The Odds and Ends of Proving: An Introduction to Mathematical Thinking

Keith Ashton Nabb

Contents

List of Figures

List of Tables

Dedication

This book is dedicated to all of those who have made a difference in my life. Personally, I thank my family−especially my dear wife Megan who offered encouragement and support at every stage of this work. Professionally, I thank the mathematics faculty at Rhode Island College and Texas Tech University for their counsel and expertise. Specifically, thanks go to Dave Abrahamson, Edward Allen, James Sedlock, Victor Shubov and the late Arthur Smith; each has been enormously influential in shaping my mathematical views. My colleagues and students at Moraine Valley have also been a continuous source of inspiration. Finally, I tip my hat to the faculty in the department of Mathematics and Science Education at the Illinois Institute of Technology; they have challenged me to question what mathematics is all about.

Note To the Reader

The thought of reading a mathematical proof for *understanding* can be an intimidating and sometimes unpleasant experience for the first time student. The general feeling may be captured in the simple question, "What if I don't get it?" Likewise, constructing/writing one's own proofs generates questions such as, "Where do I start?" and "Once I get somewhere, what do I do next?" If you have ever asked yourself these questions, I hope this book will lead you to some of the answers.

In my experience teaching at the community college level, the uncertainties above typify the feelings of the majority of students I have had. Unlike four-year institutions which offer their clientele a course on mathematical proof, two-year colleges carry the burden of teaching students specific content (e.g., calculus, differential equations or linear algebra) while incorporating unfamiliar methods of proof (e.g., mathematical induction or proof by contradiction). I have noticed that students place a premium on understanding the "proof trade," often at the expense of mathematical content. So if students spend so much energy on the proof that they miss the heart of the theorem, have we done them any favors?

My purpose in writing this book is the following. I envision that instructors at two-year colleges and high schools may find it a useful supplement for students having no previous exposure to reading/writing proofs. Because the prerequisite/corequisite is nothing more than Calculus, the choice of topics is restrictive. However, I feel the selection is broad enough to offer an authentic first taste of abstract mathematics.

Throughout this book I have taken care in using standard notation while avoiding unnecessary jargon. I dedicate the opening chapter to what I feel are some of the most important things a mathematics student should know−the "hammer and nails" of mathematics. It is from these tools which deeper, more abstract mathematics evolves. As you reach the core of the text, you will find that I have chosen topics that I feel illustrate these tools rather well. However, I have not written a book on set theory or analysis. Instead, I have tried to introduce several facets from these disciplines without getting too deeply immersed into any particular theory. An analyst at heart may find this unforgivable but this is what I feel my students need. Let the more advanced material find them when they are ready.

Finally, a brief comment about mathematics as "art and science" as this typifies my philosophy on mathematical learning. It is unfortunate that mathematics is widely conceived as a science in which answers are right or wrong. Although one could argue in favor of this view, *there is much in mathematics that is not sanctioned by this belief.* As mathematicians, teachers, and students, we should equally embrace the artistic side of mathematics. As students begin their journey here, they will undoubtedly see both sides of the coin. Conventions play a critical role in mathematics but users of the subject will need to develop styles they can call their own.

In closing, never forget that the pages of this book were written with the student in mind. I have made many efforts to include as much detail as possible but not so much as to hinder the flow of continuity. Should you have suggestions for improvement or errors to point out, please do not hesitate to contact me. I hope you find as much pleasure reading this book as I have found in writing it.

Keith Nabb
nabb@morainevalley.edu

Chapter 0

Notation

Throughout this book, we shall use the following notation.

$$\begin{aligned}
\mathbb{N} &= \{1, 2, 3, 4, \ldots\} = \text{the natural numbers} \\
\mathbb{W} &= \{0, 1, 2, 3, 4, \ldots\} = \text{the whole numbers} \\
\mathbb{Z} &= \{\ldots, -2, -1, 0, 1, 2, \ldots\} = \text{the integers} \\
\mathbb{Q} &= \text{the rational numbers} \\
\mathbb{Q}' &= \text{the irrational numbers} \\
\mathbb{R} &= \text{the real numbers}
\end{aligned}$$

Notice that unlike $\mathbb{N}$, $\mathbb{W}$, or $\mathbb{Z}$, we give no "listing" of $\mathbb{Q}$, $\mathbb{Q}'$ or $\mathbb{R}$. This is due to the fact that it is easier to *describe* these numbers rather than to list them (it is actually impossible to list them). For example, rational numbers are those numbers that can be expressed as $\frac{m}{n}$, where m and n are both integers and $n \neq 0$. A nicer way of writing this is

$$\mathbb{Q} = \left\{ \frac{m}{n} \,|\, m, n \in \mathbb{Z}, n \neq 0 \right\}.$$

The braces $\{\cdot\}$ indicate that we are speaking of a *set* (loosely defined in the first part of this book). The vertical bar means "such that" and the "$\in$" symbol means "is an element of". Thus, the above reads, "$\mathbb{Q}$ is the set of all numbers of the form $\frac{m}{n}$ such that both m and n are integers, n being nonzero." Numbers such as $\frac{15}{4}, 0.\bar{3}$, and $-5.60\bar{2}$ belong to this set.

You have probably noticed that, of all numbers you can think of, not every one belongs to $\mathbb{Q}$. For example, think of the numbers $\sqrt{2}$ and π. Try (as you may) to represent them as $\frac{m}{n}$, remembering that m and n must be *integers*. It's impossible! We can *approximate* numbers such as these with great accuracy using numbers from $\mathbb{Q}$ but we cannot *represent* them with numbers from $\mathbb{Q}$. Such numbers are called irrational, denoted by $\mathbb{Q}'$.

If we visualize the rationals and irrationals together, we obtain the real numbers. Structurally, we can see these sets below.

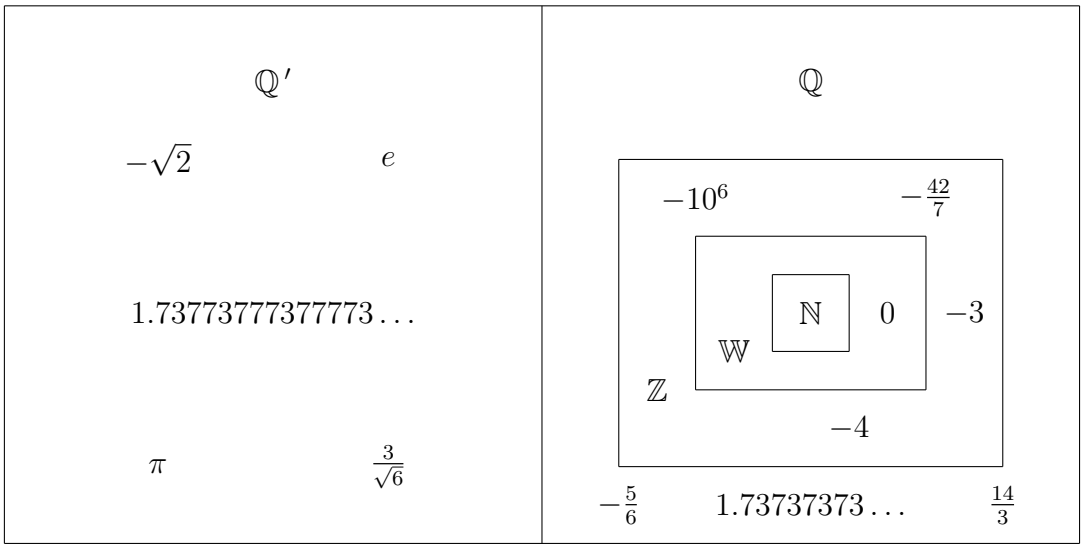

Figure 1: The Real Numbers

With these depictions, note that $\pi \in \mathbb{Q}'$ and $\pi \in \mathbb{R}$ but $\pi \notin \mathbb{Q}$. Also, we see that $\mathbb{Q} \cup \mathbb{Q}' = \mathbb{R}$. You may recall the union symbol $\cup$ from other classes; again, we will define this shortly. Additional notation we will use is the standard interval notation. If $a, b \in \mathbb{R}$ with $a < b$, then

$$\begin{aligned} [a,b] &= \{x \,|\, a \leq x \leq b\} \\ [a,b) &= \{x \,|\, a \leq x < b\} \\ (a,b] &= \{x \,|\, a < x \leq b\} \\ (a,b) &= \{x \,|\, a < x < b\}. \end{aligned}$$

The first set is regarded as *closed* (it contains the endpoints a and b) whereas the last set is said to be *open* (it contains neither a nor b). You are encouraged to construct similar "definitions" with a or b equal to $-\infty$ or ∞, respectively. Do you see that $(-\infty, \infty) = \mathbb{R}$? On another note, there may be some confusion when writing $[-\infty, \infty]$ although this is generally regarded as the *extended* real number system, denoted by $\mathbb{R}^e$. That is, $\mathbb{R}^e = \mathbb{R} \cup \{-\infty, \infty\}$. Loosely speaking, the extended real numbers is just the set of real numbers along with the symbols ∞ and $-\infty$, their meaning still unclear at this point. Again, these topics will be tackled later on; it is the notation with which we want you to become accustomed.

Moving on, you should be familiar with the greatest integer or "step" function

$$f(x) = \lfloor x \rfloor.$$

This relation assigns one value to each x in its domain (we are using all of these terms loosely so indeed it is a function). For each x, $\lfloor x \rfloor$ = the greatest integer *less than or equal to* x. Here are some examples:

$$\begin{aligned} \lfloor 5 \rfloor &= 5 \\ \lfloor 3.5 \rfloor &= 3 \\ \lfloor 0 \rfloor &= 0 \\ \lfloor -1.3 \rfloor &= -2 \end{aligned}$$

Note that $\lfloor 3.5 \rfloor = 3$ because $3 \leq 3.5$. Similarly, $\lfloor -1.3 \rfloor = -2$ since $-2 \leq -1.3$. Thus, schematically, the function moves all numbers in the left direction to the closest *integer* value. See the figure below.

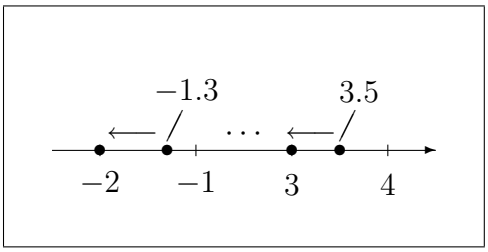

Figure 2: The Greatest Integer Function

At this point, you may wish to study the graph of this function in the $xy-$plane (see the next page). Convince yourself that this is correct. From the picture alone, we can see why it is called the "step" function.

From time to time, we will use the arrow notation "$\Rightarrow$" to indicate a direct implication. For example, we may write

$$x^2 = 4 \Rightarrow x = \pm 2$$

since the statement $x = \pm 2$ follows directly from $x^2 = 4$ (or that $x^2 = 4$ *implies* $x = \pm 2$). This notation keeps the road less congested and makes some of the presentation more readable.

Lastly, you should be familiar with dimensional extensions of the real number system. For example, we will often write $\mathbb{R}$ or $\mathbb{R}^1$ to stand for the set of real numbers. What do you think $\mathbb{R}^2$ or $\mathbb{R}^3$ means? Since $5 \in \mathbb{R}^1$ and $\pi \in \mathbb{R}^1$, you might surmise

that $(5, \pi) \in \mathbb{R}^2$. That is, $\mathbb{R}^2$ is the set of all *ordered pairs* (or you may wish to think of two-dimensional vectors) for which each entry is in $\mathbb{R}$. The definition is analogous for $\mathbb{R}^n, n \geq 3, n \in \mathbb{N}$. For example,

$$(1, 0, -1) \in \mathbb{R}^3$$

and

$$(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n,\ x_i \in \mathbb{R},\ i = 1, 2, \ldots, n.$$

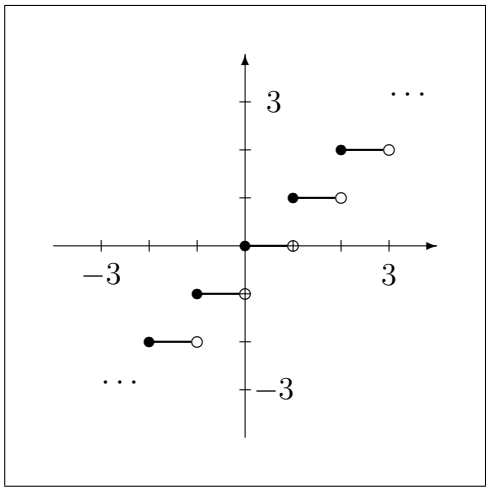

Figure 3: The Graph of $f(x) = \lfloor x \rfloor$

With these few comments alone, you are ready to move on to Chapter 1.

Chapter 1

Five Things You Need to Know Before Reading this Book!

Much of the content in this chapter should be familiar but consider all material in this section as simply indispensible knowledge. It will prove to be helpful in the study of this course as well as future courses. You should pay close attention to detail but it is more important that you leave with a knowledge of the *tools* presented here. This chapter lays a preliminary foundation for much of the material that follows.

1.1 Absolute Value

The notion of absolute value is probably something you first encountered in an elementary algebra course. Most people remember it as just being the "positive" of a number. For example, $|5| = 5$, $|-3| = 3$, etc. This, in fact, is true, but a much deeper understanding of absolute value is necessary in higher mathematics. Since absolute values are used in statements concerning limits, error bounds, and various important inequalities, one can see its immediate importance.

Precisely, the absolute value of a number x, denoted by $|x|$, is given by

$$|x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0. \end{cases}$$

At first, the definition looks rather awkward. However, it just states that if we have a nonnegative (meaning zero or positive) number x, then $|x| = x$, just whatever we started with. Hence the earlier example $|5| = 5$. On the other hand, if we have a negative number $x < 0$, then $|x| = -x$. That is, the absolute value of x equals the *opposite* of the original number. Since the original number is negative, this makes the absolute value positive. As an example, $|-3| = -(-3) = 3$. Therefore, it is okay

to think of absolute value as just the "positive" of whatever number is inside the bars $|\cdot|$. Of course, it works equally well to define the absolute value of a number as

$$|x| = \begin{cases} x, & x > 0 \\ -x, & x \leq 0 \end{cases}$$

since $|0| = \pm 0 = 0$ so we won't make a fuss here. Indeed, some textbooks take the above to be the definition.

As an alternate way to view absolute value, one can look at the graph of $y = |x|$. This should be familiar to you:

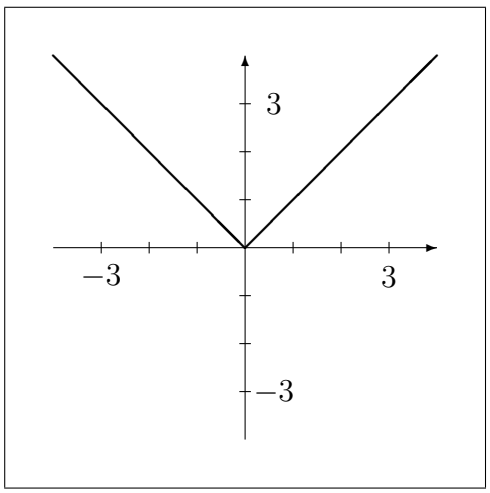

Figure 1.1: The Graph of $f(x) = |x|$

Notice that the graph lies entirely above the $x-$axis (except for the point at $x = 0$). Therefore, the function only takes on values greater than or equal to zero. That is, $y = |x| \geq 0$ as we saw earlier.

So why all the talk on absolute value? Because it quickly becomes a confusing issue for many students. You might recall the following relationships for a positive number k.

1. $|x| = k$ if and only if $x = k$ or $x = -k$.
2. $|x| \leq k$ if and only if $-k \leq x \leq k$.
3. $|x| \geq k$ if and only if $x \leq -k$ or $x \geq k$.

Remark 1.1.1 *The expression "if and only if," often abbreviated as IFF, indicates a two-way implication. For example, statement 1 is understood to convey <u>two</u> statements: (a) If $|x| = k$ then $x = k$ or $x = -k$, and (b) If $x = k$ or $x = -k$, then $|x| = k$. Analogous statements can be made for the relationships in 2 and 3.*

Similar relationships hold if the inequalities $\leq$ and $\geq$ are replaced with the strict inequalities $<$ and $>$. For further insight, we analyze statement 2 above in greater detail.

One of the best ways to think about absolute value is to view it as a "distance" function. That is, the absolute value function gives an idea about how far away something is from the origin. Thus, we can interpret the symbol $|x|$ as the "distance" from x to zero (the origin). For further emphasis, note that $|x| = |x - 0|$, so we see how both the number x and the number 0 are related to $|x|$. Sometimes we will write $|x| = |x - 0|$ or $|x| = \text{dist}(x, 0)$ for clarity. So what does $|x| \leq k$ (k positive) really mean? Well, right away we see three numbers that satisfy this inequality: $x = k, -k$ and 0. This is illustrated below.

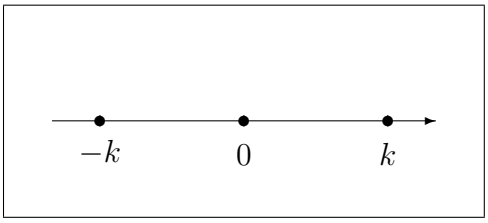

Figure 1.2: The numbers $x = -k, 0$, and k all satisfy $|x| \leq k$

Are there any other numbers that work? In other words, are there other numbers whose distance to 0 is less than or equal to k? Try $x = \frac{k}{2}$! Certainly, if $k > 0$, then $\frac{k}{2} \leq k$ so $x = \frac{k}{2}$ also works.

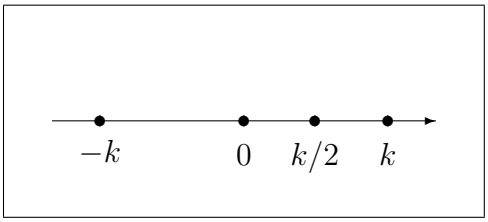

Figure 1.3: The number $x = \frac{k}{2}$ also satisfies $|x| \leq k$.

It shouldn't take long to see that $x = -\frac{k}{2}$ will also work, and, in fact, *any number* between $-k$ and k will have the property of being within k units of the origin. Thus $|x| \leq k$ is equivalent to $-k \leq x \leq k$ (see the figure on the next page). You are encouraged to give similar "verifications" for properties 1 and 3.

At this point, we give three more simple but important properties concerning the absolute value function:

1. For any two real numbers A and B, $|AB| = |A||B|$.

2. For any two real numbers A and B, $|A - B| = |B - A|$.

3. $\sqrt{A^2} = |A|$.

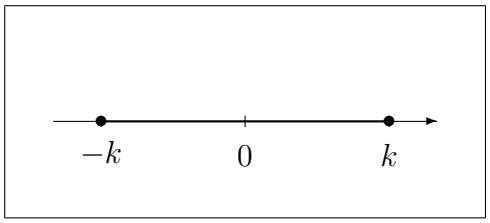

Figure 1.4: The numbers x where $|x| \leq k$

The first property seems obvious but you are encouraged to give a proof. One way to do this is to consider four separate cases: (1) $A, B \geq 0$, (2) $A \geq 0, B \leq 0$, (3) $A \leq 0, B \geq 0$, and (4) $A, B \leq 0$. We give a proof of case (3).

Proof. Let $A \leq 0$ and $B \geq 0$. Then $-A \geq 0$ so that $-AB \geq 0$. Hence

$$\begin{aligned} |A||B| &= (-A)B \\ &= -AB \\ &= |AB| \end{aligned}$$

since

$$|AB| = \begin{cases} AB, & AB \geq 0 \\ -AB, & AB \leq 0. \end{cases}$$

Here, $|AB| = -AB$ since $AB \leq 0$ follows from $-AB \geq 0$. This completes the proof (you are encouraged to look at the other cases). ■

Remark 1.1.2 *By interchanging the roles of the numbers A and B, the two cases (2) and (3) may be settled by writing a single proof.*

For the second property, we can use the first property once it has been established to be true. A proof would go something like this:

Proof. We have

$$\begin{aligned} |A-B| &= |(-1)(B-A)| \\ &= |(-1)||(B-A)| \quad \text{by property 1!} \\ &= 1 \cdot |B-A| \\ &= |B-A|. \end{aligned}$$

This completes the proof. ■

The last property is one of the most important (it is actually an algebraic *definition*). First of all, since $A^2 \geq 0$, the number i appears nowhere in the statement of the property $\sqrt{A^2} = |A|$. If we tackle a particular example, we can see why the absolute value symbol is necessary.

1. For $A = 6$, $\sqrt{6^2} = \sqrt{36} = 6$. (Notice this equals $|6|$.)
2. For $A = -6$, $\sqrt{(-6)^2} = \sqrt{36} = 6$. (Notice this equals $|-6|$.)

(Note that in the case of $\sqrt{(-6)^2}$, there is a subtle use of parentheses. If we didn't use them, $\sqrt{-6^2} = \sqrt{-36} = 6i$.) From the examples above, we see that for A either negative or positive, we take the square root of A^2 to be the *positive* number $|A|$. Thus, regardless of the sign of A, $\sqrt{A^2} = |A|$. This finding is extremely important as we will see immediately in the next section.

1.2 The Triangle Inequality

Theorem 1.2.1 *For all real numbers x and y, $|x+y| \leq |x| + |y|$.*

Proof. We know that $xy \leq |xy|$ for all $x, y \in \mathbb{R}$ since xy *could* be negative but $|xy| \geq 0$. But now we have another way of expressing $|xy|$. That is, $|xy| = \sqrt{(xy)^2} = \sqrt{x^2y^2}$. Thus, we have

$$xy \leq \sqrt{x^2y^2}.$$

Multiplying both sides by 2 and adding $x^2 + y^2$ yields

$$x^2 + 2xy + y^2 \leq x^2 + 2\sqrt{x^2y^2} + y^2,$$

or, more revealing,

$$(x+y)^2 \leq \left(\sqrt{x^2} + \sqrt{y^2}\right)^2.$$

Now, upon taking the nonnegative square root, we obtain

$$|x+y| \leq |\sqrt{x^2} + \sqrt{y^2}| = \sqrt{x^2} + \sqrt{y^2}.$$

That is,

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

so we are done here. ■

Remark 1.2.1 *Notice how many times the fact* $\sqrt{b^2} = |b|$ *was actually used in the proof above.*

Remark 1.2.2 *It is not trivial to see why the proof began the way it did! This is one of those results that is "simple to state" but "difficult to prove." An alternate (albeit tedious) route is to consider different cases for the* x *and* y *as in the sketch of the proof of* $|AB| = |A||B|$*. However, the proof presented here is a bit more elegant.*

At this point in our discussion, it is difficult to see the usefulness of such an inequality. The fact is, this inequality is indispensible to proving results related to continuity, sequences, series, and the like. We offer a few informal illustrations below.

For example, let $x = 5$ and $y = -3$. Then $|x + y| = |5 - 3| = 2$ and $|x| + |y| = 8$ so indeed $2 = |x + y| < |x| + |y| = 8$. As another example, let $x = 0$ and $y = 6$. Then $6 = |x + y| = |0 + 6| = |0| + |6| = |x| + |y| = 6$ so equality holds in this case. Try to convince yourself of its validity by producing a few more examples.

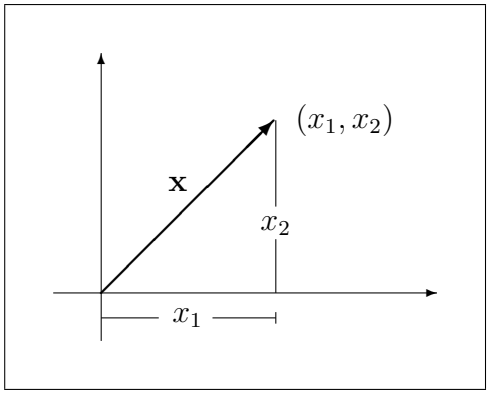

Figure 1.5: The length of $\mathbf{x}$ is $||\mathbf{x}|| = \sqrt{{x_1}^2 + {x_2}^2}$, $\mathbf{x} \in \mathbb{R}^2$.

A question that may be lurking in your mind is, "What's with the name?" The name "Triangle Inequality" is actually best seen from the point of view when x and y are two-dimensional vectors, not scalars. Recall from Calculus that a vector $\mathbf{x}$ is represented by the ordered pair of real numbers $\langle x_1, x_2 \rangle$. Likewise, let $\mathbf{y} = \langle y_1, y_2 \rangle$. Next, recall the notion of the *norm* of a vector, denoted by $||\mathbf{x}||$ and given by $||\mathbf{x}|| = \sqrt{{x_1}^2 + {x_2}^2}$. This gives the *length* or *magnitude* of the vector (it's just the Pythagorean Theorem!). See the figure above.

We should now see how to arrive at $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$. First look at the diagram below. Here, vectors $\mathbf{x}$ and $\mathbf{y}$ add to give $\mathbf{x}+\mathbf{y}$, the third side of the "triangle." Since the sum of the lengths of any two sides of a triangle always exceeds the length of the remaining side, we get $||\mathbf{x}|| + ||\mathbf{y}|| \geq ||\mathbf{x}+\mathbf{y}||$. Hence the name "triangle inequality." Furthermore, this idea generalizes to n-dimensional vectors; that is, if $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ and $\mathbf{y} = \langle y_1, \ldots, y_n \rangle$ with $||\mathbf{x}|| = \sqrt{\sum_{i=1}^{n} {x_i}^2}$ and $||\mathbf{y}|| = \sqrt{\sum_{i=1}^{n} {y_i}^2}$, then indeed $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$. We will prove this in just a moment. First, it is necessary to recall two ideas from Calculus. For $\mathbf{x}$ and $\mathbf{y}$ vectors in $\mathbb{R}^n$,

1. The dot product $\mathbf{x} \cdot \mathbf{y}$ is given by
$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= \sum_{i=1}^{n} x_i y_i \\ &= x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \end{aligned}$$

2. The angle θ between two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ is given in the relationship
$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{||\mathbf{x}||||\mathbf{y}||}, \ 0 \leq \theta \leq \pi.$$

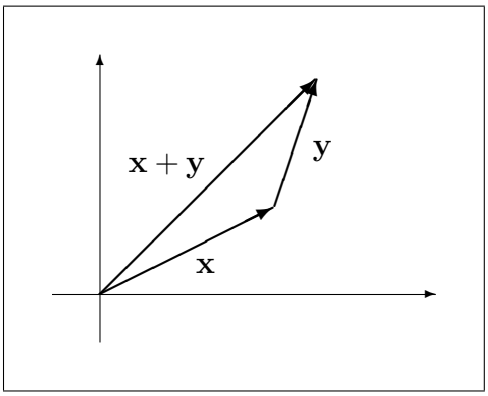

Figure 1.6: $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$

Theorem 1.2.2 *For* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

Proof. First note that
$$\begin{aligned} \mathbf{x} \cdot \mathbf{x} &= \sum_{i=1}^{n} {x_i}^2 \\ &= {x_1}^2 + {x_2}^2 + \cdots + {x_n}^2 \\ &= \left(\sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2} \right)^2 \\ &= ||\mathbf{x}||^2. \end{aligned}$$

That is, the norm squared of a vector equals the dot product with itself. Thus,

$$\begin{aligned} ||\mathbf{x}+\mathbf{y}||^2 &= (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y}) \\ &= \mathbf{x}\cdot\mathbf{x} + 2(\mathbf{x}\cdot\mathbf{y}) + \mathbf{y}\cdot\mathbf{y} \ \ (\text{Notice } \mathbf{x}\cdot\mathbf{y} = \mathbf{y}\cdot\mathbf{x}) \\ &= ||\mathbf{x}||^2 + 2(\mathbf{x}\cdot\mathbf{y}) + ||\mathbf{y}||^2 \\ &\leq ||\mathbf{x}||^2 + 2|\mathbf{x}\cdot\mathbf{y}| + ||\mathbf{y}||^2. \end{aligned} \tag{1.1}$$

We now look at $\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{||\mathbf{x}||||\mathbf{y}||}$. Taking absolute values, we have $|\cos\theta| = \frac{|\mathbf{x}\cdot\mathbf{y}|}{||\mathbf{x}||||\mathbf{y}||}$ since $||\mathbf{x}||, ||\mathbf{y}|| > 0$. But $|\cos\theta| \leq 1$ so $\frac{|\mathbf{x}\cdot\mathbf{y}|}{||\mathbf{x}||||\mathbf{y}||} \leq 1$ implies that $|\mathbf{x}\cdot\mathbf{y}| \leq ||\mathbf{x}||||\mathbf{y}||$. Using this in (1.1) above, we obtain

$$\begin{aligned} ||\mathbf{x}+\mathbf{y}||^2 &\leq ||\mathbf{x}||^2 + 2||\mathbf{x}||||\mathbf{y}|| + ||\mathbf{y}||^2 \\ &= \left(||\mathbf{x}|| + ||\mathbf{y}||\right)^2. \end{aligned}$$

Now take square roots to get $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$. ■

Remark 1.2.3 *Notice here that this is another example of "easy to state but difficult to prove." Don't focus a great deal on the proof itself though you should be able to follow it. You should be more concerned with the actual statement of the theorem as a useful tool to apply later.*

In closing, note that $|x+y| \leq |x| + |y|$ and $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ are very closely related. For example, if $\mathbf{x} = \langle x_1 \rangle \in \mathbb{R}^1$, then the definition of norm becomes $||\mathbf{x}|| = \sqrt{{x_1}^2} = |x_1|$, revealing $|x+y| \leq |x| + |y|$ as a special case of $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$, that is, when the "vectors" are just numbers (scalars).

To see a nice application of the triangle inequality in one dimension, consider the following problem.

Problem 1.2.1 *Let A, B and L be real numbers and let $\epsilon > 0$. If $|A - L| < \frac{\epsilon}{2}$ and $|B - L| < \frac{\epsilon}{2}$, show that $|A - B| < \epsilon$.*

First, let's ensure that this makes sense. The statement $\text{dist}(A, L) = |A - L| < \frac{\epsilon}{2}$ says that the numbers A and L are within $\frac{\epsilon}{2}$ units of one another. Likewise, the situation is analogous for B and L. The conclusion then says that the numbers A and B must be within ϵ units of one another. Note that if ϵ is a small number, this says that if A and L are close and B and L are close, then A and B must be close (which intuitively makes sense!). Look at the figure for a possible scenario (see the following page).

From the picture we can see that $|A - B| < \epsilon$. However, this is *not* a proof. A better approach might be trying some specific numerical values for A, B, L and ϵ. We want to show that $|A - B| < \epsilon$ will hold *provided that* $|A - L| < \frac{\epsilon}{2}$ and $|B - L| < \frac{\epsilon}{2}$.

So, for example, let $\epsilon = 1$. Then choose, say, $A = 10$ and $L = 10\frac{1}{4}$. Then indeed $|A - L| < \frac{\epsilon}{2}$. Now picking $B = 9\frac{7}{8}$, we obtain $|B - L| < \frac{\epsilon}{2}$. So we have to check: Is it true that $|A - B| < \epsilon$? Well, $|A - B| = |10 - 9\frac{7}{8}| = \frac{1}{8} < 1$ so *yes*, it follows. Still, this is *reassurance*; it is not proof! We need to show that the statement holds for *any* numbers A, B, L and $\epsilon > 0$, not just for the particular values we chose. Finally, here is the proof.

Proof. We need to show $|A - B| < \epsilon$ using the information $|A - L| < \frac{\epsilon}{2}$ and $|B - L| < \frac{\epsilon}{2}$. Here we use a trick seen frequently in higher mathematics−adding and subtracting the same quantity. From the structure of $|A - B|$, we need to add and subtract an L. Proceed as follows:

$$\begin{aligned} |A - B| &= |A \overbrace{-L + L}^{\text{zero}} - B| \\ &\leq |A - L| + |L - B| \quad \text{(Triangle Inequality)} \\ &= |A - L| + |B - L| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

So indeed $|A - B| < \epsilon$. ■

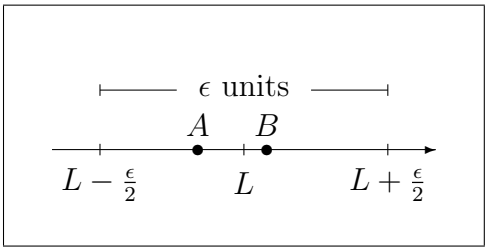

Figure 1.7: $|A - B| < \epsilon$

1.3 The Principle of Mathematical Induction

Chances are, it is likely you have heard of mathematical induction. This principle gives us a way of proving statements which hold for all *positive integers*. For example, suppose that you were asked to prove Bernoulli's inequality, which states that

$$\text{if } p > -1 \text{ then } (1 + p)^n \geq 1 + np, \, n \in \mathbb{N}.$$

From a naive point of view, we may start as follows. If $n = 1$, then $(1+p)^1 \geq 1+1p$ is valid (actually equality holds in this case). If $n = 2$, then $(1+p)^2 = 1+2p+p^2 \geq 1+2p$

since $p^2 \geq 0$. We can continue in this fashion until we get every $n \in \mathbb{N}\ldots$ or can we? The fact is, if we try to prove the statement this way, we will never in fact show that it holds *for all* positive integers n. We now look at the principle of mathematical induction; this allows us to be certain that Bernoulli's inequality (among other things) is indeed true for *all* $n \in \mathbb{N}$.

The Principle of Mathematical Induction: Let S_n be a statement in which the positive integer n appears. If

1. S_1 is true, and
2. S_k is true (for some integer k) *directly implies* that S_{k+1} is true,

then S_n is true *for all* $n \in \mathbb{N}$.

Basically, this says the following. We need to show that the statement with $n = 1$ (i.e., S_1) is true. Next, we *assume* it holds for *some* integer k (i.e., we *assume* S_k is true). This is often called the *induction hypothesis*. Finally, we prove that S_{k+1} is true as a direct consequence of the assumption that S_k is true. If we win this battle, then S_n is true for all $n \in \mathbb{N}$. Why does this work? Well, let's first try to prove Bernoulli's inequality with this tool and see why we can conclude that the statement holds *for all* $n \in \mathbb{N}$.

In Bernoulli's inequality, S_n is the statement $(1+p)^n \geq 1 + np$, $n \in \mathbb{N}$; this is what we are trying to prove. Step 1 is to prove that S_1 holds. We have already done this; if $n = 1$, then $(1+p)^1 \geq 1 + 1p$ is indeed true. Step 2 is to assume that S_k is true; that is, $(1+p)^k \geq 1 + kp$ for *some* $k \in \mathbb{N}$. We now need to prove that S_{k+1} follows; that is, $(1+p)^{k+1} \geq 1 + (k+1)p$. Well, let us begin with what we assumed to be true: $(1+p)^k \geq 1 + kp$. Since we are trying to prove a statement which has a $(1+p)^{k+1}$ on the left-hand side, it would be worthy to multiply our S_k statement by $1+p$ recalling that $1+p > 0$ by assumption (so the nature of the inequality remains unchanged). Thus,

$$\begin{aligned}(1+p)^{k+1} &= (1+p)(1+p)^k \\ &\geq (1+p)(1+kp) \text{ (using the induction hypothesis)} \\ &= 1 + (k+1)p + kp^2 \\ &\geq 1 + (k+1)p,\end{aligned}$$

since $k \in \mathbb{N}$ and $p^2 \geq 0$. Now look closely at the last inequality: $(1+p)^{k+1} \geq 1 + (k+1)p$. This is precisely the statement S_k with k replaced by $k+1$. That is, we've established S_{k+1}! As a result, we conclude that S_n is true for all $n \in \mathbb{N}$. In other words, $(1+p)^n \geq 1 + np$, $n \in \mathbb{N}$.

You may still be wondering *why* induction actually works. Many students stumble on the induction hypothesis−assuming that S_k holds for some k seems to be assuming

what we're trying to prove! In actuality, this is not the case at all. We are trying to prove that S_n holds for *all* positive integers $n-$this is quite a statement! To rephrase, we are trying to establish a truth involving the infinite. In the induction hypothesis, we are merely assuming that the statement S_k is true for just one particular positive integer $n-$that is, $n = k$. Do you see why this is a valid claim?

To see this more clearly, suppose we've established S_1 to be true. In the induction hypothesis S_k, let us momentarily assign $k = 1$ (this is fine since we *proved* S_1 was true). Now realize that we *prove* $S_{k+1} = S_{1+1} = S_2$ to be true. Okay−fine. Now in the assumption S_k, let $k = 2$ (fine again since S_2 is true). Then we *prove* $S_{k+1} = S_{2+1} = S_3$ is true. Now do you see why this works? This argument can be continued indefinitely. A lucid explanation of this is given in C.V. Eynden's *Elementary Number Theory*: Picture a string of equally spaced dominos arranged on a table. As long as we knock down the first domino, they will all fall as a consequence of our initial action−NO MATTER HOW MANY THERE ARE! To increase your comfort with induction, we give two more statements (with proofs!) using this principle. You might recognize the first from Calculus.

Problem 1.3.1 *Prove that* $\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$, $n \in \mathbb{N}$.

Proof. Our goal is to establish S_n:

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2},\ n \in \mathbb{N}.$$

First we show that S_1 is true. Here, we have $1 = \frac{1(1+1)}{2}$ and this is certainly true. We now assume that S_k holds. That is, for some $k \in \mathbb{N}$,

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}.$$

We need to show that S_{k+1} is true. That is,

$$1 + 2 + 3 + \cdots + k + (k+1) = \frac{(k+1)[(k+1)+1]}{2}.$$

(Notice that the previous statement is just S_k but with k replaced by $k + 1$.) Now, as in the proof of Bernoulli's inequality, we begin with our assumption S_k (what we

know to be true). If we add $k+1$ to both sides of S_k, then we obtain

$$\begin{aligned} 1+2+3+\cdots+k+(k+1) &= \frac{k(k+1)}{2}+(k+1) \\ &= \frac{k(k+1)}{2}+\frac{2k+2}{2} \\ &= \frac{k^2+3k+2}{2} \\ &= \frac{(k+2)(k+1)}{2} \\ &= \frac{(k+1)[(k+1)+1]}{2}. \end{aligned}$$

Hence, we've established S_{k+1}. As a consequence, $\sum_{i=1}^{n} i = 1+2+3+\cdots+n = \frac{n(n+1)}{2}$, $n \in \mathbb{N}$. ■

Remark 1.3.1 *Do you see an easier way to do the previous problem? Consider*

$$1+2+3+\cdots+(n-1)+n.$$

Now add to this (term by term) the exact same expression except written in the reverse order:

$$n+(n-1)+(n-2)\cdots+2+1.$$

That is, we have

$$\begin{array}{cccccccc} & 1 & 2 & 3 & \cdots & n-1 & n \\ + & n & n-1 & n-2 & \cdots & 2 & 1 \\ \hline & \downarrow & \downarrow & \downarrow & \cdots & \downarrow & \downarrow \\ & n+1 & n+1 & n+1 & \cdots & n+1 & n+1 \end{array}$$

Notice that there are n "$n+1$" terms so the sum is $n(n+1)$. That is,

$$2(1+2+\cdots+(n-1)+n) = n(n+1)$$

so

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

Problem 1.3.2 *Prove that $5^{2n}-1$ is a multiple of 8 for all $n \in \mathbb{N}$.*

Proof. S_1 says $5^{2(1)}-1 = 24 = 8(3)$ so this is certainly okay. Now we suppose that S_k is true. That is, we suppose that $5^{2k}-1$ is a multiple of 8 for some positive integer k. Another way of writing this is $5^{2k}-1 = 8z$ for some $k, z \in \mathbb{N}$. This is

because in order for $5^{2k}-1$ to be a multiple of 8, it must equal 8 times some integer z. We need to show that $5^{2(k+1)}-1$ is also a multiple of 8. Well, first notice that

$$5^{2(k+1)}-1 = 5^{2k}5^2 - 1.$$

Now somewhere we need to utilize $5^{2k}-1 = 8z$! Hence, we are motivated to add and subtract 5^2 and as a result

$$\begin{aligned} 5^{2k}5^2 - 1 &= 5^{2k}5^2 \overbrace{-5^2 + 5^2}^{\text{zero}} - 1 \\ &= 5^2(5^{2k}-1) + 24 \\ &= 5^2(8z) + 8(3) \\ &= 8(5^2 z + 3). \end{aligned}$$

Thus, $5^{2(k+1)}-1 = 8(5^2z+3)$, a multiple of 8, so S_{k+1} is true. Our work is done here. ■

Exercises. Prove the following by induction.

1. $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$, $n \in \mathbb{N}$.
2. $|\sin(nx)| \leq n|\sin x|$, $x \in \mathbb{R}$, $n \in \mathbb{N}$. (*Hint*: You may find the trigonometric identity $\sin(a+b) = \sin a \cos b + \cos a \sin b$ useful as well as the triangle inequality.)
3. $n^3 - n$ is a multiple of 6, $n \in \mathbb{N}$
4. $(2n)! \geq (n!)^2 2^n$, $n \in \mathbb{N}$. Recall that
$$n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1.$$
5. Prove that $|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|$ for every $x_1, x_2, \ldots, x_n \in \mathbb{R}$, $n \in \mathbb{N}$.

1.4 Proof by Contradiction

Many statements that require mathematical proof fall into the category of "if A, then B" where A and B are statements themselves. Another way of saying this is $A \Rightarrow B$. Here are a few examples:

1. If f is differentiable at $x = x_0$, then f is continuous at $x = x_0$.
2. If $p(x) = a_n x^n + \cdots + a_1 x + a_0$, then p has n roots.

3. If $\sum_{n=1}^{\infty} |a_n|$ converges, then $\sum_{n=1}^{\infty} a_n$ also converges.

Oftentimes, we can gather enough information in statement A and use our general knowledge (and various theorems) to eventually arrive at the conclusion B. Then our proof is complete. However, this is not always the case! Consider the following problem:

Problem 1.4.1 *If $p \in \mathbb{Q}$ and $q \in \mathbb{Q}'$, then $p + q \in \mathbb{Q}'$.*

How can we approach this directly? We can't possibly consider *all* rational numbers p and *all* irrational numbers q and then conclude that the sum is irrational. A different approach is necessary here.

The procedure goes something like this. We have $A \Rightarrow B$ where

$$A : p \in \mathbb{Q} \text{ and } q \in \mathbb{Q}', \text{ and}$$
$$B : p + q \in \mathbb{Q}'.$$

Let us suppose for a moment that B were false. That is, suppose that $p + q \notin \mathbb{Q}'$ which means that $p + q \in \mathbb{Q}$. (In doing so, we are still holding onto A since this is *given* information.) If we now proceed with our supposition $p+q \in \mathbb{Q}$ and get either

1. an inconsistency with statement A, or

2. an obvious absurdity (e.g., $1 < \frac{1}{2}$),

Then we can stop because here's the good news: *since all of our work was derived from the supposition $p+q \in \mathbb{Q}$, we can conclude that the statement $p+q \in \mathbb{Q}$ cannot be correct.* This method is typically called *proof by contradiction.* Let's see how this works in the proof of our problem.

Proof. Suppose that $p + q \in \mathbb{Q}$. We can still use the information $p \in \mathbb{Q}$ and $q \in \mathbb{Q}'$ but we know that this whole argument should unravel somewhere. Now since $p + q \in \mathbb{Q}$, we may write $p + q = \frac{m}{n}$ where $m, n \in \mathbb{Z}$, $n \neq 0$. But notice we have (by assumption) $p \in \mathbb{Q}$. Letting $p = \frac{a}{b}$ with $a, b \in \mathbb{Z}$, $b \neq 0$, we now obtain $\frac{a}{b} + q = \frac{m}{n}$. Thus $q = \frac{m}{n} - \frac{a}{b} = \frac{bm-an}{bn}$. Now $b, n \neq 0$ and $b, n \in \mathbb{Z}$ so that $bn \in \mathbb{Z}$. The same holds true for $bm - an$. Thus, q has representation $\frac{j}{k}$, where $j = bm - an \in \mathbb{Z}$, $k = bn \in \mathbb{Z}$, $k \neq 0$. That is, q is a *rational* number. In other words, $q \in \mathbb{Q}$. However, by assumption, $q \in \mathbb{Q}'$! So we have arrived at the conclusion that q is both irrational and rational! Since this is impossible, we conclude that our (original) supposition is incorrect. That is, supposing that $p + q \in \mathbb{Q}$ is wrong; it must be that $p + q \in \mathbb{Q}'$. This is what we were trying to prove. ■

Next, we present two different proofs of a particular result−one done in direct fashion, the other done via contradiction. This way, you can see two different methods in action.

Problem 1.4.2 *Prove that the sum of two positive even integers is even.*

1. **Proof**. (**direct proof**) Let x and y be positive even integers. Then $x = 2k$ and $y = 2m$ for $k, m \in \mathbb{N}$. We need to show that $x + y$ is even. Well,

$$\begin{aligned} x + y &= 2k + 2m \\ &= 2(k + m), \end{aligned}$$

   where $k + m \in \mathbb{N}$ since $k, m \in \mathbb{N}$. Therefore, $x + y$ is also even. ■

2. **Proof**. (**by contradiction**) As above, $x = 2k$ and $y = 2m$ for $k, m \in \mathbb{N}$. But now we *suppose* that $x + y$ is <u>not</u> even. In other words, we have $x + y \neq 2n$ for any $n \in \mathbb{N}$. Then this implies that $2(k + m) \neq 2n$ for any $n \in \mathbb{N}$. Thus $k + m \neq n$ for any $n \in \mathbb{N}$. This is a contradiction since both k and m were elements of $\mathbb{N}$ from the start. Hence, whatever we originally assumed must be incorrect. Since we assumed that $x + y$ was not even, we conclude that $x + y$ must be even which is what we were trying to show. ■

You may be wondering why we even proved this using the contradiction method. With all due respect, the direct proof was easier! Therefore, we offer one more problem which seems too impractical to approach in the "direct" fashion.

Problem 1.4.3 *There are infinitely many prime numbers.*

Remark 1.4.1 *You may recall that a prime number is a positive integer greater than one which has only itself and the number* 1 *as divisors. For example, the first few primes are* $2, 3, 5, 7, 11, \ldots$

Remark 1.4.2 *Try proving this statement directly! To do this, we would have to show that we "never run out of primes." To say the least, this is simple unreasonable (we don't have enough time to do this). A far superior starting point is proof by contradiction.*

Proof. We assume the contrary. That is, suppose that there are only a finite number of primes. Then we can list them in increasing order:

$$p_1, p_2, p_3, \ldots, p_n.$$

Now consider the new number P defined by

$$P = p_1 p_2 p_3 \cdots p_n + 1.$$

Now P can fall into one of two categories: It is either

1. prime, or

2. not prime (so one of the p_i's divides evenly into P).

First of all, P cannot be prime since p_n is our largest prime by assumption and $P > p_n$. Hence it must be that one of $p_1, p_2, p_3, \ldots, p_n$ divides exactly into P. However, this is not the case either since P/p_i has remainder 1 for $i = 1, 2, 3, \ldots, n$ (look at the way P is defined). So we've essentially done the following: P must fall into one of the categories above but we have just eliminated both possibilities! This is a contradiction. Since we assumed that there were only a finite number of primes, we can conclude that there must be an *infinite* collection of prime numbers. ■

Proof by contradiction is, no doubt, a very elegant method. It is also extremely useful to use in conjunction with uniqueness results (i.e., when you need to show that exactly one of something exists). Since this method is applied widely in the context of sequences, we review some of this material here (which you should recall from Calculus) and present two proofs using these tools.

Definition 1.4.1 *Let $L \in \mathbb{R}$. We write $\lim_{n\to\infty} s_n = L$ or $s_n \to L$ $(n \to \infty)$ and say that the limit of the sequence $\{s_n\}_{n=1}^{\infty}$ is L if, for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $|s_n - L| < \epsilon$ for $n \geq N$. If the sequence $\{s_n\}_{n=1}^{\infty}$ has a limit, it is said to converge. Otherwise, the sequence diverges.*

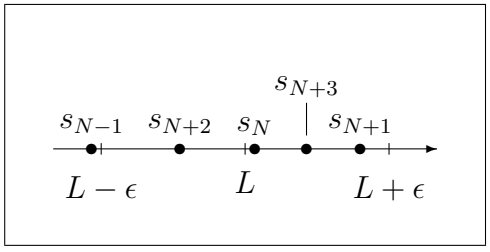

Figure 1.8: For n large enough (i.e., $n \geq N$), all of the terms of the sequence fall in the interval $(L - \epsilon, L + \epsilon)$.

Remark 1.4.3 *There are two facts that are noteworthy here.*

1. *Study what this definition really says: once we pick an $\epsilon > 0$ (any ϵ that is), then all of the terms in the sequence beyond and including the N^{th} one must satisfy $|s_n - L| < \epsilon$. That is to say, $|s_N - L| < \epsilon$, $|s_{N+1} - L| < \epsilon$, $|s_{N+2} - L| < \epsilon, \ldots$ Notice that another way of writing $|s_n - L| < \epsilon$ is $-\epsilon < s_n - L < \epsilon$ or $L - \epsilon < s_n < L + \epsilon$. In other words, $\lim_{n\to\infty} s_n = L$ means $L - \epsilon < s_n < L + \epsilon$ for $n \geq N$. See Figure 1.8.*

2. *It is appropriate to write $N = N(\epsilon)$ in most situations. That is, it is often the case that N depends on the ϵ chosen. Only in the most unusual or simplest cases does this fail to be true.*

Example 1.4.1 *$\{s_n\}_{n=1}^{\infty} = \left\{\frac{1}{n}\right\}_{n=1}^{\infty}$ has terms $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ We see that $L = 0$ here. Is it really true that for any $\epsilon > 0$ we can find an $N \in \mathbb{N}$ such that $|s_n - L| < \epsilon$ for $n \geq N$? Well, $|s_n - L| = |\frac{1}{n} - 0| = \frac{1}{n}$ and we want this to be less than ϵ. That is, we want $\frac{1}{n} < \epsilon$. Notice that this translates to $n > \frac{1}{\epsilon}$. Thus, pick $N = N(\epsilon) = \lfloor \frac{1}{\epsilon} \rfloor + 1$. See the figure below.*

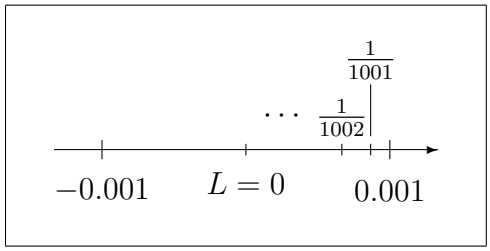

Figure 1.9: For example, for $\epsilon = 0.001$, $\frac{1}{n} < \epsilon$ for $n \geq N = 1001$.

Remark 1.4.4 *Note that $L = 0$ does not equal any term in the sequence $\left\{\frac{1}{n}\right\}_{n=1}^{\infty}$!*

Example 1.4.2 *$\{s_n\}_{n=1}^{\infty} = \{(-1)^n\}_{n=1}^{\infty}$ has terms $-1, 1, -1, 1, -1, 1, \ldots$ We see that this sequence has no limit. That is, the terms s_n never get "close" to any number L no matter how far we go into the sequence; they simply oscillate between -1 and 1 forever.*

Here are two sequence problems that benefit from proof by contradiction.

Problem 1.4.4 *Show that if a sequence has a limit, then the limit is unique.*

Proof. We need to show that if $\lim_{n\to\infty} s_n = L$, then there is only one such L that does the job! This is readily seen in the example $\{s_n\}_{n=1}^{\infty} = \left\{\frac{1}{n}\right\}_{n=1}^{\infty}$ above. The limit must be $L = 0$; it can't be anything else! However, *proving* this result involves a bit more than just strong words. We proceed by using proof by contradiction. That is, suppose we have two limits: $s_n \to L_1$ and $s_n \to L_2$ with $L_1 \neq L_2$. Since $L_1 \neq L_2$, we know that $|L_1 - L_2| > 0$ so we let $\epsilon = |L_1 - L_2|$. Now since $s_n \to L_1$, there exists an $N_1 \in \mathbb{N}$ such that $|s_n - L_1| < \frac{\epsilon}{2}$ for $n \geq N_1$. (The fact that we made $|s_n - L_1| < \frac{\epsilon}{2}$ instead of ϵ is completely arbitrary. It is done simply for elegance as we will see at the end of the proof.) Likewise, there exists an $N_2 \in \mathbb{N}$ such that $|s_n - L_2| < \frac{\epsilon}{2}$ for $n \geq N_2$. Keep in mind that it could very well be that $N_1 \neq N_2$ (this is most likely the case!). The important matter here is that we can make the

terms of $\{s_n\}_{n=1}^{\infty}$ within $\frac{\epsilon}{2}$ units of L_1 if $n \geq N_1$ and within $\frac{\epsilon}{2}$ units of L_2 if $n \geq N_2$. Now if we let $N = \max(N_1, N_2)$, then both

$$|s_n - L_1| < \frac{\epsilon}{2}$$

and

$$|s_n - L_2| < \frac{\epsilon}{2}$$

for $n \geq N$. So where is all of this going? Well, if we take $n \geq N$,

$$\begin{aligned} |L_1 - L_2| &= |L_1 - s_n + s_n - L_2| \\ &\leq |L_1 - s_n| + |s_n - L_2| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \\ &= |L_1 - L_2|. \end{aligned}$$

Notice what we have here: $|L_1 - L_2| < |L_1 - L_2|$. Definitely a contradiction! Therefore, L_1 and L_2 must be the same; i.e., $\{s_n\}_{n=1}^{\infty}$ has only one limit. ■

Remark 1.4.5 *In the previous problem, do you see why $\frac{\epsilon}{2}$ was chosen for elegance? Had we not done this, the proof would still be fine. We could have instead selected $\epsilon = \frac{|L_1 - L_2|}{2}$ and insisted on an $N_1 \in \mathbb{N}$ such that $|s_n - L_1| < \epsilon$ for $n \geq N_1$ and an $N_2 \in \mathbb{N}$ such that $|s_n - L_2| < \epsilon$ for $n \geq N_2$. The proof would then read $|L_1 - L_2| < 2\epsilon = |L_1 - L_2|$, resulting in a similar contradiction. You fill in the details.*

Problem 1.4.5 *Prove that $\{s_n\}_{n=1}^{\infty} = \{(-1)^n\}_{n=1}^{\infty}$ has no limit.*

Proof. Even though we verified this earlier, we never really *showed* it to be true. Now we do. Note that our definition requires that for any $\epsilon > 0$, we must be able to find an $N \in \mathbb{N}$ such that $|s_n - L| < \epsilon$ for $n \geq N$. Notice, for example, if we pick $\epsilon = \frac{1}{3}$ and $L = 1$, we run into problems. We cannot say that $|(-1)^n - 1| < \frac{1}{3}$ $(n \geq N)$ for *any* n. For example, if $n = 2$ then indeed $|(-1)^2 - 1| = 0 < \frac{1}{3}$. However, for $n = 3$, $|(-1)^3 - 1| = 2 > \frac{1}{3}$ so the definition is not satisfied. A similar mishap will occur if one claims the limit is $L = -1$. The best way to prove this is via contradiction. Thus, suppose that $(-1)^n \to L$. Then, given an $\epsilon > 0$, there exists an $n \in \mathbb{N}$ such that $|(-1)^n - L| < \epsilon$ $(n \geq N)$. For convenience, let us just choose an ϵ, say $\epsilon = \frac{1}{10}$. Now if n is even, we have $|1 - L| < \frac{1}{10}$; if n is odd, we have $|-1 - L| = |1 + L| < \frac{1}{10}$. Now examine this carefully. We have $\text{dist}(1, L) < \frac{1}{10}$ along with $|1 + L| = |L - (-1)| = \text{dist}(-1, L) < \frac{1}{10}$. How can this be? Look at the picture (see the next page). From this, it follows that the distance between -1 and 1 is

less than $\frac{2}{10} = \frac{1}{5}$! Since we have a contradiction, our assumption that $(-1)^n \to L$ is incorrect. In other words, $\{(-1)^n\}_{n=1}^{\infty}$ has no limit. ■

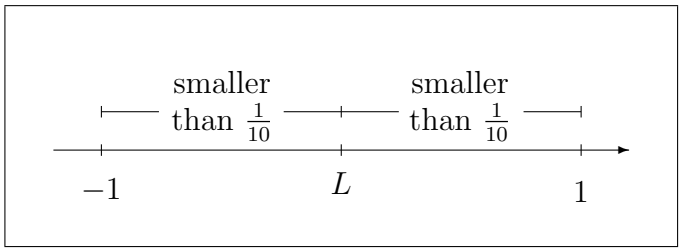

Figure 1.10: $\text{dist}(-1, 1) < \frac{2}{10}$

1.5 The Contrapositive

This section is a continuation of the previous section; everything here is discussed within the context of the if-then statement

$$\text{if } A, \text{ then } B$$

or $A \Rightarrow B$. We urge you to glance back at the three examples cited at the beginning of Section 1.4. The first example says if f is differentiable at $x = x_0$, then f is continuous at $x = x_0$ (you probably recall this fact from Calculus). Glance at the picture (see the following page). Now suppose we switch the locations of A and B and negate each−then we will have, "if f is not continuous at $x = x_0$, then f is not differentiable at $x = x_0$." (This says, "if not B, then not A," where the "not" means negation.) This new statement is called the *contrapositive* of the original statement. See below.

$$A \Rightarrow B \text{ (original statement)}$$
$$\text{not } B \Rightarrow \text{not } A \text{ (contrapositive)}$$

What do you notice about the contrapositive? It is certainly true! If f is not continuous at a point, it cannot be differentiable there. Again, recall from Calculus that in order for a function to be differentiable at a point, there can be no lapses in continuity or cusps in the graph.

The big question is this: Is the truth value of the contrapositive always the same as the truth value of the original statement? Let's delve into this, shall we? (Real justification requires examining a truth table which we do not do here.) We have $A \Rightarrow B$, our original statement. So if A happens, so does B (i.e., A *implies* B). But notice that B *can* happen without A having occurred. For example, B may be the result of many different things, A being just one of them. The bigger curiosity is

What if B doesn't happen?

Then what can be said about A? Well, if B doesn't happen, then there's no possible way that A could have happened (because if A had occurred, B would have followed as a result!). The previous sentence says "if not B, then not A." In other words, the contrapositive $(\text{not } B) \Rightarrow (\text{not } A)$ is equivalent to the original statement $A \Rightarrow B$. So why all of this discussion? Well, this is very **big** news. If $A \Rightarrow B$ and $(\text{not } B) \Rightarrow (\text{not } A)$ are equivalent statements, then this means we can prove the contrapositive of a statement instead of the original statement. Sometimes this is much easier to do!

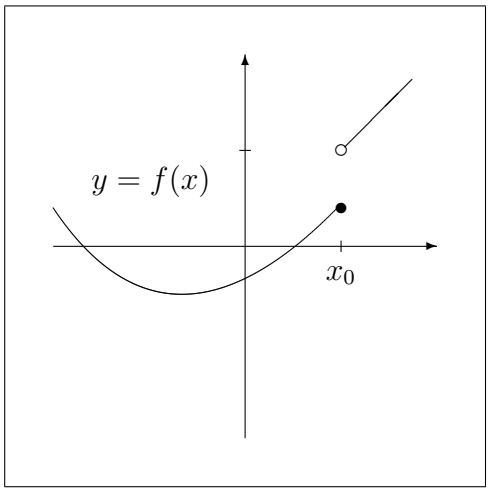

Figure 1.11: f is not continuous at $x = x_0$. Therefore f is not differentiable at $x = x_0$.

Before we move on, let us look at some simple statements along with their contrapositives so you can be convinced of their equivalence. As you look at the examples below, you should

1. set a truth value (either true or false) to the original statement, and
2. convince yourself that, based on this, the contrapositive has the same truth value as the original statement.

Example 1.5.1 *First read each statement labeled (a). Then read the contrapositive labeled (b).*

1. *(a) If it is raining, then it is cloudy.*
   (b) If it is not cloudy, then it is not raining.
2. *(a) If* $\sum_{n=1}^{\infty} |a_n|$ *converges, then* $\sum_{n=1}^{\infty} a_n$ *also converges.*

(b) *If* $\sum_{n=1}^{\infty} a_n$ *diverges, then* $\sum_{n=1}^{\infty} |a_n|$ *diverges.*

3. (a) *If your average exceeds 90, then you will earn an A.*
   (b) *If you don't earn an A, then your average is below 90.*

Remark 1.5.1 *Be sure that you realize that the logical thinking is the foundation of such claims. For example, read the following:*

1. *If the Red Sox win the World Series, then I will throw a party.*
2. *If I have not thrown a party, then the Red Sox haven't yet won the World Series.*

Do you see any problems with declaring the above statements as equivalent? How is this different from the other examples?

We now present two problems; both proofs are done by proving the contrapositive of the original statement. You are encouraged to attempt a direct proof in each case.

Problem 1.5.1 *Let* $f(x) = 5x + 2$. *If* x_1 *and* x_2 *are distinct, prove that* $f(x_1) \neq f(x_2)$.

Proof. Note that f is just a line in $\mathbb{R}^2$. You are asked to show that if you take two different x values that you obtain two different function values. This is very intuitive but sometimes the obvious things are the most difficult to prove! (Go ahead and try it.) We proceed by proving the contrapositive. That is, if $f(x_1) = f(x_2)$, then $x_1 = x_2$. We begin with $f(x_1) = f(x_2)$ so

$$5x_1 + 2 = 5x_2 + 2.$$

Hence, $5x_1 = 5x_2$ so $x_1 = x_2$. Thus, we are done! Since the contrapositive has been proved, the original statement must be true. ■

Problem 1.5.2 *If* $3k$ *is odd, then* k *is odd,* $k \in \mathbb{N}$.

Proof. Again, you may try doing this in the straightforward "direct" fashion but using the contrapositive simplifies the situation substantially. Let's prove "if k is not odd, then $3k$ is not odd." In other words, we will establish, "if k is even, then $3k$ is even." Since k is even, $k = 2n$, $n \in \mathbb{N}$. Then $3k = 3(2n) = 2(3n)$. In other words, $3k$ is a multiple of 2. Thus, $3k$ is even. The proof is complete. ■

As a finale in this preliminary chapter, we close with one more exercise which clearly benefits from using the contrapositive.

Problem 1.5.3 *Let* x *and* y *be integers such that* $x + y$ *is even. Show that* x *and* y *are either both even or both odd.*

Remark 1.5.2 *When two numbers are both even or both odd, we say that the numbers have the same parity. For example, the numbers* 3 *and* 5 *have the same parity but* 3 *and* 4 *do not.*

Proof. The contrapositive states that if x and y are of opposite parity (one even, one odd), then $x + y$ is odd. So without loss of generality, we may assume that x is even and y is odd. Hence, there exists $m, n \in \mathbb{N}$ such that $x = 2m$ and $y = 2n + 1$. Then

$$\begin{aligned} x + y &= 2m + (2n + 1) \\ &= 2(m + n) + 1 \end{aligned}$$

so $x + y$ is odd and the result follows. ■

At this point, you now have the necessary tools to proceed to Chapter 2. Best of luck!

1.6 Summary: Odds and Ends

- $|x| \geq 0$
- $|x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$
- $|x| = k$ IFF $x = k$ or $x = -k$.
- $|x| \leq k$ IFF $-k \leq x \leq k$.
- $|x| \geq k$ IFF $x \leq -k$ or $x \geq k$.
- $|AB| = |A||B|$
- $|A - B| = |B - A|$
- $\sqrt{A^2} = |A|$
- $|x + y| \leq |x| + |y|$
- $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$
- Math Induction: Let S_n be a statement in which the positive integer n appears. If

    1. S_1 is true, and

2. The assumption of S_k being true (for some integer k) *directly implies* that S_{k+1} is true,

then S_n is true *for all* $n \in \mathbb{N}$.

- Proof by contradiction: Assume the contrary and obtain either a contradiction or obvious absurdity.
- The statements $A \Rightarrow B$ and (not B) $\Rightarrow$ (not A) are contrapositives; they are equivalent statements and hence, share the same truth value.

Chapter 2

Sets

2.1 What is a Set?

A *set*, in simplest terms, is a collection of objects. The objects are typically called the *elements* of the set or its *members*. It is customary to denote a set by a capital letter. For example,

$$A = \{1, 2\}$$

is the set A containing the numbers 1 and 2. Since the objects in A are precisely these numbers, we write $1 \in A$ and $2 \in A$. As was mentioned earlier, the symbol $\in$ means "is an element of" or "belongs to." Thus, $1 \in A$ but $5 \notin A$. Here are some additional examples of sets.

1. $\mathbb{Z}$
2. $B = [0, 1]$
3. $C = \{(x, y) \mid x < 0 \text{ and } y < 0\}$
4. $D = \{x \in \mathbb{R} \mid x^2 - 5x + 6 = 0\}$

Note that there is usually more than one way to represent a set. To see this, we've already observed that $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$. B can be written as $B = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$. Notice that we cannot actually *list* all of the numbers in B contrary to the partial listing for $\mathbb{Z}$ whose interpretation is clear. Additionally, you should readily see that C is the set representing all points in the third quadrant of the Cartesian plane. Finally, wee see that $D = \{2, 3\}$.

The sets we have seen thus far are rather restrictive (that is to say, mathematical). We could have something far more general such as

$$\text{MLB} = \{\text{Astros, Cardinals, Red Sox, Cubs, Twins, } \ldots\}$$
$$\text{Red Sox} = \{\text{Millar, Damon, Schilling, Ortiz, } \ldots\}.$$

We can see that Red Sox $\in$ MLB and that Damon $\in$ Red Sox. However, although Johnny Damon plays for the Red Sox, do you see that Damon $\notin$ MLB?

Something you may be wondering about is the existence of a set containing absolutely nothing at all. We give this the symbol $\emptyset$.

Definition 2.1.1 *$\emptyset$ is the set having no members; it is called the empty or null set. In mathematical terms, $\emptyset = \{x \mid x \neq x\}$.*

Since $x \neq x$ is always false, no member x satisfying the property $x \neq x$ can belong to the set $\emptyset$. Thus, $\emptyset$ contains nothing at all! Additionally, the notion of a subset is very important.

Definition 2.1.2 *Let A and B be sets. We write $A \subset B$ and say that A is a subset of B provided that all members of A are also members of B. If $A \subset B$ but $A \neq B$, then A is said to be a proper subset of B.*

Example 2.1.1 *Let $A = \{a, b, c\}$ and $B = \{b, c\}$. Then $B \subset A$ since $b \in B$ and $c \in B$ and $b \in A$ and $c \in A$. Notice that we cannot say that $A \subset B$ since $a \in A$ yet $a \notin B$. See the diagram.*

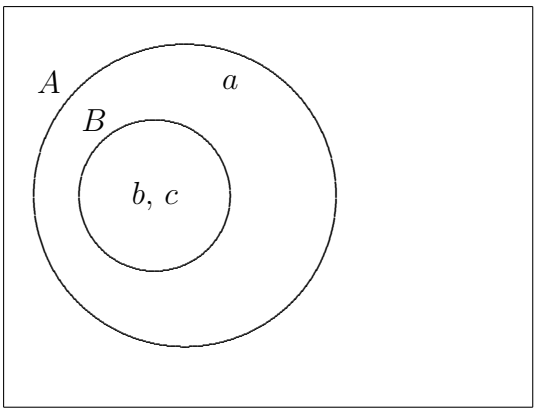

Figure 2.1: $B \subset A$

Definition 2.1.3 *Two sets A and B are said to be equal if they contain precisely the same elements. We write $A = B$.*

Remark 2.1.1 *Notice that $A = B$ really means that $A \subset B$ and $B \subset A$. Think about it!*

How does one go about *proving* that $A \subset B$? The most logical way is to take an element of A (say x) and then show that x must be in B as well. That is, show that $x \in A \Rightarrow x \in B$. This means that $A \subset B$. By reading the remark above, you can then see what needs to be done for equality of sets.

We now give two very important consequences of these definitions.

Proposition 2.1.1 *$\emptyset \subset A$ for any set A.*

Remark 2.1.2 *Notice that this says that the empty set is a subset of every set in the world! Convice yourself that this is true.*

Proof. We want to show that $\emptyset \subset A$ for any set A. Well, suppose that $\emptyset \not\subset A$ (proof by contradiction). Then $\emptyset$ would contain an element, say x, that does not belong to A. In other words, $x \in \emptyset$ but $x \notin A$. But this cannot be since $\emptyset$ *has* no elements. This is a contradiction proving that $\emptyset \subset A$. ■

Here in another (seemingly obvious) statement:

Proposition 2.1.2 *$A \subset A$ for any set A.*

Proof. This can be deduced from the fact that any set has precisely the same elements as itself! That is, $A = A$ so certainly $A \subset A$ (see the definition of equal sets). ■

The previous two propositions are elementary but they say a great deal. Given any set A, $\emptyset$ and A are subsets of A. We close this section with some illustrative examples and a proof.

Example 2.1.2 *The purpose of this example is to distinguish between the (sometimes confused) symbols $\in$ and $\subset$. Consider $S = \{a, b, \{c\}, \{d, e\}\}$. The following statements are all true; you should convince yourself of this!*

1. *The set S contains four elements: the letter a, the letter b, the set $\{c\}$, and the set $\{d, e\}$.*
2. *$\{b\} \subset S$*
3. *$\{b\} \notin S$ (nowhere do you see $\{b\}$ in S)*
4. *$d \notin S$, neither is $\{d\}$*
5. *$\emptyset \subset S$ but $\emptyset \notin S$*
6. *$\{c\} \in S \Rightarrow \{\{c\}\} \subset S$*

Example 2.1.3 *Consider $A = \{\emptyset\}$. Do you see that $A \neq \emptyset$? Why is this so? The set A contains precisely one element and this element is the set $\emptyset$. Hence, A is not empty. Therefore $A \neq \emptyset$. The bottom line is that saying $A = \emptyset$ is very different from $A = \{\emptyset\}$.*

Proposition 2.1.3 *Let A, B and C be sets. If $A \subset B$ and $B \subset C$, then $A \subset C$.*

Proof. We have $A \subset B$ and $B \subset C$. First of all, it could be that $A = \emptyset$. If so, then $A \subset C$ follows automatically since $\emptyset$ is a subset of every set. Otherwise, let $A \neq \emptyset$ and consider $x \in A$. Since $A \subset B$, $x \in B$ as well. Since $B \subset C$, $x \in C$. The string of statements connecting $x \in A$ to $x \in C$ implies, by definition, that $A \subset C$. ■

Exercises.

1. Describe the following sets.
   (a) $A = \{x \,|\, |x| \leq 7\}$
   (b) $B = \{(x, y) \,|\, x^2 + y^2 = 25\}$
   (c) $C = \{(x, y, z) \,|\, z = x + y\}$
2. True or False?
   (a) $5 \in \mathbb{Q}$
   (b) $\{5\} \in \mathbb{Q}$
   (c) $\sqrt{2} \subset \mathbb{R}$
   (d) $\{\pi, \sqrt{5}, 17\} \subset \mathbb{R}$
3. Consider the following definition:

   Definition 2.1.4 *Let A be a set. The power set of A, denoted by $\mathcal{P}(A)$, is the set containing all subsets of A.*

   For example, if $A = \{1, 2, 3\}$, then

   $$\mathcal{P}(A) = \{\emptyset, A, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}\}.$$

   (a) Let $B = \{a, b, c, d\}$. Find $\mathcal{P}(B)$.
   (b) Let $C = \{x, y\}$. Find $\mathcal{P}(C)$.
   (c) Let $D = \{a, \{b, \{c\}\}\}$. Find $\mathcal{P}(D)$.
   (d) Based on these examples, if S is a set containing n elements, how many elements does $\mathcal{P}(S)$ have?

(e) It is fairly easy to show that if $A \subset B$, then $\mathcal{P}(A) \subset \mathcal{P}(B)$.

Proof. We have $A \subset B$. Now take $X \in \mathcal{P}(A)$. We use X instead of x since X is a set (recall that the "elements" of $\mathcal{P}(A)$ are sets). We need to show that $X \in \mathcal{P}(B)$. Since $X \in \mathcal{P}(A)$, $X \subset A$. By the assumption $A \subset B$ and **Proposition 2.1.3**, $X \subset B$. But $X \subset B$ means that $X \in \mathcal{P}(B)$. We've established $X \in \mathcal{P}(A) \Rightarrow X \in \mathcal{P}(B)$ so that $\mathcal{P}(A) \subset \mathcal{P}(B)$. ■

Prove or disprove: $\mathcal{P}(A) \subset \mathcal{P}(B) \Rightarrow A \subset B$.

(f) Consider the set $[0, 1] \subset \mathbb{R}$. Note that $[0, 1] \in \mathcal{P}(\mathbb{R})$. Can you list some additional "elements" of $\mathcal{P}(\mathbb{R})$? Can you list them all? Why or why not? Is the problem any easier for $\mathcal{P}([0, 1])$? Explain.

4. Let $X = \{x \in \mathbb{Z} \,|\, |x| \leq 1\}$ and $Y = \{-1, 0, 1\}$. Prove that $X = Y$.

5. Prove that there is only one empty set. That is, let A be a set with no elements, likewise for B. Then prove that $A = B$.

6. Let $A \subset B$. If $x \notin B$ prove that $x \notin A$. *Hint:* Look at the contrapositive!

2.2 Operations on Sets

Much like operations with real numbers (addition, subtraction, . . .), there are "operations" on sets. In this section, we state all of the main definitions and prove a good number of propositions so that you can become accustomed to the language of these "set theory" proofs. Everywhere in this section, A, B, and C are assumed to be sets.

Definition 2.2.1 *The union of A and B, written $A \cup B$, is the set of all elements in either A, B or both. That is,*

$$A \cup B = \{x \,|\, x \in A \text{ or } x \in B\}.$$

Definition 2.2.2 *The intersection of A and B, written $A \cap B$, is the set of all elements in both A and B. That is,*

$$A \cap B = \{x \,|\, x \in A \text{ and } x \in B\}.$$

Remark 2.2.1 *If $A \cap B = \emptyset$, we say that A and B are disjoint.*

Definition 2.2.3 *The difference of A and B, written $A \backslash B$, is the set of all elements in A but not in B. That is,*

$$A \backslash B = \{x \,|\, x \in A \text{ and } x \notin B\}.$$

See the figure. Note that while $A\backslash B$ and $A\cap B$ are explicitly labeled, $A\cup B$ contains all such objects in A, B, or both (in the figure, this means everything contained within the circles used to represent A and B).

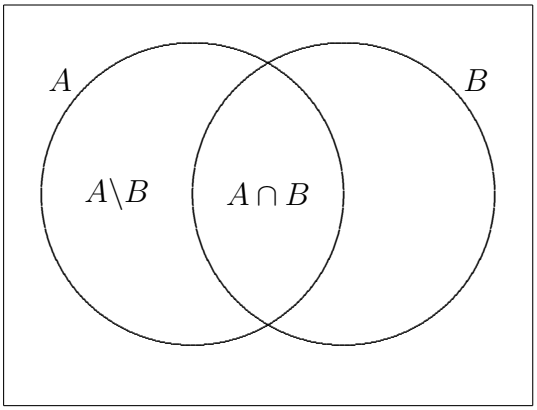

Figure 2.2: $A \cap B$ and $A\backslash B$

Example 2.2.1 *Let* $A = \{t, h, i, s\}$, $B = \{i, s\}$, *and* $C = \{f, u, n\}$.

1. $A \cup B = \{t, h, i, s\} = A$
2. $A \cap B = \{i, s\} = B$
3. $A\backslash B = \{t, h\}$
4. $B\backslash A = \emptyset$
5. $A \cup B \cup C = \{t, h, i, s, f, u, n\}$

Remark 2.2.2 *The above example illustrates some important points. First, the difference operator* $\backslash$ *is* <u>not</u> *commutative. On the other hand, the operators* $\cup$ *and* $\cap$ *commute. To see this, compare* $B \cup A$ *and* $B \cap A$ *with (1) and (2) above, respectively. Finally, much like the familiar operations of addition, subtraction, multiplication, and division, we can apply operations more than once if desired. See (5) above.*

Almost always, we speak of a <u>universal set</u> U of which A and B are both subsets. Under this assumption, it is easier to analyze the given sets and to draw pictures as an aid in visualization. In the example above, it is readily understood that U is the set containing the 26 letters of the English alphabet. We now state some more definitions.

Definition 2.2.4 *The complement of A relative to U is the set of all elements not contained in A. It is denoted by A'. That is,*

$$\begin{aligned} A' &= \{x \,|\, x \notin A\} \\ &= U \backslash A. \end{aligned}$$

Remark 2.2.3 *Notice that we now have a different way of writing $A\backslash B$. Since, by definition,*

$$A\backslash B = \{x \,|\, x \in A \text{ and } x \notin B\},$$

and $x \notin B$ means that $x \in B'$, we can write

$$\begin{aligned} A\backslash B &= \{x \,|\, x \in A \text{ and } x \in B'\} \\ &= A \cap B'. \end{aligned}$$

We now state two theorems.

Theorem 2.2.1 (**De Morgan**) *Let $A, B \subset U$. Then*

1. $(A \cup B)' = A' \cap B'$
2. $(A \cap B)' = A' \cup B'$

Proof. We will prove the second statement in detail. Recall that to show equality of sets, we need to show inclusion both ways. That is, we need to show $(A \cap B)' \subset A' \cup B'$ and $A' \cup B' \subset (A \cap B)'$.

Case 1 We will show that $(A \cap B)' \subset A' \cup B'$. Take $x \in (A \cap B)'$. Then $x \notin A \cap B$. Since x is not in $A \cap B$, one of two things can happen: either $x \in A$ or $x \notin A$. If $x \in A$ then $x \notin B$ because $x \notin A \cap B$. Now $x \notin B \Rightarrow x \in B'$. Then certainly x is in the "bigger" set $A' \cup B'$. So $x \in A' \cup B'$. Thus, $(A \cap B)' \subset A' \cup B'$. In the other case, if $x \notin A$ then $x \in A'$. Thus $x \in A' \cup B'$ so that $(A \cap B)' \subset A' \cup B'$. We're done here.

Case 2 We will show that $A' \cup B' \subset (A \cap B)'$. Take $x \in A' \cup B'$. Then $x \in A'$ or $x \in B'$. Suppose $x \in A'$ (we'll cover the other case in just a moment). Then $x \notin A$. However, if $x \notin A$, then $x \notin A \cap B$ since $A \cap B$ is "smaller" than A. Thus, $x \in (A \cap B)'$. Now consider $x \notin B$. Hence $x \notin A \cap B$ so that $x \in (A \cap B)'$. Either way, $x \in A' \cup B'$ implies that $x \in (A \cap B)'$. Therefore, $A' \cup B' \subset (A \cap B)'$.

The two inclusions tell us that $(A \cap B)' = A' \cup B'$. ■

Remark 2.2.4 *Notice that the complement has a way of changing union to intersection and vice versa. Do you see how this parallels the arithmetic operations?*

The next theorem illustrates two more familiar properties.

Theorem 2.2.2 *Let A, B and C be sets. Then*

1. $(A \cap B) \cap C = A \cap (B \cap C)$ (**Associative Law for Intersection**)
2. $(A \cup B) \cup C = A \cup (B \cup C)$ (**Associative Law for Union**)
3. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (**Distributive Law**)
4. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (**Distributive Law**)

Proof. We will prove the third equality and leave the others as exercises. We will show that

1. $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$ and
2. $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.

1. Take $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. That is, $x \in A$ and $x \in B$ or $x \in A$ and $x \in C$.
   (a) If $x \in A$ and $x \in B$, then $x \in A \cap B$. Thus, $x \in (A \cap B) \cup (A \cap C)$.
   (b) On the other hand, if $x \in A$ and $x \in C$, then $x \in A \cap C$ so that $x \in (A \cap B) \cup (A \cap C)$.

   Thus, we've proved $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$.
2. Take $x \in (A \cap B) \cup (A \cap C)$. Then $x \in A \cap B$ or $x \in A \cap C$.
   (a) Let $x \in A \cap B$. That is, let $x \in A$ and $x \in B$. Then $x \in A$ and $x \in B \cup C$ since $B \cup C$ is "bigger" than B alone. Hence, $x \in A \cap (B \cup C)$.
   (b) Let $x \in A \cap C$ so $x \in A$ and $x \in C$. Thus, $x \in A$ and $x \in B \cup C$ so $x \in A \cap (B \cup C)$.

   Hence, $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.

Using the two inclusions, we have $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. ■

Here is a nice exercise.

Problem 2.2.1 *Let $A, B \subset U = \mathbb{N}$. Prove or disprove:*

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

1. You might guess that this is false in general since De Morgan's Law states that for *any* sets A and B, $A' \cap B' = (A \cup B)'$. Let's produce a counterexample. Let $A = \{1, 2\}$ and $B = \{2\}$. Then $A' = \{3, 4, 5, \ldots\}$ and $A \cup B = \{1, 2\}$. Thus $\emptyset = A' \cap B \neq (A \cup B)' = \{3, 4, 5, \ldots\}$.

2. This almost looks true so you may try to prove it to see where the argument falls apart. To see a counterexample, let $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$. Then $A \cup B = \{1, 2, 3, 4\}$ so that $(A \cup B) \backslash A = \{4\} \neq \{2, 3, 4\} = B$.

Remark 2.2.5 *We almost always use discrete sets such as* $\{a, b, c\}$ *or* $\{1, 2, 4, 5\}$ *because they are easy to work with. Feel free to use other types of sets if you wish. For example, if* $U = \mathbb{R}$ *in the problem above, then* $A = [1, 2)$ *and* $B = \left(\frac{1}{2}, 3\right]$ *would produce a counterexample just as well for* $(A \cup B) \backslash A = B$.

We finish this section with another collection of seemingly "obvious" statements. We leave you the details of the proofs that we omit.

Proposition 2.2.1 *Let* U *denote the universal set and suppose* $A \subset U$. *Then*

1. $\emptyset \cup A = A$
2. $\emptyset \cap A = \emptyset$
3. $A = A \cup A = A \cap A$
4. $A \cup A' = U$
5. $A \cap A' = \emptyset$

Proof. We prove $A \cup A' = U$ because the proof is insightful. We need to show that $A \cup A' \subset U$ and $U \subset A \cup A'$.

1. Let $x \in A \cup A'$ so either $x \in A$ or $x \in A'$. If $x \in A$ then $x \in U$. Likewise, if $x \in A'$ then $x \in U$. Either way, $x \in U$ so that $A \cup A' \subset U$.

2. Let $x \in U$ so that either $x \in A$ or $x \notin A$. If $x \in A$ then $x \in A \cup A'$. Otherwise, let $x \notin A$. This translates to $x \in U \backslash A$; that is, $x \in A'$. Hence, $x \in A \cup A'$ so $U \subset A \cup A'$.

Thus, we have established $A \cup A' = U$. ■

Exercises.

1. Prove that $(A')' = A$.

2. Let $A = \{1, 3, 5, 7, 9\}, B = \{5, 6, 7, 8\}$, and $C = \{5\}$ while $U = \mathbb{N}$. List the members of each set below.

   (a) A'

   (b) $A \cap B$

   (c) $(A \cap B)'$

   (d) $A' \cup B'$

   (e) $(A \cup B) \cap C$

   (f) $C' \cap (B \cup A)$

   (g) $B' \cup A$

   (h) $U \backslash (B \cup A)$

3. Prove the first law of De Morgan: $(A \cup B)' = A' \cap B'$.

4. Prove **Theorem 2.2.2** part (1).

5. Prove **Theorem 2.2.2** part (2).

6. Prove **Theorem 2.2.2** part (4).

7. Prove that $A \backslash (A \cap B') = A \cap B$.

8. Prove **Proposition 2.2.1**, parts (1) and (2).

9. As a reminder, the shorthand IFF means "if and only if." Thus, when we say A IFF B, this means

   (a) if A, then B and

   (b) if B, then A

   so there are really two things to prove in these circumstances. Try proving the following:

   (a) $A \cup B = B$ IFF $A \subset B$.

   (b) $A \cap B = A$ IFF $A \subset B$.

   (c) $A \subset B$ IFF $B' \subset A'$ (similar to Exercise 6 in Section 2.1).

   (d) $A \subset B'$ IFF $A \cap B = \emptyset$.

10. The *symmetric difference* Δ is another operation defined on sets given by

$$A \Delta B = (A \backslash B) \cup (B \backslash A).$$

(a) Describe, in plain English, the effect of operator Δ on sets A and B.

(b) Prove that Δ commutes. That is, show that

$$A\Delta B = B\Delta A.$$

(c) Here is a more challenging exercise: Prove that $A\Delta B$ can also be written as

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

2.3 Special Sets and "Infinity"

In the first two sections we have merely introduced the reader to sets, some of their properties, and associated theorems. In this section, we extend these notions further. Here we present an informal discussion of infinity, the notions of increasing and decreasing sets, as well as some specific examples. The reader will notice that many of the statements made here strike a parallel chord to material previously addressed. It is the purpose of this section to deepen your understanding of what has been presented thus far in Chapter 2.

Definition 2.3.1 *Suppose that for each $n \in \mathbb{N}$ there corresponds a set A_n. Then we have*

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n=1}^{\infty} A_n = \{x \mid \exists\, n \in \mathbb{N} \ni x \in A_n\}$$

and

$$\bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n=1}^{\infty} A_n = \{x \mid \forall\, n \in \mathbb{N}, x \in A_n\}.$$

Remark 2.3.1 *The symbol $\exists$ is shorthand for the expression "there exists," $\ni$ is shorthand for "such that," and $\forall$ is shorthand for "for all."*

Remark 2.3.2 *You should look at these definitions closely; they are the same union and intersection discussed previously but the wording inside of the braces is absolutely crucial. The union is the collection of all elements x where x must be in <u>some</u> A_n. The intersection, on the other hand, requires that the x be in <u>every</u> set A_n, $n \in \mathbb{N}$.*

Definition 2.3.2 *Consider a sequence of sets $\{A_n\}_{n=1}^{\infty}$, sometimes written as*

$$A_1, A_2, A_3, \ldots, A_n, \ldots$$

$\{A_n\}_{n=1}^{\infty}$ is <u>decreasing</u> if $A_{n+1} \subset A_n$, $n \in \mathbb{N}$. Similarly, $\{A_n\}_{n=1}^{\infty}$ is <u>increasing</u> if $A_n \subset A_{n+1}$, $n \in \mathbb{N}$.

See the figures below.

Example 2.3.1 *Let* $A_n = [1, 3 - \frac{1}{n}]$. *Find* $\bigcup_{n=1}^{\infty} A_n$ *and* $\bigcap_{n=1}^{\infty} A_n$.

Note that

$$\begin{aligned}\bigcup_{n=1}^{\infty} A_n &= [1,2] \cup \left[1, \frac{5}{2}\right] \cup \left[1, \frac{8}{3}\right] \cup \cdots \\ &= [1,3)\,.\end{aligned}$$

Note that 3 is not an element of any of the A_n's. On the other hand, we see that $\bigcap_{n=1}^{\infty} A_n = A_1 = [1,2]$. □

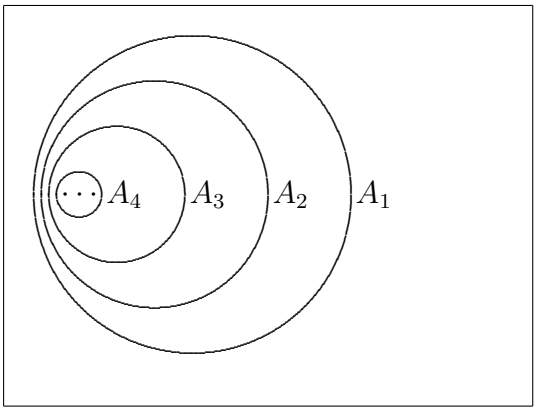

Figure 2.3: Decreasing sets get "smaller" since $A_{n+1} \subset A_n$.

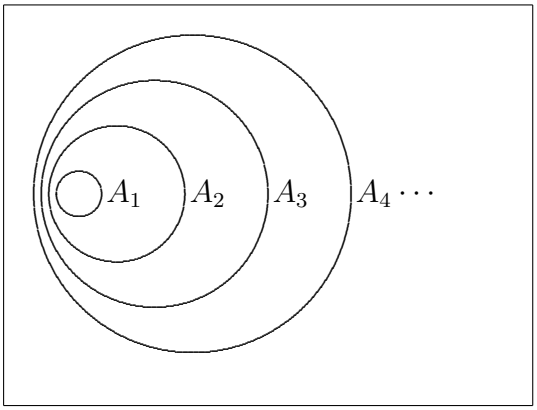

Figure 2.4: Increasing sets get "bigger" since $A_n \subset A_{n+1}$.

Remark 2.3.3 *Note that* $\{A_n\}_{n=1}^{\infty}$ *from* **Example 2.3.1** *is an increasing sequence since* $A_1 \subset A_2 \subset A_3 \subset \cdots A_n \subset A_{n+1} \subset \cdots$. *Also, it makes intuitive sense that* $\bigcap_{n=1}^{\infty} A_n = A_1$ *since the* A_n *are getting "larger" with each successive* n. *Again, see the previous figure.*

Example 2.3.2 *Give two distinct examples of decreasing sequences of (nonempty) sets whose intersection*

1. *is empty.*
2. *contains only one element.*

1. For simplicity, let us think of sets in $\mathbb{R}$. Let $A_n = \left(0, \frac{1}{n}\right) \subset \mathbb{R}$. Then we see that $A_1 = (0,1)$, $A_2 = \left(0, \frac{1}{2}\right)$, $A_3 = \left(0, \frac{1}{3}\right)$, et cetera. Certainly, they are nonempty and decreasing since

$$\cdots \subset A_{n+1} \subset A_n \subset \cdots A_3 \subset A_2 \subset A_1.$$

Next we see that

$$\begin{aligned}\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left(0, \frac{1}{n}\right) &= (0,1) \cap \left(0, \frac{1}{2}\right) \cap \left(0, \frac{1}{3}\right) \cap \cdots \\ &= \emptyset.\end{aligned}$$

Note that if ∞ were replaced by $N < \infty$, we would have

$$\bigcap_{n=1}^{N} A_n = \left(0, \frac{1}{N}\right) \neq \emptyset$$

for any $N \in \mathbb{N}$. However, for $N = \infty$, we obtain $\emptyset$. See the picture below.

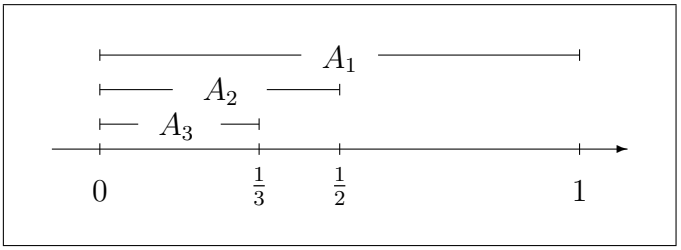

Figure 2.5: The A_n are decreasing

2. How can we adjust the A_n to accommodate? Instead, let $B_n = \left[0, \frac{1}{n}\right)$. Then

$$\bigcap_{n=1}^{\infty} B_n = \bigcap_{n=1}^{\infty} \left[0, \frac{1}{n}\right) = \{0\}.$$

That is, the only element that survives is 0. An obvious way to see this is to write $B_n = \{0\} \cup A_n = \left[0, \frac{1}{n}\right)$. Then it is clear that every B_n contains the number 0 so that at least $0 \in \bigcap_{n=1}^{\infty} B_n$. However, in this situation, $\{0\} = \bigcap_{n=1}^{\infty} B_n$.

Definition 2.3.3 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets. We say that the A_n are mutually disjoint if, for $i \neq j$, $A_i \cap A_j = \emptyset$.*

Example 2.3.3 *Consider the sets $A_n = [n, n+1)$. Then $A_1 = [1, 2)$, $A_2 = [2, 3)$, et cetera. Clearly, the A_n are mutually disjoint. For example, take $3 \neq 5$ so that $A_3 \cap A_5 = [3, 4) \cap [5, 6) = \emptyset$. Notice that $\bigcap_{n=1}^{\infty} A_n = \emptyset$ while $\bigcup_{n=1}^{\infty} A_n = [1, \infty)$.*

Example 2.3.4 *Consider the sets $B_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$. We see that the B_n are not mutually disjoint. For example, for $i > j$, $B_i \cap B_j = B_i = \left(-\frac{1}{i}, \frac{1}{i}\right) \neq \emptyset$. Also notice that $\bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$.*

Proposition 2.3.1 *Let A_n be a sequence of sets. Then we can always find mutually disjoint sets B_n such that $B_n \subset A_n$ and $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$.*

Remark 2.3.4 *Informally, this says that for any sequence of sets $\{A_n\}_{n=1}^{\infty}$ we can express its union as a collection of mutually disjoint sets B_n where the $B_n \subset A_n$. This result is very useful in advanced analysis.*

Proof. We write the proof by "constructing" the B_n sets. Let $B_1 = A_1$, $B_2 = A_2 \backslash A_1$ (so that $B_1 \cap B_2 = \emptyset$), $B_3 = A_3 \backslash (A_1 \cup A_2), \ldots, B_n = A_n \backslash \cup_{i=1}^{n-1} A_i$. Notice that constructing the sets in this way keeps the B_n mutually disjoint. Now since each of the B_n are defined as $A_n \backslash$(some set), this implies that $B_n \subset A_n$. Taking unions, we obtain $\cup_{n=1}^{\infty} B_n \subset \cup_{n=1}^{\infty} A_n$. We now need to show the reverse inclusion $\cup_{n=1}^{\infty} A_n \subset \cup_{n=1}^{\infty} B_n$. Take $x \in \cup_{n=1}^{\infty} A_n$. Then $x \in A_n$ for some $n \in \mathbb{N}$. Let $N \in \mathbb{N}$ be the smallest number for which $x \in A_N$ (in the extreme case, N would be equal to 1). Then $x \in A_N \backslash \cup_{i=1}^{N-1} A_i$ because $x \notin A_i$ for $i < N$. Hence $x \in B_N$ so certainly $x \in \cup_{n=1}^{\infty} B_n$. As a result, $\cup_{n=1}^{\infty} A_n \subset \cup_{n=1}^{\infty} B_n$ so we are done. ■

Continuing in this "constructive" flavor, we now construct one of the most famous sets known, the Cantor set K (named after Georg Cantor). We introduce this set here because it serves as a useful example in upcoming sections.

The Cantor set is a set of real numbers contained in the unit interval $[0, 1]$; that is, $K \subset [0, 1]$. Thus, if we intuitively define the "length" of an interval as

$$\text{length}([a, b]) = b - a$$

then we see that K must satisfy $\text{length}(K) \leq 1$. Of course, you may be thinking of sets such as $A = \left(\frac{1}{2}, \frac{3}{4}\right]$, $B = \{x \,|\, x \in [0, 1],\ x \in \mathbb{Q}\}$ or $C = \{x \,|\, x \in [0, 1],\ x = \frac{1}{2k},\ k \in \mathbb{N}\}$. However, the Cantor set is far more peculiar. In fact, it is so unique that none of A, B or C are subsets of K. Likewise, K is not a subset of either A, B nor C. So what is K? We construct this set below.

The Cantor set is formed by removing "open middle thirds"; that is, we are removing open intervals which account for $\frac{1}{3}$ of the entire interval.

- Step 1. Since we start with $[0,1]$, we first remove $\left(\frac{1}{3},\frac{2}{3}\right)$ so we are left with $K_1 = \left[0,\frac{1}{3}\right] \cup \left[\frac{2}{3},1\right]$. Note that length$(K_1) = \frac{2}{3}$. See Figure 2.6.
- Step 2. We continue this process. Since we have $\left[0,\frac{1}{3}\right] \cup \left[\frac{2}{3},1\right]$, we now remove their middle thirds. That is, we discard $\left(\frac{1}{9},\frac{2}{9}\right)$ and $\left(\frac{7}{9},\frac{8}{9}\right)$ and we are left with $K_2 = \left[0,\frac{1}{9}\right] \cup \left[\frac{2}{9},\frac{1}{3}\right] \cup \left[\frac{2}{3},\frac{7}{9}\right] \cup \left[\frac{8}{9},1\right]$. The size of the interval now shrinks to length$(K_2) = \frac{2}{3}\left(\frac{2}{3}\right) = \left(\frac{2}{3}\right)^2$. See Figure 2.7.
- $\vdots \qquad \vdots \qquad \vdots$
- Step n. We split all of the remaining intervals into thirds and remove their open middle thirds. We are left with the union of 2^n closed sets which we call K_n. Notice that length$(K_n) = \left(\frac{2}{3}\right)^n$.
- $\vdots \qquad \vdots \qquad \vdots$

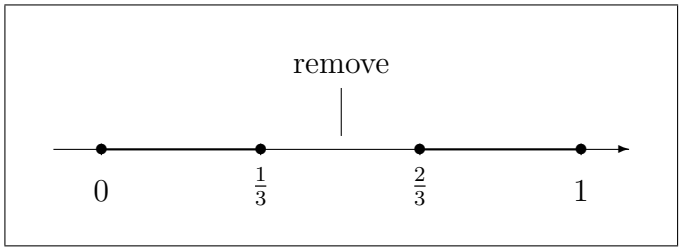

Figure 2.6: Step 1, K_1

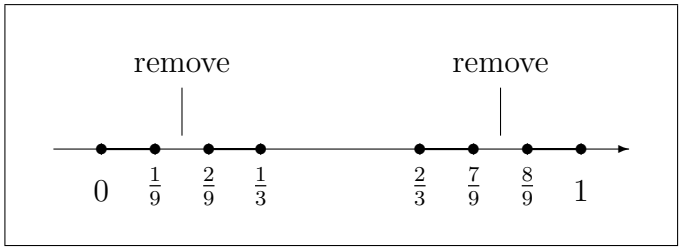

Figure 2.7: Step 2, K_2

At this point, we have not yet constructed the Cantor set K! We must continue this process indefinitely to obtain K. In other words, $K = K_\infty$ (informally). We also know that as $n \to \infty$, $\left(\frac{2}{3}\right)^n \to 0$ so that the Cantor set has length zero! However, a rather puzzling observation is that K contains all of the endpoints (an infinite number in fact) that we initiated from removing the middle thirds! This is quite peculiar but we will come to grips with this momentarily. Motivated by this construction, we have the following definition.

Definition 2.3.4 *Let* $\{K_n\}_{n=1}^{\infty}$ *be the sequence of sets immediately preceding. For example,* $K_3 = \left[0, \frac{1}{27}\right] \cup \left[\frac{2}{27}, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{7}{27}\right] \cup \left[\frac{8}{27}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{19}{27}\right] \cup \left[\frac{20}{27}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, \frac{25}{27}\right] \cup \left[\frac{26}{27}, 1\right]$. *Then the Cantor set* K *is defined as* $K = \bigcap_{n=1}^{\infty} K_n$.

Some people are naturally uneasy with the statement

$$\text{length}(K) = 0.$$

How can the Cantor set possibly have zero length if it is a union of intervals? To see this, we can consider K'. We know that $K \cup K' = [0, 1]$ since $U = [0, 1]$. Furthermore, the first interval in K' is $\left(\frac{1}{3}, \frac{2}{3}\right)$ and it has length $\frac{1}{3}$. Next, K' collects two intervals, $\left(\frac{1}{9}, \frac{2}{9}\right)$ and $\left(\frac{7}{9}, \frac{8}{9}\right)$, each with length $\frac{1}{9}$, et cetera (see Figures 2.6 and 2.7). Hence the length of K' can be found as follows:

$$\begin{aligned}
\text{length}(K') &= \frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \cdots + 2^n \cdot \frac{1}{3^{n+1}} + \cdots \\
&= \sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} \\
&= \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n \\
&= \frac{1}{3} \frac{1}{1 - \frac{2}{3}} \\
&= 1.
\end{aligned}$$

(Recall from Calculus that $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ for $|x| < 1$. This was used above.) So since $\text{length}(K') = 1$ and $K \cup K' = [0, 1]$, there is no "room" for K in the unit interval. Therefore, $\text{length}(K)$ must be zero. The notion of infinity is key here.

Remark 2.3.5 *Since* $\text{length}(K) = 0$, *this suggests that the set* K *is rather "small." Informally, it cannot contain very much if its length is zero! Sadly, this couldn't be further from the truth. We will investigate this in greater detail once we discuss countability. There, we will discover that* K *is so "large" that it is "uncountable." Strange ... but true!*

If you are not yet intrigued by some of the properties of the Cantor set then try this on for size. You probably recall the "dimensions" of simple geometric objects—a point is zero-dimensional, a line one-dimensional, a plane two-dimensional, and a solid three-dimensional. There is the interesting notion of "fractal-dimension" which can be applied to more abnormal entities such as the Cantor set (see any book on fractals or fractal geometry). It is heart-warming to know that this fractal-dimension is consistent with our ordinary definition of dimension (e.g., the fractal dimension of a line segment is 1). However, when we compute the fractal dimension of K we

obtain $\frac{\log 2}{\log 3} \approx 0.63$! This number, in some sense, is to be expected. The Cantor set is a beast caught somewhere between a point and a line. It is not quite a line but it is more than a point!

Even at this early stage, you probably realize that infinity is one of life's enigmas. In Chapter 3, we will see that there are different "sizes" of infinity and different rates at which we can approach infinity. For example, let $A_n = [-n, n]$ and $B_n = [-n^3, n^3]$. We see that even though $\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} B_n = (-\infty, \infty) = \mathbb{R}$, the structure of A_n and B_n are quite different. To see this, replace ∞ with 10 and then $\bigcap_{n=1}^{10} A_n \neq \bigcap_{n=1}^{10} B_n$. In fact, there is no $N \in \mathbb{N}$ such that $\bigcap_{n=1}^{N} A_n = \bigcap_{n=1}^{N} B_n$; the only N that "works" is $N = \infty$. This, in itself, is interesting.

Here is an example that shows how infinity can play tricks on you. You might recall from Calculus that

$$\begin{aligned}\ln 2 &= \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} \\ &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \cdots\end{aligned}$$

So we have

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots$$

$$\begin{aligned}&= (2-1) - \frac{1}{2} + \left(\frac{2}{3} - \frac{1}{3}\right) - \frac{1}{4} + \left(\frac{2}{5} - \frac{1}{5}\right) - \frac{1}{6} + \left(\frac{2}{7} - \frac{1}{7}\right) - \frac{1}{8} + \cdots \\ &= 2 - 1 + \frac{2}{3} - \frac{1}{2} + \frac{2}{5} - \frac{1}{3} + \frac{2}{7} - \frac{1}{4} + \cdots \\ &= 2\left[1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots\right] \\ &= 2 \ln 2.\end{aligned}$$

The equalities above show that $\ln 2 = 2 \ln 2$ or $1 = 2$! This is a simple consequence of $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$ being conditionally convergent−a term you need not worry about now. Simply stated, some drastic things occur at infinity! Notice that if we truncate the series at a particular point, this result will not work out. However, the fact that this series is *infinite* allows the faulty logic to seep in ...

We close this section with some additional open thoughts on infinity. Here is the (seemingly ridiculous) problem that Pickover [18] suggested sometime ago:

> Question: How many numbers contain the digit three?
>
> Answer: Virtually all of them!

Why is this answer an accurate one? you may ask. The answer to this question is couched in the idea of the infinite. To see this, we split the analysis into stages, much like the Cantor set. For example, if we consider the first 10 digits

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ \text{and}\ 10,$$

it is clear that only one number contains the digit 3. Thus, at Stage 1, 10% of the numbers contain the desired digit.

Stage 2. Consider the first $10^2 = 100$ digits given by $1, 2, 3, \ldots, 99, 100$. The percentage almost doubles to 19%. You can check this!

Stage 3. Consider the first $10^3 = 1000$ digits−now we are up to 27%.

Continuing in this fashion we obtain ...

Stage n. Consider the first 10^n digits−we reach $\left[1 - \left(\frac{9}{10}\right)^n\right] \times 100\%$.

We see that as $n \to \infty,\ 1 - \left(\frac{9}{10}\right)^n \to 1$ so we obtain 100%, an indication that nearly all numbers contain the digit 3!

Why is this counterintuitive and absurd? Well, it turns out that we are ignoring the other side of the coin here−that is, we are ignoring an entire "infinity" elsewhere in the problem. In essence, even though the quantity of numbers containing the digit three is rapidly approaching infinity, the quantity of numbers not containing the digit three is *also* approaching infinity! To see this, let A_n = the quantity of numbers containing a 3 at stage n and let B_n = the quantity of numbers not containing a 3 at stage n. Then we have

$$\begin{aligned}
(A_1, B_1) &= (1, 9) \\
(A_2, B_2) &= (19, 81) \\
(A_3, B_3) &= (271, 729) \\
\vdots \quad \vdots \quad &= \quad \vdots \quad \vdots
\end{aligned}$$

so it is readily seen that both A_n and B_n approach the infinite!

To give you a final warning on infinity, we state a few "definitions" that are conventions in mathematics.

1. $\infty + \infty = \infty$
2. $(-\infty) + (-\infty) = -\infty$
3. $\infty \cdot \infty = \infty$

Just a quick glance at the first property brings forth an interesting thought. That is, if we treat ∞ like an ordinary number, how many other numbers a have a property that looks like $a + a = a$? Other questions that you may want to ponder are how to define expressions such as $\infty \cdot 0$ or $(-\infty) + \infty$. How about division by $\pm\infty \ldots$ what is the interpretation there?

Remark 2.3.6 *Most people agree that* $\frac{1}{\infty} = 0$, *or more generally,* $\frac{a}{\pm\infty} = 0$ *for* $a \in \mathbb{R}$. *Since this offers a close relationship between zero and infinity, we must be cautious with aritmetic computations involving these quantities (see Exercise 10 in this section for example). To cite an additional example, it is again conventional to say that* $a \cdot \infty = \infty$ *for* $a \in \mathbb{R}^+$. *Therefore, a statement such as* $3 \cdot \infty = \infty = 5 \cdot \infty$ *is valid. Can we then say that* $3 = 5$ *by "cancelling" the* ∞*'s? The answer is certainly no...but why? Well, loosely speaking,* ∞ *is not a real number; rather, it is an abstract entity. As a consequence, it does not share all of the nice cancellation properties with which we are so familiar.*

The message here is that infinity is a strange beast that should be handled with great care. We'll see this shortly...

Exercises.

1. Prove De Morgan's Laws. Notice that since these involve proving statements for all $n \in \mathbb{N}$, you should use induction.
   (a) $\left(\bigcup_{n=1}^{\infty} A_n\right)' = \bigcap_{n=1}^{\infty} A_n'$
   (b) $\left(\bigcap_{n=1}^{\infty} A_n\right)' = \bigcup_{n=1}^{\infty} A_n'$

2. Consider the following sets $A_n = \left(0, \frac{1}{n}\right]$, $B_n = \left(\frac{1}{n}, n\right)$, $C_n = \left[2, 2 + \frac{1}{n}\right]$, $D_n = [-n^2, 0)$, and $E_n = (n-2, n]$. Find the union and intersection over $\mathbb{N}$ for each of the sets $A_n, \ldots, E_n$. Are the sequences of sets increasing, decreasing, or neither? Are any of the collections mutually disjoint?

3. For each sequence of sets $\{A_n\}_{n=1}^{\infty}$ listed below, find a sequence $\{B_n\}_{n=1}^{\infty}$ that satisfies **Proposition 2.3.1**. That is, find mutually disjoint B_n with $B_n \subset A_n$ and $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$.
   (a) $A_n = [n+1, n+2)$
   (b) $A_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$
   (c) $A_n = \left[1, 2 - \frac{1}{n}\right]$

4. Consider the interval $[0, 1]$. Here you will be asked to construct a Cantor-like set $\tilde{K}$. In this problem, try removing "middle-halves." In other words, $\tilde{K}_1 = \left[0, \frac{1}{4}\right] \cup \left[\frac{3}{4}, 1\right]$, $\tilde{K}_2 = \left[0, \frac{1}{16}\right] \cup \left[\frac{3}{16}, \frac{1}{4}\right] \cup \left[\frac{3}{4}, \frac{13}{16}\right] \cup \left[\frac{15}{16}, 1\right]$, et cetera. Do you

see that this problem is different in nature already? In the Cantor set K, we removed a "middle third" and we were left with two sets, each having length one-third. Here we removed the "middle-half" and are left with two sets, each having length one-fourth.

(a) Find $\tilde{K}_n$ and length$(\tilde{K}_n)$ for $n = 1, 2, 3$ and 4. Note that we did most of the problem for $n = 1$ and 2 above.

(b) Find an expression for length$(\tilde{K}_n)$, $n \in \mathbb{N}$. What happens to length$(\tilde{K}_n)$ as $n \to \infty$? Is this surprising?

(c) Prove that length$(\tilde{K}) = 0$. You may wish to do this by showing that length$(\tilde{K}') = 1$.

(d) What are your thoughts on removing "middle-fourths" instead of "middle-halves"? How about "middle-fifths"?...."middle $\frac{1}{n}$'s" for n large?

5. This question is related to Pickover's question (review the discussion on pages 45-46). Answer **true** or **false** and defend your position.

(a) Even though there is an infinity of numbers containing the digit 3, there are far more that do not contain the digit 3.

(b) We know that $(A_n, B_n) \to (\infty, \infty)$ as $n \to \infty$. $B_n \to \infty$ $(n \to \infty)$ at a more rapid pace than $A_n \to \infty$ $(n \to \infty)$.

6. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets and let B be a set. Prove the following generalizations for infinite collections of sets.

(a) $B \cup \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} (B \cup A_n)$

(b) $B \backslash (\bigcap_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} (B \backslash A_n)$

7. State and prove the analogues for Problem 6 with $\bigcup_{n=1}^{\infty} A_n$ on the left-hand side instead of $\bigcap_{n=1}^{\infty} A_n$.

8. Let $i, j \in \mathbb{N}$ with $i \leq j$. Prove the following:

(a) $\bigcup_{n=i}^{j} A_n \subset \bigcup_{n=1}^{\infty} A_n$

(b) $\bigcap_{n=1}^{\infty} A_n \subset \bigcup_{n=i}^{j} A_n$

9. The ternary expansion of $x \in \mathbb{R}$ uses only the numbers 0, 1, and 2 (similar to binary expansions using only 0 and 1). So for $x \in (0, 1)$, writing

$$x \overset{3}{=} 0.a_1a_2a_3a_4a_5 \cdots$$

means that

$$x \overset{3}{=} \sum_{n=1}^{\infty} \frac{a_n}{3^n} = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots$$

So, for example, $\frac{1}{2} \overset{3}{=} 0.\overline{1}$ since $0.\overline{1} \overset{3}{=} \sum_{n=1}^{\infty} \frac{1}{3^n} = \frac{1}{3}\sum_{n=1}^{\infty} \frac{1}{3^{n-1}} = \frac{1}{3}\sum_{m=0}^{\infty} \frac{1}{3^m} = \frac{1}{3}\frac{1}{1-\frac{1}{3}} = \frac{1}{2}$.

(a) Show that $\frac{1}{3}$ has <u>two</u> representations: $0.1\overline{0}$ and $0.0\overline{2}$.

(b) Find the two representations for $\frac{2}{3}$ (Answer: $0.2\overline{0}$ and $0.1\overline{2}$).

(c) How about the numbers $\frac{1}{9}, \frac{2}{9}, \frac{7}{9}$, and $\frac{8}{9}$? How many ternary expansions do they have? Do you see a connection to the "endpoints" from the Cantor set K?

(d) Based on the previous exercises, prove that every $x \in (0,1)$ has <u>at most</u> two ternary expansions.

(e) Prove that $x \in K$ IFF x does <u>not</u> contain a 1 in its ternary expansion (i.e., $x \overset{3}{=} 0.a_1a_2a_3a_4a_5\cdots$, $a_i = 0$ or 2, $i \in \mathbb{N}$). Notice that in the case of multiple representation, we conveniently choose the representation without the 1 in the expansion. *Hint:* Note that any $y \in \left[0, \frac{1}{3}\right]$ looks like $y \overset{3}{=} 0.0b_2b_3b_4b_5\cdots$ (e.g., $\frac{1}{3}$ has all $b_i = 2$). Also, any $z \in \left[\frac{2}{3}, 1\right]$ looks like $z \overset{3}{=} 0.2c_2c_3c_4c_5\cdots$ (e.g., $\frac{2}{3}$ has all $c_i = 0$). Then observe that any x such that $x \in \left(\frac{1}{3}, \frac{2}{3}\right)$ (not in K!) <u>must</u> look like $x \overset{3}{=} 0.1d_2d_3d_4d_5\cdots$

10. Find the error in the following argument.

Statement. Let $a = b$, $a, b \in \mathbb{R}$ and $a, b \neq 0$. Then $1 = 2$.

Proof. $a = b$ so $a^2 = ab$. Thus, $a^2 - b^2 = ab - b^2$ so that $(a+b)(a-b) = b(a-b)$. Hence, $a + b = b$ or $b + b = b$ so $2b = b$. Finally, $2 = 1$. ■

11. Comment on how the previous problem is similar to claiming that $3 \cdot \infty = \infty = 5 \cdot \infty$ implies that $3 = 5$.

12. Fill in the blanks with a number from $\mathbb{R}^e$ or write "impossible to determine." (These latter forms, you may recall, are usually called "indeterminate forms.") Give justifications for your claims.

(a) For $a < 0$, $a \cdot (-\infty) =$

(b) $-\infty - \infty =$

(c) For $a \in \mathbb{R}$, $a - \infty =$

(d) $0^{\infty} =$

(e) For $a > 1$, $a^{-\infty} =$

(f) For $0 < a < 1$, $a^{-\infty} =$

(g) For $a = 1$, $a^{\infty} =$

(h) $\infty^0 =$

(i) $\infty^{-\infty} =$

2.4 Summary: Odds and Ends

- To prove $A \subset B$, assume $x \in A$ and prove that $x \in B$. To establish $A = B$, show that $A \subset B$ and $B \subset A$.
- Given any set A, both $\emptyset$ and A are subsets.
- The power set of A, denoted by $\mathcal{P}(A)$, is the set of all subsets of A.
- Union:
  1. $A \cup B = \{x \,|\, x \in A \text{ or } x \in B\}$
  2. $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n=1}^{\infty} A_n = \{x \,|\, \exists\, n \in \mathbb{N} \ni x \in A_n\}$
- Intersection:
  1. $A \cap B = \{x \,|\, x \in A \text{ and } x \in B\}$
  2. $\bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n=1}^{\infty} A_n = \{x \,|\, \forall\, n \in \mathbb{N}, x \in A_n\}$
- $A \backslash B = \{x \,|\, x \in A \text{ and } x \notin B\}$
- $A' = U \backslash A$
- DeMorgan's Laws:
  1. $(A \cup B)' = A' \cap B'$
  2. $(A \cap B)' = A' \cup B'$

  Generalizations:
  1. $(\bigcup_{n=1}^{\infty} A_n)' = \bigcap_{n=1}^{\infty} A'_n$
  2. $(\bigcap_{n=1}^{\infty} A_n)' = \bigcup_{n=1}^{\infty} A'_n$
- Associative Laws:
  1. $(A \cap B) \cap C = A \cap (B \cap C)$

2. $(A \cup B) \cup C = A \cup (B \cup C)$

- Distributive Laws:

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

- Let A_n be a sequence of sets. Then we can always find mutually disjoint sets B_n such that $B_n \subset A_n$ and $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$.

- Let K_n be the union of 2^n closed sets found by removing "open middle thirds" at the n^{th} step. The Cantor set K is defined by $K = \cap_{n=1}^{\infty} K_n$.

Chapter 3

Functions

In this section, we begin our study of perhaps the most ubiquitous notion in mathematics, that of a function. You may remember functions such as $y = x^2$ or $f(x) = \lfloor x \rfloor$. You may also recall that a simple graph such as $x^2 + y^2 = 1$ (a circle) does not represent a function. In this section, we will first introduce the concept of a relation; then functions will be defined as special types of relations. Next, as we undertake a special class of functions (those that are one-to-one), we will begin to grasp the notion of infinity−something that we struggled with in the last section.

3.1 Relations

We begin by studying a very familiar object: the ordered pair. Recall that the coordinates $(1, 2)$ and $(2, 1)$ are both ordered pairs but they are very different! If you plot $(1, 2)$ and $(2, 1)$ on the (Cartesian) $xy-$plane, they represent different locations. Therefore, it is safe for us to say that two ordered pairs (a_1, a_2) and (b_1, b_2) are equal IFF $a_1 = b_1$ and $a_2 = b_2$.

Remark 3.1.1 *Do not confuse the notation* $(1, 2)$ *as meaning the set (or open interval)* $\{x \in \mathbb{R} \mid 1 < x < 2\}$ *as we saw earlier. The notations are identical and this is rather unfortunate. However, it should be clear from context whether we are talking about intervals or coordinates.*

Remark 3.1.2 *Notice the difference between ordered pairs and sets. For example,* $A = \{1, 2\} = \{2, 1\}$ *because* A *is a set with two members, the numbers* 1 *and* 2*. However,* $(1, 2) \neq (2, 1)$.

Definition 3.1.1 *Let* A *and* B *be sets. The Cartesian Product of* A *and* B*, denoted by* $A \times B$*, is the set of all ordered pairs* (a, b) *with* $a \in A$ *and* $b \in B$*. Symbolically,*

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

See the figure.

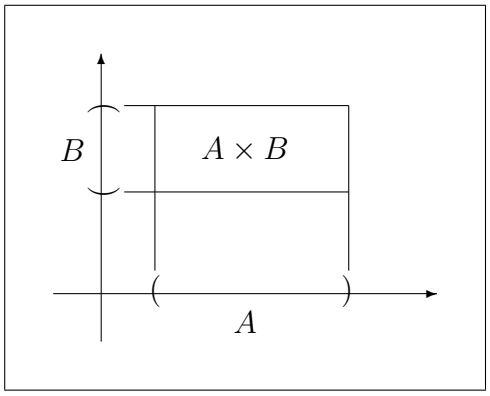

Figure 3.1: Cartesian Product $A \times B$

Example 3.1.1 *Let* $A = \{a, b, c\}$ *and* $B = \{1, 2\}$. *Then*

$$A \times B = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}$$

and

$$B \times A = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}.$$

Note that, in general, $A \times B \neq B \times A$. *See the next two figures.*

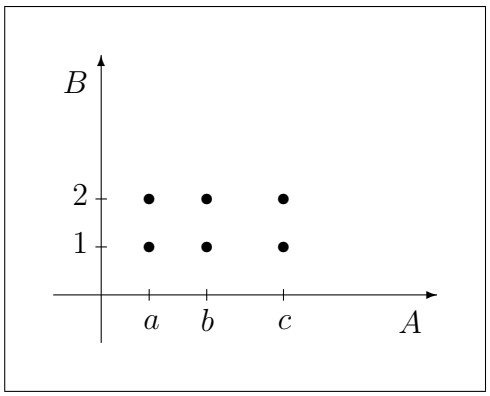

Figure 3.2: $A \times B$

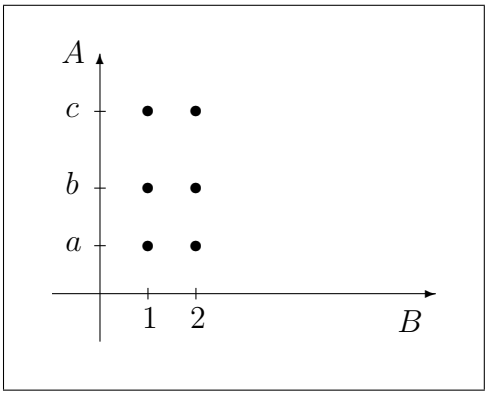

Figure 3.3: $B \times A$

Example 3.1.2 *Let A and B be as above. Let $C = \{\delta, \epsilon\}$. Find $(A \times B) \times C$ and $A \times B \times C$. Are they the same?*

Solution. From the previous example we already have $A \times B$. Therefore,

$$\begin{aligned}(A \times B) \times C = \\ \{&((a,1),\delta),((a,1),\epsilon),((a,2),\delta),((a,2),\epsilon),((b,1),\delta),((b,1),\epsilon),\\ &((b,2),\delta),((b,2),\epsilon),((c,1),\delta),((c,1),\epsilon),((c,2),\delta),((c,2),\epsilon)\}.\end{aligned}$$

We notice that $(A \times B) \times C$ is still a set of ordered pairs. On the other hand,

$$\begin{aligned}A \times B \times C = \\ \{&(a,1,\delta),(a,1,\epsilon),(a,2,\delta),(a,2,\epsilon),(b,1,\delta),(b,1,\epsilon),\\ &\{(b,2,\delta),(b,2,\epsilon),(c,1,\delta),(c,1,\epsilon),(c,2,\delta),(c,2,\epsilon)\}\end{aligned}$$

since we define $A \times B \times C = \{(a,b,c) \,|\, a \in A, b \in B, c \in C\}$. It is clear that $(A \times B) \times C \neq A \times B \times C$ since ordered pairs and ordered triples are fundamentally different.

We now prove some important results regarding the Cartesian product.

Proposition 3.1.1 *Let A, B and C be sets. Then*

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

Remark 3.1.3 *So even if the operation $\times$ fails to commute and disobeys "grouping" (evident from the previous two examples), we still have the distributive law holding for this operation.*

Proof. We prove the first statement and leave the second to the exercises. As usual, since we are showing equality of sets, we need to show inclusion both ways. So let $(x, y) \in A \times (B \cup C)$ (note that the elements in this set are ordered pairs (x, y), not singletons like x). Then $x \in A$ and $y \in B \cup C$ so that $x \in A$ and $y \in B$ or $y \in C$. Thus, we could have $x \in A$ and $y \in B$ (meaning $(x, y) \in A \times B$) or we could have $x \in A$ and $y \in C$ (meaning $(x, y) \in A \times C$). It follows that $(x, y) \in (A \times B) \cup (A \times C)$. Thus we have shown $A \times (B \cup C) \subset (A \times B) \cup (A \times C)$. To show that $(A \times B) \cup (A \times C) \subset A \times (B \cup C)$, it is a simple matter of reversing the steps that we just presented (the details are left to you). Once this is established, we have $A \times (B \cup C) = (A \times B) \cup (A \times C)$, the desired result. ■

Proposition 3.1.2 *Let A and B be sets. Then $(A \backslash B) \times B = (A \times B) \backslash (B \times B)$.*

Proof. Let $(x, y) \in (A \backslash B) \times B$ so that $x \in A \backslash B$ and $y \in B$. In other words, $x \in A$ (but $x \notin B$) and $y \in B$. So it is safe to say that $(x, y) \in A \times B$. But notice that $x \notin B$ implies that $(x, y) \notin B \times B$. Saying that $(x, y) \in A \times B$ and $(x, y) \notin B \times B$ is identical to $(x, y) \in (A \times B) \backslash (B \times B)$. Hence $(A \backslash B) \times B \subset (A \times B) \backslash (B \times B)$. In the other direction, let $(u, v) \in (A \times B) \backslash (B \times B)$. We see immediately that $(u, v) \in A \times B$ but $(u, v) \notin B \times B$. Since $(u, v) \in (A \times B)$ tells us that $u \in A$ and $v \in B$, then $(u, v) \notin B \times B$ must mean that $u \notin B$ so that $u \in A \backslash B$. Hence $(u, v) \in (A \backslash B) \times B$. Thus, $(A \times B) \backslash (B \times B) \subset (A \backslash B) \times B$ so we are done. ■

After all of this business, we arrive at the definition of a relation.

Definition 3.1.2 *Let A and B be sets. A relation R from A to B is a subset of $A \times B$.*

Remark 3.1.4 *An immediate consequence of this definition is that $A \times B$ and $\emptyset$ are both relations from A to B. See* **Proposition 2.1.1** *and* **Proposition 2.1.2**.

Example 3.1.3 *Using the A and B from our earlier example,*

$$\{(a, 1), (a, 2), (b, 2), (c, 1)\}$$

is a relation from A to B while

$$\{(2, c), (1, b)\}$$

is a relation from B to A. Notice that these collections of points are just subsets of the points given by $A \times B$ and $B \times A$, respectively (see Figures 3.2 and 3.3). If we denote the first relation (from A to B) as R, we write, for example, $(a, 1) \in R$. Another accepted notation is $aR1$. In general, writing xRy just means that "x is associated with y via the relation R."

At this point, we could go on to define the domain and range of a relation, inverse relations, and composite relations. However, we do not do this here; the groundwork of relations has been set to define the more powerful idea of function. There, we will speak of the domain and range of functions, inverse functions, and composite functions. We give a brief warning before bringing this section to a close−that is, "relation" is not synonymous with "function." You need not be concerned with the fact that we have not yet defined the word function; for now, our approach will be very informal.

Example 3.1.4 *In the last example, we had the relation*

$$R = \{(a, 1), (a, 2), (b, 2), (c, 1)\}$$

from A to B. If we associate different values with a, b, and c, we have the "graph" below:

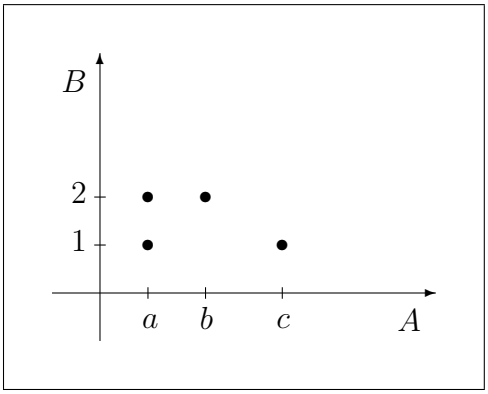

Figure 3.4: A relation from A to B

Now if you think of the AB plane as similar to the Cartesian plane, you see that this relation fails the "vertical line test" (you may recall that this test is essentially the all-or-nothing test for functions). So R is not a function.

The next example is somewhat more convincing.

Example 3.1.5 *Let $T = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 \leq 1\}$. Certainly, T is a relation since it is a subset of $\mathbb{R} \times \mathbb{R}$−the familiar $\mathbb{R}^2$ plane. In particular, we recognize region T as the unit disc (the unit circle along with its interior). See the figure on the following page.*

This is added assurance that relations are not necessarily functions.

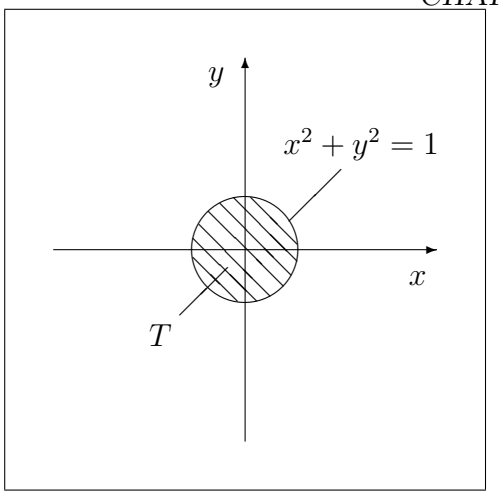

Figure 3.5: $T = \{(x, y) \in \mathbb{R} \times \mathbb{R} \,|\, x^2 + y^2 \leq 1\}$

Exercises.

1. Answer true or false. If $A = \{x, y, z\}$ and $B = \{y, z, a\}$ then

   (a) $(x, z) \in A \times B$

   (b) $(z, z) \notin A \times B$

   (c) $\{(x, y), (y, a)\} \subset A \times B$

   (d) $(y, y, y) \in (A \times B) \times A$

   (e) $\{z, a\} \in B \times A$

2. Find sets A, B and C so that $(A \times B) \times C \neq A \times (B \times C)$.

3. Prove that $A \times \emptyset = \emptyset$ for any set A.

4. Let A, B, C and D be sets. Prove that

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D).$$

5. (a) Similar to problem 4, show that

$$(A \times B) \cup (C \times D) \subset (A \cup C) \times (B \cup D).$$

   (b) Show that you cannot replace $\subset$ with $=$ in the statement above. That is, give a counterexample that disproves

$$(A \times B) \cup (C \times D) = (A \cup C) \times (B \cup D).$$

6. Prove that if A has m elements and B has n elements, then $A \times B$ has mn elements.

7. Consider the Cartesian product of a line segment and a square. What is the geometric object that results?

8. We know that, in general, $A \times B \neq B \times A$. Prove that $A \times B = B \times A$ IFF $A = B$.

9. Graph the relations below. With our work thus far, which of these would you call functions?

   (a) $R = \{(1, 2), (3, 4), (5, 6)\}$
   (b) $S = \{(x, y) \in \mathbb{R} \times \mathbb{R} \,|\, x + y = 7\}$
   (c) $T = \{(x, y) \in \mathbb{R} \times \mathbb{R} \,|\, x = 7\}$
   (d) $U = \{(x, y) \in \mathbb{R} \times \mathbb{R} \,|\, y^2 = x + 2\}$

10. Let $A = [0, 1]$, $B = \left(-\frac{1}{2}, \frac{1}{2}\right)$, and $C = \left[\frac{1}{4}, 2\right)$ so A, B, and C represent intervals. Sketch the following:

    (a) $C \times B$
    (b) $B \times C$
    (c) $(A \cup B) \times C$

11. (a) Let $A = \{a, b\}$. Find all of the relations from A to A.
    (b) Do the same for $B = \{1, 2, 3\}$.
    (c) Conjecture the number of relations from A to A if A has n elements.
    (d) Generalize: If A has m elements and B has n elements, how many relations are there from A to B?

12. Consider $A = [a, b]$ and $B = [c, d]$ where $a < b$ and $c < d$. Is it suitable to consider all relations from A to B? Why or why not?

13. Prove the second part of **Proposition 3.1.1**.

3.2 Functions and Images

We begin with the most important definition of this chapter.

Definition 3.2.1 *A function f from A into B, denoted by $f : A \to B$, is a subset of $A \times B$ with the property that for each $x \in A$ there is exactly one $y \in B$. Using different language, given $x \in A$, there exists precisely one ordered pair $(x, y) \in f$.*

It should now be clear that a function is a special type of relation−that is, a relation that does not permit $x \in A$ to pair off with two distinct elements in B. In other words, if $(x, y_1) \in f$ and $(x, y_2) \in f$ and $y_1 \neq y_2$, then the relation f is not a function.

Remark 3.2.1 *Since the statement $(x, y) \in f$ gets awkward quickly, it is more common to write $y = f(x)$.*

Definition 3.2.2 *Let f be a function. If $y = f(x)$ then y is said to be the image of x under f.*

Definition 3.2.3 *Let $f : A \to B$. The set A is called the domain of f and is denoted by $\mathcal{D}(f)$. The range of f, denoted by $\mathcal{R}(f)$, is the set of all $b \in B$ such that $b = f(a)$ for some $a \in A$. That is,*

$$\mathcal{R}(f) = \{b \in B \,|\, b = f(a) \text{ for some } a \in A\}.$$

Remark 3.2.2 *Note that $\mathcal{R}(f) \subset B$, though $\subset$ may be replaced with $=$ in some instances (see Example 3.2.2). Glance at the figure below.*

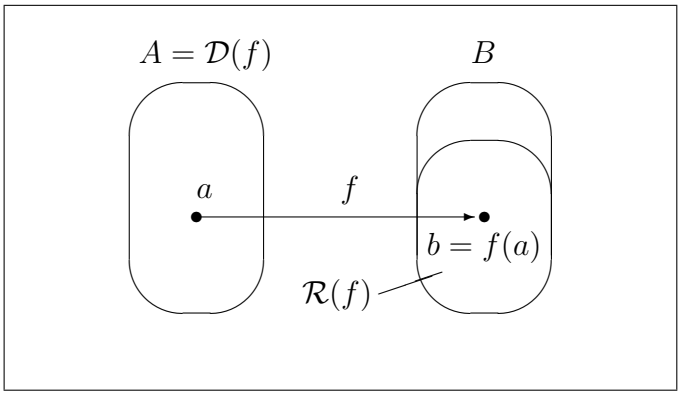

Figure 3.6: The range of f

Example 3.2.1 *Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $y = f(x) = x^2$. Note that for each $x_0 \in \mathbb{R}$, there is only one ${x_0}^2 \in \mathbb{R}$. Some ordered pairs are $(\pi, \pi^2), (-1, 1), (1, 1)$ and $(1.2, 1.44)$. We can say that 1.44 is the image of 1.2 under f. Note that $\mathcal{D}(f) = \mathbb{R}$ and $\mathcal{R}(f) = [0, \infty) \subset \mathbb{R}$. See the figure below.*

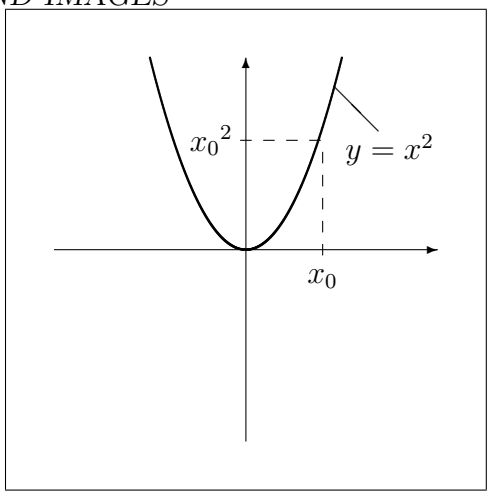

Figure 3.7: The function $y = f(x) = x^2$

Example 3.2.2 *Let $h : \mathbb{R} \to \mathbb{R}$ be defined as $h(x) = x^3$. Then $\mathcal{R}(h) = \mathbb{R}$.*

Example 3.2.3 *Let $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4, 5\}$. Consider the relation f as defined below:*

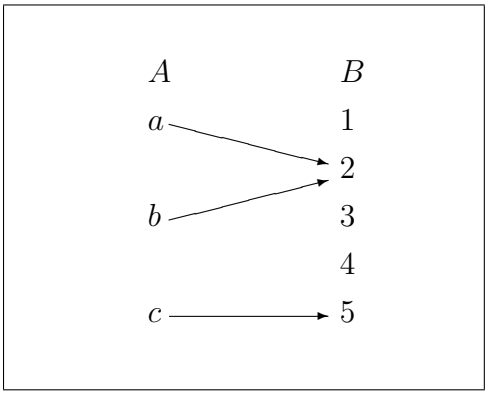

Figure 3.8: A relation f from A to B

From the diagram, $(a, 2) \in f$, $(b, 2) \in f$, and $(c, 5) \in f$. Indeed f is a function and $\mathcal{R}(f) = \{2, 5\} \subset \{1, 2, 3, 4, 5\} = B$.

Example 3.2.4 *Let $X = \{u, v, w\}$ and $Y = \{x, y, z\}$. Consider the relation g as defined below:*

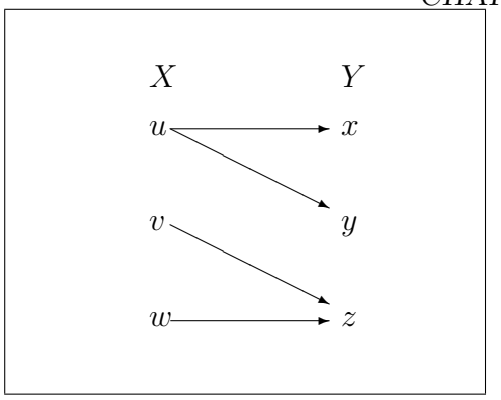

Figure 3.9: A relation g from X to Y

Note that since both (u, x) and (u, y) are in g, g is not a function.

We state two more definitions before moving on.

Definition 3.2.4 *Let $f : A \to B$ and let $C \subset A$. Then denote*

$$f(C) = \{f(x) \,|\, x \in C\}.$$

$f(C)$ is called the image of C under f.

See the figure.

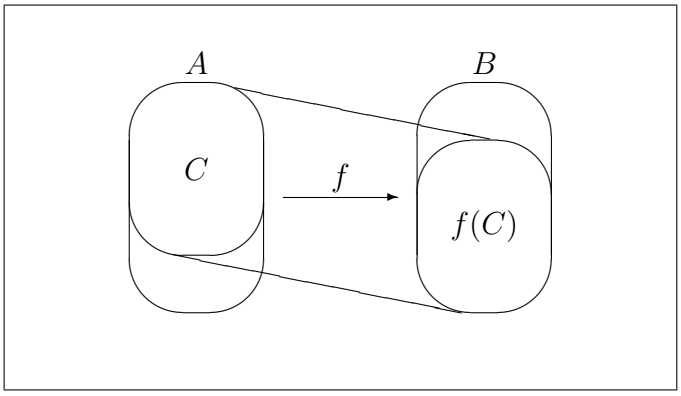

Figure 3.10: The image of C under f

Definition 3.2.5 *Let $f : A \to B$ and let $D \subset B$. Then denote*

$$f^{-1}(D) = \{x \in A \,|\, f(x) \in D\}.$$

$f^{-1}(D)$ is called the inverse image of D under f.

See the figure.

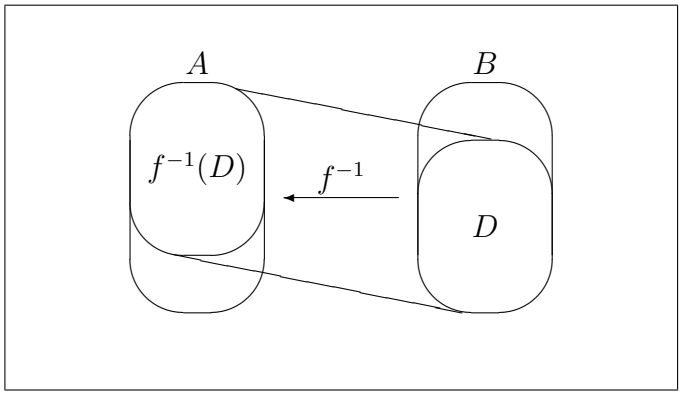

Figure 3.11: The inverse image of D under f

Remark 3.2.3 *Notice that $f^{-1}(D) \subset A$ whereas $f(C) \subset B$.*

Example 3.2.5 *Look at Example* 3.2.3. *There we have $f(\{a,b\}) = \{2\}, f^{-1}(\{2\}) = \{a,b\}, f^{-1}(\{5\}) = \{c\}$, and $f^{-1}(\{3,4\}) = \emptyset$. The equation $f^{-1}(\{2\}) = \{a,b\}$ is sometimes written as $f^{-1}(2) = \{a,b\}$ for simplicity.*

Example 3.2.6 *Look at Example* 3.2.1 *where $f(x) = x^2$. Then, for example, $f^{-1}(4) = \{2,-2\}, f\left(\left(-\frac{1}{3}, \frac{1}{2}\right)\right) = \left[0, \frac{1}{4}\right)$, and $f^{-1}(-18) = \emptyset$.*

We ask you to prove the following proposition (see the exercises).

Proposition 3.2.1 *Let $f : A \to B$ and let $X \subset A$ and $Y \subset B$. Then $f(X) \subset Y$ IFF $X \subset f^{-1}(Y)$.*

You may also try proving the following fact: Given $f : A \to B$, we have $f(x) \in B$ IFF $x \in f^{-1}(B)$. This fact is used in the proof of the following fundamental result on functions.

Theorem 3.2.1 *Let $f : A \to B$ with $X, Y \subset A$ and $C, D \subset B$. Then the following statements hold.*

1. $f(X \cup Y) = f(X) \cup f(Y)$
2. $f(X \cap Y) \subset f(X) \cap f(Y)$
3. $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$
4. $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$

Remark 3.2.4 *You may be thinking that the previous theorem holds for infinite collections of sets as well. This is, in fact, true. For example,* $f\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} f(A_n)$, *where* $A_n \subset A,\ n \in \mathbb{N}$. *See part (1) of the theorem.*

Notice the subtle inclusion in part (2) of the theorem; we leave its proof as an exercise. However, if we desire to simply show that $f(X \cap Y) \neq f(X) \cap f(Y)$ (in general), it is enough to produce a counterexample. For example, we already know that $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is a function. Now let $X = [-1, 0]$ and $Y = [0, 1]$ so that $X, Y \subset \mathbb{R}$. Then $X \cap Y = \{0\}$ so that $f(X \cap Y) = f(\{0\}) = \{0\}$. However, $f(X) = f(Y) = [0, 1] \Rightarrow f(X) \cap f(Y) = [0, 1]$. Clearly, $f(X \cap Y) \neq f(X) \cap f(Y)$.

Proof. We prove parts 1 and 4 of the theorem. The others are left as exercises.

- Part 1

  1. To prove $f(X \cup Y) \subset f(X) \cup f(Y)$, let $b \in f(X \cup Y)$. Then there is an $a \in X \cup Y$ such that $f(a) = b$. So either $a \in X$ or $a \in Y$. Therefore, either $b = f(a) \in f(X)$ or $b = f(a) \in f(Y)$. That is, $b \in f(X) \cup f(Y)$.
  2. To show $f(X) \cup f(Y) \subset f(X \cup Y)$, we let $c \in f(X) \cup f(Y)$. Then $c \in f(X)$ or $c \in f(Y)$. This says that c is the image of some point in X or some point in Y. Then certainly c is the image of some point in $X \cup Y$. That is, $c \in f(X \cup Y)$. We're finished. ■

- Part 4

  1. To show that $f^{-1}(C \cap D) \subset f^{-1}(C) \cap f^{-1}(D)$, we let $a \in f^{-1}(C \cap D)$. Then we know that $f(a) \in C \cap D$ so that $f(a) \in C$ and $f(a) \in D$. Thus, $a \in f^{-1}(C)$ and $a \in f^{-1}(D)$ and hence $a \in f^{-1}(C) \cap f^{-1}(D)$.
  2. To prove that $f^{-1}(C) \cap f^{-1}(D) \subset f^{-1}(C \cap D)$, we assume $b \in f^{-1}(C) \cap f^{-1}(D)$ so that $b \in f^{-1}(C)$ and $b \in f^{-1}(D)$. In other words, $f(b) \in C$ and $f(b) \in D$ so that $f(b) \in C \cap D$. Therefore, $b \in f^{-1}(C \cap D)$. This proves the result. ■

Here is an interesting problem.

Problem 3.2.1 *Let* $f : A \to B$, $D \subset A$. *Disprove the following statement: if* $f^{-1}(y) \not\subset D$ *then* $y \notin f(D)$.

Solution. Notice that there are two negations here ($\not\subset$ and $\notin$). Plus, we are trying to disprove this! If we write the contrapositive, we have "if $y \in f(D)$ then $f^{-1}(y) \subset D$." This seems much easier to disprove! Let us consider the familiar $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Let $D = [0, 10] \subset \mathbb{R}$. Note that $f(D) = [0, 100]$. Now letting $y = 25 \in [0, 100]$, we see that $f^{-1}(25) = \{5, -5\} \not\subset [0, 10]$. In other words, we've

produced an example where even though $y \in f(D)$, $f^{-1}(y)$ need not be a subset of D. □

Here is another theorem.

Theorem 3.2.2 *Let $f : A \to B$ with $X \subset A$ and $Y \subset B$. Then*

1. $f(f^{-1}(Y)) \subset Y$
2. $X \subset f^{-1}(f(X))$

Proof. For the first statement, let $y \in f(f^{-1}(Y))$. Then there must be an $x \in f^{-1}(Y)$ such that $y = f(x)$. Now $x \in f^{-1}(Y)$ says that $f(x) \in Y$ so there is a $\hat{y} \in Y$ such that $f(x) = \hat{y}$. However, since $y = f(x)$ then $y = f(x) = \hat{y} \in Y$. The proof of the second statement is left as an exercise. ■

Remark 3.2.5 *Many first-time readers feel that the statements above should read $f(f^{-1}(Y)) = Y$ and $X = f^{-1}(f(X))$ since f and f^{-1} have a way of "undoing" one another. The fact is, no such meaning was cast upon the symbol f^{-1} from the start! Similar to earlier remarks we have made, remember that you may convince yourself that $f(f^{-1}(Y)) = Y$ is false by simply producing a counterexample. Likewise for the second statement.*

Problem 3.2.2 *Let $f, g, h : [0, 1] \to \mathbb{R}$. Prove or disprove:* $\min(f+g, h) \leq \min(f, h)+\min(g, h)$.

Solution. Notice that this almost looks like the Triangle Inequality but the statement is false (in general). To see this, take the simplest scenario possible−constant functions. For example, let $f(x) = -1 = h(x)$ and $g(x) = 1$ for $x \in [0, 1]$. Then

$$\begin{aligned} -1 = \min(f + g, h)(x) &> \min(f, h)(x) + \min(g, h)(x) \\ &= -1 + (-1) \\ &= -2. \end{aligned}$$

This counterexample disproves the statement. □

Remark 3.2.6 *Recall that the notation* $\min(f, g)$ *means "the minimum value of f and g for x in some given interval." In the previous problem, it is clear that*

$$\min(f, g) := \min_{x \in [0,1]} (f(x), g(x)) .$$

In the next section, we tackle inverse functions along with some very special definitions.

Exercises.

1. Which of the following relations are functions? State the domain and range for those relations which are functions.

   (a) $A = \{(a, b), (b, c), (c, d), (d, e)\}$

   (b) $B = \{(1, 1), (1, 2), (1, 3)\}$

   (c) $C = \{(1, 1), (2, 1), (3, 1)\}$

   (d) $D = \{(x, y) \in \mathbb{R} \times \mathbb{R} \,|\, x = y\}$

   (e) $E = \{(x, y) \in \mathbb{R} \times \mathbb{R} \,|\, x = \tan y\}$

   (f) $F = \{(x, y) \in \mathbb{N} \times \mathbb{N} \,|\, x = y^2\}$

2. Consider the real-valued function f given by $f(x) = 2x^2+1$. Find the following.

   (a) $f(-1)$

   (b) The inverse image of 19 under f

   (c) $f^{-1}(\pi)$

   (d) $\mathcal{R}(f)$

3. The characteristic (or indicator) function is denoted by the Greek letter χ. It is defined as

$$\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

   for some prescribed set A. For example, $\chi_{[0,1]}$ looks like

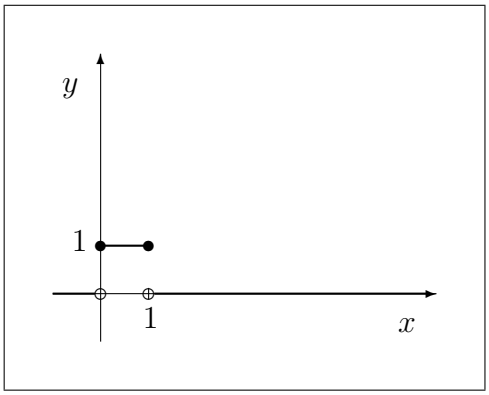

Figure 3.12: The graph of $\chi_{[0,1]}$

   If we assume $\chi_{[0,1]} : \mathbb{R} \to \mathbb{R}$. Sketch the following:

   (a) $\chi_{[-1,1]}$

   (b) $\chi_{[-1,1]'}$

(c) $\chi_{\mathbb{N}}$

(d) $\chi_{\{0\}}$

4. This exercise is meant to show the relationship between the characteristic function χ and the step (greatest integer) function $f(x) = \lfloor x \rfloor$. Consider the set

$$[0,3) = [0,1) \cup [1,2) \cup [2,3)\,.$$

Denote $A_i = [i-1, i)$ so we have $\bigcup_{i=1}^{n} A_i = [0,3)$ for $n = 3$. Now for $x \in A_i$ ($i \in \mathbb{N}$) define $g : [0,3) \to \mathbb{R}$ as $g(x) = i - 1$. Graph the function g. What happens if we let $n \to \infty$? How is the function g related to χ?

5. Here is another problem that deals with the characteristic function. Let U represent the universal set and let $A, B \subset U$. The following statements are true.

(a) $\max(\chi_A, \chi_B) = \chi_{A \cup B}$

(b) $\min(\chi_A, \chi_B) = \chi_{A \cap B}$

(c) $\chi_{A'} = 1 - \chi_A$

(d) $\chi_{A \setminus B} = \chi_A - \chi_B$, provided $B \subset A$

We prove (b). There are two cases to consider: $x \in A \cap B$ and $x \notin A \cap B$.

- Take $x \in A \cap B$. Then, for this x, $\chi_{A \cap B}(x) = 1$. However, $x \in A \cap B$ translates to $x \in A$ and $x \in B$. Thus,

$$\begin{aligned} \min(\chi_A, \chi_B) :&= \min(\chi_A(x), \chi_B(x)) \\ &= \min(1,1) \\ &= 1 \\ &= \chi_{A \cap B} \end{aligned}$$

- Suppose $x \notin A \cap B$. Then $x \in (A \cap B)'$ so $x \in A' \cup B'$ (De Morgan). Now since $x \notin A \cap B$, we know that $\chi_{A \cap B}(x) = 0$ (for this x). We also know from $x \in A'$ or $x \in B'$ that

$$\begin{aligned} \min(\chi_A, \chi_B) :&= \min(\chi_A(x), \chi_B(x)) \\ &= 0 \end{aligned}$$

since <u>at least one</u> of $\chi_A(x)$ or $\chi_B(x)$ must be zero. Hence, $\min(\chi_A, \chi_B) = 0 = \chi_{A \cap B}$. ■

Your task: prove parts (a), (c), and (d).

6. Make a conjecture about χ_U and $\chi_\emptyset$. Prove your conjectures.

7. Let $A_n = [n, n+1) \subset \mathbb{R}$, $n \in \mathbb{W}$. Consider the relation

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} n \cdot \chi_{A_n}(x) \\ &= 0 \cdot \chi_{A_0}(x) + 1 \cdot \chi_{A_1}(x) + 2 \cdot \chi_{A_2}(x) + \cdots \end{aligned}$$

   Ignore the values of χ_{A_n} where $x \notin A_n$ for all $n \in \mathbb{W}$. That is, "throw away" any parts of the graph where $\chi_{A_n}(x) = 0$ for all $n \in \mathbb{W}$. Now graph f. You should observe that f is the function $f(x) = \lfloor x \rfloor$ for $x \in [0, \infty)$.

8. Prove **Proposition 3.2.1**.

9. Prove **Theorem 3.2.1**, parts 2 and 3.

10. Prove **Theorem 3.2.2**, part 2.

11. Let $f : A \to B$ with $X, Y \subset B$. Prove that

$$f^{-1}(X) \backslash f^{-1}(Y) \subset f^{-1}(X \backslash Y).$$

12. Prove or disprove the following:

    (a) If $x \in A$, then $f(x) \in f(A)$.

    (b) If $f(x) \in f(A)$, then $x \in A$.

3.3 Construction, Properties, and Prelude to Equivalence

We begin with an example dealing with the composition of functions. (For now, we rely on your recollection of this topic.) Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = x^2$ and let $g : (0, \infty) \to \mathbb{R}$ be defined as $g(x) = \ln x$. Then you may recall that $(f \circ g)(x) = f(g(x)) = f(\ln x) = (\ln x)^2$, $x \in (0, \infty)$. That is, the composition of function g with f is the new function $(\ln x)^2$. Notice that writing $f(\ln x)$ is dangerous in some sense; that is, if the value of $\ln x$ is not in the domain of f, then $f(\ln x)$ is just nonsense. We address some of these issues in detail in this section.

Definition 3.3.1 *Let $f : A \to B$ and $g : C \to D$ be functions. Then the composition of g with f, denoted by $f \circ g$, is*

$$f \circ g = \{(c, b) \in C \times B \mid \exists\ a \in A \cap D \ni \text{ both } (c, a) \in g \text{ and } (a, b) \in f\}.$$

Remark 3.3.1 *Notice that the existence of $a \in A \cap D$ is the part of the definition which states that $\mathcal{R}(g)$ be contained in $\mathcal{D}(f)$ (however, they need not be equal). See the figure.*

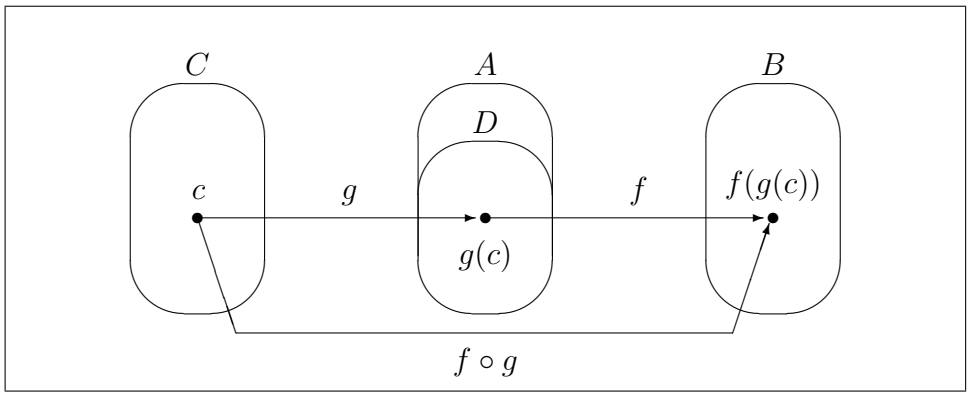

Figure 3.13: The composition of g with f

In the figure, we note that $D \subset A$. So since $g(c) \in D$, it must be that $g(c) \in A$ which indicates that writing $f(g(c))$ makes sense. Notice that if $D \not\subset A$, it would be impossible to consider $f(a)$ for $a \notin A$.

Remark 3.3.2 *In the diagram, notice that $a = g(c)$ and $b = f(a) = f(g(c))$.*

Here is a comment worth mentioning. It is not uncommon to find composition defined as

$$f \circ g = \{(c, b) \in C \times B \,|\, \text{for some } a \in A, \text{ both } (c, a) \in g \text{ and } (a, b) \in f\}$$

where $f : A \to B$ and $g : C \to A$. This definition is in no way less restrictive than the one given earlier. Some students fall into the trap of translating $g : C \to A$ as $\mathcal{R}(g) = A$. Recall that $g : C \to A$ indicates that g is a function from C into A such that $\mathcal{D}(g) = C$ and $\mathcal{R}(g) \subset A$. In other words, the reformulated definition does <u>not</u> say that $\mathcal{R}(g) = \mathcal{D}(f)$. The next example sheds some light on commutativity.

Example 3.3.1 *Let $h(x) = \cos x$ and $j(x) = x^3 - 4x$ (we know that both h and j are functions that map $\mathbb{R}$ into $\mathbb{R}$). Then*

$$(h \circ j)(x) = h(j(x)) = h(x^3 - 4x) = \cos(x^3 - 4x)$$

and

$$(j \circ h)(x) = j(h(x)) = j(\cos x) = \cos^3 x - 4\cos x.$$

Clearly, $\circ$ is not a commutative operation.

Example 3.3.2 *Let $k, l : \mathbb{N} \to \mathbb{N}$ be given by $k(x) = x + 2$ and $l(x) = 3x$. Find $\mathcal{R}(k \circ l)$ and $\mathcal{R}(l \circ k)$ and compare.*

Solution. Note that $(k \circ l)(x) = k(3x) = 3x + 2$ so

$$\mathcal{R}(k \circ l) = \{5, 8, 11, 14, \ldots\}.$$

On the other hand, $(l \circ k)(x) = l(x + 2) = 3(x + 2)$ so

$$\mathcal{R}(l \circ k) = \{9, 12, 15, 18, \ldots\}.$$

Clearly, the ranges are not equal. □

Proposition 3.3.1 *Let $f : A \to B$, $g : B \to C$, and $h : C \to D$. Then*

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

That is, composition of functions is associative.

Remark 3.3.3 *To reiterate, the domain of g is not necessarily equal to the range of f (similar statements hold for h and g). However, $\mathcal{R}(f) \subset B = \mathcal{D}(g)$.*

You may be wondering how to prove the proposition. That is, how does one show equality of functions? Remember that functions are just sets−that is, for $f : A \to B$, we define $f = \{(x, f(x)) \mid x \in A\}$. Thus, to show $f = g$ we simply need to show that $f \subset g$ and $g \subset f$.

Proof. See the diagram on the following page for some visual assistance. Since functions are sets of ordered pairs, take $(a, d) \in h \circ (g \circ f)$. This means that there must be a $c \in C$ such that $(a, c) \in g \circ f$ and $(c, d) \in h$. This is the very definition of $h \circ (g \circ f)$−that is, treat $g \circ f$ like a function. Next, we further dissect $(a, c) \in g \circ f$. This translates to there being a $b \in B$ such that $(a, b) \in f$ and $(b, c) \in g$. Now since we are heading toward $(a, d) \in (h \circ g) \circ f$, we want to settle matters between h and g first. Notice: since $(c, d) \in h$ and $(b, c) \in g$, this implies that $(b, d) \in h \circ g$. Now since $(a, b) \in f$, this, combined with $(b, d) \in h \circ g$, says that $(a, d) \in (h \circ g) \circ f$. Hence $h \circ (g \circ f) \subset (h \circ g) \circ f$. Showing that $(h \circ g) \circ f \subset h \circ (g \circ f)$ is essentially a reversal of steps; you supply the details. ■

Remark 3.3.4 *Despite the above work, some mathematicians would rather define equality of functions through the definition below.*

Definition 3.3.2 *The functions f and g are equal IFF*

1. *$\mathcal{D}(f) = \mathcal{D}(g)$ and*

2. *for any* $x \in \mathcal{D}(f) = \mathcal{D}(g)$, $f(x) = g(x)$.

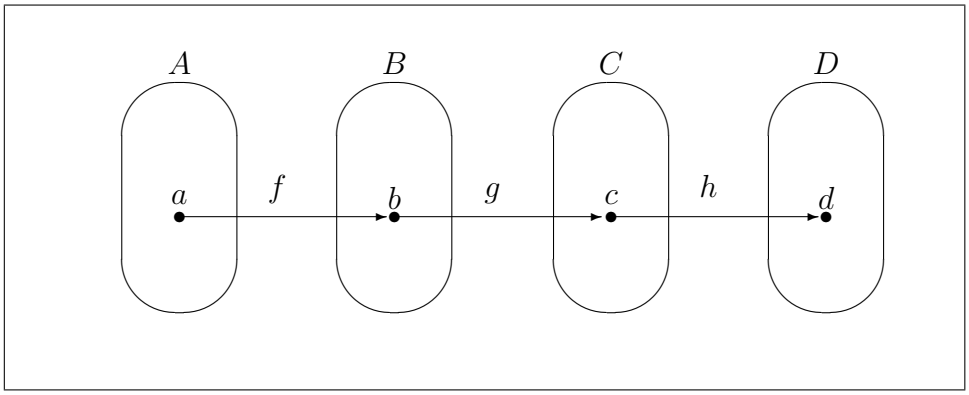

Figure 3.14: Associativity of composition

We now take a moment to discuss restrictions and extensions though we will not dwell on them. First, you should notice that when we write

$$m(x) = x^2 + \sin x, \;\; x \in [0, \infty)$$

and

$$n(x) = x^2 + \sin x, \;\; x \in [0, 1],$$

m and n are understood to be different functions (remember that equal functions have equal domains). However, one might say that n is a restriction of m or that m is an extension of n. Here is the definition.

Definition 3.3.3 *Let $f : A \to B$ and $g : C \to D$. Let $C \subset A$ with $g(x) = f(x)$ for $x \in C$. Then g is a restriction of f to C. Likewise, f is an extension of g to A. In the restriction sense, it is common to write $g = f|_C$.*

Example 3.3.3 *Let $A = \{e, \pi, \sqrt{2}, \pi^e\}$, $B = \{\alpha, \beta, \delta, \epsilon, \gamma\}$, and*

$$f = \{(e, \alpha), (\pi, \delta), (\sqrt{2}, \epsilon), (\pi^e, \delta)\}.$$

Then $f : A \to B$. Note that $f|_{\{\pi^e\}} = \{(\pi^e, \delta)\}$ and $f|_{\{e,\sqrt{2}\}} = \{(e, \alpha), (\sqrt{2}, \epsilon)\}$.

Example 3.3.4 *We are all very familiar with the function $\mathcal{I} : \mathbb{R} \to \mathbb{R}$ given by $\mathcal{I}(x) = x$. This function is often called the identity function since $\mathcal{I} = \{(x, x) \,|\, x \in \mathbb{R}\}$. Let $\mathcal{I}_1 = \mathcal{I}|_{\mathbb{Z}}$, $\mathcal{I}_2 = \mathcal{I}|_{\mathbb{R}^+}$, and $\mathcal{I}_3 = \mathcal{I}|_{[0,1]}$. See the graphs of these restrictions in the figures that follow.*

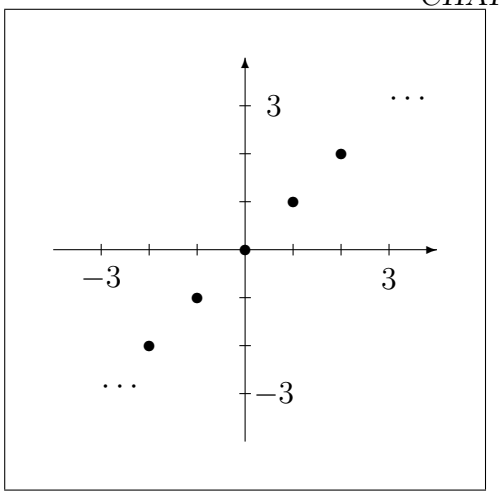

Figure 3.15: The Graph of $\mathcal{I}_1$

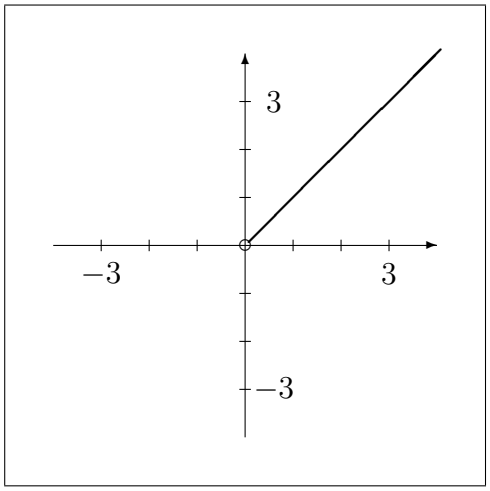

Figure 3.16: The Graph of $\mathcal{I}_2$

We now state one of the most important definitions in function theory.

Definition 3.3.4 *Let $f : A \to B$. Then f*

1. *maps A <u>onto</u> B if $\mathcal{R}(f) = B$.*

2. *is <u>one-to-one</u> if $f(a_1) = f(a_2)$ implies that $a_1 = a_2$ for $a_1, a_2 \in A$.*

3. *is <u>bijective</u> if f enjoys each of the properties above. We then say that there is a <u>one-to-one correspondence</u> between the sets A and B or that A and B are <u>equivalent</u>. We write $A \sim B$.*

Remark 3.3.5 *It is useful to know that functions that map onto B are also called surjections while one-to-one functions are sometimes called injections.*

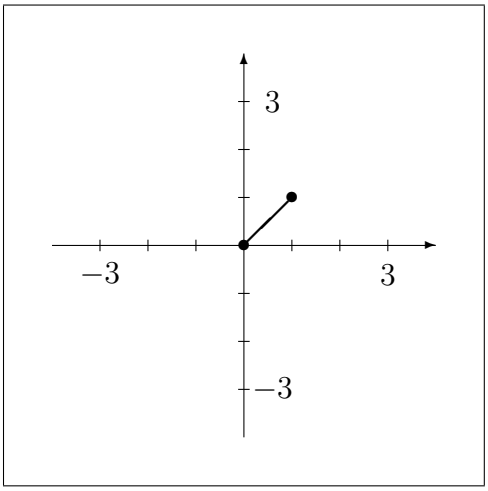

Figure 3.17: The Graph of $\mathcal{I}_3$

Remark 3.3.6 *Use of the word "onto" in the expression "A onto B" makes sense if you think about it—if $\mathcal{R}(f) = B$, then everything in B is "covered" or accounted for. In other words, to show that f is onto B, you need to consider an arbitrary $b \in B$ and show that there is an $a \in A$ for which $f(a) = b$. If f maps A onto B, we often write $f : A \Rightarrow B$. See the diagram below.*

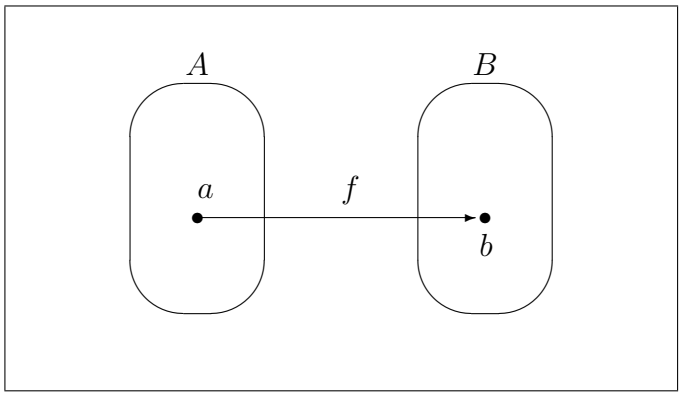

Figure 3.18: For each $b \in B$, there exists an $a \in A$ such that $f(a) = b$. Thus, f is onto B.

Notice the contrapositive of the second part of the definition above. That is, if $a_1 \neq a_2$ then $f(a_1) \neq f(a_2)$. In words, no two distinct points can map to the same place. This is another way to prove that a function is injective.

Example 3.3.5 *Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = x^2$. Notice that f is not onto $\mathbb{R}$ since $\mathcal{R}(f) = [0, \infty) \neq \mathbb{R}$. Likewise, f is not one-to-one since $f(-3) = 9 = f(3)$ but $-3 \neq 3$. See the figure below. (But take note: $f : [0, \infty) \to [0, \infty)$ with $f(x) = x^2$ is a bijection. (Verify!) It seems more than reasonable to state that $[0, \infty) \sim [0, \infty)$; see the definition of equivalent sets).*

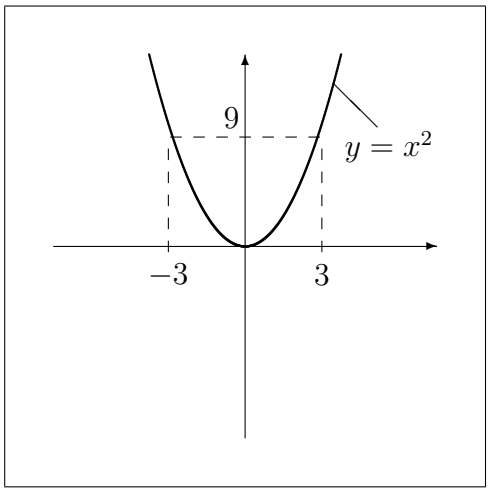

Figure 3.19: The function f is not one-to-one. A quick test for one-to-oneness is to see if the function passes a *horizontal* line test.

Example 3.3.6 *We now give an example of a function $f : [1, \infty) \Rightarrow \mathbb{N}$ that is not one-to-one. Recall that the double arrow $\Rightarrow$ means that the function f must be onto $\mathbb{N}$. A nice example to think of here is $f(x) = \lfloor x \rfloor$. To show that f is onto $\mathbb{N}$, take an $n \in \mathbb{N}$. Then there is an $x \in [1, \infty)$ such that $f(x) = n$ (for example, any x satisfying $n \leq x < n + 1$ will work). Thus f is onto $\mathbb{N}$. The above argument also shows that f is not one-to-one since we have many x's satisfying $f(x) = n$ for a fixed $n \in \mathbb{N}$. As a concrete example, note that $f(5.8) = 5 = f(5.1)$ but clearly $5.8 \neq 5.1$. A glance back at Figure 3 should convince you further.*

Example 3.3.7 *Finally, we present an example in which f is one-to-one, g is not one-to-one, yet $g \circ f$ is one-to-one. Let $f(x) = e^x$ and $g(x) = x^2$. You should show that the supposition is satisfied. Then observe that $(g \circ f)(x) = g(e^x) = e^{2x}$, which is one-to-one.*

Notice that if a function $f : A \to B$ is one-to-one then $f^{-1}(b)$ contains exactly one element (say $\{a\}$ so that $f(a) = b$). In other words, for each $b \in \mathcal{R}(f) \subset B$ there is precisely one $a \in \mathcal{D}(f)$ with the property that $f^{-1}(b) = a$. In other words, f^{-1} itself is a function. We now give the symbol f^{-1} precise meaning.

Definition 3.3.5 *Let $f : A \to B$ be one-to-one. Then $f^{-1} : \mathcal{R}(f) \to A$ is called the inverse function of f and is defined as $f^{-1}(b) = a$ provided that $f(a) = b$ for $b \in \mathcal{R}(f)$.*

Remark 3.3.7 *From this point forward, f^{-1} will indicate the inverse function of f.*

In our discussion above, since f is one-to-one, writing $f^{-1} : \mathcal{R}(f) \to A$ means that $\mathcal{R}(f^{-1}) = A$ contrary to our earlier notes. You should convince yourself of this. One way to see this is to observe that since $f : A \to B$ is one-to-one, $f : A \to \mathcal{R}(f)$ is a bijection. You can then convince yourself that $f^{-1} : \mathcal{R}(f) \to A$ is also a bijection. If still in doubt, here is another way to analyze this. We've declared that $\mathcal{R}(f^{-1}) = A$. To prove this, assume the contrary. That is, suppose that $\mathcal{R}(f^{-1}) \neq A$ (so we'll assume that $\mathcal{R}(f^{-1}) \subset A$). This would mean that there is an $a \in A$ that is "not covered" by f^{-1}. That is, there exists an $a \in A$ such that $f^{-1}(b) \neq a$ for any $b \in B$. See the diagram below.

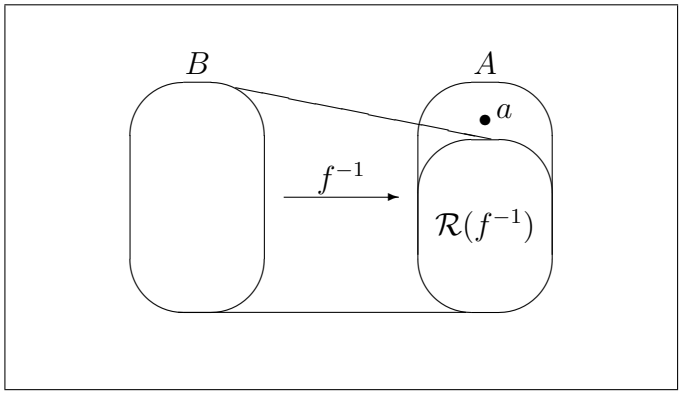

Figure 3.20: $f^{-1}(b) \neq a$ for any $b \in B$

But since $f : A \to B$, there is a $b_1 \in B$ such that $f(a) = b_1$ for this particular $a \in A$ (since $\mathcal{D}(f) = A$, $a \in A$). (See the diagram on the following page.) So with $f(a) = b_1$, use the definition of the inverse function f^{-1}. Do you see the contradiction?

Example 3.3.8 *Let $f(x) = e^x$, $x \in \mathbb{R}$. Then it follows directly that $f^{-1}(x) = \ln x$, $x > 0$. We see that $\mathcal{R}(f) = \mathcal{D}(f^{-1})$ and that $\mathcal{R}(f^{-1}) = \mathbb{R} = \mathcal{D}(f)$. Now let $a \in \mathbb{R}$; then $f(a) = e^a$ which we may denote as b so that $f(a) = b$. Then $f^{-1}(b) = \ln b = \ln(e^a) = a$, so the definition is satisfied.*

We are now in a position to state some vital theorems.

Theorem 3.3.1 *Let $f : A \to B$ be a bijection. Then $f^{-1} : B \to A$ is also a bijection.*

Proof. We leave the proof as an exercise (we practically did this in the discussion following **Remark 3.3.7**). ■

Theorem 3.3.2 *Let $f : A \to B$ and $g : C \to D$ be one-to-one. Then $g \circ f : A \to D$ is one-to-one. In words, the composition of injections is injective.*

Proof. We need to show that $g \circ f$ is one-to-one. Thus, we begin by assuming that $(g \circ f)(x_1) = (g \circ f)(x_2)$ with the intention of showing that $x_1 = x_2$ for $x_1, x_2 \in A$. So we have $g(f(x_1)) = g(f(x_2)) \Rightarrow f(x_1) = f(x_2)$ since g is an injection. Likewise, $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$ since f is an injection. Hence, we've established that $g \circ f$ is an injection. ■

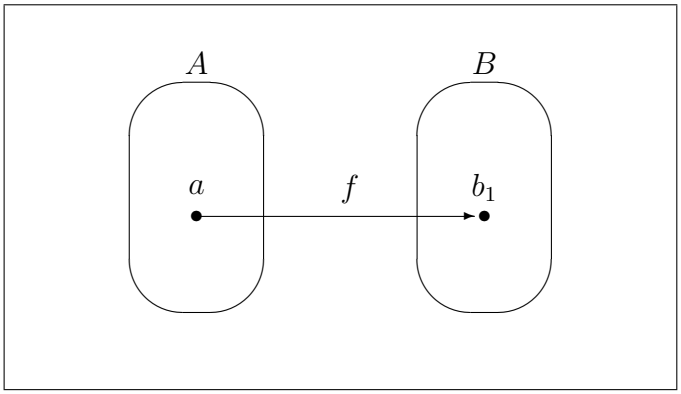

Figure 3.21: $f(a) = b_1$

Theorem 3.3.3 *Let $f : A \Rightarrow B$ and $g : B \Rightarrow C$. Then $g \circ f : A \Rightarrow C$. In words, the composition of surjections is surjective.*

Proof. By assumption, given $b \in B$, there exists an $a \in A$ such that $f(a) = b$. Also, given $c \in C$ there exists a $b \in B$ such that $g(b) = c$ (be careful to note that the b's here are most likely <u>not</u> the same). Both statements are independently valid since f and g are onto B and C, respectively. We need to show that, given $c \in C$, there is an $a \in A$ such that $(g \circ f)(a) = c$. Putting our above statements in motion, given $c \in C$, we know that there is a $b \in B$ such that $g(b) = c$, and, <u>for this $b \in B$</u>, there exists an $a \in A$ such that $f(a) = b$. In other words, given this $c \in C$, we can find an $a \in A$ with $g(f(a)) = c$. That is, $(g \circ f)(a) = c$. We are done. ■

Remark 3.3.8 *If $f : A \to B$ and $g : B \to C$ are bijections, you should now be able to make an inference on $g \circ f$.*

Proposition 3.3.2 *Let $f : A \to B$ and $g : B \to C$.*

1. *If $g \circ f : A \Rightarrow C$, then $g : B \Rightarrow C$.*
2. *If $g \circ f : A \to C$ is one-to-one, then f is one-to-one.*

Remark 3.3.9 *Note the subtle differences in the two statements (as well as the two conclusions) in the proposition!*

Proof.

1. By assumption, given $c \in C$, we can find $a \in A$ such that $g(f(a)) = c$. The question arises: how must one go about finding this $a \in A$? Since $\mathcal{D}(g) = B$, this implies that $f(a) \in B$. Now, for notational convenience, denote $f(a) = b$. We have now converted the existence of an $a \in A$ to really $b \in B$. In other words, we can rephrase the statement of this proof as "given $c \in C$, there exists $b \in B$ such that $g(b) = c$." This says that $g : B \Rightarrow C$. ■
2. By assumption we have $g(f(a_1)) = g(f(a_2)) \Rightarrow a_1 = a_2$ for $a_1, a_2 \in A$. Unfortunately, it is not exactly obvious how to proceed. Let's try the contrapositive. We have $a_1 \neq a_2 \Rightarrow g(f(a_1)) \neq g(f(a_2))$ for $a_1, a_2 \in A$. We will try to prove that f is an injection. That is, $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$ for $a_1, a_2 \in A$. We prove this by contradiction. That is, suppose there exists $a_3 \neq a_4$ yet $f(a_3) = f(a_4)$ (i.e., we are assuming that f is not one-to-one). Now since $f(a_3)$ and $f(a_4)$ are identical, we can apply g to each and obtain equality. That is, $g(f(a_3)) = g(f(a_4))$ for $a_3 \neq a_4$. Rephrasing, given $a_3 \neq a_4$, we have $(g \circ f)(a_3) = (g \circ f)(a_4)$. This says that $g \circ f$ is <u>not</u> one-to-one. This contradiction proves that f is an injection. ■

We are approaching a very important result. First, some notation.

$\mathcal{I}$ is the usual identity function; that is, $\mathcal{I}(x) = x$. Now, by its very nature, if $x \in A$, then $\mathcal{I}(x) \in A$ so $\mathcal{I} : A \to A$. To avoid having to write all of this, we simply write $\mathcal{I}_A$. In short, $\mathcal{I}_A$ is the identity function $\mathcal{I}_A(x) = x$ with $\mathcal{D}(\mathcal{I}_A) = \mathcal{R}(\mathcal{I}_A) = A$.

Proposition 3.3.3 *Let $f : A \to B$ and $g : B \to A$. Then $g = f^{-1}$ IFF $g \circ f = \mathcal{I}_A$ and $f \circ g = \mathcal{I}_B$.*

Proof. We have to prove two things:

1. If $g = f^{-1}$, then $g \circ f = \mathcal{I}_A$ and $f \circ g = \mathcal{I}_B$.
2. If $g \circ f = \mathcal{I}_A$ and $f \circ g = \mathcal{I}_B$, then $g = f^{-1}$.

1. Let $g = f^{-1}$. Then we have $f : A \to B$ and $f^{-1} : B \to A$. We know that $\mathcal{D}(f^{-1} \circ f) = \mathcal{D}(f)$ (in words, the domain of a composite function is equal to the domain of the first function applied). That is, $\mathcal{D}(f^{-1} \circ f) = A$. Now, from the definition of f^{-1}, if $x \in A$ and $(x, f(x)) \in f$, then $(f(x), x) \in f^{-1}$. Hence, $(f^{-1} \circ f)(x) = f^{-1}(f(x)) = x$. So since $(f^{-1} \circ f)(x) = x$ and $\mathcal{D}(f^{-1} \circ f) = A$, we can say that $f^{-1} \circ f = \mathcal{I}_A$. On the other hand, $\mathcal{D}(f \circ f^{-1}) = \mathcal{D}(f^{-1}) = B$ since $f^{-1} : B \to A$. Now for $y \in B$, $(y, f^{-1}(y)) \in f^{-1}$ and $(f^{-1}(y), y) \in f$. Thus, $(f \circ f^{-1})(y) = f(f^{-1}(y)) = y$. Therefore, $f \circ f^{-1} = \mathcal{I}_B$ and the first part is complete. ■

2. Assume that $g \circ f = \mathcal{I}_A$ and $f \circ g = \mathcal{I}_B$. In other words, $g(f(x)) = x$ for $x \in A$ and $f(g(y)) = y$ for $y \in B$. Now let's examine this closely. What does $g(f(x)) = x$ for $x \in A$ really mean? Clearly, $g \circ f$ is a one-to-one function (i.e., $g \circ f : A \to A$ is one-to-one). As a result, an application of **Proposition 3.3.2** implies that f is injective, in turn, guaranteeing the existence of $f^{-1} : \mathcal{R}(f) \to A$. How about $f \circ g = \mathcal{I}_B$? In detail, we have $f(g(y)) = y$ for $y \in B$ so we see that $f \circ g$ is onto B. An application of **Proposition 3.3.2** implies that f is onto B as well. Notice what we now have: f is one-to-one and onto B. In other words, f^{-1} exists and $f^{-1} : B \to A$. Now a very important fact: $f^{-1} = f^{-1} \circ \mathcal{I}_B$; that is, the identity function has the effect of leaving functions "unchanged." Notice that the B on $\mathcal{I}_B$ is crucial here because $f^{-1} : B \to A$. Then we can write

$$f^{-1} = f^{-1} \circ \mathcal{I}_B = f^{-1} \circ (f \circ g) = (f^{-1} \circ f) \circ g = \mathcal{I}_A \circ g,$$

since $\mathcal{D}(f^{-1} \circ f) = A$. Notice that we also used the associativity of $\circ$. Finally, note that $\mathcal{I}_A \circ g = g$ since again, the identity function leaves g unchanged. (We mention again that the A on $\mathcal{I}_A$ is very important since $g : B \to A$. Hence, $f^{-1} = g$ as sought. ■

Remark 3.3.10 *Notice that there is a step in the proof above where we state that $f^{-1} \circ f = \mathcal{I}_A$. Do you see why this is valid? If not, you should try proving the following corollary before continuing.*

Corollary 3.3.1 *Let $f : A \to B$. Then*

1. *$f^{-1} \circ f = \mathcal{I}_A$ and*
2. *$f \circ f^{-1} = \mathcal{I}_{\mathcal{R}(f)}$.*

We now redirect your attention to Example 3.3.8. There, $f(x) = e^x$ for $x \in \mathbb{R}$ and $g(x) = f^{-1}(x) = \ln x$ for $x > 0$. In terms of the proposition, $f : \mathbb{R} \to \mathbb{R}^+$ and $f^{-1} : \mathbb{R}^+ \to \mathbb{R}$ (though it is perfectly fine to say that $f : \mathbb{R} \to \mathbb{R}$). We have

$(g \circ f)(x) = f^{-1}(f(x)) = \ln(e^x) = x$ for $x \in \mathbb{R}$ so that $g \circ f = \mathcal{I}_{\mathbb{R}}$. Going the other way, $(f \circ g)(x) = f(f^{-1}(x)) = e^{\ln x} = x$ for $x > 0$ so $f \circ g = \mathcal{I}_{\mathbb{R}^+}$. From this come the two familiar identities $\ln(e^x) = x$ and $e^{\ln x} = x$. Match this with the proposition!

Exercises.

1. For the functions below, find $f \circ g$ and $g \circ f$. Use the understood domains for both f and g. Additionally, state the domain and range of each composite function. Is there anything special about parts (a) and (c)?

   (a) $f(x) = e^x$ and $g(x) = \ln x$

   (b) $f(x) = \sin x$ and $g(x) = 5x^2 + \cos x$

   (c) $f(x) = 5x + 3$ and $g(x) = \frac{1}{5}x - \frac{3}{5}$

   (d) $f(x) = \frac{x+3}{x-6}$ and $g(x) = 2|x|$

2. Consider $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(x) = (\cos x, e^x)$.

   (a) State the range of f.

   (b) Find $f^{-1}((1, 1))$.

   (c) Let $g : \mathbb{R}^2 \to \mathbb{R}$ be given by $g((a, b)) = a^b$. Find $(g \circ f)(5)$.

3. Let $f : A \to B$ and let f^{-1} denote the inverse of f. Prove that

   (a) $f^{-1}(f(a)) = a$ for $a \in A$.

   (b) $f(f^{-1}(b)) = b$ for $b \in \mathcal{R}(f)$.

4. Let $f : A \to B$ and $g : B \to C$. Find f and g such that

   (a) g is surjective but $g \circ f$ is <u>not</u> surjective.

   (b) f is injective but $g \circ f$ is <u>not</u> injective.

   (c) $g \circ f$ is injective but g is <u>not</u> injective.

5. Show that $f : [0, 1] \to [a, b]$ given by $f(x) = (b - a)x + a$ is a bijection. What can you say about the sets $[0, 1]$ and $[a, b]$? Does this seem at all strange?

6. Find $f : \mathbb{N} \to \mathbb{N}$ such that

   (a) f is neither one-to-one nor onto $\mathbb{N}$.

   (b) f is onto $\mathbb{N}$ but not one-to-one.

   (c) f is one-to-one but not onto $\mathbb{N}$.

7. Prove **Theorem 3.3.1**.

8. Prove **Corollary 3.3.1**.

9. Prove the following generalizations of **Corollary 3.3.1**. Let $f : A \to B$ be bijective. Show that

   (a) $f^{-1} : B \to A$ is bijective.

   (b) $f^{-1} \circ f = \mathcal{I}_A$.

   (c) $f \circ f^{-1} = \mathcal{I}_B$.

10. This problem generalizes **Theorem 3.2.2** from the previous section. Let $f : A \to B$ with $X \subset A$ and $Y \subset B$. Prove that

    (a) $f(f^{-1}(Y)) = Y$ if f is surjective.

    (b) $f^{-1}(f(X)) = X$ if f is injective.

11. State whether g is a restriction of f.

    (a) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \lfloor x \rfloor$; $g : [0, 1] \to \mathbb{R}$, $g(x) = 0$.

    (b) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$; $g : \mathbb{R}^+ \to \mathbb{R}$, $g(x) = x$.

3.4 Equivalence and Countability

Recall the definition of equivalence from the last section. Two sets A and B are equivalent (and we write $A \sim B$) if there exists a bijection $f : A \to B$. The definition is sensible enough when we have $A = B$ since it seems entirely natural for a set to be equivalent to itself. However, most people would adamantly reject a statement such as $[0, 1] \sim [0, 3]$. How could these sets be equivalent? After all, the interval $[0, 3]$ is triple the length of $[0, 1]$−hardly "equivalence" here! The fact is, $[0, 1]$ is equivalent to $[0, 3]$, or, informally, $[0, 1]$ and $[0, 3]$ are really the same "size." Not convinced? Let $f : [0, 1] \to [0, 3]$ be defined as $f(x) = 3x$; clearly f is a bijection so $[0, 1] \sim [0, 3]$.

The above example is enough to make almost anyone uneasy. What if we were to state that $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \sim \mathbb{R}$? Guess what? It's true! Let $g : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ be defined as $g(x) = \tan x$ so g is bijective. The above (counterintuitive) results need more than just strong words and instincts as a means for assertion. We warned you early on about infinity! Although this section is dedicated to the study of infinite sets, common sense tells us to address *finite* sets prior to infinite ones. So we begin.

Definition 3.4.1 *Let A be a set. Then A is finite if $A \sim \{1, 2, 3, \ldots, n\}$ for some $n \in \mathbb{N}$.*

Definition 3.4.2 *If A is not finite, we say that A is infinite.*

Example 3.4.1 *Here is an example of each type.*

1. *$A = \{a, b, c, d, e\}$ is a finite set; $n = 5$ here.*
2. *$B = \{x \mid x \text{ is prime}\} = \{2, 3, 5, 7, 11, 13, \ldots\}$ is an infinite set. We cannot say that B contains n members for any $n \in \mathbb{N}$.*

Remark 3.4.1 *Sometimes an alternate definition of infinite sets is more convenient. That is, A is infinite if, for any $n \in \mathbb{N}$, A contains a subset with n elements. Convince yourself that this makes sense.*

What follows is an intuitive result. However, the proof is insightful.

Proposition 3.4.1 *If A and B are finite sets and $A \sim B$, then sets A and B have the same number of elements.*

Proof. Let $f : A \to B$ be a bijection so that $A \sim B$. We "list" the elements in A and B:

$$A = \{a_1, a_2, a_3, \ldots, a_n\}$$

and

$$B = \{b_1, b_2, b_3, \ldots, b_m\}.$$

Furthermore, we assume that the elements in each set are distinct. That is, $a_i \neq a_j$ for $i \neq j$ and $b_i \neq b_j$ for $i \neq j$. Our goal is to show that $m = n$. We proceed by contradiction. First, suppose that $n > m$. Then there is a $b_i \in B$ ($i \in \{1, 2, \ldots, m\}$) such that $f(a_k) = b_i = f(a_j)$ where $a_k \neq a_j$. This is a contradiction to f being one-to-one. Hence it must be that $n \leq m$. Therefore, we suppose that $n < m$ (so B has more elements). Then, since f is onto B, there must be an $a_l \in A$ ($l \in \{1, 2, \ldots, n\}$) such that $f(a_l) = b_j$ and $f(a_l) = b_k$. This contradicts the fact that f was assumed to be a function. Therefore, it must be that $m = n$. ■

Remark 3.4.2 *In the preceding proof, do you see why it is fine to assume that $a_i \neq a_j$, $i \neq j$? It is redundant to repeat terms in a set since sets speak of "belongingness." For example, $A = \{1, 2, 3, 1, 5\}$ is precisely the same as $A = \{1, 2, 3, 5\}$.*

We now move on to infinite sets; matters get quite sticky at this stage. We will see that there are different "sizes" of infinite sets, the smallest being a countable set.

Definition 3.4.3 *Let A be an infinite set. A is said to be countable if $A \sim \mathbb{N} = \{1, 2, 3, 4, \ldots\}$.*

Remark 3.4.3 *The word "countable" makes sense. We know that*

$$\{1, 2, 3, \ldots\} = \mathbb{N}$$

is an infinite set but there seems to be some sort of coherent order to this type of infinity. That is, we could begin counting $1, 2, 3, \ldots$ *and if we ever became fatigued, we could kindly ask someone to take over for a while. In other words, if we had "forever" to do so, perhaps we could "count"* $\mathbb{N}$*. Try doing this for* $\mathbb{R}$ *and we may run into some difficulties.*

A direct (and logical) consequence of the definition is that if A is infinite and not countable, then we say that A is uncountable. One such example is $A = \mathbb{R}$ (we will prove this later). Here is an important proposition.

Proposition 3.4.2 $\mathbb{Z} \sim \mathbb{N}$. *In words, the set of integers is countable.*

Remark 3.4.4 *This may seem surprising at first since* $\mathbb{Z}$ *seems to be much "bigger" than* $\mathbb{N}$ *(actually, about twice the size!). This proposition really says that a set can be equivalent to a proper subset of itself!*

Proof. The goal here is to find a function $f : \mathbb{N} \to \mathbb{Z}$ that is bijective. Thus, everything in $\mathbb{Z}$ must look like $f(x),\ x \in \mathbb{N}$. The best way to approach this is to begin constructing a list. We will insert $\mathbb{N}$ into f and eventually output all of $\mathbb{Z}$:

$$\begin{aligned} f(1) &= 0 \\ f(2) &= 1 \\ f(3) &= -1 \\ f(4) &= 2 \\ f(5) &= -2 \\ f(6) &= 3 \\ \vdots &= \vdots \end{aligned}$$

It turns out there is a general rule here. Observe that f is piecewise and given by

$$f(x) = \begin{cases} 1 - \frac{x+1}{2}, & x = 1, 3, 5, 7, \ldots \\ \frac{x}{2}, & x = 2, 4, 6, 8, \ldots \end{cases}$$

We now show that f is one-to-one and onto $\mathbb{Z}$. We will assume $f(x) = f(y)$ and prove that $x = y$. First, if x and y are even, then $f(x) = f(y) \Rightarrow \frac{x}{2} = \frac{y}{2} \Rightarrow x = y$. Second, if x and y are odd, then, similarly, $1 - \frac{x+1}{2} = 1 - \frac{y+1}{2} \Rightarrow x = y$. Finally, if x and y are of different parity (so $x \neq y$), then $\text{sign}(f(x)) \neq \text{sign}(f(y))$. Hence, $f(x) \neq f(y)$. All three of these cases show that f is a one-to-one function. To show that f is onto $\mathbb{Z}$, consider $z \in \mathbb{Z}$.

1. Case 1: Let $z > 0$. Then $2z > 0$ and $2z$ is even. Thus, for $2z \in \mathbb{N}$, $f(2z) = \frac{2z}{2} = z$.

2. Case 2: Let $z \leq 0$. Then $2z \leq 0$ and $1 - 2z$ is odd. Thus, for $1 - 2z \in \mathbb{N}$, $f(1 - 2z) = 1 - \frac{(1-2z)+1}{2} = z$.

Thus, for arbitrary $z \in \mathbb{Z}$, we can find an $n \in \mathbb{N}$ such that $f(n) = z$. This shows that f is onto $\mathbb{Z}$. Hence $\mathbb{N} \sim \mathbb{Z}$ so $\mathbb{Z}$ is countable. See the diagram below for visual assistance.

$\mathbb{N}$	1	2	3	4	5	6	$\cdots$	$\cdots$	n	$\cdots$	$\cdots$
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$			$\swarrow$ $\searrow$		
$\mathbb{Z}$	0	1	-1	2	-2	3	$\cdots$		$\frac{n}{2}$ or $1 - \frac{n+1}{2}$		$\cdots$

This completes the proof. ■

After this proof, you may be wondering about the set $\mathbb{Q}$. Is it countable? We will find out shortly. First, a very powerful theorem.

Theorem 3.4.1 *Let the sets $A_1, A_2, A_3, \ldots, A_n, \ldots$ be countable. Then $\bigcup_{n=1}^{\infty} A_n$ is countable.*

Proof. Denote $A_1 = \{a_{11}, a_{12}, a_{13}, \ldots\}$, $A_2 = \{a_{21}, a_{22}, a_{23}, \ldots\}$, $\ldots$, so, in general, $A_n = \{a_{n1}, a_{n2}, a_{n3}, \ldots\}, \ldots$ In other words, $a_{ij} = j^{\text{th}}$ element in the i^{th} set. Now consider the sum of the subscripts on each element, denoted by the symbol $\sum$ so $\sum a_{11} = 2, \sum a_{21} = 3, \ldots$, and in general, $\sum a_{ij} = i + j$. Next, notice that we can group the elements in $\bigcup_{n=1}^{\infty} A_n$ by their sum:

Table 3.1: Elements in Fixed Sums

$\sum$	Elements
2	a_{11}
3	$a_{12},\ a_{21}$
4	$a_{13},\ a_{22},\ a_{31}$
5	$a_{14},\ a_{23},\ a_{32},\ a_{41}$
$\vdots$	$\vdots$

Notice that for a sum totaling m, there corresponds $m - 1$ elements from the sets $A_i,\ i \in \mathbb{N}$. Certainly, we remove any a_{ij}'s if they appear more than once (i.e., we skip those that have already been counted). So the bottom line is that we can "list"

the elements in $\bigcup_{n=1}^{\infty} A_n$ via their sum:

$$\bigcup_{n=1}^{\infty} A_n = \{\underbrace{a_{11}}_{\sum=2}, \underbrace{a_{12}, a_{21}}_{\sum=3}, \underbrace{a_{13}, a_{22}, a_{31}}_{\sum=4}, \ldots\}$$

Much like $\mathbb{N}$ this scheme will eventually "count" every element in $\bigcup_{n=1}^{\infty} A_n$. This completes the proof. See the diagram below for visual clues. ■

$\sum$	2	3	4	$\cdots$	n	$\cdots$
	$\downarrow$	$\downarrow$	$\downarrow$		$\downarrow$	
elements	$\{a_{11}\}$	$\{a_{12}, a_{21}\}$	$\{a_{13}, a_{22}, a_{31}\}$	$\cdots$	set with (at most) $n-1$ elements	$\cdots$

We now get to answer our earlier question.

Theorem 3.4.2 *$\mathbb{Q}$ is countable.*

Proof. Let $B_1 = \left\{\frac{0}{1}, \frac{1}{1}, \frac{-1}{1}, \frac{2}{1}, \frac{-2}{1}, \ldots\right\}$. That is, let B_1 be the set of integers divided by 1. Likewise, let

$$\begin{aligned} B_2 &= \left\{\frac{0}{2}, \frac{1}{2}, \frac{-1}{2}, \frac{2}{2}, \frac{-2}{2}, \ldots\right\} \\ \vdots \quad &= \qquad\qquad \vdots \\ B_n &= \left\{\frac{0}{n}, \frac{1}{n}, \frac{-1}{n}, \frac{2}{n}, \frac{-2}{n}, \ldots\right\} \\ \vdots \quad &= \qquad\qquad \vdots \end{aligned}$$

First observe that each B_i ($i \in \mathbb{N}$) is countable since each has the same number of elements as $\mathbb{Z}$ (and $\mathbb{Z}$ is countable). Now apply **Theorem 3.4.1** so that $\bigcup_{n=1}^{\infty} B_n$ is countable. Then observe that $\bigcup_{n=1}^{\infty} B_n = \mathbb{Q}$. ■

We now give an example.

Problem 3.4.1 *Let $B \subset A$ with B countable and A uncountable. Prove that $A \setminus B$ is uncountable.*

Solution. Suppose that $A \setminus B$ is countable. Then the elements in $A \setminus B$ may be listed as such:

$$A \setminus B = \{x_1, x_2, x_3, \ldots\}.$$

However, since $B \subset A$, we can write $A = B' \cup B = (A \setminus B) \cup B$ where both $A \setminus B$ and B are countable. In other words, A can be written as the union of two countable sets. Thus, by **Theorem 3.4.1**, A is countable. This is a contradiction to A being uncountable. Conclusion: $A \setminus B$ is uncountable. □

The next theorem is of tremendous importance.

Theorem 3.4.3 *Let A be a countable set with $B \subset A$. If B is an infinite set, then it is countable.*

Remark 3.4.5 *Notice that if B is finite, e.g. $B = \{b_1, b_2, \ldots, b_n\}$, then the result is not very interesting.*

Proof. A is countable so we write $A = \{a_1, a_2, \ldots, a_n, \ldots\}$. Notice that each member of B is an a_i for some $i \in \mathbb{N}$. Now we simply move from left to right in A, "picking" those that belong to B. For example, if a_4 is the first, we can rename it $\hat{a_1}$. If a_{12} is the next, we write $a_{12} = \hat{a_2}$. We continue this indefinitely so that $B = \{\hat{a_1}, \hat{a_2}, \hat{a_3}, \ldots, \hat{a_n}, \ldots\}$. Thus, B is countable. ■

Remark 3.4.6 *Note the contrapositive of the theorem. Assuming $B \subset A$, if B is uncountable, then A is uncountable as well.*

Example 3.4.2 *The following are consequences of the last theorem.*

1. *The positive rationals are countable.*
2. *The rationals in $[-15, 13)$ are countable.*
3. *The collection of all numbers of the form $\frac{1}{3^n}$ $(n \in \mathbb{Z})$ is countable.*

Before moving on to uncountable sets, we give a final theorem. You should think about how you would attempt a proof before looking at the outline given here. For this theorem, we only supply a sketch of the proof−you are asked to fill in the details in the exercises.

Theorem 3.4.4 *Let A and B be countable. Then the Cartesian product $A \times B$ is countable.*

(**Sketch of**) **Proof**. Let $A = \{a_1, a_2, a_3, \ldots\}$ and $B = \{b_1, b_2, b_3, \ldots\}$ and recall that $A \times B = \{(a, b) \,|\, a \in A,\ b \in B\}$. Let $E_1 = \{(a_1, b_1), (a_1, b_2), (a_1, b_3), \ldots\}$. That is, let $E_1 = \{(a_1, b_i) \,|\, i \in \mathbb{N}\}$. Then E_1 is countable (why?). In general, let $E_j = \{(a_j, b_i) \,|\, i \in \mathbb{N}\}$, $j \in \mathbb{N}$. Then $A \times B = \bigcup_{j=1}^{\infty} E_j$ is countable (apply **Theorem 3.4.1**). ■

We now find ourselves in a position to discuss uncountable sets. You may have noticed that all of our countable sets have been discrete in nature−that is, $\mathbb{N} =$

$\{1,2,3,\ldots\}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$, $\mathbb{Q} = \left\{ \frac{m}{n} \,|\, m, n \in \mathbb{Z},\ n \neq 0 \right\}$, etc. Thus, it may come as no surprise that *intervals* are uncountable. Below we have a very important theorem with a famous and flashy proof.

Theorem 3.4.5 *The interval* $[0,1]$ *is uncountable.*

Proof. (**Cantor's Diagonalization Argument**) We prove this by contradiction so suppose $[0,1]$ is countable. Then we can "list" the elements in $[0,1]$. That is, there is an $f : \mathbb{N} \to [0,1]$ that is a bijection. We now expand each number in $[0,1]$ as a decimal. We have the following scenario:

$$\begin{aligned} f(1) &= 0.a_{11}a_{12}a_{13}a_{14}\cdots \\ f(2) &= 0.a_{21}a_{22}a_{23}a_{24}\cdots \\ f(3) &= 0.a_{31}a_{32}a_{33}a_{34}\cdots \\ \vdots &= \quad \vdots \\ f(n) &= 0.a_{n1}a_{n2}a_{n3}a_{n4}\cdots \\ \vdots &= \quad \vdots \end{aligned}$$

where the a_{ij}'s are just whole numbers. Now consider the number $x \in (0,1)$ with decimal expansion $x = 0.b_1b_2b_3b_4\cdots$. Let this number have the special property that

$$b_i = \begin{cases} 7, & a_{ii} \neq 7 \\ 2, & a_{ii} = 7. \end{cases}$$

This, in effect, makes the number $x = 0.b_1b_2b_3b_4\cdots$ different from every number on the list. Specifically, this number x disagrees with $f(1)$ in the a_{11} digit, disagrees with $f(2)$ in the a_{22} digit, ..., disagrees with $f(n)$ in the a_{nn} digit, etc. In other words, we have found an $x \in (0,1)$ such that there is no $\tilde{n} \in \mathbb{N}$ with $f(\tilde{n}) = x$. This contradicts the fact that every number in $(0,1)$ appears on our list. Hence, $(0,1)$ (and $[0,1]$) is uncountable. ■

Corollary 3.4.1 *Any interval* $[a,b]$ *is uncountable.*

(**Sketch of**) **Proof**. To see this, consider the function $f : [0,1] \to [a,b]$ given by $f(x) = (b-a)x + a$ and show that it is a bijection. Then $[0,1] \sim [a,b]$ and since $[0,1]$ is uncountable, $[a,b]$ must be as well. ■

Remark 3.4.7 *You may wish to convince yourself that the same is true of* $[a,b)$, $(a,b]$, *and* (a,b).

Another immediate consequence:

Corollary 3.4.2 $\mathbb{R}$ *is uncountable.*

Proof. Assume the contrary (i.e., let $\mathbb{R}$ be countable). Notice that $[0,1] \subset \mathbb{R}$. Thus, $[0,1]$ must be countable since a subset of a countable set is countable. This contradiction proves the corollary. ■

Remark 3.4.8 *See problem 3 in the exercise set for another proof.*

Corollary 3.4.3 $\mathbb{Q}'$ *is uncountable.*

Proof. Exercise.

Here is a problem related to uncountable sets.

Problem 3.4.2 *Let $f : A \to B$ with $\mathcal{R}(f)$ uncountable. Prove that $\mathcal{D}(f)$ is also uncountable.*

Solution. By now, you are probably noticing a "contradiction" pattern emerging (this is why it made the top five list!). Assume $\mathcal{D}(f)$ is countable so that we may write

$$\mathcal{D}(f) = \{a_1, a_2, a_3, \ldots\}.$$

Now, by definition,

$$\begin{aligned}\mathcal{R}(f) &= \{b \in B \,|\, b = f(a) \text{ for some } a \in A\} \\ &= \{f(a_1), f(a_2), f(a_3), \ldots\}.\end{aligned}$$

Thus, $\mathcal{R}(f)$ is countable. This contradiction proves that $\mathcal{D}(f)$ is uncountable. □

At this point, you are advised to review the material on the Cantor set K. Recall that the Cantor set seems "small" since $K \subset [0,1]$ and $\text{length}(K') = 1$. This fact might further suggest to you that K is countable. This is not true as we demonstrate below.

Theorem 3.4.6 *K is uncountable.*

Proof. We will show this by proving that $K \sim [0,1]$, keeping in mind that $[0,1]$ is uncountable. So we need to find $f : K \to [0,1]$ that is both one-to-one and onto $[0,1]$. Recall that *any* $x \in K$ has a ternary expansion $x \overset{3}{=} 0.b_1b_2b_3b_4\cdots,\ b_i = 0 \text{ or } 2,\ i \in \mathbb{N}$ (see the exercise set from Section 2.3). We now define f as

$$f(x) \overset{2}{=} 0.a_1a_2a_3a_4\cdots,\ a_i = \frac{b_i}{2},\ i \in \mathbb{N}$$

so that $a_i = 0 \text{ or } 1,\ i \in \mathbb{N}$. Thus, f gives us a binary expansion of a number in $[0,1]$. That is, $f(x) = \sum_{n=1}^{\infty} \frac{a_n}{2^n} = \sum_{n=1}^{\infty} \frac{b_n}{2^{n+1}}$. We now prove that f is one-to-one and onto $[0,1]$.

1. One-to-oneness: Consider $x_1, x_2 \in K,\ x_1 \neq x_2$. Then this means a particular b_i ($i \in \mathbb{N}$) in the representations of x_1 and x_2 must disagree. Specifically, for this $i \in \mathbb{N}$, $b_i = 0$ in one of x_1 or x_2 and $b_i = 2$ in the other. Thus, $f(x_1) \neq f(x_2)$ since $a_i = 0$ in one of $f(x_1)$ or $f(x_2)$ and $a_i = 1$ in the other. Hence, f is one-to-one.

2. Onto $[0,1]$: Consider $y \overset{2}{=} 0.a_1a_2a_3a_4\cdots \in [0,1],\ a_i = 0$ or $1,\ i \in \mathbb{N}$. Now ask: is there an $x \in K$ such that $f(x) = y$? Well, consider $x \overset{3}{=} 0.(2a_1)(2a_2)(2a_3)(2a_4)\cdots$. We know that $x \in K$ since each "digit" is either 0 or 2 (since $a_i = 0$ or 1). Then

$$\begin{aligned} f(x) &\overset{2}{=} 0.\left(\frac{2a_1}{2}\right)\left(\frac{2a_2}{2}\right)\left(\frac{2a_3}{2}\right)\cdots \quad \text{(by definition)} \\ &= 0.a_1a_2a_3a_4\cdots \\ &= y. \end{aligned}$$

   Hence, f is onto $[0,1]$.

Thus $[0,1] \sim K$. Since $[0,1]$ is uncountable, so is K. ■

Remark 3.4.9 *The Cantor set is one of the most intriguing sets in mathematics. Although* $\text{length}(K) = 0$, *K is uncountable! The statements appear to contradict one another;* $\text{length}(K) = 0$ *suggests that the Cantor set contains little while its uncountability hints at its similarity to an interval. See the discussion from* **Remark 2.3.5**.

Exercises.

1. Prove that if A is countable, then $A \cup \{x\}$ is countable.

2. Fill in the details of the proof that if A and B are countable, then $A \times B$ is countable.

3. Here is a quick proof that $\mathbb{R}$ is uncountable. Let $f : (0,1) \to \mathbb{R}$ be defined as $f(x) = \tan\left[\pi\left(x - \frac{1}{2}\right)\right]$. Show that f is a bijection.

4. We know that $\mathbb{R}$ is 'bigger' than $\mathbb{N}$. A natural question that arises is: "Are there sets even "bigger" than $\mathbb{R}$?" This exercise answers this question. Recall that $\mathcal{P}(A) =$ the set of all subsets of A (thus, $A \subset \mathcal{P}(A)$). Show that $\mathbb{R} \not\sim \mathcal{P}(\mathbb{R})$ by supposing there is a bijective function $f : \mathbb{R} \to \mathcal{P}(\mathbb{R})$.

5. Let A be a countable set. Let P be the set of all *finite* subsets of A. Prove that P is countable.

6. Show that $(0,1) \sim [0,1)$ (see exercise 1).

7. Prove that $\mathbb{Q} \times \mathbb{Q}$ is countable (see exercise 2).

8. Let P_n be the set of all polynomials of degree n ($n \in \mathbb{N}$ is fixed) with integer coefficients. That is,
$$\sum_{i=0}^{n} a_i x^i \in P_n,\ a_i \in \mathbb{Z}.$$
Prove that P_n is countable.

9. Let A be an infinite set. Show that there exists a set $B \subset A$ such that B is countable. *Note*: A may not be countable for if it were, the proof would be trivial with $B = A$.

10. Show that $\mathbb{Q}'$ is uncountable.

11. Supply the details to the proof of **Corollary 3.4.1**.

3.5 Summary: Odds and Ends

- Cartesian Product: $A \times B = \{(a, b) \,|\, a \in A \text{ and } b \in B\}$
- A relation from A to B is a subset of $A \times B$.
- A function is a special type of relation. That is, a function $f : A \to B$, is a subset of $A \times B$ with the property that for each $x \in A$ there is exactly one $y \in B$.
  1. The set A is called the domain of f, denoted by $\mathcal{D}(f)$.
  2. The range of f, denoted by $\mathcal{R}(f)$, is defined by
  $$\mathcal{R}(f) = \{b \in B \,|\, b = f(a) \text{ for some } a \in A\}.$$
  3. For $C \subset A$,
  $$f(C) = \{f(x) \,|\, x \in C\}.$$
  $f(C)$ is called the image of C under f.
  4. For $D \subset B$,
  $$f^{-1}(D) = \{x \in A \,|\, f(x) \in D\}.$$
  $f^{-1}(D)$ is called the inverse image of D under f.
  5. The identity function $\mathcal{I}(x) = x$ plays a central role in function theory.
- For $f : A \to B$ and $g : C \to D$,
$$f \circ g = \{(c, b) \in C \times B \mid \exists\ a \in A \cap D \ni \text{ both } (c, a) \in g \text{ and } (a, b) \in f\}.$$

- Composition of functions is associative but not commutative.
- Equal functions have (a) equal domains and (b) equal rules (i.e., $f(x) = g(x)$). If the first assumption is relaxed, then we enter the world of restrictions and extensions.
- Let $f : A \to B$. Then f
  1. maps A onto B if $\mathcal{R}(f) = B$.
  2. is one-to-one if $f(a_1) = f(a_2)$ implies that $a_1 = a_2$ for $a_1, a_2 \in A$.
  3. is bijective if f has each of the properties above. We then say that there is a one-to-one correspondence between the sets A and B or that A and B are equivalent. We write $A \sim B$.
- Composition of injections (surjections) is injective (surjective).
- About f^{-1}:
  1. If f maps a to b then f^{-1} maps b to a. For f^{-1} to be considered a function in the sense of the definition, f must be one-to-one.
  2. If f is a bijection then so is f^{-1}.
  3. Let $f : A \to B$ and $g : B \to A$. Then $g = f^{-1}$ IFF $g \circ f = \mathcal{I}_A$ and $f \circ g = \mathcal{I}_B$.
- Countably infinite sets are equivalent to $\mathbb{N}$.
- Examples of countable sets include
  1. $\mathbb{N}$
  2. $\mathbb{Z}$
  3. $\mathbb{Q}$
  4. Unions of countable sets
  5. Subsets of countable sets
  6. Cartesian products of countable sets
- Examples of uncountable sets include
  1. $\mathbb{R}$
  2. $\mathbb{Q}'$
  3. intervals
  4. the Cantor set

Chapter 4

The Real Numbers

In order to fully understand the real number system, we begin this chapter with some simple inequalities and culminate with the Completeness Axiom−a result that tells us that there are no "gaps" in the number line when we speak of real numbers. Imagine for a moment if we only knew of *rational* numbers. Then if one were asked to solve the equation $x^2 - 5 = 0$, one might mechanically produce $x = \pm\sqrt{5}$ and say "no solution" since the numbers $\pm\sqrt{5}$ are completely foreign. Just from this example, we can see that with only consideration of the rationals on the number line, we are left with a great number of gaps (i.e., irrational numbers) that we would be ignoring (to see this, imagine having the ability to inspect the number line with a microscope). Since numbers are the very heart of quantitative argument, we cannot find an oddity such as this acceptable. Because, in fact, we know that $x^2 - 5 = 0$ has solutions and these solutions are $x = \sqrt{5}$ and $x = -\sqrt{5}$.

4.1 Preliminary Inequalities

In this section, we discuss how the real numbers have a one-to-one correspondence with the number line as we know it. That is,

- given a real number, there is a "location" for it on the number line, and
- if we "locate" a spot on the number line, there is precisely one real number that corresponds to this location.

(See the diagrams on the page that follows.) To study these ideas at a sufficient depth, we need to first speak of upper bounds and lower bounds−topics for the next section. In this section, we lay down the foundation for these notions. The first result may look strange upon first glance but it is so important that we state it as a theorem.

Theorem 4.1.1 *Let $x, y \in \mathbb{R}$. Then*

1. *$x < y + \epsilon$ for every $\epsilon > 0$ IFF $x \leq y$.*

2. *$x > y - \epsilon$ for every $\epsilon > 0$ IFF $x \geq y$.*

Remark 4.1.1 *This is one of our first encounters with the ubiquitous ϵ in mathematics. Here, as is the norm, ϵ is any positive number; notice that it can be as small as we'd like.*

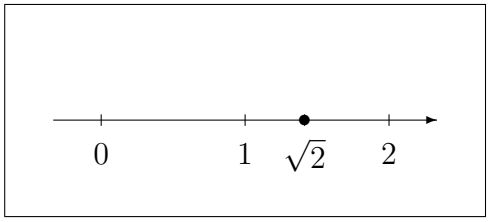

Figure 4.1: Given $\sqrt{2}$, its location can be found on the number line.

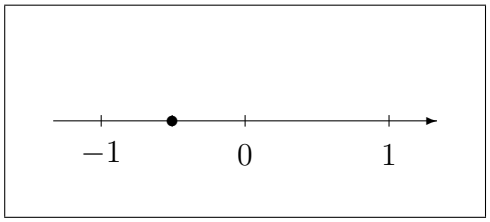

Figure 4.2: Given the location marked above, we find this corresponds to $-\frac{1}{2}$.

Proof. We will prove the first statement and place the other statement in the exercises. Both statements require two proofs since the implication can go in either direction.

- First we will prove that if $x < y + \epsilon$ for every $\epsilon > 0$, then $x \leq y$. This statement is probably easier to prove when considering the contrapositive of the statement. Hence, we will prove: if $x > y$ then $x \geq y + \epsilon$ for some $\epsilon > 0$. Note the subtlety here. It is quite clear that the negation of $x \leq y$ is $x > y$. However, the other end of the statement is somewhat tricky. If we are handed $x < y + \epsilon$ for every $\epsilon > 0$, then the negation must be that we can find an $\epsilon > 0$ such that $x \not< y + \epsilon$. That is, for some particular $\epsilon > 0$, we should get $x \geq y + \epsilon$. The proof, after all of this, is really quite simple. Since $x > y$ we can put $\epsilon = x - y > 0$. Then

  $$y + \epsilon = y + (x - y) = x.$$

  Now $y + \epsilon$ cannot exceed x since it is equal to x. Putting this into symbols, $y + \epsilon \not> x$. In other words, $x \geq y + \epsilon$.

- The converse states that $x \leq y$ implies that $x < y + \epsilon$ for each $\epsilon > 0$. So we let $x \leq y$ with $\epsilon > 0$ given. It must be that either $x < y$ or $x = y$. If $x < y$ then

$$x + 0 < y + 0 < y + \epsilon$$

so $x < y + \epsilon$ follows immediately. If $x = y$ then $x < y + \epsilon$ since $\epsilon > 0$. Either way, $x < y + \epsilon$ for all positive ϵ.

The two proofs above establish the statement. ■

The next result is something that we intuitively "verified" in Chapter 1. Now we are able to prove it.

Theorem 4.1.2 *Let $x \in \mathbb{R}$ and $k \geq 0$. Then $|x| \leq k$ IFF $-k \leq x \leq k$.*

Proof. Again, we'll split the work into two parts.

1. Suppose $|x| \leq k$. Now we know that $-|x| \leq x \leq |x|$ for any value of x since either $x = |x|$ or $x = -|x|$, depending on the sign of x. Now if $|x| \leq k$ then $-|x| \geq -k$. Thus
$$-k \leq -|x| \leq x \leq |x| \leq k$$
so that $-k \leq x \leq k$.

2. In the other direction, we begin with $-k \leq x \leq k$. Again, x can be either positive or negative. If $x \geq 0$, then $|x| = x \leq k$. If $x < 0$, then $x = -|x|$. Using $-k \leq x$, we obtain $-k \leq -|x|$ or $|x| \leq k$. Thus, regardless of the sign of x, $|x| \leq k$.

This completes the proof. ■

Theorem 4.1.3 *Suppose that $x \in \mathbb{R}$. Then $|x| < \epsilon$ for every $\epsilon > 0$ IFF $x = 0$.*

Remark 4.1.2 *The theorem above looks somewhat odd since ϵ never equals zero. In words, this theorem tells us that if a number is strictly less than every positive number (in the absolute sense), then this number must be zero. This important connection appears time and time again in higher mathematics.*

Proof. We suppose that $|x| < \epsilon$ for every $\epsilon > 0$. Using the previous theorem, we get $-\epsilon < x < \epsilon$. Now if we gaze back at **Theorem 4.1.1**, we find a pot of gold here. Rewording this result gives

$$x - y < \epsilon \iff x - y \leq 0$$

and

$$x - y > -\epsilon \iff x - y \geq 0.$$

If we then combine assumptions and combine conclusions, we get

$$-\epsilon < x - y < \epsilon \iff 0 \leq x - y \leq 0.$$

Denoting $x-y$ as the number z, the statement $-\epsilon < z < \epsilon$ leads directly to $0 \leq z \leq 0$ so $z = 0$. ■

Remark 4.1.3 *Notice that the proof of Theorem 4.1.3 does not require two parts like the previous two theorems. This is due to the fact that all statements made in the proof are IFF statements.*

We wrap things up with two inequalities that are carved into the stone of mathematical knowledge. One we recognize from early on.

Theorem 4.1.4 *Let $x, y \in \mathbb{R}$. Then*

1. $|x + y| \leq |x| + |y|$ (**The Triangle Inequality**)
2. $||x| - |y|| \leq |x - y|$ (**Reverse Triangle Inequality**)

Proof.

1. We already did this in Chapter 1. As an aside, notice that if we let $x = a - c$ and $y = c - b$ then we obtain

   $$|a - b| \leq |a - c| + |c - b|,$$

   or, more revealing,

   $$\text{dist}(a, b) \leq \text{dist}(a, c) + \text{dist}(c, b),$$

   which makes sense when we look at Figures 4.3 and 4.4 (see the following page).

2. We will use the Triangle Inequality to prove this. Since the Triangle Inequality holds for *all* real numbers x, it must also work for the real number $x-y$. Thus, we get

   $$|(x - y) + y| \leq |x - y| + |y|$$

   or

   $$|x| - |y| \leq |x - y|. \tag{4.1}$$

   Similarly, if we substitute $y - x$ for y in the Triangle Inequality, we obtain

   $$|x + (y - x)| \leq |x| + |y - x|$$

or

$$|y| \leq |x| + |y - x|.$$

Since $|y - x| = |x - y|$, the previous inequality says

$$|y| - |x| \leq |x - y|. \tag{4.2}$$

Notice that there seems to be some connection lurking between the inequalities in statements (4.1) and (4.2). Realize that $|y|-|x| \leq |x-y|$ just says $|x|-|y| \geq -|x-y|$ upon multiplying by -1. We now have

$$-|x - y| \leq |x| - |y| \leq |x - y|$$

which says $||x| - |y|| \leq |x - y|$ after applying **Theorem 4.1.2**. ■

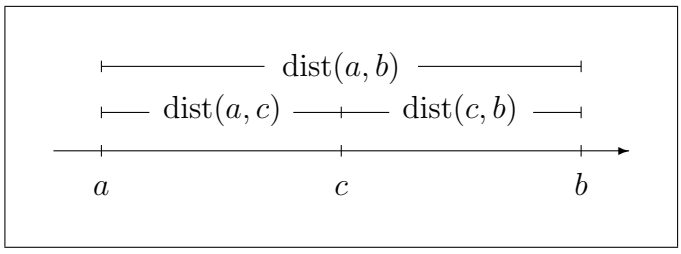

Figure 4.3: $\text{dist}(a, b) \leq \text{dist}(a, c) + \text{dist}(c, b)$ with $a < c < b$

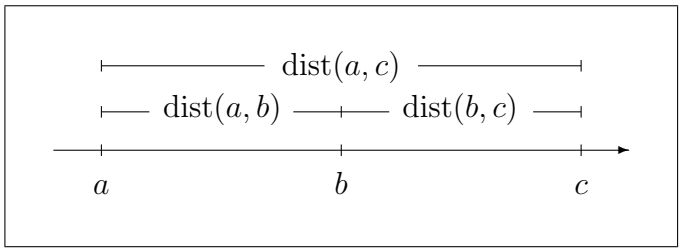

Figure 4.4: $\text{dist}(a, b) \leq \text{dist}(a, c) + \text{dist}(c, b)$ with $a < b < c$

As an aside, it is quite surprising that the "not so obvious" inequality $||x| - |y|| \leq |x - y|$ is equivalent to the self-evident inequality $|xy| \geq xy$. To see this, we begin with $||x| - |y|| \leq |x - y|$. Next use $|b| = \sqrt{b^2}$ on both sides of the equation. This gives $\sqrt{(|x| - |y|)^2} \leq \sqrt{(x - y)^2}$. Since the square root function is increasing, we must have $(|x| - |y|)^2 \leq (x - y)^2$. Then $|x|^2 - 2|x||y| + |y|^2 \leq x^2 - 2xy + y^2$ so $-2|x||y| \leq -2xy$. Thus, $|x||y| \geq xy$ or $|xy| \geq xy$.

Exercises.

1. Let $x, y \in \mathbb{R}$. Show that if $0 < x < y$, then $0 < \frac{1}{y} < \frac{1}{x}$.

2. Show that for $x, y > 1$, we have $x + y < 2xy$.

3. For any x and y, prove that $2(x^2 + y^2) \geq (x + y)^2$.

4. Let $x, y \in \mathbb{R}$. Prove that

   (a) $||x| - |y|| \leq |x \pm y|$

   (b) $\left|\frac{x}{y}\right| = \frac{|x|}{|y|}$

   (c) $\frac{|x|}{1+|x|} + \frac{|y|}{1+|y|} \geq \frac{|x+y|}{1+|x+y|}$

   (d) $|x| + |y| \leq |x + y| + |x - y|$

5. This problem shows that taking the minimum or maximum of two numbers is fundamentally linked to absolute value. That is, for $a, b \in \mathbb{R}$,

   (a) $\max\{a, b\} = \frac{|a-b|+a+b}{2}$

   (b) $\min\{a, b\} = \frac{-|a-b|+a+b}{2}$.

   To prove (a), consider two cases.

   - Case 1: Let $a \geq b$ so $a - b \geq 0$ and $\max\{a, b\} = a$. Then

   $$\frac{|a - b| + a + b}{2} = \frac{a - b + a + b}{2} = \frac{2a}{2} = a = \max\{a, b\}.$$

   - Case 2: Let $b \geq a$ so $b - a \geq 0$ and $\max\{a, b\} = b$. Then

   $$\begin{aligned}\frac{|a - b| + a + b}{2} &= \frac{|(-1)(b - a)| + a + b}{2} \\ &= \frac{|-1||b - a| + a + b}{2} \\ &= \frac{1 \cdot (b - a) + a + b}{2} \\ &= \frac{2b}{2} \\ &= b \\ &= \max\{a, b\}.\end{aligned}$$

   Your task: prove(b).

6. In the scope of the previous problem, prove that

(a) $\max\{x, -x\} = |x|$

(b) $\min\{x, -x\} = -|x|$

7. Prove that any closed interval $[a, b]$ can be expressed as

$$\{x \in \mathbb{R} \,|\, |x - m| \leq n\},$$

where m and n depend on the numbers a and b.

8. Prove the second part of **Theorem 4.1.1**.

4.2 Max and Min vs. Sup and Inf

You may recall from previous coursework that if $A = [0, 1]$, then $\max A = 1$ and $\min A = 0$. The concepts of minimum and maximum should be quite familiar to you. However, what if we consider $B = (0, 1]$? Certainly, $\max B = 1$ but it would be incorrect to say that $\min B = 0$. Notice that $0 \notin B$ so it can't be the minimum element. At this point, one would be tempted to say, "Take the smallest number greater than zero; this is $\min B$." However, the following question immediately arises: What (if any) is the smallest number greater than zero? Plus, even if we could find it, who's to say that it equals $\min B$? In contemplating this, let $\epsilon > 0$ and suppose that $\epsilon = \min B$. Since $\epsilon > 0$, we know that $0 < \frac{\epsilon}{2} < \epsilon$. Notice that $\frac{\epsilon}{2} \in B$ and it is smaller than ϵ. Hmmm, what a problem! A similar difficulty will arise when considering the discrete set $C = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots, \frac{1}{2^n}, \ldots\}$. Certainly $\max C = 1$ but $\min C$ is not obvious (in fact, it does not exist). We will revisit these problems later in this section. For now, we give some important definitions.

Definition 4.2.1 *Let $A \subset \mathbb{R}$. If there is a real number M such that $x \leq M$ for all $x \in A$, then we say that M is an upper bound for A and that A is bounded above.*

Definition 4.2.2 *Let $A \subset \mathbb{R}$. If there is a real number M such that $x \geq m$ for all $x \in A$, then we say that m is a lower bound for A and that A is bounded below.*

Remark 4.2.1 *If A is bounded below and bounded above, we naturally say that A is bounded.*

Notice that it is possible for a set to have multiple upper bounds or perhaps none at all (similarly for lower bounds). For example, $D = \{5, 2, \pi\}$ has any of the following as upper bounds: $5, 12, 10^6, \pi^e$. However, we do not say that ∞ is an upper bound because ∞ is not a real number (so it violates the definition). Interestingly enough, 5 is the least of all upper bounds. Likewise, 2 is the greatest of all lower bounds. With this in mind, we now state the familiar definitions of maximum and minimum.

Definition 4.2.3 *Let M be an upper bound for A. If $M \in A$, then M is the largest element (or maximum) of A and we write $M = \max A$. Similarly, let m be a lower bound for A. If $m \in A$, then m is the smallest element (or minimum) of A and we write $m = \min A$.*

Example 4.2.1 *Here are some miscellaneous examples:*

1. *It is straightforward to see that the following two statements are true:* $\max D = 5$ *and* $\min D = 2$.

2. *Now look back at the example with* $B = (0, 1]$. *Again,* $\max B = 1$. *Also, note that B is bounded below by any nonpositive number (for example, $-5, 0, -1$, and $-\frac{1}{10}$ are all suitable lower bounds for B). Zero is the largest of these lower bounds but $0 \notin B$. Hence,* $\min B$ *does not exist.*

3. *Let* $E = (-\infty, 1]$. *E is not bounded below so it has no minimum. On the other hand, it is bounded above with* 1 *being the smallest of these upper bounds. Since* $1 \in (-\infty, 1]$, $1 = \max E$.

Notice that a set A is bounded IFF it is contained in some interval of finite length. That is, A is bounded IFF $A \subset [a, b]$ for some real numbers a and b (so neither a nor b can equal $\pm\infty$). Since $\text{length}([a, b]) = b - a < \infty$, we can take a to be a lower bound of A and b to be an upper bound of A.

Remark 4.2.2 *Although* $[0, 1]$ *is uncountable, notice that it is bounded. On the other hand,* $\mathbb{N}$ *is countable but unbounded. To see this, recall that*

$$\mathbb{N} = \{1, 2, 3, 4, \ldots\} \subset \mathbb{R} = (-\infty, \infty).$$

Now take $x \in \mathbb{R}$. Then put $y = \lfloor x \rfloor + 1 \in \mathbb{N}$ so that $y > x$. In words, for any real number x (no matter how large), we can always find a positive integer y that is greater than it. Since x is arbitrary, this shows that $\mathbb{N}$ is not bounded above (so $\mathbb{N}$ is not bounded). In sum, this remark shows that there is no connection between countability and boundedness.

We are now in a position to give the most important definition of this section.

Definition 4.2.4 *Let $A \subset \mathbb{R}$ be bounded above. The number $L \in \mathbb{R}$ is called the least upper bound (or supremum) if*

1. *L is an upper bound of A and*

2. *no number smaller than L is an upper bound for A.*

The notation is $L = \text{lub}\, A$ *or* $L = \sup A$.

In mathematical terms, the previous definition states that for any positive number ϵ, $L - \epsilon$ cannot be an upper bound for A.

Definition 4.2.5 *Let $A \subset \mathbb{R}$ be bounded below. The number $l \in \mathbb{R}$ is called the greatest lower bound (or infimum) if*

1. *l is a lower bound of A and*
2. *no number greater than l is a lower bound for A.*

The notation is $l = \text{glb}\, A$ or $l = \inf A$.

In mathematical terms, the previous definition states that for any positive number ϵ, $l + \epsilon$ cannot be a lower bound for A. Here is a simple example.

Example 4.2.2 *Let us revisit $B = (0, 1]$. Certainly $\sup B = 1$ since*

- *1 is an upper bound for B and*
- *for $\epsilon > 0$, $1 - \epsilon$ is not an upper bound for B (that is, no number less than 1 can serve as an upper bound).*

Similarly, $\inf B = 0$ since

- *0 is a lower bound and*
- *for $\epsilon > 0$, $0 + \epsilon$ is not a lower bound. See the diagram below.*

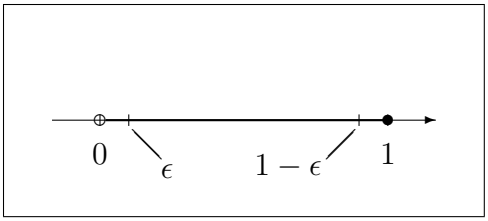

Figure 4.5: The number $1-\epsilon$ is not an upper bound for $(0, 1]$ since there are infinitely many numbers belonging to B in the interval $(1 - \epsilon, 1)$. Similarly, ϵ is not a lower bound for $(0, 1]$ since there are infinitely many numbers belonging to B in $(0, \epsilon)$.

Remark 4.2.3 *Notice that, in the previous example, $\max B = 1 = \sup B$. This is not always the case; that is, suprema and maxima do not always coincide. The same applies to infima and minima. For example, even though $\inf B = 0$, we already know that $\min B$ fails to exist.*

Before giving more examples, we state a theorem.

Theorem 4.2.1 *If a set A has a supremum, then it is unique. Likewise, if A has an infimum, it is unique.*

Proof. We prove the statement for the infimum and leave the other part as an exercise. Let $l = \inf A$ and proceed by contradiction. Suppose l' is also a greatest lower bound with l' different from l. Then there are two cases:

1. Case 1: Let $l' > l$. Then l' cannot be a lower bound for A (by definition, no number greater than l can be a lower bound). This is a contradiction to $l' = \text{glb}\, A$.

2. Case 2: Let $l' < l$. This immediately says that l' cannot be the greatest lower bound since l is a lower bound and $l > l'$. This again contradicts the hypothesis.

Hence, we obtain $l = l'$. ■

Example 4.2.3 *Recall $C = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots, \frac{1}{2^n}, \ldots\}$ from earlier in this section. It is apparent that $\sup C = 1$. We now prove that $\inf C = 0$ as you may have predicted. Suppose that $\inf C \neq 0$. Since C contains no negative elements, we further assume that $\inf C > 0$. Now denote $\inf C = \epsilon$. Since ϵ is the infimum, $x \geq \epsilon$ for all $x \in C$. Thus, $\frac{1}{2^n} \geq \epsilon$ for all $n \in \mathbb{N}$, or, rephrasing, $1 \geq \epsilon \cdot 2^n$ for all $n \in \mathbb{N}$. Do you see the contradiction? (For instance, consider $n = \lfloor \frac{\ln(1/\epsilon)}{\ln 2} \rfloor + 1 \in \mathbb{N}$ which comes from isolating $1 = \epsilon \cdot 2^n$ for n.)*

Example 4.2.4 *The set $(2, 3)$ contains neither its* inf *nor its* sup, *given by 2 and 3, respectively.*

Example 4.2.5 *Consider $E = \{5\}$, a singleton. Then $\sup E = 5 = \inf E$.*

Example 4.2.6 *Let $F = \emptyset$. Recall that $\emptyset \subset A$ for any set A. Hence $F \subset [a, b]$ for any $a, b \in \mathbb{R}$. In other words, any $b \in \mathbb{R}$ can serve as an upper bound implying that $\emptyset$ does not have a least upper bound. Similarly, $\inf F$ does not exist either.*

Remark 4.2.4 *Did you notice the peculiarity of set F above? Certainly, $\emptyset$ is bounded since $\emptyset \subset A$ for any set A. However, $\emptyset$ has no* inf *and no* sup. *Can you think of any other set that has this property?*

Before closing, we give an interesting problem.

Problem 4.2.1 *Find a set $A \subset \mathbb{R}$ that is both countable and bounded yet $\sup A \notin A$ and $\inf A \notin A$.*

Solution. A bounded set suggests that we choose something like an interval $A = (3,4)$. However, A is not countable so we make a revision: Let $A = \{x \in (3,4) \,|\, x \in \mathbb{Q}\}$. Certainly this is countable since $\mathbb{Q}$ is countable. We are sure to find a rational number in $(3,4)$ as close as we'd like to 3. A similar statement is true for the number 4. (However, the validity of these statements rests on the Completeness Axiom−the main act of the next section.) Hence, $3 = \inf A$ and $4 = \sup A$ yet $3 \notin A$ and $4 \notin A$. Many other examples are possible. □

Exercises.

1. Find the inf, sup, min, and max of each set below. Realize that not all of the sets possess these attributes.
   (a) $\{1, 4, 7, 10, \ldots\}$
   (b) $\left\{\sqrt{2}, \sqrt{2+\sqrt{2}}, \sqrt{2+\sqrt{2+\sqrt{2}}}, \ldots\right\}$
   (c) $\left\{2, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \ldots\right\}$
   (d) the Cantor set K
   (e) $\left\{(1+1)^1, \left(1+\frac{1}{2}\right)^2, \left(1+\frac{1}{3}\right)^3, \ldots\right\}$
   (f) $(3, \pi)$
   (g) $[15, 17.2)$
   (h) $\bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right)$
   (i) $\bigcup_{n=1}^{\infty} [2n, n^2+1)$
2. Let $S \neq \emptyset$ be a bounded subset of $\mathbb{R}$.
   (a) Prove that $\inf S \leq \sup S$.
   (b) Suppose that $\inf S = \sup S$. What is special about S?
3. Suppose that $A \subset B \subset \mathbb{R}$, both A and B nonempty. Prove that
$$\inf B \leq \inf A \leq \sup A \leq \sup B.$$

4.3 The Completeness Axiom

In this section, we uncover one of the most important axioms in all of mathematics, the Completeness Axiom. The elementary theorems from Calculus and higher Analysis (e.g., the Intermediate Value Theorem) are left dangling in the wind without this axiom. We state it now and give many of its far-reaching consequences.

Axiom 4.3.1 (The Completeness Axiom) *Let $A \subset \mathbb{R}$ be nonempty. If A is bounded above, then A has a least upper bound in $\mathbb{R}$. That is,* $\sup A$ *exists and it is a real number.*

Remark 4.3.1 *Because the statement is an axiom, there is no proof. In a course of set theory, the Completeness Axiom can be taken as a construction from the development of the real numbers—there, it can be taken to be a theorem, and thus, proved.*

Informally, the axiom states that $\mathbb{R}$ has no gaps; if we were to consider only $\mathbb{Q}$, this would not be true. The next problem addresses precisely this concern.

Example 4.3.1 *Consider the set given by*

$$A = \{1, 1.7, 1.73, 1.732, 1.7320, 1.73205, \ldots\}.$$

Show that the Completeness Axiom does not hold for the rational numbers $\mathbb{Q}$.

Solution. First, $A \subset \mathbb{Q}$ since all elements are rational. Also, A is bounded above (e.g., $x \leq 2$ for all $x \in A$). Finally, after pondering A, we see that A is a set in which although all members are rational, each additional element contains a subsequent digit of the irrational number $\sqrt{3}$. Since A is an infinite set, we can never "see" all of its elements; hence there is no rational number that can serve as a least upper bound for A (but there are many rationals that may serve as upper bounds for A). To see this, suppose we claim that 1.732050807 is a least upper bound for A. Noting that

$$\sqrt{3} \approx 1.732050807568877$$

(correct to the fifteenth decimal place), we see why we cannot assert 1.732050807 as the least upper bound. However, we can say that $\sqrt{3} = \text{lub}\, A$ since

1. $x \leq \sqrt{3}$ for all $x \in A$ and
2. $\sqrt{3} - \epsilon$ is not an upper bound for any $\epsilon > 0$.

Thus, although we have a set A containing all rational numbers, $\sup A$ is not rational. We see that $\sup A = \sqrt{3} \in \mathbb{Q}'$. Hence, the Completeness Axiom does not hold for $\mathbb{Q}$. □

Remark 4.3.2 *The above example reminds us that $\mathbb{Q}$ has "gaps" or "holes" in it; all of these holes are filled in by irrational numbers. This is precisely why $\mathbb{R}$ or $\mathbb{Q} \cup \mathbb{Q}'$ has no holes in it.*

We now make a brief comment (in general) about the Completeness Axiom. The fact that we stated the axiom in terms of lub's is insignificant. We could have just as well stated that "if A is bounded below, then A has a greatest lower bound in $\mathbb{R}$." However, since we did not do this, we can state this as a theorem and use the Completeness Axiom as the foundation of its proof.

Theorem 4.3.1 *Let $A \subset \mathbb{R}$ be nonempty. If A is bounded below, then A has a greatest lower bound in $\mathbb{R}$. That is,* $\inf A$ *exists and it is a real number.*

Proof. The proof is quite elegant and just borrows from the Completeness Axiom. First, denote

$$\hat{A} = \{x \in \mathbb{R} \mid -x \in A\}.$$

That is, all of the elements of $\hat{A}$ are opposites of those in A. Since A is bounded below, let m be a lower bound for A. By the very construction of $\hat{A}$, $-m$ must be an upper bound for $\hat{A}$. To see this, choose $x \in \hat{A}$ so $-x \in A$. Since m is a lower bound for A, $m \leq -x$. This is equivalent to saying that $-m \geq x$. Hence, $-m$ serves as an upper bound for $\hat{A}$. Then, by the Completeness Axiom, $\hat{A}$ must have a least upper bound, call it $\hat{m}$. See the figure below.

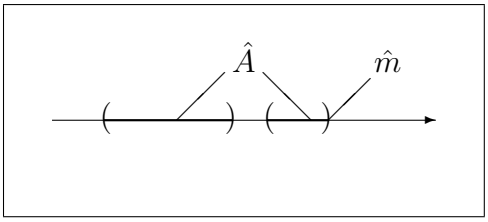

Figure 4.6: $\hat{m} = \text{lub}\,\hat{A}$

Now we show that $-\hat{m}$ is the greatest lower bound for A. Thus, we show two things:

1. $-\hat{m} \leq -x$ for all $-x \in A$ and

2. no number greater that $-\hat{m}$ is a lower bound for A.

To show 1, we note that since $\hat{m} = \sup \hat{A}$, $x \leq \hat{m}$ for all $x \in \hat{A}$. Now $x \leq \hat{m}$ says that $-x \geq -\hat{m}$. Since $-x \in A$ (for $x \in \hat{A}$), $-\hat{m} \leq -x$ tells us that $-\hat{m}$ is a lower bound for A. To prove 2, we proceed by contradiction. Fix $\epsilon > 0$ and put $M = -\hat{m} + \epsilon > -\hat{m}$. Suppose that M is a lower bound for A. Then this would mean that $M \leq -x$ for all $-x \in A$. In other words, $-\hat{m} + \epsilon \leq -x$ or $x \leq \hat{m} - \epsilon$. The previous statement says that $x \leq \hat{m} - \epsilon$ for all $x \in \hat{A}$. That is, $\hat{m} \neq \sup A$. This contradicts our earlier work so we conclude that $-\hat{m}$ is the greatest lower bound for A. ■

We next state an important and well known consequence of the Completeness Axiom−the Archimedean Property of $\mathbb{R}$. In plain English, this result asserts that there is no largest natural number n. That is, $\mathbb{N}$ is not bounded above. Surprisingly, this has a lot to do with $\mathbb{R}$.

Theorem 4.3.2 (Archimedean Property) *Let $a, b > 0$. Then there exists an $n \in \mathbb{N}$ such that $na > b$.*

Remark 4.3.3 *At first glance, the Archimedean Property doesn't seem to say much of anything significant. However, a closer look tells us that it really says that no matter how small the number a (and perhaps how large the number b), we can always find a multiple of a that exceeds b.*

Here are a few examples that illustrate the point.

Example 4.3.2

1. *Let $a = 5$ and $b = 3$. Then we can obviously choose $n = 1$ (or any integer larger than 1) so that $1 \cdot 5 > 3$.*

2. *Let $a = 3$ and $b = 5$. Then we can pick $n = 2$ (or any integer larger than 2) so that $2 \cdot 3 > 5$.*

3. *A more interesting case is when a is very small and b is very large. For example, put $a = \epsilon > 0$, where ϵ is small. Since we want b to be large, let $b = \frac{1}{\epsilon}$. Can we find a multiple of a that is greater than b? To find this $n \in \mathbb{N}$, note that we need $n\epsilon > \frac{1}{\epsilon}$ or $n > \frac{1}{\epsilon^2}$. Since $\frac{1}{\epsilon^2}$ is most likely not in $\mathbb{N}$, we choose $n = \lfloor \frac{1}{\epsilon^2} \rfloor + 1 \in \mathbb{N}$. Then certainly $na > b$.*

Proof. We prove this by contradiction. That is, suppose there is no $n \in \mathbb{N}$ such that $na > b$. In other words, for all $a, b > 0$ and $n \in \mathbb{N}$, $na \leq b$. Now consider the set

$$S = \{na \,|\, n \in \mathbb{N}\}.$$

Since $na \leq b$, this tells us that b is an upper bound for S. By the Completeness Axiom, we know that S has a least upper bound so we denote $M = \text{lub}\, S$. Now no number smaller than M can be an upper bound for S. In particular, $M - a$ is not an upper bound for S. Rephrasing, for some $n \in \mathbb{N}$, $M - a < na$ (i.e., one of the elements in S is greater than this $M - a$). This implies that $M < a + na$ or $M < (n+1)a$. Now notice that $(n+1)a \in S$ since $n + 1 \in \mathbb{N}$. Since M is less than an element of S, M can't possibly be an upper bound for S. This contradicts our earlier work so the theorem is proved. ■

We now state a result that reveals the multiple guises of the Archimedean Property; all of the statements in the following theorem are different ways of saying the same thing.

Theorem 4.3.3 *The following statements are equivalent:*

1. $\mathbb{N}$ *is not bounded above in* $\mathbb{R}$.
2. *The Archimedean Property of* $\mathbb{R}$.
3. *For any* $\epsilon > 0$, *there exists an* $n \in \mathbb{N}$ *such that* $0 < \frac{1}{n} < \epsilon$.

Proof. It is necessary to show that

$$1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$$

to establish the equivalence. We omit the proof of $1 \Rightarrow 2$ since we practically did this earlier. To prove $2 \Rightarrow 3$, we have the following. For any $a, b > 0$, there exists an $n \in \mathbb{N}$ such that $na > b$. With $a > 0$ and $b = 1$, we know that there is an $\hat{n} \in \mathbb{N}$ such that $\hat{n}a > 1$ or $\frac{1}{\hat{n}} < a$. Since $\hat{n} \in \mathbb{N}$, $\frac{1}{\hat{n}} > 0$ so that $0 < \frac{1}{\hat{n}} < a$. To prove $3 \Rightarrow 1$, suppose that 1 were false. Thus, we assume that $\mathbb{N}$ is bounded above by some $r \in \mathbb{R}$. That is, $n \leq r$ for all $n \in \mathbb{N}$. Then $\frac{1}{n} \geq \frac{1}{r}$ for all $n \in \mathbb{N}$ (since both $n, r > 0$). Finally, since $\frac{1}{r} \in \mathbb{R}$, the previous statement translates to $\frac{1}{n} > \epsilon$ (put $\epsilon = \frac{1}{r}$ and note that $\epsilon > 0$) for all $n \in \mathbb{N}$. This contradicts statement 3. ■

Remark 4.3.4 *You probably noticed the subtle step in the proof that for* $0 < n < r$, *we have* $0 < \frac{1}{r} < \frac{1}{n}$. *Although this may seem elementary, you should probably think about how to prove it. Maybe peek back at section* 4.1.

We now state a proposition assuring us of the existence of square roots−something that is clearly overlooked in elementary courses. Surprisingly enough, the proof requires the Completeness Axiom, the Archimedean Property, and one of the preliminary inequalities from the beginning of this chapter.

Proposition 4.3.1 *Given a real number* $c > 0$, c *has a positive square root.*

Remark 4.3.5 *The familiar notation for this square root is* $+\sqrt{c}$. *Furthermore, one can show that this positive square root is unique though we do not do this here.*

Proof. Consider the set

$$A = \{x \in \mathbb{R} \,|\, x \geq 0,\ x^2 \leq c\}.$$

First we note that A is nonempty (e.g., $0 \in A$). We now show that A is bounded above. Assume the contrary. That is, suppose that for $x \in A$ we have $x > \max\{c, 1\}$. In other words, we are assuming there is an element of A that exceeds the larger of c and 1. Then, since $x > \max\{c, 1\}$, we may write $x > c$ and $x > 1$. Hence,

$$x^2 = x \cdot x > x \cdot 1 = x > c.$$

Thus, we've established $x^2 > c$. Since x enjoys this property, $x \notin A$. This is a contradiction so A must be bounded above. Now, by the Completeness Axiom, A possesses a least upper bound; denote $m = \text{lub}\, A$. We now prove that $m^2 = c$. In doing this, we first prove that $m > 0$ (though this may seem obvious to you). Note that

$$\begin{aligned}(\min\{c,1\})^2 &= \min\{c,1\} \cdot \min\{c,1\} \\ &\leq \min\{c,1\} \cdot 1 \\ &= \min\{c,1\} \\ &\leq c.\end{aligned}$$

(Notice how the steps above use the facts that $\min\{c,1\} \leq 1$ and $\min\{c,1\} \leq c$. If you can't visualize this, just assign a value to c.) From the above work, $\min\{c,1\} \in A$. (Note that $x \in A$ provided that $x \geq 0$ and $x^2 \leq c$.) Thus, since $\min\{c,1\} \in A$, we must have $m > 0$ since $c > 0$ by assumption. The implication here is that we can find an $\epsilon > 0$ such that $0 < \epsilon < m$; we know this ϵ exists due to the Archimedean Property. Now we have

$$0 < m - \epsilon < m < m + \epsilon$$

so that

$$(m-\epsilon)^2 < m^2 < (m+\epsilon)^2 \tag{4.3}$$

(see the remark immediately following this proof). In a similar manner, we ask you to supply the details in showing that

$$(m-\epsilon)^2 < c < (m+\epsilon)^2.$$

Upon multiplication by -1, we get

$$-(m+\epsilon)^2 < -c < -(m-\epsilon)^2 \tag{4.4}$$

so by adding (4.3) and (4.4) we obtain

$$(m-\epsilon)^2 - (m+\epsilon)^2 < m^2 - c < (m+\epsilon)^2 - (m-\epsilon)^2$$

or

$$-\left[(m+\epsilon)^2 - (m-\epsilon)^2\right] < m^2 - c < (m+\epsilon)^2 - (m-\epsilon)^2.$$

In other words,

$$|m^2 - c| < (m+\epsilon)^2 - (m-\epsilon)^2$$

or

$$|m^2 - c| < 4m\epsilon.$$

Now notice that we can make $m^2 - c$ as small as we'd like since ϵ is arbitrarily small (in other words, we can make $|m^2 - c|$ smaller than any positive number K). So we have

$$|m^2 - c| < K$$

for arbitrarily small K. Applying **Theorem 4.1.3**, $|m^2 - c| = 0$ so that $m^2 = c$. Therefore, given $c > 0$, there exists an m such that $m^2 = c$. In other words, $m = \sqrt{c} = \text{lub}\, A$. ■

Remark 4.3.6 *You will notice that the following result was used in the proof: if $0 < a < b$, then $a^2 < b^2$. Put simply, smaller numbers have smaller squares (when dealing with positive quantities). Proving this is fairly direct. For example, we might be motivated to multiply both sides of the inequality $a < b$ by the numbers a or b to arrive at the result. Multiplying by a gives $a^2 < ab$ whereas multiplying by b gives $ab < b^2$. In both cases, the direction of the inequality is unchanged since $a, b > 0$. So it follows that $a^2 < ab < b^2$ or that $a^2 < b^2$.*

All matters aside, the previous proof is not an easy one. However, it illustrates that it is often necessary to prove auxiliary statements before reaching a desired goal. As you embark on future courses, this becomes more commonplace.

After this section, it is hoped that an appreciation has been gained for some of the delicate features of the real number system. It is very important to understand $\mathbb{R}$ before embarking on more advanced courses. You are now in a position to study some of the classical material in real variables such as limits, continuity, differentiation, Riemann integration, and infinite sequences/series. As mentioned previously, a course in theoretical Calculus will invoke the Completeness Axiom to prove the following statement:

> Let f be continuous on $[a, b]$. If $f(a) \leq k \leq f(b)$, then there exists a number $c \in [a, b]$ such that $f(c) = k$.

You might remember the above statement as the Intermediate Value Theorem, one of the many theorems from Calculus that cannot be proved without the material discussed here.

Exercises.

1. Let $x, y \in \mathbb{R}$ with $x < y$. Prove that there exists an $r \in \mathbb{Q}$ such that $x < r < y$.
2. Repeat Exercise 1 except for $z \in \mathbb{Q}'$.
3. Suppose $0 < \frac{m}{n} < 1$ where $\frac{m}{n}$ is irreducible. Prove that there exists an $N \in \mathbb{N}$ such that
$$\frac{1}{N+1} < \frac{m}{n} < \frac{1}{N}.$$

4. Prove that for any $x > 0$, there is an $n \in \mathbb{N}$ such that $n^{-1} < x < n$.

5. Let p be a prime number. Prove that there exists an $n \in \mathbb{R}^+$ such that $n^2 = p$.

6. Prove that the following statement is equivalent to the Archimedean Property: for each $x \in \mathbb{R}$, there exists an $n \in \mathbb{N}$ such that $n > x$.

4.4 Summary: Odds and Ends

- Important statements of inequality:

  1. Let $x, y \in \mathbb{R}$. Then
     (a) $x \leq y$ IFF $x < y + \epsilon$ for every $\epsilon > 0$.
     (b) $x \geq y$ IFF $x > y - \epsilon$ for every $\epsilon > 0$.
  2. Given any real number x, $|x| < \epsilon$ for every $\epsilon > 0 \iff x = 0$.
  3. Triangle Inequalities:
     (a) $|x + y| \leq |x| + |y|$ (**The Triangle Inequality**)
     (b) $||x| - |y|| \leq |x - y|$ (**Reverse Triangle Inequality**)
  4. If $0 < a < b$, then $\frac{1}{b} < \frac{1}{a}$. In words, smaller numbers have larger reciprocals.
  5. If $0 < a < b$, then $a^2 < b^2$. In words, smaller numbers have smaller squares.

- Supremum and Infimum:

  1. Let $A \subset \mathbb{R}$ be bounded above. The number $L \in \mathbb{R}$ is called the supremum (or least upper bound) if
     (a) L is an upper bound of A and
     (b) no number smaller than L is an upper bound for A.

     The notation is $L = \text{lub}\, A$ or $L = \sup A$. Furthermore, L is unique.
  2. Let $A \subset \mathbb{R}$ be bounded below. The number $l \in \mathbb{R}$ is called the infimum (or greatest lower bound) if
     (a) l is a lower bound of A and
     (b) no number greater than l is a lower bound for A.

     The notation is $l = \text{glb}\, A$ or $l = \inf A$. Furthermore, l is unique.

- Completeness Axiom: Let $A \subset \mathbb{R}$ be nonempty. If A is bounded above, then A has a least upper bound in $\mathbb{R}$. That is, $\sup A$ exists and it is a real number. Informally, this tells us that the real numbers "fill" the entire number line. Just as well, this axiom could be stated in terms of greatest lower bounds without loss of generality.

- Archimedean Property: Let $a, b > 0$. Then there exists an $n \in \mathbb{N}$ such that $na > b$. Informally, regardless of the size of the numbers a and b, one can always find a multiple of a that exceeds b. Below are two statements that are equivalent to the Archimedean Property.

  1. $\mathbb{N}$ is not bounded above in $\mathbb{R}$.
  2. For any $\epsilon > 0$, there exists an $n \in \mathbb{N}$ such that $0 < \frac{1}{n} < \epsilon$.

Chapter 5

Metric Spaces

5.1 Introduction

In our continued study of the real numbers, we now introduce the idea of an "abstract space." One type you may have encountered already is a linear space (or vector space). In this section, we study where Chapters 2 and 3 collide—a metric space is a set equipped with a special type of function. The simplest illustration arises from examination of the absolute value function. For review purposes, we have, for $x, y \in \mathbb{R}$

1. $|x| = 0$ if and only if $x = 0$
2. $|x| > 0$ for $x \neq 0$
3. $|x| = |-x|$
4. $|x + y| \leq |x| + |y|$

The first three statements are quite apparent (see Chapter 1) while we recognize the last property as the triangle inequality. Also, recall that we introduced the notation $d(x, y) = |x - y|$ in Chapter 1 and used it again in section 4.1. Adopting this notation, we can state the following:

1. $d(x, x) = 0$ since the distance between a point and itself is zero.
2. $d(x, y) > 0$ provided that $x \neq y$. That is, the distance between two distinct points is positive.
3. $d(x, y) = d(y, x)$. The distance between x and y equals the distance between y and x.
4. $d(x, y) \leq d(x, z) + d(z, y)$. This was proved in **Theorem 4.1.4**.

See the graphic below for an illustration of the fourth property with $x < y < z$.

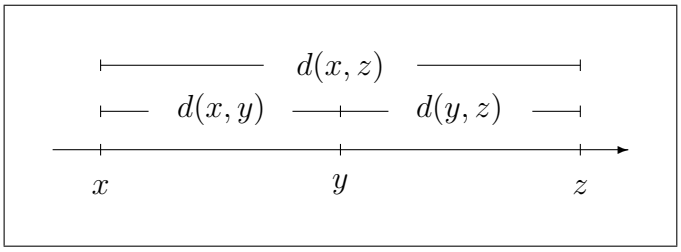

Figure 5.1: $d(x, y) \leq d(x, z) + d(z, y)$ with $x < y < z$

The elements of $\mathbb{R}$, equipped with the distance function d, is called a metric space. The general definition follows.

Definition 5.1.1 *Let A be a set. A metric or distance function for A is a function $d : A \times A \to [0, \infty)$ with the following properties:*

1. $d(x, x) = 0$
2. $d(x, y) > 0$ *for* $x \neq y$.
3. $d(x, y) = d(y, x)$.
4. $d(x, y) \leq d(x, z) + d(z, y)$.

It is standard to refer to $\langle A, d \rangle$ as a metric space.

Note that the definition above is a generalization of the example $d(x, y) = |x - y|$ with $A = \mathbb{R}$. Also, the notation $d : A \times A \to [0, \infty)$ is very important; d is a function of two variables where $(x, y) \in A \times A$ and $d(x, y) \in [0, \infty)$.

5.2 Two Famous Inequalities

In this section, we present the statements and proofs of two classical inequalities: the Schwarz inequality and the Minkowski inequality. We study them here as they prove useful in the upcoming sections.

Theorem 5.2.1 (Schwarz) *Let $a_n, b_n \in \mathbb{R}$, $n = 1, 2, 3, \ldots, N$. Then*

$$\left| \sum_{n=1}^{N} a_n b_n \right| \leq \sqrt{\sum_{n=1}^{N} {a_n}^2} \sqrt{\sum_{n=1}^{N} {b_n}^2}.$$

Proof. Let $x \in \mathbb{R}$ and formulate $\sum_{n=1}^{N}(a_n - xb_n)^2$. Certainly

$$\sum_{n=1}^{N}(a_n - xb_n)^2 \geq 0$$

so that

$$\sum_{n=1}^{N} {a_n}^2 - 2x\sum_{n=1}^{N} a_n b_n + x^2 \sum_{n=1}^{N} {b_n}^2 \geq 0.$$

Assuming $\sum_{n=1}^{N} {b_n}^2 \neq 0$ and putting $x = \frac{\sum_{n=1}^{N} a_n b_n}{\sum_{n=1}^{N} {b_n}^2}$, we obtain

$$\sum_{n=1}^{N} {a_n}^2 - 2\frac{\left(\sum_{n=1}^{N} a_n b_n\right)^2}{\sum_{n=1}^{N} {b_n}^2} + \frac{\left(\sum_{n=1}^{N} a_n b_n\right)^2}{\sum_{n=1}^{N} {b_n}^2} \geq 0$$

or

$$\sum_{n=1}^{N} {a_n}^2 \sum_{n=1}^{N} {b_n}^2 - \left(\sum_{n=1}^{N} a_n b_n\right)^2 \geq 0$$

or

$$\left(\sum_{n=1}^{N} a_n b_n\right)^2 \leq \sum_{n=1}^{N} {a_n}^2 \sum_{n=1}^{N} {b_n}^2.$$

Taking square roots, we obtain

$$\left|\sum_{n=1}^{N} a_n b_n\right| \leq \sqrt{\sum_{n=1}^{N} {a_n}^2}\sqrt{\sum_{n=1}^{N} {b_n}^2},$$

the inequality sought. ■

Remark 5.2.1 *Under certain conditions, it can be shown that the Schwarz inequality also holds for the mighty $N = \infty$.*

Theorem 5.2.2 (Minkowski) *Let $a_n, b_n \in \mathbb{R}$, $n = 1, 2, 3, \ldots, N$. Then*

$$\sqrt{\sum_{n=1}^{N}(a_n + b_n)^2} \leq \sqrt{\sum_{n=1}^{N} {a_n}^2} + \sqrt{\sum_{n=1}^{N} {b_n}^2}.$$

Remark 5.2.2 *Notice the special case when $N = 1$. The inequality reduces to $\sqrt{(a_1 + b_1)^2} \leq \sqrt{{a_1}^2} + \sqrt{{b_1}^2}$, or $|a_1 + b_1| \leq |a_1| + |b_1|$, the familiar triangle inequality.*

Proof. Since we are trying to bound $\sqrt{\sum_{n=1}^{N}(a_n+b_n)^2}$, we first work with $\sum_{n=1}^{N}(a_n+b_n)^2$. We have

$$\begin{aligned}\sum_{n=1}^{N}(a_n+b_n)^2 &= \sum_{n=1}^{N}{a_n}^2 + 2\sum_{n=1}^{N}a_nb_n + \sum_{n=1}^{N}{b_n}^2 \\ &\leq \sum_{n=1}^{N}{a_n}^2 + 2\left|\sum_{n=1}^{N}a_nb_n\right| + \sum_{n=1}^{N}{b_n}^2.\end{aligned}$$

Next, we use the Schwarz inequality on $\left|\sum_{n=1}^{N}a_nb_n\right|$ to arrive at

$$\begin{aligned}\sum_{n=1}^{N}(a_n+b_n)^2 &\leq \sum_{n=1}^{N}{a_n}^2 + 2\sqrt{\sum_{n=1}^{N}{a_n}^2}\sqrt{\sum_{n=1}^{N}{b_n}^2} + \sum_{n=1}^{N}{b_n}^2 \\ &= \left(\sqrt{\sum_{n=1}^{N}{a_n}^2} + \sqrt{\sum_{n=1}^{N}{b_n}^2}\right)^2.\end{aligned}$$

So since

$$\sum_{n=1}^{N}(a_n+b_n)^2 \leq \left(\sqrt{\sum_{n=1}^{N}{a_n}^2} + \sqrt{\sum_{n=1}^{N}{b_n}^2}\right)^2,$$

we take square roots to obtain

$$\begin{aligned}\sqrt{\sum_{n=1}^{N}(a_n+b_n)^2} &\leq \left|\sqrt{\sum_{n=1}^{N}{a_n}^2} + \sqrt{\sum_{n=1}^{N}{b_n}^2}\right| \\ &= \sqrt{\sum_{n=1}^{N}{a_n}^2} + \sqrt{\sum_{n=1}^{N}{b_n}^2},\end{aligned}$$

the desired inequality. ■

Remark 5.2.3 *It is worth mentioning (again) that the Minkowski inequality remains true for N replaced by ∞. That is, $\sqrt{\sum_{n=1}^{\infty}(a_n+b_n)^2} \leq \sqrt{\sum_{n=1}^{\infty}{a_n}^2} + \sqrt{\sum_{n=1}^{\infty}{b_n}^2}$ under certain assumptions. However, background knowledge of infinite series and higher analysis is needed to prove this.*

You should glance back at section 1.2 (The Triangle Inequality). There we proved that for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$. This is precisely the Minkowski inequality!

Take a moment to compare the two proofs. The first one uses some elementary results on vector addition, dot products, and the angle between two vectors. The proof in this section uses fundamental properties of absolute value as well as the Schwarz inequality.

Remark 5.2.4 *Although it seems fairly innocent to let $N = \infty$ in both of the inequalities discussed here, we need to be careful when making such a declaration. With all of the earlier warnings about infinity, this should not be surprising.*

5.3 Examples

In this section we give an abundance of examples of metric spaces. You will soon notice that verification of the first three properties is usually trivial; it is usually the triangle inequality that is the most difficult to prove. Hence, we will leave the simpler property verifications as exercises. In the beginning however, we give some unabridged proofs so that you may acquire a taste for the style of writing involved.

Example 5.3.1 *Let $A_1 = \mathbb{R}$ with $d_1(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right|$ for $x, y > 0$. Prove that d_1 is a metric for $(0, \infty)$.*

Solution. We verify all four properties:

1. $d_1(x, x) = \left|\frac{1}{x} - \frac{1}{x}\right| = |0| = 0$.

2. $d_1(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right| = \left|\frac{y-x}{xy}\right| = \frac{|y-x|}{|xy|} > 0$ provided that $x, y > 0$ and $x \neq y$.

3. Using some elementary properties of absolute value, we note that

$$\begin{aligned} d_1(x, y) &= \left|\frac{1}{x} - \frac{1}{y}\right| \\ &= |-1| \left|\frac{1}{x} - \frac{1}{y}\right| \\ &= \left|(-1)\left(\frac{1}{x} - \frac{1}{y}\right)\right| \\ &= \left|\frac{1}{y} - \frac{1}{x}\right| \\ &= d_1(y, x). \end{aligned}$$

4. Finally,

$$\begin{aligned} d_1(x,y) &= \left|\frac{1}{x}-\frac{1}{y}\right| \\ &= \left|\frac{1}{x}-\frac{1}{z}+\frac{1}{z}-\frac{1}{y}\right| \\ &\leq \left|\frac{1}{x}-\frac{1}{z}\right|+\left|\frac{1}{z}-\frac{1}{y}\right| \\ &= d_1(x,z)+d_1(z,y). \end{aligned}$$

Hence d_1 is a metric for $(0,\infty)$. □

Example 5.3.2 *Let $A_2 = \mathbb{R}^2$ with $\mathbf{x} = (x_1, y_1)$ and $\mathbf{y} = (x_2, y_2)$ so $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. Define $d_2(\mathbf{x}, \mathbf{y}) = |x_1 - x_2| + |y_1 - y_2|$. Show that d_2 is a valid metric.*

Solution. We prove the four properties:

1. $d_2(\mathbf{x}, \mathbf{x}) = |x_1 - x_1| + |y_1 - y_1| = 0$.
2. $d_2(\mathbf{x}, \mathbf{y}) = |x_1 - x_2| + |y_1 - y_2| > 0$ since $\mathbf{x} \neq \mathbf{y}$ implies that at least one of $x_1 \neq x_2$ or $y_1 \neq y_2$ must hold.
3. We also have $d_2(\mathbf{x}, \mathbf{y}) = |x_1 - x_2| + |y_1 - y_2| = |x_2 - x_1| + |y_2 - y_1| = d_2(\mathbf{y}, \mathbf{x})$.
4. Finally, let $\mathbf{z} = (x_3, y_3)$. Then

$$\begin{aligned} d_2(\mathbf{x},\mathbf{y}) &= |x_1 - x_2| + |y_1 - y_2| \\ &= |x_1 - x_3 + x_3 - x_2| + |y_1 - y_3 + y_3 - y_2| \\ &\leq |x_1 - x_3| + |x_3 - x_2| + |y_1 - y_3| + |y_3 - y_2| \\ &= (|x_1 - x_3| + |y_1 - y_3|) + (|x_3 - x_2| + |y_3 - y_2|) \\ &= d_2(\mathbf{x},\mathbf{z}) + d_2(\mathbf{z},\mathbf{y}). \end{aligned}$$

So d_2 is a metric. □

Remark 5.3.1 *Notice the importance of the absolute value property $|A - B| = |B - A|$ as well as the triangle inequality in the examples seen above (see Chapter 1 on absolute value).*

Example 5.3.3 *Revisit Example* 5.3.2 *but let $d_3(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - x_2|, |y_1 - y_2|\}$.*

Solution. Here we go again:

1. $d_3(\mathbf{x}, \mathbf{x}) = \max\{|x_1 - x_1|, |x_2 - x_2|\} = \max\{0, 0\} = 0$.

2. $d_3(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - x_2|, |y_1 - y_2|\} > 0$ since at least one of $|x_1 - x_2|$ or $|y_1 - y_2|$ is nonzero for $\mathbf{x} \neq \mathbf{y}$.

3. Next, we have

$$\begin{aligned} d_3(\mathbf{x}, \mathbf{y}) &= \max\{|x_1 - x_2|, |y_1 - y_2|\} \\ &= \max\{|x_2 - x_1|, |y_2 - y_1|\} \\ &= d_3(\mathbf{y}, \mathbf{x}). \end{aligned}$$

4. Lastly,

$$\begin{aligned} d_3(\mathbf{x}, \mathbf{y}) &= \max\{|x_1 - x_2|, |y_1 - y_2|\} \\ &= \max\{|x_1 - x_3 + x_3 - x_2|, |y_1 - y_3 + y_3 - y_2|\} \\ &\leq \max\{|x_1 - x_3| + |x_3 - x_2|, |y_1 - y_3| + |y_3 - y_2|\} \\ &\leq \max\{|x_1 - x_3|, |y_1 - y_3|\} + \max\{|x_3 - x_2|, |y_3 - y_2|\} \\ &= d_3(\mathbf{x}, \mathbf{z}) + d_3(\mathbf{z}, \mathbf{y}). \end{aligned}$$

Hence, d_3 is another valid metric. □

Remark 5.3.2 *Note the reasoning in showing the triangle inequality; you should study this carefully.*

Now that you have seen a few of these proofs, we omit proving the first three properties from this point forward.

Example 5.3.4 *Let $A_4 = \mathbb{R}^2$ with $\mathbf{x} = (x_1, y_1)$ and $\mathbf{y} = (x_2, y_2)$ and define $d_4(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Note that this is the standard way to define distance in $\mathbb{R}^2$ (i.e., the Pythagorean theorem). Prove the triangle inequality only.*

Solution. We need to show that

$$d_4(\mathbf{x}, \mathbf{y}) \leq d_4(\mathbf{x}, \mathbf{z}) + d_4(\mathbf{z}, \mathbf{y})$$

or

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \leq \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2} + \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}.$$

Notice how this resembles

$$\sqrt{\sum_{n=1}^{2} (a_n + b_n)^2} \leq \sqrt{\sum_{n=1}^{2} {a_n}^2} + \sqrt{\sum_{n=1}^{2} {b_n}^2}$$

which is Minkowski's inequality. If we let $a_1 = x_1 - x_3$, $a_2 = y_1 - y_3$, $b_1 = x_3 - x_2$, and $b_2 = y_3 - y_2$ then Minkowski's inequality

$$\sqrt{(a_1 + b_1)^2 + (a_2 + b_2)^2} \leq \sqrt{{a_1}^2 + {a_2}^2} + \sqrt{{b_1}^2 + {b_2}^2}$$

says

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \leq \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2} + \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2},$$

the desired result.

Remark 5.3.3 *You may wish to try a similar proof with d_4 defined as $d_4(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{n=1}^{N}(x_n - y_n)^2}$ where*

$$\mathbf{x} = (x_1, x_2, x_3, \ldots, x_N)$$

and

$$\mathbf{y} = (y_1, y_2, y_3, \ldots, y_N).$$

See the exercises.

In light of the previous examples, we now discuss the geometry involved. Looking back, we have $d_2(\mathbf{x}, \mathbf{y}) = |x_1 - x_2| + |y_1 - y_2|$, $d_3(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - x_2|, |y_1 - y_2|\}$, and $d_4(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ for $\mathbf{x} = (x_1, y_1)$ and $\mathbf{y} = (x_2, y_2)$ with $A = \mathbb{R}^2$. In the problem below, we compare these three metrics graphically.

Problem 5.3.1 *Let $S_2 = \{\mathbf{z} \in \mathbb{R}^2 \,|\, d_2(\mathbf{0}, \mathbf{z}) \leq 1\}$, $S_3 = \{\mathbf{z} \in \mathbb{R}^2 \,|\, d_3(\mathbf{0}, \mathbf{z}) \leq 1\}$, and $S_4 = \{\mathbf{z} \in \mathbb{R}^2 \,|\, d_4(\mathbf{0}, \mathbf{z}) \leq 1\}$. The symbol $\mathbf{0}$ indicates $(0, 0)$ (the origin) in $\mathbb{R}^2$. Sketch each of the sets S_2, S_3, and S_4 in $\mathbb{R}^2$.*

Solution. Since d_4 is the natural way to define distance in $\mathbb{R}^2$, we sketch S_4 first. Let $\mathbf{z} = (x, y)$. Then S_4 is the set of all points where

$$\begin{aligned} d_4(\mathbf{0}, \mathbf{z}) &= \sqrt{(0 - x)^2 + (0 - y)^2} \\ &= \sqrt{x^2 + y^2} \\ &\leq 1. \end{aligned}$$

That is, we have $x^2 + y^2 \leq 1$, the unit disc. Here is the picture:

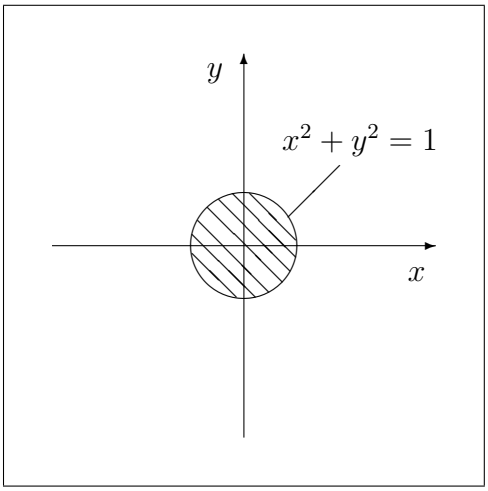

Figure 5.2: $S_4 = \{\mathbf{z} \in \mathbb{R}^2 \mid d_4(\mathbf{0}, \mathbf{z}) \leq 1\}$.

This should make sense since all of the points inside (and lying on) the unit circle $x^2 + y^2 = 1$ are within 1 unit of the origin $(0, 0)$. For S_2, we get all points satisfying $|0 - x| + |0 - y| \leq 1$ or $|x| + |y| \leq 1$. In quadrant 1 this becomes $x + y \leq 1$; in quadrant 2, $-x + y \leq 1$; in quadrant 3, $-x - y \leq 1$; in quadrant 4, $x - y \leq 1$. Geometrically, we have the following:

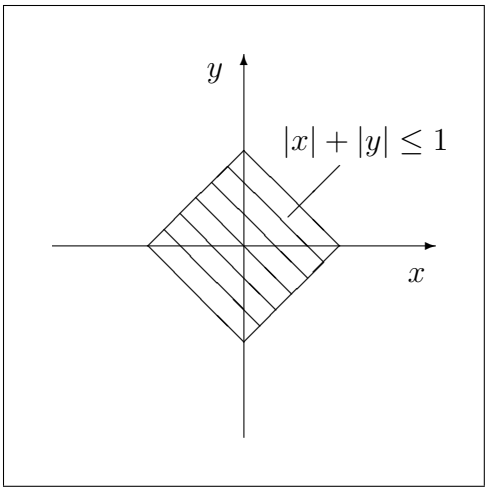

Figure 5.3: $S_2 = \{\mathbf{z} \in \mathbb{R}^2 \mid d_2(\mathbf{0}, \mathbf{z}) \leq 1\}$

Notice that for any point in the shaded region above, the inequality $|x| + |y| \leq 1$ is satisfied. Finally, S_3 will give us all points such that $\max\{|x|, |y|\} \leq 1$. In other words, x and y can be no larger than 1 and no smaller than -1. The picture is as follows:

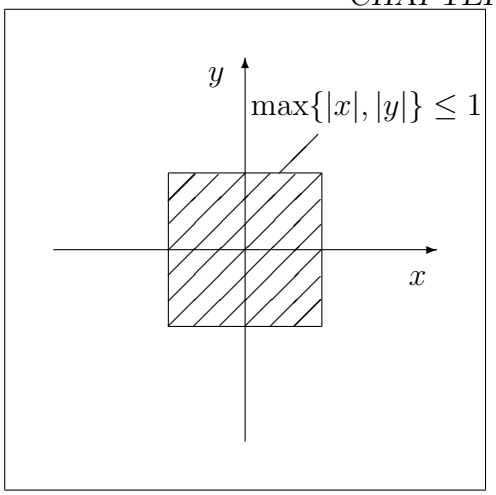

Figure 5.4: $S_3 = \{\mathbf{z} \in \mathbb{R}^2 \,|\, d_3(\mathbf{0}, \mathbf{z}) \leq 1\}$

We continue with some less conventional examples.

Example 5.3.5 *This example has some interesting features. Let d_5 be defined as*

$$d_5(x, y) = \begin{cases} 0, & x = y \\ 1, & x \neq y. \end{cases}$$

where $x, y \in \mathbb{R}$. Since it is straightforward to prove the first three properties, show the triangle inequality only. That is, show that $d_5(x, y) \leq d_5(x, z) + d_5(z, y)$.

Solution. We consider different cases.

1. Let $x \neq y$ and $x \neq z$. So we can immediately say that $y \neq z$. Thus $d_5(x, y) = 1$ and $d_5(x, z) + d_5(z, y) = 1 + 1 = 2$ so indeed $d_5(x, y) \leq d_5(x, z) + d_5(z, y)$.

2. Let $x \neq y$ but $x = z$. So $y \neq z$. Therefore, $d_5(x, y) = 1$ while $d_5(x, z) + d_5(z, y) = 0 + 1 = 1$ so $d_5(x, y) = d_5(x, z) + d_5(z, y)$.

3. Let $x \neq z$ but $y = z$. We leave the proof as an exercise.

4. Let $x = y$ but $y \neq z$. We leave the proof as an exercise.

5. Let $x = y = z$. Then we obtain $d_5(x, y) = d_5(x, z) + d_5(z, y)$ since all distances are zero.

As a result, d_5 is a metric from $\mathbb{R} \times \mathbb{R} \to [0, \infty)$. □

Remark 5.3.4 *$\langle \mathbb{R}, d_5 \rangle$ is often called the discrete metric. Note that for any two distinct points in $\mathbb{R}$, the "distance" is defined to be one. Otherwise, it is zero.*

Together, Examples 5.3.1 and 5.3.5 illustrate that it is possible for a set (in this case $\mathbb{R}$) to have more than one metric. Both d_1 and d_5 are valid metrics on $\mathbb{R}$.

Example 5.3.6 *Let A_6 = the set of all bounded sequences. In symbols, $A_6 = \{\{x_n\}_{n=1}^{\infty} \,|\, |x_n| \leq L,\, L \in \mathbb{R},\, n \in \mathbb{N}\}$. Notice that the "elements" in A_6 are actually sequences. For ease and readability, we will denote these elements by capital letters. For example, $X = \{x_n\}_{n=1}^{\infty}$, $Y = \{y_n\}_{n=1}^{\infty}$, and $Z = \{z_n\}_{n=1}^{\infty}$. Prove the triangle inequality with $d_6(X,Y) = \text{lub}_{n\in\mathbb{N}}|x_n - y_n|$.*

Solution. We will attempt to prove that $d_6(X,Y) \leq d_6(X,Z) + d_6(Z,Y)$ or

$$\text{lub}_{n\in\mathbb{N}}|x_n - y_n| \leq \text{lub}_{n\in\mathbb{N}}|x_n - z_n| + \text{lub}_{n\in\mathbb{N}}|z_n - y_n|.$$

We begin in the usual manner:

$$\begin{aligned} |x_n - y_n| &= |x_n - z_n + z_n - y_n| \\ &\leq |x_n - z_n| + |z_n - y_n| \\ &\leq \text{lub}_{n\in\mathbb{N}}|x_n - z_n| + \text{lub}_{n\in\mathbb{N}}|z_n - y_n|, \end{aligned}$$

the last statement being valid since the least upper bound of a set is at least as large as (often larger than) any element in the set. Hence we have

$$|x_n - y_n| \leq d_6(X,Z) + d_6(Z,Y)$$

for $n \in \mathbb{N}$. Now since $d_6(X,Z) + d_6(Z,Y)$ serves as an upper bound for $|x_n - y_n|$, the <u>least</u> upper bound of $|x_n - y_n|$ must be less than or equal to $d_6(X,Z) + d_6(Z,Y)$. That is,

$$\text{lub}_{n\in\mathbb{N}}|x_n - y_n| \leq d_6(X,Z) + d_6(Z,Y)$$

or

$$d_6(X,Y) \leq d_6(X,Z) + d_6(Z,Y)$$

so we are done here. □

Example 5.3.7 *We close this section with an example of a very different type. We will prove, given that d_1 and d_2 are metrics for A, $d_1 + d_2$ must also be a metric for A.*

Remark 5.3.5 *A few notes are in order here. First, we do not know A (unlike the previous six examples) nor do we know how d_1 or d_2 are defined. However, what we <u>do</u> know is that for $x, y \in A$, $(d_1 + d_2)(x,y) = d_1(x,y) + d_2(x,y)$ since $d_1, d_2 : A \times A \to [0,\infty)$. In words, d_1 and d_2 are real-valued functions. After this, the proof is quite simple.*

Solution.

1. We have

$$\begin{aligned}(d_1 + d_2)(x, x) &= d_1(x, x) + d_2(x, x) \\ &= 0 + 0 \ \ (d_1 \text{ and } d_2 \text{ are metrics}) \\ &= 0.\end{aligned}$$

2. Here,

$$\begin{aligned}(d_1 + d_2)(x, y) &= d_1(x, y) + d_2(x, y) \\ &\geq 0 + 0 \\ &= 0 \text{ for } x \neq y.\end{aligned}$$

3. Next,

$$\begin{aligned}(d_1 + d_2)(x, y) &= d_1(x, y) + d_2(x, y) \\ &= d_1(y, x) + d_2(y, x) \ \ (d_1 \text{ and } d_2 \text{ are metrics}) \\ &= (d_1 + d_2)(y, x).\end{aligned}$$

4. Finally,

$$\begin{aligned}(d_1 + d_2)(x, y) &= d_1(x, y) + d_2(x, y) \\ &\leq d_1(x, z) + d_1(z, y) + d_2(x, z) + d_2(z, y) \ \ (d_1 \text{ and } d_2 \text{ are metrics}) \\ &= d_1(x, z) + d_2(x, z) + d_1(z, y) + d_2(z, y) \\ &= (d_1 + d_2)(x, z) + (d_1 + d_2)(z, y).\end{aligned}$$

Thus, $d_1 + d_2$ is a metric for A. □

Exercises.

1. As a generalization to Example 5.3.4, show that $d_4(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{n=1}^{N}(x_n - y_n)^2}$ with $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_N)$ and $\mathbf{y} = (y_1, y_2, y_3, \ldots, y_N)$ defines a metric on $\mathbb{R}^N$.

2. Complete the proof seen in Example 5.3.5.

3. Let p and w be metrics for A. Show that $\max\{p, w\}$ is also a metric for A.

4. Let A be a metric space. Show that, for $w, x, y, z \in A$,

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

5. If d is a metric for A, show that

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

for any $x, y, z \in A$.

6. Let $x, y \in \mathbb{R}$. Show that the following define metrics on $\mathbb{R}$.

(a) $d_1(x,y) = \sqrt{|x-y|}$

(b) $d_2(x,y) = \frac{|x-y|}{|x-y|+1}$

(c) $d_3(x,y) = \min\{1, d_1(x,y)\}$

7. Show that if d defines a metric for A, so does $3d$.

8. Generalize Exercise 6(c). That is, show that if $\langle A, d\rangle$ is a metric space, then $\min\{1, d\}$ defines a metric for A.

9. Consider two metric spaces $\langle A, d_1\rangle$ and $\langle B, d_2\rangle$. Prove that $\langle A \times B, d\rangle$ is a metric space where d is defined by

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{[d_1(x_1, x_2)]^2 + [d_2(y_1, y_2)]^2}$$

for $\mathbf{x} = (x_1, y_1)$ and $\mathbf{y} = (x_2, y_2)$. Naturally, this metric space is called the Cartesian product metric space.

5.4 A Review of Limits

Since the ultimate goal is to discuss limits in metric spaces, we review some of the elementary material from Calculus−namely, the limit of a real-valued function at a point.

Definition 5.4.1 *Let f be a function defined on an open interval that contains c (though f need not be defined at c) and let $L \in \mathbb{R}$. We write*

$$\lim_{x \to c} f(x) = L$$

or

$$f(x) \to L \text{ as } x \to c$$

and say the limit of $f(x)$ as x approaches c is L if, for each number $\epsilon > 0$, there exists a corresponding number $\delta > 0$ such that if

$$0 < |x - c| < \delta$$

then

$$|f(x) - L| < \epsilon.$$

A few remarks are in order here.

Remark 5.4.1 *The symbols ϵ and δ represent small, positive numbers. The definition tells us that ϵ is a given quantity. Based on this assumption, we need to find a number $\delta = \delta(\epsilon)$ such that if $0 < |x - c| < \delta$ then $|f(x) - L| < \epsilon$. The notation $\delta = \delta(\epsilon)$ implies that δ is often a function of ϵ.*

Remark 5.4.2 *Since $|x - c|$ represents the distance from x to c and $|f(x) - L|$ represents the distance from $f(x)$ to L, we may intuitively say that $\lim_{x\to c} f(x) = L$ is really saying that if we take x very close to c (but not equal to c), we can make $f(x)$ as close to L as we'd like (in particular, within ϵ units of L.)*

Another important subtlety in the definition is the inequality $0 < |x - c| < \delta$. Since $0 < |x - c|$ we see that $x - c \neq 0$; i.e., $x \neq c$. In other words, the value $f(c)$, if it exists, is completely irrelevant to computing $\lim_{x\to c} f(x)$. Only the values of f for x close to c are important.

In an effort to slightly reformulate the definition, we revisit some of the basic inequalities involving absolute value. The inequality $|x - c| < \delta$ can be rewritten as $-\delta < x - c < \delta$ or $c - \delta < x < c + \delta$, keeping in mind that $x \neq c$. Similarly, $|f(x) - L| < \epsilon$ can be reformulated as $L - \epsilon < f(x) < L + \epsilon$. This tells us that the statement $\lim_{x\to c} f(x) = L$ is equivalent to the following: given $\epsilon > 0$, there exists a $\delta > 0$ such that if x lies in the open interval $(c - \delta, c + \delta)$, $x \neq c$, then $f(x)$ lies in the open interval $(L - \epsilon, L + \epsilon)$. See the figure.

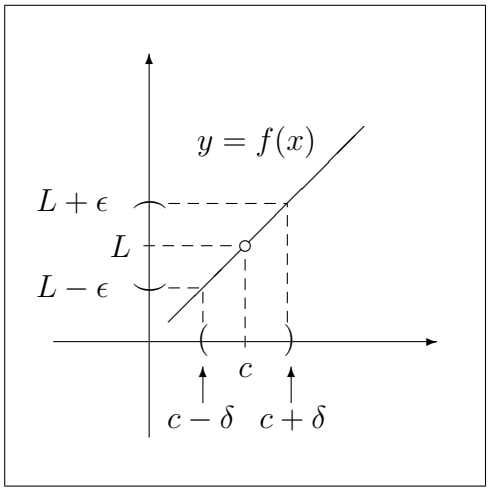

Figure 5.5: $\lim_{x\to c} f(x) = L$ seen graphically

Theorem 5.4.1 (Reformulated) *The limit of $f(x)$ as x approaches c is L if, for every $\epsilon > 0$, there is a $\delta > 0$ such that if $x \in (c - \delta, c + \delta)$, $x \neq c$, then $f(x) \in (L - \epsilon, L + \epsilon)$.*

Remark 5.4.3 *This definition will reappear in an even more general form in future analysis courses.*

You should study the following three visual examples (and match them with the definition) before moving on.

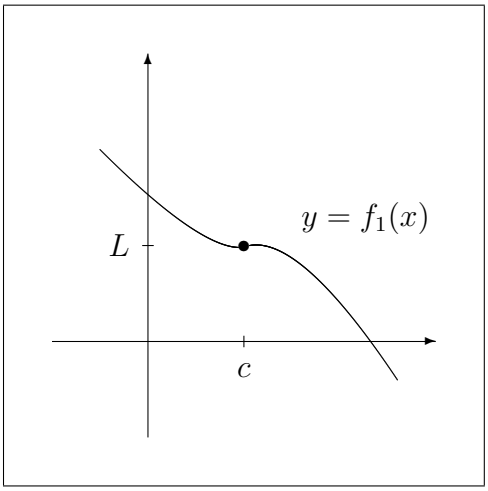

Figure 5.6: $\lim_{x \to c} f_1(x) = L$ seen graphically

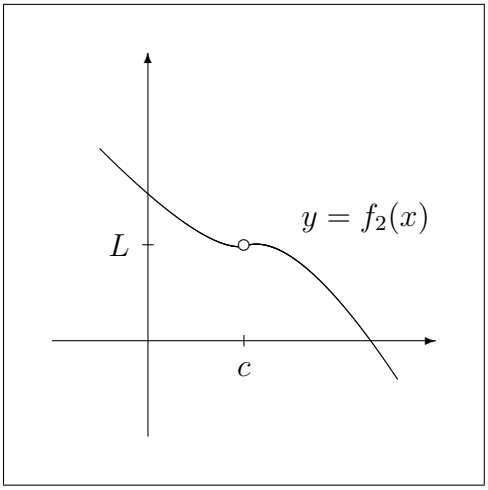

Figure 5.7: $\lim_{x \to c} f_2(x) = L$ seen graphically

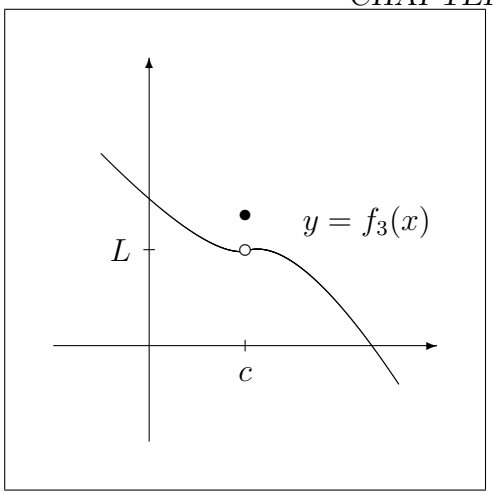

Figure 5.8: $\lim_{x\to c} f_3(x) = L$ seen graphically

Although it is clear that f_1, f_2 and f_3 are distinct functions, the validity of the statements $\lim_{x\to c} f_n(x) = L$ $(n = 1, 2, 3)$ is a direct consequence of the definition. We now present a few proofs using the ϵ-δ definition; some may be review from Calculus.

Problem 5.4.1 *Prove that* $\lim_{x\to 2}(3x - 2) = 4$.

We usually begin this process by searching for a suitable δ. Then, after this has been accomplished, we may proceed by writing a concise proof. We need to show that, for any $\epsilon > 0$, there is a δ such that if $0 < |x - 2| < \delta$ then $|(3x - 2) - 4| < \epsilon$. The latter inequality can be manipulated to read $|3(x - 2)| < \epsilon$ or $3|x - 2| < \epsilon$ or finally $|x - 2| < \frac{\epsilon}{3}$. Now based on the similarity to the inequality $|x - 2| < \delta$, the obvious choice here is to select $\delta = \frac{\epsilon}{3}$. Note the earlier comment that δ is nearly always a function of ϵ. Now for the proof.

Proof. Let $\delta = \frac{\epsilon}{3}$. Then if $|x - 2| < \delta$, we have

$$\begin{aligned} |f(x) - L| &= |(3x - 2) - 4| \\ &= 3|x - 2| \\ &< 3\delta \ \ (\text{since } |x - 2| < \delta) \\ &= 3\left(\frac{\epsilon}{3}\right) \left(\text{since } \delta = \frac{\epsilon}{3}\right) \\ &= \epsilon. \end{aligned}$$

Thus, $|f(x) - L| < \epsilon$ provided that $0 < |x - c| < \delta$. The proof is complete. ■

Remark 5.4.4 *Look at the picture that follows. Notice that we do not have to choose* $\delta = \frac{\epsilon}{3}$*; it is simply the most* <u>convenient</u> *choice to select. Any* <u>smaller</u> δ *(e.g.,* $\delta = \frac{\epsilon}{5}$*) will also work.*

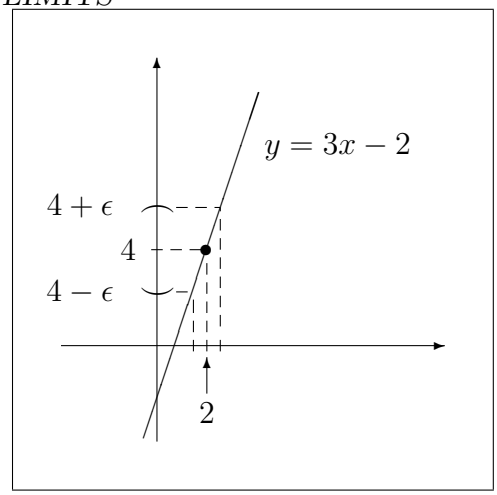

Figure 5.9: $\lim_{x\to 2}(3x-2) = 4$ seen graphically with $\delta = \frac{\epsilon}{3}$

Remark 5.4.5 *Notice that in the problem above* $\lim_{x\to c} f(x) = f(c)$. *This need not always be the case as mentioned earlier.*

Two additional notes are worthy of mention:

1. The key to writing these proofs is linking $|x-c|$ to $|f(x)-L|$. In the above example, we extracted $|x-2|$ from $|(3x-2)-4|$. Often, these extractions are more difficult to recognize.

2. It is probable that you will need to conduct some "scratch work" before writing the actual proof. It is likely that this work reveals the important link between the inequalities. We did just this in the preceding example.

The next few examples continue to emphasize these points.

Problem 5.4.2 *Prove that* $\lim_{x\to 4}(2x^2+x) = 36$.

We begin by collecting some notes to gather some important information. We need to show, for any $\epsilon > 0$, there exists a $\delta > 0$ such that if $0 < |x-4| < \delta$ then $|2x^2+x-36| < \epsilon$. First, we notice that

$$\begin{aligned} |2x^2+x-36| &= |(2x+9)(x-4)| \\ &= 2\left|x+\frac{9}{2}\right||x-4|. \end{aligned}$$

Now there is a slight problem. If we could say that $|2x^2+x-36| = C|x-4|$ where C is some constant, we would be nearly finished with a choice of $\delta = \frac{\epsilon}{C}$. However, the expression $\left|x+\frac{9}{2}\right|$ varies with x. Thus, we restrict δ to small values. For example,

$\delta \leq \frac{1}{2}$ is nice. Why such a choice? This should seem logical as we are only concerned with x values near 4. So with $\delta \leq \frac{1}{2}$ we have $|x-4| < \delta \leq \frac{1}{2}$ so $|x-4| < \frac{1}{2}$ or $\frac{7}{2} < x < \frac{9}{2}$. So with the restriction $x \in \left(\frac{7}{2}, \frac{9}{2}\right)$, we have

$$\left|x + \frac{9}{2}\right| < \frac{9}{2} + \frac{9}{2} = 9.$$

Hence,

$$\begin{aligned} |2x^2 + x - 36| &= 2\left|x + \frac{9}{2}\right| |x-4| \\ &< 2(9)|x-4| \quad \left(\text{for } |x-4| < \frac{1}{2}\right) \\ &= 18|x-4| \\ &< \epsilon \end{aligned}$$

provided $|x-4| < \frac{\epsilon}{18}$. Notice there are now two restrictions on $|x-4|$; namely $|x-4| < \frac{\epsilon}{18}$ and $|x-4| < \frac{1}{2}$. To ensure that both of these inequalities hold, we must choose $|x-4|$ to be the smaller of $\frac{\epsilon}{18}$ and $\frac{1}{2}$. Hence we choose $\delta = \min\left\{\frac{\epsilon}{18}, \frac{1}{2}\right\}$. See the figure.

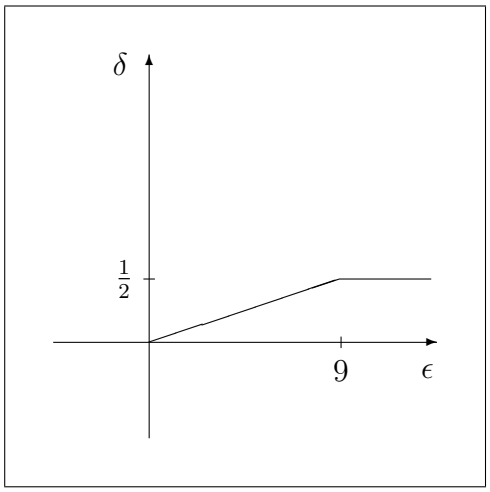

Figure 5.10: A sketch of the function $\delta(\epsilon) = \min\left\{\frac{\epsilon}{18}, \frac{1}{2}\right\}$.

Finally, here is the proof.

Proof. Given $\epsilon > 0$, let $\delta = \min\left\{\frac{\epsilon}{18}, \frac{1}{2}\right\}$. If $0 < |x-4| < \delta$ then certainly $|x-4| < \frac{1}{2}$ or $x \in \left(\frac{7}{2}, \frac{9}{2}\right)$. Hence, $\left|x + \frac{9}{2}\right| < 9$. Since $|x-4| < \frac{\epsilon}{18}$ as well, we have the following:

$$\begin{aligned} |f(x) - L| &= |2x^2 + x - 36| \\ &= 2\left|x + \frac{9}{2}\right| |x - 4| \\ &< 18|x - 4| \\ &< 18\left(\frac{\epsilon}{18}\right) \\ &= \epsilon. \end{aligned}$$

Thus, $\lim_{x\to 4}(2x^2 + x) = 36$. ■

Remark 5.4.6 *Note that δ is <u>not</u> determined when the statement $\delta \leq \frac{1}{2}$ is made. The proof is still written in complete generality as we are only imposing a <u>restriction</u> on δ. On the other hand, if we were to say, "let $\delta = 1$," this would not be valid. (Certainly, for any $\epsilon > 0$, the choice of $\delta = 1$ would not necessarily work in showing that $0 < |x - c| < \delta$ implies $|f(x) - L| < \epsilon$.)*

Remark 5.4.7 *Finally, always remember that $x \neq c$ (so $x \neq 4$ in the above example); otherwise, what happens in the proof?*

We give another example where the notion of a conjugate is important.

Problem 5.4.3 *Prove that* $\lim_{x\to 4}\sqrt{x} = 2$.

Here we set up

$$\begin{aligned} |\sqrt{x} - 2| &= |\sqrt{x} - 2|\left|\frac{\sqrt{x} + 2}{\sqrt{x} + 2}\right| \\ &= \frac{|x - 4|}{|\sqrt{x} + 2|}. \end{aligned}$$

Similar to the last example, we look at x "close" to 4; in particular, let $\delta \leq 1$ so that $|x - 4| < 1$ or $x \in (3, 5)$. Then

$$\frac{|x - 4|}{|\sqrt{x} + 2|} < \frac{|x - 4|}{\sqrt{3} + 2}.$$

Now we make the last quantity less than ϵ so that $|x-4| < \epsilon\left(\sqrt{3} + 2\right)$. What follows is the formal write-up.

Proof. Given $\epsilon > 0$, let $\delta = \min\left\{1, \epsilon\left(\sqrt{3}+2\right)\right\}$. Then for $|x-4| < 1$ we have $x \in (3,5)$ so that $|\sqrt{x}+2| > \sqrt{3}+2$. Thus, $\frac{1}{|\sqrt{x}+2|} < \frac{1}{\sqrt{3}+2}$ for $x \in (3,5)$. Also, $|x-4| < \epsilon\left(\sqrt{3}+2\right)$. So

$$\begin{aligned} |f(x)-L| &= |\sqrt{x}-2| \\ &= \frac{|x-4|}{|\sqrt{x}+2|} \\ &< \frac{|x-4|}{\sqrt{3}+2} \\ &< \frac{\epsilon\left(\sqrt{3}+2\right)}{\sqrt{3}+2} \\ &= \epsilon. \end{aligned}$$

So $\lim_{x\to 4}\sqrt{x} = 2$. ■

We believe it is equally important to show that a function may fail to have a limit as $x \to c$. In light of the definition, we now stress its negation.

Result 5.4.1 *To prove that* $\lim_{x\to c} f(x)$ *does* <u>not</u> *exist, it is sufficient to show that, for* <u>some</u> $\epsilon > 0$, *there is* <u>no</u> $\delta > 0$ *for which* $0 < |x-c| < \delta \Rightarrow |f(x)-L| < \epsilon$.

It is important to understand the negation above. The original definition demands that, for <u>every</u> positive ϵ, a delta must be found. Hence, to illustrate failure of the definition, we need only produce a <u>particular</u> ϵ for which there is no corresponding δ. Here is an example.

Problem 5.4.4 *Prove that* $\lim_{x\to 0}\sin\frac{1}{x}$ *does not exist.*

Proof. Intuitively, $\frac{1}{x} \to \pm\infty$ as $x \to 0$ so $\sin\frac{1}{x}$ oscillates rapidly as $x \to 0$. We proceed by contradiction. That is, suppose that $\sin\frac{1}{x} \to L$ as $x \to 0$ with $L \in \mathbb{R}$. Then, given $\epsilon > 0$, there exists a $\delta > 0$ such that if

$$0 < |x-0| < \delta \text{ then } \left|\sin\frac{1}{x} - L\right| < \epsilon.$$

We should analyze where the seams in this argument begin to unravel. First, note two important facts:

1. $\sin\left[\frac{\pi}{2} + 2n\pi\right] = 1$ for all $n \in \mathbb{W}$
2. $\sin\left[\frac{3\pi}{2} + 2n\pi\right] = -1$ for all $n \in \mathbb{W}$

Since we have $0 < |x| < \delta$, we have $x \in (-\delta, \delta)$, $x \neq 0$. Thus, we may look at $x \in (0, \delta)$ or $x \in (-\delta, 0)$. For simplicity sake, we take $x \in (0, \delta)$. Notice that we can find an $\hat{x} = \frac{1}{\frac{\pi}{2}+2n\pi} \in (0, \delta)$ (for n large enough) where $\sin \frac{1}{\hat{x}} = 1$. Similarly, we can find an $\tilde{x} = \frac{1}{\frac{3\pi}{2}+2n\pi} \in (0, \delta)$ (n may have to be quite large) with $\sin \frac{1}{\tilde{x}} = -1$. In other words, for $\delta > 0$, we can find $\hat{x}, \tilde{x} \in (0, \delta)$ with $\sin \frac{1}{\hat{x}} = 1$ and $\sin \frac{1}{\tilde{x}} = -1$ no matter how small δ may be. Given $\epsilon > 0$, the statement $\left|\sin \frac{1}{x} - L\right| < \epsilon$ becomes both

$$|1 - L| < \epsilon \text{ and } |-1 - L| < \epsilon.$$

To see the problem here, let $\epsilon = \frac{1}{2}$. Then $|1 - L| < \frac{1}{2} \Rightarrow |L - 1| < \frac{1}{2} \Rightarrow \frac{1}{2} < L < \frac{3}{2}$. Likewise, $|-1 - L| < \frac{1}{2} \Rightarrow |L + 1| < \frac{1}{2} \Rightarrow -\frac{3}{2} < L < -\frac{1}{2}$. Having $L \in \left(\frac{1}{2}, \frac{3}{2}\right)$ and $L \in \left(-\frac{3}{2}, -\frac{1}{2}\right)$ is impossible so no δ works for this particular $\epsilon = \frac{1}{2}$. This contradiction proves that $\lim_{x\to 0} \sin \frac{1}{x}$ does not exist. ■

Finally, we close this section with a theorem that states some basic limit laws.

Theorem 5.4.2 *Let* $\lim_{x\to c} f(x) = L$ *and* $\lim_{x\to c} g(x) = M$. *Then*

1. $\lim_{x\to c} [f(x) + g(x)] = L + M$
2. $\lim_{x\to c} [f(x) - g(x)] = L - M$
3. $\lim_{x\to c} [f(x)g(x)] = LM$
4. $\lim_{x\to c} \left[\frac{f(x)}{g(x)}\right] = \frac{L}{M},\ M \neq 0$

The proofs of 1 and 2 are rather straightforward but 3 and 4 require a bit more sophistication and ingenuity.

Proof. We prove parts 1 and 4 of the theorem. The others are left as exercises.

- Part 1 Since $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$ then for $\epsilon > 0$, we know there are numbers δ_1 and δ_2 such that

  1. if $0 < |x - c| < \delta_1$, then $|f(x) - L| < \frac{\epsilon}{2}$ and
  2. if $0 < |x - c| < \delta_2$, then $|g(x) - M| < \frac{\epsilon}{2}$.

  (Note that $\delta_1 \neq \delta_2$ is most likely the case and $\frac{\epsilon}{2}$ is just a cosmetic feature−we mentioned this earlier in the text.) So if we let $\delta = \min\{\delta_1, \delta_2\}$, then having both $0 < |x - c| < \delta_1$ and $0 < |x - c| < \delta_2$ translates to $0 < |x - c| < \delta$. So given $0 < |x - c| < \delta$, we have

$$\begin{aligned} |[f(x) + g(x)] - (L + M)| &= |f(x) - L + g(x) - M| \\ &\leq |f(x) - L| + |g(x) - M| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

so the proof is complete. ■

- Part 4 For $\tilde{\epsilon} > 0$, there exist numbers δ_1 and δ_2 such that

  1. if $0 < |x - c| < \delta_1$, then $|f(x) - L| < \frac{\tilde{\epsilon}}{2}$ and
  2. if $0 < |x - c| < \delta_2$, then $|g(x) - M| < \frac{\tilde{\epsilon}}{2}$.

  We need to show that $\left|\frac{f(x)}{g(x)} - \frac{L}{M}\right|$ can be made sufficiently small for x sufficiently close to c. Well,

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - \frac{L}{M}\right| &= \left|\frac{Mf(x) - Lg(x)}{Mg(x)}\right| \\ &= \left|\frac{Mf(x) - ML + ML - Lg(x)}{Mg(x)}\right| \\ &= \left|\frac{M[f(x) - L] + L[M - g(x)]}{Mg(x)}\right| \\ &\leq \frac{|M||f(x) - L| + |L||M - g(x)|}{|M||g(x)|}.\end{aligned}$$

  Now we pause for a moment: $|M|$ and $|L|$ are just constants while $|f(x) - L|$ and $|M - g(x)|$ can be made sufficiently small for x close to c. The question remains, *can we do a similar thing with* $|g(x)|$? Certainly, we know that for $0 < |x - c| < \delta_3$, we can make $|g(x) - M| < \hat{\epsilon} = \frac{|M|}{2}$. In other words, for x within δ_3 units of c, we can make $g(x)$ within $\frac{|M|}{2}$ units of M. Here, we are simply "choosing" a value for $\hat{\epsilon}$, namely $\hat{\epsilon} = \frac{|M|}{2}$. Next we observe that if $|g(x) - M| < \frac{|M|}{2}$, then $-|g(x) - M| > -\frac{|M|}{2}$. Holding onto this thought, we have

$$|M| - |g(x)| \leq ||M| - |g(x)|| \leq |M - g(x)| = |g(x) - M|$$

  by the reverse Triangle Inequality. The upshot is $|M| - |g(x)| \leq |g(x) - M|$ or

$$\begin{aligned}|g(x)| &\geq |M| - |g(x) - M| \\ &> |M| - \frac{|M|}{2} \\ &= \frac{|M|}{2}.\end{aligned}$$

Thus, $\frac{1}{|g(x)|} < \frac{2}{|M|}$. Finally, for $\delta = \min\{\delta_1, \delta_2, \delta_3\}$, $0 < |x - c| < \delta$ implies that

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - \frac{L}{M}\right| &\leq \frac{|M||f(x) - L| + |L||M - g(x)|}{|M||g(x)|} \\ &< \frac{|M|\tilde{\epsilon} + |L|\tilde{\epsilon}}{|M|}\frac{1}{|g(x)|} \\ &< \tilde{\epsilon}\left(\frac{|M| + |L|}{|M|}\right)\frac{2}{|M|} \\ &= \left(\frac{2\left(|M| + |L|\right)}{|M|^2}\right)\tilde{\epsilon} \\ &= \epsilon,\end{aligned}$$

completing the proof. ■

Remark 5.4.8 *Here we denote ϵ as $C\tilde{\epsilon}$ where $C = \frac{2(|M|+|L|)}{|M|^2}$. Note that ϵ is arbitrary and positive since $\tilde{\epsilon}$ is as well.*

We close by simply emphasizing the multitude of tools used in the previous two proofs: various properties of absolute value, the min function, the Triangle Inequality, addition and substraction of the same quantity, the reverse Triangle Inequality, and the list continues ...

Exercises.

1. Prove the following statements.
   (a) $\lim_{x\to 1}(x - 4) = -3$
   (b) $\lim_{x\to -1}(2x - 3) = -5$
   (c) $\lim_{x\to 0} x = 0$
2. It was mentioned earlier in the limit definition that $\delta = \delta(\epsilon)$, or that δ is frequently a function of ϵ. Give an example of a function f where $\lim_{x\to c} f(x) = L$ but δ is independent of the ϵ being used.
3. Prove the following statements.
   (a) $\lim_{x\to 16}\sqrt{x} = 4$
   (b) $\lim_{x\to 2}(2x^2 - 3x) = 2$
   (c) $\lim_{x\to 0}\frac{x+2}{x+1} = 2$
   (d) $\lim_{x\to -1}\frac{1}{x+2} = 1$

4. Prove that the following limits do not exist.

   (a) $\lim_{x\to 0} \frac{1}{x^2}$

   (b) $\lim_{x\to 0} \frac{1}{x}$

   (c) $\lim_{x\to 3} \frac{1}{(x-3)^2}$

5. Finish the proof of **Theorem 5.4.2**, parts 2 and 3.

6. Prove that if $\lim_{x\to c} f(x)$ exists, then the limit is unique. (Hint: See page 21, specifically, Problem 1.4.4.)

7. Prove that $\lim_{x\to c} f(x) = L$ if and only if $\lim_{x\to c} g(x) = 0$ where $g(x) = f(x) - L$.

5.5 Limits in Metric Spaces

In this section, we discuss the topics from the last section but in a generalized setting. Previously, attention was devoted (or restricted−depending on how you look at it) to limits in $\mathbb{R}$ using the absolute value metric. Here we consider arbitrary metric spaces $\langle A, d\rangle$. At this point, we ask you to reexamine Definition 5.4.1. Then you can see how the following definition arises.

Definition 5.5.1 *Let $\langle A_1, d_1\rangle$ and $\langle A_2, d_2\rangle$ be metric spaces. Furthermore, suppose that $f : A_1 \to A_2$ with $c \in A_1$ and $\mathcal{R}(f) \subset A_2$. We write $\lim_{x\to c} f(x) = L$ with $L \in A_2$ if, given $\epsilon > 0$, there is a $\delta > 0$ such that if $0 < d_1(x, c) < \delta$ then $d_2(f(x), L) < \epsilon$.*

Remark 5.5.1 *Note the similarities to Definition 5.4.1. To see that this new definition generalizes the old, put $A_1 = A_2 = \mathbb{R}$ and $d_1(x, y) = d_2(x, y) = |x - y|$ so that $0 < d_1(x, c) < \delta$ becomes $0 < |x - c| < \delta$ and $d_2(f(x), L) < \epsilon$ becomes $|f(x) - L| < \epsilon$.*

First we look at an example.

Example 5.5.1 *Let $f(x, y) = 2x + y^2 - 2$. Prove that $\lim_{(x,y)\to(3,2)} f(x, y) = 8$.*

Proof. Note that $f : \mathbb{R}^2 \to \mathbb{R}$. Thus, the "distance" functions d_1 and d_2 are defined differently. In $\mathbb{R}$ we have the usual $d(x, y) = |x - y|$ while in $\mathbb{R}^2$ we use $d(\langle x_1, y_1\rangle, \langle x_2, y_2\rangle) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, the familiar distance formula. We must show, given $\epsilon > 0$, there exists a $\delta > 0$ such that if $\sqrt{(x - 3)^2 + (y - 2)^2} < \delta$

then $|f(x,y)-8|<\epsilon$. First note that

$$\begin{aligned}|f(x,y)-8| &= |2x+y^2-2-8|\\ &= |2x+y^2-10|\\ &= |2x-6+y^2-4|\\ &= |2(x-3)+(y+2)(y-2)|\\ &\leq 2|x-3|+|y+2||y-2|.\end{aligned}$$

Notice that -10 was split up as -6 and -4 so that the factors $|x-3|$ and $|y-2|$ would appear. This is a good strategy since $(x,y)\to(3,2)$. Now since we want y close to 2, we can choose $y\in(1,3)$. That is, impose $|y-2|<1$ so that $|y+2|\leq 3+2=5$. (Notice what we have just done: let $\delta\leq 1$. See the figure.)

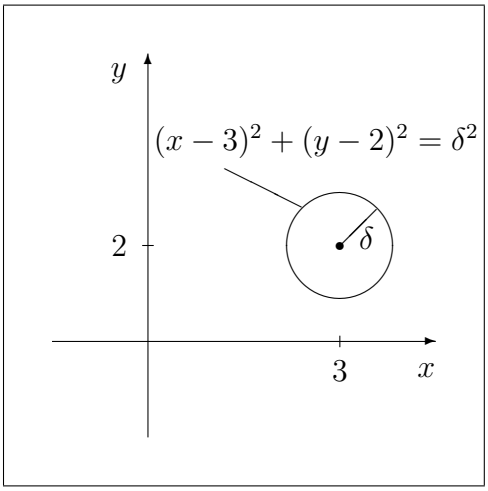

Figure 5.11: Restricting δ so that $\sqrt{(x-3)^2+(y-2)^2}<\delta\leq 1$

Thus,

$$\begin{aligned}2|x-3|+|y+2||y-2| &< 5|x-3|+5|y-2|\\ &= 5\left(|x-3|+|y-2|\right)\end{aligned}$$

for $y\in(1,3)$. Now since we need to link $|x-3|+|y-2|$ with $\sqrt{(x-3)^2+(y-2)^2}$ we use the Schwarz inequality $\sum_{n=1}^{N}|a_nb_n|\leq\sqrt{\sum_{n=1}^{N}{a_n}^2}\sqrt{\sum_{n=1}^{N}{b_n}^2}$. Putting $N=2$ and $b_n=1$, we have $\sum_{n=1}^{2}|a_n|\leq\sqrt{2}\sqrt{{a_1}^2+{a_2}^2}$ or

$$|a_1|+|a_2|\leq\sqrt{2}\sqrt{{a_1}^2+{a_2}^2}.$$

Thus,

$$\begin{aligned} |f(x,y) - 8| &< 5\left(|x-3| + |y-2|\right) \\ &\leq 5\sqrt{2}\sqrt{(x-3)^2 + (y-2)^2} \\ &< \epsilon \end{aligned}$$

where $\delta = \min\left\{1, \frac{\epsilon}{5\sqrt{2}}\right\}$. ■

Because of similarities to **Theorem 5.4.2**, we now state the analogous result for $c \in A$, where $\langle A, d\rangle$ is an arbitrary metric space.

Theorem 5.5.1 *Let $\langle A, d\rangle$ be a metric space and let $c \in A$. Furthermore, let $f, g : A \to \mathbb{R}$ so that f and g are real-valued functions. If $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$ then*

1. $\lim_{x\to c}[f(x) + g(x)] = L + M$
2. $\lim_{x\to c}[f(x) - g(x)] = L - M$
3. $\lim_{x\to c}[f(x)g(x)] = LM$
4. $\lim_{x\to c}\left[\frac{f(x)}{g(x)}\right] = \frac{L}{M},\ M \neq 0$

Remark 5.5.2 *It is important to note the utility of two metrics in the theorem−the absolute value metric as well as an arbitrary metric d.*

Proof. We prove only the first property since the proofs are similar to those seen in **Theorem 5.4.2**. We need to show that given $\epsilon > 0$, there exists a $\delta > 0$ such that if $0 < d(x,c) < \delta$ then $|[f(x) + g(x)] - (L + M)| < \epsilon$. Since we may assume $0 < d(x,c) < \delta_1 \Rightarrow |f(x) - L| < \frac{\epsilon}{2}$ and $0 < d(x,c) < \delta_2 \Rightarrow |g(x) - M| < \frac{\epsilon}{2}$, we take $\delta = \min\{\delta_1, \delta_2\}$ so that $0 < d(x,c) < \delta$ gives

$$\begin{aligned} |[f(x) + g(x)] - (L + M)| &\leq |f(x) - L| + |g(x) - M| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

Thus, $\lim_{x\to c}[f(x) + g(x)] = L + M$ as desired. ■

Compare this proof with the proof of **Theorem 5.4.2**, part 1.

Problem 5.5.1 *Let $x_n \to x$ and $y_n \to y$ be convergent sequences in the metric space $\langle A, d\rangle$. Show that $d(x_n, y_n) \to d(x, y)$ in $\mathbb{R}$.*

Proof. We have $d(x_n, x) < \frac{\epsilon}{2}$ for $n > N_1$ and $d(y_n, y) < \frac{\epsilon}{2}$ for $n > N_2$. We need to show that for some $n > N$, $|d(x_n, y_n) - d(x, y)| < \epsilon$. Notice here that we are using the $\mathbb{R}^1$ metric. We begin with

$$\begin{aligned} d(x_n, y_n) &\leq d(x_n, x) + d(x, y_n) \\ &\leq d(x_n, x) + d(x, y) + d(y, y_n) \end{aligned}$$

so

$$d(x_n, y_n) - d(x, y) \leq d(x_n, x) + d(y, y_n). \tag{5.1}$$

Notice the application of the Triangle Inequality for metric spaces (twice) in the previous step. Similarly,

$$\begin{aligned} d(x, y) &\leq d(x, x_n) + d(x_n, y) \\ &\leq d(x, x_n) + d(x_n, y_n) + d(y_n, y) \end{aligned}$$

so

$$d(x, y) - d(x_n, y_n) \leq d(x, x_n) + d(y_n, y).$$

Thus

$$d(x_n, y_n) - d(x, y) \geq -d(x, x_n) - d(y_n, y) \tag{5.2}$$

upon multiplication by -1. Putting equations 5.1 and 5.2 together, we have

$$-d(x, x_n) - d(y_n, y) \leq d(x_n, y_n) - d(x, y) \leq d(x_n, x) + d(y, y_n),$$

or, more compactly,

$$|d(x_n, y_n) - d(x, y)| \leq d(x_n, x) + d(y, y_n).$$

Hence, for $n > N = \max\{N_1, N_2\}$ we have $d(x_n, x) < \frac{\epsilon}{2}$ and $d(y_n, y) < \frac{\epsilon}{2}$ so that

$$|d(x_n, y_n) - d(x, y)| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This proves that $\{d(x_n, y_n)\}_{n=1}^{\infty}$ converges to $d(x, y)$ in $\mathbb{R}^1$. ■

Exercises.

1. Complete the proof of **Theorem 5.5.1**.
2. Consider two sequences $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$. Denote $J_n = \langle x_n, y_n \rangle \in \mathbb{R}^2$ for each $n \in \mathbb{N}$. Prove that $J_n \to \langle x, y \rangle \in \mathbb{R}^2$.
3. Prove that $\lim_{(x,y)\to(5,-2)}(x^2 - 7y + 8) = 47$.
4. Prove that $\lim_{(x,y,z)\to(1,2,0)}[x^3 y - z\sin(xy)] = 2$.

5.6 Summary: Odds and Ends

- A metric space $\langle A, d\rangle$ is a set A equipped with a distance function $d : A \times A \to [0, \infty)$ obeying the following properties:

  1. $d(x, x) = 0$
  2. $d(x, y) > 0$ for $x \neq y$.
  3. $d(x, y) = d(y, x)$.
  4. $d(x, y) \leq d(x, z) + d(z, y)$.

- Two very useful inequalities:

  1. $\left|\sum_{n=1}^{N} a_n b_n\right| \leq \sqrt{\sum_{n=1}^{N} {a_n}^2}\sqrt{\sum_{n=1}^{N} {b_n}^2}$ (**Schwarz Inequality**)
  2. $\sqrt{\sum_{n=1}^{N}(a_n + b_n)^2} \leq \sqrt{\sum_{n=1}^{N} {a_n}^2} + \sqrt{\sum_{n=1}^{N} {b_n}^2}$ (**Minkowski Inequality**)

- There are many examples of valid metrics on $\mathbb{R}$, $\mathbb{R}^2$, etc. Some examples are given below.

  1. For $\mathbb{R}$:
     (a) $d(x, y) = |x - y|$. Notice that this is the standard way to define distance in $\mathbb{R}$.
     (b) $d(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right|$
     (c) $d(x, y) = \begin{cases} 0, & x = y \\ 1, & x \neq y. \end{cases}$ This is known as the discrete metric.
  2. For $\mathbb{R}^2$:
     (a) $d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. This is the standard way to define distance in $\mathbb{R}^2$.
     (b) $d(\mathbf{x}, \mathbf{y}) = |x_1 - x_2| + |y_1 - y_2|$
     (c) $d(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - x_2|, |y_1 - y_2|\}$

- The statement $\lim_{x\to c} f(x) = L$ or $f(x) \to L$ $(x \to c)$ means, for each number $\epsilon > 0$, there exists a corresponding number $\delta > 0$ such that if $0 < |x - c| < \delta$ then $|f(x) - L| < \epsilon$. For general metric spaces $\langle A_1, d_1\rangle$ and $\langle A_2, d_2\rangle$ with $f : A_1 \to A_2$, we write $\lim_{x\to c} f(x) = L$ if, given $\epsilon > 0$, there is a $\delta > 0$ such that if $0 < d_1(x, c) < \delta$ then $d_2(f(x), L) < \epsilon$.

- Let $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$. Then

  1. $\lim_{x\to c} [f(x) + g(x)] = L + M$

2. $\lim_{x\to c}[f(x)-g(x)] = L - M$
3. $\lim_{x\to c}[f(x)g(x)] = LM$
4. $\lim_{x\to c}\left[\frac{f(x)}{g(x)}\right] = \frac{L}{M},\ M \neq 0$

This theorem is also valid in the case of a general metric space $\langle A, d\rangle$ for which $f, g : A \to \mathbb{R}$.

Chapter 6

Loose Ends

The purpose of this brief and final chapter is two-fold. First, I hope to whet the reader's appetite for more advanced mathematical topics, thus encouraging further study in these areas. Second, I hope to offer two ecclectic yet interesting proofs whose specific details borrow heavily from previous chapters.

6.1 The Irrationality of e

Depending on the purpose of study, the ubiquitous mathematical constant e can be defined in a number of ways. Three of the most popular are through (a) a limit tending to infinity, (b) an infinite series, and (c) a definite integral. We list the three representations below and follow with some comments.

Definition 6.1.1 *Depending on the theoretical path one hopes to explore, any of the following can be considered a starting point for defining the number e:*

1. $e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$
2. $e = \sum_{n=0}^{\infty} \frac{1}{n!}$
3. *e is the number for which* $\int_1^e \frac{1}{x} dx = 1$.

Each of the above has its advantages. The limit definition permits one to construct a table and to be convinced that, for large n, $\left(1 + \frac{1}{n}\right)^n \approx e$. See Table 6.1 for an illustration (values are rounded).

Although this approach is straightforward, it takes a very large n to convince the skeptic that the limit is e. Fortunately, the second definition addresses this problem. If taking the infinite series as the definition of e, one can add terms to the partial sum expression $s_n = \sum_{k=0}^{n} \frac{1}{k!}$ and likewise be swayed. Recall that $n!$, read "n factorial,"

Table 6.1: Value of $\left(1+\frac{1}{n}\right)^n$ for large n

n	$\left(1+\frac{1}{n}\right)^n$
5	2.48832
10	2.593742460
50	2.691588029
100	2.704813829
$1,000$	2.716923932
$10,000$	2.718145927
$\vdots$	$\vdots$

Table 6.2: Value of $\sum_{k=0}^{n} \frac{1}{k!}$ for various n

n	$1+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}$
3	2.666666667
5	2.716666667
8	2.718278770
$\vdots$	$\vdots$

is defined by $n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$ with $0! = 1$. See Table 6.2 for an illustration.

The noticeable difference here is that one need not add many terms before being convinced that the sum approaches e. This rapidity of convergence is a benchmark that we exploit in the paragraphs that follow. Additionally, the reader familiar with calculus will recall that

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Knowing that this converges for all real x, we can put $x = 1$ to get the definition above.

Finally, the integral definition−similar to each of the above definitions−allows for a geometric argument by way of computing the area under the curve $y = 1/x$. We don't dwell on this any further.

The mere fact that such radically different views of e are equivalent might suggest that this constant is an innocuous entity, perhaps even rational. We will utilize the second definition (the infinite series) to establish its irrationality. That is, e cannot be expressed as the ratio of two integers $\frac{m}{n}$ no matter how hard one tries. Many

people find this fact most intriguing since an infinite sum of rather ordinary-looking *rational* numbers gives rise to this wildy popular *irrational* number.

Table 6.3: Rational Approximations for e

Rational Number	Decimal Approximation
65/24	2.708333333
109601/40320	2.718278770
260412269/95800320	2.718281828
$\vdots$	$\vdots$

No matter how far we continue the process suggested by Table 6.3 (i.e., $\frac{65}{24} = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}$), we will never acquire the precise decimal representation for the numerical constant e.

Onto our main business in this section. To begin our quest, we first take

$$\begin{aligned} e &= \sum_{n=0}^{\infty} \frac{1}{n!} \\ &= 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots . \end{aligned}$$

We should convince ourselves that this sum does not grow without bound. Letting

$$\begin{aligned} s_n &= 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} \\ &= \sum_{k=0}^{n} \frac{1}{k!}, \end{aligned}$$

we can first view $\{s_n\}_{n=1}^{\infty}$ as a sequence. Then

$$\begin{aligned} s_n &= 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} \cdots + \frac{1}{n!} \\ &< 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^{n-1}} \\ &< 1 + \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k \\ &= 1 + \frac{1}{1 - \frac{1}{2}} \\ &= 3, \end{aligned}$$

so $\{s_n\}_{n=1}^{\infty}$ is bounded above by 3. It is also fairly clear that $\{s_n\}_{n=1}^{\infty}$ is monotonic (increasing, in fact) so one can establish the convergence of $\{s_n\}_{n=1}^{\infty}$ quickly. The limit turns out to be e.

The next result tells us that, for increasingly large n, the difference between the values of e and s_n is quite small. In other words, $\sum_{n=0}^{\infty} \frac{1}{n!}$ converges rapidly to the number e.

Result 6.1.1 $e - s_n < \frac{1}{n!n}$.

Proof. We examine the difference $e - s_n$ so

$$\begin{aligned}
e - s_n &= e - \sum_{k=0}^{n} \frac{1}{k!} \\
&= \sum_{k=n+1}^{\infty} \frac{1}{k!} \\
&= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \frac{1}{(n+3)!} + \cdots \\
&= \frac{1}{(n+1)!}\left[1 + \frac{1}{n+2} + \frac{1}{(n+3)(n+2)} + \frac{1}{(n+4)(n+3)(n+2)} \cdots\right] \\
&< \frac{1}{(n+1)!}\left[1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \frac{1}{(n+1)^3} + \cdots\right] \\
&= \frac{1}{(n+1)!} \sum_{m=0}^{\infty} \left(\frac{1}{n+1}\right)^m \\
&= \frac{1}{(n+1)!} \frac{1}{1 - \frac{1}{n+1}} \\
&= \frac{1}{(n+1)!} \frac{n+1}{n} \\
&= \frac{1}{n!n},
\end{aligned}$$

so we may conclude that $e - s_n < \frac{1}{n!n}$. ■

To illustrate how such a statement can be of use, view the example.

Example 6.1.1 *Applying* $e - s_n < \frac{1}{n!n}$ *with* $n = 4$, *we get* $e - s_4 < \frac{1}{4!4}$ *or*

$$e - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}\right) < \frac{1}{96}.$$

In a nutshell, this tells us that by simply adding the first five terms in the expression $\sum_{n=0}^{\infty} \frac{1}{n!}$, *we are already within* $\frac{1}{96}^{\text{th}}$ *of the "true" value of e. Note that*

$$1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} = 2.70833333\ldots.$$

We next use this result to reach our goal.

Theorem 6.1.1 *The number e is irrational.*

Proof. We assume the contrary. That is, let $e = \frac{m}{n}$ where $m, n \in \mathbb{N}$. Then by rearranging and bounding Result 6.1.1 above, we get

$$0 < n!(e - s_n) < \frac{1}{n} < 1 \tag{6.1}$$

Now since e is assumed to be rational, the number

$$n!e = n!\frac{m}{n} = (n-1)!m$$

is an integer. One can also show that $n!s_n$ is an integer since

$$n!s_n = n!\left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}\right).$$

Thus, $n!e - n!s_n$ is an integer. However, statement (6.1) indicates that this integer is between 0 and 1. This contradiction proves the theorem. ■

6.2 Adding Prime Reciprocals

This section addresses a question that appears simple on the surface. Consider the infinite set of prime numbers $p_1 = 2, p_2 = 3, p_3 = 5, p_4 = 7, \ldots$ Does the series of reciprocals $\sum_{n=1}^{\infty} \frac{1}{p_n}$ converge? Here we will take a journey into the inner workings of

$$\sum_{n=1}^{\infty} \frac{1}{p_n} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \cdots$$

and decide. To start, it might be a good idea to take an informal look into other infinite series that are qualitatively similar. For example, if we consider $\mathbb{Z}^+$ and add their reciprocals we have

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

which the reader will recognize as the harmonic series. We'll show that this diverges.

Proof. Let us consider the first n terms of the series $\sum_{n=1}^{\infty} \frac{1}{n}$ so we'll work with the partial sum

$$s_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

One argument goes something like this:

$$\begin{aligned} s_n &= 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \\ &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \cdots + \frac{1}{16}\right) + \cdots + \frac{1}{n} \\ &> 1 + \frac{1}{2} + \frac{2}{4} + \frac{4}{8} + \frac{8}{16} + \cdots + \frac{1}{n} \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2}. \end{aligned}$$

It seems clear from the above argument that we can have as many $\frac{1}{2}$ terms as we wish; just advance far enough into the partial sum. Thus, s_n is not bounded above so $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges. ■

A different exercise might be to collect a very sparse collection of $\mathbb{Z}^+$, say something like $\{2, 4, 8, 16, \ldots\}$ and add their reciprocals. This leads to

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

which can be shown to converge. We demonstrate this presently.

Proof. As in the case of the harmonic series, we work with the partial sum

$$s_n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n}.$$

It is helpful to write down the first few partial sums:

$$\begin{aligned} s_1 &= \frac{1}{2} \\ s_2 &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \\ s_3 &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8} \\ &\vdots \\ s_n &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n} = \frac{2^n - 1}{2^n} \end{aligned}$$

So with $s_n = \frac{2^n - 1}{2^n} = 1 - \frac{1}{2^n}$ we find

$$\lim_{n \to \infty} s_n = \lim_{n \to \infty} \left(1 - \frac{1}{2^n}\right) = 1,$$

so $\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n$ converges to 1. ■

Now the reason for discussing these two series is clear since $\sum_{n=1}^{\infty} \frac{1}{p_n}$ seems to fall somewhere between these two extremes. That is, the harmonic series gathers all of the reciprocals (and diverges) whereas the geometric series takes only a sparse collection (so sparse that it converges). The reader will notice that the prime numbers make themselves very present initially but seem to drop off in proportion as one ventures deeper into $\mathbb{Z}^+$. For example, see the prime numbers underlined below:

$$1, \underline{2}, \underline{3}, 4, \underline{5}, 6, \underline{7}, 8, 9, 10, \underline{11}, 12, \underline{13}, 14, 15, 16, \underline{17}, 18, \underline{19}, 20, 21, 22, \underline{23}, 24, 25, 26, 27, \ldots$$

So then our question seems very appropriate: if we add the prime reciprocals, will the series converge? (We know from earlier work that there are infinitely many prime numbers for otherwise the answer to our present question would be obvious.) Before embarking on this mission, we will briefly discuss a collection of inequalities that will help us in writing a proof on the fate of $\sum_{n=1}^{\infty} \frac{1}{p_n}$.

Without getting into too much detail, let's consider two numbers $m, n \in \mathbb{Z}^+$. It is fairly clear that

$$\frac{1}{mn} < \left(\frac{1}{m} + \frac{1}{n}\right)^2$$

since the expression $\left(\frac{1}{m} + \frac{1}{n}\right)^2 = \frac{1}{m^2} + \frac{2}{mn} + \frac{1}{n^2}$; that is, a larger multiple of $\frac{1}{mn}$ makes its presence known along with two additional terms. In a similar manner, one could argue that, for three numbers $p, q, r \in \mathbb{Z}^+$,

$$\frac{1}{pqr} < \left(\frac{1}{p} + \frac{1}{q} + \frac{1}{r}\right)^3$$

since $\left(\frac{1}{p} + \frac{1}{q} + \frac{1}{r}\right)^3$ expands to $\frac{1}{p^3} + \frac{3}{p^2q} + \frac{3}{p^2r} + \frac{3}{pq^2} + \frac{6}{pqr} + \frac{3}{pr^2} + \frac{1}{q^3} + \frac{3}{q^2r} + \frac{3}{qr^2} + \frac{1}{r^3}$. Generalizing, one obtains, for $n_1, n_2, \ldots, n_m \in \mathbb{Z}^+$, the inequality

$$\frac{1}{n_1 n_2 \cdots n_m} < \left(\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_m}\right)^m.$$

(Note that an inductive argument seems especially fitting here.) We mention these inequalities in this section because we will return to them momentarily. We now state our main claim.

Theorem 6.2.1 *The infinite series* $\sum_{n=1}^{\infty} \frac{1}{p_n}$ *diverges.*

Proof. We shall assume that the series converges and then arrive at a contradiction. Hence, we may state that $\sum_{n=1}^{\infty} \frac{1}{p_n} = L < \infty$. Furthermore, there must exist a $k \in \mathbb{N}$ such that $\sum_{n=1}^{k} \frac{1}{p_n}$ is within $\frac{1}{2}$ unit of L. Putting this in different terms, we have

$$\sum_{n=k+1}^{\infty} \frac{1}{p_n} < \frac{1}{2}.$$

Next, let's define the number $Q = (p_1)(p_2)(p_3)\cdots(p_k)$. We now multiply this number by $n \in \mathbb{Z}^+$ and add 1. A glance at the number $1 + nQ$ tells us that none of the primes $p_1, p_2, \ldots, p_k$ are factors. That is, $1 + nQ$ must be a product of other primes beyond the k^{th} one (see the proof in Problem 1.4.3 for similarities). So, depending on the specific value of n, $1 + nQ$ may look like $(p_{k+1})(p_{k+4})(p_{k+11})$ or $(p_{k+6})(p_{k+42})(p_{k+659})(p_{k+698})$, etc. Next we consider a multiple of the harmonic series, namely

$$\frac{1}{1+Q} \sum_{n=1}^{\infty} \frac{1}{n}.$$

Doing some work on this leads to

$$\begin{aligned}
\frac{1}{1+Q} \sum_{n=1}^{\infty} \frac{1}{n} &= \frac{1}{1+Q} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots\right) \\
&= \frac{1}{1+Q} + \frac{1}{2+2Q} + \frac{1}{3+3Q} + \cdots \\
&< \frac{1}{1+Q} + \frac{1}{1+2Q} + \frac{1}{1+3Q} + \cdots \\
&= \sum_{n=1}^{\infty} \frac{1}{1+nQ}.
\end{aligned}$$

That is,

$$\frac{1}{1+Q} \sum_{n=1}^{\infty} \frac{1}{n} < \sum_{n=1}^{\infty} \frac{1}{1+nQ}.$$

Then since $\frac{1}{1+Q} \sum_{n=1}^{\infty} \frac{1}{n}$ diverges, the larger sum $\sum_{n=1}^{\infty} \frac{1}{1+nQ}$ diverges just as well. At this point in the proof, we will show that $\sum_{n=1}^{\infty} \frac{1}{1+nQ}$ in fact, must converge. This contradiction should give us what we want. To do this, we split the series into terms based on the nature of the number $1 + nQ$.

1. Case 1: There could be many choices of n for which $1 + nQ$ is precisely equal to a prime number larger than p_k; we'll call these n's $n_1, n_2, n_3, \ldots$, etc. So considering this as a common class, we shall collect the terms in $\sum_{n=1}^{\infty} \frac{1}{1+nQ}$

for which this is true and sum the terms. So, for example, it might be that

$$\begin{aligned}\frac{1}{1+n_1Q}+\frac{1}{1+n_2Q}+\frac{1}{1+n_3Q}+\cdots &= \frac{1}{p_{k+3}}+\frac{1}{p_{k+12}}+\frac{1}{p_{k+38}}+\cdots\\ &\leq \frac{1}{p_{k+1}}+\frac{1}{p_{k+2}}+\frac{1}{p_{k+3}}+\cdots\\ &< \frac{1}{2},\end{aligned}$$

from our earlier assumption. Note that the $(k+3)^{\text{rd}}$, $(k+12)^{\text{th}}$, and $(k+38)^{\text{th}}$ primes are just choosen to illustrate the point.

2. Case 2: There could be choices of n for which $1+nQ=(p_v)(p_w)$; that is, $1+nQ$ is expressible as the product of two primes beyond the k^{th} one. In this case we have

$$\begin{aligned}\frac{1}{1+n_1Q}+\frac{1}{1+n_2Q}+\frac{1}{1+n_3Q}+\cdots &= \frac{1}{p_{v_1}p_{w_1}}+\frac{1}{p_{v_2}p_{w_2}}+\frac{1}{p_{v_3}p_{w_3}}+\cdots\\ &\leq \left(\frac{1}{p_{k+1}}+\frac{1}{p_{k+2}}+\frac{1}{p_{k+3}}+\cdots\right)^2\\ &< \left(\frac{1}{2}\right)^2.\end{aligned}$$

Note here that all of the p_v and p_w primes are contained in the collection $\{p_{k+1},p_{k+2},p_{k+3},\ldots\}$.

3. Case 3: Marching onward, there could be choices of n for which $1+nQ=(p_x)(p_y)(p_z)$; that is, $1+nQ$ is expressible as the product of three primes beyond the k^{th} one. Analogous to the other cases, we have

$$\begin{aligned}\frac{1}{1+n_1Q}+\frac{1}{1+n_2Q}+\frac{1}{1+n_3Q}+\cdots &\leq \left(\frac{1}{p_{k+1}}+\frac{1}{p_{k+2}}+\frac{1}{p_{k+3}}+\cdots\right)^3\\ &< \left(\frac{1}{2}\right)^3.\end{aligned}$$

The same comment as in Case 2 applies to the p_x, p_y, and p_z here.

Naturally, we consider the process above indefinitely. When the dust settles, the big picture is this: when reconsidering $\sum_{n=1}^{\infty}\frac{1}{1+nQ}$ via the classes discussed above, we

obtain

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{1+nQ} &< \frac{1}{2}+\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3+\cdots\\ &= \sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n\\ &< \infty.\end{aligned}$$

This contradiction proves that $\sum_{n=1}^{\infty}\frac{1}{p_n}$ diverges as sought. ■

Bibliography

[1] Aigner, M., & Ziegler, G. (1999). *Proofs from THE BOOK.* Berlin: Springer-Verlag.

[2] Boas, R. (1996). *A Primer of Real Functions.* Washington, D.C.: The Mathematical Association of America, Inc.

[3] Buchanan, O.L. (1970). *Limits: A Transition to Calculus.* Boston: Houghton Mifflin Company.

[4] Carothers, N., (2000). *Real Analysis.* United Kingdom: Cambridge University Press.

[5] Cloud, M., & Drachman, B. (1998). *Inequalities with Applications to Engineering.* New York: Springer-Verlag.

[6] Cupillari, A. (1993). *The Nuts and Bolts of Proofs.* Boston: PWS Publishing Company.

[7] de Souza, P.N., & Silva, J.N. (1998). *Berkeley Problems in Mathematics.* New York: Springer-Verlag.

[8] Fletcher, P., & Patty, C.W. (1996). *Foundations of Higher Mathematics.* Boston: PWS Publishing Company.

[9] Gaskill, H., & Narayanaswami, P. (1998). *Elements of Real Analysis.* New Jersey: Prentice-Hall.

[10] Goldberg, R. (1976). *Methods of Real Analysis.* New York: John Wiley & Sons, Inc.

[11] Kaplan, R., & Kaplan, E. (2003). *The Art of the Infinite.* New York: Oxford University Press, Inc.

[12] Knopp, K. (1956). *Infinite Sequences and Series.* New York: Dover Publications, Inc.

[13] Knopp, K. (1990). *Theory and Application of Infinite Series.* New York: Dover Publications, Inc.

[14] Kolmogorov, A., & Fomin, S. (1970). *Introductory Real Analysis.* New Jersey: Prentice-Hall.

[15] Lamport, L. (1994). *LaTeX: A Document Preparation System.* California: Addison-Wesley.

[16] Lay, S. (1990). *Analysis with an Introduction to Proof.* New Jersey: Prentice-Hall.

[17] Olmsted, J. (1959). *Real Variables.* New York: Appleton-Century-Crofts, Inc.

[18] Pickover, C. (1995). *Keys to Infinity.* New York: John Wiley & Sons, Inc.

[19] Polya, G. (1954). *Induction and Analogy in Mathematics.* New Jersey: Princeton University Press.

[20] Pownall, M. (1967). *A Prelude to the Calculus.* New York: McGraw Hill Book Company.

[21] Rosenlicht, M. (1986). *Introduction to Analysis.* New York: Dover Publications, Inc.

[22] Ross, K. (1980). *Elementary Analysis: The Theory of Calculus.* New York: Springer-Verlag.

[23] Royden, H. (1988). *Real Analysis.* New Jersey: Prentice-Hall.

[24] Rudin, W. (1976). *Principles of Mathematical Analysis.* New York: McGraw-Hill, Inc.

[25] Solow, D. (1990). *How to Read and Do Proofs.* New York: John Wiley & Sons.

[26] Velleman, D. (1994). *How to Prove It.* United Kingdom: Cambridge University Press.