta)) < \epsilon$.

As $\alpha - \delta$ is less than α, $\alpha - \delta$ is not an upper bound of $\{a_n \mid n \in \mathbb{N}^*\}$, so there is some $n_0 \in \mathbb{N}$ such that $a_{n_0} > \alpha - \delta$. Choose $n \geq n_0$ large enough so that $(b - a)/2^n < \delta$. Therefore, as the a_n sequence is increasing with upper bound α, we know

$$\alpha - \delta < a_{n_0} \leq a_n \leq \alpha$$

and

$$b_n = a_n + \frac{b - a}{2^n} < a_n + \delta \leq \alpha + \delta$$

This implies that $[a_n, b_n] \subseteq (\alpha-\delta, \alpha+\delta)$. Thus, by Lemma 3.63, we know that extent$(f, [a_n, b_n]) \leq$ extent$(f, (\alpha - \delta, \alpha + \delta)) < \epsilon$. However, this means that $[a_n, b_n]$ has the ϵ-extent property (just break it into 1 piece!), contradicting the choice of $[a_n, b_n]$. ■

Uniform Continuity

For the last topic of this chapter, let's take a closer look at the definition of continuity. If a real function f is continuous at some point a, then it means for any accuracy $\epsilon > 0$ we want, there is some $\delta > 0$ guaranteeing that $|f(x) - f(a)| < \epsilon$ when $|x - a| < \delta$. δ tells us how much we have to control the inputs to our function if we want the function values $f(x)$ and $f(a)$ to stay close together.

However, δ also depends on a. In some sense, δ measures how rapidly the function f is changing near a: the faster f is rising or falling near a, the tighter a control we need for δ. We illustrate this with an example.

Example 3.67:

Consider the function $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = x^2$ for all $x > 0$. Back in Example 3.21, we showed that $x^2 \to 1$ as $x \to 1$ by showing, for each $\epsilon > 0$, what to choose for the corresponding $\delta > 0$. We've seen since then that $x^2 \to a^2$ as $x \to a$ for any $a > 0$. Let's go back over the formal proof of this limit, so that we can study how the value of δ depends on a.

Let $a > 0$ and $\epsilon > 0$ be given; we wish to find $\delta > 0$ so that whenever $x \in (0, \infty)$ and $|x-a| < \delta$, then $|x^2-a^2| < \epsilon$. Writing $|x^2-a^2|$ as $|x+a||x-a|$, we want to first find an appropriate bound for $|x+a|$, and then we'll use that bound to figure out how small to make $|x-a|$. Using the Triangle Inequality, we find

$$|x + a| = |x - a + 2a| \leq |x - a| + |2a| = |x - a| + 2a$$

since $a > 0$. Thus, if we make sure that $|x-a|$ is at most a, then $|x+a| \leq 3a$.

Therefore,

$$|x^2 - a^2| = |x + a||x - a| \leq 3a|x - a|$$

In order to make this less than ϵ, we'd like $|x - a|$ to be at most $\epsilon/(3a)$. This means that we want to guarantee both $|x - a|$ less than a and less than $\epsilon/(3a)$, so we take

$$\delta = \min\left\{a, \frac{\epsilon}{3a}\right\}$$

Notice that when a is large enough (specifically, larger than $\sqrt{\epsilon/3}$), δ is $\epsilon/(3a)$. Thus, as a gets bigger and bigger, our control will become smaller and smaller, suggesting that we need tighter control on our input to achieve the same accuracy. This makes sense, since the graph of f rises faster and faster the more you move to the right, with no apparent bound on the "speed" of f's ascent (we will make this more precise in the next chapter when we study derivatives).

A good question to ask is: for a given $\epsilon > 0$, is it possible to pick one value of δ which works for ALL values of $a > 0$ at the same time? It seems that any such value of δ would have to be less than $\epsilon/(3a)$ for every a, but $\epsilon/(3a)$ can be made arbitrarily close to 0 by taking a large enough. This suggests that it is impossible to pick one value of δ which works for all a at the same time. To prove this statement more rigorously, see Exercise 3.12.7.

However, if we choose to restrict our choices of a to a closed bounded interval, then it IS possible to take one value of δ that works for all a in that interval. For instance, suppose we only want to consider values of a from $[1, 3]$. In that case, $a \leq 3$, and $\epsilon/(3a) \leq \epsilon/(3 \cdot 1) = \epsilon/3$. Thus, the choice of $\delta = \min\{3, \epsilon/3\}$ works for all values of a in $[1, 3]$, and it only depends on ϵ (as opposed to depending on ϵ and a). $\square$

The previous example motivates the following definition:

Definition 3.68. Let f be a real function, and let S be an interval contained in $\text{dom}(f)$ (it can be unbounded). We say that f is *uniformly continuous on* S if the following holds:

$$\forall \epsilon > 0 \; \exists \delta > 0 \; \forall x, y \in S \; (|x - y| < \delta \to |f(x) - f(y)| < \epsilon)$$

Thus, if f is uniformly continuous on S, then f is continuous at each point a of S (just plug in a for y). However, the choice of δ only depends on ϵ, not on any particular point where the limit is computed (here, δ is chosen BEFORE x or y is given).

To summarize Example 3.67, we have shown that the function f from the example is uniformly continuous on $[1, 3]$, but it is NOT uniformly continuous on $(0, \infty)$. In some sense, this is because there is no bound to how fast f rises when you look at all values in $(0, \infty)$, but there is such a bound on $[1, 3]$. This shows that the notion of uniform continuity on S depends highly on the set S.

To illustrate the idea behind uniform continuity, let's return to the gas tank example we discussed after the definition of extent. In that example, our gas consumption function $f : [0, 8] \to [0, \infty)$ is increasing and continuous. At any specific point a of the trip, because f is continuous at a, no matter how small the size ϵ of our gas tank is, there will always be some $\delta > 0$ which guarantees that $\text{extent}(f, (a - \delta, a + \delta)) < \epsilon$, i.e. at most ϵ gallons of gas are burned between the times $a-\delta$ and $a+\delta$. Thus, your car can successfully run for 2δ hours without needing to refuel a tank of size ϵ. However, δ normally depends on a: this makes sense, since the faster the car is burning gas at time a, the less time the car can run without refueling.

However, if f is uniformly continuous on $[0, 8]$, that means that some value of δ works for EVERY a. Thus, no matter what point in the trip you've reached, you can always run the car for 2δ hours before refueling an ϵ-sized tank. This allows you to conveniently break up your trip into segments that are 2δ hours long, rather than having to refuel more frequently during the faster portions of your trip. This suggests that uniform continuity is related to the ϵ-extent property.

In general, there is a connection between the ϵ-extent property and uniform continuity, and we will use this connection to prove that all continuous functions are uniformly continuous on closed bounded intervals (i.e. when the domain has the form $[a, b]$). Recall that continuity of a function f at a means that for all $\epsilon > 0$, there is some $\delta > 0$ such that $\text{extent}(f, (a - \delta, a + \delta)) < \epsilon$. In contrast, you can prove in Exercise 3.12.5.(b) that uniform continuity of f on an interval S, when $S \subseteq \text{dom}(f)$, is equivalent to saying that for all $\epsilon > 0$, there is a $\delta > 0$ such that ALL intervals of length δ contained in S have extent less than ϵ. (You can check that two points x and a satisfy $|x - a| < \delta$ iff the two points belong to the open interval of length δ centered at $(x + a)/2$.) Therefore, the result from Theorem 3.66, which allows us to break $[a, b]$ into subintervals of small extent, will be helpful to us in proving the following:

Theorem 3.69. *Let f be a real function continuous on a closed interval $[a, b]$. Then f is uniformly continuous on $[a, b]$.*

Strategy. Theorem 3.66 tells us that we can break up the interval $[a, b]$ into some number of equal-width subintervals, say n of them, each having extent less than ϵ. Thus, if x and y are any two points belonging to the same subinterval S, then

$$f(x) - f(y) \le \sup\{f(z) \mid z \in S\} - \inf\{f(z) \mid z \in S\} = \text{extent}(f, S) < \epsilon$$

Similarly, we also find that $f(y) - f(x) < \epsilon$, so $|f(x) - f(y)| < \epsilon$. Since each subinterval has length $(b - a)/n$, this suggests that we should pick $\delta = (b - a)/n$.

However, if $x, y \in [a, b]$ are within distance $(b-a)/n$ from each other, that doesn't mean they have to be in the same subinterval of $[a, b]$. There are two cases: either x and y belong to the same subinterval of $[a, b]$, or they belong to adjacent subintervals. In the second case, how can we estimate the distance $|f(x) - f(y)|$? Let's suppose that c is the common boundary between the two subintervals. Thus, as x and c belong to a common subinterval, $|f(x) - f(c)| < \epsilon$, and as c and y belong to a common subinterval, $|f(c) - f(y)| < \epsilon$. We can use the Triangle Inequality to obtain

$$|f(x) - f(y)| \leq |f(x) - f(c)| + |f(c) - f(y)| < 2\epsilon$$

Therefore, whether or not x and y lie in the same subinterval of $[a, b]$, we know the distance $|f(x) - f(y)|$ is less than 2ϵ. This means we should have broken up $[a, b]$ into subintervals with extent $\epsilon/2$ instead of extent ϵ.

Proof. Let f, a, b be given as described. Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ so that

$$\forall x, y \in [a, b]\ (|x - y| < \delta \to |f(x) - f(y)| < \epsilon)$$

By Theorem 3.66, there is some $n \in \mathbb{N}^*$ such that if we break $[a, b]$ into n equal-width subintervals (name them S_1, S_2, and so on up to S_n going from left to right), each subinterval has extent less than $\epsilon/2$ for f. Let δ be the width of a subinterval, so $\delta = (b - a)/n$.

Now, let $x, y \in [a, b]$ be given satisfying $|x - y| < \delta$. (WLOG we may assume $x < y$.) Let's say that x lies in the subinterval S_i where i is from 1 to n (if x is an endpoint between two subintervals, choose i to be the left one). If $i = n$, then $y > x$ implies that $y \in S_n$ as well (S_n is the rightmost subinterval). Otherwise, we note that $x + \delta \in S_{i+1}$, so as $y < x + \delta$, y is in either S_i or S_{i+1}. This gives us two cases to consider.

For the first case, suppose $y \in S_i$. Then because $x, y \in S_i$, we have $f(x) \leq \sup\{f(z) \mid z \in S_i\}$ and $f(y) \geq \inf\{f(z) \mid z \in S_i\}$, so $f(x) - f(y) \leq \text{extent}(f, S_i) < \epsilon/2$. The same reasoning, but with the roles of x and y interchanged, shows that $f(y) - f(x) \leq \text{extent}(f, S_i) < \epsilon/2$ as well. Thus, $|f(x) - f(y)| < \epsilon/2 < \epsilon$.

For the second case, suppose $y \in S_{i+1}$. Let c be the endpoint shared by S_i and S_{i+1}. Because x and c are both in S_i, $|f(x) - f(c)| < \epsilon/2$ by the proof of the first case. Also, because c and y are both in S_{i+1}, $|f(c) - f(y)| < \epsilon/2$. Therefore, by the Triangle Inequality,

$$|f(x) - f(y)| \leq |f(x) - f(c)| + |f(c) - f(y)| < 2(\epsilon/2) = \epsilon$$

In either case, we have shown $|f(x) - f(y)| < \epsilon$, as desired. ■

3.12 Exercises

1. In the proof of Lemma 3.58, we created a sequence of closed intervals $[a_n, b_n]$, where for every $n \in \mathbb{N}$, $[a_{n+1}, b_{n+1}]$ is always one of the two halves of $[a_n, b_n]$. We saw that the a_n values have a supremum α. This, in essence, says that the left endpoints of the intervals converge. This exercise will address the right endpoints.
   (a) Prove that $b_{n+1} \leq b_n$ for each $n \in \mathbb{N}$, and also show that $\{b_n \mid n \in \mathbb{N}\}$ is bounded below.
   (b) By part (a), $\inf\{b_n \mid n \in \mathbb{N}\}$ exists; call it β. Prove that $\alpha = \beta$. (Hint: If α were smaller than β, then we could find n large enough so that $(b-a)/2^n < \beta - \alpha$. How big can b_n be?) This shows that the right endpoints also converge to the same limit as the left endpoints.
   (c) Prove that if x belongs to $[a_n, b_n]$ for every $n \in \mathbb{N}$, then $x = \alpha = \beta$.

2. If we try the proof from Lemma 3.58 to the function which sends $x \in (0, 1)$ to $1/x$, where does the proof fail? (Hint: For each $n \in \mathbb{N}$, you should be able to figure out what $[a_n, b_n]$ will be.)

3. Prove that if f is continuous on a closed bounded interval $[a, b]$, then the set $\{f(x) \mid x \in [a, b]\}$ is also a closed bounded interval. (Hint: Use both the IVT and the EVT.)

4. In each part of this problem, a function f from $\mathbb{R}$ to $\mathbb{R}$, an interval $[a, b]$, and a positive number ϵ is specified. For each part, show that the interval has the ϵ-extent property by giving a number n of equal-width subintervals so that each subinterval has extent less than ϵ. Show your steps.
   (a) $f(x) = x$, $[a, b] = [0, 2]$, $\epsilon = 1$.

(b) $f(x) = x^2 - x$, $[a, b] = [-1, 1]$, $\epsilon = 0.5$.
(c) $f(x) = \sin x$, $[a, b] = [0, 2\pi]$, $\epsilon = 0.5$.
(d) $f(x) = \sqrt{x}$, $[a, b] = [0, 3]$, $\epsilon = 1$.

5. (a) Let f be a real function and let $a \in \mathrm{dom}(f)$ be given. Prove that f is continuous at a iff for all $\epsilon > 0$, there is some $\delta > 0$ so that $\mathrm{extent}(f, (a - \delta, a + \delta)) < \epsilon$.
   (b) Let f be a real function and let S be an interval contained in $\mathrm{dom}(f)$. Prove that f is uniformly continuous on S iff for all $\epsilon > 0$, there is some $\delta > 0$, such that for every interval $I \subseteq S$ of length less than δ, $\mathrm{extent}(f, I) < \epsilon$.

6. Prove Lemma 3.63.

7. Prove that the function $f : (0, \infty) \to \mathbb{R}$, defined by $f(x) = x^2$ for all $x \in (0, \infty)$, is not uniformly continuous on $(0, \infty)$. More specifically, show that when $\epsilon = 1$, no choice of δ satisfies the definition of uniform continuity.

8. Let f be a real function, and let S be an interval contained in $\mathrm{dom}(f)$. We say that f is *Lipschitz-continuous* on S with *Lipschitz constant* $L > 0$ if the following holds:

$$\forall x, y \in S \ |f(x) - f(y)| \leq L|x - y|$$

Thus, for a Lipschitz-continuous function, the distance between outputs is directly proportional to the distance between inputs.

   (a) Prove that if f is Lipschitz-continuous on S (i.e. if it has a Lipschitz constant L), then f is uniformly continuous on S. Thus, Lipschitz continuity is a stronger property than uniform continuity.
   (b) Prove that the function $f : [0, 1] \to \mathbb{R}$, defined by $f(x) = \sqrt{x}$ for all $x \in \mathbb{R}$, is not Lipschitz-continuous on $[0, 1]$. As Theorem 3.69 proves that f is uniformly continuous on $[0, 1]$, this proves that uniform continuity does not imply Lipschitz continuity. (Hint: If $L > 0$ is given, how do you pick $x, y \in [0, 1]$ such that $x < y$ and $\dfrac{\sqrt{y} - \sqrt{x}}{y - x} > L$?)

9. Prove that sin and cos are uniformly continuous on all of $\mathbb{R}$ in each of the following two ways:

   (a) Theorem 3.69 shows that sin and cos are uniformly continuous on $[-2\pi, 2\pi]$. Use this and the fact that sin and cos have period 2π to prove uniform continuity on all of $\mathbb{R}$. (Hint: If $x, y \in \mathbb{R}$ are close together, then find a multiple of 2π that can be subtracted from both x and y to make them both lie in $[-2\pi, 2\pi]$.)

   (b) For the second proof, use the trigonometric identities

$$\sin x - \sin y = 2 \sin\left(\frac{x-y}{2}\right) \cos\left(\frac{x+y}{2}\right)$$

   and

$$\cos x - \cos y = -2 \sin\left(\frac{x-y}{2}\right) \sin\left(\frac{x+y}{2}\right)$$

   which are valid for all $x, y \in \mathbb{R}$, and the inequality $|\sin\theta| \leq |\theta|$ for $\theta \in (-\pi/2, \pi/2)$, to prove that both sin and cos are Lipschitz-continuous on $\mathbb{R}$ with Lipschitz constant 1 (see Exercise 3.12.8).

10. Suppose that f is uniformly continuous on an open interval (a, b) with $a < b$. Prove that f is bounded on (a, b).

    (Hint: Suppose that, in the definition of uniform continuity, $\epsilon = 1$, and δ is chosen correspondingly. Let $c = (a+b)/2$ be the midpoint of (a, b). Suppose we break up (a, c) and (c, b) into n subintervals, where n is chosen so that each subinterval has width less than δ. If x is a point in (a, b) which is k subintervals away from c, where $0 \leq k \leq n$, then how large can $|f(x) - f(c)|$ be in terms of k?)

    Remark. It turns out that the hypotheses of this exercise imply that $\lim_{x \to a^+} f(x)$ and $\lim_{x \to b^-} f(x)$ both exist, so f can be extended to a continuous function g defined on $[a, b]$. However, the proof of that is better suited for an upper-level real analysis course.

Chapter 4

Derivatives

In calculus, the derivative is used to study rates of change. Rates of change appear in many different situations. For instance, the slope of a line measures the rate of change of vertical position with respect to horizontal position (this is more casually called "change in y over change in x" or "rise over run"). When an object is moving, its velocity measures the rate of change of its position with respect to time. Also, acceleration measures the rate of change of velocity with respect to time. For all these applications and more, derivatives come in handy. Let's illustrate the main idea with an example.

Example 4.1:

Suppose that we watch an apple fall from a tree which is 10 meters tall. After t seconds of falling, let's say the apple is $f(t)$ meters above the ground. Newton found that $f(t) = 10 - 4.9t^2$ for any time t until the apple lands. To find out when the apple lands, we set $f(t) = 0$, i.e. $10 - 4.9t^2 = 0$, and we solve for t to get $t = \pm\sqrt{10/4.9} = \pm\sqrt{100/49} = \pm 10/7$. Since we're more interested in the positive answer (a negative number for time could make sense in other situations, referring to the past, but $f(t)$ wouldn't make sense for negative t in this context), we conclude that the apple hits the ground after $10/7$ seconds. This tells us that f is a function from $[0, 10/7]$ to $[0, 10]$; let's call f the position function for the apple.

That apple falls quite quickly; it falls 10 meters in $10/7$ seconds, so the *average velocity* of the apple during the entire fall is

$$\frac{\text{change in height}}{\text{change in time}} = \frac{(0-10)\,\text{meters}}{(10/7-0)\,\text{sec}} = -7\,\text{meters/sec}$$

(Here, *velocity* measures speed AND direction, so the negative sign means that the apple is falling, rather than rising.) More generally, for any $a, b \in [0, 10/7]$ with $a < b$, the average velocity from time $t = a$ to time $t = b$ is the change in position $f(b) - f(a)$ divided by the change in time $b - a$. We compute

$$\begin{aligned} \frac{f(b) - f(a)}{b - a} &= \frac{(10 - 4.9b^2) - (10 - 4.9a^2)}{b - a} \\ &= -4.9\frac{b^2 - a^2}{b - a} \\ &= -4.9\frac{(b-a)(b+a)}{b-a} = -4.9(b+a) \end{aligned}$$

Frequently, in science, the capital Greek letter Delta, written as Δ, is used to denote changes. Thus, by using Δt to represent the change in time (i.e. $\Delta t = b - a$) and Δf to represent the change in position (i.e. $\Delta f = f(b) - f(a)$), we can write the average velocity as

$$\frac{\Delta f}{\Delta t} = -4.9(b + a)$$

This form of writing the velocity shows more clearly that velocity is a rate of change of position with respect to time.

Now, we'd like to know about the apple's *instantaneous velocity*: at any time $a \in [0, 10/7]$, how fast is the apple falling exactly at time a? In fact, how do we even assign meaning to the notion of instantaneous velocity at one point, rather than average velocity between two points?

We try the following approach. Let's suppose $a \in [0, 10/7]$ is given. If b is any time close to a, then the velocity of the apple at time a should be pretty close the average velocity from time a to time b. We found that this average velocity is $-4.9(b + a)$ meters per second. (Although we did our earlier calculations assuming that $a < b$, the calculations also work if $b < a$.) As b gets closer to a, the average velocity should become a better approximation to the instantaneous velocity at a. This suggests that we should take a limit of the average velocity as $b \to a$, and we find that

$$\lim_{b \to a} -4.9(b + a) = -9.8a$$

(Technically, if $a = 0$ or $a = 10/7$, then this limit is a one-sided limit, because we did not define f outside of the interval $[0, 10/7]$.)

This suggests that if we use $v : [0, 10/7] \to \mathbb{R}$ to represent a velocity function, where $v(a)$ is the instantaneous velocity at time a for each $a \in [0, 10/7]$, then we should let $v(a) = -9.8a$. The function v has some properties that we would expect a velocity function to have. For instance, when $a = 0$, i.e. the apple has only just started falling, we have $v(a) = 0$, signifying that the apple has not yet built up any speed. We see that v is a strictly decreasing function; this makes sense, as the apple falls faster with more time, since gravity makes the apple accelerate. As a approaches $10/7$, i.e. when the apple is close to landing, the velocity of the apple approaches $\lim_{a \to 10/7^-} v(a)$, which is $-9.8(10/7) = -14$ meters per second. Thus, this apple has a lot of momentum when it lands, so you definitely don't want this apple to land on your foot. $\square$

In calculus terminology, we say that v is the *derivative* of f. The value $v(a)$ is obtained by taking an average rate of change of f from a to b and then letting b approach a in a limit. We will study this process in general, trying to find out how properties of a derivative relate to properties of the original function used to construct it.

4.1 The Definition of Derivative

Example 4.1 motivates the following definition:

Definition 4.2. Suppose that f is a real function and $a \in \mathrm{dom}(f)$ is given. For any $x \in \mathrm{dom}(f)$ satisfying $x \neq a$, we call the ratio

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

the *difference quotient* of f from a to x. (This is also a rate of change: it is the ratio of change in output to change in input from a to x.) We say that f has a *derivative at a*, or is *differentiable at a*, if the difference quotient attains a limit as $x \to a$, i.e. if the limit

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

exists.[1] This limit is called the *derivative of f at a (with respect to the variable x)* and is written $f'(a)$ (the notation is pronounced as "f prime at

[1] Note that in order for this limit to exist, we require f to be defined in an open interval around a.

a"). If S is the set of points $a \in \mathbb{R}$ where $f'(a)$ exists, then the function $f' : S \to \mathbb{R}$ which maps each point $a \in S$ to $f'(a)$ is called the *derivative of f*.

Thus, in Example 4.1, another name for v is f'. Technically, since the domain of f in the example is a closed interval $[0, 10/7]$, the values $v(0)$ and $v(10/7)$ are really one-sided limits of the difference quotient. In other words, we say that $v(0)$ and $v(10/7)$ are *one-sided derivatives*, where $v(0)$ is the *derivative from the right* of f at 0, and $v(10/7)$ is the *derivative from the left* of f at $10/7$.

Remark. As another way to calculate the derivative, if we use the letter h to represent the change from a to x, i.e. $h = x - a$, then we get $x = a + h$ and the difference quotient becomes

$$\frac{f(a+h) - f(a)}{h}$$

Therefore, $f'(a)$ is also equal to

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

when this limit exists. This form of the derivative is frequently easier to calculate.

Remark. Note that instead of writing $f(x) = 10 - 4.9x^2$ and $f'(x) = -9.8x$, we will sometimes write $(10 - 4.9x^2)' = -9.8x$. This is technically an abuse of notation, as we blur the distinction between the function f and its value at an arbitrary x. However, when we write an expression like this, it should not cause confusion as long as we make clear which letter denotes the input variable to the function and which quantities depend on that variable.

Examples and Tangent Lines

Let's compute some derivatives of simple functions, and we'll see what the derivatives tell us about their graphs.

Example 4.3:

Let's consider when $f : \mathbb{R} \to \mathbb{R}$ is the constant function $f(x) = c$, where c is some real number. Since the graph of the function never rises nor falls, i.e. the output values always have zero change, it makes sense that the rate of change of the function should be zero. Indeed, for any $a \in \mathbb{R}$, we compute

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

Now, here's a good question to ponder: is the converse true? In other words, if a function's derivative is always zero, does that mean that the function is constant? This turns out to be surprisingly subtle! For instance, consider the Heaviside function H defined in Definition 3.42:

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

When $a < 0$, then $H(x) = H(a)$ for all $x < 0$, so the difference quotient $(H(x) - H(a))/(x - a)$ is zero for all $x < 0$. Thus, $H'(a) = 0$ when $a < 0$. Similarly, if $a, x > 0$, then $H(x) = H(a)$, so $H'(a) = 0$ as well. Thus, $H'(a)$ is zero for all $a \neq 0$, but clearly H is not a constant function on $\mathbb{R}$.

The issue here is that H is not continuous at 0; its values jump from 0 to 1 at that point. We'll soon show that this implies H is not differentiable at 0 (or see Exercise 4.2.7.(a)). It turns out that if a function has derivative zero IN AN ENTIRE INTERVAL, then the function is constant on that entire interval. We will prove this surprisingly deep result later in the chapter. □

Example 4.4:

Now, let's consider a linear function $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = mx + b$ where $x, m, b \in \mathbb{R}$. When we graph this line with the equation $y = mx + b$, the slope of this line is m. This means that for every 1 unit the line moves to the right, the line moves m units upward (if m is negative, then moving m units upward is the same as moving $|m|$ units downward). In other words, the rate of change of the y-position with respect to the x-position is m.

Thus, we'd expect our derivative to equal m. Indeed, we compute

$$f'(x) = \lim_{h\to 0} \frac{(m(x+h) + b) - (mx + b)}{h} = \lim_{h\to 0} \frac{mh}{h} = m$$

for all $x \in \mathbb{R}$. It is because of this that the derivative $f'(x)$ is sometimes called the *instantaneous slope*. □

Let's look at the notion of instantaneous slope a little more carefully, going back to the apple falling in Example 4.1. Here, the position function is $f(t) = 10 - 4.9t^2$ for $t \in [0, 10/7]$. If $a, b \in [0, 10/7]$ are given with $a \neq b$, then the average velocity of the apple from time a to time b was found to be $-4.9(b+a)$. In other words, the average slope of f from a to b is $-4.9(b+a)$, representing the slope of the solid line segment connecting the two points in Figure 4.1.

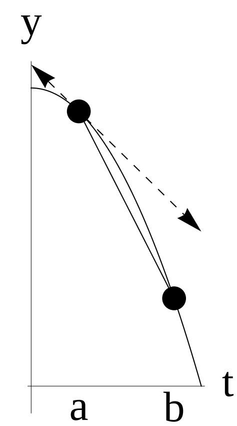

Figure 4.1: Average and instantaneous slopes

The line through the two points $(a, f(a))$ and $(b, f(b))$ is called a *secant line* to the graph of f. This secant line goes through $(a, f(a))$ and has slope $-4.9(b+a)$. As b approaches a, the slope of this line approaches $-9.8a$, which is the value of $f'(a)$. Thus, the secant lines start to look more like the line through $(a, f(a))$ with slope $-9.8a$ as the second point $(b, f(b))$ on the line approaches $(a, f(a))$. (This line is the dashed line in Figure 4.1.) In essence, the line through $(a, f(a))$ with slope $-9.8a$ goes "in the same direction" as f when $t = a$. This line is given a special name:

Definition 4.5. Suppose f is a real function, and $a \in \mathbb{R}$ is given such that f is differentiable at a. Therefore, the point $(a, f(a))$ is on the graph of f, and the instantaneous slope there is $f'(a)$. The *tangent line* to f at a is the line through $(a, f(a))$ with slope $f'(a)$. In point-slope form, this line has the equation

$$y - f(a) = f'(a)(x - a)$$

Thus, in our example, at any $a \in [0, 10/7]$, the tangent line has the equation

$$y - (10 - 4.9a^2) = -9.8a(t - a)$$

for the height y as a function of the time t. Here, the tangent line has a physical meaning. If, at the time $t = a$, gravity stops pulling on the apple, then Newton's laws of motion say that the apple will continue moving according to its current velocity until some other force acts on it. Thus, from the point $(a, f(a))$, the apple would fall downward at its current instantaneous speed of $9.8a$ meters per second. In other words, the position function of the apple without gravity's influence will follow the tangent line!

Remark. You may have heard of the terminology "tangent line" before in geometry, where a tangent line to a circle is a line touching the circle at exactly one point. For graphs other than circles, it's not as important whether a line touches the graph at exactly one point. For instance, the vertical line through the point $(a, f(a))$ touches the graph of f exactly once, but that line tells us practically nothing about the behavior of f overall! That is why the tangent line is instead defined using instantaneous slope, since that line has the same trajectory as the graph it is touching.

However, the two notions of tangent line are related. In Section 4.5, when we have more techniques for computing derivatives, we will find tangent lines to circles using our definition of a tangent line, and we'll see that our tangent lines touch the circle in exactly one spot.

Example 4.6:

Let's see another example of computing tangent lines, and we'll find another use for tangent lines as good approximations to a function.

Let's consider the function $f : [0, \infty) \to [0, \infty)$ defined by $f(x) = \sqrt{x}$ for all $x \in [0, \infty)$. Where is this function differentiable? First, let's suppose $a \in (0, \infty)$ is given. By rationalizing, we compute

$$\begin{aligned} f'(a) &= \lim_{h \to 0} \frac{\sqrt{a+h} - \sqrt{a}}{h} \cdot \frac{\sqrt{a+h} + \sqrt{a}}{\sqrt{a+h} + \sqrt{a}} \\ &= \lim_{h \to 0} \frac{(a+h) - a}{h(\sqrt{a+h} + \sqrt{a})} \\ &= \lim_{h \to 0} \frac{1}{\sqrt{a+h} + \sqrt{a}} = \frac{1}{2\sqrt{a}} \end{aligned}$$

(This is valid because $\sqrt{a}$ is not zero when $a > 0$.) Thus, the tangent line at a has the equation

$$y - \sqrt{a} = \frac{1}{2\sqrt{a}}(x - a)$$

We draw one such tangent line in Figure 4.2. The picture shows that the tangent line seems to be following the general trajectory of the graph of f at the point $(a, f(a))$. One way to think of this is to imagine yourself riding in a rollercoaster car, and the track is shaped like the graph of f. If you always keep your vision parallel with the coaster, then when the car reaches the point $(a, f(a))$, the tangent line shows your line of vision. Notice that you get the same line of vision whether you ride the car towards the right (i.e. forwards on the track) or if you ride the car towards the left (i.e. backwards). The line also shows the trajectory that the coaster would take if the car detached from the track and flew in a straight line unaffected by gravity.

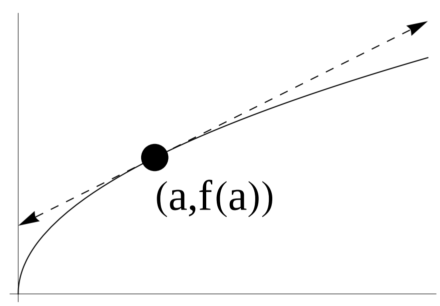

Figure 4.2: A tangent line to $y = \sqrt{x}$ at $x = a$

Because the tangent line seems to be "following the graph of f" at $x = a$, the line provides a decent way to approximate values of $\sqrt{x}$ when x is close to a. For instance, let's consider when $a = 1$. Then the tangent line has equation $y - 1 = (1/2)(x - 1)$, i.e. $y = x/2 + 1/2$. When x is close to 1, the y-coordinates of the graph $y = \sqrt{x}$ and $y = x/2 + 1/2$ shouldn't be far apart. Thus, when $x = 1.001$, we have

$$\sqrt{1.001} \approx (1.001)/2 + 1/2 = 1.0005$$

A calculator gives $\sqrt{1.001} \approx 1.000499875$, so our estimate of $\sqrt{1.001}$ using the tangent line was pretty accurate. This method of using the tangent line to make approximations for a curve is often called the process of *linear approximation*.

Another thing to notice about this function is that it is not differentiable at 0. If we try and compute the derivative there, we find

$$\lim_{h\to 0^+} \frac{\sqrt{0+h}-\sqrt{0}}{h} = \lim_{h\to 0^+} \frac{1}{\sqrt{h}} = \infty$$

(The definition of an infinite limit was given in the exercises in Section 3.4.) We should note that just saying "the formula $1/(2\sqrt{a})$ is undefined when $a = 0$" is NOT enough to prove that $f'(0)$ doesn't exist. This is because in our calculation for $f'(a)$ when $a > 0$, we used the quotient rule for limits in the last step. The quotient rule doesn't apply to limits where the denominator approaches 0, so we aren't allowed to use it when $a = 0$. This means a separate calculation of the limit for $a = 0$ is necessary, and this calculation gives the answer of ∞.

Thus, the difference quotients from $a = 0$ approach ∞ as $h \to 0$, so intuitively the graph is rising infinitely fast at $a = 0$. We say that f has a *vertical tangent* at $a = 0$. Indeed, if you imagine yourself in the rollercoaster that travels along the graph of f, at the very start of your trip, you will be facing straight up. $\square$

Two Notations

The notation f' for the derivative of f is often called *Newton notation* for the derivative. There is another common notation for derivatives called *Leibniz notation* which writes $\frac{df}{dx}$ for the derivative of f with respect to the input variable x. The idea behind this notation is as follows: if we write Δf to denote $f(x) - f(a)$, i.e. the change in f from x to a, and we write Δx to denote $x - a$, i.e. the change in the input variable x, then the difference quotient of f from a to x becomes

$$\frac{f(x)-f(a)}{x-a} = \frac{\Delta f}{\Delta x}$$

As $x \to a$, we have $\Delta x \to 0$, and thus

$$\frac{df}{dx} = \lim_{\Delta x\to 0} \frac{\Delta f}{\Delta x}$$

(so essentially, the "Greek letter d", Δ, becomes a "Roman letter d" after the limit is taken). The derivative $\frac{df}{dx}$ in Leibniz notation is frequently pronounced "$d\ f\ d\ x$", i.e. you pronounce the four letters separately. Also, the word "over" is not said when reading the symbol $\frac{df}{dx}$, as the symbol is not technically a fraction. If our input variable is t instead of x, we write dt instead of dx, and so forth.

In this notation, here are some of the derivatives we have found, where $c, m, b \in \mathbb{R}$ are constants:

$$\frac{dc}{dx} = 0 \qquad \frac{d}{dt}(10 - 4.9t^2) = -9.8t \qquad \frac{d}{dx}(mx + b) = m$$

The second of these is read as "$d\ d\ t$ of $10 - 4.9t^2$ is $-9.8t$", where the derivative is now with respect to t. In essence, the symbol $\frac{d}{dx}$ represents the operation of taking a derivative with respect to x.

To denote the value of the derivative at a point a in Leibniz notation, we write

$$\left.\frac{df}{dx}\right|_{x=a}$$

Many formulas will be quicker to write in Newton notation. However, we'll see several formulas later in this chapter which become clearer when written in Leibniz notation. This is primarily because the Leibniz notation writes the derivative to look like a quotient, reminding us that a derivative arises as a limit of a difference quotient. Also, the Leibniz notation's "denominator" reminds us what the input variable is, which can help avoid confusion in problems with multiple variables.

Higher-Order Derivatives

Since the derivative of a real function is another real function, we can certainly consider taking derivatives again. In fact, we can try taking derivatives as many times as desired.

Definition 4.7. If f is a real function and $n \in \mathbb{N}$, the n^{th} *derivative* $f^{(n)}$ of f (also called the n^{th}*-order derivative*) is the function obtained by differentiating f a total of n times. More formally, we define $f^{(n)}$ by recursion on n. As a base case, $f^{(0)}$ is f. Next, if $f^{(n)}$ has been created, then $f^{(n+1)}$ is $(f^{(n)})'$. For example,

$$f^{(3)} = (f^{(2)})' = ((f^{(1)})')' = (((f^{(0)})')')' = ((f')')'$$

For any $a \in \mathbb{R}$, if $f^{(n)}(a)$ exists, we say that f is *n-times differentiable* at a.

Note that the parentheses around n in the superscript are REQUIRED: the phrase $f^n(a)$ is interpreted as the n^{th} power of f at a instead of being interpreted as the n^{th} derivative. For convenience, $f^{(2)}$ is frequently written f'', and $f^{(3)}$ is written f'''. The notation $f^{(n)}$ is Newton notation for the n^{th} derivative. In Leibniz notation, the n^{th} derivative of f is written as

$$\frac{d^n f}{dx^n} \qquad \text{where} \qquad \frac{d^{n+1} f}{dx^{n+1}} = \frac{d}{dx}\left(\frac{d^n f}{dx^n}\right)$$

Thus, in Leibniz notation, the n^{th} derivative is written like the "n^{th} power" of the operation $\frac{d}{dx}$.

Lastly, when $a \in \mathbb{R}$ and $f^{(n)}(a)$ exists for every $n \in \mathbb{N}$, we say that f is *infinitely differentiable* at a.

Example 4.8:

With the apple in Example 4.1, when $f(t) = 10 - 4.9t^2$, we found $f'(t) = -9.8t$. Taking a second derivative is easy, because f' is a linear function, and we get $f''(t) = -9.8$. This shows that the second derivative of f is a constant.

What physical meaning does the second derivative have? We saw that the first derivative of f, i.e. the first derivative of the apple's position, corresponds to velocity. The second derivative measures rate of change of velocity, or how fast velocity changes with respect to time. This is called acceleration. Thus, since $f''(t) = -9.8$, we see that the apple is accelerating downward at the constant rate of -9.8 meters per seconds squared. This is the commonly accepted value of gravitational attraction for objects near the surface of Earth: it is historically denoted by $g = 9.8\,\text{meters/sec}^2$.

Remark. Historically, none of Newton's Laws actually gives a formula for the position of a falling object. Instead, Newton's famous Law of Universal Gravitation says that

$$F = G\frac{Mm}{R^2}$$

where F is the magnitude of the force due to gravity, G is a constant, M and m are the masses of two objects drawn together by gravity, and R is the distance between the objects. For objects near Earth's surface, the distance R is roughly the radius of Earth, so if we let M be the mass of Earth and let $g = GM/R^2$, then we get the approximation $F \approx mg$ for describing the force of gravity that pulls an object of mass m towards Earth.

Since force is mass times acceleration, this shows that near Earth's surface, gravitational acceleration is roughly the constant value $-g$ (the negative sign indicates a downward direction). In other words, we first find that $f''(t) = -9.8$. From that, we will later be able to show that if v_0 is the initial velocity of the apple (in our case, $v_0 = 0$, meaning the apple has not built up any speed yet), then f' must satisfy the formula $f'(t) = v_0 - 9.8t$. After that, we'll also see that if f_0 is the initial height of the apple (so in our case, $f_0 = 10$), then f must satisfy the formula $f(t) = f_0 + v_0 t - 4.9t^2$.

It is not hard to check that if f is defined by the formula $f(t) = f_0 + v_0 t - 4.9t^2$, then $f(0) = f_0$, $f'(0) = v_0$, and $f''(t) = -9.8$. The hard part is knowing that this is the UNIQUE function f to have all those properties. This ends up being related to our earlier question about whether a function with zero derivative must be a constant function. Later in the chapter, we'll prove some powerful theorems that will address the uniqueness of the function f above.

Returning to the example of the apple falling from the tree, since the second derivative is constant, the third derivative is $f^{(3)}(t) = 0$ for all $t \in \mathrm{dom}(f)$. Thus, since the third derivative is constant, the fourth derivative is also zero, and so forth. In summary, f is infinitely differentiable at every value in $[0, 10/7]$, but only $f^{(0)}$, $f^{(1)}$, and $f^{(2)}$ are not constantly zero. □

4.2 Exercises

1. Find each of the following derivatives in terms of the input variable x. For example, $(10 - 4.9x^2)'$ is $-9.8x$, as we showed earlier. Show your calculations using the definition of derivative.

   (a) $(x^2 + x + 1)'$.

   (b) $(x^3)'$.

   (c) $(1/x)'$ where $x \neq 0$. (Hint: Use a common denominator.)

2. For each part, find the tangent line to the function f at the specified value of a. Also, find all points where the tangent line is horizontal.

(a) $f(x) = x^3$, $a = 2$.

(b) $f(x) = x + \sqrt{x}$, $a = 1$.

(c) $f(x) = \dfrac{1-x}{2+x}$, $a = -1$.

3. By using a linear approximation (see Example 4.6) to the graph of $y = x^4$ near $a = 2$, find approximate values for $(2.001)^4$ and $(1.999)^4$.

4. Find the following second derivatives in terms of x.

   (a) $((x+1)^2)''$

   (b) $(\sqrt{x})''$

5. For each $n \in \mathbb{N}^*$, define $f_n : \mathbb{R} \to \mathbb{R}$ by $f_n(x) = x^n$ for each $x \in \mathbb{R}$. Thus, f_n is the n^{th} power function. This problem will find f_n'.

   (a) Prove that for all $x, a \in \mathbb{R}$ and all $n \in \mathbb{N}^*$,

   $$x^n - a^n = (x-a)\sum_{i=0}^{n-1} x^i a^{n-1-i}$$

   (b) Let $n \in \mathbb{N}^*$ and $a \in \mathbb{R}$ be given. Use part (a) to prove that $f_n'(a) = na^{n-1}$.

   (c) Let $n \in \mathbb{N}^*$ and $x \in \mathbb{R}$ be given. The *Binomial Theorem* says that for all $h \in \mathbb{R}$,

   $$(x+h)^n = \sum_{i=0}^{n} \binom{n}{i} x^i h^{n-i}$$

   where $\binom{n}{i}$, pronounced "n choose i", is the number

   $$\binom{n}{i} = \frac{n(n-1)(n-2)\ldots(n-i+1)}{i!} = \frac{n!}{(n-i)!\,i!}$$

   (this is frequently proven in discrete mathematics courses). Assuming the Binomial Theorem, prove $f_n'(x) = nx^{n-1}$. (You are not required to prove the Binomial Theorem, though an induction proof is possible; look up Pascal's Identity.) This provides another way to find the derivative of the n^{th} power function.

6. Let $n \in \mathbb{N}^*$ be given. Let $f : \mathbb{R} - \{0\} \to \mathbb{R}$ be defined by $f(x) = x^{-n} = 1/(x^n)$ for all $x \neq 0$. Use Exercise 4.2.5 to prove that $f'(x) = -nx^{-n-1}$ for all $x \neq 0$. (Hint: After making a common denominator, group the terms so that the difference quotient for the n^{th} power appears.)

Remark. Therefore, since $(x^0)' = 1' = 0 = 0x^{-1}$, this exercise and Exercise 4.2.5 prove that for all $n \in \mathbb{Z}$, $(x^n)' = nx^{n-1}$ at all $x \in \mathbb{R}$ where x^n is defined.

7. (a) Prove, using the definition of derivative, that the Heaviside function H is not differentiable at 0 (H is defined in Definition 3.42).
   (b) Prove, using the definition of derivative, that for all $a \in \mathbb{R}$, $\chi_{\mathbb{Q}}$ is not differentiable at a, where $\chi_{\mathbb{Q}}$ is defined in Definition 3.42. (Hint: Show that when $|h| < 1$, the difference quotient from a to $a + h$ is either 0 or has absolute value greater than 1. Why does that mean the difference quotient has no limit?)

4.3 Arithmetic Derivative Rules

Now that we've introduced the definition of derivative, we'd like some rules, much like with limits, that will make derivatives easier to compute. Some of the derivative rules are just as simple as the corresponding limit rules, but some of the derivative rules are more complicated.

How Far Off Is The Tangent Line?

Before introducing any derivative rules, let's first examine the connection between functions and tangent lines a little more closely. There are a couple good reasons for doing this. First, whenever making any kind of approximation, it is always worthwhile to make precise statements concerning how close the approximation actually is. This is what makes estimation practical and insightful, as opposed to just wishful thinking. Second, our study here generalizes in some very handy ways. One generalization, which we will study extensively in Chapter 8, deals with approximating functions by polynomials of arbitrary degree. Another generalization arises in multivariate calculus, where we take functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ (where $n, m \in \mathbb{N}^*$) and compute something called the "total derivative" by analyzing approximations by a suitable generalization of a linear function. This study is beyond the scope of this book.[2]

[2] However, after completing this book, you should study this topic, and more, in a formal multivariate calculus course.

Let's suppose f is a real function which is differentiable at $a \in \mathbb{R}$. Therefore, the instantaneous slope of f at a is $f'(a)$, meaning that the graph roughly follows the line with slope $f'(a)$ through the point $(a, f(a))$ on the graph of f. In other words, using the formula for the tangent line, we have

$$f(x) \approx f(a) + f'(a)(x-a)$$

when x is close to a. We'd like to analyze how far apart the tangent line and the original function are. Thus, we define

$$E(x) = f(x) - (f(a) + f'(a)(x-a))$$

for all $x \in \mathrm{dom}(f)$, so E is a real function with $\mathrm{dom}(E) = \mathrm{dom}(f)$. We'll call E the *error* of the linear approximation $f(a) + f'(a)(x-a)$ (or, more specifically, the error with respect to the function f at the value a).

Note that $E(a) = 0$; this makes sense because both the function f and the tangent line go through the point $(a, f(a))$, so there is no error when $x = a$. We'd like to show that, in addition, $E(x) \to 0$ as $x \to a$. In other words, we can make the error as close to 0 as we'd like by taking x close enough to a.[3] After all, we don't use approximations to estimate the value of f at a; we use them instead to estimate the values of f NEAR a.

To show this, we first note that when $x \neq a$, we can divide by $x - a$ and obtain

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

Since $f'(a)$ is, by definition, the limit of the difference quotient $(f(x) - f(a))/(x-a)$ as x approaches a, this tells us that

$$\lim_{x \to a} \frac{E(x)}{x-a} = f'(a) - f'(a) = 0$$

Using this limit, we readily find that

$$\lim_{x \to a} E(x) = \lim_{x \to a} \left(\frac{E(x)}{x-a} \right) (x-a) = 0(a-a) = 0$$

as desired.

Using this limit, we obtain a useful connection between the notions of differentiability (i.e. the notion of having a derivative) and continuity:

[3]In Chapter 8, when we study Taylor polynomials, we'll be able to obtain estimates for the size of $E(x)$. Another estimate is found in Exercise 4.10.17.

Theorem 4.9. *If $a \in \mathbb{R}$ and f is a real function which is differentiable at a, then f is continuous at a.*

Strategy. Using the notation from earlier, we write

$$f(x) = f(a) + f'(a)(x-a) + E(x)$$

We'd like to show that $f(x) \to f(a)$ as $x \to a$. The work we did earlier helps us find the limit of the right-hand side.

Proof. Let f, a be given as described. Let $E : \mathrm{dom}(f) \to \mathbb{R}$ at a be the error function described earlier. We have previously shown that

$$f(x) = f(a) + f'(a)(x-a) + E(x)$$

for all $x \in \mathrm{dom}(f)$, and also $E(x) \to 0$ as $x \to a$. Therefore,

$$\begin{aligned} \lim_{x\to a} f(x) &= \lim_{x\to a}(f(a) + f'(a)(x-a) + E(x)) \\ &= f(a) + f'(a)(a-a) + 0 = f(a) \end{aligned}$$

which proves that f is continuous at a. ■

Example 4.10:

While differentiability implies continuity, the converse is NOT true. Consider when f is the absolute-value function and $a = 0$. We know that f is continuous at 0. However, we compute the one-sided limits of the difference quotient at $a = 0$:

$$\lim_{h\to 0^+} \frac{|0+h| - |0|}{h} = \lim_{h\to 0^+} \frac{h}{h} = 1$$

and

$$\lim_{h\to 0^-} \frac{|0+h| - |0|}{h} = \lim_{h\to 0^-} \frac{-h}{h} = -1$$

which show that the two-sided limit, i.e. $f'(0)$, does not exist.

Basically, from the left, the graph looks like it has slope -1, and from the right, the graph looks like it has slope 1. Recall the rollercoaster analogy we developed, where we imagine ourselves riding on a rollercoaster shaped like the graph of the function. With that analogy, tangent lines correspond to lines of vision when we face parallel to the track. However, if a rollercoaster were riding a track shaped like the absolute-value function, then the rider would be severely whipped around at $x = 0$, rapidly switching from a diagonal downward-pointing line of vision to a diagonal upward-pointing line of vision.

This situation is frequently described by saying that the graph has a *sharp corner*. Sharp corners are one of the ways that a continuous function can fail to be differentiable. Another way is for the graph to have a vertical tangent, as we've seen with the square-root function near 0. □

At this point, we've seen that the tangent line to a function f at a is a suitable approximation to f, because it produces small errors $E(x)$ for points x near a. In fact, we know more than just $E(x) \to 0$; we've actually shown the stronger property that $E(x)/(x-a) \to 0$. This tells us that $E(x)$ is NOT proportional to $x - a$, since if $E(x)$ were roughly $r(x - a)$ for some nonzero ratio of proportion r, then the limit $E(x)/(x - a)$ would come out to r, not to 0. Instead, in some sense, $E(x)$ is significantly smaller than $x - a$. Thus, we sometimes describe $E(x)/(x - a) \to 0$ by saying that the error is "better than linear".

This raises the question: are there any other linear approximations to f which produce "better than linear" errors? Suppose we'd like to approximate $f(x)$ for x near a with a line of slope m going through the point $(a, f(a))$ on the graph of f. Thus, the line we use for approximation has the formula

$$y = f(a) + m(x - a)$$

and we use E again to stand for the error in the approximation:

$$E(x) = f(x) - (f(a) + m(x - a))$$

Our question now becomes: which choices of m make $E(x)/(x-a)$ approach 0 as $x \to a$? This useful lemma shows that the tangent line slope $f'(a)$ is the only choice:

Lemma 4.11. *Let $a, m \in \mathbb{R}$ be given, and let f be a real function defined in an interval around a. Define $E : \mathrm{dom}(f) \to \mathbb{R}$ by*

$$E(x) = f(x) - (f(a) + m(x - a))$$

for all $x \in \mathrm{dom}(f)$, so E represents the error in approximating $f(x)$ by the line through $(a, f(a))$ with slope m. Then $E(x)/(x - a) \to 0$ as $x \to a$ iff f is differentiable at a and $m = f'(a)$.

Strategy. The right-to-left direction of proof was done earlier, where we showed that using $m = f'(a)$ produces an approximation with a "better than linear" error. For the other direction, we find that

$$\frac{E(x)}{x-a} = \frac{f(x) - (f(a) + m(x-a))}{x-a} = \frac{f(x) - f(a)}{x-a} - m$$

This gives us a way to calculate the limit of the difference quotient, i.e. the derivative.

Proof. Let f, a, m, E be given as described. The right-to-left direction of proof was done in this section, so let's do the other direction. Assume that

$$\lim_{x \to a} \frac{E(x)}{x-a} = 0$$

The definition of E also shows that for all $x \in \text{dom}(f)$ with $x \neq a$,

$$\frac{E(x)}{x-a} = \frac{f(x) - f(a)}{x-a} - m$$

Therefore, we compute

$$\begin{aligned} f'(a) &= \lim_{x \to a} \frac{f(x) - f(a)}{x-a} \\ &= \lim_{x \to a} \frac{E(x)}{x-a} + m \\ &= 0 + m = m \end{aligned}$$

as desired. ■

Lemma 4.11 gives a formal way of saying that the tangent line is the best approximating line at a. In fact, in some situations, we can use this lemma to prove that $f'(a)$ exists: we find a linear approximation for f near a, and if the error is better than linear, then $f'(a)$ is the line's slope. This is esepcially handy when f is built out of simpler functions, so that a linear approximation for f can be built out of approximations for the simpler pieces. We will illustrate this process with some of our derivative laws.

Linearity of Differentiation

For our first rule for calculating derivatives, we'll show that the operation of taking derivatives is linear:

Theorem 4.12 (Linearity Rule). *For any real functions f, g and any real numbers a, c_1, c_2, if f and g are both differentiable at a, then so is the linear combination $c_1 f + c_2 g$. Furthermore,*

$$(c_1 f + c_2 g)'(a) = c_1 f'(a) + c_2 g'(a)$$

Recall that in Chapter 3, we showed that the operation of taking a limit is a linear operation as well, i.e.

$$\lim_{x \to a}(c_1 f + c_2 g)(x) = c_1 \lim_{x \to a} f(x) + c_2 \lim_{x \to a} g(x)$$

when the limits on the right side exist.

Remark. From the general "linearity rule", we get similar rules for constant multiples, addition, and subtraction by plugging in specific choices for the functions and constants. For instance, if we choose g to be the constant zero function, then Theorem 4.12 shows us that $(c_1 f)'(a) = c_1 f'(a)$, which gives us a constant multiple rule. On the other hand, if we let g stay arbitrary but instead choose $c_1 = 1$ and $c_2 = 1$, then we get $(f + g)'(a) = f'(a) + g'(a)$, which gives us an addition law. Similarly, we get $(f - g)'(a) = f'(a) - g'(a)$ by choosing $c_1 = 1$ and $c_2 = -1$.

Strategy. The definition of derivative says that

$$(c_1 f + c_2 g)'(a) = \lim_{x \to a} \frac{(c_1 f(x) + c_2 g(x)) - (c_1 f(a) + c_2 g(a))}{x - a}$$

By using some algebra to group the terms with f together and the terms with g together, we readily obtain the difference quotients for f and g.

Proof. Let f, g, a, c_1, c_2 be given as described. By definition, we have

$$\begin{aligned} (c_1 f + c_2 g)'(a) &= \lim_{x \to a} \frac{(c_1 f(x) + c_2 g(x)) - (c_1 f(a) + c_2 g(a))}{x - a} \\ &= \lim_{x \to a} \left(c_1 \cdot \frac{f(x) - f(a)}{x - a} + c_2 \cdot \frac{g(x) - g(a)}{x - a} \right) \end{aligned}$$

Because f is differentiable at a, the quotient involving f has the limit $f'(a)$ as $x \to a$. Similarly, the quotient involving g has the limit $g'(a)$. Thus, by the limit laws, we find $(c_1 f + c_2 g)'(a) = c_1 f'(a) + c_2 g'(a)$, as desired. ■

Remark. It's also instructive to see how we could have used our results about linear approximations to prove this theorem. The main idea is: if we have

$$f(x) \approx f(a) + f'(a)(x-a) \qquad \text{and} \qquad g(x) \approx g(a) + g'(a)(x-a)$$

then we also should have

$$c_1 f(x) + c_2 g(x) \approx (c_1 f(a) + c_2 g(a)) + (c_1 f'(a) + c_2 g'(a))(x-a)$$

More formally, let E_f be the better-than-linear error for f at a, and let E_g be the better-than-linear error for g at a. Some algebra shows that

$$\begin{aligned} c_1 f(x) + c_2 g(x) &= (c_1 f(a) + c_2 g(a)) + (c_1 f'(a) + c_2 g'(a))(x-a) \\ &\quad +(c_1 E_f(x) + c_2 E_g(x)) \end{aligned}$$

This yields a linear approximation for $c_1 f + c_2 g$ with slope $c_1 f'(a) + c_2 g'(a)$ whose error $E(x)$ is $c_1 E_f(x) + c_2 E_g(x)$. To apply Lemma 4.11, it remains to show that E is better than linear, i.e. that $E(x)/(x-a) \to 0$. This follows quickly from the fact that $E_f(x)/(x-a)$ and $E_g(x)/(x-a)$ approach 0.

This alternate argument can be summarized very nicely: if f and g both have good lines approximating them, then we can get a good line for $c_1 f + c_2 g$ by just scaling and adding our previous lines!

Theorem 4.12 shows that it is quite easy to take derivatives of sums and differences: we merely take the derivative of each piece individually and combine the results. For example, if we consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2 + x + 1$ for all $x \in \mathbb{R}$, then $f'(x) = (x^2)' + (x)' + (1)'$. Using Exercise 4.2.5, which proves that $(x^n)' = nx^{n-1}$ for all $n \in \mathbb{N}$, we therefore find that $f'(x) = 2x + 1$.

Another example to consider is Example 4.1, featuring the apple falling from the tree. The apple's height is $f(t) = 10 - 4.9t^2$ after t seconds. We can quickly find that the velocity of the apple, $f'(t)$, is $-4.9(t^2)'$, or $-9.8t$, without having to write a difference quotient!

Example 4.13:

Using Theorem 4.12 together with the formula $(x^n)' = nx^{n-1}$, we can compute derivatives of all polynomials quite easily. Let's say that p is a degree n polynomial where $n \in \mathbb{N}^*$, so p has the form $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, i.e $p(x) = \sum_{i=0}^{n} a_i x^i$, for all $x \in \mathbb{R}$ and some coefficients

$a_n, a_{n-1}, \cdots, a_0 \in \mathbb{R}$ with $a_n \neq 0$. (Recall that the degree of a polynomial is the highest power associated with a nonzero coefficient.) We find that

$$p'(x) = \left(\sum_{i=0}^{n} a_i x^i\right)' = \sum_{i=0}^{n} a_i (x^i)' = \sum_{i=0}^{n} i a_i x^{i-1}$$

In this derivative, the term corresponding to $i = 0$ has a coefficient of zero, so it can be removed from the sum. Hence,

$$p'(x) = \sum_{i=1}^{n} i a_i x^{i-1} = n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \cdots + 2a_2 x + a_1$$

This shows that p' is itself a polynomial. Since $na_n \neq 0$, p' is a polynomial of degree $n-1$. Thus, taking a derivative of a nonconstant polynomial (i.e. a polynomial with degree at least 1) lowers its degree by 1.

If $n-1$ is at least 1, then we can take another derivative to get p'', and p'' is a polynomial of degree $(n-1)-1$, or $n-2$. Once again, if $n-2 \geq 1$, then we can take a third derivative p''' to get a polynomial with degree $n-3$, and so forth. You can do a short induction proof to show that for any $n, k \in \mathbb{N}^*$ with $k \leq n$, the k^{th} derivative of an n^{th} degree polynomial is a polynomial of degree $n-k$.

However, what happens if you take more than n derivatives of an n^{th} degree polynomial? Well, the n^{th} derivative of an n^{th} degree polynomial has degree 0, i.e. it's a constant function. Thus, the next derivative after that is constantly zero. This shows that the $(n+1)^{\text{st}}$ derivative of an n^{th} degree polynomial, and all other derivatives after that, are constantly zero.

It turns out that, in fact, the ONLY functions whose $(n+1)^{\text{st}}$ derivatives are zero everywhere are the polynomials whose degree is at most n. (When we use $n = 0$, this fact implies that the only functions whose derivative is zero everywhere are the constant functions.) This follows from a deep result called the Mean Value Theorem, which will come in a later section of this chapter. $\square$

Products and Quotients

Next, we would like to find a way to take derivatives of products. Here, derivatives behave rather differently from limits: if f and g are differentiable at some value a, then $(fg)'(a)$ usually is NOT the same as $f'(a)g'(a)$. For instance, $(x^2)' = 2x$, but $(x)' \cdot (x') = 1 \cdot 1 = 1$. What should the formula for $(fg)'(a)$ be instead?

To gain some ideas, we can use linear approximations. If we say that

$$f(x) \approx f(a) + f'(a)(x-a) \qquad g(x) \approx g(a) + g'(a)(x-a)$$

then multiplying these expressions suggests that

$$f(x)g(x) \approx f(a)g(a) + f'(a)g(a)(x-a) + f(a)g'(a)(x-a) + f'(a)g'(a)(x-a)^2$$

This result doesn't give us a line because of the presence of the term with $(x-a)^2$. However, $(x-a)^2$ goes to 0 faster than $x-a$ does as $x \to a$, since $(x-a)^2/(x-a) \to 0$. This suggests that the term with $(x-a)^2$ gets counted as part of the linear error, and what remains is

$$f(a)g(a) + (f'(a)g(a) + f(a)g'(a))(x-a)$$

Thus, we are led to the following theorem:

Theorem 4.14 (Product Rule). *Let $a \in \mathbb{R}$ be given, and let f and g be real functions which are each differentiable at a. Then fg is differentiable at a, and*

$$(fg)'(a) = f'(a)g(a) + f(a)g'(a)$$

Remark. Here is a fun mnemonic device you may want to use to help remember this formula[4]. Let's call one of the functions "hi" and the other function "ho". If you use a "d" to represent derivative, the Product Rule becomes

$$\text{d(hi ho)} = \text{ho d(hi)} + \text{hi d(ho)}$$

This is intended to be pronounced "d hi ho is ho d hi plus hi d ho", sounding almost like a yodel. There will be an analogous device for remembering how to differentiate quotients later.

Strategy. We want to compute the limit

$$(fg)'(a) = \lim_{x \to a} \frac{f(x)g(x) - f(a)g(a)}{x-a}$$

[4]Warning: Many mnemonic devices may feel silly or embarassing to repeat to yourself, especially this one. Use this with caution.

It is not clear how to get difference quotients for f and g to appear in this expression. However, note that if the second term in the numerator were $f(a)g(x)$ instead of $f(a)g(a)$, then we could factor out $g(x)$ and get $(f(x) - f(a))g(x)$ in the numerator. This suggests that having a "mixed term" in the numerator, where one factor uses x and the other uses a, would be helpful.

To make a "mxed term" appear in the numerator, we use an old algebra trick: we add and subtract the term! Let's say, for instance, that we add and subtract $f(a)g(x)$. This turns our fraction into

$$\frac{f(x)g(x) - f(a)g(x) + f(a)g(x) - f(a)g(a)}{x-a}$$

Thus, in the first two terms, we can factor out $g(x)$, and in the last two terms, we can factor out $g(a)$. This makes difference quotients for f and g appear!

Proof. Let f, g, a be given as described. By adding and subtracting $f(a)g(x)$, we find that

$$\begin{aligned}
(fg)'(a) &= \lim_{x\to a} \frac{f(x)g(x) - f(a)g(a)}{x-a} \\
&= \lim_{x\to a} \frac{(f(x)g(x) - f(a)g(x)) + (f(a)g(x) - f(a)g(a))}{x-a} \\
&= \lim_{x\to a} \left(\frac{f(x)-f(a)}{x-a} g(x) + f(a)\frac{g(x)-g(a)}{x-a} \right)
\end{aligned}$$

The difference quotient with f approaches $f'(a)$, and the difference quotient with g approaches $g'(a)$. Also, since g is differentiable at a, it is also continuous at a by Theorem 4.9, so $g(x) \to g(a)$. This proves that $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$, as desired. ■

Example 4.15:

Theorem 4.14 allows us to obtain another proof of the derivative formula for powers, $(x^n)' = nx^{n-1}$ (where $n \in \mathbb{N}^*$), which does not require knowing any summation identities (unlike the proofs suggested in Exercise 4.2.5). We will use induction on n.

As a base case, when $n = 1$, we have proven earlier that $x' = 1$. Now, let $n \in \mathbb{N}^*$ be given, and we assume that $(x^n)' = nx^{n-1}$ as an induction hypothesis. We find the derivative of $(x^{n+1})'$ using the product rule as follows:

$$(x^{n+1})' = (x^n)' \cdot x + x' \cdot x^n = nx^{n-1}x + x^n = (n+1)x^n$$

This finishes the inductive step and proves our result.

This approach can be generalized to find a formula for the derivative of the n^{th} power of any differentiable function: see Exercise 4.4.7. $\square$

Example 4.16:

In the last section, we found that $(\sqrt{x})' = 1/(2\sqrt{x})$ at all $x > 0$ using the definition of derivative. Let's see how we can use the product rule to obtain this formula with less calculation. Let f be the square-root function, and therefore f satisfies the equation $f^2(x) = x$ for all $x \geq 0$. We take the derivative of each side with respect to x, treating the left side as the product of $f(x)$ with itself, and we get

$$f'(x)f(x) + f(x)f'(x) = \frac{d}{dx}(f^2(x)) = \frac{d}{dx}(x) = 1$$

(We used Leibniz notation here for the derivative of f^2 because it looks a little cleaner than $(f^2(x))'$.) When $x > 0$, we have $f(x) > 0$, and therefore we may solve for $f'(x)$ to obtain

$$f'(x) = \frac{1}{2f(x)} = \frac{1}{2\sqrt{x}}$$

The main tactic used here is to write an equation involving f (where the equation is simpler-looking than the formula for f itself) and then take derivatives of each side, solving for f' at the end. The equation involving f is said to define f *implicitly*, so this technique is often called *implicit differentiation*. (In contrast, a formula giving the exact values of f outright is an *explicit* definition.) We will have more to say about implicit differentiation in the next section.

The tactic of implicit differentiation works quite well with any rational power. If we let $n, m \in \mathbb{Z}$ be given with $n > 0$, and we let f be defined by $f(x) = x^{m/n}$ for all $x > 0$, then $f^n(x) = x^m$ for all $x > 0$. You can work out the rest of the details in Exercise 4.4.8. $\square$

Remark. There is a technical issue involved with using implicit differentiation. Consider what happens when trying to find the derivative of $f(x) = x^{m/n}$. When we write $f^n(x) = x^m$ and try to take the derivative of $f^n(x)$ using the product rule, one of the hypotheses of the product rule is that we have to ALREADY KNOW that f is differentiable. Therefore, we cannot actually prove that f is differentiable using implicit differentation.

Later in this chapter, we will introduce a theorem that lets us show that f' exists. Once we have that theorem, implicit differentiation provides a convenient way to obtain a formula for f'.

Now that we have analyzed derivatives of products, we'd also like derivatives of quotients. To make our calculations simpler, we'll first show how to find the derivative of a reciprocal. After this, if f and g are differentiable functions, then we will compute $(f/g)'$ by writing f/g as $f \cdot (1/g)$ and using the product rule.

In Exercise 4.2.6, we saw that $(1/x)' = (x^{-1})' = -1/x^2$ for all $x \neq 0$. A similar-looking law holds for arbitrary reciprocals, and fortunately the algebra is more straightforward than it was for the product rule:

Lemma 4.17 (Reciprocal Rule). *Let $a \in \mathbb{R}$ be given, and let g be a real function. If g is differentiable at a, and $g(a) \neq 0$, then $1/g$ is differentiable at a, with*

$$\left(\frac{1}{g}\right)'(a) = \frac{-g'(a)}{g^2(a)}$$

Strategy. As with our other proofs, we start with the definition:

$$\left(\frac{1}{g}\right)'(a) = \lim_{x\to 0} \frac{1/g(x) - 1/g(a)}{x-a}$$

To get rid of nested fractions, we will multiply and divide by something to clear out the fractions, such as $g(x)g(a)$. A little more work makes a difference quotient for g appear.

Proof. Let a, g be given as described. Since $g'(a)$ exists, g is continuous at a. Also, $g(a) \neq 0$, so $g(x)$ is not zero when x is close enough to a.

We use the definition of derivative and multiply and divide by $g(a)g(x)$ (which is valid since $g(x) \neq 0$ when x is near a) to compute

$$\begin{aligned}
\left(\frac{1}{g}\right)'(a) &= \lim_{x\to a} \frac{1/g(x) - 1/g(a)}{x-a} = \lim_{x\to a} \frac{g(a)-g(x)}{g(a)g(x)(x-a)} \\
&= \lim_{x\to a} \frac{1}{g(a)g(x)} \cdot \frac{-(g(x)-g(a))}{x-a}
\end{aligned}$$

The second fraction approaches $-g'(a)$, and the first fraction approaches $1/g^2(a)$ because g is continuous and nonzero at a. Thus, we have shown $(1/g)'(a) = -g'(a)/g^2(a)$. ■

Using Lemma 4.17, we obtain a rule for derivatives of quotients:

Corollary 4.18 (Quotient Rule). *Let $a \in \mathbb{R}$ be given, and let f and g be real functions which are both differentiable at a. If $g(a) \neq 0$, then f/g is differentiable at a, and*

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g^2(a)}$$

Remark. This "quotient rule" also has a fun mnemonic device to help remember it. If we call the numerator function "hi" and the denominator function "lo", and we use "d" to stand for derivative, then we get

$$\mathrm{d}\left(\frac{\mathrm{hi}}{\mathrm{lo}}\right) = \frac{\text{lo d(hi) - hi d(lo)}}{\text{lo lo}}$$

which is pronounced "d (lo over high) is lo d hi minus hi d lo over lo lo".

Strategy. We just need to put together our rules for products and reciprocals.

Proof. Let a, f, g be given as described. By Lemma 4.17, $1/g$ is differentiable at a, so using that in conjunction with Theorem 4.14,

$$\begin{aligned}
\left(\frac{f}{g}\right)'(a) &= \left(f \cdot \frac{1}{g}\right)'(a) \\
&= f'(a) \cdot \frac{1}{g(a)} + f(a) \cdot \left(\frac{1}{g}\right)'(a) \\
&= \frac{f'(a)g(a)}{g^2(a)} - \frac{f(a)g'(a)}{g^2(a)}
\end{aligned}$$

as desired. ■

Example 4.19:

With the aid of the quotient rule, we can easily find a formula for the derivative of x^{-n} where $n \in \mathbb{N}^*$. Lemma 4.17 tells us that

$$\left(x^{-n}\right)' = \left(\frac{1}{x^n}\right)' = \frac{-nx^{n-1}}{x^{2n}} = -nx^{-n-1}$$

for all $x \in \mathbb{R} - \{0\}$. This formula was previously found in Exercise 4.2.6 using the definition of derivative, but this approach is much quicker.

In fact, let's be more general. We've already seen that the derivative of a polynomial is another polynomial. Also, the sum, difference, or product of two polynomials is a polynomial. However, the quotient of two polynomials is not necessarily a polynomial (e.g. $1/x$): it is instead called a *rational function*. It is not hard to show that any sum, difference, product, or quotient of rational functions is also a rational function.

Using the quotient rule, we can show that the derivative of any rational function is also rational. To see this, suppose we take an arbitrary rational function and express it in the form p/q, where p and q are polynomials with no common zeroes. (This is similar to how we can always express rational numbers as fractions of integers with no common factors.) Because p and q have no common zeroes, the values x where $q(x) = 0$ are discontinuities of p/q (i.e. we don't get a "0/0" form by plugging in x). For all other values of x, we find

$$\left(\frac{p}{q}\right)'(x) = \frac{p'(x)q(x) - p(x)q'(x)}{q^2(x)}$$

The numerator and denominator of this expression are both polynomials, which shows us that the derivative of p/q is also a rational function.

Here are some more concrete examples of derivatives of rational functions:

$$\begin{aligned}
\left(\frac{x}{x+1}\right)' &= \frac{(x+1)x' - x(x+1)'}{(x+1)^2} = \frac{(x+1)-x}{(x+1)^2} = \frac{1}{(x+1)^2} \quad \text{if } x \neq -1 \\
\left(\frac{x}{1-x^2}\right)' &= \frac{(1-x^2)(1) - x(-2x)}{(1-x^2)^2} = \frac{1+x^2}{(1-x^2)^2} \quad \text{if } x \neq -1, 1
\end{aligned}$$

□

Trigonometric Derivatives

Up to this point, we've only found derivatives of polynomials, rational functions, and a few other related functions like the square-root function.

Let's find a couple other useful derivatives: the derivatives of the sine and cosine functions. Since these trigonometric functions don't arise by combining the other functions we've studied so far, we'll have to find these derivatives by using the definition of derivative.

Example 4.20:

Let $a \in \mathbb{R}$ be given; let's first find the derivative of sin at a, which is

$$\left.\frac{d}{dx}(\sin x)\right|_{x=a} = \lim_{h\to 0} \frac{\sin(a+h) - \sin a}{h}$$

(We use Leibniz notation here to avoid writing $\sin'(a)$, which looks awkward because it's easy to miss seeing the ' symbol in that location. Also, right now, we want to make the distinction between the variable x of differentiation and the position a where this derivative is taking place.)

In order to rewrite the numerator of our limit in a more usable form, we'll use the following identity which is valid for all $x, y \in \mathbb{R}$:

$$\sin y - \sin x = 2\sin\left(\frac{y-x}{2}\right)\cos\left(\frac{y+x}{2}\right)$$

(One way to prove this identity is outlined in Exercise 4.4.9.) When we apply this identity to the difference quotient, where $y = a + h$ and $x = a$, we get

$$\lim_{h\to 0} \frac{\sin(a+h) - \sin a}{h} = \lim_{h\to 0} \frac{2}{h}\sin\left(\frac{h}{2}\right)\cos\left(a + \frac{h}{2}\right)$$

Now, we notice that $h/2 \to 0$ as $h \to 0$, so we can let $t = h/2$ and rewrite the limit as

$$\lim_{t\to 0} \frac{1}{t}(\sin t)\cos(a+t)$$

(Technically, this uses the Composition Limit Theorem, where t is the inner function depending on h.) Now, we can use the limit

$$\lim_{t\to 0} \frac{\sin t}{t} = 1$$

which we found in Example 3.38, along with the fact that cos is continuous, to finish the computation:

$$\left.\frac{d}{dx}(\sin x)\right|_{x=a} = \lim_{t\to 0}\left(\frac{\sin t}{t}\right)\cos(a+t) = (1)\cos(a+0) = \cos a$$

To summarize, we have found that sin is differentiable everywhere, and $(\sin x)' = \cos x$. □

Example 4.21:

The steps used to find the derivative of cos are quite similar. We start with the identity

$$\cos y - \cos x = -2\sin\left(\frac{y-x}{2}\right)\sin\left(\frac{y+x}{2}\right)$$

which is valid for all $x, y \in \mathbb{R}$. (Exercise 4.4.9 also outlines a proof of this identity.) We use this to find

$$\frac{d}{dx}(\cos x)\bigg|_{x=a} = \lim_{h\to 0}\frac{\cos(a+h)-\cos a}{h} = \lim_{h\to 0}\frac{-2}{h}\sin\left(\frac{h}{2}\right)\sin\left(a+\frac{h}{2}\right)$$

We let $t = h/2$, so $t \to 0$ as $h \to 0$. By using the fact that $(\sin t)/t \to 1$ as $t \to 0$, as well as the fact that sin is continuous, we obtain

$$\frac{d}{dx}(\cos x)\bigg|_{x=a} = \lim_{t\to 0} -\left(\frac{\sin t}{t}\right)\sin(a+t) = -(1)\sin(a+0) = -\sin a$$

Therefore, cos is differentiable everywhere, and $(\cos x)' = -\sin x$. □

From the derivatives of sin and cos, we can obtain the derivatives of all the other common trigonometric functions by using our derivative laws. For example, we use the quotient rule to find the derivative of tan:

$$\begin{aligned}(\tan x)' &= \left(\frac{\sin x}{\cos x}\right)' = \frac{(\cos x)(\sin x)' - (\sin x)(\cos x)'}{\cos^2 x} \\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x\end{aligned}$$

because $\cos^2 x + \sin^2 x = 1$. You can compute the rest of the trigonometric derivatives in Exercise 4.4.10. For reference, we list the remaining derivatives here:

- $(\sec x)' = \sec x \tan x$
- $(\csc x)' = -\csc x \cot x$
- $(\cot x)' = -\csc^2 x$

4.4 Exercises

1. Use linear approximations and Lemma 4.11 to make another proof of the Product Rule. Your proof should use techniques similar to the argument in the remark following Theorem 4.12. (Hint: Your error $E(x)$ should be a sum of six different terms, each of which goes to zero faster than $x - a$.)

2. Compute the derivatives of the following functions of x, using any formulas and rules presented so far. Also, for each part, state where your derivative formula is valid.

   (a) $\left(\dfrac{1}{x^2+1}\right)'$

   (b) $\left(\dfrac{x^2+3x+2}{x^4+x^2+1}\right)'$

   (c) $((x+1)\sqrt{x} - \sin x)'$

   (d) $(x^5 \cos x)'$

   (e) $(\sin^2 x \cos^2 x)'$

   (f) $\left(\dfrac{x \sin x}{1+x^2}\right)'$

   (g) $\left(\dfrac{2-\sin x}{2-\cos x}\right)'$

3. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} x^2 - ax & \text{if } x < 1 \\ bx^3 & \text{if } x \geq 1 \end{cases}$$

   for all $x \in \mathbb{R}$, where $a, b \in \mathbb{R}$ are constants. Which values of a and b make f differentiable at 1, and what is $f'(1)$ in that case? (You may assume that the derivative laws from this section work with one-sided derivatives; the proofs are almost exactly the same.)

4. Let's suppose that instead of dropping an apple from 10 meters high, we throw an apple upwards from the ground with some initial velocity v_0, where $v_0 > 0$. If T is the time in seconds when the apple returns to the ground, then we get a function $f : [0, T] \to \mathbb{R}$ to describe the height of the apple from the ground, defined by

$$f(t) = v_0 t - 4.9t^2$$

   for all $t \in [0, T]$, where t is measured in seconds and $f(t)$ is measured in meters.

(a) Find T as a function of v_0.

(b) What should v_0 be if the apple lands after 5 seconds?

(c) Find the time $t \in (0, T)$ at which $f'(t) = 0$. At this time, the apple is neither falling nor rising, since the instantaneous velocity of the apple at this moment is 0. Using that information, explain why your answer for t makes sense physically.

(d) Just before the apple hits the ground, it is approaching the ground with a velocity of $\lim_{t \to T^-} f'(t)$. Find this limiting velocity as a function of v_0 only (i.e. T should not be in your final answer). Why does your answer make sense physically?

5. Prove the following generalization of the product rule by induction on $n \in \mathbb{N}^*$: if $n \in \mathbb{N}^*$, $a \in \mathbb{R}$, and $f_1, f_2, \ldots, f_n$ are real functions which are differentiable at a, then

$$\frac{(f_1 f_2 \cdots f_n)'(a)}{(f_1 f_2 \cdots f_n)(a)} = \frac{f_1'(a)}{f_1(a)} + \frac{f_2'(a)}{f_2(a)} + \cdots + \frac{f_n'(a)}{f_n(a)}$$

provided that none of the values $f_1(a), f_2(a), \ldots, f_n(a)$ are zero.

6. Recall that for $n, i \in \mathbb{N}$ with $i \le n$, the number "n choose i", written $\binom{n}{i}$, was defined in Exercise 4.2.5.(c) by

$$\binom{n}{i} = \frac{n(n-1)\cdots(n-i+1)}{i!}$$

(This has the name "n choose i" because, in discrete math classes, you can show that this number counts how many ways to choose i objects out of a collection of n distinguishable objects to form a subset.)

It is possible to prove (although you do not need to do so) that for any $n, i \in \mathbb{N}$ with $1 \le i \le n$, we have *Pascal's Identity*

$$\binom{n+1}{i} = \binom{n}{i-1} + \binom{n}{i}$$

You might also want to note that for any $n \in \mathbb{N}$, we have

$$\binom{n}{0} = \binom{n}{n} = 1$$

Using these facts, prove the following "n^{th} derivative" version of the product law by induction: if $a \in \mathbb{R}$ and f and g are any real functions which are n times differentiable at a (i.e. $f^{(n)}(a)$ and $g^{(n)}(a)$ exist), then

$$(fg)^{(n)}(a) = \sum_{i=0}^{n} \binom{n}{i} f^{(i)}(a) g^{(n-i)}(a)$$

7. Suppose that $a \in \mathbb{R}$ and that f is a real function differentiable at a. For each $n \in \mathbb{N}^*$, find a formula for $(f^n)'(a)$, i.e. the derivative of the n^{th} power at a, and prove by induction on n that your formula is correct.

8. Suppose $m, n \in \mathbb{Z}$ are given with $n > 0$. Let $f : (0, \infty) \to (0, \infty)$ be the (m/n)-th power function, i.e. we have $f(x) = x^{m/n}$ for all $x > 0$. Therefore, for any $x > 0$, we have

$$f^n(x) = x^m$$

Assume that f is differentiable on $(0, \infty)$, and use implicit differentiation (as introduced in Example 4.16), along with the result of Exercise 4.4.7, to find that

$$f'(x) = \frac{m}{n} x^{(m/n)-1}$$

This shows that for all $q \in \mathbb{Q}$, we have $(x^q)' = qx^{q-1}$. (However, we have not yet defined what it means to take a power with an IRRATIONAL exponent; that topic will wait until Chapter 7 when we study exponentiation.)

9. This exercise will outline how to come up with the key identities used to compute the derivatives of sin and cos. Our main starting point will be the angle-addition identities

$$\sin(x + y) = \sin x \cos y + \cos x \sin y$$

and

$$\cos(x + y) = \cos x \cos y - \sin x \sin y$$

which are valid for all $x, y \in \mathbb{R}$.

(a) By using the fact that $\sin(-x) = -\sin x$ and $\cos(-x) = \cos x$ for all $x \in \mathbb{R}$ (i.e. sin is an odd function, and cos is an even function), find formulas for $\sin(x - y)$ and $\cos(x - y)$, where $x, y \in \mathbb{R}$ are given.

(b) Using the identities from the beginning of this problem and the results of part (a), derive the following identities for all $a, b \in \mathbb{R}$:

$$\sin(a+b) - \sin(a-b) = 2\cos a \sin b$$

and

$$\cos(a+b) - \cos(a-b) = -2\sin a \sin b$$

(c) Use the identities from part (b) to conclude that for all $x, y \in \mathbb{R}$,

$$\sin x - \sin y = 2\cos\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right)$$

and

$$\cos x - \cos y = -2\sin\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right)$$

(Hint: Which values of a and b should you use with part (b)?)

10. Recall that

$$\cot x = \frac{\cos x}{\sin x} \qquad \sec x = \frac{1}{\cos x} \qquad \csc x = \frac{1}{\sin x}$$

Use these definitions to prove the following derivative formulas for all $x \in \mathbb{R}$ where the functions are defined:

(a) $(\cot x)' = -\csc^2 x$
(b) $(\sec x)' = \sec x \tan x$
(c) $(\csc x)' = -\csc x \cot x$

11. Use the previous exercise to compute

$$(\tan^2 x)' \qquad \text{and} \qquad (\sec^2 x)'$$

Give a simple reason explaining why your two answers are equal.

4.5 The Chain Rule

At this point, we have seen rules for computing derivatives with the major arithmetic operations. In this section, we address how to take the derivative of a composite function, i.e. a function of the form $f \circ g$. This leads us to a theorem called the Chain Rule, which will aid us tremendously in finding derivatives. The Chain Rule also is very practical for relating rates of change of different quantities, as we'll see later in this section.

Example 4.22:

Suppose that I'm eating candy. As I eat candy, my waistline expands, which causes my belly to get bigger. In essence, we have a chain of events, starting with the consumption of candy and ending with the growth of my belly.

I'd like to know: how fast is my belly growing as a function of the amount of candy I eat? In other words, I'd like to know what happens to the difference quotient

$$\frac{\text{change in belly size}}{\text{change in candy consumption}}$$

especially as my change in candy consumption approaches 0. We can write this as

$$\begin{aligned}&\frac{\text{change in belly size}}{\text{change in candy consumption}}\\ =\ &\frac{\text{change in belly size}}{\text{change in waistline}}\cdot\frac{\text{change in waistline}}{\text{change in candy consumption}}\end{aligned}$$

This is helpful because eating candy doesn't directly cause my belly to grow: it instead directly affects my waistline, and my waistline directly affects my belly weight. Thus, we write our quotient in a way that treats each step in the chain of events separately.

Let's say that x denotes the amount of candy I've eaten, measured in pounds. Also, let's write $w(x)$ for the size of my waistline, measured in inches, as a function of x. Lastly, let's write $S(w)$ to represent the size of my belly, measured in pounds, as a function of w (so here, w is being viewed as a variable). This means that as my candy consumption changes from a pounds to x pounds, the average speed at which my belly grows is

$$\begin{aligned}\frac{\text{change in belly size}}{\text{change in candy consumption}} &= \frac{S(w(x))-S(w(a))}{x-a}\\ &= \frac{S(w(x))-S(w(a))}{w(x)-w(a)}\cdot\frac{w(x)-w(a)}{x-a}\end{aligned}$$

In the first fraction, let's say we think of w as a variable of input to the size function, rather than being a function itself, and let's use b to denote $w(a)$ (so b is a constant). Then our expression becomes

$$\frac{S(w)-S(b)}{w-b}\cdot\frac{w(x)-w(a)}{x-a}$$

As $x \to a$, i.e. as the change in candy consumption goes to 0, the second fraction approaches $w'(a)$, which is the speed of waistline growth with respect to amount of candy eaten. Also, w approaches b as $x \to a$, so the first fraction approaches $S'(b)$, the speed of belly growth with respect to waistline size. Thus, with respect to candy consumption, my belly grows at the speed of $S'(b)w'(a)$, which is $S'(w(a))w'(a)$.

To get some more concrete numbers, let's say I do some measurements. (Note that these measurements are fabricated for the sake of the example and do not necessarily represent the author's true proportions!) I start with $a = 0$ pounds of candy. Next, I find that $w(x)$ is $(x/8) + 12$ when x is small, so $b = w(a) = 12$. Also, I assume that my belly looks like half a sphere, where my waistline is the radius. From this, I find that $S(w) = (2\pi w^3)/360$, since my belly size is proportional to the volume $(2\pi w^3)/3$ of my semispherical belly. With these numbers, we get $w'(a) = 1/8$ and $S'(b) = (6\pi b^2)/360$, so my belly growth speed is

$$\frac{6\pi \cdot 12^2}{360} \cdot \frac{1}{8} = \frac{3\pi}{10}$$

pounds of belly added per pound of candy eaten. $\square$

Now, let's try to apply the reasoning in Example 4.22 more generally. Let's say that f and g are real functions, and $a \in \mathbb{R}$ is given, such that $g \circ f$ is defined around a. We want to find

$$(g \circ f)'(a) = \lim_{x \to a} \frac{g(f(x)) - g(f(a))}{x - a}$$

We multiply the top and bottom by $f(x) - f(a)$ to get

$$\lim_{x \to a} \frac{g(f(x)) - g(f(a))}{f(x) - f(a)} \cdot \frac{f(x) - f(a)}{x - a}$$

The second fraction is a difference quotient of f from a to x. We want the first fraction to be a difference quotient as well, where the input is $f(x)$ instead of x. To make this clearer, let's say $y = f(x)$ and $b = f(a)$. Thus, instead of thinking as g as the outer layer of a composition, we think of g as a function of y alone. Since $y \to b$ as $x \to a$, our calculation becomes

$$(g \circ f)'(a) = \lim_{y \to b} \frac{g(y) - g(b)}{y - b} \cdot \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

From here, we see that if f and g are differentiable, our answer becomes $g'(b)f'(a)$, which is the same as $g'(f(a))f'(a)$. This suggests the following:

Theorem 4.23 (Chain Rule). *Let $a \in \mathbb{R}$ be given, and let f and g be real functions. Assume that $f'(a)$ and $g'(f(a))$ both exist. Then $g \circ f$ is differentiable at a, and $(g \circ f)'(a) = g'(f(a))f'(a)$.*

Remark. In the case where f and g are differentiable on their entire domains, Leibniz notation gives the Chain Rule the appearance of a simple algebraic identity:

$$\frac{dg}{dx} = \frac{dg}{df} \cdot \frac{df}{dx}$$

However, this way of writing the Chain Rule does not say where each derivative in the right-hand side is being evaluated.

Strategy. Let's look back at the work preceding the statement of Theorem 4.23. Most of the major steps can be made more precise with a little work, but we run into one issue: the fraction $(g(f(x)) - g(f(a)))/(f(x) - f(a))$ doesn't make sense when $f(x) = f(a)$. Let's call a choice of x which makes $f(x)$ equal to $f(a)$ a "forbidden input". This raises the question: how many choices of x are forbidden? After all, in order for the expression

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

to be defined, we need the body of the limit to be defined for x near a. In other words, we'd like there to exist some distance $\delta > 0$ so that no value of x within distance δ of a is forbidden, except for possibly $x = a$.

Unfortunately, certain choices for the function f can rule out enough options for x that no such δ exists. Consider when $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2 \sin(1/x)$ when $x \neq 0$, and $f(0) = 0$. An argument using the Squeeze Theorem shows that f is continuous at 0. In fact, it is possible to show that $f'(0)$ exists and equals 0; we'll study this example in a later section. However, there are lots of inputs x where $f(x) = 0$; for any integer $n \neq 0$, we have $f(1/(n\pi)) = 0$. In essence, f oscillates quite wildly near 0, so when $a = 0$, there are forbidden inputs arbitrarily close to a.

To get around this difficulty, we need a way of writing the difference quotient which makes sense even for the forbidden values of x. To do this, let's first consider the simpler difference quotient $(g(y) - g(b))/(y - b)$, where $b = g(a)$. This quotient is valid for all values of y except for b, and as $y \to b$, this quotient approaches $g'(b)$. Thus, intuitively, the difference quotient "should be" $g'(b)$ when $y = b$. To formalize this, we'll make a new function D, standing for "difference quotient", defined as follows:

$$D(y) = \begin{cases} \frac{g(y)-g(b)}{y-b} & \text{if } y \neq b \\ g'(b) & \text{if } y = b \end{cases}$$

For all values of y except for b, $D(y)$ is the difference quotient, but unlike the difference quotient, D is continuous at b! (In essence, the difference quotient had a "hole" at $y = b$, and D has "patched up the hole".) Therefore, we may safely write

$$g(y) - g(b) = D(y)(y - b)$$

for all values of y. When we plug in $y = f(x)$ and $b = f(a)$, we therefore obtain

$$g(f(x)) - g(f(a)) = D(f(x))(f(x) - f(a))$$

which is valid for ALL x, even the "forbidden" values! This means that, in the proof steps outlined before the statement of Theorem 4.23, instead of taking the fraction $(g(f(x)) - g(f(a)))/(x - a)$ and multiplying and dividing by $f(x) - f(a)$, we instead replace the numerator with the expression involving D. Thus, $(f(x) - f(a))/(x - a)$ appears in our work, like we wanted, without the risk of performing any forbidden divisions, and we may safely take limits as $x \to a$.

Proof. Let a, f, g be given as described. First, we define a function $D : \text{dom}(g) \to \mathbb{R}$ by saying

$$D(y) = \begin{cases} \frac{g(y)-g(f(a))}{y-f(a)} & \text{if } y \neq f(a) \\ g'(f(a)) & \text{if } y = f(a) \end{cases}$$

for all $y \in \text{dom}(g)$. Notice that when $y \in \text{dom}(g)$ and $y \neq f(a)$, $D(y)$ is just the difference quotient of g from $f(a)$ to y. Also, by definition of the derivative $g'(f(a))$, D is continuous at $f(a)$.

The equation

$$g(f(x)) - g(f(a)) = D(f(x))(f(x) - f(a))$$

is valid for all $x \in \text{dom}(f)$: when $f(x) = f(a)$, both sides are just 0, and when $f(x) \neq f(a)$, both sides are equal by the definition of D. Therefore,

$$\begin{aligned} (g \circ f)'(a) &= \lim_{x \to a} \frac{g(f(x)) - g(f(a))}{x - a} \\ &= \lim_{x \to a} D(f(x)) \cdot \frac{f(x) - f(a)}{x - a} \end{aligned}$$

Because $f'(a)$ exists, f is continuous at a, and we've also seen that D is continuous at $f(a)$. Therefore, by the Composition Limit Theorem, $D \circ f$ is continuous at a, so we get

$$\begin{aligned} (g \circ f)'(a) &= D(f(a)) \left(\lim_{x \to a} \frac{f(x) - f(a)}{x - a} \right) \\ &= g'(f(a))f'(a) \end{aligned}$$

as desired. ■

Example 4.24:

With the Chain Rule, many derivatives become easier to calculate. For example, if we consider the function mapping $x \in \mathbb{R}$ to $\sqrt{x^2+1}$, this is the composition of an inner function $f(x) = x^2 + 1$ and an outer function $g(y) = \sqrt{y}$. Since $g'(y) = 1/(2\sqrt{y})$ for all $y > 0$, and $f'(x) = 2x$, and we note that $f(x) \geq 1$ always, we find that

$$(\sqrt{x^2+1})' = \frac{1}{2\sqrt{(x^2+1)}}(x^2+1)' = \frac{2x}{2\sqrt{x^2+1}} = \frac{x}{\sqrt{x^2+1}}$$

As another simple example, if we return to the function from Example 4.22, we compute

$$\left((x^2+1)^{10}\right)' = 10(x^2+1)^9 \cdot (2x) = 20x(x^2+1)^9$$

It's much easier to compute this derivative with the Chain Rule instead of by expanding out the 10th power. □

Example 4.25:

For something more complicated, let's consider the function which takes $x \in \mathbb{R}$ to $\sin(\sin(\sin x))$. This function is really a composition of three sin functions. However, the Chain Rule only handles a composition of two functions. Therefore, to find this derivative, we first treat $\sin \circ \sin \circ \sin$ as $\sin \circ (\sin \circ \sin)$ and get

$$(\sin(\sin(\sin x)))' = \cos(\sin(\sin x))(\sin(\sin x))'$$

(so here, in the notation of the Chain Rule, g is sin and f is $\sin \circ \sin$). After this step, what remains is the derivative of a composite of two functions, and the Chain Rule gives us the final answer of

$$(\sin(\sin(\sin x)))' = \cos(\sin(\sin x)) \cos(\sin x) \cos x$$

In general, the rule of thumb to remember is "take derivatives from the outside going in". Each time you apply the Chain Rule, you "peel off" one layer from the outside of the composition, which is hopefully simple to differentiate, and then you handle the remaining layers with more applications of the Chain Rule if necessary or other such rules. For instance, here is a derivative which uses the Chain Rule three times to handle a composition of four functions:

$$(\sin(\cos(x^2))^3)' = 3(\sin(\cos(x^2))^2 \cdot \cos(\cos(x^2)) \cdot (-\sin(x^2)) \cdot 2x$$

(The outermost layer is the cubing function, followed by sin, followed by cos, and ending with the squaring function. We've only written the final answer here; the intermediate steps take a long time to write and are not generally required.) Here is a different example, which uses the Chain Rule once, then uses the product rule, and one of the parts of the product uses the Chain Rule again:

$$(\sqrt{x\sin(1-x)})' = \frac{1}{2\sqrt{x\sin(1-x)}} \cdot (\sin(1-x) + x\,(\cos(1-x)\cdot -1))$$

For some of these long calculations, it is useful to simplify the final result, but frequently there is rather little simplification possible. □

Implicit Differentiation

Implicit differentiation was introduced in Example 4.16. The main idea is that if f is a function of x, then sometimes it is easier to find an equation involving $f(x)$, rather than finding an explicit formula for $f(x)$. Implicit differentiation works by taking derivatives of both sides of the equation.

For instance, in Example 4.16, we saw that the square root function f satisfies the equation

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

for all $x \geq 0$. We took derivatives of both sides to get

$$2f(x)f'(x) = 1$$

and we used that to solve for $f'(x)$. We say that the equation $f^2(x) = x$ *implicitly* defines f. (We'll sometimes call that equation an "implicit equation" for f.) In that example, we used the product rule to find the derivative of

f^2. Another option is to use the Chain Rule, looking at f^2 as a composition of the squaring function with f. Frequently, in implicit equations, the Chain Rule becomes very handy for computing derivatives, as we'll demonstrate with a couple of examples.

Example 4.26:

This example will show how to find tangent lines to a circle. We start with the equation for a circle with radius r, centered at the origin, which is

$$x^2 + y^2 = r^2$$

This implicit equation can be solved for y to obtain the solutions

$$y = \pm\sqrt{r^2 - x^2}$$

which are valid when $x \in [-r, r]$. Thus, if we want to view y as a function of x, then there are two natural solutions for y.

In order to find tangent lines to the circle, we want the slope of the tangent line, which is given by the derivative $\frac{dy}{dx}$. We can compute $\frac{dy}{dx}$ for either of the two solutions for y written above, but we get an ugly and non-intuitive expression. Instead, let's use implicit differentiation to get a much simpler expression.

When we take the implicit equation for the circle and take the derivative with respect to x of each side, we get

$$\begin{aligned}
\frac{d}{dx}(x^2 + y^2) &= \frac{d}{dx}(r^2) \\
2x + (2y)\frac{dy}{dx} &= 0 \\
\frac{dy}{dx} &= \frac{-x}{y}
\end{aligned}$$

Note that, as a function of x, y^2 is the composition of y and squaring, and that r^2 is a constant with respect to x. Using Leibniz notation here helps remind us that the variable of differentiation is x and not y or r (so that we don't accidentally write the derivative of y^2 as $2y$ instead of $2y\frac{dy}{dx}$).

Our equation for $\frac{dy}{dx}$ is valid at all points where $y \neq 0$. What do we do when $y = 0$, i.e. when $x = r$ or $x = -r$? We notice that as x gets closer to

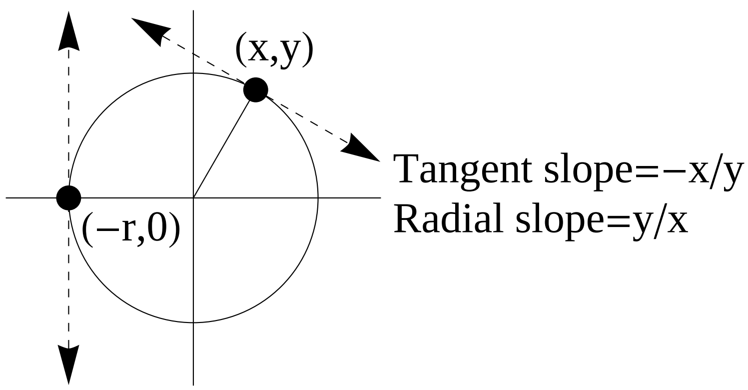

Figure 4.3: A circle and tangent lines to it

r or $-r$, y gets closer to 0, so the ratio $-x/y$ approaches either ∞ or $-\infty$. Hence, the points where $y = 0$ have vertical tangents. See Figure 4.3.

What are the benefits of using implicit differentiation with this example? The first is that the calculations are simple. Second, the equation

$$\frac{dy}{dx} = \frac{-x}{y}$$

works for EITHER solution for y in terms of x! Thus, we don't have to compute derivatives separately for each of the two solutions; we just find one answer in terms of x and y and then plug in the appropriate y.

Third, when using implicit differentiation, we find that our answer has a nice geometrical meaning. We found that the slope of the tangent line through (x, y) on the circle is $-x/y$. If you consider a radial segment that connects the center of the circle to the point (x, y), the slope of that segment is y/x, i.e. rise over run. If neither x nor y is 0, then y/x and $-x/y$ are negative reciprocals! That means that the tangent line and radial segment are perpendicular. Thus, our notion of tangent line in calculus agrees with the notion of tangent line to a circle in Euclidean geometry. □

Example 4.27:

For a longer example of implicit differentation, let's suppose that y is a function of x implicitly defined by the equation

$$y \sin y + 20x^3y^2 = \sqrt{x}$$

The derivative of the left side with respect to x is the following (broken down into smaller steps for clarity):

$$\begin{aligned}
\frac{d}{dx}(y\sin y + 20x^3y^2) &= \frac{dy}{dx}\sin y + y\frac{d}{dx}(\sin y) \\
&+ 20\left(\left(\frac{d}{dx}(x^3)\right)y^2 + x^3\frac{d}{dx}(y^2)\right) \\
&= \frac{dy}{dx}\sin y + y\cos y\frac{dy}{dx} + 20\left(3x^2y^2 + x^3\left(2y\frac{dy}{dx}\right)\right) \\
&= \frac{dy}{dx}\left(\sin y + y\cos y + 40x^3y\right) + 60x^2y^2
\end{aligned}$$

The right side of our implicit equation is much simpler, with derivative $1/(2\sqrt{x})$. Therefore, by setting the two derivatives equal to each other and solving for $\frac{dy}{dx}$, we get

$$\frac{dy}{dx} = \frac{\frac{1}{2\sqrt{x}} - 60x^2y^2}{\sin y + y\cos y + 40x^3y}$$

Going back to the implicit equation, we notice that choosing $x = 0$ and $y = 0$ satisfies the equation, but then our expression for the slope is undefined. Another point that satisfies the implicit equation is given by

$$y = \pi \qquad x = \left(\frac{1}{20\pi^2}\right)^{2/5}$$

At this point, the slope of the tangent line is

$$m = \frac{\left(\frac{1}{20\pi^2}\right)^{-1/5} - 60\left(\frac{1}{20\pi^2}\right)^{4/5}\pi^2}{0 + \pi(-1) + 40\left(\frac{1}{20\pi^2}\right)^{6/5}\pi}$$

(simplify this if you dare!), and thus we find that the tangent line through the point has the equation

$$y - \pi = m\left(x - \left(\frac{1}{20\pi^2}\right)^{2/5}\right)$$

$\square$

Remark. There are a couple subtle points worth mentioning in these examples. First of all, when writing an implicit equation, we're assuming that there really ARE solutions for y as functions of x. In our first example with the circle, we can find two solutions explicitly, but for the second example, we aren't necessarily sure whether any solutions exist or what their domains are. There are powerful theorems in three-dimensional calculus which address the question of whether implicit equations have solutions, but they are beyond the scope of this book.

The other subtle point we should note is that in order to take the derivative of both sides of the equation, we have to already know that both sides are differentiable. In our circle example, we had explicit solutions for y, so we already knew y was differentiable, but in many other examples, we have to assume that y is differentiable. This means that implicit differentiation will not work for EVERY solution of y in an equation, but it will work for the differentiable solutions. Since differentiability is a nice property for a function to have, in many cases this assumption is justified in practice.

Implicit equations can pop up in many different problems, allowing us to define complicated functions in terms of simpler-looking equations. For the rest of this section, we'll demonstrate two major applications of implicit differentiation: related rates and derivatives of inverse functions.

Related Rates

In many physical situations, we'd like to know the relationship between different rates of change. Some examples include: if a spherical balloon is being filled with air at some fixed rate, then how fast is its radius growing? If the top of a ladder slides down a wall at some fixed speed, then how fast is the bottom of the ladder sliding away from the wall? If I open my eyes at some fixed angular speed (i.e. my vision angle is increasing at some fixed speed), then how quickly is my total area of vision increasing?

All of these questions, and more, are called *related rates* problems. By using the Chain Rule and implicit differentiation, we can solve many different related rates problems. In this section, we'll show how to solve the examples we mentioned in the last paragraph, and some other problems will be provided as exercises.

Example 4.28:

Let's suppose that we're blowing air into a balloon, which always takes the shape of a sphere as we inflate it. At some point in time, the balloon has a 10 cm radius, and we are inflating it with air at the rate of 3 cubic centimeters per second. At that moment, how fast is the radius of the balloon growing?

To attack this problem, let's first decide on some notation to describe the problem. Let's say the balloon's volume is V (measured in cubic centimeters), the radius is r (measured in cm), and time is t (measured in seconds). With this notation, we are given that at some fixed time, we have $r = 10$ and $\frac{dV}{dt} = 3$. We would like to know $\frac{dr}{dt}$.

V and r are both functions of time, but V is also a function of r, according to the standard volume formula for a sphere:

$$V(t) = \frac{4}{3}\pi r^3(t)$$

(We write V as $V(t)$ and r as $r(t)$ for now to remind ourselves that V and r depend on t. In later problems, we won't bother to write (t) everywhere.)

To find $\frac{dr}{dt}$, we should take derivatives of both sides of this equation with respect to t, and we use the Chain Rule:

$$\frac{dV}{dt} = (4\pi r^2(t))\frac{dr}{dt}$$

(Technically, this assumes that r is a differentiable function of t, but that assumption is generally considered acceptable in related rates problems.)

From this, we plug in our given information to get

$$3 = 4\pi(10)^2\frac{dr}{dt}$$

which leads us to

$$\frac{dr}{dt} = \frac{3}{400\pi}$$

Thus, the radius is currently growing at the speed of $3/(400\pi)$ centimeters per second.

Note that for this problem, we didn't even need to know how r behaves as a function of time. All we needed was the simple relationship between the two quantities V and r, which both depend on t. $\square$

Example 4.28 illustrates a useful sequence of steps we can take when presented with a related rates problem:

1. Identify the relevant quantities for the problem, and give them names and units. Usually, most of your quantities are functions of time.

2. Find out which quantity you're trying to get as your answer.

3. Find some equation(s) describing how your quantities relate to one another. In many cases, drawing a picture will help.

4. Take derivatives of each side of your equations, and solve for your unknown values. The Chain Rule frequently comes in handy with the derivatives.

Let's try out these steps with the next example.

Example 4.29:

Let's suppose that we have a 10-foot ladder which is leaning against a wall as in Figure 4.4. The top of the ladder is sliding down the wall at the constant speed of 1 foot per second, which causes the bottom of the ladder to slide away from the wall as well. We'd like to answer the following question: when the top of the ladder is 6 feet off the ground, how fast is the bottom sliding away from the wall?

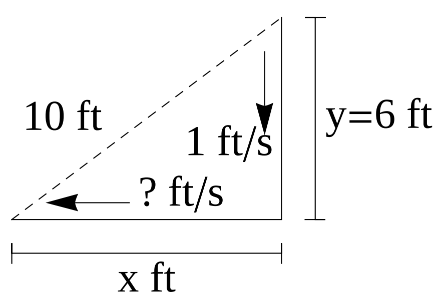

Figure 4.4: A ladder sliding down a wall

For our first step, let's label some relevant quantities. Let's say that t represents time in seconds. We'll use y to denote the height of the top of the ladder from the ground, and we'll use x to denote the distance of the bottom of the ladder from the wall. We are given that at this moment, $y = 6$ and $\frac{dy}{dt} = -1$. (The negative sign is needed, since y is decreasing as time passes.) We'd like to know $\frac{dx}{dt}$.

The key relationship we need is that the ladder forms a right triangle with the wall and the floor, so the Pythagorean Theorem tells us that

$$x^2 + y^2 = 10$$

We can solve for x in terms of y before differentiating, or we can differentiate implicitly. Let's try implicit differentation with respect to t, giving us

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0$$

When we plug in our known values and solve, we get

$$\frac{dx}{dt} = \frac{(6)(1)}{x}$$

Thus, to finish this problem, we need to know the value for x at this moment. To do this, we go back to the relationship between x and y. Plugging in 6 for y, we find $x^2 = 64$, or $x = \pm 8$. For this problem, only the positive solution makes sense physically, so $x = 8$. Plugging this in to our expression for $\frac{dx}{dt}$, we have the answer that the bottom is sliding away at a speed of 3/4 feet per second. □

Remark. Suppose we take the previous example and try to generalize it by considering the same situation, but instead of the top of the ladder starting 6 feet high, it starts h feet high. We'll still treat h as a constant with respect to time, but we'd like to see how our final answer for $\frac{dx}{dt}$ depends on h. You can do mostly the same steps as in Example 4.29, and you find that

$$\frac{dx}{dt} = \frac{h}{\sqrt{100 - h^2}}$$

When we plug in $h = 0$ (i.e. the ladder is already lying flat on the ground), this gives us an answer of 0, which makes sense since the ladder is done sliding. However, if we plug in values of h that are near 10, then our answer becomes very large, even larger than the speed of light! This is clearly an impractical answer.

The issue is that to model this problem, we assumed that the top of the ladder is falling at a constant speed of 1 foot per second. This is a reasonable assumption when the ladder has already slid down the wall a fair amount, like in Example 4.29. However, when h is close to 10, the ladder is nearly vertical, and then a falling speed of 1 foot per second is inaccurate. This goes to show that when modeling the real world with mathematics, it is important to know when your model is actually an accurate approximation of reality.

Example 4.30:

In this example, suppose I was sleepwalking, but I slowly wake up, opening my eyes. As I open my eyes, my angle of vision becomes greater, at the rate of 0.8 radians per second. To simplify matters, let's suppose I have such bad tunnel vision that I cannot see to the sides, so my field of vision is essentially a two-dimensional area.

I'd like to know: if I'm standing 10 feet away from a wall, facing that wall, and my eyes are currently open at the angle of $\pi/6$, then how fast is the area of my field of vision growing? Assume that the ceiling is high enough that my field of vision is not reaching the ceiling. This situation is depicted in Figure 4.5, where the triangle represents my area of vision. (In case you are concerned about your artistic ability when drawing figures, drawings for calculus problems frequently don't have to be very artistic if you don't want to be artistic! It is important, however, that your drawing convey the essential details of the problem.)

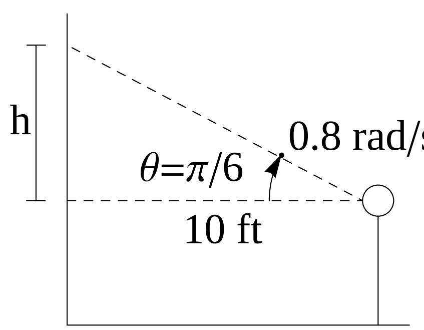

Figure 4.5: A sketch of my field of vision as I open my eyes

We'll use t to represent time in seconds again. We'd like a variable to stand for my current vision angle; let's use θ. The last relevant variable is the area of my current field of vision, which we'll call A. If we wish, we may also use a variable, like h, to represent the height of my field of vision along the wall. (However, note that my height is not relevant to computing the area of my field of vision, because we've assumed that my top "vision boundary" doesn't hit the ceiling.) Thus, we are given that $\theta = \frac{\pi}{6}$ and $\frac{d\theta}{dt} = 0.8$. We'd like to know $\frac{dA}{dt}$.

Now we need to find relationships between our quantities. First, we note that A is the area of a right triangle with height h and base length 10, so

$$A = \frac{(10)h}{2} = 5h$$

However, since we don't care about $\frac{dh}{dt}$, we'd like to replace h with a function of θ. To do this, we use familiar facts about the trigonometric functions. In this case, h is the side length opposite θ, and 10 is the adjacent side length, and $\tan\theta$ is the ratio of the two sides. Thus,

$$\tan\theta = \frac{h}{10}$$

By solving this equation for h and putting the result in the area equation, we find

$$A = 50\tan\theta$$

(You may not bother to introduce h as a variable, preferring to immediately write the relationship of A and θ. This is fine, provided that people can follow along with your work.)

We are now in a good position to take derivatives with respect to t. We obtain

$$\frac{dA}{dt} = (50\sec^2\theta)\frac{d\theta}{dt}$$

Plugging in our known values, we find $\sec(\pi/6) = 1/(\cos(\pi/6)) = 2/\sqrt{3}$ and thus

$$\frac{dA}{dt} = 50\left(\frac{2}{\sqrt{3}}\right)^2 (0.8) = \frac{160}{3}$$

The area is growing at the speed of 160/3 feet squared per second. □

It turns out that the speed we found at the end of Example 4.30 can also help us approximate how much the area of my field of vision changes after small amounts of time pass. The main idea is to think of $\frac{dA}{dt}$ as a fraction, where dA and dt are respectively small changes in area and time. Thus, we get

$$dA = \frac{160}{3} dt$$

For instance, if 1 second passes (i.e. $dt = 1$), then the area grows by roughly 160/3 feet squared (the change is not EXACTLY 160/3, because the speed is only 160/3 at a specific instant, as opposed to having that constant speed for the entire second). At the current time, the area is $50 \tan(\pi/6)$, which is $50/\sqrt{3}$, so in one second the area will be roughly $50/\sqrt{3} + 160/3$.

When dA and dt are treated in this way, as small changes, they are called *differentials*. In general, for any function f and a number a, the derivative $\frac{df}{dx}(a)$ can be thought of as a ratio of differentials, saying how much a small change dx in x causes an approximate change df in the value of $f(x)$. (Technically, df is really the change in the linear approximation of f at a, which is almost the same as the actual change in f's value if dx is small.)

It is also possible to think of differentials as errors in measurement. To illustrate this, let's look back at Example 4.28, involving the volume of a balloon. We know $V = (4/3)\pi r^3$, which yields

$$\frac{dV}{dr} = 4\pi r^2 \qquad \text{i.e.} \qquad dV = 4\pi r^2 dr$$

This can be interpreted as follows. Let's say the true radius of the balloon is r, but instead our measurement of the radius has an error of dr (so we measured something between $r - dr$ and $r + dr$). In other words, our measurement yielded a change of dr in radius from the true value of r. This causes the volume to change by approximately dV, so dV roughly represents the error we'll get when trying to measure the true volume of the balloon. Note that dV increases as r increases, signifying that even if you only make a small error dr in measuring the radius, the corresponding error dV in measuring the volume can be quite significant if the radius is large.

Inverse Functions: Flipping $\frac{dy}{dx}$ Upside-Down

For our last major application of the Chain Rule and implicit differentiation, let's look at the problem of finding the derivative of an inverse function. Suppose that f is a differentiable function with an inverse g: how do we find

g'? One of our first results about inverse functions, Lemma 1.49, gives us an idea: it tells us that for all $x \in \mathrm{dom}(f)$ and $y \in \mathrm{ran}(f)$,

$$f(x) = y \leftrightarrow x = g(y)$$

Thus, if we think of f as taking x to y, then the inverse g takes y back to x. This gives us an equation which implicitly defines g, namely:

$$f(g(y)) = y$$

Hence, y is now our variable of interest.

We'd like to find $g'(y)$. If we take the equation $f(g(y)) = y$, and we differentiate both sides with respect to y, the Chain Rule gives us

$$f'(g(y))g'(y) = 1$$

which we solve to get $g'(y) = 1/f'(g(y))$ when $f'(g(y))$ is not zero.

Unfortunately, while this method works well to calculate g', it doesn't actually prove that g' exists, since we had to assume g was differentiable to use the Chain Rule. However, we still have the following theorem:

Theorem 4.31. *Let $a \in \mathbb{R}$ be given, and let f be a real function defined in an open interval around a. Assume that f is invertible, with inverse g, and let $b = f(a)$. If $f'(a)$ exists and is not zero, then $g'(b)$ exists, and*

$$g'(b) = \frac{1}{f'(a)} = \frac{1}{f'(g(b))}$$

Remark. If we think of f as taking x to y and g as taking y back to x, then in Leibniz notation, this theorem says

$$\frac{dx}{dy} = \frac{1}{\left(\frac{dy}{dx}\right)}$$

which looks like a simple algebraic identity.

Strategy. We'd like to find the following limit:

$$\lim_{y \to b} \frac{g(y) - g(b)}{y - b}$$

Since f and g are inverse functions, we know that $g(b) = a$. Using the idea from Lemma 1.49, let's set $x = g(y)$, so $f(x) = y$ and our fraction becomes

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

In essence, we have just flipped the roles of "input" and "output" (which is what inverse functions do anyway). This new fraction is the reciprocal of the difference quotient for f from x to a.

However, let's keep in mind that our limit is letting y approach b. In order for the difference quotient for f from x to a to be useful for us, we'd like x to approach a. How will we show this?

In terms of the original variable y with which we started, saying that x approaches a is the same as saying that $g(y)$ approaches $g(b)$, i.e. that g is continuous at b. Since f is differentiable at a, f is also continuous at a, and the proof of Theorem 3.54 can be adapted to show that therefore g is continuous at b. (That theorem shows that if f is continuous on an INTERVAL, then g is continuous on an interval as well, but the same idea works if f is only continuous at a single point.)

Proof. Let a, b, f, g be given as described. Let $y \in \text{dom}(g)$ with $y \neq b$ be given, and define $x = g(y)$. Because f and g are inverse functions, we know that $y = f(x)$ and $a = g(b)$. Therefore,

$$\frac{g(y)-g(b)}{y-b} = \frac{x-a}{f(x)-f(a)}$$

Since $y \neq b$ and g is injective (because it is invertible), $g(y) \neq g(b)$, i.e. $x \neq a$, so we may write our difference quotient as

$$\frac{1}{\left(\frac{f(x)-f(a)}{x-a}\right)}$$

Because f is continuous at a (because f is differentiable at a), g is continuous at b by a minor modification of Theorem 3.54. Thus, by the Composition Limit Theorem, as $y \to b$, we have $g(y) \to g(b)$, i.e. $x \to a$, and thus

$$g'(b) = \lim_{y \to b} \frac{g(y)-g(b)}{y-b} = \lim_{x \to a} \frac{1}{\left(\frac{f(x)-f(a)}{x-a}\right)}$$

Because the denominator approaches $f'(a)$, which is not zero, the quotient rule for limits tells us that $g'(b) = 1/f'(a)$, as desired. ■

Remark. To see another reason why this theorem makes sense, recall that the graphs of inverse functions look like mirror images reflected over the line $y = x$. Thus, if we look at a tangent line to f, and we take its mirror image, the tangent slope gets reflected as well. See Figure 4.6, which depicts $f(x) = x^3 - 1$ and its inverse $g(x) = (x+1)^{1/3}$, with tangent lines touching each. (We write g as a function of x here, to make it less confusing for us to graph it.)

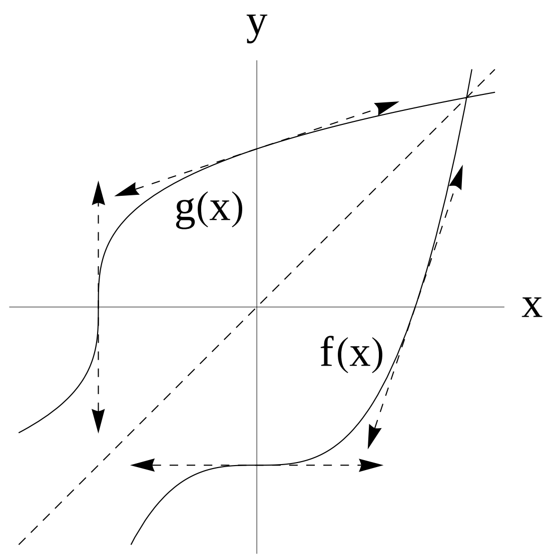

Figure 4.6: A function and its inverse, with tangent lines

This picture also suggests that whenever $f'(a)$ is zero, yielding a horizontal tangent, then $g'(b)$ has a vertical tangent. To see this, we can perform most of the same calculations as in the proof of Theorem 4.31, but this time we find that

$$\lim_{y \to b} \frac{g(y) - g(b)}{y - b} = \lim_{x \to a} \frac{1}{\left(\frac{f(x) - f(a)}{x - a}\right)}$$

is a limit of the form $1/0$. This goes to either ∞ or $-\infty$, based on whether the difference quotient for f approaches 0 from positive values or from negative values. Either way, this means that the tangent line is infinitely steep, i.e. vertical.

Example 4.32:

Using Theorem 4.31, we can finally determine where the n^{th} root functions are differentiable, for any $n \geq 2$. We note that the n^{th} root function is the inverse of the n^{th} power function: if $f(x) = x^n$ and $g(y) = y^{1/n}$ (valid for $x, y \in [0, \infty)$ when n is even, and valid for $x, y \in \mathbb{R}$ when n is odd), then f and g are inverses. We know that $f'(x) = nx^{n-1}$, which is zero only when $x = 0$, which means that g has a vertical tangent when $y = 0^{1/n} = 0$. Otherwise, for all other $y \in \text{dom}(g)$, we have

$$g'(y) = \frac{1}{f'(g(y))} = \frac{1}{n(y^{1/n})^{n-1}} = \frac{1}{n} y^{(1/n)-1}$$

This shows that n^{th} roots are differentiable. After that, it follows that any rational power function is differentiable. This is because if m/n is a rational number (where $m \in \mathbb{Z}$ and $n \in \mathbb{N}^*$), then $x^{m/n} = (x^{1/n})^m$, i.e. the m/nth power is the composition of n^{th} root and m^{th} power. This finally fills in the gap in the proof from Exercise 4.4.8. $\square$

Example 4.33:

We've also seen another useful collection of inverse functions by this point: the inverse trigonometric functions. We'll show how to differentiate the inverse sine function here, and we'll leave the rest of the inverse trigonometric derivatives as exercises.

Let f be the sine function restricted to the domain $[-\pi/2, \pi/2]$, i.e. $f = \sin \upharpoonright [-\pi/2, \pi/2]$. This is the domain we used when we defined arcsin in Chapter 3: arcsin has domain $[-1, 1]$ and range $[-\pi/2, \pi/2]$. We see that f is differentiable on $(-\pi/2, \pi/2)$ (f does not have a TWO-sided derivative at the endpoints), and $f' = (\sin)' = \cos$. Thus, for any $y \in (f(-\pi/2), f(\pi/2))$, i.e. $y \in (-1, 1)$, we have

$$(\arcsin y)' = \frac{1}{\cos(\arcsin y)}$$

This is well-defined because when $y \in (-1, 1)$, $\cos(\arcsin y) > 0$.

It turns out that we can find a much better formula for the derivative by using some trigonometric identities. We'd like to write $\cos(\arcsin y)$ in terms of $\sin(\arcsin y)$, so that the sin and arcsin cancel. To do this, we know that

$$\cos(\arcsin y) = \pm\sqrt{1 - \sin^2(\arcsin y)}$$

Since $\cos(\arcsin y) > 0$ for $y \in (-1, 1)$, we take the positive square root. Thus, we obtain the simpler formula for the derivative of arcsin:

$$(\arcsin y)' = \frac{1}{\sqrt{1 - (\sin(\arcsin y))^2}} = \frac{1}{\sqrt{1 - y^2}}$$

$\square$

Using similar methods, you can prove the following formulas in Exercise 4.6.21:

$$(\arccos y)' = \frac{-1}{\sqrt{1 - y^2}} \qquad (\arctan y)' = \frac{1}{1 + y^2}$$

The first formula is valid when $|y| < 1$, and the second is valid for all $y \in \mathbb{R}$.

4.6 Exercises

For the first two exercises, we'll develop another proof of the Chain Rule using linear approximations. Recall that in Lemma 4.11, the error function E for approximating f at a satisfies the property that $E(x)/(x-a) \to 0$ as $x \to a$. Because the fraction $E(x)/(x-a)$ is not defined when $x = a$, we can run into difficulties like we had with the fraction $(f(g(x)) - f(g(a)))/(g(x) - g(a))$ in our proof of the Chain Rule. We can fix this issue using a similar tactic that the function D was used for in the proof of the Chain Rule.

1. Prove the following alternate form of Lemma 4.11:

   Lemma 4.34. *Let $a \in \mathbb{R}$ be given, and let f be a real function defined in an interval around a. Then f is differentiable at a if and only if there exists an $m \in \mathbb{R}$ and a function $\hat{E} : \mathrm{dom}(f) \to \mathbb{R}$ for which $\hat{E}(a) = 0$, $\hat{E}$ is continuous at a, and*

   $$f(x) = f(a) + m(x - a) + \hat{E}(x)(x - a)$$

   is true for all $x \in \mathrm{dom}(f)$. In this case, $m = f'(a)$.

   (Hint: Use Lemma 4.11. For the left-to-right direction, if E is the error function from Lemma 4.11, then it's not hard to find a formula for $\hat{E}(x)$ when $x \neq a$. How do you "patch up" that formula to make sense for all $x \in \mathrm{dom}(f)$?)

2. Using Lemma 4.34, give another proof of the Chain Rule using linear approximations.

For the next seven exercises, compute the following derivatives. You may assume that you are only considering values of $x \in \mathbb{R}$ which satisfy the requirements of our derivative rules.

3. $(\sin(2x))'$
4. $(\sqrt{1-x^2})'$
5. $(x\sqrt{1+x^2})'$
6. $\left(\dfrac{\sin^2 x}{\sin(x^2)}\right)'$
7. $(\cos(1-\sin(3x)))'$
8. $(\sin(\cos^2 x) \cdot \cos(\sin^2 x))'$
9. $\left(\left(\dfrac{1-x^2}{1+x^2}\right)^3\right)'$

For the next three exercises, assume that y is a differentiable function of x, and use the provided implicit equation to find $\dfrac{dy}{dx}$ in terms of x and y.

10. $x^3 + y^3 = 1$
11. $\sin x \cos y = \cos x \sin y$
12. $(xy+1)^2 = \sqrt{y}$
13. In Exercise 4.6.10, assume y'' exists, and find it in terms of x and y.
14. Let $k, M \in \mathbb{R}$ be given with $k, M > 0$, and suppose that $P : \mathbb{R} \to \mathbb{R}$ is a differentiable function which satisfies the equation

$$\frac{dP}{dt} = kP(t)(M - P(t))$$

for all $t \in \mathbb{R}$. (This kind of equation involving a derivative of an unknown function is called a *differential equation*. This particular equation is called the *logistic equation*, though it has other names as well, and it arises in situations where $P(t)$ is supposed to represent the population of a species at time t.)

From this equation, find the second derivative $\dfrac{d^2P}{dt^2}$ in terms of $P(t)$, k, and M. For which values of $P(t)$ is the second derivative equal to 0?

15. Each edge of a cube is expanding at the rate of 1 cm per second. At a specific moment, all the side lengths are s.

    (a) How fast is the volume of the cube changing at this time? Express your answer as a function of s.

    (b) How fast is the surface area changing? Express your answer as a function of s.

    (c) Let's say we're measuring the side length of this cube, and instead of getting the true side length of s, we have a measurement error of $ds = 0.1$ cm. (See the remarks about differentials following Example 4.30.) What are the approximate errors we'll obtain measuring the volume and the surface area?

16. A particle is traveling along the curve $y = x^2$, where its position (x, y) starts at the origin $(0, 0)$. The particle moves rightward with the speed of $\dfrac{dx}{dt} = 1$ unit per second. When the particle is at $(2, 4)$ (i.e. when $x = 2$ and $y = 4$), how fast is the particle's distance to the origin changing?

17. A large weight is hanging from a rope that is supported at two corners, as shown in Figure 4.7 (the rope is drawn with dashed lines). The weight dangles 10 feet below the ceiling and is slowly being slid along the rope to the right, at the rate of 1 foot per minute. As the weight slides, it always stays 10 feet below the ceiling. Eventually, the two angles α and β in the drawing are both equal to $\pi/4$. At that moment, how fast are α and β changing?

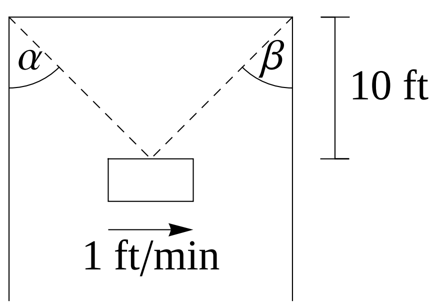

Figure 4.7: A weight hanging from a rope

18. A boat is currently 5 miles north of the shoreline, and the boat sails due east at 40 miles per hour, parallel to the shoreline. There is a lighthouse on this shoreline. When the boat is 12 miles west and 5 miles north of the lighthouse, how fast is the distance between the boat and the lighthouse changing?

19. A paper cup is shaped like a circular cone with radius 3 inches and height 6 inches. The cup is being filled with water at the rate of 0.5 cubic inches per second, as drawn in Figure 4.8 (the dashes represent the top of the water in the cup). At the point when the water is filled 3 inches deep, how fast is the depth changing?

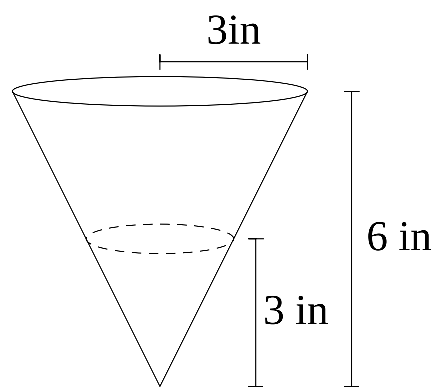

Figure 4.8: A cup being filled with water

In case you need it, the volume V of a circular cone with radius r and height h is $V = \pi r^2 h/3$. (Hint: What is the ratio of the radius of the water to the depth of the water?)

20. Consider Example 4.29, where the 10-foot ladder is sliding down a wall at a constant rate of 1 foot per second. In that example, we calculated the speed $\dfrac{dx}{dt}$ at which the ladder's bottom slides away from the wall. At that moment, what is the acceleration of the ladder's bottom away from the wall? (In other words, what's the second derivative?)

Hint: When solving for $\dfrac{dx}{dt}$ in the example, do not plug in known values yet! After differentiating with respect to t one more time, you may plug in your known values.

21. Prove that the following derivative formulas are correct using Theorem 4.31 and trigonometric identities:
   (a) $(\arccos y)' = \dfrac{-1}{\sqrt{1-y^2}}$ for all $y \in (-1,1)$
   (b) $(\arctan y)' = \dfrac{1}{1+y^2}$ for all $y \in \mathbb{R}$

4.7 Some Unusual Functions II

At this point, now that we are studying differentiability, we'd like to return to the examples from the Some Unusual Functions section in the previous chapter and analyze the differentiability of those functions. We will also introduce some new examples with surprising behavior.

Examples with $\chi_{\mathbb{Q}}$

Recall that in the previous chapter, we defined the characteristic function of the rationals, $\chi_{\mathbb{Q}}$, as follows: for all $x \in \mathbb{R}$, $\chi_{\mathbb{Q}}(x) = 1$ if $x \in \mathbb{Q}$, and otherwise $\chi_{\mathbb{Q}}(x) = 0$. We saw that this function is discontinuous everywhere. However, if we multiply $\chi_{\mathbb{Q}}$ by a continuous function f, then the result can have continuity points. For example, when $f : \mathbb{R} \to \mathbb{R}$ is the identity function mapping each $x \in \mathbb{R}$ to itself, then we've seen that $(f\chi_{\mathbb{Q}})$, which takes $x \in \mathbb{R}$ to $x\chi_{\mathbb{Q}}(x)$, is continuous at 0.

Is this function differentiable at 0? Because $\chi_{\mathbb{Q}}$ is not differentiable anywhere (because it is not continuous anywhere), we can't use the product rule to help us. Instead, we use the definition of derivative and find

$$(f\chi_{\mathbb{Q}})'(0) = \lim_{x\to 0} \frac{x\chi_{\mathbb{Q}}(x) - 0\chi_{\mathbb{Q}}(0)}{x-0} = \lim_{x\to 0} \chi_{\mathbb{Q}}(x)$$

Since this limit doesn't exist, $f\chi_{\mathbb{Q}}$ is not differentiable at 0. However, if we instead consider the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$ for all $x \in \mathbb{R}$, then we find

$$(g\chi_{\mathbb{Q}})'(0) = \lim_{x\to 0} \frac{x^2\chi_{\mathbb{Q}}(x) - 0^2\chi_{\mathbb{Q}}(0)}{x-0} = \lim_{x\to 0} x\chi_{\mathbb{Q}}(x) = 0$$

Thus, this function is differentiable at 0, with the value of 0. This gives us an example of a function with exactly one differentiability point.

To better understand the difference between these two functions ($f\chi_{\mathbb{Q}}$ and $g\chi_{\mathbb{Q}}$), we can draw a rough picture. In Figure 4.9, we draw dashes representing the curves $y = x$, $y = x^2$, and $y = 0$. The basic idea is that the values of $f\chi_{\mathbb{Q}}$ keep jumping between the curves $y = x$ and $y = 0$, and the values of $g\chi_{\mathbb{Q}}$ keep jumping between the curves $y = x^2$ and $y = 0$.

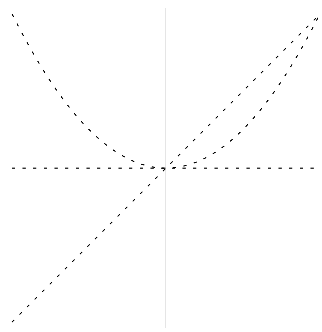

Figure 4.9: A rough picture of $x\chi_{\mathbb{Q}}(x)$ and $x^2\chi_{\mathbb{Q}}(x)$

From the picture, we see why our two functions are continuous at 0: as $x \to 0$, all three curves tend to the value 0. (The actual proof of continuity here uses the Squeeze Theorem.) For the function $f\chi_{\mathbb{Q}}$, the graph keeps jumping between a curve of slope 1 and a curve of slope 0, so there is no "limit slope" (i.e. no derivative). However, for $g\chi_{\mathbb{Q}}$, the graph keeps jumping between slope $2x$ and slope 0 (since $(x^2)' = 2x$), and as $x \to 0$, both $2x$ and 0 approach 0. This shows why $f\chi_{\mathbb{Q}}$ is not differentiable at 0 but $g\chi_{\mathbb{Q}}$ is.

Our picture suggests that any function of the form $f\chi_{\mathbb{Q}}$, where f is any function, keeps jumping between slope $f'(x)$ (where $f'(x)$ exists) and slope 0. (You can explore this further in Exercise 4.8.2.) Although $f\chi_{\mathbb{Q}}$ isn't necessarily a constant function, it "looks like" the constant function 0 for each irrational input (and since the irrationals are dense, there are infinitely many irrationals in any interval). This suggests that if $f\chi_{\mathbb{Q}}$ is differentiable anywhere, then its derivative should be the derivative of a constant, which is 0. Thus, we have

Lemma 4.35. *Let $a \in \mathbb{R}$ be given and f be a real function defined around a. If $f\chi_{\mathbb{Q}}$ is differentiable at a, and $(f\chi_{\mathbb{Q}})(a) = 0$, then $(f\chi_{\mathbb{Q}})'(a) = 0$.*

Strategy. One way to show $(f\chi_{\mathbb{Q}})'$ is 0 is to show that it can't be anything else. Let's say $L = (f\chi_{\mathbb{Q}})'(a)$, and let's assume that $L \neq 0$ and find a contradiction.

Since the difference quotient of $f\chi_{\mathbb{Q}}$ from a to x approaches L, these difference quotients must also be nonzero when x is close enough to a. More specifically, eventually the difference quotients have to be within distance $\epsilon = |L|/2$ of L, i.e. their values have to lie in the interval $(L - |L|/2, L + |L|/2)$.

However, the difference quotient of $f\chi_{\mathbb{Q}}$ from a to x is

$$\frac{f(x)\chi_{\mathbb{Q}}(x) - f(a)\chi_{\mathbb{Q}}(a)}{x-a}$$

Since we assumed that $f(a)\chi_{\mathbb{Q}}(a) = 0$, this shows that for any $x \in \mathbb{R} - \mathbb{Q}$, the difference quotient is 0. Since $0 \notin (L - |L|/2, L + |L|/2)$, and there are irrational numbers x arbitrarily close to a, we have our contradiction.

Note: As shown by the example when a is 0 and $f(x)$ is x, the assumption of $(f\chi_{\mathbb{Q}})(a) = 0$ is NOT enough to prove that $(f\chi_{\mathbb{Q}})'(a)$ exists! This is why our proof needs to use the assumption that $(f\chi_{\mathbb{Q}})'(a)$ exists. (The tactic we used, which shows that a limit is 0 by assuming it is nonzero and using half its absolute value as ϵ, is quite common in analysis.)

Proof. Let a, f be given as described. Let $L = (f\chi_{\mathbb{Q}})'(a)$. Assume that $L \neq 0$, and we will obtain a contradiction.

Since, by definition, we have

$$L = \lim_{x\to a} \frac{f(x)\chi_{\mathbb{Q}}(x) - f(a)\chi_{\mathbb{Q}}(a)}{x-a} = \lim_{x\to a} \frac{f(x)\chi_{\mathbb{Q}}(x)}{x-a}$$

(because we assumed $(f\chi_{\mathbb{Q}})(a) = 0$), this means that when $\epsilon = |L|/2$, there is some $\delta > 0$ such that

$$\forall x \in \mathrm{dom}(f)\ \left(0 < |x-a| < \delta \to \left|\frac{f(x)\chi_{\mathbb{Q}}(x)}{x-a} - L\right| < |L|/2\right)$$

(In fact, because f is defined around a, we may assume that δ is small enough so that the entire interval $(a - \delta, a + \delta)$ is contained in $\mathrm{dom}(f)$.)

Because the irrational numbers are dense, we may choose some $x \in \mathbb{R} - \mathbb{Q}$ which belongs to $(a, a+\delta)$. Thus, $\chi_{\mathbb{Q}}(x) = 0$ and $0 < |x - a| < \delta$, so we have

$$\left|\frac{f(x)\cdot 0}{x-a} - L\right| < |L|/2$$

However, this tells us that $|L| < |L|/2$, which is a contradiction. ■

Because Lemma 4.35 requires that you assume your function $f\chi_{\mathbb{Q}}$ is differentiable at a, it cannot be used to prove that $(f\chi_{\mathbb{Q}})'(a)$ exists. However, it is useful sometimes for showing that $(f\chi_{\mathbb{Q}})'(a)$ does NOT exist, because all you have to do is show that the derivative cannot be 0. We will use this to our advantage soon.

Oscillating Examples

The examples involving $\chi_{\mathbb{Q}}$ are interesting because the function values keep "jumping" between two different curves. Thus, for those functions, many limits fail to exist. Another way to produce limits that fail to exist is by having a function oscillate quite rapidly, such as the function taking $x \in \mathbb{R} - \{0\}$ to $\sin(1/x)$. As x approaches 0, the values of this function keep wavering between -1 and 1, and Exercise 3.6.8 proves more formally that the function has no limit at 0. However, the function is differentiable at every other value of x, and the Chain Rule yields

$$\left(\sin\left(\frac{1}{x}\right)\right)' = \cos\left(\frac{1}{x}\right) \cdot \frac{-1}{x^2}$$

However, just like we did with $\chi_{\mathbb{Q}}$, maybe we can multiply $\sin(1/x)$ by a function that helps "smooth it out". As one such example, consider $x\sin(1/x)$. Since $0 \le |x\sin(1/x)| \le |x|$, the Squeeze Theorem shows that $|x\sin(1/x)|$ approaches 0 as $x \to 0$. That means that the following function $g : \mathbb{R} \to \mathbb{R}$ is continuous everywhere:

$$\forall x \in \mathbb{R}\ g(x) = \begin{cases} x\sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

In essence, by multiplying by x, we squeeze the graph between the curves $y = x$ and $y = -x$, which causes it to have a limit at 0. See the left graph of Figure 4.10 (the dashed lines represent the curves $y = x$ and $y = -x$).

Now that we know g is continuous at 0, we'd like to know if g is differentiable at 0. Because $\sin(1/x)$ is not differentiable at 0, we cannot use the product rule, so we must use the definition of derivative again. We compute

$$g'(0) = \lim_{x\to 0} \frac{x\sin\left(\frac{1}{x}\right) - 0}{x - 0} = \lim_{x \to 0} \sin\left(\frac{1}{x}\right)$$

and this limit does not exist. Therefore, g is not differentiable at 0.

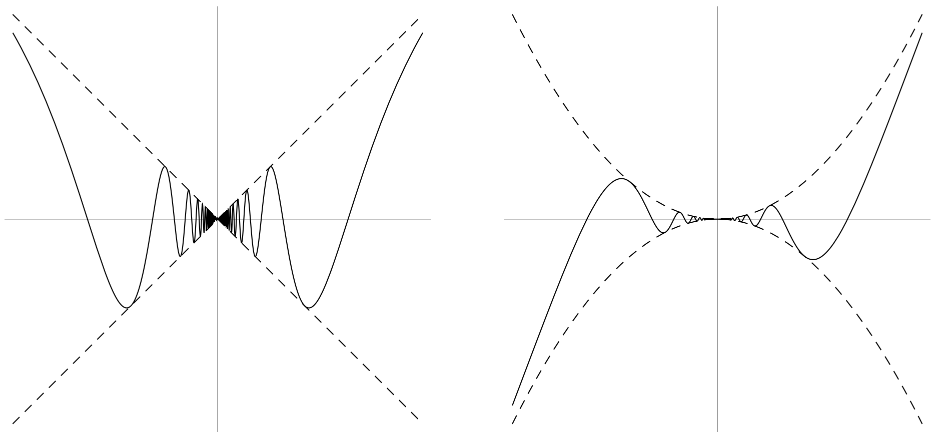

Figure 4.10: The graphs of $g(x) = x\sin(1/x)$ (left) and $h(x) = x^2\sin(1/x)$ (right)

What if we instead try multiplying $\sin(1/x)$ by a different function, like x^2? As with the function g above, we can define $h : \mathbb{R} \to \mathbb{R}$ by $h(x) = x^2\sin(1/x)$ when $x \neq 0$ and $h(0) = 0$, and we can show that h is continuous everywhere. In this situation, however, we have

$$h'(0) = \lim_{x\to 0} \frac{x^2\sin\left(\frac{1}{x}\right) - 0}{x - 0} = \lim_{x\to 0} x\sin\left(\frac{1}{x}\right) = 0$$

The function h is displayed in the right graph of Figure 4.10.

The function h has a rather interesting property. If we compute $h'(x)$ for nonzero values of x, we get

$$\begin{aligned}\left(x^2\sin\left(\frac{1}{x}\right)\right)' &= 2x\sin\left(\frac{1}{x}\right) + x^2\left(\cos\left(\frac{1}{x}\right)\cdot\frac{-1}{x^2}\right)\\ &= 2x\sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)\end{aligned}$$

Therefore,

$$h'(x) = \begin{cases} 2x\sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) & \text{if } x \neq 0\\ 0 & \text{if } x = 0\end{cases}$$

Now, as $x \to 0$, $x\sin(1/x)$ approaches 0, but $\cos(1/x)$ does not have a limit (the proof is nearly the same as with $\sin(1/x)$). We claim this implies that

$h'(x)$ cannot have a limit as $x \to 0$. If such a limit existed, we could write

$$\cos\left(\frac{1}{x}\right) = 2x\sin\left(\frac{1}{x}\right) - h'(x)$$

for all $x \neq 0$ and take the limit of the right-hand side using the difference rule for limits. This would mean that $\cos(1/x)$ has a limit as $x \to 0$, which is not true.

Therefore, h is a function which is differentiable everywhere, but whose DERIVATIVE is discontinuous at 0. The lesson here is rather subtle: if you want to take a derivative of some function f at a, but your derivative rules don't work at a, then you are NOT guaranteed to get the correct answer by finding $f'(x)$ for other values of x and then taking a limit as $x \to a$. (However, under certain conditions, this approach does work: see Exercise 4.10.19.) For these situations, you should use the definition of derivative. Later, in Chapter 8, we'll see that in some cases, we can find derivatives much more easily by approximating the function we'd like to differentiate.

The Ruler Function, Revisited

Let's return to another example we defined in the Some Unusual Functions section from Chapter 3: the ruler function $r : \mathbb{R} \to \mathbb{R}$, defined by

$$\forall x \in \mathbb{R}\ r(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ 1/q & \text{if } x = p/q \text{ in lowest terms, } q > 0 \end{cases}$$

In Theorem 3.44, we saw that for all $a \in \mathbb{R}$, $r(x) \to 0$ as $x \to a$. Therefore, r is continuous at precisely the irrational numbers.

We'd like to answer the question: is r differentiable at any irrational inputs? It turns out the answer is "no": we present a rather recent proof. [5]

Theorem 4.36. *The ruler function is not differentiable anywhere.*

Strategy. Suppose that $r'(a)$ exists for a contradiction, where a is some real number. Since the ruler function is only continuous at irrational numbers, a must be irrational. Therefore, $r(a) = 0$.

[5] Credit goes to the article "Nondifferentiability of the Ruler Function" by William Dunham, available through http://www.jstor.org, for the major ideas of the proof.

Also, because $r(x)$ is zero for every $x \notin \mathbb{Q}$, it follows that $r = r\chi_{\mathbb{Q}}$. This means that Lemma 4.35 is applicable to r. From the lemma, we find that since we assume $r'(a)$ exists, it must be zero. Thus, no matter what distance ϵ we specify, there exists some δ such that all x within distance δ of a satisfy

$$\left|\frac{r(x) - r(a)}{x - a}\right| < \epsilon$$

In other words, the difference quotients near a have small magnitude, signifying r has arbitrarily small slopes near a.

To obtain a contradiction, let's try to find a difference quotient with large magnitude (i.e. larger than ϵ) by finding a suitable value of x within distance δ of a. Note that if x is irrational, then $r(x) = 0$, which makes the difference quotient 0; hence we are interested in looking for a rational value of x.

This raises the question: which rational numbers can be found within distance δ of a? As the definition of the ruler function indicates, the denominator of our rational number is particularly important. Thus, for each $n \in \mathbb{N}^*$, let's consider the rational numbers whose denominator is n: does such a number lie in the interval $(a - \delta, a + \delta)$?

The rational numbers with denominator $1/n$ are spaced at distance $1/n$ from each other. Thus, since a is irrational, a is strictly between two consecutive multiples of $1/n$, say $m/n < a < (m + 1)/n$. (A formula for m can be found in terms of the floor function.) In particular, we have $a - m/n < 1/n$, so $a - 1/n < m/n < a$. To guarantee that m/n is in the interval $(a - \delta, a + \delta)$, we choose n large enough to satisfy $1/n < \delta$. Let's take $x = m/n$.

Now, m/n might not be a representation in lowest terms. If we write m/n in lowest terms as p/q, then $r(m/n) = 1/q$. Since q is a divisor of n, $q \le n$, so $1/q \ge 1/n$. Since m/n is within distance $1/n$ of a, $|m/n - a| \le 1/n$, and thus

$$\left|\frac{f\left(\frac{m}{n}\right) - f(a)}{\frac{m}{n} - a}\right| = \frac{\left|\frac{1}{q} - 0\right|}{\left|\frac{m}{n} - a\right|} \ge \frac{1/n}{1/n} = 1$$

This shows that when x is m/n, the difference quotient has magnitude at least 1. This causes a contradiction when ϵ is less than or equal to 1.

Proof. Let $a \in \mathbb{R}$ be given, and assume for contradiction that the ruler function r is differentiable at a. By Theorem 3.44, because r is continuous at a, a must be irrational. Therefore, $r(a) = 0$.

Because r is zero at all irrationals, we see that $r = r\chi_{\mathbb{Q}}$. Hence, since $r(a) = 0$, Lemma 4.35 tells us that $r'(a) = 0$. By the definition of derivative, this means that when $\epsilon = 1$, there is some $\delta > 0$ such that

$$\forall x \in \mathbb{R} \ \left(0 < |x-a| < \delta \to \left|\frac{r(x)-r(a)}{x-a}\right| < 1\right)$$

To obtain a contradiction, we will find some $x \in (a-\delta, a+\delta)$ for which $|r(x)-r(a)|/|x-a| \geq 1$. Since $r(a) = 0$, this is equivalent to $|r(x)| \geq |x-a|$.

First, choose $n \in \mathbb{N}^*$ large enough to satisfy $1/n < \delta$. Take $m = \lfloor na \rfloor$. By the definition of the floor function, $m \leq na < m+1$, so $m/n \leq a < (m+1)/n$, which implies that $a - m/n < 1/n$. (Also, since a is irrational, $m/n \neq a$.) Thus, $m/n > a - 1/n > a - \delta$, showing that $|m/n - a| < \delta$.

Let's take $x = m/n$. If x has the lowest-terms form p/q, then we must have $q \leq n$ (since q is a divisor of n), so

$$r(x) = \frac{1}{q} \geq \frac{1}{n} \geq a - \frac{m}{n} = |x-a|$$

Thus, we have found a value $x \in (a-\delta, a+\delta)$ satisfying $|r(x)| \geq |x-a|$. ∎

The examples we considered using $\chi_{\mathbb{Q}}$ had only finitely many points of continuity. In contrast, Theorem 4.36 tells us that the ruler function, a function with infinitely many continuity points, has no points of differentiability! The proof basically shows that no matter which $a \in \mathbb{R} - \mathbb{Q}$ you consider, as x approaches a, the average slopes jump between having value 0 (as shown in Lemma 4.35) and magnitudes greater than or equal to 1, so they do not settle down to any limit.

The Staircase Function, Revisited

Given our analysis of the ruler function, you may be curious about where the staircase function s from Chapter 3 is differentiable. Exercise 3.8.11 shows that s is continuous at the irrational numbers. However, as of the writing of this book, the author does not know whether anyone has been able to discover where s is differentiable![6] It turns out that there is a theorem due to Lebesgue that says that, roughly, any monotone function must

[6] If you manage to solve this problem, then the author would love to hear your answer! Of course, you will get credit for your answer.

be differentiable "almost everywhere", but this theorem doesn't tell us specifically where the differentiability occurs.

To illustrate the main challenge in solving this problem, recall that the staircase function was built out of approximation functions; we named the n^{th} such approximation g_n. The function g_n has jumps at n different rationals and is flat everywhere else, so $g_n'(x) = 0$ at all $x \in \mathbb{R} - \mathbb{Q}$. As n gets larger, the functions g_n get closer to s, in the sense that at every $x \in \mathbb{R}$, $g_n(x)$ gets arbitrarily close to $s(x)$. However, this does NOT necessarily imply that the values $g_n'(x)$ become arbitrarily close to $s'(x)$. We'll demonstrate why this is the case with a simpler example that we consider next.

Another Function Defined By Approximations

At this point, let's show how we can build a function f which is the sup of a set of approximating functions f_n, i.e. $f(x) = \sup\{f_n(x) \mid n \in \mathbb{N}^*\}$ for each $x \in \mathbb{R}$, but $f'(x)$ will not be the sup of the values $\{f_n'(x) \mid n \in \mathbb{N}^*\}$. This cautionary example demonstrates that even when f_n is arbitrarily close to f, that does not mean that f_n' is arbitrarily close to f'.

In Chapter 3, we demonstrated a way we could list the rational numbers as r_1, r_2, r_3, and so forth, so that $\mathbb{Q} = \{r_n \mid n \in \mathbb{N}^*\}$. Define $h_n : \mathbb{R} \to \mathbb{R}$ by $h_n(x) = 1 - |x - r_n|$. The graph of h_n is a "spike" with the tip at r_n. We'll define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$\forall x \in \mathbb{R}\ f_n(x) = \sup\{h_1(x), h_2(x), \ldots, h_n(x)\}$$

(since we are taking a sup of a finite set, we could also write max instead of sup if we want). A graph of h_1, h_2, h_3, and f_3 is provided in Figure 4.11, where the first three rationals in our list are $r_1 = -1$, $r_2 = 0$, and $r_3 = 1$. (Here, the solid graph is the graph of f_3, and the dashed sections represent the portions of the h_1 through h_3 graphs which are not part of the graph of f.) The idea is that after drawing the graphs of h_1 through h_n, the function f_n takes the "top edge" of all the graphs when drawn on the same axes.

Note that for all $n \in \mathbb{N}^*$ and all $x \in \mathbb{R}$, $f_{n+1}(x) \geq f_n(x)$, because $f_{n+1}(x)$ takes the max of a bigger set than $f_n(x)$ does. Also, we have $f_n(x) \leq 1$. Thus, as n grows, the $f_n(x)$ values increase but remain bounded, so they have a supremum. In fact, the values of $f_n(x)$ become arbitrarily close to the supremum (as was also the case with the staircase function and its approximations).

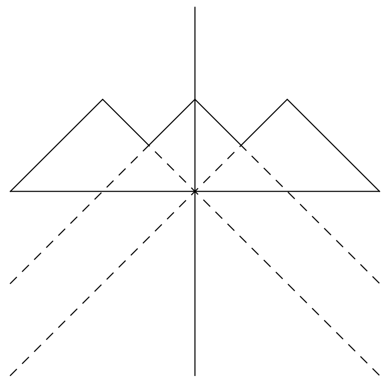

Figure 4.11: The function f_3, made up of spikes h_1 through h_3

We define $f(x) = \sup\{f_n(x) \mid n \in \mathbb{N}^*\}$. You can verify that this means the same thing as saying $f(x) = \sup\{h_n(x) \mid n \in \mathbb{N}^*\}$: see Exercise 4.8.5. Thus, in a sense, f represents the "top edge" after we add infinitely many spikes, one for each rational number. Let's call f the "sup of spikes" function.

Now, how do f and the f_n functions relate to each other? Each spike has one sharp corner where it is not differentiable, but every other point on the spike has derivative 1 or -1. Using this, you can show in Exercise 4.8.6 that f_n has $2n-1$ sharp corners, but everywhere else, the derivative is either 1 or -1. In fact, all the sharp corners will occur at rational points, so for every irrational number x, $f_n'(x)$ is either 1 or -1. In essence, as n grows larger, the graph of f_n becomes more and more jagged.

However, when we look at the limit function $f(x)$, we find that f is quite simple indeed:

Theorem 4.37. *The "sup of spikes" function f is the constant function 1.*

You can prove this theorem in Exercise 4.8.7. The basic idea is that for any fixed $x \in \mathbb{R}$, as we add in more and more spikes, we find rationals r_n which are arbitrarily close to x, so $|x-r_n|$ can become as small as we want. In essence, although the f_n's become more and more jagged, any jagged section of the graph is pushed closer and closer to the line $y = 1$, so in the limit, the graph becomes completely smooth.

In particular, Theorem 4.37 implies that for all $x \in \mathbb{R}$, $f'(x) = 0$, which is NOT a limit of numbers which are either 1 or -1. This shows that the $f_n'(x)$ values do not approach $f'(x)$ as n grows large, which is what we wanted to demonstrate.

For another nice example of a function that is a limit of approximations, and which is smoother than its approximations, see Exercise 4.8.8.

4.8 Exercises

1. Let $A \subseteq \mathbb{R}$ be given. Using Exercise 3.8.2, prove that χ_A has derivative 0 at every point of continuity. By looking at the graph of χ_A, explain why this result makes sense.

2. Let $f : \mathbb{R} \to \mathbb{R}$ be a function which is differentiable everywhere. By Exercise 3.8.3, for each $a \in \mathbb{R}$, $f\chi_{\mathbb{Q}}$ is continuous at a iff $f(a) = 0$.

   Thus, let $a \in \mathbb{R}$ be given which satisfies $f(a) = 0$. Prove that $(f\chi_{\mathbb{Q}})'(a)$ exists iff $f'(a) = 0$. (Hint: How can you write the difference quotient for $f\chi_{\mathbb{Q}}$ from a to x in terms of the difference quotient for f? Can you reuse the result of Exercise 3.8.3?)

3. (a) Suppose that $p : \mathbb{R} \to \mathbb{R}$ is a polynomial, and $a \in \mathbb{R}$ is a zero of p (i.e. $p(a) = 0$). Factor as many powers of $x - a$ out of $p(x)$ as possible, so that for some $r \in \mathbb{N}^*$, we obtain

   $$p(x) = (x - a)^r q(x)$$

   where $q(x)$ is a polynomial satisfying $q(a) \neq 0$. (We say that a is a root *of multiplicity* r.) Prove that $p'(a) = 0$ iff $r \geq 2$.

   (Thus, a zero of p is also a zero of p' iff the zero is repeated in p.)

   (b) Use the previous part, as well as the results of Exercises 4.8.2 and 3.8.3, to prove the following: If S and T are any finite subsets of $\mathbb{R}$ with $S \subseteq T$, then there exists a function $f_{S,T} : \mathbb{R} \to \mathbb{R}$ such that $f_{S,T}$ is continuous precisely at the points of T and differentiable precisely at the points of S.

   (For example, $x\chi_{\mathbb{Q}}(x)$ is a valid choice for $f_{\emptyset,\{0\}}$, and $x^2\chi_{\mathbb{Q}}(x)$ is a valid choice for $f_{\{0\},\{0\}}$.)

4. Prove that the square root of the ruler function, which maps any $x \in \mathbb{R}$ to $\sqrt{r(x)}$, is not differentiable anywhere. (Hint: Note that $\sqrt{r(x)} \geq r(x)$ and imitate the proof of Theorem 4.36.)

5. Using the definitions of h_n, f_n, and f from the "sup of spikes" section, prove that for all $x \in \mathbb{R}$, $f(x) = \sup\{h_n(x) \mid n \in \mathbb{N}^*\}$.

6. This exercise proves some properties of the f_n functions from the "sup of spikes" section. Let $n \in \mathbb{N}^*$ be given. For this particular n, the definition of f_n uses the rational numbers r_1 through r_n. Let's say we write that list of n rational numbers in increasing order as $r_{n,1}$ through $r_{n,n}$, so $r_{n,1}$ is the smallest of them, $r_{n,2}$ is the second smallest, etc. Thus, $r_{n,i}$ is the i^{th} smallest rational in the list r_1 through r_n, where $i \in \mathbb{N}^*$ and $i \le n$.
   (a) Prove that the graph of f_n is *piecewise linear*, which means that f_n can be written with a piecewise definition where each piece is a linear function.
   (b) Explain why f_n is continuous everywhere.
   (c) Where are the sharp corners of f_n?
   (d) Where is the derivative f_n' equal to 1? Where is the derivative equal to -1?

7. Prove Theorem 4.37. (Hint: The result of Exercise 4.8.5 might be helpful, but it is not necessary.)

8. For each $n \in \mathbb{N}^*$, define $g_n : \mathbb{R} \to \mathbb{R}$ by
$$\forall x \in \mathbb{R}\ g_n(x) = \frac{\lfloor nx \rfloor}{n}$$
   (a) Graph g_2, g_3, and g_4.
   (b) For each $n \in \mathbb{N}^*$, note that g_n is monotone increasing. Where are its discontinuities? What are the gaps at those discontinuities?
   (c) For each $n \in \mathbb{N}^*$, where is g_n differentiable? What is the derivative at those points?
   (d) Prove that for all $n \in \mathbb{N}^*$, $|x - g_n(x)| \le 1/n$.

   Remark. From part (d), it follows that as n gets larger, the values of $g_n(x)$ become arbitrarily close to x. Therefore, the "limit" function g for this example is the identity function, with derivative 1 everywhere. Notice that this is NOT a limit of the derivative you find in part (c).

4.9 Extreme Values and the MVT

At this point, we have introduced useful rules for calculating derivatives, and we've seen some interesting examples. However, we still have not addressed the question: what does a function's derivative tell us about the function? Similarly, what do the tangent lines to a function tell us about the graph of the function?

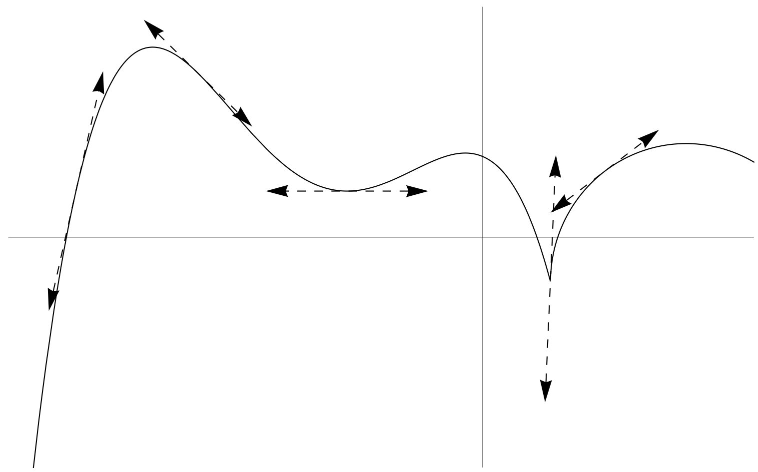

Figure 4.12: A graph with tangent lines drawn at several places

To illustrate the main ideas, see Figure 4.12. In this figure, we have the graph of a function f, together with a few dashed tangent lines drawn touching the graph. Intuitively, the tangent lines "follow the direction" of the graph. It makes sense to conjecture that when the tangent lines go upward (i.e. f' is positive), f is also going upward (i.e. f is strictly increasing), and when the tangent lines go downward (i.e. f' is negative), f is strictly decreasing. This gives us an important relationship between a function and its derivative.

This drawing also points out an important component of this graph: the graph has a few different "peaks", where the curve stops rising and turns around, as well as a few "valleys", where the curve stops falling and turns around. One of these valleys is the sharp corner, but apart from that, it looks like the graph "flattens out" at a peak or a valley. This leads us to conjecture that if f is differentiable at a peak or a valley, then the tangent line is horizontal. Let's first study this conjecture more carefully.

Peaks and Valleys: Relative Extrema

In Chapter 3, when we studied the Extreme Value Theorem (or the EVT), we were introduced to the notion of *absolute extreme values*, which represent the largest and smallest values that a function's range takes. However, as Figure 4.12 indicates, most peaks do not attain the absolute maximum value of the function; instead, the function merely "tops out" there briefly. Similarly, not all valleys correspond to absolute minimum values.

This definition is a more formal way to talk about peaks and valleys:

Definition 4.38. Let f be a real function, and let $a \in \text{dom}(f)$ be given. We say that f attains a *relative maximum* at a if there is some open interval $I \subseteq \text{dom}(f)$ such that $a \in I$ and $f(a) = \max\{f(x) \mid x \in I\}$. Similarly, we say that f attains a *relative minimum* at a if there is some open interval $I \subseteq \text{dom}(f)$ such that $a \in I$ and $f(a) = \min\{f(x) \mid x \in I\}$. If f attains either a relative maximum or relative minimum at a, then $f(a)$ is called a *relative extreme value* or *relative extremum*.

When f attains a relative extremum at a, $f(a)$ is the biggest or smallest value in the range among inputs that are near a, as opposed to among all inputs. To use the rollercoaster analogy for tangent lines, the coaster is about to turn around at these extreme values.[7]

Remark. You may be tempted to say that any absolute extremum is automatically a relative extremum, but there is an important scenario where that statement is false. Suppose that $f : [a, b] \to \mathbb{R}$ has a closed bounded interval as its domain. Any open interval that contains a cannot be contained in $[a, b]$, because such an interval necessarily contains points smaller than a. Similarly, there is no open interval which contains b and is contained inside $[a, b]$. Thus, the endpoints of $[a, b]$ cannot yield relative extrema.[8]

This means that when the domain is a closed bounded interval $[a, b]$, the endpoints of the interval could yield absolute extrema without yielding relative extrema. This becomes useful to keep in mind when finding maximum or minimum values of functions, as we'll see in some examples a little later.

Now, we'll prove that at any relative extremum of a function, either the function flattens out or the function fails to be differentiable (e.g. a sharp corner):

[7] Chances are, somebody is screaming at those points. This definitely gives a new meaning to the word "extreme" used to describe these points!

[8] However, a and b could be *one-sided relative extrema*. We encourage you to formulate a precise definition of this concept.

Theorem 4.39 (Fermat's Theorem). *Let $a \in \mathbb{R}$ be given, and let f be a real function defined around a. If f attains a relative extreme value at a, and $f'(a)$ exists, then $f'(a) = 0$.*

Strategy. If f attains a relative extremum at a, then the function can't be rising on both sides of a, nor can it be falling on both sides. This gives us the idea to show, via a proof by contradiction, that each of the statements "$f'(a) > 0$" and "$f'(a) < 0$" is impossible.

First, suppose that $f'(a) > 0$ were true. In that case, since the difference quotients of f from a to x approach $f'(a)$, the difference quotients must be positive if x is close enough to a, i.e. there is some $\delta > 0$ so that

$$\frac{f(x) - f(a)}{x - a} > 0 \qquad \text{if } 0 < |x - a| < \delta$$

This implies that for x close enough to a, $f(x) > f(a)$ when $x > a$ (which is impossible if f has a relative maximum at a), and $f(x) < f(a)$ when $x < a$ (which is impossible if f has a relative minimum at a). This finishes the contradiction when $f'(a) > 0$. The case where $f'(a) < 0$ can be handled similarly.

Proof. Let a, f be given as described. Since f attains a relative extremum at a, we may choose some open interval $I \subseteq \text{dom}(f)$ so that $a \in I$ and either $f(a) = \max\{f(x) \mid x \in I\}$ or $f(a) = \min\{f(x) \mid x \in I\}$. Furthermore, we assume that $f'(a)$ exists, so to prove $f'(a) = 0$, it suffices to show that "$f'(a) > 0$" and "$f'(a) < 0$" are each impossible.

For the first case, let's assume that $f'(a) > 0$. By the definition of derivative, when $\epsilon = f'(a)/2$, there exists some $\delta > 0$ so that

$$\forall x \in \text{dom}(f) \ \left(0 < |x - a| < \delta \to \left|\frac{f(x) - f(a)}{x - a} - f'(a)\right| < \frac{f'(a)}{2}\right)$$

In fact, we may choose such a value for δ small enough so that $(a-\delta, a+\delta) \subseteq I$ (because I is an open interval). Thus, for any $x \in I$, we have

$$\frac{f(x) - f(a)}{x - a} \in \left(\frac{f'(a)}{2}, \frac{3f'(a)}{2}\right)$$

so that, in particular, the difference quotient is positive. This means that for all $x \in I$, $f(x) > f(a)$ is true iff $x > a$ is true. In particular, when

$x = a + \delta/2$, $f(x) > f(a)$ (which is a contradiction if f attains a relative maximum at a), and when $x = a - \delta/2$, $f(x) < f(a)$ (which is a contradiction if f attains a relative minimum at a). This finishes the first case.

For the second case, when $f'(a) < 0$, we could either use a similar style of proof as in the first case. or we could merely apply the argument for the first case to the function $-f$ (because $-f$ also has a relative extremum at a, and $(-f)'(a) > 0$). ■

Note that Theorem 4.39 says that IF $f'(a)$ exists at a relative extremum, THEN $f'(a) = 0$. It is easy to mix this up with the converse statement, which says that if $f'(a) = 0$, then f has a relative extremum at a. This converse is false: see Exercise 4.10.1. To help distinguish between points where f' is 0 and points where a relative extremum is obtained, we use the term *critical value* to denote a value $a \in \mathbb{R}$ where either $f'(a) = 0$ or where $f'(a)$ does not exist.

The critical values represent the POSSIBLE places where one could find a relative extremum, but not all critical values necessarily yield relative extrema. Nevertheless, for many functions, Theorem 4.39 is still very useful for finding extreme values of a function, as the following example indicates.

Example 4.40:

Suppose that $f : [1, 2] \to \mathbb{R}$ is defined by $f(x) = (x^3 - 6x)^{-2}$ for all $x \in [1, 2]$. Note that f is continuous on $[1, 2]$: the denominator $x^3 - 6x$ is zero when $x = 0$ or when $x = \pm\sqrt{6}$, and none of these values are in the interval $[1, 2]$. Thus, the Extreme Value Theorem asserts that f attains an absolute maximum and an absolute minimum. Using Theorem 4.39, we can determine where those absolute extremes are located.

To do this, we find the critical values. For any $x \in (1, 2)$, the Chain Rule tells us that

$$f'(x) = -2(x^3 - 6x)^{-3} \cdot (3x^2 - 6) = \frac{-6(x^2 - 2)}{(x^3 - 6x)^3}$$

Since this derivative is defined for all $x \in (1, 2)$, the critical values occur when $f'(x) = 0$, i.e. $x^2 - 2 = 0$, or $x = \pm\sqrt{2}$. Since $-\sqrt{2}$ is not in the interval $(1, 2)$, $x = \sqrt{2}$ is our only critical value.

Now, according to the remark following Definition 4.38, the absolute extreme values of the function either occur at the endpoints of $[1, 2]$ or they occur at relative extrema. Therefore, we check the values of f at our endpoints and at our critical values and see that

$$\begin{aligned} f(1) &= \frac{1}{(1-6)^2} = \frac{1}{25} \\ f(\sqrt{2}) &= \frac{1}{(2\sqrt{2}-6\sqrt{2})^2} = \frac{1}{32} \\ f(2) &= \frac{1}{(8-12)^2} = \frac{1}{16} \end{aligned}$$

Hence, the absolute maximum is at $x = 2$, and the absolute minimum is at $x = \sqrt{2}$. $\square$

The previous example illustrates a general strategy called the *closed interval method* for finding extrema of a continuous function f defined on a closed bounded interval:

1. Determine where f' fails to exist.
2. For the values where f' exists, determine which values $x \in \text{dom}(f')$ satisfy $f'(x) = 0$.
3. Plug all the crticial values from Steps 1 and 2 into f, as well as plugging in the endpoints of $\text{dom}(f)$, to find the largest and smallest values of $\text{ran}(f)$.

Remark. At this point, you might be wondering: "Can't I just find maximum and minimum values by drawing a graph?" In some cases, drawing a graph is a pretty good way to analyze how a function behaves and to see where the function takes large or small values.

However, there are several issues with drawing a graph to find extrema. One issue is precision: any drawing of a graph only plots finitely many points of the graph and then approximates the rest by a smooth curve, so if your drawing doesn't use enough points, then your drawing might miss some key features of the function. Another related issue is that a graph is an approximation: usually a graph will not tell you the EXACT value of x where f attains an extremum. (In the previous example, especially, you might be able to use a graph to find that the minimum is attained when $x \approx 1.4142$, but the graph would not tell you that the exact value of x is $\sqrt{2}$.)

When using derivatives, however, we are often able to find the peaks and valleys of a function easily and accurately. We will soon prove a theorem that will help us decide whether a critical value is a relative extremum, and this theorem will be very helpful to us in analyzing functions.

In many practical problems, we would like to find the largest or smallest value possible for some quantity. Examples include determining which shape a gallon of water should take in order to minimize surface area, determining how light should travel from a point in the sky to a point in a pool to minimize travel time, or determining how many toys a factory should produce to obtain the maximum profit. These are called *optimization* problems, and these are solved by finding absolute extremes of functions. We'll go over a simple example here, and we'll provide other problems as exercises.

Example 4.41:

Suppose that you have 40 yards of fence, and you'd like to use that fence to enclose a rectangular area. We'd like to know: which dimensions of this rectangle will produce the largest area possible?

First, just like we did with related rates problems, we should identify the relevant quantities of the problem. A rectangle is specified by its width w and height h, which we'll measure in yards. We would like to maximize the area A of the rectangle, measured in square yards.

Second, we need to know some relationships between these quantities. For a rectangle, we know that $A = wh$. However, this writes A as a function of two variables, w and h, and in order to use the closed interval method, we need A to be a differentiable function of one variable. Thus, we need to rewrite one variable in terms of the other. To do this, we use the fact that the perimeter of the rectangle should use up all 40 yards of fence, i.e. $2w + 2h = 40$. Therefore, $h = 20 - w$, and thus $A(w) = w(20 - w)$ is now a function of w. We treat A as having domain $[0, 20]$. (Arguably, the values 0 and 20 may not make much sense as possible values of w, since length measurements should normally be positive, but it makes our life easier to have a closed interval domain for A, and furthermore we can think of the values of 0 and 20 as producing a rectangle of no thickness, i.e. 0 area.)

Since $A(w) = w(20 - w) = 20w - w^2$, we find that $A'(w) = 20 - 2w$ for all $w \in (0, 20)$. The only critical value occurs when $20 - 2w = 0$, i.e. when $w = 10$. Plugging in our endpoints and critical value to A, we find $A(0) = A(20) = 0$ and $A(10) = 100$. Therefore, the maximum area of 100 square yards is obtained by choosing w to be 10 yards and h to be $20 - 10$, or 10, yards. In other words, a square yields the maximum area out of all the possible rectangles. □

Rolle's Theorem and the Mean Value Theorem

Now that we have the closed interval method, we should take a closer look at functions of the form $f : [a, b] \to \mathbb{R}$, where the domain is a closed bounded interval. In the last chapter, we considered when f is continuous on $[a, b]$ and proved some useful theorems, such as the Intermediate Value Theorem, the Extreme Value Theorem, and a uniform continuity theorem. Now, let's consider when f is continuous on $[a, b]$ and differentiable on (a, b) (remember that we only have one-sided derivatives at the endpoints), and let's see what we can infer from the differentiability assumption.

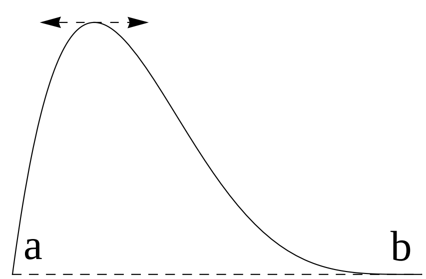

Figure 4.13: A function f with $f(a) = f(b)$ and a flat tangent line

As a starting point in our analysis, Theorem 4.39 gives us a way to find places where the derivative is 0 by looking for relative extrema. Note that not all continuous and differentiable functions from $[a, b]$ to $\mathbb{R}$ have relative extrema: if the function is strictly monotone, then it obtains its absolute extrema at the endpoints of $[a, b]$ and has no relative extrema in the interior (a, b). However, if the function satisfies $f(a) = f(b)$, then the function cannot be strictly monotone, and intuitively if such a function's graph rises or falls, then it must eventually "turn around" to get back to the value of $f(b)$. This point where the graph turns around is a relative extremum, as pictured in Figure 4.13. This suggests the following theorem:

Theorem 4.42 (Rolle's Theorem). *Let $a, b \in \mathbb{R}$ with $a < b$ be given, and let $f : [a, b] \to \mathbb{R}$ be a function which is continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists some $c \in (a, b)$ such that $f'(c) = 0$.*

Strategy. We'd like to find some place where $f'(c) = 0$. One way to do this is to find a relative extremum of f, since Theorem 4.39 says that if f has a relative extremum at c, where $c \in (a, b)$, then $f'(c) = 0$. However, does any relative extremum exist?

We know that ABSOLUTE extrema exist by the EVT, because f is continuous on $[a, b]$. If either an absolute minimum or an absolute maximum occurs in (a, b), then that point will also be a relative extremum and we'll be done. If this is not the case, then both absolute extrema occur at the endpoints. However, since $f(a) = f(b)$, this means that the maximum and minimum values of the range are the same, so the function is constant! In this case, we trivially have $f'(c) = 0$ for EVERY $c \in (a, b)$.

Proof. Let a, b, f be given as described. Because f is continuous on $[a, b]$, by the Extreme Value Theorem (Theorem 3.59), f attains an absolute maximum and an absolute maximum. There are two cases to consider.

In the first case, both absolute extrema occur at the endpoints. Hence, either $f(a) = \min(\text{ran}(f))$ and $f(b) = \max(\text{ran}(f))$ or vice versa, but in either case, since $f(a) = f(b)$, we have $\min(\text{ran}(f)) = \max(\text{ran}(f))$. This means that f is a constant function, so $f'(c) = 0$ for any $c \in (a, b)$.

In the second case, the first case does not hold, so some absolute extremum c must occur in (a, b). That absolute extremum is also a relative extremum. Because f is differentiable on (a, b), f is differentiable at c, so Theorem 4.39 implies that $f'(c) = 0$. ■

Rolle's Theorem tells us that if a differentiable function repeats values, then the function must have a derivative of 0 at some point. The contrapositive of this statement is useful: it says that if the derivative is never 0, then the function never repeats values, i.e. it is injective. Here is an application of this.

Example 4.43:

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^5 - x^3 + x + 2$ for all $x \in \mathbb{R}$, so f is a polynomial of degree 5. This means that f can have at most 5 different real roots. We'll prove that f has exactly one real root.

First, why does f have any roots at all? We note that $f(x) < 0$ if x is a negative number with large magnitude; for instance, $f(-10) = -10^5 + 10^3 - 10 + 2 = -100000 + 1000 - 10 + 2$, which is clearly negative. Similarly, $f(x) > 0$ if x is positive and has large magnitude; for instance, $f(10) =$

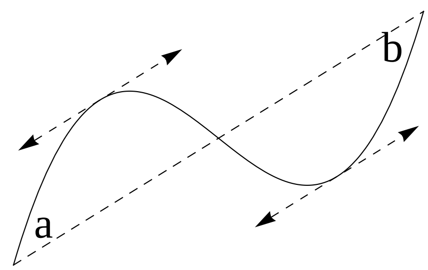

Figure 4.14: A secant line from $(a, f(a))$ to $(b, f(b))$ and parallel tangent lines

$10^5 - 10^3 + 10 + 2$, which is clearly positive. Thus, since f is continuous on $[-10, 10]$, Bolzano's Theorem (or the Intermediate Value Theorem) says that f has a root in $(-10, 10)$.

Second, why does f have only one root? If we assume that f has two roots, say $f(a) = f(b) = 0$ with $a < b$, then since f is differentiable everywhere (and thus on (a, b)), Rolle's Theorem implies that $f'(c) = 0$ for some $c \in (a, b)$. However, for all $x \in \mathbb{R}$,

$$f'(x) = 5x^4 - 3x^2 + 1 = 5(x^2)^2 - 3(x^2) + 1$$

so f' is a quadratic function in x^2. Applying the quadratic formula, we find no real solutions for x^2, because $(-3)^2 - 4(5)(1) < 0$. Thus, f' has no zeroes, and the statement $f'(c) = 0$ is impossible.

(In fact, Exercise 4.8.3.(a) implies that the unique root of f must be a single root, i.e. it has multiplicity 1.) $\square$

As Figure 4.13 indicates, another way we can view Rolle's Theorem is that it says somewhere in (a, b), the tangent line is parallel to the secant line which connects $(a, f(a))$ to $(b, f(b))$, when that secant line is flat. In other words, the average rate of change from a to b equals the instantaneous rate of change somewhere in (a, b). This raises the question: if the secant line from a to b is not flat, is there still some tangent line parallel to it?

As Figure 4.14 indicates, it seems plausible that there exists some tangent line parallel to the secant line. In other words, the average slope from a to b is equal to the instantaneous slope somewhere in (a, b). For instance, if you drive a car 90 miles over the course of 2 hours, so your average speed is 45mph, then at some point in the trip your speedometer must have read 45mph; after all, you couldn't have been going slower than 45mph the whole time, nor could you have been faster the whole time! The next theorem, often abbreviated as the MVT, makes this idea precise:

Theorem 4.44 (Mean Value Theorem). *Let $a, b \in \mathbb{R}$ with $a < b$ be given, and let $f : [a, b] \to \mathbb{R}$ be a function which is continuous on $[a, b]$ and differentiable on (a, b). Then there exists some $c \in (a, b)$ satisfying*

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

or equivalently $f(b) - f(a) = f'(c)(b - a)$.

Remark. When looking at linear approximations, we established that $f(x) - f(a) \approx f'(a)(x - a)$, where the error in the approximation is "better than linear". The MVT says that $f(x) - f(a) = f'(c)(x - a)$ for some c between a and x. This has the same form as $f'(a)(x - a)$, but the derivative is evaluated at a different point. This means that the error of the linear approximation is proportional to how far apart $f'(a)$ and $f'(c)$ are: see Exercise 4.10.17 for a deeper study of this.

Strategy. The secant line from a to b has the following equation in point-slope form:

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

We'd like to find a point where the graph of f has the same slope as this line. In essence, at such a point, the graph of f and the secant line are moving side by side, rather than approaching or retreating from each other. Another way to look at this is: we'd like to find a point where the distance between the graph of f and the secant line is not changing. This gives us the idea to consider the function

$$h(x) = f(x) - \left(f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right)$$

which is the difference between the function f and the secant line. By design, the secant line goes through $(a, f(a))$ and $(b, f(b))$, so we have $h(a) = h(b) = 0$. This means we can apply Rolle's Theorem to h, and the rest follows quickly.

Proof. Let a, b, f be given as described. Define the function $h : [a, b] \to \mathbb{R}$ by

$$h(x) = f(x) - \left(f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right)$$

for all $x \in [a, b]$. Since h is just the difference of f and a linear function, h is continuous on $[a, b]$ and differentiable on (a, b). Furthermore, $h(a) = h(b) = 0$, so by Rolle's Theorem, there exists some $c \in (a, b)$ satisfying $h'(c) = 0$. Since

$$h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0$$

we have $f(b) - f(a) = f'(c)(b - a)$, as desired. ■

Applications of the MVT

Because the Mean Value Theorem relates average slopes and instantaneous slopes, it is probably the most important theorem that indicates how properties of the derivative relate to properties of the original function. For instance, inequalities involving f' can be used to find inequalities involving f, as the following indicates:

Example 4.45:

Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable everywhere and satisfies $|f'(x)| < 1$ for all $x \in \mathbb{R}$. Let's consider the question: if $f(-1) = 2$, then what can we conclude about the value of $f(1)$?

Since we have information about f', we have information about how fast the function f changes. This gives us the idea to consider the CHANGE in f from -1 to 1, i.e. $f(1) - f(-1)$. By the MVT, there exists some $c \in (-1, 1)$ which satisfies $f(1) - f(-1) = f'(c)(1 - (-1))$. By plugging in $f(-1) = 2$ and taking absolute values, we therefore find

$$|f(1) - 2| = 2|f'(c)| < 2 \cdot 1$$

This means that $f(1)$ is less than 2 units away from 2, so $f(1) \in (0, 4)$.

One way to think of this is that the condition $|f'(x)| < 1$ imposes a "speed limit" on f: the speed at which f rises or falls must be less than 1. In fact, by doing similar steps to what we did above, we find that for any $x \in \mathbb{R}$, $f(x)$ lies between $2-|x+1|$ and $2+|x+1|$. This situation is drawn in Figure 4.15, where the dashed lines represent $2 - |x + 1|$ and $2 + |x + 1|$, and the graph of f must travel "between the speed limits". □

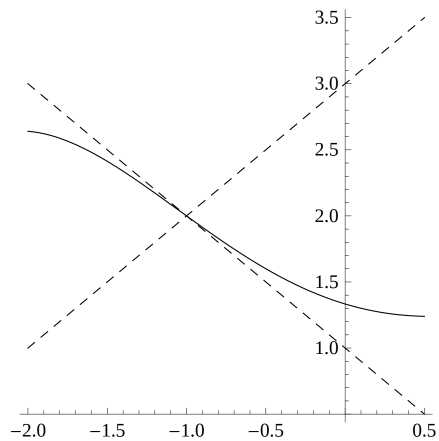

Figure 4.15: A function f with $f(-1) = 2$ and $|f'(x)| < 1$ for all x

Example 4.46:

We've seen in Example 3.38 that for all $x \in (0, \pi/2)$, $\sin x < x$. We can prove something more general with the Mean Value Theorem. Suppose that $a, b \in \mathbb{R}$ are given with $a < b$; let's see how much the sin function changes from a to b. Using the MVT, there is some $c \in (a, b)$ such that

$$|\sin b - \sin a| = |(\cos c)(b - a)| \leq |b - a|$$

If we restrict the domain of sin, we can say even more. For instance, if we take a and b from the interval $[0, 1]$, then we have $|\sin b - \sin a| = |\cos c|\,|b - a| < |b - a|$ since $0 < \cos c < 1$ for any $c \in (0, 1)$. Also, $\sin a$ and $\sin b$ are in $[0, 1]$ as well, so we may apply sin to them (treating them as the "new" a and b) and conclude $|\sin(\sin b) - \sin(\sin a)| < |\sin b - \sin a|$. Doing this repeatedly yields

$$\begin{aligned} |b - a| &> |\sin b - \sin a| \\ &> |\sin(\sin b) - \sin(\sin a)| \\ &> |\sin(\sin(\sin b)) - \sin(\sin(\sin a))| > \cdots \end{aligned}$$

In essence, every time sin is applied to two values in $[0, 1]$, the distance between the values goes down. This is sometimes described by saying that sin is a *contraction* on $[0, 1]$. There is a theorem proven in upper-level analysis courses that says if f is a contraction on a closed bounded interval, and x is given in that interval, then the sequence $x, f(x), f(f(x)), \ldots$ (obtained by repeatedly applying f) approaches a limit L satisfying $f(L) = L$. In our case, the sequence $a, \sin a, \sin(\sin a), \ldots$ approaches the limit 0 (since 0 is the only value $L \in [0, 1]$ satisfying $\sin L = L$).

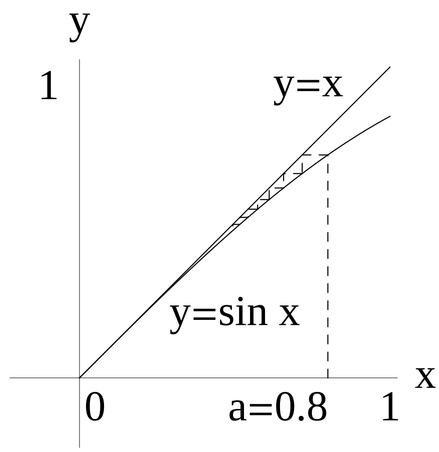

Figure 4.16: Repeatedly applying sin to $a = 0.8$

To illustrate how a contraction works, look at Figure 4.16. Here, we have graphs $y = \sin x$ and $y = x$ on $[0, 1]$. In this figure, suppose that $a = 0.8$, which we illustrate by positioning ourselves at $(0.8, 0)$. Next, to apply sin, we move up to the graph of sin, making our y-coordinate equal to $\sin(0.8)$. Next, we move left to the graph of $y = x$, so that our x-coordinate is now $\sin(0.8)$. Then, we move back to the graph of sin, so our y-coordinate is now $\sin(\sin(0.8))$, and so on. Our path of motion is displayed in the figure by a collection of dashed line segments starting at $(a, 0)$.

The picture conveys that the dashed lines are repeatedly "bouncing" between the two graphs, moving down and left. Because sin is a contraction on $[0, 1]$, the vertical distance between the two graphs is shrinking as we move more to the left, so eventually the bouncing settles down at the origin. This illustrates the fact that repeatedly applying sin to a makes the values approach 0. $\square$

The previous examples show how the Mean Value Theorem can be used to study changes in distance. Another very important use of the MVT is to help us determine whether a function rises or falls, as this theorem shows:

Theorem 4.47. *Let $a, b \in \mathbb{R}$ be given, and suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b). Then the following holds:*

1. *If $f'(x) > 0$ for all $x \in (a, b)$, then f is strictly increasing on $[a, b]$.*

2. *If $f'(x) < 0$ for all $x \in (a, b)$, then f is strictly decreasing on $[a, b]$.*

3. *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant on $[a, b]$.*

Strategy. If $x, y \in [a, b]$ are given with $x < y$, we'd like to know whether $f(x) < f(y)$, $f(x) > f(y)$, or $f(x) = f(y)$. We can use the MVT with the difference $f(y)-f(x)$ to find out whether that difference is positive, negative, or zero.

Proof. Let a, b, f be given as described in the first sentence of the theorem. Let $x, y \in [a, b]$ with $x < y$ be given. By the MVT, there exists $c \in (x, y)$ such that $f(y) - f(x) = f'(c)(y - x)$. Since $y - x$ is positive, we find that $f(y) > f(x)$ if $f'(c) > 0$, $f(y) < f(x)$ if $f'(c) < 0$, and $f(y) = f(x)$ if $f'(c) = 0$, proving all three parts of the theorem. ■

We already used the closed interval method to find out relative extrema of functions. Now, equipped with Theorem 4.47, we may also find out where a function rises and falls, as shown in the next example.

Example 4.48:

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x/(x^2 + 1)$ for all $x \in \mathbb{R}$. Note that the domain of this function is not a closed bounded interval, so the closed interval method does not tell us where the function obtains maximum or minimum values. However, we can still obtain useful information from the derivative. We compute, for all $x \in \mathbb{R}$,

$$f'(x) = \frac{(x^2 + 1)(1) - (2x)(x)}{(x^2 + 1)^2} = \frac{1 - x^2}{(x^2 + 1)^2}$$

We have $f'(x) = 0$ when $x = \pm 1$. These roots of the derivative split the real line into three intervals: $(-\infty, -1)$, $(-1, 1)$, and $(1, \infty)$. We note that $f'(x) < 0$ for $x \in (-\infty, -1) \cup (1, \infty)$ and $f'(x) > 0$ for $x \in (-1, 1)$. Therefore, by Theorem 4.47, f is strictly decreasing on $(-\infty, -1] \cup [1, \infty)$ and strictly increasing on $[-1, 1]$. (Technically, the theorem should be modified to account for infinite intervals, but you should be able to perform this modification yourself.) Since f is decreasing to the left of $x = -1$ and increasing to the right, f has a relative minimum at $x = -1$. Similarly, f is increasing to the left of $x = 1$ and decreasing to the right, so f has a relative maximum at $x = 1$.

It also helps to note a few useful things about f before we draw its graph. First of all, the denominator of $f(x)$ is always positive, so $f(x) < 0$ when $x < 0$ and $f(x) > 0$ when $x > 0$. At the relative extrema, we have $f(-1) = -1/2$ and $f(1) = 1/2$. There is a zero of f at $x = 0$. Since $f(-x) = -f(x)$ for all $x \in \mathbb{R}$, f is an odd function, so its graph is rotationally symmetric about the origin. Lastly, as x goes to either ∞ or $-\infty$ (as defined in the Exercises of Section 3.4), we have $f(x) \to 0$. All of this leads to the drawing in Figure 4.17.

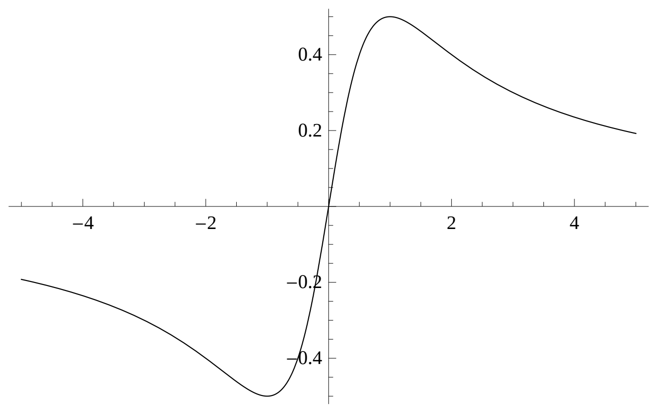

Figure 4.17: A sketch of $f(x) = \dfrac{x}{x^2+1}$

In fact, since f is increasing on $[0, 1]$, decreasing on $[1, \infty)$, and takes negative values on $(-\infty, 0]$, we see that f obtains its ABSOLUTE maximum at $x = 1$. Similarly, f attains its absolute minimum at $x = -1$. $\square$

In that example, we made use of the sign of f' to determine which critical values were relative maxima and which were relative minima. That process can be summarized in the following theorem, which you can prove as Exercise 4.10.22:

Theorem 4.49 (First Derivative Test). *Let $a \in \mathbb{R}$ be given, and let f be a real function which is continuous on an open interval I containing a.*

1. *If for all $x \in I$, $f'(x) > 0$ when $x < a$ and $f'(x) < 0$ when $x > a$, then f has a relative maximum at a.*

2. *If for all $x \in I$, $f'(x) < 0$ when $x < a$ and $f'(x) > 0$ when $x > a$, then f has a relative minimum at a.*

Therefore, armed with Theorem 4.39 to tell us about critical values and Theorems 4.47 and 4.49 to tell us about where a function rises and falls, we can easily find important features of a function's graph. These theorems address the question we posed at the beginning of this section, asking how the derivative of a function influences the function itself.

In the exercises, you can analyze another useful tool: the second derivative f''. Whereas the first derivative tells us whether the original function rises or falls, the second derivative tells us whether the first derivative rises or falls. In other words, the second derivative tells us whether the function becomes steeper or shallower, and thus it tells us information about the way the function "curves". More details are in the exercises.

4.10 Exercises

1. Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ and a number $a \in \mathbb{R}$ such that f is differentiable everywhere, $f'(a) = 0$, but f does not attain a relative extremum at a.

For the next three exercises, use the closed interval method to find all critical values, the absolute maximum of the function, and the absolute minimum.

2. $f : [-5, 5]$, $f(x) = \dfrac{1}{4}x^4 + x^3 - 2x^2$
3. $f : [0, 2]$, $f(x) = x - \sqrt{x}$
4. $f : [-\pi, \pi]$, $f(x) = \dfrac{\sin x}{2 + \cos x}$

For the next three exercises, solve the given optimization problem. You may use the closed interval method where it applies, or you may use Theorems 4.47 and 4.49.

5. A can is shaped like a circular cylinder, with a base of radius r and a height of h. If the volume of the can is one cubic meter, then which choices of h and r make the surface area smallest?

6. Let's suppose you are walking from the beach back to your car. You start 100 feet west and 20 feet south of your car. The northernmost border of the beach is road, but everything else is sand. You can walk 6 feet per second on the road, but on the sand, you can only walk 3 feet per second.

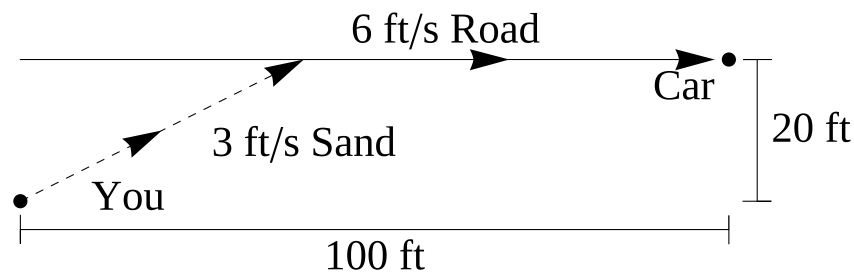

Figure 4.18: Walking from the beach to your car

You would like to arrive at your car as quickly as possible. Thus, as drawn in Figure 4.18 by the paths with the arrows, you need to decide how far you wish to walk on the sand before walking the rest of the distance on the road. Which path should you take to minimize the time you need to walk to your car?

7. Suppose you would like to build a cone-shaped drinking cup out of a circular piece of paper of radius r. To do this, you cut out a circular sector from the paper with angle θ, and then you tape the two edges together that you just cut. In Figure 4.19, the picture on the left shows the circle (with the sector that you cut out), and the picture on the right shows the resulting cone. Note that the radius of the circle becomes the slant height of the cone.

   Which choice of θ yields the maximum volume of your cone-shaped cup? (Hint: The top of your cone is a circle. Find the circumference of that circle first, from which you can get the cone's radius, labeled in the figure as w.)

8. This exercise outlines an informal proof of *Snell's Law* in physics. The main idea is that when light moves from one medium (like air), to another medium (like water), the path of the light bends because the speed of light is different in the two media. Let's consider the situation

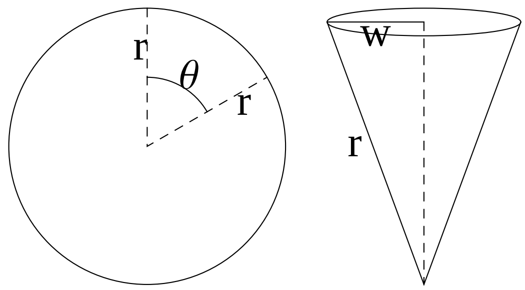

Figure 4.19: Making a cone out of a circle

drawn in Figure 4.20, where light passes through the air with speed v_1 at an angle of α_1, and then the light proceeds through the water with speed v_2 at an angle of α_2. We suppose that the light travels L feet horizontally in total from the top-right corner to the bottom-left corner. Since light travels along the path that minimizes total travel time, we want to find which path the light will take in order to minimize its total time T of travel.

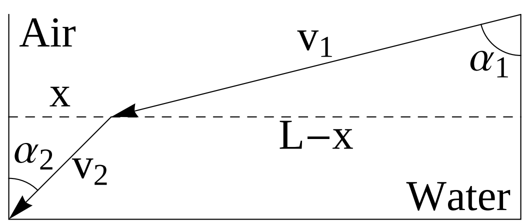

Figure 4.20: Snell's Law

(a) Suppose that x is the horizontal distance from the left end to the point where the light crosses the water. Find a formula for T in terms of x, v_1, v_2, α_1, α_2, and L.

(b) Find $T'(x)$, and show that it equals 0 when

$$\frac{\sin \alpha_1}{v_1} = \frac{\sin \alpha_2}{v_2}$$

(Note that all our variables except T and x are constants with respect to x.)

The formula at the end of part (b) is Snell's Law.

9. Prove that the function $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = 1.1x - \sin x + 2$ for all $x \in \mathbb{R}$, has exactly one root.

10. Suppose $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = 1 - x^{2/3}$ for all $x \in \mathbb{R}$. Show that $f(-1) = f(1) = 0$, but there is no $c \in (-1, 1)$ such that $f'(c) = 0$. Why doesn't this contradict Rolle's Theorem (Theorem 4.42)?

11. Prove the following by induction on $n \in \mathbb{N}^*$: if $f : \mathbb{R} \to \mathbb{R}$ is n-times differentiable everywhere (i.e. $f^{(n)}$ exists), and f has at least $n + 1$ distinct roots, say $a_1 < a_2 < \cdots < a_{n+1}$ satisfy $f(a_1) = f(a_2) = \cdots = f(a_{n+1}) = 0$, then there exists some $c \in (a_1, a_{n+1})$ satisfying $f^{(n)}(c) = 0$.

12. Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and twice differentiable (i.e. f'' exists) on (a, b). Suppose that the line from $(a, f(a))$ to $(b, f(b))$ (in other words, the secant line from a to b) intersects the graph of f at $(c, f(c))$ for some $c \in (a, b)$. (For example, see the picture from Figure 4.14.) Prove that there exists some $t \in (a, b)$ satisfying $f''(t) = 0$. (Hint: Use the MVT multiple times.)

13. This exercise outlines a proof of the following result, called the *intermediate value property for derivatives* (and is also known as *Darboux's Theorem*):

    Theorem 4.50. *Suppose that f is a real function which is differentiable on an open interval I. For any $a, b \in I$ with $a < b$, f' attains every value between $f'(a)$ and $f'(b)$ somewhere in (a, b). In other words, if k is any number between $f'(a)$ and $f'(b)$ then there exists some $c \in (a, b)$ so that $f'(c) = k$.*

    Note that this theorem does NOT assume that f' is continuous, so this theorem is not the same as the Intermediate Value Theorem. In this exercise, we provide a five-step proof outline. (Also, see the next exercise.)

    (a) First, define $g : [a, b] \to \mathbb{R}$ by

    $$g(x) = \begin{cases} \frac{f(x)-f(a)}{x-a} & \text{if } x \neq a \\ f'(a) & \text{if } x = a \end{cases}$$

    Thus, g calculates difference quotients of f which use the left endpoint a. Prove that g attains every value between $f'(a)$ and $g(b)$ somewhere in (a, b).

(b) Suppose that $k \in \mathbb{R}$ is between $f'(a)$ and $g(b)$. Therefore, by part (a), there is some $d \in (a, b)$ such that $g(d) = k$. Use the MVT to show that there exists some $c \in (a, d)$ such that $f'(c) = k$.

(c) Next, define $h : [a, b] \to \mathbb{R}$ by

$$h(x) = \begin{cases} \frac{f(x)-f(b)}{x-b} & \text{if } x \neq b \\ f'(b) & \text{if } x = b \end{cases}$$

Thus, h calculates difference quotients of f which use the right endpoint b. Prove that for all values k between $f'(b)$ and $h(a)$, there is some $d \in (a, b)$ such that $h(d) = k$.

(d) Using the MVT and the choice of d from part (c), prove that there exists $c \in (d, b)$ such that $f'(c) = k$.

(e) Note that $g(b) = h(a)$. Using this fact, finish the proof of Theorem 4.50.

14. In this exercise, we outline another proof of the IVT for derivatives, as mentioned in the previous exercise. Let f, I, a, b, k be given as described in that theorem. Suppose WLOG that $f'(a) < k < f'(b)$ (otherwise, we consider $-f$ instead of f).

We define $g : I \to \mathbb{R}$ as follows for all $x \in I$:

$$g(x) = f(x) - kx$$

Note that g is differentiable on I, $g'(a) = f'(a) - k < 0$ and $g'(b) = f'(b) - k > 0$. Hence, to prove our theorem, we need only show that g' has a root in (a, b).

(a) Explain why g attains an absolute minimum on $[a, b]$. Let's suppose this minimum is attained at $(c, g(c))$.

(b) Using the fact that $g'(a) < 0$, prove that $c \neq a$. (Hint: Theorem 4.39 doesn't technically apply, but its main idea does.)

(c) Using the fact that $g'(b) > 0$, prove that $c \neq b$.

(d) From the previous parts, we see $c \in (a, b)$. Explain why $g'(c) = 0$. This finishes the proof.

15. Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$. Recall from Exercise 3.12.8 that f is Lipschitz-continuous on $[a, b]$ if there exists some $L > 0$ such that for all $x, y \in [a, b]$, $|f(x) - f(y)| \leq L|x - y|$.

(a) Prove that if f is differentiable on (a, b), and the derivative is bounded on (a, b), then f is Lipschitz-continuous on $[a, b]$.
(b) Give an example to show that if f is Lipschitz-continuous on $[a, b]$, then that does not imply that f is differentiable on (a, b).
(c) Show that if f is Lipschitz-continuous on $[a, b]$, and f is differentiable on (a, b), then f' is bounded on (a, b).

16. Let $f : \mathbb{R} \to \mathbb{R}$ be a function with the following property: there exists some $r > 1$ and some $L > 0$ such that for all $x, y \in \mathbb{R}$,

$$|f(x) - f(y)| \leq L|x - y|^r$$

Certainly f is Lipschitz-continuous on $\mathbb{R}$. However, we can show a stronger property. Prove that f is a constant function!

17. This exercise analyzes the error from a linear approximation more closely. Let $a \in \mathbb{R}$ be given, and let f be a real function which is twice differentiable (i.e. f'' exists) around a. For all $x \in \text{dom}(f)$, define $L : \text{dom}(f) \to \mathbb{R}$ by

$$L(x) = f(a) + f'(a)(x - a)$$

so that L is the linear approximation of f at a. Therefore, the error function E in the linear approximation is $E = f - L$.

(a) Let x near a be given (i.e. x is close enough to a so that f'' exists for all values from x to a). Prove that there exists some $c \in \text{dom}(f)$ between x and a such that

$$E(x) = (f'(c) - f'(a))(x - a)$$

(b) Use part (a) to prove that for all x near a, there exists some $d \in \text{dom}(f)$ between x and a such that

$$|E(x)| \leq |f''(d)||x - a|^2$$

Therefore, if M is an upper bound on the values $|f''(d)|$ for all d between x and a, then $|E(x)| \leq M|x - a|^2$.
(c) Suppose that f is the square root function and $a = 1$. Use a linear approximation to estimate $\sqrt{1.001}$, and use part (b) to find a bound on the error of this approximation.

Remark. In Chapter 8, we'll show that if a function has an $(n+1)^{\text{st}}$ derivative near a, then we can approximate the function with a polynomial of degree n, and the error will be bounded by something of the form

$$|f^{(n+1)}(d)|\frac{|x-a|^{n+1}}{(n+1)!}$$

18. Let's consider the following generalization of the MVT, which will be useful when we study L'Hôpital's Rule in Chapter 7:

Theorem 4.51 (Cauchy's Mean Value Theorem). *Let $a, b \in \mathbb{R}$ be given with $a < b$, and let $f, g : [a, b] \to \mathbb{R}$ be given. Assume that f and g are both continuous on $[a, b]$ and differentiable on (a, b). Also assume that g' is nonzero on (a, b) (thus, by Rolle's Theorem, $g(a)$ can't equal $g(b)$). Then there is some $c \in (a, b)$ satisfying*

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

or equivalently $(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$.

Note that the MVT is a special case of Theorem 4.51 where $g(x) = x$ for all $x \in [a, b]$.

(a) Here is an attempt at a proof of Theorem 4.51:

Proof. Let a, b, f, g be given as described. By the MVT, there exists some $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$. By the MVT again, there exists some $c \in (a, b)$ such that $g(b) - g(a) = g'(c)(b - a)$. Since $g'(c) \neq 0$, some algebra shows that

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

■

What mistake does this proof make?

(b) Give a correct proof of Cauchy's Mean Value Theorem. (Hint: Use the approach that the proof of the MVT uses, but modify the h function to use $g(x)$ instead of x in certain places.)

19. (a) Suppose that f is differentiable on an open interval (a, b) and that f is continuous on $[a, b]$. Prove that if $f'(x)$ approaches a limit L as $x \to a^+$, then

$$\lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} = L$$

(Hint: Use the MVT and the Squeeze Theorem.)

(b) Come up with an analogous left-handed version of part (a), and prove it.

(c) Use parts (a) and (b) to prove the following: if $a \in \mathbb{R}$ and f is a real function which is continuous around a and differentiable near a, and if $f'(x)$ has a limit as $x \to a$, then $f'(a)$ exists and is that limit.

20. Show that for all distinct $a, b \in [1, 2]$, $|\sqrt{b} - \sqrt{a}| < |b - a|$. In other words, the square root function is a contraction on $[1, 2]$ (see Example 4.46).

21. Suppose that $f : [a, b] \to [a, b]$ is a contraction, so for all $x, y \in [a, b]$ distinct, $|f(y) - f(x)| < |y - x|$. Show that there cannot exist two distinct values $L_1, L_2 \in [a, b]$ such that $f(L_1) = L_1$ and $f(L_2) = L_2$. (A value of x satisfying $f(x) = x$ is said to be a *fixed point* of f, so this exercise shows contractions have at most one fixed point.)

22. Prove the First Derivative Test (Theorem 4.49). (Hint: Take any $c, d \in I$ so that $c < a < d$, and consider how f rises or falls on the intervals $[c, a]$ and $[a, d]$.)

For the next three exercises, a real function is given. Use the techniques used in Example 4.48 to draw a graph of the function, making sure to indicate where the function is rising or falling as well as labelling any relative extrema.

23. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = (x - 1)^2(x + 2)$

24. $f : \mathbb{R} - \{0\} \to \mathbb{R}$, $f(x) = x + \dfrac{1}{x^2}$

25. $f : \mathbb{R} - \{-3, 3\} \to \mathbb{R}$, $f(x) = \dfrac{x^2 - 4}{x^2 - 9}$

For the rest of the exercises in this section, we'll discuss the use of the second derivative f''. The sign of f'' tells us whether f' is increasing or decreasing. When f' is increasing on an open interval (a, b), the graph of f rises faster, forming an upward-cup shape, and we say that f' is *concave up* on (a, b). Similarly, if f' is decreasing on (a, b), then the graph cups downwards, and we say f is *concave down*[9] on (a, b). Lastly, if f' has a relative extremum at a, so that the *concavity* of f changes at a, then we say that a is an *inflection point* of f.

For example, the function defined by $y = x^2$ satisfies $y'' = 1$, so it is always concave up, producing the distinctive "cup" shape that we associate with parabolas. In contrast, the cubic function $y = x(x-1)(x+1) = x^3 - x$ has $y'' = 6x$, so it is concave down when $x < 0$ and concave up when $x > 0$, with an inflection point at $x = 0$. The graph of $y = x^3 - x$ is shown in Figure 4.21.

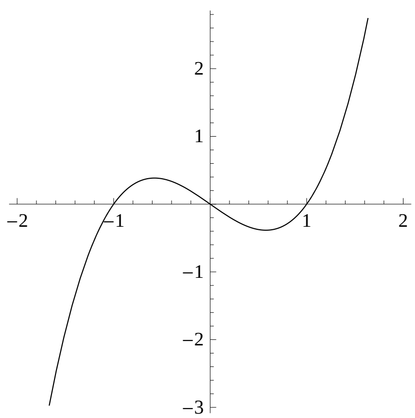

Figure 4.21: A graph of $y = x(x-1)(x+1) = x^3 - x$

26. Prove that if f is a real function which is twice differentiable on (a, b), and $f''(x) > 0$ for all $x \in (a, b)$, then f is concave up on (a, b). Similarly, prove that if $f''(x) < 0$ for all $x \in (a, b)$, then f is concave down on (a, b).

27. Prove this theorem, which sometimes is useful for determining relative extrema:

[9] Sometimes the term *convex* is used instead of "concave up", so that the term *concave* means "concave down".

Theorem 4.52 (Second Derivative Test). *Let $a \in \mathbb{R}$ be given, let an open interval I containing a be given, and let f be a real function which is twice differentiable on I. Suppose that $f'(a) = 0$.*

(a) If $f''(a) > 0$, then f has a relative minimum at a.

(b) If $f''(a) < 0$, then f has a relative maximum at a.

(Hint: Use the First Derivative Test.)

28. Go back to all the previous exercises in this section involving sketching graphs, and find out where the functions are concave up and concave down. Incorporate that information into your drawings.

29. Suppose that $p : \mathbb{R} \to \mathbb{R}$ is a cubic polynomial with three distinct real zeroes r_1, r_2, r_3. Show that p has an inflection point at the average of the three zeroes $(r_1 + r_2 + r_3)/3$. (Hint: Write $p(x)$ in factored form.)

30. Define $f(x) = \sqrt{x}$ for all $x > 0$. Thus, the tangent line at $a = 1$ is $L(x) = 1 + (1/2)(x - 1)$, and the error is $E(x) = f(x) - L(x)$. Use the First Derivative Test to find the absolute maximum of $E(x)$, and use this to prove that $f(x) \le L(x)$ for all $x > 0$.

    (Remark: This result can be generalized to show that when f is concave down, the tangent line lies above the graph.)

31. Suppose that f is continuous on $[a, b]$, twice-differentiable on (a, b), and $f''(x) > 0$ for all $x \in (a, b)$. Let L be the secant line from $(a, f(a))$ to $(b, f(b))$. Let E be the linear error, so $E(x) = f(x) - L(x)$ for all $x \in [a, b]$.

    (a) Where is $E(x)$ maximized? How does this relate to the MVT? (Hint: How many roots can E' have?)

    (b) Use part (a) to prove that $f(x) < L(x)$ for all $x \in (a, b)$.

    This proves that a concave-upwards graph lies below its secant line.

Chapter 5

Integrals

There are three major operations that are performed in calculus, and at this point we've seen two of them: limits and derivatives. We use limits to determine what values a function approaches, which leads to the notion of continuity. We use derivatives, which are based on limits, to determine how fast a function is changing, which lets us study velocities and relative extrema. In this chapter, we'll study the third operation, the integral, which determines the area that a function's graph encloses. Limits will also be useful for the definition of an integral, though we will mainly use the related notions of suprema and infima in our definition.

We will introduce the integral so that we can answer the following question: if $f : [a, b] \to \mathbb{R}$ is given, and $f(x) \geq 0$ for all $x \in [a, b]$, then how do we find the area between the graph of f and the x-axis on $[a, b]$? In this situation, we say that we are finding the area *under the graph of f on* $[a, b]$, and we will temporarily denote this quantity by $A(f; [a, b])$. When the graph of f is just a collection of line segments, then the region under the graph is a polygon, and the area becomes straightforward to calculate. On the other hand, when the graph f is irregularly curved, it's not clear how exactly how to measure $A(f; [a, b])$. We'll demonstrate our main strategy in an example.

Example 5.1:

Let's say that $f : [0, 1] \to \mathbb{R}$ is defined by $f(x) = x^2$ for all $x \in [0, 1]$. How would we determine $A(f; [0, 1])$, as displayed in the leftmost graph in Figure 5.1? To solve this problem, we will approximate $A(f; [0, 1])$ by using shapes with which we are familiar, like rectangles. The idea is that just like we computed derivatives by approximating the instantaneous slope with difference quotients and taking a limit, we will compute the area by approximating with collections of rectangles and taking a limit. In essence, we'll chop up the domain $[0,1]$ into pieces to try and approximate each piece of the graph of f with a rectangle.

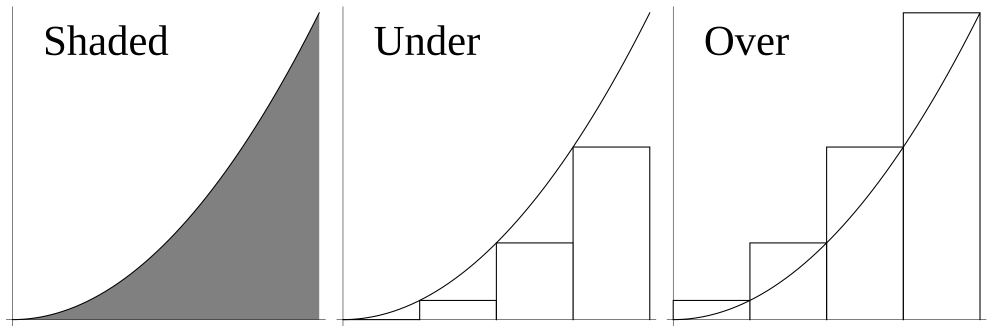

Figure 5.1: The area under the graph of $f(x) = x^2$ and two estimates

To see how this works, look at the middle and rightmost graphs in Figure 5.1. In the middle graph, we have approximated $A(f;[0,1])$ by using four rectangles, each having the same width of $1/4$ and having their upper-left corner on the graph of f. Since f is a strictly increasing function, the total area of these four rectangles is smaller than $A(f;[0,1])$, as the picture makes evident. In the rightmost graph, we use rectangles with the same width, but now their upper-right corners are on the graph of f. In this scenario, the total area of these four rectangles is larger than $A(f;[0,1])$.

In the middle picture, the heights of the four rectangles are $f(0)$, $f(1/4)$, $f(1/2)$, and $f(3/4)$ (using the fact that the rectangle's height is measured from its upper-left corner). In the rightmost picture, the heights of the four rectangles are $f(1/4)$, $f(1/2)$, $f(3/4)$, and $f(1)$. Therefore, using the formula $f(x) = x^2$, and the fact that each rectangle has width $1/4$, we have

$$\begin{aligned} & \frac{1}{4}\left(0^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{3}{4}\right)^2\right) \\ < \ & A(f;[0,1]) \\ < \ & \frac{1}{4}\left(\left(\frac{1}{4}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{3}{4}\right)^2 + 1^2\right) \end{aligned}$$

which simplifies to $7/32 < A(f;[0,1]) < 15/32$. This gives us the estimate that $A(f;[0,1]) \approx 11/32$ (the midpoint of the interval $(7/32, 15/32)$), with error at most $4/32$, or $1/8$.

Note that the maximum possible error of $1/8$ is a pretty large proportion of our estimate $11/32$. This means that our estimate, using approximations with four rectangles, might not be very precise. After all, as Figure 5.1 shows, you can see plenty of extra area covered by the rectangles in the bottom-left graph, as well as plenty of uncovered area in the bottom-right graph. This suggests that in order to get a better estimate, we should chop up the domain $[0,1]$ into more pieces, so we use thinner rectangles. Let's try using eight rectangles, each with width $1/8$, instead of 4 rectangles with width $1/4$, from which we get the picture in Figure 5.2.

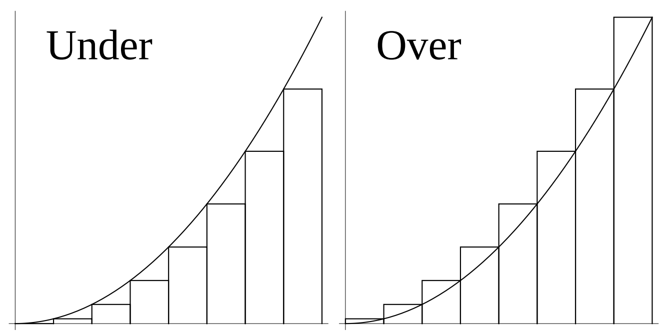

Figure 5.2: Approximating the area under the graph of $f(x) = x^2$ with eight rectangles

This picture suggests that using eight rectangles will yield a better approximation. Indeed, by computing the area under these rectangles in a similar manner, we find

$$\begin{aligned} & \frac{1}{8}\left(0^2 + \left(\frac{1}{8}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{3}{8}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{5}{8}\right)^2 + \left(\frac{3}{4}\right)^2 + \left(\frac{7}{8}\right)^2\right) \\ < \; & A(f;[0,1]) \\ < \; & \frac{1}{8}\left(\left(\frac{1}{8}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{3}{8}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{5}{8}\right)^2 + \left(\frac{3}{4}\right)^2 + \left(\frac{7}{8}\right)^2 + 1^2\right) \end{aligned}$$

which simplifies to $35/128 < A(f;[0,1]) < 51/128$. This gives us the new estimate of $A(f;[0,1]) \approx 43/128$, with maximum error at most $8/128$, or $1/16$. This error is twice as small as the error when we used four rectangles. In essence, our approximation using eight rectangles was twice as good.

You may suspect that if we use more rectangles, then the approximation will be even better. This raises the question: if we use an arbitrary number of rectangles, say n of them, then what estimate do we obtain as a function of n? With that in mind, for each $n \in \mathbb{N}^*$, let's use L_n to denote the underestimate obtained when we make n rectangles, each of width $1/n$, with their upper-left corners on the graph of f (L stands for "lower estimate"), and we'll use U_n to denote the overestimate from n equal-width rectangles with their upper-right corners on the graph of f (U stands for "upper estimate").

For each i from 1 to n, the i^{th} rectangle from the left in the overestimate has height $f(i/n)$, and the i^{th} rectangle in the underestimate has height $f((i-1)/n)$. Since each rectangle has width $1/n$, we obtain the following:

$$L_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i-1}{n}\right)^2 < A(f;[0,1]) < U_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n}\right)^2$$

To simplify this, we may pull $1/n^2$ out of each sum, and we can also throw out the $i = 1$ term from the left sum (because when $i = 1$, $((i-1)/n)^2$ is zero). After this, we also reindex the left sum to start from 1, giving us

$$\frac{1}{n^3}\sum_{i=1}^{n-1} i^2 < A(f;[0,1]) < \frac{1}{n^3}\sum_{i=1}^{n} i^2$$

In order to find out what U_n and L_n approach as n grows large, we should find a better way to write the sum of the first n squares. In Exercise 1.9.1.(a), you were asked to prove

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

by induction on $n \in \mathbb{N}^*$. Using this formula, our inequalities now become

$$\frac{(n-1)(n)(2n-1)}{6n^3} < A(f;[0,1]) < \frac{(n)(n+1)(2n+1)}{6n^3}$$

By splitting the three powers of n in the denominator between the three factors, we can rewrite this as

$$\frac{1}{6}\left(1-\frac{1}{n}\right)(1)\left(2-\frac{1}{n}\right) < A(f;[0,1]) < \frac{1}{6}(1)\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right)$$

As n grows arbitrarily large, $1/n$ goes to zero, so both the lower and upper bound approach the value $1/3$. By the Squeeze Theorem, we must have $A(f; [0, 1]) = 1/3$. $\square$

Example 5.1 suggests a strategy for finding the area $A(f; [a, b])$. We make approximations to the area, both from below and from above, with collections of rectangles. We choose the approximations so that we can find upper and lower estimates as close as we'd like to each other, i.e. they approach a common limit. That value of the limit is the area under the function. We will call that area *the integral of f from a to b*. In this chapter, our aim is to develop a formal definition which explains this process, as well as to prove useful properties of the integral, culminating with the *Fundamental Theorem of Calculus* which explains the relationship between derivatives and integrals.

5.1 Step Functions

First, we'd like to describe more specifically what kinds of collections of rectangles we will use when approximating area. We are interested in rectangles whose base is on the x-axis. The bases should not overlap (though they may touch at their boundaries), and the entire interval $[a, b]$ should be covered. One convenient way to describe where the bases of the rectangles lie is with the following definition:

Definition 5.2. Let $a, b \in \mathbb{R}$ be given with $a \le b$. A *partition* of $[a, b]$ is a finite subset of $[a, b]$ which includes a and b. For convenience, when we write a partition P as $\{x_0, x_1, \ldots, x_n\}$ for some $n \in \mathbb{N}$, we implicitly assume that

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

unless otherwise specified.

In this notation, we say that x_1 through x_{n-1} are the *subdivision points* of the partition, and x_0 and x_n are the endpoints. (The members of P are called *partition points*.) For each $i \in \mathbb{N}$ from 1 to n, the interval (x_{i-1}, x_i) is called the i^{th} *open subinterval of P*, and $[x_{i-1}, x_i]$ is called the i^{th} *closed subinterval of P*. (We also say that P partitions $[a, b]$ *into n subintervals*.) The *width* of the i^{th} subinterval is $x_i - x_{i-1}$, sometimes denoted as Δx_i (i.e. as the i^{th} change in x of the partition).

With this definition, we can say that we want the bases of our rectangles to be the closed subintervals of a partition on $[a, b]$. In Example 5.1, when n rectangles are used, the bases of the rectangles correspond to the partition $\{0, 1/n, 2/n, \ldots, 1\}$. In these partitions, every subinterval has the same width, but in general, we might want to allow the subintervals to have different widths. That way, if one particular part of the graph needs to be subdivided into thinner rectangles, we can use thin rectangles in that one part without requiring the rest of the rectangles to be as thin.

This next definition helps us describe the heights of our rectangles:

Definition 5.3. Let $a, b \in \mathbb{R}$ be given with $a \leq b$, and suppose that $P = \{x_0, x_1, \ldots, x_n\}$ is a partition of $[a, b]$. We say that a real function s is a *step function with respect to the partition* P if s is defined on $[a, b] - P$ and, for each $i \in \mathbb{N}$ from 1 to n, s is constant on the open subinterval (x_{i-1}, x_i). In other words, there is some $a_i \in \mathbb{R}$ such that for every $x \in (x_{i-1}, x_i)$, we have $s(x) = a_i$.

In Example 5.1 with n subintervals, our underestimate and overestimate each correspond to step functions with respect to $\{0, 1/n, 2/n, \ldots, 1\}$. For the underestimate, we use the step function $s_n : [0, 1] \to \mathbb{R}$ whose value on the i^{th} subinterval is $((i-1)/n)^2$ for each i from 1 to n. For the overestimate, we use the step function $t_n : [0, 1] \to \mathbb{R}$ whose value on the i^{th} subinterval is $(i/n)^2$ for each i. Thus, the area of the underestimate is the area under the graph of s_n, i.e. $A(s_n; [0, 1])$, and the area of the overestimate is $A(t_n; [0, 1])$.

Remark. In general, whenever a function is changed at a single point, the region under the function's graph either gains or loses a vertical line segment. Line segments have zero width, so they have zero area. Applying this reasoning inductively, we see that if we change a function at FINITELY many points, then the area does not change.

This is why our definition of step function does not require the function to be defined at the partition points: the area under the graph is unaffected by the values at the partition points, so those points have little relevance. (This is also why partitions are required to be FINITE subsets of $[a, b]$.)

In our definition of step function, we allow a step function to take negative values. When a step function is negative on an open subinterval, our convention is to give the area of the corresponding rectangle a negative sign. Note the following example:

Example 5.4:

If $s : [0, 10] \to \mathbb{R}$ is the step function defined by

$$\forall x \in [0,1]\ s(x) = \begin{cases} 1 & \text{if } x \in [0,3) \\ -2 & \text{if } x \in [3,7) \\ 3 & \text{if } x \in [7,10] \end{cases}$$

pictured in Figure 5.3, then s is a step function with respect to $\{0, 3, 7, 10\}$, with three subintervals, and the total area under the step function is

$$(1)(3-0) + (-2)(7-3) + (3)(10-7) = 4$$

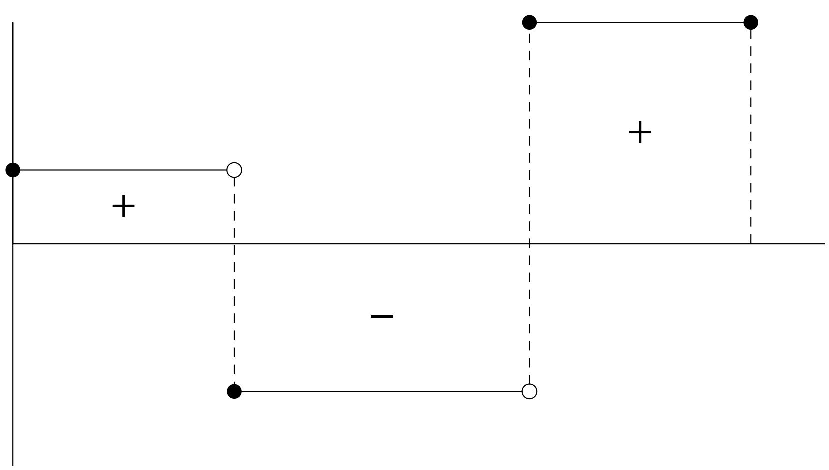

Figure 5.3: A step function s on $[0, 1]$ with positive and negative areas marked

This quantity we computed, which counts rectangles below the x-axis as having negative area, could be called "*signed area*" of s, as opposed to "*unsigned area*" of s which counts all rectangles as having positive area. (The "unsigned area" is $1(3)+2(4)+3(3) = 20$.) Note that the unsigned area of s is just the signed area of $|s|$, the absolute value of s. Thus, in general, we will focus on computing signed areas. $\square$

Rather than use the term "signed area", we instead use the following special name:

Definition 5.5. Let $a, b \in \mathbb{R}$ be given with $a \leq b$, and suppose that $P = \{x_0, x_1, \ldots, x_n\}$ is a partition of $[a, b]$ and s is a step function with respect to

P. The *integral of s from a to b (with respect to P)*, also called the *integral of s over $[a, b]$ with respect to P*, is

$$\sum_{i=1}^{n} a_i(x_i - x_{i-1}) = \sum_{i=1}^{n} a_i \cdot \Delta x_i$$

where a_i is the value taken by s on the open subinterval (x_{i-1}, x_i), and Δx_i is the width of the i^{th} subinterval. The function s is called the *integrand*, the endpoint a is called the *lower limit of integration*, and the endpoint b is called the *upper limit of integration*.

For the time being, we will write the integral as

$$\int_P s \qquad \text{or} \qquad \int_P s(x)\,dx$$

and we read this second notation as "the integral of $s(x)\,dx$ from a to b (with respect to P)". The "dx" is a symbol we include to tell us that we are using the variable x as an input variable to s. The variable x is a dummy variable, and we can just as easily write

$$\int_P s(t)\,dt \qquad \text{or} \qquad \int_P s(z)\,dz$$

or use any other variable in place of x, so long as that variable is not currently being used somewhere else.

Remark. The symbol used to denote the integral is a stretched letter "S", standing for "summation". The notation is intended to look similar to the summation that is used to define the integral of a step function. For instance, the dx symbol is analogous to the width Δx_i of the i^{th} subinterval. Once we define the integral more generally later in this chapter, and when we study techniques of integration in Chapter 6, the dx symbol will be more useful.

With this notation, when s is the function defined in Example 5.4, we showed that $\displaystyle\int_P s(x)\,dx = 4$ where $P = \{0, 3, 7, 10\}$. In Example 5.1, if we let s_n and t_n refer to respectively the underestimating and overestimating step functions, and we let P_n be the partition $\{0, 1/n, 2/n, \ldots, 1\}$, then we showed that

$$\int_{P_n} s_n(x)\,dx = L_n = \frac{(n-1)(n)(2n-1)}{6n^3}$$

and

$$\int_{P_n} t_n(x)\,dx = U_n = \frac{(n)(n+1)(2n+1)}{6n^3}$$

Remark. Note that the definition of the integral only uses the values a_i that the step function takes on open subintervals; the values $s(x_i)$ of the step function at partition points do not appear in the definition. This makes sense, since as we've remarked earlier, the values $s(x_0)$, $s(x_1)$, etc. through $s(x_n)$ do not affect the area under the graph of s.

Example 5.6:

The floor function, defined in Theorem 2.37, is constant on any interval of the form $(z, z+1)$ where $z \in \mathbb{Z}$. Thus, for any $a, b \in \mathbb{R}$ with $a \le b$, the floor function is a step function on $[a, b]$.[1] For example, let's say that f is the floor function restricted to $[0, \pi]$. Then f is a step function with respect to the partition $P = \{0, 1, 2, 3, \pi\}$, and

$$\begin{aligned}\int_P f &= \int_P \lfloor x \rfloor\,dx = (0)(1-0) + (1)(2-1) + (2)(3-2) + (3)(\pi - 3)\\ &= 3\pi - 6\end{aligned}$$

Similarly, suppose $g : [0, \pi] \to \mathbb{R}$ is defined by $g(x) = \lfloor x + 0.5 \rfloor$ for any $x \in [0, \pi]$ (so g is a shifted version of the floor function, restricted to $[0, \pi]$). g is also a step function, but not with respect to P. (We will also say that g is *not compatible* with P.) Instead, g is a step function with respect to $Q = \{0, 0.5, 1.5, 2.5, \pi\}$, and we have

$$\begin{aligned}\int_Q g &= \int_Q \lfloor x + 0.5 \rfloor\,dx\\ &= (0)(0.5-0) + (1)(1.5-0.5) + (2)(2.5-1.5) + (3)(\pi - 2.5)\\ &= 3\pi - 4.5\end{aligned}$$

If we consider the sum $f + g$, then $f + g$ is a step function, but it's not compatible with P or with Q. Instead, $f + g$ rises by 1 every half a unit, and we see that $f + g$ is compatible with the partition $P \cup Q =$

[1] Sometimes we extend our definitions to say that a function is a *step function on* $\mathbb{R}$ when it is a step function on all closed bounded intervals.

$\{0, 0.5, 1, 1.5, 2, 2.5, 3, \pi\}$. Since every subinterval of this partition has width 0.5 except the last one, which has width $\pi - 3$, we find that

$$\begin{aligned}
\int_{P \cup Q} (f+g) &= \int_{P \cup Q} \lfloor x \rfloor + \lfloor x + 0.5 \rfloor \, dx \\
&= (0+1+2+3+4+5)(0.5) + (6)(\pi - 3) \\
&= 6\pi - 10.5
\end{aligned}$$

Note that $6\pi - 10.5$ is the sum of $3\pi - 6$ and $3\pi - 4.5$, so we found that the integral of the sum of our two step functions equals the sum of their integrals. We'll prove this property in general soon. $\square$

How Much Does The Partition Matter?

As Example 5.6 indicates, it seems that the integral of a sum of step functions should be the sum of the corresponding integrals. This is a very plausible property for the integral to have, because we designed the integral to measure area, and when we add two functions, the areas under the functions should also be added. However, there is a cumbersome technical concern: to make this property precise, and to prove it, our definition of the integral requires that we specify partitions for each step function.

This raises the question: if a step function is compatible with multiple partitions, then does the choice of partition affect the integral? After all, the choice of partition corresponding to a step function is not unique. For instance, in Example 5.4, the step function s is compatible with $P = \{0, 0.3, 0.7, 1\}$, but it is also compatible with $Q = \{0, 0.2, 0.3, 0.5, 0.7, 0.9, 1\}$. This is because each open subinterval of Q is contained in an open subinterval of P, and s is constant on each open subinterval of P. In fact, the same reasoning shows that s is compatible with any partition that contains $\{0, 0.3, 0.7, 1\}$ as a subset.

This leads to the following definition:

Definition 5.7. Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and suppose P and P_2 are partitions of $[a, b]$. We say that P_2 *refines* P, or *is a refinement of* P, if $P \subseteq P_2$. Sometimes we also say that P_2 is *finer* than P, or P is *coarser* than P_2. (Think of P as chopping $[a, b]$ into pieces, and then P_2 chops up the pieces of P even more.)

If P_2 is finer than P, then every subinterval of P_2 is a subset of a subinterval of P. Therefore, if f is a step function which is compatible with P, then f is also compatible with P_2. In particular, if f and g are step functions, where f is compatible with P and g is compatible with Q, then $f + g$ is a step function compatible with $P \cup Q$ because $P \cup Q$ refines both P and Q. We call $P \cup Q$ the *least common refinement* of P and Q.

As our main lemma for analyzing how the choice of partition affects an integral, we have the following property:

Lemma 5.8. *Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let s be a step function with respect to some partition P of $[a, b]$. If $c \in [a, b] - P$, then*

$$\int_P s = \int_{P \cup \{c\}} s$$

In other words, the value of an integral does not change if we refine the partition by adding one more point.

Strategy. Let's say that $P = \{x_0, x_1, \ldots, x_n\}$, where $a = x_0 < x_1 < \cdots < x_n = b$. Because c is in $[a, b]$ but not in P, there must be some k from 1 to n such that $c \in (x_{k-1}, x_k)$, i.e. c belongs to the k^{th} open subinterval of P. Let's say that $s(x) = a_k$ for all $x \in (x_{k-1}, x_k)$.

If we compare the summations used in the definitions of our two integrals, then they are the same for every subinterval of P except for the k^{th} subinterval. In the integral with respect to P, the sum uses the term $a_k(x_k - x_{k-1})$, and in the integral with respect to $P \cup \{c\}$, the sum uses $a_k(c - x_{k-1})$ and $a_k(x_k - c)$. (In other words, in the refinement, (x_{k-1}, x_k) is broken into the two pieces (x_{k-1}, c) and (c, x_k), each of which creates a term in the sum.) However, since

$$a_k(c - x_{k-1}) + a_k(x_k - c) = a_k(x_k - x_{k-1})$$

both integrals compute the same value.

Proof. Let a, b, s, P, c be given as described. The strategy described is a valid proof of this lemma. However, for the readers who feel that the strategy is too handwavy about specifying where the two summations differ, we provide a more pedantic approach to the proof here.

Let's write P in the form $\{x_0, x_1, \ldots, x_n\}$ for some $n \in \mathbb{N}$, where

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

Because $c \in [a,b] - P$, there exists a unique $k \in \{1, 2, \ldots, n\}$ such that $c \in (x_{k-1}, x_k)$. (This is because if you consider the set of all i from 1 to n such that $x_i > c$, then this set is finite and nonempty, so it has a minimum element k.) We write $P \cup \{c\}$ as $\{y_0, y_1, \ldots, y_{n+1}\}$ by defining y_j for each j from 1 to $n+1$ as follows:

$$y_j = \begin{cases} x_j & \text{if } j < k \\ c & \text{if } j = k \\ x_{j-1} & \text{if } j > k \end{cases}$$

Therefore, we have $a = y_0 < y_1 < \cdots < y_{n+1} = b$, where $y_k = c$.

For each i from 1 to n, let a_i be the value that s takes on (x_{i-1}, x_i). For each j from 1 to $n+1$, let b_j be the value that s takes on (y_{j-1}, y_j). When $j < k$, we have $b_j = a_j$, and when $j > k+1$, $b_j = a_{j-1}$. Also, we have $b_k = b_{k+1} = a_k$ because c is in (x_{k-1}, x_k). By the definition of the integral of a step function, we have

$$\begin{aligned} \int_{P\,\cup\,\{c\}} s(y)\,dy &= \sum_{j=0}^{n+1} b_j(y_j - y_{j-1}) \\ = & \sum_{j=0}^{k-1} b_j(y_j - y_{j-1}) + b_k(y_k - y_{k-1}) \\ & + b_{k+1}(y_{k+1} - y_k) + \sum_{j=k+2}^{n+1} b_j(y_j - y_{j-1}) \end{aligned}$$

Now, we change the variable used in the summations to i instead of j (this is purely for cosmetic reasons, but it helps us avoid getting the two partitions confused), we replace the b_j and y_j values with the corresponding a_i and x_i values (as we defined earlier), and we also shift the last summation down by one. This gives us

$$\begin{aligned} & \sum_{i=0}^{k-1} a_i(x_i - x_{i-1}) + a_k(c - x_{k-1}) \\ & + a_k(x_k - c) + \sum_{i=k+1}^{n} a_i(x_i - x_{i-1}) \\ = & \sum_{i=0}^{n} a_i(x_i - x_{i-1}) = \int_P s(x)\,dx \end{aligned}$$

as desired. ■

From this lemma, you can prove the following corollary as Exercise 5.2.7:

Corollary 5.9. *Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let P and P_2 be partitions of $[a, b]$ such that P_2 refines P. If s is a step function with respect to P (so that it is also a step function with respect to P_2), then*

$$\int_P s = \int_{P_2} s$$

Remark. From this corollary, it follows that if s is a step function which is compatible with two partitions P and Q of $[a, b]$, then

$$\int_P s = \int_{P \cup Q} s = \int_Q s$$

because $P \cup Q$ refines both P and Q. Therefore, the integral of a step function s on $[a, b]$ does NOT depend on which compatible partition of $[a, b]$ is used to evaluate the integral. Because of this, the limits of integration, a and b, are usually written in the integral notation instead of the partition itself, and we will write

$$\int_a^b s \qquad \text{or} \qquad \int_a^b s(x)\,dx$$

instead of writing $\int_P s$ as before.[2]

Properties of Step Function Integrals

Now that we have shown that the value of a step function integral doesn't depend on which compatible partition is used, we are able to prove some useful results related to integrals of step functions. First, as Example 5.6 suggests, we'd like to prove that the integral of a sum of step functions equals the sum of the integrals. In fact, we'll prove the stronger property that integration is a linear operation, just like the operations of limit and derivative:

[2]In fact, most books on calculus don't introduce any notation like $\int_P$, preferring instead to make a short remark about how the partition "doesn't matter", but we wanted to give this subtle point about partitions more careful attention.

Theorem 5.10 (Linearity property). *Let $a, b, c, d \in \mathbb{R}$ with $a \leq b$ be given, and let s, t be step functions on $[a, b]$. Then*

$$\int_a^b (cs + dt) = c \int_a^b s + d \int_a^b t$$

Strategy. Normally, it is tricky to add two step functions because they may have been defined with respect to different partitions. However, because of the remark after Corollary 5.9, we may use one partition for both functions (the common refinement). After that, the rest is straightforward calculation.

Proof. Let a, b, c, d, s, t be given as described. We may suppose that s and t are both compatible with a partition $P = \{x_0, x_1, \dots, x_n\}$ of $[a, b]$. Let's say that for each i from 1 to n, s takes the constant value a_i on the i^{th} subinterval of P, and t takes the constant value b_i on the i^{th} subinterval. Therefore, $cs + dt$ takes the constant value $ca_i + db_i$ on the i^{th} subinterval of P, and we compute

$$\begin{aligned}
\int_a^b (cs + dt) &= \sum_{i=1}^{n} (ca_i + db_i)(x_i - x_{i-1}) \\
&= c \sum_{i=1}^{n} a_i(x_i - x_{i-1}) + d \sum_{i=1}^{n} b_i(x_i - x_{i-1}) \\
&= c \int_a^b s + d \int_a^b t
\end{aligned}$$

as desired. ■

As another useful property of integrals, you can prove in Exercise 5.2.9 that a larger step function has more area underneath it:

Theorem 5.11 (Comparison property). *Let $a, b \in \mathbb{R}$ be given satisfying $a \leq b$, and let s, t be step functions on $[a, b]$. If $s(x) \leq t(x)$ for all $x \in [a, b]$, then*

$$\int_a^b s \leq \int_a^b t$$

(We say that integrals preserve inequalities*.) In fact, this is also true for strict inequalities: if $a < b$ and $s(x) < t(x)$ for all $x \in [a, b]$, then*

$$\int_a^b s < \int_a^b t$$

Another useful property of the integral is motivated by Figure 5.4. If we take an interval $[a, b]$ and split it into two pieces at the point c, then this splits the area into two pieces whose sum is the total area under $[a, b]$. More formally, you can prove this property in Exercise 5.2.10:

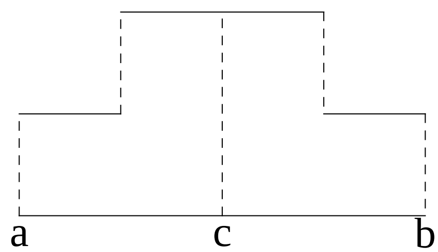

Figure 5.4: Demonstration of $\int_a^b s = \int_a^c s + \int_c^b s$

Theorem 5.12 (Interval addition property). *Let $a, b, c \in \mathbb{R}$ be given satisfying $a \le c \le b$, and let s be a step function on $[a, b]$. Then*

$$\int_a^b s = \int_a^c s + \int_c^b s$$

Remark. One subtlety we have not addressed yet, but which is useful now, is what happens when you integrate with the same upper and lower limit. When you apply the definition of integral to compute $\int_a^a s$, you have only one possible partition of $[a, a]$ ($[a, a]$ is the same thing as $\{a\}$), and that partition has a single interval of width zero. Thus,

$$\int_a^a s = 0$$

This result is also obtained by plugging in $c = a$ into Theorem 5.12.

Furthermore, Theorem 5.12 suggests a way to define $\int_a^b s$ when $a > b$, i.e. when the ends are "in the wrong order". If we ignore the restriction on the ordering of a, b, and c in the theorem, and we replace b with a and c with b, we obtain

$$0 = \int_a^a s = \int_a^b s + \int_b^a s$$

Because of this, we DEFINE $\displaystyle\int_a^b s$ to be $\displaystyle -\int_b^a s$. In essence, when we write the endpoints in the other order, we flip the sign of the integral, as we're "measuring the area backwards." Furthermore, with this definition, you can prove as part of Exercise 5.2.11 that the formula

$$\int_a^b s = \int_a^c s + \int_c^b s$$

holds for ALL choices of $a, b, c \in \mathbb{R}$ so long as s is a step function on $[\min\{a,b,c\}, \max\{a,b,c\}]$ (so that all three integrals are defined). In fact, all of the integral properties in this section, except for the comparison property, are valid when the ends are "in the wrong order", as you can verify.

One last property we'll present here deals with the area under a step function reflected horizontally. In Figure 5.5, we see a step function s on $[a, b]$ as well as the graph of its horizontal reflection which takes each $x \in [-b, -a]$ to $s(-x)$. It seems quite plausible that the two graphs have the same area underneath, and you can prove this in Exercise 5.2.12:

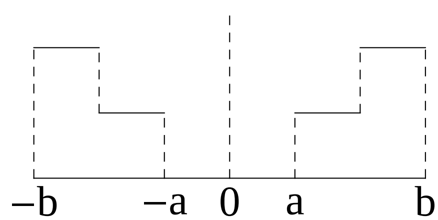

Figure 5.5: A step function and its horizontal reflection

Theorem 5.13 (Horizontal reflection property). *Let $a, b \in \mathbb{R}$ be given with $a \le b$, and let s be a step function on $[a, b]$. If $t : [-b, -a] \to \mathbb{R}$ is defined by $t(x) = s(-x)$ for all $x \in [-b, -a]$, then t is a step function on $[-b, -a]$ and*

$$\int_a^b s = \int_{-b}^{-a} t \qquad i.e. \qquad \int_a^b s(x)\,dx = \int_{-b}^{-a} s(-x)\,dx$$

There are a few other properties of the integral of a step function which you can prove in the exercises.

5.2 Exercises

1. Using the same approach as Example 5.1, find the area under the graph of $y = x^2$ from 0 to b for any $b > 0$.

2. Using the same approach as Example 5.1, find the area under the graph of $y = x^3$ from 0 to b for any $b > 0$. You may use the following identity without proof (although its proof is quite simple and is done by induction): for any $n \in \mathbb{N}^*$,

$$\sum_{i=1}^{n} i^3 = \left(\frac{n(n+1)}{2}\right)^2$$

3. In each part of this exercise, a formula defining a step function s is given, and an interval $[a, b]$ is given. For each, give a partition of $[a, b]$ with which s is compatible.
   (a) $s(x) = \lfloor x + \pi \rfloor$, $[a, b] = [0, \pi]$
   (b) $s(x) = \lfloor x^2 \rfloor$, $[a, b] = [0, 2]$
   (c) $s(x) = \lfloor x^2 + 2x \rfloor$, $[a, b] = [-3, 1]$
   (d) $s(x) = \lfloor x \rfloor + \lfloor 3x \rfloor$, $[a, b] = [-1, 1]$

4. For each of the functions in the previous exercise, compute its integral over the specified interval $[a, b]$.

5. Prove by induction that for all $n \in \mathbb{N}^*$,

$$\int_0^{n^2} \lfloor \sqrt{x} \rfloor \, dx = \frac{n(n-1)(4n+1)}{6}$$

6. Prove that for all $n \in \mathbb{N}^*$ and all $x \in \mathbb{R}$,

$$\lfloor nx \rfloor = \sum_{i=0}^{n-1} \left\lfloor x + \frac{i}{n} \right\rfloor$$

   (Hint: Consider the fractional part of x, namely $x - \lfloor x \rfloor$. By studying this fractional part, determine which terms of the sum are one larger than the other terms.)

7. Using Lemma 5.8, prove Corollary 5.9. (Hint: Use induction on the size of $P_2 - P$.)

8. Use Corollary 5.9 to prove that if s and t are two step functions on $[a, b]$ which differ at only finitely many inputs, then $\int_a^b s = \int_a^b t$.

9. Prove Theorem 5.11.

10. Prove Theorem 5.12. (Hint: If s is compatible with a partition P of $[a, c]$ and with a partition Q of $[c, b]$, then what is a partition of $[a, b]$ with which s is compatible?)

11. Recall that in the remark following Theorem 5.12, we defined the integral from a to b of a step function when $a > b$.

    (a) Use Theorem 5.12 to prove that for any $a, b, c \in \mathbb{R}$, if s is a step function on $[\min\{a, b, c\}, \max\{a, b, c\}]$, then

    $$\int_a^b s = \int_a^c s + \int_c^b s$$

    (b) Give a counterexample to show that Theorem 5.11 does not hold if $a > b$, i.e. find values $a, b \in \mathbb{R}$ with $a > b$, as well as step functions $s, t : [b, a] \to \mathbb{R}$ such that $s(x) < t(x)$ for all $x \in [b, a]$ but $\int_a^b s \geq \int_a^b t$.

    (c) Based on your counterexample from part (b), modify Theorem 5.11 so that it is true for any $a, b \in \mathbb{R}$, and prove your modified version.

12. Prove Theorem 5.13. (Hint: If s is compatible with a partition P of $[a, b]$, then what is a partition of $[-b, -a]$ with which t is compatible?)

13. This theorem says that the area under a step function does not change if you shift the function horizontally:

    Theorem 5.14 (Shifting property). *Let $a, b, c \in \mathbb{R}$ with $a \leq b$ be given, and let s be a step function on $[a, b]$. If $t : [a - c, b - c] \to \mathbb{R}$*

is defined by $t(x) = s(x + c)$ *for all* $x \in [a - c, b - c]$, *then* t *is a step function on* $[a - c, b - c]$ *and*

$$\int_a^b s = \int_{a-c}^{b-c} t \qquad \textit{i.e.} \qquad \int_a^b s(x)\, dx = \int_{a-c}^{b-c} s(x + c)\, dx$$

Prove this theorem.

14. This theorem says that if you stretch a function horizontally by a factor of k, the area grows by a factor of k:

    Theorem 5.15 (Stretching property). *Let* $a, b, k \in \mathbb{R}$ *be given with* $a \le b$ *and* $k > 0$ *be given, and let* s *be a step function on* $[a, b]$. *If* $t : [ka, kb] \to \mathbb{R}$ *is defined by* $t(x) = s(x/k)$ *for all* $x \in [ka, kb]$, *then* t *is a step function on* $[ka, kb]$ *and*

    $$\int_a^b s = \frac{1}{k}\int_{ka}^{kb} t \qquad \textit{i.e.} \qquad \int_a^b s(x)\, dx = \frac{1}{k}\int_{ka}^{kb} s\left(\frac{x}{k}\right) dx$$

    (a) Prove this theorem.
    (b) State an analogous version of this theorem for $k < 0$, and prove it. (Hint: Theorem 5.13 might be useful.)

15. Suppose that we defined the integral of a step function s on $[a, b]$ by instead using the formula

    $$\int_a^b s = \sum_{i=1}^{n} a_i^2 (x_i - x_{i-1})$$

    You may take for granted that this formula gives the same value for any partition $\{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ with which s is compatible. Using this altered definition, which theorems from Theorem 5.10 to Theorem 5.15 are true? Prove the true theorems, and provide counterexamples for the false theorems.

16. Do the same as the previous exercise, except use the definition

    $$\int_a^b s = \sum_{i=1}^{n} a_i (x_i^3 - x_{i-1}^3)$$

5.3 The General Definition of Integral

Now that we have introduced the integral of a step function, which tells us how much area lies underneath the graph of a step function, we'd like to make a definition that works for other real functions. Let's suppose that we have a real function f defined on $[a,b]$, and we'd like to formally define $\int_a^b f$ to measure what $A(f;[a,b])$ should be. (More precisely, we'd like to measure the "signed area" if f takes negative values.) We'd like our definition to exhibit the properties of area that we proved as theorems in the last section, such as the comparison property, the linearity property, the interval addition property, and the horizontal reflection property. This raises the question: what do those properties tell us about how to define the integral?

Upper And Lower Integrals

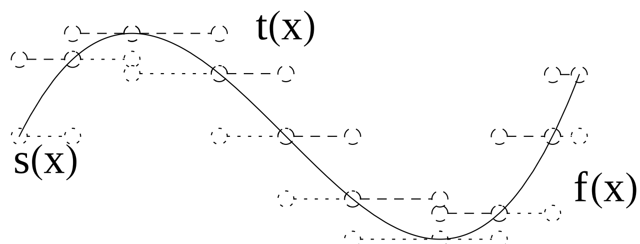

Figure 5.6: A function f between two step functions s (dotted) and t (dashed)

We'll use the strategy outlined at the beginning of this chapter in Example 5.1. Consider the picture in Figure 5.6. In this picture, we have drawn a real function f on $[a,b]$, as well as two step functions s and t. Here, $s(x) \le f(x) \le t(x)$ for all $x \in [a,b]$, which we describe by saying that s is a *lower step function* for f on $[a,b]$ and that t is an *upper step function* for f on $[a,b]$. Therefore, the comparison property indicates that

$$\int_a^b s \le \int_a^b f \le \int_a^b t$$

should be true. In other words, the step function s gives us an underestimate of the area, and the step function t gives us an overestimate of the area.

We can apply this reasoning with ANY lower and upper step functions. More precisely, if we write S to denote the set of lower step functions for f on $[a, b]$, and we write T to denote the set of upper step functions for f on $[a, b]$, then for all $s \in S$ and all $t \in T$,

$$\int_a^b s \le \int_a^b f \le \int_a^b t$$

Which choices of s and t give us the best possible estimates? We obtain the best estimate when $\displaystyle\int_a^b s$ is as large as possible and when $\displaystyle\int_a^b t$ is as small as possible. However, there might not exist a choice of s which yields a maximum value of $\displaystyle\int_a^b s$, nor might there exist a choice of t which yields a minimum value of $\displaystyle\int_a^b t$. In other words, if we write

$$L = \left\{ \int_a^b s \mid s \in S \right\} \qquad U = \left\{ \int_a^b t \mid t \in T \right\}$$

so that L is the set of underestimates from lower step functions, and U is the set of overestimates from upper step functions, then we're not guaranteed that $\max L$ or $\min U$ exist. After all, if f isn't a step function, then we would not expect a perfect estimate by a step function when measuring area, and we'd also expect that every estimate by a step function can be made better by using a finer partition.

Instead of using $\max L$ and $\min U$ for our best estimates, what's our next best option? We can try using least upper bound and greatest lower bound, i.e. we can use $\sup L$ and $\inf U$. One advantage of using suprema and infima is that in Chapter 2, we introduced simple conditions for determining if a set has a supremum or an infimum. Let's see if L and U satisfy these conditions.

First, we need L and U to be nonempty sets. In other words, we need f to have a lower step function and an upper step function, i.e. we want S and T to be nonempty. This may sound trivial, but there is a very important situation where L or U can be empty: when f is unbounded. Since a step function can only take finitely many values in its range, any step function must be a bounded function. Therefore, in order for f to have lower and upper step functions, f must be bounded. Conversely, if f is bounded, say there is some $M > 0$ so that $|f(x)| \le M$ for all $x \in [a, b]$, then the constant function with value M is an upper step function for f, and similarly the constant function with value $-M$ is a lower step function for f, showing that L and U are nonempty.

Second, let's suppose f is bounded, and let's find out if L has an upper bound and if U has a lower bound. We've shown above that

$$\int_a^b s \leq \int_a^b t$$

for all $s \in S$ and all $t \in T$. This means that every member of L is less than or equal to every member of U. In particular, since we know that L and U are nonempty, each member of L is a lower bound for U, and also each member of U is an upper bound for L.

Therefore, $\sup L$ and $\inf U$ exist. In fact, the inequality above implies, via Theorem 2.33, that $\sup L \leq \inf U$. We summarize the discussion with the following definitions:

Definition 5.16. Let $a, b \in \mathbb{R}$ satisfying $a \leq b$ be given, and let f be a real function whose domain contains $[a, b]$. When f is bounded on $[a, b]$, we define

$$L(f; [a,b]) = \sup \left\{ \int_a^b s \;\middle|\; \begin{array}{l} s \text{ is a step function on } [a,b], \\ \forall x \in [a,b]\ s(x) \leq f(x) \end{array} \right\}$$

to be the *lower integral of f on* $[a, b]$, and we define

$$U(f; [a,b]) = \inf \left\{ \int_a^b t \;\middle|\; \begin{array}{l} t \text{ is a step function on } [a,b], \\ \forall x \in [a,b]\ t(x) \geq f(x) \end{array} \right\}$$

to be the *upper integral of f on* $[a, b]$.

Thus, $L(f; [a, b])$ is, in some sense, the best lower estimate of the area that we can obtain from step functions, and $U(f; [a, b])$ is the best upper estimate of the area. Ideally, these two estimates are the same, which means that the area under the curve HAS to be that common value. In that case, we use their common value as the definition of our integral:

Definition 5.17. Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let f be a real function with $\text{dom}(f) \supseteq [a, b]$, and suppose f is bounded on $[a, b]$. We say that f *is integrable on* $[a, b]$ if $L(f; [a, b]) = U(f; [a, b])$, in which case we define *the integral of f on* $[a, b]$ as

$$\int_a^b f = L(f; [a,b]) = U(f; [a,b])$$

We also write

$$\int_a^b f(x)\,dx$$

to represent the integral of f on $[a,b]$, much like we did for integrals of step functions. If $L(f;[a,b]) \neq U(f;[a,b])$, we say that f is *not integrable on* $[a,b]$.

(Note that if f is a step function, then f is both a lower and an upper step function for itself, so this definition of integral yields the same result as the previous definition for step functions.)

Remark. The integral which we have defined is often called the *Riemann integral*. It requires that the function f be bounded on a closed bounded interval and that its lower and upper integral be equal. There are more general definitions of integration which allow us to integrate more functions, such as *Lebesgue integration*, but those topics are more complicated and are often covered in graduate courses. We will revisit this topic when we study improper integration in Chapter 9.

Calculating An Integral By Definition

The definition of the integral is rather hard to use, because it involves showing that a supremum of a set of integrals of step functions is equal to an infimum of another set of integrals of step functions. To make our job more manageable, we'd like a way to find these suprema and infima without analyzing every single lower and upper step function. One approach is to make a sequence of lower and upper step functions and take a limit, as the following examples show:

Example 5.18:

Let's return to the process we used at the beginning of this chapter in Example 5.1. In that example, we were trying to find the area under the function $f:[0,1]\to\mathbb{R}$ defined by $f(x)=x^2$ for each $x\in[0,1]$. To do this, for each $n\in\mathbb{N}^*$ we made a lower step function $s_n:[0,1]\to\mathbb{R}$ and an upper step function $t_n:[0,1]\to\mathbb{R}$ by breaking the interval $[0,1]$ into n pieces of width $1/n$ and picking appropriate heights for the step functions. When we did this, we found

$$\int_0^1 s_n = \frac{(n-1)n(2n-1)}{6n^3} = \frac{1}{6}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right)$$

and
$$\int_0^1 t_n = \frac{n(n+1)(2n+1)}{6n^3} = \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right)$$
Therefore, since $L(f;[0,1])$ is, by definition, an upper bound on the integrals of lower step functions, and $U(f;[0,1])$ is a lower bound on the integrals of upper step functions, we have

$$\frac{1}{6}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right) \le L(f;[0,1]) \le U(f;[0,1]) \le \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right)$$

Note that as n grows arbitrarily large, the outer parts of this inequality both approach the limit of $1/3$. Therefore, by the Squeeze Theorem, we must have $L(f;[0,1]) = U(f;[0,1]) = 1/3$, which proves that $\int_0^1 x^2\,dx = 1/3$.

This illustrates that we don't have to look at all lower or upper step functions to find the values of $L(f;[0,1])$ and $U(f;[0,1])$; we just need to look at enough step functions so as to "squeeze" the values of $L(f;[0,1])$ and $U(f;[0,1])$ together. □

Example 5.19:

Suppose that $b \in \mathbb{R}$ with $b > 0$ is given, and let's define $f : [0,b] \to \mathbb{R}$ by $f(x) = x$ for each $x \in [0,b]$. Let's use the definition of integral to find

$$\int_0^b f = \int_0^b x\,dx$$

Since the region under f is a triangle with the base length and height equal to b, we should expect that the answer is the area of that triangle, $b^2/2$.

Just as in the previous example, for each $n \in \mathbb{N}^*$ we'll make a lower step function s_n and an upper step function t_n which are compatible with the partition of $[0,b]$ broken into n equal-width pieces. The open subintervals of this partition have the form $((i-1)b/n, ib/n)$ where i is an integer from 1 to n. On this subinterval, the minimum value that f takes is $(i-1)b/n$ (i.e. the value at the left endpoint) and the maximum value that f takes is ib/n (i.e. the value at the right endpoint), so therefore for all x in this interval, we define
$$s_n(x) = \frac{(i-1)b}{n} \qquad t_n(x) = \frac{ib}{n}$$

It doesn't matter what we assign for values of s_n and t_n at the partition points, so long as s_n stays a lower step function for f and t_n stays an upper step function for f.

Since each subinterval has width b/n, we use the definition of the integral of a step function, as well as the formula for the sum of the first n positive integers, to obtain

$$\int_0^1 t_n(x)\,dx = \sum_{i=1}^{n}\left(\frac{ib}{n}\right)\left(\frac{b}{n}\right) = \frac{b^2}{n^2}\sum_{i=1}^{n} i = \frac{b^2}{n^2}\cdot\frac{n(n+1)}{2} = \frac{b^2}{2}(1)\left(1+\frac{1}{n}\right)$$

A similar computation yields

$$\int_0^1 s_n(x)\,dx = \frac{b^2}{2}\left(1-\frac{1}{n}\right)(1)$$

Therefore,

$$\frac{b^2}{2}\left(1-\frac{1}{n}\right) \le L(f;[0,b]) \le U(f;[0,b]) \le \frac{b^2}{2}\left(1+\frac{1}{n}\right)$$

As n grows arbitrarily large, $L(f;[0,b])$ and $U(f;[0,b])$ are squeezed to equal $b^2/2$, as expected. □

These examples suggest that there is a more "limit-like" definition of integral we can use to compute integrals. The following theorem provides such a definition, which imitates the ϵ-δ style of the definition of limit:

Theorem 5.20. *Let $a, b, I \in \mathbb{R}$ with $a \le b$ be given, and let f be a real function defined and bounded on $[a,b]$. Then f is integrable on $[a,b]$ with $\int_a^b f = I$ iff for every $\epsilon > 0$, there exists a lower step function s and an upper step function t for f on $[a,b]$ such that*

$$I - \epsilon < \int_a^b s \le \int_a^b t < I + \epsilon$$

Strategy. Our goal is to prove is an "iff" statement, so we have to provide two directions of proof. For the first direction, suppose that f is integrable. In that case, we know that $L(f;[a,b]) = U(f;[a,b])$. Since $L(f;[a,b])$ is the least upper bound of the integrals of lower step functions, there exist lower step functions s whose integrals are arbitrarily close to $L(f;[a,b])$. A similar argument works for $U(f;[a,b])$ to produce an upper step function t.

For the other direction, we get to assume that we can "squeeze" upper and lower step function integrals between $I-\epsilon$ and $I+\epsilon$ for every ϵ. Since $L(f;[a,b])$ is no smaller than any lower step function integral, and $U(f;[a,b])$ is no larger than any upper step function integral, these values are also squeezed between $I-\epsilon$ and $I+\epsilon$. We can just let ϵ approach 0 to show $L(f;[a,b]) = U(f;[a,b])$.

Proof. Let a, b, I, f be given as described. For convenience, let's say that L is the set of integrals of lower step functions for f on $[a,b]$ and that U is the set of integrals of upper step functions for f on $[a,b]$. Thus, $L(f;[a,b]) = \sup L$ and $U(f;[a,b]) = \inf U$.

For the first direction of the proof, assume that $\int_a^b f = I$. Therefore, by definition of the integral, we have $L(f;[a,b]) = U(f;[a,b]) = I$. Let $\epsilon > 0$ be given. By definition of $L(f;[a,b])$, $L(f;[a,b]) - \epsilon$ is smaller than $\sup L$, so there must exist some value in L greater than $L(f;[a,b]) - \epsilon$. In particular, there is some lower step function s for f on $[a,b]$ such that

$$L(f;[a,b]) - \epsilon < \int_a^b s \le L(f;[a,b])$$

Similarly, since $U(f;[a,b]) + \epsilon$ is greater than $\inf U$, there exists a value in U less than $U(f;[a,b]) + \epsilon$, so there is some upper step function t for f on $[a,b]$ such that

$$U(f;[a,b]) \le \int_a^b t < U(f;[a,b]) + \epsilon$$

Since $L(f;[a,b]) = U(f;[a,b]) = I$ by assumption, we put this all together and obtain

$$I - \epsilon < \int_a^b s \le I \le \int_a^b t < I + \epsilon$$

as desired.

For the other direction of the proof, suppose that for each $\epsilon > 0$, we can find a lower step function s and an upper step function t satisfying

$$I - \epsilon < \int_a^b s \le \int_a^b t < I + \epsilon$$

By the definitions of $L(f;[a,b])$ and $U(f;[a,b])$, we also know that

$$\int_a^b s \leq L(f;[a,b]) \leq U(f;[a,b]) \leq \int_a^b t$$

Putting this all together, we have

$$I - \epsilon < L(f;[a,b]) \leq U(f;[a,b]) < I + \epsilon$$

Furthermore, $L(f;[a,b])$ and $U(f;[a,b])$ do not depend on ϵ (whereas s and t do). Thus, we may take the limit as $\epsilon \to 0$ of all sides of the inequality to obtain

$$I \leq L(f;[a,b]) \leq U(f;[a,b]) \leq I$$

(note that when taking limits, inequalities may not be strict anymore). Thus, $L(f;[a,b]) = U(f;[a,b]) = I$, proving that f is integrable with integral I. ■

A small variant on this theorem, which resembles the tactics we've used so far, can be proven as Exercise 5.4.1:

Corollary 5.21. *Let $a, b, I \in \mathbb{R}$ with $a \leq b$ be given, and let f be a real function defined and bounded on $[a,b]$. Then f is integrable on $[a,b]$ with $\int_a^b f = I$ iff there exists a sequence $s_1, s_2, \ldots$ of lower step functions for f on $[a,b]$ and a sequence $t_1, t_2, \ldots$ of upper step functions for f on $[a,b]$ such that*

$$\lim_{n\to\infty} \int_a^b s_n = \lim_{n\to\infty} \int_a^b t_n = I$$

In other words, when n is a sufficiently large integer, the integrals of s_n and t_n are as close to I as desired.

One last remark is appropriate in this section about computing integrals by definition. When trying to write a proof, it can be useful to know whether a function is integrable, without knowing what the value of the integral is. The basic idea is that for an integrable function, the lower and upper step functions can be chosen so that their integrals are arbitrarily close to each other. The following "limit-like" theorem makes this more formal, and you can prove it in Exercise 5.4.2:

Theorem 5.22. *Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let f be a real function defined and bounded on $[a, b]$. Then f is integrable on $[a, b]$ iff for all $\epsilon > 0$, there exists a lower step function s for f on $[a, b]$ and an upper step function t for f on $[a, b]$ such that*

$$\int_a^b (t - s) = \int_a^b t - \int_a^b s < \epsilon$$

(Note that since $s(x) \leq t(x)$ on $[a, b]$, the integral of $(t - s)$ must be nonnegative.)

To illustrate what Theorem 5.22 is saying, we take the picture from Figure 5.6 and shade the region between the upper and lower step functions, as you can see in Figure 5.7. The shaded region, whose area is the integral of $t - s$, represents how "far apart" s and t are in measuring the area of f. Thus, Theorem 5.22 says that this "error area" between the upper and lower estimates can be made arbitrarily small for an integrable function.

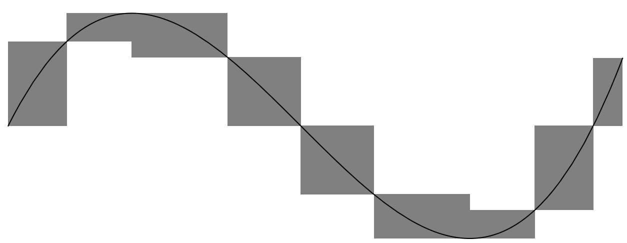

Figure 5.7: $\int_a^b (t - s)$ shaded

Revisiting Properties Of The Integral

Since the general definition of the integral is, in essence, a limit of integrals of step functions, it should be no surprise that the properties of integrals that we proved for step functions still hold. This also makes sense, because the properties we proved for step function integrals are properties that we expect the area under a graph to have. You can prove most of these theorems in the exercises, though we will present the proof of the interval addition property to demonstrate how we put our most recent theorems about computing integrals to good use.

Theorem 5.23 (Linearity). *Let $a, b, c, d \in \mathbb{R}$ with $a \le b$ be given, and let f, g be two real functions which are both integrable on $[a, b]$. Then $cf + dg$ is also integrable on $[a, b]$, with*

$$\int_a^b (cf + dg) = c\int_a^b f + d\int_a^b g$$

Remark. The linearity property shows that the sum or difference of integrable functions on an interval $[a, b]$ produces an integrable function on $[a, b]$. It is also true that the product of two integrable functions is integrable, but it is NOT the case in general that

$$\int_a^b (fg) = \left(\int_a^b f\right)\left(\int_a^b g\right)$$

(For example, when $a = 0$, $b = 1$, and $f(x) = g(x) = x$ for all $x \in [0, 1]$, then $\displaystyle\int_0^1 x^2\,dx = 1/3$ and $\displaystyle\left(\int_0^1 x\,dx\right)^2 = 1/4$.) It is also true that if an integrable function has a positive lower bound for its range, then the reciprocal of the function is also integrable. You can prove these results in the exercises in Section 5.8.

Theorem 5.24 (Comparison). *Let $a, b \in \mathbb{R}$ with $a \le b$ be given, and let f, g be two real functions which are both integrable on $[a, b]$. If $f(x) \le g(x)$ for all $x \in [a, b]$, then*

$$\int_a^b f \le \int_a^b g$$

Remark. Suppose we consider a "strict-inequality version" of the comparison property, where we assume that $f(x) < g(x)$ for all $x \in [a, b]$ (where $a < b$) and conclude that $\displaystyle\int_a^b f < \int_a^b g$. This is much harder to prove than for step functions, because inequalities that are strict for lower and upper step functions might no longer remain strict when limits are taken. In general, the process of taking limits can destroy strict inequalities: for instance, for every $n \in \mathbb{N}^*$, we know that $-1/n$ is strictly smaller than $1/n$, but as n grows arbitrarily large, $-1/n$ and $1/n$ both approach 0.

In the exercises, you will be able to prove that this strict-inequality version is true when $g - f$ has a point of continuity in $[a, b]$. It turns out to be true that any integrable function must have a point of continuity, but the author is not aware of any short proof of this which uses the material we've covered so far. In many applications, when we want to make use of the strict-inequality version, our integrands f and g will be continuous, so we will not have to prove the strict-inequality version in complete generality.

Theorem 5.25 (Interval Addition). *Let $a, b, c \in \mathbb{R}$ be given with $a \leq c \leq b$, and let f be a real function. If f is integrable on $[a, c]$ and also integrable on $[c, b]$, then f is integrable on $[a, b]$, and*

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

Remark. As with the integral of step functions, Theorem 5.25 motivates the following definitions for "out-of-order" limits of integration when f is integrable on $[\min\{a, b\}, \max\{a, b\}]$:

$$\int_a^a f = 0 \qquad \int_a^b f = -\int_b^a f \text{ when } a > b$$

Also, all the properties in this section, except for the comparison property, are still true if we drop the hypotheses that say the limits of integration have to be in "the correct order". This is proven in the exact same way it was proven for integrals of step functions.

Strategy. Since we'd like to prove that f is integrable on $[a, b]$, and furthermore we know which value the integral should equal, Theorem 5.20 works quite well. Let $\epsilon > 0$ be given. We'd like to make a lower step function s and an upper step function t whose integrals over $[a, b]$ are less than ϵ away from $\displaystyle\int_a^c f + \int_c^b f$.

Since f is integrable on $[a, c]$, we can use Theorem 5.20 to find lower and upper step functions for f on $[a, c]$, say s_1 and t_1, whose integrals are as close to $\displaystyle\int_a^c f$ as we want. We can do the same thing with f on $[c, b]$ to get another lower and upper step function, say s_2 and t_2. Thus, in some sense, we've found step functions which approximate f quite well on $[a, c]$, as well as some other step functions which approximate f well on $[c, b]$.

How do we make step functions which approximate f well on $[a, b]$? We stick our other functions together! More formally, we can make $s(x) = s_1(x)$

when $x < c$ and $s(x) = s_2(x)$ when $x > c$, so that s is a lower step function for f on $[a, b]$. (It doesn't matter what we pick for the value of $s(c)$.) Similarly, we make an upper step function t.

Since s is built out of s_1 and s_2, it makes sense to use the interval addition property for integrals of step functions to analyze $\int_a^b s$ to see how close it is to $\int_a^c f + \int_c^b f$. After that, we have some simple calculations using the Triangle Inequality. The same steps work for the upper step function as well.

Proof. Let a, b, c, f be given as described. Because f is integrable on $[a, c]$ and on $[c, b]$, f is bounded on $[a, c]$ and $[c, b]$, so f is also bounded on $[a, b]$. Therefore, if we let

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

we may use Theorem 5.20 to prove that f is integrable on $[a, b]$ with $\int_a^b f = I$. Hence, let $\epsilon > 0$ be given; we wish to find a lower step functions s for f on $[a, b]$ and an upper step function t for f on $[a, b]$ so that

$$I - \epsilon < \int_a^b s \leq \int_a^b t < I + \epsilon$$

First, because f is integrable on $[a, c]$, by Theorem 5.20, we may choose a lower step function s_1 for f on $[a, c]$ and an upper step function t_1 for f on $[a, c]$ such that

$$\int_a^c f - \frac{\epsilon}{2} < \int_a^c s_1 \leq \int_a^c t_1 < \int_a^c f + \frac{\epsilon}{2}$$

Similarly, we may choose a lower step function s_2 for f on $[c, b]$ and an upper step function t_2 for f on $[c, b]$ such that

$$\int_c^b f - \frac{\epsilon}{2} < \int_c^b s_2 \leq \int_c^b t_2 < \int_c^b f + \frac{\epsilon}{2}$$

Now, we define functions $s, t : [a, b] \to \mathbb{R}$ for all $x \in [a, b]$ as follows:

$$s(x) = \begin{cases} s_1(x) & \text{if } x \in [a, c) \\ f(c) & \text{if } x = c \\ s_2(x) & \text{if } x \in (c, b] \end{cases} \qquad t(x) = \begin{cases} t_1(x) & \text{if } x \in [a, c) \\ f(c) & \text{if } x = c \\ t_2(x) & \text{if } x \in (c, b] \end{cases}$$

You may check that s is a lower step function for f on $[a,b]$ and t is an upper step function for f on $[a,b]$.

By adding our inequalities together concerning s_1, s_2, t_1, and t_2, we get

$$\begin{aligned} &\int_a^c f + \int_c^b f - \left(\frac{\epsilon}{2} + \frac{\epsilon}{2}\right) < \int_a^c s_1 + \int_c^b s_2 \\ \leq\ &\int_a^c t_1 + \int_c^b t_2 < \int_a^c f + \int_c^b f + \left(\frac{\epsilon}{2} + \frac{\epsilon}{2}\right) \end{aligned}$$

Since $s(x) = s_1(x)$ for all $x \in (a,c)$, we have $\displaystyle\int_a^c s = \int_a^c s_1$. Similar statements hold for the interval (c,b) and for the upper step functions. Thus, by the interval addition property for integrals of step functions, and our choice of I, we see that

$$I - \epsilon < \int_a^b s \leq \int_a^b t < I + \epsilon$$

as desired. ■

Theorem 5.26 (Horizontal reflection). *Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let f be a real function integrable on $[a,b]$. If we define $g : [-b,-a] \to \mathbb{R}$ by $g(x) = f(-x)$ for all $x \in [-b,-a]$, then g is integrable on $[-b,-a]$, and we have*

$$\int_a^b f = \int_{-b}^{-a} g \qquad i.e. \qquad \int_a^b f(x)\,dx = \int_{-b}^{-a} f(-x)\,dx$$

5.4 Exercises

1. Prove Corollary 5.21. (Hint: For one direction of the proof, to find s_n and t_n, apply Theorem 5.20 with a choice of ϵ depending on n.)

2. Prove Theorem 5.22. (Hint: It will help to show that for any lower step function s and upper step function t, we have $0 \leq U(f;[a,b]) - L(f;[a,b]) \leq \displaystyle\int_a^b t - \int_a^b s$.)

3. This theorem is sometimes helpful when we want to cut an interval up into pieces:

Theorem 5.27 (Subinterval property). *Let $a, b, c, d \in \mathbb{R}$ be given such that $a \le c \le d \le b$. Thus, $[c, d] \subseteq [a, b]$. If f is a real function which is integrable on $[a, b]$, then f is also integrable on $[c, d]$.*

Use Theorem 5.22 to prove this theorem. (Hint: Suppose that the step functions s and t approximate f well on $[a, b]$. How would you make step functions that approximate f well on $[c, d]$? You'll also want to use the interval addition and comparison properties for step function integrals.)

4. Consider the function $f : [0, 1] \to \mathbb{R}$ defined as follows for any $x \in [0, 1]$:

$$f(x) = \begin{cases} 1 & \text{if } 1/n < x \le 1/(n-1), \text{ where } n \in \mathbb{N}^* \text{ and } n \text{ is even} \\ 0 & \text{if } 1/n < x \le 1/(n-1), \text{ where } n \in \mathbb{N}^* \text{ and } n \text{ is odd} \\ 0 & \text{if } x = 0 \end{cases}$$

Thus, f is almost like a step function, but instead of being constant on finitely many subintervals, f is constant on each of the intervals $(1/2, 1]$, $(1/3, 1/2]$, and so on. In some sense, f is a "step function with an infinite partition", and $f(x)$ oscillates rapidly between 0 and 1 as x approaches 0. A rough graph of f, with some parts left out and replaced with ellipses, is in Figure 5.8.

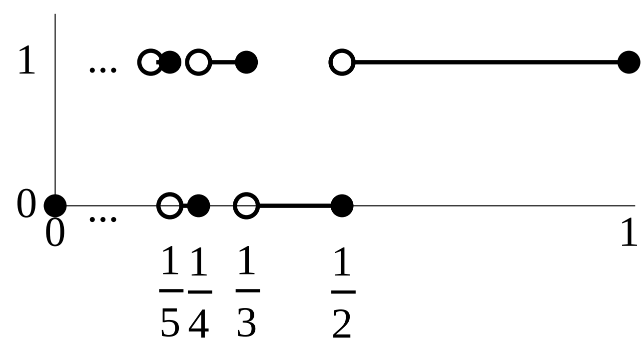

Figure 5.8: A partial picture of f from Exercise 5.4.4

Prove, using Theorem 5.22, that f is integrable on $[0, 1]$. (The value of its integral ends up being rather interesting! However, that value doesn't make the proof easier.)

5. Prove the linearity property, Theorem 5.23, by proving these three special cases:
   (a) First, prove the special case where $c = d = 1$, so that you have to show
   $$\int_a^b (f+g) = \int_a^b f + \int_a^b g$$
   (b) Next, prove the special case where $d = 0$ and $c \geq 0$, so that you have to show
   $$\int_a^b (cf) = c\int_a^b f$$
   (c) Last, prove the special case where $d = 0$ and $c < 0$. The linearity property follows from applying these three cases.

6. Prove the comparison property, Theorem 5.24. (Hint: You might find Exercise 2.7.8 useful.)

7. Prove the horizontal reflection property, Theorem 5.26.

8. Let $b \in \mathbb{R}$ be given with $b \geq 0$, and let f be a real function which is integrable on $[-b, b]$. Using the horizontal reflection and interval addition properties, prove the following:
   (a) If f is an odd function, meaning that $f(-x) = -f(x)$ for all $x \in \mathrm{dom}(f)$, then $\displaystyle\int_{-b}^b f = 0$.
   (b) If f is an even function, meaning that $f(-x) = f(x)$ for all $x \in \mathrm{dom}(f)$, then $\displaystyle\int_{-b}^b f = 2\int_0^b f$.

   Also, explain why these results are plausible geometrically.

9. Prove that if $a, b \in \mathbb{R}$ are given with $a \leq b$, and f is a real function such that f and $|f|$ (i.e. the absolute-value function composed with f) are both integrable on $[a, b]$, then
   $$\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx$$
   Note that when f is a step function, this result is equivalent to the Triangle Inequality, so this result can be considered more general than the Triangle Inequality. (Hint: You might want to use Lemma 3.2.)

Note: In Exercise 5.6.5, you'll see that the integrability of f implies the integrability of $|f|$.

10. For any $a, b \in \mathbb{R}$, use the result of Example 5.19 to prove that

$$\int_a^b x\,dx = \frac{b^2}{2} - \frac{a^2}{2}$$

Also, explain why this is plausible geometrically. (Hint: First handle the special case when $a = 0$ and $b < 0$ using the horizontal reflection property.)

11. In Theorem 5.14, the shifting property for integrals of step functions, we showed that shifting a step function horizontally does not change the area underneath it.

    (a) Prove the shifting property for arbitrary integrable functions:

    Theorem 5.28 (Shifting). *Let $a, b, c \in \mathbb{R}$ with $a \le b$ be given, and suppose that f is a real function which is integrable on $[a, b]$. If we define $g : [a + c, b + c] \to \mathbb{R}$ by $g(x) = f(x - c)$ for all $x \in [a + c, b + c]$, then g is integrable on $[a + c, b + c]$, and*

$$\int_a^b f = \int_{a+c}^{b+c} g \qquad i.e. \qquad \int_a^b f(x)\,dx = \int_{a+c}^{b+c} f(x - c)\,dx$$

    (b) Let's say that $f : \mathbb{R} \to \mathbb{R}$ is given and $p > 0$ is given such that f is *periodic with period* p, meaning that for all $x \in \mathbb{R}$, $f(x + p) = f(x)$. Also assume that f is integrable on $[0, p]$. Using the shifting property and the subinterval property (from Theorem 5.27), prove that for any $a \in \mathbb{R}$, f is integrable on $[a, a + p]$ and

$$\int_a^{a+p} f = \int_0^p f$$

    Thus, when a function is periodic, you obtain the same integral when you integrate over any periodic interval.

12. In Theorem 5.15, the stretching property, we showed that stretching a step function by a factor of k horizontally makes the area multiply by a factor of k.

(a) Prove the stretching property for arbitrary integrable functions:

Theorem 5.29 (Stretching). *Let $a, b, k \in \mathbb{R}$ be given with $a \le b$ and $k > 0$, and suppose that f is a real function integrable on $[a, b]$. If we define $g : [ka, kb] \to \mathbb{R}$ by $g(x) = f(x/k)$ for all $x \in [ka, kb]$, then g is integrable on $[ka, kb]$ and*

$$\int_a^b f = \frac{1}{k}\int_{ka}^{kb} g \qquad \textit{i.e.} \qquad \int_a^b f(x)\,dx = \frac{1}{k}\int_{ka}^{kb} f\left(\frac{x}{k}\right)\,dx$$

(b) State a version of the stretching property for $k < 0$, and prove it.

(c) Use properties of the integral, including the stretching property, along with the result $\displaystyle\int_0^1 x^2\,dx = \frac{1}{3}$ from Example 5.18, and prove that for all $a, b \in \mathbb{R}$,

$$\int_a^b x^2\,dx = \frac{b^3}{3} - \frac{a^3}{3}$$

(d) Let $r > 0$ be given. Assuming that all functions involved in this problem are integrable, prove that

$$\int_{-r}^{r} \sqrt{r^2 - x^2}\,dx = r^2 \int_{-1}^{1} \sqrt{1 - x^2}\,dx$$

Also answer the following question: in terms of areas under functions, what is this result saying geometrically?

13. This exercise will outline a proof of a strict-inequality version of the comparison property for continuous functions. Let $a, b \in \mathbb{R}$ with $a < b$ be given.

(a) Suppose that h is a real function which is integrable on $[a, b]$. Also assume that $h(x) > 0$ for all $x \in [a, b]$. If there exists some $c \in [a, b]$ such that h is continuous at c, then $\displaystyle\int_a^b h > 0$. (Hint: First, show that if x is close enough to c, say $|x - c| < \delta$ for some $\delta > 0$, then $h(x) \ge h(c)/2$. Consider integrating h on $[c-\delta, c+\delta] \cap [a, b]$.)

(b) Now, suppose that f and g are real functions which are integrable on $[a, b]$. If f and g are continuous on $[a, b]$ and $f(x) < g(x)$ for all $x \in [a, b]$, then use the previous part to prove that $\displaystyle\int_a^b f < \int_a^b g$.

5.5 Some Unusual Functions III

Let's return to some of the examples we have seen in Some Unusual Functions and Some Unusual Functions II, but now we'd like to discover the intervals over which these functions are integrable. We continue to study these examples because they present good exercises in using the formal definition of the integral, as well as keeping us open-minded about the kinds of functions we can study. We'll also end this section with a very useful theorem about integrating monotone functions, which will allow us to compute some integrals more easily.

Characteristic Functions

Recall that if $A \subseteq \mathbb{R}$, then the function χ_A, the characteristic function of A, gives the value 1 on inputs that belong to A and gives 0 on inputs that do not belong to A. Some characteristic functions are quite simple, such as $\chi_{[0,1]}$, which is a step function on any interval $[a, b]$. In fact, step functions are nothing more than linear combinations of characteristic functions of intervals: if $s : [a, b] \to \mathbb{R}$ is a step function compatible with the partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$, with the property that for each i from 1 to n, s takes the value a_i on the i^{th} open subinterval of P, then you can check that

$$s = \sum_{i=1}^{n} a_i \chi_{(x_{i-1},x_i)} + \sum_{i=0}^{n} s(x_i)\chi_{[x_i,x_i]}$$

(the second part of the sum handles the values at partition points).

These are not the only types of characteristic functions which are integrable on an interval $[a, b]$. For example, in Exercise 5.4.4, you proved that the characteristic function

$$\chi_{(1/2,1] \cup (1/4,1/3] \cup (1/6,1/5] \cup \cdots}$$

is integrable on $[0, 1]$. Intuitively, although this characteristic function is not a step function, it is very close to being a step function. However, some characteristic functions bounce between 0 and 1 so wildly that we cannot approximate them well by step functions:

Theorem 5.30. *The characteristic function of the rationals, $\chi_{\mathbb{Q}}$, is not integrable on $[a, b]$ for any $a, b \in \mathbb{R}$ with $a < b$.*

Strategy. To show that $\chi_{\mathbb{Q}}$ is not integrable on $[a, b]$, we need to show that $L(\chi_{\mathbb{Q}}, [a, b]) \neq U(\chi_{\mathbb{Q}}, [a, b])$. This means that the lower step functions for $\chi_{\mathbb{Q}}$ cannot produce integrals arbitrarily close to the integrals for the upper step functions for $\chi_{\mathbb{Q}}$.

Suppose that s is a lower step function for $\chi_{\mathbb{Q}}$ on $[a, b]$, and that t is an upper step function for $\chi_{\mathbb{Q}}$ on $[a, b]$. How large can the values of s be, and how small can the values of t be? In any subinterval, there are rational points and there are also irrational points, because both $\mathbb{Q}$ and $\mathbb{R} - \mathbb{Q}$ are dense. Thus, $\chi_{\mathbb{Q}}$ takes each of the values 0 and 1 in that subinterval. This means that the lower step function must take values no larger than 0, and the upper step function must take values no smaller than 1. This tells us that

$$\int_a^b s \leq \int_a^b 0 = 0 \qquad \text{and} \qquad \int_a^b t \geq \int_a^b 1 = b - a$$

Proof. Let a, b be given as described. Let s be an arbitrary lower step function for $\chi_{\mathbb{Q}}$ on $[a, b]$, and let t be an arbitrary upper step function for $\chi_{\mathbb{Q}}$ on $[a, b]$. Suppose that s and t are both compatible with the partition $\{x_0, x_1, \ldots, x_n\}$ of $[a, b]$.

For any i from 1 to n, let's say that s takes the value a_i on (x_{i-1}, x_i), and t takes the value b_i on (x_{i-1}, x_i). This subinterval contains both rational and irrational points. Therefore, on (x_{i-1}, x_i), $\chi_{\mathbb{Q}}$ takes the values 0 and 1. Thus, we must have $a_i \leq 0$ and $b_i \geq 1$, meaning that

$$\int_a^b s = \sum_{i=1}^n a_i(x_i - x_{i-1}) \leq \sum_{i=1}^n 0(x_i - x_{i-1}) = 0$$

and

$$\int_a^b t = \sum_{i=1}^n b_i(x_i - x_{i-1}) \geq \sum_{i=1}^n 1(x_i - x_{i-1}) = b - a$$

Since this is true for all choices of s and t, this means that $L(\chi_{\mathbb{Q}}; [a, b]) \leq 0$ and $U(\chi_{\mathbb{Q}}; [a, b]) \geq b - a$, so we must have $L(\chi_{\mathbb{Q}}; [a, b]) \neq U(\chi_{\mathbb{Q}}; [a, b])$. (In fact, you can easily show that the lower integral is 0, and the upper integral is $b - a$, but we do not need those specific values.) ■

This theorem gives us our first major example of a non-integrable function. Note that $\chi_{\mathbb{Q}}$ is discontinuous everywhere, so what happens if we multiply $\chi_{\mathbb{Q}}$ by a continuous function, which gives the product some continuity points? We'll present an example here, and you can handle this question more generally in Exercise 5.6.3.

Example 5.31:

Let's suppose that $f : [0,1] \to \mathbb{R}$ is defined by $f(x) = x\chi_{\mathbb{Q}}(x)$ for all $x \in [0,1]$. We have seen in Chapter 3 that f is continuous at 0 but not continuous anywhere else. Let's address the question of whether f is integrable on $[0,1]$.

We'd like to use the strategy of the proof of Theorem 5.30, but we need to modify the argument because f is not constant at rational inputs. Intuitively, since f jumps between the values 0 and x, we'd like to say that lower and upper step functions have to respectively take values below 0 and above x. Let's show this more formally. Suppose we have a lower step function s and an upper step function t, compatible with the partition $\{x_0, x_1, \ldots, x_n\}$ of $[0,1]$, and let's look at the i^{th} subinterval (x_{i-1}, x_i), where i is from 1 to n.

On this subinterval, the range of f consists of 0 and the rational numbers between x_{i-1} and x_i. Thus, 0 is the infimum of the range on this interval, and x_i is the supremum of the range on this interval. Thus, s must take a value not exceeding 0, and t must take a value at least as large as x_i. Since $x_i \geq x$ for all $x \in (x_{i-1}, x_i)$, this shows that at any $x \in [0,1]$ which is not a partition point, $t(x) \geq x$ and $s(x) \leq 0$. Thus,

$$\int_0^1 s \leq 0 \qquad \text{and} \qquad \int_0^1 t \geq \int_0^1 x\,dx = \frac{1}{2}$$

This shows that $L(f; [0,1]) \leq 0$ and $U(f; [0,1]) \geq 1/2$, so f is not integrable on $[0,1]$. $\square$

The Ruler Function

Although the ruler function takes different values at rational and irrational inputs, the ruler function is continuous at all irrational numbers. In fact, at every $a \in \mathbb{R}$, we found that $r(x) \to 0$ as $x \to a$. The proof of this that we gave uses the fact that no matter which $\epsilon > 0$ is specified, only finitely many inputs near a cause the ruler function to attain a value larger than ϵ. This same tactic leads to the following result:

Theorem 5.32. *For all $a, b \in \mathbb{R}$ with $a < b$, the ruler function r is integrable on $[a,b]$, and*

$$\int_a^b r = 0$$

Strategy. Theorem 5.20 is a good tool to use: for each $\epsilon > 0$, we need to find a lower step function s for r and an upper step function t for r whose integrals on $[a, b]$ are less than ϵ away from 0. For the lower step function, this is easy: the function which is constantly zero suffices. However, how do we pick an upper step function whose integral is less than ϵ?

The simplest idea is to pick t to be nearly a constant function. If t has the value $\epsilon/(b-a)$ at all points except for finitely many partition points, then its integral over $[a, b]$ will have the value ϵ. This raises the question: which values of x in $[a, b]$ satisfy $r(x) \geq \epsilon/(b-a)$? These values must be rational and have the lowest-terms form p/q where $1/q \geq \epsilon/(b-a)$, so that $q \leq (b-a)/\epsilon$. This means that only finitely many values of x in $[a, b]$ satisfy $r(x) \geq \epsilon/(b-a)$, and the value of an integral doesn't change if you alter finitely many values of a function.

Proof. Let a, b be given as described. Since for all $x \in [a, b]$, $0 \leq r(x) \leq 1$, r is bounded on $[a, b]$. We will use Theorem 5.20 with the value $I = 0$, so it suffices to let $\epsilon > 0$ be given and find a lower step function s for r on $[a, b]$ and an upper step function t for r on $[a, b]$ such that

$$-\epsilon < \int_a^b s \leq \int_a^b t < \epsilon$$

First, define s to be constantly zero on $[a, b]$, so that clearly s is a lower step function for r and has integral zero on $[a, b]$. To define t, let $P = \left\{x \in [a, b] \mid r(x) \geq \frac{\epsilon}{2(b-a)}\right\} \cup \{a, b\}$. Clearly P only contains a, b, and some rational numbers. If P contains a rational number which has the form p/q in lowest terms, then we must have $r(p/q) = 1/q \geq \epsilon/(2(b-a))$. Only finitely many positive integers q have that property, and for each such q, only finitely many integers p have the property that $p/q \in [a, b]$. (In fact, there are at most $q(b-a)$ choices for p.) Thus, P is a finite set and is a partition of $[a, b]$.

If we define t to have the constant value $\epsilon/(2(b-a))$ on the open subintervals of P, and define $t(x) = 1$ for all $x \in P$, then t is, by definition of P, an upper step function for r on $[a, b]$. Furthermore,

$$\int_a^b t = \int_a^b \frac{\epsilon}{2(b-a)}\, dx = \left(\frac{\epsilon}{2(b-a)}\right)(b-a) = \frac{\epsilon}{2} < \epsilon$$

which finishes the proof. ∎

Note that in this proof, in order to make the integral of $t - s$ on $[a, b]$ less than ϵ, we chose s and t so that $t(x) - s(x)$ is less than $\epsilon/(b-a)$ at every $x \in [a, b]$. This is a common trick for proofs involving integration.

The Staircase Function

The staircase function s is an interesting example to consider for several reasons. First, like the ruler function, the staircase function is continuous at each irrational number and discontinuous at each rational number. Second, the staircase function is strictly increasing. Third, the staircase function is a limit of approximations: recall that for each $x \in \mathbb{R}$ we defined

$$s(x) = \sup_{n \in \mathbb{N}^*} g_n(x) = \lim_{n \to \infty} g_n(x)$$

where the n^{th} approximation $g_n(x)$ is

$$g_n(x) = \sum_{i=1}^{n} \left(\frac{1}{2}\right)^i H(x - r_i)$$

(H is the Heaviside function $\chi_{[0,\infty)}$, and r_i is the i^{th} rational number in our enumeration).

Note that for each $n \in \mathbb{N}$ and for any closed and bounded interval $[a, b]$, g_n is a step function, because H is a step function and thus g_n is a linear combination of step functions. Also, $s(x) \geq g_n(x)$ for all $x \in [a, b]$, so each g_n is a lower step function for s on $[a, b]$. What would be a corresponding sequence of upper step functions for s on $[a, b]$? One possibility is to use the constant function with value 1 as an upper step function, but that's a poor choice because it doesn't become arbitrarily close to the values of s. Instead, to get a better choice, let's analyze how well g_n approximates s.

To do this, let's ask: how large is $s(x) - g_n(x)$? Note that $s(x) - g_n(x)$ is the limit as $m \to \infty$ of $g_m(x) - g_n(x)$. When $m \in \mathbb{N}^*$ and $m \geq n$, we obtain the following by canceling terms in a summation and noting that the function H is bounded above by 1:

$$\begin{aligned} g_m(x) - g_n(x) &= \sum_{i=1}^{m} \left(\frac{1}{2}\right)^i H(x - r_i) - \sum_{i=1}^{n} \left(\frac{1}{2}\right)^i H(x - r_i) \\ &= \sum_{i=n+1}^{m} \left(\frac{1}{2}\right)^i H(x - r_i) \leq \sum_{i=n+1}^{m} \left(\frac{1}{2}\right)^i \end{aligned}$$

To compute this sum, we pull out $n+1$ factors of $1/2$ from each term in the sum, and then we use Exercise 1.9.2 to find

$$\begin{aligned}
\left(\frac{1}{2}\right)^{n+1} \sum_{i=n+1}^{m} \left(\frac{1}{2}\right)^{i-n-1} &= \left(\frac{1}{2}\right)^{n+1} \sum_{i=0}^{n-m-1} \left(\frac{1}{2}\right)^{i} \\
&= \left(\frac{1}{2}\right)^{n+1} \frac{1-\left(\frac{1}{2}\right)^{n-m}}{1-\frac{1}{2}} \\
&< \left(\frac{1}{2}\right)^{n+1} \frac{1}{1-\frac{1}{2}} = \left(\frac{1}{2}\right)^{n}
\end{aligned}$$

Thus, since this inequality is true for all $m \geq n$, in the limit as $m \to \infty$, we have $s(x) - g_n(x) \leq (1/2)^n$. Note that this inequality depends only on n and not on x. We describe this situation by saying that the g_n functions approach s *uniformly*; we'll study this situation more in Chapter 10.

This inequality helps us, because we can express it in the equivalent form $g_n(x) + (1/2)^n \geq s(x)$. In other words, if we define $t_n(x) = g_n(x) + (1/2)^n$ for each $x \in \mathbb{R}$ and $n \in \mathbb{N}^*$, then t_n is an upper step function for s on $[a,b]$. Furthermore, the difference between our lower and upper step functions, $t_n - g_n$, approaches 0 as n grows arbitrarily large.

We are now ready to prove that s is integrable on $[a,b]$ by using Theorem 5.22. Let $\epsilon > 0$ be given. We'll use the tactic that we used when analyzing the ruler function: we'll look for upper and lower step functions whose values differ by less than $\epsilon/(b-a)$. Thus, pick $n \in \mathbb{N}^*$ large enough so that $(1/2)^n < \epsilon/(b-a)$, and we'll use the lower step function g_n and the upper step function t_n. Therefore, we have

$$\int_a^b (t_n - g_n) = \int_a^b \left(\frac{1}{2}\right)^n dx < \int_a^b \frac{\epsilon}{b-a}\, dx = \epsilon$$

as desired. As a byproduct of this proof, we also find that the integral $\displaystyle\int_a^b s$ is the limit of $\displaystyle\int_a^b g_n$, because for all $n \in \mathbb{N}^*$,

$$0 \leq \int_a^b (s - g_n) \leq \int_a^b (t_n - g_n) = 2^{-n}(b-a)$$

which approaches 0 as $n \to \infty$. (Unfortunately, this doesn't give us an easy-to-compute formula for the value of the integral, but it does give us a way to find good approximations to the value of the integral.)

A similar process can be used whenever a function is a uniform limit of step functions, as you will be asked to show in Exercise 5.6.6.

Integrability Of Monotone Functions

We have another approach available for showing that the staircase functions is integrable on any interval. Unlike the other examples we have analyzed in this section, the staircase function is increasing. Increasing functions are rather easy to approximate with step functions, because if you use the partition with i^{th} open subinterval (x_{i-1}, x_i), then you know that the function's value at x_{i-1} will be a lower bound, and the function's value at x_i will be an upper bound. In other words, we know that the left endpoint of the subinterval always can be used to make a lower step function, and the right endpoint can be used to make an upper step function.

We can use this idea to prove the following theorem:

Theorem 5.33. *Let $a, b \in \mathbb{R}$ be given with $a \leq b$, and let f be a real function which is defined on $[a, b]$. If f is monotone on $[a, b]$ (so therefore it is bounded by the values $f(a)$ and $f(b)$), then f is integrable on $[a, b]$.*

Strategy. Let's consider the case when f is increasing, because in the case when f is decreasing, $-f$ is increasing. We'd like to use Theorem 5.22 by showing that for any $\epsilon > 0$, we can find an upper step function t and a lower step function s for f on $[a, b]$ whose integrals on $[a, b]$ are less than ϵ apart. To do this, we'll use the approach we used when calculating the integrals of x and x^2: using equal-width partitions and left endpoints or right endpoints.

Thus, suppose $n \in \mathbb{N}^*$, and we make the partition consisting of n equal-width subintervals, so each subinterval has width $(b-a)/n$. Let's say the i^{th} subinterval is $[x_{i-1}, x_i]$ for each i from 1 to n. For any $x \in [x_{i-1}, x_i]$, since f is increasing, we have $f(x_{i-1}) \leq f(x) \leq f(x_i)$. This means that if we make t have the value $f(x_i)$ and s have the value $f(x_{i-1})$ on (x_{i-1}, x_i), then t is an upper step function and s is a lower step function.

When we calculate the integral of $t-s$, the width of the partition intervals can be factored out of the sum (because all the widths are equal). Thus, we have

$$\int_a^b (t-s) = \frac{b-a}{n} \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))$$

In this sum, lots of cancelation occurs (this kind of sum is called a *telescoping sum*). You should check for yourself that this simplifies to

$$\frac{b-a}{n}(f(x_n) - f(x_0))$$

Since $b = x_n$ and $a = x_0$, this integral of $t-s$ can be made as small as desired by choosing a large enough value of n.

Proof. Let a, b, f be given as described. WLOG, we may assume that f is increasing. This is because when f is decreasing, we can apply the proof for the increasing case to $-f$ and use the linearity property.

Since f is increasing on $[a, b]$, we have $f(a) \le f(x) \le f(b)$ for all $x \in [a, b]$, so f is bounded on $[a, b]$. We shall use Theorem 5.22, so let $\epsilon > 0$ be given. We wish to construct a lower step function s and an upper step function t for f on $[a, b]$ so that the integral of $t - s$ on $[a, b]$ is less than ϵ.

Choose $n \in \mathbb{N}^*$ large enough so that

$$\frac{(f(b) - f(a))(b-a)}{n} < \epsilon$$

(which is possible because the left fraction goes to 0 as $n \to \infty$). Let P be the partition $\{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ which uses equal-width subintervals: more precisely, for each i from 0 to n, $x_i = a + i(b-a)/n$. Now, for each i from 1 to n, we define s and t for each x in the i^{th} open subinterval (x_{i-1}, x_i) of P by

$$s(x) = f(x_{i-1}) \qquad t(x) = f(x_i)$$

Because f is increasing, we have $s(x) \le f(x) \le t(x)$ for all $x \in (x_{i-1}, x_i)$. Also, define $s(x) = t(x) = f(x)$ for each $x \in P$, so that s is a lower step function and t is an upper step function for f on $[a, b]$.

By our choice of n, s, and t, we have

$$\begin{aligned}
\int_a^b (t-s) &= \sum_{i=1}^{n} (f(x_i) - f(x_{i-1})) \left(\frac{b-a}{n}\right) \\
&= \frac{b-a}{n} \sum_{i=1}^{n} (f(x_i) - f(x_{i-1})) \\
&= \frac{b-a}{n}(f(x_n) - f(x_0)) \\
&= \frac{(f(b) - f(a))(b-a)}{n} < \epsilon
\end{aligned}$$

as desired. ■

In fact, not only does this theorem prove that monotone functions are integrable, it gives us a good way to find the integral of a monotone function. In the process of proving the theorem, we created a lower and upper step function, and we showed that the integral of the difference approaches 0. Thus, the value $\int_a^b f$ should be the only number between the integrals of these upper and lower estimates. You can use this strategy to prove the following corollary in Exercise 5.6.7:

Corollary 5.34. *Let $a, b, I \in \mathbb{R}$ with $a \leq b$ be given, and let f be a real function defined and bounded on $[a, b]$. For each $n \in \mathbb{N}^*$ and each i from 0 to n, define*

$$x_{i,n} = a + \frac{i(b-a)}{n}$$

(i.e. $x_{i,n}$ is the i^{th} partition point in an equal-widths partition of $[a, b]$ using n subintervals). If f is increasing on $[a, b]$ and

$$\sum_{i=1}^{n} f(x_{i-1,n}) \left(\frac{b-a}{n}\right) \leq I \leq \sum_{i=1}^{n} f(x_{i,n}) \left(\frac{b-a}{n}\right)$$

is true for every $n \in \mathbb{N}^*$, then $\int_a^b f = I$. Similarly, if f is decreasing on $[a, b]$ and

$$\sum_{i=1}^{n} f(x_{i,n}) \left(\frac{b-a}{n}\right) \leq I \leq \sum_{i=1}^{n} f(x_{i-1,n}) \left(\frac{b-a}{n}\right)$$

is true for every $n \in \mathbb{N}^*$, then $\int_a^b f = I$.

Normally, when computing integrals, we have to find the candidate number I between the upper and the lower estimates, and we also have to prove that the difference of the estimates has an integral going to 0. With the aid of Corollary 5.34, we can skip showing that the difference goes to 0, and we only have to find the number I which satisfies the inequalities of the corollary. This makes computing some integrals much easier, as you'll see with some of the following exercises.

5.6 Exercises

1. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined for all $x \in \mathbb{R}$ by

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ -x & \text{if } x \in \mathbb{R} - \mathbb{Q} \end{cases}$$

Prove that f is not integrable on $[0, 1]$. (Note: There is a quick way to solve this problem using the linearity property with a previously studied example.)

2. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined for all $x \in \mathbb{R}$ by

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ x^2 & \text{if } x \in \mathbb{R} - \mathbb{Q} \end{cases}$$

Prove that f is not integrable on $[0, 1]$.

3. Let $a, b \in \mathbb{R}$ with $a < b$ be given, and suppose that $f : [a, b] \to \mathbb{R}$ is given. If there exists some $c \in [a, b]$ such that $f(c) \neq 0$ and f is continuous at c, then prove that $f\chi_{\mathbb{Q}}$ is not integrable on $[a, b]$. (Hint: WLOG you may suppose $f(c) > 0$. Try to use the main idea from Exercise 5.4.13.)

4. Show that for any $k \in \mathbb{Q}$ with $k > 0$, the k^{th} power of the ruler function r^k is integrable on any interval $[a, b]$, with $\displaystyle\int_a^b r^k(x)\,dx = 0$.

5. Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let f be a real function which is integrable on $[a, b]$. Therefore, since f is bounded on $[a, b]$, $|f|$ is also bounded on $[a, b]$. This exercise outlines a proof that $|f|$ is integrable on $[a, b]$. (Using the interval addition property, this is easy to prove if f only changes sign finitely many times, but this proof will work even for oscillating integrable functions!)

   (a) Define the *positive part* $f_+ : \text{dom}(f) \to \mathbb{R}$ of f (this is read "f plus") by $f_+(x) = \max\{f(x), 0\}$ for all $x \in \text{dom}(f)$. Thus, when $f(x)$ is nonnegative, $f_+(x) = f(x)$, but when $f(x)$ is negative, $f_+(x)$ is zero. Prove that f_+ is integrable on $[a, b]$.

(Hint: If t and s are respectively upper and lower step functions for f on $[a, b]$, compatible with a partition P, then you want to make upper and lower step functions t_2 and s_2 for f_+ on $[a, b]$ such that $t_2 - s_2 \le t - s$. Define t_2 and s_2 by cases for each subinterval of P based on whether t is positive on that subinterval.)

(b) Define the *negative part* $f_- : \text{dom}(f) \to \mathbb{R}$ of f (this is read "f minus") by $f_-(x) = \max\{-f(x), 0\}$ for all $x \in \text{dom}(f)$. Thus, when $f(x)$ is nonpositive, $f_-(x) = |f(x)|$ (the absolute value is used by convention), but when $f(x)$ is positive, $f_-(x) = 0$. Prove that f_- is integrable on $[a, b]$.

(Hint: It's possible to reuse the work from part (a).)

(c) Note that for all $x \in \text{dom}(f)$, $f(x) = f_+(x) - f_-(x)$ (which is easily proven by cases). Find a similar expression for $|f(x)|$ in terms of $f_+(x)$ and $f_-(x)$, and use it to prove that $|f|$ is integrable on $[a, b]$.

6. Let $a, b \in \mathbb{R}$ with $a \le b$ be given. Let $f : [a, b] \to \mathbb{R}$ be given. Suppose that for each $\epsilon > 0$, there is a step function $g_\epsilon : [a, b] \to \mathbb{R}$ on $[a, b]$ satisfying
$$|f(x) - g_\epsilon(x)| \le \epsilon$$
for all $x \in [a, b]$ (i.e. we say the difference $f - g_\epsilon$ is bounded *uniformly*, since the bound doesn't depend on x). Prove that f is integrable on $[a, b]$. (Hint: g_ϵ may not be an upper or a lower step function, but there are upper and lower step functions for f which are close to g_ϵ.)

7. Prove Corollary 5.34.

8. Let $a, b \in \mathbb{R}$ and $p \in \mathbb{N}$ be given. This exercise will outline how to use Corollary 5.34 to find the integral of x^p on $[a, b]$. From this, we can use linearity to integrate any polynomial.

(a) Note that for any $x, y \in \mathbb{R}$, the following identity holds:
$$y^{p+1} - x^{p+1} = (y - x) \sum_{i=0}^{p} y^i x^{p-i}$$
(You can verify this by distributing $y - x$ through the summation and seeing what cancels.) Use this to show that when $0 \le x < y$, we have
$$(p+1)x^p \le \frac{y^{p+1} - x^{p+1}}{y - x} \le (p+1)y^p$$

(b) Using part (a), show that for any $n \in \mathbb{N}^*$ and any i from 1 to n,

$$\left(\frac{i-1}{n}\right)^p \cdot \frac{1}{n} \leq \frac{1}{p+1} \cdot \left(\left(\frac{i}{n}\right)^{p+1} - \left(\frac{i-1}{n}\right)^{p+1}\right) \leq \left(\frac{i}{n}\right)^p \cdot \frac{1}{n}$$

(c) For each $n \in \mathbb{N}^*$, by adding up the inequalities in part (b) for each i from 1 to n, show that

$$\sum_{i=1}^{n} \left(\frac{i-1}{n}\right)^p \cdot \frac{1}{n} \leq \frac{1}{p+1} \leq \sum_{i=1}^{n} \left(\frac{i}{n}\right)^p \cdot \frac{1}{n}$$

Therefore, by Corollary 5.34, we have $\displaystyle\int_0^1 x^p\,dx = \frac{1}{p+1}$ (because the function taking $x \in [0,1]$ to x^p is increasing).

(d) Use the properties of the integral with part (c) to show that for all $a, b \in \mathbb{R}$,

$$\int_a^b x^p\,dx = \frac{b^{p+1} - a^{p+1}}{p+1}$$

9. Let $a, b \in \mathbb{R}$ be given with $0 < a < b$. Therefore, if $f : [a,b] \to \mathbb{R}$ is defined by $f(x) = 1/\sqrt{x}$ for all $x \in [a,b]$, then f is decreasing on $[a,b]$.

(a) By performing a rationalization, show that whenever $x, y \in \mathbb{R}$ satisfy $0 < x < y$,

$$\frac{1}{2\sqrt{y}} \leq \frac{\sqrt{y} - \sqrt{x}}{y - x} \leq \frac{1}{2\sqrt{x}}$$

(b) Use part (a) to show that for any $n \in \mathbb{N}^*$ and any i from 1 to n,

$$\frac{1}{\sqrt{x_{i,n}}}\left(\frac{b-a}{n}\right) \leq 2(\sqrt{x_{i,n}} - \sqrt{x_{i-1,n}}) \leq \frac{1}{\sqrt{x_{i-1,n}}}\left(\frac{b-a}{n}\right)$$

where $x_{i,n} = a + \dfrac{i(b-a)}{n}$.

(c) Use part (b) and Corollary 5.34 to prove that

$$\int_a^b \frac{1}{\sqrt{x}}\,dx = 2(\sqrt{b} - \sqrt{a})$$

10. This exercise will find the integral of cos. It will also demonstrate how the Mean Value Theorem can be useful in computing integrals.

    (a) Use the MVT to show that if $x, y \in [0, \pi]$ satisfy $x < y$, then
    $$\cos y < \frac{\sin y - \sin x}{y - x} < \cos x$$

    (b) Use part (a) and Corollary 5.34 to prove that for all $a, b \in [0, \pi]$ with $a \leq b$,
    $$\int_a^b \cos x \, dx = \sin b - \sin a$$

    (c) Using a similar approach to parts (a) and (b), prove that for all $a, b \in [\pi, 2\pi]$ with $a \leq b$,
    $$\int_a^b \cos x \, dx = \sin b - \sin a$$

11. The previous exercise established the validity of the formula
    $$\int_a^b \cos x \, dx = \sin b - \sin a$$
    whenever $a < b$ and $a, b \in [0, \pi]$ or $a, b \in [\pi, 2\pi]$. This exercise shows that this formula is valid for ANY $a, b \in \mathbb{R}$.

    (a) First, use the periodicity of sin and cos to show that the formula is valid whenever $a < b$ and a and b lie between the same multiples of π, i.e. there is some $k \in \mathbb{Z}$ so that $a, b \in [k\pi, (k+1)\pi]$.
    (b) Next, use interval addition and part (a) to prove that the formula is valid for any $a, b \in \mathbb{R}$ with $a < b$.
    (c) Lastly, prove that the formula is valid when $a \geq b$ (i.e. when the integral limits are out of order).

5.7 Continuity and Integrals

In this section, we'll explore the relationship between integrability and continuity. As we've seen in Some Unusual Functions III, for each $a, b \in \mathbb{R}$ with $a < b$, there are integrable functions on $[a, b]$ which are not continuous on

$[a, b]$. The ruler function and the staircase function are two such examples, which are both continuous at irrational points and discontinuous at rational points. In contrast, $\chi_{\mathbb{Q}}$ is discontinuous everywhere and is not integrable on any interval $[a, b]$ with $a < b$. These examples suggest the question: does a continuous function have to be integrable?

To answer this question, let's see what properties of continuous functions might be useful for proving integrability. To get ideas, let's analyze an integrability proof we've already done. Consider the proof we gave of Theorem 5.33, which shows that an increasing function f on $[a, b]$ is integrable on $[a, b]$. The main goal of that proof is to find upper and lower step functions t and s so that the integral of $t - s$ is less than ϵ. It can be helpful to note that the integral of $t - s$ corresponds to the area between the graphs of t and s, as you can see in Figure 5.9.

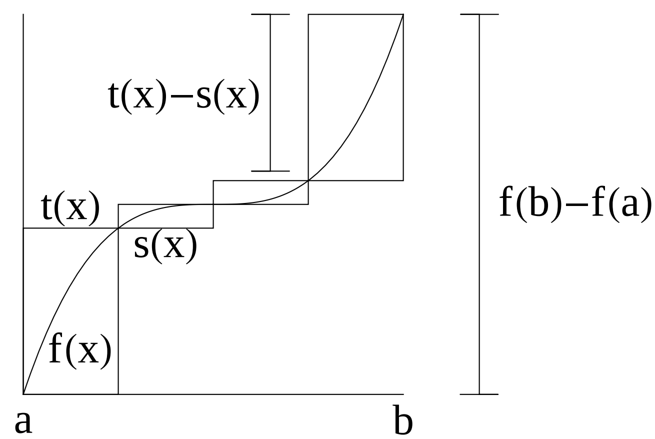

Figure 5.9: Illustrating $\displaystyle\int_a^b (t - s)$ for a monotone function

In the proof of Theorem 5.33, we showed that no matter how many pieces n we use in our partition defining t and s, the integral of $t - s$ is equal to $(f(b) - f(a))$ times the width of a partition subinterval. In other words, for all our partitions, when we add up the heights of the rectangles which represent the area between t and s, we always get the same total height of $f(b) - f(a)$. In Figure 5.9, $f(x)$, $t(x)$, and $s(x)$ are drawn, and rectangles representing the area between t and s are also drawn with solid lines.

In that proof, because the total height of our rectangles never changes, we may make the total area between t and s small by adjusting the widths of the rectangles used. Hence, by choosing a large enough n, we make the partition width small enough so that the total area between t and s, which is $(f(b) - f(a))(b-a)/n$, becomes as close to 0 as desired. In other words, in order to make the total area between t and s small, the proof for monotone functions makes the rectangles "thin".

Now, there are two ways to make a rectangle with small area: the rectangle can be "thin" (i.e. have small width), or it can be "short" (i.e. have small height). The proof for monotone functions made rectangles as thin as desired; maybe we can make a proof that makes rectangles as short as desired. This means we should try to establish some bound on the height between t and s. In general, when we are trying to define t and s on some open subinterval (x_{i-1}, x_i) of a partition, we know that we must have $s(x) \le f(x) \le t(x)$ for all $x \in (x_{i-1}, x_i)$. Thus, if we give $t(x)$ the constant value b_i on (x_{i-1}, x_i), and we give $s(x)$ the constant value a_i, we must have

$$a_i \le \inf\{f(x) \mid x \in (x_{i-1}, x_i)\} \le \sup\{f(x) \mid x \in (x_{i-1}, x_i)\} \le b_i$$

Hence,

$$b_i - a_i \ge \sup\{f(x) \mid x \in (x_{i-1}, x_i)\} - \inf\{f(x) \mid x \in (x_{i-1}, x_i)\}$$

This says that the height between t and s on this subinterval, which is $b_i - a_i$, cannot be smaller than $\text{extent}(f, (x_{i-1}, x_i))$. Hence, if we want to make our rectangles short, then we need our function to have some kind of property where the partition subintervals have small extents. One such property is the ϵ-extent property, as defined in Definition 3.64, and we showed in Theorem 3.66 that if f is continuous on $[a, b]$, then $[a, b]$ has that property for f. This leads us to the following statement:

Theorem 5.35. *Let $a, b \in \mathbb{R}$ with $a \le b$ be given, and let f be a real function such that for every $\epsilon > 0$, $[a, b]$ has the ϵ-extent property for f. Then f is integrable on $[a, b]$. In particular, if f is continuous on $[a, b]$, then f is integrable on $[a, b]$.*

Strategy. Most of the strategy has been mentioned already, but we'll point out here how to convert that intuition into a proof. Our aim is to take an arbitrary $\epsilon > 0$ and make the integral of $t - s$ less than ϵ. As we've seen before, one way to do this is to make $t(x) - s(x)$ less than $\epsilon/(b - a)$ for each $x \in [a, b]$. Since we've seen that $t(x) - s(x)$ can be, at its smallest, the extent of f on a partition subinterval, we wish to make a partition of $[a, b]$ such that for each subinterval (x_{i-1}, x_i), we have extent$(f, (x_{i-1}, x_i)) < \epsilon/(b-a)$. This is precisely what the $\epsilon/(b-a)$-extent property gives us.

Proof. Let a, b, f be given as described. We will use Theorem 5.22 to prove that f is integrable on $[a, b]$. Thus, let $\epsilon > 0$ be given, and we'll find an upper step function t for f on $[a, b]$ and a lower step function s for f on $[a, b]$ satisfying

$$\int_a^b (t - s) < \epsilon$$

(Note that finding an upper and lower step function for f on $[a, b]$ implies that f is bounded on $[a, b]$.)

Because $[a, b]$ has the $\epsilon/(b-a)$-extent property for f, there exists some $n \in \mathbb{N}^*$ such that when $[a, b]$ is broken into n equal-width subintervals, each closed subinterval has extent less than $\epsilon/(b-a)$ for f. Let $P = \{x_0, x_1, \ldots, x_n\}$ be the partition of $[a, b]$ into n equal-width subintervals. For each i from 1 to n, we define

$$a_i = \inf\{f(x) \mid x \in [x_{i-1}, x_i]\} \qquad b_i = \sup\{f(x) \mid x \in [x_{i-1}, x_i]\}$$

Then, define $s(x) = a_i$ and $t(x) = b_i$ for all $x \in (x_{i-1}, x_i)$. Also, we define $s(x) = t(x) = f(x)$ for each $x \in P$. Therefore, s is a lower step function for f on $[a, b]$ and t is an upper step function for f on $[a, b]$.

Now, we claim that for all $x \in [a, b]$, $t(x) - s(x) < \epsilon/(b-a)$. This is certainly true when $x \in P$. Otherwise, if $x \in (x_{i-1}, x_i)$ for some i from 1 to n, we have

$$\begin{aligned} t(x) - s(x) &= b_i - a_i \\ &= \sup\{f(x) \mid x \in [x_{i-1}, x_i]\} - \inf\{f(x) \mid x \in [x_{i-1}, x_i]\} \\ &= \text{extent}(f, [x_{i-1}, x_i]) \\ &< \frac{\epsilon}{b-a} \end{aligned}$$

by the definition of extent and by the choice of n.

Therefore,

$$\int_a^b (t - s) < \int_a^b \frac{\epsilon}{b-a}\, dx = \epsilon$$

as desired. ■

Example 5.36:

Theorem 5.35 says that if a function f is continuous on a closed bounded interval $[a, b]$, then it is integrable on $[a, b]$. It is important that f be continuous on a closed interval. If we only know that f is continuous on (a, b), then f might not be bounded on (a, b)!

For instance, consider $f : [0, 1] \to \mathbb{R}$, defined for all $x \in [0, 1]$ by

$$f(x) = \begin{cases} 1/\sqrt{x} & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Since $f(x) \to \infty$ as $x \to 0^+$, f is not continuous at 0, but it is continuous on $(0, 1]$. Also, f is not bounded on $[0, 1]$, so f is not integrable on $[0, 1]$. However, for any $a \in (0, 1]$, f is integrable on $[a, 1]$, and Exercise 5.6.9 shows that

$$\int_a^1 f = \int_a^1 \frac{1}{\sqrt{x}}\, dx = 2(\sqrt{1} - \sqrt{a})$$

As $a \to 0^+$, this quantity approaches 2. This limit of 2 is called the *improper integral* of f on $[0, 1]$, obtained by taking a limit of "proper" integrals. We'll revisit this topic in Chapter 9. $\square$

Example 5.37:

In the previous example, our function is unbounded. It turns out that if a function $f : [a, b] \to \mathbb{R}$ is continuous everywhere except one point, AND if f is bounded, then f is integrable on $[a, b]$. You can prove this in Exercise 5.8.1, but we'll illustrate the basic idea with an example here.

Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

for all $x \in [0, 1]$. Therefore, $f(x)$ has no limit as $x \to 0^+$, because f oscillates wildly. However, unlike the previous example, f is bounded: $|f(x)| \le 1$ for all $x \in [0, 1]$. We'll show that f is integrable on $[0, 1]$ by using Theorem 5.22. Thus, we let $\epsilon > 0$ be given, and we find upper and lower step functions t and s such that the area between t and s is less than ϵ.

Because f is continuous on $(0, 1]$, we know that f is integrable on intervals that don't contain 0. In other words, as long as we stay some distance $\delta > 0$

away from 0, f is integrable on $[\delta, 1]$. This suggests that our proof should handle the intervals $[0, \delta]$ and $[\delta, 1]$ separately, i.e. the "points near 0" and the "points away from 0". More formally, we'll find t, s, and δ so that

$$\int_0^\delta (t-s) < \frac{\epsilon}{2} \qquad \text{and} \qquad \int_\delta^1 (t-s) < \frac{\epsilon}{2}$$

so that

$$\int_0^1 (t-s) = \int_0^\delta (t-s) + \int_\delta^1 (t-s) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

(This is yet another form of "$\epsilon/2$" argument.)

On $[\delta, 1]$, because f is integrable on $[\delta, 1]$, we know we can find upper and lower step functions t_1 and s_1 for f on $[\delta, 1]$ with less than $\epsilon/2$ area between them. However, what can we do about the interval $[0, \delta]$? Since $f(x)$ oscillates wildly between -1 and 1 as x gets near 0, it seems that we can't do much better than an upper step function which is constantly 1, and similarly we'll use a lower step function which is constantly -1. Thus, we define, for all $x \in [0, 1]$,

$$s(x) = \begin{cases} -1 & \text{if } x \in [0, \delta) \\ s_1(x) & \text{if } x \in [\delta, 1] \end{cases} \qquad t(x) = \begin{cases} 1 & \text{if } x \in [0, \delta) \\ t_1(x) & \text{if } x \in [\delta, 1] \end{cases}$$

From this, we obtain

$$\int_0^1 (t(x) - s(x))\,dx = \int_0^\delta (1 - (-1))\,dx + \int_\delta^1 (t_1(x) - s_1(x))\,dx < 2\delta + \frac{\epsilon}{2}$$

Since we want 2δ to be less than $\epsilon/2$, we shall choose $\delta = \epsilon/4$. This completes the proof. $\square$

Average Value Of A Function

One useful application of the integral, apart from measuring area, is to make a meaningful definition of the average value of a real function. To see how this works, let's look at the usual notion of the average of a collection of numbers, and we'll see how the average can be viewed in terms of areas.

If a_1 through a_n are n real numbers, where $n \in \mathbb{N}^*$, then the average of a_1 through a_n is

$$\frac{1}{n} \sum_{i=1}^{n} a_i$$

In other words, we add up all the numbers, and then we divide by our count of how many numbers there are. Another way to look at this is to make a step function s on $[0,1]$ compatible with the partition $P_n = \{0, 1/n, 2/n, \ldots, 1\}$ such that for each i from 1 to n, our step function takes the value a_i on $((i-1)/n, i/n)$. Then

$$\int_0^1 s = \sum_{i=1}^{n} a_i \cdot \frac{1}{n}$$

so the area under s is the same as the average value of the a_i's. In this case, the interval widths of $1/n$ represent the condition that each a_i has equal weight when computing the average.

Another way to look at this quantity is that it measures the average height of the step function s on $[0,1]$. In essence, if we take the area under s and smooth it out to form one rectangle of width 1, then the average of the a_i's is the height of that rectangle. See the left picture in Figure 5.10, where the average value is conveyed by the dashed horizontal line.

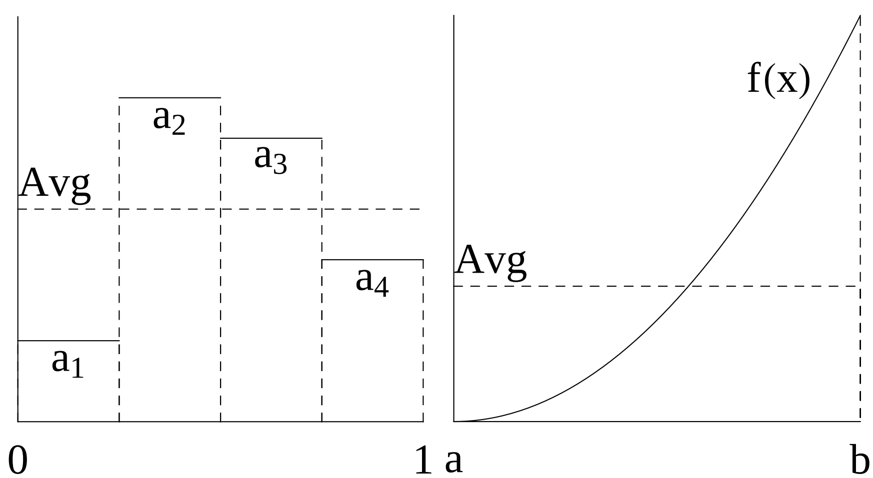

Figure 5.10: The average value of a function in terms of area

We can apply this "smoothing out" idea to continuous functions on a closed interval $[a,b]$. If f is a real function which is continuous on $[a,b]$, then $\displaystyle\int_a^b f$ is the area under f. If we smooth that area out to form one rectangle of width $b-a$, then the height of that rectangle should be the average value of the function. See the right picture in Figure 5.10, where it seems like the area above the dashed line under f can be "shoved" into the space below the dashed line and above f.

This discussion motivates the definition:

Definition 5.38. Let $a, b \in \mathbb{R}$ with $a < b$ be given, and let f be a real function which is continuous on $[a, b]$. The *average value* of f on $[a, b]$ is

$$A_a^b(f) = \frac{1}{b-a} \int_a^b f$$

In general, when taking an average of finitely many numbers, the average value does not have to equal any of the numbers. However, when averaging a continuous function, since continuous functions don't jump values, we can show that the average value IS attained by the function:

Theorem 5.39 (Mean Value Theorem for Integrals). *Let $a, b \in \mathbb{R}$ be given satisfying $a < b$, and let f be a real function which is continuous on $[a, b]$. Then for some $c \in [a, b]$, we have*

$$A_a^b(f) = \frac{1}{b-a} \int_a^b f(x)\,dx = f(c)$$

Strategy. Intuitively, the average value $A_a^b(f)$ should be between the largest and smallest values that f takes on the interval $[a, b]$. (Note that such maximum and minimum values exist because of the Extreme Value Theorem.) We can do a simple calculation to verify that this intuition is correct; namely, if $f(m)$ is the minimum value of f and $f(M)$ is the maximum value of f, then $f(m) \le A_a^b(f) \le f(M)$. From that, the Intermediate Value Theorem should give us our value of c.

Proof. Let a, b, f be given as described. By the Extreme Value Theorem, f attains an absolute maximum and absolute minimum on $[a, b]$. Suppose that f attains its absolute minimum at $m \in [a, b]$ and attains its absolute maximum at $M \in [a, b]$. Therefore, for all $x \in [a, b]$, $f(m) \le f(x) \le f(M)$, and the comparison property of integrals gives us

$$f(m)(b-a) \le \int_a^b f(x)\,dx \le f(M)(b-a)$$

Dividing throughout by $b - a$, we find that $A_a^b(f)$ is in $[f(m), f(M)]$. If $A_a^b(f)$ is equal to $f(m)$, then we take $c = m$ and are done. If $A_a^b(f)$ is equal to $f(M)$, then we take $c = M$ and are done. Otherwise, $A_a^b(f)$ is between $f(m)$ and $f(M)$, so the Intermediate Value Theorem tells us that there is some c between m and M such that $A_a^b(f) = f(c)$. ■

As a quick example, we have previously shown that

$$\int_0^1 x\,dx = \frac{1}{2} \qquad \text{and} \qquad \int_0^1 x^2\,dx = \frac{1}{3}$$

This means that the average value of x on $[0, 1]$ is $1/2$ and the average value of x^2 on $[0, 1]$ is $1/3$. (Intuitively, more of the area under x^2 is close to 0, but then the graph rises quickly.)

Example 5.40:

You were asked to show in Exercise 5.6.10 that for any $a, b \in \mathbb{R}$ with $a < b$,

$$\int_a^b \cos x\,dx = \sin b - \sin a$$

Therefore, by the MVT for integrals, we know that for some $c \in [a, b]$,

$$\frac{\sin b - \sin a}{b - a} = \cos c$$

Since cos is the derivative of sin, the conclusion we reached looks like the conclusion of the Mean Value Theorem for derivatives. Indeed, this is only the beginning of the relationship between integrals and derivatives. We'll establish a much stronger connection, called the Fundamental Theorem of Calculus, at the end of this chapter. □

There is another notion of average that we shall consider here, which is the notion of weighted average. If we have n real numbers a_1 through a_n, where $n \in \mathbb{N}^*$, and we have n positive real numbers w_1 through w_n called weights, then the weighted average of the a_i's with respect to the weights w_i is

$$\frac{\displaystyle\sum_{i=1}^{n} a_i w_i}{\displaystyle\sum_{i=1}^{n} w_i}$$

The usual average of n numbers is the same as the weighted average where each weight is $1/n$. The idea here is that a number with higher weight has a greater effect in computing the weighted average.

One useful application of weighted average is the famous Law of the Lever. Suppose that we have a seesaw whose center is located at $x = 0$. We have n objects placed at positions a_1 through a_n, where negative positions correspond to the left side of the seesaw and positive positions correspond to the right. If for each i from 1 to n, the i^{th} object weighs w_i pounds, then the i^{th} object exerts torque $a_i w_i$ on the seesaw. (If this number is positive, then the torque rotates the seesaw clockwise, otherwise the rotation is counterclockwise.) The Law of the Lever states that the seesaw is perfectly balanced iff

$$\sum_{i=1}^{n} a_i w_i = 0$$

In general, the weighted average A of the a_i's determines the average position of an object on the seesaw. Thus, the seesaw is balanced only if the average position is directly centered on the center. Otherwise, if W is the total weight of all the objects, then the seesaw behaves the same as if one object of weight W sits on the seesaw at position A, because

$$\sum_{i=1}^{n} a_i w_i = AW$$

We can generalize the definition of weighted average to continuous functions as follows:

Definition 5.41. Let $a, b \in \mathbb{R}$ with $a < b$ be given, and suppose that f and w are real functions which are continuous on $[a, b]$. Also suppose that $w(x) \geq 0$ for all $x \in [a, b]$ and $\displaystyle\int_a^b w(x)\, dx > 0$. Then the *weighted average* of f with respect to the *weight function* w on $[a, b]$ is

$$A_a^b(f, w) = \frac{\displaystyle\int_a^b f(x)w(x)\, dx}{\displaystyle\int_a^b w(x)\, dx}$$

Think of the integral of w as the "total weight" of the function f distributed over the interval $[a, b]$.

You can also show a weighted version of the MVT for integrals is true in Exercise 5.8.4:

Theorem 5.42 (Weighted MVT for Integrals). *Let $a, b \in \mathbb{R}$ with $a < b$ be given, and let f and w be real functions which are continuous on $[a, b]$. Suppose also that w is nonnegative on $[a, b]$ and that $\int_a^b w(x)\,dx > 0$. Then the weighted average of f with respect to w on $[a, b]$ is attained by f, i.e. there exists some $c \in [a, b]$ such that*

$$A_a^b(f, w) = \frac{\displaystyle\int_a^b f(x)w(x)\,dx}{\displaystyle\int_a^b w(x)\,dx} = f(c)$$

The weighted MVT for integrals is a useful tool for bounding the values of some integrals, by writing them as a product of two functions f and w where w is easy to integrate. We give one example here and leave some others as exercises.

Example 5.43:

Let's define $f, w : [0, 1] \to \mathbb{R}$ by $f(x) = 1/\sqrt{1 + x}$ and $w(x) = x^9$ on $[0, 1]$. Note that on $[0, 1]$, $w(x) \geq 0$, and in Exercise 5.6.8, we showed that the integral of w on $[0, 1]$ is $1/10$. Therefore, for some $c \in [0, 1]$, the weighted MVT tells us that

$$\int_0^1 \frac{x^9}{\sqrt{1+x}}\,dx = \frac{1}{\sqrt{1+c}}\int_0^1 x^9\,dx = \frac{1}{10\sqrt{1+c}}$$

Since $0 \leq c \leq 1$, we conclude

$$\frac{1}{10\sqrt{2}} \leq \int_0^1 \frac{x^9}{\sqrt{1+x}}\,dx \leq \frac{1}{10}$$

□

5.8 Exercises

For these exercises, it will be helpful to recall these formulas for integration found in the last exercises: for all $a, b \in \mathbb{R}$ and all $p \in \mathbb{N}$,

$$\int_a^b x^p\,dx = \frac{b^{p+1} - a^{p+1}}{p+1} \qquad \int_a^b \cos x\,dx = \sin b - \sin a$$

1. Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and suppose that f is a real function bounded on $[a, b]$. Assume that f is discontinuous at only finitely many points in $[a, b]$, i.e. for some $k \in \mathbb{N}^*$ the discontinuities are $\{x_1, x_2, \ldots, x_k\}$. Prove that f is integrable on $[a, b]$.

   (Hint: Consider modifying the strategy used in Example 5.37 by placing an interval of width δ around each discontinuity. Define your upper and lower step functions in pieces, like the example did. At the end, you should be able to discover a value of δ which makes the proof work.)

   Remark. There is a theorem which says that a bounded function on an interval $[a, b]$ is integrable iff its set of discontinuities has measure 0. The precise definition of measure 0, as well as the proof, is beyond the scope of this book, but the main idea is that a measure 0 set has "no length".

2. Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and suppose that f and g are continuous functions on $[a, b]$. Prove the following properties of the average value, as defined by Definition 5.38:

   (a) $A_a^b(f+g) = A_a^b(f) + A_a^b(g)$.
   (b) For any real number k, $A_a^b(kf) = kA_a^b(f)$.
   (c) If $f(x) \leq g(x)$ for all $x \in [a, b]$, then $A_a^b(f) \leq A_a^b(g)$.

3. Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let f and w be continuous functions on $[a, b]$. Also assume that $w(x) > 0$ for all $x \in [a, b]$, so that $\displaystyle\int_a^b w > 0$. Let $c \in (a, b)$ be given.

   (a) Find real numbers M, N which do not depend on f (but they may depend on c) so that

   $$A_a^b(f) = MA_a^c(f) + NA_c^b(f)$$

   and prove that your answer is correct.

   (b) Find real numbers M, N which do not depend on f (but they may depend on c and w) so that

   $$A_a^b(f, w) = MA_a^c(f, w) + NA_c^b(f, w)$$

   where $A_a^b(f, w)$ is the weighted average defined in Definition 5.41.

Thus, in essence, the average from a to b is a weighted average of the averages from a to c and from c to b.

4. Prove the weighted MVT for integrals, Theorem 5.42.

5. Use the strategy in Example 5.43, as well as the fact that $\sqrt{1-x^2} = (1-x^2)/\sqrt{1-x^2}$ for all $x \in (-1,1)$, to show that

$$\frac{11}{24} \le \int_0^{1/2} \sqrt{1-x^2}\,dx \le \frac{11}{24}\sqrt{\frac{4}{3}}$$

(Hint: You should choose $w(x)$ to be a polynomial!)

6. Use the strategy in Example 5.43, and the fact that $(1+x^6) = (1+x^2)(1-x^2+x^4)$ for all $x \in \mathbb{R}$, and show that for every $b > 0$,

$$\frac{1}{1+b^6}\left(b - \frac{b^3}{3} + \frac{b^5}{5}\right) \le \int_0^b \frac{dx}{1+x^2} \le b - \frac{b^3}{3} + \frac{b^5}{5}$$

7. (a) Use the weighted MVT for integrals to prove that there is some $c \in [0, \pi/2]$ satisfying

$$\int_0^{\pi/2} x^3 \cos x\,dx = c^3$$

(b) By using different choices for f and w from what you used in part (a), prove that there is some $d \in [0, \pi/2]$ satisfying

$$\int_0^{\pi/2} x^3 \cos x\,dx = \frac{\pi^4 \cos d}{64}$$

(c) In the next chapter, we will develop techniques for showing that

$$\int_0^{\pi/2} x^3 \cos x\,dx = \frac{\pi^3}{8} - 3\pi + 6$$

(You do not yet need to know how to obtain this number.) Using this number, find the values of c and d from parts (a) and (b).

8. Let $a, b \in \mathbb{R}$ with $a \leq b$ be given, and let f and g be two functions which are nonnegative and integrable on $[a, b]$. This exercise outlines a proof that fg is integrable on $[a, b]$. (Note that this exercise does not assume that f and g are continuous!)

   The proof will use Theorem 5.22. Let $\epsilon > 0$ be given. Since f and g are nonnegative and integrable on $[a, b]$, they are bounded on $[a, b]$, so there exist constants $M, N > 0$ such that for all $x \in [a, b]$, $0 \leq f(x) \leq M$ and $0 \leq g(x) \leq N$.

   (a) Since f is integrable on $[a, b]$, we know that for any $\delta > 0$, there exists an upper step function t_1 for f on $[a, b]$ and a lower step function s_1 for f on $[a, b]$ satisfying

   $$\int_a^b (t_1 - s_1) < \delta$$

   Prove that such a choice for t_1 and s_1 exists which also satisfies

   $$0 \leq s_1(x) \leq t_1(x) \leq M$$

   for every $x \in [a, b]$. (Hint: If you make an upper step function smaller and a lower step function larger, then what happens to the integral of their difference?)

   (b) Similarly to part (a), for each $\delta > 0$, we find an upper step function t_2 for g on $[a, b]$ and a lower step function s_2 for g on $[a, b]$ such that

   $$0 \leq s_2(x) \leq t_2(x) \leq N$$

   for every $x \in [a, b]$, and also

   $$\int_a^b (t_2 - s_2) < \delta$$

   Therefore, $s_1 s_2$ is a lower step function for fg on $[a, b]$, and $t_1 t_2$ is an upper step function for fg on $[a, b]$. Show that for every $x \in [a, b]$,

   $$t_1(x)t_2(x) - s_1(x)s_2(x) \leq M(t_2(x) - s_2(x)) + N(t_1(x) - s_1(x))$$

   (Hint: Add and subtract $t_1(x)s_2(x)$.)

(c) By using the results of parts (a) and (b) with appropriate choices of δ (the choices will depend on ϵ, M, and N), show that

$$\int_a^b (t_1 t_2 - s_1 s_2) < \epsilon$$

which finishes the proof.

9. Using Exercise 5.8.8, we can prove that the product of ANY two integrable functions on $[a,b]$ is also integrable on $[a,b]$. The proof starts as follows. Let $a,b \in \mathbb{R}$ with $a \le b$ be given, and suppose f and g are real functions which are integrable on $[a,b]$. Therefore, f and g are bounded, so there exist constants $M,N > 0$ such that for all $x \in [a,b]$, $|f(x)| \le M$ and $|g(x)| \le N$.

   Hence, $f+M$ and $g+N$ are nonnegative integrable functions on $[a,b]$, by the linearity property. By writing fg in terms of $(f+M)(g+N)$, use Exercise 5.8.8 and the linearity property to prove that fg is integrable on $[a,b]$.[3]

10. Let $a,b \in \mathbb{R}$ with $a \le b$ be given. Let f be a nonnegative real function which is integrable on $[a,b]$, and assume that $1/f$ is bounded on $[a,b]$. This exercise outlines a proof that $1/f$ is integrable on $[a,b]$.

    We will use Theorem 5.22. Let $\epsilon > 0$ be given. Since f and $1/f$ are both nonnegative and bounded, there exist constants $M,N > 0$ so that for all $x \in [a,b]$, $0 \le f(x) \le M$ and $0 \le 1/f(x) \le N$.

    (a) Prove that for any $\delta > 0$, there exists an upper step function t_1 for f on $[a,b]$ and a lower step function s_1 for f on $[a,b]$ which satisfy

    $$\frac{1}{N} \le s_1(x) \le t_1(x) \le M$$

    for every $x \in [a,b]$, and

    $$\int_a^b (t_1 - s_1) < \delta$$

    (The steps will be quite similar to part (a) of Exercise 5.8.8.)

[3] An alternate proof involves breaking up f and g into their positive and negative parts, as defined in Exercise 5.6.5.

(b) From part (a), it follows that $1/s_1$ is an upper step function for $1/f$ on $[a,b]$, and $1/t_1$ is a lower step function for $1/f$ on $[a,b]$. Show that for any $x \in [a,b]$,

$$\frac{1}{s_1(x)} - \frac{1}{t_1(x)} \leq N^2(t_1(x) - s_1(x))$$

(c) By using the results of parts (a) and (b) with an appropriate choice of δ, show that

$$\int_a^b \left(\frac{1}{s_1} - \frac{1}{t_1}\right) < \epsilon$$

which finishes the proof.

11. Let f be the function from Exercise 5.4.4, and let r be the ruler function. We have seen that f and r are integrable on $[0,1]$. Prove that $f \circ r$ is not integrable on $[0,1]$, thus the composition of integrable functions does not have to be integrable. (Hint: Note that $f(r(x)) = 1$ iff x is a rational number which can be written with an odd denominator. Prove that the set of all such rational numbers is dense.)

5.9 Area Functions and the FTC

In this section, we will introduce the Fundamental Theorem of Calculus (or FTC for short). This theorem gives us a connection between all three of the major definitions of calculus: continuity, differentiation, and integration. The FTC also provides an extremely useful tool for calculating integrals, which we will use extensively in the next chapter to develop techniques for integration.

In essence, the FTC tells us that differentation and integration are inverse operations. For example, we have found that for any $a, b \in \mathbb{R}$,

$$\int_a^b \cos x = \sin b - \sin a$$

ie. the integral of cos is in terms of sin, and the derivative of sin is cos. Similarly,

$$\int_a^b x^2\,dx = \frac{b^3}{3} - \frac{a^3}{3}$$

i.e. the integral of x^2 is in terms of $x^3/3$, and the derivative of $x^3/3$ is x^2. However, we are treating these examples quite informally: when a and b are constants, the integral from a to b is a number, not a real function which can be differentiated. In order to obtain a real function which can be differentiated, we will introduce the idea of an area function.

Area Functions

We have developed the definition of the integral to measure area under a function from a to b. If a and b are held constant, then the area is also a constant number. However, we can make a new function by making the upper limit of integration a variable. In other words, if f is a real function and $a \in \mathbb{R}$, we can make a new function g by defining

$$g(x) = \int_a^x f = \int_a^x f(t)\,dt$$

which is defined whenever f is integrable on $[\min\{a, x\}, \max\{a, x\}]$. We'll call g an *area function of* f (more specifically, it is the area function *starting from* a). We can think of $g(x)$ as measuring the area under f from a to x when $x \geq a$ (when $x < a$, we "measure the area backwards" and introduce a minus sign).

Example 5.44:

As a simple example, suppose that you have a triangular lawn, which is shaped like the area under the graph of $y = t$ for $t \in [0, 100]$, where y and t are measured in feet. Thus, the total area of your lawn is $100^2/2 = 5000$ square feet. Let's say that you mow your lawn for 100 minutes in a way such that for each $x \in [0, 100]$, after x minutes you have mowed from $t = 0$ up to $t = x$. (In other words, you are approximately mowing the lawn in vertical stripes going from left to right.) Thus, if g is the area function of $y = t$ starting from 0, then after x minutes you have mowed

$$g(x) = \int_0^x t\,dt = \frac{x^2}{2}$$

square feet of area for each $x \in [0, 100]$.

Note that we are not cutting our lawn in the usual way that most people cut their lawns. Most people cut their lawns in a way so that they cover the area of their lawn at a constant rate, i.e. every minute they cut 10 square feet of lawn. In contrast, we are cutting our lawn in a way so that we move from left to right at the constant rate of 1 foot per minute! Hence, as we move farther to the right, the lawn gets wider, so we have to cut area more quickly in order to keep moving right at the same pace.

In fact, we can measure how fast we are mowing the lawn by taking the derivative of g: $g'(x) = x$ for each $x \in (0, 100)$. Near the left side of the lawn, i.e. when x is close to 0, we are cutting so slowly that we're hardly cutting any grass at all. However, as we approach the right side of the lawn, i.e. as x approaches 100, we are cutting lawn at 100 square feet a minute! $\square$

As illustrated by the lawn-mowing example, it can be helpful to think of area functions as describing how much area is covered as a function of the amount of time elapsed. This analogy suggests a few properties that area functions should have. For instance, the area under a positive function is increasing as we move from left to right. One way to formalize this is as follows: if $f(x) \geq 0$ for all $x \in [x_1, x_2]$, and g is an area function of f, then we use the interval addition property to find

$$g(x_2) - g(x_1) = \int_a^{x_2} f - \int_a^{x_1} f = \int_{x_1}^{x_2} f \geq 0$$

i.e. g is increasing from x_1 to x_2. (You can similarly show that if f is negative on an interval, then area functions decrease on that interval.) Here's another plausible property: the area under a large function grows faster than the area under a small function. For instance, we saw in our lawn-mowing example that the right side of the lawn is being mowed more quickly than the left side.

As a first property that we can prove to get started, we'll show that area functions don't "jump values", i.e. they are continuous:

Theorem 5.45. *Let $a, b, c \in \mathbb{R}$ be given with $b \leq a \leq c$, and suppose that f is a real function which is integrable on $[b, c]$. Thus, by the subinterval property (Theorem 5.27), f is integrable on every subinterval of $[b, c]$. Define $g : [b, c] \to \mathbb{R}$ for all $x \in [b, c]$ by*

$$g(x) = \int_a^x f(t)\, dt$$

i.e. g is the area function of f starting from a. Then g is continuous on $[b, c]$.

Strategy. We want to show that $g(y) \to g(x)$ as $y \to x$. Equivalently, we want to show that $g(y) - g(x) \to 0$ as $y \to x$. Thus, we want to estimate

$$|g(y) - g(x)| = \left|\int_a^y f(t)\,dt - \int_a^x f(t)\,dt\right| = \left|\int_x^y f(t)\,dt\right|$$

Because f is integrable on $[b, c]$ (and hence on $[\min\{x, y\}, \max\{x, y\}]$), f is bounded on $[b, c]$, so we can use that bound to estimate $g(y) - g(x)$ in terms of y and x.

Proof. Let a, b, c, f, g be given as described. Because f is integrable on $[b, c]$, it is bounded on $[b, c]$, so choose some $M > 0$ so that for all $x \in [b, c]$, $|f(x)| \le M$. Now, let $x, y \in [b, c]$ be given with $x \le y$. We compute

$$\begin{aligned}
|g(y) - g(x)| &= \left|\int_a^y f(t)\,dt - \int_a^x f(t)\,dt\right| \\
&= \left|\int_x^y f(t)\,dt\right| \\
&\le \int_x^y |f(t)|\,dt \\
&\le \int_x^y M\,dt = M(y - x)
\end{aligned}$$

where we used Exercise 5.4.9 in the first inequality.

We can do the same style of calculation when $y < x$ to obtain $|g(y) - g(x)| \le M|y - x|$, so we have

$$|g(y) - g(x)| \le M|y - x|$$

for all $x, y \in [b, c]$. Because $M|y - x| \to 0$ as $y \to x$, the Squeeze Theorem implies that $g(y) \to g(x)$ as $y \to x$. Thus, g is continuous at x. ∎

Remark. The proof above actually shows that when we consider an area function on a closed bounded interval $[b, c]$, the area function is actually Lipschitz-continuous, as defined in Exercise 3.12.8. This can be useful, because Lipschitz continuity is a stronger property than uniform continuity.

Also, the proof demonstrated a couple useful tricks when dealing with area functions. First, the result from Exercise 5.4.9 is often quite helpful for moving absolute-value bars inside of an integral. Second, in order to use that result most effectively, we simplified the subtraction $g(y)-g(x)$ into one integral, rather than using the Triangle Inequality to split into a sum of two separate absolute values. We will use these tricks again quite soon.

First Part of the FTC: The Derivative of An Area Function

Now, we have seen that intuitively, there is a relationship between the height of a function f and how fast the area under f grows as we move from left to right. In fact, we'll be able to show that under certain hypotheses, if g is ANY area function of f, then the derivative of g IS f! This makes sense, since when our function is $f(x)$ units high, g should intuitively be increasing at $f(x)$ square units per unit of time.

More formally, let's compute a difference quotient

$$\frac{g(x+h)-g(x)}{h} = \frac{1}{h}\left(\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt\right) = \frac{1}{h}\int_x^{x+h} f(t)\,dt$$

When $h > 0$ and f is continuous on $[x, x+h]$, this expression is just our definition of the average value $A_x^{x+h}(f)$. By the MVT for integrals, we know that this average value is attained at some $c \in [x, x+h]$, i.e. $A_x^{x+h}(f) = f(c)$. As $h \to 0^+$, we will have $c \to x$ because $x \le c \le x+h$, so $f(c) \to f(x)$. A similar argument applies when $h < 0$ and $h \to 0^-$.

The argument above is a formal proof that $g'(x) = f(x)$ when f is continuous on an open interval which contains x (because we need continuity on $[x, x+h]$ to use the MVT for integrals). However, it would be nice to be able to deduce $g'(x) = f(x)$ when f is continuous at x but not necessarily anywhere else. It turns out that a modification of the proof, using tactics from the proof we gave of Theorem 5.45, can prove this slightly stronger result:

Theorem 5.46 (Fundamental Theorem of Calculus, Part 1). *Let a, $b, c \in \mathbb{R}$ be given with $b \le a \le c$, and let f be a real function which is integrable on $[b, c]$. Thus, f is integrable on every subinterval of $[b, c]$, and we may define $g : [b, c] \to \mathbb{R}$ by*

$$g(x) = \int_a^x f(t)\,dt$$

for all $x \in [b, c]$. Then for any $x \in (b, c)$, if f is continuous at x, then g is differentiable at x with $g'(x) = f(x)$. (Recall that g can't have a two-sided derivative at b or at c.)

Strategy. To show that the difference quotient of g from x to $x+h$ approaches $f(x)$, we want to show that for any $\epsilon > 0$,

$$\left|\frac{g(x+h)-g(x)}{h} - f(x)\right| < \epsilon$$

provided h is close enough to 0. (Let's assume $h > 0$ for now... the proof for $h < 0$ is similar.) To be able to simplify the absolute value, let's manipulate things to produce one integral, like we did in our proof of Theorem 5.45. First, by making a common denominator of h, we obtain

$$\frac{g(x+h)-g(x)}{h} - f(x) = \frac{1}{h}\left(\int_x^{x+h} f(t)\,dt - hf(x)\right)$$

Now, how can we rewrite this difference as one integral from x to $x+h$? We note that

$$hf(x) = \int_x^{x+h} f(x)\,dt$$

(note that t is the variable of integration here, not x, so this is the integral of a constant function). Therefore, we obtain

$$\frac{g(x+h)-g(x)}{h} - f(x) = \frac{1}{h}\int_x^{x+h} (f(t)-f(x))\,dt$$

To show that this quantity has an absolute value less than ϵ, we want a bound for $|f(t)-f(x)|$ when $t \in [x, x+h]$. In other words, we'd like to know: how close are $f(t)$ and $f(x)$? This is where we use the assumption that f is continuous at x: so long as t is picked close enough to x, we can make $|f(t)-f(x)|$ as small as desired.

Proof. Let a, b, c, f, g be given as described. Let $x \in (b, c)$ be given, and assume that f is continuous at x. We will use the definition of derivative to show that $g'(x) = f(x)$. Thus, let $\epsilon > 0$ be given, and we need to find $\delta > 0$ so that whenever $h \in \mathbb{R}$ satisfies $0 < |h| < \delta$, we have

$$\left|\frac{g(x+h)-g(x)}{h} - f(x)\right| < \epsilon$$

Because f is continuous at x, we may choose $\delta > 0$ so that for all $t \in \mathrm{dom}(f)$ satisfying $|t-x| < \delta$, we have $|f(t)-f(x)| < \epsilon/2$. Since $x \in (b, c)$, we may suppose that δ is small enough so that $(x-\delta, x+\delta) \subseteq [b,c]$. Therefore, for any $h \in \mathbb{R}$ satisfying $0 < |h| < \delta$, $g(x)$ and $g(x+h)$ are both defined.

For any $h \in (0,\delta)$, since $g(x+h)-g(x)$ is the integral from x to $x+h$ of f, the calculations in the strategy show that

$$\begin{aligned}
\left|\frac{g(x+h)-g(x)}{h} - f(x)\right| &= \frac{1}{h}\left|\int_x^{x+h} (f(t)-f(x))\,dt\right| \\
&\leq \frac{1}{h}\int_x^{x+h} |f(t)-f(x)|\,dt \\
&\leq \frac{1}{h}\int_x^{x+h} \frac{\epsilon}{2}\,dt \\
&= \frac{1}{h}\left(\frac{h\epsilon}{2}\right) \\
&< \epsilon
\end{aligned}$$

where we used the result from Exercise 5.4.9 in the first inequality. When $h \in (-\delta, 0)$, we obtain the same result using almost exactly the same steps, except that we must swap the order of the integral limits to use Exercise 5.4.9. Thus, we obtain

$$\begin{aligned}
\left|\frac{g(x+h)-g(x)}{h} - f(x)\right| &= \frac{1}{|h|}\left|\int_x^{x+h} (f(t)-f(x))\,dt\right| \\
&= \frac{1}{|h|}\left|-\int_{x+h}^{x} (f(t)-f(x))\,dt\right| \\
&\leq \frac{1}{|h|}\int_{x+h}^{x} |f(t)-f(x)|\,dt \\
&\leq \frac{1}{|h|}\left(\frac{|h|\epsilon}{2}\right) < \epsilon
\end{aligned}$$

as desired. ■

Remark. Note that in the proof above, $g(x+h)-g(x)$ doesn't depend on a, the starting point used for the area function g. In fact, if h is another area function for f which is defined on $[b,c]$, let's suppose that h starts from $d \in [b,c]$. Then for all $x \in [b,c]$,

$$g(x) - h(x) = \int_a^x f(t)\,dt - \int_d^x f(t)\,dt = \int_d^a f(t)\,dt$$

This difference does not depend on x, i.e. it is a constant. Thus, any two area functions for f differ by a constant, which implies that they have the same derivative: $g'(x) = h'(x) = f(x)$.

Example 5.47:

The first part of the FTC can also be used to compute derivatives of variants of area functions. For instance, consider an expression of the form

$$h(x) = \int_x^0 f(t)\,dt$$

In this expression, the variable x is the lower limit instead of the upper limit. To find h', we should rewrite h with the variable in the upper limit. We flip the bounds by using a minus sign, and thus we find

$$h'(x) = \left(-\int_0^x f(t)\,dt\right)' = -f(x)$$

by Part 1 of the FTC.

Another variant to consider is something like

$$h(x) = \int_0^{x^2} f(t)\,dt$$

Here, instead of having x as the upper limit, we have x^2. If g represents the area function of f starting from 0, i.e. $g(x) = \displaystyle\int_0^x f(t)\,dt$, then $h(x) = g(x^2)$. Thus, to find h', we can use the Chain Rule, where g is the outer function and the squaring funtion is the inner function. This gives us

$$h'(x) = g'(x^2)(x^2)' = 2xf(x^2)$$

Putting these ideas all together, we can differentiate more complicated examples:

$$\begin{aligned}\frac{d}{dx}\left(\int_x^{x^2} f(t)\,dt\right) &= \frac{d}{dx}\left(\int_0^{x^2} f(t)\,dt - \int_0^x f(t)\,dt\right) \\ &= 2xf(x^2) - f(x)\end{aligned}$$

In the first step, we had to use interval addition to split up the integral into two integrals, because Part 1 of the FTC requires the lower limit to be a constant. □

Example 5.48:

Part 1 of the FTC can also be used in certain problems to solve for unknown functions. For instance, let's consider the problem of finding functions $f : \mathbb{R} \to \mathbb{R}$ which satisfy the following equation for all $x \in \mathbb{R}$:

$$\int_0^{2x} f(t)\,dt = 2\int_0^x f(t)\,dt$$

(We assume that f is integrable on every closed bounded interval.) One kind of solution is to choose f to be a function which has integral 0 on every interval: the ruler function, for instance, is one such solution, as is the function which is constantly 0. However, these solutions may not very satisfactory, especially since the ruler function has lots of discontinuities. Let's see what the continuous solutions for f are.

If we assume f is continuous, then we can take the derivative with respect to x of each side of our original equation and get

$$2f(2x) = 2f(x)$$

(In essence, the derivative is "canceling out" the integrals). Therefore, any continuous solution must satisfy $f(2x) = f(x)$ for EVERY $x \in \mathbb{R}$.

This suggests that f should be a constant function. If we take f to be the function which is constantly C, for some $C \in \mathbb{R}$, then

$$\int_0^{2x} C\,dt = 2xC \qquad \text{and} \qquad 2\int_0^x C\,dt = 2xC$$

which shows that constant functions are solutions to our problem. In fact, when f is continuous, you can show in Exercise 5.10.6 that f MUST be a constant function! $\square$

Second Part of the FTC: Antiderivatives

Part 1 of the FTC tells us an important property about area functions, but it also tells us an important property about derivatives. Namely, we now have an answer to the following problem: if f is a continuous function on $[a, b]$, then does there exist a function $g : [a, b] \to \mathbb{R}$ with $g' = f$ on (a, b)? We say that such a function g is an *antiderivative* of f on (a, b). Thanks to the FTC, we now know that any area function of f is an antiderivative.

This suggests the question: what are ALL the antiderivatives of f on (a,b)? Because the derivative of a constant is always zero, if g is one antiderivative for f, then so is $g+C$ for any constant C. In fact, the converse holds as well: if g and h are two antiderivatives for f on (a,b), then g and h must differ by a constant! Why is this? Since $h' = g' = f$ on (a,b), and $h-g$ is continuous on $[a,b]$, we have $(h-g)'(x) = 0$ for all $x \in (a,b)$, so Theorem 4.47 tells us that $h-g$ is constant on $[a,b]$.

We summarize this discussion with the following theorem, which we have just proven:

Theorem 5.49. *Let $a,b \in \mathbb{R}$ with $a < b$ be given, and let f be a continuous function on $[a,b]$. If we define $g : [a,b] \to \mathbb{R}$ by*

$$g(x) = \int_a^x f(t)\,dt$$

for all $x \in [a,b]$, then g is continuous on $[a,b]$ and $g'(x) = f(x)$ for all $x \in [a,b]$, so g is an antiderivative of f on (a,b). Furthermore, any other continuous function h on $[a,b]$ is an antiderivative of f on (a,b) iff $h-g$ is constant on $[a,b]$.

Because of this theorem, we frequently write the statement "g is an antiderivative of f" using the following notation:

$$\int f(x)\,dx = g(x) + C$$

This is meant to state that all antiderivatives of f are obtained by adding a suitable constant C to g. (The letter used for the constant does not have to be "C", but it is often some capital letter early in the alphabet for convention's sake.) The limits of integration are not written, because changing the lower limit still produces an antiderivative of f. We say that $g(x)+C$ is the *indefinite integral of f*, which is understood to be true on any closed intervals where f is continuous. In contrast, the usual integral notation with limits is sometimes called the *definite integral*.

For example, we know that for all $x > 0$ and $p \in \mathbb{Q}$ with $p \neq -1$, we have $(x^{p+1})' = (p+1)x^p$. Thus,

$$\int x^p\,dx = \frac{x^{p+1}}{p+1} + C \qquad \text{where } p \in \mathbb{Q}, p \neq -1$$

Since sin has derivative cos, and cos has derivative $-\sin$, we have

$$\int \cos x \, dx = \sin x + C \qquad \int \sin x \, dx = -\cos x + C$$

Similarly, we obtain

$$\int \frac{1}{\sqrt{1-x^2}} \, dx = \arcsin x + C$$

This last formula is only supposed to be used on closed intervals which are contained in $(-1, 1)$, because $1/\sqrt{1-x^2}$ is continuous on $(-1, 1)$.

Remark. When using indefinite integral notation, the "$+C$" is really meant to convey an entire family of functions, as opposed to specifying one particular value of C. Mistakes can arise if this distinction is forgotten when two antiderivatives are computed. For example, since $(x)' = (x+1)' = 1$ for all $x \in \mathbb{R}$, we can say

$$\int 1 \, dx = x + C \qquad \text{and} \qquad \int 1 \, dx = x + 1 + C$$

from which we find "$x + C = x + 1 + C$". However, it is NOT allowed to cancel C from each side and obtain "$x = x + 1$". This is because the symbol C does not represent one number; it instead is shorthand for representing all possible constants.

We have seen, from Part 1 of the FTC, that we can use definite integrals to find antiderivatives (by using area functions). For Part 2 of the FTC, we would like to show that conversely, you can obtain definite integrals from antiderivatives. As suggested by previously proven formulas like

$$\int_a^b x^2 \, dx = \frac{b^3}{3} - \frac{a^3}{3} \qquad \int_a^b \cos x \, dx = \sin b - \sin a$$

it seems that we should plug in the values a and b into an antiderivative and subtract the results.

Theorem 5.50 (Fundamental Theorem of Calculus, Part 2). *Let a, $b \in \mathbb{R}$ be given with $a < b$, and let f be a continuous function on $[a, b]$. Suppose that F is any continuous function on $[a, b]$ which is an antiderivative of f on (a, b). Then*

$$\int_a^b f(x) \, dx = F(b) - F(a)$$

Strategy. If we let g be the area function of f starting from a, then we know that g is also an antiderivative of f on (a, b). Thus, F and g differ by a constant. By plugging in a, we can find out what that constant is.

Proof. Let a, b, f, F be given as described. Because f is continuous on $[a, b]$, we may define $g : [a, b] \to \mathbb{R}$ by

$$g(x) = \int_a^x f(t)\,dt$$

for all $x \in [a, b]$. By Theorem 5.45, g is continuous on $[a, b]$. By Part 1 of the FTC, Theorem 5.46, g is an antiderivative of f on (a, b).

Therefore, by Theorem 5.49, F and g differ by a constant, i.e. there is some $C \in \mathbb{R}$ such that $F(x) - g(x) = C$ for all $x \in [a, b]$. When $x = a$, we know that $g(a) = 0$, which tells us that $C = F(a)$. Thus,

$$g(b) = \int_a^b f(x)\,dx = F(b) - C = F(b) - F(a)$$

as desired. ■

As a result of Part 2 of the FTC, we frequently use the notation

$$\int_a^b f(x)\,dx = F(x)\big|_a^b$$

where the vertical bar on the right represents the expression $F(b) - F(a)$ obtained by plugging in b, plugging in a, and subtracting. We read this as "the integral from a to b of $f(x)\,dx$ is $F(x)$ *evaluated at* a *and* b". For instance,

$$\int_a^b \sin x\,dx = -\cos x\big|_a^b = -(\cos b - \cos a)$$

and

$$\int_0^1 (x^4 + x^2)\,dx = \left(\frac{x^5}{5} + \frac{x^3}{3}\right)\Bigg|_0^1 = \frac{1}{5} + \frac{1}{3}$$

Remark. In order to use Part 2 of the FTC, you need the integrand to be continuous on $[a, b]$. For instance, you might be tempted to use the fact that

$1/(-2x^2)$ has derivative $1/x^3$ whenever $x \neq 0$, which leads you to try using the FTC to conclude that

$$\int_{-1}^{1} \frac{dx}{x^3} = \left. \frac{1}{-2x^2} \right|_{-1}^{1} = \frac{1}{-2} - \frac{1}{-2} = 0$$

However, this is invalid because the function taking x to $1/x^3$ is not continuous at 0. In fact, that function is unbounded on $[-1,1]-\{0\}$, so the function is not integrable on $[-1,1]$.

In summary, the two parts of the FTC say roughly that the derivative of an integral is the original function, and that integrating a derivative gives you back the original function evaluated at two points. This is formally what people mean when they say that integration and differentiation are inverse operations. What this also means is that the problem of integrating functions usually turns into the problem of finding antiderivatives of functions. Since we have computed many useful derivatives, and we have also proven some useful rules for computing derivatives, we should be able to use these things to our advantage when computing antiderivatives. That will be the focus of the next chapter.

5.10 Exercises

1. Define $g : \mathbb{R} \to \mathbb{R}$ for all $x \in \mathbb{R}$ by
$$g(x) = \int_0^x \lfloor t \rfloor \, dt$$
In other words, g is the area function corresponding to the floor function and the starting point of 0.
   (a) For each $n \in \mathbb{N}$, what is $g(n)$? Prove your answer.
   (b) Since the floor function is continuous precisely at the real numbers which are not integers, Part 1 of the FTC says that $g'(x) = \lfloor x \rfloor$ for all $x \notin \mathbb{Z}$. When $x \in \mathbb{Z}$, however, what is $g'(x)$?
   (c) Define $h : \mathbb{R} \to \mathbb{R}$ for all $x \in \mathbb{R}$ by
$$h(x) = \int_0^x g(t)\, dt = \int_0^x \left(\int_0^t \lfloor s \rfloor \, ds \right) dt$$
and find $h(2)$ and $h(-2)$.

2. Find the following derivatives for all $x \in \mathbb{R}$:

(a) $\displaystyle \frac{d}{dx}\int_{-x}^{1} \cos(t^3)\sin(t)\,dt$

(b) $\displaystyle \frac{d}{dx}\int_{x^3}^{x^2} \frac{t^6}{1+t^4}\,dt$

(c) $\displaystyle \frac{d}{dx}\int_{0}^{x} x\sin t\,dt$

(Note: Part 1 of the FTC only applies to integrands which do not depend on x!)

3. Suppose that $g : \mathbb{R} \to \mathbb{R}$ is continuous everywhere. Prove that
$$\frac{d}{dx}\int_0^x x^2 g(t)\,dt = 2x\int_0^x g(t)\,dt + x^2 g(x)$$

4. Find a continuous function $f : \mathbb{R} \to \mathbb{R}$ satisfying the following equations for all $x \in \mathbb{R}$. Make sure to plug your function back into the equation and verify that it is actually a solution to the equation.

   (a) $\displaystyle \int_0^x f(t)\,dt = x^2(1+x)$

   (b) $\displaystyle \int_0^{f(x)} t^2\,dt = x^2(1+x)$

   (c) $\displaystyle \int_0^x f(t)\,dt = \int_x^1 t^2 f(t)\,dt + \frac{x^{16}}{8} + \frac{x^{18}}{9} + c$, where c is a constant. Also, find the value of c.

5. Find a continuous function $f : \mathbb{R} \to \mathbb{R}$ and a value $a \in \mathbb{R}$ which satisfies
$$\int_a^{x^2} f(t)\,dt = \frac{x^6}{3} - 9$$
for all $x \in \mathbb{R}$.

6. In Example 5.48, we showed that constant functions are solutions to the equation
$$\int_0^{2x} f(t)\,dt = 2\int_0^x f(t)\,dt$$
for all $x \in \mathbb{R}$. Prove that if $f : \mathbb{R} \to \mathbb{R}$ is a continuous solution to the above equation, then f is a constant function! (Hint: For any $x \in \mathbb{R}$, show that $f(x) = f(0)$ by considering $f(x/2)$, $f(x/4)$, $f(x/8)$, $\ldots$.)

7. In this exercise, we will find the continuous functions $f : (0,\infty) \to \mathbb{R}$ which satisfy
$$\int_x^{kx} f(t)\,dt = \int_1^k f(t)\,dt \qquad \text{for all } x, k > 0$$

(a) For each constant $C \in \mathbb{R}$, use the stretching property from Theorem 5.29 to prove that if $f(x) = C/x$ for all $x > 0$, then f satisfies the equation above.

(b) Show that if $f : (0, \infty) \to \mathbb{R}$ is a continuous solution to the equation above, then for some $C \in \mathbb{R}$, we must have $f(x) = C/x$ for all $x > 0$.

8. Find the following indefinite integrals, using the notation we introduced before Part 2 of the FTC:

(a) $\displaystyle\int 3x^3 - x\,dx$

(b) $\displaystyle\int \frac{x^4 + x - 3}{x^3}\,dx$
(Assume $x \neq 0$.)

(c) $\displaystyle\int \sec x \tan x + \sec^2 x\,dx$
(Assume $x \neq \pi/2 + k\pi$ for any $k \in \mathbb{Z}$.)

(d) $\displaystyle\int \sin^2\left(\frac{x}{2}\right)\,dx$
(Hint: Half-angle identity)

9. Using the answers from the previous exercise, compute the following definite integrals:

(a) $\displaystyle\int_2^4 3x^3 - x\,dx$

(b) $\displaystyle\int_{-2}^{-1} \frac{x^4 + x - 3}{x^3}\,dx$

(c) $\displaystyle\int_0^{\pi/4} \sec x \tan x + \sec^2 x\,dx$

(d) $\displaystyle\int_0^{\pi} \sin^2\left(\frac{x}{2}\right)\,dx$

10. This exercise outlines an alternate proof of Part 2 of the FTC which doesn't use Part 1. The approach we will use is quite similar to the approach outlined for computing the integral of cos in Exercise 5.6.10.

Let $a, b \in \mathbb{R}$ with $a < b$ be given, and let f be a real function which is continuous on $[a, b]$. Suppose that $F : [a, b] \to \mathbb{R}$ is continuous and satisfies $F'(x) = f(x)$ for all $x \in (a, b)$, so F is an antiderivative of f on (a, b). Since f is continuous on $[a, b]$, f is integrable on $[a, b]$, so we know that there exists only one number $I \in \mathbb{R}$ which satisfies

$$\int_a^b s \leq I \leq \int_a^b t$$

for all upper step functions t for f on $[a, b]$ and all lower step functions s for f on $[a, b]$. In particular, $I = \displaystyle\int_a^b f$. To show that $I = F(b) - F(a)$, we will show that $F(b) - F(a)$ satisfies the inequality above for every suitable t and s.

(a) Let t and s be respectively upper and lower step functions for f on $[a, b]$, compatible with a common partition $P = \{x_0, x_1, \ldots, x_n\}$. Let i be given from 1 to n, and suppose that $s(x) = a_i$ and $t(x) = b_i$ for all $x \in (x_{i-1}, x_i)$. Use the MVT for derivatives to prove that for all i from 1 to n,

$$a_i(x_i - x_{i-1}) \leq F(x_i) - F(x_{i-1}) \leq b_i(x_i - x_{i-1})$$

(b) Use part (a) to show that

$$\int_a^b s \leq F(b) - F(a) \leq \int_a^b t$$

for all s and t as described in Part (a). Thus, we have proven

$$I = \int_a^b f(x)\,dx = F(b) - F(a)$$

Chapter 6

Techniques of Integration

At the end of the last chapter, we introduced the Fundamental Theorem of Calculus (FTC), which proves in essence that derivatives and integrals "undo each other". Most importantly, it means that we have an easier way of finding integrals: when f is continuous on $[a, b]$, rather than computing

$$\int_a^b f(x)\,dx$$

with upper and lower integrals, we instead look for an antiderivative F such that $F' = f$ on $[a, b]$ and compute $F(b) - F(a)$.

For instance, if $F(x) = x \sin x$ for all $x \in \mathbb{R}$, then $F'(x) = x \cos x + \sin x$ by the Product Rule. It follows that

$$\int_a^b (x \cos x + \sin x)\,dx = x \sin x\Big|_a^b = b \sin b - a \sin a$$

for any $a, b \in \mathbb{R}$. We can also summarize this with Leibniz notation by writing

$$\int (x \cos x + \sin x)\,dx = x \sin x + C$$

where C is an arbitrary constant.

In Chapter 4, we proved many laws for computing derivatives, such as the Product Rule and the Chain Rule. In order to compute integrals, i.e. find antiderivatives, we want to develop techniques which "undo" the derivative laws. Unfortunately, it is harder to undo the derivative laws than to apply those laws. For instance, suppose we want to integrate

$$f(x) = x \cos x + \sin x$$

to get an antiderivative F. Perhaps we suspect that the Product Rule was used to obtain F'. Hence, F has the form gh for some functions g and h, and $f = F' = g'h + gh'$. However, it's quite tricky to determine what g and h could be. Also, if you try to determine F by integrating $x \cos x$ and $\sin x$ separately, then you probably won't be able to obtain $x \cos x$ via the Product Rule for derivatives.

In fact, it turns out that there are functions which are continuous at almost every point, and thus the FTC guarantees that they have antiderivatives, but the antiderivatives cannot be built out of the functions commonly used in mathematics. Informally, we say that such functions don't have *elementary* antiderivatives[1]. A couple well-known examples of functions which have been proven to not have elementary antiderivatives are

$$\frac{\sin x}{x} \qquad \text{and} \qquad \cos(x^2)$$

Although we know antiderivatives of $\sin x$ and x, namely $-\cos x$ and $(1/2)x^2$ respectively, there is no "Quotient Rule" for antiderivatives that lets us obtain an elementary antiderivative of $(\sin x)/x$ out of antiderivatives for the individual parts.

In summary, integration is a much harder task than differentiation. We will develop some techniques/rules for finding antiderivatives, but these techniques will not be foolproof, and several techniques may need to be tried before one is found that works. These techniques tend to work quite well with some specific types of integrals which occur frequently, such as products of trigonometric functions, so we will also devote attention to some of those types of integrals.

6.1 The Substitution Rule

For our first major rule of integration, we aim to undo the Chain Rule. The Chain Rule tells us how to differentiate composite functions. For instance, if $f(x) = \sqrt{x}$ for $x \geq 0$ and $g(x) = 3x+1$ for $x \in \mathbb{R}$, then $(f \circ g)(x) = \sqrt{3x+1}$, and

$$\frac{d}{dx}\left(\sqrt{3x+1}\right) = (f \circ g)'(x) = f'(g(x))g'(x) = \frac{3}{2\sqrt{3x+1}}$$

[1] Proofs of this use the theory of *differential fields*, which will not be covered here.

Therefore, $\sqrt{3x+1}$ is an antiderivative of $\dfrac{3}{2\sqrt{3x+1}}$, and in Leibniz notation we write this as

$$\int \frac{3}{2\sqrt{3x+1}}\, dx = \sqrt{3x+1} + C$$

More generally, suppose that F is an antiderivative of f, which means $F' = f$. Thus, $(F \circ g)'(x) = F'(g(x))g'(x) = f(g(x))g'(x)$. Therefore, we have

$$\int f(g(x))g'(x)\, dx = F(g(x)) + C$$

For another way to see this, let's denote $g(x)$ by a separate variable, say we write $u = g(x)$. Then $\dfrac{du}{dx} = g'(x)$. By treating du and dx as differentials (see Section 4.5), informally we can write $du = g'(x)\, dx$. Hence, we informally replace $g(x)$ by u and $g'(x)\, dx$ by du in our integration problem, yielding

$$\int f(g(x))g'(x)\, dx = \int f(u)\, du$$

Now, this looks like an integration problem in u! As f has antiderivative F, we have

$$\int f(u)\, du = F(u) + C$$

which equals $F(g(x)) + C$ when we put the result back in terms of x.

This is the idea of the *Substitution Rule*: substitute a new variable u in place of $g(x)$. When we do this, we turn our problem of integrating $(f \circ g) \cdot g'$ into the simpler problem of integrating f! After we are done integrating f, we plug $g(x)$ back in place of u to obtain a final answer in terms of the original variable x. Sometimes, this method is called "*u-substitution*" since the letter u traditionally denotes the function used in the substitution.

Example 6.1:

To see an example of substitution in practice, let's compute

$$\int 2x \sin\left(x^2\right)\, dx$$

Since our integrand involves the composition of sin and x^2, we try letting u equal the inner part of the composition, i.e. $u = x^2$. Then $du = 2x\, dx$.

Therefore,

$$\int 2x \sin\left(x^2\right)\, dx = \int \sin\left(x^2\right)(2x\,dx) = \int \sin u\,du$$

At this point, our problem reduces to finding an antiderivative for sin. Since $-\cos u$ has derivative $\sin u$, we find

$$\int \sin u\,du = -\cos u + C = -\cos\left(x^2\right) + C$$

Note that we can check our answer quite easily: to see that $-\cos(x^2)$ is an antiderivative of $2x\sin(x^2)$, we compute

$$\frac{d}{dx}\left(-\cos\left(x^2\right)\right) = -\left(-\sin\left(x^2\right)\right)\cdot(2x) = 2x\sin\left(x^2\right)$$

as desired. □

More formally, using the remarks at the beginning of this section, you can prove the following theorem to justify our work:

Theorem 6.2 (Substitution Rule). *Let f, g be real functions such that f is continuous and g' exists and is continuous. Let F be an antiderivative of f. Then*

$$\int f(g(x))g'(x)\,dx = \int f(u)\,du = F(u) + C = F(g(x)) + C$$

where we perform the substitution $u = g(x)$ and $du = g'(x)\,dx$.

The substitution rule is a good technique to try when integrating compositions, if we can write the composition in the form $f(g(x))\cdot g'(x)$. Sometimes this takes a little extra work, as this example shows:

Example 6.3:

Let's try to find

$$\int \sqrt{2x+4}\,dx$$

This integrand is a composition of $f(x) = \sqrt{x}$ and $g(x) = 2x + 4$. Since it's easier to integrate f on its own, we try letting $u = 2x+4$. Then $du = 2\,dx$.

Technically, substitution tells us how to integrate $f(g(x))g'(x)$, which is $2\sqrt{2x+4}$, but our current integrand does not have that factor of 2 present. To fix this, we can multiply and divide by 2 and pull a $1/2$ factor outside of the integral. This gives us

$$\frac{1}{2}\int 2\sqrt{2x+4}\,dx = \frac{1}{2}\int \sqrt{u}\,du = \frac{1}{2}\frac{u^{3/2}}{3/2} + C = \frac{1}{3}(2x+4)^{3/2} + C$$

Commonly, instead of multiplying and dividing by 2, we instead write $du = 2\,dx$ as $dx = du/2$ and compute

$$\int \sqrt{2x+4}\,dx = \int \sqrt{u}\left(\frac{du}{2}\right) = \frac{1}{2}\int \sqrt{u}\,du$$

as before. Formally, this computation is shorthand for the multiplying and dividing by 2 that occurs above. □

Example 6.4:

Here, let's compute

$$\int \cos^2 x \, \sin x \, dx$$

In this problem, you may view $\cos^2 x$ as a composition where cos is the inner function. Also, you can notice that the derivative of cos is $-\sin$, which is a constant times the $\sin x$ term we have in our integrand. Thus, we choose

$$u = \cos x \quad du = -\sin x\,dx \quad dx = \frac{-du}{\sin x}$$

This gives us

$$\begin{aligned}\int \cos^2 x \sin x\,dx &= \int u^2 \sin x\left(\frac{-du}{\sin x}\right)\\ &= \int -u^2\,du = \frac{-u^3}{3} + C = \frac{-\cos^3 x}{3} + C\end{aligned}$$

where we once again use shorthand instead of first multiplying and dividing our integrand by -1. □

When performing a substitution of $u = g(x)$, we need the new integrand to be a function of u alone. However, sometimes when we try a u-substitution, we still have copies of x left over after plugging in u and du. The substitution rule can still apply if we manage to convert those remaining x's into an expression in terms of u, as the following example shows:

Example 6.5:

Let's try and compute

$$\int x(x+2)^{5/2}\,dx$$

To handle the composition, we try $u = x+2$ and $du = dx$. However, when we plug in u and du, we obtain the expression

$$\int xu^{5/2}\,du$$

This expression isn't well-defined by our definition of $\int$.

In this situation, we note that x can be written in terms of u: $x = u - 2$. When we plug that into our previous expression using both x's and u's, we obtain

$$\begin{aligned}
\int (u-2)u^{5/2}\,du &= \int (u^{7/2} - 2u^{5/2})\,du \\
&= \frac{u^{9/2}}{9/2} - 2\left(\frac{u^{7/2}}{7/2}\right) + C \\
&= \frac{2}{9}(x+2)^{9/2} - \frac{4}{7}(x+2)^{7/2} + C
\end{aligned}$$

More formally, to justify this process by the substitution rule, we are writing our original integrand in the form

$$\int x(x+2)^{5/2}\,dx = \int ((x+2)-2)(x+2)^{5/2}\,dx = \int f(g(x))g'(x)\,dx$$

where $f(u) = (u-2)u^{5/2}$ and $g(x) = x+2$. However, as shorthand, we may "solve for x in terms of u" like we did above. $\square$

Substitution can be used to simplify many integrals or manipulate them into alternate useful forms. As with any skill, you will become better at recognizing how to use it through practice.

Substitution and Definite Integrals

Since the substitution technique helps find antiderivatives, it can also be used to help calculate definite integrals. For instance, in Example 6.4, we found

$$\int \cos^2 x \sin x\,dx = \frac{-\cos^3 x}{3} + C$$

and thus, by the Fundamental Theorem,

$$\int_0^{\pi/2} \cos^2 x \sin x \, dx = \left. \frac{-\cos^3 x}{3} \right|_0^{\pi/2} = \frac{\cos^3 0 - \cos^3(\pi/2)}{3} = \frac{1}{3}$$

However, if we try to perform the substitution $u = \cos x$ in the definite integral leaving the limits alone, then we end up with the incorrect statement

$$\int_0^{\pi/2} \cos^2 x \sin x \, dx = \int_0^{\pi/2} -u^2 \, du = \left. \frac{-u^3}{3} \right|_0^{\pi/2} = \frac{-\pi^3}{24}$$

The problem here is that the limit values are values for x, not values for u.

How should we perform this calculation instead? One way is to write the limits of integration in a way that reminds us they are limits for x, not u. This would be written as

$$\begin{aligned} \int_0^{\pi/2} \cos^2 x \sin x \, dx &= \int_{x=0}^{x=\pi/2} -u^2 \, du \\ &= \left. \frac{-u^3}{3} \right|_{x=0}^{x=\pi/2} = \left. \frac{-\cos^3 x}{3} \right|_0^{\pi/2} \end{aligned}$$

By writing "$x =$" in the limits, we remind ourselves that we cannot plug in those values until we have returned our expression to a function of x. However, this is clumsy notation, and we have to do the extra work to going back to x.

Instead, when we change variables from x to u, we should also change the limit values to values for u! When $x = 0$, we have $u = \cos 0 = 1$, and when $x = \pi/2$, $u = \cos(\pi/2) = 0$. Thus, we write our work as

$$\int_0^{\pi/2} \cos^2 x \sin x \, dx = \int_1^0 -u^2 \, du = \left. \frac{-u^3}{3} \right|_1^0$$

Note that we never even have to return to the original variable x!

To justify this limit change more formally, we have this theorem:

Theorem 6.6 (Substitution with Definite Integrals). *Let $a, b \in \mathbb{R}$ be given with $a \leq b$, and let f, g be real functions. If S is the range of g on $[a, b]$, then suppose that f is continuous on S and g' exists and is continuous on $[a, b]$. Let F be an antiderivative of f on S. Then*

$$\int_a^b f(g(x))g'(x) \, dx = \int_{g(a)}^{g(b)} f(u) \, du = F(u)|_{g(a)}^{g(b)} = F(g(b)) - F(g(a))$$

where we perform the substitution $u = g(x)$ and $du = g'(x)\,dx$.

Strategy. Essentially, we are just saying that when you change from x to $u = g(x)$, you change the limits a and b to $g(a)$ and $g(b)$ respectively. To see why this is valid, we will evaluate the integrals of $(f \circ g) \cdot g'$ and of f using the Fundamental Theorem. We already have an antiderivative of f: F. The Substitution Rule shown previously gives us an antiderivative of $(f \circ g) \cdot g'$: $F \circ g$. Using these antiderivatives, we can check that we get the same answer either way we compute this integral.

Proof. Let a, b, f, g, F be given as described. First, we address some subtle details concerning where our functions are continuous (i.e. we check to see whether the FTC is allowed). Since g is continuous on $[a, b]$, S is a closed bounded interval containing $g(a)$ and $g(b)$. (See Exercise 3.12.3.) Thus, by the Composition Limit Theorem, $(f \circ g) \cdot g'$ is continuous on $[a, b]$. Also, since f is continuous on S, in particular it is continuous from $g(a)$ to $g(b)$.

Therefore, we may use the FTC to compute each of the integrals

$$\int_a^b f(g(x))g'(x)\,dx \qquad \text{and} \qquad \int_{g(a)}^{g(b)} f(u)\,du$$

Since F is an antiderivative of f, the integral on the right equals

$$F(u)\Big|_{g(a)}^{g(b)} = F(g(b)) - F(g(a))$$

By the Chain Rule, $(F \circ g)' = (F' \circ g) \cdot g' = (f \circ g) \cdot g'$, so $F \circ g$ is an antiderivative of $(f \circ g) \cdot g'$. Therefore, the integral on the left equals

$$F(g(x))\Big|_a^b = F(g(b)) - F(g(a))$$

Our two expressions are equal. ■

Example 6.7:

Let's compute

$$\int_{\pi^2/4}^{\pi^2} \frac{\sin\sqrt{x}}{\sqrt{x}}\,dx$$

Since there is a composition of sin and square root, we try $u = \sqrt{x}$, so $du = \dfrac{1}{2\sqrt{x}}\,dx$. Let's rewrite this as $dx = 2\sqrt{x}\,du$ to make it easy to plug in for dx. Also, when $x = \pi^2/4$, $u = \pi/2$, and when $x = \pi^2$, $u = \pi$. Thus, we compute

$$\begin{aligned}\int_{\pi^2/4}^{\pi^2} \frac{\sin\sqrt{x}}{\sqrt{x}}\,dx &= \int_{\pi/2}^{\pi} \frac{\sin u}{\sqrt{x}}(2\sqrt{x}\,du) \\ &= \int_{\pi/2}^{\pi} 2\sin u\,du = -2\cos u\big|_{\pi/2}^{\pi} \\ &= -2(\sin\pi - \sin(\pi/2)) = 2\end{aligned}$$

Note that the dx or du term tells us which variable is being used for the limits, and it helps us perform substitution correctly. □

As a final note on substitution, we can even use substitution to prove some theorems, because substitution gives us a way of writing integrals in other forms. For instance, if f is continuous on $[a+c, b+c]$, then we see that

$$\int_a^b f(x+c)\,dx = \int_{a+c}^{b+c} f(u)\,du$$

by substituting $u = x + c$! This gives us a quick and easy proof of the Shifting Property for integrals (Theorem 5.28) when f is continuous. We provide another example.

Example 6.8:

Let's suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous and odd, meaning that $f(-x) = -f(x)$ for all $x \in \mathbb{R}$. Define $F : \mathbb{R} \to \mathbb{R}$ by $F(x) = \int_0^x f(t)\,dt$. The Fundamental Theorem shows that F is an antiderivative of f. Let's prove that F is even, i.e. $F(-x) = F(x)$ for all $x \in \mathbb{R}$.

Let $x \in \mathbb{R}$ be given. We'd like to rewrite

$$F(-x) = \int_0^{-x} f(t)\,dt$$

in terms of $F(x)$. Since we know that f is odd, how can we use this fact to our advantage? To do so, we'd like an expression where f is applied to something with a minus sign. Because of this, we try the substitution $u = -t$ and $du = -dt$. When $t = 0$, we have $u = 0$, and when $t = -x$, we have $u = x$. Also, we have $t = -u$ and $dt = -du$, so we find

$$\int_0^{-x} f(t)\,dt = \int_0^{x} f(-u)(-du)$$

Because f is odd, $f(-u) = -f(u)$, so we obtain

$$\int_0^x f(-u)(-du) = \int_0^x f(u)\,du$$

Recall that the variable of integration can be anything which isn't already used somewhere else. Thus, by the definition of F, our final integral is equal to $F(x)$, proving $F(-x) = F(x)$. $\square$

6.2 Integration By Parts

Our next rule for integration aims to undo the Product Rule. As an example, if $f(x) = x$ and $g(x) = \sin x$ for $x \in \mathbb{R}$, then $(fg)(x) = x \sin x$ and

$$\frac{d}{dx}(x \sin x) = f'(x)g(x) + f(x)g'(x) = \sin x + x \cos x$$

Since $\sin x$ has antiderivative $-\cos x$, this means

$$\frac{d}{dx}(x \sin x - \cos x) = x \cos x$$

and therefore

$$\int x \cos x\,dx = x \sin x - \cos x + C$$

More generally, suppose that f and g are differentiable functions. The Product Rule tells us that $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$. Therefore, by integrating, we have

$$f(x)g(x) + C = \int f'(x)g(x)\,dx + \int f(x)g'(x)\,dx$$

This is useful because this formula allows us to write the integral of fg' in terms of the integral of $f'g$ and vice versa. More specifically, we have

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx$$

Hence, we have turned the problem of integrating fg' into the problem of integrating $f'g$. (There is no need to write the arbitrary constant of integration C here, because that constant will appear after integrating $f'g$. We only need to write that constant after we have finished finding an antiderivative of fg'.)

This is the main formula underlying the technique called *integration by parts*. When performing integration by parts, we try to integrate a product by expressing it in the form fg' (we say f and g' are the *parts* of the integrand) and using the formula above. This can be very useful when the form of $f'g$ is simpler than that of fg'. For instance, in the example above, $(fg')(x)$ is $x \cos x$ and $(f'g)(x)$ is $\sin x$, so $f'g$ has an antiderivative we already know.

Example 6.9:

To see this technique in practice, let's compute

$$\int x(x+2)^{5/2}\,dx$$

We showed how to use substitution to evaluate this in Example 6.5, but we can also use parts to evaluate this. We'd like to write this integrand in the form fg', such that the form of $f'g$ is much simpler. This suggests that we choose f and g so that

$$f(x) = x \qquad g'(x) = (x+2)^{5/2}$$

For g, we can take any antiderivative g', which can be found by a simple substitution of $u = x + 2$ (we leave out the details). Thus, we have

$$f'(x) = 1 \qquad g(x) = \frac{2}{7}(x+2)^{7/2}$$

(Note that substitutions which are this simple are not usually shown with all details, because it is very easy to differentiate g to check your answer.) Note that f' is much simpler than f (at the very least, it's a polynomial of smaller degree than f), and g has about the same level of complexity as g'. Hence, it is easier to integrate $f'g$.

By integration by parts, we have

$$\begin{aligned}\int x(x+2)^{5/2}\,dx &= f(x)g(x) - \int f'(x)g(x)\,dx \\ &= \frac{2}{7}x(x+2)^{7/2} - \frac{2}{7}\int (x+2)^{7/2}\,dx\end{aligned}$$

This last integral can be computed by a substitution of $u = x + 2$, which gives us the final answer of

$$\frac{2}{7}x(x+2)^{7/2} - \frac{2}{7}\left(\frac{2}{9}(x+2)^{9/2}\right) + C = \frac{2}{7}x(x+2)^{7/2} - \frac{4}{63}(x+2)^{9/2} + C$$

(Note that we only need to include the arbitrary constant C of integration at the very end, once we have found some antiderivative of our original integrand $x(x+2)^{5/2}$.)

This answer looks quite different from the answer in Example 6.5:

$$\frac{2}{9}(x+2)^{9/2} - \frac{4}{7}(x+2)^{7/2} + C$$

However, it turns out to be the same function. If we take the answer from Example 6.5 and subtract the answer we found in this example, and we pull out common factors, then we get

$$\frac{2}{7}(x+2)^{9/2} - \frac{4}{7}(x+2)^{7/2} - \frac{2}{7}x(x+2)^{7/2} = \frac{2}{7}(x+2)^{7/2}\left((x+2) - 2 - x\right) = 0$$

This illustrates that integration problems can show two quantities are equal, despite them not looking very similar at first glance. □

Remark. When performing integration by parts, it can make a big difference how the parts are chosen. If we try to perform integration by parts in the previous example with the choices

$$f(x) = (x+2)^{5/2} \qquad g'(x) = x$$

leading to

$$f'(x) = \frac{5}{2}(x+2)^{3/2} \qquad g(x) = \frac{x^2}{2}$$

then it is harder to integrate $f'g$ than to integrate fg'. In essence, we chose our parts in the previous example so as to "eliminate the x's" and only have powers of $(x+2)$ left over.

Here is the formal theorem justifying our steps, which you should try to prove for yourself.

Theorem 6.10 (Integration By Parts). *Let f, g be real functions such that f' and g' exist and are continuous. If H is an antiderivative of $f'g$, then*

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx = f(x)g(x) - H(x) + C$$

In particular, if f' and g' are continuous on an interval $[a, b]$, then

$$\begin{aligned} \int_a^b f(x)g'(x)\,dx &= f(x)g(x)\Big|_a^b - \int_a^b f'(x)g(x)\,dx \\ &= \left(f(x)g(x) - H(x)\right)\Big|_a^b \end{aligned}$$

Frequently, a notation similar to that for u-substitution is used with integration by parts. We write

$$u = f(x) \qquad v = g(x) \qquad du = f'(x)\,dx \qquad dv = g'(x)\,dx$$

(Hence, like in u-substitution, we treat du and dv as if they are differentials.) With this notation, the formula for integration by parts becomes

$$\int u\,dv = uv - \int v\,du$$

This notation helps make clear that when performing integration by parts, one part (the "u" part) is differentiated and the other (the "dv" part) is integrated.

Example 6.11:

Here, let's try and compute

$$\int x^2 \sin x\,dx$$

This is a good candidate for integration by parts because there is a product of terms, where one of them (x^2) becomes simpler when differentiated and the other ($\sin x$) doesn't become uglier when integrated. We choose

$$u = x^2 \qquad dv = \sin x\,dx$$

so that

$$du = 2x\,dx \qquad v = -\cos x$$

Hence, our problem becomes

$$uv - \int v\,du = -x^2 \cos x + 2 \int x \cos x\,dx$$

The integral of $x \cos x$ was done at the beginning of this section, using parts with $u = x$ and $dv = \cos x\,dx$. Our final answer is

$$-x^2 \cos x + 2\,(x \sin x - \cos x) + C = -x^2 \cos x + 2x \sin x - 2 \cos x + C$$

□

As Example 6.11 shows, if a polynomial is present in the integrand, then we frequently choose it to be u. This is because differentiating a polynomial lowers its degree. In fact, for degree d polynomials, you often find yourself performing integration by parts d times. However, this is only a guideline and not a strict requirement, because after all, you still need to make sure to choose dv to be something you can actually integrate.

Example 6.11 also shows that sometimes integration by parts should be performed multiple times. This is a worthwhile thing to try as long as each time integration by parts is performed, the resulting problem does not look worse. Here's an example where this is useful.

Example 6.12:

Let's integrate

$$\int \sin(2x) \cos x\,dx$$

This is not well-suited for substitution, because if we try to substitute $u = 2x$, then the $\cos x$ term becomes $\cos(u/2)$, so we don't make our problem any simpler. However, we can try parts with

$$u = \sin(2x) \qquad dv = \cos x\,dx$$

This gives us $du = 2\cos(2x)\,dx$ and $v = \sin x$, so

$$\int \sin(2x) \cos x\,dx = \sin(2x) \sin x - 2 \int \cos(2x) \sin x\,dx$$

This new integration problem is like the old one, but the roles of cos and sin are switched. If we try parts again with

$$u = \cos(2x) \qquad dv = \sin x\,dx$$

then $du = -2\sin(2x)$, $v = -\cos x$, and thus

$$\begin{aligned}\int \sin(2x)\cos x\,dx &= \sin(2x)\sin x \\ &- 2\left(-\cos(2x)\cos x - 2\int \sin(2x)\cos x\,dx\right) \\ &= \sin(2x)\sin x + 2\cos(2x)\cos x + 4\int \sin(2x)\cos x\,dx\end{aligned}$$

We have now returned full-circle to our original problem. This is useful, because we can collect the terms of $\int \sin(2x)\cos x\,dx$ to the same side, and we get

$$(-3)\int \sin(2x)\cos x\,dx = \sin(2x)\sin x + 2\cos(2x)\cos x$$

By dividing by -3 and adding in our arbitrary constant of integration, we find

$$\int \sin(2x)\cos x\,dx = \frac{1}{3}\sin(2x)\sin x + \frac{2}{3}\cos(2x)\cos x + C$$

□

Remark. Unfortunately, there are times when you return to the original integral when doing integration by parts, but this doesn't help you. For instance, suppose we try to integrate

$$\int \frac{1}{x}\,dx$$

by parts, using $u = 1/x$ and $dv = dx$. We obtain $du = -1/x^2\,dx$ and $v = x$, leading to

$$\int \frac{1}{x}\,dx = 1 + \int \frac{x}{x^2}\,dx = 1 + \int \frac{1}{x}\,dx$$

At this point, it is tempting to cancel the integral from each side and obtain $0 = 1$, but that isn't valid. The issue is that when integrating by parts, there is an integration constant C behind the scenes, which must be introduced when we finally obtain an antiderivative. To be precise, let's suppose $A(x)$ is an antiderivative of $1/x$. Thus, the integral on the left is $A(x) + C$ for some constant C, and the integral on the right is $A(x) + D$ for some constant D (which doesn't have to equal C). Hence, we have

$$A(x) + C = 1 + A(x) + D \qquad \leftrightarrow \qquad C = 1 + D$$

This equation isn't a contradiction, since it can be satisfied for any arbitrary choice of C by picking $D = C - 1$. However, this doesn't give us any information about $A(x)$, so we didn't learn anything about an antiderivative of $1/x$.

If we try to solve this problem with definite integrals instead, say from a to b where $0 < a < b$ (hoping to get rid of issues involving constants like C), then we get

$$\int_a^b \frac{1}{x}\,dx = 1\Big|_a^b + \int_a^b \frac{1}{x}\,dx$$

Note that $1|_a^b$ is equal to $1 - 1$, which is 0. Thus, once again, we don't get a contradiction, but we also don't learn anything useful.

You can try other choices of parts by writing $1/x$ as $(x^n)(x^{-(n+1)})$ for any $n \in \mathbb{N}^*$, but all of these choices will run into the same problem. In fact, in the next chapter, we'll study an antiderivative of $1/x$ called the *natural logarithm* of x, and we'll discover important properties of the logarithm that set it apart from other functions we have studied so far.

Example 6.13:

As one last example of the use of integration by parts, we consider integrating

$$\int \sin^n x\,dx$$

when $n \in \mathbb{N}^*$ and $n \geq 2$. (We already know antiderivatives when $n = 0$ or $n = 1$.)

Since $\sin^n x$ is a product of n sines, and we only know how to integrate one power of $\sin x$, let's make $dv = \sin x\,dx$, so that $u = \sin^{n-1} x$ has all the other powers. Thus, $v = -\cos x$ and $du = (n-1)\sin^{n-2} x \cos x\,dx$ by the Chain Rule. This gives

$$\int \sin^n x\,dx = -\sin^{n-1} x \cos x + (n-1)\int \sin^{n-2} x \cos^2 x\,dx$$

At first, this looks worse than the original problem with which we started. However, we know that we can write $\cos^2 x$ as $1 - \sin^2 x$, so that we get

$$\begin{aligned}
\int \sin^n x\,dx &= -\sin^{n-1} x \cos x + (n-1)\int \sin^{n-2} x(1 - \sin^2 x)\,dx \\
&= -\sin^{n-1} x \cos x + (n-1)\int \sin^{n-2} x\,dx \\
&- (n-1)\int \sin^n x\,dx
\end{aligned}$$

By collecting the integrals of $\sin^n x\,dx$ on the same side, and dividing by n, we obtain

$$\int \sin^n x\,dx = -\frac{\sin^{n-1} x\cos x}{n} + \frac{n-1}{n}\int \sin^{n-2} x\,dx$$

Therefore, we have shown how to integrate $\sin^n x$ in terms of the integral for $\sin^{n-2} x\,dx$. By repeating this process, i.e. using this formula recursively, we can eventually reduce our problem down to integrating the 0^{th} or 1^{st} power of $\sin x$. Hence, this kind of formula is often called a *reduction formula*, since it is an inductive formula for reducing the complexity of an integral. □

6.3 Exercises

Compute the indefinite or definite integrals in Exercises 1 through 12 by using substitution.

1. $\displaystyle\int \frac{x}{(x^2+1)^2}\,dx$

2. $\displaystyle\int_0^\pi \sin x\sqrt{\cos x+3}\,dx$

3. $\displaystyle\int \frac{\sin(x^{1/3})}{x^{2/3}}\,dx$

4. $\displaystyle\int_{-1}^4 x\sqrt{x+5}\,dx$

5. $\displaystyle\int x^3\sqrt{x^2+5}\,dx$

   (Hint: Use the same idea as the previous problem.)

6. $\displaystyle\int_0^1 \frac{x+1}{(x^2+2x+2)^3}\,dx$

7. $\displaystyle\int \sin x\cos^2 x\,dx$

8. $\displaystyle\int_0^\pi \sin^3 x\,dx$

   (Hint: Relate this to the previous problem.)

9. $\displaystyle\int \sin(3x)\cos(3x)\,dx$

10. $\displaystyle\int \frac{\sin x+\cos x}{(\sin x-\cos x)^2}\,dx$

11. $\displaystyle\int \frac{1}{\sqrt{(1/x^2)-1}}\,dx$

    (Hint: Make a common denominator.)

12. $\displaystyle\int_0^{\sqrt{3}} (x^2+1)^{-3/2}\,dx$

(Hint: Use $u = (x^2+1)^{-1/2}$, solve for x in terms of u, and relate this to the previous problem. Also, note that $x \geq 0$.)

Compute the integrals in Exercises 13 through 17 by using parts (occasionally substitution will also be useful to help integrate a part).

13. $\displaystyle\int_0^{\pi} (x-2)^2 \cos x\,dx$

14. $\displaystyle\int (3x+1)\sqrt{x+4}\,dx$

15. $\displaystyle\int x \sin x \cos x\,dx$

16. $\displaystyle\int \sin(ax)\cos(ax)\,dx$ where a is a nonzero constant

17. $\displaystyle\int \sin(ax)\cos(bx)\,dx$ where a, b are distinct nonzero constants

For Exercises 18 through 20, let $L(x) = \displaystyle\int_1^x \frac{dt}{t}$. Thus, $L(x)$ is an antiderivative of $1/x$.

18. (a) Show that
$$\int \tan x\,dx = -L(\cos x) + C$$
(Hint: Write things in terms of $\sin x$ and $\cos x$.)
(b) Show that
$$\int \cot x\,dx = L(\sin x) + C$$

19. (a) Compute $\displaystyle\int \sec x\,dx$ in terms of L. (Hint: Multiply and divide by $\sec x + \tan x$ and use Exercise 4.4.10 to help with a substitution.)
(b) Compute $\displaystyle\int \csc x\,dx$.

20. By using integration by parts and substitution, show that

$$\int \arctan x\,dx = x\arctan x - \frac{1}{2}L(x^2+1) + C$$

Exercise 4.6.21 will probably be useful.

21. Use the reduction formula from Example 6.13 to compute the following integrals:

$$\int_0^{\pi/2} \sin^2 x\,dx \qquad \int_0^{\pi/2} \sin^3 x\,dx \qquad \int_0^{\pi/2} \sin^4 x\,dx$$

22. Use parts to prove the reduction formula

$$\int \cos^n x\,dx = \frac{\cos^{n-1} x \sin x}{n} + \frac{n-1}{n}\int \cos^{n-2} x\,dx$$

for all $n \in \mathbb{N}$ with $n \geq 2$.

23. Use the result from the previous exercise to compute the following:

$$\int_0^{\pi/2} \cos^2 x\,dx \qquad \int_0^{\pi/2} \cos^3 x\,dx \qquad \int_0^{\pi/2} \cos^4 x\,dx$$

24. This problem shows five distinct ways of computing $\displaystyle\int \sin x \cos x\,dx$!
    (a) Show two different ways of evaluating this integral using substitution.
    (b) The two answers from part (a) don't look the same, but they differ by a constant. What is that constant?
    (c) Show two different ways of evaluating our integral by parts.
    (d) Using an identity for $\sin 2x$, show a fifth way of evaluating our integral.

25. (a) Use integration by parts to show that

$$\int \sqrt{1-x^2}\,dx = x\sqrt{1-x^2} + \int \frac{x^2}{\sqrt{1-x^2}}\,dx$$

(b) In the second integral from the previous part, write x^2 as $(x^2 - 1) + 1$ and show that

$$\int \sqrt{1-x^2}\, dx = \frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\int \frac{1}{\sqrt{1-x^2}}\, dx$$

(Note that by Example 4.33, the last integral is $\arcsin x + C$).

26. Use substitution to prove that for all $b > 0$,

$$\int_b^1 \frac{dx}{1+x^2} = \int_1^{1/b} \frac{dx}{1+x^2}$$

(You do not need to know an antiderivative of $1/(1+x^2)$ to do this problem, although it turns out that $\arctan x$ is one such antiderivative.)

27. If $m, n \in \mathbb{N}^*$, then show that

$$\int_0^1 x^m(1-x)^n\, dx = \int_0^1 x^n(1-x)^m\, dx$$

(This integral occurs in statistics in connection with the Beta distribution.)

28. (a) Show that if f is any continuous function on $[0,1]$, then

$$\int_0^\pi x f(\sin x)\, dx = \frac{\pi}{2}\int_0^\pi f(\sin x)\, dx$$

(Hint: Choose $u = \pi - x$.)

(b) Use part (a) to prove that

$$\int_0^\pi \frac{x\sin x}{1+\cos^2 x}\, dx = \pi\int_0^1 \frac{dx}{1+x^2}$$

29. (a) Let $a \in \mathbb{R}$ be given with $a > 0$. The integral $\int_0^a \sqrt{a^2-x^2}\, dx$ represents the area of a shape. What is that shape, and what is the area?

(b) Prove the reduction formula

$$\int (a^2-x^2)^n\, dx = \frac{x(a^2-x^2)^n}{2n+1} + \frac{2a^2 n}{2n+1}\int (a^2-x^2)^{n-1}\, dx$$

for all $n > 1$. (Hint: After using parts, write x^2 as $(x^2-a^2)+a^2$.)

(c) Use parts (a) and (b) to compute $\int_0^a (a^2-x^2)^{5/2}\, dx$.

6.4 Trigonometric Integrals

In this section, we'll present some strategies for dealing with integrals involving trigonometric functions. These kinds of integrals occur frequently (as we'll see in the next section). In most of these problems, the overall idea is to apply identities in order to eventually make a u-substitution.

Sines and Cosines

First, let's consider integrands consisting of products of sines and cosines.

Example 6.14:

As an example to illustrate the main idea, let's consider

$$\int \cos^3 x\,dx$$

In the exercises from the last section, we showed one way to integrate this by using parts, but it involves a complicated reduction formula. Let's show a quicker way to integrate this by substitution.

Since we have a composition (a power of $\cos x$), we could consider substituting $u = \cos x$. That doesn't currently seem applicable, because we get $du = -\sin x\,dx$, and there is no factor of $\sin x$ lying around to be used for du. However, this gives us the idea that we'd like to have a mixture of sine and cosine factors in our integrand, so that we have something to take care of du. To accomplish this, we use the Pythagorean identity

$$\sin^2 x + \cos^2 x = 1$$

to rewrite our integrand as

$$\int \cos^3 x\,dx = \int \cos x(1 - \sin^2 x)\,dx$$

By doing this, we now have two powers of $\sin x$, suggesting $u = \sin x$ instead, and fortunately we have a factor of $\cos x$ available to take care of $du = \cos x\,dx$. Thus,

$$\int \cos x(1 - \sin^2 x)\,dx = \int (1 - u^2)\,du = u - \frac{u^3}{3} + C = \sin x - \frac{\sin^3 x}{3} + C$$

□

To summarize the tactic from Example 6.14, when we have powers of sine and/or cosine, we want to use one of the substitutions

$$u = \sin x \qquad \text{or} \qquad u = \cos x$$

to evaluate the integral. In order to do this, we need a spare factor of $\sin x$ or $\cos x$ available for du. We then want to write the rest of the integrand in terms of u, and we use the Pythagorean identity to convert between pairs of sines and pairs of cosines as necessary.

Example 6.15:

As another example, consider

$$\int_0^{\pi/2} \cos^2 x \sin^5 x \, dx$$

We want to substitute either $u = \sin x$ or $u = \cos x$. If we try $u = \sin x$, then we set aside one $\cos x$ for du and convert the remaining factor of $\cos x$ into terms of $\sin x$ by writing $\cos x = \sqrt{1 - \sin^2 x}$. However, this means that we get

$$\int_0^{\pi/2} \sin^5 x \sqrt{1 - \sin^2 x} \cos x \, dx = \int_0^1 u^5 \sqrt{1 - u^2} \, du$$

and that square root is difficult to integrate.

However, if we try $u = \cos x$ instead, then we set aside one $\sin x$ for du and convert the remaining four factors of $\sin x$ into terms of $\cos x$. Since $\sin^4 x = (\sin^2 x)^2 = (1 - \cos^2 x)^2$, this conversion does not involve any square roots. After substitution of $u = \cos x$, we obtain

$$\int_0^{\pi/2} \cos^2 x (1 - \cos^2 x)^2 \sin x \, dx = -\int_1^0 u^2 (1 - u^2)^2 \, du$$

This is easy to integrate by distributing, and we get

$$\int_0^1 u^2 - 2u^4 + u^6 \, du = \left(\frac{u^3}{3} - \frac{2u^5}{5} + \frac{u^7}{7} \right) \Bigg|_0^1 = \frac{1}{3} - \frac{2}{5} + \frac{1}{7} = \frac{8}{105}$$

This goes to show that when choosing whether to substitute for $\sin x$ or for $\cos x$, it is easier to set aside a factor from an ODD power and then convert the remaining even number of factors to the other type. □

Occasionally, this tactic of leaving aside one factor and converting pairs to the other type applies to integrands which don't look like powers of sine times powers of cosine. To make this tactic apply, we try writing other trigonometric functions in terms of sine and cosine, as the following example demonstrates.

Example 6.16:

For example, consider

$$\int \frac{\sec^5 x}{\csc^3 x}\,dx$$

Since $\sec x = 1/(\cos x)$ and $\csc x = 1/(\sin x)$, our integral is the same as

$$\int \frac{\sin^3 x}{\cos^5 x}\,dx$$

Here, we want to substitute $u = \cos x$. (This is because we can only set aside a factor from the numerator to help with $du = -\sin x\,dx$. A factor of $\cos x$ from the denominator doesn't help.) Setting one power of $\sin x$ aside and converting the rest to cosines, we get

$$\begin{aligned}\int \frac{1-\cos^2 x}{\cos^5 x}\sin x\,dx &= -\int \frac{1-u^2}{u^5}\,du = \int u^{-3} - u^{-5}\,du \\ &= \frac{u^{-2}}{-2} - \frac{u^{-4}}{-4} + C = \frac{-1}{2\cos^2 x} + \frac{1}{4\cos^4 x} + C\end{aligned}$$

□

Our tactic of setting one factor aside won't work if we have even powers of sine and of cosine. In this case, we can use the half-angle identities

$$\sin^2 x = \frac{1}{2}(1-\cos(2x)) \qquad \cos^2 x = \frac{1}{2}(1+\cos(2x))$$

By doing this, we obtain a problem with smaller powers of $\cos(2x)$, and the $2x$ can easily be handled by a substitution. We demonstrate with an example.

Example 6.17:

Let's solve

$$\int_{\pi}^{2\pi} \cos^4 x\,dx$$

without using the complicated reduction formula. Since $\cos^4 x = (\cos^2 x)^2$, we can use the half-angle identities to rewrite our integral as

$$\int_{\pi}^{2\pi} \left(\frac{1}{2}(1+\cos(2x))\right)^2 dx = \int_{\pi}^{2\pi} \frac{1}{4} + \frac{\cos(2x)}{2} + \frac{1}{4}\cos^2(2x)\, dx$$

We can certainly integrate constants, and we can integrate $\cos(2x)$ easily as well with a substitution for $2x$ (that substitution is frequently done without writing full details for convenience). To handle $\cos^2(2x)$, we treat it as another even power of cosine and use a half-angle identity again to obtain

$$\cos^2(2x) = \frac{1}{2}(1+\cos(4x))$$

Putting this all together, and using a little mental math to perform substitutions of $2x$ and $4x$, we get

$$\begin{aligned} & \int_{\pi}^{2\pi} \frac{1}{4} + \frac{\cos(2x)}{2} + \frac{1}{8}(1+\cos(4x))\, dx \\ = & \left(\frac{x}{4} + \frac{\sin(2x)}{4} + \frac{1}{8}\left(x + \frac{\sin(4x)}{4}\right)\right)\Bigg|_{\pi}^{2\pi} \\ = & \left(\frac{3x}{8} + \frac{\sin(2x)}{4} + \frac{\sin(4x)}{32}\right)\Bigg|_{\pi}^{2\pi} \end{aligned}$$

Note that when plugging in π and 2π to the parts of the antiderivatives with the sines, we'll obtain 0. Thus, the only part that remains is $3x/8$ evaluated at π and 2π, and our answer is $3\pi/8$. $\square$

Secants and Tangents

Another commonly-occurring type of trigonometric integral involves powers of secant times powers of tangent. These two types of trigonometric functions go well together because their derivatives involve each other: $(\tan x)' = \sec^2 x$ and $(\sec x)' = \sec x \tan x$. We can also convert between pairs of tangents and pairs of secants using the identity

$$\sec^2 x = \tan^2 x + 1$$

These formulas tell us how to perform substitutions, as the following two examples show:

Example 6.18:

Let's compute

$$\int \tan^2 x \sec^4 x \, dx$$

Suppose that instead of using our formulas above with secants and tangents, we instead tried to express this in terms of sines and cosines. We would obtain

$$\int \frac{\sin^2 x}{\cos^6 x} \, dx$$

We'd use a half-angle identity in the numerator and denominator, obtaining a denominator of

$$\cos^6 x = (\cos^2 x)^3 = \frac{1}{8}(1 - 3\cos(2x) + 3\cos^2(2x) - \cos^3(2x))$$

After substituting for $\cos(2x)$, the resulting integral will still have a complicated polynomial in its denominator, and we do not have techniques for integrating a function like that. Hence, instead of using sines and cosines, let's use our formulas for secants and tangents.

We would like to either use $u = \sec x$ and set aside $\sec x \tan x$ for du, or we would like to use $u = \tan x$ and set aside $\sec^2 x$ for du. Since we have an even number of secants, let's use $u = \tan x$ and set aside $du = \sec^2 x \, dx$. We have two secants left over, which we can convert to $\tan^2 x + 1$. This gives us

$$\begin{aligned} \int \tan^2 x(\tan^2 x + 1)\sec^2 x \, dx &= \int u^2(u^2+1) \, du = \int u^4 + u^2 \, du \\ &= \frac{u^5}{5} + \frac{u^3}{3} + C = \frac{\tan^5 x}{5} + \frac{\tan^3 x}{3} + C \end{aligned}$$

In general, this approach of using $u = \tan x$ works well when you have an even number of secants. □

Example 6.19:

Let's integrate

$$\int_0^{\pi/4} \tan^3 x \sec x \, dx$$

Since we don't have an even number of secants, we can't set aside $\sec^2 x$ for a substitution of $u = \tan x$. Instead, we try $u = \sec x$ and set aside

$du = \sec x \tan x\, dx$. After doing this, we have two tangents left over, which can be converted to $\sec^2 x - 1$. This gives

$$\begin{aligned} \int_0^{\pi/4} (\sec^2 x - 1) \sec x \tan x\, dx &= \int_1^{\sqrt{2}} u^2 - 1\, du = \left(\frac{u^3}{3} - u \right) \Big|_1^{\sqrt{2}} \\ &= \left(\frac{2\sqrt{2}}{3} - \sqrt{2} \right) - \left(\frac{1}{3} - 1 \right) = \frac{2 - \sqrt{2}}{3} \end{aligned}$$

In general, this approach works well when you have a positive number of secants and an odd number of tangents. □

The previous two examples show general tactics for handling integrands with an even number of secants or an odd number of tangents. There are two cases that can't be handled by these tactics. The first is integrating $\tan x$ on its own: this is handled by Exercise 6.3.18. The other is when there are an odd number of secants (or zero) and an even number of tangents. To handle this, usually the tangents are all converted to secants, and then a reduction formula is used to integrate the secants. See Exercises 6.3.19 and 6.6.11.

Cosecants and Cotangents

Integrands involving $\csc x$ and $\cot x$ are handled in the same way as integrals with secants and tangents, using the formulas

$$(\cot x)' = -\csc^2 x \qquad (\csc x)' = -\csc x \cot x \qquad \csc^2 x = \cot^2 x + 1$$

You can practice integrals with $\csc x$ and $\cot x$ in the exercises.

6.5 Trigonometric Substitutions

In this section, we will handle several new kinds of integrands by turning them into integrals of trigonometric functions. To illustrate how this is done, and why it is useful, let's consider the integral

$$\int \sqrt{1 - x^2}\, dx$$

This kind of integral appears when dealing with circles, since the circle $x^2 + y^2 = 1$ can be represented by the two functions $y = \pm\sqrt{1 - x^2}$. If this were the integral of $x\sqrt{1-x^2}$, then a substitution of $u = 1 - x^2$ would work perfectly. However, without that factor of x, we need to do more work. (It is still possible to solve this problem with substitution, in a manner similar to the solution to Exercise 6.3.12, but the approach we will present is more elegant.)

We can get some intuition by thinking about circles. If (x, y) forms an angle of θ with the positive x-axis on the circle $x^2 + y^2 = 1$, then $x = \cos\theta$ and $y = \sin\theta$. Thus, if we try to replace x with $\cos\theta$ in our integral, then we get

$$\sqrt{1-x^2} = \sqrt{1-\cos^2\theta} = \sqrt{\sin^2\theta} = |\sin\theta|$$

Note that by replacing x with a trigonometric function, we can use an identity to simplify the square root greatly. This suggests that we should substitute $x = \cos\theta$ to replace the integral with respect to x by an integral with respect to θ.

For now, let's proceed informally (we'll justify all the details later). Suppose we make $x = \cos\theta$, so that $dx = -\sin\theta\, d\theta$ (once again treating dx and $d\theta$ as differentials). Also, let's temporarily ignore the absolute-value bars we obtained in our last calculation (this is also justified later), and we get

$$\int \sqrt{1-x^2}\, dx = -\int \sqrt{1-\cos^2\theta}\,\sin\theta\, d\theta = -\int \sin^2\theta\, d\theta$$

We learned how to solve this in the last section by using a half-angle identity. We obtain

$$-\frac{1}{2}\int 1 - \cos(2\theta)\, d\theta = \frac{\sin(2\theta)}{4} - \frac{\theta}{2} + C = \frac{(\sin\theta)(\cos\theta)}{2} - \frac{\theta}{2} + C$$

(We used a double-angle identity in the last step.)

At this point, we need to return to the original variable of integration, which is x. We know $\cos\theta = x$, so $\sin\theta = \sqrt{1-\cos^2\theta} = \sqrt{1-x^2}$. Also, $\theta = \arccos x$, the inverse of the cosine function. Thus, our final answer is

$$\int \sqrt{1-x^2}\, dx = \frac{x\sqrt{1-x^2}}{2} - \frac{\arccos x}{2} + C$$

(You should differentiate the answer to check that it is correct!) Note that, in particular, this formula lets us compute

$$\int_0^1 \sqrt{1-x^2}\, dx = \left(\frac{1\sqrt{1-1^2}}{2} - \frac{\arccos 1}{2}\right) - \left(\frac{0\sqrt{1-0^2}}{2} - \frac{\arccos 0}{2}\right) = \frac{\pi}{4}$$

Hence, using calculus, we can prove that the area under a quarter circle of radius 1 really is $\pi/4$, and thus the area of a circle of radius 1 is π. (Also see Exercise 6.6.21.)

Ignoring the formalities for now, let's show how we can compute other integrals with similar tactics which replace x with a trigonometric function. There are three main types of situations where these tactics are frequently used, and we'll present one example of each.

Example 6.20:

Let's compute

$$\int_0^{1/2} x^3\sqrt{1-x^2}\,dx$$

Because the expression $\sqrt{1-x^2}$ simplifies greatly when x is $\sin\theta$ or $\cos\theta$, let's try setting $x = \sin\theta$ and $dx = \cos\theta\,d\theta$. When $x = 0$, we have $\theta = \arcsin 0 = 0$, and when $x = 1/2$, we have $\theta = \arcsin(1/2) = \pi/6$. Thus,

$$\int_0^{1/2} x^3\sqrt{1-x^2}\,dx = \int_0^{\pi/6} \sin^3\theta\sqrt{1-\sin^2\theta}\cos\theta\,d\theta = \int_0^{\pi/6} \sin^3\theta\cos^2\theta\,d\theta$$

We can find this integral by setting one copy of $\sin\theta$ aside for a u-substitution of $u = \cos\theta$. When $\theta = 0$, we have $u = 1$, and when $\theta = \pi/6$, we have $u = \sqrt{3}/2$. We get

$$\begin{aligned}
\int_0^{\pi/6} (1-\cos^2\theta)\cos^2\theta\sin\theta\,d\theta &= -\int_1^{\sqrt{3}/2} (1-u^2)u^2\,du \\
&= \int_1^{\sqrt{3}/2} u^4 - u^2\,du \\
&= \left.\left(\frac{u^5}{5} - \frac{u^3}{3}\right)\right|_1^{\sqrt{3}/2} \\
&= \left(\frac{3^2\sqrt{3}}{2^5\cdot 5} - \frac{3\sqrt{3}}{2^3\cdot 3}\right) - \left(\frac{1}{5} - \frac{1}{3}\right) \\
&= \frac{2}{15} - \frac{11\sqrt{3}}{160}
\end{aligned}$$

Notice that unlike the previous example, we were solving a definite integral, so we could change limits and not have to return to the original variable x. $\square$

Example 6.21:

Let's compute

$$\int \frac{1}{x^2\sqrt{x^2+1}}\,dx$$

In this situation, we'd like to replace x with a trigonometric function so that x^2+1 becomes a perfect square. Since $\tan^2\theta + 1 = \sec^2\theta$, we try $x = \tan\theta$, so $dx = \sec^2\theta\,d\theta$. As before, let's just write $\sqrt{\tan^2\theta+1}$ as $\sec\theta$, ignoring absolute-value bars. This gives us

$$\int \frac{1}{\tan^2\theta\sqrt{\tan^2\theta+1}}\sec^2\theta\,d\theta = \int \frac{\sec\theta}{\tan^2\theta}\,d\theta$$

Since this is a fraction, as opposed to a power of secant times a power of tangent, we try the tactic of expressing everything in terms of sines and cosines. This gives us

$$\int \frac{\cos\theta}{\sin^2\theta}\,d\theta$$

which can be easily solved by substituting $u = \sin\theta$. Thus, our problem becomes

$$\int \frac{1}{u^2}\,du = \frac{-1}{u} + C = \frac{-1}{\sin\theta} + C$$

It remains to write this expression in terms of x again. The easiest way to do this is to use a "right triangle method" as follows. Since $x = \tan\theta$, and the tangent gives us the length of the opposite side of a right triangle over the length of the adjacent side, let's draw a right triangle with opposite side x and adjacent side 1. See Figure 6.1.

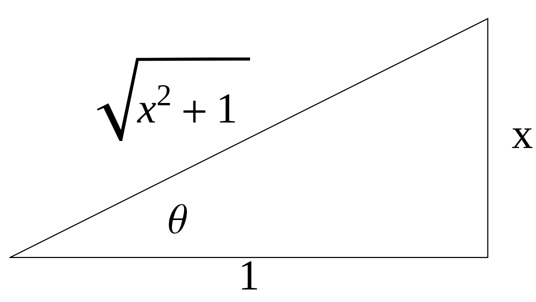

Figure 6.1: A right triangle illustrating $x = \tan\theta$

Now, $\sin\theta$ is the opposite side over the hypotenuse. The hypotenuse is found by the Pythagorean theorem to be $\sqrt{x^2+1}$, so we obtain

$$\frac{-1}{\sin\theta} + C = \frac{-\sqrt{x^2+1}}{x} + C$$

as our final answer. $\square$

Example 6.22:

Let's integrate

$$\int_{\sqrt{2}}^{2} \frac{x^3}{\sqrt{x^2-1}}\,dx$$

We'd like to pick a substitution for x which turns x^2-1 into a perfect square. Since we know that $\sec^2\theta - 1 = \tan^2\theta$, we try setting $x = \sec\theta$ and $dx = \sec\theta\tan\theta\,d\theta$. This will give us $\sqrt{\sec^2\theta - 1} = \tan\theta$ (as before, we ignore absolute-value bars for now.) When $x = \sqrt{2}$, we have $\sec\theta = 1/(\cos\theta) = \sqrt{2}$, so $\theta = \pi/4$, and when $x = 2$, we have $\theta = \pi/3$. This gives us

$$\int_{\sqrt{2}}^{2} \frac{x^3}{\sqrt{x^2-1}}\,dx = \int_{\pi/4}^{\pi/3} \frac{\sec^3\theta}{\sqrt{\sec^2\theta-1}} \sec\theta\tan\theta\,d\theta = \int_{\pi/4}^{\pi/3} \sec^4\theta\,d\theta$$

In the last section, we learned that to integrate an even power of secant, you separate out $\sec^2\theta$, convert the remaining secants to tangents, and make the substitution $u = \tan\theta$. This gives us

$$\begin{aligned}
\int_{\pi/4}^{\pi/3} (\tan^2\theta + 1)\sec^2\theta\,d\theta &= \int_1^{\sqrt{3}} u^2 + 1\,du \\
&= \left(\frac{u^3}{3} + u\right)\Bigg|_1^{\sqrt{3}} = \left(\frac{3\sqrt{3}}{3} + \sqrt{3}\right) - \left(\frac{1}{3} + 1\right) \\
&= 2\sqrt{3} - \frac{4}{3}
\end{aligned}$$

$\square$

The Formal Details

At this point, let's be more formal about the technique we have been using. Suppose that we are trying to compute

$$\int_a^b f(x)\,dx$$

where f is continuous on $[a,b]$. (We choose to analyze definite integrals here, so that we can be more precise about the values our variables take.) When we do a trigonometric substitution, we set $x = g(\theta)$ and $dx = g'(\theta)\,d\theta$ for some trigonometric function g. We require g to be invertible, so that we may change our limits from $x = a$ and $x = b$ to $\theta = g^{-1}(a)$ and $\theta = g^{-1}(b)$ respectively. Hence, we get

$$\int_a^b f(x)\,dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(\theta))g'(\theta)\,d\theta$$

Essentially, we are performing substitution in reverse, since technically the substitution rule specified in Theorem 6.6 is used to go from the integral in terms of θ to the integral in terms of x. In order for this theorem to apply, we need $f(g(\theta))$ to be continuous from $g^{-1}(a)$ to $g^{-1}(b)$, which is guaranteed by the Composition Limit Theorem as long as g is continuous from $g^{-1}(a)$ to $g^{-1}(b)$. We also need g to have a continuous derivative between $g^{-1}(a)$ and $g^{-1}(b)$ in order to apply substitution.

In summary, we have proven the following theorem:

Theorem 6.23 (Inverse Substitution). *Let $a, b \in \mathbb{R}$ be given with $a \le b$. Let f and g be real functions satisfying:*

- *f is continuous on $[a,b]$*
- *g is invertible and g^{-1} is continuous on $[a,b]$ (thus g is continuous on some closed interval containing $g^{-1}(a)$ and $g^{-1}(b)$)*
- *g' exists and is continuous between $g^{-1}(a)$ and $g^{-1}(b)$*

Then

$$\int_a^b f(x)\,dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(\theta))g'(\theta)\,d\theta$$

In our previous examples, we used this theorem with g being either sine, cosine, tangent, or secant. In order for the theorem to apply, we need g to be restricted to a domain on which it is invertible. Hence, we use the same domain for g that we used when defining the inverse trigonometric functions in Chapter 3 when g is sine, cosine, or tangent. (When g is secant, we traditionally choose the domain of $[0,\pi] - \{\pi/2\}$.) This leads to the following guidelines of which substitutions to pick[2]:

When integrating this form	Try this	Values for θ (i.e. domain of g)
$\sqrt{1-x^2}$	$x = \sin\theta$	$[-\pi/2, \pi/2]$
$\sqrt{x^2+1}$	$x = \tan\theta$	$(-\pi/2, \pi/2)$
$\sqrt{x^2-1}$	$x = \sec\theta$	$[0,\pi/2) \cup [\pi, 3\pi/2)$

There is another useful thing to notice about the table above. Suppose we are in a situation where we want to substitute $x = \sin\theta$, as the first row of the table describes. Since $\theta \in [-\pi/2, \pi/2]$, we have $\cos\theta \geq 0$, so $\sqrt{1-\sin^2\theta} = \cos\theta$. A similar remark holds for the other two rows of the table. This is why we were able to avoid writing absolute-value bars in our examples, because θ came from a range of values which made them unnecessary. This fully justifies the computations we have done earlier, and the process of using the substitutions from this table to help integrate square roots is often called the technique of *trigonometric substitution*.

Also, trigonometric substitutions can be used with some forms which are not listed in the table. In general, anything involving squared variables under a square root probably works well with trigonometric substitutions. We provide one example, and you can practice others in the exercises.

Example 6.24:

Let's integrate

$$\int x^3(4x^2+9)^{3/2}\,dx$$

Since $(4x^2+9)^{3/2}$ can be written as $(\sqrt{4x^2+9})^3$, a trigonometric substitution is probably useful. We want to pick a substitution for x so that $4x^2+9$ becomes a perfect square. We note that if $4x^2 = 9\tan^2\theta$, then $4x^2+9 = 9\tan^2\theta + 9 = 9\sec^2\theta$.

Because of this, we set $x = (3/2)\tan\theta$ and $dx = (3/2)\sec^2\theta\,d\theta$. With this choice of x, $(4x^2+9)^{3/2}$ becomes $(\sqrt{9\sec^2\theta})^3 = 27\sec^3\theta$. (Note that

[2]It is also possible to choose $\cos\theta$ instead of $\sin\theta$ in the first row, and so on, but doing this changes the values that θ can take and tends to introduce more minus signs than when $\sin\theta$ is used. Thus, we suggest using $\sin\theta$. This is also why we don't suggest using $\cot\theta$ in the second row or $\csc\theta$ in the third row.

according to our table, we choose θ to be in $(-\pi/2, \pi/2)$, so it follows that $\sec\theta \ge 0$.) We hence obtain

$$\int x^3(4x^2+9)^{3/2}\,dx = \int \left(\frac{27}{8}\tan^3\theta\right)\left(27\sec^3\theta\right)\left(\frac{3}{2}\sec^2\theta\right)\,d\theta$$

At this point, since we have an odd number of tangents, we set aside $\sec\theta\tan\theta$, convert the remaining tangents to secants, and substitute $u = \sec\theta$. Thus,

$$\begin{aligned} & \frac{2187}{16}\int(\sec^2\theta - 1)\sec^4\theta(\sec\theta\tan\theta)\,d\theta \\ = \ & \frac{2187}{16}\int(u^2-1)u^4\,du = \frac{2187}{16}\int u^6 - u^4\,du \\ = \ & \frac{2187}{16}\left(\frac{u^7}{7} - \frac{u^5}{5}\right) + C = \frac{2187}{16}\left(\frac{\sec^7\theta}{7} - \frac{\sec^5\theta}{5}\right) + C \end{aligned}$$

To finish this problem, we need to return to the original variable x. Since $x = (3/2)\tan\theta$, we have $\tan\theta = (2x)/(3)$, so we may picture a right triangle with θ as one of its angles, opposite side length $2x$, and adjacent side length 3. Thus, the hypotenuse has length $\sqrt{(2x)^2+3^2} = \sqrt{4x^2+9}$. This tells us that $\sec\theta$ is hypotenuse over adjacent, i.e. $\sec\theta = (\sqrt{4x^2+9})/3$. Hence, our final answer (left somewhat unsimplified) is

$$\frac{2187}{16}\left(\frac{(4x^2+9)^{7/2}}{3^7\cdot 7} - \frac{(4x^2+9)^{5/2}}{3^5\cdot 5}\right) + C$$

□

6.6 Exercises

For Exercises 1 through 10, evaluate the indefinite or definite integrals.

1. $\displaystyle\int \sin^2 x\cos^3 x\,dx$

2. $\displaystyle\int_0^{\pi/2} \sin^4 x\,dx$

3. $\displaystyle\int \sin x\sec^2 x\,dx$

4. $\displaystyle\int (1+\cos\theta)^2\,d\theta$

5. $\displaystyle\int_0^{\pi} \cos^2 x \sin(2x)\, dx$
(Hint: First use a double-angle identity.)

6. $\displaystyle\int_0^{\pi/3} \tan x \sec^4 x\, dx$

7. $\displaystyle\int \tan^2(3x)\, dx$

8. $\displaystyle\int \sec^6 x\, dx$

9. $\displaystyle\int \cot x \csc^3 x\, dx$

10. $\displaystyle\int \cot^4 x \csc^4 x\, dx$

11. (a) Prove that for all $n \ge 2$, we have the following reduction formula:

$$\int \sec^n x\, dx = \frac{1}{n-1} \tan x \sec^{n-2} x + \frac{n-2}{n-1} \int \sec^{n-2} x\, dx$$

This formula can be used to integrate any odd power of secant.

(b) Prove that for all $n \ge 2$,

$$\int \csc^n x\, dx = \frac{-1}{n-1} \cot x \csc^{n-2} x + \frac{n-2}{n-1} \int \csc^{n-2} x\, dx$$

12. (a) Using the identity

$$\sin a \sin b = \frac{1}{2}(\cos(a-b) - \cos(a+b))$$

which is true for all $a, b \in \mathbb{R}$, evaluate $\displaystyle\int \sin(2x)\sin(5x)\, dx$.

(b) Using the identity

$$\sin a \cos b = \frac{1}{2}(\sin(a-b) + \sin(a+b))$$

which is true for all $a, b \in \mathbb{R}$, evaluate $\displaystyle\int \sin(3x)\cos x\, dx$.

13. This exercise uses the identities from the previous exercise, along with the identity

$$\cos a \cos b = \frac{1}{2}(\cos(a+b) + \cos(a-b))$$

which is true for all $a, b \in \mathbb{R}$.

(a) Show that for all $m, n \in \mathbb{N}^*$, we have

$$\begin{aligned}
\int_{-\pi}^{\pi} \sin(mx)\cos(nx)\,dx &= 0 \\
\int_{-\pi}^{\pi} \sin(mx)\sin(nx)\,dx &= \begin{cases} 0 & \text{if } m \neq n \\ \pi & \text{if } m = n \end{cases} \\
\int_{-\pi}^{\pi} \cos(mx)\cos(nx)\,dx &= \begin{cases} 0 & \text{if } m \neq n \\ \pi & \text{if } m = n \end{cases}
\end{aligned}$$

(Note: You will want to use the fact that $m + n \neq 0$ since m and n are positive.)

(b) A *trigonometric polynomial* is a function $f : \mathbb{R} \to \mathbb{R}$ of the form

$$f(x) = C + \sum_{k=1}^{n} (a_k \sin(kx) + b_k \cos(kx))$$

where C, a_k, and b_k are constants for each k from 1 to n. (These are important in the theory of *Fourier series*.) Use part (a) to show that for each k from 1 to n,

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx \qquad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\,dx$$

as well as

$$C = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx$$

For Exercises 14 through 18, evaluate the following integrals. Remember that for indefinite integrals, the final answer needs to be in terms of the original variable (so drawing a right triangle might help with trigonometric substitutions).

14. $\displaystyle\int \frac{1}{x^2\sqrt{25 - x^2}}\,dx$

15. $\displaystyle\int \frac{\sqrt{x^2 - 1}}{x^4}\,dx$

16. $\displaystyle\int_1^{\sqrt{3}} \frac{\sqrt{x^2 + 1}}{x^4}\,dx$

17. $\displaystyle\int (16 - x^2)^{5/2}\,dx$

18. $\displaystyle\int_0^1 x\sqrt{1-x^4}\,dx$
(Hint: The x on the outside goes with dx.)

19. Compute the integral
$$\int_1^2 \sqrt{2x-x^2}\,dx$$
by first completing the square.

20. By using the same tactic as the last problem, compute
$$\int \frac{dx}{4x^2+4x+2}$$

21. When a and b are positive constants, the points $(x,y) \in \mathbb{R}^2$ which satisfy
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$
form an ellipse centered at the origin, whose horizontal semiaxis has length a and whose vertical semiaxis has length b. Therefore, the top curve of this ellipse satisfies
$$y = b\sqrt{1-\frac{x^2}{a^2}}$$
Compute
$$\int b\sqrt{1-\frac{x^2}{a^2}}\,dx$$
and use this to show that the total area enclosed by the ellipse is πab. When $a = b$, so that the ellipse is a circle, this proves that a circle with radius r has area πr^2.

6.7 Computing Some Areas

In the last chapter, our main motivation for the definition of the integral was to establish a way to compute areas. More precisely, when f is a nonnegative function which is integrable on $[a,b]$, the value $\displaystyle\int_a^b f(x)\,dx$ is the area between the x-axis and the graph of f on $[a, b]$. To develop this definition, we approximated the area from above and below by step functions. This led to the lower integral $L(f; [a, b])$ as the supremum of lower estimates by step functions and the upper integral $U(f; [a, b])$ as the infimum of upper estimates by step functions. In some sense, $L(f; [a, b])$ and $U(f; [a, b])$ are respectively the "ideal" lower and upper estimates, so when they are equal, that should be the exact value of the area.

It turns out that there are many problems where an answer can be obtained in a similar manner: you define a simpler kind of problem (such as the area of a step function), you use this simple kind to make approximations, and then you see whether lower and upper estimates agree on one value. For example, in physics, integrals are used to compute the total distance an object travels if its speed function is known (where in the simpler problem, we consider constant speeds), as well as the total work done by a force over a distance (where in the simpler problem, we consider constant forces). In statistics, an event is described by a certain probability density, and the total probability that the event occurs is found by integrating the density (where in the simpler problem, we consider constant densities, which correspond to all outcomes being equally likely).

Many of these applications are well-addressed in books specific to the subject matter in question, so we will not treat them here. Instead, we will focus on computing areas and volumes in the next two sections. These problems are quite similar to what we've already done, and they illustrate how to use integration to approach physical problems.

Example 6.25:

As a demonstration, let's try and compute the area of the region between the y-axis and the curve $x = 1 - y^2$. This region is pictured in Figure 6.2. (We will explain the vertical lines in the figure in a second.)

Clearly the y-axis and $x = 1 - y^2$ intersect when $x = 0$, i.e. $1 - y^2 = 0$. Thus, an equivalent way to state our problem is that we'd like to compute the area to the left of $x = 1 - y^2$ and to the right of the y-axis from $y = -1$ to $y = 1$. If we take our figure and swap the roles of x and y (which amounts to performing a reflection over the line $y = x$), then the area we want to compute should be the same as the area under $y = 1 - x^2$ and above the x-axis from $x = -1$ to $x = 1$, which we know is

$$\int_{-1}^{1} 1 - x^2 \, dx$$

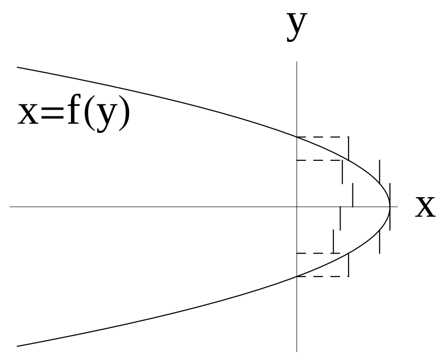

Figure 6.2: $x = f(y) = 1 - y^2$, with step functions $s(y)$ and $t(y)$ (s is on the left and t is on the right)

Intuitively, this integral should give the area we want, because a reflection will take the region to the left of $x = 1 - y^2$ and map it to a congruent region under $y = 1 - x^2$. Another more formal way of justifying this, though, is to use the approach described earlier, which involves approximations. Let $f(y) = 1 - y^2$ for all $y \in [-1, 1]$, and say that s and t are step functions such that $s(y) \le f(y) \le t(y)$ for all $y \in [-1, 1]$. (The vertical lines in the figure represent the graphs of s and t as functions of y.)

We can approximate the area to the left of $f(y)$ by saying that it is underestimated by the area to the left of $s(y)$ and overestimated by the area to the left of $t(y)$. Say that s and t are both compatible with a partition $\{y_0, y_1, \ldots, y_n\}$ of $[-1, 1]$, where for each i from 1 to n, s takes the value s_i and t takes the value t_i on the i^{th} open subinterval. Then, for each i, the region to the left of the i^{th} part of s is a rectangle with height $y_i - y_{i-1}$ and width s_i. (Think of each rectangle as being a small horizontal strip of area, as shown in Figure 6.2 surrounded by dashed lines.) A similar statement holds for t. This means that if A is the area of our region, then A satisfies

$$\sum_{i=1}^{n} s_i(y_i - y_{i-1}) \le A \le \sum_{i=1}^{n} t_i(y_i - y_{i-1})$$

In other words,

$$\int_{-1}^{1} s(y)\, dy \le A \le \int_{-1}^{1} t(y)\, dy$$

for every lower step function s and upper step function t for f. This means $A \ge L(f; [-1, 1])$ and $A \le U(f; [-1, 1])$. Because f is integrable, the only choice for A satisfying this is precisely

$$\int_{-1}^{1} f(y)\,dy = \int_{-1}^{1} 1 - y^2\,dy = \left(y - \frac{y^3}{3}\right)\Big|_{-1}^{1} = \frac{4}{3}$$

□

Remark. Another way to do the problem in the previous example is to say that the region we are measuring has a top curve of $y = \sqrt{x+1}$ and a bottom curve of $y = -\sqrt{x+1}$ from $x = 0$ to $x = 1$. This produces a slightly more complicated integral for our answer. In general, it's useful to keep in mind that some regions are best measured by using integrals with respect to x, i.e. slicing the region into vertical strips, and some regions are best measured by using integrals with respect to y, i.e. slicing the region into horizontal strips.

Area Between Curves

We have seen how to use integrals to find the area underneath one curve. How would we compute the area between two curves? The following example demonstrates the main idea:

Example 6.26:

Let $f(x) = x^2$ and $g(x) = 2x - x^2$ for all $x \in \mathbb{R}$. When we draw the graphs of f and g, we see that they enclose a small region, as shown in Figure 6.3. How can we find the area of that region?

First, we note that the two graphs intersect when $x^2 = 2x - x^2$, i.e. when $2x^2 - 2x = 2x(x-1) = 0$. Thus, the intersections occur at $x = 0$ and $x = 1$. Between $x = 0$ and $x = 1$, we have $g(x) > f(x)$, so we can obtain the area between the curves by taking the area under g on $[0, 1]$ and subtracting the area under f on $[0, 1]$. (This is because the area under both curves will cancel in the subtraction.) This gives us

$$\int_0^1 2x - x^2\,dx - \int_0^1 x^2\,dx = \int_0^1 2x - 2x^2\,dx = \left(x^2 - \frac{2x^3}{3}\right)\Big|_0^1 = 1 - \frac{2}{3} = \frac{1}{3}$$

□

The previous example indicates that if you want to find the area between f and g on $[a, b]$, when $0 \le f(x) \le g(x)$ for all $x \in [a, b]$, then you take

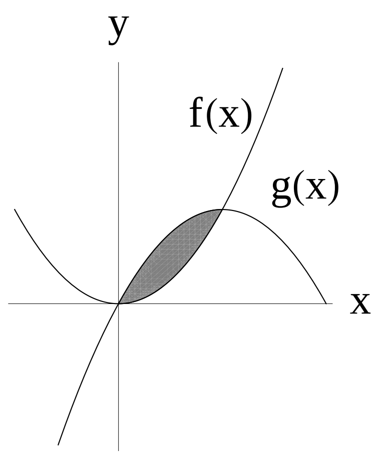

Figure 6.3: The region between $f(x) = x^2$ and $g(x) = 2x - x^2$

the area under the larger function and subtract the area under the smaller function, i.e. you compute $\int_a^b g(x)\,dx - \int_a^b f(x)\,dx$. This is the same as $\int_a^b (g(x) - f(x))\,dx$.

It turns out that this formula works even when f and g are allowed to take negative values on $[a, b]$. One way to see this is to take the graphs of f and g and shift them upwards by a large enough constant C so that $f(x) + C, g(x) + C \geq 0$ for all $x \in [a, b]$. (This is always possible because f and g are bounded functions if they are integrable.) Shifting upwards doesn't change the area between the curves, so the area is

$$\int_a^b (g(x) + C) - (f(x) + C)\,dx = \int_a^b g(x) - f(x)\,dx$$

We'd like to demonstrate another argument for why this formula works, since this argument can be generalized to other situations. In this argument, we make approximations for the area between f and g using step functions. Suppose that s is a step function below f, and t is a step function above g. Thus, if s is close to f and t is close to g, then the area between s and t is approximately the area between f and g. (See Figure 6.4.)

Let's say that s and t are compatible with a partition $\{x_0, x_1, \cdots, x_n\}$ of $[a, b]$, and for each i from 1 to n, s takes the value s_i and t takes the value t_i on the i^{th} subinterval. The region between s and t on the i^{th} subinterval

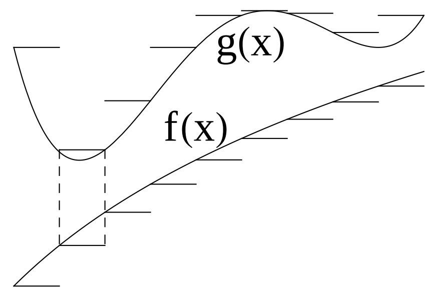

Figure 6.4: Using step functions to approximate area between curves

is a rectangle with height $t_i - s_i$ and width $x_i - x_{i-1}$. (This vertical strip is surrounded by dashed sides in Figure 6.4). Thus, adding this up for all subintervals, the area between s and t is

$$\sum_{i=1}^{n}(t_i - s_i)(x_i - x_{i-1}) = \int_a^b t(x) - s(x)\,dx$$

This yields an overestimate of the true area between f and g, i.e. $t - s$ is a lower step function for $g - f$. This means that the area between f and g is at most $U(g - f; [a, b])$. Similarly, with some care, you can obtain underestimates by using upper step functions for f and lower step functions for g, showing that the area between f and g is at least $L(g - f; [a, b])$. Therefore, when $g - f$ is integrable, the area is the integral of $g - f$, which is the answer we expected.

Intuitively, when we are computing the area between two curves, we are taking infinitesimal vertical slices of height $g(x) - f(x)$ and adding them up on $[a, b]$. The same kind of argument we just did also works for functions of y, which corresponds to taking horizontal slices of width $g(y) - f(y)$ and adding them up. This next example shows a problem which can be done either with vertical or horizontal slices.

Example 6.27:

We define $f, g : [0, 4] \to [0, 2]$ by

$$f(x) = \sqrt{x} \qquad g(x) = \begin{cases} 0 & \text{if } x < 2 \\ x - 2 & \text{if } x \geq 2 \end{cases}$$

We would like to find the area between f and g. To do that, we should first find out where the curves intersect. When $x < 2$, we have $\sqrt{x} = 0$ only when $x = 0$. When $x \geq 2$, we have $\sqrt{x} = x - 2$ iff $x = (x-2)^2 = x^2 - 4x + 4$ iff $x^2 - 5x + 4 = (x-1)(x-4) = 0$. Since $(x-1)(x-4) = 0$ when $x = 1$ or $x = 4$, the only choice satisfying $x \geq 2$ is $x = 4$. Thus, the two curves intersect at $(0,0)$ and $(4,2)$.

One way to find the area between these curves is to note that we always have $f(x) \geq g(x)$ on $[0,4]$. Thus, the area is the integral from 0 to 4 of $f(x) - g(x)$. Since g is defined piecewise, we need to compute our integral in two pieces as

$$\begin{aligned}
& \int_0^2 \sqrt{x} - 0\,dx + \int_2^4 \sqrt{x} - (x-2)\,dx \\
= & \left(\frac{2}{3}x^{3/2}\right)\Big|_0^2 + \left(\frac{2}{3}x^{3/2} + 2x - \frac{x^2}{2}\right)\Big|_2^4 \\
= & \frac{2}{3}(4)^{3/2} + 2(4-2) - \frac{1}{2}(4^2 - 2^2) = \frac{10}{3}
\end{aligned}$$

Another way to do this to view our curves as functions of y, i.e. slice the problem horizontally. When $y = \sqrt{x}$, we have $x = y^2$, and when $y = x - 2$ for $x \in [2,4]$, we have $x = y + 2$. Thus, the area we want is the area between $y + 2$ and y^2 as y goes from 0 to 2. (Note that $y + 2 \geq y^2$ for $y \in [0,2]$.) We compute

$$\int_0^2 (y+2) - y^2\,dy = \left(\frac{y^2}{2} + 2y - \frac{y^3}{3}\right)\Big|_0^2 = \frac{2^2}{2} + 2(2) - \frac{2^3}{3} = \frac{10}{3}$$

For this problem, the horizontal approach is a little easier because it does not involve splitting an integral into two parts, and the integrand is a polynomial. □

Area in Polar Coordinates

The functions that we have seen in this section have either specified y as a function of x or the other way around. In other words, we have specified points on curves by horizontal and vertical positions. This is not the only way to describe points in the real plane, however. As shown in Figure 6.5, another way of specifying a point P in the plane is to give P's distance r to

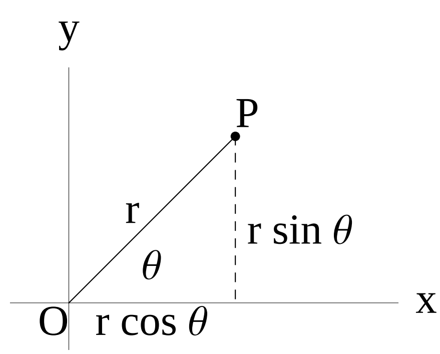

Figure 6.5: A point P in polar coordinates

the origin O and to give an angle θ that the segment OP makes with the x-axis. We say that the pair (r, θ) is a representation of P *in polar coordinates*, where θ is a *polar angle* of P.

In general, polar angles are not unique, because if $P = (r, \theta)$ in polar coordinates, then P also equals $(r, \theta + 2\pi)$ and so forth. For any point which isn't the origin, there is a unique way to write P as (r, θ) with $r > 0$ and $\theta \in [0, 2\pi)$. (The origin can have any polar angle.) By convention, if $r < 0$, then (r, θ) is the same as $(-r, \theta + \pi)$: intuitively, to walk a negative distance in the direction of θ, you walk a positive distance in the direction opposite to θ.

Example 6.28:

Polar coordinates allow us to specify curves by making r a function of θ. For example, the curve $r(\theta) = a$, where a is a positive constant, yields a circle of radius a.

As a more interesting example, consider the function $r(\theta) = \sin\theta$ for $\theta \in [0, \pi]$. Evaluating r at some simple values, we find that $r(0) = 0$, $r(\pi/2) = 1$, and $r(\pi) = 0$. Also, r increases on $[0, \pi/2]$ and decreases on $[\pi/2, \pi]$, leading to the graph on the left of Figure 6.6.

What happens if we take this function r and extend its domain to $[0, 2\pi]$? Recall $\sin\theta = -\sin(\theta - \pi)$ for any $\theta \in \mathbb{R}$. Thus, we find that for any $\theta \in [\pi, 2\pi]$, the point $(r(\theta), \theta)$ in polar coordinates is the same as $(-r(\theta - \pi), \theta)$. According to our convention for negative radii, this point is the same as $(r(\theta - \pi), \theta + \pi)$, which is the same as $(r(\theta - \pi), \theta - \pi)$ since changing a polar angle by 2π does not affect the point being plotted. This means that the graph of r on $[\pi, 2\pi]$ is the SAME as on $[0, \pi]$, so if we draw r on $[0, 2\pi]$,

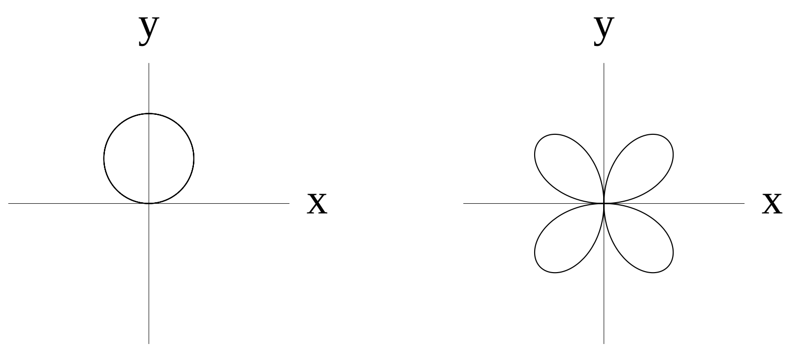

Figure 6.6: The graphs of $r = \sin\theta$ (left) and $r = \sin(2\theta)$ (right)

then we end up tracing the curve twice! (Try this with a graphing calculator using its polar mode.)

From the graph, it seems that the curve $r = \sin\theta$ produces a circle of radius $1/2$ centered at $(0, 1/2)$. We can prove this as follows. In general, when a point is represented as (x, y) in the usual Cartesian coordinates or as (r, θ) in polar coordinates, we know $x = r\cos\theta$ and $y = r\sin\theta$. (Also, $r^2 = x^2 + y^2$ by the distance formula.) The circle centered at $(0, 1/2)$ of radius $1/2$ has the equation $x^2 + (y - 1/2)^2 = 1/4$. Therefore, we have

$$\begin{aligned} x^2 + \left(y - \frac{1}{2}\right)^2 &= (x^2 + y^2) - y + \frac{1}{4} \\ &= r^2 - r\sin\theta + \frac{1}{4} = r(r - \sin\theta) + \frac{1}{4} \end{aligned}$$

This shows that $x^2 + (y - 1/2)^2 = 1/4$ iff $r = 0$ or $r = \sin\theta$, and the only point with $r = 0$ is the origin. This proves that the graph of $r = \sin\theta$ is the circle of radius $1/2$ centered at $(0, 1/2)$.

For another example, consider the curve $r = \sin(2\theta)$ for $\theta \in [0, 2\pi]$ as pictured in the graph of the right of Figure 6.6. Here, we see that $r = 0$ whenever θ is a multiple of $\pi/2$, so the curve touches the origin four times. On $[0, \pi/2]$, the curve traces out the upper-right shape looking a little like a rose petal. Next, on $[\pi/2, \pi]$, we have $\sin(2\theta) = -\sin(2(\theta - \pi/2))$, and you find that the curve traces out the same shape but rotated, making the lower-right rose petal. Similarly, on each of $[\pi, 3\pi/2]$ and $[3\pi/2, 2\pi]$, another petal is drawn. Thus, this graph is commonly called the *four-leaf rose*. (You should experiment with the graphs of $\sin(k\theta)$ for different constant values of $k \in \mathbb{N}^*$!) $\square$

Suppose f is a real function on $[a,b]$ for some $a,b \in \mathbb{R}$ with $a \leq b$. we would like to know: how do you measure the area swept out by the curve $r = f(\theta)$ in polar coordinates as θ ranges from a to b? To approach this problem, we will use our tactic of approximating f by step functions s, t such that $s(\theta) \leq f(\theta) \leq t(\theta)$ for all $\theta \in [a,b]$. To make our explanation a little simpler, let's suppose that $s(\theta) \geq 0$ for all $\theta \in [a,b]$. Suppose that s and t are both compatible with a partition $\{\theta_0, \theta_1, \ldots, \theta_n\}$ of $[a,b]$, and for each i from 1 to n, let's say s takes value s_i and t takes value t_i on the i^{th} subinterval.

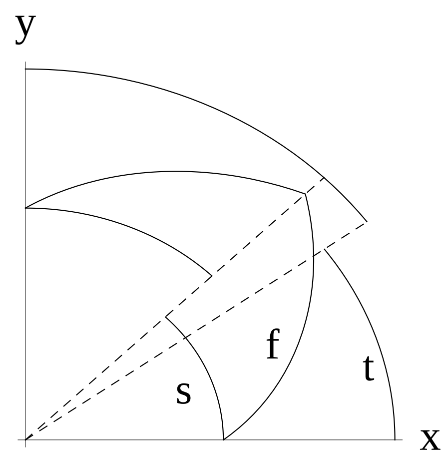

Figure 6.7: Approximating the area swept by $f(\theta)$ with $s(\theta)$ and $t(\theta)$

How do we use s and t to approximate the area swept out by f? On each subinterval of our partition, s and t are constant, so they sweep out curves of constant radius, i.e. circular sectors. (See Figure 6.7, which uses a partition into 2 pieces. The dashed sides represents the boundaries of a circular sector.) For each i, the i^{th} subinterval covers an angle of $\theta_i - \theta_{i-1}$, and thus the area of s's circular sector is $s_i^2(\theta_i - \theta_{i-1})/2$, while t's sector has area $t_i^2(\theta_i - \theta_{i-1})/2$. (You can think of each circular sector as being a "radial strip" of area, as opposed to the vertical and horizontal strips we have previously used.) Adding these up for all i, if A represents the area swept out by f, then as s underestimates the area and t overestimates the area, we have

$$\sum_{i=1}^{n} \frac{1}{2} s_i^2(\theta_i - \theta_{i-1}) \leq A \leq \sum_{i=1}^{n} \frac{1}{2} t_i^2(\theta_i - \theta_{i-1})$$

i.e.

$$\int_a^b \frac{1}{2}s^2(\theta)\,d\theta \le A \le \int_a^b \frac{1}{2}t^2(\theta)\,d\theta$$

When f^2 is integrable, the only choice for A which satsfies this inequality for all s and t is

$$A = \int_a^b \frac{1}{2}f^2(\theta)\,d\theta$$

Example 6.29:

Using the formula we just derived, we can return to the curves from Example 6.28 and find the areas associated with each. If $r = \sin\theta$ for $\theta \in [0,\pi]$, then the area swept out on $[0,\pi]$ is

$$\int_0^\pi \frac{1}{2}\sin^2\theta\,d\theta = \frac{1}{4}\int_0^\pi 1-\cos(2\theta)\,d\theta = \frac{1}{4}\left(\theta - \frac{\sin(2\theta)}{2}\right)\Bigg|_0^\pi = \frac{\pi}{4}$$

Note that if we integrate this from 0 to 2π instead, then we get $\pi/2$ as our answer, which is twice the value of $\pi/4$. This makes sense, because when we graph $r = \sin\theta$ from 0 to 2π, we trace the curve twice, so we trace out its area twice as well.

To find the area enclosed by the four-leaf rose, it is not hard to see that we need to only compute the area enclosed by one leaf and multiply by 4. By using the substitution $u = 2\theta$, and the work we just did above, we find

$$\int_0^{\pi/2} \frac{1}{2}\sin^2(2\theta)\,d\theta = \frac{1}{4}\int_0^\pi \sin^2 u\,du = \frac{1}{8}\left(u - \frac{\sin(2u)}{2}\right)\Bigg|_0^\pi = \frac{\pi}{8}$$

and thus the total area enclosed by the four-leaf rose is $\pi/2$. $\square$

6.8 Some Basic Volumes

We can extend the ideas we have developed about computing areas to help us compute volumes of three-dimensional solids. The main idea is to start with shapes that have very simple volumes, and then we use those shapes to approximate others. In two dimensions, we approximated the area of regions by using rectangles, because a rectangle has a constant height, i.e. every vertical slice through a rectangle has the same length. In contrast, if we slice a three-dimensional solid vertically (i.e. we slice it with a plane whose x-coordinate never changes, i.e. a translate of the yz-plane), then we obtain a two-dimensional shape called a *cross section*. The simplest three-dimensional shapes are those for which every cross section has the same area.

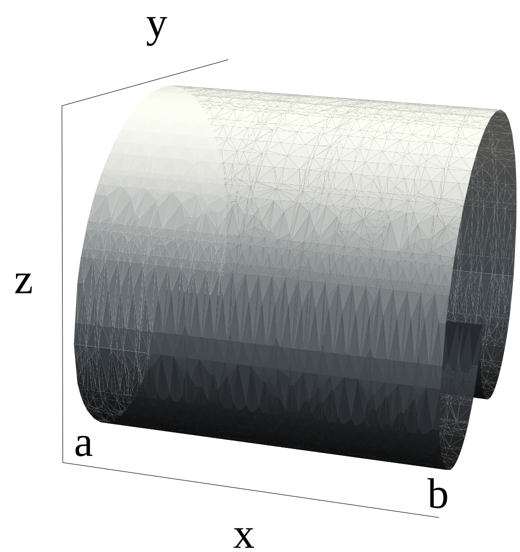

Figure 6.8: A solid with constant cross sections

First, for simplicity, consider a solid like in Figure 6.8, where every cross section is the same. When the cross sections are all translates of one another, as in this figure (where each cross section is roughly heart-shaped), we say that the solid is a *cylindrical solid*. Our solid lies from $x = a$ to $x = b$. Let's say $A(x)$ represents the area of the cross section we obtain when we take a slice at x. Because all the cross sections are the same, $A(x)$ is a constant, say $A(x) = C$ for all $x \in [a, b]$. Therefore, the volume of this solid should be C times the width of the solid, i.e. $V = C(b - a)$.

A few examples of cylindrical solids are especially well-known. When the cross sections are all congruent rectangles, and the sections are translated in the direction of the x-axis, our solid is a rectangular box! The volume of a box with dimensions l, w, and h is the product of the base area (the bases are the cross sections at the ends) times the third dimension, i.e. $V = lwh$. As another example, when the cross sections are all congruent circles, the solid is a *circular cylinder* (like a can). If the base has radius r and the can has height h (think of standing the can straight up), then the volume of the can is h times the area of the base, i.e. $\pi r^2 h$. As another example, when the cross sections are all congruent copies of one polygon, like a triangle, we also call the solid a *prism*.

Remark. It turns out that we don't need the cross sections to be EXACTLY the same to use the $C(b-a)$ formula for volume. All that matters is the area of the cross sections, not their specific shapes. Thus, any shape with every cross section having area C from $x = a$ to $x = b$ will also have volume $C(b-a)$.

Once we know how to compute the volume of cylindrical solids, it is straightforward to compute the volume of a solid which can be broken into finitely many cylindrical solids. More formally, let's say we have a solid and a partition $\{x_0, \ldots, x_n\}$ of $[a, b]$ such that for each i from 1 to n, from $x = x_{i-1}$ to $x = x_i$, the solid looks like a cylindrical solid. (You can think of the solid as being n slices of bread stuck together, where the different slices may have different shapes and areas.) Let's say that $A(x)$ represents the area of the cross section we obtain by slicing at x for each $x \in [a, b]$. For each i, on (x_{i-1}, x_i), our solid is a cylindrical solid, so $A(x)$ equals some constant A_i on (x_{i-1}, x_i), i.e. A is a step function compatible with our partition. The volume of that i^{th} solid is $A_i(x_i - x_{i-1})$. Adding these up for all i, we find that the total volume of the solid from $x = a$ to $x = b$ is

$$\sum_{i=1}^{n} A_i(x_i - x_{i-1}) = \int_a^b A(x)\,dx$$

Now, let's consider an arbitrary solid whose cross sections lie from $x = a$ to $x = b$, and let's say $A(x)$ is the area of the cross section we get by slicing at x for each $x \in [a, b]$. As the previous calculation suggests, let's approximate A with step functions s and t such that $s(x) \le A(x) \le t(x)$ for all $x \in [a, b]$. Corresponding to s, we can imagine a cylindrical solid whose cross sections have area $s(x)$ at each x, and we can do something similar for t. Because $s \le A \le t$ on $[a, b]$, it's plausible that the volume V of our arbitrary solid should be between the volumes of the solids corresponding to s and t. Thus,

$$\int_a^b s(x)\,dx \le V \le \int_a^b t(x)\,dx$$

Since this is the case for all lower step functions s and upper step functions

t, this means that when A is integrable, we have

$$V = \int_a^b A(x)\, dx$$

Intuitively, this means that to compute the volume, you "add up" all the cross-sectional areas obtained by slicing from $x = a$ to $x = b$. We demonstrate with a couple examples.

Example 6.30:

Let's use our volume formula above to compute the volume of a square pyramid whose base has side lengths of L and whose height is h. To obtain convenient cross sections, let's orient the pyramid on its side, as shown in Figure 6.9. Thus, the tip of the pyramid is at $\mathrm{O} = (0, 0, 0)$ and the base is at $x = L$. Note the OP segment in particular: in the xz-plane (so $y = 0$ throughout), it has slope L/h, so its equation is $z = Lx/h$.

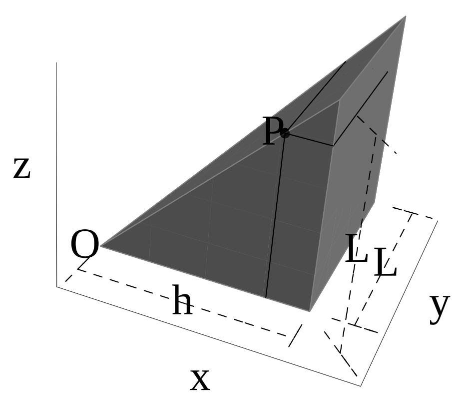

Figure 6.9: A square pyramid on its side with cross sections in two directions

For each $x \in [0, L]$, we need to figure out the area $A(x)$ of the cross section obtained by slicing vertically at x. These cross sections are squares, so we need to find the side length of the square. To do this, consider the point P in the figure, which is a corner of the cross section (outlined by black lines in the yz-plane). The z-coordinate of P is the side length of the square cross section. As $z = Lx/h$ on that line, we have

$$A(x) = \left(\frac{Lx}{h}\right)^2 = \frac{L^2}{h^2}x^2$$

and therefore
$$V = \int_0^h \frac{L^2}{h^2}x^2\,dx = \frac{L^2}{h^2} \cdot \frac{h^3}{3} = \frac{L^2h}{3}$$

In other words, the volume of this pyramid is the base area (L^2) times $h/3$. This formula can be proven in general for a pyramid with a base of any shape. In general, if the base of a pyramid of height h has area B, then for any $x \in [0, h]$, the cross section at x is geometrically similar to the base, with all side lengths being proportional with ratio x/h. Thus, $A(x) = B(x/h)^2$, and the same calculations show $V = Bh/3$. A special case of this is a cone where the base has radius r, in which case we have $V = \pi r^2 h/3$. $\square$

Remark. This is not the only way to compute the volume of the square pyramid. The way we used involves taking vertical slices, i.e. at constant values of x. Just like how we can compute areas by integrating by x or by y, we can compute the square pyramid's volume by taking slices in a different direction. Let's say we instead take slices at constant values of z: one such "horizontal slice" in the xy-plane is outlined in Figure 6.9.

Therefore, we want to compute a cross-sectional area $A(z)$ for each $z \in [0, L]$. The cross sections are trapezoids; let's say the point $P = (x, 0, z)$ in the figure is the lower-left corner of the cross section (it has the smallest x and y values of that section). The two bases run parallel to the y-axis. The base on the right (with the larger x-coordinate) has length L. The other base has length proportional to z. When $z = 0$ at the very bottom of the pyramid, that base has length 0 (i.e. the trapezoid becomes a triangle). When $z = L$ at the very top of the pyramid, that base has length L (i.e. the trapezoid becomes a line segment). Thus, the base on the left has length z.

Lastly, the width of the trapezoid is the length of the segment parallel to the x axis, $h - x$. As we found in the previous example, $z = Lx/h$, so $x = hz/L$. Since the area of a trapezoid is the average of its base lengths times the width, we have

$$A(z) = \frac{1}{2}(L + z)\left(h - \frac{hz}{L}\right) = \frac{1}{2}\left(Lh - hz + hz - \frac{hz^2}{L}\right) = \frac{1}{2}\left(Lh - \frac{hz^2}{L}\right)$$

and so

$$V = \frac{1}{2}\int_0^L Lh - \frac{hz^2}{L}\,dz = \frac{1}{2}\left(Lhz - \frac{hz^3}{3L}\right)\Big|_0^L = \frac{1}{2}\left(L^2h - \frac{L^3h}{3L}\right) = \frac{L^2h}{3}$$

We obtain the same answer we got in the example. However, this approach took more work. This shows that much like with area problems, volume problems can be sliced in several different ways, and some ways are easier than others.

Example 6.31:

Let's calculate the volume of a sphere of radius r. If we place the center of the sphere at the origin for simplicity, then the points (x, y, z) of the sphere satisfy the equation $x^2 + y^2 + z^2 = r^2$, and we have cross sections for each $x \in [-r, r]$.

For any $x \in [-r, r]$, we have $y^2 + z^2 = r^2 - x^2$, where we think of x as a constant for this section (i.e. this cross section lies in a translate of the yz-plane). This means that when we slice at x, the cross section is a circle of radius $\sqrt{r^2 - x^2}$. Thus, $A(x) = \pi(\sqrt{r^2 - x^2})^2$ and

$$V = \int_{-r}^{r} \pi(r^2 - x^2)\, dx = \pi \left(r^2 x - \frac{x^3}{3} \right) \Big|_{-r}^{r} = \frac{4}{3}\pi r^3$$

□

Solids of Revolution

A sphere is a special case of a type of solid called a *solid of revolution*. A solid of revolution is obtained by taking a two-dimensional shape and then rotating it about an axis. For instance, the sphere $x^2 + y^2 + z^2 = r^2$ can be obtained by taking the circle $x^2 + y^2 = r^2$ and spinning it about the x-axis. Since solids of revolution are obtained by spinning, they tend to produce circular cross sections, so it is not hard to compute their volumes.

Example 6.32:

Let's take the region under $y = x^2$ from $x = 0$ to $x = 1$ and rotate it about the x-axis. To find the volume of this shape, for each $x \in [0, 1]$ we want to know the cross-sectional area $A(x)$ that is produced when we slice at x. The cross section is a solid circle produced by taking the segment from $(x, 0)$ to (x, x^2) and spinning it, so the radius of that circle is x^2. Therefore,

$$A(x) = \pi(x^2)^2 = \pi x^4$$

and thus

$$V = \int_0^1 \pi x^4 \, dx = \left(\frac{\pi x^5}{5} \right) \Bigg|_0^1 = \frac{\pi}{5}$$

□

As in the previous example, when a two-dimensional region is spun about an axis that it touches, we get solid circles as our cross sections. Imagining these sections as little disks of volume, this approach for computing the volume is sometimes called the *disk method*. Some solids of revolution do not have disk-shaped cross sections, as this next example demonstrates:

Example 6.33:

Let's take the region between $y = x$ and $y = \sqrt{x}$ (we've seen that these intersect at $(0, 0)$ and $(1, 1)$) and rotate it about the y-axis. Since the axis of rotation is the y-axis, instead of the x-axis, it makes sense to slice cross sections in terms of y. Hence, we rewrite our curves in terms of y as $x = y$ and $x = y^2$.

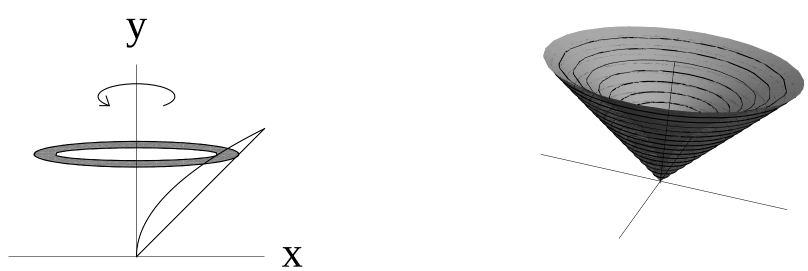

Figure 6.10: Rotating $x = y$ and $x = y^2$ about the y-axis (3D figure on right)

Now, for any $y \in [0, 1]$, we want a cross-sectional area $A(y)$. We obtain our cross section by rotating the segment from (y^2, y) to (y, y) about the y-axis (recall that $y^2 \le y$ for $y \in [0, 1]$, so the point (y^2, y) is further to the left). The shape which is obtained is a "doughnut" whose inner circle has radius y^2 and whose outer circle has radius y. See Figure 6.10 for an illustration, where the dashed curves denote the inner and outer circles of our cross section. (The figure also has a curved arrow denoting the axis of rotation.) Frequently, this kind of cross section is called a *washer*, since it looks like washers which are used in construction, and this method of computing volume is called the *washer method*.

The area of our washer is the area of the outer circle minus the area of the inner circle, so

$$A(y) = \pi y^2 - \pi(y^2)^2 = \pi(y^2 - y^4)$$

We thus have

$$V = \int_0^1 \pi(y^2 - y^4)\,dy = \pi\left(\frac{y^3}{3} - \frac{y^5}{5}\right)\Bigg|_0^1 = \frac{2\pi}{15}$$

□

Example 6.34:

Let's take the region between $y = x$ and $y = \sin x$ on $[0, \pi/2]$ and rotate it about the axis $y = 2$. Because we are rotating about a horizontal axis, let's cut our region into vertical cross sections, so for each $x \in [0, \pi/2]$, we want the area of a cross section $A(x)$. The region we are rotating to form the cross section lies in between (x, x) and $(x, \sin x)$, so we will obtain washers as cross sections. See Figure 6.11.

Note that for all $x \in [0, \pi/2]$, $\sin x \le x \le 2$ (we showed this while proving that $(\sin x)/x \to 1$ as $x \to 0$). This means that the point (x, x) is closer to the axis $y = 2$ (its distance from that axis is $2 - x$) than the point $(x, \sin x)$ (whose distance is $2 - \sin x$). Therefore, the washer has inner radius $2 - x$ and outer radius $2 - \sin x$, so

$$A(x) = \pi(2 - \sin x)^2 - \pi(2 - x)^2 = \pi(4\sin x + \sin^2 x + 4x - x^2)$$

We can therefore compute the volume as

$$\begin{aligned}
V &= \pi \int_0^2 4\sin x + \sin^2 x + 4x - x^2\,dx \\
&= \pi\left(-4\cos x + 2x^2 - \frac{x^3}{3}\right)\Bigg|_0^2 + \pi\int_0^2 \sin^2 x\,dx \\
&= \pi\left(-4\cos 2 + 8 - \frac{8}{3}\right) - \pi\,(-4) + \frac{\pi}{2}\int_0^2 1 - \cos(2x)\,dx \\
&= \frac{28\pi}{3} - 4\pi\cos 2 + \frac{\pi}{2}\left(x - \frac{\sin(2x)}{2}\right)\Bigg|_0^2 \\
&= \frac{28\pi}{3} - 4\pi\cos 2 + \pi - \frac{\pi\sin 4}{4} = \pi\left(\frac{31}{3} - 4\cos 2 - \frac{\sin 4}{4}\right)
\end{aligned}$$

□

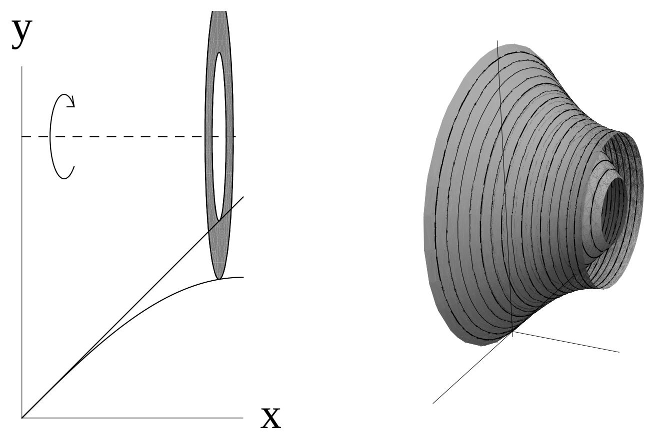

Figure 6.11: A solid of revolution rotated about $y = 2$ (3D figure on right)

Volumes of Revolution with Shells

Sometimes, it is difficult to use disks or washers to compute the volume of a solid of revolution. For instance, suppose we want to find the volume when the region under the graph of $y = 2x + x^3$ on $[0, 1]$ is rotated about the y-axis. Hence, the cross sections are sliced horizontally, and in order to find the area of one, we need to be able to solve for x in terms of y. This is quite difficult to do for the curve $y = 2x + x^3$. Hence, we'd like an alternate method for volumes of revolution which will not require us to solve for x in terms of y.

Before describing the method in general, let's consider the simplified problem of rotating the region under a constant function $f(x) = C$ from $x = a$ to $x = b$ about the y-axis (we may suppose $0 \le a < b$). The three-dimensional shape we obtain is called a *cylindrical shell* (similar to the outer layers of a can). We can find the volume of this shell by using washers. For each y from 0 to C, if we make a horizontal slice at y, then our cross section is a washer with outer radius b and inner radius a. Thus, we have $A(y) = \pi(b^2 - a^2)$ and thus the volume is $\pi C(b^2 - a^2)$. We can rewrite that as

$$V = \pi C(b^2 - a^2) = \pi C(b + a)(b - a) = 2\pi C \left(\frac{b + a}{2}\right)(b - a)$$

This form of the answer has an intuitive physical explanation, as is illustrated in Figure 6.12. Let's say we denote $(b + a)/2$, the average of a and b,

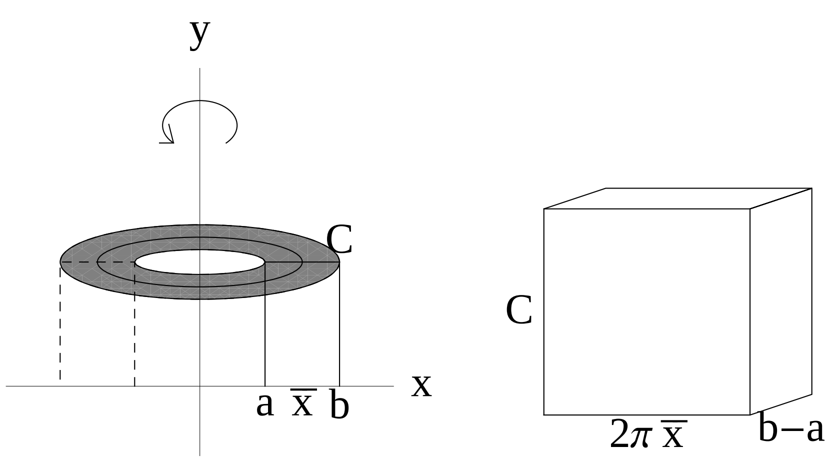

Figure 6.12: A cylindrical shell (left), and an unrolled shell (right)

by $\overline{x}$. Looking down from the top, our shell looks like a bunch of concentric circles, and the average radius of these circles is $\overline{x}$. If we cut the shell and unroll it flat, then when $b-a$ is small, our shell roughly looks like a box whose length is the circumference of the circle with radius $\overline{x}$.[3] Thus, $C(2\pi\overline{x})(b-a)$ represents the volume of that unrolled shell.

Now that we know how to obtain the volume of one shell, corresponding to a function of constant height, we can find the volume obtained from rotating a step function. Suppose s is a step function on $[a,b]$ compatible with a partition $\{x_0, x_1, \dots, x_n\}$, where s takes the value s_i on (x_{i-1}, x_i) (we may once again suppose $0 \le a < b$). For each i from 1 to n, if we rotate the region in the i^{th} subinterval about the y-axis, then our previous calculations show that we obtain a shell of volume $2\pi\overline{x}_i s_i(x_i - x_{i-1})$, where $\overline{x}_i = (x_i + x_{i-1})/2$. Thus, the total volume of the rotated solid is the sum of the volumes of these n shells:

$$V = \sum_{i=1}^{n} 2\pi\overline{x}_i s_i(x_i - x_{i-1})$$

Unfortunately, this is NOT the same as the integral

$$V = \int_a^b 2\pi x s(x)\,dx$$

[3] We will not prove this, but it is shown in many other books how to use integrals to compute the lengths of curves, and this is used to prove the circumference of a circle of radius r is $2\pi r$.

because $xs(x)$ is usually not a step function. However, the summation and the integral are pretty close when the subintervals have small length, because if $x_i - x_{i-1}$ is small, then $\overline{x}_i$ is pretty close to any value in (x_{i-1}, x_i). We will not go through the details, but it can be shown that for any desired accuracy $\epsilon > 0$, there is some $\delta > 0$ (depending on ϵ and s) such that whenever we refine the partition used for s so that each subinterval has width at most $\delta > 0$, and we compute the s_i and $\overline{x}_i$ values accordingly, we have

$$\left| \int_a^b 2\pi x s(x)\, dx - \sum_{i=1}^{n} 2\pi \overline{x}_i s_i (x_i - x_{i-1}) \right| < \epsilon$$

Thus, the volume can be approximated as well as we like by an integral.

We are now ready to handle arbitrary functions. Suppose that f is a real function on $[a, b]$ with $0 \le a < b$. When f is continuous on $[a, b]$, we know that f is integrable on $[a, b]$ by Theorem 5.35, and in fact the proof of that theorem shows that f can be approximated as well as we like by step functions. More specifically, for any $\epsilon > 0$, there is a step function s on $[a, b]$ such that $|f(x) - s(x)| < \epsilon$ for all $x \in [a, b]$. It follows that the integrals

$$\int_a^b 2\pi x s(x)\, dx \qquad \text{and} \qquad \int_a^b 2\pi x f(x)\, dx$$

can be made arbitrarily close to one another (we omit the details). Using those ideas, it can therefore be proven that the volume obtained by rotating the region under f about the y-axis is

$$V = \int_a^b 2\pi x f(x)\, dx$$

When we use this formula to compute volume, we say we are using the *shell method*. The shell method is useful when it is difficult to solve the equation $y = f(x)$ for x in terms of y. When using the shell method, think of x as being the radius of a typical shell, $f(x)$ as being the height of a shell, and dx as being the infinitesimal thickness of a shell (fulfilling the same role as $x_i - x_{i-1}$ for step functions).

Example 6.35:

Now that we have developed the shell method, let's use it to find the volume obtained by rotating the region under $y = 2x + x^3$ on $[0, 1]$ about the y-axis.

We compute

$$V = \int_0^1 2\pi x(2x + x^3)\,dx = 2\pi \int_0^1 2x^2 + x^4\,dx = 2\pi \left(\frac{2}{3}x^3 + \frac{1}{5}x^5\right)\Bigg|_0^1 = \frac{26\pi}{15}$$

□

As with other volume calculations, the shell method does not have to solely be used with a function of x being rotated about the y-axis. In general, when we rotate a region about an axis, the shell method is convenient when we want to slice a region in a direction parallel to the axis of rotation (whereas with the washer method, we slice the region perpendicular to the axis of rotation).

Example 6.36:

Let's consider the region between the graph of $y = \sqrt{x}$ and the x-axis on $[0, 1]$ and rotate it about the axis $y = -1$. We can find the volume of this solid by using both washers and shells. A picture is shown in Figure 6.13, where a typical point (x, y) on the curve $y = \sqrt{x}$ (or $x = y^2$) is marked, and vertical and horizontal slices are both drawn with some lengths marked.

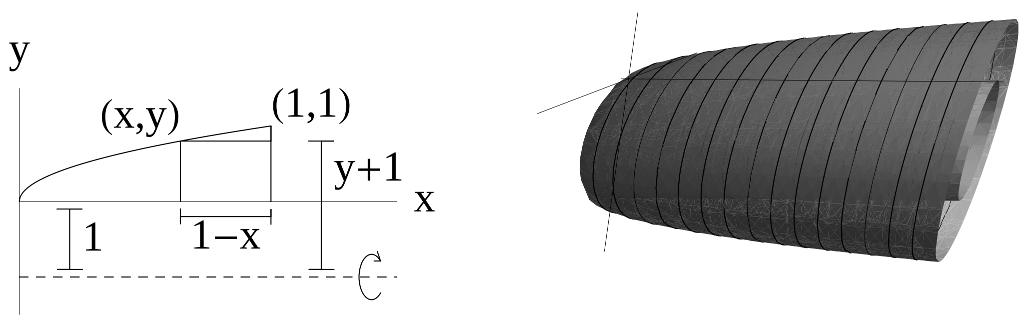

Figure 6.13: The region under $y = \sqrt{x}$ (i.e. $x = y^2$) rotated about $y = -1$, using washers and shells

If we use washers, then since our axis is horizontal, we want to slice our region vertically. If we slice at $x \in [0, 1]$, then we obtain a washer whose outer radius is $y + 1 = \sqrt{x} + 1$ and whose inner radius is 1. Thus, we obtain

$$\begin{aligned} V &= \int_0^1 \pi(\sqrt{x} + 1)^2 - 1^2\,dx = \pi \int_0^1 x + 2\sqrt{x}\,dx \\ &= \pi \left(\frac{x^2}{2} + \frac{2x^{3/2}}{3/2}\right)\Bigg|_0^1 = \frac{11\pi}{6} \end{aligned}$$

If we use shells, however, then we are slicing our region horizontally, so we should solve for x in terms of y to get $x = y^2$ for $y \in [0, 1]$. When we slice at y, we obtain a shell with radius $y + 1$ (the vertical distance between our slice and the axis) and of width $1 - x = 1 - y^2$. Thus, the volume of a shell is $2\pi(y + 1)(1 - y^2)\, dy$ where we think of dy as being a negligible thickness. Integrating, we obtain

$$\begin{aligned} V &= \int_0^1 2\pi(y+1)(1-y^2)\, dy = 2\pi \int_0^1 1 + y - y^2 - y^3\, dy \\ &= 2\pi \left(y + \frac{y^2}{2} - \frac{y^3}{3} - \frac{y^4}{4} \right)\Bigg|_0^1 = \frac{11\pi}{6} \end{aligned}$$

so we get the same answer this way as well. □

6.9 Exercises

For Exercises 1 through 5, find the area enclosed between the following curves by slicing with respect to x or y. In these problems, it's a good idea to make a rough sketch and find points of intersection.

1. $y = x \sin x$ and $y = x$ for $x \in [0, \pi/2]$
2. $y = \sqrt{1 - x^2}$ and $y = |x|$ (See also Exercise 6.9.12.)
3. $x = y^{1/3}$ and $x = \sqrt{y}$
4. $y = \arcsin x$ and $y = x$ for $x \in [0, 1]$ (Hint: Use integration by parts.)
5. $y = \sin x$ and $y = \cos x$ for $x \in [0, 2\pi]$ (Hint: You'll want to break this into three regions.)
6. For each $c > 0$, the curves $y = x^2$ and $y = cx^3$ intersect when $x = 0$ and $x = 1/c$. Find the value of c if the area between these curves is $2/3$.
7. Redo Exercises 2 through 4 above, except use a different direction of slicing the area from the one you used the first time. (Your integral may need to be computed in two pieces.)

8. Suppose $0 < r < R$. Consider the circle C_1 given by the equation $x^2 + y^2 = R^2$, and we make a second circle C_2 of radius r such that C_2's center is inside C_1 and C_1 intersects C_2 when $x = \pm r$. (See Figure 6.14.) The crescent-shaped region inside C_2 but outside C_1 is called a *lune*. Find the area of the lune.

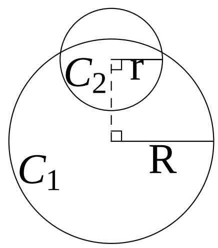

Figure 6.14: A lune between circles of radii r and R

For Exercises 9 through 11, compute the areas enclosed by the following curves in polar form where θ goes from 0 to 2π:

9. $r = 2\,|\cos\theta|$
10. $r = 1 + \cos\theta$
11. $r = \sqrt{|\cos\theta|}$

12. Solve Exercise 6.9.2 by writing the region in polar form and finding its area that way.

13. A solid has, as a base in the xy-plane, the circle $x^2 + y^2 = 1$. For each $x \in [-1, 1]$, the cross section is a square perpendicular to the x-axis (i.e. in a translate of the yz-plane). Find the volume of the solid.

14. A solid has, as a base in the xy-plane, the points $(x, y, 0)$ satisfying $0 \le y \le 4 - x^2$. (In other words, the base is the region under the parabola $4 - x^2$ and above the x-axis.) For each x, the vertical cross section is an equilateral triangle. Find the volume of the solid.

15. For the sphere $x^2 + y^2 + z^2 = r^2$, where $r > 0$, the *cap* of height h is the set of points (x, y, z) inside the sphere with $z \ge r - h$, where $0 < h < r$. Find the volume of this cap.

16. If $0 < r < R$ and $h > 0$, then a *frustrum* of a cone with top radius r, bottom radius R, and height h is the solid obtained by taking a circular cone of base radius R and removing from the top a circular cone of base radius r, where h is the perpendicular distance between the bases. (See Figure 6.15.) Find the volume of the frustrum. (You should double-check that your answer gives the volume of a cone of base radius R when $r = 0$).

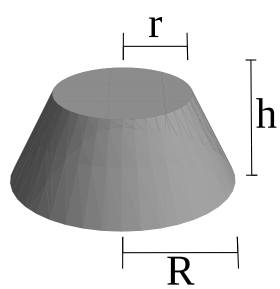

Figure 6.15: A frustrum of a cone with top radius r, bottom radius R, and height h

17. Suppose $0 < h < r$, and a solid is created by taking a sphere of radius r and drilling out a circular hole throughout the sphere of radius h. (More precisely, remove the portion of the sphere which intersects a circular cylinder of base radius h and length $2r$ which has the same center as the center of the sphere.) What is the volume of the solid?

For Problems 18 through 20, find the volume of the given solid of revolution using either washers or shells.

18. The solid obtained by rotating the region between $y = \sin x$ and $y = \cos x$ for $x \in [0, \pi/4]$ about the x-axis

19. The solid obtained by rotating the region between $y = 1 - \cos x$ and $y = 0$ for $x \in [0, \pi/2]$ about the y-axis

20. The solid obtained by rotating the region between $y = x$ and $y = \sqrt{x}$ about $x = -1$

21. Let $0 < r < R$ be given. The *torus* with radii r and R is the solid created by taking the circle $(x - R)^2 + z^2 = r^2$ in the xz-plane (i.e. where $y = 0$) and rotating it about the z-axis. A torus looks like a three-dimensional bagel or doughnut.

    (a) Show that the torus consists of the points (x, y, z) satisfying the equation
    $$z^2 + \left(\sqrt{x^2 + y^2} - R\right)^2 \leq r^2$$

    (b) Using the equation from part (a), compute the volume of the torus by taking cross sections for each value of z. This is essentially computing the volume by using washers.

    (c) Compute the volume of the torus by using shells.

    (d) A theorem of Pappus says that the volume of a solid of revolution is equal to Ad, where A is the area of the two-dimensional region being rotated and d is the distance traveled by that region's center of mass as it completes a full rotation. Use this theorem to compute the volume of the torus. You may use the formula for the circumference of a circle to find d.

6.10 Approximating Integrals

In this section, we mention some ways in which we can approximate the value of a definite integral $\int_a^b f(x)\,dx$ where f is continuous on $[a, b]$. There are several reasons why this is useful. First of all, when we are integrating a function without an elementary antiderivative, frequently we can't obtain an exact value of the integral, so an approximation may be the next best bet. Second, integral approximations can be computed quickly, by hand or by machine, whereas finding antiderivatives requires more ingenuity (by guessing the correct substitution, for instance). Third, if you don't know an exact formula for your function, but you know a few points on that function's graph, then you can still use those points to produce a numerical approximation. This is especially useful when the points come from experimentally-measured data.

Riemann Sums

As a starting point for approximating the integral of f, the simplest kind of approximation is obtained by replacing f with a constant. Recall that the Mean Value Theorem for Integrals says that

$$\int_a^b f(x)\,dx = f(c)(b-a) \qquad \text{for some } c \in [a,b]$$

i.e. $f(c)$ is the average value of f on $[a,b]$. This means that the best constant to use when approximating f is $f(c)$. However, since we might not know exactly where c is, we can try using some other point in $[a,b]$ to obtain a guess at the average value. This gives us the approximation

$$\int_a^b f(x)\,dx \approx f(x^*)(b-a)$$

where x^* is some point we choose in $[a,b]$ called a *sample point*. Different approximation schemes pick different choices of x^*.

However, $f(x^*)$ will not usually be equal to the average value $f(c)$. Do we have any guarantee of how apart $f(x^*)$ and $f(c)$ are? In the worst case, the distance between them is the distance between the maximum and minimum values of f on $[a,b]$, which we defined as $\text{extent}(f,[a,b])$. This shows that the approximation tends to be worse for functions which span a large extent. Furthermore, that error will be multiplied by $b-a$, the width of the interval, so approximations over large intervals tend to be worse than approximations over small intervals.

To address these issues, rather than use one value from f for the WHOLE interval $[a,b]$, we break up $[a,b]$ into pieces and use a value from f for each piece. Let's say that we partition $[a,b]$ as $\{x_0, x_1, \ldots, x_n\}$, and for each i from 1 to n, let's write Δx_i to denote the width $x_i - x_{i-1}$ of the i^{th} subinterval. On the i^{th} subinterval, we choose some sample point x_i^* in $[x_{i-1}, x_i]$. This gives us the approximation

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^n f(x_i^*)\Delta x_i$$

A sum of this form is often called a *Riemann sum* for f on $[a,b]$. Frequently, to simplify calculations, all of the widths are chosen to be the same, yielding the approximation

$$\frac{b-a}{n}\sum_{i=1}^n f(x_i^*)$$

Theorem 3.66 guarantees that by making n large enough, the extent of f on each piece will become arbitrarily small, and it follows that the Riemann sum will become arbitrarily close to the exact value of the integral. Now, we define a few approximation schemes based on this idea.

Definition 6.37. Let $a, b \in \mathbb{R}$ with $a < b$, $f : [a, b] \to \mathbb{R}$, and $n \in \mathbb{N}^*$ be given. Suppose we partition $[a, b]$ as $\{x_0, x_1, \ldots, x_n\}$ where for each i from 0 to n, $x_i = a + i(b - a)/n$ (so the widths are all $(b - a)/n$).

The *left-endpoint approximation* to the integral of f on $[a, b]$ is

$$L_n(f; [a, b]) = \frac{b-a}{n} \sum_{i=1}^{n} f(x_{i-1})$$

i.e. $x_i^* = x_{i-1}$, the left endpoint of the i^{th} subinterval. When we use this formula, we frequently say we are using the *Left-Endpoint Rule* for approximating.

The *right-endpoint approximation* is

$$R_n(f; [a, b]) = \frac{b-a}{n} \sum_{i=1}^{n} f(x_i)$$

i.e. $x_i^* = x_i$, the right endpoint of the i^{th} subinterval. We similarly say that we are using the *Right-Endpoint Rule* when we use this method.

The *midpoint approximation* is

$$M_n(f; [a, b]) = \frac{b-a}{n} \sum_{i=1}^{n} f\left(\frac{x_{i-1} + x_i}{2}\right)$$

i.e. x_i^* is the midpoint of the i^{th} subinterval. We say we are using the *Midpoint Rule* when we use this method.

Sometimes we will just write L_n, R_n, or M_n for these quantities if f and $[a, b]$ are understood by context.

Example 6.38:

Let's estimate

$$\int_{\pi}^{2\pi} \frac{\sin x}{x}\, dx$$

It is known that this integrand does not have an elementary antiderivative.

Using $n = 4$, we compute

$$\begin{aligned} L_4 &= \frac{\pi}{4}\left(\frac{\sin\pi}{\pi} + \frac{4\sin(5\pi/4)}{5\pi} + \frac{2\sin(3\pi/2)}{3\pi} + \frac{4\sin(7\pi/4)}{7\pi}\right) \\ &= -\frac{1}{4}\left(\frac{24\sqrt{2}}{35} + \frac{2}{3}\right) \approx -0.409103 \end{aligned}$$

and

$$\begin{aligned} R_4 &= \frac{\pi}{4}\left(\frac{4\sin(5\pi/4)}{5\pi} + \frac{2\sin(3\pi/2)}{3\pi} + \frac{4\sin(7\pi/4)}{7\pi} + \frac{\sin(2\pi)}{2\pi}\right) \\ &= L_4 \approx -0.409103 \end{aligned}$$

and

$$\begin{aligned} M_4 &= \frac{\pi}{4}\left(\frac{8\sin(9\pi/8)}{9\pi} + \frac{8\sin(11\pi/8)}{11\pi} + \frac{8\sin(13\pi/8)}{13\pi} + \frac{8\sin(15\pi/8)}{15\pi}\right) \\ &\approx -0.446179 \end{aligned}$$

Note that L_4 and R_4 are the same because $\sin\pi = \sin(2\pi) = 0$. In fact, in this example, we will have $L_n = R_n$ for all values of n. To see how the approximations behave with increasing values of n, here is a table giving some more approximations:

n	L_n	M_n
4	-0.409103	-0.446179
8	-0.427641	-0.436861
16	-0.432251	-0.434553
32	-0.433402	-0.433977

Further testing suggests that the correct value to six decimal places is $A \approx -0.433785$. This gives us a way of estimating the error of these approximations as a function of n. Let's define $E_{L_n} = A - L_n$, and similarly $E_{R_n} = A - R_n$ and $E_{M_n} = A - M_n$. Thus, E_{L_n} is the amount you need to add to L_n to get the exact value, i.e. the error of using L_n as an approximation. We have the following table:

n	E_{L_n}	E_{M_n}
4	-0.024682	0.012394
8	-0.006144	0.003076
16	-0.001534	0.000768
32	-0.000383	0.000192

Note that every time n doubles, the errors are divided roughly by 4. This suggests that the error for these schemes is proportional to $1/n^2$. In numerical analysis courses, it is possible to prove this conjecture, but the proof will not be covered here. □

Example 6.39:

For another example, let's consider approximating

$$\int_0^1 \frac{1}{1+x^2}\,dx$$

In this case, we know the exact value of this integral is $\arctan(1) - \arctan 0 = \pi/4$. Thus, by approximating this integral, we can obtain estimates for the value of π.

We have

$$L_4 = \frac{1}{4}\left(\frac{1}{1} + \frac{1}{1+(1/4)^2} + \frac{1}{1+(1/2)^2} + \frac{1}{1+(3/4)^2}\right) \approx 0.845294$$

and

$$R_4 = \frac{1}{4}\left(\frac{1}{1+(1/4)^2} + \frac{1}{1+(1/2)^2} + \frac{1}{1+(3/4)^2} + \frac{1}{2}\right) \approx 0.720294$$

and

$$M_4 = \frac{1}{4}\left(\frac{1}{1+(1/8)^2} + \frac{1}{1+(3/8)^2} + \frac{1}{1+(5/8)^2} + \frac{1}{1+(7/8)^2}\right) \approx 0.786700$$

It is known that $\pi/4$ is roughly 0.785398, so the midpoint approximation is by far the most accurate. This makes sense since the function $1/(1+x^2)$ is decreasing on $[0,1]$, so when we pick a sample point to guess where the average value is obtained on a subinterval, the midpoint is a better guess than each endpoint. □

As the previous example shows, when a function is monotone, the left and right endpoint rules perform rather poorly. In fact, you can show in Exercise 6.11.7 that one endpoint is guaranteed to overestimate the exact value of the integral and the other is guaranteed to underestimate the exact value. Thus, perhaps the average of the two estimates does a better job. This leads to the following approximation scheme:

Definition 6.40. Suppose that a, b, f, n, and a partition $\{x_0, x_1, \dots, x_n\}$ are given the same as in Definition 6.37. The *trapezoid approximation* is

$$\begin{aligned} T_n(f;[a,b]) &= \frac{L_n(f;[a,b]) + R_n(f;[a,b])}{2} \\ &= \frac{b-a}{2n}\left(f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)\right) \end{aligned}$$

When we use this formula, we say we are using the *Trapezoid Rule.*

Figure 6.16 illustrates why T_n is called the trapezoid approximation. While a Riemann sum uses a rectangle of height $f(x_i^*)$ to approximate the integral for the ith subinterval, where x_i^* is some sample point, the trapezoid approximation uses a trapezoid with the two bases $f(x_{i-1})$ and $f(x_i)$. Thus, the i^{th} subinterval is approximated with the area $\dfrac{b-a}{2n}(f(x_{i-1}) + f(x_i))$, and adding this up for all values of i gives us the value of T_n.

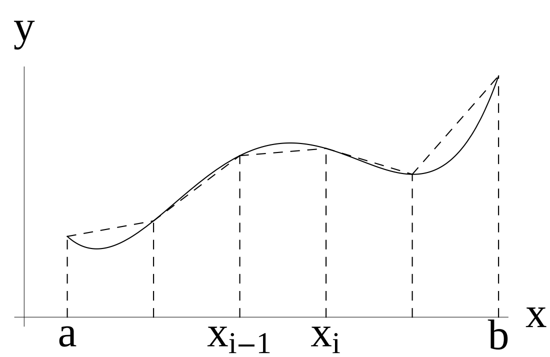

Figure 6.16: Using trapezoids to approximate the area under f

If we go back to Example 6.39, then we have

$$T_4 = \frac{1}{8}\left(\frac{1}{1} + \frac{2}{1+(1/4)^2} + \frac{2}{1+(1/2)^2} + \frac{2}{1+(3/4)^2} + \frac{1}{2}\right) \approx 0.782794$$

This is a much better estimate of $\pi/4$ than either L_n or R_n, though the midpoint approximation is still better. In fact, note that $|E_{M_n}| \approx 0.001302$ is roughly half of $|E_{T_n}| \approx 0.002604$. It can also be shown in numerical analysis courses that in general, you can expect the midpoint approximation to be twice as good as the trapezoid approximation, and also that the trapezoid error is proportional to $1/n^2$. A partial proof of this is outlined in Exercise 6.11.13.

Example 6.41:

Suppose we want to measure the area of an irregularly-shaped pond. Let's say the pond stretches 100 feet from its west end to its east end. We start at the west end of the pond, we walk east, and for every 10 feet we walk, we make a measurement of the distance between the north and south shores of the pond. In other words, if we think of the pond as occupying a region in the xy-plane from $x = 0$ to $x = 100$, then we measure the height of that region for each x coordinate which is a multiple of 10.

Let's say that $f(x)$ is the distance between the shores as we move x feet east. We summarize all our data in this table, where all measurements are in feet:

x	$f(x)$	x	$f(x)$	x	$f(x)$
0	25	40	50	80	24
10	37	50	45	90	15
20	45	60	46	100	10
30	53	70	25		

The exact area of the pond is $\displaystyle\int_0^{100} f(x)\,dx$, measured in square feet. We cannot find this exact integral because we only have finitely many values of $f(x)$ at our disposal. However, we can use approximation techniques. For instance, if we break the interval $[0, 100]$ into pieces of width 20, then we can use either midpoint or trapezoid approximations. We get

$$M_5 = 20\,(f(10) + f(30) + f(50) + f(70) + f(90)) = 3500$$

and

$$T_5 = \frac{20}{2}\,(f(0) + 2f(20) + 2f(40) + 2f(60) + 2f(80) + f(100)) = 3650$$

We can also compute

$$T_{10} = \frac{10}{2}\left(f(0) + 2\sum_{i=1}^{9} f(10i) + f(100)\right) = 3575$$

(We cannot compute M_{10} with this data because we would need the midpoints of intervals of width 10, so we'd need data for $x = 5$, 15, and so forth.) Thus, we estimate that the area of the pond is 3575 square feet. To make better approximations, we need to collect more data points. □

Simpson's Rule

While the trapezoid and midpoint rules generally work well and are easy to compute, because they use straight line segments to approximate the graph of f on each subinterval, they tend to perform poorly for functions which are highly curved. There is another approximation rule which tends to work better for highly-curved functions. It is called Simpson's Rule, named after the English mathematician Thomas Simpson. His main idea was to use parabolas instead of trapezoids to approximate the area under subintervals.

Since a parabola is the graph of a quadratic function, we need three points to specify a parabola. To be more precise, let's say we break $[a, b]$ into $2n$ pieces with the partition $\{x_0, x_1, \ldots, x_{2n}\}$. We can use the points at x_0, x_1, and x_2 on f's graph to make one parabola. We can use the points at x_2, x_3, and x_4 to make another parabola. Continuing in this manner, for each i from 1 to n, we make a parabola using the points at x_{2i-2}, x_{2i-1}, and x_{2i}. (This is why our partition uses an even number of subintervals.) See Figure 6.17.

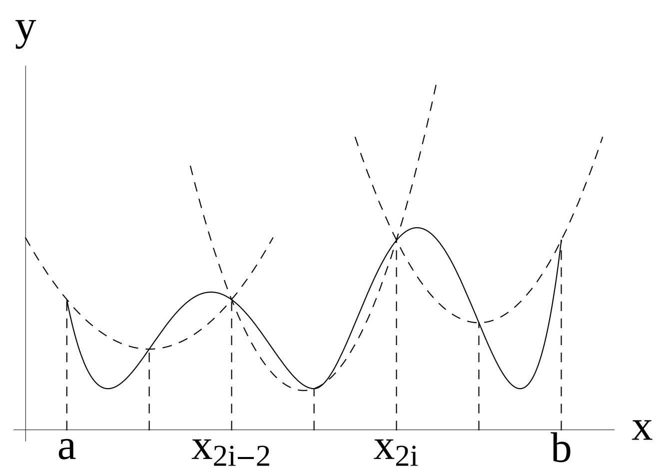

Figure 6.17: An area approximated with three parabolas

To obtain Simpson's Rule, we first want to discover the area under one parabola. We show one way of doing this: another way is outlined in Exercise 6.11.12. To make our calculations simpler, we may suppose that we are finding the area under the parabola through x_0, x_1, and x_2, where $x_1 = 0$ (this is a safe assumption because translating the parabola horizontally does not change the area underneath it). Thus, let's say that $x_2 = h$ and $x_0 = -h$.

We want to find an equation for a parabola, i.e. a function $p(x)$ of the form ax^2+bx+c for some constants a, b, c, such that $p(-h) = f(x_0)$, $p(0) = f(x_1)$, and $p(h) = f(x_2)$. This gives us the three equations

$$\begin{aligned} a(-h)^2 + b(-h) + c &= ah^2 - bh + c = f(x_0) \\ a(0)^2 + b(0) + c &= c = f(x_1) \\ a(h)^2 + b(h) + c &= ah^2 + bh + c = f(x_2) \end{aligned}$$

By subtracting the first equation from the third, we obtain

$$2bh = f(x_2) - f(x_0)$$

so

$$b = \frac{f(x_2) - f(x_0)}{2h}$$

We plug the values of b and c into the first equation and find

$$ah^2 - \left(\frac{f(x_2) - f(x_0)}{2h}\right) h + f(x_1) = f(x_0)$$

so

$$a = \frac{f(x_0) - 2f(x_1) + f(x_2)}{2h^2}$$

Now we want to integrate $p(x)$ from $-h$ to h. Since the bx term of $p(x)$ is an odd function, its integral on the interval $[-h, h]$ vanishes. Also, ax^2 and c are even functions, so their integral on $[-h, h]$ is twice their integral on $[0, h]$. Thus, we get

$$\begin{aligned} \int_{-h}^{h} p(x)\,dx &= \int_{-h}^{h} ax^2 + bx + c\,dx \\ &= 2\int_0^h \left(\frac{f(x_0) - 2f(x_1) + f(x_2)}{2h^2}\right) x^2 + f(x_1)\,dx \\ &= \left(\frac{f(x_0) - 2f(x_1) + f(x_2)}{h^2}\right) \frac{h^3 - 0^3}{3} + 2f(x_1)(h - 0) \\ &= \frac{h}{3}\left(f(x_0) + 4f(x_1) + f(x_2)\right) \end{aligned}$$

Since h is the width of a subinterval, $(b-a)/(2n)$ (since we have $2n$ subintervals in our partition), the same steps show that the area under the i^{th} parabola is

$$\frac{b-a}{6n}\left(f(x_{2i-2}) + 4f(x_{2i-1}) + f(x_{2i})\right)$$

for each i from 1 to n. Adding these up for all i gives us the following:

Definition 6.42. Let a,b,f,n be given as in Definition 6.37, and we break $[a, b]$ into $2n$ equal-width subintervals with the partition $\{x_0, x_1, \ldots, x_{2n}\}$. Then *Simpson's Rule* gives us the approximation

$$\begin{aligned}
& S_{2n}(f; [a, b]) \\
= \; & \frac{b-a}{6n}((f(x_0) + 4f(x_1) + f(x_2)) + (f(x_2) + 4f(x_3) + f(x_4)) + \cdots \\
+ \; & (f(x_{2n-2} + 4f(x_{2n-1}) + f(x_{2n}))) \\
= \; & \frac{b-a}{6n}(f(x_0) + 4f(x_1) + 2f(x_2) + \cdots + 4f(x_{2n-1}) + f(x_{2n}))
\end{aligned}$$

In other words, the width $(b - a)/(2n)$ is divided by 3, $f(x_0)$ and $f(x_{2n})$ get coefficients of 1, and the rest of the coefficients alternate between 4 and 2.

Simpson's Rule often works very well. For instance, if we go back to Example 6.39, then with $2n = 4$, we get

$$S_4 = \frac{1}{12}\left(\frac{1}{1} + \frac{4}{1+(1/4)^2} + \frac{2}{1+(1/2)^2} + \frac{4}{1+(3/4)^2} + \frac{1}{2}\right) \approx 0.785392$$

which has an error of $E_{S_4} \approx 0.000006$. This is a very small error for only using four subintervals! If we try Simpson's Rule with $2n = 10$ in Example 6.41, then we obtain the estimate of

$$S_{10} = \frac{10}{3}\left(f(0) + 4\sum_{i=0}^{4} f(20i + 10) + 2\sum_{i=1}^{4} f(20i) + f(100)\right) = 3550$$

It turns out that the error with Simpson's Rule is proportional to $1/n^4$, so the error decreases much more rapidly with Simpson's Rule than with the Midpoint Rule as n grows large, but we will not prove this here.

6.11 Exercises

For Exercises 1 through 5, a real function f and an interval $[a, b]$ are provided. Using those, compute $L_n(f; [a, b])$, $R_n(f; [a, b])$, $M_n(f; [a, b])$, $T_n(f; [a, b])$, and $S_n(f; [a, b])$ for $n = 6$ and $n = 12$.

1. $f(x) = 2x + 1$, $[a, b] = [0, 6]$
2. $f(x) = \sqrt{1 - x^2}$, $[a, b] = [-1, 1]$
3. $f(x) = \sin(x^2)$, $[a, b] = [0, \sqrt{\pi}]$
4. $f(x) = (1 + x^2)^{1/3}$, $[a, b] = [0, 6]$
5. $f(x) = 5x^3 - 4x + 1$, $[a, b] = [-3, 3]$
6. For Exercises 1, 2, and 5 above, compute the exact values of $\int_a^b f(x)\,dx$ and find the errors of your approximations from those exercises.
7. Suppose that f is a real function which is increasing on $[a, b]$. Prove that for all $n \in \mathbb{N}^*$,
$$L_n(f; [a, b]) \leq \int_a^b f(x)\,dx \leq R_n(f; [a, b])$$
What do you conclude instead if f is decreasing on $[a, b]$?
8. Certain approximations can do a very poor job if sample points are chosen poorly. Consider approximating the integral
$$\int_0^{20} \cos(\pi x)\,dx$$
Find the exact value for this integral, and also find L_{10}, R_{10}, and T_{10}. Can you explain why the error is so large?
9. Suppose that for some $m, c \in \mathbb{R}$, we have $f(x) = mx + c$ for all $x \in [a, b]$. In other words, f is a linear (or constant) function on $[a, b]$. Prove that for all $n \in \mathbb{N}^*$,
$$E_{M_n}(f; [a, b]) = E_{T_n}(f; [a, b]) = 0$$
so the midpoint and trapezoid rules produce zero error. (Suggestion: Prove this for $n = 1$ first, and then use that result to prove the error is 0 for any n.)
10. Suppose that f is a polynomial of degree at most 3 on $[a, b]$. Prove that for all $n \in \mathbb{N}^*$,
$$E_{S_{2n}}(f; [a, b]) = 0$$

11. Show that for any $a, b \in \mathbb{R}$ with $a < b$, $f : [a, b] \to \mathbb{R}$ and any $n \in \mathbb{N}^*$,

$$S_{2n}(f; [a, b]) = \frac{1}{3}T_n(f; [a, b]) + \frac{2}{3}M_n(f; [a, b])$$

Thus, Simpson's Rule gives a weighted average of the Trapezoid and Midpoint Rules.

12. This exercise shows another way to obtain the formula for Simpson's Rule. As we did with our previous derivation, let's suppose for simplicity that we are obtaining Simpson's Rule for $n = 1$, and for some $h > 0$ we have $x_0 = -h$, $x_1 = 0$, and $x_2 = h$. (Thus, h is the interval width.)

We would like to approximate the average value of f on $[-h, h]$ with a linear combination of $f(x_0)$, $f(x_1)$, and $f(x_2)$. This means we want to say

$$\frac{1}{2h}\int_{-h}^{h} f(x)\,dx \approx \alpha f(-h) + \beta f(0) + \gamma f(h)$$

for some constants $\alpha, \beta, \gamma \in \mathbb{R}$. In order to pick these constants, we impose the requirement that whenever f is a polynomial of degree at most 2, i.e. $f(x)$ has the form $ax^2 + bx + c$ for some $a, b, c \in \mathbb{R}$, then

$$\frac{1}{2h}\int_{-h}^{h} f(x)\,dx = \alpha f(-h) + \beta f(0) + \gamma f(h)$$

In other words, we require that our approximation have zero error for quadratic polynomials.

Prove that this requirement is satisfied iff $\alpha = \gamma = 1/6$ and $\beta = 2/3$. Thus, we obtain

$$\int_{-h}^{h} f(x)\,dx \approx \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2))$$

as desired.

13. In this exercise, we outline a proof of the following bound on the error from the Trapezoid Rule:

Theorem 6.43. *Let $a, b \in \mathbb{R}$ be given, and suppose f is a real function such that f'' is continuous on $[a, b]$. Let K be any number such that $|f''(x)| \leq K$ for all $x \in [a, b]$. Then for all $n \in \mathbb{N}^*$,*

$$|E_{T_n}(f; [a, b])| \leq \frac{K(b-a)^3}{12n^2} = \frac{Kn}{12}\left(\frac{b-a}{n}\right)^3$$

Hence, our error bound is proportional to $1/n^2$ and also proportional to the size of the second derivative. This makes sense, since the second derivative measures the concavity of f, i.e. how "curved" the graph of f is.

Our approach will be to first prove this bound for $n = 1$ by approximating f with a linear function, and then we will use that result to finish the proof for all n.[4]

(a) First, let's consider $n = 1$ for now. Let $P : [a, b] \to \mathbb{R}$ be defined by

$$P(x) = f(a)\frac{x-b}{a-b} + f(b)\frac{x-a}{b-a}$$

for all $x \in [a, b]$. (This is called the *Lagrange polynomial* of order 1 through the points $(a, f(a))$ and $(b, f(b))$.) Note that P is a linear function. Prove that $P(a) = f(a)$ and $P(b) = f(b)$.

(b) In the next few parts, we prove the following claim: for any $x \in [a, b]$, there is some $\xi \in [a, b]$ (ξ, the Greek letter "xi", depends on x) such that

$$f(x) - P(x) = \frac{f''(\xi)}{2}(x-a)(x-b)$$

This claim is trivial if $x = a$ or $x = b$ (by part (a)), so suppose $x \neq a$ and $x \neq b$. Define

$$\Phi(x) = \frac{f(x) - P(x)}{(x-a)(x-b)}$$

and define

$$g(t) = f(t) - P(t) - (t-a)(t-b)\Phi(x)$$

[4] This proof is adapted from the arguments found at http://math.fullerton.edu/ mathews/n2003/trapezoidalrule/TrapezoidalRuleProof.pdf and http://www.math-linux.com/spip.php?article71.

for all $t \in [a, b]$.

Prove that for all $t \in [a, b]$,

$$g''(t) = f''(t) - 2\Phi(x)$$

(Recall that x is a constant with respect to t.) Thus, g'' is continuous on $[a, b]$.

(c) From parts (a) and (b), note that $g(a) = g(b) = 0$, and also $g(x) = 0$ by the definition of $\Phi(x)$. Use Rolle's Theorem three times to prove that for some $\xi \in (a, b)$, $g''(\xi) = 0$. (You'll want to use the same idea as in Exercise 4.10.11.) From this, prove that

$$f(x) - P(x) = \frac{f''(\xi)}{2}(x - a)(x - b)$$

finishing our claim from part (b).

(d) Since ξ depends on x in part (b), let's write $\xi(x)$ instead of ξ. From the earlier parts, we know that

$$\int_a^b f(x)\,dx = \int_a^b P(x)\,dx + \int_a^b f''(\xi(x))\frac{(x-a)(x-b)}{2}\,dx$$

(the second integral exists because f and P are both integrable on $[a, b]$). Show that

$$\int_a^b P(x)\,dx = \frac{b-a}{2}(f(b) + f(a))$$

which equals $T_1(f; [a, b])$.

(e) By using the fact that $|f''|$ is bounded by K, prove that

$$\left| \int_a^b f''(\xi(x))\frac{(x-a)(x-b)}{2}\,dx \right| \le K \int_a^b \frac{(x-a)(b-x)}{2}\,dx$$

Compute this integral on the right and show that it equals

$$\frac{K(b-a)^3}{12}$$

(Suggestion: To make the work look better, substitute $u = (x-a)$, so that $b - a$ occurs in the integrand. You could also use $u = (x - a)/(b - a)$.)

(f) It follows from part (e) that

$$\begin{aligned}|E_{T_1}(f;[a,b])| &= \left|\int_a^b f(x)\,dx - T_1(f;[a,b])\right| \\ &= \left|\int_a^b f''(\xi(x))\frac{(x-a)(x-b)}{2}\,dx\right| \le \frac{K(b-a)^3}{12}\end{aligned}$$

which proves Theorem 6.43 when $n = 1$.

Now, we let $n \in \mathbb{N}$ be given. Let $\{x_0, x_1, \ldots, x_n\}$ be the partition of $[a,b]$ used in computing T_n. Apply the above result about E_{T_1} to each subinterval, and prove that

$$|E_{T_n}| \le \frac{Kn}{12}\left(\frac{b-a}{n}\right)^3 = \frac{K(b-a)^3}{12n^2}$$

finishing the proof of Theorem 6.43. (Hint: Think of each subinterval as its own subproblem of approximating with T_1 where the width is $(b-a)/n$.)

Chapter 7

Log, e, and Related Results

At this point, we have studied many different kinds of real functions. Starting with the basic arithmetic operations, we developed polynomials and rational functions. We proved the existence of nth roots of positive numbers for any $n \in \mathbb{N}$, and we used that to develop rational exponents. We also worked with the trigonometric functions. These functions all go under the category of elementary functions because of their simplicity and their widespread use. (We have also analyzed some more esoteric examples, like $\chi_{\mathbb{Q}}$ and the ruler function, though these are not considered elementary.)

In this chapter, we introduce and develop important properties of two more types of elementary functions: logarithms and exponentials. You have probably seen these kinds of functions before. Suppose that b is a constant greater than 1. Informally, the exponential function with base b is b^x where b is a positive constant. Also, the logarithm to the base b of x, written $\log_b x$, is the value y such that $b^y = x$. In other words, the function $\log_b x$ is the inverse of the function b^x.

As a practical example, exponential functions play a role when accumulating interest in a bank account. Suppose that an account starts with P dollars (P standing for Principal), and every month the account earns 3% interest. Thus, after one month, $1.03P$ dollars are in the account, and after two months, $1.03(1.03P) = (1.03)^2P$ dollars are in the account. In general, for any $n \in \mathbb{N}$, 1.03^nP dollars are in the account after n months. If we want to consider interest as being slowly built up over time, rather than being computed all at once at the end of the month, then we may say that for any positive number of months x (so $x = 1.5$ would mean a month and a half, and so forth), the account has 1.03^xP dollars after x months. In particular, if we want to know how long it takes for our money to double, i.e. we want to know the value of x for which $1.03^x P = 2P$, then $1.03^x = 2$ and $x = \log_{1.03} 2$ by the definition of logarithm. (It turns out $\log_{1.03} 2 \approx 23.45$, so you'd have to wait roughly two years to double your money!)

Logarithms and exponentials are also widely used because of some key properties they possess. For exponentials, we have

$$b^{x+y} = b^x b^y \qquad (b^x)^y = b^{xy} \qquad (ab)^x = a^x b^x$$

whenever $x, y > 0$ and $a, b > 1$. Also, for logarithms, we have

$$\log_b(xy) = \log_b x + \log_b y \qquad \log_b(a^x) = x \log_b a$$

Essentially, logarithms can be used to turn multiplication problems into addition problems, which is helpful since addition is usually easier to compute. Before computers were popular, and books consisting of tables with values of logarithms were in widespread use, one fast and accurate way to multiply two large numbers x and y would be to look up $\log_{10} x$ and $\log_{10} y$ in a logarithm table, add the results to get $\log_{10} x + \log_{10} y = \log_{10}(xy)$, and then use a logarithm table to convert back to xy.

For another application of this identity, consider a slide rule, which is essentially a ruler-like stick where the number x is marked $\log_2 x$ inches away from the end of the stick. (Obviously, you can't mark EVERY real number on a slide rule, but many numbers can be marked from 1 up to 2^L, where L is the total length in inches of the rule.) To multiply x and y using a slide rule, you would look up x on the rule at distance $d_x = \log_2 x$, look up y at distance $d_y = \log_2 y$, and add the distances. After that, you look at distance $d_x + d_y = \log_2 x + \log_2 y = \log_2(xy)$, and the number xy is marked at that spot on the rule. In fact, slide rules come with two marked sticks which can slide over each other, making it easy to add distances on the rules.

Before we can use logarithms and exponentials satisfactorily, however, we should have a more formal definition. The main issue with our current attempt at a definition is that if we say $f(x) = b^x$, then up to this point we have only defined f when x is rational. Now f is a strictly increasing function on $\mathbb{Q}$ (as you can show with a little work), so a modification of Theorem 3.48 suggests that we define

$$b^x = \sup\{b^y \mid y \in \mathbb{Q}, y < x\}$$

when x is irrational. While it is possible to use this definition to prove all the usual properties of exponential functions, including that f is a continuous bijection (so that it has a well-defined continuous inverse $\log_b x$), this process is tedious and unenlightening.

Instead, we will see that using techniques from calculus, we can first obtain a very simple and elegant definition of logarithms. From this definition, we will define exponential functions as inverses of logarithms. This method will quickly yield proofs of the main properties of these functions, and we will also have easy computations of derivatives and antiderivatives of these functions. Furthermore, this method leads us naturally to an important constant which is traditionally called e, and we will some places where e plays a crucial role.

7.1 The Definition Of The Logarithm

In order to define logarithmic functions, our approach will be based on the following question: what are the main properties we want logarithms to have? After all, new functions are introduced and made popular because of important properties they have. In particular, let's focus on finding real functions f which satisfy the functional equation

$$f(xy) = f(x) + f(y)$$

whenever x, y, and xy are in $\text{dom}(f)$. As we've seen, this property is useful, for instance, for slide rules and for computing multiplications in terms of additions.

The first thing we notice is that if $0 \in \text{dom}(f)$, then by taking $y = 0$, we obtain $f(0) = f(x) + f(0)$, i.e. $f(x) = 0$ for every $x \in \text{dom}(f)$. Thus, the only function f satisfying our functional equation which has 0 in its domain is the constant zero function. Since this function is not particularly interesting, let's consider solutions for f for which $\text{dom}(f) = \mathbb{R} - \{0\}$.

The next simple number we can try is 1. When we plug in $x = y = 1$ into the functional equation, we get $f(1) = f(1) + f(1)$, so $f(1) = 0$. Similarly, when we plug in $x = y = -1$, we get $f(1) = f(-1) + f(-1)$, so $f(-1) = 0$ as well. Now that we know $f(-1) = 0$, we find that $f(-x) = f(-1) + f(x) = f(x)$ for any $x \in \mathbb{R} - \{0\}$. Thus, any solution to our functional equation must be an even function. Lastly, when $y = 1/x$, we have $f(1) = f(x) + f(1/x) = 0$, so $f(1/x) = -f(x)$.

At this point, there aren't any more really useful numbers we can plug in, so there's not much more we can say about arbitrary solutions to $f(xy) =$

$f(x) + f(y)$. However, if we restrict our attention to differentiable solutions for f, then we can say more. (Besides, in practical situations, we'd like to work with differentiable functions anyway.) Let's suppose that $f'(x)$ exists for all $x \neq 0$. Therefore, if we think of y as a fixed constant and differentiate both sides of our functional equation with respect to x, then we get

$$\frac{d}{dx} f(xy) = \frac{d}{dx}(f(x) + f(y)) \qquad \text{i.e.} \qquad yf'(xy) = f'(x)$$

by the Chain Rule. In particular, when we take $x = 1$, we get

$$f'(y) = \frac{f'(1)}{y} \text{ for all } y \neq 0$$

We now know f', so we can obtain f by integrating. More specifically, when $x > 0$, our calculations show that f' is continuous between 1 and x, so the FTC says

$$f(x) = f(x) - f(1) = \int_1^x f'(t)\,dt = f'(1) \int_1^x \frac{dt}{t}$$

(since $f(1) = 0$). This tactic does not work for computing $f(x)$ when $x < 0$ since f' is not continuous between x and 1 (in particular, neither f nor f' are defined at 0). However, when $x < 0$, since f is even, we know $-x > 0$ and

$$f(x) = f(-x) = f'(1) \int_1^{-x} \frac{dt}{t}$$

by the previous calculation. Combining these results into one formula, and relabeling $f'(1)$ as a constant C, we find

$$f(x) = C \int_1^{|x|} \frac{dt}{t}$$

for all $x \in \mathbb{R} - \{0\}$.

In summary, we have proven that IF the functional equation $f(xy) = f(x) + f(y)$ has any differentiable solutions on $\mathbb{R} - \{0\}$, THEN $f(x)$ must be a constant multiple of the integral of $1/t\,dt$ from 1 to $|x|$. (Don't forget this includes the constant zero function.) However, we have not yet shown that this integral actually IS a solution to the functional equation. We will do so shortly.

The Natural Logarithm

The previous work leads us to the following choice of definition:

Definition 7.1. For all $x \in (0, \infty)$, the *natural logarithm* of x is

$$\log x = \int_1^x \frac{dt}{t}$$

This is called the natural logarithm since it uses the "most natural" choice of $C = 1$ in the equation we found above. (We will explain other logarithms a little later.)

Remark. For technical reasons, we restrict log to the domain of $(0, \infty)$. This is because if we were to define log on $\mathbb{R} - \{0\}$ using the formula we found earlier, then $\log(-x) = \log x$ for all $x \neq 0$, so log would not be a bijection. We want log to be a bijection so we can find its inverse later.

From this definition, we immediately find $\log 1 = 0$ and $(\log x)' = 1/x$ for all $x > 0$ by the first part of the FTC. (Note that for all $n \in \mathbb{N}$ with $n \neq 1$, we already know that an antiderivative of $1/x^n\, dx$ is $1/(n-1)x^{n-1}$, but this formula is invalid when $n = 1$.) Thus, log is differentiable and hence continuous on $(0, \infty)$. Also, since this derivative is positive, log is strictly increasing on $(0, \infty)$. Furthermore, $(\log x)'' = -1/x^2$, which is negative, so log is always concave downward. (See the discussion before Exercise 4.10.26.)

Now, we need to verify that our important property actually holds for the log function:

Theorem 7.2. *For all $x, y > 0$, $\log(xy) = \log x + \log y$.*

Strategy. We want to show

$$\int_1^{xy} \frac{dt}{t} = \int_1^x \frac{dt}{t} + \int_1^y \frac{dt}{t}$$

By subtracting the integral from 1 to x from each side, and using the interval addition property, this is equivalent to showing

$$\int_x^{xy} \frac{dt}{t} = \int_1^y \frac{dt}{t}$$

Since the only difference between these integrals is a change of lower and upper limits, this suggests that we can use substitution to show these are equal.

Proof. Let $x, y > 0$ be given. As shown in the strategy, we need only prove

$$\int_x^{xy} \frac{dt}{t} = \int_1^y \frac{dt}{t}$$

By either substituting $u = t/x$ (remember that t is the variable of integration and x is a constant), or by using the Stretching Property of Theorem 5.29, we find

$$\int_x^{xy} \frac{dt}{t} = \int_1^y \frac{x}{xu}\,du = \int_1^y \frac{du}{u}$$

as desired. ∎

Since $(\log x)' = 1/x$, the graph of the natural logarithm rises very steeply near $x = 0$ and "flattens out" as $x \to \infty$. Also, log is strictly increasing and continuous. This raises the question: is log bounded? We can use the "log of a product" property from Theorem 7.2 to show that log is unbounded above and below, as follows.

Since log is strictly increasing and $\log 1 = 0$, we have $\log 2 > 0$. Now, $\log(2^2) = \log 2 + \log 2 = 2\log 2$, and $\log(2^3) = \log(2^2) + \log 2 = 3\log 2$ using our product property. This is easily generalized by induction to $\log(2^n) = n\log 2$ for all $n \in \mathbb{N}$. (In fact, the same argument easily shows $\log(x^n) = n\log x$ for all $x > 0$ and all $n \in \mathbb{N}$.) Since $\mathbb{N}$ has no upper bound, neither does the set of numbers of the form $n\log 2$ for $n \in \mathbb{N}$, so the range of the log function is unbounded above.

Next, for any $x > 0$, we can use the product property with x and $1/x$ and obtain $\log 1 = 0 = \log x + \log(1/x)$. Thus, $\log(x^{-1}) = -\log x$ for all $x > 0$. (In fact, it follows from this and our previous work that $\log(x^z) = z\log x$ for all $z \in \mathbb{Z}$.) This shows that the range of log is also unbounded below, and in fact, the set of numbers of the form $\log(1/2^n)$ for $n \in \mathbb{N}$ is unbounded below! We may conclude that

$$\lim_{x\to\infty} \log x = \infty \qquad \lim_{x\to 0^+} \log x = -\infty$$

(See the discussion surrounding Definition 3.30.)

As a result, for any $y \in \mathbb{R}$, because the range of log is unbounded above and below, there exist $a, b > 0$ with $\log a < y < \log b$. Because log is continuous, the Intermediate Value Theorem says there is some $x \in (a, b)$ with $\log x = y$. This means that log has $\mathbb{R}$ as its range. Since log is also strictly increasing, and hence injective, we summarize our results with the following theorem:

Theorem 7.3. log *is a strictly increasing continuous bijection from* $(0, \infty)$ *to* $\mathbb{R}$.

In particular, there is exactly one $x > 0$ with $\log x = 1$. This number is denoted by e. In the exercises, you can obtain some estimates for e; it is known that $e \approx 2.71828$. (In the next chapter, we will also be able to prove that e is irrational.)

Other Logarithm Functions

We developed the log function by focusing on functions with the product property from Theorem 7.2. From the work at the beginning of this section, we can show that on $(0, \infty)$, the only differentiable functions with this product property have the form

$$f(x) = C \log x$$

for some constant C. The choice of $C = 0$ is boring, as it produces the zero function. However, when $C \neq 0$, f is easily shown to be a strictly monotone continuous bijection from $(0, \infty)$ to $\mathbb{R}$. Thus, when $C \neq 0$, we will also call $C \log x$ a logarithm function (whereas log is the natural logarithm function).

There is a more useful way of describing different logarithm functions. If f is a logarithm function, then there is a unique $b \in (0, \infty)$ such that $f(b) = 1$ because f is a bijection with range $\mathbb{R}$. By the product property, it follows that $f(b^2) = f(b) + f(b) = 2$, $f(b^3) = f(b^2) + f(b) = 3$, and so on. For this reason, b is called the *base* of the logarithm, and we write $f(x) = \log_b x$. If we also write $f(x) = C \log x$, then $f(b) = 1 = C \log b$. Hence, we must have $\log b \neq 0$, so $b \neq 1$ and $C = 1/\log b$. We summarize with the following definition:

Definition 7.4. For any $b > 0$ with $b \neq 1$, we define the *logarithm to the base* b by

$$\log_b x = \frac{\log x}{\log b}$$

for any $x > 0$. (On the right side, the log function without a subscript means natural logarithm.)

Thus, the natural logarithm is the same as the logarithm to the base e.

Remark. In different disciplines, people have different standards for what log without a subscript means. In mathematics, log usually means $\log_e$, because the derivative of $\log_e x$ is just $1/x$. (You can see that $(\log_b x)' = 1/(x \log_e b)$.) However, in computer science, since binary numbers are very common, log often means $\log_2$, the *binary logarithm*. It is also common, especially on many calculators, to use log to represent $\log_{10}$, the *decimal logarithm*, because we write numbers in base 10.

For our purposes, we will write log to represent the natural logarithm. It is useful to note, however, that when people use different bases for log, the notation "$\ln x$" is commonly used to represent the natural logarithm. This notation is especially popular on calculators.

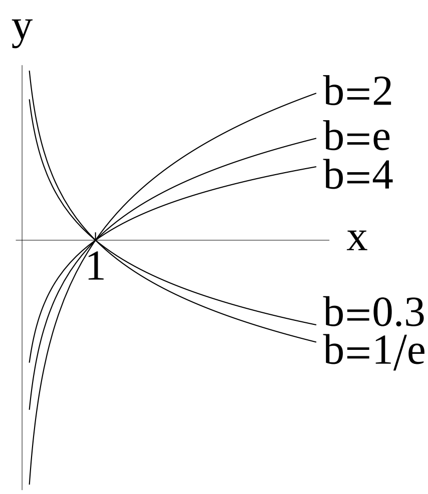

Figure 7.1: Graphs of $\log_b x$ for several values of b

The graphs of several different logarithm functions are shown in Figure 7.1. Several useful properties of logarithms are made more evident from the graphs. When $b > 1$, we have $\log b > 0$, so $\log_b x$ is strictly increasing and concave downward. When $0 < b < 1$, we have $\log b < 0$, so $\log_b x$ is strictly decreasing and concave upward. (In fact, because $\log(1/b) = -\log b$ when $b > 1$, the functions $\log_b x$ and $\log_{1/b} x$ are negations of each other.) All the logarithms have the same root of $x = 1$. When $1 < b_1 < b_2$, we have $0 < \log b_1 < \log b_2$ and hence $\log_{b_2} x < \log_{b_1} x$, so a larger base yields a smaller logarithm. Lastly, all of these functions are unbounded above and below.

7.2 Exercises

1. Prove that for all $x, y > 0$, $\log(x/y) = \log x - \log y$.

2. Prove that for all $x > 0$ and all $r \in \mathbb{Q}$, $\log(x^r) = r \log x$ by using the following steps:
   (a) Prove this result when $r \in \mathbb{N}$.
   (b) Prove this result when $r \in \mathbb{Z}$.
   (c) Prove this result when $r = 1/n$ for some $n \in \mathbb{N}^*$.
   (d) Finally, prove this result for all $r \in \mathbb{Q}$.

3. (a) Define
$$f(x) = \frac{\log x}{x}$$
   for all $x > 0$. Determine where f is increasing and where f is decreasing. (If you'd like, to aid part (b), you can also determine where f is concave up, and where f is concave down, but this is not required.)
   (b) Use the results from part (a) to sketch a graph of f.
   (c) From the results of part (a), prove that $\log x \le x/e$ for all $x > 0$.

4. For any $a, b > 0$ with $a, b \neq 1$, prove the *Change of Base formula*:
$$\log_a x = \frac{\log_b x}{\log_b a}$$
   Note that this also implies the useful result
$$\log_a b = \frac{1}{\log_b a}$$
   which allows us to switch the base and argument in a logarithm.

5. Let $f(x) = 1 - (1/x)$ and $g(x) = x - 1$ for all $x > 0$. Note that $f(1) = \log 1 = g(1) = 0$. Prove that for all $x > 0$ with $x \neq 1$,
$$f(x) < \log x < g(x)$$
   In particular, when $x = e$, we get the estimate $1 - (1/e) < 1 < e - 1$, so $e > 2$. (Hint: Compare the derivatives of f, log, and g, and use the Mean Value Theorem.)

6. For any $n \in \mathbb{N}^*$, define the n^{th} *harmonic number* by

$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$

By using integral approximations, prove that for all $n \in \mathbb{N}^*$ with $n > 1$,

$$H_n - 1 < \log n < H_{n-1}$$

In particular, this implies $1/2 < \log 2 < 1$ and $13/12 < \log 4 < 11/6$, so e must be between 2 and 4.

7. (a) Use the Trapezoid Rule with $n = 5$ to approximate

$$\log 3 = \int_1^3 \frac{dt}{t}$$

(b) Theorem 6.43 with $a = 1$, $b = 3$, and $n = 5$ implies that

$$|E_{T_5}| \le \frac{K(3-1)^3}{12(5)^2} = \frac{2K}{75}$$

where K is a bound on $|(1/x)''|$ for all $x \in [1, 3]$. Find one such bound, and use the error estimate with your value from (a) to prove that $\log 3 > 1$. Hence, $e < 3$.

(c) Use T_5 to approximate $\log 2.5$, find an error estimate as in part (b) using Theorem 6.43, and use it to prove that $e > 2.5$.

8. Define $f : \mathbb{R} - \{0\} \to \mathbb{R}$ by $f(x) = \log |x|$ for all $x \in \mathbb{R} - \{0\}$. Prove that for all $x \in \mathbb{R} - \{0\}$, $f'(x) = 1/x$. As a result, we have the following formula, valid on intervals which do not contain 0:

$$\int \frac{dx}{x} = \log |x| + C$$

9. Let $f : (0, \infty) \to \mathbb{R}$ be a continuous function, and define $F : (0, \infty) \to \mathbb{R}$ for all $x \in (0, \infty)$ by

$$F(x) = \int_1^x f(t)\, dt$$

Thus, $F'(x) = f(x)$ for all $x > 0$.

Now, assume that f has the following two properties: first, $f(2) = 2$, and second, for all $y > 0$, the integral

$$\int_x^{xy} f(t)\,dt$$

is the same for every value of $x \in (0, \infty)$. (In other words, this quantity is only a function of y.) More precisely, for all positive x_1, x_2, y, we have

$$\int_{x_1}^{x_1 y} f(t)\,dt = \int_{x_2}^{x_2 y} f(t)\,dt$$

Under these assumptions, find the value of $F(x)$ for each $x > 0$, and prove your answer is correct. (Hint: Start by writing a functional equation that F satisfies.)

10. Suppose that $f : (0, \infty) \to \mathbb{R}$ is continuous and satisfies

$$\int_1^{xy} f(t)\,dt = y \int_1^{x} f(t)\,dt + x \int_1^{y} f(t)\,dt$$

for all $x, y > 0$. Assume also that $f(1) = 3$. Let $F(x) = \displaystyle\int_1^x f(t)\,dt$ for all $x > 0$. Prove that for some $C > 0$ and all $x > 0$, $F(x) = Cx \log x$, and use this to find $f(x)$. (Hint: Consider what property $F(x)/x$ satisfies.)

7.3 The Exponential Function

At this point, we have introduced the natural logarithm function, log. This function satisfies the following useful properties for all $x, y > 0$ and all $q \in \mathbb{Q}$:

$$\log(xy) = \log x + \log y \qquad \log\left(\frac{x}{y}\right) = \log x - \log y \qquad \log(x^q) = q \log x$$

Also, log is a strictly increasing bijection from $(0, \infty)$ to $\mathbb{R}$, so it has a strictly increasing inverse from $\mathbb{R}$ to $(0, \infty)$. This function gets a special name:

Definition 7.5. The *(natural) exponential function* $\exp : \mathbb{R} \to (0, \infty)$ is the inverse of log. Therefore, for all $x \in \mathbb{R}$ and all $y > 0$, we have

$$y = \exp(x) \leftrightarrow \log y = x$$

In particular, we have $\exp(0) = 1$ and $\exp(1) = e$.

The function exp satisfies some properties which mirror the properties that log satisfies:

Theorem 7.6. *For all $x, y \in \mathbb{R}$ and all $q \in \mathbb{Q}$, we have*

$$\exp(x+y) = \exp(x)\exp(y) \qquad \exp(x-y) = \frac{\exp(x)}{\exp(y)} \qquad \exp(qx) = (\exp(x))^q$$

Strategy. Since exp is the inverse of log, we know that for all $x \in \mathbb{R}$ and $a > 0$, $a = \exp(x)$ iff $\log a = x$. Therefore, we can write x as $\log a$ and y as $\log b$ for some $a, b > 0$. After that, we can use the properties of log.

Proof. We only prove the first of our three properties; the others are left for Exercise 7.4.1. Let $x, y \in \mathbb{R}$ be given, and define $a = \exp(x)$ and $b = \exp(y)$. Therefore, $x = \log a$ and $y = \log b$. Thus,

$$x + y = \log a + \log b = \log(ab)$$

Applying exp to both sides, we get $\exp(x + y) = ab$ because exp and log are inverses, so $\exp(x + y) = \exp(x)\exp(y)$. ∎

It is also worth noting that since $\log x \to -\infty$ as $x \to 0^+$, by letting $y = \log x$, we have

$$\lim_{y\to-\infty} \exp(y) = \lim_{x\to 0^+} \exp(\log x) = \lim_{x\to 0^+} x = 0$$

by a variant of the Composition Limit Theorem (note that since exp is the inverse of a continuous function, it is continuous by a variant of Theorem 3.54). Thus, exp has the x-axis as a horizontal asymptote. Similarly, since $\log x \to \infty$ as $x \to \infty$, we find that

$$\lim_{y\to\infty} \exp(y) = \infty$$

These limits could have also been proven from the fact that exp is strictly increasing and has $(0, \infty)$ as its range.

Irrational Powers and Other Exponential Functions

One consequence of Theorem 7.6 is that whenever $x \in \mathbb{Q}$, we have

$$\exp(x) = \exp(x \cdot 1) = (\exp(1))^x = e^x$$

Based on this formula, we choose to define irrational powers of e by defining

$$e^x = \exp(x) \text{ for ALL } x \in \mathbb{R}$$

This is the most logical way to extend the notation e^x to irrational values of x, since exp is continuous.

Remark. We will still occasionally write $\exp(x)$ instead of e^x to denote the natural exponential function. For instance, writing exp is often preferred with complicated exponents: below, the left expression is easier to read.

$$\exp(x^2 + 5) \qquad e^{x^2+5}$$

Now, we can use our definition with base e to define arbitrary powers with ANY positive base. Suppose $b > 0$ is given. We know that $b = \exp(\log b)$, so for any $x \in \mathbb{Q}$, Theorem 7.6 tells us

$$\exp(x \log b) = (\exp(\log b))^x = b^x$$

This suggests that we define

$$b^x = \exp(x \log b) = e^{x \log b} \text{ for all } x \in \mathbb{R} \text{ and } b > 0$$

We say that this is the *exponential function with base b*.

The function taking $x \in \mathbb{R}$ to b^x is continuous at all values of $x \in \mathbb{R}$. For any $b > 0$, we have $b^0 = 1$, so every exponential function has a graph passing through $(0, 1)$ (much like every logarithm function has a graph passing through $(1, 0)$). When $b = 1$, we have $\log b = 0$, so $1^x = e^{x \cdot 0} = e^0 = 1$ for all $x \in \mathbb{R}$. Otherwise, we have $\log b > 0$ iff $b > 1$, so b^x is strictly increasing iff $b > 1$, and b^x is strictly decreasing iff $b < 1$. See Figure 7.2.

Because of the monotonicity of b^x, we say that b^x *grows exponentially* when $b > 1$, and we say b^x *decays exponentially* when $b < 1$. Whenever $b \neq 1$, the function b^x is also a bijection from $\mathbb{R}$ to $(0, \infty)$ with the x-axis as a horizontal asymptote. It is also useful to note that the exponential functions with $b > 1$ are all horizontal contractions / expansions of each other, and the exponential functions with $b < 1$ are reflections of the exponential functions with $b > 1$ over the y-axis because

$$b^x = \exp(x \log b) = \exp(-x \log(1/b)) = \left(\frac{1}{b}\right)^{-x}$$

As the figure shows, 2^x and $(1/2)^x$ are reflections of one another over the y-axis.

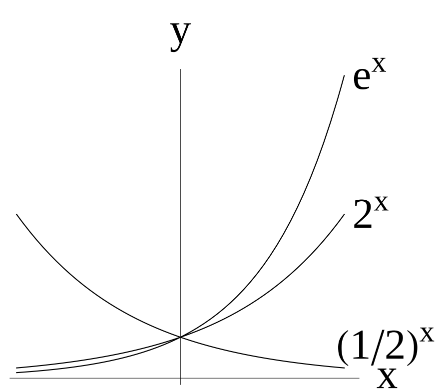

Figure 7.2: Graphs of some exponential functions

For every $b > 0$ with $b \neq 1$, the functions b^x and $\log_b x$ are inverses of each other. This is because for any $a > 0$ and any $x \in \mathbb{R}$,

$$a = b^x \leftrightarrow a = e^{x \log b} \leftrightarrow \log a = x \log b \leftrightarrow x = \frac{\log a}{\log b} = \log_b a$$

Thus, $\log_b(b^x) = x$ for every $x \in \mathbb{R}$, and $b^{\log_b x} = x$ for every $x > 0$. You can prove a few more useful properties of exponential functions and logarithms in the exercises.

The Derivative of exp

Next, we find the derivative of exp, since after all, we originally defined log in terms of its derivative. Recall that by Theorem 4.31, the inverse of a differentiable function f is differentiable and satisfies

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$$

for all $x \in \text{dom}(f^{-1})$ with $f'(f^{-1}(x)) \neq 0$. As exp has domain $\mathbb{R}$, we find that for any $x \in \mathbb{R}$,

$$\frac{d}{dx}e^x = \frac{1}{\frac{d\log}{dx}(e^x)} = \frac{1}{1/e^x} = e^x$$

(since $e^x \neq 0$ for all x). More generally, for any $b > 0$, the Chain Rule yields

$$\frac{d}{dx}b^x = \frac{d}{dx}e^{x\log b} = e^{x\log b}\log b = b^x \log b$$

for all $x \in \mathbb{R}$. This means that exponential functions are proportional to their own derivatives, which makes them very useful for modeling physical situations in which a quantity grows at a rate proportional to its size. (For instance, at the beginning of this chapter, we used $(1.03)^x$ to measure the amount of money in a bank account after x months. When accumulating interest, having more money means you accumulate money faster.)

Remark. Using the derivative of the exponential function, we can finally prove that $(x^r)' = rx^{r-1}$ for ALL $r \in \mathbb{R}$ and all $x > 0$. See Exercise 7.4.17.

We present some examples using this derivative:

Example 7.7:

Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined for all $x \in \mathbb{R}$ by

$$f(x) = e^{x^2} = \exp(x^2)$$

Using the Chain Rule, we find

$$f'(x) = \exp(x^2)\frac{d}{dx}(x^2) = 2xe^{x^2}$$

As a side note, it is known that f does not have an elementary antiderivative.
$\square$

Example 7.8:

Consider $f : (0, \infty) \to \mathbb{R}$ defined by

$$f(x) = x^x$$

for all $x > 0$. We would like to compute $f'(x)$ for all $x > 0$. Note that the expression x^x is neither an exponential function of the form b^x for some constant $b > 0$, nor is it a power function of the form x^r for some constant $r \in \mathbb{R}$, so our previous derivative formulas do not immediately apply.

As a result, we must return to the definition of x^x as $\exp(x \log x)$. Once we write our function this way, however, it is clear that the Chain Rule can be used to find the derivative:

$$f'(x) = \exp(x \log x) \frac{d}{dx}(x \log x) = x^x(\log x + 1)$$

This shows that when we have an irrational power, especially one where neither the base nor the exponent is constant, it is often very useful to write the power in terms of the exp function.[1] This is how we formally establish where many powers are differentiable. □

The fact that exp is its own derivative is not just coincidence. In fact, the exponential and the constant-zero functions are essentially the only types of functions which are proportional to their derivatives. More precisely, we have the following result:

Theorem 7.9. *Let $c \neq 0$ be given, and let $f : \mathbb{R} \to \mathbb{R}$ be a function which is differentiable everywhere, never equal to zero, and satisfies*

$$f'(x) = cf(x)$$

for all $x \in \mathbb{R}$. Then for some constant $A \in \mathbb{R} - \{0\}$, we have $f(x) = Ae^{cx}$ for all $x \in \mathbb{R}$.

Remark. The equation

$$f'(x) = cf(x)$$

involves f as an unknown function. Any equation involving an unknown function and some of its derivatives is called a *differential equation*. There are many tricks and theorems known for solving differential equations, but we will not focus on them in this book.[2] However, we will mention that the strategy we present in this theorem's proof can be adapted to solve any

[1] You can also use the technique of logarithmic differentiation, which is introduced in a later section of this chapter.

[2] There are some more remarks on differential equations near the end of Chapter 10, when we use power series to solve some equations.

differential equation in which all the occurrences of the unknown f end up on the same side and all the occurrences of x end up on the opposite side (these are called *separable* differential equations). For a quite different strategy, see Exercise 7.4.19.

Strategy. Suppose we write our differential equation with Leibniz notation as

$$\frac{df}{dx} = cf$$

Since $f \neq 0$, it is tempting to treat dx as a differential and solve to obtain

$$\frac{df}{f} = c\,dx$$

(essentially "separating" f and x on opposite sides of the equation). Still working informally, since there is a differential on each side, we write integral signs on each side to get

$$\int \frac{df}{f} = \int c\,dx \qquad \text{i.e.} \log|f| + C_1 = cx + C_2$$

where C_1 and C_2 are constants of integration. From here, we can solve for f.

Let's take this informal approach and make it more formal. We know

$$\frac{f'(x)}{f(x)} = c$$

since $f(x) \neq 0$. By the Chain Rule, we see that the left side is the derivative of $\log|f(x)|$. Hence, we integrate both sides in order to solve for $f(x)$.

Proof. Let c, f be given as described. As the strategy mentions, we suppose that $f(x) \neq 0$ for all $x \in \mathbb{R}$. It follows that since f is differentiable and hence continuous, f never changes sign. Also, we know

$$\frac{f'(x)}{f(x)} = c$$

for all $x \in \mathbb{R}$.

Since $\log|f(x)|$ has derivative $f'(x)/f(x)$ by the Chain Rule, we can integrate each side of our equation and obtain

$$\int \frac{f'(x)}{f(x)}\,dx = \int c\,dx \to \log|f(x)| + C_1 = cx + C_2$$

for some constants C_1 and C_2. Writing $D = C_2 - C_1$ for convenience, and using the fact that log and exp are inverses, we get

$$\log|f(x)| = cx + D \to |f(x)| = e^{cx+D} = e^{cx}e^{D}$$

Since f never changes sign, exactly one of the following is true for all $x \in \mathbb{R}$:

$$f(x) = e^{D}e^{cx} \qquad \text{or} \qquad f(x) = -e^{D}e^{cx}$$

Hence, our theorem is proven by choosing $A = e^D$ or $A = -e^D$ (note that $A \neq 0$). ■

Integrals With exp

Because $\exp'(x) = \exp(x)$ for all $x \in \mathbb{R}$, we immediately find that

$$\int e^x\,dx = e^x + C$$

This makes exponentials very useful when integrating by substitution or by parts. We give some examples.

Example 7.10:

Let's integrate

$$\int_0^1 x^2 e^{x^3}\,dx$$

Because of the composition $\exp(x^3)$ present in the integrand, we try substituting $u = x^3$ and $du = 3x^2\,dx$. This gives

$$\int_0^1 \frac{1}{3}e^u\,du = \left(\frac{e^u}{3}\right)\Bigg|_0^1 = \frac{e-1}{3}$$

□

Example 7.11:

Let's compute

$$\int xe^{2x}\,dx$$

Here, we try integration by parts with $u = x$ and $dv = e^{2x}\,dx$. This is because $du = dx$, so du is much simpler than u, and $v = e^{2x}/2$ (found by a mental substitution), so v is not more complicated than dv. (In general, exponentials are usually good choices for dv when integrating by parts because the integral of an exponential is not more complicated than the original function.) This gives

$$\int xe^{2x}\,dx = \frac{xe^{2x}}{2} - \frac{1}{2}\int e^{2x}\,dx = \frac{xe^{2x}}{2} - \frac{e^{2x}}{4} + C$$

□

Example 7.12:

Let's compute

$$\int e^x \sin x\,dx$$

Here, we try integration by parts. Either choice of e^x or $\sin x$ works fine for u, since when the other is picked as dv, the antiderivative v is not more complicated than dv. We will try $u = \sin x$ and $dv = e^x\,dx$, so that $du = \cos x\,dx$ and $v = e^x$. Thus,

$$\int e^x \sin x\,dx = e^x \cos x - \int e^x \cos x\,dx$$

This does not look any simpler than our original problem, but we have seen before that sometimes performing integration by parts multiple times in these situations can help us solve for our final integral as an unknown. Hence, we'll try integration by parts again with $u = \cos x$ and $dv = e^x\,dx$. (Note: If I were to pick $u = e^x$ and $dv = \cos x\,dx$ this time, I would merely undo the previous use of integration by parts and arrive where I started. Hence, I need to keep e^x as the dv part.) This gives $du = -\sin x\,dx$ and $v = e^x$, so

$$\begin{aligned}\int e^x \sin x\,dx &= e^x \cos x - \int e^x \cos x\,dx \\ &= e^x \cos x - \left(-e^x \sin x + \int e^x \sin x\,dx\right)\end{aligned}$$

We have obtained the integral of $e^x \sin x$ on both sides of our equation, so by some algebra, we get

$$\int e^x \sin x = \frac{e^x}{2}(\cos x + \sin x) + C$$

$\square$

7.4 Exercises

1. Finish the proof of Theorem 7.6 (we only proved the first of the three properties). You may use the first property if you wish.

For Exercises 2 through 8, prove the following properties of exponential functions for all $a, b > 0$ and all $x, y \in \mathbb{R}$. When solving these exercises, you may use the results of earlier exercises.

2. $a^{x+y} = a^x a^y$.
3. $a^{x-y} = \dfrac{a^x}{a^y}$ (In particular, $a^{-x} = 1/a^x$).
4. $(ab)^x = a^x b^x$.
5. $\left(\dfrac{a}{b}\right)^x = \dfrac{a^x}{b^x}$.
6. $\log_b(a^x) = x \log_b a$ when $b \neq 1$.
7. $a^{xy} = (a^x)^y$.
8. If $a < b$, then $a^x < b^x$ when $x > 0$ and $a^x > b^x$ when $x < 0$.

For Exercises 9 through 16, an expression defining $f(x)$ is given. Determine all x for which the function is differentiable, and compute $f'(x)$ for those values of x.

9. $f(x) = e^{1/x}$
10. $f(x) = 2^{x^2}$ (i.e. the exponent of 2 is x^2)
11. $f(x) = e^{e^x}$
12. $f(x) = x^{\log x}$
13. $f(x) = (\log x)^x$
14. $f(x) = (\sin x)^{\cos x}$
15. $f(x) = x^{x^x}$ (i.e. the exponent of x is x^x)
16. $f(x) = x^{1/(\log x)}$

17. The formula $(x^r)' = rx^{r-1}$ was proven earlier for $r \in \mathbb{Q} - \{0\}$. Prove it for all $r \in \mathbb{R} - \{0\}$ and all $x > 0$.

18. Let $f : \mathbb{R} \to \mathbb{R}$ be a function which is differentiable everywhere and satisfies
$$f(x + y) = f(x)f(y)$$
for all $x, y \in \mathbb{R}$. We will prove that f must be an exponential function, or constantly one or constantly zero, by performing the following steps (these steps are similar to how we arrived at our definition of log):
    (a) Prove that if $f(t) = 0$ for any $t \in \mathbb{R}$, then $f(x) = 0$ for all $x \in \mathbb{R}$.
    (b) Part (a) shows that the constant-zero function is the only possibility for f with any zeroes, so for the remainder of the problem, suppose that $f(x) \neq 0$ for all $x \in \mathbb{R}$. Now prove that $f(0) = 1$.
    (c) Prove that $\mathrm{ran}(f) \subseteq \mathbb{R}^+$. (Hint: f is continuous.)
    (d) Prove that there is some $c \in \mathbb{R}$ such that for all $x \in \mathbb{R}$, $f'(x) = cf(x)$.
    (e) Use part (d) to prove that if f is not the constant-one function, then $c \neq 0$. It now follows by Theorem 7.9 that for some $A \neq 0$, we have $f(x) = Ae^{cx}$ for all $x \in \mathbb{R}$. Part (b) implies that $A = 1$, finishing the proof.

19. Give a different proof of Theorem 7.9 by defining $g(x) = f(x)/e^{cx} = f(x)e^{-cx}$ and proving that g is a constant function.

For Exercises 20 through 23, compute the integral in the problem.

20. $\displaystyle\int_0^{\log 2} \frac{e^x}{e^x + 1}\, dx$

21. $\displaystyle\int e^{\cos x} \sin x\, dx$

22. $\displaystyle\int e^{\sqrt{x}}\, dx$ (Hint: Use a substitution followed by parts.)

23. $\displaystyle\int_0^1 x^3 e^{x^2}\, dx$ (Hint: This is similar to the previous problem.)

For Exercises 24 and 25, suppose a and b are nonzero constants, and compute the integral.

24. $\displaystyle\int e^{ax} \sin(bx)\, dx$

25. $\displaystyle\int e^{ax} \cos(bx)\, dx$

26. Prove the following reduction formula for all $a, n \in \mathbb{R} - \{0\}$:

$$\int x^n e^{ax}\, dx = \frac{x^n e^{ax}}{a} - \frac{n}{a} \int x^{n-1} e^{ax}\, dx$$

27. (a) For all $t > 0$, since exp is strictly increasing, we know that

$$e^t > 1 > e^{-t}$$

Integrate to prove that for all $x > 0$,

$$e^x > 1 + x \qquad \text{and} \qquad e^{-x} > 1 - x$$

(b) Using the idea and result of part (a), prove

$$e^x > 1 + x + \frac{x^2}{2} \qquad \text{and} \qquad e^{-x} < 1 - x + \frac{x^2}{2}$$

for all $x > 0$.

(c) More generally, prove that for all $n \in \mathbb{N}^*$ and all $x > 0$, we have

$$e^x > \sum_{i=0}^{n} \frac{x^i}{i!}$$

and

$$e^{-x} > \sum_{i=0}^{n} \frac{(-1)^i x^i}{i!} \text{ if } n \text{ is odd} \qquad e^{-x} < \sum_{i=0}^{n} \frac{(-1)^i x^i}{i!} \text{ if } n \text{ is even}$$

7.5 Derivatives and Integrals with Logs

Logarithms are useful in many calculus problems because of their product property and the simple form of their derivative. In this section, we will demonstrate some examples of working with logarithms in calculus problems. We start with a few examples of differentiating with logarithms.

Example 7.13:

Consider the expression

$$\log(\log x)$$

defined when $\log x > 0$, i.e. when $x > 1$. Using the Chain Rule, for any $x > 1$, we compute

$$\frac{d}{dx}(\log(\log x)) = \frac{1}{\log x}\left(\frac{1}{x}\right) = \frac{1}{x \log x}$$

□

Example 7.14:

Consider the expression

$$\log \sqrt{4 - x^2}$$

defined when $x^2 < 4$, i.e. when $x \in (-2, 2)$. One way to compute this derivative is with the Chain Rule twice, yielding

$$\frac{d}{dx} \log \sqrt{4 - x^2} = \frac{1}{\sqrt{4 - x^2}}\left(\frac{1}{2}(4 - x^2)^{-1/2}\right)(-2x) = \frac{-x}{4 - x^2}$$

However, an easier way is to use Exercise 7.2.2 (the so-called "log of a power" property) to note that

$$\log \sqrt{4 - x^2} = \log\left((4 - x^2)^{1/2}\right) = \frac{1}{2}\log(4 - x^2)$$

and therefore we compute

$$\frac{d}{dx}\left(\frac{1}{2}\log(4 - x^2)\right) = \frac{1}{2}\left(\frac{1}{4 - x^2}\right)(-2x) = \frac{-x}{4 - x^2}$$

as before. □

Logarithmic Differentiation

As Example 7.14 indicates, using properties of logarithms can help simplify calculations of derivatives. In fact, for some functions f, the product property of log makes it so that $\log f(x)$ is easier to differentiate than $f(x)$! We demonstrate with the following example.

Example 7.15:

Suppose that $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = x^2(x-1)^3 \sin x$$

We'd like to compute f'. Clearly the Product Rule proves f' exists, but it is messy to use the Product Rule to compute f' since we have to use that rule twice (because there are three simple functions being multiplied). However, because of the product property of logarithms, as well as the power property from Exercise 7.2.2, we note that $\log |f(x)|$ is fairly simple whenever $f(x) \neq 0$:

$$\begin{aligned} \log |f(x)| &= \log |x|^2 + \log |x-1|^3 + \log |\sin x| \\ &= 2\log |x| + 3\log |x-1| + \log |\sin x| \end{aligned}$$

(We take absolute values with f so that the logarithm will be defined at all points where $f(x)$ is not 0, as opposed to only the points where $f(x)$ is positive.)

Suppose we write $g(x) = \log |f(x)|$ when $f(x) \neq 0$. Since f is known to be differentiable, by the Chain Rule and the result of Exercise 7.2.8, we get

$$g'(x) = \frac{2}{x} + \frac{3}{x-1} + \frac{1}{\sin x}(\cos x) = \frac{2}{x} + \frac{3}{x-1} + \frac{\cos x}{\sin x}$$

However, the Chain Rule also tells us that

$$g'(x) = \frac{d}{dx}\left(\log |f(x)|\right) = \frac{1}{f(x)} f'(x)$$

Therefore, by setting our expressions for g' equal and multiplying by $f(x)$, we obtain

$$\begin{aligned} f'(x) &= f(x)\left(\frac{2}{x} + \frac{3}{x-1} + \frac{\cos x}{\sin x}\right) \\ &= x^2(x-1)^3 \sin x \left(\frac{2}{x} + \frac{3}{x-1} + \frac{\cos x}{\sin x}\right) \\ &= 2x(x-1)^3 \sin x + 3x^2(x-1)^2 \sin x + x^2(x-1)^3 \cos x \end{aligned}$$

Although we derived this formula when $f(x) \neq 0$, you can check with the Product Rule that this formula is also correct when $f(x) = 0$. $\square$

The approach used in Example 7.15 is called *logarithmic differentiation*. It finds the derivative f' by computing $g(x) = \log |f(x)|$, finding g', noticing that $g' = f'/f$ by the Chain Rule, and thus concluding $f' = f \cdot g'$. This approach is useful whenever g is simpler than f because of properties of log. In particular, logarithmic differentiation tends to work well with products, quotients, and powers (using the results of Exercises 7.2.1 and 7.2.2).

Remark. There are a couple technical points that should be noted. First, you need to already know f is differentiable in order to use the Chain Rule with $\log|f(x)|$. However, in many cases, it is easy to show that f is differentiable by using the rules we have proven before. Logarithmic differentiation does not prove that f' exists: it provides a convenient way to compute f'.

Second, since $g(x) = \log|f(x)|$ is only defined when $f(x) \neq 0$, logarithmic differentiation only produces the derivative $f'(x) = f(x)g'(x)$ at points where $f(x) \neq 0$. However, even if $a \in \mathbb{R}$ satisfies $f(a) = 0$, we can still compute

$$\lim_{x \to a} f'(x) = \lim_{x \to a} f(x)g'(x)$$

provided that $f(x) \neq 0$ when x is close enough to a but not equal to a. If this limit exists, then by Exercise 4.10.19, $f'(a)$ is that limit. In many situations, when you multiply f and g', cancelation occurs, and after the cancelation the resulting expression is frequently continuous at a anyway! In those situations, you may safely plug a into the expression to get $f'(a)$, even when $f(a) = 0$.

Example 7.16:

Suppose f is defined by

$$f(x) = \frac{x\tan^2 x}{1 + x^2}$$

at all $x \in \mathbb{R}$ where $\tan x$ is defined. For any such x, we would like to compute $f'(x)$. Because f is built out of products, quotients, and powers of simpler functions, logarithmic differentiation is a good approach to use.

Using our logarithm properties for products, powers, and quotients, we obtain

$$g(x) = \log|f(x)| = \log|x| + 2\log|\tan x| - \log|1 + x^2|$$

(Note that we do not have a property that allows us to simplify $\log|1 + x^2|$.) Therefore,

$$g'(x) = \frac{1}{x} + \frac{2}{\tan x}(\sec^2 x) - \frac{1}{1 + x^2}(2x)$$

and thus

$$\begin{aligned}
f'(x) &= f(x)g'(x) \\
&= \left(\frac{x\tan^2 x}{1+x^2}\right)\left(\frac{1}{x} + \frac{2\sec^2 x}{\tan x} - \frac{2x}{1+x^2}\right) \\
&= \frac{\tan^2 x}{1+x^2} + \frac{2x\sec^2 x\tan x}{1+x^2} - \frac{2x^2\tan^2 x}{(1+x^2)^2}
\end{aligned}$$

Since this expression is continuous at all x where $\tan x$ is defined, the earlier remark shows that this formula works for all x where $\tan x$ is defined. □

Logarithms in Substitutions

There are plenty of substitution problems which cause the final integrand to resemble $1/u$. In these problems, we can use log to obtain an antiderivative. In fact, the result of Exercise 7.2.8 shows that

$$\int \frac{1}{x}\,dx = \log|x| + C$$

whenever we integrate over an interval which does not contain 0. Here are some examples of this.

Example 7.17:

Let's compute

$$\int_0^1 \frac{x}{1+x^2}\,dx$$

When we use the substitution $u = 1 + x^2$, we get $du = 2x\,dx$, and the limits change to 1 and 2. This yields

$$\int_0^1 \frac{x}{1+x^2}\,dx = \frac{1}{2}\int_1^2 \frac{du}{u} = \frac{1}{2}\log|u|\Big|_1^2 = \frac{1}{2}(\log 2 - \log 1) = \frac{\log 2}{2}$$

□

Remark. When we use the FTC to integrate $1/x$, we need $1/x$ to be continuous on the interval of integration. Thus, when we obtain the antiderivative $\log|x|$, we can only plug in limits if our interval of integration doesn't have 0 in it. To see what can go wrong if this restriction is forgotten, consider the expression

$$\int_{-3}^{1} \frac{1}{x}\,dx$$

It is tempting to say that this equals $\log|1| - \log|-3| = -\log 3$. However, in fact, $1/x$ is unbounded on $[-3, 1]$, so it is not integrable on $[-3, 1]$! (This expression is an *improper integral*, since the integrand has a vertical asymptote. We will study improper integrals in Chapter 9.)

Example 7.18:

Let's compute

$$\int \frac{dx}{x \log x \log(\log x)}$$

In order for this denominator to be defined and nonzero, we need x and $\log x$ to be positive (so that $\log x$ and $\log(\log x)$ is defined) and also $x, \log x \neq 1$ (so $\log x, \log(\log x) \neq 0$). As $\log x > 0$ when $x > 1$, and $\log x = 1$ when $x = e$, this integral is well-defined over any closed bounded subinterval of $(1, \infty)$ which does not contain e.

The presence of the composition $\log(\log x)$ suggests trying $u = \log x$. As $du = dx/x$, this gives us

$$\int \frac{dx}{x \log x \log\log x} = \int \frac{du}{u \log u}$$

The presence of the $1/u$ in the integrand suggests that a substitution for $\log u$ will work. If we substitute $v = \log u$, we get $dv = du/u$ and thus

$$\int \frac{du}{u \log u} = \int \frac{dv}{v} = \log|v| + C = \log|\log u| + C = \log|\log\log x| + C$$

Alternately, if we chose the substitution $u = \log\log x$ in the first place, then we would immediately obtain the integral of $1/u$, which gives us $\log|u| + C = \log|\log\log x| + C$. $\square$

It's worth noting that we have actually used logarithms in substitutions before now, but we had to express the results in terms of a function

$$L(x) = \int_1^x \frac{dt}{t}$$

for which we did not have an alternate name. For instance, see Exercises 6.3.18 and 6.3.19, where we integrated $\tan x$ and $\sec x$ respectively. From those exercises, you can find

$$\int \tan x\,dx = \log|\sec x| + C \qquad \int \cot x\,dx = -\log|\csc x| + C$$

and

$$\int \sec x\,dx = \log|\sec x + \tan x| + C \qquad \int \csc x\,dx = -\log|\csc x + \cot x| + C$$

Logarithms with Parts

Because $\log x$ has derivative $1/x$, which is simpler than $\log x$, logarithms can also appear in problems with integration by parts. For instance:

Example 7.19:

Let's compute

$$\int \log x\,dx$$

The only tool we can really try here is integration by parts. Thus, one part needs to be chosen as u, which gets differentiated, and the rest is chosen as dv, which gets integrated. We try $u = \log x$ and $dv = dx$, so that $du = dx/x$ and $v = x$. Thus,

$$\int \log x\,dx = x\log x - \int dx = x\log x - x + C$$

In fact, by performing nearly the same steps, you can also show

$$\int \log|x|\,dx = x\log|x| - x + C$$

□

Frequently, when performing integration by parts with logarithms, the logarithm is chosen to be the differentiated part u. Doing this frequently makes the resulting integrand much simpler. (This is similar to why when we see an inverse trigonometric function, it is usually chosen as u.)

Example 7.20:

We compute the integral

$$\int_1^e x^2 \log x\, dx$$

Since a logarithm is present, and no substitution seems useful, we do parts with $u = \log x$ and $dv = x^2\, dx$. Thus, $du = dx/x$ and $v = (x^3)/3$, yielding

$$\int_1^e x^2 \log x\, dx = \left.\frac{x^3 \log x}{3}\right|_1^e - \frac{1}{3}\int_1^e x^2\, dx = \left.\left(\frac{x^3 \log x}{3} - \frac{x^3}{9}\right)\right|_1^e = \frac{2e^3+1}{9}$$

For contrast, suppose we switch the roles of our two parts, choosing $u = x^2$ and $dv = \log x\, dx$ (since we frequently pick polynomials as u in other parts problems). We get $du = 2x\, dx$ and $v = x \log x - x$ from the result of Example 7.19. Thus,

$$\begin{aligned}\int_1^e x^2 \log x\, dx &= x^2(x\log x - x)\big|_1^e - \int_1^e 2x(x \log x - x)\, dx \\ &= x^2(x\log x - x)\big|_1^e - 2\int_1^e x^2 \log x + \int_1^e 2x^2\, dx\end{aligned}$$

At this point, we can compute the integral of $2x^2$ easily, and we can collect the terms with our unknown integral on the same side. This gives us

$$3\int_1^e x^2 \log x\, dx = \left.\left(x^2(x\log x - x) + \frac{2x^3}{3}\right)\right|_1^e = \frac{2e^3}{3} - \left(-1 + \frac{2}{3}\right) = \frac{2e^3+1}{3}$$

so we get the same answer as in the other approach. However, this approach is much harder, and we already needed to know how to integrate $\log x$ for this approach to work. □

7.6 Partial Fractions

As a very useful application of logarithms in calculus, we will show how we can obtain antiderivatives for every rational function (recall that rational functions are fractions of polynomials). The technique we will use is called *integration by partial fractions*. This technique works by breaking up a fraction into simpler pieces, based on the factors of the denominator.

More precisely, suppose that p and q are polynomials, and we would like to compute

$$\int \frac{p(x)}{q(x)}\, dx$$

for intervals where q is not zero. We want to find polynomials $f_1, f_2, \ldots, f_m$ and $g_1, g_2, \ldots, g_m$ so that

$$\frac{p(x)}{q(x)} = \frac{f_1(x)}{g_1(x)} + \frac{f_2(x)}{g_2(x)} + \cdots + \frac{f_m(x)}{g_m(x)}$$

where, for each i from 1 to m, $g_i(x)$ is a factor of $q(x)$. In fact, we'll be able to make sure that each g_i is either a power of a linear function or a power of a quadratic function. We call the fraction f_i/g_i a *partial fraction*, and we call the sum of these fractions a *partial-fraction decomposition* for p/q. This decomposition is useful because each partial fraction is simpler to integrate.

Here is the main theorem guiding our work, which we will not prove: [3]

Theorem 7.21. *Suppose that p and q are polynomials with $\deg(p) < \deg(q)$. For some $m \in \mathbb{N}^*$, there exist polynomials $f_1, f_2, \ldots, f_m, g_1, g_2, \ldots, g_m$ such that*

$$\frac{p(x)}{q(x)} = \sum_{i=1}^{m} \frac{f_i(x)}{g_i(x)}$$

for every $x \in \mathbb{R}$ where $q(x) \neq 0$. Furthermore, for each i from 1 to m, $g_i(x)$ is a divisor of $q(x)$, and exactly one of these two statements holds:

1. *f_i and g_i have the form*

$$f_i(x) = A \qquad g_i(x) = (ax+b)^r$$

for all $x \in \mathbb{R}$, where $A, a, b \in \mathbb{R}$, $a \neq 0$, and $r \in \mathbb{N}^*$. We say g_i is a linear factor *if $r = 1$ or a* repeated linear factor *if $r > 1$.*

2. *f_i and g_i have the form*

$$f_i(x) = Ax + B \qquad g_i(x) = (ax^2 + bx + c)^r$$

for all $x \in \mathbb{R}$, where $A, B, a, b, c \in \mathbb{R}$, $a \neq 0$, $r \in \mathbb{N}^*$, and $b^2 - 4ac < 0$. (Thus, the equation $ax^2 + bx + c = 0$ has no real solutions.) We say g_i is an irreducible quadratic factor *if $r = 1$ or a* repeated irreducible quadratic factor *if $r > 1$.*

[3]For the interested reader, to prove this theorem, you need the Fundamental Theorem of Algebra and some results concerning the Euclidean Algorithm (for polynomials) for computing greatest common divisors. The Fundamental Theorem of Algebra is generally proven in graduate-level algebra courses, and while the results about the Euclidean Algorithm are not particularly difficult, they are outside the scope of this book.

Furthermore, if g_1 through g_m are all distinct, then there is a unique way to pick f_1 through f_m satisfying the above properties.

Example 7.22:

As an example, let's show how to compute

$$\int \frac{dx}{x^2 - 1}$$

by using partial fractions. Since $x^2 - 1$ factors as $(x-1)(x+1)$ (i.e. $g_1(x) = x - 1$ and $g_2(x) = x + 1$), Theorem 7.21 guarantees that there exist unique constants $f_1(x) = A$ and $f_2(x) = B$ satisfying

$$\frac{1}{x^2 - 1} = \frac{A}{x - 1} + \frac{B}{x + 1}$$

(The theorem doesn't rule out the case that A or B is zero.)

We would like to solve for A and B. First, we multiply $x^2 - 1 = (x - 1)(x + 1)$ to both sides of our equation, and we obtain

$$1 = A(x + 1) + B(x - 1)$$

(However, see Exercise 7.7.26 for a subtlety concerning this step.) At this point, we have two ways to proceed. One way is to collect similar terms together on each side, so we write our equation in the form

$$0x + 1 = (A + B)x + (A - B)$$

In order for these two polynomials to be equal for all $x \in \mathbb{R}$ (technically, they only need to be equal for two values of x, since two points determine a line), we need the coefficients to match. Thus, $A + B = 0$ and $A - B = 1$. When we add these equations, we get $2A = 1$, so $A = 1/2$, and when we subtract the equations, we get $2B = -1$, so $B = -1/2$.

The other way to find A and B is to plug in convenient values of x. If we plug in $x = -1$, then the equation $1 = A(x+1)+B(x-1)$ becomes $1 = -2B$, so $B = -1/2$. If we plug in $x = 1$, then we get $1 = 2A$, so $A = 1/2$.

Using either approach, we have successfully rewritten our original integral using two partial fractions. By applying some simple substitutions, we readily find that

$$\int \frac{dx}{x^2 - 1} = \frac{1}{2}\left(\int \frac{dx}{x - 1} - \int \frac{dx}{x + 1}\right) = \frac{1}{2}\left(\log|x - 1| - \log|x + 1|\right) + C$$

This can also be written as $(1/2)\log|(x-1)/(x+1)|+C$. When $(x-1)/(x+1)$ is positive, i.e. when $|x|>1$, we can use the power property from Exercise 7.2.2 and write this as

$$\log\left|\sqrt{\frac{x-1}{x+1}}\right|+C$$

□

Improper Rational Functions and Long Division

It is important to note that Theorem 7.21 applies only to rational functions for which the numerator has a smaller degree than the denominator, i.e. fractions of the form $p(x)/q(x)$ with $\deg(p(x)) < \deg(q(x))$. These kinds of rational functions are called *proper rational functions*. When $\deg(p(x)) \geq \deg(q(x))$, we instead say the function is an *improper rational function*.

How do we integrate an improper rational function $p(x)/q(x)$? The first thing we do is divide $p(x)$ by $q(x)$ to obtain a quotient and remainder. One convenient way of doing this is by using *long division* for polynomials. We will demonstrate the process of long division with an example; this will also illustrate the process better than a long step-by-step list of instructions.

Example 7.23:

As a demonstration, suppose we'd like to divide $p(x) = x^4 - x^3$ by $q(x) = x^2 + 2$. We set up our problem with a notation similar to long division of whole numbers:

$$x^2+0x+2\;\overline{\big)\,x^4-x^3+0x^2+0x+0}$$

(We have also added terms with coefficients of 0, so that $p(x)$ has a term for each power of x up to the leading term. This will make the work simpler.)

For the first step of long division, we divide the leading term of our dividend $p(x)$ with the leading term of our divisor $q(x)$. In this case, we divide x^4 by x^2 to obtain x^2. We think of this as saying "x^2 copies of $q(x)$ go into x^4". We write this quotient, x^2, above the x^4 term, and then we subtract $x^2q(x)$ from the dividend. This gives us

$$\begin{array}{rl} & x^2 \\ x^2+0x+2 & \overline{\big)\,x^4-x^3+0x^2+0x+0} \\ & \underline{-x^4-0x^3-2x^2} \\ & \quad\;\; -x^3-2x^2+0x+0 \end{array}$$

Now, we think of $-x^3 - 2x^2 + 0x + 0$ as being the new dividend, and we repeat the previous step. Thus, $-x^3$ divided by x^2 is $-x$, so we write $-x$ above the x^3 term and we subtract $-xq(x)$ from our dividend. (In other words, we add $xq(x)$.) This gives

$$
\begin{array}{r|l}
 & \;x^2 - x \\ \cline{2-2}
x^2 + 0x + 2 & \;x^4 - x^3 + 0x^2 + 0x + 0 \\
 & \underline{-x^4 - 0x^3 - 2x^2} \\
 & \qquad - x^3 - 2x^2 + 0x + 0 \\
 & \qquad\quad \underline{x^3 + 0x^2 + 2x} \\
 & \qquad\qquad\quad - 2x^2 + 2x + 0
\end{array}
$$

Doing this one more time, with our new dividend of $-2x^2 + 2x + 0$, we write -2 up above the x^2 term (this shows why it was useful to write $0x^2$ in $p(x)$, so that we have a convenient place to write this -2) and we subtract $-2q(x)$ from our dividend. This yields

$$
\begin{array}{r|l}
 & \;x^2 - x - 2 \\ \cline{2-2}
x^2 + 0x + 2 & \;x^4 - x^3 + 0x^2 + 0x + 0 \\
 & \underline{-x^4 - 0x^3 - 2x^2} \\
 & \qquad - x^3 - 2x^2 + 0x + 0 \\
 & \qquad\quad \underline{x^3 + 0x^2 + 2x} \\
 & \qquad\qquad\quad - 2x^2 + 2x + 0 \\
 & \qquad\qquad\qquad \underline{2x^2 + 0x + 4} \\
 & \qquad\qquad\qquad\qquad\quad 2x + 4
\end{array}
$$

At this point, we notice that the leading term of our dividend has smaller degree than the leading term of $q(x)$, so we stop. We conclude that $x^2 + 2$ goes into $x^4 - x^3$ a total of $x^2 - x - 2$ times (this is the polynomial we wrote at the top), with a remainder of $2x + 4$. You can check this result by seeing that

$$x^4 - x^3 = (x^2 + 2)(x^2 - x - 2) + (2x + 4)$$

and thus

$$\frac{x^4 - x^3}{x^2 + 2} = x^2 - x - 2 + \frac{2x + 4}{x^2 + 2}$$

□

Note that at the end of the previous example, we were able to write our improper rational function $p(x)/q(x)$ as a polynomial plus a proper rational function. This is the reason we do long division. We can integrate the polynomial easily, and then we break the proper rational function into partial fractions. Sometimes, however, the proper rational function is already in partial-fractions form, as this example shows:

Example 7.24:

Let's compute

$$\int \frac{2x^2 - 2x + 1}{2x - 1}\, dx$$

Because this is an improper rational function, we first perform long division. This gives us

$$\begin{array}{r|l} & x - \frac{1}{2} \\ \hline 2x - 1 & 2x^2 - 2x + 1 \\ & \underline{-2x^2 + x } \\ & - x + 1 \\ & \underline{ x - \frac{1}{2}} \\ & \frac{1}{2} \end{array}$$

(Note that in the second step, $-x$ divided by $2x$ yields $-1/2$; the quotients don't have to be integers!) Thus, our quotient is $x - (1/2)$ and our remainder is $1/2$. This means that

$$\int \frac{2x^2 - 2x + 1}{2x - 1}\, dx = \int x - \frac{1}{2} + \frac{1/2}{2x - 1}\, dx$$

We integrate the polynomial $x - (1/2)$ easily, and we perform a mental substitution of $u = 2x - 1$ for the proper rational function. Thus, our integral is

$$\frac{x^2}{2} - \frac{x}{2} + \frac{1}{4} \log |2x - 1| + C$$

□

The Different Types of Partial Fractions

We now turn our attention to integrating proper rational functions. Theorem 7.21 states that our partial fractions can come either from linear factors (possibly repeated) or irreducible quadratic factors (possibly repeated). When trying to decompose a proper rational function into partial fractions, we usually factor the denominator completely, write down all the forms that a partial fraction can take with some unknown coefficients, and then we solve for those unknowns. Let's take a look at a few examples concerning different types of partial-fraction decompositions.

Example 7.25:

Our first example will use only non-repeated linear factors. Let's integrate

$$\int \frac{6x^2 - 7x - 10}{x^3 - 3x^2 - 10x}$$

Note that this is proper, so we do not need to do long division. First, the denominator can be written as $x(x^2-3x-10) = x(x+2)(x-5)$. Therefore, each of our partial fractions comes from a linear factor, and we can write

$$\frac{6x^2 - 7x - 10}{x(x+2)(x-5)} = \frac{A}{x} + \frac{B}{x+2} + \frac{C}{x-5}$$

for some constants A, B, C. (Note that these constants might not be the same, so we need to use different names to represent each one.) To solve for these constants, we first multiply both sides by $x(x+2)(x-5)$ to get

$$6x^2 - 7x - 10 = A(x+2)(x-5) + Bx(x-5) + Cx(x+2)$$

As in Example 7.22, there are two methods we can use to solve for these coefficients. In the first method, we expand out the right side and collect powers of x together. This gives us

$$\begin{aligned} 6x^2 - 7x - 10 &= A(x^2 - 3x - 10) + B(x^2 - 5x) + C(x^2 + 2x) \\ &= (A + B + C)x^2 + (-3A - 5B + 2C)x + (-10A) \end{aligned}$$

Therefore, we set the coefficients equal and obtain

$$6 = A + B + C \qquad -7 = -3A - 5B + 2C \qquad -10 = -10A$$

The last equation immediately yields $A = 1$. Plugging that into the others gives

$$5 = B + C \qquad -4 = -5B + 2C$$

By doubling the first of these equations and subtracting the second, we get

$$14 = 7B$$

so $B = 2$. Plugging that in to $5 = B + C$, we find $C = 3$.

The other way to find A, B, C is to plug values of x into the equation

$$6x^2 - 7x - 10 = A(x+2)(x-5) + Bx(x-5) + Cx(x+2)$$

We plug in values which cause the linear factors to become zero. For instance, when we plug in $x = 0$, we get $-10 = -10A$, so $A = 1$. When we plug in $x = -2$, we get $28 = 14B$, so $B = 2$. When we plug in $x = 5$, we get $105 = 35C$, so $C = 3$. As you can see, this second method works very quickly for non-repeated linear factors (though as we'll soon see, it causes some difficulties for other types of factors).

We have now shown that

$$\begin{aligned} \int \frac{6x^2 - 7x - 10}{x(x+2)(x-5)}\,dx &= \int \frac{1}{x}\,dx + \int \frac{2}{x+2}\,dx + \int \frac{3}{x-5}\,dx \\ &= \log|x| + 2\log|x+2| + 3\log|x-5| + C \end{aligned}$$

(Note that the C here is a constant of integration, not the same C we used for partial-fraction decomposition. Sometimes, to avoid this confusion, people write A_1, A_2, etc. to denote the constants in the decomposition.) $\square$

Example 7.26:

In this example, we consider an integral where we have a repeated linear factor. Let's integrate

$$\int \frac{3x^3 - 4x^2 + 3x - 6}{x^4 - 2x^3}\,dx = \int \frac{3x^3 - 4x^2 + 3x - 6}{x^3(x-2)}\,dx$$

Here, the x^3 is a repeated factor. Its divisors include x, x^2, and x^3. Theorem 7.21 tells us to consider partial fractions of the form A/x^r, where r could be 1, 2, or 3. Therefore, we put in one partial fraction for each possibility of r, each with their own constant, and write

$$\frac{3x^3 - 4x^2 + 3x - 6}{x^4 - 2x^3} = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x^3} + \frac{D}{x-2}$$

(In general, when you have a repeated factor with a power of r, make one partial fraction for EACH power up to r, and give each fraction its own unknown constants.)

Once again, we multiply $x^4 - 2x^3 = x^3(x-2)$ to both sides and get

$$3x^3 - 4x^2 + 3x - 6 = Ax^2(x-2) + Bx(x-2) + C(x-2) + Dx^3$$

If we try the second method from the previous example, then plugging in $x = 0$ gives us $-6 = -2C$, so $C = 3$, and plugging in $x = 2$ gives us $8 = 8D$, so $D = 1$. After that, there is no value left to plug in which immediately tells us A or B in the same manner.

However, we can still plug in some other simple values of x to get equations involving the constants. Since we have two constants left to determine, we should plug in two values of x to get two equations. Suppose we plug in $x = 1$ and $x = -1$, giving us

$$-4 = -A - B - C + D \qquad \text{and} \qquad -16 = -3A + 3B - 3C - D$$

After plugging in our known values of $C = 3$ and $D = 1$ to this, this system becomes

$$-2 = -A - B \qquad \text{and} \qquad -6 = -3A + 3B$$

At this point, we can triple the equation $-2 = -A - B$ and add it to $-6 = -3A + 3B$ to get $-12 = -6A$, so $A = 2$. Using that in the equation $-2 = -A - B$, we get $B = 0$.

Thus, even when you have factors other than non-repeated linear factors, you can still use the "plug in values of x" method to obtain simple equations involving your constants. In this example, we saw that two values of x, 0 and 2, immediately gave us two of our constants; this is because we had two types of linear factors. However, the other two constants had to be found by plugging in two more values and getting two equations involving the remaining unknowns, so this method resembles the "expand and collect powers" method when you have factors with large values of r. (Note that the "expand and collect powers" method also works on this example, as you should check for yourself.)

We have now found that

$$\begin{aligned} \int \frac{3x^3 - 4x^2 + 3x - 6}{x^3(x-2)}\,dx &= \int \frac{2}{x}\,dx + \int \frac{3}{x^3}\,dx + \int \frac{1}{x-2}\,dx \\ &= 2\log|x| - \frac{3}{2x^2} + \log|x-2| + C \end{aligned}$$

□

Our remaining examples involve irreducible quadratic factors. Partial fractions involving irreducible quadratic factors are not as easy to integrate as those involving linear factors. Our approach for integrating

$$\int \frac{Ax+B}{(ax^2+bx+c)^r}\,dx$$

depends on whether or not b is zero.

If $b=0$, then we split the integral into two pieces as

$$\int \frac{Ax}{(ax^2+c)^r}\,dx + \int \frac{B}{(ax^2+c)^r}\,dx$$

The first piece is computed by making the substitution $u = ax^2+c$. The second piece is integrated by a trigonometric substitution involving tangent. (See Exercise 7.7.27, or see Exercise 7.7.28 for an alternate approach.)

On the other hand, if $b \neq 0$, then we complete the square for ax^2+bx+c and write it in the form $a(x+p)^2+q$ for some $p,q \in \mathbb{R}$. This yields

$$\int \frac{Ax+B}{(a(x+p)^2+q)^r}$$

We perform the substitution $u = x+p$, and then the resulting integral in u is solved by using the tactics we gave for the case where $b=0$.

Example 7.27:

In this problem, let's compute

$$\int \frac{3x^2+5x+4}{x^3+2x^2+2x}\,dx = \int \frac{3x^2+5x+4}{x(x^2+2x+2)}\,dx$$

(Note that x^2+2x+2 has discriminant $2^2-4(1)(2)<0$, so it is irreducible.) We express the partial-fraction decomposition as

$$\frac{3x^2+5x+4}{x(x^2+2x+2)} = \frac{A}{x} + \frac{Bx+C}{x^2+2x+2}$$

for some unknown constants A,B,C. Multiplying $x(x^2+2x+2)$ to both sides, we get

$$3x^2+5x+4 = A(x^2+2x+2)+(Bx+C)x$$

In this problem, it is easier to collect powers of x. We get $3x^2 + 5x + 4 = (A+B)x^2 + (2A+C)x + (2A)$, so we have

$$3 = A + B \qquad 5 = 2A + C \qquad 4 = 2A$$

The last equation yields $A = 2$. Plugging that into the other equations, we immediately find $B = 1$ and $C = 1$. Therefore, we have found

$$\begin{aligned} \int \frac{3x^2 + 5x + 4}{x(x^2 + 2x + 2)}\, dx &= \int \frac{2}{x}\, dx + \int \frac{x+1}{x^2 + 2x + 2}\, dx \\ &= 2\log|x| + \int \frac{x+1}{x^2 + 2x + 2}\, dx \end{aligned}$$

To handle the remaining partial fraction, we complete the square to get

$$x^2 + 2x + 2 = (x^2 + 2x + 1) + 1 = (x+1)^2 + 1$$

Thus, we substitute $u = x + 1$, so $du = dx$ and we get

$$\int \frac{(x+1)}{(x+1)^2 + 1}\, dx = \int \frac{u}{u^2 + 1}\, du$$

We solve this by substituting $v = u^2 + 1$, so $dv = 2u\, du$ and we get

$$\int \frac{1}{2v}\, dv = \frac{1}{2}\log|v| + C$$

Going back to the original variable of x and putting everything together, our final answer is

$$2\log|x| + \frac{1}{2}\log|(x+1)^2 + 1| + C = 2\log|x| + \frac{1}{2}\log|x^2 + 2x + 2| + C$$

$\square$

Example 7.28:

Let's integrate

$$\int \frac{x^4 + 2x^2 + x}{(x-1)(x^2+1)^2}\, dx$$

Here, $(x^2+1)^2$ is a repeated irreducible quadratic factor with $r=2$, so we will have a partial fraction with $(x^2+1)^1$ and one with $(x^2+1)^2$ in our decomposition. This gives

$$\frac{x^4+2x^2+x}{(x-1)(x^2+1)^2} = \frac{A}{x-1} + \frac{Bx+C}{x^2+1} + \frac{Dx+E}{(x^2+1)^2}$$

We multiply $(x-1)(x^2+1)^2$ to both sides and we get

$$x^4+2x^2+x = A(x^2+1)^2 + (Bx+C)(x-1)(x^2+1) + (Dx+E)(x-1)$$

For variety, let's try the method of plugging in values of x. When we plug in $x=1$, we get $4=4A$, so $A=1$. However, no other value of x immediately gives us any other constants. We try plugging in $x=0$, $x=-1$, $x=2$, and $x=-2$, and we get the four equations

$$\begin{aligned}
0 &= A+(C)(-1)(2)+(E)(-1) \\
&= A-2C-E \\
2 &= 4A+(-B+C)(-2)(2)+(-D+E)(-2) \\
&= 4A+4B-4C+2D-2E \\
26 &= 25A+(2B+C)(1)(5)+(2D+E)(1) \\
&= 25A+10B+5C+2D+E \\
22 &= 25A+(-2B+C)(-3)(5)+(-2D+E)(-3) \\
&= 25A+30B-15C+6D-3E
\end{aligned}$$

By plugging in our known value of $A=1$, we get the equations

$$\begin{aligned}
-1 &= -2C-E \\
-2 &= 4B-4C+2D-2E \\
1 &= 10B+5C+2D+E \\
-3 &= 30B-15C+6D-3E
\end{aligned}$$

First, we eliminate E from these equations by adding the first and third equations, adding twice the third equation to the second, and adding three times the third equation to the fourth. This yields

$$\begin{aligned}
0 &= 10B+3C+2D \\
0 &= 24B+6C+6D \\
0 &= 60B+12D
\end{aligned}$$

(At this point, you can see that these equations are satisfied with $B = C = D = 0$, but let's show the rest of the steps for completeness.) Next, we eliminate C from the equations by subtracting twice the first of these equations from the second. We get

$$\begin{aligned} 0 &= 4B + 2D \\ 0 &= 60B + 12D \end{aligned}$$

By subtracting 6 times the first of these equations from the second, we get $0 = 36B$, so $B = 0$. Putting that into $0 = 4B + 2D$, we get $D = 0$. Putting $B = 0$ and $D = 0$ into $0 = 10B + 3C + 2D$, we get $C = 0$. Lastly, putting all these values into $-1 = -2C - E$ yields $E = 1$.

Therefore, we have found

$$\int \frac{x^4 + 2x^2 + x}{(x-1)(x^2+1)^2} = \int \frac{1}{x-1}\,dx + \int \frac{1}{(x^2+1)^2}\,dx$$

The first integral is $\log|x-1| + C$. For the second integral, we substitute $x = \tan\theta$, so $dx = \sec^2\theta\,d\theta$ and thus

$$\begin{aligned} \int \frac{1}{x-1}\,dx + \int \frac{1}{(x^2+1)^2}\,dx &= \log|x-1| + \int \frac{\sec^2\theta}{(\tan^2\theta + 1)^2}\,d\theta \\ &= \log|x-1| + \int \frac{\sec^2\theta}{\sec^4\theta}\,d\theta \\ &= \log|x-1| + \int \cos^2\theta\,d\theta \\ &= \log|x-1| + \frac{1}{2}\int 1 + \cos(2\theta)\,d\theta \\ &= \log|x-1| + \frac{\theta}{2} + \frac{\sin(2\theta)}{4} + C \\ &= \log|x-1| + \frac{\theta}{2} + \frac{\sin\theta\cos\theta}{2} + C \end{aligned}$$

At this point, we need to return to the original variable of x. Since $x = \tan\theta$, we have $\theta = \arctan x$. Also, imagining a right triangle with an angle of θ with opposite side x and adjacent side 1 (so that $\tan\theta = x$), we find that the hypotenuse has length $\sqrt{x^2+1}$. Thus, $\sin x = x/\sqrt{x^2+1}$ and $\cos x = 1/\sqrt{x^2+1}$, so our final answer is

$$\log|x-1| + \frac{\arctan x}{2} + \frac{x}{2(x^2+1)} + C$$

□

Although integration by partial fractions tends to involve a lot of computation, most of the computation should be pretty mechanical, i.e. the steps don't require ingenuity. The main steps are to perform long division if necessary, identify the types of partial fractions needed (making sure to represent a repeated factor of power r with r different fractions, one for each power up to r), solve for some constants, and then integrate each partial fraction using the methods we have discussed. In fact, many computer programs can integrate by partial fractions automatically, as long as they can factor the denominator. Also, some integration problems can be handled by turning them into partial-fractions problems via substitution: for instance, see Exercises 7.7.29 and 7.7.30.

7.7 Exercises

For Exercises 1 through 8, an expression defining a function f is given in terms of x. Determine the set of $x \in \mathbb{R}$ for which this function is well-defined. Also, determine where the function is differentiable and compute its derivative, using logarithmic differentiation if desired.

1. $f(x) = \log(1 - x^2)$
2. $f(x) = \log\left(\frac{1-x}{1+x}\right)$
3. $f(x) = \log|\sec x + \tan x|$
4. $f(x) = \log(x^2 \log x)$
5. $f(x) = \sin(\log x) - \cos(\log x)$
6. $f(x) = -x\log x - (1-x)\log(1-x)$
7. $f(x) = \frac{x^{1/3}\log x}{(\sin x + 2)^2}$
8. $f(x) = x(x+1)^2(x-2)^3(x+3)^4$

For Problems 9 through 13, evaluate the integral.

9. $\int_1^2 \frac{3x^2 + 2x}{x^3 + x^2}\, dx$
10. $\int \frac{\log|x|}{x\sqrt{1 + \log|x|}}\, dx$
11. $\int_1^e x^n \log x\, dx$ where $n \in \mathbb{R} - \{0\}$
12. $\int \frac{\log x}{x}\, dx$
13. $\int \frac{\log x}{x^n}\, dx$ where $n \in \mathbb{R} - \{1\}$

14. Prove the following reduction formula for all $m, n \in \mathbb{R}$ with $n \neq 0$ and $m \neq -1$ (you may assume $\log x > 0$):

$$\int x^m \log^n x \, dx = \frac{x^{m+1} \log^n x}{m+1} - \frac{n}{m+1} \int x^m \log^{n-1} x \, dx$$

For Exercises 15 through 22, evaluate the integral using the method of partial fractions. Sometimes you will need to perform long division as part of the process.

15. $\displaystyle\int \frac{x^2}{x-1} \, dx$

16. $\displaystyle\int_1^3 \frac{x^3 + x^2 + 2x + 1}{x^2 + x} \, dx$

17. $\displaystyle\int \frac{ax+1}{x^2 - bx} \, dx$
for any $a, b \in \mathbb{R} - \{0\}$

18. $\displaystyle\int_2^4 \frac{3.5x^2 + 0.5x - 1}{x^2(x-1)^2} \, dx$

19. $\displaystyle\int \frac{4x^2 + 5x + 16}{x^3 + 4x} \, dx$

20. $\displaystyle\int \frac{3x^2 + x + 3}{(2x-1)(x^2 - 2x + 5)} \, dx$

21. $\displaystyle\int_1^{\sqrt{3}} \frac{3}{x^4 + 5x^2 + 4} \, dx$

22. $\displaystyle\int \frac{2x^4 + 35x^2 + 162}{x(x^2+9)^2} \, dx$

23. We computed the integral in Example 7.22 by using partial fractions. Compute that integral by using a trigonometric substitution instead, and show that you get the same answer when $|x| > 1$.

24. Compute $\displaystyle\int \frac{1}{1+e^x} \, dx$ in each of the following three ways (you should also check to see that you get the same answer each way, even if the answers don't look the same at first):
    (a) By multiplying the numerator and denominator by e^{-x}
    (b) By writing $\dfrac{1}{1+e^x}$ as $1 - \dfrac{e^x}{1+e^x}$
    (c) By substituting $u = e^x$ and performing integration by partial fractions

25. Rational functions in e^x (i.e. fractions whose numerators and denominators consist of linear combinations of integer powers of e^x) can be integrated by using the substitution $u = e^x$ and $du = e^x \, dx = u \, dx$

to turn the problem into an integration by partial fractions. Use this approach to compute the following integrals:

(a) $\displaystyle\int \frac{1}{1+e^{2x}}\,dx$

(b) $\displaystyle\int_{\log 2}^{\log 3} \frac{e^x+1}{e^x-1}\,dx$

(c) $\displaystyle\int \frac{e^{4x}+e^{3x}}{e^x+2}\,dx$

26. When trying to solve for a partial-fraction decomposition of the form

$$\frac{p(x)}{q(x)} = \sum_{i=1}^{n} \frac{f_i(x)}{g_i(x)}$$

as Theorem 7.21 guarantees, we frequently start by multiplying $q(x)$ to both sides to get rid of denominators. In particular, for each i from 1 to n, let's say h_i is the unique polynomial satisfying $g_i h_i = q$ (such an h_i exists because g_i is a divisor of q). We get

$$\frac{p(x)}{q(x)} = \sum_{i=1}^{n} \frac{f_i(x)}{g_i(x)} \leftrightarrow p(x) = \sum_{i=1}^{n} f_i(x)h_i(x)$$

This turns our rational equation above into a polynomial equation. However, the rational equation is only defined for values of $x \in \mathbb{R}$ for which $q(x) \neq 0$, and the polynomial equation is defined for all $x \in \mathbb{R}$. Thus, it is not clear whether the rational equation and the polynomial equation are logically equivalent at values of $x \in \mathbb{R}$ satisfying $q(x) = 0$.

Prove that if the rational equation is satisfied for all $x \in \mathbb{R}$ with $q(x) \neq 0$, then the corresponding polynomial equation is satisfied for all $x \in \mathbb{R}$. This allows us to plug in whatever value of x we wish into the polynomial equation when solving for the unknown coefficients. (Hint: You should use the fact that a polynomial of degree d is uniquely determined by its values at any $d+1$ points.)

27. Recall the reduction formula for powers of cosine from Exercise 6.3.22, valid for all $n \geq 2$:

$$\int \cos^n x\,dx = \frac{\cos^{n-1} x \sin x}{n} + \frac{n-1}{n}\int \cos^{n-2} x\,dx$$

Use this formula and a trigonometric substitution to compute

$$\int \frac{1}{(x^2+1)^3}\,dx$$

28. This exercise develops a reduction formula for integrating repeated irreducible quadratic factors.
    (a) By integrating by parts with $dv = dx$, prove that for all $a \in \mathbb{R}$ and all $n \neq -1/2$,

    $$\int (x^2+a^2)^n\,dx = \frac{x(x^2+a^2)^n}{2n+1} + \frac{2na^2}{2n+1}\int (x^2+a^2)^{n-1}$$

    (Hint: After one step, you will want to write x^2 as $(x^2+a^2)-a^2$.)
    (b) Use part (a) to conclude that for all $a \in \mathbb{R}$ and all $r \in \mathbb{R} - \{-3/2, 1\}$,

    $$\begin{aligned}\int \frac{1}{(x^2+a^2)^r}\,dx &= \frac{x(x^2+a^2)^{-(r-1)}}{2(r-1)a^2}\\ &+ \frac{2r-3}{2(r-1)a^2}\int \frac{1}{(x^2+a^2)^{r-1}}\,dx\end{aligned}$$

    (Hint: Note that the formula from part (a) reduces the power on (x^2+a^2). This should suggest which value to use for n.)
    (c) Use part (b) to compute

    $$\int \frac{1}{(x^2+1)^3}\,dx$$

29. By using integration by parts, followed by integration by partial fractions if necessary, evaluate

    $$\int \log|x^2+x|\,dx$$

    and

    $$\int \log\left|\frac{1+x}{1-x}\right|\,dx$$

30. In this exercise, we introduce a useful trick called the *Weierstrass substitution* (named after Karl Weierstrass) for turning fractions with $\sin x$ and $\cos x$ into rational functions.[4] The Weierstrass substitution is

$$t = \tan\left(\frac{x}{2}\right)$$

where $|x| < \pi$. Thus, $x = 2\arctan t$ and hence

$$dx = \frac{2}{1+t^2}\,dt$$

(a) By using double-angle identities, prove that

$$\sin x = \sin\left(2\left(\frac{x}{2}\right)\right) = \frac{2t}{1+t^2}$$

and

$$\cos x = \cos\left(2\left(\frac{x}{2}\right)\right) = \frac{1-t^2}{1+t^2}$$

(Hint: Note that $1 + t^2 = \sec^2(x/2)$.)

(b) Use the Weierstrass substitution and the results from part (a) to evaluate

$$\int \frac{\sin x}{1+\cos x}\,dx$$

(Don't forget to return to the original variable x in your final answer!)

(c) Use the Weierstrass substitution to evaluate

$$\int \frac{1}{1+\cos^2 x}\,dx$$

(Hint: $t^4 + 1 = (t^2 + t\sqrt{2} + 1)(t^2 - t\sqrt{2} + 1)$.)

[4] For the curious reader, the Weierstrass substitution historically arises from the problem of trying to describe all points with rational coordinates on the unit circle. It turns out that except for $(-1, 0)$, they all have the form

$$(\cos x, \sin x) = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)$$

for $|x| < \pi$ and $t \in \mathbb{Q}$.

7.8 How Quickly log and exp Grow

We have seen that logarithms and exponentials both go to ∞ as their argument goes to ∞ (when their bases are greater than 1). However, as Figures 7.1 and 7.2 seem to indicate, logarithms grow rather slowly and exponentials grow rather quickly. This raises the question: among exponentials, power functions, and logarithms, which of these functions go to ∞ the fastest as their input grows large, and which go to ∞ the slowest?

To answer these questions, we want to develop inequalities between exponentials, logarithms, and power functions. One such inequality is proven in Exercise 7.4.27:

Theorem 7.29. *For all $x > 0$ and all $n \in \mathbb{N}$, we have*

$$e^x > \frac{x^n}{n!}$$

The main idea behind the proof of this theorem, as is the case with many proofs of inequalities relating exponentials and polynomials, is to start with $e^x > 1$ when $x > 0$ and repeatedly integrate. (This approach is especially useful for exponentials since e^x is its own derivative.) We can use this inequality to prove

Theorem 7.30. *For all $a > 0$ and all $b > 1$, we have*

$$\lim_{x\to\infty} \frac{x^a}{b^x} = 0$$

This is informally described as saying that "exponentials dominate powers" (i.e. b^x approaches ∞ much faster than any x^a does).

Strategy. To show that the fraction x^a/b^x goes to 0, we find something larger which also approaches 0. Since $b^x = e^{x\log b}$, Theorem 7.29 roughly says that b^x is larger than a constant times x^n for any $n \in \mathbb{N}$. Thus, x^a/b^x is less than something like x^a/x^n. When n is larger than a, this bound approaches 0.

Proof. Let $a > 0$ and $b > 1$ be given. Choose some $n \in \mathbb{N}$ larger than a. Thus, for any $x > 0$, since $b^x = \exp(x \log b)$ and $\log b > 0$, Theorem 7.29 yields

$$0 < \frac{x^a}{b^x} < \frac{n!\, x^a}{(x\log b)^n} = \frac{n!}{(\log b)^n} x^{a-n}$$

Note that $n!/(\log b)^n$ is a constant with respect to x. Also, since $n > a$, $x^{a-n} \to 0$ as $x \to \infty$. Therefore, by the Squeeze Theorem, $x^a/b^x \to 0$ as $x \to \infty$ as well. ■

As a result of this theorem, any exponential of the form b^x with $b > 1$, no matter how close b is to 1, grows much faster than any power of x of the form x^a, no matter how large a is. Many limits involving exponentials can be manipulated to a form for which Theorem 7.30 apples, as we shall see with a couple of examples.

Example 7.31:

Let's analyze

$$\lim_{x\to\infty} \frac{x^2\, 2^x}{e^x}$$

Since $2^x/e^x = (2/e)^x = 1/(e/2)^x$, and $e > 2$, we can apply Theorem 7.30 with $a = 2$ and $b = e/2$ to conclude that the limit is 0. □

Our next example makes use of a very simple but handy tactic. Any limit of the form

$$\lim_{t\to 0^+} f(t)$$

is the same as

$$\lim_{x\to\infty} f(1/x)$$

by setting $x = 1/t$. This can often be useful for altering the form of a limit, as we see in this example:

Example 7.32:

Consider the following limit

$$\lim_{t\to 0^+} \frac{e^{-1/t}}{t}$$

This limit has a 0/0 form, because $-1/t \to -\infty$ as $t \to 0^+$, and exp approaches 0 as its argument approaches $-\infty$. Using our tactic described above, we let $x = 1/t$, so $x \to \infty$ as $t \to 0^+$. Our limit becomes

$$\lim_{x\to\infty} \frac{e^{-x}}{1/x}$$

by a variant of the Composition Limit Theorem. We simplify this to get

$$\lim_{x\to\infty} \frac{x}{e^x} = 0$$

$\square$

Now, we would like to develop some similar theorems concerning logarithms and powers, where we aim to show that logarithms grow much more slowly than powers. We have already shown in Exercises 7.2.3 and 7.2.5 that $\log x < x - 1$ and $\log x < x/e$ for all $x > 1$. We can also try to take the inequality from Theorem 7.29 and take logarithms of both sides, yielding

$$x > \log\left(\frac{x^n}{n!}\right) = n \log x - \log(n!)$$

so that

$$\log x < \frac{x}{n} + \frac{\log(n!)}{n}$$

However, this formula is not very useful for us, since it only compares $\log x$ to a linear function, whereas we'd like a comparison of $\log x$ with arbitrary powers of x.

We will prove an inequality comparing logarithms with arbitrary powers a little later, but even without such an inequality, we can still show that logarithms grow more slowly than powers:

Theorem 7.33. *For any $a > 0$ and $b > 1$, we have*

$$\lim_{x\to\infty} \frac{\log_b x}{x^a} = 0$$

We sometimes describe this by saying "powers dominate logarithms".

Strategy. The key idea is to make a change of variable to turn this limit into a limit involving exponentials. Namely, if $u = \log_b x$, then $x = b^u$, and $u \to \infty$ as $x \to \infty$. The new limit with respect to u can be found by Theorem 7.30.

Proof. Let $a > 0$ and $b > 1$ be given. If we let $u = \log_b x$, then $x = b^u$ and $u \to \infty$ as $x \to \infty$. Therefore, we have

$$\lim_{x\to\infty} \frac{\log_b x}{x^a} = \lim_{u\to\infty} \frac{u}{(b^u)^a}$$

Since $(b^u)^a = b^{au} = (b^a)^u$, and $b^a > 1$, our limit is 0 by Theorem 7.30. $\blacksquare$

Remark. It immediately follows from this theorem that powers of logarithms are also dominated by powers. This is because when $a, c > 0$ and $b > 1$, we have

$$\lim_{x\to\infty} \frac{(\log_b x)^c}{x^a} = \lim_{x\to\infty} \left(\frac{\log_b x}{x^{a/c}}\right)^c = 0^c = 0$$

because the function which takes c^{th} powers is continuous on $[0, \infty)$.

Because logarithms have many useful algebraic properties, allowing us to simplify many powers with them (such as x^x, for instance), Theorem 7.33 can be useful in a variety of limits. We give a couple examples.

Example 7.34:

We consider the limit

$$\lim_{x\to 0^+} x \log x$$

As $x \to 0^+$, we have $\log x \to -\infty$, so our limit has the form $0 \cdot (-\infty)$. Since the factor with limit 0 makes our product smaller in magnitude, and the factor with limit $-\infty$ makes our product larger in magnitude, it is unclear what the limit should be. (This is similar to why the limit of a $0/0$ form cannot be found without some rewriting; in fact, our limit can be written as a $0/0$ form by writing $x/(1/\log x)$.)

To handle this limit, we use a tactic from earlier: we let $u = 1/x$, so that $u \to \infty$ as $x \to 0^+$. Therefore, $x = 1/u$, and we get

$$\lim_{u\to\infty} \frac{1}{u} \log\left(\frac{1}{u}\right)$$

To make this simpler, a logarithm property lets us write this as

$$= \lim_{u\to\infty} \frac{-\log u}{u}$$

which is 0 by our theorem. (See Exercise 7.10.4 as well.) $\square$

Example 7.35:

Now, let's compute

$$\lim_{x\to 0^+} x^x$$

It is useful to rewrite x^x as $\exp(x \log x)$, like we did when computing the derivative of x^x. In this form, since exp is continuous, we see that we are really trying to compute

$$\exp\left(\lim_{x \to 0^+} x \log x\right)$$

By the previous example, this equals $e^0 = 1$. □

Although we now have proven a useful limit involving powers and logarithms in Theorem 7.33, it is still useful to have a convenient inequality relating powers and logarithms. The following inequality is usually adequate for many computations:

Theorem 7.36. *For all $a > 0$ and all $x > 1$, we have*

$$\log x < \frac{x^a}{a}$$

Strategy. There are a couple of useful ways to prove this theorem. One way involves showing that $\log x$ grows more slowly than x^a/a; you can develop such a proof yourself in Exercise 7.10.9. Another way returns to the definition of the logarithm as an integral:

$$\log x = \int_1^x \frac{1}{t}\,dt$$

By replacing $1/t$ by a larger quantity, we obtain a larger integral. In particular, we can replace $1/t$ by any power of t with exponent greater than -1. This is useful since it is easy to integrate every power of t except for the -1^{st} power.

Proof. Let $a > 0$ and $x > 1$ be given. For any $t \in (1, x)$, we have

$$\frac{1}{t} \leq \frac{t^a}{t} = t^{a-1}$$

because $t > 1$ and $a > 0$. Since $x \geq 1$, we may use the Comparison Property for integrals to state that

$$\log x = \int_1^x \frac{1}{t}\,dt \leq \int_1^x t^{a-1}\,dt = \left.\frac{t^a}{a}\right|_1^x = \frac{x^a - 1}{a} < \frac{x^a}{a}$$

as desired. ■

Remark. The inequality from Theorem 7.36 becomes less precise for small values of a, because when x is fixed and $a \to 0^+$, $x^a/a \to \infty$. However, when a is constant and x grows large, the theorem is very useful. For instance, see Exercise 7.10.8.

At this point, we have introduced useful inequalities and limits explaining how log and exp behave as their arguments go to ∞. (We have also seen a little bit about how $\log x$ behaves as $x \to 0^+$ and how e^x behaves as $x \to -\infty$.) In other words, we now have some useful information about the asymptotes of log and exp. However, there are many simple limits involving log and exp which do not involve their asymptotes, such as the following:

$$\lim_{x\to 1} \frac{\log x}{x-1} \qquad \lim_{x\to 0} \frac{e^x - 1}{x}$$

We have not yet developed any general tactics for analyzing these types of limits. (It turns out that each of these two limits is actually a difference quotient, so the limits are respectively the derivative of log at 1 and the derivative of exp at 0, but there are similar limits which cannot be computed this easily without better tools.) In the next section, we will introduce a powerful technique for handling limits in $0/0$ form, which can be used to evaluate the two limits above (as well as many other limits).

7.9 L'Hôpital's Rule and Indeterminate Forms

In this section, we present L'Hôpital's Rule[5], which relates the limit of a fraction f/g to the limit of derivatives f'/g'. However, before presenting this rule, it's worth asking: if $f(x)/g(x)$ is a $0/0$ form as $x \to a$, then why should f/g and f'/g' be related in the first place?

Let's suppose f and g are real functions defined near $a \in \mathbb{R}$, and let's also suppose for now that f and g are also defined at a with $f(a) = g(a) = 0$. Recall from Chapter 4 (especially Lemma 4.11) that f can be approximated by a linear function going through the point $(a, f(a))$ whose slope is the derivative at a. This means

$$f(x) \approx f(a) + f'(a)(x-a) = f'(a)(x-a)$$

[5] The rule is named after the Marquis de L'Hôpital, a French nobleman who wrote a textbook which popularized differential calculus. His book included this theorem, though most of the proof of the theorem was actually done by L'Hôpital's teacher, Johann Bernoulli.

for x close to a. Furthermore, the error E_f of this approximation, defined by $E_f(x) = f(x) - f'(a)(x-a)$, is "better than linear", meaning that $E_f(x)/(x-a)$ goes to 0 as $x \to a$, so $E_f(x)$ also goes to 0 as $x \to a$. Similar statements can also be made for g with an error function E_g.

From all this, we find that if $f(x)$ and $g(x)$ both approach 0 as $x \to a$, and if f and g are both differentiable at a, then we have

$$\frac{f(x)}{g(x)} \approx \frac{f'(a)(x-a)}{g'(a)(x-a)} = \frac{f'(a)}{g'(a)}$$

for x close to a. More precisely, when $x \to a$ and $g'(a) \neq 0$,

$$\frac{f(x)}{g(x)} = \frac{f'(a)(x-a) + E_f(x)}{g'(a)(x-a) + E_g(x)} \to \frac{f'(a)}{g'(a)}$$

Informally, this happens because if you zoom in on the graphs of f and g near a, then the graphs look a lot like straight lines with slopes $f'(a)$ and $g'(a)$ respectively.

However, this argument requires knowing that $f'(a)$ and $g'(a)$ exist. We can make a more general argument by looking at average slopes near a instead of instantaneous slopes. Let's say that for each $x \neq a$, we write $\Delta x = x - a$, $\Delta f = f(x) - f(a)$, and $\Delta g = g(x) - g(a)$. Since $f(a) = g(a) = 0$, we have

$$\frac{f(x)}{g(x)} = \frac{\Delta f}{\Delta g} = \frac{\Delta f/\Delta x}{\Delta g/\Delta x}$$

It is plausible that $\Delta f/\Delta x$ is close to $\dfrac{df}{dx}$, i.e. $f'(x)$, and similarly for g. This suggests that $f(x)/g(x) \approx f'(x)/g'(x)$, so if f'/g' approaches a limit as $x \to a$, then f/g should approach that same limit. This leads us to the statement of L'Hôpital's Rule (though L'Hôpital's Rule is even more general since it doesn't assume f or g is defined at a).

Theorem 7.37 (L'Hôpital's Rule for $0/0$). *Let $a \in \mathbb{R}$ be given, and let f and g be real functions defined near a. Suppose that $f(x), g(x) \to 0$ as $x \to a$. Also assume that f and g are differentiable near a, and suppose that for some $L \in \mathbb{R}$,*

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = L$$

(so, in particular, g' is nonzero near a). Then $f(x)/g(x)$ has a limit as $x \to a$, and

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} = L$$

Remark. L'Hôpital's Rule also works for one-sided limits, and it also works when a is ∞ or $-\infty$. (You can prove the $a = \pm\infty$ case in Exercise 7.10.30.) In fact, a small modification of the proof we present here will also prove L'Hôpital's Rule works even if L is ∞ or $-\infty$! (See Exercise 7.10.31.)

Strategy. To simplify our work, we will actually prove the theorem for the right-hand limit, i.e. as $x \to a^+$. This is because a right-hand limit will consider intervals $(a, a+\delta)$ with a as an endpoint (as opposed to sets of the form $(a-\delta, a+\delta) - \{a\}$ which have a missing from the middle). It is simple to modify our argument to work for left-hand limits as well. Note that f and g might not be defined at a, but since the limit of f/g doesn't depend on whether $f(x)/g(x)$ is defined at $x = a$, we may WLOG redefine $f(a)$ and $g(a)$ to be 0. This makes f and g continuous at a from the right.

Now, using the notation we introduced before L'Hôpital's Rule, we have

$$\frac{f(x)}{g(x)} = \frac{\Delta f}{\Delta g} = \frac{\Delta f/\Delta x}{\Delta g/\Delta x}$$

for any $x > a$ which is close to a. The numerator and the denominator of this fraction are average changes of f and g. To connect these average changes with instantaneous changes, the Mean Value Theorem is very helpful. In particular, if x is close enough to a, then f and g are differentiable on (a, x) and continuous on $[a, x]$, so therefore $\Delta f/\Delta x = f'(c)$ for some $c \in (a, x)$, and $\Delta g/\Delta x = g'(d)$ for some $d \in (a, x)$. Thus,

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(d)}$$

Unfortunately, although c and d both approach a as $x \to a$ (since c and d are between a and x), we cannot say that $f'(c)/g'(d)$ will approach L unless we know whether c and d can be chosen equal for every x close enough to a. To handle this, in Exercise 4.10.18, we introduced a stronger version of the MVT called Cauchy's Mean Value Theorem (Theorem 4.51). It tells us that some $c \in (a, x)$ can always be picked to satisfy

$$\frac{\Delta f}{\Delta g} = \frac{f'(c)}{g'(c)}$$

Thus, as $x \to a^+$, we will have $c \to a^+$, so that $f'(c)/g'(c)$ and $f(x)/g(x)$ will approach L.

(In the formal argument, we will use the definition of limit instead of saying the casual-sounding phrase "c approaches a as x approaches a". Although the proof can be completed using this phrase with the Composition Limit Theorem, treating c as a function of x, an argument in the δ-ϵ style feels more rigorous.)

Proof. Let a, f, g, L be given as described. We will prove the right-hand version of L'Hôpital's Rule here, as previously mentioned. Let $\epsilon > 0$ be given; we wish to find $\delta > 0$ so that

$$\forall x \in (a, a+\delta)\ \left|\frac{f(x)}{g(x)} - L\right| < \epsilon$$

We may redefine $f(a) = g(a) = 0$ without affecting whether our desired conclusion is true; this choice of $f(a)$ and $g(a)$ means that $f(x) \to f(a)$ and $g(x) \to g(a)$ as $x \to a^+$, i.e. f and g are continuous from the right at a.

Since $f'(x)/g'(x) \to L$ as $x \to a^+$, there is some $\delta > 0$ such that

$$\forall t \in (a, a+\delta)\ \left|\frac{f'(t)}{g'(t)} - L\right| < \epsilon$$

In particular, this means that $f'(t)/g'(t)$ is defined for all $t \in (a, a+\delta)$, so f' and g' exist with $g' \neq 0$ on $(a, a+\delta)$. Let $x \in (a, a+\delta)$ be given. Thus, f and g are differentiable on (a, x) and continuous on $[a, x]$.

We should check that $g(x) \neq 0$, so that $f(x)/g(x)$ is defined. If it were 0, then since $g(x) = g(a) = 0$, Rolle's Theorem (or the MVT) would imply that for some $t \in (a, x)$, $g'(t) = 0$. Since no such t exists, we know $g(x) \neq 0$.

Now, since f and g are continuous on $[a, x]$ and differentiable on (a, x), Cauchy's MVT (Theorem 4.51) implies that there exists some $t \in (a, x)$ such that

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(t)}{g'(t)}$$

Since $x \in (a, a+\delta)$, we have $t \in (a, a+\delta)$ as well. It follows that since $f(a) = g(a) = 0$,

$$\left|\frac{f(x)}{g(x)} - L\right| = \left|\frac{f(x) - f(a)}{g(x) - g(a)} - L\right| = \left|\frac{f'(t)}{g'(t)} - L\right| < \epsilon$$

as desired. ■

Example 7.38:

L'Hôpital's Rule can be used to find the limits mentioned at the end of the previous section. For the first limit, if $f(x) = \log x$ and $g(x) = x - 1$ for all $x > 0$, then $f'(x) = 1/x$ and $g'(x) = 1$, so that

$$\lim_{x\to 1} \frac{\log x}{x-1} = \lim_{x\to 1} \frac{1/x}{1} = 1$$

For the second limit, we can use L'Hôpital's Rule to conclude

$$\lim_{x\to 0} \frac{e^x - 1}{x} = \lim_{x\to 0} \frac{e^x}{1} = e^0 = 1$$

Note that we write these calculations as if the work flows from left to right. However, you need to know if the limit of f'/g' exists before you can use L'Hôpital's Rule, so the very first = sign in these calculations is actually the LAST step. Sometimes, to keep this in mind and make our work easier to follow, we will sometimes write "L'H" over an = sign whose justification comes from L'Hôpital's Rule. Thus, our last computation can be written as

$$\lim_{x\to 0} \frac{e^x - 1}{x} \overset{\text{L'H}}{=} \lim_{x\to 0} \frac{e^x}{1} = e^0 = 1$$

□

Example 7.39:

Some limits we found in Chapter 3 are particularly simple to calculate with L'Hôpital's Rule. For instance,

$$\lim_{x\to 0} \frac{\sin x}{x} \overset{\text{L'H}}{=} \lim_{x\to 0} \frac{\cos x}{1} = \cos 0 = 1$$

To be technical, we used the $(\sin x)/x$ limit to find the derivative of sin in the first place. Thus, we cannot use only L'Hôpital's Rule to calculate this limit, or else we have circular reasoning. We still need the work we did in Chapter 3, using the Squeeze Theorem, to compute this limit in the first place. □

Remark. Note that L'Hôpital's Rule says that if f'/g' approaches a limit, and f/g is a $0/0$ form, then f/g approaches that same limit. The converse does not necessarily hold: namely, f/g might have a limit where f'/g' doesn't. For instance, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Back in Chapter 4 (see Figure 4.10), we saw that $f'(0) = 0$ but $f'(x)$ has no limit as $x \to 0$. Therefore,

$$\lim_{h\to 0} \frac{f(h) - f(0)}{h} \text{ exists, but } \lim_{h\to 0} \frac{f'(h)}{1} \text{ does not exist.}$$

Hence, L'Hôpital's Rule cannot be used here to determine the limit of $(f(h) - f(0))/h$.

It is also important to make sure that your original limit really has a $0/0$ form before applying L'Hôpital's Rule. For instance, if we apply L'Hôpital's Rule without checking this condition, then we obtain

$$\lim_{x\to 1} \frac{x^2 - 1}{x^3 - 3x^2 + 1} \stackrel{\text{L'H}}{=} \lim_{x\to 1} \frac{2x}{3x^2 - 6x} = \frac{2}{3-6} = \frac{-2}{3}$$

but the correct value of the first limit is 0.

Example 7.40:

Occasionally, we can use L'Hôpital's Rule multiple times to find a limit. This is useful when each use of L'Hôpital's Rule looks like it is simplifying the fraction at hand. For instance, let's compute

$$\lim_{x\to 0} \frac{\sin x - x}{x^2}$$

Because this is a $0/0$ form, L'Hôpital's Rule suggests that we consider the limit

$$\lim_{x\to 0} \frac{\cos x - 1}{2x}$$

This limit is still in $0/0$ form, but the denominator is simpler than before. Using L'Hôpital's Rule one more time, we get

$$\lim_{x\to 0} \frac{-\sin x}{2} = 0$$

These steps can be put together and written as

$$\lim_{x\to 0} \frac{\sin x - x}{x^2} \overset{\text{L'H}}{=} \lim_{x\to 0} \frac{\cos x - 1}{2x} \overset{\text{L'H}}{=} \lim_{x\to 0} \frac{-\sin x}{2} = 0$$

□

Remark. When using L'Hôpital's Rule multiple times, it's a good idea to see if your limits are becoming simpler with each step. If not, then you should probably use some other tactics before trying L'Hôpital's Rule. For instance, consider the limit from Example 7.32:

$$\lim_{x\to 0^+} \frac{e^{-1/x}}{x}$$

If we try L'Hôpital's Rule with this, the limit becomes

$$\lim_{x\to 0^+} \frac{(1/x^2)e^{-1/x}}{1} = \lim_{x\to 0^+} \frac{e^{-1/x}}{x^2}$$

which looks worse. In fact, for any $n > 0$, we find that trying L'Hôpital's Rule with $\exp(-1/x)/(x^n)$ yields

$$\lim_{x\to 0^+} \frac{(1/x^2)e^{-1/x}}{nx^{n-1}} = \lim_{x\to 0^+} \frac{e^{-1/x}}{nx^{n+1}}$$

which has a worse denominator. Therefore, applying L'Hôpital's Rule over and over will never get rid of the $0/0$ form here.

Example 7.41:

There are situations in which L'Hôpital's Rule applies but where using it may take longer to find a solution. For instance, consider

$$\lim_{x\to 0} \frac{x - \tan x}{x - \sin x}$$

If we take derivatives of the top and bottom to use L'Hôpital's Rule, we get

$$\lim_{x\to 0} \frac{1 - \sec^2 x}{1 - \cos x}$$

At this point, we still have a 0/0 form. One way to proceed is to use L'Hôpital's Rule again, giving

$$\lim_{x\to 0} \frac{-(2\sec x)(\sec x \tan x)}{\sin x} = \lim_{x\to 0} \frac{-2\sec^2 x \sin x}{\sin x \cos x} = \frac{-2\sec^2(0)}{\cos 0} = -2$$

Note that we make sure to rewrite the expression and get $\sin x$ to cancel. If we didn't bother to do that, then the result after yet another use of L'Hôpital's Rule would look even more complicated and would be harder to solve for the final answer!

An alternative to a second use of L'Hôpital's Rule is to multiply numerator and denominator by $\cos^2 x$ (to cancel out the $\sec^2 x$) and then factor a difference of squares:

$$\begin{aligned} \frac{1-\sec^2 x}{1-\cos x} \cdot \frac{\cos^2 x}{\cos^2 x} &= \frac{\cos^2 x - 1}{(1-\cos x)\cos^2 x} \\ &= \frac{(\cos x - 1)(\cos x + 1)}{(1-\cos x)\cos^2 x} \\ &= \frac{-(\cos x + 1)}{\cos^2 x} \end{aligned}$$

(An equivalent way to do this algebra is to write $\sec^2 x$ as $1/(\cos^2 x)$ and simplify the resulting nested fraction.) This also gives us the limit of -2 as $x \to 0$, and arguably this approach is easier than a second use of L'Hôpital's Rule. (After all, you should frequently simplify expressions, especially trigonometric expressions, before using L'Hôpital's Rule again.) $\square$

Other Kinds of Indeterminate Forms

We have spent a lot of time discussing how to attack limits which attain a 0/0 form. This kind of form is often called an *indeterminate form*, because the form tells you no information about what the limit actually is. For instance, for each $A \in \mathbb{R}$, the limit

$$\lim_{x\to 0} \frac{Ax}{x}$$

is a 0/0 indeterminate form, and its limit is A, so the limit of a 0/0 form could be any real number. There isn't even any guarantee that an expression with an indeterminate form has a limit at all: looking at

$$\lim_{x\to 0} \frac{x}{x^2} \text{ and } \lim_{x\to 0} \frac{x\sin(1/x)}{x}$$

the expression on the left diverges to $\pm\infty$, and the expression on the right oscillates wildly.

Another indeterminate form is the ∞/∞ *form*, i.e. a fraction where the numerator and denominator each approach either ∞ or $-\infty$ (conventionally, signs are ignored when writing types of forms). You can construct examples similar to the ones above which show that an ∞/∞ form could produce any possible limit. It turns out that L'Hôpital's Rule also applies to ∞/∞ forms, though the proof is rather different and is outlined in Exercise 7.10.32:

Theorem 7.42 (L'Hôpital's Rule for ∞/∞). *Let $a \in \mathbb{R}$ be given, and let f, g be real functions defined near a. Suppose that $f(x)$ and $g(x)$ each approach ∞ or $-\infty$ as $x \to a$. Also assume that f and g are differentiable near a, and suppose that for some $L \in \mathbb{R}$,*

$$\lim_{x\to a} \frac{f'(x)}{g'(x)} = L$$

Then $f(x)/g(x)$ has a limit as $x \to a$, and

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} = L$$

Similar versions of this theorem also hold for one-sided limits, when a is ∞ or $-\infty$, as well as when L is ∞ or $-\infty$. Proofs for these versions are constructed by the same modifications as with the $0/0$ theorem.

Example 7.43:

With L'Hôpital's Rule for ∞/∞, there is a trivial proof of Theorem 7.33: for any $a > 0$ and $b > 1$,

$$\lim_{x\to\infty} \frac{\log_b x}{x^a} \overset{\text{L'H}}{=} \lim_{x\to\infty} \frac{1/(x\log b)}{ax^{a-1}} = \lim_{x\to\infty} \frac{1}{ax^a \log b} = 0$$

because our original limit is an ∞/∞ form. There is a similar simple proof of Theorem 7.30 for computing

$$\lim_{x\to\infty} \frac{x^a}{b^x} = 0$$

when $b > 1$ and $a \in \mathbb{N}$: you use L'Hôpital's Rule a times. (L'Hôpital's Rule does not prove this theorem when a is not a natural number, but once you know this theorem for natural numbers, it is not hard to prove it for arbitrary positive real numbers a.) $\square$

We can switch between $0/0$ and ∞/∞ forms by noting that

$$\frac{f}{g} = \frac{1/g}{1/f}$$

so that if f and g both approach 0, then $1/g$ and $1/f$ both approach $\pm\infty$, and vice versa. Sometimes this, or other ways of rewriting, helps simplify a calculation, as the following example shows:

Example 7.44:

Consider

$$\lim_{x\to\infty} \frac{1/x}{1/\log x}$$

This is currently written as a $0/0$ form. Trying L'Hôpital's Rule on this yields

$$\lim_{x\to\infty} \frac{-1/x^2}{-1/(x(\log x)^2)} = \lim_{x\to\infty} \frac{(\log x)^2}{x}$$

If we don't use Theorem 7.33 here, and instead we keep applying L'Hôpital's Rule, then you can check that we need two more applications of the rule. You also need to do a fair bit of work computing these derivatives.

Instead of doing these complicated steps, we rewrite our original limit as

$$\lim_{x\to\infty} \frac{\log x}{x}$$

which is an ∞/∞ form. We have seen that L'Hôpital's Rule works much better with this expression and gives the limit 0.

For a similar example, recall the limit

$$\lim_{x\to 0^+} \frac{e^{-1/x}}{x}$$

We have seen that in this $0/0$ form, L'Hôpital's Rule does not help. However, if we set $u = 1/x$, then we get

$$\lim_{u\to\infty} \frac{e^{-u}}{1/u} = \lim_{u\to\infty} \frac{u}{e^u}$$

which is an easy ∞/∞ form to which you can apply L'Hôpital's Rule. $\square$

A related indeterminate form is the $0 \cdot \infty$ *form*, which describes a product of the form fg where f approaches 0 and g approaches ∞ or $-\infty$. By writing

$$fg = \frac{f}{1/g} = \frac{g}{1/f}$$

we can turn $0 \cdot \infty$ forms into either $0/0$ forms or ∞/∞ forms (pick whichever form makes the rest of the work easier). We present one example.

Example 7.45:

We previously found the limit

$$\lim_{x \to 0^+} x \log x$$

by substituting $u = 1/x$ and doing some rewriting to use Theorem 7.33. We can also rewrite this $0 \cdot \infty$ form (note that $\log x$ goes to $-\infty$) as an ∞/∞ form as follows:

$$\lim_{x \to 0^+} \frac{\log x}{1/x} \stackrel{\text{L'H}}{=} \lim_{x \to 0^+} \frac{1/x}{-1/x^2} = \lim_{x \to 0^+} -x = 0$$

We can describe this tactic by saying that we moved x into the denominator as $1/x$ to apply L'Hôpital's Rule. If we instead try to move $\log x$ to the denominator, to obtain

$$\lim_{x \to 0^+} \frac{x}{1/\log x}$$

then the limit we obtain from L'Hôpital's Rule is

$$\lim_{x \to 0^+} \frac{1}{-1/(x(\log x)^2)} = \lim_{x \to 0^+} -x(\log x)^2$$

which is worse than our original problem. In general, you should move polynomials into the denominator instead of logarithms, because $1/\log x$ has a more complicated derivative than $1/x$. $\square$

The next type of indeterminate form we will mention is the $\infty - \infty$ *form*. This is obtained by subtracting two functions which both tend to ∞, or equivalently, by adding two functions which tend to opposite signs of ∞. When trying to handle a limit with this form, we usually try and write our limit as a product or quotient, such as by making a common denominator, factoring, or rationalizing. Here are some examples.

Example 7.46:

Consider the limit

$$\lim_{x\to 0} \frac{1}{x} - \frac{1}{\sin x}$$

This has the form $\infty - \infty$. Since we are subtracting fractions here, we try making a common denominator:

$$\lim_{x\to 0} \frac{\sin x - x}{x \sin x}$$

This has changed our limit to a 0/0 form, and L'Hôpital's Rule yields

$$\lim_{x\to 0} \frac{\cos x - 1}{\sin x + x \cos x}$$

This is still in 0/0 form, and we try L'Hôpital's Rule once more to get the answer:

$$\lim_{x\to 0} \frac{-\sin x}{\cos x + \cos x - x \sin x} = \frac{-\sin 0}{2\cos 0 - 0 \sin 0} = 0$$

$\square$

Example 7.47:

Consider the limit

$$\lim_{x\to\infty} x - \sqrt{x}$$

which has the form $\infty - \infty$. To handle this, we factor $\sqrt{x}$ out of both terms to get $\sqrt{x}(\sqrt{x} - 1)$. Written this way, both factors tend to ∞, so the product tends to ∞. $\square$

Example 7.48:

Consider the limit

$$\lim_{x\to\infty} \sqrt{x^2+1} - x$$

Here, because of the square root, we try rationalization. Thus, we multiply and divide by $\sqrt{x^2+1} + x$ and get

$$\lim_{x\to\infty} \frac{(x^2+1) - (x)^2}{\sqrt{x^2+1} + x} = \lim_{x\to\infty} \frac{1}{\sqrt{x^2+1} + x} = 0$$

because the denominator tends to ∞. $\square$

Lastly, we will consider some indeterminate forms involving exponentials. Consider an expression of the form f^g. When f tends to ∞ and g tends to 0, we say f^g is an ∞^0 *form*. Informally, this is an indeterminate form because the base is trying to make the expression large, but the power is trying to make the expression small, so it's not clear what the final limit is. To be more precise, for any $r > 0$, we can make an ∞^0 form with limit r as follows:

$$\lim_{x\to\infty} x^{\log r/\log x} = \lim_{x\to\infty} e^{(\log r/\log x)\log x} = e^{\log r} = r$$

Another indeterminate form involving exponentials is the 0^0 form. The base tends to 0, so the base is trying to make the exponential small, but the power is also tending to 0, so the power is trying to make the exponential close to 1. We have seen an example of this before in Example 7.35: there, we showed

$$\lim_{x\to 0^+} x^x = \lim_{x\to 0^+} \exp(x\log x) = e^0 = 1$$

However,

$$\lim_{x\to 0^+} 0^x$$

is another 0^0 form with limit 0.

One more type of indeterminate form is the 1^∞ *form*, which looks like f^g where f tends to 1 and g tends to ∞. Informally, this is indeterminate because a base of 1 causes all powers to yield 1, but the power is also tending to ∞. Consider the following example:

Example 7.49:

For any $a \in \mathbb{R}$, consider the limit

$$\lim_{x\to 0}(1+ax)^{1/x}$$

which has the form 1^∞. Since this is a power whose base and exponent both depend on x, we first rewrite it with a base of e by the definition of irrational exponents. Our expression becomes

$$\lim_{x\to 0}\exp((1/x)\log(1+ax)) = \exp\left(\lim_{x\to 0}\frac{\log(1+ax)}{x}\right)$$

The limit inside the exp is a 0/0 form and can be evaluated easily with L'Hôpital's Rule. We get

$$\exp\left(\lim_{x\to 0}\frac{a}{1+ax}\right) = e^a$$

Another way we could find this limit is to let $f(x) = (1+ax)^{1/x}$ for all $x > 0$, and then we take a logarithm to bring the exponent down:

$$\log f(x) = \frac{\log(1+ax)}{x}$$

As before, we then show that $\log f(x) \to a$ as $x \to 0$, so that $f(x)$ approaches e^a. $\square$

Remark. The limit in the previous example can be expressed in several different forms by changing the variable. For instance, if $a \neq 0$ and we let $t = ax$, so that $1/x = a/t$ and $t \to 0$ as $x \to 0$, then our limit becomes

$$\lim_{t \to 0} (1+t)^{a/t} = e^a$$

Also, if we let $u = 1/x$ in the limit from the example (or $u = 1/t$ in the limit above), we obtain

$$\lim_{u \to \infty} \left(1 + \frac{a}{u}\right)^u = \lim_{u \to \infty} \left(1 + \frac{1}{u}\right)^{au} = e^a$$

In particular, we obtain

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = e$$

These forms of the limit are very useful when calculating compound interest. To see why, suppose that we are investing P dollars (P stands for "Principal") at a rate of interest r (where $r = 0.1$ means 10%, $r = 0.5$ means 50%, and so on), but we compound this interest n times a year for t years. Without compounding, each year we would multiply our current amount of money by $1 + r$, so that after t years we would have $P(1+r)^t$ dollars in the bank. However, when we compound the interest n times a year, we accrue interest n times as often, but the rate of interest is divided by n to compensate. Thus, we'd earn

$$P\left(1 + \frac{r}{n}\right)^{nt}$$

dollars after t years. (You can show in Exercise 7.10.28 that this is an increasing function of n when t and r are fixed, so compounding more often gives more money.)

If we were hypothetically able to compound our interest infinitely often, i.e. we let n tend to ∞, then our amount of money becomes

$$\lim_{n\to\infty} P\left(1+\frac{r}{n}\right)^{nt} = P\left(\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^{n}\right)^{t} = Pe^{rt}$$

This formula (sometimes affectionately called the "Pert formula") tells you how much money you make when interest is *compounded continuously.* Note that as a function of time t, we have

$$\frac{d}{dt}(Pe^{rt}) = rPe^{rt}$$

so that the rate at which you earn money is proportional to how much money you have.

In general, an indeterminate form with an exponential can be handled either by writing it in terms of exp or by first taking a logarithm. In either case, we are left with a limit of $0\cdot\infty$ form, which can be handled using methods discussed previously. Note that expressions of the form ∞^1 and 0^∞ are NOT considered indeterminate forms, since you can show that any expression of the form ∞^1 tends to ∞ and any expression of the form 0^∞ tends to 0. (After all, an indeterminate form needs to have the property that the form alone cannot tell you what the limit is.)

We summarize the indeterminate forms we have discussed in the following table. The simplest kinds, which can frequently be handled by L'Hôpital's Rule, are the $0/0$ and the ∞/∞ forms. The rest of the forms are usually turned into simpler forms, as the table indicates:

Form	What to change it to	Suggested ways to handle it
$0/0$	∞/∞ if needed	Algebra, L'Hôpital's Rule, or change to $u=1/x$ if needed
∞/∞	$0/0$ if needed	
$0\cdot\infty$	$0/0$ or ∞/∞	Move something to the denominator
$\infty-\infty$	$0/0$, ∞/∞, or $0\cdot\infty$	Common denominator, Factoring, Rationalization, Algebra
0^0	$0\cdot\infty$	Take a logarithm, or write in terms of exp
∞^0	$0\cdot\infty$	
1^∞	$0\cdot\infty$	

You can practice computing many limits with these indeterminate forms in the exercises.

7.10 Exercises

For Exercises 1 through 7, use Theorems 7.30 and 7.33 to compute the following limits (i.e. do not use L'Hôpital's Rule for these problems):

1. $\lim\limits_{x\to\infty} \dfrac{x^3\, e^x}{e^{3x}}$
2. $\lim\limits_{x\to\infty} \dfrac{(e^x + x)^2}{e^{2x}}$
3. $\lim\limits_{x\to 0^+} \dfrac{e^{-1/x^2}}{x^a}$ for any $a > 0$
4. $\lim\limits_{x\to 0^+} x^a \log x$ for any $a > 0$
5. $\lim\limits_{x\to\infty} \dfrac{\log\log x}{x}$
6. $\lim\limits_{x\to 0} x^5 \log\left(\dfrac{1}{x^2}\right)$
7. $\lim\limits_{x\to 0^+} (2x)^{3x}$
8. Give a separate proof of Theorem 7.33 using Theorem 7.36.
9. Give another proof of Theorem 7.36 by using the following steps. First, let $a > 0$ be given, and consider the expression
$$f(x) = \frac{x^a}{a} - \log x$$
defined for all $x > 0$. Show that $f(1) \geq 0$ and that $f'(t) > 0$ for all $t > 1$. Lastly, prove that $f(x) > 0$ for all $x > 1$, which proves the theorem.
10. Theorem 7.36 only applies to values of x greater than 1. Prove the following inequality for all $x \in (0, 1)$ and all $a > 0$:
$$\log x > \frac{-1}{ax^a}$$
Thus, as $x \to 0^+$, $|\log x|$ tends to ∞ more slowly than any function of the form $1/x^a$. (Hint: Let $u = 1/x$.)

For Exercises 11 through 27, compute the following limits. You may use L'Hôpital's Rule for either $0/0$ or ∞/∞ forms when applicable, but please indicate clearly when the rule is being used. Also note that L'Hôpital's Rule isn't always the best way to approach $0/0$ or ∞/∞ forms.

11. $\lim_{x\to 0} \frac{a^x - b^x}{x}$
for any $a, b > 0$ with $a \neq b$

12. $\lim_{x\to 0^+} \frac{e^x - 1 - x}{x^2}$

13. $\lim_{x\to 0} \frac{\log(\cos(ax))}{\log(\cos(bx))}$
for any $a, b \neq 0$

14. $\lim_{x\to 0} \frac{\log(x+1)}{x}$

15. $\lim_{x\to 0} \frac{\sin x - x}{x^3}$

16. $\lim_{x\to 0} \frac{e^x - e^{-x} - 2\sin x}{2x^3}$

17. $\lim_{x\to 1} \frac{x^x - x}{1 - x + \log x}$

18. $\lim_{x\to 0} \sin^2 x \tan(x + \pi/2)$

19. $\lim_{x\to \infty} x^{1/4} \sin(1/\sqrt{x})$

(Hint: This limit is easier to compute if you make a change of variable.)

20. $\lim_{x\to 0^+} \sin x \log x$

21. $\lim_{x\to 1^-} \log x \log(1-x)$

(Hint: After a rewrite and one use of L'Hôpital's Rule, you do not have to perform L'Hôpital's Rule a second time on EVERY part of the fraction you obtain. You can put some terms off to the side first.)

22. $\lim_{x\to -\infty} xe^x$

23. $\lim_{x\to \infty} \sqrt{x^4 + x^2} - x^2$

(Hint: Get this into another form, but don't use L'Hôpital's Rule. Instead, multiply and divide by something.)

24. $\lim_{x\to -\infty} e^{-x} + x$

(Hint: A previous exercise in this section might be useful.)

25. $\lim_{x\to \infty} x^{1/x}$

26. $\lim_{x\to 0^+} (\sin x)^x$

27. $\lim_{x\to 1} (2 - x)^{1/(e^x - e)}$

28. In this exercise, suppose that r and t are positive constants, and we will prove that the function $f : (0, \infty) \to \mathbb{R}$ defined by

$$f(n) = \left(1 + \frac{r}{n}\right)^{nt} \text{ for all } n > 0$$

is a strictly increasing function. $f(n)$ represents the amount of money earned from \$1 by compounding n times per year at an interest rate of r over t years. Intuitively, when $n \in \mathbb{N}^*$, $n > 1$, and you compound n times per year instead of once per year, the interest earned from the first compounding accumulates additional interest, which shows that $f(n) > f(1)$. However, to prove more formally that f is strictly increasing on its domain, we will show f' is positive on its domain.

(a) Show that for all $n > 0$,

$$f'(n) = t\left(1 + \frac{r}{n}\right)^{nt}\left(\log\left(1 + \frac{r}{n}\right) - \frac{r}{n+r}\right)$$

(b) From Exercise 7.2.5, we find that for all $x > 0$ with $x \neq 1$,

$$\log x > 1 - \frac{1}{x}$$

Use this and part (a) to prove that $f'(n) > 0$ for all $n > 0$.

29. In this exercise, we use the fact that

$$\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = e$$

to approximate e. Let's use the definition of f from the previous exercise with $r = t = 1$. Because f is strictly increasing, we know that

$$f(1) \leq f(2) \leq f(3) \leq \cdots \leq e = \sup\{f(n) \mid n \in \mathbb{N}^*\}$$

(a) Using the Binomial Theorem (which can be found in Exercise 4.2.5.(c)), prove that for all $n \in \mathbb{N}^*$,

$$f(n) = \left(1 + \frac{1}{n}\right)^n = \sum_{i=0}^{n}\left(\frac{1}{i!}\prod_{j=0}^{i-1}\left(1 - \frac{j}{n}\right)\right)$$

Note that when $i = 0$, the product has no terms in it and is considered to be 1, so we can also write this

$$f(n) = 1 + \sum_{i=1}^{n} \frac{1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{i-1}{n}\right)}{i!}$$

(b) Use part (a) to conclude that for all $n \in \mathbb{N}^*$,

$$f(n) \le \sum_{i=0}^{n} \frac{1}{i!}$$

NOTE: It follows from Exercise 7.4.27 that

$$\sum_{i=0}^{n} \frac{1}{i!} < e$$

Therefore, since $f(n) \to e$ as $n \to \infty$, the Squeeze Theorem tells us that this summation approaches e as $n \to \infty$. We will also see a different proof of this limit in the next chapter.

(c) In Example 1.59, we showed that $n! > 2^n$ when $n \in \mathbb{N}^*$ and $n \ge 4$. Use this and a geometric series with part (b) to prove that for all $n \in \mathbb{N}^*$ with $n \ge 4$, $f(n) < 2.8$. Therefore, $e < 2.8$.

30. Prove L'Hôpital's Rule for $0/0$, Theorem 7.37, in the case where a is ∞ or $-\infty$. (Hint: Make a change of variable.)

31. Prove L'Hôpital's Rule for $0/0$, Theorem 7.37, in the case where L is ∞ or $-\infty$. (This will involve small modifications of the original proof.)

32. This exercise will outline a proof for L'Hopital's Rule for ∞/∞, Theorem 7.42. For simplicity, we will consider the case where $a, L \in \mathbb{R}$ (i.e. neither is $\pm\infty$) and we approach a from the right. The proofs of the other cases can be done in the same manner as the previous two exercises. We can also suppose WLOG that $f(x), g(x) \to \infty$ as $x \to a^+$. (This is because if a function approaches $-\infty$, then its negation approaches ∞, and we can factor out minus signs.)

    First, we present the main strategy (you will prove the formal details in the parts following this strategy). We want to prove that $f(x)/g(x) \to$

L by the definition of limit, so for all $\epsilon > 0$, we want to find $\delta > 0$ so that

$$\forall x \in (a, a+\delta) \ \left(\left| \frac{f(x)}{g(x)} - L \right| < \epsilon \right)$$

To do this, suppose that z is some value greater than a but close to a. Let $\Delta f = f(x) - f(z)$ and $\Delta g = g(x) - g(z)$. We will make sure that $\Delta f, \Delta g \neq 0$ provided x is close enough to a. Then the Triangle Inequality guarantees

$$\left| \frac{f(x)}{g(x)} - L \right| \leq \left| \frac{f(x)}{g(x)} - \frac{\Delta f}{\Delta g} \right| + \left| \frac{\Delta f}{\Delta g} - L \right|$$

so we will aim to make each absolute value on the right side less than $\epsilon/2$. We can show $\Delta f / \Delta g$ is close to L by using the MVT, and we can also note that

$$\frac{f(x)}{g(x)} - \frac{\Delta f}{\Delta g} = \left(\frac{\Delta f}{\Delta g} \cdot \frac{f(x)}{\Delta f} \cdot \frac{\Delta g}{g(x)} \right) - \frac{\Delta f}{\Delta g} = \frac{\Delta f}{\Delta g} \left(\frac{f(x)}{\Delta f} \cdot \frac{\Delta g}{g(x)} - 1 \right)$$

It remains to show this expression is less than $\epsilon/2$. Because f and g tend to ∞, we can show that $f(x)/\Delta f$ and $g(x)/\Delta g$ are close to 1. We will also use the fact that $\Delta f / \Delta g$ is close to L.

Now, we start the proof.

(a) Let $\epsilon > 0$ be given. Because we assume $f'(x)/g'(x) \to L$ as $x \to a^+$, there is some $\delta_1 > 0$ so that

$$\forall x \in (a, a+\delta_1) \ \left(\left| \frac{f'(x)}{g'(x)} - L \right| < \frac{\epsilon}{2} \right)$$

Let $z = a + \delta_1$. Prove that there exists $\delta_2 \in (0, \delta_1)$ such that for all $x \in (a, a+\delta_2)$, $f(x) > |f(z)|$ and $g(x) > |g(z)|$.

(b) Now let $x \in (a, a+\delta_2)$ be given. Let $\Delta f = f(x) - f(z)$ and $g(x) - g(z)$. By part (a), $\Delta f, \Delta g \neq 0$. Prove that

$$\left| \frac{\Delta f}{\Delta g} - L \right| < \frac{\epsilon}{2}$$

Also prove the following consequence:

$$\left| \frac{\Delta f}{\Delta g} \right| < |L| + \frac{\epsilon}{2}$$

(c) Now, as suggested by the strategy, we can write

$$\left|\frac{f(x)}{g(x)} - \frac{\Delta f}{\Delta g}\right| = \left|\frac{\Delta f}{\Delta g}\right| \left|\frac{f(x)}{\Delta f} \cdot \frac{\Delta g}{g(x)} - 1\right|$$

(Since $f(x) > |f(z)|$ and $g(x) > |g(z)|$ by part (a), it follows that $f(x), g(x) \neq 0$.) Prove that

$$\lim_{x \to a^+} \frac{f(x)}{\Delta f} = \lim_{x \to a^+} \frac{\Delta g}{g(x)} = 1$$

Recall that z is constant with respect to x. (Hint: For each fraction, multiply and divide by something.)

(d) From part (c) and the Product Rule for limits, it follows that there exists $\delta_3 \in (0, \delta_2)$ such that

$$\forall x \in (a, a+\delta_3) \left(\left|\frac{f(x)}{\Delta f} \cdot \frac{\Delta g}{g(x)} - 1\right| < \frac{\epsilon/2}{|L| + \epsilon/2}\right)$$

Let $\delta = \delta_3$. Use the previous parts to show that

$$\forall x \in (a, a+\delta) \left(\left|\frac{f(x)}{g(x)} - L\right| < \epsilon\right)$$

finishing the proof.

Chapter 8

Approximating with Taylor Polynomials

Up to this point, we've seen that we can use lines to approximate smooth functions. This raises a few questions. For instance, could we use parabolas for approximation? What about cubic functions? If so, then how well do these approximations work? Does a parabola produce better approximations than a line?

To illustrate these ideas, let's revisit the idea of a tangent line. More specifically, if f is a real function which is differentiable at some $a \in \mathbb{R}$, then the tangent line to f at a is defined by the equation

$$y - f(a) = f'(a)(x - a)$$

Let's call this function f_1, so $f_1(x) = f(a) + f'(a)(x-a)$ for all $x \in \mathbb{R}$. Thus, f_1 is a linear function (or constant function if $f'(a) = 0$) which seems to be a good approximation for f near a, i.e.

$$f(x) \approx f_1(x) = f(a) + f'(a)(x - a)$$

when x is close to a. Intuitively, this is because at a, f_1 and f have the same value and the same slope, so their graphs are "going in the same direction." Linear approximations are very useful because linear functions are easy to use in calculations.

As an example, consider when $f(x) = e^x$ for all $x \in \mathbb{R}$ and $a = 0$. We have $f' = f$, so $f'(0) = f(0) = e^0 = 1$, so $f_1(x) = 1 + x$ for all $x \in \mathbb{R}$. You can see graphs of f and f_1 in Figure 8.1 (we will introduce the function f_2 shortly).

We see that the graphs of f_1 and f are close for values of x near 0, but they grow farther apart as x gets further from 0. This makes sense, because f_1 is a line whereas f is always concave upward: we have $f''(x) = e^x > 0$ for all $x \in \mathbb{R}$.

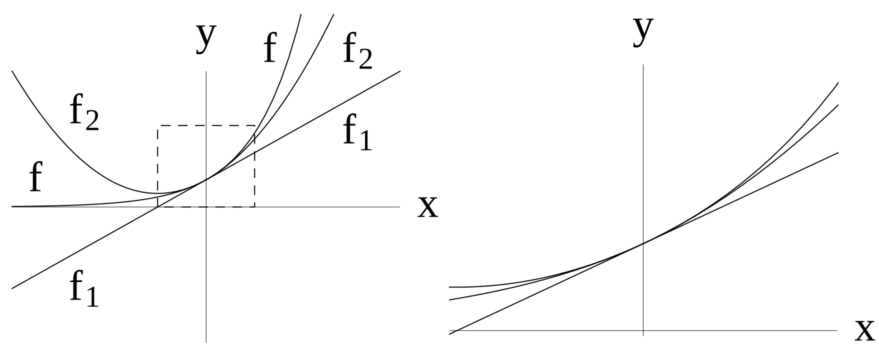

Figure 8.1: Graphs of $f(x) = e^x$, $f_1(x) = 1 + x$, $f_2(x) = 1 + x + x^2/2$ (left), and a zoomed-in graph of the dashed box (right)

This suggests that in order to make a better approximation to f which is still easy to calculate, we should use a function with the same value, the same slope, AND the same concavity as f at 0. Hence, we find a quadratic polynomial f_2 such that $f_2(0) = f(0)$, $f_2'(0) = f'(0)$, and $f_2''(0) = f''(0)$. Since $f''(0) = f'(0) = f(0) = e^0 = 1$, you can readily check that the function defined by

$$f_2(x) = 1 + x + \frac{x^2}{2}$$

for all $x \in \mathbb{R}$ is a quadratic polynomial satisfying these conditions. The graph of f_2 is included as well in Figure 8.1, from which you see that f_2 does a better job than f_1 at approximating f. Also, since f_2 is a quadratic polynomial, it is still easy to use in calculations.

Continuing this train of thought, we can make a cubic polynomial f_3 such that $f_3(0) = f(0)$, $f_3'(0) = f'(0)$, $f_3''(0) = f''(0)$, and $f_3'''(0) = f'''(0)$. f_3 should provide an even better approximation to f than f_2. (Although we have not assigned any physical meaning to the third derivative of a function, like we did with the first and second derivatives, it should still be plausible that f_3 "behaves more like f at 0" than f_2 does, because it agrees with f

at one more derivative.) In fact, you can check by induction that for each $n \in \mathbb{N}$, there is a unique polynomial f_n of degree n satisfying $f_n^{(k)}(0) = f^{(k)}(0)$ for all k from 0 to n: it is defined by

$$f_n(x) = \sum_{i=0}^{n} \frac{x^i}{i!}$$

for all $x \in \mathbb{R}$. Some basic inequalities relating these f_n polynomials to f were given in Exercise 7.4.27.

Since polynomials are easy to differentiate and integrate, and they are fairly easy to use in calculations, these f_n polynomials are convenient approximations to the exponential function f. Intuitively, when we approximate $f(x)$ with $f_n(x)$, our approximation becomes more accurate either when x gets closer to 0 or when n gets larger. Hence, there is a tradeoff: for values of x which are very close to a, a small value of n (i.e. a low-degree polynomial) will suffice for a good approximation, but as x gets further from a, you need to use higher values of n (i.e. more complicated polynomials) to maintain the same level of accuracy.

In general, for any real function f and any $a \in \mathbb{R}$, as long as f has enough derivatives at a (we will formalize "enough" soon), we can introduce approximation polynomials in a similar manner, by making our approximations agree with f and its first few derivatives at a. These polynomials are called *Taylor polynomials*[1]. Our aim for this chapter is to explain this process of computing Taylor polynomials, formalize the notion of "good approximation" by finding bounds on the error obtained by using Taylor polynomials, and to demonstrate applications of Taylor polynomials.

8.1 Computing Taylor Polynomials

Let's show how we can build Taylor polynomials. Suppose that $a \in \mathbb{R}$, $n \in \mathbb{N}$, and f is a real function with n derivatives at a, i.e. $f^{(k)}(a)$ exists for all k from 0 to n. For simplicity, let's first work with the case when $a = 0$. We would like to find a polynomial p of degree at most n such that $f^{(k)}(0) = p^{(k)}(0)$ for all k from 0 to n.

[1]Taylor polynomials are named after Brook Taylor, an English mathematician who popularized their use.

Since p has degree at most n, there exist some $a_0, a_1, \ldots, a_n \in \mathbb{R}$ with

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

for all $x \in \mathbb{R}$. (To simplify our writing, we will use the convention that x^0 is always 1, even when x is 0.) We would like to solve for the a_i constants. To start, since $p(0)$ must equal $f(0)$, and $p(0) = a_0$ (all the other terms disappear), we must have $a_0 = f(0)$. Next, since $p'(0)$ must equal $f'(0)$, we compute

$$p'(x) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1} = \sum_{i=1}^{n} i a_i x^{i-1}$$

so that $p'(0) = a_1 = f'(0)$. Similarly, to obtain a_2, we take another derivative to get

$$p''(x) = 2a_2 + 3(2)a_3 x + \cdots + n(n-1)a_n x^{n-2} = \sum_{i=2}^{n} i(i-1)a_i x^{i-2}$$

yielding $p''(0) = 2a_2 = f''(0)$ and $a_2 = f''(0)/2$.

In general, you can prove by induction on k from 0 to n that

$$p^{(k)}(x) = \sum_{i=k}^{n} i(i-1)\cdots(i-k+1)a_i x^{i-k}$$

When we plug in $x = 0$, every term disappears except for the $i = k$ term, yielding $p^{(k)}(0) = k(k-1)\cdots(1)a_k = k!\, a_k$. Therefore, in order for this to equal $f^{(k)}(0)$, we must choose $a_k = f^{(k)}(0)/k!$. This proves that there is exactly one polynomial p of degree at most n which agrees with the first n derivatives of f at 0, and it is defined by

$$p(x) = \sum_{i=0}^{n} \frac{f^{(i)}(0)}{i!} x^i$$

for all $x \in \mathbb{R}$. (This polynomial will have degree less than n if $f^{(n)}(0) = 0$.)

Now, what do we do if a is not zero? We reduce to the case where a is zero by shifting everything over by a units horizontally. More precisely, if f is n-times differentiable at a, then define $g(x) = f(x + a)$. The Chain Rule shows that g is n-times differentiable at 0, and $g^{(k)}(0) = f^{(k)}(a)$ for all k from 0 to n. Thus, applying our previous result to g, we find that the polynomial

$$q(x) = \sum_{i=0}^{n} \frac{g^{(i)}(0)}{i!} x^i$$

satisfies $q^{(k)}(0) = g^{(k)}(0)$ for each k from 0 to n. If we set $p(x) = q(x - a)$, then we have

$$p(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!} (x - a)^i$$

and $p^{(k)}(a) = f^{(k)}(a)$ for each k from 0 to n. In fact, it is easy to check that p is the unique polynomial of degree at most n which agrees with f and its first n derivatives at a.

This leads to the following definition:

Definition 8.1. Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and let f be a real function which is n-times differentiable at a. The *Taylor polynomial for f of order n centered at a* is the unique polynomial of degree at most n which agrees with f and its first n derivatives at a: its value at any $x \in \mathbb{R}$ is

$$\sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!} (x - a)^i$$

We will use the notation "$T_n f(x; a)$" to represent this. We will sometimes write this as "$T_n(f(x); a)$" instead, though this is a slight abuse of notation (conceptually, the function f and the value $f(x)$ are not the same thing, but this alternate notation ignores that distinction). In the special case where a is zero, we will often drop the "$; a$" and write "$T_n f(x)$". This special case is sometimes also called the *Maclaurin polynomial*[2]. For each k from 0 to n, the fraction $f^{(k)}(a)/k!$ is sometimes called the *Taylor coefficient for f at a of order k*.

Remark. We have used the symbol T_n before when discussing the Trapezoid Rule. Hopefully, the use of T_n will be clear from context (in particular, we will not bring up the Trapezoid Rule again in this chapter). Sometimes we will write "$T_n f$" to denote the Taylor polynomial of order n (here, $T_n f$ is a

[2]In honor of Colin Maclaurin, a Scottish mathematician who further popularized these polynomials.

function, not a value) instead of writing "$T_n f(x;a)$". In these situations, the value of a is usually understood from context or is assumed to be 0. Also, with Taylor polynomials, the second argument a in "$T_n f(x;a)$" is a number, whereas with the Trapezoid Rule, the second argument $[a,b]$ in "$T_n(f;[a,b])$" is an interval.

Example 8.2:

When $f = \exp$ and $a = 0$, then $f^{(k)}(x) = e^x$ for all $k \in \mathbb{N}$, so $f^{(k)}(0) = 1$. This means that for all $n \in \mathbb{N}$ and all $x \in \mathbb{R}$,

$$T_n \exp(x) = \sum_{i=0}^{n} \frac{1}{i!} x^i$$

as previously shown. Since this polynomial is centered at 0, it agrees with exp and its first n derivatives at 0, so informally it should yield good approximations for values of e^x where x is near 0. For example, if we use $n = 5$ and $x = 1$, then we get the approximation

$$e \approx T_5 \exp(1) = 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} = \frac{163}{60} \approx 2.716667$$

This is a reasonably good approximation, even though you might not consider 1 as "being close" to 0.

For any $a \in \mathbb{R}$, we can also easily compute the Taylor polynomial centered at a. We have $f^{(k)}(a) = e^a$ for all $k \in \mathbb{N}$, so we have

$$T_n \exp(x;a) = \sum_{i=0}^{n} \frac{e^a}{i!}(x-a)^i$$

However, in order to use this formula for approximations, we need to first obtain a good approximation of e^a. As a result, only $a = 0$ seems to yield a convenient and practical expression for approximations. (You could try something like $a = \log 2$, so that $e^{\log 2}$ would be simple, but then you'd have difficulty approximating $(x - \log 2)$.)

In general, when computing Taylor polynomials, a different center might be useful because it allows us to approximate some quantities more effectively. For instance, if we want to approximate $f(2)$, then we expect to get a better approximation using $a = 1$ as our center instead of $a = 0$, because 2 is closer to 1 than to 0, and Taylor polynomials tend to be more accurate when their input is close to their center. However, when we change the center for a Taylor polynomial, we might make the polynomial more difficult to use. This shows that although we can theoretically make a Taylor polynomial centered at any a, in practice we tend to use only choices of a leading to approximations which are easy to calculate. $\square$

Example 8.3:

As another example of a simple Taylor polynomial, let's consider when $f(x) = \sin x$ and $a = 0$. The derivatives of sin follow a very simple pattern: $f'(x) = \cos x$, $f''(x) = -\sin x$, $f'''(x) = -\cos x$, and then $f^{(4)} = f$, so the pattern repeats every four derivatives. This pattern seems quite obvious, though it is worth noting that any rigorous argument involving a pattern generally needs an induction proof. One way of proving this by induction would be to prove that for all $n \in \mathbb{N}$, $f^{(4n)} = \sin$, $f^{(4n+1)} = \cos$, $f^{(4n+2)} = -\sin$, and $f^{(4n+3)} = -\cos$. (We leave the details to the reader.)

As a result, we find that $f^{(n)}(0) = 0$ whenever n is even, because $\sin 0 = 0$. On the other hand, $f^{(1)}(0) = \cos 0 = 1$, $f^{(3)}(0) = -\cos 0 = -1$, and in general you can check that $f^{(2k+1)}(0) = (-1)^k$. (Powers of -1 are very useful for describing quantities that keep switching sign.) Therefore, the Taylor coefficients of even order are 0, and the Taylor coefficients of odd order keep switching signs. We have, for any $n \in \mathbb{N}$,

$$T_{2n+1} \sin(x) = \sum_{i=0}^{n} \frac{(-1)^i}{(2i+1)!} x^{2i+1}$$

Note that only odd powers of x occur with nonzero coefficients, so the Taylor polynomial is an odd function. This makes sense, since sin is also an odd function. For more on this, see Exercise 8.2.13.

For convenience in writing the formula, we have written this formula for Taylor polynomials of odd order only. However, since the even-order coefficients are 0, we see that $T_{2n+2} \sin = T_{2n+1} \sin$ for any $n \in \mathbb{N}$. Essentially, this means that any 3^{rd}-order approximation is already a 4^{th}-order approximation, any 5^{th}-order approximation is already a 6^{th}-order approximation, and so forth, so it feels like you gain an order of accuracy for free. (This will be useful to note when we present error bounds in a later section.)

As an example, we can use this to approximate $\sin 1$. We get

$$T_3 \sin(1) = 1 - \frac{1}{6} \approx 0.833333$$

and

$$T_5 \sin(1) = 1 - \frac{1}{6} + \frac{1}{120} \approx 0.841667$$

which is rather close to the value $\sin 1 \approx 0.841471$ obtained by a calculator.

Similarly, we can obtain the Taylor polynomials at $a = 0$ for $f(x) = \cos x$. The pattern we get tells us that $f^{(n)}(0) = 0$ when n is odd and $f^{(2k)}(0) = (-1)^k$. Thus, we have

$$T_{2n} \cos(x) = \sum_{i=0}^{n} \frac{(-1)^i}{(2i)!} x^{2i}$$

Even though this formula only expresses Taylor polynomials of even order, you can also see that $T_{2n+1} \cos = T_{2n} \cos$ for any $n \in \mathbb{N}$. $\square$

With the previous example, since cos is the derivative of sin, you might presume that the Taylor polynomials for cos are the derivatives of the Taylor polynomials for sin. This is true: note that for any $n \in \mathbb{N}$,

$$\frac{d}{dx} \sum_{i=0}^{n} \frac{(-1)^i}{(2i+1)!} x^{2i+1} = \sum_{i=0}^{n} \frac{(-1)^i}{(2i+1)!} (2i+1) x^{2i} = \sum_{i=0}^{n} \frac{(-1)^i}{(2i)!} x^{2i}$$

so that $(T_{2n+1} \sin(x))' = T_{2n} \cos(x)$. In general, we have the following simple properties of T_n:

Theorem 8.4. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and let f and g be real functions which are n-times differentiable at a.*

1. *For any constants $A, B \in \mathbb{R}$,*

$$T_n(Af + Bg)(x; a) = A\, T_n f(x; a) + B\, T_n g(x; a)$$

   In other words, T_n is a linear operation.

2. *If $n > 0$, then*

$$T_{n-1} f'(x; a) = \frac{d}{dx} T_n f(x; a)$$

   Roughly speaking, the Taylor polynomial of a derivative is the derivative of a Taylor polynomial.

3. *If x is close enough to a so that f is continuous and n-times differentiable from a to x (note that when $n > 0$, n-times differentiability implies continuity), then we have*

$$T_{n+1}\left(\int_a^x f(t)\,dt; a\right) = \int_a^x T_n f(t;a)\,dt$$

Roughly speaking, the Taylor polynomial of an integral is the integral of a Taylor polynomial, when the integral starts at the center.

Strategy. All of the parts of this theorem make statements of the form

$$T_N F(x;a) = P(x)$$

for some function F, some $N \in \mathbb{N}$, and some polynomial P. By definition, the Taylor polynomial for F of order N at a is the unique polynomial P satisfying $P^{(k)}(a) = F^{(k)}(a)$ for each k from 0 to n. Each part of this theorem presents a possibility for P; we just need to check the derivatives of P to see if they agree with the derivatives of F at a.

Proof. For all the parts of this theorem, let a, n, f, g be given as described. For part 1, also suppose that $A, B \in \mathbb{R}$ are given. Since f and g are n-times differentiable at a, so is $Af + Bg$. Also, for each k from 0 to n, we have

$$(Af + Bg)^{(k)}(a) = Af^{(k)}(a) + Bg^{(k)}(a)$$

by linearity of derivatives. By the definitions of $T_n f$ and $T_n g$, this is equal to

$$(A\,T_n f + B\,T_n g)^{(k)}(a)$$

(we have omitted the $(x;a)$'s for convenience). Thus, $A\,T_n f + B\,T_n g$ agrees with $Af + Bg$ and its first n derivatives at a. Also, note that $A\,T_n f + B\,T_n g$ is a linear combination of two polynomials of degree at most n, so it is a polynomial of degree at most n. Thus, by definition, it must be $T_n(Af + Bg)$.

For part 2, suppose $n > 0$. Since f is n-times differentiable at a, f' is $(n-1)$-times differentiable at a. Also, for any k from 0 to $(n-1)$, the definition of $T_n f$ tells us that

$$(f')^{(k)}(a) = f^{(k+1)}(a) = (T_n f)^{(k+1)}(a) = ((T_n f)')^{(k)}(a)$$

Thus, $(T_n f)'$ agrees with f' and its first $n-1$ derivatives at a. Furthermore, $T_n f$ has degree at most n, so $(T_n f)'$ is a polynomial of degree at most $n-1$. This shows $(T_n f)' = T_{n-1} f'$, as desired.

For part 3, suppose that x is close enough to a so that f is continuous and n-times differentiable from a to x. We define

$$P(x) = \int_a^x T_n f(t; a)\, dt$$

For any t from a to x, f is continuous at t, so the FTC tells us that $P'(x) = T_n f(x; a)$. Thus, for any k from 1 to $n+1$, we have

$$P^{(k)}(a) = (T_n f(x; a))^{(k-1)}(a) = f^{(k-1)}(a) = \left(\int_a^x f(t; a)\, dt \right)^{(k)} (a)$$

Also, when $k = 0$, we clearly have

$$P(a) = \int_a^a T_n f(t; a)\, dt = 0 = \int_a^a f(t)\, dt$$

(This is why we need the lower limit of integration to be a). Thus, P agrees with $\displaystyle\int_a^x f(t)\, dt$ and its first $n+1$ derivatives at a. Since P is obtained by integrating a polynomial of degree at most n, P is a polynomial of degree at most $n+1$. Thus, by definition, P is $T_{n+1}\left(\displaystyle\int_a^x f(t)\, dt; a \right)$, as desired. ■

Theorem 8.4 allows us to take Taylor polynomials we have already computed and manipulate them to obtain Taylor polynomials for new functions. For instance, consider the following example.

Example 8.5:

Let $a = 0$ and define $f : \mathbb{R} - \{1\} \to \mathbb{R}$ by

$$f(x) = \frac{1}{1-x}$$

for all $x \neq 1$. It is not hard to see that f is infinitely differentiable at $a = 0$. In fact, every derivative of f is a rational function whose denominator is zero only at $x = 1$: we compute

$$f'(x) = \frac{1}{(1-x)^2} \qquad f''(x) = \frac{2}{(1-x)^3} \qquad f'''(x) = \frac{6}{(1-x)^4}$$

and in general, you can prove by induction that

$$f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}}$$

for all $k \in \mathbb{N}$ and all $x \neq 1$. Therefore, $f^{(k)}(0) = k!$, so the k^{th} Taylor coefficient equals $k!/k! = 1$. This tells us that for all $n \in \mathbb{N}$,

$$T_n f(x) = \sum_{i=0}^{n} x^i$$

We have seen this kind of summation before. For instance, the staircase function uses this kind of summation with $x = 1/2$. In this summation, the ratio of any two consecutive terms is constant with respect to i: for any i from 0 to $n-1$, we have $x^{i+1}/x^i = x$. We describe this by saying that this is a *geometric summation*. In general, Exercise 1.9.2 tells us that the sum of a geometric summation is

$$\sum_{i=0}^{n} x^i = \frac{x^{n+1}-1}{x-1}$$

for $x \neq 1$. Hence, we have

$$f(x) - T_n f(x) = \frac{1}{1-x} - \sum_{i=0}^{n} x^i = \frac{x^{n+1}}{1-x}$$

so in this case, we know the exact error of our Taylor approximation. We will return to this example in the upcoming section on error analysis.

Now that we have the Taylor polynomial for $1/(1-x)$, we can use the fact that $1/(1-x)$ is the derivative of $-\log(1-x)$, along with Parts 1 and 3 of Theorem 8.4, to prove

$$T_{n+1}(\log(1-x)) = -\int_0^x \sum_{i=0}^{n} t^i \, dt = \sum_{i=0}^{n} \frac{-x^{i+1}}{i+1}$$

By following similar steps with the function $1/(1+x)$, we get

$$T_n\left(\frac{1}{1+x}\right) = \sum_{i=0}^{n} (-1)^i x^i$$

and

$$T_{n+1}(\log(1+x)) = \sum_{i=0}^{n} \frac{(-1)^i x^{i+1}}{i+1}$$

We can use these Taylor polynomials to estimate natural logarithms, thus giving us another way of approximating logarithms (previously, we used integral approximation techniques to estimate logarithms). For instance, by using $x = 1$ and $n = 4$, we get

$$\log 2 \approx \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} = \frac{47}{60} \approx 0.783333$$

□

Some Plausible Conjectures About Taylor Polynomials

Looking back at Example 8.5, suppose we define $f(x) = 1/(1-x)$ and

$$P(x) = T_n f(x) = \sum_{i=0}^{n} x^i$$

We also found that

$$T_n\left(\frac{1}{1+x}\right) = \sum_{i=0}^{n} (-x)^i = P(-x)$$

This makes sense, as $1/(1+x) = f(-x)$. This raises the question: can we obtain new Taylor polynomials by substituting into old Taylor polynomials?

More precisely, we are led to the following conjecture:

Conjecture 8.6. *Let $a, b \in \mathbb{R}$ and $m, n \in \mathbb{N}$ be given, let g be a polynomial of degree at most m with $g(b) = a$, and let f be a real function which is n-times differentiable at a. Define $P(x) = T_n f(x; a)$. Then*

$$T_{mn}(f(g(x)); b) = P(g(x))$$

Note that in this conjecture, we would expect $P(g(x))$ to be a Taylor polynomial centered at b, because $g(x) \approx a$ when $x \approx b$ by continuity of g. Also, note that $P(g(x))$ has degree at most mn, hence why we conjecture that it is the Taylor polynomial of order mn.

Unfortunately, this conjecture is false in general. Part of the issue is that the Chain Rule only guarantees that $f \circ g$ is n-times differentiable at b, not mn-times differentiable. However, even when $f \circ g$ is mn-times differentiable, the conjecture fails: see Exercise 8.2.12. However, some special cases of the conjecture are simple to prove, such as the following theorem:

Theorem 8.7. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and let f be a real function which is n-times differentiable at a. For any constant $c \in \mathbb{R}$, we have*

$$T_n(f(x+c); a-c) = T_n f(x; a)$$

(in particular, when $c = a$, we have $T_n(f(x+a); 0) = T_n f(x; a)$), and when $c \neq 0$, we have

$$T_n(f(cx); a/c) = T_n f(cx; a)$$

Therefore, putting these two statements together, we may conclude that Conjecture 8.6 is true when g is a non-constant linear function.

You are asked to prove this theorem in Exercise 8.2.9.

Remark. In particular, Theorem 8.7 justifies substituting $-x$ into the Taylor polynomial for $1/(1-x)$ to obtain the Taylor polynomial for $1/(1+x)$. The theorem also allows us to exclusively focus on Taylor polynomials centered at 0, because $T_n f(x; a) = T_n(f(x+a); 0)$. Thus, in many proofs or computations, we can WLOG assume $a = 0$ to make our work look simpler.

It is possible to modify Conjecture 8.6 to obtain a true statement which is very general, but the proof is quite involved. We will return to this after we have presented some more information about the error of a Taylor approximation: see Exercise 8.5.16.

We can obtain some other plausible conjectures about Taylor polynomials by considering different ways of manipulating polynomials. For instance, what happens when you compute the Taylor polynomial of a polynomial? If p is a polynomial of degree n, then for any $a \in \mathbb{R}$ and any $m \in \mathbb{N}$ with $m \geq n$, it is easy to see by definition that $T_m p(x; a) = p(x)$. When $m < n$, however, $p(x)$ cannot be its own Taylor polynomial of order m. In this case, you can prove in Exercise 8.2.10 that $T_m p(x; a)$ is obtained by taking $p(x)$ and "chopping off" terms of high degree:

Theorem 8.8. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and let P be a polynomial of degree n expressed in the form*

$$P(x) = \sum_{i=0}^{n} p_i(x-a)^i$$

for some constants $p_0, p_1, \ldots, p_n \in \mathbb{R}$ with $p_n \neq 0$. Prove that for all $m \in \mathbb{N}$, if $m < n$ then

$$T_m P(x; a) = \sum_{i=0}^{m} p_i(x-a)^i$$

We sometimes call $T_m P(x; a)$ the truncation of $P(x)$ of order *m.*

For another example of a plausible conjecture, let's consider what happens when you multiply two Taylor polynomials. Suppose we define $P(x) = T_n f(x; a)$ and $Q(x) = T_m g(x; a)$, where $a \in \mathbb{R}$, $n, m \in \mathbb{N}$, and f and g are real functions such that $f^{(n)}(a)$ and $g^{(m)}(a)$ exist. Intuitively, we have $P(x) \approx f(x)$ and $Q(x) \approx g(x)$ when $x \approx a$, so $P(x)Q(x) \approx f(x)g(x)$. This leads to the following conjecture:

Conjecture 8.9. *Let $a \in \mathbb{R}$ and $n, m \in \mathbb{N}$ be given, and suppose f and g are real functions such that $f^{(n)}(a)$ and $g^{(m)}(a)$ exist. If $P(x) = T_n f(x; a)$ and $Q(x) = T_m g(x; a)$, then*

$$T_{n+m}(f(x)g(x); a) = P(x)Q(x)$$

Unfortunately, this conjecture is false in general. For a counterexample, we set $a = 0$, $n = 1$, $f(x) = 1 - x$, $g(x) = 1/(1-x)$, and we let $m \in \mathbb{N}$ be arbitrary. We have $P(x) = T_1 f(x) = f(x)$ (as f is trivially a polynomial of degree at most 1 which agrees with f and its first derivative at 0). Also, we've previously shown

$$Q(x) = T_m g(x) = \sum_{i=0}^{m} x^i = \frac{1 - x^{m+1}}{1 - x}$$

Thus, $P(x)Q(x) = 1 - x^{m+1}$. However, $f(x)g(x) = 1$, so we clearly have $T_{m+1}(f(x)g(x)) = 1$, and $P(x)Q(x) \neq 1$ for any $x \in (0, 1)$.

In this counterexample, note that although $1 - x^{m+1}$ and 1 are not the same, they do match on every term of order at most m. In general, you can prove the following modification of Conjecture 8.9 in Exercise 8.2.14:

Theorem 8.10. *Let a, n, m, f, g, P, Q be given as in Conjecture 8.9. Suppose WLOG that $m \le n$. Then*

$$T_m(f(x)g(x); a) = T_m(P(x)Q(x); a)$$

Thus, by Theorem 8.8, $T_m(f(x)g(x); a)$ is obtained by truncating $P(x)Q(x)$ to order m.

This theorem gives us a way of computing Taylor polynomials of products. Unfortunately, because we truncate, we lose information about orders which are higher than m. It is useful to have some result about Taylor polynomials of products which allows us to increase the order of our Taylor polynomials. One such important result, which we will use in a later section, is presented in this theorem, whose proof is outlined in Exercise 8.2.15:

Theorem 8.11. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and suppose f is a real function which is n-times continuously differentiable at a. Define $g : \mathrm{dom}(f) \to \mathbb{R}$ by $g(x) = (x - a)f(x)$ for all $x \in \mathrm{dom}(f)$. Then g is $(n + 1)$-times differentiable at a, and*

$$T_{n+1}g(x; a) = (x - a)T_n f(x; a)$$

Informally, this says that powers of $x - a$ can be pulled out of a Taylor polynomial.

8.2 Exercises

For Exercises 1 through 7, a value $a \in \mathbb{R}$ and an equation defining a function f are given (the domain of f is considered to be all values of x for which the equation is meaningful). For any $n \in \mathbb{N}$, compute $T_n f(x; a)$ at all points x where the Taylor polynomial exists. You may use previous examples and exercises in your work.

1. $a = 0,\ f(x) = b^x$ where $b > 0$ is a constant
2. $a = \pi,\ f(x) = \sin x$
3. $a = 0,\ f(x) = \dfrac{1}{b - x}$ where $b \in \mathbb{R} - \{0\}$ is a constant
4. $a = 1,\ f(x) = \log x$
5. $a = 0,\ f(x) = \log \sqrt{\dfrac{1 + x}{1 - x}}$
6. $a = 0,\ f(x) = \cos(2x)$
7. $a = 0,\ f(x) = \sin^2 x$ (Hint: Use a half-angle identity.)

8. Prove that for any $x > 0$ and any $n \in \mathbb{N}$ with $n \geq 2$,

$$T_n(\sqrt{x}; 1) = 1 + \frac{1}{2}(x-1) + \sum_{i=2}^{n} \frac{(-1)^{i-1}\,(2i-2)!}{2^{2i-1}\,(i-1)!\,i!}(x-1)^i$$

(Hint: Let $f(x) = \sqrt{x}$ and prove a formula for $f^{(n)}(x)$ by induction.)

9. Prove Theorem 8.7.

10. Prove Theorem 8.8.

11. For any $\alpha \in \mathbb{R}$ and any $k \in \mathbb{N}$, define

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!}$$

(thus, when $\alpha \in \mathbb{N}$, this gives the same value as the binomial coefficient $\binom{n}{k}$ defined in Exercise 4.2.5.(c)).

(a) Prove that for any $n \in \mathbb{N}$,

$$T_n\left((1+x)^\alpha; 0\right) = \sum_{i=0}^{n} \binom{\alpha}{i} x^i$$

(In particular, when $n = 0$, $\binom{\alpha}{0} = 1$ and $T_0((1+x)^\alpha; 0) = 1$.)

(b) Using part (a) and Theorem 8.8, give a proof of the Binomial Theorem, which says that for all $a, b \in \mathbb{R} - \{0\}$ and all $n \in \mathbb{N}^*$,

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

12. This exercise provides a counterexample to Conjecture 8.6 using functions which are infinitely differentiable at $a = 1$. Let $f, g : [0, \infty) \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$ and $g(x) = x^2$ for all $x \geq 0$. Exercise 8.2.8 shows that

$$T_2 f(x; 1) = 1 + \frac{1}{2}(x-1) - \frac{1}{8}(x-1)^2$$

for all $x > 0$ (note that x cannot be 0 since f is not differentiable at 0). Define $P(x)$ to be $T_2 f(x; 1)$. (Thus, using the notation of the conjecture, we have $n = m = 2$ and $a = b = 1$.)

(a) Show that

$$P(g(x)) = 1 + (x-1) - \frac{1}{2}(x-1)^3 - \frac{1}{8}(x-1)^4$$

(Hint: It may be helpful to note that $x^2 - 1 = (x-1)(x+1) = (x-1)((x-1)+2.)$

(b) Compute $T_4(f \circ g)(x; 1)$, and show that it does not equal $P(g(x))$ at any positive value of x except $x = 1$.

13. Prove that if f is an odd function which is n-times differentiable at 0, then $f^{(2k)}(0) = 0$ for all $k \leq n/2$. Thus, $T_n f(x)$ is also an odd function. Similarly, prove that if f is an even function which is n-times differentiable at 0, then $f^{(2k+1)}(0) = 0$ for all $k \leq (n-1)/2$, which shows that $T_n f(x)$ is an even function.

   (You do not have to prove the fact that the derivative of an odd function is even and vice versa, though the proof is not very difficult.)

14. Prove Theorem 8.10. It is probably useful, though not necessary, to use the result of Exercise 4.4.6, which states that

$$(fg)^{(n)}(x) = \sum_{i=0}^{n} \binom{n}{i} f^{(i)}(x) g^{(n-i)}(x)$$

15. This exercise outlines a proof of Theorem 8.11. Let a, n, f, g be given as described. In fact, the remark after Theorem 8.7 allows us to assume WLOG that $a = 0$, so $g(x) = xf(x)$.

   (a) Prove by induction on k that for all k from 1 to n,

$$g^{(k)}(x) = xf^{(k)}(x) + kf^{(k-1)}(x)$$

   for all $x \in \mathrm{dom}(f^{(k)})$. Note that this claim holds for $k = 0$ if we interpret $0f^{(-1)}(x)$ as 0 (even though the notation $f^{(-1)}$ is undefined).

   (b) Part (a) implies that $g^{(k)}(0) = kf^{(k-1)}(0)$ for all k from 0 to n (using the same convention for $k = 0$ as in part (a)). Prove that this formula holds as well for $k = n + 1$, i.e.

$$g^{(n+1)}(0) = (n+1)f^{(n)}(0)$$

Note that you may NOT plug in $k = n + 1$ into part (a) because you may not assume that f is $(n+1)$-times differentiable. Instead, you should use the definition of derivative.

(c) Use the previous parts to prove that $T_{n+1}g(x) = xT_n f(x)$, finishing the proof of Theorem 8.11.

16. Taylor polynomials do not always yield good approximations, even when the function is infinitely differentiable. For instance, in this exercise, we present an example of a nonzero infinitely-differentiable function whose Taylor polynomials are all constantly zero at $a = 0$. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} \exp\left(\frac{-1}{x^2}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

We will prove $f^{(n)}(0) = 0$ for every $n \in \mathbb{N}$. However, we cannot directly compute $f^{(n)}(0)$ with the Chain Rule, because the inner function $-1/x^2$ is not differentiable at 0.

(a) First, prove that for any $a \in \mathbb{N}$,

$$\lim_{x \to 0} x^{-a} \exp(-1/x^2) = 0$$

In particular, this proves that f is continuous at 0.

(b) Prove that for all $n \in \mathbb{N}$, there exists a polynomial P_n such that for every $x \neq 0$,

$$f^{(n)}(x) = P_n(1/x) \cdot \exp(-1/x^2)$$

To clarify, let's say $P_n(t)$ is the polynomial $p_0 + p_1 t + \cdots + p_k t^k$ (k, and the values of the p_i constants, will depend on n). Then $P_n(1/x)$ replaces t with $1/x$.

(c) Use parts (a) and (b) to prove by induction on $n \in \mathbb{N}$ that $f^{(n)}(0) = 0$.

(Hint: Use the definition of derivative in the inductive step.)

8.3 Taylor's Theorem

We have now introduced the notion of a Taylor polynomial, though apart from some wishful thinking, we have not shown why Taylor polynomials tend to produce good approximations. We need some sort of bound on the error, i.e. some sort of guarantee of reliability. For instance, suppose my friend asked me to loan him "about \$200". Without further information, the phrase "about \$200" could mean "somewhere between \$150 and \$250" or it could mean "between \$195 and \$205". I would highly prefer to loan \$205 instead of \$250, so it is important for me to know how accurate my friend's estimate of \$200 is.

In this section, we study the error of a Taylor polynomial approximation. In particular, we will be able to obtain a couple convenient forms for representing that error. From these forms, we will be able to find bounds on the error. These error bounds are what make Taylor polynomials so useful, even for theoretical problems (as you will see in the next section). To start, we introduce the notion of error more formally:

Definition 8.12. Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and let f be a real function which is n-times differentiable at a. For any x where $T_n f(x; a)$ exists, the error $E_n f(x; a)$ is defined as

$$E_n f(x; a) = f(x) - T_n f(x; a)$$

so that $f(x) = T_n f(x; a) + E_n f(x; a)$. As with the notation "$T_n f(x; a)$", we will occasionally write "$E_n(f(x); a)$". We may also drop "$; a$" when a is zero.

Our main goal for this section is to find convenient ways of expressing $E_n f(x; a)$ which will help us produce bounds. In particular, we'll be able to study how the choice of n affects the error bound, so that we can decide in an approximation problem how large to make n for our desired level of accuracy.

To get started, let's consider the simplest case, when $n = 0$. Since $E_0 f(x; a) = f(x) - f(a)$, we know by the FTC that

$$E_0 f(x; a) = f(x) - f(a) = \int_a^x f'(t)\, dt$$

when f is continuously differentiable from a to x. This formula makes sense, because if you approximate f with a constant, then the error of your approximation should depend on how steep f is, i.e. how large $|f'|$ is. From this formula, we note that if f' is bounded from a to x, say $|f'(t)| \le M$ for all t between a and x, then we get

$$|E_0 f(x; a)| \le M|x - a|$$

Next, let's analyze $E_1 f(x;a)$. It is plausible to guess that the error of a linear approximation depends on how curved f is, i.e. how large $|f''|$ is. In fact, in Exercise 4.10.17, you were asked to show that when f'' exists near a, for any x near a there exists some c between a and x satisfying

$$|E_1 f(x;a)| \le |f''(c)||x-a|^2$$

To get a more precise formula for $E_1 f(x;a)$, note that

$$f(x) = f(a) + E_0 f(x;a) = f(a) + \int_a^x f'(t)\,dt$$

and also

$$f(x) = f(a) + f'(a)(x-a) + E_1 f(x;a)$$

by definition of $E_1 f(x;a)$. Since we have the intuition that E_1 should be related to f'', and f'' is the derivative of f', we should rewrite the integral of $f'(t)\,dt$ using integration by parts to try and produce $f'(a)(x-a)$ along the way. Let's set $u = f'(t)$ and $dv = dt$. Now, $du = f''(t)\,dt$, but we can choose v to be any antiderivative of 1, say $v = t + C$ for some constant C which does not depend on t. We get

$$\begin{aligned} f(x) &= f(a) + f'(t)(t+C)\Big|_a^x - \int_a^x f''(t)(t+C)\,dt \\ &= f(a) + f'(x)(x+C) - f'(a)(a+C) - \int_a^x f''(t)(t+C)\,dt \end{aligned}$$

Since we want $f'(a)(x-a)$ to appear in our final expression, we try $C = -x$ (note that $-x$ is a constant with respect to t). This causes the $f'(x)(x+C)$ term to disappear, and we get

$$\begin{aligned} f(x) &= f(a) - f'(a)(a-x) - \int_a^x f''(t)(t-x)\,dt \\ &= f(a) + f'(a)(x-a) + \int_a^x f''(t)(x-t)\,dt \end{aligned}$$

Therefore, we have shown that

$$E_1 f(x;a) = \int_a^x f''(t)(x-t)\,dt$$

To get $E_2 f(x;a)$, you can try another integration by parts, using a similar choice of parts, and you'll find that

$$E_2 f(x;a) = \int_a^x f'''(t) \frac{(x-t)^2}{2}\, dt$$

By doing these steps repeatedly, and noticing a pattern, we are led to the following theorem, which is sometimes called *Taylor's Theorem*[3]:

Theorem 8.13 (Taylor's Theorem). *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and let f be a real function which has a continuous $(n+1)^{\text{st}}$ derivative around a, say $f^{(n+1)}$ is continuous on $(a-\delta, a+\delta)$ for some $\delta > 0$. Then for all $x \in (a-\delta, a+\delta)$, we have*

$$E_n f(x;a) = \int_a^x f^{(n+1)}(t) \frac{(x-t)^n}{n!}\, dt$$

Remark. To remember this formula, note that the last term of $T_n f(x;a)$ is

$$\frac{f^{(n)}(a)}{n!}(x-a)^n$$

The formula for $E_n f(x;a)$ looks nearly the same, but all the a's are replaced with t's, the derivative is one order higher (which makes sense since the 0^{th} error is related to the 1^{st} derivative, the 1^{st} error is related to the 2^{nd} derivative, etc.), and the result is integrated with respect to t from a to x.

It is also worth noting that the result of this theorem does NOT imply

$$\frac{d}{dx} E_n f(x;a) = f^{(n+1)}(x) \frac{(x-x)^n}{n!} = 0$$

This is because if you use Part 1 of the FTC to take the derivative of an integral, then the integrand must not depend on x. However, the integrand in the expression for $E_n f(x;a)$ does depend on x.

Strategy. The idea is to repeatedly perform integration by parts, which we formalize by writing a proof by induction. Suppose that as an inductive hypothesis, we know

$$E_n f(x;a) = \int_a^x f^{(n+1)}(t) \frac{(x-t)^n}{n!}\, dt$$

[3]Sometimes, the name "Taylor's Theorem" is instead used to describe any one of a collection of inequalities we will soon see which are corollaries of this theorem.

Let's choose the parts $u = f^{(n+1)}(t)$ and $dv = (x-t)^n/(n!)\,dt$. We get $du = f^{(n+2)}(t)\,dt$ and $v = -(x-t)^{n+1}/((n+1)!)$ (note that we get a minus sign because v is a function of t, not of x). This turns our integral into

$$\left.\frac{-f^{(n+1)}(t)(x-t)^{n+1}}{(n+1)!}\right|_a^x + \int_a^x f^{(n+2)}(t)\frac{(x-t)^{n+1}}{(n+1)!}\,dt$$

Simplfying this, we obtain the last term of $T_{n+1}f(x;a)$, plus $E_{n+1}f(x;a)$.

Proof. We prove the theorem by induction on $n \in \mathbb{N}$, where a, n, f, and δ are given as described. As a base case, when $n = 0$, the FTC tells us that

$$f(x) = f(a) + \int_a^x f'(t)\,dt$$

whenever f' is continuous from a to x. Since $f(a) = T_0 f(x;a)$, this shows

$$E_0 f(x;a) = \int_a^x f^{(0+1)}(t)\frac{(x-t)^0}{0!}\,dt$$

proving our base case.

Now let $n \in \mathbb{N}$ be given, and assume that our theorem is true for n as an inductive hypothesis. For our inductive step, suppose that $f^{(n+2)}$ is continuous on $(a-\delta, a+\delta)$ for some $\delta > 0$. Let $x \in (a-\delta, a+\delta)$ be given. Because $f^{(n+1)}$ is continuous on $(a-\delta, a+\delta)$, the inductive hypothesis says that

$$f(x) = T_n f(x;a) + \int_a^x f^{(n+1)}(t)\frac{(x-t)^n}{n!}\,dt$$

We use integration by parts on this error integral, with the parts $u = f^{(n+1)}(t)$ and $v = -(x-t)^{n+1}/(n+1)!$, so that $du = f^{(n+2)}(t)\,dt$ and $dv = (x-t)^n/n!\,dt$. This yields

$$\begin{aligned}
f(x) &= T_n f(x;a) - \left.\left(f^{(n+1)}(t)\frac{(x-t)^{n+1}}{(n+1)!}\right)\right|_a^x + \int_a^x f^{(n+2)}(t)\frac{(x-t)^{n+1}}{(n+1)!}\,dt \\
&= T_n f(x;a) + f^{(n+1)}(a)\frac{(x-a)^{n+1}}{(n+1)!} + \int_a^x f^{(n+2)}(t)\frac{(x-t)^{n+1}}{(n+1)!}\,dt \\
&= T_{n+1} f(x;a) + \int_a^x f^{(n+2)}(t)\frac{(x-t)^{n+1}}{(n+1)!}\,dt
\end{aligned}$$

where the last line follows from the definition of $T_{n+1}f(x;a)$. This finishes the inductive step. ■

Taylor's Theorem is very useful because it shows that the error of an n^{th}-order Taylor approximation depends on the $(n+1)^{\text{st}}$ derivative of the function. This means that we can use bounds on the $(n+1)^{\text{st}}$ derivative to get bounds on the error. We demonstrate with some examples.

Example 8.14:

By now, we have seen that for any $n \in \mathbb{N}$ and any $x \in \mathbb{R}$,

$$T_n \exp(x) = \sum_{i=0}^{n} \frac{x^i}{i!}$$

In particular, when $x = 1$, we can use these Taylor polynomials to approximate e^1, which is e. For instance, we've seen that $e \approx T_5 \exp(1) = 163/60 \approx 2.716667$. Now, we can use Taylor's Theorem to estimate the error from this approximation. We get

$$E_5 \exp(1) = \int_0^1 \exp^{(6)}(t) \frac{(1-t)^5}{5!}\, dt$$

Since exp is its own derivative, $\exp^{(6)}(t) = e^t$. Since t goes from 0 to 1, e^t goes from 1 to e. Since we do not yet know a precise value for e (after all, we're trying to approximate it!), for now let's use the simple bounds of $1 \leq e^t \leq 3$. Thus,

$$\begin{aligned}
\int_0^1 \frac{(1-t)^5}{5!}\, dt \leq \quad & E_5 \exp(1) & \leq \int_0^1 3\frac{(1-t)^5}{5!}\, dt \\
-\frac{(1-t)^6}{6!}\bigg|_0^1 \leq \quad & E_5 \exp(1) & \leq -3\frac{(1-t)^6}{6!}\bigg|_0^1 \\
\frac{1}{6!} = \frac{1}{720} \leq \quad & E_5 \exp(1) & \leq \frac{3}{6!} = \frac{1}{240}
\end{aligned}$$

This shows that the error is between 0.001389 and 0.0041667. In particular, the error is definitely positive, so we know that $T_5 \exp(1)$ is an underestimate to the true value of e. Since $e = T_5 \exp(1) + E_5 \exp(1)$, this tells us that

$$2.718056 \approx \frac{163}{60} + \frac{1}{720} \leq e \leq \frac{163}{60} + \frac{1}{240} \approx 2.720833$$

However, these estimates don't even tell us the correct second decimal place of e. Using a larger value of n should give a better estimate for e, but how large of a value should we use? In general, if we use $T_n \exp(1)$ to approximate e, then the same reasoning as above shows that

$$\begin{aligned} \int_0^1 \frac{(1-t)^n}{n!}\,dt \le \quad E_n \exp(1) \quad &\le \int_0^1 3\frac{(1-t)^n}{n!}\,dt \\ \frac{1}{(n+1)!} \le \quad E_n \exp(1) \quad &\le \frac{3}{(n+1)!} \end{aligned}$$

Therefore, if we want the error to be less than some positive number ϵ, we should choose n large enough so that $3/((n+1)!) < \epsilon$, i.e. $(n+1)! > 3/\epsilon$. (Note that since $(n+1)!$ grows very rapidly as n gets large, this means that the Taylor approximation errors approach 0 quite rapidly, so small values of n can still produce very good approximations.)

For instance, suppose that we want at least four decimal places of accuracy. Note that the fourth decimal place can be affected by the fifth, because of rounding. Thus, to guarantee at least four correct decimal places, we want our error to be less than 5 units in the fifth decimal place, so we take $\epsilon = 5 \times 10^{-5}$. Testing several values of n, we find that $9! > 3/\epsilon$, so any value of n satisfying $n+1 \ge 9$, such as $n = 8$, should work. This gives us

$$e \approx T_8 \exp(1) = \sum_{i=0}^{8} \frac{1}{i!} = \frac{109601}{40320} \approx 2.718279$$

with error at most $\frac{3}{9!} \approx 8.267 \times 10^{-6}$. $\square$

Example 8.15:

As a more complicated example, let's show how we can approximate the integral

$$\int_0^1 \exp(t^2)\,dt$$

by using Taylor's Theorem. (We can't expect to compute the integral exactly because $\exp(t^2)$ has no elementary antiderivative.) One advantage of using Taylor's Theorem here, instead of using integral approximation techniques from the end of Chapter 5, is that we can use Taylor's Theorem to obtain a whole sequence of approximations to this integral (one approximation for each value of n). This allows us much more control over the accuracy of our approximations.

First, we aim to approximate $\exp(t^2)$. Rather than computing the Taylor polynomial for $\exp(t^2)$ by taking derivatives (it's quite difficult to identify a pattern for the n^{th} derivative of this), let's reuse our knowledge about exp that we found in the previous example. Suppose we define $f : \mathbb{R} \to \mathbb{R}$ by $f(t) = \exp(t^2)$ for all $t \in \mathbb{R}$. Let's also say that $T(x) = T_n \exp(x)$ and $E(x) = E_n \exp(x)$. Since for any $n \in \mathbb{N}$ and any $x \in \mathbb{R}$, we have

$$\exp(x) = T(x) + E(x)$$

we can plug in $x = t^2$ to obtain

$$f(t) = \exp(t^2) = T(t^2) + E(t^2)$$

This lets us use our bounds on E to obtain bounds for $E(t^2)$. Hence, we can obtain good approximations to $f(t)$ without knowing whether or not $T(t^2)$ is a Taylor polynomial for $f(t)$. (In fact, this method even works if you substitute something which isn't a polynomial into T; for instance, you can also say

$$\exp(\sin t) = T(\sin t) + E(\sin t)$$

with this method, and then you proceed to find bounds on $E(\sin t)$.)

The work from the previous example shows that

$$\frac{(t^2)^{n+1}}{(n+1)!} \leq E_n \exp(t^2) \leq \exp(t^2)\frac{(t^2)^{n+1}}{(n+1)!}$$

(It turns out that for any fixed t, these error bounds approach 0 as $n \to \infty$, though this is not obvious: see Exercise 8.5.14.) Thus,

$$\sum_{i=0}^{n} \frac{t^{2i}}{i!} + \frac{t^{2n+2}}{(n+1)!} \leq \exp(t^2) \leq \sum_{i=0}^{n} \frac{t^{2i}}{i!} + \exp(t^2)\frac{t^{2n+2}}{(n+1)!}$$

We can now use our inequalities for $f(t)$ to get inequalities for the integral:

$$\int_0^1 \sum_{i=0}^{n} \frac{t^{2i}}{i!} + \frac{t^{2n+2}}{(n+1)!}\, dt \leq \int_0^1 \exp(t^2)\, dt \leq \int_0^1 \sum_{i=0}^{n} \frac{t^{2i}}{i!} + \exp(t^2)\frac{t^{2n+2}}{(n+1)!}\, dt$$

The left side of this is the integral of a polynomial, which is easy to compute. For the right side, however, we do not have an elementary antiderivative for

$\exp(t^2)$. Instead, we use the naive inequality $\exp(t^2) \le 3$ for $t \in [0,1]$. This gives us

$$\begin{aligned} & \sum_{i=0}^{n} \frac{1}{i!\,(2i+1)} + \frac{1}{(n+1)!\,(2n+3)} \\ \le \int_0^1 \exp(t^2)\,dt \quad \le \quad & \sum_{i=0}^{n} \frac{1}{i!\,(2i+1)} + \frac{3}{(n+1)!\,(2n+3)} \end{aligned}$$

For instance, when $n = 4$, we get a lower bound of $15203/10395 \approx 1.462530$ and an upper bound of $12175/8316 \approx 1.464045$. Thus, the approximation of 1.46 is correct to two decimal places. □

The Lagrange Form

Occasionally, some other ways of writing the error in a Taylor approximation are more practical. We present one other way, which is called *Lagrange's form of the error*[4]. This form does not use any integrals, instead expressing the error in terms of $f^{(n+1)}$ at some unknown point from a to x.

To see the main idea behind Lagrange's form, recall that

$$E_0 f(x;a) = \int_a^x f'(t)\,dt$$

when f' is continuous from a to x. The Mean Value Theorem for Integrals says that this integral equals $f'(c)(x-a)$ for some value c from a to x. We aim to get a similar-looking formula for all values of n.

In general, we have

$$E_n f(x;a) = \int_a^x f^{(n+1)}(t)\frac{(x-t)^n}{n!}\,dt$$

In order to "pull $f^{(n+1)}$ out of the integral", the Weighted Mean Value Theorem for Integrals (Theorem 5.42) comes in handy. For simplicity, let's first consider when $x \ge a$. The Weighted MVT says that when F and G are two continuous functions on $[a,x]$, and G is nonnegative on $[a,x]$, then there is some $c \in [a,x]$ satisfying

$$\int_a^x F(t)G(t)\,dt = F(c)\int_a^x G(t)\,dt$$

[4]Due to Joseph Louis Lagrange.

Applying this with $F(t) = f^{(n+1)}(t)$ and $G(t) = (x-t)^n/n!$ gives

$$\begin{aligned} E_n f(x;a) &= \int_a^x f^{(n+1)}(t) \frac{(x-t)^n}{n!}\, dt \\ &= f^{(n+1)}(c) \int_a^x \frac{(x-t)^n}{n!}\, dt = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \end{aligned}$$

for some c from a to x (c depends on f, n, a, and x).

Now, let's consider when $x < a$. Here, we write our error as

$$E_n f(x;a) = -\int_x^a f^{(n+1)}(t) \frac{(x-t)^n}{n!}\, dt$$

so that the integral limits are in the correct order. Also, note that for all $t \in (x,a)$, $(x-t)^n$ is positive iff n is even. Thus, when n is even, the same choices of F and G work as above. However, when n is odd, we move the minus sign inside the integral and choose $F(t) = f^{(n+1)}(t)$ and $G(t) = -(x-t)^n/n!$ in the Weighted MVT. This yields

$$\begin{aligned} E_n f(x;a) &= \int_x^a f^{(n+1)}(t) \frac{-(x-t)^n}{n!}\, dt \\ &= f^{(n+1)}(c) \int_x^a \frac{-(x-t)^n}{n!}\, dt = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \end{aligned}$$

Thus, in all cases, we have proven the following:

Theorem 8.16 (Lagrange's Form of the Error). *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and suppose that f is a real function with $f^{(n+1)}$ continuous on some open interval containing a. Then for all x in that interval, there exists some c from a to x such that*

$$E_n f(x;a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

(c depends on n and x, and it could be equal to a or to x).

In this form, the error looks a lot like the term of order $n+1$ in a Taylor polynomial for f, except that the derivative is evaluated at c instead of a. Informally, we can say the error behaves a lot like the "first term left out" of $T_n f(x;a)$.

All of our examples earlier can be analyzed with Lagrange's form instead of with Taylor's Theorem; the work would be mostly the same, but sometimes Lagrange's form makes the work a little cleaner. Lagrange's form is often easier to use when $x < a$, since you don't have to worry about an integral with its limits out of order (though you still have to be careful, since $(x-a)^{n+1}$ will be negative if n is even). For illustration, we present a more theoretical example using Lagrange's form.

Example 8.17:

Define $f(x) = \sqrt{1+x}$ for all $x > -1$. The result of Exercise 8.2.8 shows that

$$T_1(\sqrt{1+x}) = 1 + \frac{x}{2}$$

You can calculate $f''(x) = -(1+x)^{-3/2}/4$, so by Lagrange's form, we have

$$\sqrt{1+x} = 1 + \frac{x}{2} + \frac{f''(c)}{2}x^2 = 1 + \frac{x}{2} - \frac{x^2}{8(1+c)^{3/2}}$$

for some c from 0 to x. In particular, $1+c \geq 0$, so this proves that $\sqrt{1+x} \leq 1 + x/2$.

If we have some further restrictions on x, then we can say even more. For instance, suppose that $x > 0$. Thus, in our formula above, $c \in [0, x]$, so $f''(c)$ is in $[-1/4, -(1+x)^{-3/2}/4]$. This yields

$$1 + \frac{x}{2} - \frac{x^2}{8} \leq \sqrt{1+x} \leq 1 + \frac{x}{2} - \frac{x^2}{8(1+x)^{3/2}}$$

For example, when $x = 1$, we obtain

$$1 + \frac{1}{2} - \frac{1}{8} \leq \sqrt{2} \leq 1 + \frac{1}{2} - \frac{1}{8(2^{3/2})} < 1 + \frac{1}{2} - \frac{1}{24}$$

where we use the simple inequality $2^{3/2} < 3$ (because $2^3 = 8 < 3^2 = 9$). Hence, $\sqrt{2}$ is somewhere from $11/8 = 1.375$ to $35/24 \approx 1.458333$.

This shows how Taylor errors can be used to prove some inequalities. This approach also generalizes rather well: for instance, you can make a similar proof that for all $\alpha \in (0,1)$ and all $x > -1$, $(1+x)^\alpha \leq 1 + \alpha x$. $\square$

We can summarize many of our findings about the Taylor error in this convenient theorem, which you should prove using Lagrange's form and Taylor's Theorem:

Corollary 8.18. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and suppose f is a real function such that $f^{(n+1)}$ is continuous on some open interval $(a-\delta, a+\delta)$ where $\delta > 0$. Suppose that $m, M \in \mathbb{R}$ satisfy $m \le f^{(n+1)}(c) \le M$ for all $c \in (a-\delta, a+\delta)$. Then for all $x \in (a, a+\delta)$, we have*

$$m\frac{(x-a)^{n+1}}{(n+1)!} \le E_n f(x;a) \le M\frac{(x-a)^{n+1}}{(n+1)!}$$

and for all $x \in (a-\delta, a)$, we have

$$m\frac{|x-a|^{n+1}}{(n+1)!} \le (-1)^{n+1} E_n f(x;a) \le M\frac{|x-a|^{n+1}}{(n+1)!}$$

(Thus, the sign of the error can depend on the parity of n when $x < a$.) In particular, if $K \ge 0$ satisfies $|f^{(n+1)}(c)| \le K$ for all $c \in (a-\delta, a+\delta)$, then

$$|E_n f(x;a)| \le K\frac{|x-a|^{n+1}}{(n+1)!}$$

It is worth noting that in this corollary, the bounds m, M, and K depend on n, a, and δ. For this reason, we will sometimes write subscripts of n on these bounds (i.e. writing K_n instead of K). Suppose we fix the value of x and consider what happens as we change n. If K_n grows slowly as n increases, then the Taylor error E_n will approach 0 quickly. However, if K_n increases rapidly, then the error bounds from the corollary become much less useful. Indeed, there are real functions f and numbers $a \in \mathbb{R}$ such that f is infinitely differentiable at a, but the Taylor polynomial $T_n f(x;a)$ only approaches $f(x)$ for x in some small open interval containing a. See, for instance, Exercises 8.5.8 and Exercises 8.5.20 through 8.5.22.

8.4 More About Order Of Error

In the recently mentioned corollary, we analyzed how $E_n f(x;a)$ depends on n when x is held fixed. In this section, we will consider the converse problem: how does $E_n f(x;a)$ depend on x when n is fixed? Lagrange's form of the error is particularly clarifying here: we have

$$E_n f(x;a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

for some c from a to x. This equation tells us that when n is fixed, the error looks a lot like $(x-a)^{n+1}$.

Back in Chapter 4, we studied the derivative by analyzing it in terms of linear approximations. We established that a linear approximation using $f'(a)$ as its slope has a "better than linear" error, in the sense that $E_1 f(x;a)/(x-a) \to 0$ as $x \to a$. For an arbitrary order n, our work above with Lagrange's form suggests the following statement about "better than n^{th}-order" errors:

Theorem 8.19. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and suppose f is a real function for which $f^{(n+1)}$ exists and is continuous around a. Then the n^{th}-order error $E_n f(x;a)$ goes to 0 "faster than n^{th} order" as x approaches a, in the sense that*

$$\lim_{x \to a} \frac{E_n f(x;a)}{(x-a)^n} = 0$$

Remark. Note that if a function $g(x)$ approaches 0 faster than n^{th} order as $x \to a$, then it also approaches faster than $(n-1)^{\text{st}}$ order. This is because

$$\lim_{x \to a} \frac{g(x)}{(x-a)^{n-1}} = \lim_{x \to a} (x-a) \frac{g(x)}{(x-a)^n} = 0 \cdot 0 = 0$$

Intuitively, this means that higher orders of error tell us more information than low orders of error. We will return to this idea more rigorously near the end of this chapter.

Strategy. Our work before this theorem suggests that Lagrange's form of the error is good to use here: it tells us that

$$\frac{E_n f(x;a)}{(x-a)^n} = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)$$

when x is near a (but not equal to a) and c is from a to x. Certainly $x-a \to 0$ as $x \to a$, and $(n+1)!$ is a constant with respect to x.

However, what happens to $f^{(n+1)}(c)$ as $x \to a$? Since c is from a to x, and $f^{(n+1)}$ is continuous at a, we would expect that $c \to a$ and hence $f^{(n+1)}(c) \to f^{(n+1)}(a)$. We can use this idea more formally in a couple ways. The way we will use is to treat c as a function of x and use the Squeeze Theorem with it, followed by using the Composition Limit Theorem to analyze $f^{(n+1)}(c)$ (where c is the inner function). The other way (which you can explore in

Exercise 8.5.10) is to use the fact that $f^{(n+1)}(x)$ is close to $f^{(n+1)}(a)$ when x is close to a, so we can find a constant K such that $|f^{(n+1)}(x)| \le K$ whenever x is close enough to a. This constant can be used with Corollary 8.18 to simplify our limit calculation.

Proof. Let a, n, f be given as described. By Lagrange's form of the error, for each x near a but not equal to a, we may choose a value $c(x)$ from a to x satisfying

$$\frac{E_n f(x;a)}{(x-a)^n} = \frac{f^{(n+1)}(c(x))}{(n+1)!}(x-a)$$

Now, when $x > a$, we have $a \le c(x) \le x$, so the Squeeze Theorem shows that $c(x) \to a$ as $x \to a^+$. Similarly, when $x < a$, we have $x \le c(x) \le a$, so $c(x) \to a$ as $x \to a^-$.

Therefore, $c(x) \to a$ as $x \to a$. Because $f^{(n+1)}$ is continuous at a, we may use the Composition Limit Theorem to deduce

$$\lim_{x \to a} \frac{f^{(n+1)}(c(x))}{(n+1)!}(x-a) = \frac{f^{(n+1)}(a)}{(n+1)!} \cdot 0 = 0$$

as desired. ∎

Motivated by "better than linear" errors of linear approximations (as mentioned in Lemma 4.11), we have shown that Taylor polynomials of order n produce "better than n^{th}-order" errors as $x \to a$. Now, Lemma 4.11 actually says something stronger about linear approximations: it says that the ONLY linear approximation to $f(x)$ at a with a "better than linear" error is $T_1 f(x;a)$. We aim to prove that a similar statement holds for n^{th}-order Taylor polynomials. More precisely, we obtain the following conjecture:

Conjecture 8.20. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and suppose f is a real function which is n-times differentiable at a. Suppose that P and E are real functions satisfying*

$$f(x) = P(x) + E(x)$$

for all x near a, where P is a polynomial of degree at most n and E converges faster than n^{th}-order to 0 as $x \to a$. (In other words, $E(x)/(x-a)^n \to 0$ as $x \to a$.) Then $P(x) = T_n f(x;a)$.

We will soon see a proof of this conjecture which requires an extra assumption: $E(x)/(x-a)^n$ has n derivatives at a.[5] With this conjecture, we will be able to compute some Taylor polynomials very efficiently. For example, if we let $f(x) = 1/(1-x)$, then

$$f(x) = 1 + x + \cdots + x^n + \frac{x^{n+1}}{1-x}$$

Setting $E(x) = x^{n+1}/(1-x)$, it is easy to see that $E(x)/x^n \to 0$ as $x \to 0$ and that $E(x)/x^n$ has n derivatives at 0. Thus, our conjecture shows that $1 + x + \cdots + x^n$ is $T_n f(x)$. This is a much quicker derivation of $T_n f(x)$ than the derivation in Example 8.5.

In order to prove Conjecture 8.20, we first introduce a helpful condition for establishing that a polynomial is a Taylor polynomial for f. Let's say $f = P + E$ as above. In order for $P(x)$ to be $T_n f(x;a)$, we must have $P^{(k)}(a) = f^{(k)}(a)$ for all k from 0 to n, which is equivalent to saying $E^{(k)}(a) = 0$. This leads to the following theorem (you should fill in the few remaining details):

Theorem 8.21. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and let f, P, and E be real functions such that f and E are n-times differentiable at a and P is a polynomial. Suppose that for all $x \in \mathrm{dom}(E)$,*

$$f(x) = P(x) + E(x)$$

If $E^{(k)}(a) = 0$ for all k from 0 to n (equivalently, if $T_n E(x;a)$ is the zero polynomial), then $T_n f(x;a) = T_n P(x;a)$. In particular, by Theorem 8.8, if $\deg(P) \le n$, then $P(x) = T_n f(x;a)$ and $E(x) = E_n f(x;a)$.

Theorem 8.21 is useful because certain common types of functions E occurring in practice can be shown to satisfy $T_n E(x;a) = 0$ for all k from 0 to n. Our main goal is to prove that this condition is satisfied when E converges to 0 faster than n^{th}-order. (You can see another type of function E with $T_n E(x;a) = 0$ in Exercise 8.5.16.) To make our work a little simpler, let's introduce a new function $g : \mathrm{dom}(E) \cup \{a\} \to \mathbb{R}$ as follows:

$$g(x) = \begin{cases} \frac{E(x)}{(x-a)^n} & \text{if } x \neq a \\ 0 & \text{if } x = a \end{cases}$$

[5]The author is currently unaware whether there is a proof of the conjecture which does not use an assumption like this.

Note that $g(a)$ is defined to be 0 so that g is continuous at a (because E is better than n^{th}-order). Also, we assume $g^{(n)}(a)$ exists.

We may write
$$E(x) = (x-a)^n g(x)$$
for all $x \in \mathrm{dom}(g)$. This form of writing E should remind you of Conjecture 8.9 concerning products of Taylor polynomials. While that conjecture is not true in general, in that discussion, we mentioned a result for pulling powers of $x-a$ out of a Taylor polynomial: Theorem 8.11. That theorem is the key to the following result:

Theorem 8.22. *Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ be given, and suppose that g is a real function which is n-times differentiable at a with $g(a) = 0$ (when $n = 0$, we assume g is continuous at a). Then the function $E : \mathrm{dom}(g) \to \mathbb{R}$ defined for all $x \in \mathrm{dom}(g)$ by*
$$E(x) = (x-a)^n g(x)$$
satisfies $T_n E(x; a) = 0$, or equivalently $E^{(k)}(a) = 0$ for all k from 0 to n.

Remark. This theorem, together with Theorem 8.21, finishes the proof of Conjecture 8.20.

Strategy. Theorem 8.11 tells us that
$$T_n(x-a)^n g(x) = (x-a)T_{n-1}(x-a)^{n-1}g(x)$$
By applying the theorem n times in total, we finally obtain
$$T_n(x-a)^n g(x) = (x-a)^n T_0 g(x)$$
Lastly, since g is continuous at a, $T_0 g(x) = g(a) = 0$.

Proof. Let a, g be given as described. The proof is by induction on n, where n and E are given as described. The base case of $n = 0$ is clear, because $T_0(x-a)^0 g(x) = T_0 g(x) = g(a) = 0$ as g is continuous at a.

Now let $n \in \mathbb{N}$ be given, and suppose that $T_n(x-a)^n g(x) = 0$ as an inductive hypothesis. For the inductive step, we want to compute $T_{n+1}(x-a)^{n+1}g(x)$. Since $g^{(n+1)}(a)$ exists, the function $(x-a)^n g(x)$ has n continuous derivatives at a. By Theorem 8.11, and using the inductive hypothesis,
$$T_{n+1}(x-a)^{n+1}g(x) = (x-a)T_n(x-a)^n g(x) = (x-a) \cdot 0 = 0$$

as desired. ■

We now present a couple examples demonstrating the power of Conjecture 8.20 for finding some new Taylor polynomials.

Example 8.23:

Recall the following equation from Example 8.5:

$$\frac{1}{1-x} = \sum_{i=0}^{n} x^i + \frac{x^{n+1}}{1-x}$$

After the statement of Conjecture 8.20, we saw how the conjecture proved that this formula allowed us to readily compute $T_n(1/(1-x))$. We can also use this formula, and a little algebra, to compute some other Taylor polynomials.

For instance, if we multiply throughout by x, we get

$$\frac{x}{1-x} = \left(x + x^2 + \cdots + x^{n+1}\right) + \frac{x^{n+2}}{1-x} = P(x) + E(x)$$

where $P(x) = x + x^2 + \cdots + x^{n+1}$ and $E(x) = x^{n+2}/(1-x)$. If we let $g(x) = E(x)/x^{n+1} = x/(1-x)$, then g has $n+1$ derivatives at 0 and $g(0) = 0$. Therefore, Conjecture 8.20 proves that

$$T_{n+1}\left(\frac{x}{1-x}\right) = P(x) = x + x^2 + \cdots + x^{n+1}$$

with very little effort. (Note that there is another easy way to get this Taylor polynomial: we can write $x/(1-x) = 1/(1-x) - 1$ and use linearity of Taylor polynomials.)

Similarly, if we replace x with $-x^2$ in our equation for $1/(1-x)$, we get

$$\frac{1}{1+x^2} = \sum_{i=0}^{n} (-1)^i x^{2i} + \frac{(-1)^{n+1} x^{2n+2}}{1+x^2}$$

Here, $E(x) = (-1)^{n+1}x^{2n+2}/(1+x^2)$ and $g(x) = E(x)/x^{2n+1} = (-1)^{n+1}x/(1+x^2)$. As g has $2n+1$ derivatives at 0 with $g(0) = 0$, we get

$$T_{2n+1}\left(\frac{1}{1+x^2}\right) = \sum_{i=0}^{n} (-1)^i x^{2i}$$

(note that this is also the Taylor polynomial of order $2n$, since the $(2n+1)$st Taylor coefficient is 0, so it's like we gained an order of approximation for free, much like we did with the Taylor polynomials for sin and cos). By integrating, Theorem 8.4 tells us

$$T_{2n+2} \arctan(x) = \sum_{i=0}^{n} (-1)^i \frac{x^{2i+1}}{2i+1}$$

□

We can even use Conjecture 8.20 to obtain errors of Taylor polynomials:

Example 8.24:

In the previous example, we showed that

$$\frac{1}{1+t^2} = P(t) + E(t)$$

where $P(t) = \sum_{i=0}^{n} (-1)^i t^{2i}$ and $E(t) = (-1)^{n+1} \frac{t^{2n+2}}{1+t^2}$. Not only did we learn that $P(t) = T_{2n+1}(1/(1+t^2))$, but we also learned from Conjecture 8.20 that $E(t) = E_{2n+1}(1/(1+t^2))$. Therefore, by integrating from 0 to x, we obtain

$$\arctan x = T_{2n+2} \arctan(x) + E_{2n+2} \arctan(x)$$

with

$$T_{2n+2} \arctan(x) = \int_0^x P(t)\, dt = \sum_{i=0}^{n} (-1)^i \frac{x^{2i+1}}{2i+1}$$

and

$$E_{2n+2} \arctan(x) = \int_0^x E(t)\, dt = \int_0^x (-1)^{n+1} \frac{t^{2n+2}}{1+t^2}\, dt$$

We can try to compute this error integral by partial fractions, but when we perform long division, we get

$$\frac{t^{2n+2}}{1+t^2} = \sum_{i=0}^{n} (-1)^{n-i} t^{2i} + \frac{(-1)^{n+1}}{1+t^2}$$

so we end up recreating $T_{2n+2} \arctan$ after integrating. Instead, we'll use a simple inequality to make our work easier. Since the denominator is the part of the fraction which makes the integral difficult, we'll use the naive inequality $1+t^2 \geq 1$ to get $t^{2n+2}/(1+t^2) \leq t^{2n+2}$. Now, we may be tempted to write

$$E_{2n+2}\arctan(x) \leq \int_0^x (-1)^{n+1} t^{2n+2}\,dt = (-1)^{n+1}\frac{x^{2n+3}}{2n+3}$$

but this has a couple flaws. First, if x is negative, then the integral limits are out of order, so we need to introduce another minus sign to get the limits in the correct order before using the Comparison Property for Integrals. Second, if n is even, then multiplying by $(-1)^{n+1}$ reverses any inequality we are making. However, you can see that in all cases, we still get a bound on the magnitude of the error:

$$|E_{2n+2}\arctan(x)| \leq \frac{|x|^{2n+3}}{2n+3}$$

This error bound is very convenient for approximations. For instance, since $\arctan 1$ is $\pi/4$, we can approximate π as

$$\pi \approx 4T_{2n+1}\arctan(1) = 4\left(1 - \frac{1}{3} + \frac{1}{5} + \cdots + \frac{(-1)^n}{2n+1}\right)$$

for any $n \in \mathbb{N}$, with

$$|E_{2n+1}\arctan(1)| \leq \frac{4}{2n+3}$$

Thus, if we use $n = 10$, then we get $\pi \approx 3.232316$, where the error has magnitude at most $4(1/23) \approx 0.173913$. Thus, π is between 3.058403 and 3.232316.

In general, if we use an arbitrary n and $x = 1$, then our error is at most $4/(2n+3)$, which can be made arbitrarily small by choosing n large enough. However, this error estimate goes to 0 much more slowly than the error estimates from other examples, since those earlier estimates had factorials in their denominators. Essentially, this happens because the derivatives of exp at 0 are all equal to 1, but the derivatives of $\arctan x$ at 0 are proportional to $n!/(2n+1)$ (i.e. they grow large very quickly).

In fact, when $|x| > 1$, because exponentials dominate polynomials, the estimate $|x|^{2n+3}/(2n+3)$ goes to ∞ as $n \to \infty$! This shows that the Taylor polynomials for arctan are not practical for approximating at values of x with $|x| > 1$. When $|x| = 1$, the error estimate goes to 0 slowly as $n \to$

∞. When $|x| < 1$, the error behaves a lot like $|x|^{2n+3}$ and so it goes to 0 rapidly. This realization allows us to use trigonometric identities to make other approximation schemes to approximate π much more effectively; for instance, see Exercise 8.5.18. $\square$

8.5 Exercises

In Exercises 1 through 7, there are two parts. In part (a), prove the error bound for all $n \in \mathbb{N}$ and all specified x. In part (b), use the error bound and the Taylor polynomials found in previous examples to compute an approximation within the specified accuracy (i.e. with error smaller than the number specified). You may use any of the techniques discussed in this section, such as Taylor's Theorem, Lagrange's form of the error, or Corollary 8.18. You may also use the results of previous exercises.

1. (a) $|E_{2n}\sin(x)| \le \dfrac{|x|^{2n+1}}{(2n+1)!}$ for all $x \in \mathbb{R}$
   (b) Approximate $\sin(0.5)$ to within $1/1000$.
   (Suggestion: Do not solve for n; just check several values to see what error bounds they produce.)

2. (a) $|E_{2n+1}\cos(x)| \le \dfrac{|x|^{2n+2}}{(2n+2)!}$ for all $x \in \mathbb{R}$
   (b) Approximate $\cos(2)$ to within $1/10$.

3. (a) $\left|E_n\left(\dfrac{1}{1-x}\right)\right| < |x|^{n+1}$ for all $x < 0$
   (b) Use a Taylor approximation to approximate $1/1.1$ to within $1/100$. How does the error bound compare with the actual error?

4. (a) $0 < (-1)^{n+1}E_n\left(\dfrac{1}{1+x}\right) < x^{n+1}$ for all $x > 0$
   (b) Use a Taylor approximation to approximate $1/1.8$ to within $1/3$. How does the error bound compare with the actual error?

5. (a) $0 < (-1)^{n+1}E_{n+1}(\log(1+x)) < \dfrac{x^{n+2}}{n+2}$ for all $x > 0$
   (Hint: Use the previous problem.)

(b) Approximate $\log 2$ to within $1/10$. In your approximation, is the error negative or positive?

6. (a) $\left|E_{2n+1}\left(\frac{x}{1-x^2}\right)\right| \le \frac{4}{3}|x|^{2n+3}$ for all $x \in (-0.5, 0.5)$

   (b) Use a Taylor approximation to approximate $0.25/(1-0.25^2)$ to within $1/100$. How does the error bound compare with the actual error?

7. (a) $\left|E_{2n+2}\left(\log(1-x^2)\right)\right| \le \frac{8}{3} \cdot \frac{x^{2n+4}}{2n+4}$ for all $x \in [0, 0.5)$

   (Hint: Use the previous problem.)

   (b) Approximate $\log(15/16)$ to within $1/10000$.

8. Define $f : (0, \infty) \to \mathbb{R}$ by $f(x) = 1/(1+x)$ for all $x > 0$. Prove that for all $n \in \mathbb{N}$, $T_n f(1)$ is either 0 or 1. Thus, the error of the Taylor approximations does not approach 0.

9. (a) Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = (\sin x)/x$ if $x \neq 0$ and $f(0) = 1$ (thus, f is continuous at 0). Prove that for all $n \in \mathbb{N}$, there is a function $E : \mathbb{R} \to \mathbb{R}$ such that for all $x \in \mathbb{R}$,

$$f(x) = \sum_{i=0}^{n} (-1)^i \frac{x^{2i}}{(2i+1)!} + E(x)$$

   and $|E(x)| \le \frac{|x|^{2n+2}}{(2n+3)!}$.

   (Hint: Define $E(x)$ in terms of $E_{2n+2}\sin(x)$ when $x \neq 0$, and then figure out the proper value for $E(0)$.)

   (b) Use part (a) to approximate the value of $\int_0^1 f(t)\,dt$ with four decimal places of accuracy.

10. Provide an alternate proof of Theorem 8.19 using the second approach outlined in the strategy. More specifically, prove that $f^{(n+1)}$ is bounded by some constant K for points near a, then use Corollary 8.18, and lastly use that result to prove that $E_n f(x; a)/(x-a)^n \to 0$ as $x \to a$.

11. We have previously seen in Exercise 7.2.5 that for all $x > 1$, $\log x < x - 1$. Prove that for all $x \in (1, 2)$,

$$\log x > x - 1 - \frac{(x-1)^2}{2}$$

(Hint: Write x as $1 + u$ and consider $E_2 \log(1 + u)$.)

12. Use a Taylor approximation to prove that

$$\cos x \geq 1 - \frac{x^2}{2} + \frac{x^4}{48}$$

for all $x \in (-\pi/3, \pi/3)$. (Hint: Use a 3$^{\text{rd}}$-order approximation and Lagrange's form.)

13. Use a Taylor approximation to prove that $|\sin x| \leq |x|$ for all $x \in \mathbb{R}$. (Hint: When $|x| \leq 1$, use Lagrange's form of the error and consider cases based on the sign of x. There are two possible choices to use for the order of the Taylor approximation, and each one works for the proof, though the arguments are a little different for each order.)

14. Prove that for all $b > 1$, $b^n/(n!) \to 0$ as $n \to \infty$. We sometimes describe this result as "factorials dominate exponentials." (Hint: When $n > 2b$, show that the ratio between $b^{n+1}/(n+1)!$ and $b^n/n!$ is less than $1/2$.)

15. Use the previous exercise to prove that for any $x \in \mathbb{R}$,

$$E_n \exp(x) \to 0 \text{ as } n \to \infty$$

Thus, we have $T_n \exp(x) \to e^x$ for any $x \in \mathbb{R}$, so Taylor polynomials produce arbitrarily good approximations for the exponential function.

16. In this exercise, we outline a proof of a general "substitution rule" for Taylor polynomials, i.e. a correction to Conjecture 8.6:

Theorem 8.25. *Let $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$ be given. Suppose that f is a real function which is n-times differentiable at a, and g is a polynomial*[6]

[6] Although the statement of Theorem 8.25 requires g to be a polynomial, the result actually holds for any function g which is n-times differentiable at b. The proof is similar, but part (a)'s result needs to be structured in a different way which is much harder to express, let alone prove by induction.

with $g(b) = a$. By the Chain Rule, we know that $f \circ g$ is n-times differentiable at b. Define $P : \mathbb{R} \to \mathbb{R}$ by $P(x) = T_n f(x; a)$ for all $x \in \mathbb{R}$. Then we have

$$T_n(f \circ g)(x; b) = T_n(P \circ g)(x; b)$$

In other words, by the result of Theorem 8.8, to get the n^{th}-order Taylor polynomial for $f \circ g$ at b, we write $P(g(x))$ in terms of powers of $(x - b)$ and then truncate to order n.

(Note that by the result of Exercise 8.2.12, we cannot guarantee that $P(g(x))$ yields a Taylor polynomial for $f \circ g$ at b with a higher order than n.)

(a) First, define $P(x) = T_n f(x; a)$ and $E(x) = E_n f(x; a)$. Thus,

$$f(x) = P(x) + E(x)$$

As a result, if we define $H(x) = E(g(x))$, then we have

$$f(g(x)) = P(g(x)) + H(x)$$

Prove by induction on k that for all k from 0 to n, $H^{(k)}$ has the form

$$H^{(k)}(x) = \sum_{i=0}^{k} p_{k,i}(x) E^{(i)}(g(x))$$

where $p_{k,i}$ is a polynomial for each i from 0 to k.

(b) Use part (a) to prove that $H^{(k)}(b) = 0$ for all k from 0 to n. Thus, by Theorem 8.21, we have $T_n(f \circ g)(x; b) = T_n(P \circ g)(x; b)$.

17. Use Theorem 8.25 from the previous exercise to compute the following Taylor polynomials: for all $n, m \in \mathbb{N}^*$,

$$T_{mn}(\exp(x^m)) = \sum_{i=0}^{n} \frac{x^{mi}}{i!}$$

$$T_{m(2n+1)}(\sin(x^m)) = \sum_{i=0}^{n} (-1)^i \frac{x^{m(2i+1)}}{(2i+1)!}$$

$$T_{2mn}(\cos(x^m)) = \sum_{i=0}^{n} (-1)^i \frac{x^{2mi}}{(2i)!}$$

18. Previously, we computed approximations to π by using Taylor polynomials for $\arctan 1 = \pi/4$. This method produced large errors of approximation. To get an improved method, we will instead express π in terms of values of the form $\arctan x$ with $|x| < 1$. This is because when $|x| < 1$, the error bound of $|x|^{2n+3}/(2n+3)$ we found with arctan converges to 0 rather quickly.

    The heart of our method is the following identity of John Machin, which we will prove:

    $$\pi = 16 \arctan\left(\frac{1}{5}\right) - 4 \arctan\left(\frac{1}{239}\right)$$

    (a) Prove the trigonometric identity

    $$\tan(x+y) = \frac{\tan x + \tan y}{1 - \tan x \tan y}$$

    for all $x, y \in \mathbb{R}$ for which $\tan x$, $\tan y$, and $\tan(x+y)$ are defined.

    (b) Now, let $\alpha = \arctan(1/5)$ and $\beta = 4\alpha - \pi/4$. (α is chosen this way because $\pi/4$ is roughly $0.8 = 4/5 \approx 4\alpha$.) Use the identity from part (a) with $x = y = \alpha$, then with $x = y = 2\alpha$, and lastly with $x = 4\alpha$ and $y = -\pi/4$, to obtain $\tan(2\alpha) = 5/12$, $\tan(4\alpha) = 120/119$, and $\tan\beta = 1/239$. From this, it follows that

    $$\pi = 16\alpha - 4\beta = 16 \arctan\left(\frac{1}{5}\right) - 4 \arctan\left(\frac{1}{239}\right)$$

    proving Machin's identity.

    (c) Obtain an approximation for $16 \arctan(1/5)$ which has error magnitude less than 5×10^{-8}.

    (d) Obtain an approximation for $4 \arctan(1/239)$ which has error magnitude less than 5×10^{-8}. (Note that since $1/239$ is much closer to 0 than $1/5$ is, you will be able to use a much smaller order for your Taylor polynomial!)

    Using the results of parts (b), (c), and (d), you should be able to obtain the approximation $\pi \approx 3.141592682$, with error magnitude less than 10^{-7}.

19. This exercise outlines an alternate proof of Lagrange's form which has slightly weaker hypotheses. This proof is useful for a couple reasons. First, it will also give us a way to obtain other forms of the Taylor error. Second, this proof can be generalized in multivariable calculus to obtain a version of Taylor's Theorem in multiple variables.

    Suppose that $a \in \mathbb{R}$ and $n \in \mathbb{N}$ are given, and f is a real function which is $(n+1)$-times differentiable on an interval $[a-\delta, a+\delta]$ (where $\delta > 0$). However, we do not require $f^{(n+1)}$ to be continuous on $[a-\delta, a+\delta]$. Let $x \in [a-\delta, a+\delta]$ be given, and define $F : [a-\delta, a+\delta] \to \mathbb{R}$ by

    $$F(t) = \sum_{i=0}^{n} \frac{f^{(i)}(t)}{i!}(x-t)^i$$

    for all $t \in [a-\delta, a+\delta]$.

    (a) Show that $F(x) = f(x)$ and $F(a) = T_n f(x; a)$. Thus, $F(x) - F(a) = E_n f(x; a)$. (Intuitively, $F(t)$ provides a "smooth transition" from $T_n f(x; a)$ to $f(x)$ as t goes from a to x.)

    (b) Prove that F is differentiable from a to x, and

    $$\frac{dF}{dt} = \frac{(x-t)^n}{n!} f^{(n+1)}(t)$$

    (c) When G is any continuously differentiable function from a to x, satisfying $G'(c) \neq 0$ for all c between a and x, Cauchy's Mean Value Theorem implies

    $$G'(c)(F(x) - F(a)) = F'(c)(G(x) - G(a))$$

    for some c between a and x. By using this together with part (b) and the choice of $G(t) = (x-t)^{n+1}$, prove the Lagrange form of the error:

    $$E_n f(x; a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

    (Note that other forms of the Taylor error can be obtained with different choices for G.)

In the next three exercises, we analyze two different schemes for approximating logarithms. Our first scheme only produces arbitrarily good approximations of $\log y$ when $y \in (0, 2]$ (i.e. for any $y \in (0, 2]$, by taking larger and larger orders of approximation, we can make the error as small as desired). However, our second scheme produces arbitrarily good approximations of $\log y$ for any $y \geq 1$ (note that values of y in $(0, 1)$ can be handled by the identity $\log y = -\log(1/y)$).

20. For our first scheme, define $f : (-1, \infty) \to \mathbb{R}$ by $f(x) = \log(1 + x)$ for all $x > -1$.

    (a) By doing steps like those in Example 8.23, prove that for all $x > -1$ and all $n \in \mathbb{N}^*$,

    $$T_{n-1}f'(x) = T_{n-1}\left(\frac{1}{1+x}\right) = \sum_{i=0}^{n-1}(-1)^i x^i$$

    and

    $$E_{n-1}f'(x) = E_{n-1}\left(\frac{1}{1+x}\right) = \frac{(-1)^n x^n}{1+x}$$

    Therefore, by integrating, we obtain

    $$T_n f(x) = \int_0^x T_{n-1}\left(\frac{1}{1+t}\right)\,dt = \sum_{i=1}^{n}(-1)^{i-1}\frac{x^i}{i}$$

    and

    $$E_n f(x) = \int_0^x E_{n-1}\left(\frac{1}{1+t}\right)\,dt = \int_0^x \frac{(-1)^n t^n}{1+t}\,dt$$

    (b) Prove that for all $n \in \mathbb{N}^*$, we have the following: if $x \geq 0$ then

    $$\frac{1}{1+x}\cdot\frac{x^{n+1}}{n+1} \leq |E_n f(x)| \leq \frac{x^{n+1}}{n+1}$$

    and if $-1 < x < 0$, then

    $$\frac{|x|^{n+1}}{n+1} \leq |E_n f(x)| \leq \frac{1}{1+x}\cdot\frac{|x|^{n+1}}{n+1}$$

    (c) Use part (b) to conclude that when $x \in (-1, 1]$, $E_n f(x) \to 0$ as $n \to \infty$. Thus, we have $T_n f(x) \to \log(1 + x)$ as $n \to \infty$, so by making n as large as desired, we can obtain arbitrarily precise approximations to $\log(1 + x)$.

(d) To approximate $\log y$ where $y \in (0, 2]$, we can use $T_n f(x)$ with $x = y - 1$. Using this idea, find the value of $\log 1.5$ correct to three decimal places.

21. Let f be the same function as in Exercise 8.5.20. The goal of this exercise is to show that $T_n f(x)$ does not do a good job at approximating $\log(1 + x)$ for any $x > 1$. More precisely, let $x > 1$ be given, and we want to show $T_n f(x) \not\to \log x$ as $n \to \infty$. In fact, we will show that $T_n f(x)$ has no limit as $n \to \infty$!

   (a) Suppose that, for contradiction, there exists some $L \in \mathbb{R}$ so that $T_n f(x) \to L$ as $n \to \infty$. Prove that $(-1)^{n-1}x^n/n \to 0$ as $n \to \infty$. (Hint: $(-1)^{n-1}x^n/n = T_n f(x) - T_{n-1} f(x)$.)

   (b) From Exercise 7.4.27, we know that for any $y > 0$, $e^y > 1+y$. Use this to prove that
   $$\frac{x^n}{n} > \log x$$
   for any $n \in \mathbb{N}^*$.

   (c) Use parts (a) and (b) to show that $T_n f(x)$ cannot approach a limit as $n \to \infty$. Thus. our scheme from Exercise 8.5.20 only produces good approximations for $\log(1+x)$ when $x \le 1$ (i.e. for $\log y$ when $y \le 2$.)

22. In this problem, we introduce a different approximation scheme that avoids the problem demonstrated in Exercise 8.5.21. We define $g : (-1, 1) \to \mathbb{R}$ by
$$g(x) = \log\left(\frac{1+x}{1-x}\right)$$
for all $x \in (-1, 1)$. With this scheme, we approximate $\log y$ for any $y \ge 1$ by first writing y in the form $(1 + x)/(1 - x)$ for some $x \in [0, 1)$ and then using Taylor polynomials for $g(x)$. (Note: We will not need to use negative values of x for our approximations, but g needs to be defined on both sides of 0 in order for $T_n g(x; 0)$ to exist.)

   (a) Prove that for every $y \ge 1$, there exists a unique $x \in [0, 1)$ with $(1 + x)/(1 - x) = y$. It follows that the function which takes $x \in [0, 1)$ to $(1 + x)/(1 - x)$ is a bijection from $[0, 1)$ to $[1, \infty)$.

(b) Prove that for all $x \in [0,1)$, if f is defined as in Exercise 8.5.20, then $g(x) = f(x) - f(-x)$. Use this to show that for all $n \in \mathbb{N}^*$,

$$T_{2n}g(x) = 2\left(\sum_{i=0}^{n-1} \frac{x^{2i+1}}{2i+1}\right)$$

and

$$|E_{2n}g(x)| \le \frac{2-x}{1-x} \cdot \frac{x^{2n+1}}{2n+1}$$

(c) Use part (b) to prove that for any $x \in [0,1)$, $E_{2n}g(x) \to 0$ as $n \to \infty$. Thus, using the result in part (a), we can use Taylor polynomials for $\log((1+x)/(1-x))$ to obtain arbitrarily good approximations for $\log y$ for any $y \ge 1$.

(d) Using the previous parts of this problem, approximate $\log 3$ correct to three decimal places.

8.6 Some Applications of Taylor Polynomials

Taylor polynomials do more than just give us numerical approximations for values like e or π. They give us ways to replace complicated expressions with simple expressions, as long as we permit a little error. For instance, let's say you are investing one dollar for n years at a rate of interest x, where $x > 0$ and $n \in \mathbb{N}^*$. After n years, you will have $(1+x)^n$ dollars. If x is close to 0, then the following approximation is pretty accurate:

$$(1+x)^n \approx T_1(1+x)^n = 1 + nx$$

Note that $1 + nx$ is the amount of money you would have if you received x dollars every year for n years, starting with one dollar. Thus, when x is small, there is little difference between accumulating interest and merely receiving a lump sum each year. In fact, when $x > 0$, then you can see from Lagrange's form that

$$E_1(1+x)^n = \frac{n(n-1)}{2}(1+c)^{n-2}x^2$$

for some $c \in (0,x)$. This error gives us the useful inequality

$$(1+x)^n \ge 1 + nx + \frac{n(n-1)}{2}x^2$$

when $x > 0$ and $n \geq 2$. (It turns out you can prove this inequality without the use of calculus; try it!)

This example only uses a linear approximation to a complicated quantity. However, Taylor polynomials can also be used with higher orders to obtain even better accuracy. Because we have the flexibility to choose whichever order we like, we can customize the order of Taylor polynomials to suit our needs. This gives us a more complete picture of how a function behaves near a point, much like a derivative at a point gives us a sense of how much the function rises or falls. In this section, based on these ideas, we will present a few important examples of the use of Taylor approximation.

Kinetic Energy: Einstein vs. Newton

One of the first problems one might study in physics is determining how much energy a moving object carries as a result of its motion. This is called its *kinetic energy*. Kinetic energy is important, for instance, when analyzing how much fuel a car or a rocket is expected to burn to move at a designated speed. For many years, the main formula for calculating kinetic energy was the following formula, due to Isaac Newton:

$$K_N(v) = \frac{1}{2}mv^2$$

Here, v is the velocity of the object, m is its mass, and $K_N(v)$ is its kinetic energy (we have chosen to use the symbol K_N, where the subscript N stands for Newton). Newton's formula tells us that kinetic energy depends quadratically on speed: thus, for instance, if two objects have the same mass but one moves twice as fast as the other, then the faster object will have four times as much energy.

In 1905, however, the world of physics changed dramatically when Albert Einstein published his famous paper on the theory of special relativity. One of the findings of this paper was that Newton's formula for kinetic energy did not properly model the correct behavior at very high speeds. Instead, Einstein had a different kinetic energy formula, which we will denote as K_E with a subscript of E for Einstein:

$$K_E(v) = \frac{mc^2}{\sqrt{1-(v^2/c^2)}} - mc^2$$

Here, c is the speed of light, which is about $299,792,458$ meters per second or $670,616,629$ miles per hour.

In Einstein's formula, the first term represents the total energy E that an object with mass[7] m carries when moving with velocity v. For instance, when $v = 0$ (for an object at rest), we have the famous result $E = mc^2$. Also, we can see that $K_E(v) \to \infty$ as $v \to c^-$, so Einstein predicted that an object's kinetic energy becomes unbounded as its speed approaches the speed of light. In contrast, Newton's formula says that $K_N(v) \to mc^2/2$ as $v \to c^-$, so kinetic energy remains bounded with this formula.

How can we reconcile the differences between these two formulas? Since the dependence of $K_E(v)$ on v is somewhat complicated, let's use Taylor polynomials to make approximations for $K_E(v)$. The hardest part to evaluate in $K_E(v)$ is the expression $(1 - (v^2/c^2))^{-1/2}$. To handle this, we introduce the function $f(x) = (1-x)^{-1/2}$, so that

$$K_E(v) = mc^2\left(f\left(\frac{v^2}{c^2}\right) - 1\right)$$

Since $f'(x) = (1-x)^{-3/2}/2$, the linear approximation to f centered at 0 is

$$T_1 f(x) = f(0) + f'(0)x = 1 + \frac{x}{2}$$

Using this approximation in $K_E(v)$, we get

$$\begin{aligned} K_E(v) &= mc^2\left(f\left(\frac{v^2}{c^2}\right) - 1\right) \\ &\approx mc^2\left(\left(1 + \frac{v^2}{2c^2}\right) - 1\right) \\ &= \frac{1}{2}mv^2 = K_N(v) \end{aligned}$$

Thus, Newton's formula can be considered a Taylor approximation of Einstein's formula. When v^2/c^2 is close to 0, i.e. at low speeds, Newton's formula is quite accurate, but Newton's formula becomes less accurate at high speeds, when $v \approx c$. To be more precise, we can analyze the error from the Taylor approximation. For any $x \in (0,1)$, Lagrange's form says that there exists some ξ from 0 to x such that

$$E_1 f(x) = \frac{f''(\xi)}{2}x^2 = \frac{3}{8}(1-\xi)^{-5/2}x^2$$

[7] Technically, m is the *rest mass*, i.e. the mass when the object is at rest. Einstein also postulated that the mass of an object effectively changes as a function of its speed, but for simplicity we will assume m is constant.

which is a value from $3x^2/(8(1-x)^{5/2})$ to $3x^2/8$. Therefore, since $K_E(v) - K_N(v) = mc^2E_1 f(v^2/c^2)$, we have

$$\frac{3mv^4}{8c^2(1-(v^2/c^2))^{5/2}} \le K_E(v) - K_N(v) \le \frac{3mv^4}{8c^2}$$

For example, let's say our object in question is an average automobile, with a mass of $m = 2000$ kilograms.[8] If the car is driving at highway speeds, so $v = 90$ kilometers per hour (which is 25 meters per second), then $K_E(v) - K_N(v)$ is bounded by approximately

$$\frac{3 \cdot 2000 \cdot 25^4}{8(3 \times 10^8)^2} \approx 3 \times 10^{-9}$$

Thus, these measurements differ by only about 3 nJ (nanoJoules)! To get a sense of how tiny this quantity is, a Joule is the amount of energy needed to move 1 kg a distance of 1 meter, i.e. approximately the energy used to lift a small apple from the ground to the height of an average person's torso. Our error in measurement is approximately 1 billion times less energy![9] This illustrates that for most common objects, moving at common speeds, the difference between Einstein's formula and Newton's formula is negligible.

Proving e Is Irrational

As a more surprising application of Taylor polynomials, we can prove

Theorem 8.26. *$e \notin \mathbb{Q}$.*

Strategy. The main idea is that if e is rational, then $n!\,e$ is also an integer when n is a sufficiently large natural number. However, $e \approx T_n \exp(1)$, and $n!\,T_n \exp(1)$ is also an integer because

$$T_n \exp(1) = \sum_{i=0}^{n} \frac{1}{i!} = \frac{1}{n!}\left(\sum_{i=0}^{n} \frac{n!}{i!}\right)$$

[8] Thanks to http://cars.lovetoknow.com/How_Much_Does_My_Car_Weigh for some sample data.

[9] As another illustration, a mosquito's kinetic energy is approximately 53 times bigger than this error! (Thanks to http://en.wikipedia.org/wiki/Joule for this information.)

(Note that for each i from 0 to n, $n!/i!$ is the same as $(i+1)(i+2)\cdots(n)$, so it is an integer.) Thus,

$$n!\,E_n \exp(1) = n!\,e - n!\,T_n \exp(1)$$

and $n!\,E_n \exp(1)$ is also an integer. However, $E_n \exp(1)$ approaches 0 very quickly, and we can show that when n is large enough, $|n!\,E_n \exp(1)|$ is too small to be an integer.

Proof. Let's make the argument from the strategy more precise. Suppose e is rational, so that $e = p/q$ for some $p, q \in \mathbb{N}^*$. Hence, $n!\,e$ is an integer for any $n \in \mathbb{N}^*$ with $n \geq q$. For any such value of n, $n!\,E_n \exp(1)$ is also an integer as discussed in the strategy. We will show that $0 < n!\,E_n \exp(1) < 1$ when n is large enough, contradicting the fact that $n!\,E_n \exp(1)$ is an integer.

To do this, we obtain some bounds on $E_n \exp(1)$ by using Corollary 8.18 with $a = 0$ and $x = 1$. Thus, we want bounds on $\exp^{(n+1)}$ over the interval $[0,1]$. The $(n+1)^{\text{st}}$ derivative of exp is exp, so for any $t \in [0,1]$, we have $\exp(0) \leq \exp^{(n+1)}(t) \leq \exp(1)$, i.e. $1 \leq \exp^{(n+1)}(t) \leq e$. Hence, Corollary 8.18 tells us that

$$\frac{1}{(n+1)!} \leq E_n \exp(1) \leq \frac{e}{(n+1)!}$$

Thus, when $n \in \mathbb{N}^*$ and $n \geq \max\{2, q\}$, we have

$$\frac{1}{3} \leq \frac{1}{n+1} \leq n!\,E_n \exp(1) \leq \frac{e}{n+1} < 1$$

This shows that $n!\,E_n \exp(1)$ is an integer between 0 and 1, which is impossible. Hence, e cannot be rational. ∎

Remark. The argument used to prove Theorem 8.26 suggests the following question: can π be proven irrational with a similar argument? After all, for every $n \in \mathbb{N}^*$, we have $\pi/4 \approx T_{2n} \arctan(1)$. Unfortunately, this argument won't work here; let's follow its steps and see which part fails.

We start by finding an expression for $T_{2n} \arctan(1)$, and we discover (you should fill in the details) that $(2n-1)!\,T_{2n} \arctan(1) \in \mathbb{Z}$. If π is assumed to be rational (so that $\pi/4$ is also rational) and n is large enough, then $(2n-1)!\,\pi/4 \in \mathbb{Z}$. Thus, $(2n-1)!\,E_{2n} \arctan(1)$ is also an integer.

However, the error bounds obtained from Example 8.24 show that

$$|(2n-1)!\,E_{2n} \arctan(1)| \leq \frac{(2n-1)!}{2n+1}$$

As n grows large, this bound grows infinitely large. Thus, this bound cannot be used to show that $|(2n-1)!\, E_{2n} \arctan(1)| < 1$. This is where the proof breaks down.

The reason this proof attempt does not work here is that although the errors $E_{2n} \arctan(1)$ approach 0 as n grows large, they approach 0 much more slowly than the errors $E_n \exp(1)$ approach 0. In particular, $n!\, E_n \exp(1)$ goes to 0, but $(2n-1)!\, E_{2n} \arctan(1)$ might not go to 0. To prove π is irrational, we want a better approximation for π than $4T_{2n} \arctan(1)$.

This leads to another attempt by using a better approximation: Machin's identity from Exercise 8.5.18. Hence, we have

$$\pi \approx 16T_{2n} \arctan\left(\frac{1}{5}\right) - 4T_{2n} \arctan\left(\frac{1}{239}\right)$$

However, this attempt does not work either. The first problem is that although the approximations are rational, their denominators are even larger than $(2n-1)!$. (Naively, you can write the approximation as a fraction with denominator $5^{2n-1}239^{2n-1}(2n-1)!$.) Second, let's say that E_n^* is the error of this approximation. Although E_n^* approaches 0 more rapidly than $E_{2n} \arctan(1)$ does as n grows large, when you create an upper bound for $(2n-1)!\, E_n^*$, that upper bound still grows arbitrarily large. Hence, it is unlikely that you can show $0 < |(2n-1)!\, E_n^*| < 1$ with this method.

It turns out that proofs of π's irrationality tend to use different tactics: one such proof is outlined in Exercise 8.7.23. However, modifications of the proof of Theorem 8.26 can prove the irrationality of several other numbers: see Exercises 8.7.3 and 8.7.4.

Little-o Notation And Limits

Our last major application of Taylor polynomials will be to use them to compute limits. This is where our theorems regarding the order of a Taylor error come into play. As an example of how this works, let's revisit the limit

$$\lim_{x \to 0} \frac{\sin x}{x}$$

We originally proved this limit is 1 by using geometry and the Squeeze Theorem. Later, we were able to show that L'Hôpital's Rule could easily compute the limit as well. Now, with the aid of Taylor polynomials, we have another way to proceed.

Intuitively, when x is close to 0, $\sin x \approx T_1 \sin(x) = x$, so $(\sin x)/x \approx x/x = 1$. This is too informal, but we can make it more formal by introducing the error of the approximation, $E_1 \sin(x)$. We have

$$\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{T_1 \sin(x) + E_1 \sin(x)}{x} = \lim_{x \to 0} \frac{x}{x} + \frac{E_1 \sin(x)}{x}$$

Certainly we have $x/x = 1$ when $x \neq 0$, and we also saw in Theorem 8.19 that $E_1 \sin(x)$ is "better than 1^{st}-order" as $x \to 0$, i.e. $E_1 \sin(x)/x \to 0$. This proves that $(\sin x)/x \to 1$, as desired.

Before we proceed to more examples of the use of Taylor polynomials in limits, note that the only information we needed about $E_1 \sin(x)$ was merely that $E_1 \sin(x)/x \to 0$. In particular, we didn't need to know any other details about $E_1 \sin(x)$ that Taylor's Theorem or Lagrange's form provides. Because of this, it is helpful to introduce a shorthand notation which places our focus solely on this limiting behavior:

Definition 8.27. Let $a \in \mathbb{R}$, and suppose f and g are real functions defined near a, where $g(x) \neq 0$ for all x near a (except possibly at a). We say that $f(x)$ *is little-o* (as in the letter "o"[10]) *of* $g(x)$ *as* $x \to a$, also written as "$f(x) = o(g(x))$ as $x \to a$", to mean

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 0$$

Intuitively, this means that $f(x)$ has higher order than $g(x)$ as $x \to a$. For instance, whenever we said "$E_n f(x)$ is better than n^{th}-order as $x \to a$" earlier, we could say the same thing more compactly as "$E_n f(x) = o((x-a)^n)$".

Remark. It is rather common to omit the phrase "as $x \to a$" when describing little-o notation if the value of a is clear from context. Similarly, we sometimes will write a whole string of calculations involving little-o and only write "as $x \to a$" at the end. This helps us focus on the main computations.

Also, it is important to note that the notation for little-o is misleading, particularly the use of the equals sign. For instance, if $a \in \mathbb{R}$, and f_1, f_2, and g are functions satisfying $f_1(x) = o(g(x))$ and $f_2(x) = o(g(x))$ as $x \to a$, then we may not conclude that $f_1(x) = f_2(x)$. After all, knowing $f_1(x)/g(x) \to 0$

[10] This notation is due to E. Landau. You can look up several variants of this notation, such as big-O and big-Ω; these variants are useful for describing growth rates, especially in computer science.

and $f_2(x)/g(x) \to 0$ doesn't tell us much about how $f_1(x)$ and $f_2(x)$ are related to each other.

One possible interpretation of the notation "$o(g(x))$ as $x \to a$" is to think of $o(g(x))$ as a set: it can be considered as the set of all functions f for which $f(x)/g(x) \to 0$ as $x \to a$. With this interpretation, the expression "$f(x) = o(g(x))$" should really be "$f(x) \in o(g(x))$", which avoids the confusion from the previous paragraph. However, this notation is not in common usage.

The current notation is still popular because it helps us use little-o expressions in formulas. When we write an expression $o(g(x))$ in a formula, we mean "some function f for which $f(x) = o(g(x))$". For instance, the statement

$$\sin x + \cos x = (x + o(x)) + (1 + o(1)) \qquad \text{as } x \to 0$$

is really shorthand for "$\sin x + \cos x$ has the form $(x + f_1(x)) + (1 + f_2(x))$ for some functions f_1 and f_2 satisfying $f_1(x) = o(x)$ and $f_2(x) = o(1)$". (Note that saying $f_2(x) = o(1)$ is the same as saying that $f_2(x) \to 0$ as $x \to 0$.) Similarly, the statement

$$\sin x - x = o(o(x^2)) \qquad \text{as } x \to 0$$

means "$\sin x - x = o(f(x))$ for some function f satisfying $f(x) = o(x^2)$".

Little-o notation can be used to write some of our theorems more compactly. Theorem 8.19 states that when a real function f has a continuous $(n+1)^{\text{st}}$ derivative at a number $a \in \mathbb{R}$, then $E_n f(x; a) = o((x-a)^n)$ as $x \to a$. Conversely, Theorem 8.22 basically states that when $a \in \mathbb{R}$ and E is a function satisfying $E(x) = o((x-a)^n)$ as $x \to a$, then $T_n E(x; a) = 0$.

Many convenient properties about little-o notation, which make it easier to compute with little-o expressions, are given in this theorem:

Theorem 8.28. *Let $a, c \in \mathbb{R}$ be given. Suppose f and g are real functions defined near a with $f(x), g(x) \neq 0$ for x near a. As $x \to a$, we have*

1. *$o(g(x)) + o(g(x)) = o(g(x))$ and $o(g(x)) - o(g(x)) = o(g(x))$. (In other words, the sum and difference of any two functions which are $o(g(x))$ are also $o(g(x))$.)*

2. *If $c \neq 0$, then $o(cg(x)) = o(g(x))$. (In other words, any function which is $o(cg(x))$ is also $o(g(x))$. This basically means that constant factors don't affect the order of a function.)*

3. *$f(x) \cdot o(g(x)) = o(f(x)g(x))$. (In other words, when f is multiplied by a function which is $o(g(x))$, the result is $o(f(x)g(x))$.)*

4. *$o(g(x))/f(x) = o(g(x)/f(x))$. (In other words, when a function which is $o(g(x))$ is divided by $f(x)$, the result is $o(g(x)/f(x))$.)*

5. *$o(o(g(x))) = o(g(x))$. (In other words, if two functions f_1, f_2 satisfy $f_1(x) = o(f_2(x))$ and $f_2(x) = o(g(x))$, then $f_1(x) = o(g(x))$.)*

The following consequence of Part 5 above is also worth noting: if $g(x) = o(1)$ and f_1 is a function satisfying $f_1(x) = o(g(x))$, then $f_1(x) = o(o(1)) = o(1)$ and hence $f_1(x) \to 0$ as $x \to a$.

Strategy. The hardest part about proving this theorem is understanding what the theorem actually says. Once that is done, many of the proofs come from applying the definition of little-o and using limit laws. Part 5 is a little trickier, but it can also be proven by using the definitions: if $f_1(x) = o(f_2(x))$ and $f_2(x) = o(g(x))$, i.e. $f_1(x)/f_2(x)$ and $f_2(x)/g(x)$ both go to 0, then

$$\frac{f_1(x)}{g(x)} = \frac{f_1(x)}{f_2(x)} \cdot \frac{f_2(x)}{g(x)}$$

also goes to 0.

Proof. Let a, c, f, g be given as described. For part 1, suppose that f_1 and f_2 are functions satisfying $f_1(x) = o(g(x))$ and $f_2(x) = o(g(x))$ as $x \to a$. Thus, $f_1(x)/g(x)$ and $f_2(x)/g(x)$ approach 0 as $x \to a$. It follows that

$$\lim_{x\to a} \frac{f_1(x) + f_2(x)}{g(x)} = \lim_{x\to a} \frac{f_1(x)}{g(x)} + \frac{f_2(x)}{g(x)} = 0 + 0 = 0$$

so $f_1(x) + f_2(x) = o(g(x))$. The same basic argument shows $f_1(x) - f_2(x) = o(g(x))$.

For part 2, if $f_1(x) = o(cg(x))$, then we multiply and divide by c to obtain

$$\lim_{x\to a} \frac{f_1(x)}{g(x)} = c \lim_{x\to a} \frac{f_1(x)}{cg(x)} = c \cdot 0 = 0$$

showing that $f_1(x) = o(g(x))$.

For part 3, suppose that f_1 and f_2 are functions satisfying $f_1(x) = f(x)f_2(x)$ where $f_2(x) = o(g(x))$. Then

$$\lim_{x\to a} \frac{f(x)f_2(x)}{f(x)g(x)} = \lim_{x\to a} \frac{f_2(x)}{g(x)} = 0$$

since $f(x) \neq 0$ for x near a. This proves $f_1(x) = f(x)f_2(x) = o(f(x)g(x))$. Part 4 works nearly the same way: if $f_1(x) = f_2(x)/f(x)$ where $f_2(x) = o(g(x))$, then

$$\lim_{x\to a} \frac{f_2(x)/f(x)}{g(x)/f(x)} = \lim_{x\to a} \frac{f_2(x)}{g(x)} = 0$$

proving part 4.

Lastly, for part 5, suppose f_1 and f_2 are functions satisfying $f_1(x) = o(f_2(x))$ and $f_2(x) = o(g(x))$. Therefore, $f_1(x)/f_2(x)$ and $f_2(x)/g(x)$ both approach 0 as $x \to a$. Furthermore, by the definition of $f_1(x) = o(f_2(x))$, $f_2(x) \neq 0$ for x near a. Thus, we may multiply and divide by $f_2(x)$ when x is near a to obtain

$$\lim_{x\to a} \frac{f_1(x)}{g(x)} = \lim_{x\to a} \frac{f_1(x)}{f_2(x)} \cdot \frac{f_2(x)}{g(x)} = 0 \cdot 0 = 0$$

proving that $f_1(x) = o(g(x))$. ■

Example 8.29:

In many cases, we use Theorem 8.28 when $g(x)$ has the form $(x-a)^n$ for some $n \in \mathbb{N}$. In these cases, it is useful to note that whenever $m \in \mathbb{N}$ with $m < n$, you can easily show that $(x-a)^n = o((x-a)^m)$ as $x \to a$. (Hence, it is more informative to say a function is little-o of a high power than to say it is little-o of a low power. We sometimes describe this by saying that the low power is a *weaker order* than the high power, or that the high power is a *stronger order* than the low power.) By part 5 above, $o(g(x)) = o((x-a)^n) = o((x-a)^m)$.

This remark is especially useful when combined with part 1. For instance, we can show that $o(x^5) + o(x^3) = o(x^3)$ as $x \to 0$ by the calculations

$$o(x^5) + o(x^3) = o(o(x^3)) + o(x^3) = o(x^3) + o(x^3) = o(x^3)$$

(the second equality follows from part 5 and the third follows from part 1). We cannot, in general, replace the x^3 in the last step with a stronger order. For instance, $x^6 + x^4 = o(x^5) + o(x^3)$, but $x^6 + x^4$ is not $o(x^4)$ because

$(x^6+x^4)/x^4$ approaches 1 as $x \to 0$. Intuitively, we see that in general, when adding two little-o expressions with different orders, you can only conclude that the result is little-o of the weaker order (so in the above calculation, x^3 is a weaker order than x^5).

For some extra practice, you should use the remarks above and Theorem 8.28 to check the following calculations by filling in any missing details or steps:

$$\begin{aligned} \sin x + \log(1+x) &= \left(x - \frac{x^3}{6} + o(x^3)\right) + \left(x - \frac{x^2}{2} + o(x^2)\right) \\ &= 2x - \frac{x^2}{2} - \frac{x^3}{6} + o(x^2) \\ &= 2x - \frac{x^2}{2} + o(x^2) \end{aligned}$$

and

$$1 - \cos x = \frac{x^2}{2} + o(x^3)$$

(For the second calculation, it helps to note that $T_2 \cos(x) = T_3 \cos(x)$ when $a = 0$.) $\square$

With the tools provided by Theorem 8.28 and Example 8.29, we can use little-o notation to compute many limits. Let's look at some examples.

Example 8.30:

Let's compute the limit

$$\lim_{x\to 0} \frac{\sin x - x}{x^3}$$

This limit can be computed with three applications of L'Hôpital's Rule, but let's instead show how little-o notation solves this problem quickly. Little-o notation works well here because we know Taylor polynomials for $\sin x$ with center 0. If we try a linear approximation for $\sin x$, so that $\sin x = x + o(x)$, then we get

$$\frac{\sin x - x}{x^3} = \frac{x + o(x) - x}{x^3} = \frac{o(x)}{x^3} = o(x^{-2})$$

as $x \to 0$ (the last step uses part 4 of Theorem 8.28). Unfortunately, this does not help us find our limit, because $x^{-2} \to \pm\infty$ as $x \to 0$. Hence, knowing that a function is $o(x^{-2})$ does not uniquely tell us that function's limit!

To avoid this problem, we'd like our final result to only use little-o expressions of functions which approach 0 as $x \to 0$. Thus, we should use a Taylor polynomial for $\sin x$ which has error with order at least $o(x^3)$ to cancel the x^3 in the denominator of our limit. Hence, let's use $T_3 \sin(x)$ to approximate $\sin x$, so $\sin x = x - (x^3)/(3!) + o(x^3)$. This gives us

$$\frac{\sin x - x}{x^3} = \frac{x - \frac{x^3}{6} + o(x^3) - x}{x^3} = \frac{-1}{6} + o(1)$$

Hence

$$\lim_{x\to 0} \frac{\sin x - x}{x^3} = \lim_{x\to 0} \frac{-1}{6} + o(1) = \frac{-1}{6}$$

by the definition of $o(1)$. This shows that when using little-o notation with limits, some choices of Taylor order will not give you the answer, whereas higher orders will work. (You should check to see that if you use $T_5 \sin(x)$ instead, the answer will not change.)

It is also useful to note that since $T_3 \sin(x) = T_4 \sin(x)$ (the 4^{th}-order Taylor coefficient is 0), so we actually have $\sin x = x - (x^3)/6 + o(x^4)$. Because of this, we can compute another limit:

$$\lim_{x\to 0} \frac{\sin x - x}{x^4} = \lim_{x\to 0} \frac{\frac{-x^3}{6} + o(x^4)}{x^4} = \lim_{x\to 0} \frac{-1}{6x} + o(1)$$

We conclude that this limit is infinite because $o(1) \to 0$ and $-1/(6x) \to \pm\infty$ as $x \to 0$. $\square$

Example 8.31:

Little-o calculations are especially useful when dealing with multiple approximations at the same time. For instance, let's consider

$$\lim_{x\to 0} \frac{\sin x - x\cos x}{x^2}$$

For our first attempt, let's try the simplest approximation possible for cosine: $\cos x = 1 + o(1)$. Thus, $x \cos x = x + o(x)$ by part 3 of Theorem 8.28. To cancel the x in this expression, we try a linear approximation with the sine: $\sin x = x + o(x)$. Thus,

$$\frac{\sin x - x\cos x}{x^2} = \frac{(x + o(x)) - x(1 + o(1))}{x^2} = \frac{o(x)}{x^2} = o(x^{-1})$$

As in the previous example, this final result does not help us.

Thus, we should use higher-order Taylor polynomials. Let's try second-order for $\cos x$ and third-order for $\sin x$: $\sin x = x - (x^3)/6 + o(x^3)$ and $x \cos x = x(1 - (x^2)/2 + o(x^2))$. When we pick these orders, the only terms which do not cancel contain factors of x^2, so we get

$$\begin{aligned}\lim_{x \to 0} \frac{\sin x - x\cos x}{x^2} &= \lim_{x \to 0} \frac{\left(x - \frac{x^3}{6} + o(x^3)\right) - \left(x - \frac{x^3}{2} + o(x^3)\right)}{x^2} \\ &= \lim_{x \to 0} \frac{x}{3} + o(x) = 0\end{aligned}$$

In fact, the same work shows a more useful limit:

$$\lim_{x \to 0} \frac{\sin x - x\cos x}{x^3} = \lim_{x \to 0} \frac{1}{3} + o(1) = \frac{1}{3}$$

You should try L'Hôpital's Rule as well with this second limit, and you'll see that little-o notation works much more quickly. $\square$

Example 8.32:

As another example involving multiple approximations, let's look at

$$\lim_{x \to 0} \frac{a^x - b^x}{x}$$

where $a, b \in \mathbb{R}^+$ and $a \neq b$. You were asked to compute $T_n a^x$ in a previous exercise, though we can also note that a^x can be written in terms of exp by its definition: $a^x = \exp(x \log a)$. Therefore, we may use the linear approximation to exp, along with Theorem 8.7 to handle the $\log a$ constant, to conclude

$$a^x = e^{x \log a} = 1 + (x \log a) + o(x \log a) = 1 + x \log a + o(x)$$

(the last step uses part 2 of Theorem 8.28 when $a \neq 1$). Similarly, we obtain $b^x = 1 + x \log b + o(x)$. Thus,

$$\lim_{x \to 0} \frac{a^x - b^x}{x} = \lim_{x \to 0} \frac{x(\log a - \log b) + o(x)}{x} = \lim_{x \to 0} (\log a - \log b) + o(1)$$

so our limit is $\log a - \log b = \log(a/b)$. $\square$

Example 8.33:

Little-o notation even helps with indeterminate forms which are not $0/0$-form. For instance, let's consider

$$\lim_{x\to 0^+} x^{(x^x-1)}$$

i.e. the base is x and the exponent is one less than x^x. (We can modify the definition of little-o to handle one-sided limits; the properties it satisfies are basically the same.) Let's first focus on the exponent, $x^x - 1$. By definition, $x^x = \exp(x\log x)$.

To handle this, we first handle exp on its own: $\exp(t) = 1 + t + o(t)$ as $t \to 0$. We'd like to replace t with $x\log x$. We note that $x\log x \to 0$ as $x \to 0^+$, as we showed in Chapter 7 (unfortunately, we cannot use little-o work to get this limit, because $\log x$ does not have a Taylor polynomial centered at 0). This allows us to plug in $x\log x$ for t and conclude

$$\exp(x\log x) = 1 + x\log x + o(x\log x) \qquad \text{as } x \to 0^+$$

(see Exercise 8.7.7 for more of the formal details).

Therefore,

$$x^{x^x-1} = x^{x\log x + o(x\log x)} = \exp((x\log x + o(x\log x))\log x)$$

where we again use the definition of exponentiation to a base which is not e. The methods from Chapter 7 also prove that $x\log^2 x \to 0$ as $x \to 0^+$, so by plugging in $x\log^2 x + o(x\log^2 x)$ for t in our Taylor expression for $\exp(t)$ above, we get

$$x^{x^x-1} = 1 + (x\log^2 x + o(x\log^2 x)) + o(x\log^2 x + o(x\log^2 x))$$

Since every term approaches 0 as $x \to 0^+$ except for the term 1, we have $x^{x^x-1} \to 1$. $\square$

As we have seen, Taylor polynomials with little-o error estimates frequently yield convenient expressions for evaluating limits. You will get more examples to try in the exercises. Frequently, little-o expressions provide quicker and more convenient approaches to limits than L'Hôpital's Rule (though as we saw with $x\log x$ in the previous example, we cannot expect to avoid all usage of L'Hôpital's Rule).

8.7 Exercises

In the first two exercises, we present some other uses of Taylor polynomial approximations in science.

1. When a string or a cable is supported on two ends, and the middle hangs down, the resulting shape is often called a *catenary*. Let's say that the lowest point of the catenary is $(0,0)$. We model the catenary as a curve of the form

$$y = f(x) = \frac{a}{2}\left(e^{x/a} + e^{-x/a}\right) - a$$

where a is a positive constant describing how much the catenary "opens up". Note that $f(0) = 0$ and $f(a) = a(e+1/e)/2 - a \approx 0.543a$. A couple different catenaries are shown in Figure 8.2.

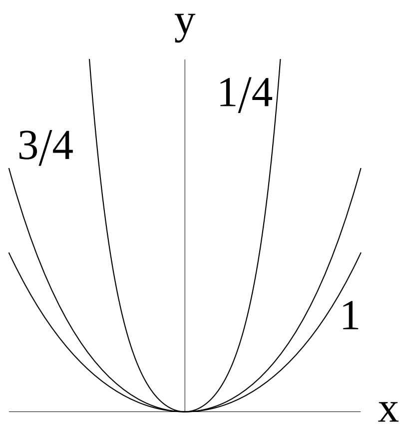

Figure 8.2: Three catenaries, labeled with their value of a

These pictures suggest that a catenary looks nearly like a parabola when x is close to 0. Use Taylor polynomials to prove that when x is close to 0,

$$f(x) \approx \frac{x^2}{2a}$$

(This confirms our intuition that a smaller value of a leads to a steeper catenary.) Also, prove the following error estimate for all $x \in [-a, a]$:

$$\left| f(x) - \frac{x^2}{2a} \right| \le \frac{e|x|^3}{3a^2} \le \frac{ea}{3}$$

2. An electric *dipole* consists of two electric charges which have the same magnitude but opposite sign (such as a proton and an electron a small distance apart). Let's suppose the charges have values q and $-q$ and are separated by a distance d. We consider a point P as in Figure 8.3, so P is collinear with the charges q and $-q$, P is closer to q than to $-q$, and P is distance D from q. It is known that the electric field at the point P is
$$E(D) = \frac{q}{D^2} - \frac{q}{(D+d)^2}$$

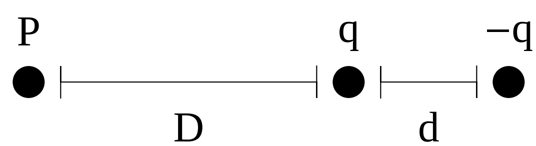

Figure 8.3: A dipole

Prove that as D goes to ∞, $E(D)$ is approximately proportional to $1/D^3$. (Hint: Pull a factor of $1/D^2$ out of $E(D)$, and try to approximate a function $f(x)$ where $x = d/D$, since x is close to 0.)

3. Suppose that we use the proof tactic from Theorem 8.26 to prove that e^2 is irrational. Show that $n!\, E_n \exp(2)$ does not approach 0 as $n \to \infty$, and hence the proof tactic fails.

4. It is known that e^m is irrational for every $m \in \mathbb{N}^*$, but we will not prove this fact (the proof uses tactics somewhat similar to the proof of π's irrationality at the end of this section). Use this fact to prove the following:
   (a) Whenever $q \in \mathbb{Q}$ and $q \neq 0$, $e^q \notin \mathbb{Q}$.
   (b) Use part (a) to prove that $\log q$ is irrational whenever $q \in \mathbb{Q}$ and $q > 1$.

5. Suppose that $a \in \mathbb{R}$ and g is a real function defined near a with $g(x) \neq 0$ for x near a. If $g(x) \to 0$ as $x \to a$, prove that
$$o(g(x) + o(g(x))) = o(g(x)) \qquad \text{as } x \to a$$
In other words, if f and h are real functions defined near a such that $g(x)+h(x) \neq 0$ for x near a, $f(x) = o(g(x)+h(x))$, and $h(x) = o(g(x))$, then $f(x) = o(g(x))$.

6. Prove that if $a \in \mathbb{R}$ and f and g are real functions defined near a with $f(x), g(x) \neq 0$ for x near a, then $o(f(x)) \cdot o(g(x)) = o(f(x)g(x))$ as $x \to a$. Also, use this to show that

$$\log^2(1+x) = x^2 - x^3 + o(x^3) \qquad \text{as } x \to 0$$

7. Suppose that $a, b \in \mathbb{R}$, and f, g, and h are real functions defined near a such that $g(t), h(t) \neq 0$ for t near a. Also, suppose that x is a real function defined near b such that $x(t) \neq a$ for t near b, and $x(t) \to a$ as $t \to b$. Prove that

$$f(t) = g(t) + o(h(t)) \qquad \text{as } t \to a$$

implies

$$f(x(t)) = g(x(t)) + o(h(x(t))) \qquad \text{as } t \to b$$

This allows us to substitute functions into little-o expressions. (Hint: Use the definition of little-o together with the Composition Limit Theorem.)

8. (a) Suppose that $a \in \mathbb{R}$ and g is a real function defined near a with $g(x) \neq 0$ for x near a. Prove that if $g(x) \to 0$ as $x \to a$, then for any $n \in \mathbb{N}$,

$$\frac{1}{1-g(x)} = \sum_{i=0}^{n} g^i(x) + o(g^n(x)) \qquad \text{as } x \to a$$

(Hint: Consider $T_n(1/(1-t))$ centered at $t = 0$.)

(b) Use part (a) to prove that

$$\frac{1}{\cos x} = 1 + \frac{x^2}{2} + o(x^2) \qquad \text{as } x \to 0$$

(Hint: Approximate $\cos x$ first.)

(c) Use part (b) to prove that

$$\tan x = x + \frac{x^3}{3} + o(x^3) \qquad \text{as } x \to 0$$

(Hint: You may want to use the result of Exercise 8.7.6.)

In Exercises 9 through 21, use little-o notation to compute the following limits. To aid your work, you may use the following limits from Chapter 7, which are true for all $a > 0$ and all $b > 1$:

$$\lim_{x\to\infty} \frac{x^a}{e^x} = 0 \qquad \lim_{x\to\infty} \frac{\log^b x}{x^a} = 0 \qquad \lim_{x\to 0^+} x^a \log^b x = 0$$

9. $\displaystyle\lim_{x\to 0} \frac{\sin(ax)}{\sin(bx)}$

   where $a, b > 0$ are constants

10. $\displaystyle\lim_{x\to 0} \frac{\sin x - x}{\arctan x - x}$

11. $\displaystyle\lim_{x\to 0} \frac{1-\cos(x^2)}{x^2 \sin(x^2)}$

12. $\displaystyle\lim_{x\to\infty} x^{1/4}\left(e^{-1/\sqrt{x}} - 1\right)$

    (Hint: Change variables.)

13. $\displaystyle\lim_{x\to \pi/2} \frac{\cos x}{x - \pi/2}$

    (Hint: What should be the center of your Taylor polynomial?)

14. $\displaystyle\lim_{x\to 0^+} \log x \log(1-x)$

    (Hint: Only one of these logarithms can be approximated with Taylor polynomials.)

15. $\displaystyle\lim_{x\to 0^+} \sin x \log x$

    (Hint: Which function can be approximated near $x = 0$?)

16. $\displaystyle\lim_{x\to 0^+} x^{\sin x}$

    (Hint: Use the previous exercise.)

17. $\displaystyle\lim_{x\to 0^+} (\sin x)^x$

    (Hint: After an approximation, you'll want to use log properties.)

18. $\displaystyle\lim_{x\to 0} \frac{\log(\cos(ax))}{\log(\cos(bx))}$

    (Hint: After one set of approximations, you'll want to do another set.)

19. $\displaystyle\lim_{x\to 1} x^{1/(1-x)}$

    (Hint: It might be easier to solve this with a variable approaching 0.)

20. $\lim_{x\to 1} \left(\frac{1}{\log x} - \frac{1}{x-1}\right)$

21. $\lim_{x\to 0} \left(\frac{(1+x)^{1/x}}{e}\right)^{1/x}$

    (Hint: First show that $(1+x)^{1/x} = \exp(1 - x/2 + o(x)) = e \cdot (1 - x/2 + o(x))$. You might want to use the result of Exercise 8.7.5 for some steps.)

22. Find the value of $c > 0$ so that

    $$\lim_{x\to\infty} \left(\frac{x+c}{x-c}\right)^x = 4$$

    (Hint: $(x+c)/(x-c) = 1 + 2c/(x-c)$, and $2c/(x-c)$ approaches 0 as $x \to \infty$.)

23. In this exercise, we outline a proof that π is irrational.[11] For those who are curious, this proof uses techniques coming from the theory of Fourier series, as opposed to the theory of Taylor polynomials.

    We start by assuming for a contradiction that π is rational, so we may write $\pi = a/b$ where $a, b \in \mathbb{N}^*$. Let $n \in \mathbb{N}^*$ be given. We define $f, F : \mathbb{R} \to \mathbb{R}$ for all $x \in \mathbb{R}$ by

    $$f(x) = \frac{x^n(a-bx)^n}{n!}$$

    and

    $$F(x) = \sum_{j=0}^{n} (-1)^j f^{(2j)}(x)$$

    (Since f is a polynomial, f is infinitely differentiable at every point.)

    (a) Prove that f is a polynomial of degree $2n$ whose coefficients are integers divided by $n!$ (i.e. $n!\,f$ has integer coefficients), and also $f(0) = f(\pi) = 0$.

    (b) Prove that for all $x \in (0, \pi)$,

    $$0 < f(x) < \frac{\pi^n a^n}{n!}$$

[11] The website http://www.lrz-muenchen.de/~hr/numb/pi-irr.html provided this proof, which it says is due to John Niven.

(c) Prove that for all j from 0 to $n-1$, $f^{(j)}(0) = f^{(j)}(\pi) = 0$. (Hint: What kinds of terms appear in $f^{(j)}(x)$? You may find Exercise 4.4.6 useful.)

(d) Prove that for all $j \in \mathbb{N}$ with $j \geq n$, $f^{(j)}(0)$ and $f^{(j)}(\pi)$ are integers. (Your work from part (c) should be useful here.) It follows from this and part (c) that $F(0)$ and $F(\pi)$ are integers.

(e) Prove that for all $x \in \mathbb{R}$, $F(x) + F''(x) = f(x)$.

(f) Using part (e), prove that

$$\frac{d}{dx}(F'(x)\sin x - F(x)\cos x) = f(x)\sin x$$

for all $x \in \mathbb{R}$. Use this to show that

$$\int_0^\pi f(x)\sin x\,dx = F(0) + F(\pi)$$

(g) Use part (b) to show that

$$0 < \int_0^\pi f(x)\sin x\,dx < \frac{2\pi^n a^n}{n!}$$

(h) Finally, use the fact that factorials dominate exponentials (see Exercise 8.5.14), along with parts (d), (f), and (g), to show that if n is large enough, then $F(0) + F(\pi)$ is an integer between 0 and 1. This yields the desired contradiction.

Chapter 9

Sequences and Series

At many different points in this book, we have created lists of mathematical objects indexed by natural numbers. We call such a list a *sequence* of mathematical objects. For instance, here are some examples:

- In Lemma 3.58, we proved that any continuous real function f on a closed interval $[a, b]$ is bounded on $[a, b]$. This is the argument that the proof used, called the bisection method. We started with $[a_0, b_0] = [a, b]$, assuming for contradiction that f was unbounded on $[a, b]$. Thus, f was unbounded on one of the halves of $[a_0, b_0]$, and we took $[a_1, b_1]$ to be one such half. After that, we took $[a_2, b_2]$ to be one of the halves of $[a_1, b_1]$ on which f was unbounded, and so on. In this way, we developed a sequence of closed intervals

$$I_0 = [a_0, b_0], I_1 = [a_1, b_1], I_2 = [a_2, b_2], \ldots$$

so that for any $n \in \mathbb{N}$, I_{n+1} was one of the halves of I_n on which f was unbounded. Intuitively, the I_n intervals closed in on a single point as n grew large. In fact, this point was $\sup\{a_n \mid n \in \mathbb{N}\}$.

- In Chapter 5, we used sequences of step functions to help study integrability. For instance, in Example 5.1, we found the area under the function $f : [0, 1] \to \mathbb{R}$, defined by $f(x) = x^2$ for all $x \in \mathbb{R}$, as follows. For each $n \in \mathbb{N}^*$, we made step functions $s_n, t_n : [0, 1] \to \mathbb{R}$ which broke the interval $[0, 1]$ into n equal-width subintervals, such that s_n was below f and t_n was above f. Furthermore, we showed that by making n sufficiently large, we could get the areas under s_n and t_n to be as close as desired to 1/3. In other words, the sequences of numbers

$$\int_0^1 s_1(x)\,dx, \int_0^1 s_2(x)\,dx, \int_0^1 s_3(x)\,dx, \ldots$$

and

$$\int_0^1 t_1(x)\,dx, \int_0^1 t_2(x)\,dx, \int_0^1 t_3(x)\,dx, \ldots$$

both approached 1/3. Based on this, we concluded that f was integrable on $[0,1]$ with value 1/3.

- In Chapter 8, we introduced Taylor polynomials. In the special case where f was a real function which is infinitely differentiable at some $a \in \mathbb{R}$, we obtained a sequence of polynomials

  $$T_0 f(x;a), T_1 f(x;a), T_2 f(x;a), \ldots$$

  meant to approximate f well at points near a. Note that each Taylor polynomial could also be written in the form

  $$T_n f(x;a) = \sum_{i=0}^{n} t_i(x;a) \qquad \text{where} \qquad t_i(x;a) = \frac{f^{(i)}(a)}{i!}(x-a)^i$$

  Here, $t_i(x;a)$ can be thought of as the "i^{th} Taylor term" and $T_n f(x;a)$ is the sum of the Taylor terms up to the n^{th}.

These examples demonstrate how useful sequences are in mathematics. Furthermore, we see that there are many different ways of creating sequences. In our first example, I_{n+1} is created from I_n, so intervals in this sequence are recursively defined from the previous ones (as opposed to giving one formula that defines all the I_n intervals at once). In our second example, the sequence s_n of step functions is used to create the sequence of numbers $\displaystyle\int_0^1 s_n(x)\,dx$. Lastly, in our third example, we see that the T_n sequence of functions is built by taking sums from the t_i sequence of functions.

In fact, the process used in our third example, where we take a sequence and add up its members, is very common. The new sequence created this way is called the *series* whose terms are the *partial sums* of the old sequence. A series gives us a way to reason about "adding infinitely many numbers". Some series give us well-defined sums, and some do not.

To illustrate this, we consider two series. Our first series is made from the terms $a_n = 2^{-n}$ for each $n \in \mathbb{N}$. Thus, the n^{th} partial sum is

$$s_n = \sum_{i=0}^{n} a_i = 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n}$$

As this sum is a geometric summation with ratio $1/2$, Exercise 1.9.2 tells us that $s_n = 2 - 2^{-n}$ for each $n \in \mathbb{N}$. As $2 - 2^{-n} \to 2$ as $n \to \infty$, this suggests that the sum with "infinitely many terms" should equal 2:

$$\sum_{i=0}^{\infty} 2^{-i} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^i} + \cdots = 2$$

We say that this series has sum 2. This is an example of a *convergent* series, meaning that the partial sums approach a finite limit.

For our second series, suppose we instead use the sequence $a_n = 1$ for all $n \geq 0$ to build partial sums. Then s_n, the n^{th} partial sum of the a_n sequence, equals $n+1$ (recall that our sequence of a_n terms starts with $n = 0$). Hence, $s_n \to \infty$ as $n \to \infty$, which suggests that

$$\sum_{i=0}^{\infty} 1 = 1 + 1 + \cdots + 1 + \cdots = \infty$$

Unlike the previous series, this series has an infinite sum. We say that this series *diverges*. (More specifically, this series diverges to ∞.)

In this chapter, we will take these ideas, make them more precise, and explore the consequences, particularly with series. Series are usually quite tricky because it can be difficult to obtain a convenient expression for the partial sums s_n from which one can analyze the limiting behavior. Thus, in many cases we will not be able to find the exact value of the sum of a series though we will be able to determine whether the series converges. Even if we don't know the exact sum of a series, knowing whether the series converges is still useful, because as we'll see, convergent series can be manipulated algebraically in many ways as if they were finite sums.

9.1 Sequences

Before we proceed any further, we should establish a precise definition of what a sequence is. It can be instructive to ask: what properties should this definition have? One important property is that order of terms in a sequence DOES matter: the sequence $1, 2, 1, 2, \ldots$ is not the same as the sequence $2, 1, 2, 1, \ldots$. Another important property is that terms can occur multiple times in a sequence (unlike in a set, where multiplicity does not matter). Thus, a sequence is like an ordered pair, but with infinitely many coordinates (it could be considered an ordered "∞-tuple"). To make this more precise, we use functions:

Definition 9.1. If S is any set, then an *(infinite) sequence of members of* S is a function $a : \mathbb{N} \to S$. For each $n \in \mathbb{N}$, $a(n)$ is called the n^{th} *term of the sequence*, which is often written instead as a_n. We also say that a_n is the term with *index* n. Frequently, for our purposes, S will either be $\mathbb{R}$ or a set of real-valued functions.

More generally, for any sets I and S, a *sequence from S indexed by I* is a function $a : I \to S$, where we often write a_i instead of $a(i)$ for the term with *index* i when $i \in I$. I is called the *index set* of the sequence. We will also use the notation $(a_i)_{i \in I}$, or sometimes simply (a_i) if the context is clear, to refer to the entire sequence.

(As with many other notations, such as with summations or integrals, the subscript is a dummy variable and can be replaced with any other variable which is not already used. Thus, $(a_n)_{n \in \mathbb{N}}$ and $(a_k)_{k \in \mathbb{N}}$ mean the same thing.)

The *range* of the sequence $(a_i)_{i \in I}$ is $\text{ran}(a) = \{a_i \mid i \in I\}$. Thus, the range is a set, not a sequence. We will occasionally write this as $\{a_i\}_{i \in I}$, or more simply write $\{a_i\}$ if the context is clear.

With sequences, many notational shorthands are very common. When the index set is $\mathbb{N}$, we may write $(a_n)_{n=0}^{\infty}$ instead of $(a_n)_{n \in \mathbb{N}}$, using a notation similar to that for summations. Analogously, if the index set has the form $\{k, k+1, k+2, \ldots\}$ for some $k \in \mathbb{N}$, then we may write $(a_n)_{n=k}^{\infty}$, and we say that the sequence *starts with index* k. If the index set has the form $\{i, i+1, i+2, \ldots, j\}$ for some $i, j \in \mathbb{N}$ with $i \leq j$, then we may write $(a_n)_{n=i}^{j}$, and we say that this is a *finite sequence*. In most cases, our sequences will take one of these forms.

Example 9.2:

We recast our examples from the beginning of this chapter with this new notation. In the first example, $(I_n)_{n=0}^{\infty}$ is an infinite sequence of closed bounded intervals (so the set S can be taken to be the collection of subsets of $[a, b]$). In the second example, we have sequences $(s_n)_{n=1}^{\infty}$ and $(t_n)_{n=1}^{\infty}$ of step functions, as well as sequences

$$\left(\int_0^1 s_n\right)_{n=1}^{\infty} \text{ and } \left(\int_0^1 t_n\right)_{n=1}^{\infty}$$

of real numbers. In the third example, we have sequences of functions $(t_i)_{i=0}^{\infty}$ and $(T_n f)_{n=0}^{\infty}$, though for any fixed $x \in \mathrm{dom}(f)$, $(t_i(x;a))_{i=0}^{\infty}$ and $(T_n f(x;a))_{n=0}^{\infty}$ are real-valued sequences.

As another simple example, the real-valued sequence $(1)_{n=0}^{\infty}$ has every term equal to 1. This is an infinite sequence with a one-element range. $\square$

Remark. There are many different popular notations associated with sequences. For instance, some books write $\{a_n\}_{n=0}^{\infty}$ or $\langle a_n\rangle_{n=0}^{\infty}$ instead of $(a_n)_{n=0}^{\infty}$. We will use parentheses to denote sequences in this book, so that curly braces can be used in connection with sets, i.e. (a_n) is the sequence whereas $\{a_n\}$ is the range.

Also, when working with a sequence $(a_n)_{n=0}^{\infty}$, mathematicians tend to forget about the underlying function $a : \mathbb{N} \to \mathbb{R}$. As a result, the variable a is considered free. For example, we can write phrases like "if $a_n = 2^{-n}$, and $a = 0$, then a_n approaches a as n gets large".

Useful Properties of Sequences

There are many useful properties of functions that are also used with sequences. For instance, a sequence $(a_n)_{n\in\mathbb{N}}$ is *increasing* if $i < j$ implies $a_i \le a_j$ for all $i, j \in \mathbb{N}$. (If $i < j$ implies $a_i < a_j$ for all $i, j \in \mathbb{N}$, we instead say the sequence is *strictly increasing*.) The definition of a decreasing or strictly decreasing sequence is similar. Likewise, you should be able to formulate definitions for a sequence to be *bounded (above or below)*.

Intuitively, to check whether a sequence (a_n) is increasing, we only need to verify whether each term in a sequence is below the next term. In other words, we should show $a_0 \le a_1$, $a_1 \le a_2$, and so on. More formally, you should prove this result in Exercise 9.2.1:

Theorem 9.3. *Let $(a_n)_{n=0}^{\infty}$ be an infinite real-valued sequence. (a_n) is increasing iff $a_n \le a_{n+1}$ for all $n \in \mathbb{N}$. Similar statements hold for strictly increasing, decreasing, and strictly decreasing sequences as well.*

We put this theorem to good use with a couple examples.

Example 9.4:

Consider the sequence defined by

$$a_n = \frac{n}{n^2+1}$$

for each $n \in \mathbb{N}^*$. Writing out a few terms, we find that $a_1 = 1/2$, $a_2 = 2/5$, $a_3 = 3/10$, and so on. From this, we suspect that (a_n) is strictly decreasing. (It is easy to check that this sequence is bounded below by 0. Once we prove that the sequence is decreasing, it follows it is also bounded above by a_1.)

We have several different ways of showing that (a_n) is strictly decreasing. One way is to show directly that for any $n \in \mathbb{N}^*$, $a_n > a_{n+1}$. We do some algebra with inequalities to find that

$$\begin{aligned}
& a_n > a_{n+1} \\
\leftrightarrow \quad & \frac{n}{n^2+1} > \frac{n+1}{(n+1)^2+1} \\
\leftrightarrow \quad & n((n+1)^2+1) > (n+1)(n^2+1) \\
\leftrightarrow \quad & n^3+2n^2+2n > n^3+n^2+n+1 \\
\leftrightarrow \quad & n^2+n > 1
\end{aligned}$$

which is readily seen to be true for any $n \geq 1$.

For a different approach for showing that (a_n) is strictly decreasing, we recall that a real function whose derivative is always negative on an interval is strictly decreasing. (a_n) is not a function with an interval for a domain (as the index set is $\mathbb{N}^*$, which contains no nonempty open intervals), so we can't compute the derivative of (a_n) directly. However, we can make a corresponding real function

$$f(x) = \frac{x}{x^2+1} \text{ for all } x > 0$$

so that $f(n) = a_n$ for each $n \in \mathbb{N}^*$. Now, f is differentiable at every positive x, yielding

$$f'(x) = \frac{(1)(x^2+1)-(x)(2x)}{(x^2+1)^2} = \frac{1-x^2}{(1+x^2)^2}$$

Since the denominator is always positive, and the numerator is negative when $x > 1$, this proves that f is strictly decreasing on $[1, \infty)$. It follows that (a_n) is strictly decreasing as well. $\square$

Example 9.5:

Consider the following *recursive definition* (this means the same thing as inductive definition) for a sequence:

$$a_0 = 1 \qquad a_{n+1} = 2 - \frac{1}{2a_n} \text{ for any } n \in \mathbb{N}$$

For any $k \in \mathbb{N}^*$, this formula computes a_k in k steps, by first computing a_1 from a_0 (using $n = 0$ in the definition), then a_2 from a_1 (using $n = 1$ in the definition), and so on until a_k is computed from a_{k-1}. The first few terms of this sequence are

$$1, \frac{3}{2}, \frac{5}{3}, \frac{17}{10}, \ldots$$

Unlike the previous example, it is not clear how the values a_n of this sequence can be obtained by plugging n into a simple real function f. (In other words, we say that we do not have a *closed form* for a_n.) Nevertheless, we can still reason about this sequence by using induction, since the definition is well-suited for aiding the inductive step of a proof.

As a first claim about this sequence, we claim that a_n is never zero. This is important, because if $a_n = 0$, then a_{n+1} involves division by zero and is not well-defined. We can try to prove $a_n \neq 0$ by induction on $n \in \mathbb{N}$. The base case is obvious, but when we do the inductive step, we find that

$$a_{n+1} \neq 0 \leftrightarrow 2 - \frac{1}{2a_n} \neq 0 \leftrightarrow a_n \neq \frac{1}{4}$$

At this point, we are stuck, because we can't use our inductive hypothesis of $a_n \neq 0$ to conclude $a_n \neq \frac{1}{4}$.

To get around this difficulty, we surprisingly find it easier to prove a STRONGER property by induction[1]. The first few terms of this sequence suggest that (a_n) is strictly increasing. Thus, we will prove simultaneously that $a_n \neq 0$ and $a_n < a_{n+1}$ to use Theorem 9.3. Note that once we know that (a_n) is strictly increasing, it follows that

$$a_n > a_{n-1} > \cdots > a_0 = 1 > 0$$

for all $n \in \mathbb{N}$, so $a_n \neq 0$.

[1]Occasionally, it is easier to prove a stronger property than to prove a weaker one. This phenomenon is sometimes called the *Inventor's Paradox*.

Thus, we aim to prove $P(n)$ by induction on $n \in \mathbb{N}$, where $P(n)$ is "$0 < a_n < a_{n+1}$". For the base case of $n = 0$, we find $a_0 = 1$ and $a_1 = 3/2 > 1$, so the base case is done. Now let $n \in \mathbb{N}$ be given, and suppose $0 < a_n < a_{n+1}$ as an inductive hypothesis. We wish to prove that $0 < a_{n+1} < a_{n+2}$ in our inductive step.

We already know that $a_{n+1} > 0$, so it remains to show $a_{n+2} > a_{n+1}$. Because a_{n+1} is positive, a_{n+2} is well-defined. We start from our hypothesis of $a_n < a_{n+1}$ and perform algebra on both sides, imitating the steps of our definition of (a_n):

$$\begin{aligned} & a_n < a_{n+1} \\ \leftrightarrow\quad & 2a_n < 2a_{n+1} \\ \leftrightarrow\quad & \frac{1}{2a_n} > \frac{1}{2a_{n+1}} \\ \leftrightarrow\quad & 2 - \frac{1}{2a_n} < 2 - \frac{1}{2a_{n+1}} \\ \leftrightarrow\quad & a_{n+1} < a_{n+2} \end{aligned}$$

(Note that the third line, which takes reciprocals, uses our hypothesis that a_n and a_{n+1} are positive.) This finishes the inductive step. Hence, now we know that (a_n) is well-defined. As a bonus, we also know that (a_n) is strictly increasing. $\square$

Remark. The work from the previous example can be made a little more efficient by introducing a real function to help with the inductive step. Suppose we define $f : \mathbb{R} - \{0\} \to \mathbb{R}$ by

$$f(x) = 2 - \frac{1}{2x} \text{ for all } x \neq 0$$

Thus, $a_{n+1} = f(a_n)$ for all $n \in \mathbb{N}$. We say that (a_n) is obtained by starting with $a_0 = 1$ and *iterating* f, so $a_n = (f \circ f \circ \cdots \circ f)(a_0)$, where we have composed f with itself n times.

It is not hard to check, by similar steps to the previous induction proof, that f is strictly increasing. (Alternately, you can prove $f'(x) > 0$ for all $x \neq 0$.) It follows that $a_n < a_{n+1}$ implies $f(a_n) < f(a_{n+1})$, i.e. $a_{n+1} < a_{n+2}$. Hence, to finish the proof that $0 < a_n < a_{n+1}$ for all $n \in \mathbb{N}$, it remains to show the base case, i.e. that $0 < a_0 < a_1 = f(a_0)$.

(Note that if a_0 were chosen differently so that $a_0 > f(a_0)$, then you would be able to show that $a_n > a_{n+1}$ for all n such that a_n and a_{n+1} are defined. Thus, this new (a_n) is strictly decreasing, provided that it is well-defined. For more on this, see Exercise 9.2.18.)

Limits of Sequences

Apart from questions of whether a sequence increases or is bounded, the next most natural question to ask about a sequence is probably the following: does the sequence tend towards any finite limit as the index goes to ∞? Up to this point, we have dealt with limits of real functions, where the notation

$$\lim_{x\to\infty} f(x)$$

requires f to be defined on some unbounded interval (a,∞) (where $a \in \mathbb{R}$). However, our sequences have only been defined for natural numbers. Thus, we introduce a definition of a sequence limit, which is meant to be very similar to the usual limit definition in most respects:

Definition 9.6. Let $L \in \mathbb{R}$ and $k \in \mathbb{N}$ be given, and suppose $(a_n)_{n=k}^{\infty}$ is a real-valued sequence. We say that (a_n) *has limit L as n approaches* ∞ (alternately, we say that (a_n) *converges to* L), written as

$$\lim_{n\to\infty} a_n = L$$

or as "$a_n \to L$ as $n \to \infty$", to mean that for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ (which depends on ϵ) such that $N \geq k$ and

$$\forall n \in \mathbb{N} \ \ (n > N \to |a_n - L| < \epsilon)$$

In other words, no matter how close we want the sequence to be to L, a_n is close enough for all "sufficiently large" values of n. N represents a "cutoff" such all that larger values of n are guaranteed to yield terms a_n which are sufficiently close to L.

If no such L satisfying this definition exists, we say that (a_n) *has no limit as* $n \to \infty$ (or that (a_n) *diverges*).

As a simple example, we can prove that $1/n \to 0$ as $n \to \infty$. To see this, let $\epsilon > 0$ be given. For any $n \in \mathbb{N}^*$, we have

$$\left|\frac{1}{n} - 0\right| < \epsilon \leftrightarrow n > \frac{1}{\epsilon}$$

and therefore the definition of the limit is satisfied when we choose $N = \lceil 1/\epsilon \rceil$. Note that only very minor changes to this proof are needed to also show that $1/x \to 0$ as $x \to \infty$, where x is a REAL number tending to ∞. This suggests the following result:

Theorem 9.7. *Let $a \in \mathbb{R}^+$ and $L \in \mathbb{R}$ be given, and suppose that f is a real function defined on (a, ∞) with*

$$\lim_{x\to\infty} f(x) = L$$

where x is a real variable. Suppose the sequence $(a_n)_{n=\lceil a \rceil}^{\infty}$ is defined by $a_n = f(n)$ for $n \in \mathbb{N}$ with $n \geq a$. Then

$$\lim_{n\to\infty} a_n = L$$

in the sense of Definition 9.6.

Strategy. Intuitively, if $f(x) \to L$ as $x \to \infty$, then $f(x)$ can be made as close to L as desired by taking x past some cutoff. The same cutoff, or any higher cutoff, should still work in the special case where x is a natural number.

Proof. Let $a, L, f, (a_n)$ be given as described. Let $\epsilon > 0$ be given. We need to find $N \in \mathbb{N}$ with $N \geq \lceil a \rceil$ such that

$$\forall n \in \mathbb{N} \;\; (n > N \to |a_n - L| < \epsilon)$$

Because we know $f(x) \to L$ as $x \to a$, there is some $M \in \mathbb{R}$ such that

$$\forall x \in (a, \infty) \;\; (x > M \to |f(x) - L| < \epsilon)$$

(In fact, WLOG we may assume $M \geq a$, since otherwise we may replace M with $\max\{M, a\}$.) Let $N = \lceil M \rceil$. Thus, whenever $n \in \mathbb{N}$ and $n > N$, we have $n > M \geq a$ and $n \geq \lceil a \rceil$. It follows that $|f(n) - L| < \epsilon$ and $f(n) = a_n$, so $|a_n - L| < \epsilon$, as desired. ■

This theorem essentially says that if an infinite sequence is given by a real function which has a limit, then the sequence also has that same limit. This allows us to reuse lots of work that we have spent on limits in previous chapters. Also, this theorem is useful because certain techniques for limit computation, such as L'Hôpital's Rule, are only valid on functions defined

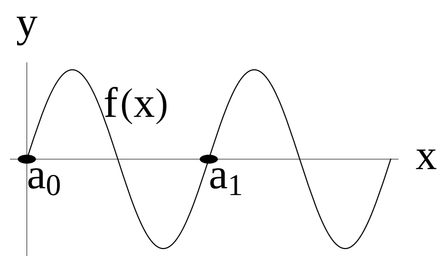

Figure 9.1: The graphs of $f(x) = \sin(2\pi x)$ (curve) and $a_n = \sin(2\pi n)$ (solid dots)

on intervals. (A sequence is not differentiable, so you can't use L'Hôpital's Rule with it!)

However, it is important to note that Theorem 9.7 does not have an "iff" in it. The theorem says that IF $f(x)$ has a limit, THEN $a_n = f(n)$ has the same limit. The converse is false: consider

$$f(x) = \sin(2\pi x) \text{ for all } x \in \mathbb{R}$$

When $n \in \mathbb{N}$, we have $a_n = f(n) = \sin(2\pi n) = 0$, so $a_n \to 0$ as $n \to \infty$. However, $f(x)$ has no limit as $x \to \infty$, since the graph of f oscillates. See Figure 9.1, where the solid dots represent the sequence (a_n) and the curved graph represents the graph of $f(x)$. The issue is that the sequence limit only looks at the graph "every so often", so the sequence does not detect the function's oscillation.

Remark. As a result of this example, it is important to realize that real-valued limits and sequence limits for a variable approaching ∞ are not the same thing. However, traditionally, the same notation with the lim symbol is used for each. We will adopt the custom that if we write a limit with a variable such as n, k, i, or some other letter near the middle of the alphabet, then we are using a sequence limit. If we use a variable near the end of the alphabet, such as x, t, or u, then we are using a real-valued limit. Hopefully, the context will also make it clear which type of limit is being used (and sometimes, by Theorem 9.7, the specific type of limit doesn't make a difference!).

Although not all our theorems for real-valued limits work for sequences, most of them do. You can prove that the usual limit laws, such as laws for sums, products and quotients, all work for sequences with nearly the same proofs as their real-valued counterparts. There is also a definition of an infinite limit for sequences: we say $a_n \to \infty$ as $n \to \infty$ if for all $M \in \mathbb{R}$, there exists $N \in \mathbb{N}$ (which depends on M) such that

$$\forall n \in \mathbb{N} \ (n > N \to a_n > M)$$

(The intuition is that the sequence values become arbitrarily high; for any fixed y-value cutoff M, we can always make a_n larger than M by taking big enough values of n.) The definition of "$a_n \to -\infty$ as $n \to \infty$" is similar. With this definition, we can state and prove a couple variants of the Composition Limit Theorem for sequences: see Exercises 9.2.16 and 9.2.17.

One other particularly important limit law which is provable for sequences is the Squeeze Theorem. This can be used to obtain limits in some important cases where other laws fail to apply, as the following example shows:

Example 9.8:

In this example, we present an important property about factorials. Previously, we have seen that factorials dominate exponentials: Exercise 8.5.14 shows that for any fixed $b > 1$, $b^n/n! \to 0$ as $n \to \infty$. This tells us that factorials grow very rapidly for large values of n. Now, we show that the sequence (n^n) grows even faster: in fact, $n! = o(n^n)$ as $n \to \infty$, i.e.

$$\lim_{n\to\infty} \frac{n!}{n^n} = 0$$

Because factorials are recursively defined ($0! = 1$ and $(n+1)! = (n+1) \cdot n!$), most of the limit laws don't apply to them. However, we note that we can write our fraction as

$$\frac{n!}{n^n} = \frac{n \cdot (n-1) \cdot \dots \cdot 1}{n \cdot n \cdot \dots \cdot n} = 1 \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{1}{n} \le \frac{1}{n}$$

where we bound all the fractions except the last by 1. Because $1/n \to 0$ as $n \to \infty$, and $n!/n^n > 0$ for all $n \in \mathbb{N}$, the Squeeze Theorem proves $n!/n^n \to 0$ as $n \to \infty$.

In fact, this tactic can be used to prove an even stronger property: $n!/n^n$ goes to 0 more quickly than ANY fixed power of n.[2] To see this, for any fixed $k \in \mathbb{N}$, whenever $n \ge k$ we have

$$\frac{n!}{n^n} = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{k}{n} \cdot \frac{k-1}{n} \cdot \dots \cdot \frac{1}{n} \le \frac{k!}{n^k}$$

[2] In fact, a result called Stirling's Formula shows that $n!$ is asymptotic to $(n/e)^n\sqrt{2\pi n}$, meaning that their ratio approaches 1 as $n \to \infty$.

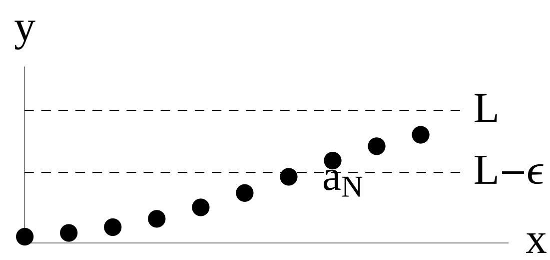

Figure 9.2: An illustration of the BMCT

where we bound the first $n-k$ fractions by 1. (Although $k!$ grows faster than n^k as a function of k, recall that k is a constant here.) This stronger estimate will come in handy in a later section when we look at series. □

The BMCT

In general, it can be quite difficult to tell whether a sequence converges to a limit, especially if the sequence is recursively-defined (which makes it harder to use limit techniques that only apply to closed-form real functions). However, the situation becomes much simpler when we look at a monotone sequence, one which is either always increasing or always decreasing. To see why this is the case, recall that monotone real functions have a nice property: Theorem 3.48 proves that one-sided limits always exist for monotone functions. The main idea of that theorem's proof leads us to the following very important theorem, the Bounded Monotone Convergence Theorem (or BMCT):

Theorem 9.9 (Bounded Monotone Convergence Theorem). *Let (a_n) be an infinite real sequence, and suppose that (a_n) is monotone. Then (a_n) converges (i.e. a_n has a finite limit as $n \to \infty$) iff (a_n) is bounded.*

Strategy. WLOG, we may focus on the case where (a_n) is increasing. The main idea is presented in Figure 9.2. If (a_n) is unbounded, then the terms rise arbitrarily high and go to ∞. Otherwise, any upper bound on (a_n) places a "ceiling" on the terms above which the dots representing (a_n) cannot go. Out of all such bounds, the one that does the best job at staying close to the sequence, i.e. the "tightest ceiling", is the least upper bound, $L = \sup\{a_n\}$.

Now, we want to show that (a_n) converges to L. Thus, for any $\epsilon > 0$, we eventually want the terms to stay between $L-\epsilon$ and $L+\epsilon$. Actually, we just need SOME term to be greater than $L - \epsilon$, because all terms afterward will also be greater than $L - \epsilon$ as (a_n) is increasing.

This is where we use the fact that L is the tightest bound. As $L = \sup\{a_n\}$, $L - \epsilon$ is NOT an upper bound of $\{a_n\}$, so there must be some $N \in \mathbb{N}$ such that $a_N > L - \epsilon$. Hence, for all $n > N$, $a_n \geq a_N > L - \epsilon$.

Proof. Let (a_n) be given as described. WLOG, we may assume (a_n) is increasing, because if (a_n) is decreasing, then $(-a_n)$ is increasing and we can apply this proof to $(-a_n)$.

For one direction, suppose that (a_n) is unbounded. Since (a_n) is increasing, it must hence be unbounded above, so for any $M \in \mathbb{R}$, there is some $N \in \mathbb{N}$ with $a_N > M$. (Otherwise, M would be an upper bound.) Because (a_n) is increasing, for all $n \in \mathbb{N}$ with $n > N$, we have $a_n \geq a_N > M$. This means that $a_n \to \infty$ as $n \to \infty$, so (a_n) does not have a finite limit.

For the other direction, suppose that (a_n) is bounded. Therefore, $\sup\{a_n\}$ exists; let's call it L. We claim that $a_n \to L$ as $n \to \infty$.

To prove this, let $\epsilon > 0$ be given. Because $L - \epsilon < L$, and L is the least upper bound of $\{a_n\}$, $L - \epsilon$ is not an upper bound of $\{a_n\}$, so there is some $N \in \mathbb{N}$ with $a_N > L - \epsilon$. Hence, for all $n \in \mathbb{N}$ with $n > N$, $a_n \geq a_N > L - \epsilon$. However, we also have $a_n \leq L$ because L is an upper bound. Putting these conclusions together, we obtain $|a_n - L| < \epsilon$ for all $n > N$, as desired. ∎

The BMCT is often very useful for recursive sequences, because monotonicity and boundedness can be checked by induction, whereas solving for N in terms of ϵ in the limit definition is hard. In fact, in many recursive sequences, it is not hard to find what the limit should be (i.e. determining a candidate for the limit); rather, the difficulty comes from showing that this candidate really is the limit. In these situations, the BMCT can be useful, as we demonstrate with a couple examples.

Example 9.10:

Let's return to the sequence (a_n) from Example 9.5:

$$a_0 = 1 \qquad a_{n+1} = 2 - \frac{1}{2a_n} \text{ for any } n \in \mathbb{N}$$

We previously proved that (a_n) is well-defined and strictly increasing. Now we'd like to know: does the sequence converge? And if so, what is the limit?

To prove convergence, the BMCT tells us that it only remains to check that (a_n) is bounded. The recursive definition suggests that 2 is an upper bound of (a_n). In fact, since a_n is always positive, it follows that

$$a_{n+1} = 2 - \frac{1}{2a_n} < 2$$

for all $n \in \mathbb{N}$. Combined with the fact that $a_0 = 1 < 2$, we have shown that 2 is an upper bound for (a_n). As (a_n) is increasing, $a_0 = 1$ is a lower bound.

Thus, we now know that $a_n \to L$ as $n \to \infty$ for some $L \in \mathbb{R}$. (In fact, we also know that $1 \le L \le 2$ because L is the least upper bound of $\{a_n\}$.) How do we find L? The key tactic here is to take a limit of both sides of the recursive definition, because a_n and a_{n+1} both approach L as $n \to \infty$. (This is justified by a variant of the Composition Limit Theorem, since $n+1 \to \infty$ as $n \to \infty$.) This tells us that

$$\begin{aligned}
\lim_{n\to\infty} a_{n+1} &= \lim_{n\to\infty} \left(2 - \frac{1}{2a_n}\right) \\
L &= 2 - \frac{1}{2L} \\
2L^2 &= 4L - 1 \\
2L^2 - 4L + 1 &= 0 \\
L &= \frac{4 \pm \sqrt{8}}{4} = 1 \pm \frac{\sqrt{2}}{2}
\end{aligned}$$

Note that this gives us two candidates for the limit. However, because the limit must be in $[1, 2]$, the minus sign does not make sense. Thus, the limit is $1 + (\sqrt{2})/2$. $\square$

Remark. Suppose we take the sequence from the previous example and apply the definition a few times. Thus, we write a_{n+1} in terms of a_n,

$$a_{n+1} = 2 - \frac{1}{2a_n}$$

then we replace a_n with an expression in terms of a_{n-1},

$$\begin{aligned}
a_{n+1} &= 2 - \frac{1}{2\left(2 - \frac{1}{2a_{n-1}}\right)} \\
&= 2 - \frac{1}{4 - \frac{1}{a_{n-1}}}
\end{aligned}$$

and so forth:

$$a_{n+1} \quad = \quad 2 - \cfrac{1}{4 - \frac{1}{2 - \frac{1}{2a_{n-2}}}}$$

(This is called *unrolling the recurrence.*) If we were to unroll infinitely many times, then we would obtain an expression like

$$2 - \cfrac{1}{4 - \frac{1}{2 - \frac{1}{4 - \frac{1}{2 - \ddots}}}}$$

where the left sides of the subtractions alternate between 2 and 4. This kind of object, involving a sequence of nested fractions, is called a *continued fraction*. To find the value of a continued fraction, the continued fraction is expressed as the limit of a recursive sequence, and the limit is proven to exist by using the BMCT or a similar tool, much like we did with the previous example. Continued fractions play a major role in several areas of mathematics, such as number theory and numerical analysis.

In Exercise 9.2.22, you will be able to develop properties of another continued fraction:

$$1 + \cfrac{1}{1 + \frac{1}{1 + \ddots}}$$

Example 9.11:

In this example, we use recursive sequences to obtain a well-defined value for the expression

$$\sqrt{3 + \sqrt{3 + \sqrt{3 + \cdots}}}$$

To approach this, let's say that for each $n \in \mathbb{N}^*$, a_n is the expression where the square roots go n levels deep. Thus, we have the recurrence

$$a_1 = \sqrt{3} \qquad a_{n+1} = \sqrt{3 + a_n} \text{ for all } n \in \mathbb{N}^*$$

It is plausible to say that the value of our "infinite expression" should be the limit of this sequence (a_n).

Before we establish that the sequence has a limit in the first place, let's first address the question: what would that limit have to be? (After all, we are most interested in the value of that infinite expression.) Let's say that $a_n \to L$ as $n \to \infty$. Thus, $a_{n+1} \to L$ also, so we find

$$\begin{aligned}
\lim_{n\to\infty} a_{n+1} &= \lim_{n\to\infty} \sqrt{3+a_n} \\
L &= \sqrt{3+L} \\
L^2 &= L+3 \\
L^2 - L - 3 &= 0 \\
L &= \frac{1 \pm \sqrt{13}}{2}
\end{aligned}$$

Since we would expect our limit to be positive, we should use the plus sign, so $L = (1/2)(1+\sqrt{13})$.

Thus, if our (a_n) sequence has a limit, then that limit must be L. To show that (a_n) is convergent, we will use the BMCT. To make our work a little more straightforward, it is helpful to note that our definition of (a_n) is obtained by iterating the function

$$f(x) = \sqrt{3+x} \text{ for all } x > 0$$

starting with the value $\sqrt{3}$. In other words, $a_{n+1} = f(a_n)$ for all $n \in \mathbb{N}$. Note that we found $L = \sqrt{3+L}$ above, i.e. $L = f(L)$, so L is a *fixed point* of f.

Now, we prove monotonicity. Note that

$$a_1 = \sqrt{3+\sqrt{3}} > \sqrt{3} = a_0$$

and that f is strictly increasing. These facts yield a straightforward proof by induction that $a_n < a_{n+1}$ for all $n \in \mathbb{N}$ (fill in the details). Thus, (a_n) is strictly increasing.

Showing that (a_n) is bounded is a little trickier, because we have to first guess the bound. However, here's where our knowledge that $L = f(L)$ comes in handy. Because f is strictly increasing, whenever $a_n < L$, we have

$$a_{n+1} = f(a_n) < f(L) = L$$

This suggests that L is an upper bound of (a_n). (This is why it was useful to compute L first before proceeding with the BMCT!) You can check that $a_0 < L$, and hence a simple proof by induction shows that $a_n < L$ for all $n \in \mathbb{N}$. This proves that (a_n) is bounded above by L. Since (a_n) is also bounded below (by a_0), we know that (a_n) converges by the BMCT. $\square$

9.2 Exercises

1. Prove Theorem 9.3. (Hint: Suppose that $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$. Prove that $a_n \leq a_{n+m}$ for all $n, m \in \mathbb{N}$ by induction on m.)

In Exercises 2 through 13, a formula defining a real-valued sequence $(a_n)_{n=1}^{\infty}$ is given. For that sequence, determine whether it is increasing, decreasing, both, or neither. Furthermore, determine if (a_n) is bounded below or above, and provide appropriate bounds when they exist.

2. $a_n = \cos(n\pi + \pi/2)$
3. $a_n = \dfrac{n}{2^n}$
4. $a_n = \sqrt{n+1} - \sqrt{n}$
5. $a_n = \dfrac{n!}{2^n}$
6. $a_n = \dfrac{\sin n}{n}$
7. $a_n = \dfrac{n+3}{n-1/2}$
8. $a_n = \dfrac{e^{1/n} - 1}{1/n}$

   (Hint: A Taylor estimate shows $e^{1/n} - 1 \approx 1/n$. Use the integral form of the approximation's error; it tells you more about the monotonicity than the Lagrange form does.)

9. $a_1 = \sqrt{2}$

   $a_{n+1} = \sqrt{2a_n}$ for $n \in \mathbb{N}^*$
10. $a_1 = 2$

    $a_{n+1} = (a_n)^{1/n}$ for $n \in \mathbb{N}^*$
11. $a_1 = 0$

    $a_{n+1} = \dfrac{1}{2}(a_n + 5)$ for $n \in \mathbb{N}^*$
12. $a_1 = 1/2$

    $a_{n+1} = \dfrac{1}{1 - a_n}$ for $n \in \mathbb{N}^*$
13. $a_1 = 1/3$

    $a_{n+1} = \dfrac{1}{1 - a_n}$ for $n \in \mathbb{N}^*$
14. The last two exercises suggest an interesting pattern. Define $f : \mathbb{R} - \{1\} \to \mathbb{R}$ by

$$f(x) = \frac{1}{1-x}$$

for all $x \neq 1$.

(a) Show that $f'(x) > 0$ for all $x \neq 1$, but f is not an increasing function on $\mathbb{R} - \{1\}$. Why doesn't this contradict Theorem 4.47?

(b) Show that whenever $x \in \mathbb{R} - \{0, 1\}$, $f(f(f(x))) = x$.

(c) If $(a_n)_{n=1}^{\infty}$ is a sequence obtained by iterating f, i.e. $a_{n+1} = f(a_n)$ for all $n \in \mathbb{N}^*$, then which values of a_1 cause the sequence to converge?

15. For each of the sequences (a_n) from Exercises 9 through 13, determine $\lim_{n\to\infty} a_n$ if it exists. (Use the BMCT.)

16. Suppose $(a_n)_{n=0}^{\infty}$ is a real-valued sequence and $(n_k)_{k=0}^{\infty}$ is a strictly increasing sequence from $\mathbb{N}$. The sequence $(a_{n_k})_{k=0}^{\infty}$ is called a *subsequence of* (a_n) (note that if a is the function underlying (a_n) and n is the function underlying (n_k), then $a \circ n$ is the function underlying (a_{n_k})). Prove the following:

    Theorem 9.12. *Let $(a_n)_{n=0}^{\infty}$ be a real-valued sequence. If $a_n \to L$ as $n \to \infty$, where L is either a real number, ∞, or $-\infty$, then every subsequence of (a_n) also converges to L. In other words, $a_{n_k} \to L$ as $k \to \infty$ for every strictly increasing sequence $(n_k)_{k=0}^{\infty}$ from $\mathbb{N}$.*

    Remark. This theorem implies that if $a_n \to L$, then $a_{n+1} \to L$ as well as $n \to \infty$ (by choosing $n_k = k+1$ for each $k \in \mathbb{N}$). However, note that a divergent sequence may have convergent subsequences! For instance, if $a_n = (-1)^n$, then the subsequence (a_{2n}) is constantly 1, but (a_n) has no limit.

17. (a) Prove the following variant of the Composition Limit Theorem:

    Theorem 9.13. *Let $L \in \mathbb{R}$ be given, and let $(a_n)_{n=0}^{\infty}$ be a real sequence with $a_n \to L$ as $n \to \infty$. Suppose also that f is a real function such that $f(a_n)$ is defined for all $n \in \mathbb{N}$ and f is continuous at L. Then $f(a_n) \to f(L)$ as $n \to \infty$.*[3]

[3] It turns out that the converse of this theorem is true: namely, if f has the property that $f(a_n) \to f(L)$ whenever $a_n \to L$, then f is continuous at L. We will not have use for this result (though the proof is a good exercise in using a contradiction argument), but this result shows that continuity could have been defined in terms of sequences back in Chapter 3.

(b) Prove this modification of Theorem 9.13:

Theorem 9.14. *Let $(a_n)_{n=0}^{\infty}$ be a real sequence with $a_n \to L$ as $n \to \infty$, where L is ∞ or $-\infty$. Suppose that f is a real function such that $f(a_n)$ is defined for all $n \in \mathbb{N}$ and $f(x) \to M$ as $x \to L$, where M is either a real number, ∞, or $-\infty$. Then $f(a_n) \to M$ as $n \to \infty$.*

18. Suppose that f is the function from Example 9.5:

$$f(x) = 2 - \frac{1}{2x} \text{ for all } x \neq 0$$

Note that f is strictly increasing from $\mathbb{R} - \{0\}$ to $\mathbb{R}$.

Suppose we define $(a_n)_{n=0}^{\infty}$ by starting with $a_0 = 2$ and iterating f, i.e. $a_{n+1} = f(a_n)$ for all $n \in \mathbb{N}$. (In the example, we used $a_0 = 1$ instead.) Prove that a_n is well-defined for each $n \in \mathbb{N}$ and (a_n) is strictly decreasing. What is the limit of (a_n) as $n \to \infty$?

19. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = e^x - 1$$

Note that f is strictly increasing on $\mathbb{R}$.

(a) Define $(a_n)_{n=0}^{\infty}$ by starting with $a_0 = -1/2$ and then iterating f, i.e. $a_{n+1} = f(a_n)$ for all $n \in \mathbb{N}$. Prove that (a_n) is strictly increasing and bounded. What is its limit?

(b) Now, suppose $a_0 = 1/2$ and $a_{n+1} = f(a_n)$ for all $n \in \mathbb{N}$. Prove that (a_n) is strictly increasing. Why does it follow that (a_n) diverges, and hence (a_n) must be unbounded?[4]

20. A sequence $(a_n)_{n=0}^{\infty}$ defined by the equation

$$a_0 = x \qquad a_{n+1} = ma_n + b \text{ for all } n \in \mathbb{N}$$

where $m, b, x \in \mathbb{R}$, is said to be obtained by a *linear recurrence* (the sequence is obtained by iterating a linear function). In this exercise, we come up with a condition for when linear recurrences converge. To simplify matters, we assume $m \neq 1$.

[4]This exercise shows that the starting point of a sequence can have a large effect on the behavior. For some especially fascinating applications of this, look up chaos theory and fractals. There are also applications in biology concerning the life cycles of some organisms.

(a) If (a_n) converges, then what must the limit be? Call that value L.

(b) Now define $b_n = a_n - L$ for all $n \in \mathbb{N}$, using your value of L from part (a). (This is a pretty common trick with recursive sequences: make a new sequence which measures displacement from the candidate limit.) What is b_{n+1} in terms of b_n?

(c) Use your answer from part (b) to obtain a closed form for a_n for each $n \in \mathbb{N}$. In other words, obtain a formula for a_n which does not use summation symbols $\sum$, product symbols $\prod$, or earlier terms in the sequence.

(d) Based on part (c), come up with a condition in terms of x, m, and b that describes exactly when (a_n) converges. Also, what happens when $m = 1$?

21. Suppose that $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=0}^{\infty}$ are real-valued sequences. We define $(c_n)_{n=0}^{\infty}$ by

$$c_{2k} = a_k \qquad c_{2k+1} = b_k$$

for all $k \in \mathbb{N}$. Thus, (c_n)'s first few terms are

$$a_0, b_0, a_1, b_1, a_2, b_2, \ldots$$

We say that (c_n) is obtained by *interleaving* (a_n) and (b_n).

Prove that if (a_n) and (b_n) both converge to the same limit, then (c_n) also converges to that limit.

22. In this exercise, we will prove the following:

$$1 + \cfrac{1}{1 + \cfrac{1}{1 + \ddots}} = \frac{1 + \sqrt{5}}{2}$$

This value is called the *golden ratio*, often denoted by ϕ, and it appears in many natural processes.

The continued fraction on the left looks like it can be obtained by starting with 1 and applying the function

$$f(x) = 1 + \frac{1}{x} \text{ for all } x > 0$$

infinitely many times. More formally, let $(a_n)_{n=0}^{\infty}$ be the sequence obtained by starting with $a_0 = 1$ and iterating f, i.e. $a_{n+1} = f(a_n)$ for all $n \in \mathbb{N}$. Thus, the first few terms are

$$a_0 = 1, a_1 = 1 + \frac{1}{1}, a_2 = 1 + \frac{1}{1 + \frac{1}{1}}, a_3 = 1 + \frac{1}{1 + \frac{1}{1+\frac{1}{1}}}$$

We define the value of our continued fraction to be the limit of (a_n), provided this limit exists.

(a) Prove that if (a_n) converges to a limit L, then $L = f(L)$. Use this to show that
$$L = \frac{1 \pm \sqrt{5}}{2}$$
(Note that L must use the plus sign and not the minus sign, since $f : \mathbb{R}^+ \to \mathbb{R}^+$.)

(b) Note that f is strictly decreasing on $\mathbb{R}^+$. Use this to prove that
$$a_0 < a_2 < a_4 < \cdots$$
i.e. the subsequence $(a_{2k})_{k=0}^{\infty}$ is strictly increasing. (The term "subsequence" was defined in Exercise 9.2.16.)

(c) Similarly, prove that
$$a_1 > a_3 > a_5 > \cdots$$
i.e. the subsequence $(a_{2k+1})_{k=0}^{\infty}$ is strictly decreasing.

(d) Use the previous parts to prove that for all $n \in \mathbb{N}$, $a_0 \le a_n \le a_1$. Thus, (a_n) is bounded.

(e) Although (a_n) is NOT monotone, parts (b) and (c) show that the subsequences (a_{2k}) and (a_{2k+1}) are monotone. Thus, the BMCT implies that they converge. Prove that they each converge to $L = (1 + \sqrt{5})/2$.

(f) Finally, use the result of Exercise 9.2.21 to prove that $a_n \to L$ as $n \to \infty$.

23. Prove that if $(a_n)_{n=0}^{\infty}$ is any convergent real sequence, i.e. (a_n) has a finite limit, then (a_n) is bounded. (Hint: Show that all but finitely many terms are close to the limit, and use the Triangle Inequality.)

9.3 Basic Results About Series

Now that we have built up some useful information about sequences, we can formally introduce the main topic for the rest of this chapter: infinite series. As mentioned in this chapter's introduction, an infinite series gives us a way to talk about a sum of infinitely many terms. However, we should first ask: what should an "infinite sum" mean? Is it possible to add infinitely many positive numbers and achieve a finite sum?

Consider the expression

$$\frac{1}{1} + \frac{1}{1} + \frac{1}{2} + \frac{1}{6} + \cdots + \frac{1}{n!} + \cdots$$

There are an infinite number of positive terms, so the sum keeps growing as we take more terms. However, this doesn't necessarily mean that the sum grows to ∞. Let's see what happens when we add the first few terms. In other words, for any $n \in \mathbb{N}$, we consider the n^{th} *partial sum*

$$s_n = \sum_{i=0}^{n} \frac{1}{i!}$$

The first few terms of s_n are

$$1, 2, \frac{5}{2} = 2.5, \frac{8}{3} \approx 2.666667, \frac{65}{24} \approx 2.708333, \cdots$$

It is easy to see that (s_n) is strictly increasing. Also, by inspecting a few values, it seems that (s_n) is bounded above by 2.72. Thus, the BMCT tells us that (s_n) should converge to a value L satisfying $L \leq 2.72$. Since s_n can be made arbitrarily close to L by taking n sufficiently large, we get the idea that when we take EVERY term of the infinite sum, the sum should be L. Thus, it makes sense to say

$$\sum_{n=0}^{\infty} \frac{1}{n!} = L$$

In fact, in this case, we can find out what L is. Note that s_n is the same as $T_n \exp(1)$, i.e. it is the n^{th}-order approximation to $e^1 = e$. We showed in the previous chapter that $E_n \exp(1) \to 0$ as $n \to \infty$, i.e. $e - s_n \to 0$, so $L = e$! (This also shows a useful connection between Taylor polynomials and infinite series, which we will explore in more detail in the next chapter after we've built up more knowledge about infinite sums.)

However, many infinite sums are not fortunate enough to have simple answers, or even to have finite answers at all. For instance, the series

$$1 + 2 + 4 + \cdots + 2^n + \cdots$$

has the following first few partial sums:

$$1, 3, 7, 15, 31, \cdots$$

It is not hard to prove that the partial sums diverge to ∞, so this infinite sum can be thought of as being infinitely large. Hence, unlike the first sum, this sum is considered ill-defined. As another example, you can see that the sum

$$\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2} + \cdots$$

has partial sums bounded by 1.65. This infinite sum is well-defined, but it is not clear what the limit of the partial sums equals.[5] The problem with this last example is that it is difficult to obtain any simple formula for the n^{th} partial sum. Thus, in many situations, we will only be able to study whether a sum makes sense, not bothering to find the exact value of the sum.

The Formal Definition

To make the intuitive ideas above precise, we define an infinite sum as a sequence of partial sums, as follows:

Definition 9.15. Let $k \in \mathbb{N}$ and an infinite sequence $(a_n)_{n=k}^{\infty}$ of real numbers be given. The *infinite series with terms* (a_n), denoted by

$$\sum_{n=k}^{\infty} a_n$$

is the sequence $(s_n)_{n=k}^{\infty}$ defined by

$$s_n = \sum_{i=k}^{n} a_i$$

[5] Amazingly, this sum equals $\pi^2/6$; this result was first shown by Leonhard Euler.

for each $n \geq k$. s_n is called the n^{th} *partial sum* of the series. (Occasionally, we define $s_n = 0$ whenever $n < k$, so that s_n is defined for all $n \in \mathbb{N}$.) Note that the partial sums can also be defined recursively by

$$s_k = a_k \qquad s_{n+1} = s_n + a_{n+1} \text{ for all } n \geq k$$

(As with the definition of sequence, the letter n is a dummy variable and can be replaced with any other available variable in the summation expression.) Occasionally, the series will also be denoted as $\sum a_n$ if the starting index is clear from context or unimportant. (For instance, see Example 9.16.)

If the sequence (s_n) of partial sums converges to a limit $L \in \mathbb{R}$, then we say the series *converges with sum* L and write

$$\sum_{n=k}^{\infty} a_n = L$$

Thus, we have

$$\sum_{n=k}^{\infty} a_n = \lim_{N\to\infty} \sum_{n=k}^{N} a_n$$

Otherwise, if the sequence (s_n) diverges, then we say that the series *diverges*. (However, if $s_n \to \infty$ as $n \to \infty$, then we sometimes say that the series diverges to ∞ to be more precise about the nature of the divergence.)

Remark. We use a summation symbol both for representing the series and the value of the series (i.e. the sum). However, we will usually use the symbol to refer to the sum. For instance, suppose (a_n) and (b_n) are two sequences. When we write

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n$$

we mean that one series converges iff the other converges, and in that case they both converge to the same sum. We do not usually write this expression to mean that the sequences of partial sums are the same.

Example 9.16:

As a very simple example, we note that any finite sum can be viewed as an infinite series: for instance, the finite sum $a_1 + a_2 + \cdots + a_n$ is the same as $\displaystyle\sum_{i=1}^{\infty} a_i$ if we define $a_i = 0$ whenever $i > n$. Conversely, if $(a_i)_{i=1}^{\infty}$ is any infinite sequence with only finitely many nonzero terms, then its series clearly converges: when a_N is the last nonzero term, the partial sums $(s_n)_{n=1}^{\infty}$ satisfy $s_n = s_N$ for all $n \geq N$.

In fact, as far as convergence is concerned, finitely many terms make no difference. To be precise, let's say that $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are two different infinite sequences which only differ at finitely many terms. Let $N \in \mathbb{N}$ be the largest index for which $a_N \neq b_N$. Thus, if (s_n) and (t_n) respectively denote the partial sums of (a_n) and (b_n), then for any $n > N$ we have

$$\begin{aligned} s_n &= \sum_{i=1}^{n} a_i = \sum_{i=1}^{N} a_i + \sum_{i=N+1}^{n} a_i \\ &= \left(\sum_{i=1}^{N} b_i + \sum_{i=1}^{N} (a_i - b_i) \right) + \sum_{i=N+1}^{n} b_i \\ &= \sum_{i=1}^{n} b_i + \sum_{i=1}^{N} (a_i - b_i) = t_n + \sum_{i=1}^{N} (a_i - b_i) \end{aligned}$$

because $a_i = b_i$ whenever $i > N$. Note that

$$\sum_{i=1}^{N} (a_i - b_i)$$

is a constant with respect to n. Thus, s_n and t_n differ by a constant for any $n > N$, so (s_n) converges iff (t_n) converges. In fact, if both converge, then we have

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n + \sum_{i=1}^{N} (a_i - b_i)$$

On a related note, the starting point of a series doesn't affect whether the series converges. To be precise, let's say that $k, l \in \mathbb{N}$ satisfy $k < l$, and $(a_i)_{i=k}^{\infty}$ is a given real sequence. For each $n \geq l$, let

$$s_n = \sum_{i=k}^{n} a_i \text{ and } t_n = \sum_{i=l}^{n} a_i$$

so s_n and t_n are partial sums. Note that s_n and t_n differ by

$$s_n - t_n = \sum_{i=k}^{l-1} a_i$$

which is independent of n. Thus, (s_n) converges iff (t_n) converges, i.e.

$$\sum_{i=k}^{\infty} a_i \text{ and } \sum_{i=l}^{\infty} a_i$$

both converge or both diverge. For this reason, we sometimes write a series in the form $\sum a_n$, omitting a starting index for the sum, when we are only discussing convergence and the actual value of the sum is unimportant. □

It is easiest to analyze a series when there is a convenient closed-form expression for the partial sums. This makes calculating a limit of partial sums simple. We present a few examples of such series.

Example 9.17:

Consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2+n}$$

The first few partial sums of this series are

$$\begin{aligned} \frac{1}{2} &= \frac{1}{2} \\ \frac{1}{2}+\frac{1}{6} &= \frac{2}{3} \\ \frac{1}{2}+\frac{1}{6}+\frac{1}{12} &= \frac{3}{4} \end{aligned}$$

and so forth. From this, you can conjecture that for any $n \in \mathbb{N}^*$, the n^{th} partial sum is $s_n = n/(n+1)$. This can be readily proven by induction. Since $n/(n+1) \to 1$ as $n \to \infty$, it follows that our series converges with sum 1.

For a different way of looking at this series, we note that the term $1/(n^2+n)$ can be broken into simpler pieces using partial fractions:

$$\frac{1}{n^2+n} = \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$$

Therefore, for any $N \in \mathbb{N}^*$, we have

$$\begin{aligned} s_N &= \sum_{n=1}^{N} \left(\frac{1}{n} - \frac{1}{n+1}\right) \\ &= \left(\frac{1}{1}-\frac{1}{2}\right) + \left(\frac{1}{2}-\frac{1}{3}\right) + \cdots + \left(\frac{1}{N}-\frac{1}{N+1}\right) \\ &= 1 - \frac{1}{N+1} \end{aligned}$$

because all terms except for 1 and $-1/(N+1)$ cancel. This form makes it especially convenient to find the limit of the partial sums. □

The previous example suggests a general approach which works well for some series. Suppose we have a series

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

and we find some sequence $(b_n)_{n=1}^{\infty}$ such that for each $n \in \mathbb{N}$, $a_n = b_n - b_{n+1}$. This choice of b_n causes the N^{th} partial sum of our series to be

$$\sum_{n=1}^{N} (b_n - b_{n+1}) = b_1 - b_{N+1}$$

due to cancelation. In this situation, we say that we have written our series as a *telescoping series* (because the partial sums can be "collapsed" to $b_1 - b_{N+1}$, much like a navigator's telescope can be collapsed down to a small size when not in use). When a series can be written as a telescoping series for a convenient choice of (b_n), we can readily determine the sum of the series, as this theorem states:

Theorem 9.18. *Let $k \in \mathbb{N}$ be given, and suppose that $(a_n)_{n=k}^{\infty}$ and $(b_n)_{n=k}^{\infty}$ are infinite real-valued sequences. Also suppose that for all $n \geq k$, $a_n = b_n - b_{n+1}$. Then the telescoping series*

$$\sum_{n=k}^{\infty} a_n = \sum_{n=k}^{\infty} (b_n - b_{n+1})$$

converges iff the sequence (b_n) converges, and in that case we have

$$\sum_{n=k}^{\infty} a_n = b_k - \lim_{n \to \infty} b_n$$

You should prove this theorem yourself. There are also some other variants on telescoping series that you can explore in Exercise 9.4.14.

Example 9.19:

Consider the series

$$\sum_{n=1}^{\infty} \log\left(\frac{n+1}{n}\right)$$

Note that the quotient property of logarithms allows us to write $\log((n+1)/n) = \log(n+1) - \log n$. This puts our series into the telescoping form from Theorem 9.18 with $b_n = -\log n$. Since $\log n \to \infty$ as $n \to \infty$, our series diverges. In fact, the proof of Theorem 9.18 readily shows that our series diverges to ∞.

Another way of solving this problem, which amounts to essentially the same work, is to compute the n^{th} partial sum directly as

$$\log\frac{2}{1} + \log\frac{3}{2} + \cdots + \log\frac{n+1}{n} = \log\left(\frac{2}{1}\cdot\frac{3}{2}\cdot\cdots\cdot\frac{n+1}{n}\right) = \log\frac{n+1}{1}$$

using the product property of log. Since $\log(n+1)$ diverges to ∞ as $n \to \infty$, so does our series. $\square$

Interestingly, the result of the previous example tells us information about another series. Notice that

$$\log\left(\frac{n+1}{n}\right) = \log\left(1 + \frac{1}{n}\right) \approx \frac{1}{n}$$

where we approximate with the Taylor polynomial $T_1 \log(1+x)$ at $x = 1/n$. This approximation should be pretty accurate since $1/n$ is close to 0 when n is large. In particular, we can conjecture that

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

diverges to ∞ because the series from the previous example does.

This conjecture can be proven in a few different ways. Let's say t_n is the n^{th} partial sum of $\sum(1/n)$ and s_n is the n^{th} partial sum of $\sum \log((n+1)/n)$. One way of proving this conjecture is to analyze the Taylor error to show that $t_n \geq s_n$ when n is large. Since $s_n \to \infty$, this implies that $t_n \to \infty$. Another way of proving this conjecture uses completely different inequalities to find something smaller than t_n which still approaches ∞. For more details, see Exercise 9.4.15.

Geometric Series

At this point, we'll study a particular type of series, called a *geometric series*, for which we can compute closed-form formulas for the partial sums. These series are the natural extensions of geometric summations that we saw back in Example 8.5. Geometric series appear in a number of contexts, and we will make use of them throughout this chapter.

To be precise, suppose $r \in \mathbb{R}$ is given. A sequence (a_n) is called a *geometric sequence with ratio r* if each term except the first is obtained from the previous term by multiplying by r, i.e. $a_{n+1} = ra_n$. The corresponding series is called a *geometric series*. Thus, if our sequence starts with index k, then our geometric series takes the form

$$a_k + ra_k + r^2 a_k + \cdots = \sum_{n=k}^{\infty} r^{n-k} a_k$$

It is also possible to express this sum as

$$\sum_{i=0}^{\infty} r^i a_k$$

(Formally, these two series are not the same when $k \neq 0$ because the partial sums of the first series start with index k, and the partial sums of the second series start with index 0. However, these two sequences of partial sums are merely shifted versions of one another, so if one converges, then the other also converges to the same limit.)

First, we note that when $a_k = 0$, every term in our series is 0, so the sum is trivially 0. Hence, let's consider what happens when $a_k \neq 0$. Next, when $r = 1$, every term equals a_k, so the series diverges to either ∞ or $-\infty$ (the sign of the infinity equals the sign of a_k). Thus, for the remainder of this discussion, let's consider when $a_k \neq 0$ and $r \neq 1$.

The n^{th} partial sum is

$$s_n = \sum_{i=0}^{n} r^i a_k = a_k(1 + r + \cdots + r^n)$$

for $n \in \mathbb{N}$. In Exercise 1.9.2, and when studying geometric summations in Example 8.5, we have seen a closed form for this sum:

$$s_n = a_k \frac{r^{n+1} - 1}{r - 1} = a_k \left(\frac{1}{1-r} - \frac{r^{n+1}}{1-r} \right)$$

(Alternately, $a_k r^n$ can be written as $a_k(r^n/(1-r) - r^{n+1}/(1-r))$, which allows us to obtain a telescoping form with $b_n = a_k r^n/(1-r)$.)

Thus, our geometric series converges iff r^{n+1} converges as $n \to \infty$. When $|r| < 1$, we have $r^{n+1} \to 0$. When $|r| > 1$, $|r^{n+1}| \to \infty$, so r^{n+1} diverges in this case. Lastly, if $|r| = 1$ and $r \neq 1$, then $r = -1$, and $(-1)^{n+1}$ diverges as $n \to \infty$ because it alternates between -1 and 1. Thus, we have found that the geometric series converges iff $|r| < 1$, in which case the sum is

$$\sum_{n=0}^{\infty} a_k r^n = \frac{a_k}{1-r}$$

Another way to say this is that the sum of a geometric series, when it converges, is its first nonzero term divided by one minus the common ratio.

It turns out we've already seen this sum before. In the case when $k = 0$ and $a_0 = 1$, the sum is $1/(1-r)$. We already know that

$$T_n\left(\frac{1}{1-x}\right) = 1 + x + \cdots + x^n$$

for any $n \in \mathbb{N}$ and any $x \in \mathbb{R} - \{1\}$. Thus, our work above shows that the Taylor polynomials for $1/(1-x)$ converge to $1/(1-x)$ iff $|x| < 1$. In other words, Taylor approximations to $1/(1-x)$ are only useful for points which are less than distance 1 from the center of the approximation.

Some Properties of Series

As series are merely sequences of finite sums, it makes sense that many limit laws relating to sums would apply to series. For instance, we have the following linearity law:

Theorem 9.20. *Let $k \in \mathbb{N}$ and $\alpha, \beta \in \mathbb{R}$ be given, and let $(a_n)_{n=k}^{\infty}$ and $(b_n)_{n=k}^{\infty}$ be infinite sequences of real numbers. If*

$$\sum_{n=k}^{\infty} a_n \text{ and } \sum_{n=k}^{\infty} b_n$$

both converge, then so does the series with terms $(\alpha a_n + \beta b_n)_{n=k}^{\infty}$, and

$$\sum_{n=k}^{\infty} (\alpha a_n + \beta b_n) = \alpha \sum_{n=k}^{\infty} a_n + \beta \sum_{n=k}^{\infty} b_n$$

Strategy. When we discuss the convergence of a series, we are really discussing the convergence of its sequence of partial sums. Thus, all we have to do is write everything in terms of the partial sums, and then we can use our limit laws for sequences.

Proof. Let $k, \alpha, \beta, (a_n), (b_n)$ be given as described. For each $n \geq k$, define

$$s_n = \sum_{i=k}^{n} a_i \text{ and } t_n = \sum_{i=k}^{n} b_i$$

Thus, s_n and t_n are respectively the n^{th} partial sums of (a_n) and (b_n). Hence, the n^{th} partial sum of $(\alpha a_n + \beta b_n)$ is

$$\sum_{i=k}^{n} (\alpha a_i + \beta b_i) = \alpha \sum_{i=k}^{n} a_i + \beta \sum_{i=k}^{n} b_i = \alpha s_n + \beta t_n$$

By assumption, (s_n) and (t_n) converge as $n \to \infty$. Thus, by the limit laws for sequences, $(\alpha s_n + \beta t_n)$ also converges as $n \to \infty$, and

$$\begin{aligned} \sum_{n=k}^{\infty} (\alpha a_n + \beta b_n) &= \lim_{n\to\infty} (\alpha s_n + \beta t_n) \\ &= \alpha \lim_{n\to\infty} s_n + \beta \lim_{n\to\infty} t_n \\ &= \alpha \sum_{n=k}^{\infty} a_n + \beta \sum_{n=k}^{\infty} b_n \end{aligned}$$

as desired. ■

Example 9.21:

Theorem 9.20 tells us that a linear combination of convergent series is also convergent. In particular, the sum or difference of two convergent series is convergent. As an example, we can use this theorem to calculate

$$\sum_{n=0}^{\infty} \left(\frac{1+2^n}{3^n} \right) = \sum_{n=0}^{\infty} \left(\frac{1}{3} \right)^n + \sum_{n=0}^{\infty} \left(\frac{2}{3} \right)^n = \frac{1}{1-1/3} + \frac{1}{1-2/3} = \frac{9}{2}$$

In fact, linearity suggests another way to compute the sum of the geometric series

$$\sum_{i=0}^{\infty} r^i a_k$$

As before, let's consider when $r \neq 1$ and $a_k \neq 0$. The only term in this sum without any powers of r is the $i = 0$ term. We take that term out to the side, and we factor a single r out of the remaining terms, to get[6]

$$\sum_{i=0}^{\infty} r^i a_k = a_k + r \sum_{i=1}^{\infty} r^{i-1} a_k$$

By reindexing the second summation with $n = i - 1$, so that n starts at 0, we obtain[7]

$$\sum_{i=0}^{\infty} r^i a_k = a_k + r \sum_{n=0}^{\infty} r^n a_k$$

At this point, the sum we are trying to find appears on both sides of the equation (recall that i and n are dummy indeces). We move them to the same side and obtain

$$\begin{aligned} \left(\sum_{i=0}^{\infty} r^i a_k\right)(1-r) &= a_k \\ \sum_{i=0}^{\infty} r^i a_k &= \frac{a_k}{1-r} \end{aligned}$$

since $r \neq 1$. This agrees with our previous computation.

This approach for calculating the sum is usually easier to remember. However, the approach does have one flaw: you are asked in Exercise 9.4.16 to determine the problem with this computation. □

Remark. Theorem 9.20 addresses what happens when convergent series are added together. This suggests the question: what happens if you add two divergent series together, or if you add a convergent and a divergent series together? It turns out that when you add two divergent series together, the result may or may not diverge. For instance, we know that $\sum 1$ diverges to ∞, and

$$\sum_{n=0}^{\infty} (1+1)$$

[6]Technically, the step that pulls the $i = 0$ term out to the side is treating the original series as a sum of two series: the first series has its 0^{th} term equal to a_k, but all other terms are equal to 0. This detail can safely be ignored in practice.

[7]Formally, this step uses the Composition Limit Theorem with the sequence of partial sums, treating n as a function of i. Usually, we do not need to be so formal about index changes like these.

diverges to ∞, but

$$\sum_{n=0}^{\infty}(1+(-1))$$

converges to 0.

However, when a convergent series and a divergent series are added, the result is always a divergent series. To see this, suppose $\sum a_n$ converges (the starting index of the sum isn't important here) but $\sum b_n$ diverges. If $\sum(a_n+b_n)$ were convergent, then we could write

$$\sum b_n = \sum(a_n+b_n) - \sum a_n$$

and hence $\sum b_n$ would converge as well by Theorem 9.20. This proves by contradiction that $\sum(a_n+b_n)$ must diverge.

Example 9.22:

To demonstrate the remarks made above, we note that

$$\sum_{n=0}^{\infty}\frac{4^n+2}{3^n} = \sum_{n=0}^{\infty}\left(\left(\frac{4}{3}\right)^n + 2\left(\frac{1}{3}\right)^n\right)$$

diverges, because it is the sum of a divergent geometric series and a convergent geometric series.

However, consider the series

$$\sum_{n=0}^{\infty}(2^n+(-2)^n)$$

Since both $\sum 2^n$ and $\sum(-2)^n$ diverge, we cannot immediately use Theorem 9.20 to conclude whether our series converges. Instead, we try to write this series in a better form. Note that when n is odd, $(-2)^n = -(2^n)$, so $2^n + (-2)^n = 0$. However, when n is even, $(-2)^n = 2^n$, so $2^n + (-2)^n = 2^{n+1}$. Therefore, our series can be written as[8]

$$2^1+0+2^3+0+2^5+\cdots = \sum_{n=0}^{\infty}2^{2n+1} = \sum_{n=0}^{\infty}(2\cdot 4^n)$$

which diverges to ∞. $\square$

[8] We are being pretty informal with the first equals sign in the above calculation, but it is not hard to check that $\sum 2^{2n+1}$ diverges iff the original series diverges. This check is frequently ignored in practice.

The Divergence Test and Related Remarks

At this point, it is prudent to mention that although the linearity property works for infinite series, as Theorem 9.20 demonstrates, certain other properties of finite sums do NOT carry over to infinite series. The next example contains a famous invalid calculation illustrating these ideas:

Example 9.23:

Consider the following calculation:

$$\begin{aligned} 0 &= 0+0+\cdots \\ &= (1+(-1))+(1+(-1))+\cdots \\ &= 1+(-1)+1+(-1)+\cdots \\ &= 1+((-1)+1)+((-1)+1)+\cdots \\ &= 1+0+0+\cdots = 1 \end{aligned}$$

Some step of this calculation must be invalid, since $0 \neq 1$. Which steps are problematic? The first and last steps are fine; it is safe to replace 0 with $(1+(-1))$ or with $((-1)+1)$, since those are just equal numbers. However, we claim that the transition from the second to the third line is problematic, as is the transition from the third line to the fourth.

To shed more light on what is happening here, let's be more formal about the transition from the second line to the third line. The series on the second line has terms $(a_n)_{n=0}^{\infty}$ with $a_n = 1+(-1)$ for each $n \in \mathbb{N}$. The series on the third line has terms $(b_n)_{n=0}^{\infty}$, where

$$b_n = \begin{cases} 1 & \text{if } n \text{ is even} \\ -1 & \text{if } n \text{ is odd} \end{cases}$$

for each $n \in \mathbb{N}$. Thus, if we let s_n be the n^{th} partial sum of $\sum a_n$, and we let t_n be the n^{th} partial sum of $\sum b_n$, then $s_n = 0$, whereas $t_n = 1$ for even n and $t_n = 0$ for odd n. This shows that (s_n) converges to 0 but (t_n) diverges, alternating between 1 and 0.

Hence, by removing and regrouping parentheses, we have changed the structure of the partial sums. This means that in an infinite series, the sums are NOT associative in general. Associativity is a property which holds for finite sums but not for infinite sums.

Another property which holds for finite sums but fails for infinite sums is commutativity. Unfortunately, we cannot yet produce a good counterexample where a series converges but a rearrangement of the terms diverges (we will have such counterexamples later, when we cover the Riemann Rearrangement Theorem, Theorem 9.79). However, we can demonstrate an example where rearranging the terms of a divergent series causes divergence in a different way. As shown above, the series

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

diverges because its partial sums alternate between 1 and 0. Now, suppose we rearrange the terms to produce

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

(we write 1 followed by -1, then 1 twice followed by -1 twice, and so on, incrementing the number of consecutive 1's and -1's each time). With this rearrangement, the partial sums are not bounded! (For each $N \in \mathbb{N}$, you can find some $n > N$ such that the n^{th} partial sum is N.) $\square$

In the first part of the previous example, where we had the paradoxical calculation to show that $0 = 1$, we created a series whose partial sums alternate between 1 and 0. Thus, the partial sums don't converge to any one value, because they change by 1 at every step. Intuitively, in order for a series to converge, the partial sums must become arbitrarily close to the final sum, so they change less and less as n grows large. This leads to the following handy condition:

Theorem 9.24 (Divergence Test). *Let (a_n) be an infinite real sequence. If $\sum a_n$ converges (for any starting index), then $a_n \to 0$ as $n \to \infty$.*

Strategy. The main realization here is that a term in the series is also the difference between two partial sums: if (s_n) is the sequence of partial sums, then

$$s_{n+1} = s_n + a_{n+1}$$

Since our series is convergent, the partial sums approach a finite limit. In fact, s_n and s_{n+1} both approach the same limit, so it follows that $a_{n+1} \to 0$.

Proof. Let (a_n) be given as described, and assume $\sum a_n$ converges. Let (s_n) be the sequence of partial sums. By assumption, there is some $L \in \mathbb{R}$ such that $s_n \to L$ as $n \to \infty$. Now, by definition, we have

$$s_{n+1} = s_n + a_{n+1}$$

for all $n \in \mathbb{N}$ such that a_{n+1} is defined. Therefore, since we also have $s_{n+1} \to L$ as $n \to \infty$, we have

$$\lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} (s_{n+1} - s_n) = L - L = 0$$

It follows as well that $a_n \to 0$ as $n \to \infty$. ■

Example 9.25:

The Divergence Test earns its name because it is usually a quick way to check whether a series is divergent before doing any more advanced testing. For instance, the series

$$\sum_{n=0}^{\infty} e^{1/n}$$

is immediately known to be divergent, because $e^{1/n} \to 1$ as $n \to \infty$. Since it is usually very quick to apply the Divergence Test (i.e. to see whether the terms approach 0), you should often apply the Divergence Test first when studying a series. After all, if the terms do go to 0, then you have only spent a little effort on the Divergence Test, whereas if the terms don't go to 0 and you try other techniques, then you waste much more effort.

For a harder, more subtle example, consider

$$\sum_{n=0}^{\infty} \sin n$$

We know that $\sin x$ has no limit as $x \to \infty$ when x is REAL. This doesn't immediately prove that $\sin n$ has no limit when n is an INTEGER tending to ∞. As it turns out, we don't need to prove that $\sin n$ has no limit to show our series diverges by the Divergence Test (though see Exercise 9.4.17). We just need to show $\sin n \not\to 0$.

To prove this, consider the sector of the unit circle between angles $\pi/6$ and $5\pi/6$, as displayed in Figure 9.3. For all $\theta \in [\pi/6, 5\pi/6]$, we have $\sin\theta \geq$

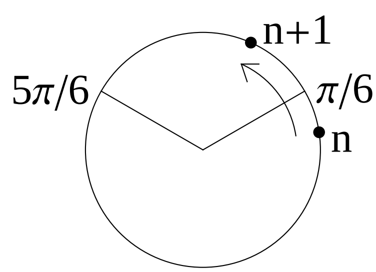

Figure 9.3: Illustration that n lands between $\pi/6$ and $5\pi/6$ infinitely often

$1/2$. To show that $\sin n \not\to 0$, it suffices to show that $\sin n \geq 1/2$ infinitely often, which means showing that the angle n lands in this sector of the circle infinitely often.

We can think of n as representing a point on the circle with polar angle of n radians, so as we proceed from n to $n+1$, we take a point on the circle and rotate it counterclockwise by 1 extra radian, as shown in the figure. Our sector covers an angle of $2\pi/3$ radians. Since $2\pi/3 > 1$, we intuitively see that as this point keeps rotating around the circle, it can't "jump over" the sector, so it must land in the sector at least once for each full revolution around the circle. This shows that n lies in the sector infinitely often. (You can formalize this more in Exercise 9.4.17.) $\square$

Remark. The Divergence Test says that if $\sum a_n$ converges, then $a_n \to 0$. However, the converse is not true! For instance, we saw in Example 9.19 that

$$\sum_{n=1}^{\infty} \log\left(\frac{n+1}{n}\right)$$

diverges, but $\log((n+1)/n) \to 0$. We will see many more examples soon of series which diverge but have terms approaching 0.

At this point, we have a few basic tools for computing sums of series. These include:

- Analysis of Taylor errors
- Finding a simple telescoping form
- Proving a formula for the partial sums by induction
- Building a series out of linear combinations of simpler series or basic examples (such as a geometric series)

- Using the Divergence Test

These basic approaches can only go so far, since we have so few basic examples at the moment. As a result, there are many series for which these approaches are not currently helpful. For these series, we usually can't expect to easily find the value of the sum if the series converges. Instead, for the rest of this chapter, we will focus our efforts on finding ways of proving convergence or divergence, as well as obtaining estimates for the error in approximating a series's sum with partial sums.

9.4 Exercises

1. Consider the series
$$\sum_{i=0}^{\infty}(-1)^i \frac{\pi^{2i+1}}{2^{2i+1}(2i+1)!}$$
For any $n \in \mathbb{N}$, its n^{th} partial sum is $T_{2n+1}\sin(\pi/2)$. Prove that $E_{2n+1}\sin(\pi/2) \to 0$ as $n \to \infty$, and use this to prove that the infinite series has 1 as its sum.

2. Consider the series
$$\sum_{i=0}^{\infty} \frac{(-1)^{i-1}}{(2i)!}$$
As in the previous exercise, the partial sums come from Taylor approximations. Identify which Taylor approximations they are, prove that the error converges to 0, and use this to find the sum of the series.

3. (a) Suppose $(s_n)_{n=0}^{\infty}$ is a real-valued infinite sequence. Prove that there exists exactly one real sequence $(a_n)_{n=0}^{\infty}$ such that for each $n \in \mathbb{N}$,
$$s_n = \sum_{i=0}^{n} a_i$$
Thus, any sequence can be considered to be the partial sums of a series. (Hint: Define a_n inductively.)

   (b) Suppose $(a_n)_{n=0}^{\infty}$ is given. Prove that for any $b_0 \in \mathbb{R}$, there exists exactly one sequence $(b_n)_{n=1}^{\infty}$ such that for all $n \in \mathbb{N}$, $a_n = b_n -$

b_{n+1}. Thus, every series can be expressed in telescoping form, since

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} (b_n - b_{n+1})$$

(However, if there is no simple closed-form expression for b_n, then this telescoping form is not useful because it becomes difficult to calculate a limit of b_n as $n \to \infty$!)

For Exercises 4 through 12, compute the sum of the series if it converges, or establish that the series diverges. You may solve these problems by either guessing a formula for the partial sums which is proven by induction, or by using any theorems or series we have established in this chapter.

4. $\displaystyle\sum_{n=0}^{\infty} (-1)^n n$

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

6. $\displaystyle\sum_{n=1}^{\infty} \frac{6n^2 - 4n - 1}{(3n+1)(3n-2)}$

7. $\displaystyle\sum_{n=0}^{\infty} \frac{3^n + 4^n}{5^{2n+1}}$

8. $\displaystyle\sum_{n=0}^{\infty} \frac{4^n + 3^{2n}}{6^n}$

9. $\displaystyle\sum_{n=0}^{\infty} n^{1/n}$

10. $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n^2 - 1}$

11. $\displaystyle\sum_{n=1}^{\infty} \frac{\sqrt{n+1} - \sqrt{n}}{\sqrt{n^2 + n}}$

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

13. Use partial fractions to prove that for all $n \in \mathbb{N}^*$,

$$\frac{n}{(n+1)(n+2)(n+3)} = \frac{-1}{2}\left(\frac{1}{n+1} - \frac{1}{n+2}\right) + \frac{3}{2}\left(\frac{1}{n+2} - \frac{1}{n+3}\right)$$

Use this result to compute the sum of

$$\sum_{n=1}^{\infty} \frac{n}{(n+1)(n+2)(n+3)}$$

14. In this exercise, we consider some modifications of Theorem 9.18. Let $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=0}^{\infty}$ be given.
   (a) Suppose that for each $n \in \mathbb{N}$, $a_n = b_n - b_{n+2}$. Prove that
   $$\sum_{n=0}^{\infty} a_n$$
   converges iff $b_n + b_{n+1}$ converges as $n \to \infty$. (In particular, if $b_n \to 0$, then $b_n + b_{n+1} \to 0$.) If $b_n + b_{n+1} \to L$ as $n \to \infty$, then what is the value of our series?
   (b) Apply your result from part (a) to the two series
   $$\sum_{n=0}^{\infty} \frac{1}{\sqrt{n} + \sqrt{n+2}} \qquad \text{and} \qquad \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)(n+3)}$$
   Find the sums of the series above which converge.
   (c) Suppose instead that for some fixed $k \in \mathbb{N}^*$, we have $a_n = b_n - b_{n+k}$ for all $n \in \mathbb{N}$. Generalize your results from part (a) to this case, and use them to find the sum of the series
   $$\sum_{n=1}^{\infty} \frac{\log\left(n^{n+3}/(n+3)^n\right)}{n^2+3n}$$

15. This exercise presents two ways of proving formally that $\sum(1/n)$ diverges. For part (a), suppose that s_n is the n^{th} partial sum of $\sum \log((n+1)/n)$ and t_n is the n^{th} partial sum of $\sum(1/n)$.
   (a) Using Exercises 7.2.5 and 8.5.11, we know that for all $x \in (1, 2)$,
   $$x - 1 > \log x > x - 1 - \frac{(x-1)^2}{2}$$
   Use this to establish that $t_n \geq s_n \geq t_n/2$ for all $n \geq 2$. Conclude by explaining why s_n diverges to ∞ iff t_n diverges to ∞.
   (b) For a different way of analyzing the series $\sum(1/n)$, prove by induction that for all $n \in \mathbb{N}^*$,
   $$\sum_{i=1}^{2^n} \frac{1}{i} \geq \frac{n}{2} + 1$$
   Use this to prove that $\sum(1/n)$ diverges.

16. What is wrong with the argument from Example 9.21 that tries to calculate the sum of the geometric series? (Note that there must be a flaw, or else that argument would establish that

$$\sum_{i=0}^{\infty} 2^i$$

converges to $1/(1-2)$, which is negative!)

17. This exercise aims to formally prove that $\sin n \not\to 0$ as $n \to \infty$. The key challenge of writing this argument precisely is specifying exactly what it means to take a number representing an angle and rotate it counterclockwise by 1 radian. To do this, we introduce the function $\rho : [0, 2\pi) \to [0, 2\pi)$ defined for any $x \in [0, 2\pi)$ by

$$\rho(x) = \begin{cases} x+1 & \text{if } x < 2\pi - 1 \\ x+1-2\pi & \text{if } x \geq 2\pi - 1 \end{cases}$$

Intuitively, ρ takes a number representing an angle and adds 1 to it, adjusting it if necessary to lie in the correct range. (Sometimes people say that ρ adds 1 "modulo 2π", suggesting a similarity with arithmetic in $\mathbb{Z}_n$.)

Next, define the sequence $(a_n)_{n=0}^{\infty}$ inductively by

$$a_0 = 0 \qquad a_{n+1} = \rho(a_n) \text{ for } n \in \mathbb{N}$$

Intuitively, a_n is supposed to represent the same angle as n.

(a) Prove by induction that for all $n \in \mathbb{N}$, $a_n - n$ has the form $2\pi k_n$ where k_n is an integer. Use this to conclude that

$$\sin n = \sin a_n$$

Hence, it suffices to prove $\sin a_n \not\to 0$.

(b) Now, we will prove that a_n is in $[\pi/6, 5\pi/6]$ for infinitely many values of n. Prove that $a_1 \in [\pi/6, 5\pi/6]$ and that for all $n \in \mathbb{N}$, if $a_n \in [\pi/6, 5\pi/6]$, then at least one of a_{n+6} or a_{n+7} is in $[\pi/6, 5\pi/6]$. (Hint: Consider cases based on whether a_n is larger than $\pi/6 + (2\pi - 6)$.)

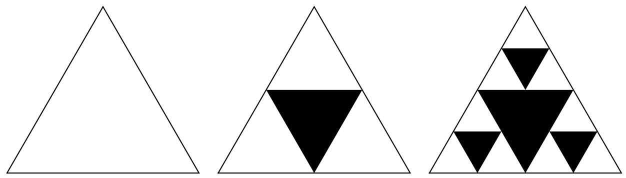

Figure 9.4: Three stages of the construction of the Sierpinski triangle (solid black triangles represent removed regions)

(c) Use part (b) to prove that $\sin a_n \geq 1/2$ for infinitely many values of $n \in \mathbb{N}$. Therefore, $\sin a_n \not\to 0$.

(d) Let's go further with this analysis. Use the idea of part (b) to prove that $a_n \in [7\pi/6, 11\pi/6]$ for infinitely many values of n. Why does this, together with parts (a) and (c), prove that $\sin n$ has no limit as $n \to \infty$?[9]

18. Find the value of the repeating decimal $1.2454545\ldots$ (i.e. the digit pattern 45 keeps repeating) by expressing it as the sum of a geometric series.

19. As an application of geometric sequences and series, we consider a shape called the *Sierpinski triangle*[10]. To create it, we start with an equilateral triangle with area 1, which is the leftmost shape in Figure 9.4. Next, we break the triangle into four equilateral triangles with half the side length and remove the interior of the middle triangle (i.e. you do not remove the boundary of the middle triangle). This gives us the middle shape in Figure 9.4. Next, in each of the three triangles remaining, we similarly break them into 4 triangles with half the side length and remove the middle one. This yields the rightmost shape in Figure 9.4.)

[9] In fact, it is possible to prove that for any $k \in \mathbb{N}^*$, if the interval $[0, 2\pi]$ is broken into k equal subintervals $[0, 2\pi/k]$, $[2\pi/k, 4\pi/k]$, and so on, then each subinterval contains infinitely many terms of (a_n). (In some sense, this says that the integers are uniformly spread out over the unit circle.) If you wish to try proving this yourself, then it is suggested that you first try to prove there exist two integers m and n with $m < n$ and $|a_m - a_n| \leq 2\pi/k$.

[10] The shape is named after the Polish mathematician Waclaw Sierpinski, who is also credited with creating other self-similar sets and simple fractals.

Doing this "middle-removal" step repeatedly, let's say that S_n is the shape that remains after n steps. For instance, our figure shows S_0, S_1, and S_2. When you perform this "middle-removal" step infinitely many times, what remains is the set

$$\bigcap_{n=0}^{\infty} S_n = S_0 \cap S_1 \cap S_2 \cap \cdots$$

called the *Sierpinski triangle*.

(a) Prove that the area enclosed by the triangles in S_n is $(3/4)^n$ and the total perimeter of the triangles in S_n is

$$\left(\frac{3}{2}\right)^n \text{Perimeter}(S_0)$$

It follows that in the limit, the Sierpinski triangle has area 0 but infinite perimeter!

(b) Alternately, we can measure the area of the Sierpinski triangle by looking at the removed areas. Prove that for each $n \in \mathbb{N}$, the area removed from S_0 to create S_n is

$$\sum_{i=1}^{n} 3^{i-1} \left(\frac{1}{4}\right)^i$$

Show that as $n \to \infty$, we obtain an infinite series with sum 1. Thus, as we found before, all of the area of the original triangle is gone in the Sierpinski triangle.[11]

9.5 Improper Integrals and the Integral Test

As the end of the last section indicates, we aim to develop ways to test whether a series converges without knowing its exact sum. This means the same thing as determining whether the sequence of partial sums has a limit without actually finding the value of the limit. This raises the question:

[11] Nevertheless, the Sierpinski triangle still has infinitely many points in it. In fact, it can be shown to have infinitely many points which aren't on the boundary of any removed triangle!

what tools do we have for finding whether a limit exists without actually computing the limit?

The only such theorem we have at our disposal is the Bounded Monotone Convergence Theorem, or the BMCT. If we are going to use the BMCT on a sequence of partial sums, then we should next ask: when is our sequence of partial sums monotone? To be more precise, let's say s_n is the n^{th} partial sum of the series $\sum a_n$. Therefore, $s_{n+1} = s_n + a_{n+1}$ for all n. This shows us that $s_{n+1} \geq s_n$ iff $a_{n+1} \geq 0$. As a result, we obtain the following:

Theorem 9.26. *Let $k \in \mathbb{N}$ be given, and suppose that $(a_n)_{n=k}^{\infty}$ is an infinite real-valued sequence where all the terms (except for possibly a_k) have the same sign. Then the partial sums $(s_n)_{n=k}^{\infty}$ corresponding to $\sum a_n$ form a monotone sequence: they are increasing if $a_n \geq 0$ for all $n > k$, and they are decreasing if $a_n \leq 0$ for all $n > k$. Thus, $\sum a_n$ converges iff the partial sums (s_n) are bounded.*

Because of this theorem, we will focus for now on *nonnegative series*, i.e. series of the form $\sum a_n$ where $a_n \geq 0$ for each n. For a nonnegative series, the series converges iff the partial sums are bounded above. However, where will we look for bounds? We have several ways to find bounds which we will discuss in the next few sections.

Our first major way of finding bounds uses an idea we have seen back when we studied integration: there's a connection between summations of positive numbers and areas under positive step functions. We illustrate this connection with two important examples.

Example 9.27:

First, let's consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

For each $i \in \mathbb{N}^*$, suppose we think of $1/i^2$ as representing the area underneath a rectangle with base length 1 and height $1/i^2$. Let's say this is rectangle $\#i$. If we put rectangles $\#1$ through $\#n$ side by side, so the heights range from $1/1^2$ to $1/n^2$, then the total area of our rectangles is the n^{th} partial sum s_n. We can fit these rectangles side by side under the graph of $y = 1/x^2$, as shown in Figure 9.5.

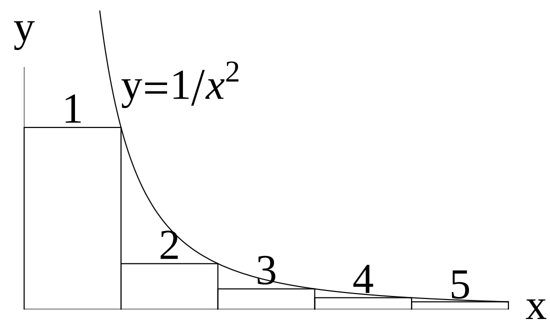

Figure 9.5: Rectangles underneath $y = 1/x^2$ representing s_5 (where rectangle $\#i$ carries the number i on top)

From the figure, we see that s_n is related to a right-endpoint approximation of the area underneath $y = 1/x^2$. More precisely, we have

$$s_n = \sum_{i=1}^{n} \frac{1}{i^2} = \frac{1}{1^2} + \sum_{i=2}^{n} \frac{1}{i^2} < 1 + \int_1^n \frac{dx}{x^2}$$

(Note that we don't want to compare the partial sum to an integral from 0 to n, because $1/x^2$ is unbounded on $(0, n)$! That is why the area of rectangle $\#1$ is pulled out of the sum first.) Thus, we have

$$s_n < 1 + \left(\frac{-1}{x} \right) \Bigg|_1^n = 2 - \frac{1}{n}$$

In particular, this shows that $s_n \leq 2$ for all $n \in \mathbb{N}^*$. Since our partial sums are bounded above, our series converges to a value which is at most 2.

Going further with this, we can use the same idea to get a lower bound on the sum of our series! Instead of placing the rectangles under the graph of $y = 1/x^2$, we place them over the graph by making the upper-left corners touch the graph instead of the upper-right corners. This is illustrated in Figure 9.6.

Therefore, s_n is a left-endpoint approximation to the area under $y = 1/x^2$ on $[1, n+1]$ (Note that rectangle $\#i$ will have its base on the x-axis covering $[i, i+1]$ this time.) This gives us

$$s_n > \int_1^{n+1} \frac{1}{x^2}\, dx = 1 - \frac{1}{n+1}$$

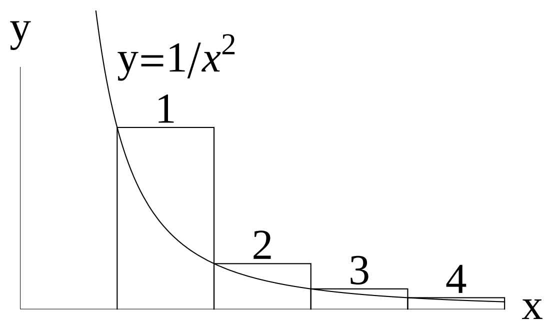

Figure 9.6: Rectangles above $y = 1/x^2$ representing s_5

Since the value of our sum is the limit of s_n as $n \to \infty$, this shows that our sum is at least 1. This is not a particularly good lower bound: we will show how we can get better lower bounds from this method near the end of this section. $\square$

Example 9.28:

For a slightly different kind of example, with a different conclusion, let's consider the series

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

This famous series is called the *harmonic series*, and its n^{th} partial sum is called the n^{th} *harmonic number* H_n. In the remarks following Example 9.19, and in Exercise 9.4.15, we showed that the harmonic series diverges to ∞. Let's show how the ideas from the previous example can be used to prove that this series diverges.

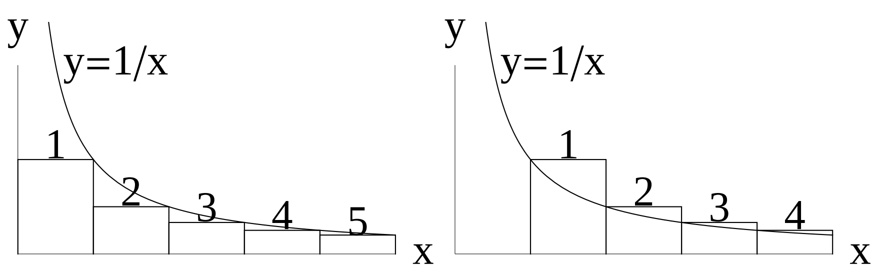

Figure 9.7: The 5^{th} harmonic number H_5, related to the curve $y = 1/x$

As before, we think of H_n as representing the area of n rectangles, but now the i^{th} rectangle has height $1/i$. By placing these rectangles with their upper-right corners touching the graph of $y = 1/x$, we get the graph on the left side of Figure 9.7. By placing these rectangles with their upper-left corners touching the graph of $y = 1/x$, we get the graph on the right side of Figure 9.7. (The graphs are similar to those in the previous example, but $y = 1/x$ decreases to 0 more slowly as $x \to \infty$ than $y = 1/x^2$ does. Also, as $x \to 0^+$, $y = 1/x$ goes to ∞ faster than $y = 1/x^2$ does.)

The two illustrations give us an underestimate and an overestimate of s_n:

$$\int_1^{n+1} \frac{dx}{x} < H_n < \frac{1}{1} + \int_1^n \frac{dx}{x}$$

so that H_n is between $\log(n+1)$ and $1 + \log n$. Both of these quantities approach ∞ as $n \to \infty$. In particular, since the underestimate goes to ∞, the partial sums are unbounded above. $\square$

Comparing these two examples, we see that each involves an overestimate and an underestimate to partial sums by thinking of them as areas of a collection of rectangles. We compare these areas to integrals of a similar real function ($1/x^2$ for the first example, $1/x$ for the second). In our first example, we proved convergence by showing the upper bounds were below a convergent limit of integrals. In our second example, we proved divergence by showing the lower bounds were above a divergent limit of integrals. In either case, there is a connection between a limit of integrals and the sum of the series. Before making that connection formal, let's first study limits of integrals in more detail.

Improper Integrals Of Type 1

Looking back at Figure 9.5, we saw

$$\sum_{i=2}^{n} \frac{1}{i^2} < \int_1^n \frac{dx}{x^2}$$

and then we took a limit as $n \to \infty$ of both sides. It is tempting to describe this limit with the following notation:

$$\sum_{i=2}^{\infty} \frac{1}{i^2} \le \int_1^\infty \frac{dx}{x^2}$$

Up to this point, we defined integrals only over closed bounded intervals. However, the above work suggests that we can extend the definition of integral to include infinite intervals. This leads us to the following type 1 integrals (we will consider type 2 at the end of this section):

Definition 9.29. Let $a \in \mathbb{R}$ be given and let $f : [a, \infty) \to \mathbb{R}$ be a function such that f is integrable on $[a, b]$ for all $b > a$. Then the *improper integral (of type 1) of f over* $[a, \infty)$ is

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx$$

(where b is real). When this limit exists, we say the improper integral *converges*. Otherwise, we say that the improper integral *diverges*.

Similarly, if $b \in \mathbb{R}$ is given and $f : (-\infty, b] \to \mathbb{R}$ is integrable on $[a, b]$ for all $a < b$, then the *improper integral of f over* $(-\infty, b]$ is

$$\int_{-\infty}^b f(x)\,dx = \lim_{a\to-\infty} \int_a^b f(x)\,dx$$

Intuitively, when f is positive, we can think of the improper integral over $[a, \infty)$ as the total area under the graph of f to the right of $x = a$. Sometimes this area is finite (when the integral converges), and sometimes this area is infinite. We present a few examples.

Example 9.30:

Consider

$$\int_0^\infty e^{-x}\,dx$$

We calculate this as follows:

$$\begin{aligned}\int_0^\infty e^{-x}\,dx &= \lim_{b\to\infty} \int_0^b e^{-x}\,dx \\ &= \lim_{b\to\infty} -e^{-x}\Big|_0^b \\ &= \lim_{b\to\infty} 1 - e^{-b} = 1\end{aligned}$$

Similarly, we have

$$\int_{-\infty}^0 e^{-x}\,dx = \lim_{a\to-\infty} e^{-a} - 1 = \infty$$

so the improper integral diverges on $(-\infty, 0]$. This makes sense, because the graph of e^{-x} blows up as $x \to -\infty$, whereas $e^{-x} \to 0$ as $x \to \infty$. $\square$

Example 9.31:

Let's generalize the examples with $1/x$ and $1/x^2$ from earlier and consider

$$\int_1^\infty \frac{dx}{x^p}$$

where p is a constant. We have the following antiderivative for $x \in [1, \infty)$:

$$\int x^{-p}\,dx = \begin{cases} \dfrac{x^{-p+1}}{-p+1} & \text{if } p \neq 1 \\ \log x & \text{if } p = 1 \end{cases}$$

Therefore,

$$\int_1^\infty x^{-p}\,dx = \begin{cases} \displaystyle\lim_{b\to\infty} \frac{b^{-p+1}-1}{-p+1} & \text{if } p \neq 1 \\ \displaystyle\lim_{b\to\infty} \log b & \text{if } p = 1 \end{cases} = \begin{cases} \dfrac{1}{p-1} & \text{if } p > 1 \\ \infty & \text{if } p = 1 \\ \infty & \text{if } p < 1 \end{cases}$$

since b^{-p+1} converges iff $-p+1 < 0$.

This example is interesting because it exhibits a "cutoff" between convergence and divergence. The function $1/x$ converges to 0 as $x \to \infty$, but it doesn't quite converge to 0 fast enough to have a finite improper integral. However, $1/x^p$ for any $p > 1$ has a finite improper integral. (See Exercise 9.7.30 for an interesting application of this!) $\square$

Example 9.32:

Consider the improper integral

$$\int_0^\infty \cos(\pi x)\,dx$$

For any $b > 0$, we have

$$\int_0^b \cos(\pi x)\,dx = \left.\frac{\sin(\pi x)}{\pi}\right|_0^b = \frac{\sin(\pi b)}{\pi}$$

since $\sin 0 = 0$. As $b \to \infty$, the function $\sin(\pi b)$ oscillates between -1 and 1, so our improper integral diverges.

However, it is worth noting that if we consider a SEQUENCE limit instead of a limit of a real variable, then

$$\lim_{n\to\infty} \int_0^n \cos(\pi x)\, dx = \lim_{n\to\infty} \frac{\sin(\pi n)}{\pi} = 0$$

because $\sin(\pi n) = 0$ for every integer n. This illustrates why we use a limit of a REAL variable to define improper integrals. $\square$

Remark. As with infinite series, the starting point of an improper integral does not affect whether the improper integral converges (although it does affect the value when the integral converges). To see this, suppose that $a \in \mathbb{R}$ is given, and f is a real function which is integrable on $[a, b]$ for every $b > a$. Suppose that $c > a$ is also given. For every $b > c$, we have

$$\int_a^b f(x)\, dx - \int_c^b f(x)\, dx = \int_a^c f(x)\, dx$$

so these integrals differ by something which does not depend on b. Thus, each improper integral converges as $b \to \infty$ iff the other does.

With this in mind, we can define improper integrals over $(-\infty, \infty)$ as

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{c} f(x)\, dx + \int_c^{\infty} f(x)\, dx$$

for any real value c, provided that the improper integrals on the right side converge. If either of the integrals on the right side diverges, then we say the integral over $(-\infty, \infty)$ diverges as well. As with the work above, you can check that the value of c is unimportant: if any one value of c causes the right side to have a well-defined sum, then all values of c cause the right side to have the same well-defined sum.

For instance,

$$\int_{-\infty}^{\infty} e^{-|x|}\, dx = \int_{-\infty}^{0} e^{x}\, dx + \int_0^{\infty} e^{-x}\, dx$$

converges (see Example 9.30), and its value is 2. Also, see Exercise 9.7.29.

The Integral Test And p-Series

Previously, we saw a connection between limits of integrals and limits of partial sums. Now, we want to make that connection precise using improper integrals of type 1. Suppose that $a > 0$ and we have a positive function $f : [a, \infty) \to \mathbb{R}$ which is integrable on $[a, b]$ for all $b > a$. We want to establish inequalities that relate

$$\int_a^\infty f(x)\,dx \qquad \text{and} \qquad \sum_{n=\lceil a \rceil}^\infty f(n)$$

(For instance, when considering $\sum(1/n^2)$, we can use $f(x) = 1/x^2$ and $a = 1$.) These inequalities will tell us a way to determine if partial sums of $\sum f(n)$ are bounded, so that we can use Theorem 9.26 to determine whether the series converges.

Before proceeding to the main result, the Integral Test, it will be worth our while to mention a useful theorem that, much like Theorem 9.26 for series, allows us to judge convergence or divergence of an improper integral when the integrand doesn't change sign:

Theorem 9.33. *Let $a \in \mathbb{R}$ be given, and suppose $f : [a, \infty) \to \mathbb{R}$ is integrable over $[a, b]$ for each $b > a$. Also, suppose that f never changes sign. Then the improper integral*

$$\int_a^\infty f(x)\,dx$$

converges iff the set of partial integrals

$$S = \left\{ \int_a^b f(x)\,dx \mid b \in (a, \infty) \right\}$$

is bounded. Furthermore, when S is bounded and $f(x) \geq 0$ for all $x \in [a, \infty)$, the value of the improper integral is $\sup S$. *Similarly, when S is bounded and $f(x) \leq 0$ for all $x \in [a, \infty)$, the value of the improper integral is* $\inf S$.

You can prove Theorem 9.33 in Exercise 9.7.25 (the proof should be quite similar to the proof of the BMCT). The use of this theorem is that it allows us to prove an improper integral converges by finding bounds. This leads us to the following test, motivated by the examples at the beginning of the section:

Theorem 9.34 (Integral Test). *Let $k \in \mathbb{N}$ be given, and let $f : [k, \infty) \to \mathbb{R}$ be given such that f is integrable over $[k, b]$ for all $b > k$. Also, suppose that $f(x) \geq 0$ for all $x \geq k$ and f is decreasing on $[k, \infty)$. Then*

$$\int_k^\infty f(x)\,dx \qquad \text{and} \qquad \sum_{n=k}^\infty f(n)$$

either both converge or both diverge.

Strategy. The main idea is much like the picture in Figure 9.7. Each term $f(n)$ of the series represents a rectangle $\#n$ with height $f(n)$ and width 1. If we place rectangles $\#k$ through $\#n$ under the graph of $y = f(x)$ by placing their upper-right corners on the graph, then we find

$$s_n = \sum_{i=k}^n f(i) = f(k) + \sum_{i=k+1}^n f(i) \leq f(k) + \int_k^n f(x)\,dx$$

Thus, if the integrals are all bounded (which follows from convergence of the improper integral by Theorem 9.33), then the partial sums s_n are also bounded. At this point, we may invoke Theorem 9.26.

On the other hand, we can also try placing the rectangles with their upper-left corners on the graph of $y = f(x)$, so that the rectangles lie over the graph. This gives us

$$\int_k^{n+1} f(x)\,dx \leq \sum_{i=k}^n f(k) = s_n$$

Thus, if the partial sums are all bounded, then we can show all the integrals are bounded.

Formally, instead of talking about "placing rectangles under/over the graph", we can use step functions. More specifically, let's say we're trying to show that the integral over $[k, n+1]$ is at most s_n. Since we want each rectangle to be one unit wide, we take a partition of $[k, n+1]$ into subintervals of length 1. Thus, a typical open subinterval has the form $(i, i+1)$ where $i \in \mathbb{N}$ and $k \leq i \leq n$. The rectangle which lies over the graph of $y = f(x)$ in this subinterval has height $f(i)$, because f is decreasing. Thus, for any $x \in (i, i+1)$, we get an upper step function by using the value $f(\lfloor x \rfloor)$ as the value on the subinterval. Similar computations show that a lower step function uses the value $f(\lceil x \rceil)$.

Proof. Let k, f be given as described. For each $n \in \mathbb{N}$ with $n \geq k$, and for each $b \in \mathbb{R}$ with $b > k$, we define

$$s_n = \sum_{i=k}^{n} f(i) \qquad \text{and} \qquad I(b) = \int_k^b f(x)\,dx$$

so s_n is the n^{th} partial sum and $I(b)$ is the partial integral up to b. Because $f(x) \geq 0$ for all $x \geq k$, I is an increasing function of b and (s_n) is increasing, so the partial integrals and partial sums are certainly bounded below. By Theorems 9.26 and Theorem 9.33, it remains to prove that (s_n) is bounded above if and only if all the partial integrals $I(b)$ are bounded above when $b > k$. In fact, since $I(b) \leq I(\lceil b \rceil)$ for all $b > k$, it suffices to show (s_n) is bounded above iff $(I(n))_{n=k}^{\infty}$ is bounded above.

In order to do this, our goal is to establish the following inequalities for all $n \in \mathbb{N}$ with $n \geq k$:

$$\int_k^{n+1} f(x)\,dx \leq s_n \leq f(k) + \int_k^n f(x)\,dx$$

Why does this suffice? If $(I(n))$ is bounded above, say by M, then the second inequality above shows that $s_n \leq f(k) + M$ for each n, so (s_n) is bounded above. Conversely, if (s_n) is bounded above, say by some number K, then the first inequality above shows that $I(n) \leq I(n+1) \leq K$ for all n.

To prove the first inequality of our goal, define $U : [k, n+1] \to \mathbb{R}$ by $U(x) = f(\lfloor x \rfloor)$ for all $x \in [k, n+1]$. Thus, U is a step function compatible with the partition of $[k, n+1]$ in which all subintervals have length 1. (More precisely, they have the form $[i, i+1]$ for $i \in \mathbb{N}$ from k to n.) Furthermore, because $\lfloor x \rfloor \leq x$ for all x, and f is decreasing, U is an upper step function for f on $[k, n+1]$. Thus, by the definition of integral, we have

$$\begin{aligned} \int_k^{n+1} f(x)\,dx &\leq \int_k^{n+1} U(x)\,dx \\ &= \sum_{i=k}^{n} f(i)((i+1) - i) = s_n \end{aligned}$$

To prove the second inequality of our goal, define $L : [k, n] \to \mathbb{R}$ by $L(x) = f(\lceil x \rceil)$ for all $x \in [k, n]$. Thus, L is a step function compatible with the partition of $[k, n]$ in which all subintervals have length 1. (More precisely, they have the form $[i-1, i]$ for $i \in \mathbb{N}$ from $k+1$ to n.) Because $x \leq \lceil x \rceil$ for all x, and f is decreasing, L is a lower step function for f on $[k, n]$. Thus, we have

$$\begin{aligned} \int_k^n f(x)\,dx &\geq \int_k^n L(x)\,dx \\ &= \sum_{i=k+1}^{n} f(i)(i-(i-1)) = s_n - f(k) \end{aligned}$$

as desired. ∎

When a series comes from a function which we know how to integrate, the Integral Test is often useful. For instance, consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

where $p \in \mathbb{R}$ is a constant. This series is sometimes called the *p-series*. For $p \leq 0$, the terms do not converge to 0, and this series diverges by the Divergence Test. When $p > 0$, we saw in Example 9.31 that

$$\int_1^{\infty} \frac{dx}{x^p}$$

converges iff $p > 1$. Thus, by applying the Integral Test to $f(x) = 1/x^p$ (note that f is positive and decreasing for $p > 0$), we find

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \text{ converges iff } p > 1$$

We will make frequent use of p-series in later sections, since much like geometric series, they can be used for a range of basic examples with which we compare other series.

Remark. The p-series occur in a very famous function, the *Riemann zeta function*:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \text{ for } s > 1$$

("zeta" is the Greek letter used for its name). Euler found many connections between the ζ function and number theory, and he computed the values of

ζ at even positive integers. Bernhard Riemann found a connection between the distribution of prime numbers and the zeroes of the zeta function when s is allowed to be complex. However, these results are beyond the scope of this book.[12]

Example 9.35:

For another example where we can use the Integral Test, consider

$$\sum_{n=1}^{\infty} \frac{\log n}{n}$$

This looks like it comes from the function $f(x) = (\log x)/x$, defined for $x > 0$. Note that $f(x) > 0$ when $x > 1$. However, it is not immediately clear whether f is decreasing. To determine where f decreases, we compute the derivative:

$$f'(x) = \frac{1 - \log x}{x^2}$$

This computation shows that f is decreasing on (e, ∞). Thus, in order to use f in the Integral Test, we need to pick a starting point larger than e. In particular, we will use the Integral Test to determine whether $\sum (\log n)/n$ converges starting from $k = 3$. (Remember, though, that the starting point doesn't affect whether the series converges!)

Now, we compute the improper integral via a substitution of $u = \log x$ and $du = dx/x$:

$$\begin{aligned}
\int_3^\infty \frac{\log x}{x}\,dx &= \lim_{b\to\infty} \int_3^b \frac{\log x}{x}\,dx \\
&= \lim_{b\to\infty} \int_{\log 3}^{\log b} u\,du = \lim_{b\to\infty} \left.\frac{u^2}{2}\right|_{\log 3}^{\log b} \\
&= \lim_{b\to\infty} \frac{\log^2 b - \log^2 3}{2} = \infty
\end{aligned}$$

Therefore, by the Integral Test, $\sum (\log n)/n$ diverges to ∞. We could have also discovered this by realizing that $\sum 1/n$ diverges to ∞ and noticing that $\sum (\log n)/n$ has larger terms when $n > e$. $\square$

[12] You may check out http://en.wikipedia.org/wiki/Riemann_zeta_function for more information about the zeta function.

An Error Analysis Of The Integral Test

The Integral Test can tell us more than just whether some series converge or diverge. For the series that converge, the Integral Test can give us estimates of the sum of the series! The main idea is to use the inequalities with s_n from the proof of the Integral Test to obtain estimates of the error we get by approximating the sum of the series with s_n.

To be specific, let's say that we are applying the Integral Test with a positive decreasing function $f : [k, \infty) \to \mathbb{R}$, which yields a series $\sum f(n)$ with partial sums $(s_n)_{n=k}^{\infty}$. Suppose that the series $\sum f(n)$ and the improper integral of f over $[k, \infty)$ both converge: let's say

$$\sum_{n=k}^{\infty} s_n = S \qquad \int_k^{\infty} f(x)\,dx = \lim_{b\to\infty} I(b) = I$$

where $I(b)$ is the partial integral over $[k, b]$. For any $N \in \mathbb{N}$ with $N \geq k$, we would like to find bounds on $S - s_N$, the error in approximating the sum S of the series with the N^{th} partial sum s_N.

We note that

$$S - s_N = \sum_{i=k}^{\infty} a_i - \sum_{i=k}^{N} a_i = \sum_{i=N+1}^{\infty} a_i$$

To estimate this, we consider the Integral Test applied with the starting point of $N + 1$. The proof shows that for all $n \in \mathbb{N}$ with $n \geq N + 1$,

$$\int_{N+1}^{n+1} f(x)\,dx \leq \sum_{i=N+1}^{n} f(i) \leq f(N+1) + \int_{N+1}^{n} f(x)\,dx$$

In fact, since $f(N + 1)$ is the smallest value f takes on $[N, N + 1]$ (as f is decreasing), we have

$$\int_{N+1}^{n+1} f(x)\,dx \leq \sum_{i=N+1}^{n} f(i) \leq \int_{N}^{N+1} f(x)\,dx + \int_{N+1}^{n} f(x)\,dx = \int_{N}^{n} f(x)\,dx$$

We may take a limit as $n \to \infty$ of all the parts of this inequality (because they all converge), and we obtain

$$\int_{N+1}^{\infty} f(x)\,dx \leq \sum_{i=N+1}^{\infty} f(i) = S - s_N \leq \int_{N}^{\infty} f(x)\,dx$$

It is straightforward to check that

$$\int_N^\infty f(x)\,dx = I - I(N) \qquad \int_{N+1}^\infty f(x)\,dx = I - I(N+1)$$

Because $I(b) \to I$ as $b \to \infty$, it follows that $I - I(N)$ and $I - I(N+1)$ can be made as close to 0 as desired by taking N large enough. We summarize this in the following corollary:

Corollary 9.36 (Integral Test Error Estimates). *Let $k, f, (s_n)$ be given as described in the Integral Test, and suppose that the improper integral of f over $[k, \infty)$ converges. Therefore, the series $\sum f(n)$ converges, say S is its sum. Then we have*

$$\int_{N+1}^\infty f(x)\,dx \le S - s_N \le \int_N^\infty f(x)\,dx$$

for all $N \in \mathbb{N}$ with $N \ge k$. Furthermore, all of the parts of this inequality go to 0 as $N \to \infty$.

Example 9.37:

We already saw that the p-series with $p = 2$,

$$S = \sum_{n=1}^\infty \frac{1}{n^2}$$

converges by the Integral Test. Let's use Corollary 9.36 to estimate its sum. For instance, we'd like to know which partial sum s_N gives us an estimate with error less than 0.1.

For any $N \in \mathbb{N}$, we know

$$\begin{aligned}\int_N^\infty \frac{dx}{x^2} &= \lim_{b\to\infty} \int_N^b \frac{dx}{x^2} = \lim_{b\to\infty} \left.\frac{-1}{x}\right|_N^b \\ &= \lim_{b\to\infty} \left(\frac{1}{N} - \frac{1}{b}\right) = \frac{1}{N}\end{aligned}$$

Therefore, the corollary tells us that the error $S - s_N$ is between $1/(N+1)$ and $1/N$. To guarantee that the error is at most $1/10$, we can choose $N = 10$. Therefore,

$$S \approx s_{10} = \frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{10^2} \approx 1.55$$

However, note that $1/N$ converges to 0 pretty slowly as $N \to \infty$, so you need a pretty large value of N with this method to get a good estimate. We can get a better estimate by using our error bounds more efficiently. Since we know that the error is at least $1/(N+1)$, $s_N + 1/(N+1)$ is a better estimate of S than s_N is. (Essentially, we take s_N and shift it up by the lower bound of our error.) Since

$$\frac{1}{N+1} \le S - s_N \le \frac{1}{N}$$

we find that

$$0 \le S - \left(s_N + \frac{1}{N+1}\right) \le \frac{1}{N} - \frac{1}{N+1} = \frac{1}{N(N+1)}$$

Therefore, when approximating by $s_N + 1/(N+1)$, the error is guaranteed to be proportional to $1/N^2$, which is much better than $1/N$. When $N = 10$, this tells us $S \approx s_{10} + 1/11 \approx 1.64$ with error at most $1/110$.

In fact, we can do even better: rather than shifting up by the lower bound of the error, let's shift up by the average of our lower and upper error bounds! This means that we estimate S with $S + (1/N + 1/(N+1))/2$. By doing similar steps to those above, we find

$$\frac{-1}{2N(N+1)} \le S - \left(s_N + \frac{1}{2}\left(\frac{1}{N} + \frac{1}{N+1}\right)\right) \le \frac{1}{2N(N+1)}$$

i.e.

$$\left|S - \left(s_N + \frac{1}{2}\left(\frac{1}{N} + \frac{1}{N+1}\right)\right)\right| \le \frac{1}{2N(N+1)}$$

Using $N = 10$ here, we find that $S \approx 1.645$, with an error of magnitude at most $1/220$. (With this approach, to guarantee an error with magnitude less than $1/10$, the choice of $N = 2$ suffices!) $\square$

Improper Integrals Of Type 2

As we have now seen, type 1 improper integrals can be very helpful for analyzing series. However, there are other ways in which an integral can be improper. For instance, the integrand can have a vertical asymptote, i.e.a place where the integrand goes to $\pm\infty$.[13] How would we make a meaningful definition of integral for these types of functions?

For instance, let's suppose that f is a real function defined on some interval $[a, b)$, but f has a vertical asymptote at b. Intuitively, if we integrate f over $[a, c]$ instead of over $[a, b]$, where $c \in (a, b)$, then we'd like to have

$$\int_a^c f(x)\,dx \approx \int_a^b f(x)\,dx \qquad \text{when } c \approx b$$

This suggests that as c gets closer and closer to b, the integral over $[a, c]$ does a better and better job at estimating the integral over $[a, b]$. Thus, we are led to the following definition:

Definition 9.38. Let $a, b \in \mathbb{R}$ be given, and let $f : [a, b) \to \mathbb{R}$ be a function which is integrable on $[a, c]$ for all $c \in (a, b)$. Also, suppose that f has a vertical asymptote at b. Then the *improper integral (of type 2) of f over $[a, b]$* is

$$\int_a^b f(x)\,dx = \lim_{c \to b^-} \int_a^c f(x)\,dx$$

(where c is real). When this limit exists, we say the improper integral *converges*, otherwise it *diverges*.

Similarly, if f is defined on $(a, b]$ and integrable on $[c, b]$ for all $c \in (a, b)$, but has a vertical asymptote at a, then the *improper integral of f over $[a, b]$* is

$$\int_a^b f(x)\,dx = \lim_{c \to a^+} \int_c^b f(x)\,dx$$

Note that for both of these definitions, we require that we have exactly one vertical asymptote at a limit of integration.

Example 9.39:

For any $p \in \mathbb{R}$, consider

$$\int_0^1 \frac{dx}{x^p}$$

[13]Integrals of this type tend not to be useful in analyzing series, but we mention them here since they are a natural generalization of type 1 integrals. Furthermore, there are some particularly useful functions which are written as type 2 improper integrals, as we will explore in the next section.

Note that when $p \leq 0$, the integrand is bounded on $[0, 1]$, and the integral is proper. For $p > 0$, the integral is improper of type 2 because we have a vertical asymptote at $x = 0$. We have the following antiderivative for $x \in (0, 1)$:

$$\int \frac{dx}{x^p} = \begin{cases} \dfrac{x^{-p+1}}{-p+1} & \text{if } p \neq 1 \\ \log x & \text{if } p = 1 \end{cases}$$

Therefore,

$$\int_0^1 \frac{dx}{x^p} = \begin{cases} \displaystyle\lim_{a \to 0^+} \frac{1 - a^{-p+1}}{-p+1} & \text{if } p \neq 1 \\ \displaystyle\lim_{a \to 0^+} -\log a & \text{if } p = 1 \end{cases} = \begin{cases} \frac{1}{1-p} & \text{if } 0 < p < 1 \\ \infty & \text{if } p = 1 \\ \infty & \text{if } p > 1 \end{cases}$$

since a^{-p+1} converges as $a \to 0^+$ iff $-p + 1 > 0$.

Thus, in this example, we have convergence of the improper integral iff $p < 1$. Contrast this with Example 9.31, which shows the type 1 integral over $[1, \infty)$ converges iff $p > 1$! $\square$

There are other extensions to the definition of improper integral which allow us to integrate functions with more asymptotes. The main idea is to keep using the interval addition property to break up an integral until each subproblem has only one vertical asymptote located at a limit of integration. Rather than write a complicated formal statement to describe this precisely, we demonstrate one example:

Example 9.40:

The integral

$$\int_{-1}^{2} \frac{dx}{x(x-1)}$$

is improper because it has vertical asymptotes at 0 and 1. We split this up as

$$\int_{-1}^{0} \frac{dx}{x(x-1)} + \int_{0}^{1/2} \frac{dx}{x(x-1)} + \int_{1/2}^{1} \frac{dx}{x(x-1)} + \int_{1}^{2} \frac{dx}{x(x-1)}$$

so that each of the four subproblems has only one asymptote occurring at a limit of integration. Our original integral is said to converge iff each of the four subproblems converges. (The choice of 1/2 as a splitting point is just for convenience; you can check that any point in $(0,1)$ can be used without affecting the final answer.)

We find an antiderivative by using partial fractions (some details are left to the reader):

$$\int \frac{dx}{x(x-1)} = \int \left(\frac{1}{x-1} - \frac{1}{x}\right) dx = \log|x-1| - \log|x| + C$$

Since $\log|x|$ approaches $-\infty$ as $x \to 0^-$, the subproblem

$$\int_{-1}^{0} \frac{dx}{x(x-1)}$$

diverges. As a result, our original integral is divergent. (In fact, each of the four subproblems is divergent!) □

We also remark that we can make other examples of improper integrals by mixing type 1 and type 2 definitions. For instance, consider

$$\int_{0}^{\infty} \frac{e^{-x}}{\sqrt{x}}\, dx$$

This integrand has a vertical asymptote at 0, so the problem needs to be broken into two subproblems, such as one over $[0,1]$ (of type 2) and one over $[1,\infty)$ (of type 1). However, it is very difficult to obtain an antiderivative for this integrand. We will return to this problem in the next section, when we have another tool for analyzing convergence of improper integrals.

9.6 Comparison-Based Tests

In the last few sections, we presented some theorems that help us analyze series (such as Theorem 9.26 which simplifies our work in checking nonnegative series), and we also studied two important basic types of series. The first is the geometric series $\sum r^n$, which converges iff $|r| < 1$. The second is the p-series $\sum n^{-p}$, which converges iff $p > 1$. Many other series can be analyzed by relating them to one of these basic types.

As a simple example, consider the nonnegative series

$$\sum_{n=2}^{\infty} \frac{1}{n^2+1}$$

Suppose that the partial sums of this sequence are $(s_n)_{n=2}^{\infty}$; we'd like to know whether (s_n) is bounded above. Note that our series resembles $\sum 1/n^2$. In fact, for any $n \geq 2$, we have $1/(n^2+1) < 1/n^2$. It follows that

$$\frac{1}{2^2+1} + \frac{1}{3^2+1} + \cdots + \frac{1}{n^2+1} < \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$$

In other words, if $(t_n)_{n=2}^{\infty}$ is the sequence of partial sums of $\sum 1/n^2$, then $s_n \leq t_n$ for each $n \geq 2$. Since $\sum 1/n^2$ converges, (t_n) is bounded above, so (s_n) is also bounded above. Thus, our original series converges.

By modifying this reasoning, you can readily prove the following theorem in Exercise 9.7.26:

Theorem 9.41 (Comparison Test). *Let $k \in \mathbb{N}^*$ be given, and let $(a_n)_{n=k}^{\infty}$, $(b_n)_{n=k}^{\infty}$ be two nonnegative sequences. If $a_n \leq b_n$ for all $n \geq k$, and $\sum b_n$ converges, then $\sum a_n$ converges as well. (We say that $\sum a_n$ is* dominated *by $\sum b_n$, or that $\sum b_n$* dominates *$\sum a_n$.) Equivalently (by contrapositive), if $b_n \leq a_n$ for all $n \geq k$, and $\sum b_n$ diverges, then $\sum a_n$ diverges as well.*

The main idea of the Comparison Test (sometimes abbreviated CT) is that we can analyze a nonnegative series $\sum a_n$ by comparing it against a previously-analyzed nonnegative series $\sum b_n$. An informal way of summarizing the theorem is "smaller than finite is finite, bigger than infinite is infinite". In other words, if a series $\sum a_n$ is dominated by a series $\sum b_n$ with a finite sum, then $\sum a_n$ also must have a finite sum. Intuitively, when $a_n \leq b_n$ for all large n, a_n goes to 0 at least as quickly as b_n does.

We present some examples using the Comparison Test. Note that the previously-known series we use in these examples are frequently p-series or geometric series. Also, CT usually takes much less work to apply than the Integral Test, since we don't have to check in CT whether terms are decreasing, nor do we have to compute any integrals. (The challenge in applying CT is picking the correct inequality to use!)

Example 9.42:

At the beginning of Section 9.3, we saw that

$$\sum_{n=0}^{\infty} \frac{1}{n!} = e$$

We can also prove that this series converges by using CT. Let $a_n = 1/n!$ for $n \in \mathbb{N}$. Intuitively, we know that a_n approaches 0 quite rapidly, so we suspect that $\sum a_n$ converges. This means we want to find some larger sequence (b_n) such that $\sum b_n$ converges.

In fact, we have

$$\frac{1}{n!} \leq \frac{1}{2^{n-1}}$$

for all $n \geq 1$, which can readily be proven by induction on n. (Also, see Example 1.59.) Thus, we can apply the Comparison Test with $b_n = 1/2^{n-1}$ and $k = 1$. Since $\sum b_n$ is a geometric series with ratio $1/2$, $\sum b_n$ converges, and $\sum a_n$ converges as well by CT. (Note, though, that CT only tells us convergence, not the value of the sum, so our previous work actually established a stronger result.)

It is worth noting that it can be unclear at first which inequalities will be useful for CT. For instance, it is quite simple to notice that

$$\frac{1}{n!} \leq \frac{1}{n}$$

for all $n \in \mathbb{N}$ with $n \geq 1$. This suggests that we try CT with $a_n = 1/n!$ and $b_n = 1/n$. However, $\sum b_n$ diverges, so CT tells us no information about $\sum a_n$. (In other words, we only showed that $\sum a_n$ is "no bigger than infinite", but that doesn't tell us whether it is finite!) $\square$

Example 9.43:

For another example involving factorials, we saw in Example 9.8 that

$$\frac{n!}{n^n} \leq \frac{k!}{n^k}$$

for any $k \in \mathbb{N}$ and any $n \in \mathbb{N}$ with $n \geq k$. When $k \geq 2$, the series $\sum k!/n^k$ is a convergent p-series (where n is the index of summation, and k is a constant with respect to n). Therefore, $\sum n!/n^n$ converges by the Comparison Test. (Note that the choice of $k = 1$ would not yield a useful conclusion, by the same reasoning as the last example.) $\square$

Example 9.44:

Consider

$$\sum_{n=1}^{\infty} \frac{|\sin n|}{n^p}$$

where $p > 0$ is a constant. (Note that we need the absolute-value bars in the numerator to have a nonnegative series!) Perhaps the most straightforward inequality for trigonometric functions is $|\sin n| \le 1$. Thus, our series is dominated by $\sum 1/n^p$. When $p > 1$, CT implies that our series converges.

However, when $p \le 1$, the inequality $|\sin n|/n^p \le 1/n^p$ cannot be used with CT to prove convergence, because $\sum 1/n^p$ diverges. Thus, we may suspect that our original series diverges for $p \le 1$. This leads us to try establishing some inequality of the form

$$\frac{|\sin n|}{n^p} \ge \frac{C}{n^p}$$

for some positive constant C. Unfortunately, though, no such inequality can be proven for all n, since it turns out $\sin n$ can be made arbitrarily close to 0 infinitely often. (More precisely, for every $\epsilon > 0$ and for every $k \in \mathbb{N}$, there is some $n \in \mathbb{N}$ with $n \ge k$ and $|\sin n| < \epsilon$.) Nevertheless, when $C = 1/2$, it is possible to show that this inequality holds for enough values of n to make the partial sums unbounded: see Exercise 9.7.31. □

Example 9.45:

For an example involving terms which go to 0 slowly, consider

$$\sum_{n=2}^{\infty} \frac{1}{\log^p n}$$

where p is a positive constant. On first glance, we might try to compare this with $1/n^p$ and consider cases based on whether $p > 1$. However, upon further consideration, we recall that $\log n$ goes to ∞ more slowly than any power of n: to be precise, Theorem 7.36 says that for any $a > 0$ and any $n > 1$, we have

$$\log n < \frac{n^a}{a}$$

Therefore, for all $a, p > 0$, we have

$$\frac{1}{\log^p n} > \frac{a^p}{n^{ap}}$$

Choosing a to be $1/p$, we find that our series dominates $\sum((1/p)^p)/n$, where n is the index of summation. Since $(1/p)^p$ is a constant with respect to n, our series diverges by CT. □

Limit Comparison Test

Sometimes, we would like to apply the Comparison Test, but the inequality we find goes in the wrong direction to be helpful. For instance, consider

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n^2+1}}$$

Since $\sqrt{n^2+1} > \sqrt{n^2} = n$ for all $n \in \mathbb{N}^*$, our series is dominated by $\sum 1/n$. However, $\sum 1/n$ diverges, so this comparison does not help us. (We get no information by showing the series is dominated by a divergent series.)

Since $\sqrt{n^2+1}$ is approximately equal to n when n is large, this suggests that we should be able to compare our series to $\sum 1/n$ and conclude divergence. Thus, for large values of n, we would like to establish

$$\frac{1}{\sqrt{n^2+1}} \geq \frac{C}{n}$$

where C is some positive constant. This is equivalent to showing that for large enough values of n,

$$\frac{n}{\sqrt{n^2+1}} \geq C$$

for some $C > 0$. Luckily, we know a useful piece of information about the ratio $n/\sqrt{n^2+1}$: we can show that it approaches 1 as $n \to \infty$. Thus, for any $\epsilon > 0$, if n is sufficiently large, then the ratio is between $1-\epsilon$ and $1+\epsilon$. In particular, choosing $\epsilon = 1/2$, we find that for n sufficiently large,

$$\frac{n}{\sqrt{n^2+1}} \geq \frac{1}{2}$$

This allows us to use CT to prove divergence of our original series.

This example suggests that when we want to compare a nonnegative sequence (a_n) with another nonnegative sequence (b_n), the ratios (a_n/b_n) can provide useful information for the Comparison Test. For this reason, we introduce the following definition:

Definition 9.46. Let (a_n) and (b_n) be two infinite real-valued sequences with $a_n, b_n \neq 0$ for all n sufficiently large. We say that (a_n) and (b_n) are *asymptotically equal* if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1$$

This is also written as "$a_n \sim b_n$ as $n \to \infty$".

Intuitively, if two sequences (a_n) and (b_n) are asymptotically equal, then they are approximately the same for large values of n. (In our previous example, $1/\sqrt{n^2+1} \sim 1/n$.) More generally, if a_n/b_n approaches some $C \in (0,\infty)$ as $n \to \infty$, then $a_n \sim Cb_n$ and (a_n) and (b_n) are essentially proportional. This suggests that $\sum a_n$ and $\sum Cb_n$ should have the same convergence status. The Limit Comparison Test (or LCT for short) makes this precise:

Theorem 9.47 (Limit Comparison Test). *Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$, $(b_n)_{n=k}^{\infty}$ be two positive sequences. If*

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

converges to a positive real number, then $\sum a_n$ and $\sum b_n$ both converge or both diverge. In particular, if $a_n \sim b_n$ as $n \to \infty$, then both $\sum a_n$ and $\sum b_n$ converge or both diverge.

Remark. The LCT requires the ratio a_n/b_n to approach a positive number. However, we can still get some useful information from other values of the limit. If $a_n/b_n \to 0$ (i.e. $a_n = o(b_n)$), then convergence of $\sum b_n$ implies convergence of $\sum a_n$, but not necessarily conversely. Similarly, if $a_n/b_n \to \infty$, then divergence of $\sum b_n$ implies divergence of $\sum a_n$, but not conversely. You can prove these results in Exercise 9.7.27.

Strategy. Let's say that $a_n/b_n \to L$ as $n \to \infty$ where $L > 0$. Intuitively, this means $a_n/b_n \approx L$, so $a_n \approx Lb_n$. More precisely, for any $\epsilon > 0$, we have $a_n/b_n \in (L-\epsilon, L+\epsilon)$ for sufficiently large n. Hence,

$$(L-\epsilon)b_n < a_n < (L+\epsilon)b_n$$

If we pick ϵ so that $L - \epsilon > 0$, then we have two possible comparisons to use with CT, depending on whether we wish to establish convergence or divergence of $\sum a_n$.

Proof. Let $k, (a_n), (b_n)$ be given as described. Assume that

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

is in $(0,\infty)$: call that limit L. By the definition of limit, there is some $N \in \mathbb{N}$ with $N \geq k$ such that for all $n \in \mathbb{N}$ with $n \geq N$, we have

$$\left|\frac{a_n}{b_n} - L\right| < \frac{L}{2}$$

It follows by some simple inequalities that

$$\frac{L}{2}b_n < a_n < \frac{3L}{2}b_n$$

Now, we wish to show $\sum a_n$ and $\sum b_n$ either both converge or both diverge. For the first case, suppose that $\sum b_n$ converges. Then $\sum(3L/2)b_n$ is a convergent series with positive terms, so $\sum a_n$ converges as well by the Comparison Test. For the second case, suppose that $\sum b_n$ diverges. Then $\sum(L/2)b_n$ is a divergent series with positive terms, so $\sum a_n$ diverges as well by the Comparison Test. ■

Frequently, the Limit Comparison Test is easier to apply to a series than the Comparison Test. We offer the following examples:

Example 9.48:

Suppose we consider

$$\sum_{n=1}^{\infty} \frac{n^3+2n^2+1}{2n^5+n^3-1}$$

Let's say a_n is the n^{th} term of this series. For both the numerator and denominator, the highest-degree terms are the most important as n grows large. Thus, we claim that $a_n \sim b_n$ where $b_n = n^3/(2n^5) = 1/(2n^2)$. To verify this, we calculate

$$\begin{aligned}
\lim_{n\to\infty} \frac{a_n}{b_n} &= \lim_{n\to\infty} \frac{n^3+2n^2+1}{n^3} \cdot \frac{2n^5}{2n^5+n^3-1} \\
&= \lim_{n\to\infty} \frac{1+2/n+1/n^3}{1} \cdot \frac{2}{2+1/n^2-1/n^5} \\
&= 1 \cdot 1 = 1
\end{aligned}$$

(note that we split up the parts of a_n and b_n appropriately to make it easy to multiply and divide by appropriate powers of n). Since $\sum 1/(2n^2)$ is convergent, the LCT implies that $\sum a_n$ converges as well.

Note that because we took a limit to use the LCT, we could focus on the leading terms right away and ignore everything else. In contrast, if we were to use the Comparison Test, then we would have to show $a_n \le Cb_n$ for some constant C, which is a fairly technical task: we cannot just naively make the numerator bigger and the denominator smaller since there are sums in both the numerator and denominator. The LCT also saves us from worrying about which direction of inequality we can establish. □

Example 9.49:

For a similar example, consider

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n(n+1)}}$$

Here, we claim that the terms are asymptotically equal to $1/\sqrt{n^2}$, i.e. $1/n$. To see this, we compute

$$\lim_{n\to\infty} \frac{1/\sqrt{n(n+1)}}{1/\sqrt{n^2}} = \sqrt{\lim_{n\to\infty} \frac{n^2}{n(n+1)}} = \sqrt{\lim_{n\to\infty} \frac{1}{1+1/n}} = 1$$

Since $\sum 1/n$ diverges, the LCT says our original series does too. □

Example 9.50:

Consider the following two series:

$$\sum_{n=1}^{\infty} \frac{\cos(1/n)}{n} \qquad \sum_{n=1}^{\infty} \frac{\sin(1/n)}{n}$$

Since $\cos(1/n) \to 1$ as $n \to \infty$, the first series diverges by using the Limit Comparison Test with $b_n = 1/n$. However, in the second series, $\sin(1/n) \to 0$ as $n \to \infty$, and we are not allowed to use the LCT with $b_n = 0/n$. To deal with this, we should analyze how quickly $\sin(1/n)$ goes to 0.

To simplify our notation, let $x = 1/n$, so that $x \to 0^+$ as $n \to \infty$. We already know $(\sin x)/x \to 1$ as $x \to 0$. Thus, $\sin x \sim x$ as $x \to 0^+$, i.e. $\sin(1/n) \sim 1/n$ as $n \to \infty$. Hence, the terms of our second series are asymptotically equal to $1/n^2$, and $\sum 1/n^2$ converges, so our second series converges by the LCT.

It is worth mentioning that we can also use Taylor approximations here! We know that $T_1 \sin(x) = x$, so $\sin x = x + o(x)$ as $x \to 0^+$. From here, it is easy to see that $\sin x \sim x$ as $x \to 0^+$. In fact, we can use any Taylor approximation to get an estimate: for instance, if we instead use $T_3 \sin(x)$, so $\sin x = x - x^3/6 + o(x^3)$, then we can show that $\sin x \sim x - x^3/6$ as $x \to 0^+$. However, since $x^3/6 = o(x)$ as $x \to 0^+$ (and in fact, $x \sim x - x^3/6$ as $x \to 0^+$), this 3$^{\text{rd}}$-order approximation does not yield any more useful information than the 1$^{\text{st}}$-order approximation did. □

The Limit Comparison Test is a very useful consequence of the Comparison Test, and it makes many problems easier, but the Comparison Test still has its uses. For instance, problems involving $\sin n$ and $\cos n$, such as Example 9.44, are easier to analyze with inequalities because $\sin n$ and $\cos n$ oscillate as $n \to \infty$. Also, the Limit Comparison Test can fail to yield a useful conclusion with some series where the terms have parts going to ∞ at very different speeds, as the next example indicates:

Example 9.51:

Consider the series

$$\sum_{n=1}^{\infty} \frac{n}{2^n}$$

This series is difficult to analyze with the Limit Comparison Test. Suppose $a_n = n/2^n$. If we try choosing $b_n = 1/2^n$, then $a_n/b_n = n \to \infty$. If we try choosing $b_n = n$, then $a_n/b_n = 1/2^n \to 0$. When we look at our collection of basic examples, such as the p-series and the geometric series, there seems to be no good choice for b_n which makes a_n/b_n approach a positive finite ratio.

Instead, let's use the Comparison Test. (The Integral Test also works, though the integration problem uses integration by parts, which arguably requires more work than the method we will demonstrate with CT.) Recall that Theorem 7.30 roughly says that any power of n goes to ∞ much more slowly than any exponential. Thus, not only does $n/2^n$ go to 0, but n/b^n also goes to 0 for any $b > 1$. This is useful because we can write

$$\frac{n}{2^n} = \frac{n}{b^n(2/b)^n} = \frac{n/b^n}{(2/b)^n}$$

where the numerator goes to 0 and the denominator grows exponentially.

In particular, let's choose $b = \sqrt{2}$ (though any $b \in (1, 2)$ will work). Since $n/(\sqrt{2})^n \to 0$, we have the following when n is sufficiently large:

$$\frac{n}{2^n} = \frac{n/(\sqrt{2})^n}{(\sqrt{2})^n} \le \frac{1}{(\sqrt{2})^n}$$

This means that $\sum n/2^n$ is dominated by $\sum 1/(\sqrt{2})^n$, which is a geometric series with ratio $1/\sqrt{2}$. Hence, our series converges by CT. In fact, you can use this method of "breaking a large denominator into two moderately large pieces" (in this case, breaking 2^n into $(\sqrt{2})^n(\sqrt{2})^n$) in Exercise 9.7.28

to show that when p is any polynomial, and b is any number bigger than 1, $\sum |p(n)|/b^n$ converges. $\square$

Comparison Tests For Improper Integrals

The main idea behind the Comparison Test is that a convergent positive series can be used to provide an upper bound for the partial sums of another positive series. Analogously, when dealing with improper integrals, a convergent improper integral of a positive function can be used to bound partial integrals of other improper integrals. More precisely, you can prove the following:

Theorem 9.52 (Comparison Test for Integrals). *For all parts of this theorem, suppose that I is a (possibly unbounded) interval of $\mathbb{R}$, and $f, g : I \to \mathbb{R}$ are positive on I. Also, suppose that $f(x) \le g(x)$ for all $x \in I$.*

1. *If I has the form $[a, \infty)$ for some $a \in \mathbb{R}$, then*

$$\int_a^\infty g(x)\,dx \text{ converges} \to \int_a^\infty f(x)\,dx \text{ converges}$$

   A similar statement holds when I has the form $(-\infty, b]$ for some $b \in \mathbb{R}$.

2. *If I has the form $[a, b)$ for some $a, b \in \mathbb{R}$, such that f or g has a vertical asymptote at b, then*

$$\int_a^b g(x)\,dx \text{ converges} \to \int_a^b f(x)\,dx \text{ converges}$$

   The same statement holds when I has the form $(a, b]$ for some $a, b \in \mathbb{R}$ such that f or g has a vertical asymptote at a.

In summary, if the larger function has a convergent improper integral, then the smaller function does too. As a contrapositive, if the smaller function has a divergent improper integral, then the larger function does too.

There is also a Limit Comparison Test for improper integrals, which is proven using the Comparison Test in the same way as for series. We state this theorem only for type 1 improper integrals, though it is not hard to state and prove a version of this theorem for type 2 integrals as well:

Theorem 9.53 (Limit Comparison Test for Integrals). *Let $a \in \mathbb{R}$ be given, and let $f, g : [a, \infty) \to \mathbb{R}$ be real functions which are positive on $[a, \infty)$. Suppose also that*

$$\lim_{x \to \infty} \frac{f(x)}{g(x)}$$

converges to a positive real number. Then the improper integrals

$$\int_a^\infty f(x)\, dx \qquad \int_a^\infty g(x)\, dx$$

both converge or both diverge. Similar statements hold for other types of improper integrals.

Remark. When f and g are decreasing functions, one way to prove these theorems is to apply the Integral Test and then use the corresponding theorems for series. However, these theorems for integrals are more general: they do not require f or g to be decreasing. Instead of using the Integral Test, the proofs of these theorems merely imitate the arguments used in proving the CT and LCT for series.

As with CT and LCT for series, the CT and LCT for improper integrals allow us to analyze convergence or divergence of integrals which are too difficult to evaluate directly. We present a few examples.

Example 9.54:

Consider the improper integral

$$\int_0^\infty e^{-x^2}\, dx$$

This integrand does not have an elementary antiderivative. However, we can analyze this by using CT for integrals. Note that e^{-x^2} goes to 0 very rapidly; in fact, for any $x \geq 1$, we have $x^2 \geq x$, so $e^{-x^2} \leq e^{-x}$.

Therefore, our improper integral is dominated by e^{-x} on $[1, \infty)$. As Example 9.30 shows (or we can use the Integral Test),

$$\int_1^\infty e^{-x}\, dx \text{ converges.}$$

Therefore,

$$\int_0^\infty e^{-x^2}\,dx = \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx$$

converges, as it is the sum of an ordinary integral and an integral which is convergent by the Comparison Test. □

Example 9.55:

Consider the improper integral

$$\int_0^\infty \frac{e^{-x}}{\sqrt{x}}$$

This integrand has a vertical asymptote at 0, so we must break up this integral. One way to do this is to write

$$\int_0^1 \frac{e^{-x}}{\sqrt{x}}\,dx + \int_1^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$$

In order for this to converge, each of these two integrals must converge.

First, let's consider the integral over $[1,\infty)$. As in the previous example, the integrand is dominated by e^{-x}. (It is useful to think of this as "the exponential goes to 0 faster than $1/\sqrt{x}$, so we bound $1/\sqrt{x}$ by a constant to focus on the dominant part e^{-x}".) Thus, this integral converges by CT.

Second, let's consider the integral over $[0,1]$. This time, the $\sqrt{x}$ in the denominator is the part that causes problems when x is close to 0, so $1/\sqrt{x}$ is the dominant part. Hence, we find a bound on e^{-x} to simplify the problem. Note that $e^{-x} \le 1$ on $[0,1]$, so our integrand is dominated by $1/\sqrt{x}$ on $[0,1]$. By the result of Example 9.39, we find that

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx \text{ converges.}$$

Therefore, our original integral also converges by CT.

The strategy presented in this example is typical for improper integrals involving a few different types of functions in the integrand. The main idea is that in an improper integral, usually some of the functions in the integrand can be bounded, allowing us to focus on the "dominant" functions. For instance, in the integral over $[1,\infty]$ above, $1/\sqrt{x}$ was not the dominant part, so it was bounded by 1. In the integral over $[0,1]$, e^{-x} was not the dominant part, so it was bounded by 1 as well. $\square$

To end this section, let's address one particularly difficult, but interesting, type of improper integral. Sometimes, an improper integral can naturally be broken into pieces indexed by integers. For instance,

$$\int_1^\infty f(x)\,dx$$

can be broken into pieces of the form

$$\int_n^{n+1} f(x)\,dx$$

for each $n \in \mathbb{N}^*$. Suppose f is positive on $[1,\infty)$, and suppose we find numbers $a_n, b_n \in \mathbb{R}$ depending only on n satisfying

$$a_n \le \int_n^{n+1} f(x)\,dx \le b_n$$

It follows that

$$\sum_{i=1}^n a_i \le \sum_{i=1}^n \int_i^{i+1} f(x)\,dx = \int_1^{n+1} f(x)\,dx \le \sum_{i=1}^n b_i$$

Letting n approach ∞, we find that the improper integral over $[1,\infty)$ falls in between the series $\sum a_n$ and $\sum b_n$. Therefore, we can use information about series to study improper integrals, whereas beforehand we used improper integrals to study series! We illustrate this idea with one example: see Exercise 9.7.32 as well.

Example 9.56:

Consider the improper integral

$$\int_\pi^\infty \frac{|\sin x|}{x}\,dx$$

This integral resembles the series

$$\sum_{n=1}^\infty \frac{|\sin n|}{n}$$

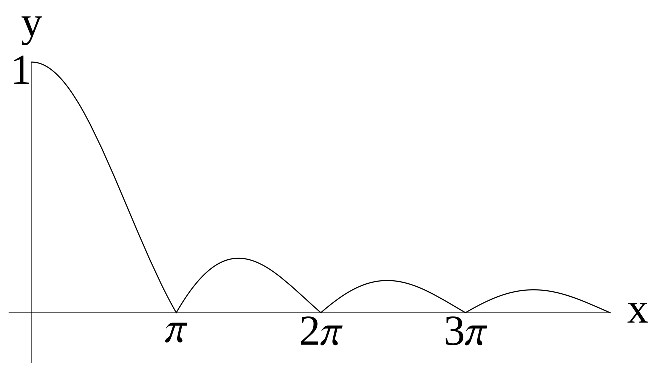

Figure 9.8: Part of the graph $y = |\sin x|/x$ on $(0, \infty)$

from Example 9.44. However, we cannot use the Integral Test to relate these two problems, because our integrand is not a decreasing function of x. We need a new approach to analyze our integral.

To get an idea of how to proceed, we consider the graph of $|\sin x|/x$ in Figure 9.8. Note that the curve touches the x-axis precisely at the integer multiples of π. This suggests that on $[\pi, \infty)$, the area under the curve can be broken into "hills", the first on $[\pi, 2\pi]$, the second on $[2\pi, 3\pi]$, and so forth. Thus, we define

$$I_n = \int_{n\pi}^{(n+1)\pi} \frac{|\sin x|}{x}\, dx$$

for each $n \in \mathbb{N}^*$ to be the area of the n^{th} "hill". It is not hard to show that our improper integral on $[\pi, \infty)$ converges iff the series $\sum I_n$ converges.

To determine whether the series converges, we would like to estimate I_n to apply either CT or LCT. In the integral defining I_n, the integrand does not have an elementary antiderivative, so we should bound either the numerator or the denominator. Suppose we try and bound the numerator by using the inequality $|\sin x| \le 1$. Then

$$I_n \le \int_{n\pi}^{(n+1)\pi} \frac{dx}{x} = \log|(n+1)\pi| - \log|n\pi| = \log\left(\frac{n+1}{n}\right)$$

In Example 9.19, we found that the series $\sum \log((n+1)/n)$ diverges (it telescopes, though you could also apply the LCT to compare it with $1/n$). Hence, we have dominated $\sum I_n$ by a divergent series, which does not tell us anything useful.

Therefore, instead of bounding $|\sin x|$ in our integrand, let's bound $1/x$. (After all, $\sin x$ should not be ignored as it was in our first approach, since $\sin x$ creates the "hills" in the first place!) On $[n\pi, (n+1)\pi]$, $1/x$ lies in the interval $[1/((n+1)\pi), 1/(n\pi)]$. Thus,

$$\frac{1}{(n+1)\pi} \int_{n\pi}^{(n+1)\pi} |\sin x| \, dx \le I_n \le \frac{1}{n\pi} \int_{n\pi}^{(n+1)\pi} |\sin x| \, dx$$

At this point, we need to know whether $|\sin x|$ is $\sin x$ or $-\sin x$, which leads us to consider cases based on the parity of n. You can check that in either case,

$$\int_{n\pi}^{(n+1)\pi} |\sin x| \, dx = 2$$

Therefore, I_n lies in $[2/((n+1)\pi), 2/(n\pi)]$. Note that $\sum 2/((n+1)\pi)$ and $\sum 2/(n\pi)$ both diverge, since they are constants times the p-series $\sum 1/n$. As a result, since $\sum I_n$ dominates $\sum 2/((n+1)\pi)$, our original improper integral diverges by the Comparison Test. $\square$

9.7 Exercises

In Exercises 1 through 9, an improper integral is presented. Determine if the integral converges, and if so, compute its value.

1. $\displaystyle\int_0^\infty te^{-t}\,dt$
2. $\displaystyle\int_3^\infty \frac{dx}{x^2+9}$
3. $\displaystyle\int_0^\infty \frac{x+2}{\sqrt{x^2+4x+3}}\,dx$
4. $\displaystyle\int_1^2 \frac{\log(t-1)}{t-1}\,dt$
5. $\displaystyle\int_{-1}^1 \frac{x}{\sqrt{x+1}}\,dx$
6. $\displaystyle\int_0^e \log t\,dt$
7. $\displaystyle\int_{-2}^2 \frac{1}{\sqrt{(x-2)(x+2)}}\,dx$
8. $\displaystyle\int_0^2 \frac{x}{|x^2-3x+2|}\,dx$
9. $\displaystyle\int_0^\infty e^{-t} - t^{-1/3}\,dt$

10. For which values of $p \in (0, \infty)$ does the series

$$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$$

converge?

11. (a) Make a function $f : [0, \infty) \to [0, \infty)$ such that f is continuous at all $x \in [0, \infty)$,

$$\int_0^{\infty} f(x)\, dx$$

converges, but

$$\lim_{x \to \infty} f(x)$$

does not exist. Thus, the Divergence Test (Theorem 9.24) does not generalize to improper integrals.

(b) Suppose that $a \in \mathbb{R}$ and $g : [a, \infty) \to \mathbb{R}$ are given, and suppose that g is positive and decreasing. Prove that if

$$\int_a^{\infty} g(x)\, dx$$

converges, then $g(x) \to 0$ as $x \to \infty$.

In Exercises 12 through 24, determine whether the series or improper integral converges or diverges. You may use any technique or test covered in this chapter up to now, but cite which techniques or tests you use and demonstrate why they apply.

12. $\displaystyle\sum_{n=1}^{\infty} \frac{n}{(4n-3)(4n-1)}$

13. $\displaystyle\sum_{n=2}^{\infty} \frac{\sqrt{n}\,|\cos(n\sqrt{n})|}{n^2 - 2}$

14. $\displaystyle\sum_{n=1}^{\infty} \left(1 - \cos\left(\frac{1}{n}\right)\right)$

15. $\displaystyle\sum_{n=1}^{\infty} \left(e^{-1/n} - 1\right)$

16. $\displaystyle\sum_{n=0}^{\infty} \frac{2 + (-1)^n}{2^n}$

17. $\displaystyle\sum_{n=0}^{\infty} (2 + (-1)^n)(1.1)^n$

18. $\displaystyle\sum_{n=1}^{\infty} \frac{\log n}{n\sqrt{n+1}}$

19. $\displaystyle\sum_{n=3}^{\infty} \frac{2^n}{n!}$

20. $\displaystyle\int_1^{\infty} \frac{e^{-\sqrt{x}}}{x}\,dx$

21. $\displaystyle\int_0^1 x\,|\log x|\,dx$

22. $\displaystyle\int_0^{\pi} \frac{\cos t}{\sqrt{t}}\,dt$

    (Note: This integrand changes sign! This means you should split this integral into pieces which are treated separately.)

23. $\displaystyle\int_0^1 \frac{\log x}{1-x}\,dx$

    (Hint: At the $x = 1$ end, consider a Taylor approximation.)

24. $\displaystyle\int_0^1 \frac{\log x}{\sqrt{x}}\,dx$

    (Hint: You may want to use the result of Exercise 7.10.10 to bound $\log x$ when $x \approx 0$.)

25. Prove Theorem 9.33.

26. Prove the Comparison Test, Theorem 9.41.

27. In this exercise, we present two extensions of the Limit Comparison Test (Theorem 9.47). In each of these, $k \in \mathbb{N}$ is given, and $(a_n)_{n=k}^{\infty}$ and $(b_n)_{n=k}^{\infty}$ are two positive sequences.
    (a) Prove that if
    $$\lim_{n\to\infty} \frac{a_n}{b_n} = 0$$
    and $\sum b_n$ converges, then $\sum a_n$ also converges. Also, provide an example to show the converse is not true, i.e. an example where $a_n/b_n \to 0$, $\sum a_n$ converges, but $\sum b_n$ diverges.
    (b) Prove that if
    $$\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$$
    and $\sum b_n$ diverges, then $\sum a_n$ also diverges. Also, provide an example to show the converse is not true, i.e. an example where $a_n/b_n \to \infty$, $\sum a_n$ diverges, but $\sum b_n$ converges.

28. Prove that for all polynomials p and all numbers $b > 1$, the series

$$\sum_{n=1}^{\infty} \frac{|p(n)|}{b^n}$$

converges.

Hint: Consider Example 9.51. It also helps to show that $|p(n)|$ can be dominated by a single power of n.

29. (a) Consider the two expressions

$$\int_{-\infty}^{\infty} \sin x\, dx \qquad \text{and} \qquad \lim_{b\to\infty} \int_{-b}^{b} \sin x\, dx$$

Show that the integral on the left diverges but that the limit on the right exists. This limit is called the doubly-infinite *Cauchy Principal Value*.

(b) Consider the two expressions

$$\int_{-1}^{1} \frac{dx}{x} \qquad \text{and} \qquad \lim_{\epsilon\to 0^+} \left(\int_{-1}^{-\epsilon} \frac{dx}{x} + \int_{\epsilon}^{1} \frac{dx}{x} \right)$$

Show that the integral on the left diverges but that the limit on the right exists. This limit is called the *Cauchy Principal Value* of the integral from $[-1, 1]$ about 0.

30. In Exercise 9.4.19, we saw an example of a two-dimensional shape with infinite perimeter but finite area. In this exercise, we introduce a shape with infinite surface area but finite volume! Let S be the solid of revolution made by taking the area under the graph of $y = \frac{1}{x}$ on $[1, \infty)$ and spinning it about the x-axis. More formally, for any $b > 1$, let S_b be the solid of revolution obtained by spinning the area under $y = 1/x$ for $x \in [1, b]$ about the x-axis, and then

$$S = \bigcup_{b>1} S_b = \{(x, y, z) \in \mathbb{R}^3 \mid \exists b > 1\ (x, y, z) \in S_b\}$$

The volume of S is defined to be the limit of the volume of S_b as $b \to \infty$, and similarly the surface area of S is the limit of the surface area of S_b

as $b \to \infty$. Because this shape looks like an infinitely-stretching horn, it is called *Gabriel's Horn* or *Torricelli's Trumpet*.[14]

(a) Compute the volume of S_b for every $b > 1$, and show that Gabriel's Horn has volume π.

(b) It is known that when the area under the graph of a function $y = f(x)$ is rotated about the x-axis on $[a, b]$, the surface area of the resulting solid is

$$A = 2\pi \int_a^b f(x)\sqrt{1 + (f'(x))^2}\, dx$$

Use this formula to prove that the surface area of S_b tends to ∞ as $b \to \infty$. Thus, Gabriel's Horn has infinite surface area!

Remark. Gabriel's Horn leads to the *painter's paradox*: although it takes an infinite amount of paint to paint the outside of Gabriel's Horn, only finitely much paint needs to be poured to fill the inside of the horn! There are several arguments for settling this paradox, but we will not explore them here.

31. In this exercise, we prove more formally that $\sum |\sin n|/n$ diverges. It follows by CT that $\sum |\sin n|/n^p$ diverges for any $p \le 1$.

First, for any $n \in \mathbb{N}^*$, define

$$s_n = \sum_{i=1}^{n} \frac{|\sin i|}{i}$$

Second, from Exercise 9.4.17, part (b), we know that for any $n \in \mathbb{N}^*$, if $\sin n \ge 1/2$, then at least one of $\sin(n+6)$ or $\sin(n+7)$ is at least $1/2$ as well.

Use these results to prove that for all $n \in \mathbb{N}^*$,

$$s_{7n} \ge \frac{1}{14} \sum_{i=1}^{n} \frac{1}{i}$$

then use this to conclude that $\sum |\sin n|/n$ diverges.

[14] According to http://en.wikipedia.org/wiki/Gabriel's_Horn, the mathematician Evangelista Torricelli discovered this solid, though the name Gabriel's Horn refers to the archangel Gabriel in the Bible.

32. Prove that
$$\int_1^\infty \frac{|\cos x|}{x}\,dx$$
diverges. (Hint: You can use similar tactics to Example 9.56, though there is also another quick approach reusing that example.)

33. The *gamma function* Γ (this is a capital "gamma") is defined by
$$\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}\,dt$$
for all $s > 0$.

   (a) Prove that the improper integral defining $\Gamma(s)$ converges for all $s > 0$. (Hint: Consider Example 9.51.)

   (b) Prove, using integration by parts, that $\Gamma(s+1) = s\Gamma(s)$ for all $s > 0$.

   (c) Use part (b) to prove that $\Gamma(n) = (n-1)!$ for all $n \in \mathbb{N}^*$. Thus, the gamma function is one way to extend the factorial function to a function which is defined on $(0,\infty)$. (In fact, Γ is continuous on $(0,\infty)$.)

34. Suppose that $(a_n)_{n=1}^\infty$ is a DECREASING sequence of positive real numbers. This exercise outlines a proof of *Cauchy's Condensation Test*:
$$\sum_{n=1}^\infty a_n \text{ converges} \qquad \leftrightarrow \qquad \sum_{n=0}^\infty 2^n a_{2^n} \text{ converges}$$
This is called the Condensation Test because it essentially "condenses" the partial sum $a_1 + a_2 + a_3 + \cdots + a_{2^n}$ with 2^n terms down to a partial sum $a_1 + 2a_2 + 4a_4 + \cdots + 2^n a_{2^n}$ with $n+1$ terms.

   (a) First, prove that for all $n \in \mathbb{N}^*$,
$$\sum_{i=1}^{2^{n+1}-1} a_i \le \sum_{i=0}^{n} 2^i a_{2^i}$$
(Hint: Try this with small examples, using the fact that the terms decrease. It might be useful to note that $1+2+\cdots+2^n = 2^{n+1}-1$.)

(b) Next, prove that for all $n \in \mathbb{N}$,

$$\sum_{i=0}^{n} 2^i a_{2^i} \le \sum_{i=1}^{2^n} (2a_i)$$

(c) Use parts (a) and (b) to prove the Condensation Test.

35. This exercise shows some applications of the Condensation Test from the previous exercise.
    (a) Give a proof using the Condensation Test that $\sum n^{-p}$ converges iff $p > 1$.
    (b) Use the Condensation Test to show that $\sum (\log n)^{-p}$ diverges for all $p \in \mathbb{R}$. (Hint: Recall $\log n = (\log_2 n)(\log 2)$.)
    (c) Suppose $a, b, c \in (0, \infty)$ are constants. Use the Condensation Test to discover when

$$\sum_{n=3}^{\infty} \frac{1}{n^a (\log_2 n)^b (\log_2 \log_2 n)^c}$$

converges.

9.8 The Root and Ratio Tests

In this section, we present our final tests that apply to nonnegative series. These tests work particularly well when a series is "almost geometric". Intuitively, when we have a positive series $\sum a_n$, these tests try to find a value r so that a_n behaves a lot like r^n.

As an example, consider the series from Example 9.51:

$$\sum_{n=1}^{\infty} \frac{n}{2^n}$$

Let's say $a_n = n/2^n$ and $b_n = 1/2^n$ for $n \ge 1$. We would like to say that a_n behaves like b_n, because n goes to ∞ much more slowly than 2^n does. However, a_n is not asymptotically equal to b_n; in fact, a_n/b_n goes to ∞! We should ask ourselves: if a_n is not asymptotically equal to b_n, then how do we justify that a_n is "a lot like" b_n?

We want to show that a_n behaves essentially like an exponential function. One key property that an exponential function r^n has (where $r \in \mathbb{R}$ is a constant) is that its n^{th} root is constant: $(r^n)^{1/n}$ is always r. Thus, let's analyze $a_n^{1/n} = n^{1/n}/2$. Since $n^{1/n} \to 1$ as $n \to \infty$, we can say that $a_n^{1/n}$ is very close to $1/2$ when n is large. This gives us one way to justify our claim from the previous paragraph: $n/2^n$ should behave a lot like $1/2^n$ because $(n/2^n)^{1/n}$ is very close to $1/2$.

For another approach, we note another key property of the exponential function r^n: the ratio between consecutive terms, i.e. r^{n+1}/r^n, is always r. This suggests that we analyze the ratio

$$\frac{a_{n+1}}{a_n} = \frac{n+1}{2^{n+1}} \cdot \frac{2^n}{n} = \frac{n+1}{n} \cdot \frac{1}{2}$$

and this ratio is very close to $1/2$ when n is large. This gives another way to justify our claim: $n/2^n$ should behave a lot like $1/2^n$ because the terms are roughly cut in half when n goes up by one.

These two approaches provide ways of relating a positive sequence to a geometric sequence. We will soon show that each of these approaches can be formalized as a series test, and we'll demonstrate examples where these tests are useful.

Root Test

Our first test in this section formalizes the first approach above:

Theorem 9.57 (Root Test). *Let $k \in \mathbb{N}$ be given and let $(a_n)_{n=k}^{\infty}$ be a real-valued sequence with $a_n > 0$ for all $n \geq k$. Suppose that*

$$\lim_{n\to\infty} a_n^{1/n} = r$$

where r is either a nonnegative real number or ∞.

1. *If $r < 1$, then the series $\sum a_n$ converges.*
2. *If $r > 1$, then the series $\sum a_n$ diverges. (This includes the case where $r = \infty$.)*
3. *If $r = 1$, then the series $\sum a_n$ may converge or may diverge. In this case, a different test must be used.*

Strategy. Suppose that $a_n^{1/n} \to r$ as $n \to \infty$. Hence, for large n, $a_n^{1/n} \approx r$. We'd like to say that $a_n \approx r^n$, but our example at the beginning of this section shows that such a leap of logic may be unjustified. (For another example, although most people would agree that 1.99 is close to 2, 1.99^n is not close to 2^n for large values of n. See Exercise 9.9.1.)

Let's be more precise about how close $a_n^{1/n}$ is to r. For any $\epsilon > 0$, when n is large enough, we have

$$\left|a_n^{1/n} - r\right| < \epsilon \qquad \text{i.e.} \qquad r - \epsilon < a_n^{1/n} < r + \epsilon$$

Thus, a_n lies between $(r-\epsilon)^n$ and $(r+\epsilon)^n$. This gives us two comparisons which we can use with the Comparison Test.

When $r < 1$, $r + \epsilon < 1$ when ϵ is small enough, so our series is dominated by $\sum (r+\epsilon)^n$. When $r > 1$, $r - \epsilon > 1$ when ϵ is small enough, so our series dominates $\sum (r-\epsilon)^n$. However, when $r = 1$, neither comparison gives us a conclusion via the Comparison Test, because $r - \epsilon < 1$ and $r + \epsilon > 1$.

Proof. Let $k, r, (a_n)$ be given as described. First, suppose that $r \in [0, \infty) - \{1\}$ (i.e. r is not ∞ or 1). Choose $\epsilon = |r-1|/2$. Because $a_n^{1/n} \to r$ as $n \to \infty$, there is some $N \in \mathbb{N}$ with $N \geq k$ such that for all $n \in \mathbb{N}$ with $n > N$, we have

$$r - \epsilon < a_n^{1/n} < r + \epsilon \qquad \to \qquad (r-\epsilon)^n < a_n < (r+\epsilon)^n$$

If $r < 1$, then $r + \epsilon = r + (1-r)/2 = 1 - (1-r)/2 < 1$, so $\sum a_n$ is dominated by the convergent geometric series $\sum (r+\epsilon)^n$. By CT, $\sum a_n$ converges in this case. If $r > 1$, then $r - \epsilon = r - (r-1)/2 = 1 + (r-1)/2 > 1$, so $\sum a_n$ dominates the divergent geometric series $\sum (r-\epsilon)^n$. Thus, by CT, $\sum a_n$ diverges in this case.

Lastly, we consider the cases $r = \infty$ and $r = 1$. When $r = \infty$, a similar argument to above shows that $a_n^{1/n} > 2$ for all n large enough, so $\sum a_n$ dominates $\sum 2^n$ and hence diverges. To see that $r = 1$ provides no useful information, consider the p-series $\sum n^{-p}$ where p is a constant. We already know that $\sum n^{-p}$ converges iff $p > 1$. However, we note that for both $p = 1$ and $p = 2$, $(n^{-p})^{1/n} \to 1$ as $n \to \infty$. ■

Remark. When $r > 1$ in the proof above, there is another way to show that $\sum a_n$ diverges. Because $a_n > (r-\epsilon)^n$ for large enough values of n, and $r - \epsilon > 1$ with our choice of ϵ, a_n cannot approach 0. Thus, by the Divergence Test, $\sum a_n$ diverges.

Example 9.58:

The Root Test makes it easy to see whether

$$\sum_{n=2}^{\infty} \left(\frac{1}{\log n}\right)^n$$

converges. We set $a_n = (1/\log n)^n$, so that $a_n^{1/n} = 1/\log n \to 0$ as $n \to \infty$. Because $0 < 1$, the Root Test says that $\sum a_n$ converges.

(For another way to approach this series, note that $\log n > 2$ whenever $n > e^2$. Thus, $\sum a_n$ is dominated by $\sum 1/2^n$. In fact, generalizing this argument, for any $M > 1$, $\log n > M$ when $n > e^M$, so $\sum a_n$ is dominated by $\sum 1/M^n$.) $\square$

Example 9.59:

Consider

$$\sum_{n=1}^{\infty} a_n \qquad \text{where} \qquad a_n = \left(\frac{n}{n+1}\right)^{(n^2)}$$

Note that for this example, it's not easy to tell whether $a_n \to 0$, because as $n \to \infty$, a_n has the indeterminate form 1^∞. However, the Root Test works well here. We compute

$$\begin{aligned} a_n^{1/n} &= \left(\frac{n}{n+1}\right)^n = \frac{1}{((n+1)/n)^n} \\ &= \frac{1}{\left(1+\frac{1}{n}\right)^n} \to \frac{1}{e} \end{aligned}$$

as $n \to \infty$. Because $1/e < 1$, the Root Test tells us that $\sum a_n$ converges.

Compare this example with the series

$$\sum_{n=1}^{\infty} b_n \qquad \text{where} \qquad b_n = \left(\frac{n}{n+1}\right)^n$$

The Root Test is ineffective here, because $b_n^{1/n} \to 1$. However, since $b_n = a_n^{1/n}$, the calculations above show that $b_n \to 1/e$, so b_n does not approach 0. Hence, the Divergence Test tells us that $\sum b_n$ diverges. $\square$

It is also possible to use the Root Test on series whose terms do not look like n^{th} powers, though it can be difficult to analyze the n^{th} root of such terms. Here is one example which illustrates this:

Example 9.60:

Consider the series

$$\sum_{n=1}^{\infty} a_n \qquad \text{where} \qquad a_n = \frac{n!}{n^n}$$

We have studied a_n previously in Examples 9.8 and 9.43. There, we saw that for any fixed $p > 1$, $a_n < p!/n^p$ for large enough n. Thus, our series converges, since for any $p > 1$, $\sum a_n$ is dominated by a constant times $\sum n^{-p}$. Now, let's use the Root Test to see whether our series is dominated by a geometric series. (This is useful information because a geometric series converges more quickly than a p-series, since exponentials dominate polynomials.)

To use the Root Test, we need to know something about the behavior of $(n!)^{1/n}$. There is a famous result called *Stirling's Formula* which says that[15]

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

as $n \to \infty$. From this, we can deduce (see Exercise 9.9.1) that $(n!)^{1/n} \sim n/e$. There is also a simple argument, outlined in Exercise 9.9.17, which proves $(n!)^{1/n} \sim n/e$ without using Stirling's Formula.

Therefore, we know that $(n!)^{1/n}/(n/e) \to 1$ as $n \to \infty$. Using this, we can compute the limit of $a_n^{1/n}$ by multiplying and dividing by n/e:

$$\begin{aligned} a_n^{1/n} &= \frac{(n!)^{1/n}}{n} = \frac{(n!)^{1/n}}{n} \cdot \frac{n/e}{n/e} \\ &= \frac{(n!)^{1/n}}{n/e} \cdot \frac{n/e}{n} \to 1 \cdot \frac{1}{e} \end{aligned}$$

This shows that our series $\sum a_n$ behaves a lot like $\sum 1/e^n$. More precisely, the proof of the Root Test shows that $\sum a_n$ is dominated by any sum of the form $\sum r^n$ where $1/e < r < 1$. $\square$

Ratio Test

[15] There are other variants of Stirling's Formula; the variant we present here is accurate enough for our purposes.

There is another convergence test, called the Ratio Test, which analyzes the convergence of positive series by trying to relate them to geometric series. This test uses the second approach outlined at the beginning of the section, which says that a_n behaves like r^n if a_{n+1}/a_n is close to r for large values of n. More formally, we have

Theorem 9.61 (Ratio Test). *Let $k \in \mathbb{N}$ be given and let $(a_n)_{n=k}^{\infty}$ be a real-valued sequence with $a_n > 0$ for all $n \geq k$. Suppose that*

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

where r is either a nonnegative real number or ∞.

1. *If $r < 1$, then the series $\sum a_n$ converges.*
2. *If $r > 1$, then the series $\sum a_n$ diverges. (This includes the case where $r = \infty$.)*
3. *If $r = 1$, then the series $\sum a_n$ may converge or may diverge. In this case, a different test must be used.*

Strategy. The basic idea of this test is that if $a_{n+1}/a_n \to r$, then each term is approximately r times the previous term. Thus, $a_{k+1} \approx ra_k$, $a_{k+2} \approx r^2 a_k$, $a_{k+3} \approx r^3 a_k$, and so on. In fact, for any $m \in \mathbb{N}^*$, we have

$$\frac{a_{k+m}}{a_k} = \frac{a_{k+1}}{a_k} \cdot \frac{a_{k+2}}{a_{k+1}} \cdot \cdots \cdot \frac{a_{k+m}}{a_{k+m-1}} \approx r \cdot r \cdot \cdots \cdot r = r^m$$

so that $a_{k+m} \approx r^m a_k$. This suggests that we can compare $\sum a_n$ to $a_k \sum r^n$ (since a_k is a constant with respect to n).

As with the Root Test, we can be more precise by establishing inequalities showing how close a_{n+1}/a_n is to r. For any $\epsilon > 0$, we have

$$\left|\frac{a_{n+1}}{a_n} - r\right| < \epsilon \qquad \text{i.e.} \qquad r - \epsilon < \frac{a_{n+1}}{a_n} < r + \epsilon$$

when n is large enough, say when $n \geq N$. Then the same derivation above shows that

$$a_N(r-\epsilon)^m < a_{N+m} < a_N(r+\epsilon)^m$$

for any $m \in \mathbb{N}^*$. When $r \neq 1$, then we can choose ϵ small enough so that $(r - \epsilon, r + \epsilon)$ does not contain 1, and the proof proceeds much like the proof of the Root Test.

Proof. Let $k, r, (a_n)$ be given as described. First, suppose that $r \in [0, \infty) - \{1\}$, so r is not ∞ or 1. Choose $\epsilon = |r-1|/2$. Because $a_{n+1}/a_n \to r$, there is some $N \in \mathbb{N}$ with $N > k$ such that for all $n \in \mathbb{N}$ with $n \geq N$, we have

$$r - \epsilon < \frac{a_{n+1}}{a_n} < r + \epsilon$$

By the choice of ϵ, all of these numbers are positive. Hence, we can deduce that for all $m \in \mathbb{N}^*$,

$$(r-\epsilon)^m < \frac{a_{N+m}}{a_N} < (r+\epsilon)^m$$

(the proof by induction on m is left to the reader). Therefore,

$$a_N(r-\epsilon)^{n-N-1} < a_n < a_N(r+\epsilon)^{n-N-1}$$

for all $n > N+1$ (by choosing $m = n - N - 1$).

If $r < 1$, then $r + \epsilon < 1$ as well, so $\sum a_n$ is dominated by a positive constant times $\sum (r+\epsilon)^n$ (recall that N does not depend on n). Thus, by CT, $\sum a_n$ converges in this case. If $r > 1$, then $r - \epsilon > 1$ as well, so $\sum a_n$ dominates a positive constant times $\sum (r-\epsilon)^n$. As a result, CT tells us that $\sum a_n$ diverges in this case.

Lastly, when $r = \infty$, the same proof style as above shows that $a_{n+1}/a_n > 2$ for all n large enough, so $\sum a_n$ dominates a constant times $\sum 2^n$ and hence diverges. For $r = 1$, consider the p-series $\sum n^{-p}$. For $p = 1$, we have

$$\lim_{n\to\infty} \frac{1/(n+1)}{1/n} = \lim_{n\to\infty} \frac{n}{n+1} = 1$$

and for $p = 2$, we have

$$\lim_{n\to\infty} \frac{1/(n+1)^2}{1/n^2} = \lim_{n\to\infty} \frac{n^2}{n^2\left(1+\frac{1}{n}\right)^2} = 1$$

Both of these values of p produce $r = 1$, although $\sum n^{-1}$ diverges and $\sum n^{-2}$ converges. ■

The Ratio Test is frequently more practical to apply than the Root Test. It generally works well with large products, such as exponentials or factorials.

For instance, if we use the Ratio Test with the example at the beginning of this section, where $a_n = n/2^n$, we get

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{n+1}{2^{n+1}} \cdot \frac{2^n}{n} = \lim_{n\to\infty} \frac{n+1}{2n} = \frac{1}{2}$$

which is less than 1. Here are some other examples.

Example 9.62:

Consider the series

$$\sum_{n=1}^{\infty} \frac{n!}{n^n}$$

from Example 9.60. If we set $a_n = n!/n^n$ and try the Ratio Test, we find

$$\begin{aligned} \frac{a_{n+1}}{a_n} &= \frac{(n+1)!}{(n+1)^{n+1}} \cdot \frac{n^n}{n!} \\ &= \frac{n!}{(n+1)^n} \cdot \frac{n^n}{n!} = \left(\frac{n}{n+1}\right)^n \\ &= \frac{1}{\left(1+\frac{1}{n}\right)^n} \to \frac{1}{e} \end{aligned}$$

Since $1/e < 1$, the Ratio Test tells us that $\sum a_n$ converges.

In fact, note that the Ratio Test produced the same limit that the Root Test did in Example 9.60. This is not a coincidence: see Exercises 9.9.22 and 9.9.23. $\square$

Example 9.63:

Consider the series

$$\sum_{n=1}^{\infty} \frac{n^2}{n^4+1}$$

If we try the Ratio Test with $a_n = n^2/(n^4+1)$, then

$$\begin{aligned} \frac{a_{n+1}}{a_n} &= \frac{(n+1)^2}{(n+1)^4+1} \cdot \frac{n^4+1}{n^2} \\ &= \frac{(n+1)^2}{n^2} \cdot \frac{n^4+1}{(n+1)^4+1} \\ &= \frac{n^2\left(1+\frac{1}{n}\right)^2}{n^2} \cdot \frac{n^4\left(1+\frac{1}{n^4}\right)}{n^4\left(\left(1+\frac{1}{n}\right)^4+\frac{1}{n^4}\right)} \\ &\to 1 \end{aligned}$$

as $n \to \infty$ (note that the second-to-last step involved factoring out the dominant powers of n from the top and bottom of each fraction). Because this limit is 1, the Ratio Test does not help. However, this series can quickly be shown to converge via CT or LCT.

In general, the Ratio Test or the Root Test will always produce a limit of 1 with any rational function, as you can prove in Exercise 9.9.18. This seems quite plausible, because the Ratio and Root Tests look for geometric behavior, but polynomials grow more slowly than exponentials. The Ratio and Root Tests are primarily useful with products for this reason. □

Example 9.64:

Consider the series

$$\sum_{n=1}^{\infty} a_n \qquad \text{with} \qquad a_n = \frac{(n!)^2}{(2n)!}$$

Because the terms involve factorials, which are large products, the Ratio Test is a good test to try. We compute

$$\begin{aligned} \frac{a_{n+1}}{a_n} &= \frac{((n+1)!)^2}{(2n+2)!} \cdot \frac{(2n)!}{(n!)^2} \\ &= \frac{((n+1)n!)^2}{(n!)^2} \cdot \frac{(2n)!}{(2n+2)(2n+1)(2n)!} \\ &= \frac{(n+1)^2}{(2n+2)(2n+1)} \to \frac{1}{4} \end{aligned}$$

Since $1/4 < 1$, our series converges.

Another option is to apply the Root Test. In Example 9.60, we mentioned that $(n!)^{1/n} \sim n/e$. It follows that $((2n)!)^{1/(2n)} \sim (2n)/e$, which leads to $((2n)!)^{1/n} \sim 4n^2/e^2$, as you can check. Putting these results together, you can readily show that $a_n^{1/n} \to 1/4$, as with the Ratio Test.

It turns out that we can also use Stirling's Formula to get a good approximation for a_n: see Exercise 9.9.19. □

Example 9.65:

We can also use the Ratio Test to analyze some series where we do not have a simple closed form for the terms. For instance, let's consider $\sum a_n$ where

$$a_n = \frac{1 \cdot 4 \cdot 7 \cdot \cdots \cdot (3n-2)}{1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1)} = \prod_{i=1}^{n} \frac{3i-2}{2i-1}$$

for all $n \in \mathbb{N}^*$. Because our terms are products, we try the Ratio Test. We have

$$\frac{a_{n+1}}{a_n} = \frac{1 \cdot 4 \cdot \dots \cdot (3n-2)(3(n+1)-2)}{1 \cdot 3 \cdot \dots \cdot (2n-1)(2(n+1)-1)} \cdot \frac{1 \cdot 3 \cdot \dots \cdot (2n-1)}{1 \cdot 4 \cdot \dots \cdot (3n-2)} = \frac{3n+2}{2n+1}$$

which approaches $3/2$ as $n \to \infty$. Since $3/2 > 1$, the Ratio Test tells us our series diverges.

In contrast, consider when

$$a_n = \frac{1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)}{2 \cdot 4 \cdot 6 \cdot \dots \cdot (2n)}$$

A little work shows that $a_{n+1}/a_n = (2n+1)/(2n+2) \to 1$ as $n \to \infty$. Thus, the Ratio Test cannot be used to determine whether $\sum a_n$ converges. This series requires some more advanced techniques: see Exercise 9.9.20. □

Suggestions For Picking A Test

In the last few sections, we have presented a number of different tests that we can use to check whether a series with positive terms converges or diverges. With so many tests available, it helps to have guidelines that describe how to pick a test. Certainly, practice makes perfect: the more examples you see and try, the more intuition you will obtain for analyzing series. However, we also offer the following rules of thumb for determining whether a positive series $\sum a_n$ converges:

1. First, it's worth checking whether $a_n \to 0$ if this check can be done quickly. If $a_n \not\to 0$, then the Divergence Test says $\sum a_n$ diverges.

   Examples: $\sum e^{1/n}$, $\sum (2 \cdot 4 \cdot \dots \cdot (2n))/(1 \cdot 3 \cdot \dots \cdot (2n-1))$ (In these series, the terms are all greater than 1!)

2. If a_n consists of products, such as factorials or exponentials, then the Ratio Test is often helpful. The Root Test is also worth trying if it looks like it produces simpler computations than the Ratio Test does.

   Examples: $\sum 2^n/n!$, $\sum \exp(-n^2)$

3. When dealing with polynomial expressions, or with functions that can be approximated well by a Taylor polynomial (for instance, $\sin(1/n)$ is approximated well by Taylor polynomials for sin centered at 0), the Limit Comparison Test is frequently handy.

   Examples: $\sum (n+2)/(n^2+4)$, $\sum \arctan(1/n)$

4. When there are obvious inequalities which simplify a problem (such as $|\cos n| \leq 1$), the Comparison Test is appropriate. Sometimes, when applying an inequality, the inequality goes in the wrong direction for CT to yield a conclusion. In those cases, you can try to introduce an extra constant factor (for instance, $1/(n+1)$ is not bigger than $1/n$, but it is bigger than $1/(2n)$ for n large), or you can try the LCT.

   Examples: $\sum |\sin n| / n^2$, $\sum (2 + (-1)^n)/3^n$

5. The Integral Test is often used as a last resort, because it requires both an integral computation and also showing that a function is decreasing. However, in certain situations, particularly when a substitution is obvious, the Integral Test can be promising.

   Examples: $\sum (\log n)/n$, $\sum ne^{-n}$ (Both of these series can also be handled by CT, similarly to Example 9.51.)

6. Lastly, sometimes multiple tests can be put together, each handling a specific step.

   Example: One way of analyzing the series $\sum (2 + (-1)^n \sin n)/(n^2 - 1)$, where $n > 1$, is to first note that
   $$\frac{2 + (-1)^n \sin n}{n^2 - 1} \leq \frac{2+1}{n^2-1} = \frac{3}{n^2-1}$$
   and then we note that $3/(n^2 - 1) \sim 3/n^2$. Thus, the final answer uses both LCT and CT.

It is also worth noting that this chapter is not the final word on series tests for positive series. There are other series tests that have been developed for handling cases where our tests have difficulty (especially for cases where the Ratio or Root Tests produce a limit of 1). However, these tests are generally difficult to motivate, difficult to apply, and difficult to prove.[16]

In contrast, we have restricted our attention to tests that have simple motivations, can be usually applied with little effort, and that work well with our basic examples of p-series and geometric series. All of these tests come originally from one central idea: to show a positive series converges, we only need to show its partial sums are bounded above. This central idea does not work with series where the terms change sign, so we will make some remarks about those series in the last section of this chapter.

[16] The interested reader may want to search in the literature for *Raabe's Test* and *Gauss's Test*.

9.9 Exercises

1. Suppose that $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are both sequences of positive real numbers.
   (a) Prove that if $a_n \sim b_n$ as $n \to \infty$, then $a_n^{1/n} \sim b_n^{1/n}$.
   (b) Show that the converse of part (a) is not true: namely, produce an example of (a_n) and (b_n) such that $a_n^{1/n} \sim b_n^{1/n}$ but $a_n \not\sim b_n$. If you try to imitate the proof of part (a) to prove the converse, then where does the proof fail?
   (c) Use part (a) to prove that if p is any polynomial with $p(n) > 0$ for all $n \in \mathbb{N}^*$, then $p(n)^{1/n} \to 1$. (Hint: Consider the leading term.)

For Exercises 2 through 15, determine whether the series converges or diverges. You may use any technique or test covered in this chapter up to now, but cite which techniques or tests you use and demonstrate why they apply.

2. $\displaystyle\sum_{n=1}^{\infty} \frac{e^{2n}}{(2e)^n}$

3. $\displaystyle\sum_{n=1}^{\infty} \frac{n^{2n}}{(2n)^n}$

4. $\displaystyle\sum_{n=1}^{\infty} \frac{(n!)^2}{2^{(n^2)}}$

5. $\displaystyle\sum_{n=3}^{\infty} \frac{(n-2)3^n}{(n-1)\left(4-\frac{1}{n}\right)^n}$

6. $\displaystyle\sum_{n=1}^{\infty} \frac{10^n}{(n+1)4^{2n+1}}$

7. $\displaystyle\sum_{n=1}^{\infty} \frac{n!\,(2n)!}{(3n)!}$

8. $\displaystyle\sum_{n=1}^{\infty} \frac{(n+1)!\,(n-1)!}{(2n)!}$

9. $\displaystyle\sum_{n=1}^{\infty} \frac{1 \cdot 3 \cdot \cdots \cdot (2n-1)}{(2n-1)!}$

10. $\displaystyle\sum_{n=1}^{\infty} \frac{2 \cdot 4 \cdot \cdots \cdot (2n)}{n!}$

11. $\displaystyle\sum_{n=1}^{\infty} \frac{2^n n!}{5 \cdot 8 \cdot \cdots \cdot (3n+2)}$

12. $\displaystyle\sum_{n=2}^{\infty} \frac{1}{(\log n)^{1/n}}$

13. $\displaystyle\sum_{n=2}^{\infty} (n^{1/n}-1)^n$

14. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{(\arctan n)^n}$

15. $\displaystyle\sum_{n=1}^{\infty} \frac{n^{(n+(1/n))}}{\left(n+\frac{1}{n}\right)^n}$

(Hint: Pull $n^{1/n}$ out to the side, since it converges to 1. One option is to rewrite everything else in terms of exponentials and logarithms, and then use a Taylor approximation of $\log(1+x)$ when x is near 0.)

16. For which values of $k \in \mathbb{N}^*$ does the series

$$\sum_{n=1}^{\infty} \frac{(n!)^2}{(kn)!}$$

converge?

17. This exercise outlines a proof of $(n!)^{1/n} \sim n/e$ as $n \to \infty$ which does not use Stirling's Formula. The main idea is to find inequalities for $\log(n!) = \log 1 + \log 2 + \cdots + \log n$ by using integrals.

(a) By making appropriate step functions below and above the graph of $f(x) = \log x$, show that for any $n \in \mathbb{N}^*$,

$$\sum_{i=1}^{n-1} \log i \le \int_1^n \log x \, dx \le \sum_{i=2}^{n} \log i$$

(b) Use part (a) to show that

$$\log((n-1)!) \le n \log n - n + 1 \le \log(n!)$$

and consequently

$$\frac{n!}{n} \le \frac{e \cdot n^n}{e^n} \le n!$$

(c) Use the inequalities from part (b) to deduce

$$e\left(\frac{n}{e}\right)^n \le n! \le ne\left(\frac{n}{e}\right)^n$$

From this, and the Squeeze Theorem, prove that $(n!)^{1/n} \sim n/e$.

18. Prove that if p is any polynomial with $p(n) > 0$ for all $n \in \mathbb{N}^*$, then $p(n+1)/p(n) \to 1$ as $n \to \infty$. Thus, the Ratio Test should not be used on series whose terms are rational functions (i.e. quotients of polynomials).

19. Use Stirling's Formula
$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$
to prove that
$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \sim \frac{2^{2n}}{\sqrt{\pi n}}$$
as $n \to \infty$. In particular, this implies that the n^{th} root of $(2n)!/(n!)^2$ approaches 4, so this exercise proves a stronger result than Example 9.64. (Also, see the next exercise.)

20. This exercise outlines two proofs that when $s > 0$ is a constant,
$$\sum_{n=1}^{\infty} a_n^s \qquad \text{where} \qquad a_n = \frac{1 \cdot 3 \cdot \cdots \cdot (2n-1)}{2 \cdot 4 \cdot \cdots \cdot (2n)}$$
converges iff $s > 2$. The first proof, in parts (a) and (b), shows that a_n is asymptotically equal to a constant times $n^{-1/2}$. This proof also uses Stirling's Formula. The second proof, in parts (c) and (d), shows that a_n^2 is bounded by C/n and D/n for some constants $C, D > 0$. This proof does not use Stirling's Formula.

    (a) For any $n \in \mathbb{N}^*$, show that
$$a_n = \frac{(2n)!}{2^{2n}(n!)^2}$$
    (Hint: Since the denominator of a_n is the product of the first n odd numbers, multiply and divide by the product of the first n even numbers.)

    (b) Use part (a) and the result of Exercise 9.9.19 to prove that $\sum a_n^s$ converges iff $s > 2$. This finishes the first proof.

    (c) For the second proof, for any $n \in \mathbb{N}^*$, let's write a_n^2 as
$$\frac{1 \cdot 1 \cdot 3 \cdot 3 \cdot \cdots \cdot (2n-1) \cdot (2n-1)}{2 \cdot 2 \cdot 4 \cdot 4 \cdot \cdots \cdot (2n) \cdot (2n)}$$
    Recall that for any $x > 1$, $x^2 > (x-1)(x+1)$. Use this to prove that
$$\frac{1}{4n} \le a_n^2 \le \frac{1}{2n}$$

(Hint: For one of the inequalities, pull out factors of the form x^2 from the numerator. For the other, pull out factors of the form x^2 from the denominator.)

(d) Use the result of part (c) to prove that $\sum a_n^s$ converges iff $s > 2$.

21. The following amazing formula for $1/\pi$ is due to Ramanujan:

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{n=0}^{\infty} \frac{(4n)!\ (1103 + 26390n)}{(n!)^4 396^{4n}}$$

Verify that the series on the right-hand side converges.

22. In this exercise, we outline a proof of the following result:

Theorem 9.66. *Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$ be an infinite sequence of positive real numbers. Suppose that*

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

where r is either a nonnegative real number or ∞. Then $a_n^{1/n} \to r$ as $n \to \infty$.

For simplicity, we will prove this result when $r \in (0, \infty)$; the cases $r = 0$ and $r = \infty$ can be proven similarly. Thus, suppose that $a_{n+1}/a_n \to r$.

(a) The proof of the Ratio Test shows that for any $\epsilon \in (0, r)$, for some $N \in \mathbb{N}^*$ with $N \geq k$, we have

$$a_N(r - \epsilon)^m < a_{N+m} < a_N(r + \epsilon)^m$$

for all $m \in \mathbb{N}^*$. Use this to find $C, D > 0$ such that for all $n \in \mathbb{N}^*$ with $n > N$, we have

$$C(r - \epsilon)^n < a_n < D(r + \epsilon)^n$$

(C and D may depend on ϵ and N, but not on n or m.)

(b) Use part (a) to show that for all $\delta \in (0, 1)$, there exists some $M \in \mathbb{N}^*$ such that $M \geq k$ and for all $n \in \mathbb{N}^*$ with $n > M$, we have

$$(1 - \delta)^2 r < a_n^{1/n} < (1 + \delta)^2 r$$

(Hint: Choose $\epsilon = \delta r$, and use the fact that $C^{1/n} \to 1$ as $n \to \infty$ to show that $C^{1/n} > 1 - \delta$ for large values of n. Do something similar with $D^{1/n}$. This will tell you how to find M.)

(c) Use part (b) to prove that $a_n^{1/n} \to r$.

23. In this exercise, we show that the converse of Theorem 9.66 is false, i.e. $a_n^{1/n} \to r$ does not necessarily imply $a_{n+1}/a_n \to r$. (Thus, the Root Test is a stronger test than the Ratio Test.) Consider $(a_n)_{n=1}^{\infty}$ defined by

$$a_{2n} = (0.4)^n(1.6)^n \qquad a_{2n-1} = (0.4)^{n-1}(1.6)^n$$

for all $n \in \mathbb{N}^*$. Prove that $a_n^{1/n} \to 0.8$ as $n \to \infty$ (and thus $\sum a_n$ converges by the Root Test), but a_{n+1}/a_n has no limit as $n \to \infty$.

9.10 Series With Mixed Terms

Most of the series tests that were developed earlier in this chapter only apply to series with positive terms. (The Divergence Test applies to any series, however.) These tests were developed around the idea of finding upper bounds on the partial sums of a positive series, because the partial sums are monotone increasing. By factoring out -1 from a series, we can also use these tests to handle series whose terms are all negative. In fact, since finitely many terms never affect convergence of a series, our tests can be applied to any series whose terms change sign only finitely many times.

However, when presented with a series with infinitely many positive terms and infinitely many negative terms, we cannot establish monotonicity of partial sums, so many of our earlier tests are not applicable. Because of this, these kinds of series (which we will call *series with mixed terms*) are more difficult to study. For simplicity, we will mainly focus on one special type of mixed series: *alternating series*, where the terms alternate in sign. We will also introduce the notion of *absolute convergence*, which lets us relate a mixed series $\sum a_n$ to the positive series of absolute values $\sum |a_n|$. Lastly, we will explore a curious phenomenon concerning rearrangements of series that can arise with mixed series.

Alternating Series

An alternating series has the form $\sum a_n$ where the terms of (a_n) alternate in sign. Thus, we either have $a_n = (-1)^n|a_n|$ for all n or $a_n = (-1)^{n-1}|a_n|$ for all n, depending on the sign of the first term. For instance, consider the following series, which is called the *alternating harmonic series*:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

For each $n \in \mathbb{N}^*$, let s_n be the n^{th} partial sum of the series.

Because $((-1)^{n-1}/n)_{n=1}^{\infty}$ is alternating, the partial sums alternately rise and fall. Intuitively, we can imagine this by thinking of the partial sums placed on a number line. Before adding any terms, we start at 0. Next, we add 1 to get s_1, moving 1 unit to the right. Next, we add $-1/2$ to get s_2, moving $1/2$ a unit to the left. Continuing in this manner, our partial sums continue jumping (or zig-zagging) to the left and to the right. A rough picture illustrating is presented in Figure 9.9.

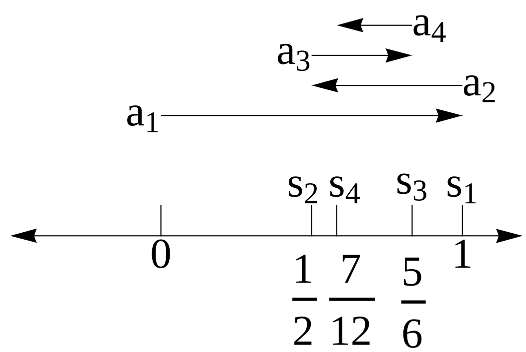

Figure 9.9: The first few partial sums of the alternating harmonic series

Note that since $(1/n)$ is decreasing, our partial sums jump less with each step. This has some important consequences. For instance, $s_3 < s_1$, because we move $1/2$ of a unit to the left from s_1 to s_2 but only $1/3$ of a unit to the right from s_2 to s_3. Similar logic shows that $s_5 < s_3$, $s_7 < s_5$, and so forth. We can also show $s_4 > s_2$, $s_6 > s_4$, and so forth. Also, note that all the even-indexed partial sums are smaller than all the odd-indexed partial sums.

Putting these remarks together, we find that if our series converges, then its sum should lie in $[s_2, s_1]$, and in $[s_4, s_3]$, and in $[s_6, s_5]$, etc. In general, the closed interval $[s_{2n}, s_{2n-1}]$ has length $|s_{2n} - s_{2n-1}| = 1/(2n)$, which goes to 0 as $n \to \infty$. This intuitively means that these intervals $[s_{2n}, s_{2n-1}]$ "zoom in" on one point, which should be the sum of our series.

When we take the reasoning above and generalize it to an alternating series $\sum(-1)^{n-1}|a_n|$, we obtain the following result due to Leibniz:

Theorem 9.67 (Alternating Series Test). *Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$ be a given real-valued sequence. Suppose that (a_n) is alternating. If $(|a_n|)$ is decreasing and $a_n \to 0$ as $n \to \infty$, then $\sum a_n$ converges.*

Strategy. WLOG, we may suppose $k = 1$ and $a_n = (-1)^{n-1}|a_n|$, i.e. $a_1 \geq 0$. Let's say $(s_n)_{n=1}^{\infty}$ is the sequence of partial sums. As suggested by the argument earlier, we can show that the subsequence (s_{2n}) is increasing and the subsequence (s_{2n-1}) is decreasing. Furthermore, these subsequences are bounded, as for all $m, n \in \mathbb{N}^*$, we can show $s_{2m} \leq s_{2n-1}$.

As a result, both of these subsequences converge by the BMCT. Let's say $s_{2n} \to L_-$ and $s_{2n-1} \to L_+$ as $n \to \infty$. (The $-$ and $+$ are used because we expect the s_{2n} values to approach the sum from the left side, and we expect the s_{2n-1} values to approach the sum from the right side.) Because $a_n \to 0$, we can show that $L_- = L_+$. Thus, since the even-indexed and the odd-indexed partial sums both approach the same limit, the sequence of partial sums approaches that limit as well by Exercise 9.2.21.

Proof. Let k and (a_n) be given as described, and assume that (a_n) is alternating, $(|a_n|)$ decreases, and $a_n \to 0$ as $n \to \infty$. WLOG, we may also suppose $k = 1$ and $a_1 \geq 0$, so that $a_n = (-1)^{n-1}|a_n|$ for each $n \in \mathbb{N}^*$. (Thus, odd-indexed terms are nonnegative, and even-indexed terms are nonpositive.) Define $(s_n)_{n=1}^{\infty}$ to be the sequence of partial sums of $\sum a_n$.

First, for any $n \in \mathbb{N}^*$,

$$s_{2n+2} = s_{2n} + a_{2n+1} + a_{2n+2} = s_{2n} + (|a_{2n+1}| - |a_{2n+2}|) \geq s_{2n}$$

and

$$s_{2n+1} = s_{2n-1} + a_{2n} + a_{2n+1} = s_{2n-1} + (-|a_{2n}| + |a_{2n+1}|) \leq s_{2n+1}$$

because $(|a_n|)$ is decreasing. Thus, $(s_{2n})_{n=1}^{\infty}$ is increasing and $(s_{2n-1})_{n=1}^{\infty}$ is decreasing. We claim that from this, it follows that both of these sequences are bounded. For any $n \in \mathbb{N}^*$, since $s_{2n} = s_{2n-1} + a_{2n} \leq s_{2n-1}$, we have

$$s_2 \leq s_{2n} \leq s_{2n-1} \leq s_1$$

This shows that both sequences are bounded below by s_2 and above by s_1, proving the claim. (In fact, the same proof shows that for any $m \in \mathbb{N}^*$, $s_{2m} \leq s_{2n} \leq s_{2n-1} \leq s_{2m-1}$ for any $n \in \mathbb{N}^*$ with $n \geq m$.)

By the BMCT, there exist $L_-, L_+ \in \mathbb{R}$ such that $s_{2n} \to L_-$ and $s_{2n-1} \to L^+$ as $n \to \infty$. In fact, the proof of the BMCT shows that

$$L_- = \sup\{s_{2n} \mid n \in \mathbb{N}\} \qquad L_+ = \inf\{s_{2n-1} \mid n \in \mathbb{N}\}$$

Now, $s_{2n} = s_{2n-1} + a_{2n}$. By taking the limit as $n \to \infty$ of this equality, we obtain $L_- = L_+ + 0$, so $L_- = L_+$. For simplicity, let's denote this common value of L_- and L_+ by L.

Lastly, note that we have proven that (s_{2n}) and (s_{2n-1}) both approach the same limit L. By Exercise 9.2.21, $s_n \to L$ as $n \to \infty$ as well. Thus, our alternating series converges to L. ■

The Alternating Series Test, sometimes abbreviated as AST, is a handy tool for approaching many alternating series. For instance, it immediately implies that the alternating harmonic series converges. Let's say the sum is L, and the partial sums are $(s_n)_{n=1}^{\infty}$ as before. We know that $s_n \to L$, but how quickly does s_n converge to L?

To answer this question, we use more information from the proof of the AST. For any $n \in \mathbb{N}^*$, we can show $s_{2n} \le L \le s_{2n-1}$. Using $n = 2$, we have $7/12 \le L \le 5/6$, so if we guess L is the average of these two bounds, then we find $L \approx 17/48$, and our error is at most $1/2(5/6 - 7/12) = 1/2(1/4) = 1/8$. Using $n = 3$, you can compute $37/60 \le L \le 47/60$, leading us to the estimate $L \approx 7/10$ with error at most $1/12$.

This kind of reasoning leads to the following corollary on error estimates, which you can prove in Exercise 9.12.5:

Corollary 9.68. *Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$ be a real-valued sequence. Suppose that (a_n) satisfies the hypotheses of the AST. Let*

$$L = \sum_{n=k}^{\infty} a_n$$

For any $n \ge k$, if s_n is the n^{th} partial sum of $\sum a_n$, then

$$|L - s_n| \le |a_{n+1}|$$

In other words, the error in approximating by s_n is bounded by the first term of the series that s_n "leaves out". As a result, the estimate $L \approx (s_n + s_{n+1})/2$, which approximates L by the average of s_n and s_{n+1}, has error magnitude at most $|a_{n+1}|/2$.

Remark. It turns out that with the alternating harmonic series, for any $n \in \mathbb{N}^*$, s_n is $T_n \log(1+x)$ evaluated at 1. Recall from Exercise 8.5.20 that for any $x \in [0,1]$,

$$|E_n \log(1+x)| \le \frac{x^{n+1}}{n+1} \le \frac{1}{n+1}$$

Applying this for $x = 1$ and letting n approach ∞, this shows that $L = \log 2$. In fact, the error bound from Taylor estimates agrees with the error bound from Corollary 9.68. (This might not happen in general, though, with approximations whose Taylor polynomials yield alternating series.)

There are other ways to obtain the sum of the alternating harmonic series, as well as other similar series, by a more in-depth study of the harmonic series. See Exercises 9.12.22 and 9.12.24.

Example 9.69:

As another example of the use of the Alternating Series Test, consider

$$\sum_{n=2}^{\infty} a_n \qquad \text{where} \qquad a_n = \frac{(-1)^n(n^2+1)}{n^3-1}$$

This is certainly an alternating series, and it's straightforward to show that $|a_n| \to 0$ as $n \to \infty$. (Note that $|a_n| \to 0$ iff $a_n \to 0$ as $n \to \infty$.) It remains to see whether $(|a_n|)$ decreases. One way to verify that $(|a_n|)$ decreases is to show directly that $|a_{n+1}| < |a_n|$ for all $n \in \mathbb{N}$ with $n \ge 2$; this involves a fair bit of algebra. As an alternative, we can consider the corresponding real-valued function $f(x) = (x^2+1)/(x^3-1)$ and compute

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

Since $f'(x) < 0$ for all $x > 1$, f is strictly decreasing on $(1, \infty)$. Since $|a_n| = f(n)$ for all $n \in \mathbb{N}$ with $n \ge 2$, $(|a_n|)$ is strictly decreasing as well. Thus, AST applies, and $\sum a_n$ converges. $\square$

As the previous example shows, most of the work in applying AST comes from verifying that $(|a_n|)$ decreases. This raises the question: for an alternating series $\sum a_n$ to converge, is it required that $(|a_n|)$ be decreasing? After all, we must require $a_n \to 0$, by the Divergence Test. The following example shows that the answer to our question is "no":

Example 9.70:

One way to construct a convergent alternating series $\sum a_n$ where $(|a_n|)$ is not decreasing is to take two positive convergent series and interleave their terms with alternating signs. For instance, if we interleave $\sum n^{-2}$ and $\sum 2^{-n}$ in this manner, then we obtain

$$1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{4} + \frac{1}{9} - \frac{1}{8} + \cdots = \sum_{n=1}^{\infty} a_n \qquad \text{with} \qquad a_{2n} = -2^{-n}, a_{2n-1} = \frac{1}{n^2}$$

For this series, $(|a_n|)_{n=k}^{\infty}$ is not decreasing for any $k \in \mathbb{N}^*$, because $1/n^2$ converges to 0 more slowly than 2^{-n} does. However, you can check that $\sum a_n$ converges, and in fact

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{1}{n^2} - \sum_{n=1}^{\infty} \frac{1}{2^n}$$

(To prove this formally, you should prove a relationship between the partial sums of the three series in the equation above. This relationship will have to consider two cases based on whether the partial sum on the left uses an even or odd number of terms.)

On the other hand, we can also use this interleaving idea to construct a divergent alternating series $\sum a_n$ where $(|a_n|)$ is not decreasing. To do this, we interleave two positive series with alternating signs, where exactly one of our positive series converges. For instance, if we interleave $\sum n^{-1}$ and $\sum 2^{-n}$ in this manner, then we obtain

$$1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{8} + \cdots = \sum_{n=1}^{\infty} a_n \qquad \text{with} \qquad a_{2n} = -2^{-n}, a_{2n-1} = \frac{1}{n}$$

Here, $\sum a_n$ cannot converge, because if it did, then $\sum a_n + \sum 2^{-n}$ would also converge, but this sum is the divergent series $\sum n^{-1}$.

It is worth noting that any alternating series can be obtained by interleaving two positive series. However, with many examples, the two interleaved series both diverge. (For instance, this is the case for the alternating harmonic series.) We will return to this idea when studying rearrangements at the end of this section. □

For a more involved example of an alternating series $\sum a_n$ where $(|a_n|)$ does not decrease, you should see Exercise 9.12.4.

Absolute and Conditional Convergence

Example 9.70 introduces an example of a series where the positive and the negative terms each yield convergent series. In particular, $\sum |a_n|$ converges for this example. In contrast, the alternating harmonic series $\sum (-1)^{n-1}/n$ has the property that it converges but $\sum |(-1)^{n-1}/n|$ diverges. To explore this idea more, we introduce the following definitions:

Definition 9.71. Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$ be a real-valued sequence. The series

$$\sum_{n=k}^{\infty} a_n$$

is said to be *absolutely convergent* if $\sum |a_n|$ converges. For instance, if $a_n \geq 0$ for each n, then $\sum a_n$ converges iff it converges absolutely.

If $\sum a_n$ converges but does not converge absolutely, then we say that $\sum a_n$ is *conditionally convergent*.

Informally, when a series $\sum a_n$ converges conditionally, this means that the magnitudes $(|a_n|)$ approach 0 too slowly for $\sum |a_n|$ to converge, but the positive and negative terms cancel each other enough to obtain convergence of the original series. This is demonstrated well by the zig-zagging partial sums for the alternating harmonic series in Figure 9.9 earlier. In contrast, an absolutely convergent series does not need to rely on any cancelation in order to obtain a convergent sum. More precisely, we have the following:

Theorem 9.72. *Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$ be a real-valued sequence. If $\sum a_n$ converges absolutely, i.e. $\sum |a_n|$ converges, then $\sum a_n$ also converges, and*

$$\left| \sum_{n=1}^{\infty} a_n \right| \leq \sum_{n=1}^{\infty} |a_n|$$

In other words, the Triangle Inequality holds for absolutely-convergent series.

Strategy. The most important part of this proof is the following inequality, which holds for any $x \in \mathbb{R}$:

$$-|x| \leq x \leq |x|$$

Informally, we'd like to apply this to each a_n and add the results for n from k to ∞ to get

$$-\sum_{n=k}^{\infty}|a_n| \le \sum_{n=k}^{\infty} a_n \le \sum_{n=k}^{\infty}|a_n|$$

However, there are two issues with this step. The first issue is that each side of the inequality is really a limit (since a series is a limit of partial sums), and we're not allowed to apply limits throughout an inequality unless we already know the limits exist. In other words, using this step to prove $\sum a_n$ converges is begging the question. Secondly, the main result that gives us inequalities between series, the Comparison Theorem, only applies to series with nonnegative terms, and $\sum a_n$ may have some negative terms.

To deal with these issues, we can rewrite our main inequality in a form where each part is nonnegative. To do this, we take $-|a_n| \le a_n \le |a_n|$ and add $|a_n|$ throughout to obtain

$$0 \le a_n + |a_n| \le 2|a_n|$$

Now, if we let $b_n = a_n + |a_n|$ and $c_n = 2|a_n|$, then we see that (b_n) and (c_n) are nonnegative sequences, and (c_n) dominates (b_n). Also, $\sum c_n$ converges by assumption. Thus, $\sum b_n$ converges by the Comparison Test. It follows that since $\sum b_n$ and $\sum |a_n|$ converge, so does $\sum(b_n - |a_n|) = \sum a_n$.

Now that we know $\sum a_n$ converges, the step that we wanted to do earlier is justified. This is how we establish the inequality stated in the theorem.

Proof. Let k and (a_n) be given as described. For every $n \in \mathbb{N}$ with $n \ge k$, we have

$$-|a_n| \le a_n \le |a_n| \to 0 \le a_n + |a_n| \le 2|a_n|$$

Thus, if we let $b_n = a_n + |a_n|$ and $c_n = 2|a_n|$, then $b_n, c_n \ge 0$, $b_n \le c_n$, and $\sum c_n$ converges by assumption. By the Comparison Test, $\sum b_n$ converges. Therefore, $\sum(b_n - |a_n|)$ converges, as it is the difference of two convergent series, and this series equals $\sum a_n$.

Lastly, to establish the inequality stated in the theorem, we know from the Triangle Inequality that for every $M \in \mathbb{N}$ with $M \ge k$,

$$\left|\sum_{n=k}^{M} a_n\right| \le \sum_{n=k}^{M} |a_n|$$

Because both sides have a finite limit as $M \to \infty$, taking a limit as $M \to \infty$ finishes the proof. ■

Theorem 9.72 tells us that absolute convergence is stronger than convergence. Because of this, it is often a good idea to test a mixed series for absolute convergence first. When testing for absolute convergence, we consider a nonnegative series $\sum |a_n|$, so we may use all of the tests for nonnegative series from earlier in this chapter. If the series does not converge absolutely, then we test for conditional convergence, often by using the AST.

Example 9.73:

For each $p \in \mathbb{R}$, consider the alternating p-series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^p}$$

The absolute-value series corresponding to this is $\sum 1/n^p$, which converges iff $p > 1$. Thus, this series is absolutely convergent iff $p > 1$. Another simple case to consider is when $p \leq 0$. In this case, $1/n^p$ does not approach 0, so the Divergence Test says this series diverges for $p \leq 0$.

Lastly, we consider when $0 < p \leq 1$. In this case, it is easy to check that $1/n^p$ decreases and converges to 0. Thus, the AST shows our series converges conditionally for $p \in (0, 1]$. □

Example 9.74:

For each $p \in \mathbb{R}$, consider the series

$$\sum_{n=1}^{\infty} \frac{\sin n}{n^p}$$

This series is a mixed series, but it is not alternating. Certainly, if $p \leq 0$, then the terms do not converge to 0, so in that case the series diverges by the Divergence Test. Hence, from now on we consider when $p > 0$.

When we take absolute values of the terms, we find

$$\left| \frac{\sin n}{n^p} \right| \leq \frac{1}{n^p}$$

Thus, when $p > 1$, our series converges absolutely by the Comparison Test. This does not, however, imply that the absolute-value series diverges for

$p \le 1$; the Comparison Theorem does not give you any information for a series which is dominated by a divergent series. Nevertheless, it is true that our absolute-value series diverges for $p \le 1$, but the argument is fairly difficult (see Exercise 9.7.31 for an outline).

Thus, for $p \in (0,1]$, the question remains: does our series converge conditionally? Because the series is not alternating, we cannot use the AST. It turns out that the series converges conditionally, though most proofs of this use some theory from complex-valued series, which we do not have enough time to cover in this book. For more information, look up the complex exponential function $e^{i\theta}$ and a test called Dirichlet's Test. $\square$

9.11 Series Rearrangements

Absolutely convergent series and conditionally convergent series have a very surprising difference related to rearranging their terms. Before mentioning precisely what that difference is, though, let's see some examples of what can happen when terms are rearranged in a series. In Example 9.23, we studied the divergent series $\sum(-1)^n$ and realized that we could make the partial sums diverge in different ways by ordering the terms differently. For instance, the ordering

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

which alternates between positive and negative terms, causes the partial sums to alternate between 1 and 0. On the other hand, the ordering

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

which takes more and more positive terms before balancing them out with negative terms, has an unbounded sequence of partial sums!

For a more surprising example, we offer the following:

Example 9.75:

We'll show that the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots = \log 2$$

can produce a different sum by reordering its terms! First, suppose we halve all the terms of this series, so that we get

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \cdots = \frac{1}{2}\log 2$$

Now, add these series for $\log 2$ and $(\log 2)/2$ together. The $1/2$'s cancel, and the two copies of $-1/4$ combine to form $-1/2$. Similarly, the $1/6$'s cancel, and the $-1/8$'s combine to form $-1/4$. Continuing with this pattern, we get

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \cdots = \frac{3}{2}\log 2$$

This shows that if we rearrange the alternating harmonic series by taking two positive terms for each negative term, the sum becomes $3/2$ times its original value! A similar tactic allows us to obtain, for each even $k \in \mathbb{N}^*$, a rearrangement of the alternating harmonic series which converges to $(1 + 1/k)\log 2$: see Exercise 9.12.23.

In fact, if you consider the rearrangement of the alternating harmonic series which uses p positive terms followed by m negative terms, where $p, m \in \mathbb{N}^*$, then the value of the sum can be found in terms of p and m. However, the approach used to compute this sum is more difficult: see Exercise 9.12.24. $\square$

Thus, we see that for some series, the order in which the terms are presented affects the sum. The mathematician Bernhard Riemann (the same Riemann from Riemann sums in integration) realized that rearrangements affect all conditionally convergent series (they do not affect any absolutely convergent series, as you can prove in Exercise 9.12.25). This result is formalized in the *Riemann Rearrangement Theorem*, also known as *Riemann's Series Theorem*.

Before presenting the Riemann Rerrangement Theorem in Theorem 9.79, it is worthwhile to define rearrangement more precisely and also introduce some useful results. Without loss of generality, let's focus only on series whose terms start with index 1.

Definition 9.76. Let $(a_n)_{n=1}^{\infty}$ be a real-valued sequence. For any bijection $\sigma : \mathbb{N}^* \to \mathbb{N}^*$, the sequence $(a_{\sigma(n)})_{n=1}^{\infty}$ is called a *rearrangement* of (a_n). In other words, a rearrangement is a reordering in which each index occurs exactly once. Similarly, the *rearrangment of the series* $\sum a_n$ is the series $\sum a_{\sigma(n)}$.

Certainly, the sum of a mixed series will depend on the interaction between the positive and negative terms. To make this interaction easier to analyze, let's introduce some useful notation to help us separate the positive and the negative terms. For any $x \in \mathbb{R}$, define

$$x^+ = \max\{x, 0\} = \begin{cases} |x| & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases} \qquad x^- = -\min\{x, 0\} = \begin{cases} |x| & \text{if } x < 0 \\ 0 & \text{if } x \ge 0 \end{cases}$$

x^+ and x^- are respectively called the *positive part* and *negative part* of x. The main idea behind these definitions is that x^+ is nonzero only when x is positive, and x^- is nonzero only when x is negative. When x^+ or x^- is nonzero, it equals x's magnitude.

Why are these parts useful? Suppose $\sum a_n$ is a given mixed series. The sum of its positive terms is

$$\sum_{n=1}^{\infty} a_n^+$$

(each negative a_n has $a_n^- = 0$, so negative terms don't affect this sum), and the sum of its negative terms is

$$-\sum_{n=1}^{\infty} a_n^-$$

Thus, the x^+ and x^- notation provides a convenient way to separate these two sums.

Furthermore, for any $x \in \mathbb{R}$, you can check that

$$x = x^+ - x^- \qquad |x| = x^+ + x^-$$

By solving these equations for x^+ and x^-, we obtain

$$x^+ = \frac{|x| + x}{2} \qquad x^- = \frac{|x| - x}{2}$$

This establishes plenty of useful relationships between $\sum a_n^+$, $\sum a_n^-$, $\sum a_n$, and $\sum |a_n|$, which are summarized in the following theorem:

Theorem 9.77. *Let $(a_n)_{n=1}^{\infty}$ be a real-valued sequence such that $\sum a_n$ converges.*

1. *$\sum a_n$ converges absolutely iff $\sum a_n^+$ and $\sum a_n^-$ both converge.*

2. *$\sum a_n$ converges conditionally iff $\sum a_n^+$ and $\sum a_n^-$ both diverge to ∞.*

Strategy. All of the parts of this theorem come from the relationships above. For instance, if $\sum a_n^+$ and $\sum a_n^-$ converge, then $\sum |a_n|$ converges because $|a_n| = a_n^+ + a_n^-$. The other direction of Part 1 is similar. For Part 2, if $\sum a_n$ converges conditionally, then $\sum a_n^+$ is the sum of a convergent and a divergent series, because $a_n^+ = (|a_n| + a_n)/2$. Thus, $\sum a_n^+$ diverges. The rest of this part is proven similarly.

Proof. Let $(a_n)_{n=1}^\infty$ be given such that $\sum a_n$ converges. For Part 1, if $\sum a_n^+$ and $\sum a_n^-$ both converge, then $\sum |a_n|$ converges as well because $|a_n| = a_n^+ + a_n^-$ for each $n \in \mathbb{N}^*$. Conversely, if $\sum |a_n|$ converges, then both $\sum a_n^+$ and $\sum a_n^-$ converge because we have

$$a_n^+ = \frac{|a_n| + a_n}{2} \qquad \text{and} \qquad a_n^- = \frac{|a_n| - a_n}{2}$$

and $\sum a_n$ also converges.

(For an alternate proof of this direction, note that for each $n \in \mathbb{N}^*$, $0 \le a_n^+$, $0 \le a_n^-$, and $a_n^+ + a_n^- = |a_n|$. Thus, $\sum a_n^+$ and $\sum a_n^-$ are nonnegative series which are dominated by $\sum |a_n|$, so they converge by CT.)

For Part 2, first suppose that $\sum a_n^+$ and $\sum a_n^-$ diverge. By Part 1, $\sum a_n$ does not converge absolutely. Because $\sum a_n$ is assumed to converge, it must converge conditionally, proving one direction. For the other direction, if $\sum a_n$ converges and $\sum |a_n|$ diverges, then the equations for a_n^+ and a_n^- above show that $\sum a_n^+$ and $\sum a_n^-$ are sums of a convergent and a divergent series. Therefore, they must diverge, and since they are nonnegative series, they diverge to ∞. ■

The Rearrangement Theorem And A "Tug Of War" Game

To illustrate the importance of Theorem 9.77, we present a "tug of war" metaphor. Let's suppose that two players named P and M (for "Plus" and "Minus") are holding onto opposite ends of a rope. This rope is hovering over a number line, and to start, the center of the rope hovers over 0. Player P wants to pull the rope farther to the right, increasing the coordinate of the rope's center, whereas Player M wants to pull the rope farther to the left, decreasing the coordinate of the rope's center. Also, Player P starts with a list of all the positive terms of a mixed series $\sum a_n$ (i.e. P starts with all the

a_n^+ terms), and Player N starts with a list of all the negative terms (i.e. all the a_n^- terms).

The players play a game with an infinite sequence of moves. To make a move, one of the players looks at their list of numbers, selects the first one on the list, moves the rope by that amount, and then removes the number they just used from their list. Thus, if Player P makes a move, then the rope moves to the right by a_n^+ units for some n. Similarly, if Player M makes a move, then the rope moves left by a_n^- units for some n. A player may make multiple moves in a row before giving the other player a turn; we only require that each player must give up their turn after finitely many moves.

Why does this game relate to rearrangements? On each move, some term of the original series is selected by a player; let's say that on the n^{th} move, the term $a_{\sigma(n)}$ is selected. (Thus, $a_{\sigma(n)} = a^+_{\sigma(n)}$ when Player P chooses, and $a_{\sigma(n)} = -a^-_{\sigma(n)}$ when Player N chooses.) Also, let's say S_n is the position of the rope's center after n moves. The game starts with the rope's center hovering over 0, so $S_0 = 0$. Also, if the rope hovers over S_n after n moves, then on the $(n+1)^{\text{st}}$ move, the rope is pulled by the amount $a_{\sigma(n+1)}$, so $S_{n+1} = S_n + a_{\sigma(n+1)}$. Therefore, we have

$$S_n = \sum_{i=1}^{n} a_{\sigma(i)}$$

so S_n is the n^{th} partial sum of the rearranged series with terms $(a_{\sigma(n)})_{n=1}^{\infty}$! Hence, as the players play their game, they are defining a rearranged version of the original series.

Example 9.78:

For a concrete example, consider the rearrangement of the alternating harmonic series from Example 9.75 which converges to $(3/2)\log 2$. This rearrangement is built by taking two positive terms, then one negative term, and then repeating. In terms of the "tug of war" game, this corresponds to Player P making two moves each turn and Player M making one move each turn.

Player P's list starts with the terms $a_1^+ = a_1 = 1$, $a_3^+ = a_3 = 1/3$, $a_5^+ = a_5 = 1/5$, and so forth. Player M's list starts with the terms $a_2^- = -a_2 = 1/2$, $a_4^- = -a_4 = 1/4$, $a_6^- = -a_6 = 1/6$, and so forth. The first few terms used in the rearrangement of the series (i.e. the first few moves made in the game) are, in order,

$$a_1, a_3, a_2, a_5, a_7, a_4, \ldots$$

which leads to $\sigma(1) = 1$, $\sigma(2) = 3$, $\sigma(3) = 2$, $\sigma(4) = 5$, $\sigma(5) = 7$, $\sigma(6) = 4$, and so forth. In general, you can prove by induction that for any $n \in \mathbb{N}^*$,

$$\sigma(3n-2) = 4n-3 \qquad \sigma(3n-1) = 4n-1 \qquad \sigma(3n) = 2n$$

and you can verify that σ is a bijection from $\mathbb{N}^*$ to $\mathbb{N}^*$.

Note that with this particular game, Player P always took two moves per turn, and Player M always took one move per turn. However, in general, the number of moves per turn does not have to be constant for either player. All that matters is that the number of moves per turn is always finite, so that each player will get infinitely many turns and be able to use every number on their list eventually. □

With this metaphorical game, we can explain the difference between absolutely convergent and conditionally convergent. For an absolutely convergent series, each player has a maximum total length that they can ever hope to pull (these total lengths are $\sum a_n^+$ for P and $\sum a_n^-$ for M). As the game proceeds, each player approaches their total limit of how much they can pull, so they become unable to pull the rope very far late in the game. Even when a player uses a lot of moves per turn, if that player is close to their total limit, then the rope won't budge very far after all these moves. This suggests that no matter who tugs when, the rope will settle down to a limit position at the end of the game. (In fact, the limit position will be the same for each game played with this series!)

In contrast, for a conditionally convergent series, $\sum a_n^+$ and $\sum a_n^-$ are both infinite, so neither player has a limit on their total amount to pull. This means that no matter how many moves a player has made, they still have "infinite strength" and can move the rope as far as desired given enough moves in a row. Although the moves become smaller and smaller as the game proceeds (because $a_n \to 0$ for a convergent series), to compensate, the players will take more and more moves before giving the other player a turn.

Although Riemann probably did not use the phrase "tug of war" to describe conditionally convergent series, his famous Rearrangement Theorem, which we are now ready to present, does use an algorithm which is similar to our game:

Theorem 9.79 (Riemann Rearrangement Theorem). *Let $(a_n)_{n=1}^\infty$ be a given real sequence such that $\sum a_n$ converges conditionally. Then for any $L \in \mathbb{R}$, there exists a bijection $\sigma : \mathbb{N}^* \to \mathbb{N}^*$ such that the rearranged series* *converges to L, i.e.*

$$\sum_{n=1}^{\infty} a_{\sigma(n)} = L$$

Strategy. To illustrate the main idea, we'll continue with the tug-of-war metaphor. Suppose that the two players start the game centered at 0, and a marker is placed on L. Player P wants to be on the right side of this marker, and Player M wants to be on the left side. Thus, for the first turn, Player P keeps pulling the rope to the right until L is passed, i.e. until the rope lies to the right of L. (If $L < 0$, then Player P makes only one move, just so their turn does something.) After that, Player M makes some moves pulling the rope until L is passed again, i.e. until the rope center lies to the left of L. At this point, control goes back to Player P, who pulls until L is passed again, and so forth. During the game, the two players pass L infinitely often.

Let's be more precise about how we keep track of the game in progress. At the start of the game, let's say that Player P's list consists of the subsequence $a_{p_1}^+$, $a_{p_2}^+$, $a_{p_3}^+$, and so forth. (Thus, $(a_{p_i})_{i=1}^\infty$ enumerates the subsequence of positive terms of $(a_n)_{n=1}^\infty$.) Similarly, Player M's list consists of the subsequence $a_{m_1}^-$, $a_{m_2}^-$, and so forth, yielding a subsequence $(a_{m_j})_{j=1}^\infty$ of negative terms of $(a_n)_{n=1}^\infty$. With this notation, we can keep track of how many moves each player has made with two variables i and j. At the start of the game, i and j are both 1. When Player P moves, they select a_{p_i} from their list, and then i is incremented. When Player M moves, they select a_{m_j} from their list, and then j is incremented. Thus, when Player P moves, $i - 1$ is the number of moves they have made beforehand, and similarly for Player M.

To see that this is a valid game, we need to know that each player only moves finitely many times before relinquishing their turn. This is where we use the fact that $\sum a_n^+$ and $\sum a_n^-$ are infinite. When it is Player P's turn, P wants to move the rope as far as needed to pass L. The rope only needs to move a finite distance before the turn ends. The total amount that P could possibly move the rope, $a_{p_i} + a_{p_{i+1}} + a_{p_{i+2}} + \cdots$, is infinite. Thus, only finitely many moves from P are needed to pass L. A similar argument works for Player M.

As this game is played, a rearrangement of the original series is built. As the game proceeds, the moves played appear further down the players' original lists (because each player makes at least one move per turn). Since $a_n \to 0$ as $n \to \infty$, it follows a_{p_i} and a_{m_j} go to 0 as $i, j \to \infty$ as well. Thus, the players overshoot L by less and less per turn, and the rope eventually settles in on L.

To make a formal proof, we just have to phrase the ideas from the game precisely. We will use the sequences (p_i) and (m_j). We'll maintain variables i and j to denote the progress of the game. We'll keep track of how many moves n have been made in total, using this to build a rearrangement $\sigma(n)$. We'll keep track of the sum we have built so far using the variable S. Also, we will keep track of whose turn it is with a variable; let's say $T = 1$ when it is P's turn and $T = -1$ when it is M's turn. With these values, we can describe the algorithm without any mention of the game.

Proof. Let $(a_n)_{n=1}^{\infty}$ be given as described. Also, let $L \in \mathbb{R}$ be given. We describe an algorithm for producing a bijection $\sigma : \mathbb{N}^* \to \mathbb{N}^*$ such that

$$\sum_{n=1}^{\infty} a_{\sigma(n)} = L$$

Before starting the algorithm, let $(a_{p_i})_{i=1}^{\infty}$ be the subsequence of (a_n) consisting of the positive terms. In other words, $(p_i)_{i=1}^{\infty}$ is a strictly increasing sequence from $\mathbb{N}^*$, and for each $n \in \mathbb{N}^*$, if $a_n > 0$, then there is some $i \in \mathbb{N}^*$ such that $n = p_i$. Similarly, let $(a_{m_j})_{j=1}^{\infty}$ be the subsequence of (a_n) consisting of the negative terms. (We may WLOG suppose that no terms of (a_n) are zero, because terms of zero do not affect the sum of a series.) Note that

$$\sum_{n=1}^{\infty} a_n^+ = \sum_{i=1}^{\infty} a_{p_i} = \infty \qquad \text{and} \qquad \sum_{n=1}^{\infty} a_n^- = \sum_{j=1}^{\infty} -a_{m_j} = \infty$$

because of Theorem 9.77 (since $\sum a_n$ converges conditionally).

Now, we give the steps of the algorithm, which uses the variables i, j, n, S, and T. i and j keep track of how many terms of our subsequences have been used, whereas n keeps track of the total number of terms used. S will maintain the partial sum of the rearrangement built so far, and T will either be 1 if we are adding positive terms or -1 if we are adding negative terms.

1. As initial values, set i, j, and n to 1, set S to 0, and set T to 1.

2. If T is 1, then add a_{p_i} to the current value of S, set $\sigma(n)$ equal to p_i, and increment the values of i and n. Otherwise, if T is -1, then add a_{m_j} to the current value of S, set $\sigma(n)$ equal to m_j, and increment the values of j and n.

3. If T is 1 and $S > L$, then set T to -1. Otherwise, if T is -1 and $S < L$, then set T to 1.

4. Go back to Step 2. (Hence, this algorithm will never terminate.)

At this point, we establish a useful property of this algorithm. Every time Step 2 is reached, we have the following:

$$S = \sum_{k=1}^{n-1} a_{\sigma(k)} = \sum_{k=1}^{i-1} a_{p_i} + \sum_{k=1}^{j-1} a_{m_j}$$

(This property is called an *invariant* of the algorithm.) You can check that the invariant holds the first time Step 2 is reached, due to the assignments in Step 1. Also, each time Step 2 is run, the invariant remains true because of how $\sigma(n)$ is assigned before n is incremented (along with i or j). This proves that our invariant always holds.

Next, we claim that at any point in the algorithm, T changes after finitely many more steps. To see this, let's first consider the case when T is 1, and suppose for contradiction that T never changes again. Thus, i always increases and j never changes in Step 2 as the algorithm continues. Since $\sum a_{p_i} = \infty$, there is some $I \in \mathbb{N}^*$ for which

$$\sum_{k=1}^{I-1} a_{p_k} > L - \sum_{k=1}^{j-1} a_{m_j}$$

When i takes the value I, our invariant shows that $S > L$, and Step 3 changes the value of T, contradicting our assumption. This proves the claim in the case where T is 1, and a similar proof works for the case when T is -1.

As a result, because T changes infinitely often, Step 2 guarantees that i and j are incremented infinitely often. Thus, every p_i and every m_j will occur in the range of σ as the algorithm proceeds. This proves σ is surjective. Furthermore, because $(p_i)_{i=1}^{\infty}$ and $(m_j)_{j=1}^{\infty}$ are strictly increasing sequences with disjoint ranges, σ will never repeat values: i.e. we cannot find $m, n \in \mathbb{N}^*$ such that $m \neq n$ and $\sigma(m) = \sigma(n)$. Thus, σ is injective as well, so it is bijective.

We have proven that this algorithm produces a rearrangement $(a_{\sigma(n)})_{n=1}^{\infty}$ of our original sequence. It remains to show that the partial sums of this rearrangement converge to L. Thus, let $\epsilon > 0$ be given; we need to show when n is sufficiently large,

$$\left| L - \sum_{k=1}^{n-1} a_{\sigma(k)} \right| = |L - S| < \epsilon$$

To see this, note that the largest S gets is just after a term a_{p_i} is added in Step 2 to make $S > L$, before Step 3 changes the value of T to -1. Similarly, the smallest S gets is just after a term a_{m_j} is added in Step 2 to make $S < L$, before Step 3 changes the value of T to 1. Thus, to guarantee that S lies between $L - \epsilon$ and $L + \epsilon$, we need only require that $|a_{p_i}|, |a_{m_j}| < \epsilon$ for i and j large enough. This follows because $a_n \to 0$ as $n \to \infty$ (since $\sum a_n$ converges), so $a_{p_i} \to 0$ as $i \to \infty$ and $a_{m_j} \to 0$ as $j \to \infty$ as well. ■

Remark. It is worth noting that there are many different rearrangements of a conditionally-convergent series which produce the same sum. For instance, in the proof we just gave, each player stopped moving immediately after they passed L. We could make a different game where each player stops moving two moves after they pass L. You can check that this would yield a different rearranged series with the sum L.

Also, it is possible to make rearrangements which cause the sum to diverge to ∞ or $-\infty$. See Exercise 9.12.27.

To demonstrate the algorithm used in the proof of Theorem 9.79, let's consider running that algorithm with the alternating harmonic series and $L = 1$ (so $a_n = (-1)^{n-1}/n$ for each $n \in \mathbb{N}^*$). Thus, the positive terms are a_1, a_3, a_5, and so forth, and the negative terms are a_2, a_4, a_6, and so on. In particular, $p_i = 2i - 1$ and $m_j = 2j$ for each $i, j \in \mathbb{N}^*$.

We start with $i, j, n = 1$, $S = 0$, and $T = 1$. Because $0 < L$, we use the term $a_{p_i} = a_1 = 1/1$ and increment i and n. Now, $S = 1 = L$, so Step 2 will use the term $a_{p_i} = a_3 = 1/3$ and increment i and n. Now, $S = 4/3 > L$, so Step 3 changes T to -1. The next run of Step 2 adds $a_{m_j} = a_2 = -1/2$ to S, incrementing j and n. Now, $S = 5/6 < 1$, so Step 3 changes T back to 1.

After this, $1/5$ is added, which makes S larger than 1 again. Thus, $-1/4$ is added, making S smaller than 1 again. Next, $1/7$ and $1/9$ are both added before S becomes greater than 1, and so forth. In summary, the first few terms of the rearranged series which converges to 1 are

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} - \frac{1}{4} + \frac{1}{7} + \frac{1}{9} - \frac{1}{6} + \cdots$$

After adding $-1/6$, when the algorithm reaches Step 4, it has the values $n = 9$, $i = 6$, $j = 4$, $T = 1$, and $S = 1097/1260 < L$. (Note that in the terms above, $i - 1$ of them are positive and $j - 1$ of them are negative, leading to $n - 1$ total terms.) Try computing a few more terms of this rearrangement on your own!

9.12 Exercises

In Exercises 1 through 3, use the Alternating Series Test to show the series converges, and find an $n \in \mathbb{N}^*$ such that the n^{th} partial sum approximates the sum of the series to within $1/100$.

1. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1} \log n}{n}$

2. $\displaystyle\sum_{n=1}^{\infty} (-1)^n \log(1 + 1/n)$

3. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n (n+2)}{n^3}$

4. For each $p \in (0, \infty)$, consider the series

$$\sum_{n=1}^{\infty} (-1)^{n-1} b_n \qquad \text{with} \qquad b_n = \frac{1}{n^p + (-1)^n}$$

(Note that if $p \leq 0$, then b_n does not approach 0 as $n \to \infty$.)

(a) Prove that this series converges absolutely iff $p > 1$.

(b) Now suppose $p \in (0, 1]$. Show that $(b_n)_{n=1}^{\infty}$ is not a decreasing sequence by proving that for all $n \in \mathbb{N}^*$, $b_{2n+1} > b_{2n}$. Thus, the Alternating Series Test does not apply. (Hint: The difference $(2n + 1)^p - (2n)^p$ can be approximated using the Mean Value Theorem.)

(c) Prove that for $p \in (0, 1]$,

$$\lim_{n \to \infty} \frac{b_{2n+1} - b_{2n}}{n^{-2p}} = L$$

exists and is positive. Thus, $b_{2n+1} - b_{2n} \sim Ln^{-2p}$.

(d) Use parts (a) and (c) to prove that our series converges conditionally iff $p \in (1/2, 1]$.

5. Prove Corollary 9.68.

6. Prove the following extension of the Root Test for mixed series:

   Theorem 9.80. *Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$ be a real-valued sequence. Suppose that*
   $$\lim_{n\to\infty} |a_n|^{1/n} = r$$
   where r is either a nonnegative real number or ∞. The series $\sum a_n$ converges when $r < 1$, and the series $\sum a_n$ diverges when $r > 1$ (this includes the case when $r = \infty$).

   (Hint: The only part requiring a new idea is when $r > 1$. Use the Divergence Test.)

7. Prove the following extension of the Ratio Test for mixed series:

   Theorem 9.81. *Let $k \in \mathbb{N}$ be given, and let $(a_n)_{n=k}^{\infty}$ be a real-valued sequence with $a_n \neq 0$ for all $n \geq k$. Suppose that*
   $$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = r$$
   where r is either a nonnegative real number or ∞. The series $\sum a_n$ converges when $r < 1$, and the series $\sum a_n$ diverges when $r > 1$ (this includes the case when $r = \infty$).

For Exercises 8 through 20, determine if the series is absolutely convergent, conditionally convergent, or divergent. You may use any tests from this chapter, including the extended Root and Ratio Tests from the previous two exercises, but you should justify why each test you use applies.

8. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\sqrt{n}}$

9. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n\sqrt{n}}{n+100}$

10. $\displaystyle\sum_{n=1}^{\infty} \frac{\cos^3 n}{n^{3/2}}$

11. $\displaystyle\sum_{n=1}^{\infty} (-1)^n \left(1 + \frac{1}{n}\right)^n$

12. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n^{1/n}}$

13. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{\log(e^n + e^{-n})}$

14. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)(-4)(-7)\cdots(-(3n+1))}{(2)(4)(6)\cdots(2n)}$

15. $\displaystyle\sum_{n=1}^{\infty}(-1)^n \sin\left(\frac{1}{n}\right)$

16. $\displaystyle\sum_{n=1}^{\infty}(-1)^n \left(1-\cos\left(\frac{1}{n}\right)\right)$

17. $\displaystyle\sum_{n=1}^{\infty}(-1)^n \left(e^{1/n}-1\right)$

18. $\displaystyle\sum_{n=1}^{\infty}(-1)^{n(n-1)/2}2^{-n}$

19. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n\log^2(n+1)}$

20. $\displaystyle\sum_{n=1}^{\infty} a_n \qquad \text{with} \qquad a_n = \begin{cases} 1/n & \text{if } n=m^2 \text{ for some } m \in \mathbb{N}^* \\ 1/n^2 & \text{otherwise} \end{cases}$

21. Use the idea of Theorem 9.77 to prove that for all real-valued sequences $(a_n)_{n=1}^{\infty}$, if exactly one of the series $\sum a_n^+$ and $\sum a_n^-$ converges, then $\sum a_n$ diverges.

22. When discussing the Integral Test, we showed the n^{th} harmonic number

$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$

is closely related to $\log n$. In this exercise, we outline a proof that $H_n - \log n$ approaches a limit γ as $n \to \infty$. This limit is called the *Euler-Mascheroni constant*[17]

$$\gamma \approx 0.5772156649\ldots$$

(a) Prove that for all $n \in \mathbb{N}^*$,

$$H_n - \log n = \sum_{i=1}^{n} \frac{1}{i} - \sum_{j=1}^{n-1} \int_j^{j+1} \frac{dx}{x}$$

[17] Interestingly, it is unknown, as of the writing of this book, whether γ is rational! See http://en.wikipedia.org/wiki/Euler-Mascheroni_constant for more details.

Use this to conclude that $H_n - \log n$ is the $(2n-1)^{\text{st}}$ partial sum of the alternating series

$$\sum_{n=1}^{\infty}(-1)^{n-1}a_n \qquad \text{with} \qquad a_{2n-1} = \frac{1}{n}, a_{2n} = \int_n^{n+1} \frac{dx}{x}$$

(b) Prove that the alternating series from part (a) converges. Use this to conclude that $H_n - \log n$ converges as $n \to \infty$.

Remark. From this exercise, it follows that H_n can be written as

$$H_n = \log n + \gamma + o(1) \text{ as } n \to \infty$$

using little-o notation. This becomes quite useful for analyzing variants on the harmonic series.

23. Use an argument similar to Example 9.75 to construct, for each even $k \in \mathbb{N}^*$, a rearrangement of the alternating harmonic series that converges to $(1 + 1/k)\log 2$.

24. At the end of Exercise 9.12.22, it is shown that

$$H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} = \log n + \gamma + o(1)$$

as $n \to \infty$. Let's use this to analyze rearrangements of the alternating harmonic series. For each $n \in \mathbb{N}^*$, let $a_n = (-1)^{n-1}/n$.

(a) Prove that for each $n \in \mathbb{N}^*$,

$$\sum_{i=1}^{n} \frac{1}{2i} = \frac{H_n}{2}$$

and

$$\sum_{i=1}^{n} \frac{1}{2i-1} = H_{2n} - \frac{H_n}{2}$$

(b) Suppose we take the rearrangement of the alternating harmonic series consisting of repeating p positive terms followed by m negative terms indefinitely, where $p, m \in \mathbb{N}^*$. (In Example 9.75, the rearrangement corresponds to $p = 2$ and $m = 1$.) Use part (a) to show that for each $n \in \mathbb{N}^*$, the $(p+m)n^{\text{th}}$ partial sum of this rearranged series is

$$H_{2pn} - \frac{H_{pn} + H_{mn}}{2}$$

From this, use the remark at the beginning of this exercise to show this equals

$$\log(2) + \frac{1}{2}\log(p/m) + o(1)$$

as $n \to \infty$.

(c) Using the rearrangement from part (b), show that the rearranged series converges to $\log 2 + (\log(p/m))/2$. (Hint: Part (b) only proves that the partial sums of the form $s_{(p+m)n}$ converge to that limit! Can you show that the partial sums of the form $s_{(p+m)n+1}$ have the same limit? What about the partial sums of the form $s_{(p+m)n+2}$, and so on? After this, adapt the idea of Exercise 9.2.21.)

25. Let $(a_n)_{n=1}^{\infty}$ be an infinite real-valued sequence such that $\sum |a_n|$ converges. By Theorem 9.77, we may define

$$L_+ = \sum_{n=1}^{\infty} a_n^+ \qquad L_- = \sum_{n=1}^{\infty} a_n^-$$

with $L_+, L_- \in [0, \infty)$. Prove that every rearrangement of $\sum a_n$ converges to $L_+ - L_-$. (Hint: For each $\epsilon > 0$, take enough terms from $\sum a_n^+$ so that the partial sum is within $\epsilon/2$ of the total sum. Do a similar step with $\sum a_n^-$. How many terms should be taken from the rearranged series?)

26. Consider the series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{\sqrt{n}}$$

This series is conditionally convergent. Find the first 20 terms in a rearrangement of the series which converges to the sum 2. Also, find the first 20 terms in a rearrangement of the series which converges to the sum -1.

27. Prove that every conditionally convergent series has a rearrangement which produces the sum ∞, as well as a rearrangement producing the sum $-\infty$. (Hint: Adapt the proof of the Rearrangement Theorem allowing L to change at certain points in the algorithm.)

28. Previously, in Example 9.56, we saw that

$$\int_1^\infty \frac{|\sin x|}{x}\,dx$$

diverges. In this exercise, we will outline a proof that

$$\int_1^\infty \frac{\sin x}{x}\,dx$$

converges. (Thus, this improper integral converges conditionally.) The idea of this argument, much like the argument in Example 9.56, is to consider "hills" of positive and negative area. For each $n \in \mathbb{N}^*$, define

$$I_n = \int_{n\pi}^{(n+1)\pi} \frac{\sin x}{x}\,dx$$

i.e. I_n represents the area of the n^{th} hill. (Note that I_n is positive iff n is even.)

(a) Use the results of Example 9.56 to show that $(|I_n|)_{n=1}^\infty$ is decreasing and $I_n \to 0$ as $n \to \infty$. Thus, by the Alternating Series Test, $\sum I_n$ converges.

(b) Prove that

$$\lim_{b\to\infty} \int_1^b \frac{\sin x}{x}\,dx = \int_1^\pi \frac{\sin x}{x}\,dx + \lim_{N\to\infty} \sum_{n=1}^N I_n$$

This proves that our improper integral converges. (Hint: Compare the partial integral up to b with the partial sums using $\lfloor b/\pi \rfloor$ or $\lceil b/\pi \rceil$ terms.)

Chapter 10

Uniform Convergence and Power Series

In the previous chapter, we studied sequences of real numbers, particularly in the context of infinite series. Along the way, we demonstrated some techniques for analyzing convergence of sequences and series, such as the BMCT and many series tests. A key theme in that chapter is approximation: the terms of a sequence become close to the limit, and partial sums of a series become close to the sum of the series. Similarly, in this chapter, we study sequences and series of real-valued functions, and we analyze how close functions in a sequence become to a "limiting function".

It's worth noting that we've already seen some important examples of sequences of real-valued functions, although we may not have referred to them as such. For instance, we studied geometric series, showing that

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$$

for every $r \in (-1,1)$. Rather than think of r as a fixed constant, we can think of it as a variable in a function. More precisely, if we define $f : (-1,1) \to \mathbb{R}$ by

$$f(r) = \frac{1}{1-r} \text{ for all } r \in (-1,1)$$

and for each $n \in \mathbb{N}$, we define $f_n : (-1,1) \to \mathbb{R}$ by

$$f_n(r) = \sum_{i=0}^{n} r^i \text{ for all } r \in (-1,1)$$

(so $f_n(r)$ is the n^{th} partial sum of the geometric series), then our geometric series formula says that $f_n(r) \to f(r)$ for every $r \in (-1,1)$ as $n \to \infty$. In this way, the sequence of functions $(f_n)_{n=0}^{\infty}$ approaches the function f.

As another example, we have the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \text{ for all } s > 1$$

As with the geometric series, $\zeta(s)$ is the limit of a sequence of partial sums. More precisely, if

$$\zeta_n(s) = \sum_{i=0}^{n} \frac{1}{i^s}$$

then $\zeta_n(s) \to \zeta(s)$ as $n \to \infty$ for all $s > 1$. Since we don't have a convenient way to find exact values of $\zeta(s)$, the next best thing to do is to ask how similar ζ is to each ζ_n. This raises questions such as: Since each ζ_n is continuous, does it follow that ζ is? Is ζ bounded on $(1,\infty)$ because each ζ_n is? Is ζ differentiable or integrable? Later in this chapter, we will be able to answer many of these questions.

Last, but certainly not least, Taylor polynomials provide a very good example of sequences of functions. When $a \in \mathbb{R}$ is given, and f is a real function which is infinitely differentiable at a, we can form the sequence of Taylor polynomials $(T_n f(x;a))_{n=0}^{\infty}$. Intuitively, $T_n f(x;a)$ is supposed to be a good approximation to $f(x)$ when x is near a, and increasing n is supposed to make the approximation better. More precisely, if S is the set of all $x \in \mathbb{R}$ satisfying $E_n f(x;a) \to 0$ as $n \to \infty$ (recall that $E_n f(x;a)$ is the n^{th}-order Taylor error), then for all $x \in S$,

$$\begin{aligned} f(x) &= \lim_{n\to\infty} T_n f(x;a) \\ &= \lim_{n\to\infty} \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i!}(x-a)^i \end{aligned}$$

This last expression for $f(x)$, which uses an infinite series of powers of $x-a$, is called a *power series* centered at a. We will be able to prove a useful result about the set S later in this chapter: S must be a (possibly infinite) interval centered at a.

To generalize these examples, suppose that we have a sequence of real-valued functions $(f_n)_{n=0}^{\infty}$, all with the same domain D. The main question we ask is: does this sequence of functions converge to a limit function f, in the sense that $f_n(x) \to f(x)$ as $n \to \infty$ for all $x \in D$? When f_n approaches f in this manner, we say that f is the *pointwise limit* of the sequence (f_n) on D. Intuitively, you may suspect that when f_n converges to f, f_n behaves much like f when n is sufficiently large. More formally, we'd like to know which properties of the f_n functions carry over to f, such as continuity, differentiability, integrability, boundedness, and so forth.

As we will soon see, however, there is very little that can be proven in general when f is the pointwise limit of a sequence (f_n). The main issue is that the speed at which $f_n(x)$ converges to $f(x)$ may vary wildly based on x. To address this, we introduce another way in which functions can converge to other functions, called *uniform convergence*. In addition to pointwise convergence, uniform convergence requires $f_n(x)$ to approach $f(x)$ in a manner independent of x. We will show that when f_n converges to f uniformly, many properties of the f_n functions carry over to f. This is useful in many applications, where we create a function f with some desired properties by building it as a uniform limit of a sequence (f_n) of approximating functions. For instance, at the end of this chapter, we'll show how to solve some differential equations and how to construct a function which is continuous on all of $\mathbb{R}$ but not differentiable at any point.

10.1 Perils of Pointwise Convergence

To start our discussion, let's first make a precise definition of what it means for a sequence of functions to converge. The simplest possible definition is the notion of pointwise convergence:

Definition 10.1. Let $D \subseteq \mathbb{R}$ and $k \in \mathbb{N}$ be given, and let $(f_n)_{n=k}^{\infty}$ be an infinite sequence of real functions defined on D. Also, let f be a real-valued function defined on D. We say that (f_n) *converges to f pointwise on D as $n \to \infty$* if

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for all $x \in D$. f is called the *(pointwise) limit* of (f_n) on D, and we write

$$f_n \to f \text{ on } D \text{ as } n \to \infty$$

The set D is sometimes not mentioned when D is clear from the context.

The definition of pointwise convergence is simple. However, it's often not useful for proving theorems, since when $f_n \to f$ pointwise, very few properties satisfied by the f_n functions can be guaranteed to carry over to f. In the rest of this section, we provide some examples to support this idea. In the next section, we will introduce the notion of uniform convergence to address most of the issues presented by these examples.

First, we show that in general, a pointwise limit of bounded functions might not be bounded:

Example 10.2:

Consider the geometric series from this chapter's introduction. Thus, for each $n \in \mathbb{N}$ and each $x \in (-1, 1)$, we define

$$f_n(x) = \sum_{i=0}^{n} x^i \qquad f(x) = \frac{1}{1-x}$$

We know that $f_n \to f$ on $(-1, 1)$. In fact, the equation defining f_n actually describes a continuous function on $[-1, 1]$, so each f_n is bounded on $(-1, 1)$. (In fact, it is easy to show that $\sup\{|f_n(x)| \mid x \in (-1, 1)\} = n + 1$ for all $n \in \mathbb{N}$.) However, f is unbounded on $(-1, 1)$.

The issue with this example is that the bound on the range of f_n depends on n. Specifically, $\sup\{|f_n(x)| \mid x \in (-1, 1)\}$ approaches ∞ as $n \to \infty$. On the other hand, it is not too difficult to prove that f is bounded when each f_n is bounded by the SAME bound. More precisely, you can prove in Exercise 10.3.2 that if there is a constant $M > 0$ satisfying

$$\forall x\, \forall n \in \mathbb{N}\, |f_n(x)| \leq M$$

then $|f(x)| \leq M$ for all x as well. $\square$

Next, we proceed to what is arguably the largest problem with the notion of pointwise convergence: it does not preserve continuity. In other words, there are examples of sequences of real functions (f_n) where each $f_n : D \to \mathbb{R}$ is continuous on D but the pointwise limit is not continuous on D. For instance, we offer the following examples:

Example 10.3:

For each $n \in \mathbb{N}^*$, we define $f_n : [0, 1] \to \mathbb{R}$ by

$$f_n(x) = x^n$$

for each $x \in [0,1]$. It is not hard to check that for each $x \in [0,1]$,

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x = 1 \end{cases}$$

Clearly, the limit function is discontinuous at 1, although each f_n is continuous on $[0,1]$.

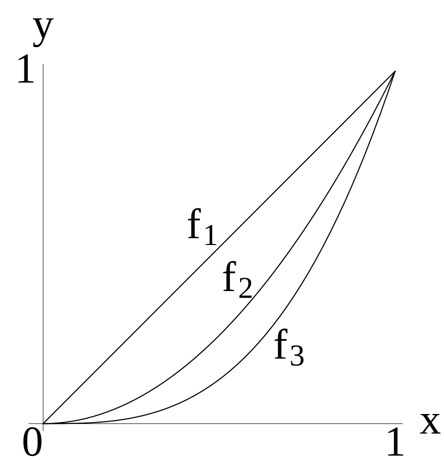

Figure 10.1: Graphs of $f_n(x) = x^n$ for some values of n

Informally, why does this behavior occur? Some insight may be obtained from the graphs in Figure 10.1. The closer x is to 0, the faster x^n converges to 0. (In fact, when $0 < x < y < 1$, $x^n = o(y^n)$ as $n \to \infty$.) Thus, as n grows, points with x close to 1 are "lagging behind" the values where x is close to 0. Once we reach the limit f, every point $x \in [0,1)$ has "caught up" and obtained its limit of 0. This sudden shift from "lagging behind" to "catching up" causes the limit function's graph to jump at 1. $\square$

Example 10.4:

Suppose $f_n : \mathbb{R} \to \mathbb{R}$ is defined for each $n \in \mathbb{N}^*$ and $x \in \mathbb{R}$ by

$$f_n(x) = \cos(\pi x)^{2n}$$

A similar argument as in the previous example shows that

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Z} \\ 1 & \text{if } x \in \mathbb{Z} \end{cases}$$

Thus, this example shows that the pointwise limit of continuous functions can have infinitely many discontinuities![1] $\square$

For a somewhat different example, we demonstrate one way to start with a limit function in mind and then create a sequence of functions with that limit. This is useful for generating counterexamples to claims about pointwise convergence.

Example 10.5:

Recall the Heaviside function $H : \mathbb{R} \to \mathbb{R}$ is defined for all $x \in \mathbb{R}$ by

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

In this example, we show a way of making a sequence of continuous functions (f_n) which converge pointwise to H on $\mathbb{R}$.

The key idea here is that H has only one discontinuity, and it's at 0. Thus, to make a continuous function which is a good approximation of H, we can make the function equal to H until we get a certain distance from 0, and then we connect the pieces of the graph together with a simple curve. One such function is defined by the equation

$$f(x) = \begin{cases} H(x) & \text{if } x \geq 0 \text{ or } x < -1 \\ x+1 & \text{if } x \in [-1, 0) \end{cases}$$

Intuitively, this choice of f agrees with H everywhere except for an interval of length 1, then it uses a straight segment to join the two parts of the Heaviside function together. For convenience, we chose to make the function linear on $[-1, 0)$, though any continuous function on $[-1, 0]$ with the value 0 at -1 and the value 1 at 0 could be used instead.

Now, we use this idea to make a sequence of approximating functions. For each $n \in \mathbb{N}^*$, we make the n^{th} function agree with H everywhere except for an interval of length $1/n$ (this way, the "exception interval" becomes arbitrarily small as n approaches ∞), and then we use a segment to join the two parts of the curve. This leads to the following equation for all $x \in \mathbb{R}$:

$$f_n(x) = \begin{cases} 0 & \text{if } x < -1/n \\ nx+1 & \text{if } x \in [-1/n, 0) \\ 1 & \text{if } x \geq 0 \end{cases}$$

[1]Remark: It is unknown to the author whether it is possible to have a pointwise limit of continuous functions be nowhere continuous.

It is not hard to check that each f_n is continuous on $\mathbb{R}$. It is also clear that $f_n \to H$ on $[0, \infty)$. To prove that $f_n \to H$ on $(-\infty, 0)$, we take an arbitrary $x < 0$ and aim to show that $f_n(x) \to 0$ as $n \to \infty$. Intuitively, we see that as n grows larger, the "exception interval" for f_n gets smaller, closing in on 0, so eventually x lies outside this "exception interval". More formally, whenever n is large enough to satisfy $1/n < |x|$ (i.e. $n > 1/|x|$), we have $x < -1/n$ and thus $f_n(x) = 0 = H(x)$.

This finishes the proof that $f_n \to H$ on $\mathbb{R}$. Note that the fact that f_n is linear on $(-1/n, 0)$ never gets used in the proof! Informally, the behavior on the "exceptional interval" is irrelevant, because as n approaches ∞, the "exceptional interval" becomes compressed down to the single-element set $\{0\}$. Thus, we can choose to define f_n on $(-1/n, 0)$ however we like, provided that f_n stays continuous on $\mathbb{R}$. $\square$

Based on the previous examples, it may come as no surprise that pointwise limits don't preserve integrability either. In other words, a pointwise limit of functions which are integrable on an interval $[a, b]$ might not be integrable on $[a, b]$. For instance, we have the following:

Example 10.6:

Perhaps our most well-studied example of a non-integrable function is the characteristic function of the rationals $\chi_{\mathbb{Q}}$, defined for all $x \in \mathbb{R}$ by

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

We aim to show that $\chi_{\mathbb{Q}}$ can be expressed as the pointwise limit of integrable functions.

To do this, we first note what makes $\chi_{\mathbb{Q}}$ non-integrable on any interval $[a, b]$ with $a < b$. $\chi_{\mathbb{Q}}$ jumps between 0 and 1 very rapidly, because $\mathbb{Q}$ and $\mathbb{R} - \mathbb{Q}$ are both dense in $\mathbb{R}$. This forces every lower step function to lie below $y = 0$ and every upper step function to lie above $y = 1$.

This suggests that one way to approximate $\chi_{\mathbb{Q}}$ is to use only finitely many jumps, i.e. make a function which takes the value 1 at only finitely many rationals. Since the value of an integral is never affected by finitely many points, the resulting function is integrable. To formalize this idea, let's enumerate the rationals as $(r_n)_{n=1}^{\infty}$ (see the definition of the staircase function in Chapter 3 for more details). Now, for each $n \in \mathbb{N}^*$, we define

$f_n = \chi_{\{r_1,r_2,\ldots,r_n\}}$, i.e.

$$f_n(x) = \chi_{\{r_1,r_2,\ldots,r_n\}}(x) = \begin{cases} 1 & \text{if } x \in \{r_1, r_2, \ldots, r_n\} \\ 0 & \text{otherwise} \end{cases}$$

Each function of this sequence takes the value 0 everywhere except at finitely many points, so

$$\int_a^b f_n(x)\,dx = 0$$

for all $a, b \in \mathbb{R}$ and all $n \in \mathbb{N}^*$. Also, we claim that $f_n \to \chi_{\mathbb{Q}}$ on $\mathbb{R}$. It is obvious that $f_n \to \chi_{\mathbb{Q}}$ on $\mathbb{R} - \mathbb{Q}$. For each $x \in \mathbb{Q}$, there is some $N \in \mathbb{N}^*$ such that $x = r_N$, and thus whenever $n \geq N$, we have $f_n(x) = 1 = \chi_{\mathbb{Q}}(x)$. This proves that $f_n \to \chi_{\mathbb{Q}}$ on $\mathbb{Q}$ as well. $\square$

Lastly, we present another issue that can arise with integrals and pointwise convergence. Even when each f_n is integrable, $f_n \to f$, and f is integrable, the integral of f does not have to be related to the integrals of the f_n functions. In particular, it is possible to have

$$\int_a^b \left(\lim_{n\to\infty} f_n(x)\right) dx \neq \lim_{n\to\infty} \left(\int_a^b f_n(x)\,dx\right)$$

In other words, the integral of a limit might not equal the limit of integrals. The following example demonstrates this:

Example 10.7:

In this example, we create a sequence of functions $(f_n)_{n=1}^\infty$ on $\mathbb{R}$ such that $f_n \to 0$ pointwise but the integrals of f_n do not approach 0. To do this, we use an idea similar to the one from Example 10.5: we make f_n equal to 0 everywhere except for an "exception interval" of length $1/n$, and then in that interval, we make a function which has area 1 underneath it. One way of doing this is with the following definition for all $x \in \mathbb{R}$ and all $n \in \mathbb{N}^*$:

$$f_n(x) = \begin{cases} 0 & \text{if } x < 0 \text{ or } x > 1/n \\ 4n^2 x & \text{if } x \in [0, 1/(2n)) \\ 4n - 4n^2 x & \text{if } x \in [1/(2n), 1/n] \end{cases}$$

(Equivalently, we could define $f_n(x) = 2n - 2n^2|x - 1/(2n)|$ for $x \in [0, 1/n]$.)

We constructed this function so that the area between the graph of f_n and the x-axis forms a triangle with height $2n$ and base $1/n$: see Figure 10.2. Therefore, for each $n \in \mathbb{N}^*$, we have

$$\int_0^1 f_n(x)\,dx = 1$$

As n grows larger, the triangle grows taller and skinnier, but its area does not change. Essentially, the triangle is getting pressed against the y-axis as n grows. It is also possible to view this as a "sliding mass" example, where the mass of the triangle (i.e. the area under the curve) slides closer to $x = 0$ as n grows. In the limit, the triangle has become completely flattened, and the limiting curve has zero area.

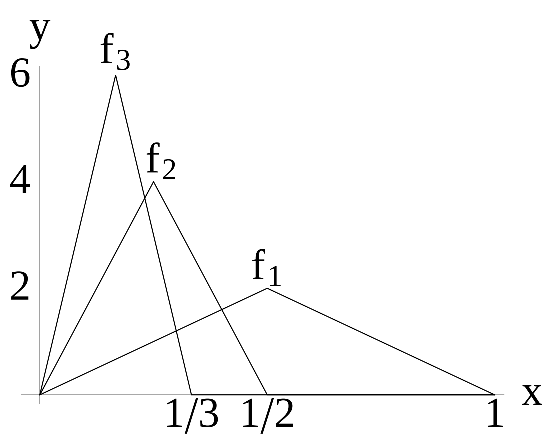

Figure 10.2: Several graphs from Example 10.7 with a triangle of area 1

Now, we prove that $f_n \to 0$ pointwise. This proof proceeds in the same way as in Example 10.5. Certainly, $f_n(x) = 0$ for every $x \le 0$. Also, when $x > 0$, then $f_n(x) = 0$ for all $n > 1/x$ (i.e. once n is greater than $1/x$, x does not lie in f_n's "exception interval"). Note that this proof doesn't use any information about how f_n is defined on $(0, 1/n)$, so we could replace the triangle shape in that interval with any other curve with area 1 underneath to generate a similar counterexample. □

10.2 Uniform Convergence

As we have seen in the previous section, when $f_n \to f$ pointwise on a set D, there is rather little that can be proven about f. However, if we require that $f_n(x)$ converge to $f(x)$ at roughly the same rate for all $x \in D$, then we are able to show much more about f. More precisely, we have the following definition:

Definition 10.8. Let $D \subseteq \mathbb{R}$ and $k \in \mathbb{N}$ be given, and let $(f_n)_{n=k}^{\infty}$ be an infinite sequence of real-valued functions defined on D. Also, let f be a real-valued function defined on D. We say that (f_n) *converges to f uniformly on D as $n \to \infty$* if the following holds: for every $\epsilon > 0$, there is an $N \in \mathbb{N}$ (depending ONLY on ϵ) such that

$$|f(x) - f_n(x)| < \epsilon$$

for all $x \in D$ and for all $n \in \mathbb{N}$ with $n > N$. We call f the *uniform limit of* (f_n) on D and write

$$f_n \to f \text{ uniformly on } D \text{ as } n \to \infty$$

As with pointwise convergence, the set D is sometimes not mentioned when it is understood from context.

Several remarks are in order concerning this definition and how it differs from pointwise convergence. If we write out the definition of pointwise convergence completely, using quantifiers, then we obtain

$$\forall \epsilon > 0\ \forall x \in D\ \exists N \in \mathbb{N}\ \forall n > N\ (|f(x) - f_n(x)| < \epsilon)$$

In contrast, the definition of uniform convergence yields

$$\forall \epsilon > 0\ \exists N \in \mathbb{N}\ \forall x \in D\ \forall n > N\ (|f(x) - f_n(x)| < \epsilon)$$

There is only one difference: uniform convergence has $\exists N$ before $\forall x$, whereas pointwise convergence has $\forall x$ before $\exists N$. However, this is a rather large difference. With uniform convergence, N has to be chosen independently of x, whereas N can depend on x with pointwise convergence. (In particular, uniform convergence implies pointwise convergence, but not conversely.) By forcing N to be chosen independently of x, uniform convergence forces the

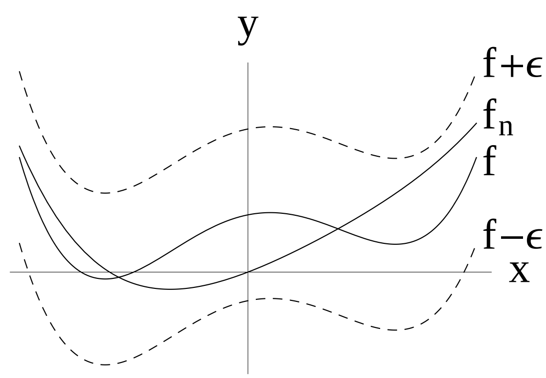

Figure 10.3: A function f_n lying between $f - \epsilon$ and $f + \epsilon$

entire graph of f_n to lie within ϵ units of the graph of f when $n > N$. This is illustrated in Figure 10.3, where functions f_n and f are graphed, and the dotted curves represent $f \pm \epsilon$.

To gain a better appreciation of uniform convergence, let's revisit a couple of the examples from the previous section and see whether they exhibit uniform convergence:

Example 10.9:

In Example 10.3, we defined

$$f_n(x) = x^n \qquad f(x) = \begin{cases} 0 & \text{if } x \in [0,1) \\ 1 & \text{if } x = 1 \end{cases}$$

for all $x \in [0,1]$. We saw that $f_n \to f$ pointwise on $[0,1]$ as $n \to \infty$. Intuitively, this convergence should not be uniform because when x is close to 1, x^n approaches 0 much more slowly than when x is close to 0.

More precisely, we can show that uniform convergence fails by showing that the definition can't be satisfied with $\epsilon = 1/2$. If the convergence were uniform, then we could choose some $N \in \mathbb{N}^*$ such that $|f(x) - f_N(x)| < 1/2$ for all $x \in [0,1]$. In particular, this means that $|x^N| < 1/2$ for all $x \in [0,1)$. However, this is not possible because $x^N \to 1$ as $x \to 1^-$. Alternately, we can choose x to be $(3/4)^{1/N}$ as a counterexample (note that this choice of x can depend on N).

Another way to think about this is that since each f_n maps the interval $[0,1)$ to $[0,1)$, we cannot force the entire graph to lie within 1/2 a unit of

0. In fact, we can't force the entire graph to lie within ϵ units of 0 for any $\epsilon < 1$, because for each $n \in \mathbb{N}^*$,

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| = \sup_{x\in[0,1)} |x^n| = 1$$

(think of this supremum as the "maximum distance"[2] between f_n and f). When convergence is uniform, the distance between f_n and f must approach 0, but in this example the distance is always 1. See Exercise 10.3.4 for more details. $\square$

Example 10.10:

In Example 10.2, we defined

$$f_n(r) = \sum_{i=0}^{n} r^i \qquad f(r) = \sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$$

for all $r \in (-1, 1)$. We saw $f_n \to f$ pointwise on $(-1, 1)$. To see whether this convergence is uniform, we analyze how far apart f_n and f for each $r \in (-1, 1)$ and $n \in \mathbb{N}^*$:

$$\begin{aligned} f(r) - f_n(r) &= \sum_{i=0}^{\infty} r^i - \sum_{i=0}^{n} r^i = \sum_{i=n+1}^{\infty} r^i \\ &= r^{n+1}\left(1 + r + r^2 + \cdots\right) = \frac{r^{n+1}}{1-r} \end{aligned}$$

Note that as $r \to 1^-$, $r^{n+1}/(1-r) \to \infty$. Therefore, the set $\{|f(r) - f_n(r)| \mid r \in (-1, 1)\}$ is unbounded, so f_n and f are always "infinite distance apart". Thus, no choice of N satisfies the definition of uniform convergence for ANY $\epsilon \in (0, \infty)$. In fact, in Exercise 10.3.10, you can show that a uniform limit of bounded functions must be bounded. It follows that convergence cannot be uniform here, because each f_n is bounded on $(-1, 1)$ and f is not bounded on $(-1, 1)$.

However, we can obtain uniform convergence on smaller domains than $(-1, 1)$. Let's say $b \in (0, 1)$ is given, and we consider f_n and f restricted to

[2]This notion of "maximum distance" is called the L_∞ norm of $f_n - f$. There are also other ways to assign a notion of distance between two functions, but we will not cover them here.

$(-b, b)$. (Intuitively, we know that f is unbounded when r is near 1 or -1, so we restrict the domain to stay a fixed distance away from 1 and -1.) The previous calculations show that for all $r \in (-b, b)$,

$$|f(r) - f_n(r)| = \frac{|r|^{n+1}}{|1-r|} \leq \frac{b^{n+1}}{1-b}$$

Now, the maximum distance between f and f_n is bounded. Furthermore, since $b < 1$, $b^{n+1} \to 0$ as $n \to \infty$. Thus, the maximum distance between f and f_n approaches 0 on $(-b, b)$, so $f_n \to f$ uniformly on $(-b, b)$! More formally, for any $\epsilon > 0$, choose N large enough so that $b^{N+1}/(1-b) < \epsilon$, and this choice of N satisfies the definition of uniform convergence. (Note that this argument cannot prove uniform convergence on $(-1, 1)$ because N depends on b as well as ϵ.) $\square$

The previous example illustrates a very important point. Sometimes, when we consider convergence of a sequence of functions on some interval, we may not obtain uniform convergence because of undesired behavior at the ends of the interval. In these situations, if we restrict the domain to stay away from the ends of the interval, then we frequently obtain uniform convergence. We illustrate this idea with a couple examples, and we will also return to this point in the upcoming section on power series.

Example 10.11:

First, for each $n \in \mathbb{N}^*$ and $x \in [0, \infty)$, let's define $f_n, f : [0, \infty) \to \mathbb{R}$ by

$$f_n(x) = \frac{x}{n+x} \qquad f(x) = 0$$

It is clear that for each $x \in \mathbb{R}$, $f_n(x) \to f(x)$ as $n \to \infty$. In other words, $f_n \to f$ pointwise. However, is this convergence uniform on $[0, \infty)$? Note that the maximum distance between f_n and f is

$$\sup_{x \in [0,\infty)} \frac{x}{n+x} = 1$$

because f_n is strictly increasing with a horizontal asymptote at $y = 1$. (The easiest way to see $f_n(x)$ is strictly increasing is to use long division to write it as $1 - n/(n+x)$.) Thus, we do not have uniform convergence on $[0, \infty)$.

However, let's try restricting our domain to an interval of the form $[0, b]$ where $b > 0$. (This stays away from the "end" of ∞ which was making uniform convergence fail.) On $[0, b]$, the distance between f_n and f is

$$\sup_{x \in [0,b]} \frac{x}{n+x} = \frac{b}{n+b}$$

Since $b/(n+b)$ approaches 0 as $n \to \infty$, f_n converges to f uniformly on $[0, b]$ for any $b > 0$. Thus, f_n converges to f uniformly on arbitrarily long finite intervals but not on an infinite interval! □

Example 10.12:

In contrast to the f_n functions from Example 10.11, let's consider defining $g_n, g : [0, \infty) \to \mathbb{R}$ for all $x \in [0, \infty)$ and all $n \in \mathbb{N}^*$ by

$$g_n(x) = \frac{x}{n+x^2} \qquad g(x) = 0$$

Here, we have $g_n \to g$ pointwise on $[0, \infty)$, but unlike the f_n functions from Example 10.11, the g_n functions have horizontal asymptotes at $y = 0$. In fact, g_n is not strictly increasing on $[0, \infty)$. Several graphs of the g_n functions are drawn in Figure 10.4.

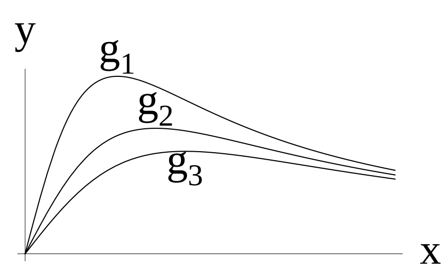

Figure 10.4: Several graphs of $g_n(x) = x/(n + x^2)$

From these graphs, we see that each g_n rises until it attains a maximum and then falls off gradually towards 0. This suggests that the distance between g_n and g,

$$\sup_{x \in [0,\infty)} \frac{x}{n+x^2}$$

is attained when g_n reaches its maximum.

Therefore, we aim to maximize g_n. To do this, we first find the derivative:

$$g_n'(x) = \frac{(n+x^2)(1) - (x)(2x)}{(n+x^2)^2} = \frac{n-x^2}{(n+x^2)^2}$$

Since the denominator is always positive, we find that $g_n'(x) > 0$ when $x < \sqrt{n}$ and $g_n'(x) < 0$ when $x > \sqrt{n}$. Thus, the absolute maximum of g_n is attained at $\sqrt{n}$. Let's say $x_n = \sqrt{n}$; thus, for all $x \in [0,\infty)$, we have

$$0 \le g_n(x) \le g_n(x_n) = \frac{\sqrt{n}}{n + (\sqrt{n})^2} = \frac{1}{2\sqrt{n}}$$

Since $1/(2\sqrt{n}) \to 0$ as $n \to \infty$, this proves that $g_n \to g$ uniformly on $[0,\infty)$. (Alternately, for any $\epsilon > 0$, you can show that the choice of $N = 1/(4\epsilon^2)$ satisfies the definition of uniform convergence.) $\square$

Series of Functions and the M-Test

Many useful examples of sequences of functions come from series of functions, much like how series of real numbers provided many examples of sequences of real numbers. However, series of functions are even more important since we can discuss questions such as continuity of the series, integrability of the series, and so forth. Therefore, we take some time to discuss series of functions, and we prove an extremely useful test for analyzing when a series converges uniformly.

First, we should be precise about what a series of functions is. A series of functions takes the form

$$\sum_{i=k}^{\infty} a_i$$

where $k \in \mathbb{N}$ and a_i is a real function for each $i \in \mathbb{N}$ with $i \ge k$. For each $n \in \mathbb{N}$, the n^{th} partial sum of the series is the function f_n satisfying

$$f_n(x) = \sum_{i=k}^{n} a_i(x)$$

for all x in the domains of a_k through a_n. (Technically, the series is the sequence of partial sums, as is the case with series of real numbers.) We say that our series converges pointwise on D to a function $f : D \to \mathbb{R}$ if

$$f(x) = \lim_{n\to\infty} f_n(x) = \sum_{i=k}^{\infty} a_i(x)$$

for each $x \in D$. (In other words, our series converges pointwise iff $f_n \to f$ pointwise.) We say that our series converges uniformly on D if $f_n \to f$ uniformly on D.

Much like with series of real numbers, we would like a way to test whether a series of functions converges uniformly without requiring an exact formula for the n^{th} partial sum f_n. (Pointwise convergence can be established using the series tests from the previous chapter, because for each fixed $x \in D$, the series $\sum a_n(x)$ is just a series of real numbers.) To make such a test for uniform convergence, it makes sense to try replacing the functions $a_i(x)$ with values that do not depend on x. In particular, if we had a bound M_i on a_i for each i, i.e. $|a_i(x)| \le M_i$ for all i and all $x \in D$, then we would have

$$|f_n(x)| \le \sum_{i=k}^{n} |a_i(x)| \le \sum_{i=k}^{n} M_i$$

for all $x \in D$.

In particular, if $\sum M_i$ converges, then $\sum |a_i(x)|$ converges for each $x \in D$ by the Comparison Test. Going further with this reasoning, we obtain the following handy result due to Weierstrass:

Theorem 10.13 (M-Test). *Let $D \subseteq \mathbb{R}$ be given, and let $(a_n)_{n=1}^{\infty}$ be a sequence of real functions defined on D. Suppose that for each $n \in \mathbb{N}^*$, there is a constant $M_n \in \mathbb{R}$ such that*

$$|a_n(x)| \le M_n$$

for all $x \in D$. If the series $\sum M_n$ converges, then the series $\sum a_i$ converges absolutely and uniformly on D.

Strategy. Let's say f_n is the n^{th} partial sum of our series $\sum a_i$. The remarks before the statement of the M-Test show that $\sum a_i$ converges pointwise on D. Let's say that f is the sum of the series on D, i.e. the pointwise limit of $(f_n)_{n=1}^{\infty}$. To determine whether $f_n \to f$ uniformly, we want to analyze the distance between f_n and f. This leads us to consider $|f_n(x) - f(x)|$.

We find that for each $x \in D$,

$$|f(x) - f_n(x)| = \left| \sum_{i=1}^{\infty} a_i(x) - \sum_{i=1}^{n} a_i(x) \right| = \left| \sum_{i=n+1}^{\infty} a_i(x) \right|$$

(Sometimes we refer to the finite sum up to n as the "head" of the series up to n terms, whereas the infinite sum from $i = n + 1$ to ∞ is called the "tail" of the series.) We want to show that this tail can be made arbitrarily small by choosing a large enough value of n.

The Triangle Inequality for series (i.e. Theorem 9.72) shows that

$$\left|\sum_{i=n+1}^{\infty} a_i(x)\right| \leq \sum_{i=n+1}^{\infty} |a_i(x)| \leq \sum_{i=n+1}^{\infty} M_i$$

for all $x \in D$. Note that this bound only depends on n, not on x. Therefore, it remains to show that the last series in this inequality can be made arbitrarily small. However, we can write this last series as the tail of the series $\sum M_i$ and use the hypothesis that $\sum M_i$ converges.

Proof. Let D and $(a_n)_{n=1}^{\infty}$ be given as described. Choose a bound $M_n \in \mathbb{R}$ for each $n \in \mathbb{N}^*$ satisfying $|a_n(x)| \leq M_n$ for all $x \in D$. Also, for each $n \in \mathbb{N}^*$ and $x \in D$, we define $f_n : D \to \mathbb{R}$ and $s_n, s \in \mathbb{R}$

$$f_n(x) = \sum_{i=1}^{n} a_i(x) \qquad s_n = \sum_{i=1}^{n} M_i \qquad s = \sum_{i=1}^{\infty} M_i$$

Because $0 \leq |a_i(x)| \leq M_i$ for all $i \in \mathbb{N}^*$ and all $x \in D$, and $\sum M_i$ converges, for each $x \in D$ the series

$$\sum_{i=1}^{\infty} |a_i(x)|$$

converges absolutely by the Comparison Test. Therefore, we may define $f : D \to \mathbb{R}$ for all $x \in D$ by

$$f(x) = \sum_{i=1}^{\infty} a_i(x)$$

as absolute convergence implies convergence by Theorem 9.72.

This proves that $f_n \to f$ pointwise on D as $n \to \infty$. This also proves that our series $\sum a_i$ converges absolutely on D. Now, to prove uniform convergence, let $\epsilon > 0$ be given. We need to find $N \in \mathbb{N}^*$ such that $n > N$ implies $|f_n(x) - f(x)| < \epsilon$ for all $x \in D$. First, we note that because $\sum M_i$

converges, $s_n \to s$ as $n \to \infty$, so there is some $N \in \mathbb{N}^*$ such that $n > N$ implies

$$s - s_n = \sum_{i=1}^{\infty} M_i - \sum_{i=1}^{n} M_i = \sum_{i=n+1}^{\infty} M_i < \epsilon$$

(We do not need absolute values because each M_i must be nonnegative.) With this choice of N, we compute, for any $n \in \mathbb{N}^*$ with $n > N$ and any $x \in D$,

$$\begin{aligned} |f(x) - f_n(x)| &= \left| \sum_{i=1}^{\infty} a_i(x) - \sum_{i=1}^{n} a_i(x) \right| \\ &= \left| \sum_{i=n+1}^{\infty} a_i(x) \right| \leq \sum_{i=n+1}^{\infty} |a_i(x)| \\ &\leq \sum_{i=n+1}^{\infty} M_i < \epsilon \end{aligned}$$

as desired. ■

When analyzing a series of functions, usually some series test from the previous chapter is first used to determine where the series converges pointwise. After that, we can try the M-Test to determine whether the convergence is uniform. The following two examples are typical:

Example 10.14:

Let's consider the following series

$$f(x) = \sum_{n=0}^{\infty} (n+1)x^n$$

The Root Test or the Ratio Test[3] readily shows that this series converges pointwise when $|x| < 1$ and diverges when $|x| > 1$. Also, the series diverges when $|x| = 1$ because the terms do not approach 0.

At this point, we'd like to know if the series converges uniformly on $(-1, 1)$. To use the M-Test, for each $n \in \mathbb{N}$, we find a bound for $|(n+1)x^n|$

[3] Technically, we are using the stronger versions of the tests from Exercises 9.12.6 and 9.12.7.

on $(-1,1)$. (The fact that this series starts from $n=0$ instead of $n=1$ does not matter for the M-Test.) The best bound we can pick is

$$M_n = \sup_{x\in(-1,1)} |(n+1)x^n| = n+1$$

However, since $\sum M_n$ diverges (in particular, M_n does not approach 0 as $n\to\infty$), the M-Test does not apply on the interval $(-1,1)$.[4]

Essentially, this series fails to converge uniformly on $(-1,1)$ for the same reason that the sequence in Example 10.9 fails to converge uniformly on $[0,1]$: when $|x|$ is very close to 1, x^n converges to 0 very slowly as $n\to\infty$. This suggests that in order to obtain uniform convergence, we should restrict the domain to stay a fixed distance away from the points 1 and -1. Thus, for each $b\in(0,1)$, let's consider the domain $D_b=(-b,b)$. On D_b, we can choose a better bound for $|(n+1)x^n|$:

$$M_n = \sup_{x\in(-b,b)} |(n+1)x^n| = (n+1)b^n$$

Note that $\sum M_n = f(b)$, and f converges at b because $0<b<1$. Thus, by the M-Test, the series $f(x)$ converges uniformly on $(-b,b)$ for each $b\in(0,1)$. This is a very similar situation to Example 10.10, where we don't have uniform convergence on the entire interval of pointwise convergence, but we get uniform convergence on many subintervals. □

Example 10.15:

Let's consider the series

$$f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2\, 2^{3n}}$$

The Root Test readily shows that this series converges pointwise when $|x|<8$ and diverges when $|x|>8$. When $|x|=8$, we find that the series converges absolutely because it is dominated by $\sum 1/n^2$. Thus, this series converges pointwise on $[-8,8]$.

In fact, in contrast to the previous example, we also have uniform convergence on the whole interval of pointwise convergence. To see this, for each

[4] In fact, the series does not converge uniformly on $(0,1)$. When f_n is the n^{th} partial sum, you can check that the distance between f and f_n is at least $n+2$ with a similar argument.

$n \in \mathbb{N}^*$ and each $x \in [-8, 8]$, we have

$$\left|\frac{x^n}{n^2\ 2^{3n}}\right| \leq \frac{8^n}{n^2\ 8^n} = \frac{1}{n^2}$$

Therefore, we choose $M_n = 1/n^2$. Since $\sum M_n$ converges, our series for $f(x)$ converges uniformly on $[-8, 8]$. $\square$

These examples demonstrate how the M-Test is useful for telling us when certain series of functions converge uniformly. In fact, the proof of the M-Test also suggests a way to find a bound on the maximum distance between f_n and f, where f_n is the n^{th} partial sum of the series. However, we still don't yet know much information about the series sum f, such as whether f is continuous, differentiable, integrable, etc. We now turn our attention to answering these questions.

Uniform Convergence, Continuity, and Integration

If you go back to the examples from earlier where a pointwise limit of continuous functions fails to be continuous, then you will find that in those examples, the convergence is not uniform. (We studied one such example in detail in Example 10.9 above.) This is because unlike pointwise convergence, uniform convergence preserves continuity of functions. This is one of the primary reasons why uniform convergence is more useful than pointwise convergence in proofs:

Theorem 10.16. *Let $D \subseteq \mathbb{R}$ be an interval with more than one point (the interval may be open, half-open, or closed), and let $(f_n)_{n=1}^{\infty}$ be a sequence of real-valued functions defined on D. Also, let $f : D \to \mathbb{R}$ be given, and assume that $f_n \to f$ uniformly on D as $n \to \infty$. For any $a \in D$, if f_n is continuous at a for each $n \in \mathbb{N}^*$, then f is also continuous at a.*

In other words,

$$\begin{aligned} f(a) &= \lim_{x \to a} f(x) \\ &= \lim_{x \to a} \left(\lim_{n \to \infty} f_n(x) \right) = \lim_{n \to \infty} \left(\lim_{x \to a} f_n(x) \right) \\ &= \lim_{n \to \infty} f_n(a) \end{aligned}$$

Thus, as the middle row shows, we can interchange the limits involving n and involving x when convergence is uniform.

Strategy. We want to show that $f(x) \to f(a)$ as $x \to a$. By definition, this means that for all $\epsilon > 0$, we need to find some $\delta > 0$ such that[5]

$$\forall x \in D \ \ (|x - a| < \delta \to |f(x) - f(a)| < \epsilon)$$

Informally, we want $f(x)$ to be arbitrarily close to $f(a)$ provided that x is close enough to a.

We know that $f_n \to f$ uniformly, so $f(x)$ is close to $f_n(x)$ and $f(a)$ is close to $f_n(a)$ when n is large enough. Also, if x is close to a, then $f_n(x)$ is close to $f_n(a)$ because f_n is continuous at a. Informally, this means

$$f(x) \approx f_n(x) \approx f_n(a) \approx f(a)$$

More formally, we can estimate the distance $|f(x) - f(a)|$ by using the three distances $|f(x) - f_n(x)|$, $|f_n(x) - f_n(a)|$, and $|f_n(a) - f(a)|$. In fact, the Triangle Inequality yields

$$|f(x) - f(a)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)|$$

Since we want $|f(x) - f_n(x)| < \epsilon$, this suggests that we should try to make each distance on the right side less than $\epsilon/3$. (In fact, this kind of proof is fairly common in analysis dealing with uniformity, so this proof is sometimes called "an $\epsilon/3$ argument".)

First, let's look at the distances $|f(x)-f_n(x)|$ and $|f(a)-f_n(a)|$. Because $f_n \to f$, each of these distances can be made less than $\epsilon/3$ for appropriate choices of n. We want one choice of n that will work for all possible x AT THE SAME TIME. This is where uniform convergence becomes important. Because of uniform convergence, we can choose an N such that $n > N$ implies $|f(x) - f_n(x)| < \epsilon/3$ for every $x \in D$ (in particular, $|f(a) - f_n(a)| < \epsilon/3$).

Once we have chosen our value of N, we can control the last distance $|f_n(x) - f_n(a)|$. Because f_n is continuous at a, there is some $\delta > 0$ such that $|x - a| < \delta$ implies $|f_n(x) - f_n(a)| < \epsilon/3$. δ depends on n in general, so we may not be able to find a value of δ which works for all $n > N$. However, we only need $|f_n(x) - f_n(a)|$ to be less than $\epsilon/3$ for ONE value of $n > N$. For simplicity, we choose $n = N + 1$ and take its corresponding value of δ.

In summary, we choose N first using uniform convergence. After that, we choose $n = N + 1$, so n depends on N. Lastly, we pick δ based on n using the continuity of f_n. Putting this all together, each of our three distances is less than $\epsilon/3$.

[5]It is worth noting that this is the definition for two-sided continuity if a is not an endpoint of D, and it is the definition for one-sided continuity if a is an endpoint of D.

Proof. Let D, $(f_n)_{n=1}^{\infty}$, and f be given as described. Suppose that $f_n \to f$ uniformly on D, and let $a \in D$ be given such that f_n is continuous at a for each $n \in \mathbb{N}^*$. Let $\epsilon > 0$ be given; we need to find $\delta > 0$ such that

$$\forall x \in D \ \ (|x-a| < \delta \to |f(x) - f(a)| < \epsilon)$$

By the Triangle Inequality, for any $x \in D$, we have

$$|f(x) - f(a)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)|$$

Because $f_n \to f$ uniformly on D, there is $N \in \mathbb{N}^*$ such that for all $n \in \mathbb{N}^*$ with $n > N$ and all $x \in D$,

$$|f_n(x) - f(x)| < \frac{\epsilon}{3}$$

Let's choose $n = N + 1$. Now, f_n is continuous at a, so there is some $\delta > 0$ such that

$$\forall x \in D \ \left(|x-a| < \delta \to |f_n(x) - f_n(a)| < \frac{\epsilon}{3}\right)$$

Using the choices of n and δ and the bound on $|f(x) - f(a)|$ above, we get

$$|f(x) - f(a)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

for all $x \in D$ with $|x - a| < \delta$, as desired. ■

With Theorem 10.16, which says that uniform convergence preserves continuity, we can now address the question of continuity for series of functions. In Example 10.14, for instance, we showed that the series

$$f(x) = \sum_{n=0}^{\infty} (n+1)x^n$$

converges uniformly on $(-b, b)$ for every $b \in (0, 1)$. Since each partial sum of the series is a polynomial, each partial sum is continuous on $(-b, b)$ for $b \in (0, 1)$. It follows that f is continuous on $(-b, b)$ for each $b \in (0, 1)$ and hence f is continuous on $(-1, 1)$ (since every $x \in (-1, 1)$ belongs to an interval of the form $(-b, b)$: choose $b = (1 + |x|)/2$). Similarly, the series from Example 10.15,

$$f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2\ 2^{3n}}$$

converges uniformly on $[-8, 8]$ and is hence continuous on $[-8, 8]$.

Next, we address integrability. It turns out that a uniform limit of functions integrable on an interval $[a, b]$ is also integrable on $[a, b]$; the proof is left for Exercise 10.3.14. However, even more than that is true:

Theorem 10.17. *Let $a, b \in \mathbb{R}$ be given with $a < b$, and let $(f_n)_{n=1}^{\infty}$ be a sequence of real-valued functions defined on $[a, b]$. Also, let $f : [a, b] \to \mathbb{R}$ be given, and suppose that $f_n \to f$ uniformly on $[a, b]$ as $n \to \infty$. If f_n is integrable on $[a, b]$ for each $n \in \mathbb{N}^*$, then so is f (by Exercise 10.3.14).*

Moreover, let $c \in [a, b]$ be given, and define $g_n, g : [a, b] \to \mathbb{R}$ for each $n \in \mathbb{N}^*$ and $x \in [a, b]$ by

$$g_n(x) = \int_c^x f_n(t)\, dt \qquad g(x) = \int_c^x f(t)\, dt$$

Then $g_n \to g$ uniformly as $n \to \infty$ on $[a, b]$. In particular, for any $c, x \in [a, b]$,

$$\int_c^x f(t)\, dt = \int_c^x \left(\lim_{n\to\infty} f_n(t) \right) dt = \lim_{n\to\infty} \left(\int_c^x f_n(t)\, dt \right)$$

We say that the integral of the limit is the limit of the integrals. In other words, we can interchange integration and the limit involving n.

Strategy. We want to show that

$$\int_c^x f_n(t)\, dt \to \int_c^x f(t)\, dt \text{ uniformly as } n \to \infty$$

This means that for each $\epsilon > 0$, we want to show that when n is large enough, we have

$$\left| \int_c^x f(t)\, dt - \int_c^x f_n(t)\, dt \right| < \epsilon$$

for all $x \in [a, b]$. Some simple inequalities yield

$$\left| \int_c^x f(t)\, dt - \int_c^x f_n(t)\, dt \right| = \left| \int_c^x (f(t) - f_n(t))\, dt \right| \le \int_c^x |f(t) - f_n(t)|\, dt$$

(Technically, this is only valid when the integration limits are in the correct order, i.e. when $x \ge c$, but a similar calculation can be done when $x < c$ by switching the limits and introducing minus signs.)

We'd like this integral to be less than ϵ. Because $f_n \to f$, for any individual $t \in [c, x]$, we can make $|f(t) - f_n(t)|$ arbitrarily small by taking n large enough. In general, n might depend on t, but since $f_n \to f$ uniformly, we can choose an n which works for all t simultaneously! Thus, for each $\delta > 0$, we can find $N \in \mathbb{N}^*$ large enough so that $n > N$ implies $|f(t) - f_n(t)| < \delta$ for all $t \in [a, b]$. This yields

$$\int_c^x |f(t) - f_n(t)|\, dt \le \int_c^x \delta\, dt = \delta(x - c) \le \delta(b - a)$$

Since we want to show the integral is less than ϵ, we may choose δ to be $\epsilon/(b - a)$. (Actually, since the inequalities in our work above may not be strict, we choose δ to be slightly smaller than $\epsilon/(b - a)$.)

Proof. Let $a, b, (f_n)_{n=1}^\infty$, and f be given as described. Also, assume f_n is integrable on $[a, b]$ for each $n \in \mathbb{N}^*$, and let $c \in [a, b]$ be given. We know that f is integrable on $[a, b]$ by Exercise 10.3.14. Let $\epsilon > 0$ be given; we aim to find $N \in \mathbb{N}^*$ so that for all $n \in \mathbb{N}^*$ with $n > N$ and all $x \in [a, b]$, we have

$$|g(x) - g_n(x)| = \left| \int_c^x f(t)\, dt - \int_c^x f_n(t)\, dt \right| < \epsilon$$

First, because $f_n \to f$ uniformly on $[a, b]$ as $n \to \infty$, we may choose $N \in \mathbb{N}^*$ such that for all $n \in \mathbb{N}^*$ with $n > N$ and all $t \in [a, b]$, we have

$$|f(t) - f_n(t)| < \frac{\epsilon}{2(b - a)}$$

Next, let $x \in [a, b]$ be given. When $x \ge c$, we have

$$\begin{aligned} |g(x) - g_n(x)| &= \left| \int_c^x (f(t) - f_n(t))\, dt \right| \le \int_c^x |f(t) - f_n(t)|\, dt \\ &\le \int_c^x \frac{\epsilon}{2(b - a)}\, dt = \frac{\epsilon}{2(b - a)}(x - c) \\ &\le \frac{\epsilon}{2(b - a)}(b - a) = \frac{\epsilon}{2} < \epsilon \end{aligned}$$

as desired. Similarly, when $x \le c$, we have

$$\begin{aligned} |g(x) - g_n(x)| &= \left| -\int_x^c (f(t) - f_n(t))\, dt \right| \le \int_x^c |f(t) - f_n(t)|\, dt \\ &\le \int_x^c \frac{\epsilon}{2(b - a)}\, dt = \frac{\epsilon}{2(b - a)}(c - x) \\ &\le \frac{\epsilon}{2(b - a)}(b - a) = \frac{\epsilon}{2} < \epsilon \end{aligned}$$

as needed. ■

As with Theorem 10.16 about continuity and uniform convergence, Theorem 10.17 is often useful with series of functions. More precisely, let's say

$$f(x) = \sum_{i=1}^{\infty} a_i(x)$$

is a series of functions which are integrable on $[a, b]$. For each $n \in \mathbb{N}^*$, let f_n be the n^{th} partial sum of this series. If this series converges uniformly on an interval D containing a and b, i.e. $f_n \to f$ uniformly on D as $n \to \infty$, then we find

$$\int_a^b f(t)\,dt = \lim_{n\to\infty} \left(\int_a^b f_n(t)\,dt \right)$$

or in other words,

$$\int_a^b \left(\sum_{i=1}^{\infty} a_i(t) \right) dt = \lim_{n\to\infty} \sum_{i=1}^{n} \left(\int_a^b a_i(t)\,dt \right) = \sum_{i=1}^{\infty} \left(\int_a^b a_i(t)\,dt \right)$$

This is often summarized as saying that uniform convergence lets us interchange integrals and summations. Alternately, we say that a uniformly convergent series may be integrated *term by term*. We present a couple examples to demonstrate this idea.

Example 10.18:

As we saw in Example 10.14, the series

$$f(x) = \sum_{n=0}^{\infty} (n+1)x^n$$

converges uniformly on $(-b, b)$ for every $b \in (0, 1)$. Hence, since each $x \in (-1, 1)$ belongs to some interval of the form $(-b, b)$ for $b \in (0, 1)$, we may integrate f term by term and obtain

$$\int_0^x f(t)\,dt = \sum_{n=0}^{\infty} \int_0^x (n+1)t^n\,dt = \sum_{n=0}^{\infty} x^{n+1}$$

At this point, we note that $\sum x^{n+1}$ is the geometric series with ratio x starting with the term x^1. Thus, the sum of that series is

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

(This geometric series can also be seen to equal $1/(1-x)-1$, since it is obtained from the usual geometric series that starts with x^0 by subtracting x^0.) Since f is continuous at x, we can differentiate both sides and use the Fundamental Theorem of Calculus to obtain

$$f(x) = \frac{1}{(1-x)^2}$$

for all $x \in (-1,1)$.

This shows that we can find the exact values of some series by using term-by-term integration! $\square$

Example 10.19:

For a similar example, consider

$$f(x) = \sum_{n=0}^{\infty} nx^n$$

A simple use of the Root Test, together with the M-Test, shows that this series converges uniformly on $(-b,b)$ for every $b \in (0,1)$. We can find the exact value of this series in a few different ways. One way is to use the series from the previous example and subtract $\sum x^n$, which yields

$$\sum_{n=0}^{\infty} nx^n = \sum_{n=0}^{\infty}(n+1)x^n - \sum_{n=0}^{\infty} x^n = \frac{1}{(1-x)^2} - \frac{1}{1-x} = \frac{x}{(1-x)^2}$$

Another way to compute $f(x)$ is to immediately integrate term by term from 0 to x, find the value of the integral, and then differentiate (which is what the previous example did). However, the integral is

$$\int_0^x f(t)\,dt = \sum_{n=0}^{\infty} \frac{n}{n+1}x^{n+1}$$

which is not very convenient for finding exact sums because it is not geometric. We'd prefer to have an integrated series which is easier to evaluate.

To accomplish this, we note something clever: we can write $f(x)$ as

$$f(x) = x + 2x^2 + 3x^3 + \cdots = x\left(1 + 2x + 3x^2 + \cdots\right) = x\sum_{n=1}^{\infty} nx^{n-1}$$

(note that we may factor out x because x is constant with respect to the summation index n), and let's consider the series

$$g(x) = \sum_{n=1}^{\infty} nx^{n-1}$$

This series converges uniformly on the same intervals that f does, and it's easier to integrate g term by term than f, because the integral of nt^{n-1} from 0 to x is x^n; there's no expression like $n/(n+1)$ getting in the way. Thus, we find

$$\int_0^x g(t)\,dt = \sum_{n=1}^{\infty} x^n = \left(\sum_{n=0}^{\infty} x^n\right) - 1 = \frac{1}{1-x} - 1$$

and differentiating both sides yields

$$g(x) = \frac{1}{(1-x)^2}$$

Finally,

$$f(x) = xg(x) = \frac{x}{(1-x)^2}$$

The lesson to take from this example is that if the form of a series does not make term-by-term integration convenient, then we can sometimes improve the form of the integrated series by factoring out powers of x or by adding and subtracting other known series. By using these tactics, it is possible to obtain exact sums for any series of the form $\sum p(n)x^n$ where p is a polynomial. A few other examples of this are in the exercises. $\square$

Uniform Convergence and Differentiation

Based on our most recent theorems, you might guess that uniform convergence preserves differentiability. Unfortunately, however, this is not generally true. When $f_n \to f$ uniformly on an interval D, and each f_n is differentiable on D, it doesn't necessarily follow that f is differentiable on D. Furthermore, when f is differentiable on D, we might not have $f_n' \to f'$ pointwise on D. In other words, a derivative of a limit is not necessarily a limit of derivatives. Here are some examples:

Example 10.20:

For a rather simple example, consider $f_n : \mathbb{R} \to \mathbb{R}$ defined for each $n \in \mathbb{N}^*$ and each $x \in \mathbb{R}$ by

$$f_n(x) = \frac{\sin(nx)}{n}$$

Since $|f_n(x)| \leq 1/n$, it is easy to show that $f_n \to 0$ uniformly on $\mathbb{R}$. However, $f_n'(x) = \cos(nx)$, so $f_n'(0) = 1$ for all $n \in \mathbb{N}^*$.

For a similar example with series of functions, consider

$$f(x) = \sum_{n=1}^{\infty} \frac{\sin(nx)}{n^2}$$

We can use the M-Test with $M_n = 1/n^2$ for each $n \in \mathbb{N}^*$ to prove that this series converges uniformly to a continuous function on $\mathbb{R}$. In other words, if f_n denotes the n^{th} partial sum, then $f_n \to f$ uniformly on $\mathbb{R}$. However, if we try to differentiate this series term by term, then we obtain the series

$$g(x) = \sum_{n=1}^{\infty} \frac{\cos(nx)}{n}$$

and $g(0)$ diverges (it's the harmonic series). In other words, the sequence of derivatives $(f_n')_{n=1}^{\infty}$ does not approach a finite limit at $x = 0$.[6] $\square$

Example 10.21:

In this example, let $f(x) = |x|$ for all $x \in \mathbb{R}$. We construct a sequence of differentiable functions $(f_n)_{n=1}^{\infty}$ on $\mathbb{R}$ which converge uniformly to f on $\mathbb{R}$. This shows that a uniform limit of differentiable functions might not be differentiable, since the absolute-value function is not differentiable at 0.

The main idea behind the construction is similar to the idea behind the construction from Example 10.5 with the Heaviside function. We want f_n to behave a lot like f except for points near 0; near 0, we use a curve which is smoother than the "sharp corner" of f. To do this, for each $n \in \mathbb{N}^*$, we make $f_n(x) = |x|$ for all x outside of an "exception interval" $[-1/n, 1/n]$, and on $[-1/n, 1/n]$, we connect the graph with a parabola whose center is on

[6]It is worth noting that this argument does not prove that $f'(0)$ doesn't exist. The argument only proves that $f'(0)$, if it exists, cannot be the limit of $f_n'(0)$. It is actually quite difficult (beyond the scope of this book) to analyze whether $f'(0)$ exists for this example.

the y-axis. This means that our definition takes the following form for some $a_n, c_n \in \mathbb{R}$:

$$f_n(x) = \begin{cases} |x| & \text{if } x < -1/n \text{ or } x > 1/n \\ a_n x^2 + c_n & \text{if } x \in [-1, 1] \end{cases}$$

(We will find a_n and c_n soon.) For example, the graph of f_1 is provided in Figure 10.5.

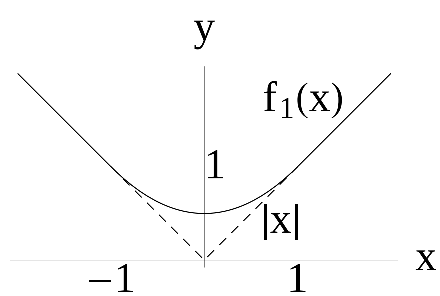

Figure 10.5: A graph of f_1 with exception interval $[-1, 1]$ (and the graph of $|x|$ with dashed lines)

In order for f_n to be continuously differentiable on $\mathbb{R}$, we need to pick a_n and c_n so that f_n is continuous and differentiable at $x = \pm 1/n$. For $x = 1/n$, this requires

$$\lim_{x \to 1/n^-} f_n(x) = \lim_{x \to 1/n^+} f_n(x) \qquad \text{and} \qquad \lim_{x \to 1/n^-} f_n'(x) = \lim_{x \to 1/n^+} f_n'(x)$$

which leads to the equations

$$\frac{a_n}{n^2} + c_n = \frac{1}{n} \qquad \text{and} \qquad \frac{2a_n}{n} = 1$$

These equations have the unique solution $a_n = n/2$ and $c_n = 1/(2n)$. With these values of a_n and c_n, you can check that f_n is also continuously differentiable at $-1/n$.

Now that we have constructed the sequence of f_n functions, we should prove that $f_n \to f$ uniformly on $\mathbb{R}$ as $n \to \infty$. To do this, we need to study the maximum distance between f_n and f:

$$\sup_{x \in \mathbb{R}} ||x| - f_n(x)|$$

Certainly $||x| - f_n(x)| = 0$ when x is not in $[-1/n, 1/n]$. Also, when $x \in [-1/n, 1/n]$, it is clear that $f_n(x)$ and $|x|$ are both in $[0, 1/n]$, so $||x| - f_n(x)| \leq 1/n$. Thus, the maximum distance between f_n and f is at most $1/n$, and this goes to 0 independently of x. Thus, $f_n \to f$ uniformly.

It is worth noting that the fact that f_n is a parabola on $[-1/n, 1/n]$ hardly gets used in the proof! All that is used from the definition on $[-1/n, 1/n]$ is the fact that $f_n(x) \in [0, 1/n]$ for $x \in [-1/n, 1/n]$. Thus, we can make similar examples to this one by defining f_n however we like on $[-1/n, 1/n]$, provided that we require f_n to be continuously differentiable on $\mathbb{R}$ and $f_n(x) \in [0, 1/n]$ for all $x \in [-1/n, 1/n]$. $\square$

These examples demonstrate that differentiation is more difficult to analyze than continuity and integration for uniform convergence. The last example we just showed makes this particularly prominent, because we can make f_n oscillate several times between 0 and $1/n$ in the interval $[-1/n, 1/n]$ and still obtain uniform convergence! The issue is that when a function oscillates in a small area (e.g. x goes from $-1/n$ to $1/n$ and y goes from 0 to $1/n$), the value of the function and the integral of the function don't change very much, but the slope of the function can change very quickly.

Certain special types of series of functions do behave nicely with respect to differentiation, however. In the next section, we will introduce power series, and among other results, we will prove that it is safe to differentiate a power series term by term. This, along with other nice properties, make power series particularly easy to use, as we will see.

10.3 Exercises

1. Suppose that $(f_n)_{n=1}^{\infty}$ and $(g_n)_{n=1}^{\infty}$ are both sequences of real-valued functions defined on a set $D \subseteq \mathbb{R}$. Suppose that $f_n \to f$ and $g_n \to g$ uniformly on D as $n \to \infty$. Prove that $f_n + g_n \to f + g$ uniformly on D as $n \to \infty$. (Hint: Imitate the proof of the addition limit law.)[7]

2. Suppose that $(f_n)_{n=1}^{\infty}$ is a sequence of real-valued functions defined on a set $D \subseteq \mathbb{R}$, and suppose $f_n \to f$ pointwise on D as $n \to \infty$. Prove that if there is some $M \in [0, \infty)$ such that
$$\forall x \in D \; \forall n \in \mathbb{N}^* \; |f_n(x)| \leq M$$

[7] Almost all of the limit laws for sequences apply to uniform convergence as well. The proofs for uniform convergence are quite similar to the original proofs of the limit laws.

then $|f(x)| \le M$ for all $x \in D$ as well.

3. Let $D \subseteq \mathbb{R}$ be given, let $(f_n)_{n=1}^{\infty}$ be a sequence of real-valued functions defined on D, and let $f : D \to \mathbb{R}$ be given.
   (a) Suppose that $D = D_1 \cup D_2$, where $D_1, D_2 \subseteq \mathbb{R}$. Prove that $f_n \to f$ uniformly on D as $n \to \infty$ if and only if $f_n \to f$ uniformly on both D_1 and D_2 as $n \to \infty$.
   (b) Show that the analogue of part (a) for infinite unions is false by creating functions $f, (f_n)_{n=1}^{\infty}$ and subsets $D, (D_n)_{n=1}^{\infty}$ of $\mathbb{R}$ such that
$$D = \bigcup_{i=1}^{\infty} D_i$$
   and for all $i \in \mathbb{N}^*$,
$$f_n \to f \text{ uniformly on } D_i \text{ as } n \to \infty$$
   but f_n does not converge to f uniformly on D as $n \to \infty$.

4. Let D be a given nonempty subset of $\mathbb{R}$, let $f : D \to \mathbb{R}$ be given, and let $(f_n)_{n=1}^{\infty}$ be a sequence of real-valued functions defined on D. Prove that $f_n \to f$ uniformly on D as $n \to \infty$ iff for all $\epsilon > 0$, there is some $N \in \mathbb{N}^*$ such that for all $n \in \mathbb{N}$, $n > N$ implies
$$\sup\{|f_n(x) - f(x)| \mid x \in D\} < \epsilon$$
   (In particular, the supremum on the left side exists whenever $n > N$.) This supremum is called the L_∞ *distance* from f to f_n on D.

5. For each $n \in \mathbb{N}^*$ and each $x \in [0, 1]$, define $f_n : [0, 1] \to \mathbb{R}$ by
$$f_n(x) = nx(1 - x^2)^n$$
   Prove that $f_n \to 0$ pointwise on $[0, 1]$ as $n \to \infty$, but that
$$\lim_{n\to\infty} \int_0^1 f_n(t)\, dt \neq 0$$
   It follows by Theorem 10.17 that f_n does not converge uniformly to 0.

6. Let $r : \mathbb{R} \to \mathbb{R}$ be the ruler function (as defined in Chapter 3). For each $n \in \mathbb{N}^*$ and each $x \in \mathbb{R}$, define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} \dfrac{1}{q} & \text{if } x \in \mathbb{Q} \text{ and } x = p/q \text{ in lowest terms with } q \le n \\ 0 & \text{otherwise} \end{cases}$$

Prove that $f_n \to r$ uniformly on $\mathbb{R}$ as $n \to \infty$. Use this to prove that the ruler function is continuous at all irrational points and has integral 0 on every interval.

7. Let $(r_i)_{i=1}^{\infty}$ be an enumeration of the rationals, and for each $n \in \mathbb{N}^*$ and each $x \in \mathbb{R}$, define $g_n : \mathbb{R} \to \mathbb{R}$ by

$$g_n(x) = \sum_{i=1}^{n} \left(\frac{1}{2}\right)^i H(x - r_i)$$

where H is the Heaviside function. Define $s : \mathbb{R} \to \mathbb{R}$ for all $x \in \mathbb{R}$ by

$$s(x) = \sup\{g_n(x) \mid n \in \mathbb{N}^*\}$$

Thus, s is the staircase function from Chapter 3.

Prove that $g_n \to s$ uniformly on $\mathbb{R}$ as $n \to \infty$, and therefore $s(x)$ is the uniform sum of the series

$$\sum_{i=1}^{\infty} \frac{1}{2^i} H(x - r_i)$$

Use this to prove that s is continuous at all irrational points.

8. Let $g : \mathbb{R} \to [0, \infty)$ be given, and for each $n \in \mathbb{N}^*$ and each $x \in \mathbb{R}$, define $f_n : \mathbb{R} \to [0, \infty)$ by

$$f_n(x) = \sqrt{\frac{1}{n} + g(x)}$$

Prove that $f_n(x) \to \sqrt{g(x)}$ uniformly on $\mathbb{R}$ as $n \to \infty$. (Note that when $g(x) = x^2$, this provides another way to create a sequence of continuous differentiable functions which converges uniformly to $|x|$.)

9. For each $n \in \mathbb{N}^*$ and each $x \in \mathbb{R}$, define $f_n : \mathbb{R} \to \mathbb{R}$ by
$$f_n(x) = \frac{n^2 x}{1 + n^3 x^2}$$
   (a) Prove that $f_n \to 0$ pointwise on $\mathbb{R}$ as $n \to \infty$.
   (b) Prove that f_n does not converge to 0 uniformly on $(-b, b)$ for any $b \in (0, \infty)$ as $n \to \infty$.
   (c) Prove that $f_n \to 0$ uniformly on (a, ∞) for any $a \in (0, \infty)$ as $n \to \infty$.

10. Prove that a uniform limit of bounded functions is bounded. More specifically, suppose $D \subseteq \mathbb{R}$, $f : D \to \mathbb{R}$, and a sequence $(f_n)_{n=1}^{\infty}$ of real-valued functions defined on D are given. Suppose that f_n is bounded on D for each $n \in \mathbb{N}^*$, and suppose $f_n \to f$ uniformly on D as $n \to \infty$. Prove that f is bounded on D.

11. Suppose that $D \subseteq \mathbb{R}$ and a sequence of real functions $(a_i)_{i=0}^{\infty}$ defined on D is given such that
$$\sum_{i=0}^{\infty} a_i(x)$$
converges uniformly on D. Prove that $a_i \to 0$ uniformly on D as $i \to \infty$. (This is essentially a uniform version of the Divergence Test.)

12. A *two-dimensional sequence* is a sequence indexed by a subset of $\mathbb{N} \times \mathbb{N}$. For instance, we can make the two-dimensional sequence $(a_{i,j})_{(i,j) \in \mathbb{N}^* \times \mathbb{N}^*}$ by defining
$$a_{i,j} = \frac{i+1}{ij}$$
for all $i, j \in \mathbb{N}^*$. With two-dimensional sequences, we can take limits involving each variable separately, treating the other variable as a constant. For instance, with our definition of $a_{i,j}$ above, we have
$$\lim_{i \to \infty} a_{i,j} = \frac{1}{j} \qquad \text{and} \qquad \lim_{j \to \infty} a_{i,j} = 0$$
(the first limit treats j as a constant, and the second limit treats i as a constant). From this, we see that
$$\lim_{i \to \infty} \lim_{j \to \infty} a_{i,j} = \lim_{i \to \infty} 0 = 0$$
$$\text{and} \quad \lim_{j \to \infty} \lim_{i \to \infty} a_{i,j} = \lim_{j \to \infty} \frac{1}{j} = 0$$

Create a two-dimensional sequence $(a_{i,j})_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*}$ where

$$\lim_{i\to\infty}\lim_{j\to\infty} a_{i,j} \neq \lim_{j\to\infty}\lim_{i\to\infty} a_{i,j}$$

This shows that we cannot always interchange limits with multiple dimensions. (Hint: One way to do this is to define the sequence by cases based on whether i or j is bigger.)

13. Prove that the Riemann zeta function $\zeta : (1,\infty) \to \mathbb{R}$, defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

for all $s > 1$, is continuous on $(1,\infty)$. (Hint: Consider intervals of the form $[a,\infty)$ for $a > 1$.)

14. Prove the following theorem:

Theorem 10.22. *Let $a, b \in \mathbb{R}$ be given with $a < b$, and let $(f_n)_{n=1}^{\infty}$ be a sequence of real-valued functions defined on $[a,b]$. Also, let $f : [a,b] \to \mathbb{R}$ be given, and suppose that $f_n \to f$ uniformly on $[a,b]$ as $n \to \infty$. If f_n is integrable on $[a,b]$ for each $n \in \mathbb{N}^*$, then so is f.*

(Hint: Use Theorem 5.22, so that for each $\epsilon > 0$, you want to find step functions $s, t : [a,b] \to \mathbb{R}$ such that $s(x) \le f(x) \le t(x)$ for all $x \in [a,b]$ satisfying

$$\int_a^b (t(x) - s(x))\,dx < \epsilon$$

Now, for each $\delta > 0$, we can choose n large enough so that $|f_n(x) - f(x)| < \delta$ for all $x \in [a,b]$. Use this to show that upper and lower step functions for f_n can be changed slightly to produce upper and lower step functions for f. Based on this, what do you pick for δ in terms of ϵ, a, and b?)

15. Use the ideas from Examples 10.18 and 10.19 to find a closed-form value of the following series for each $x \in (-1,1)$:

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n+1}$

(c) $\displaystyle\sum_{n=0}^{\infty} (n+2)(n+1)x^n$

(d) $\displaystyle\sum_{n=0}^{\infty} n^2 x^n$

16. When $(f_n)_{n=1}^{\infty}$ and f are all real functions defined and integrable on an interval $[a,b]$, we say that $f_n \to f$ *in the* L_1 *sense on* $[a,b]$ *as* $n \to \infty$ if

$$\lim_{n\to\infty} \left(\int_a^b |f_n(x) - f(x)|\, dx \right) = 0$$

Thus, Theorem 10.17 shows that when $f_n \to f$ uniformly on $[a,b]$ as $n \to \infty$, then $f_n \to f$ also in the L_1 sense.

Construct a counterexample to show that $f_n \to f$ in the L_1 sense does not imply that $f_n \to f$ uniformly. Thus, the converse of Theorem 10.17 fails.

REMARK: It turns out if $f_n \to f$ in the L_1 sense, then it doesn't even follow that $f_n \to f$ pointwise!

17. We have seen that uniform convergence does not necessarily preserve differentiability. However, there are cases where uniform convergence behaves well with derivatives. One common case is handled by the following theorem:

Theorem 10.23. *Let $a, b \in \mathbb{R}$ be given with $a < b$, and let $(f_n)_{n=1}^{\infty}$ be a sequence of real functions defined on $[a,b]$. Also, let $f, g : [a,b] \to \mathbb{R}$ be given. Assume that for each $n \in \mathbb{N}^*$, f_n is continuously differentiable on $[a,b]$ (i.e. f_n' is continuous). Also assume that $f_n \to f$ pointwise and $f_n' \to g$ uniformly on $[a,b]$ as $n \to \infty$. Then $f_n \to f$ uniformly on $[a,b]$, and $f'(x) = g(x)$ for all $x \in [a,b]$.*

Prove this theorem.

(Hint: Integrate each f_n' from a to x, and use the Fundamental Theorem of Calculus together with Theorem 10.17.)

10.4 Power Series

In this section, we look at a special type of series of functions: power series. Power series behave a lot like polynomials that have "infinite degree". As we will shortly show, power series also have many nice properties in common with polynomials, such as being easy to differentiate and integrate.

To get started, here is a definition:

Definition 10.24. Let $a \in \mathbb{R}$ be given, and let $(c_n)_{n=0}^{\infty}$ be an infinite sequence of real numbers. The *power series centered at a with coefficients* (c_n) is the infinite series of functions

$$\sum_{n=0}^{\infty} c_n(x-a)^n$$

(We use the convention that $(x-a)^0 = 1$ even when $x = a$, so the value of the series at $x = a$ is c_0.) For each $n \in \mathbb{N}$, the expression $c_n(x-a)^n$ is the n^{th} *term* of the power series, and c_n is the n^{th} *coefficient*.

A power series always converges at its center: the value of $\sum c_n(x-a)^n$ at $x = a$ is c_0. To find out which other values of x yield convergence, the Root Test and Ratio Test are frequently useful. Here are some examples:

Example 10.25:

Consider the power series

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Let's say that for each $n \in \mathbb{N}$, $a_n(x)$ is the n^{th} term $x^n/(n!)$. Whenever $x \in \mathbb{R} - \{0\}$, the Ratio Test on $(|a_n(x)|)$ yields

$$\left|\frac{a_{n+1}(x)}{a_n(x)}\right| = \frac{|x|^{n+1}}{(n+1)!} \cdot \frac{n!}{|x|^n} = \frac{|x|}{n+1} \to 0$$

as $n \to \infty$. Therefore, our series converges absolutely for all $x \in \mathbb{R}$. $\square$

Example 10.26:

Consider the series

$$\sum_{n=0}^{\infty} n!\, x^n$$

This series has value 1 when $x = 0$. However, for all $x \neq 0$, the Ratio Test on the absolute values of the terms yields

$$\frac{(n+1)!\, |x|^{n+1}}{n!\, |x|^n} = (n+1)|x| \to \infty$$

as $n \to \infty$. Therefore, our series diverges for all $x \neq 0$. This is an example of a power series which converges at only one point. □

Example 10.27:

Let's consider the series

$$\sum_{n=1}^{\infty} \frac{(x-3)^n}{n}$$

This series has center 3. At $x = 3$, the value is 0 (i.e. the 0^{th} coefficient is 0). For all other values of x, the Root Test on the absolute values of the terms yields

$$\frac{|x-3|}{n^{1/n}} \to |x-3|$$

as $n \to \infty$. Thus, our series converges absolutely when $|x - 3| < 1$ and diverges when $|x - 3| > 1$.

In other words, we know we have absolute convergence on $(2, 4)$ (the numbers which are less than 1 unit away from 3) and divergence outside of $[2, 4]$. It remains to check the endpoints of this interval, $x = 2$ and $x = 4$. When $x = 4$, so that $x-3 = 1$, our series becomes $\sum 1/n$, which is divergent. When $x = 2$, so that $x - 3 = -1$, our series becomes $\sum(-1)^n/n$, which converges by the Alternating Series Test.

In summary, our series converges on $[2, 4)$ (and 2 is the only point with conditional convergence) and diverges elsewhere. We see that this "convergence set" is an interval centered at 3 with radius 1. □

Interval of Convergence

In our examples of power series, we saw that the set of values where the series converges is an interval centered at the same center of the series. In our first example, the convergence set is $\mathbb{R}$, which can be considered an interval of infinite radius. In our second example, the convergence set is $\{0\}$, which can be considered an interval of radius 0. Our last example featured an interval of radius 1 centered at 3, so we had convergence at points close to 3 and divergence at points far away from 3.

In general, a power series is more likely to converge at points which are close to the center. To see why this is plausible, let's consider a power series of the form

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

Let's say we know this series converges at $y \in \mathbb{R}$, and suppose $x \in \mathbb{R}$ is given which satisfies $|x-a| \le |y-a|$. (In other words, x is at least as close to a as y is.) Then, for each $n \in \mathbb{N}$, we have $|c_n(x-a)^n| \le |c_n(y-a)^n|$. It follows that if f converges absolutely at y, then it also converges absolutely at x by the Comparison Test. Thus, if f converges absolutely at y, then it converges absolutely at all points closer to the center than y.

It turns out that we can prove a similar statement even when the convergence at y is conditional, and we can also make a statement concerning uniform convergence:

Lemma 10.28. *Let $a \in \mathbb{R}$ be given, let $(c_n)_{n=0}^{\infty}$ be a sequence of real numbers, and let f be the power series defined by*

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

for all $x \in \mathbb{R}$ where the series converges. Suppose that $y \in \mathbb{R} - \{a\}$ is given such that f converges at y. Let $s = |y-a|$. Then f converges absolutely and uniformly on any interval of the form $(a-r, a+r)$ where $0 < r < s$. It follows that f converges absolutely on $(a-s, a+s)$.

Strategy. Let's say r is given with $0 < r < s$. As the remarks before this lemma show, when $x \in (a-r, a+r)$, we have

$$|c_n(x-a)^n| \le |c_n| r^n \le |c_n| s^n = |c_n(y-a)^n|$$

This suggests choosing $M_n = |c_n| r^n$ and trying to apply the M-Test. However, this only works if $\sum |c_n(y-a)^n|$ converges, i.e. if f converges absolutely at y. When f converges conditionally at y, we need to try something different.

Since we know that $|c_n(x-a)^n|$ is smaller than $|c_n(y-a)^n|$, we ask: HOW much smaller is it? One way to address this question is to consider proportions (since proportions are important in the Limit Comparison Test and the Ratio Test). We see that

$$\frac{|c_n(x-a)^n|}{|c_n(y-a)^n|} = \left(\frac{|x-a|}{|y-a|}\right)^n < \left(\frac{r}{s}\right)^n$$

so

$$|c_n(x-a)^n| \leq |c_n(y-a)^n| \left(\frac{r}{s}\right)^n$$

Thus, since $r < s$, not only is $|c_n(x-a)^n|$ smaller than $|c_n(y-a)^n|$, it is exponentially smaller. We'd like to focus on this exponential factor by creating a comparison between $\sum |c_n(x-a)^n|$ and $\sum (r/s)^n$.

In order to do this, however, we need to know how large $|c_n(y-a)^n|$ is. Here is where we use the fact that $f(y)$ converges. Since $\sum c_n(y-a)^n$ converges, the terms of that series approach 0, so $|c_n(y-a)^n|$ must be close to 0 for n large. This means that when n is large enough, we have

$$|c_n(x-a)^n| = |c_n(y-a)^n| \left(\frac{r}{s}\right)^n \leq \left(\frac{r}{s}\right)^n$$

and we can use the M-Test with $M_n = (r/s)^n$.

Proof. Let $a, (c_n), y$ be given as described. Let $s = |y-a|$, and let $r \in (0, s)$ be given. We will prove that f converges uniformly on $(a-r, a+r)$ by the M-Test.

First, because $f(y)$ converges, the series

$$\sum_{n=0}^{\infty} c_n(y-a)^n$$

converges. Therefore, its terms approach 0 as $n \to \infty$. In particular, there is some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n > N$, we have $|c_n(y-a)^n| < 1$. It follows that for all $n \in \mathbb{N}$ with $n > N$, and for all $x \in (a-r, a+r)$, we have

$$\begin{aligned} |c_n(x-a)^n| &= |c_n(y-a)^n| \cdot \frac{|c_n(x-a)^n|}{|c_n(y-a)^n|} \\ &= |c_n(y-a)^n| \left(\frac{|x-a|}{|y-a|}\right)^n \\ &\le \left(\frac{r}{s}\right)^n \end{aligned}$$

Define the sequence $(M_n)_{n=0}^\infty$ as follows for each $n \in \mathbb{N}$:

$$M_n = \begin{cases} \left(\frac{r}{s}\right)^n & \text{if } n > N \\ |c_n r^n| & \text{if } n \le N \end{cases}$$

Therefore, $|c_n(x-a)^n| \le M_n$ for all $n \in \mathbb{N}$ and all $x \in (a-r, a+r)$. Because $0 < r < s$, $\sum M_n$ converges. Thus, by the M-Test, $\sum c_n(x-a)^n$ converges absolutely and uniformly on $(a-r, a+r)$, as desired. ■

This lemma is the key result used to prove the following statement, which formally establishes that power series converge on an interval:

Theorem 10.29 (Power Series Interval of Convergence). *Let $a \in \mathbb{R}$ and a sequence $(c_n)_{n=0}^\infty$ of real numbers be given. Let f be the power series*

$$f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n$$

Then exactly one of these three cases holds:

1. *The series $f(x)$ only converges for $x = a$. In this case, we set $R = 0$.*

2. *The series $f(x)$ converges absolutely for all $x \in \mathbb{R}$. In this case, we set $R = \infty$.*

3. *There exists a constant $R \in (0, \infty)$ such that $f(x)$ converges absolutely when $|x-a| < R$ and diverges when $|x-a| > R$. (When $|x-a| = R$, the series may converge absolutely, converge conditionally, or diverge.)*

In all cases, our series converges on an interval centered at a of radius R called the interval of convergence *of f. (R is called the* radius of convergence *of f.) Furthermore, in cases 2 and 3, f converges uniformly on every interval of the form $(a-r, a+r)$ with $0 < r < R$.*

Strategy. Lemma 10.28 shows that if $f(y)$ converges, then f converges absolutely on $(a-s, a+s)$ where $s = |y-a|$. Another way to phrase this is that if f converges at a point which is distance s from the center, then it also converges at points which are less than distance s away from the center as well. Based on this, we are led to consider: for which values of s can we find a point where the series converges?

More formally, we want to consider the set

$$S = \{|y-a| \mid y \in \mathbb{R}, f(y) \text{ converges}\}$$

Our earlier remarks show that for every $s \in S$, the interval $[0, s]$ is contained in S. Thus, since S only contains nonnegative numbers, S is an interval with 0 as its lowest endpoint.

This raises the question: does S have a top endpoint? More generally, does S have an upper bound? If S doesn't have an upper bound, then the work in the last paragraph shows that $S = [0, \infty)$, which corresponds to case 2. Otherwise, S has an upper bound, and hence it has a least upper bound. Intuitively, it makes sense to choose $R = \sup S$, since the radius of convergence should correspond to the "farthest you can travel" from the center before you reach points where f diverges. From here, everything follows by the properties of supremum and Lemma 10.28.

Proof. Let $a, (c_n), f$ be given as described, and define

$$S = \{|y-a| \mid y \in \mathbb{R}, f(y) \text{ converges}\}$$

Note that S only contains nonnegative numbers. Also, since any power series converges at its center, we know $f(a)$ converges, so $0 \in S$. There are three mutually-exclusive possibilities to consider: either $S = \{0\}$, S is unbounded above, or S is bounded above with $S \neq \{0\}$.

In the first possibility, when $S = \{0\}$, this means that $f(x)$ only converges at $x = a$. This is case 1 from the theorem.

In the second possibility, suppose S is unbounded above. Therefore, for any $x \in \mathbb{R}$, there is some $s \in S$ with $s > |x-a|$. Hence, by definition of S, there is some $y \in \mathbb{R}$ such that $|y-a| = s$ and $f(y)$ converges. By Lemma 10.28, f converges absolutely at x. Thus, f converges absolutely on $\mathbb{R}$, and case 2 is satisfied.

For the third possibility, suppose that $S \neq \{0\}$ and S is bounded above. We define $R = \sup S$. Note that S has a positive number in it, so $R > 0$. We claim that R satisfies the conditions of case 3.

Let $x \in \mathbb{R}$ be given. If $|x-a| > R$, then by definition of supremum, $|x-a| \notin S$. It follows, by the definition of S, that f diverges at x.

Otherwise, if $|x-a| < R$, then $|x-a|$ is not an upper bound of S, so there is some $s \in S$ with $s > |x-a|$. Also, there is some $y \in \mathbb{R}$ such that $|y-a| = s$ and $f(y)$ converges, by definition of S. Thus, by Lemma 10.28, f converges at x, as desired. This verifies the conditions of case 3.

Lastly, we address uniform convergence. Suppose that we are in case 2 or 3 and $r \in \mathbb{R}$ is given with $0 < r < R$. The arguments from these cases show that there is some $s \in S$ with $r < s$, and hence there is some $y \in \mathbb{R}$ with $|y-a| = s$ and $f(y)$ convergent. By Lemma 10.28, f converges uniformly on $(a-r, a+r)$. ■

Theorem 10.29 helps us analyze power series, because it describes very specifically the sets on which they converge. This occasionally comes in handy for series which do not work well with the Root or Ratio Tests. For instance, we offer these examples:

Example 10.30:

Consider the power series
$$\sum_{n=2}^{\infty} \frac{x^n}{\log n}$$
centered at 0. We can test the absolute-value series using the Ratio Test, yielding the ratio of
$$\left| \frac{x^{n+1}}{\log(n+1)} \cdot \frac{\log n}{x^n} \right| = |x| \frac{\log n}{\log(n+1)}$$
To determine the limit as $n \to \infty$, L'Hôpital's Rule works[8], and we obtain the limit of $|x|$. Thus, our series converges absolulely for $|x| < 1$ and diverges for $|x| > 1$, so the series has radius 1. (The endpoints of the interval of convergence, i.e. $x = \pm 1$, need to be checked using other tests; you can show that we get conditional convergence at -1 and divergence at 1.)

Another, perhaps easier, way to analyze this series is to note that $x^n / \log n$ satisfies the following for all $x \in \mathbb{R}$ and all $n \in \mathbb{N}^*$ with $n > e$:
$$\frac{|x|^n}{n} \le \frac{|x|^n}{\log n} \le |x|^n$$

[8] Remember that this limit involves n, so when applying L'Hôpital's Rule, you want to differentiate with respect to n, not with respect to x.

Both $\sum x^n$ and $\sum x^n/n$ have radius 1 (as we have shown with earlier examples), so this suggests that $\sum x^n/\log n$ also has radius 1. To prove this more formally, we use two instances of the Comparison Test. When $|x| < 1$, the comparison to $|x|^n$ shows that $\sum |x|^n/\log n$ converges. When $|x| > 1$, the comparison to $|x|^n/n$ shows that $\sum |x|^n/\log n$ diverges.

Therefore, the series $\sum x^n/\log n$ converges absolutely for $|x| < 1$ and does not converge absolutely for $|x| > 1$. Hence its radius of convergence must be 1 by Theorem 10.29. One nice advantage of using Theorem 10.29 is that although the work above did not prove divergence of $\sum x^n/\log n$ for $|x| > 1$ (it only proved that the absolute-value series $\sum |x|^n/\log n$ diverges for $|x| > 1$), we obtain divergence for $|x| > 1$ automatically from Theorem 10.29. $\square$

Example 10.31:

Consider the power series

$$f(x) = \sum_{n=0}^{\infty}(1 + (-2)^n)x^n$$

It is possible to analyze this series with the Ratio Test or the Root Test, though the work is a little difficult, especially with the Root Test. Instead, we try a different approach. It seems plausible to split up 1 and $(-2)^n$ in the above series and write

$$f(x) = \sum_{n=0}^{\infty} x^n + \sum_{n=0}^{\infty}(-2x)^n$$

but this equation requires both $\sum x^n$ and $\sum(-2x)^n$ to converge in order to be valid.[9] Nevertheless, we can still gain useful knowledge about $f(x)$ by studying $\sum x^n$ and $\sum(-2x)^n$.

We already know that the geometric series $\sum x^n$ converges absolutely for $|x| < 1$ and diverges otherwise. Similarly, $\sum(-2x)^n$ converges absolutely when $|-2x| < 1$ and diverges otherwise, so this series has radius of convergence 1/2. Thus, since $\sum x^n$ and $\sum(-2x)^n$ both converge for $|x| < 1/2$, so

[9] Actually, this equation also makes sense when exactly one of $\sum x^n$ and $\sum(-2x)^n$ converges, because adding the terms of a convergent series and a divergent series must produce a divergent result.

does $f(x)$ as shown in the equation above. Similarly, if $1/2 < |x| < 1$, then $f(x)$ is the sum of a convergent and a divergent series, so $f(x)$ diverges.

Putting this together, since $f(x)$ converges for $|x| < 1/2$, the radius of convergence for f is at least $1/2$. Since $f(x)$ diverges when $1/2 < |x| < 1$, the radius of convergence is at most $1/2$. Thus, the radius is exactly $1/2$. Furthermore, the equation $f(x) = \sum x^n + \sum(-2x)^n$ shows that $f(1/2)$ and $f(-1/2)$ both diverge.

(One useful consequence of Theorem 10.29 applied to this example is that $f(x)$ also diverges when $|x| \geq 1$, since the radius of convergence is $1/2$. This conclusion does NOT follow immediately from writing $f(x) = \sum x^n + \sum(-2x)^n$, since adding the terms of two divergent series may or may not yield a divergent result.) $\square$

Remark. When reasoning about power series and their radii of convergence, the following corollary to Theorem 10.29 is often handy:

Corollary 10.32. *Let $a \in \mathbb{R}$ and a sequence $(c_n)_{n=0}^{\infty}$ of real numbers be given. Let f be the power series*

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

and let R be its radius of convergence.

Let $x \in \mathbb{R}$ be given. If $f(x)$ converges, then $|x-a| \leq R$. If $f(x)$ diverges, then $R \leq |x-a|$.

You can prove this corollary in Exercise 10.5.12.

This corollary helps to shorten the work needed for finding a radius of convergence by plugging in appropriate choices of x. For instance, let's revisit Example 10.30, where we analyzed the power series $\sum x^n/\log n$. We saw this series has radius at least 1 by comparing against the series $\sum x^n$. To see that the series has radius at most 1, we note that for any $b > 1$, $\sum b^n/\log n$ diverges because $b^n/\log n \to \infty$ as $n \to \infty$. Thus, the radius cannot be larger than b for any $b > 1$, so the radius is at most 1.

Integrating and Differentiating Power Series

Let's consider a power series f with center of a and radius of convergence R, so f has the form

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

Thanks to Theorem 10.29, we know that f converges uniformly on intervals of the form $(a-r, a+r)$ for $0 < r < R$. (For now, we consider only the cases when $R \neq 0$, since an interval of convergence consisting of a single point is uninteresting.) This means that our results about uniform series of functions apply to power series. Since each partial sum of a power series is a polynomial, which is continuous and integrable, it follows by Theorems 10.16 and 10.17 that for each r with $0 < r < R$, f is both continuous and integrable on $(a-r, a+r)$.

Since every x belonging to $(a-R, a+R)$ (note that when $R = \infty$, we treat this interval as $\mathbb{R}$) also belongs to some interval of the form $(a-r, a+r)$ with $0 < r < R$, we see that f is continuous on $(a-R, a+R)$. Also, for every $x \in (a-R, a+R)$, we have

$$\begin{aligned} \int_a^x f(t)\, dt &= \int_a^x \sum_{n=0}^{\infty} c_n (t-a)^n\, dt \\ &= \sum_{n=0}^{\infty} c_n \int_a^x (t-a)^n\, dt = \sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1} \end{aligned}$$

by integrating term by term.

This means that within the interior of the interval of convergence (the endpoints are a more subtle matter), a power series can be integrated as if it were a polynomial! Performing this integration yields a new power series whose $(n+1)^{\text{st}}$ coefficient is $c_n/(n+1)$ (and whose 0^{th} coefficient is 0). This raises the question: what is the radius of the integrated power series? Is it bigger or smaller than R?

It is not hard to show that the radius of the integrated series is at least as large as R. The work we just presented shows that the integrated series

$$\sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1}$$

converges absolutely for every $x \in (a-R, a+R)$. (In fact, by Theorem 10.17, this integrated series converges uniformly on $(a-r, a+r)$ for any r with $0 < r < R$.) It follows by Corollary 10.32 that the radius of the integrated series cannot be any smaller than R. With a little more work, we get the very important result that the radius of the integrated series is R:

Theorem 10.33. *Let $a \in \mathbb{R}$ and a sequence of real numbers $(c_n)_{n=0}^{\infty}$ be given. Let f be the power series centered at a with coefficients (c_n). Let R be the radius of convergence of f. Let g be the power series defined by*

$$g(x) = \sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1}$$

for all $x \in \mathbb{R}$ where the series converges. Then g has radius of convergence R, and

$$\int_x^y f(t)\,dt = g(t)|_x^y = g(y) - g(x)$$

for all $x, y \in (a-R, a+R)$.

Remark. Although the integrated series has the same radius, the behavior might change at the endpoints of the interval of convergence. For instance, the geometric series

$$\sum_{n=0}^{\infty} x^n$$

has radius 1 and diverges when $|x| = 1$, but the integrated series

$$\sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = \sum_{n=1}^{\infty} \frac{x^n}{n}$$

converges conditionally at $x = -1$.

Strategy. Most of the parts of this theorem were already proven in the remarks before the statement of the theorem. All that remains is showing that the radius of g is at most R. In other words, if R_g is the radius of g, then we've seen that $R_g \geq R$, so now we want $R_g \leq R$.

There are two possible ways to proceed. One way is to show that $g(x)$ diverges outside of f's interval of convergence. In other words, we prove that $|x-a| > R$ implies divergence of $g(x)$. (It follows by Corollary 10.32 that R_g cannot be larger than R.) However, this is a difficult approach to use, because although we know that $f(x)$ diverges when $|x-a| > R$, we don't have much information about HOW it diverges. In particular, we don't know whether the partial sums of $f(x)$ blow up in magnitude or not.

Instead, we'll use a different approach: we'll show that $f(x)$ converges inside of g's interval of convergence. In other words, we'll show that when

$|x-a| < R_g$, $f(x)$ converges. We try this strategy because Theorem 10.29 tells us that $g(x)$ converges ABSOLUTELY for $|x-a| < R_g$, which is a nice fact we can exploit.

Thus, let's suppose $|x-a| = r < R_g$, so that

$$g(x) = \sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1}$$

converges absolutely. We'd like to show that

$$f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n$$

also converges absolutely by comparing its terms against the terms of $g(x)$. First, let's try using the main idea from the proof of Lemma 10.28: analyze the proportion between the terms of the two series. This leads us to consider the ratio

$$\left|\frac{c_n(x-a)^n}{c_n(x-a)^{n+1}/(n+1)}\right| = \frac{n+1}{|x-a|} = \frac{n+1}{r}$$

(We can suppose that $x \neq a$, since any power series automatically converges at its center). Unfortunately, $(n+1)/r$ goes to ∞ as $n \to \infty$.

How do we deal with this issue? Looking back at the proof from Lemma 10.28, we notice one important feature of that proof which we have not imitated: in the lemma, the ratio was computed with two points x and y satisfying $|x-a| < |y-a|$. Because $|x-a| < |y-a|$ in that argument, an exponential term like $(|x-a|/|y-a|)^n$ appeared in the ratio, and this term was crucial for obtaining a comparison with a convergent series. We'd like a similar term to appear in our proof.

Therefore, let's take some value y such that $|x-a| < |y-a| < R_g$, and let's define $s = |y-a|$. We know that $g(y)$ converges absolutely by definition of R_g. Now, we try to compare the magnitudes of $f(x)$'s n^{th} term and $g(y)$'s n^{th} term (where we suppose $c_n \neq 0$ for simplicity):

$$\begin{aligned} |c_n(x-a)^n| \le \left|\frac{c_n}{n+1}(y-a)^{n+1}\right| &\leftrightarrow |x-a|^n \le \frac{|y-a|^{n+1}}{n+1} \\ \leftrightarrow \quad r^n \le \frac{s^{n+1}}{n+1} &\leftrightarrow n+1 \le \frac{s^{n+1}}{r^n} = s\left(\frac{s}{r}\right)^n \end{aligned}$$

Since $0 < r < s$, $(s/r)^n$ approaches ∞ much faster than $n+1$ does! Thus, our comparison works for large enough values of n.

Proof. Let $a, (c_n), f, R, g$ be given as described. Let R_g be the radius of convergence of g. The remarks made before the statement of the theorem prove that $R_g \geq R$. It remains to show that $R_g \leq R$, and then all the remaining statements of the theorem are proven as described before the statement of the theorem. We will prove that for all $x \in \mathbb{R}$ with $0 < |x-a| < R_g$, $f(x)$ converges, so that $R \geq |x-a|$ by Corollary 10.32. Once this is done, we know $R \geq |x-a|$ whenever $|x-a| < R_g$, so it follows that $R \geq R_g$ as desired.

Let $x \in \mathbb{R}$ be given with $0 < |x-a| < R_g$. Choose some $y \in \mathbb{R}$ such that $|x-a| < |y-a| < R_g$ (for instance, $y = a + (R_g + |x-a|)/2$ works), and define $r = |x-a|$ and $s = |y-a|$. Because $|y-a| < R_g$,

$$g(y) = \sum_{n=0}^{\infty} \frac{c_n(y-a)^{n+1}}{n+1}$$

converges absolutely by definition of R_g. We will use this fact to show that $f(x)$ converges absolutely by the Comparison Test. Namely, we will prove that when n is large enough,

$$|c_n(x-a)^n| \leq \left| \frac{c_n}{n+1}(y-a)^{n+1} \right|$$

When $c_n = 0$, our goal inequality is automatically satisfied. When $c_n \neq 0$, we replace $|x-a|$ by r and $|y-a|$ by s, and we rewrite our goal inequality as follows:

$$\begin{aligned} & |c_n(x-a)^n| \leq \left| \frac{c_n}{n+1}(y-a)^{n+1} \right| \leftrightarrow |x-a|^n \leq \frac{|y-a|^{n+1}}{n+1} \\ \leftrightarrow \quad & r^n \leq \frac{s^{n+1}}{n+1} \leftrightarrow 1 \leq \frac{s^{n+1}}{r^n(n+1)} = \frac{s(s/r)^n}{n+1} \end{aligned}$$

Because $0 < r < s$, and exponentials dominate polynomials, $s(s/r)^n/(n+1) \to \infty$ as $n \to \infty$. Therefore, we may pick some $N \in \mathbb{N}$ such that whenever $n > N$, $s(s/r)^n/(n+1) > 1$. As a result, for all $n > N$, the n^{th} term of $f(x)$ has a smaller magnitude than the n^{th} term of $g(y)$. Because $g(y)$ converges absolutely, $f(x)$ converges absolutely by the Comparison Test. ■

When f is a power series with radius of convergence R and center a, Theorem 10.33 doesn't just tell us that we can integrate f term by term on $(a-R, a+R)$. It also tells us that the radius of convergence doesn't grow when we integrate! (In fact, most of the work of the previous theorem was spent proving that the radius doesn't grow.) Equivalently, if we start with the integrated series g from the theorem, and we differentiate term by term to get back f, then the radius doesn't shrink! This leads to the following corollary:

Corollary 10.34. *Let $a \in \mathbb{R}$ and a sequence of real numbers $(c_n)_{n=0}^{\infty}$ be given. Let f be the power series centered at a with coefficients (c_n). Let R be the radius of convergence of f. Let g be the power series defined by*

$$g(x) = \sum_{n=1}^{\infty} nc_n(x-a)^{n-1} = \sum_{n=0}^{\infty}(n+1)c_{n+1}(x-a)^n$$

for all $x \in \mathbb{R}$ where the series converges. Then g has radius of convergence R, and $f'(x) = g(x)$ for all $x \in (a-R, a+R)$. We say that g is obtained by differentiating f term by term.

Strategy. The main idea is that since g can be obtained from f by differentiating terms, f can be obtained from g by integrating terms. More specifically, let's say R_g is the radius of convergence of g, and consider the series

$$\int_a^x g(t)\,dt = \sum_{n=1}^{\infty}\left(\int_a^x nc_n(t-a)^{n-1}\,dt\right) = \sum_{n=1}^{\infty} c_n(x-a)^n$$

defined for all $x \in (a-R_g, a+R_g)$. This integrated series has radius R_g by Theorem 10.33, but we also see that the integral is $f(x) - f(a) = f(x) - c_0$. Thus, $f'(x) = g(x)$ by the Fundamental Theorem of Calculus.

Proof. Let $a, (c_n), f, R, g$ be given as described. Let R_g be the radius of convergence of g. For any $x \in (a-R_g, a+R_g)$, Theorem 10.33 tells us that

$$\int_a^x g(t)\,dt = \sum_{n=1}^{\infty}\left(\int_a^x nc_n(t-a)^{n-1}\,dt\right) = \sum_{n=1}^{\infty} c_n(x-a)^n$$

and this series has radius of convergence R_g. Furthermore, we see that

$$\int_a^x g(t)\,dt = \left(\sum_{n=0}^{\infty} c_n(x-a)^n\right) - c_0 = f(x) - f(a)$$

Hence, f has radius of convergence R_g, so $R_g = R$.

Lastly, because g has radius R, Theorem 10.29 tells us that g is continuous on $(a-R, a+R)$. Therefore, for any $x \in (a-R, a+R)$, the Fundamental Theorem of Calculus lets us differentiate our previous equality with respect to x and obtain $g(x) = f'(x)$, as desired. ■

Corollary 10.34 is a result that sets power series apart from many other series of functions. We have previously seen that differentiation does not behave well with uniform convergence: a uniform limit of differentiable functions can fail to be differentiable, or perhaps the derivative of the limit fails to equal the limit of derivatives. In contrast, Corollary 10.34 proves that a power series is differentiable inside its interval of convergence, the derivative can be computed very easily (term by term), and the derivative is a power series with the same radius! Essentially, power series enjoy all the properties of differentiation that we would ideally like to have.

Example 10.35:

Consider the geometric series

$$f(x) = \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

which has center 0 and radius 1. By Corollary 10.34, we may differentiate this series term by term on $(-1, 1)$ and obtain

$$f'(x) = \frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} nx^{n-1}$$

(Note that this starts at $n = 1$ because the $n = 0$ term of $f(x)$ is a constant and hence disappears in the derivative.) The series for f' also has radius 1, so we may differentiate again on $(-1, 1)$ to get

$$f''(x) = \frac{2}{(1-x)^3} = \sum_{n=2}^{\infty} n(n-1)x^{n-2}$$

(As before, the $n = 1$ term of f' disappears in f''.)

We can continue taking derivatives for each $x \in (-1, 1)$, and we obtain the formula

$$f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}} = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)x^{n-k}$$

for all $x \in (-1,1)$ and all $k \in \mathbb{N}$ (which can be readily proven by induction on k). As a result, we obtain power series representations for $1/(1-x)^{k+1}$ for every $k \in \mathbb{N}$. In fact, using the binomial coefficients $\binom{n}{k}$, we can write this as

$$\frac{1}{(1-x)^{k+1}} = \sum_{n=k}^{\infty} \frac{n(n-1)\cdots(n-k+1)}{k!} x^{n-k} = \sum_{n=k}^{\infty} \binom{n}{k} x^{n-k}$$

We can also shift the index of the series to start at 0 and obtain

$$\frac{1}{(1-x)^{k+1}} = \sum_{n=0}^{\infty} \binom{n+k}{k} x^n$$

This can be used to deduce identities about binomial coefficients: see Exercise 10.5.25. $\square$

Computing Sums of Some Power Series

Because we can integrate and differentiate power series term by term, we can find closed-form sums of several power series. We have demonstrated the main ideas before in Examples 10.18 and 10.19. Now, we can get a few more formulas by using differentiation. We present some examples now, and we will have more examples when we talk about Taylor series in the next section.

Example 10.36:

In this example, we find a couple power series whose sum is

$$f(x) = \frac{1}{2-x}$$

The main idea is to relate $f(x)$ to the geometric series

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

for $x \in (-1,1)$.

One way to proceed is to note that $2-x = 1-(x-1)$. Hence,

$$f(x) = \frac{1}{1-(x-1)} = \sum_{n=0}^{\infty} (x-1)^n$$

which is valid when $x - 1 \in (-1, 1)$, i.e. $x \in (0, 2)$. This writes $f(x)$ as a series with center 1 and radius 1.

Another way of solving this problem is to factor 2 out of the denominator to get

$$f(x) = \frac{1}{2}\left(\frac{1}{1 - (x/2)}\right) = \frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^n = \sum_{n=0}^{\infty}\frac{x^n}{2^{n+1}}$$

This series has center 0 and radius 2, since it converges for $|x/2| < 1$ and diverges for $|x/2| \geq 1$. Since this second series converges in more places than our first series, this second series is probably a better choice of series. □

Example 10.37:

In this example, we'd like to obtain the exact value of

$$\sum_{n=1}^{\infty}\frac{1}{n2^n}$$

There are techniques in discrete math for finding this sum, but they are generally difficult. Instead, let's show how we can find this sum using power series. The main trick is to note that if we define

$$f(x) = \sum_{n=1}^{\infty}\frac{x^n}{n}$$

then our desired sum is just $f(1/2)$. (This is valid because f has radius 1 about the center 0, as you can check.)

Now, $f(x)$ is quite similar to $\sum x^n$ except for the n in the denominator. We can get rid of that n by taking a derivative term by term, obtaining

$$f'(x) = \sum_{n=1}^{\infty} x^{n-1} = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

Therefore, by integrating both sides, we get

$$f(x) - f(0) = \int_0^x f'(t)\,dt = \int_0^x \frac{dt}{1-t} = -(\log(1-x) - \log 1)$$

This is valid for all $x \in (-1, 1)$.

Since f is centered at 0 and has zero constant term, we see that $f(0) = 0$. Thus, $f(x) = -\log(1-x)$ and hence $f(1/2) = -\log(1/2) = \log 2$. □

10.5 Exercises

1. Determine where the series

$$f(x) = \sum_{n=1}^{\infty} \frac{(\cos(\pi x))^n}{n}$$

converges absolutely, converges conditionally, or diverges. This shows that concepts like "interval of convergence" or "center" do not necessarily make sense for series which are not power series.

For Exercises 2 through 11, a power series is given. Determine the radius of convergence of each series. Also, for Exercises 2 through 5, when the radius is finite, determine the interval of convergence (i.e. check the endpoints).

2. $\displaystyle\sum_{n=0}^{\infty}(-1)^n x^{2n}$

3. $\displaystyle\sum_{n=0}^{\infty}\frac{n^2 x^n}{3^{n+2}}$

4. $\displaystyle\sum_{n=2}^{\infty}\frac{x^n}{n^2 \log n}$

5. $\displaystyle\sum_{n=0}^{\infty}\frac{n!\ x^n}{(2n)!}$

6. $\displaystyle\sum_{n=0}^{\infty}\frac{n^n x^n}{n!}$

7. $\displaystyle\sum_{n=0}^{\infty}(2+(-1)^n)^n x^n$

(Hint: Consider Example 10.31.)

8. $\displaystyle\sum_{n=0}^{\infty}3^{\sqrt{n}}x^n$

9. $\displaystyle\sum_{n=0}^{\infty}a^{(n^2)}x^n$ where $a > 0$

(This answer has several cases depending on how large a is.)

10. $\displaystyle\sum_{n=1}^{\infty}\left(\frac{a^n}{n}+\frac{b^n}{n^2}\right)x^n$

where $a, b > 0$

11. $\displaystyle\sum_{n=0}^{\infty}(\sin n)x^n$

12. Prove Corollary 10.32.

13. Prove that if $(c_n)_{n=0}^{\infty}$ is a bounded sequence of real numbers, then the power series $\sum_{n=0}^{\infty} c_n x^n$ has radius at least 1.

14. Suppose that $(c_n)_{n=0}^{\infty}$ is a sequence of real numbers and the power series

$$\sum_{n=0}^{\infty} c_n x^n$$

has radius $R > 0$. For each $k \in \mathbb{N}^*$, what is the radius of

$$\sum_{n=0}^{\infty} c_n x^{kn}?$$

(Hint: Corollary 10.32 might be useful.)

15. Suppose that $a \in \mathbb{R}$ is given and $(c_n)_{n=0}^{\infty}$ is a sequence of real numbers, and suppose that

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

is a power series with radius of convergence $R > 0$.

 (a) Prove that f is infinitely differentiable at a and, for all $n \in \mathbb{N}$, we have

$$c_n = \frac{f^{(n)}(a)}{n!}$$

 (b) Suppose that $(d_n)_{n=0}^{\infty}$ is given such that

$$g(x) = \sum_{n=0}^{\infty} d_n (x-a)^n$$

 is a power series with radius of convergence R, and assume $f(x) = g(x)$ for all $x \in (a-R, a+R)$. Prove that $c_n = d_n$ for every $n \in \mathbb{N}$. (Thus, if two power series represent the same function on $(a-R, a+R)$, then their coefficients must match!)

16. Suppose that f is a power series of the form

$$f(x) = \sum_{n=0}^{\infty} c_n x^n$$

with radius of convergence $R > 0$. Prove that f is an even function on $(-R, R)$ (i.e. $f(-x) = f(x)$ for all $x \in (-R, R)$) iff $c_n = 0$ for all odd $n \in \mathbb{N}$. Similarly, prove that f is an odd function on $(-R, R)$ (i.e. $f(-x) = -f(x)$ for all $x \in (-R, R)$) iff $c_n = 0$ for all even $n \in \mathbb{N}$.

(Hint: If f is even, then show

$$f(x) - \sum_{n=0}^{\infty} c_{2n} x^{2n} = \sum_{n=0}^{\infty} c_{2n+1} x^{2n+1}$$

is both even and odd. You may want to use the result of Exercise 10.5.15.)

In Exercises 17 through 22, an expression defining a real function of x is given. Find a power series with center 0 whose sum is the specified real function, and find the radius of convergence of the power series. (For instance, see Example 10.36.)

17. $\dfrac{1}{1+x}$

18. $\log(1+x)$

19. $\log\left(\dfrac{1+x}{1-x}\right)$

 (See also Exercise 8.5.22.)

20. $\dfrac{1}{1+x^2}$

21. $\arctan x$

22. $\dfrac{x}{1+x-2x^2}$

 (Hint: Use partial fractions.)

In Exercises 23 and 24, a series is given. Use power series methods, much like in Example 10.37, to obtain the value of the sum.

23. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n2^n}$

24. $\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{(2n+1)3^n}$

 (Hint: Integrate a power series where x has been replaced with x^2.)

25. In Example 10.35, we saw that for each $k \in \mathbb{N}$ and all $x \in (-1, 1)$, we have

$$\frac{1}{(1-x)^{k+1}} = \sum_{n=0}^{\infty} \binom{n+k}{k} x^n$$

Use this identity, and the result of Exercise 10.5.15, to prove the following:

(a) For all $n, k \in \mathbb{N}$,

$$\binom{n+k+1}{k+1} = \binom{n+k}{k} + \binom{n+k}{k+1}$$

(Hint: Write $1/(1-x)^{k+1}$ as $(1-x)/(1-x)^{k+2}$.)

(b) For all $n, k \in \mathbb{N}$,

$$(k+1)\binom{n+k+1}{k+1} = (n+1)\binom{n+k+1}{k}$$

(Hint: Differentiate $1/(1-x)^{k+1}$.)

10.6 Some Applications of Power Series

In the last section, we introduced power series and saw that they are easy to integrate and differentiate within their interval of convergence. It follows that a power series with positive radius is infinitely differentiable at its center. In fact, in Exercise 10.5.15, we saw that when a function is given by a power series with positive radius, the coefficients of that series are uniquely determined. For convenience, we restate that result here:

Theorem 10.38. *Let $a \in \mathbb{R}$ be given, let f be a real function which is infinitely differentiable at a, and suppose that there exist $(c_n)_{n=0}^{\infty}$ and $R > 0$ (R might be ∞) satisfying*

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

for all $x \in (a-R, a+R)$. Then

$$c_n = \frac{f^{(n)}(a)}{n!}$$

for all $n \in \mathbb{N}$.

Remark. When a, f, and R satisfy the hypotheses of Theorem 10.38, we say that f is *analytic* on $(a-R, a+R)$. In other words, an analytic function is a function whose values are given by a power series.

In this section, we explore a few uses for power series. First, we will explore the relationship between power series and Taylor polynomials of functions, which lets us find useful representations for many common functions. Next, we will use power series to solve some differential equations, where term-by-term differentation of powers series comes in handy. Lastly, we will show how the uniqueness of power series coefficients can be used to help compute some sequences from discrete mathematics!

Taylor Series

Recall from Chapter 8 that when $a \in \mathbb{R}$, $n \in \mathbb{N}$, and f is a real function which is n-times differentiable at a, the n^{th}-order Taylor polynomial for f at a is defined by

$$T_n f(x; a) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i$$

for all $x \in \mathbb{R}$. When we compare this with the formula from Theorem 10.38, we see that $T_n f$ is the n^{th} partial sum of f's power series! This should be no surprise, because $T_n f$ is specifically designed to agree with f and its first n derivatives at a, whereas the power series for f is an "infinite-degree polynomial" which agrees with ALL the derivatives of f at a.

This suggests that when f is a function is infinitely differentiable at a, we can form the power series Tf given by the limit of $T_n f$ as $n \to \infty$:

$$Tf(x; a) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

This is called the *Taylor series* for f at a.[10] Theorem 10.38 tells us that IF f is analytic, i.e. if f can be represented by a power series, THEN the power series for f has to be the Taylor series Tf.

However, this raises the question: when does the Taylor series actually converge to $f(x)$? In other words, if f is infinitely differentiable, then is it analytic? There are two issues to consider here. First, $Tf(x; a)$ might not converge at all values where $f(x)$ is defined. For instance, when $f(x) = 1/(1-x)$, f is defined on $\mathbb{R} - \{1\}$, but

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

[10] When $a = 0$, the series is also soemtimes called the *Maclaurin series* for f at a, in honor of the Scottish mathematician Colin Maclaurin.

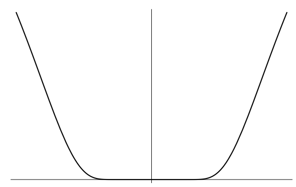

Figure 10.6: A non-analytic function $y = e^{-1/x^2}$ for $x \neq 0$, $y(0) = 0$

only when $|x| < 1$. Second, it might be possible for $Tf(x; a)$ to converge to a different value from $f(x)$. For instance, Exercise 8.2.16 shows that when

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

we have $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$ and hence $Tf(x; 0) = 0$ for all $x \in \mathbb{R}$. Thus, $Tf(x; 0) \neq f(x)$ except when $x = 0$. The graph of $y = f(x)$ is displayed in Figure 10.6; intuitively, the graph is so flat near 0 that the Taylor series is "fooled" into thinking the function is constant!

To deal with these issues, our most important tool is the Taylor error E_n, defined by $E_n f(x; a) = f(x) - T_n f(x; a)$. We have $Tf(x; a) = f(x)$ precisely when $E_n f(x; a) \to 0$ as $n \to \infty$. To analyze $E_n f(x; a)$, we can use Lagrange's form of the error (Theorem 8.16): for all $x \in \mathbb{R}$, there is some c between x and a (which depends on x, a, and n) such that

$$E_n f(x; a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - a)^{n+1}$$

This suggests that if $f^{(n)}$ grows more slowly than $n!$, then $E_n f \to 0$. For instance, you can prove the following in Exercise 10.7.1:

Theorem 10.39. *Let $a \in \mathbb{R}$ and $r \in (0, \infty)$ be given, and let f be a real function which is infinitely differentiable on $(a - r, a + r)$. Suppose that $C, A > 0$ are given such that*

$$|f^{(n)}(x)| \leq CA^n$$

for all $x \in (a-r, a+r)$. (Thus, the n^{th} derivative grows at most exponentially, independently of x.) Then $E_n f(x; a) \to 0$ uniformly on $(a - r, a + r)$ as $n \to \infty$, i.e. $T_n f(x; a) \to f(x)$ uniformly on $(a - r, a + r)$ as $n \to \infty$.

This theorem provides a useful way to check many Taylor series for functions we studied when computing Taylor polynomials in Chapter 8. When applying this theorem, r is often chosen to be an arbitrary positive number less than the radius of convergence of the Taylor series. For instance, we offer the following examples:

Example 10.40:

Suppose $f(x) = e^x$ for all $x \in \mathbb{R}$. We have

$$Tf(x;0) = \lim_{n\to\infty} T_n f(x;0) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

The Ratio Test shows that this series has infinite radius of convergence.

We'd like to show that this Taylor series converges to e^x by using Theorem 10.39. For any $r \in (0,\infty)$, and for all $x \in (-r,r)$, we have

$$|f^{(n)}(x)| = |e^x| \le e^r \le CA^n$$

where we choose $C = e^r$ and $A = 1$. (These are not the only choices of C and A which work.) Thus, the Taylor series for f converges uniformly to e^x on $(-r,r)$ for any $r \in (0,\infty)$. (However, note that the convergence is NOT uniform on $\mathbb{R}$, because $x^n/n!$ does not converge uniformly to 0 on $\mathbb{R}$ as $n \to \infty$; see Exercise 10.3.11).

Therefore, for any $x \in \mathbb{R}$, we have

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

We can use this to obtain Taylor series for other related functions. For instance, by replacing x with x^2, we obtain

$$e^{(x^2)} = \sum_{n=0}^{\infty} \frac{x^{2n}}{n!}$$

If we write this in the form $\sum c_n x^n$, then we find that for all $k \in \mathbb{N}$,

$$c_{2k+1} = 0 \qquad c_{2k} = \frac{1}{k!}$$

Therefore, when $g(x) = e^{(x^2)}$, Theorem 10.38 tells us that $g^{(2k+1)}(0) = 0$ and $g^{(2k)}(0) = (2k)!/k!$ for all $k \in \mathbb{N}$. This result provides an easy way of obtaining derivatives of g at 0 without repeatedly using the Chain Rule!

Furthermore, we can integrate the Taylor series for g term by term and obtain

$$\int_0^x e^{(t^2)}\,dt = \sum_{n=0}^{\infty} \int_0^x \frac{t^{2n}}{n!}\,dt = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{n!\,(2n+1)}$$

This is particularly useful since $e^{(t^2)}$ does not have an elementary derivative, so we can't obtain a convenient closed form for its integrals. However, we do have a series for the integral, and this series can be used to approximate the value of the integral by partial sums. $\square$

Example 10.41:

Let's consider $f(x) = \sin x$ for all $x \in \mathbb{R}$. We have

$$T\sin(x;0) = \lim_{n\to\infty} T_n \sin(x;0) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}$$

This series has infinite radius of convergence, as you can check. Also, for any $n \in \mathbb{N}$, $f^{(n)}$ is either $\pm\sin$ or $\pm\cos$, so $|f^{(n)}(x)| \le 1$ for all $x \in \mathbb{R}$. Therefore, we can choose $C = A = 1$ in Theorem 10.39, and we choose r to be any positive number, and we find that the Taylor series converges uniformly to $\sin x$ on $(-r, r)$.

Similarly, the series

$$T\cos(x;0) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

converges uniformly to $\cos x$ on every interval of the form $(-r, r)$. This can either be proven by the same approach as for sin, or it can be proven by differentiating the series for $\sin x$ term by term.

As in the previous example, we can obtain new series by substituting in different expressions for x, because the Taylor series for sine and cosine converge everywhere. For instance, we obtain

$$x\sin(x^2) = x\sum_{n=0}^{\infty} \frac{(-1)^n (x^2)^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{4n+3}}{(2n+1)!}$$

We can also integrate term by term to obtain formulas like

$$\int_0^x \cos(t^4)\,dt = \sum_{n=0}^{\infty} \int_0^x \frac{(-1)^n t^{8n}}{(2n)!}\,dt = \sum_{n=0}^{\infty} \frac{(-1)^n x^{8n+1}}{(2n!)\ (8n+1)}$$

□

In the last two examples, we presented examples of functions f and Taylor series which converge to f on $\mathbb{R}$. When the Taylor series converges on a finite interval, however, the endpoints usually require more subtle arguments. The following example illustrates the main idea:

Example 10.42:

Consider the Taylor series representation

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

which is valid for $|x| < 1$. By substituting values for x which have magnitude less than 1, we can obtain other formulas. For instance, by plugging in $-3x$, we get

$$\frac{1}{1+3x} = \sum_{n=0}^{\infty} (-3)^n x^n$$

whenever $|-3x| < 1$, i.e. $|x| < 1/3$. Also,

$$\frac{1}{1+x^2} = \sum_{n=0}^{\infty} (-1)^n x^{2n}$$

when $|-x^2| < 1$, i.e. $|x| < 1$. We can also integrate term by term: for instance, by plugging in $-x$ and then integrating, we get

$$\int_0^x \frac{1}{1+t}\,dt = \sum_{n=0}^{\infty} \int_0^x (-1)^n t^n\,dt$$

whenever $|x| < 1$, i.e.

$$\log(1+x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{n+1}}{n+1}$$

Let $f(x) = 1/(1+x)$ and $g(x) = \log(1+x)$. We've seen that $T_n f \to f$ and $T_n g \to g$ on $(-1,1)$ as $n \to \infty$. (It turns out that this convergence is not uniform on $(-1,1)$ but is uniform on $(-r,r)$ whenever $r \in (0,1)$.) We also note that the Taylor series for g converges at $x = 1$: $Tg(1)$ is the alternating harmonic series $\sum(-1)^n/(n+1)$. This raises the question: does the Taylor series $Tg(x)$ converge to $g(x)$ when $x = 1$?

To analyze this, we need to study the Taylor error $E_n g(x)$. Since g is obtained by integrating f, $E_n g$ can be obtained by integrating $E_{n-1}f$ (as integrating increases the degree by one). Furthermore, we actually know the exact value of $E_{n-1}f$: since

$$\frac{1}{1-x} = 1 + x + \cdots + x^{n-1} + \frac{1-x^n}{1-x}$$

we have $E_{n-1}f(x;0) = (1-x^n)/(1-x)$. After applying some inequalities and integrating (see Exercise 8.5.20 for more details), we eventually obtain

$$|E_n g(x)| \le \frac{|x|^{n+1}}{n+1}$$

Thus, for $x = 1$, we have $E_n g(1) \to 0$ as $n \to \infty$. This means that the Taylor series for g is also valid at the endpoint $x = 1$, giving the sum of $\log 2$. $\square$

By using similar techniques, you can show that almost all of the Taylor polynomials we found in Chapter 8 lead to Taylor series expansions which are valid in an interval with positive radius. We summarize some of the most important examples in the table below, all of which are studied by either using Theorem 10.39 or by manipulating the geometric series for $1/(1-x)$. All of these examples have their center at 0.

It is worth noting two subtle points here. First, when the radius R is finite, techniques like those in Example 10.42 are needed to analyze the endpoints of the interval of convergence. Second, these series do not usually converge uniformly on $(-R,R)$, but they do converge uniformly on $(-r,r)$ whenever $0 < r < R$. Informally, as long as you stay away from the ends of the interval, you can show uniform convergence (often by using the M-Test).

Function	Taylor Series	Where it is Valid
e^x	$\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n!}$	$\mathbb{R}$
$\sin x$	$\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}$	$\mathbb{R}$
$\cos x$	$\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$	$\mathbb{R}$
$\dfrac{1}{1-x}$	$\displaystyle\sum_{n=0}^{\infty} x^n$	$(-1, 1)$
$\log(1+x)$	$\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{n+1}}{n+1}$	$(-1, 1]$
$\arctan x$	$\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1}$	$[-1, 1]$

Differential Equations

A *differential equation* is an equation involving an unknown function and some of its derivatives. For instance, when y is an unknown function of x, the equation

$$y' = y$$

is a differential equation. We say this equation is *first-order* because the highest order of derivative in the equation is the first derivative y'. Similarly, the equation

$$y'' + y = 0$$

is a *second-order* differential equation since its highest-order derivative is the second derivative of y.

Now, let's show how the techniques from this chapter can be used to find solutions to differential equations. The basic idea is to assume y can be written as a power series $\sum c_n x^n$, plug this into the equation, and find conditions that the sequence (c_n) must satisfy[11]. We demonstrate the method with a few examples:

Example 10.43:

Consider the differential equation $y' = y$. Let's say we try and represent y as a power series

$$y(x) = \sum_{n=0}^{\infty} c_n x^n$$

where the coefficients (c_n) are to be determined. (We also ignore the issue of radius of convergence for now; once we find the coefficients, we can go back and figure out the radius.) Differentiating term by term, we have

$$y'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1}$$

Therefore, since $y' = y$, we have

$$\sum_{n=1}^{\infty} n c_n x^{n-1} = \sum_{n=0}^{\infty} c_n x^n$$

To write both sides in the same form, where x^n is on both sides, we shift the left sum to get

$$\sum_{n=0}^{\infty} (n+1) c_{n+1} x^n = \sum_{n=0}^{\infty} c_n x^n$$

By Theorem 10.38, the coefficients must match on each side of the equation, so for every $n \in \mathbb{N}$, we must have $(n+1)c_{n+1} = c_n$. In other words,

$$c_{n+1} = \frac{c_n}{n+1}$$

This gives us a recurrence for computing the coefficients c_1, c_2, c_3, and so forth from a value of c_0 (c_0 can be an arbitrary real number). The recurrence tells us that

$$c_1 = \frac{c_0}{1} \qquad c_2 = \frac{c_1}{2} = \frac{c_0}{2!} \qquad c_3 = \frac{c_2}{3} = \frac{c_0}{3!} \qquad \cdots$$

[11] This method generally finds all solutions to many differential equations; this is justified by uniqueness theorems proven in upper-level classes.

In general, an easy induction proof shows that $c_n = c_0/n!$ for any $n \in \mathbb{N}$. Therefore, our solution is

$$y(x) = \sum_{n=0}^{\infty} c_0 \frac{x^n}{n!} = c_0 e^x$$

Thus, we have obtained an infinite collection of solutions to our differential equation, one for each value of c_0. These solutions have infinite radius of convergence, so they satisfy $y'(x) = y(x)$ for all $x \in \mathbb{R}$.

By following the same approach, you can show that if $y(x) = \sum c_n x^n$ solves the differential equation $y' = ky$, where k is a constant, then $y(x) = c_0 e^{kx}$. (Compare this with the statement of Theorem 7.9.) $\square$

Example 10.44:

Consider the equation

$$y'' = -y$$

Let's represent y as a power series by $y(x) = \sum c_n x^n$. Therefore, we have

$$y'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1} \qquad y''(x) = \sum_{n=2}^{\infty} n(n-1) c_n x^{n-2}$$

and hence $y'' = -y$ implies

$$\sum_{n=2}^{\infty} n(n-1) c_n x^{n-2} = \sum_{n=0}^{\infty} -c_n x^n$$

By shifting the left sum to start from 0, so that we have terms of x^n on both sides, we get

$$\sum_{n=0}^{\infty} (n+2)(n+1) c_{n+2} x^n = \sum_{n=0}^{\infty} -c_n x^n$$

Matching coefficients, we find

$$c_{n+2} = \frac{-c_n}{(n+2)(n+1)}$$

for all $n \in \mathbb{N}$. By iterating this recurrence, we can obtain c_2 from c_0, c_4 from c_2, etc., as well as obtain c_3 from c_1, c_5 from c_3, etc. Thus, c_0 and c_1 can take arbitrary values, but after that, a routine induction shows that for each $n \in \mathbb{N}$,

$$c_{2n} = \frac{(-1)^n c_0}{(2n)!} \qquad c_{2n+1} = \frac{(-1)^n c_1}{(2n+1)!}$$

Putting this all together, we get

$$y(x) = c_0 \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} + c_1 \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = c_0 \cos x + c_1 \sin x$$

Notice that in this example, we have an infinite family of solutions specified by two constants c_0 and c_1, whereas the previous example had only one constant c_0. This is because this example has a second-order equation whereas the other has a first-order equation. In general, we typically find that all solutions to differential equations of order k are built by taking combinations of k separate basic[12] solutions using k arbitrary constants. (In this example, $\sin x$ and $\cos x$ are the basic solutions.) $\square$

Example 10.45:

Now, for an example where x makes an appearance in the differential equation, consider

$$y'(x) = xy(x)$$

Suppose that $y(x) = \sum c_n x^n$ as usual. Thus, using our earlier formula for y',

$$\sum_{n=1}^{\infty} nc_n x^{n-1} = x \sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty} c_n x^{n+1}$$

If we shift the right sum to start from 2, then we get an equation with x^{n-1} on both sides:

$$\sum_{n=1}^{\infty} nc_n x^{n-1} = \sum_{n=2}^{\infty} c_{n-2} x^{n-1}$$

Here, we note that the left side has a term for $n = 1$ whereas the right side does not. In other words, the right side has a coefficient of 0 for x^0. Since the coefficients of the power series must match on both sides, the left

[12] We will not take the time to formalize the notion of "basic solution"; this task is for a linear algebra course or a differential equations course.

side also has a coefficient of 0 for x^0, i.e. $1c_1x^0 = 0x^0$. This tells us that c_1 must be zero.

For all values of $n \geq 2$, however, setting the coefficients equal yields $nc_n = c_{n-2}$, so $c_n = c_{n-2}/n$. This recurrence lets us obtain c_2 from c_0, c_4 from c_2, and so on. You can show that for each $n \in \mathbb{N}$,

$$c_{2n} = \frac{c_{2n-2}}{2n} = \frac{c_{2n-4}}{(2n)(2(n-1))} = \cdots = \frac{c_0}{2^n\ n!}$$

On the other hand, since c_3 is obtained from c_1 using the recurrence, then c_5 is obtained from c_3, and so forth, you can also show that $c_{2n+1} = 0$ for each $n \in \mathbb{N}$. Putting this all together, we get

$$y(x) = \sum_{n=0}^{\infty} \frac{c_0 x^{2n}}{2^n\ n!} = c_0 \sum_{n=0}^{\infty} \frac{(x^2/2)^n}{n!} = c_0 e^{(x^2/2)}$$

□

As the examples demonstrate, power series yield a powerful method for finding solutions to differential equations. However, they can only be used to find analytic solutions to differential equations. For instance, consider the equation

$$x^2 y''(x) = y(x)$$

If $y(x)$ has the form $\sum c_n x^n$, then

$$x^2 \sum_{n=2}^{\infty} n(n-1)c_n x^{n-2} = \sum_{n=2}^{\infty} n(n-1)x^n = \sum_{n=0}^{\infty} c_n x^n$$

Matching coefficients, we have find that $n(n-1)c_n = c_n$ for all $n \geq 2$, and also $c_0 = c_1 = 0$. However, for $n \geq 2$, $n(n-1)c_n = c_n$ is only possible when $c_n = 0$. Thus, y is constantly zero, which is not a particularly interesting solution to the differential equation.

The issue here is that $y(x) = 0$ is the only analytic solution to $x^2y''(x) = y(x)$. It turns out that the two basic solutions to $x^2y''(x) = y(x)$ are not analytic: see Exercise 10.7.23. Nevertheless, the power series method is generally effective on many differential equations encountered in practice. Even if the power series does not have a convenient closed form like the ones from our earlier examples, having an answer in the form of a series can still be useful. For instance, see Exercise 10.7.26.

Solving Recurrences: The Method of Generating Functions

When solving differential equations with power series, we tend to get a recurrence for finding the coefficients (c_n) of the power series. Thus, by using power series, we turn a problem involving derivatives into a problem involving recurrences, i.e. a calculus problem becomes a discrete math problem. In the examples we covered, the recurrence was quite simple, and a closed-form formula could be proven by induction. However, what do we do with more complicated recurrences?

It turns out that a simple trick can be used to turn discrete math problems into calculus problems! The main idea is: when we have a recurrence describing a sequence $(a_n)_{n=0}^{\infty}$, we can create the power series

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

The function f is called the *generating function* for the sequence (a_n). We assume that f has some positive radius R of convergence (the exact value of this radius doesn't actually matter). Roughly, our main approach is to first obtain a closed form for f, and then we use this closed form to find a nice expression for the Taylor series, telling us the coefficients (a_n).

To illustrate how this works in practice, we present two examples. You can try more examples in the exercises.

Example 10.46:

Let's show how generating functions can be used to solve the recurrence

$$a_{n+1} = 2a_n + 1 \text{ for } n \in \mathbb{N} \qquad a_0 = 1$$

Let $f(x) = \sum a_n x^n$ be the generating function with radius $R > 0$. Informally, we use our recurrence to compute

$$\begin{aligned} f(x) &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \\ &= 1 + (2a_0 + 1)x + (2a_1 + 1)x^2 + (2a_2 + 1)x^3 + \cdots \\ &= 1 + 2x(a_0 + a_1 x + a_2 x^2 + \cdots) + x(1 + x + x^2 \cdots) \\ &= 1 + 2xf(x) + x\left(\frac{1}{1-x}\right) \end{aligned}$$

Thus, $f(x)$ is written in terms of itself, and we can solve for $f(x)$.

Let's do the work above more formally. Since $a_{n+1}x^{n+1} = (2a_n + 1)x^n$ for $n \in \mathbb{N}$, and $a_0 x^0 = a_0$, we write f in the form

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n x^n = a_0 + \sum_{n=0}^{\infty} a_{n+1} x^{n+1}$$

applying an index shift so that a_{n+1} appears. By plugging in our recurrence and our base case, this tells us that

$$f(x) = 1 + \sum_{n=0}^{\infty} (2a_n + 1)x^{n+1} = 1 + 2\sum_{n=0}^{\infty} a_n x^{n+1} + \sum_{n=0}^{\infty} x^{n+1}$$

which is valid when $|x| < \min\{R, 1\}$. Lastly, by pulling a factor of x out of the summations, we obtain

$$f(x) = 1 + 2x\sum_{n=0}^{\infty} a_n x^n + x\sum_{n=0}^{\infty} x^n = 1 + 2xf(x) + x\left(\frac{1}{1-x}\right)$$

By collecting the terms involving $f(x)$ to the same side, we obtain

$$f(x)(1-2x) = 1 + \frac{x}{1-x} = \frac{1}{1-x}$$

so we obtain the closed form

$$f(x) = \frac{1}{(1-x)(1-2x)}$$

whenever $|x| < \min\{R, 1\}$. This closed form looks related to the geometric series for $1/(1-x)$ and $1/(1-2x)$. To write f in terms of these series, we can use partial fractions!

Thus, let's write

$$\frac{1}{(1-x)(1-2x)} = \frac{A}{1-x} + \frac{B}{1-2x}$$

where A and B are to be determined. By clearing the denominators, we obtain

$$1 = A(1-2x) + B(1-x)$$

When $x = 1$, we find $A = -1$, and when $x = 1/2$, we find $B = 2$.

As a result, we know

$$f(x) = \frac{-1}{1-x} + \frac{2}{1-2x} = -\sum_{n=0}^{\infty} x^n + 2\sum_{n=0}^{\infty}(2x)^n = \sum_{n=0}^{\infty}(2^{n+1}-1)x^n$$

Since $f(x) = \sum a_n x^n$, we find $a_n = 2^{n+1} - 1$ for all $n \in \mathbb{N}$. $\square$

In Example 10.46, it is probably not hard to see a pattern for a_n and prove it by induction. However, this next example probably doesn't have as obvious a pattern:

Example 10.47:

Let's consider the recurrence

$$a_{n+2} = 5a_{n+1} - 6a_n \text{ for } n \in \mathbb{N} \qquad a_0 = 2, a_1 = 5$$

As before, let's set $f(x) = \sum a_n x^n$. Our first step is to find a closed form for f. In order to use our knowledge about a_0 and a_1, we set the terms a_0 and $a_1 x$ aside from $f(x)$, so that

$$f(x) = a_0 + a_1 x + \sum_{n=2}^{\infty} a_n x^n = a_0 + a_1 x + \sum_{n=0}^{\infty} a_{n+2} x^{n+2}$$

(we also shifted our index so that a_{n+2} appears). Plugging in our recurrence, we find

$$\begin{aligned} f(x) &= 2 + 5x + \sum_{n=0}^{\infty}(5a_{n+1} - 6a_n)x^{n+2} \\ &= 2 + 5x + 5\sum_{n=0}^{\infty} a_{n+1}x^{n+2} - 6\sum_{n=0}^{\infty} a_n x^{n+2} \end{aligned}$$

By pulling out factors of x and shifting indeces so that our terms have the form $a_n x^n$, we find

$$f(x) = 2 + 5x + 5x\sum_{n=1}^{\infty} a_n x^n - 6x^2 \sum_{n=0}^{\infty} a_n x^n$$

Lastly, we note that the sum from $n = 1$ is almost equal to $f(x)$, except that it's missing the $n = 0$ term a_0, which is 2. Thus, we find

$$f(x) = 2 + 5x + 5x\,(f(x) - 2) - 6x^2 f(x)$$

Solving for $f(x)$ yields $f(x)(1-5x-6x^2) = 2-5x$, so

$$f(x) = \frac{2-5x}{1-5x-6x^2} = \frac{2-5x}{(1-2x)(1-3x)}$$

Next, we aim to find a convenient Taylor series for f. As in the last example, we will do this by using partial fractions. Thus, let's say

$$\frac{2-5x}{(1-2x)(1-3x)} = \frac{A}{1-2x} + \frac{B}{1-3x}$$

which leads to

$$2-5x = A(1-3x) + B(1-2x) = (A+B) + (-3A-2B)x$$

Thus, $A+B = 2$ and $-3A-2B = -5$. Solving these equations yields $A = B = 1$. (Alternately, we could plug in $x = 1/3$ and $x = 1/2$ to obtain these values.) As a result,

$$f(x) = \frac{1}{1-2x} + \frac{1}{1-3x} = \sum_{n=0}^{\infty}(2x)^n + \sum_{n=0}^{\infty}(3x)^n = \sum_{n=0}^{\infty}(2^n+3^n)x^n$$

It follows that $a_n = 2^n + 3^n$ for all $n \in \mathbb{N}$. $\square$

The study of generating functions is surprisingly beautiful and provides a concrete link between discrete mathematics and techniques of analysis. However, in the interest of time and space, we have only presented the basics. For more details, you may be interested to check out the book *generatingfunctionology* (it's all one word!) by Herbert Wilf: the book is available for free at his website.

10.7 Exercises

1. Use Lagrange form to prove Theorem 10.39.

In Exercises 2 through 7, a real function f of the variable x is specified. Find the Taylor series Tf for this function centered at 0, and determine which values of $x \in \mathbb{R}$ satisfy $Tf(x) = f(x)$.

2. $f(x) = \dfrac{e^x + e^{-x}}{2}$

3. $f(x) = \begin{cases} \dfrac{\sin x}{x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$

4. $f(x) = a^x$ where $a > 0$

5. $f(x) = \sin^2 x$
(Hint: Use a half-angle identity.)

6. $f(x) = \dfrac{1}{1 + 4x^2}$

7. $f(x) = \dfrac{2x}{(2 - x)^2}$

8. Prove that the Taylor series for $\arctan x$ converges to $\arctan x$ when $|x| = 1$. (Hint: Use an argument like the one from Example 10.42, but use the results from Example 8.24 instead of from Exercise 8.5.20.)

In Exercises 9 through 14, a series is given. Use power series methods to obtain the value of the sum.

9. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n!}$

10. $\displaystyle\sum_{n=0}^{\infty} \frac{2^n}{(n+3)!}$

11. $\displaystyle\sum_{n=0}^{\infty} \frac{(-2)^n}{(2n+1)!}$

12. $\displaystyle\sum_{n=3}^{\infty} \frac{(-2)^{2n+1}}{5^n\,(2n+1)}$

13. $\displaystyle\sum_{n=2}^{\infty} \frac{n(-\pi^2)^n}{(2n)!}$

14. $\displaystyle\sum_{n=0}^{\infty} \frac{1}{(2n+1)!}$
(Hint: Combine e^x and e^{-x} somehow.)

15. Define $f : (-1, 1) \to \mathbb{R}$ by $f(x) = x \log(1 + x^2)$ for all $x \in (-1, 1)$. For each $n \in \mathbb{N}$, compute $f^{(n)}(0)$. (Hint: See Example 10.40.)

In Exercises 16 through 19, a differential equation is specified for y in terms of x. First, find the solutions to the differential equation in the form of a power series, then find the closed-form sum of the power series.

16. $y'(x) = 2xy(x)$

17. $y'(x) - y(x) = x$

18. $(1 - x)y'(x) = y(x)$

19. $y'(x) - y(x) = e^x$

20. Find all solutions to the differential equation

$$y''(x) + xy'(x) - y(x) = 0$$

satisfying $y'(0) = 0$. Leave the answer in the form of a series (so do not find a closed-form for the sum). Your answer should involve one unknown constant.

21. Find the solution to the differential equation

$$(1 - x^2)y''(x) - 2xy'(x) + 6y(x) = 0$$

which satisfies $y(0) = 1$ and $y'(0) = 0$.

22. Find all solutions to the differential equation

$$y'''(x) = y(x)$$

but leave the answer in the form of a series. (In other words, do not find a closed-form value for the sum.) Your answer should involve three unknown constants, since this is a third-order differential equation.

23. Consider the differential equation

$$x^2y''(x) = y(x)$$

We previously saw that its only analytic solution was the constant zero function. In this exercise, we use a different approach for solving the equation: we guess that y has the form $y(x) = x^p$ for some $p \in \mathbb{R}$.

   (a) Find all values of $p \in \mathbb{R}$ such that the function $y(x) = x^p$ satisfies the differential equation $x^2y''(x) = y(x)$ for all $x > 0$. There should be two such values: we'll call them α and β.

   (b) Prove that for any $A, B \in \mathbb{R}$, the function

$$y(x) = Ax^\alpha + Bx^\beta$$

   satisfies our differential equation for all $x > 0$, where α and β were found in part (a). This shows that when we take two solutions to our differential equation and make a linear combination of them, we obtain another solution.

(c) Explain why x^α and x^β are not analytic on any interval of the form $(-r, r)$ for $r > 0$, where α and β come from part (a).

24. Recall from Exercise 8.2.11 that for any $\alpha \in \mathbb{R}$ and any $n \in \mathbb{N}$, we defined

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)}{n!}$$

That exercise also shows that

$$T_n\left((1+x)^\alpha; 0\right) = \sum_{i=0}^{n} \binom{\alpha}{i} x^i$$

Let $\alpha \in \mathbb{R}$ be given. In this exercise, we outline a proof of *Newton's Binomial Theorem*, which states that

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$$

for all $x \in (-1, 1)$. The series on the right is called the *binomial series* with power α.

(a) Show that the binomial series with power α has radius 1.

(b) Show that for all $x \in (-1, 1)$, the differential equation

$$y'(x) = \frac{\alpha y(x)}{1+x}$$

is satisfied when $y(x)$ is $(1+x)^\alpha$ or when $y(x)$ is the binomial series for α. (You may not assume that these two functions are the same!)

(c) Here is a uniqueness theorem for differential equations which we will not prove:

Theorem 10.48. *Let D be an open interval in $\mathbb{R}$, let P and Q be continuous functions on D, and let $a \in D$ and $b \in \mathbb{R}$ be given. Then the differential equation*

$$y'(x) = P(x)y(x) + Q(x)$$

has a unique solution y on D satisfying $y(a) = b$.

Use this theorem and part (b) to prove Newton's Binomial Theorem.

25. The differential equation

$$x^2y''(x) + xy'(x) + (x^2 - k^2)y(x) = 0$$

is called *Bessel's Equation* (k is a constant). Let's suppose $k \in \mathbb{N}$ is fixed, and suppose that y is a solution to Bessel's Equation of the form

$$y(x) = \sum_{n=0}^{\infty} c_n x^n$$

(a) Prove that $-k^2c_0 = 0$, $(1-k^2)c_1 = 0$, and $(n^2-k^2)c_n + c_{n-2} = 0$ for all $n \in \mathbb{N}$ with $n \geq 2$.

(b) Use part (a) to prove that $c_n = 0$ whenever $n \in \mathbb{N}$ and $n < k$.

(c) Use part (a) to prove that for all $n \in \mathbb{N}$,

$$c_{2n+k} = \frac{(-1)^n\ k!\ c_k}{2^{2n}\ n!\ (n+k)!}$$

(d) From parts (b) and (c), it follows that a solution to Bessel's Equation is uniquely determined by the values of c_k and c_{k+1}. Show that one solution to Bessel's Equation is

$$y(x) = \sum_{n=0}^{\infty} (-1)^n \frac{(x/2)^{2n+k}}{n!\ (n+k)!}$$

(In other words, find out which choices of c_k and c_{k+1} produce this formula.)

26. For any $k \in \mathbb{N}$, the *Bessel function* J_k of order k is defined by

$$J_k(x) = \sum_{n=0}^{\infty} (-1)^n \frac{(x/2)^{2n+k}}{n!\ (n+k)!}$$

Among other uses, the Bessel functions appear when studying wave propogation and signal processing.[13] The Bessel function J_k also solves Bessel's Equation, as seen in Exercise 10.7.25. In this exercise, we develop a few more properties of J_k.

[13] See http://en.wikipedia.org/wiki/Bessel_function for more details.

(a) Prove that J_k has infinite radius of convergence for each $k \in \mathbb{N}$.
(b) Prove that for all $x \in \mathbb{R}$ and $k \in \mathbb{N}^*$, $(x^k J_k(x))' = x^k J_{k-1}(x)$.
(c) Prove that for all $x \in \mathbb{R}$ and $k \in \mathbb{N}$, $(x^{-k} J_k(x))' = -x^{-k} J_{k+1}(x)$.

27. Suppose that $(a_n)_{n=0}^\infty$ satisfies the recurrence

$$a_{n+2} = 5a_{n+1} - 6a_n \text{ for } n \in \mathbb{N}$$

(this is the same recurrence as in Example 10.47), but a_0 and a_1 are unspecified. Use generating functions to find the solution to this recurrence. (Your final answer should express a_n in a closed form using a_0, a_1, and n.)

28. The *Fibonacci numbers* $(f_n)_{n=0}^\infty$ are defined by the recurrence

$$f_{n+2} = f_{n+1} + f_n \text{ for } n \in \mathbb{N} \qquad f_0 = 0, f_1 = 1$$

Use generating functions to find a closed form for f_n in terms of n.

29. Find the unique sequence $(a_n)_{n=0}^\infty$ satisfying the conditions

$$a_{n+2} = 4a_{n+1} - 4a_n \qquad a_0 = 3, a_1 = 8$$

(Hint: Near the end of the problem, recall that $1/(1-x)^2$ is the derivative of $1/(1-x)$.)

10.8 Weierstrass's Function

Previously, we've seen that a uniform limit of continuous functions is continuous, but a uniform limit of differentiable functions might not be differentiable. In this section, we take these ideas to the extreme: we make a function, due to Karl Weierstrass, which is everywhere continuous but nowhere differentiable![14] Intuitively, the graph of such a function is connected (because of the continuity) but it is far from being smooth: there are jagged corners everywhere. For instance, see Figure 10.7.

Before we show how this function is built, it is worth making a few remarks about the nature of the function itself. We've seen examples of nowhere differentiable functions before, like the ruler function, but these examples have

[14]Historically, Weierstrass was not the first to create a continuous nowhere-differentiable function, but it is believed that his function was the first one published.

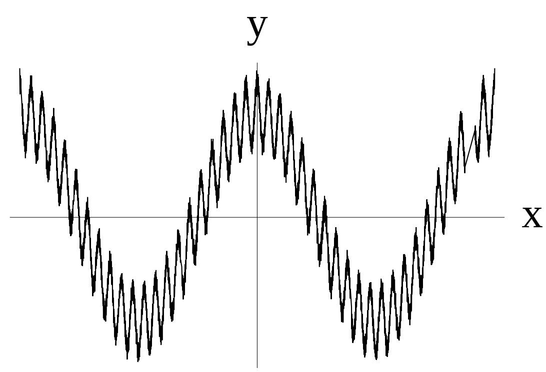

Figure 10.7: A graph of a continuous nowhere-differentiable function

plenty of discontinuities. (For instance, the ruler function is discontinuous at rational points.) Furthermore, when we introduced some functions which are continuous but not differentiable everywhere, like the absolute-value function, we saw these functions only have a few "sharp corners" or "vertical tangents" every so often. It is not too hard to imagine a continuous function with infinitely many corners (for instance, a periodic function whose graph consists of triangles), but it is difficult to imagine how one could build a function with a sharp turn or vertical tangent at EVERY point. The fact that such a function exists is highly surprising indeed![15]

A Short Introduction to Fourier Series

How might we build a continuous nowhere-differentiable function? The main idea is to build this function as a uniform series of continuous functions, which ensures that our function is continuous. However, we cannot use power series to make a continuous nowhere-differentiable function, since a power series is differentiable inside its interval of convergence. Instead, we use *Fourier series*: a typical Fourier series has the form

$$a_0 + \sum_{n=0}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$$

where $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are sequences of real constants.

[15]In fact, there are many of these functions, as it can be shown (using a result called the Baire Category Theorem) that every continuous bounded function can be written as a uniform limit of continuous nowhere-differentiable functions!

Basically, a Fourier series uses cosines and sines instead of powers of x. The partial sums of a Fourier series are called *trigonometric polynomials*. Note as well that each trigonometric function in this series is periodic with period 2π (the shortest period for the functions $\cos(nx)$ and $\sin(nx)$ is $2\pi/n$, which divides 2π evenly). It follows that each trigonometric polynomial has period 2π, so our series is periodic with period 2π as well. Thus, Fourier series are used for studying periodic functions, and in fact they turn out to be a very useful tool when studying waves and partial differential equations.

One way to think of a Fourier series is to think of each $\cos(nx)$ or $\sin(nx)$ term as a simple wave or a signal at a certain frequency. When n is larger, the wave oscillates more tightly. Multiplying the wave by a constant, like a_n or b_n, changes the amplitude of the wave. (If the constant is negative, then the wave is reflected vertically as well.) When we add these amplified waves together to make a trigonometric polynomial, we are layering the waves on top of each other; in physics, this is called superposition. When we add many different signals with rapidly growing frequencies, the resulting wave can be very "noisy", leading to the jagged picture from Figure 10.7.

Due to this, Fourier series do not usually satisfy the nice differentiability properties that power series do. For instance, we've seen that the Fourier series

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n^2}$$

converges uniformly on $\mathbb{R}$ (by the M-Test with $M_n = 1/n^2$), and hence this series is continuous on $\mathbb{R}$, but this series cannot be differentiated term by term: the Fourier series

$$\sum_{n=1}^{\infty} \left(\frac{\sin(nx)}{n^2}\right)' = \sum_{n=1}^{\infty} \frac{\cos(nx)}{n}$$

diverges at $x = 0$. (In fact, whenever x is not an integer multiple of 2π, this differentiated series converges conditionally, but the proof uses techniques we do not have time to cover.)

There are some special cases in which Fourier series can be differentiated term by term. As an example, consider the series

$$f(x) = \sum_{n=1}^{\infty} \frac{\sin(nx)}{n^3}$$

and let g be its term-by-term derivative

$$g(x) = \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^2}$$

It is not hard to show that f and g converge uniformly on $\mathbb{R}$. As a result, for any $a, x \in \mathbb{R}$, we may integrate g term by term and obtain

$$\int_a^x g(t)\,dt = \sum_{n=1}^{\infty} \int_a^x \frac{\cos(nt)}{n^2}\,dt = \sum_{n=1}^{\infty} \left(\frac{\sin(nx)}{n^3} - \frac{\sin(na)}{n^3} \right) = f(x) - f(a)$$

By the Fundamental Theorem of Calculus, it follows that $f(x) - f(a)$ is differentiable at x and $f'(x) = g(x)$.

This argument can be generalized to show that when f is a Fourier series whose term-by-term derivative g converges uniformly, it follows that $f' = g$. (Also, see Exercise 10.3.17.) However, when the term-by-term derivative g does NOT converge uniformly, that does not necessarily imply that f fails to be differentiable. Nevertheless, this does suggest that to make a continuous nowhere-differentiable function, we could make a Fourier series which converges absolutely but for which the term-by-term derivative diverges. This is the idea we pursue next.

The Main Result

Based on our recent discussion of properties of Fourier series, Weierstrass considered the following definition for a function, where a and b are positive constants:

$$W(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

(The choice of πx instead of x in the series makes things a little more convenient, since the period of $\cos(\pi x)$ is 2, which is an integer.) In essence, W is a superposition of infinitely many waves, but the frequences are growing exponentially (as opposed to growing linearly in many other Fourier series), so W should become jagged quite quickly if b is large.

More precisely, let's analyze W and its term-by-term derivative. Since

$$|a^n \cos(b^n \pi x)| \le a^n$$

for each $n \in \mathbb{N}$ and each $x \in \mathbb{R}$, the M-Test shows that W converges uniformly and is continuous when $a < 1$. Also, W has term-by-term derivative

$$-\pi \sum_{n=0}^{\infty} (ab)^n \sin(b^n \pi x)$$

In order to make this series diverge, b should be large enough to satisfy $ab \geq 1$. In fact, Weierstrass showed that W is nowhere differentiable when a and b satisfy slightly stronger conditions:[16]

Theorem 10.49. *Let $a, b \in (0, \infty)$ be given such that $a < 1$, $ab > 1+(3\pi/2)$, and b is an odd integer. Define the Fourier series W by*

$$W(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

Then $W(x)$ converges for all $x \in \mathbb{R}$, W is continuous on $\mathbb{R}$, and $W'(x)$ does not exist for any $x \in \mathbb{R}$.

Strategy. We've already shown that W is well-defined and continuous on $\mathbb{R}$. It remains to show that W is nowhere-differentiable. This proof is definitively harder than most of the other proofs in this book, as it uses a few clever tricks that we have not used much before. There are also some rather long calculations. For these reasons, we will not try and explain how someone could have created this proof in the first place, but we will merely present an outline of the main steps of the argument.[17]

Our goal is to take an arbitrary $x_0 \in \mathbb{R}$ and show that this limit of difference quotients does not exist:

$$\lim_{y \to x_0} \frac{W(y) - W(x)}{y - x_0} = \lim_{y \to x_0} \sum_{n=0}^{\infty} a^n \frac{\cos(b^n \pi y) - \cos(b^n \pi x_0)}{y - x_0}$$

In fact, we will show that this limit fails to exist in a rather special way: the difference quotients are unbounded above and below as $y \to x_0$. To do this,

[16] In fact, G.H. Hardy proved years later that the conditions $ab \geq 1$ and $0 < a < 1$ imply that W is nowhere differentiable, but that proof is beyond the scope of this textbook.

[17] This argument is due to Johan Thim, from his Masters Thesis at http://www.scribd.com/doc/29256978/Continuous-Nowhere-Differentiable-Functions-Johan-Thim. According to Thim, this argument is close to Weierstrass's original proof.

we will find sequences $(y_m)_{m=0}^\infty$ and $(z_m)_{m=0}^\infty$ such that x_0 is not in $\{y_m\}_{m=0}^\infty$ or $\{z_m\}_{m=0}^\infty$, y_m and z_m tend to x_0 as $m \to \infty$,

$$\lim_{m\to\infty} \left| \frac{W(y_m) - W(x_0)}{y_m - x_0} \right| = \lim_{m\to\infty} \left| \frac{W(z_m) - W(x_0)}{z_m - x_0} \right| = \infty$$

and the difference quotients at y_m and at z_m have opposite signs. In essence, these sequences show that the difference quotients oscillate wildly "from $-\infty$ to ∞" as y_m and z_m get closer to x_0.

How do we find these points y_m and z_m? We look at the m^{th} wave in the series: $a^m \cos(b^m \pi x)$. This wave has period $2/b^m$, and the point x_0 lies between two local maxima (the "crests" of the wave), as well as lying between two local minima (the "troughs" of the wave); see Figure 10.8. We want to pick y_m and z_m to be local extrema so that the difference quotients of y_m and z_m have opposite signs and are as steep as possible. Hence, the proof finds the closest extremum point to x_0 (marked with a hollow dot in the figure) and then chooses y_m and z_m to be the extrema a half-period away in either direction. It follows that $y_m < x_0 < z_m$, with y_m and z_m being at most one period away from x_0, so y_m and z_m approach x_0 as $m \to \infty$.

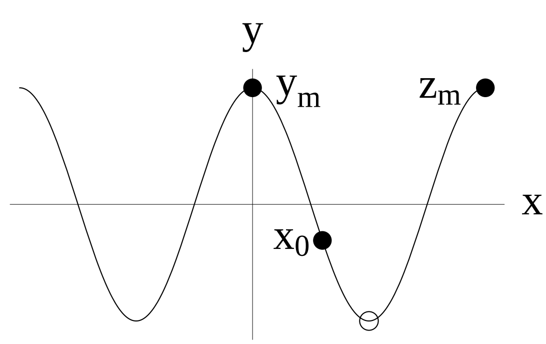

Figure 10.8: The m^{th} wave $a^m \cos(b^m \pi x)$ with x_0 and some extrema labelled

Now, we analyze the difference quotient at y_m. We've seen that this difference quotient is an infinite series. The next trick of the argument is to split that series into the terms with $n < m$ and the terms with $n \geq m$; in essence, we study the "low-frequency waves" (with $n < m$) and the "high-frequency waves" (with $n \geq m$) separately. This means that $(W(y_m) -$

$W(x_0))/(y_m - x_0)$ is $S_{\text{low}}(y_m, m) + S_{\text{high}}(y_m, m)$, where

$$S_{\text{low}}(x, m) = \sum_{n=0}^{m-1} a^n \frac{\cos(b^n \pi x) - \cos(b^n \pi x_0)}{x - x_0}$$

and

$$S_{\text{high}}(x, m) = \sum_{n=m}^{\infty} a^n \frac{\cos(b^n \pi x) - \cos(b^n \pi x_0)}{x - x_0}$$

Intuitively, we expect S_{low} to be small and S_{high} to be large, as the sharpest slopes should come from waves that oscillate tightly. We use some inequalities involving trigonometric identities, using the assumptions that b is an odd integer and $ab > 1+(3\pi/2)$. Eventually, we find that $|S_{\text{low}}(y_m, m)+S_{\text{high}}(y_m, m)|$ is proportional to $(ab)^m$, which shows that it approaches ∞ and $m \to \infty$. The same steps work for z_m, though we find that the difference quotient at z_m has the opposite sign from the difference quotient at y_m.

Proof. Let a, b, W be given as described. The remarks before the statement of the theorem prove that W is continuous on $\mathbb{R}$. Now, let $x_0 \in \mathbb{R}$ be given, and we will show that $W'(x_0)$ does not exist. More specifically, we will create sequences $(y_m)_{m=0}^{\infty}$ and $(z_m)_{m=0}^{\infty}$ of points such that

1. For all $m \in \mathbb{N}$, $y_m < x_0 < z_m$.
2. $y_m \to x_0$ and $z_m \to x_0$ as $m \to \infty$.
3. The difference quotients

$$\frac{W(y_m) - W(x_0)}{y_m - x_0} \text{ and } \frac{W(z_m) - W(x_0)}{z_m - x_0}$$

have opposite signs and have magnitudes at least as large as $C(ab)^m$, where C is some positive constant independent of m.

It follows from conditions 2 and 3 of this list that W has unbounded difference quotients near x_0 and hence is not differentiable at x_0.

To construct these sequences, let $m \in \mathbb{N}$ be given. Let α_m be the nearest integer to $b^m x_0$; more specifically, choose α_m to be the unique integer satisfying $b^m x_0 - \alpha_m \in (-1/2, 1/2]$. Also, define $\Delta_m = b^m x_0 - \alpha_m$. (Note that since the graph of $y = \cos(\pi x)$ has local extrema at the integers, the m^{th} wave $a^m \cos(b^m \pi x)$ has its closest extremum to x_0 at $x = \alpha_m / b^m$.) Define

$$y_m = \frac{\alpha_m - 1}{b^m} \qquad z_m = \frac{\alpha_m + 1}{b^m}$$

It follows that

$$y_m - x_0 = -\frac{1 + \Delta_m}{b^m} \text{ and } z_m - x_0 = \frac{1 - \Delta_m}{b^m}$$

Since $|\Delta_m| \leq 1/2$, it follows that $y_m < x_0 < z_m$ and that $y_m, z_m \to x_0$ as $m \to \infty$. This proves conditions 1 and 2 from our conditions on y_m and z_m.

It remains to prove condition 3. First, for any $x \in \mathbb{R}$, define

$$S_{\text{low}}(x, m) = \sum_{n=0}^{m-1} a^n \frac{\cos(b^n \pi x) - \cos(b^n \pi x_0)}{x - x_0}$$

and

$$\begin{aligned} S_{\text{high}}(x, m) &= \sum_{n=m}^{\infty} a^n \frac{\cos(b^n \pi x) - \cos(b^n \pi x_0)}{x - x_0} \\ &= \sum_{n=0}^{\infty} a^{m+n} \frac{\cos(b^{m+n} \pi x) - \cos(b^{m+n} \pi x_0)}{x - x_0} \end{aligned}$$

where we do an index shift in the last sum for convenience. (It is straightforward to see that these series converge absolutely for all $x \neq x_0$ by the Comparison Test.) Therefore, we have

$$\begin{aligned} \frac{W(y_m) - W(x_0)}{y_m - x_0} &= \sum_{n=0}^{\infty} a^n \frac{\cos(b^n \pi y_m) - \cos(b^n \pi x_0)}{y_m - x_0} \\ &= S_{\text{low}}(y_m, m) + S_{\text{high}}(y_m, m) \end{aligned}$$

and a similar statement holds for the difference quotient at z_m.

Next, we analyze S_{low}. For any $x \in \mathbb{R} - \{x_0\}$ and any $n \in \mathbb{N}$, applying the Mean Value Theorem to cos on the interval from $b^n \pi x$ to $b^n \pi x_0$ yields

$$\cos(b^n \pi x) - \cos(b^n \pi x_0) = b^n \pi (x - x_0) \sin(c)$$

for some c between $b^n\pi x$ and $b^n\pi x_0$. Thus, $|\cos(b^n\pi x) - \cos(b^n\pi x_0)| \leq b^n\pi|x - x_0|$, and hence

$$|S_{\text{low}}(x,m)| \leq \sum_{n=0}^{m-1} a^n \frac{b^n\pi|x - x_0|}{|x - x_0|} = \sum_{n=0}^{m-1} \pi(ab)^n$$

Since $ab > 1 + (3\pi/2)$, $ab \neq 1$, so we may use the partial sum formula for a geometric series to write

$$|S_{\text{low}}(x,m)| \leq \frac{\pi((ab)^m - 1)}{ab - 1} \leq \frac{\pi(ab)^m}{ab - 1}$$

Now, we analyze S_{high}. Let $n \in \mathbb{N}$ be given. Recall that b is an odd integer, so b^n is also an odd integer. Also, note that $\cos(\pi x) = (-1)^x$ when x is an integer. Therefore,

$$\begin{aligned}
\cos(b^{m+n}\pi y_m) &= \cos\left(b^{m+n}\pi\frac{\alpha_m - 1}{b^m}\right) \\
&= \cos\left(b^n\pi(\alpha_m - 1)\right) = (-1)^{(b^n(\alpha_m - 1))} \\
&= \left((-1)^{(b^n)}\right)^{\alpha_m - 1} = (-1)^{\alpha_m - 1} = -(-1)^{\alpha_m}
\end{aligned}$$

The same calculations with z_m show that $\cos(b^{m+n}\pi z_m) = (-1)^{\alpha_m + 1} = -(-1)^{\alpha_m}$. Next, we rewrite $\cos(b^{m+n}\pi x_0)$ by using a trigonometric identity, the facts that $\cos(\pi x) = (-1)^x$ and $\sin(\pi x) = 0$ when x is an integer, and the definitions of y_m, z_m, α_m, and Δ_m (recalling that α_m is an integer):

$$\begin{aligned}
\cos(b^{m+n}\pi x_0) &= \cos\left(b^{m+n}\pi\frac{\alpha_m + \Delta_m}{b^m}\right) \\
&= \cos(b^n\pi(\alpha_m + \Delta_m)) \\
&= \cos(b^n\pi\alpha_m)\cos(b^n\pi\Delta_m) - \sin(b^n\pi\alpha_m)\sin(b^n\pi\Delta_m) \\
&= \left((-1)^{(b^n)}\right)^{\alpha_m}\cos(b^n\pi\Delta_m) - 0 \cdot \sin(b^n\pi\Delta_m) \\
&= (-1)^{\alpha_m}\cos(b^n\pi\Delta_m)
\end{aligned}$$

Using the definition of y_m and the previous results, we obtain

$$\begin{aligned}
S_{\text{high}}(y_m, m) &= \sum_{n=0}^{\infty} a^{m+n}\frac{-(-1)^{\alpha_m} - (-1)^{\alpha_m}\cos(b^n\pi\Delta_m)}{-(1 + \Delta_m)/b^m} \\
&= (ab)^m(-1)^{\alpha_m}\sum_{n=0}^{\infty} a^n\frac{1 + \cos(b^n\pi\Delta_m)}{1 + \Delta_m}
\end{aligned}$$

Since $|\cos(b^n\pi\Delta_m)| \le 1$, each term in this sum is nonnegative. Thus, we get a lower bound on that sum by throwing out all terms except for the term with $n = 0$. Also, since $|\Delta_m| \le 1/2$, it follows that $\cos(\pi\Delta_m) \ge 0$, so

$$\sum_{n=0}^{\infty} a^n \frac{1+\cos(b^n\pi\Delta_m)}{1+\Delta_m} \ge \frac{1+\cos(\pi\Delta_m)}{1+\Delta_m} \ge \frac{1}{1+1/2} = \frac{2}{3}$$

Similarly, we find that

$$\begin{aligned} S_{\text{high}}(z_m, m) &= \sum_{n=0}^{\infty} a^{m+n} \frac{-(-1)^{\alpha_m} - (-1)^{\alpha_m}\cos(b^n\pi\Delta_m)}{(1-\Delta_m)/b^m} \\ &= -(ab)^m(-1)^{\alpha_m} \sum_{n=0}^{\infty} a^n \frac{1+\cos(b^n\pi\Delta_m)}{1-\Delta_m} \end{aligned}$$

and

$$\sum_{n=0}^{\infty} a^n \frac{1+\cos(b^n\pi\Delta_m)}{1-\Delta_m} \ge \frac{1+\cos(\pi\Delta_m)}{1-\Delta_m} \ge \frac{1}{1-(-1/2)} = \frac{2}{3}$$

Finally, we collect all these results together. To make it easier to combine results, let's write our inequalities as equalities with extra variables. More specifically, our work with S_{low} shows that there exist $\epsilon_1, \epsilon_2 \in [-1, 1]$ satisfying

$$S_{\text{low}}(y_m, m) = \epsilon_1 \frac{\pi(ab)^m}{ab-1} \qquad S_{\text{low}}(z_m, m) = \epsilon_2 \frac{\pi(ab)^m}{ab-1}$$

Also, we can find $\eta_1, \eta_2 \in [1, \infty)$ satisyfing

$$\sum_{n=0}^{\infty} a^n \frac{1+\cos(b^n\pi\Delta_m)}{1+\Delta} = \eta_1 \cdot \frac{2}{3} \qquad \sum_{n=0}^{\infty} a^n \frac{1+\cos(b^n\pi\Delta_m)}{1-\Delta} = \eta_2 \cdot \frac{2}{3}$$

Therefore, we find that

$$\begin{aligned} \frac{W(y_m) - W(x_0)}{y_m - x_0} &= S_{\text{low}}(y_m, m) + S_{\text{high}}(y_m, m) \\ &= \frac{\epsilon_1\pi(ab)^m}{ab-1} + (-1)^{\alpha_m}(ab)^m \frac{2\eta_1}{3} \\ &= (-1)^{\alpha_m}(ab)^m \left(\frac{2\eta_1}{3} + (-1)^{\alpha_m}\frac{\epsilon_1\pi}{ab-1}\right) \end{aligned}$$

and similarly

$$\frac{W(z_m) - W(x_0)}{z_m - x_0} = -(-1)^{\alpha_m}(ab)^m \left(\frac{2\eta_2}{3} - (-1)^{\alpha_m} \frac{\epsilon_1 \pi}{ab - 1} \right)$$

Now, because $ab > 1 + (3\pi/2)$, we have $\pi/(ab - 1) < 2/3$. Thus, in our expressions for the difference quotients at y_m and at z_m, the number inside the large parentheses is at least $2/3 - \pi/(ab - 1)$, which is positive. It follows that the two difference quotients have opposite signs. Furthermore, when $C = 2/3 - \pi/(ab - 1)$, our difference quotients have magnitude at least $C(ab)^m$. Finally, we have proven condition 3 from our conditions on y_m and z_m, so we are done. ■

Remark. The proof we just finished uses two sequences of points, approaching x_0 from different sides, to show that the difference quotients are unbounded. It turns out that we only need one of these sequences if we want to prove that the difference quotients are unbounded. However, note that we don't actually know the sign of either of our difference quotients, because we don't know whether α_m is even or odd. Hence, if we worked with only one sequence of points, we wouldn't know if the difference quotients go to ∞, go to $-\infty$, or oscillate between the two infinities. (Note that if the difference quotients go to ∞ or go to $-\infty$, then we have a vertical tangent at x_0.)

In contrast, by working with two difference quotients of opposite signs, we show that the difference quotients are unbounded above AND below. This tells us that we do not have a vertical tangent at x_0; instead, secant slopes oscillate extremely wildly near x_0. In fact, this construction tells us that W has another nonintuitive property:

Corollary 10.50. *Let a, b, W be given as in Theorem 10.49. Then W is not monotone on any nonempty open interval.*

The proof of this corollary is quite simple. Given any nonempty open interval I, take a point $x_0 \in I$, and construct the sequences (y_m) and (z_m) from the proof of Theorem 10.49. Because y_m and z_m approach x_0 as $m \to \infty$, when m is large enough, y_m and z_m also belong to I. Since the difference quotient from y_m to x_0 and the difference quotient from x_0 to z_m have different signs, it follows that W is not monotone on $[y_m, z_m]$ and hence is not monotone on I.

Corollary 10.50 says that the Weierstrass function W is nowhere monotone. In contrast, most of the continuous functions we studied before are piecewise monotone. Intuitively, most continuous functions "change direction" only finitely many times, but W changes direction all the time!

Concluding Remarks

The Weierstrass function is the last example this book covers. In many ways, it is very deserving of this honor, since in order to introduce it and analyze it, we use many different tools and techniques from this book. In the process, we also encounter many important reasons why people study analysis, as opposed to merely studying the formulas of calculus. These reasons include:

- *Analysis rigorously justifies results.* The theorems of calculus are quite powerful and beautiful, but much like any powerful tool, we need control to use the tool properly and purposefully. By writing careful, precise definitions, and using rigorous proofs of results, we get control and comfort in using our tools of calculus effectively.

- Although many results from calculus arise from natural observations and intuition, *there are plenty of important examples which are surprising and non-intuitive.* When dealing with such examples, we need the rigor of analysis to be able to make sense out of them. These examples challenge our conventional ways of thinking about calculus, leading to a deeper understanding. If used correctly, rigor does not detract from intuition: it ultimately enhances and refines intuition.

- For many students, the main work in mathematics prior to analysis uses equations and exact solutions. However, *analysis also considers inequalities and error estimates very important.* For instance, Taylor's Theorem is particularly fundamental because it not only asserts that Taylor polynomials are generally good approximations, it also makes specific statements about HOW good they are. As another example, in our proof that the Weierstrass function is nowhere differentiable, we use complicated inequalities to obtain our main results.

Even outside of analysis, there is much to be gained by pursuing mathematics with care and detail. However, being rigorous does not mean that examples are "uninteresting" or "too concrete". After all, definitions and theorems are only useful when they model examples, address questions motivated by prior experience, or provide more complete pictures as to how an area of study works. There is a cycle at work here: we use examples and questions to motivate definitions, we proceed from these definitions to theorems and proofs, and then we use the theorems to create new examples to start the cycle over again.

It is worth noting that different areas of mathematics tend to use different proof styles and have different conventions. This causes some confusion over standards, but with practice that confusion tends to become negligible. Ultimately, mathematicians from different areas devise results that target the questions they find interesting. This doesn't mean that these different areas are incompatible: there are many important results that solve a problem in one area of mathematics by converting it to an equivalent problem in another area. (Generating functions are a particularly good example of this.)

Ultimately, I hope that you, as the student, will take some of these lessons to heart in your future mathematics education. There is a vast amount of mathematics to explore, and you should be prepared for it upon finishing this book. You will be able to explore more beautiful and complex results with careful reasoning and structuring as your guides.

Index