Chapter 1

Writing Proofs

Proof serves many purposes simultaneously. In being exposed to the scrutiny and judgment of a new audience, the proof is subject to a constant process of criticism and revalidation. Errors, ambiguities, and misunderstandings are cleared up by constant exposure. Proof is respectability. Proof is the seal of authority. Proof, in its best instances, increases understanding by revealing the heart of the matter. Proof suggests new mathematics. The novice who studies proofs gets closer to the creation of new mathematics. Proof is mathematical power, the electric voltage of the subject which vitalizes the static assertions of the theorems.

Finally, proof is ritual, and a celebration of the power of pure reason. Such an exercise in reassurance may be necessary in view of all the messes that clear thinking clearly gets us into.

Philip J. Davis and Reuben Hersh

One of the goals for this course is to improve your ability both to discover and to express (in oral or written form) proofs of mathematical assertions. This section presents some guidelines that will be useful in each of these tasks, especially in preparing written proofs. These notes on writing proofs were originally produced by M. Stob. This version has been revised somewhat by R. Pruim

1.1 Proofs in Mathematics

I have made such wonderful discoveries that I am myself lost in astonishment; out of nothing I have created a new and another world.

John Bolyai

Mathematicians prove their assertions. This distinguishes mathematics from all other sciences and, indeed, all other intellectual pursuits. Indeed,
one definition of mathematics is that it is “proving statements about abstract objects.” You probably first met this conception of mathematics in your secondary school geometry course. While Euclid wasn’t the first person to prove mathematical propositions, his treatment of geometry was the first systematization of a large body of mathematics and has served as a model of mathematical thought for 2200 years. To understand the role of proof in mathematics, we can do no better than to start with Euclid. The basic ingredients of Euclidean geometry are three.

**Primitive terms.** Point, line, plane are all primitive terms in modern treatments of Euclidean geometry. We usually call these undefined terms and do not attempt to give definitions for them. (Euclid did give definitions for all his terms; we discuss the role of definition in mathematics later.) Of course a logical system must start somewhere; it is impossible to define all terms without falling prey to either circularity or infinite regress. David Hilbert, in his *Foundations of Geometry*, simply starts as follows:

> Let us consider three distinct systems of things. The things composing the first system, we will call points and designate them by the letters $A$, $B$, $C$, …; those of the second, we will call straight lines, and designate them by the letters $a$, $b$, $c$ …; and those of the third system, we will call planes and designate them by the Greek letters $\alpha$, $\beta$, $\gamma$, ….

Hilbert says nothing about what the “things” are.

**Axioms.** An axiom is a proposition about the objects in question which we do not attempt to prove but rather which we accept as given. One of Euclid’s axioms, for example, was “It shall be possible to draw a straight line joining any two points.” Aristotle describes the role of axioms.

> It is not everything that can be proved, otherwise the chain of proof would be endless. You must begin somewhere, and you start with things admitted but undemonstrable. These are first principles common to all sciences which are called axioms or common opinions.

Euclid and Aristotle thought of axioms as propositions which were “obviously” true. But it is not necessary to think of them as either true or false. Rather, they are propositions which we agree, perhaps for the sake of argument, to accept as given.

**Rules of inference.** Axioms and previously proved theorems are combined to prove new theorems using the laws of logic. The nature of such rules of inference is best illustrated by an example.
All men are mortal
Socrates is a man
Socrates is mortal

In this example, the axioms (called premises or hypotheses) are written above the line and the theorem (called the conclusion) is written below the line. The whole argument is called a deduction. This particular argument is an example of a rule of inference which is now usually called Universal Instantiation. Two important features of this argument characterize a rule of inference. First, the relationship of the conclusion to the hypotheses is such that we cannot fail to accept the truth of the conclusion if we accept the truth of the hypotheses. Of course we do not have to accept the truth of the hypotheses and so are not compelled to believe the conclusion. But a rule of inference must necessarily produce true conclusions from true hypotheses. Second, this relationship does not depend on the concepts mentioned (humanity, mortality, Socratiety) but only on the form of the propositions. Hilbert said it this way: “One must be able at any time to relace ‘points, lines, and planes’ with ‘tables, chairs, and beer mugs.’” The next nonsen- sical argument is as valid as the previous one.

All beer mugs are chairs
Socrates is a beer mug
Socrates is a chair

While neither hypothesis is true and certainly the conclusion is false (perhaps your favorite chair is named Socrates but the Socrates I am thinking of was not a chair), this argument too is a perfectly acceptable example of universal instantiation.

This then is how mathematics is created. Starting from axioms which are propositions about certain objects which may be undefined, the rules of inference are used to prove new theorems. Anyone accepting the truth of the axioms, must accept the truth of all the theorems.

Exercises

1. Morris Kline says that

Mathematics is a body of knowledge. But it contains no truths.

In a similar vein, Hilbert said

It is not truth but only certainty that is at stake.

What do they mean?
1.2 Written Proofs

*I like being a writer; it’s the paperwork I dislike.*

**Peter De Vries**

This chapter is specifically concerned with written proofs. While a proof might be thought of as an abstract object existing only in our minds, the fact is that mathematics advances only as far as proofs are communicated. And writing remains the principal means of such communication. So to be a mathematician, you need to learn how to prove things but also to write those proofs clearly and correctly. Learning to write proofs also makes reading other people’s proofs easier.

What does a written proof look like? Like any other piece of prose, a written proof is organized into sentences and paragraphs. Like all other correct writing, those sentences are written following all the standard conventions of the language in which they are written. Because of the precision of the thoughts that a proof must convey, it is especially important that the prose be clear and correct. In order to help the reader follow the argument being presented, it is helpful to use a style that emphasizes the organization and flow of the argument.

It is useful to think of a proof as consisting of three sorts of sentences. First, some sentences express the result of a deduction. That is they result from other sentences by applying one of the rules of inference as described in the last section. These sentences can be recognized by the fact that they start with words like “therefore,” “thus,” “so,” or “from this it follows that.” Of course such a sentence will not normally be the first sentence of the proof and such a sentence normally depends on one or more earlier sentences of the proof.

A second sort of sentence is used to state a given fact. These sentences normally are restatements of hypotheses of the theorem being proved or recall facts that the reader should know already. An example of the latter type is the sentence “Recall that \( \sqrt{2} \) is an irrational number.” Again, these sentences should contain some sort of cue that they are stating hypotheses or previously proved facts rather than asserting something new.

Finally, a third sort of sentence appearing in a proof is what might be called “glue” or the prose that holds the other sentences together. These sentences are used for purposes such as to inform the reader of the structure of the argument (“Next we will show that \( G \) is abelian.”) or to establish notation (“Let \( X = Y \oplus Z \).”). This glue may also include longer passages that outline an entire argument in advance, summarize it at the conclusion, provide motivation for the methods being employed, or describe an example. It is a common error of beginning proof-writers to use too few of these sentences so that the structure of the argument is not clear.

This description of what a proof should look like is at odds with some of the ideas of proof that students take with them from high-school math-
ematics and perhaps even calculus. Some students think that a proof, like the solution to any homework problem, should be a series of equations with few, if any, words. Now an equation is a perfectly acceptable sentence. The equation “$x + 2 = 3$” is a sentence and could, perhaps, appear in a proof. Mathematical notation serves as abbreviation for words and phrases which would be tedious to write many times. However this equation almost certainly should not appear as a sentence of a proof, for it is not clear whether it expresses a hypothesis or a conclusion. Probably what is meant is “Suppose that $x$ is a real number such that $x + 2 = 3.$” or “Therefore, it follows from the last equation that $x + 2 = 3.$” Equations must be tied together with words.

A second misconception of proof is a side-effect of studying geometry in high school. High school geometry courses often teach that a proof should be arranged in two columns; in one column we write assertions and in the other we write the reason that the corresponding assertion is true. Look in a mathematics journal and you will see that no mathematician writes proofs in this manner. The main reason for teaching “two-column” proofs is that they sometimes help beginning proof writers concentrate on the logical structure of the argument without having to attend to the details of grammatical writing. A two-column proof reinforces the notion that each assertion in a proof should serve a definite purpose and must be justified. For realistic theorems however (unlike the baby theorems of high school geometry), the two-column proof becomes cumbersome and hard to read. The two-column proof, among other defects, is missing the “glue” described above. These informative sentences are crucial to helping a reader understand a complicated proof. We want the reader to understand not only the justification of each step but also the logical structure of the argument as a whole. Nevertheless, a two-column proof is sometimes a good place to start on scratch paper.

1.3 Proverbs

Here I have written out for you sayings full of knowledge and wise advice

Proverbs 22:20

Proof-writing is, to some extent at least, a creative activity. There is no step-by-step recipe that, if followed carefully, is guaranteed to produce a proof. Proof-writing is different in this way than differentiation in calculus. Machines can differentiate as well as you can, but have not made much progress in producing proofs. However, experienced proof-writers do have certain principles by which their search for a proof is at least implicitly guided. In these notes we try to summarize at least some of these principles in the form of proverbs. As a vehicle for introducing the first few of these proverbs, we consider the proof of the following (easy) theorem.
Theorem 1.1  The square of an even natural number is even.

How does one get started writing a proof? The following proverb suggests the answer.

Proverb 1.2  The form of the theorem suggests the outline of the proof.

In fact

Proverb 1.3  You can write the first line of a proof, even if you don’t understand the theorem.

In a later section, we make a more detailed study of theorem forms. However Theorem 1.1 has a common and easily recognizable form. Here is an abstract version of that form.

Theorem 1.4  If object x has property P then x has property Q.

or more simply

Theorem 1.5  If x is a P then x is a Q.

It will become clear to you in your study of abstract mathematics that many theorems have exactly this form, though you might have to rewrite the theorem a bit to see this clearly. Here are some examples from all sorts of different mathematical specialities.

Theorem 1.6  Every differentiable function is continuous.

Theorem 1.7  Every semistable elliptic curve is modular.

Theorem 1.8  The lattice of computably enumerable sets is distributive.

Theorem 1.9  If I is a prime ideal of a ring R, then R/I is an integral domain.

The first of these theorems is one you know and love from your calculus class. The second is the famous theorem of Wiles proved in 1993 (it implies Fermat’s last theorem). The third comes from a field of mathematical logic known as recursion theory. The last theorem is one that you would meet in an abstract algebra class. Notice that Theorem 1.9 is really about two objects and both properties P and Q express relationships between these two objects. Note too that while Theorem 1.8 appears to be about only one object (the lattice of recursively enumerable sets), the other three theorems are about an unspecified number of objects. In fact, our Theorem 1.1 is about infinitely many objects.

So if Proverb 1.2 is correct, the outline of the proof Theorem 1.4 should suggest itself. In fact, by Proverb 1.3, we should be at least able to write
the first sentence of the proof. Indeed, here is a reasonable way to begin theorems such as Theorem 1.4.

Proof (of Theorem 1.4). Suppose $x$ has property $P$. We must show that $x$ has property $Q$. □

The second sentence might be a bit pedantic, but the first clearly sets the stage. Thus, our proof of Theorem 1.1 begins as follows.

Proof (of Theorem 1.1). Suppose that $n$ is an even natural number ... □

Actually, we can also predict the last line of theorems with this form. We would expect the proof of Theorem 1.1 to look like this.

Proof (of Theorem 1.1). Suppose that $n$ is an even natural number. ... Thus $n^2$ is even. □

So what remains in the proof of the theorem is to fill in the ... . We need to reason from the fact that $x$ has property $P$ to the fact that $x$ has property $Q$. In our particular example, and in many similar cases, the next proverb supplies the key.

Proverb 1.10 Use the definition.

In some cases it will be obvious that an $x$ with property $P$ also has property $Q$. However, if it is that obvious, you probably will not be asked to supply a proof in the first place. So the first step is usually to determine what it means to have property $P$. All mathematical terms, except the primitive ones, have precise definitions. In our case, the following definition is the important one.

Definition 1.11 A natural number $n$ is even if there is a natural number $k$ such that $n = 2k$.

Using the definition gives us the second line of our proof. It also gives us a hint as to what the next to last line of the proof might be.

Proof (of Theorem 1.1). Suppose that $n$ is an even natural number. Then there is a natural number $k$ such that $n = 2k$. ... Thus there is a natural number $l$ such that $n^2 = 2l$. Thus $n^2$ is even. □

The only work in this theorem is filling in the remaining gap. And this work is simply a computation. Thus the complete proof is

Proof (of Theorem 1.1). Suppose that $n$ is an even natural number. Then there is a natural number $k$ such that $n = 2k$. So $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$. Thus there is a natural number $l$ (namely $2k^2$) such that $n^2 = 2l$. Thus $n^2$ is even. □

Notice that the end of each of these proofs above is marked by the symbol □. This or some other symbol is often used in books and articles to mark the ends of proofs. Especially in older writing, the letters QED (which stand...
for a Latin phrase meaning ‘that which was to be proven has been shown’) are often used. Even when using such a marker, the last sentence or two of a proof should also indicate somehow that the conclusion of the argument has been reached, but the additional marker is also helpful if, for example, the reader wants to skip over the proof on a first reading or get an estimate on its length.

Here is one more (easy) theorem of this type will serve to reinforce the point of this section.

**Theorem 1.12** *The base angles of an isosceles triangle are congruent.*

Recasting this theorem in the form of Theorem 1.4, we might write this theorem as

**Theorem 1.13** *If a triangle is isosceles, then the base angles of that triangle are congruent.*

A diagram, Figure 1.3, serves to illustrate the statement of the theorem.

![Diagram](image)

**Proof.** Suppose that $\triangle ABC$ is isosceles. That is, suppose that $AB$ and $BC$ are the same length. (Note the clever use of the definition of isosceles triangle.) We must show that the base angles $\alpha$ and $\gamma$ are equal. To do that notice that triangles $\triangle ABC$ and $\triangle CBA$ are congruent by side-angle-side. Furthermore, $\alpha$ and $\gamma$ are corresponding angles of these two triangles. Thus $\gamma$ and $\alpha$ are congruent (equal) since corresponding parts of congruent triangles are congruent.

**Exercises**

1. Write good first and last lines to the proofs of
   (a) Theorem 1.6
   (b) Theorem 1.7
   (c) Theorem 1.8
   (d) Theorem 1.9.

2. (a) Write a good definition of “odd natural number.”
(b) Prove that the square of an odd number is odd.

3. Prove the following theorem

**Theorem 1.14** If a triangle has two congruent angles, then the triangle is isosceles.

(This theorem is the converse of Theorem 1.12. It is always difficult to know what you can assume in geometry, but the proof of this Theorem is similar to that of Theorem 1.12 — it only uses simple facts about congruent triangles.)

### 1.4 Contraposition and Contradiction

“*There is no use trying,*” said Alice; “*one can’t believe impossible things.*”

“I dare say you haven’t had much practice,” said the Queen,

“When I was your age, I always did it for half an hour a day. Why, sometimes I’ve believed as many as six impossible things before breakfast.”

**Lewis Carroll**

In the last section, we saw how to produce a reasonable outline of a proof of a theorem with the following form.

**Theorem 1.15** For all \(x\), if \(x\) has property \(P\) then \(x\) has property \(Q\).

However,

**Proverb 1.16** There is more than one way to prove a theorem.

In this section, we look at two alternative outlines of the proof of Theorem 1.15. The first of these is called “proof by contraposition.” It depends on the following lemma.

**Lemma 1.17** Suppose that \(\alpha\) and \(\beta\) are propositions. Then the statement

\[
\text{If } \alpha \text{ then } \beta \tag{1.1}
\]

is logically equivalent to the statement

\[
\text{If } \beta \text{ is not true then } \alpha \text{ is not true.} \tag{1.2}
\]
The statement 1.2 is called the contrapositive of the statement 1.1. Here logically equivalent means that 1.1 is true just in case 1.2 is true. So to prove a statement of form 1.1, it is sufficient to prove 1.2. This fact gives us the first of two alternate outlines for a proof of Theorem 1.15.

\textit{Proof (of Theorem 1.15).} We prove the theorem by contraposition. Suppose that \( x \) does not have property \( Q \). Then \( x \) does not have property \( P \).

As a specific example of a proof with this outline, consider the following theorem.

\textbf{Theorem 1.18} \textit{Suppose that} \( n \) \textit{is a natural number such that} \( n^2 \) \textit{is odd. Then} \( n \) \textit{is odd.}

\textit{Proof.} We prove the theorem by contraposition. Suppose that \( n \) is not odd. That is suppose that \( n \) is even. Then \( n^2 \) is even (by Theorem 1.1). Thus \( n^2 \) is not odd.

How does one decide whether to prove a theorem directly or by contraposition? Often, the form of the properties \( P \) and \( Q \) give us a clue. If \( Q \) is a negative sort of property, (such as \( x \) is not divisible by 3), it may very well be easier to work with the hypothesis that \( x \) does not have property \( Q \). Of course we could always try to prove the theorem both ways and see which one works.

Another alternate way to prove a theorem is to give a proof by contradiction. In a proof by contradiction, we suppose that the theorem is false and derive some sort of false statement from that assumption. If mathematics is consistent and if our reasoning is sound, this means that our assumption that the theorem is false is in error. So the outline of a proof by contradiction of Theorem 1.15 is as follows.

\textit{Proof (of Theorem 1.15)} \textit{.} We prove the theorem by contradiction. So suppose that \( x \) has property \( P \) but that \( x \) does not have property \( Q \). Then \( 1=2 \) (or any other obviously false statement). Thus our assumption is in error and it must be the case that if \( x \) has property \( P \), \( x \) also has property \( Q \).

The most famous theorem which is usually proved by contradiction is the following.

\textbf{Theorem 1.19} \( \sqrt{2} \) \textit{is irrational.}

We first rephrase the theorem so that it has the form of Theorem 1.15.

\textbf{Theorem 1.20} \textit{If} \( x \) \textit{is any real number such that} \( x^2 = 2 \), \textit{then} \( x \) \textit{is not rational.}
This form is logically superior to that of Theorem 1.19 since it doesn’t assume the existence or uniqueness of a number $\sqrt{2}$. This form also suggests either contraposition or contradiction since the property “is not rational” is not as easy to work with as the property “is rational.”

**Proof.** We proof Theorem 1.20 by contradiction. So suppose that $x$ is a number such that $x^2 = 2$ and $x$ is rational. Then there are natural numbers $p$ and $q$ such that $p$ and $q$ have no common divisors and $x = \frac{p}{q}$. Then

$$2 = x^2 = \left(\frac{p}{q}\right)^2$$

Therefore $p^2 = 2q^2$. This implies that $p^2$ is even and so $p$ is even (See exercise 1 below.) Since $p$ is even, $p = 2k$ for some natural number $k$. Thus $2q^2 = p^2 = 4k^2$ or $q^2 = 2k^2$. But this implies that $q^2$ is even and so that $q$ is even. Thus $p$ and $q$ are both even but this cannot be true since we assumed that $p$ and $q$ had no common divisors. \(\square\)

Students sometimes confuse proofs by contradiction and contraposition. The proofs look similar in the beginning since each begins by assuming that $x$ does not have property $Q$. But a proof by contradiction also assume that $x$ has property $P$ and proves the negation of some true proposition while a proof by contraposition does not make this assumption and simply proves that $x$ does not have property $P$.

**Exercises**

1. Prove that if $p$ is a natural number such that $p^2$ is even then $p$ is even.

2. Prove that $\sqrt{3}$ is irrational.

3. Prove that $\sqrt{6}$ is irrational.

4. Write the contrapositive of the following theorems.
   (a) Theorem 1.6
   (b) Theorem 1.7
   (c) Theorem 1.8
   (d) Theorem 1.9.

5. Write the assumption made in a proof by contradiction of the following theorems.
   (a) Theorem 1.6
   (b) Theorem 1.7
   (c) Theorem 1.8
   (d) Theorem 1.9.
1.5 Ten Rules

When I read some of the rules for speaking and writing the English language correctly . . . I think any fool can make a rule and every fool will mind it.

Henry Thoreau

As in most writing, there are certain mistakes that occur over and over again in mathematical writing. As well, there are certain conventions that mathematicians adhere to. In this section, we present ten rules to follow in writing proofs. If you follow these rules, you will go a long way towards making your writing clear and correct. Most of these rules come in one form or another from the wonderful book, Mathematical Writing, by Donald Knuth.

**Rule 1** Use the present tense, first person plural, active voice.

Bad: It will now be shown that . . .
Good: We show . . .

**Rule 2** Choose the right technical term.
For example, not all formulas are equations.

**Rule 3** Don’t start a sentence with a symbol.

Bad: \( x^n - a \) has \( n \) distinct roots.
Good: The polynomial \( x^n - a \) has \( n \) distinct roots.

**Rule 4** Respect the equal sign.

Bad: \( x^2 = 4 = |x| = 2 \).
Good: If \( x^2 = 4 \), then \( |x| = 2 \).

**Rule 5** Normally, if “if”, then “then”.

Bad: If \( x \) is positive, \( x > 0 \).
Good: If \( x \) is positive, then \( x > 0 \).

**Rule 6** Don’t omit “that”.

Bad: Assume \( x \) is positive.
Good: Assume that \( x \) is positive.
Rule 7 Identify the type of variables.

Bad: For all \( x, y \), \(|x + y| \leq |x| + |y|\).

Good: For all real numbers \( x, y \), we have \(|x + y| \leq |x| + |y|\).

There is an exception to this rule. In many situations, the notation used implies the type of a variable. For example, it is usually understood that \( n \) denotes a natural number. In Real Analysis, \( x \) almost always denotes a real number. In these cases, the type can be suppressed in the interest of saving ink and avoiding the distraction of extra words.

Rule 8 Use “that” and “which” correctly.

Bad: The least integer which is greater than \( \sqrt{27} \) . . .

Good: The least integer that is greater than \( \sqrt{27} \) . . .

The distinction is this: ‘that’ introduces a modifying clause that adds information necessary to specify the object modified; ‘which’ introduces a modifying clause that adds additional information about an already specified object.

Rule 9 A variable used as an appositive need not be set off by commas.

Bad: Consider the group, \( G \), that . . .

Good: Consider the group \( G \) that . . .

Rule 10 Don’t use symbols such as \( \exists \), \( \forall \), \( \vee \), \( > \) in text; replace them by words. Symbols may, of course, be used in formulas.

Bad: Let \( S \) be the set of numbers \( < 1 \).

Good: Let \( S \) be the set of numbers less than 1.

1.6 Forms of Theorems

And if you go in, should you turn left or right . . .
or right-and-three-quarters? Or maybe not quite?
Or go around back and sneak in from behind?
Simple it’s not, I’m afraid you will find,
for a mind-maker-upper to make up his mind.
Not all theorems have the form of Theorem 1.4. In this section we look at a few of the more common forms.

1.7 Biconditional Theorems

A biconditional theorem is one with form

**Theorem 1.21** $P$ is true if and only if $Q$ is true.

The proof of Theorem 1.21, as you would suspect, has two parts.

*Proof (of Theorem 1.21).* First, suppose that $P$ is true. \ldots Then $Q$ is true.

Next, suppose that $Q$ is true. \ldots Then $P$ is true. \hfill $\square$

Often, one of the two directions of a biconditional is best proved by contraposition.

1.8 Existence Theorems

An existence theorem is one with the form

**Theorem 1.22** There is an $x$ with property $P$.

Here is a specific example.

**Theorem 1.23** There is a real number $x$ such that $x^2 + 6x - 17 = 0$.

The most common form of proof of a theorem with the form of Theorem 1.22 is

*Proof.* Let $x$ be defined (constructed, given) by \ldots Then $x$ has property $P$ because \ldots \hfill $\square$

For example, here is a complete proof of Theorem 1.23.

*Proof.* Let $x$ be given by

$$x = -3 + \sqrt{26}.$$ 

Then $x^2 + 6x - 17 = (-3 + \sqrt{26})^2 + 6(-3 + \sqrt{26}) - 17 = 9 - 6\sqrt{26} + 26 - 18 + 6\sqrt{26} - 17 = 0$. \hfill $\square$

It is sometimes possible to prove Theorem 1.22 by contradiction. Such a proof would look like

*Proof (of Theorem 1.22).* We prove the theorem by contradiction. So suppose that no such $x$ exists. Thus no $x$ has property $P$. \ldots Then $0=1$. Therefore our assumption that no such $x$ exists must be in error so indeed such an $x$ does exist. \hfill $\square$

Such a proof by contradiction is rather peculiar since we now know that $x$ exists but the proof does not give us any way to find $x$. 
1.9 Uniqueness

A uniqueness theorem is one with form

**Theorem 1.24** There is a unique \( x \) with property \( P \).

This theorem actually says two things; there is an \( x \) with property \( P \) but there is only one \( x \) with that property. Thus, the proof has two parts; an existence part (and we have already discussed that) and a uniqueness part. The proof of Theorem 1.24 therefore looks like this.

**Proof. Existence.** Define \( x \) by \( \ldots \) Then \( x \) has property \( P \) because \( \ldots \)

**Uniqueness.** Suppose that \( x \) and \( y \) have property \( P \). \( \ldots \) Then \( x = y \). \( \Box \)

Often, the uniqueness portion of Theorem 1.24 is proved by contradiction. Then the second part of the proof looks like

**Proof. Uniqueness.** We prove uniqueness by contradiction. So suppose \( x \) and \( y \) have property \( P \) and \( x \neq y \). \( \ldots \) Then 0=1. Therefore our assumption that \( x \neq y \) must be in error. \( \Box \)

The next theorem is a typical (though simple) example of a uniqueness theorem. Recall that an additive identity is a number 0 with the property that \( 0 + x = x + 0 = x \) for all real numbers \( x \).

**Theorem 1.25** The additive identity for the set of real numbers is unique.

**Proof.** Suppose that 0 and \( 0' \) are numbers satisfying the defining condition for being an additive identity, namely that

\[
0 + x = x + 0 = x \quad \text{for all } x \quad (1.3)
\]

and

\[
0' + x = x + 0' = x \quad \text{for all } x \quad (1.4)
\]

Then by 1.3 (with \( x = 0' \)) we have that \( 0 + 0' = 0' \) and by 1.4 (with \( x = 0 \)) we have that \( 0 + 0' = 0 \). Thus \( 0 = 0' \). \( \Box \)

1.10 Universal Statements

A universal statement is one of form

**Theorem 1.26** For all \( x \), \( x \) has property \( R \).

In fact, our original theorem form, Theorem 1.4, can best be understood as a universal statement.

**Theorem 1.27** For all \( x \), if \( x \) has property \( P \) then \( x \) has property \( Q \).
Consider again our proof of Theorem 1.4. We write it somewhat differently.

**Proof.** Let $x$ be an arbitrary object with property $P$. Then ... Then $x$ has property $Q$. 

The key feature of this proof that allows us to claim that the Theorem is true of all *objects*, is that we did not assume any special properties of $x$ (other than $P$). That is, though we named $x$, $x$ was otherwise a “generic” object. You will recall this sort of reasoning from geometry. If you were asked to prove something for all triangles, you probably drew a triangle and labeled the vertices $A$, $B$, and $C$, but realized that you were not allowed to assume anything special about the triangle (say, from the diagram). Your proof had to work for equilateral triangles as well as triangles that were not even isosceles. So the general proof of a universal statement like Theorem 1.26 goes as follows.

**Proof.** Let $x$ be given. ... Then $x$ has property $R$. 

Here, the ... is filled in by an argument which assumes nothing in particular about $x$. As a simple concrete example, consider the following.

**Theorem 1.28** For all real numbers $x$, $x^2 + x + 1 \geq 0$.

**Proof.** Let $x$ be an arbitrary real number.

$$x^2 + x + 1 = (x^2 + x + \frac{1}{4}) + \frac{3}{4} = (x + \frac{1}{2})^2 + \frac{3}{4}$$

But $(x + 1/2)^2 \geq 0$ so that $(x + 1/2)^2 + 3/4 > 0$. Thus $x^2 + x + 1 > 0$. 

In this proof, we used no special property of $x$ and so we were entitled to conclude that the property in question held for all $x$.

To prove a theorem like Theorem 1.26 by contradiction, we have to realize that to deny that *every* $x$ has property $R$ is to assert that *some* $x$ does not have property $R$. So a proof by contradiction of Theorem 1.26 might look like this.

**Proof.** We prove the theorem by contradiction. So suppose that there is an $x$ which does not have property $R$. ... Then $0=1$. Therefore, no such $x$ exists, i.e., all $x$ have property $R$. 

The middle part of this proof would presumably be filled in by an argument showing that $x$ has some impossible property.
Chapter 2

Cardinality

In this chapter we want to investigate the sizes of sets, especially infinite sets. We will discover that infinite sets can come in different sizes and make an important distinction between the smallest infinite sets (called countably infinite sets) and larger infinite sets (called uncoutable sets).

In Section 2.2 we will make precise what we mean by the size of a set and present some of the most important basic results on cardinality (the technical name for the size of a set). In section 2.3, we will illustrate Cantor’s zig-zag and diagonalization techniques, two important techniques for establishing the cardinality of a set. But first, we turn our attention to some motivating examples from everyday life.

2.1 Counting, Children, and Chairs

We will motivate the definitions of cardinality by considering some simpler situations that involve counting of everyday objects, like blocks and people and chairs. If you have ever seen a young child attempt to count, you know that this is something that must be learned, and that young children must pass through several phases as they learn to count. In each phase, some new skill must be learned or some error corrected.

The first step is to memorize a list of words: one, two, three, four . . . 1 Even once the list can be repeated, in order and without error, the child cannot really count anything until she understands how to associate these words with the objects being counted.

And so the child is taught to say these words while pointing to objects or pictures in a book. This phase of associating the number-words with objects develops slowly, and at first the child is likely to make one or both of two errors that lead to a miscount. The first error is to skip some of the objects and leads to an undercount. The second error is to say more than

---

1Later, for counting larger sets, it will be necessary to grasp the pattern by which new names for numbers are generated, but this generally comes after a child has learned to count with the memorized list of numbers to, say, twenty.
one number while pointing to the same object and leads to an overcount. So a typical young child presented with 7 objects might obtain a “count” of ten by randomly pointing at objects and saying numbers until they reach ten (a natural stopping point). Some objects will have been pointed at more than once, of course. Perhaps others were never pointed at.

Counting has only been mastered when the child understands that each object must be associated with exactly one number from the list (no skips, no double counts). The last number spoken is then called the size of the collection of objects.\(^2\)

So we see that counting is the association of elements in two sets (in our example above a set of number-words and a set of objects) such that each element of one set is paired with exactly one element of the other set. We could visualize counting three blocks as follows:

\[
\begin{array}{c|c}
1 & A \\
2 & F \\
3 & D \\
\end{array}
\]

Now suppose you are having a dinner party. At some point shortly before dinner you begin to fear that you do not have the correct number of chairs at the table. There are two ways to check. One is to count the chairs and the people and see if the numbers match. But there is another (albeit potentially more embarassing) method. Simply invite everyone to the table. If every person has a chair and every chair has one person sitting on it, then the number of chairs and people is the same, even though we don’t know just what that number is without some additional work. On the other hand, if there is an unoccupied chair, or if there is a person left standing . . .

The point here is that even though we do not know names for “infinite numbers” (although they do exist) and do not have time to count elements in infinite sets (or very large finite sets) by pointing to them and reciting a memorized list of words, we can still compare two sets to see if they are the same size by pairing up elements in one set with elements in the other. Provided we do so while avoiding the errors of skipping elements or double-matching elements, we will know our two sets have the same size. We will formalize this in more mathematical language in the next section.

\(^2\)A good indication that all this is in place is when a child says something like “one, two, three; three blocks”. The repetition of “three” indicating that the size of the collection of blocks has been associated with the last number used to count them. Thus “three” is actually being used to in two related but somewhat different ways – no wonder it takes some practice to learn to count.
2.2 Basic Cardinality Results

Mathematically, the “pairing up” of elements between two sets is done by means of a function \( f \) from \( A \) to \( B \) (written \( f : A \rightarrow B \)), which we can think of informally as an assignment of an element in \( B \) to each element of \( A \).

We want our function to avoid the errors of skipping or double-counting, so we introduce the following definitions:

**Definition 2.1 (one-to-one)** A function \( f : A \rightarrow B \) is one-to-one if for any \( a_1, a_2 \in A \), if \( a_1 \neq a_2 \), then \( f(a_1) \neq f(a_2) \).

**Definition 2.2 (onto)** A function \( f : A \rightarrow B \) is onto if for any \( b \in B \), there is some \( a \in A \) such that \( f(a) = b \).

Notice that a one-to-one function is a function that does not double-count and an onto function is one that does not skip. One-to-one functions are sometimes called *injective* functions or *injections*. Onto functions are sometimes called *surjective* functions or *surjections*. A function that is both one-to-one and onto is called a *bijection*.

With all this background, it is pretty clear what it means to say that two sets have the same size:

**Definition 2.3 (Same-sized sets)** Two sets \( A \) and \( B \) have the same cardinality (written \( |A| = |B| \)) if there is a function \( f : A \rightarrow B \) such that

- \( f \) is one-to-one, and
- \( f \) is onto.

Note that although the notation suggests that we have implicitly defined \( |A| \) and \( |B| \) (the sizes of the sets \( A \) and \( B \)), we have not really done so. This can be done by specifying the list of infinite numbers (called cardinals) to be used once we finish with the the natural numbers, but it is not necessary for our purposes.\(^3\)

**Example 2.4** Let \( E \) be the non-negative even integers \((E = \{0, 2, 4, 6, 8, \ldots \})\). Then \( |N| = |E| \), since \( f : n \mapsto 2n \) is one-to-one and onto.

Notice that the example above shows that infinite sets behave a bit differently from finite sets. The natural numbers include all of the evens, and much more besides; nevertheless, these two sets have the same size. In fact, this can be taken as the definition of an infinite set:

\(^3\)The study of cardinal numbers and cardinal arithmetic is part of a branch of logic known as set theory. It turns out that the properties of cardinals are closely related to properties of sets and in fact depend to some extent on the axioms one chooses to use for set theory.
Definition 2.5 (Finite, infinitite) A set $A$ is infinite if it has a proper subset $B \subseteq A$ such that $|A| = |B|$. Otherwise, $A$ is finite.

If when we pair up the elements of $A$ with elements of $B$ we use up all of $A$ but perhaps have skipped over some elements of $B$, then $A$ cannot be larger than $B$:

Definition 2.6 (No bigger than) A set $A$ is no bigger than the set $B$ (written $|A| \leq |B|$), if there is a function $f : A \to B$, such that

- $f$ is one-to-one.

The notation chosen above suggests that “same size as” and “no bigger than” behave in nice ways. The next two theorems demonstrate that this is the case:

Theorem 2.7 (Properties of $=$) “Same size as” is an equivalence relation. That is,  

1. “Same size” is reflexive: $|A| = |A|$.
2. “Same size” is symmetric: If $|A| = |B|$, then $|B| = |A|$.
3. “Same size” is transitive: If $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.

Theorem 2.8 (Properties of $\leq$)

1. “No bigger than” is reflexive: $|A| \leq |A|$.
2. “No bigger than” is transitive: If $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$.
3. If $A \subseteq B$, then $|A| \leq |B|$.
4. $|B| \leq |A|$ if and only if there is a function $f : A \to B$ that is onto.
5. If $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

Thus the use of $=$ and $\leq$ is justified in our notation. This also suggests some extensions to the notation:

- $|B| \geq |A|$ means $|A| \leq |B|$,
- $|A| \neq |B|$ means it is not the case that $|A| = |B|$,
- $|A| < |B|$ means $|A| \leq |B|$ but $|A| \neq |B|$.

Definition 2.9 (Countable) If $|A| = |\mathbb{N}|$, then we say that $A$ is countably infinite. A countable set is any set that is either finite or countably infinite. In other words, a countable set is the same size as some subset of $\mathbb{N}$. 
Most of the important infinite sets we will encounter in this class will be countably infinite. The following theorem establishes some nice properties of countable sets.

**Theorem 2.10 (Properties of countable sets)**

1. The following sets are all countable: \( \mathbb{N} \), \( \mathbb{Z} \) (the integers), \( \mathbb{Q} \) (the rationals), the set of even integers.

2. The following sets are uncountable: \( \mathbb{R} \) (the real numbers), \([0, 1]\) (the reals in the interval from 0 to 1).

3. A finite union of countable sets is countable:
   If \( A \) and \( B \) are countable, then \( A \cup B \) is also countable.

4. The cross product of countable sets is countable:
   Let \( A \) and \( B \) be countable, then \( A \times B \) is countable.

5. A countable union of countable sets is countable:
   Suppose that for each natural number \( n \), \( A_n \) is countable. Then \( A = \bigcup_{n=0}^{\infty} A_n \) is also countable.

6. The set of all finite sequences from a countable set is countable:
   Let \( A^{\geq n} \) denote the set of all sequences of \( n \) items from \( A \). (For example, \((1,4,3,0) \in \mathbb{N}^{\geq 4}\).) Let \( A^* = \bigcup_{n=0}^{\infty} A^{\geq n} \). Then \( A^{\geq n} \) and \( A^* \) are countable.

7. Every infinite set has a countably infinite subset.

Finally, Cantor’s diagonalization argument can be used to establish the following general fact:

**Theorem 2.11 (Power set)** Let \( \mathcal{P}(A) \) denote the power set of \( A \) (i.e., the set of all subsets of \( A \)). Then \( |A| < |\mathcal{P}(A)| \).

Among other things, this shows that there is no largest size of set.

You are asked to prove 3 of Theorem 2.10 in the exercises. The proofs the rest of the this theorem and of Theorem 2.11 will have to wait until we have seen Cantor’s two powerful techniques of zig-zag and diagonalization.

**Exercises**

1. Prove Theorem 2.7.

2. Prove Theorem 2.8.

3. Prove part 3 of Theorem 2.10.

4. Show that \( \mathbb{Z} \) is countable.

5. Assuming that \( \mathbb{R} \) is uncountable and \( \mathbb{Q} \) is countable, determine whether the set of irrationals is countable or uncountable. Justify your claim.
2.3 Cantor’s Two Ideas

Zig-Zag

Several of the countability results of the previous section can be proven using a zig-zag argument that goes back to Cantor. We will use this method here to prove part 4 of Theorem 2.10.

Proof (of Theorem 2.10, part 4). Let $A$ and $B$ be countably infinite. We must show that $A \times B$ is countably infinite. (Strictly speaking, we need to deal with the cases where one or both of the sets are finite, too, but we will only do the case where both are infinite here.)

Since $A$ and $B$ are countably infinite, there are one-to-one, onto functions $f : \mathbb{N} \to A$ and $g : \mathbb{N} \to B$. So $A \times B = \{(f(i), g(j)) | i, j \in \mathbb{N}\}$. The key idea of Cantor is to arrange the elements of $A \times B$ in a rectangular grid filling one quadrant of the plane:

$$(f(0), g(0)) \leftrightarrow (f(0), g(1)) \quad (f(0), g(2)) \leftrightarrow (f(0), g(3))$$

$$(f(1), g(0)) \quad (f(1), g(1)) \quad (f(1), g(2))$$

$$(f(2), g(0)) \quad (f(1), g(2))$$

$$(f(3), g(0))$$

$$(f(4), g(0))$$

As the picture indicates, we can enumerate $A \times B$ beginning in the upper left-hand corner and following the arrows. In fact, by slightly modifying the zig-zag argument (travel each diagonal from top to bottom instead of zig-zagging), it is not too hard to give an exact formula for a one-to-one, onto function mapping $A \times B$ to $\mathbb{N}$ (or vice versa).

It is worthwhile to give another proof of this.
Proof (of Theorem 2.10, part 4). This time we will make use of part 5 of Theorem 2.8. For this we need to exhibit one-to-one functions $\alpha : A \times B \to \mathbb{N}$ and $\beta : \mathbb{N} \to A \times B$. The following two functions can easily be shown to be one-to-one (for the first we use the fact that prime factorizations are unique):

$$\alpha : (f(i), g(j)) \mapsto 2^i3^j$$
$$\beta : n \mapsto (f(n), g(0))$$

\hfill \square

Diagonalization

Cantor’s diagonalization idea is even cleverer than the previous idea. We will use it here to prove Theorem 2.11.

Proof (of Theorem 2.11). Let $A$ be any set, we need to show that $|A| < |\mathcal{P}(A)|$. First notice that clearly $|A| \leq |\mathcal{P}(A)|$, since

$$x \mapsto \{x\}$$

is one-to-one.

The heart of the matter is to show that there is no function $f : A \to \mathcal{P}(A)$ that is onto. We will do this using the method of “defeating an arbitrary example”. For any function $f : A \to \mathcal{P}(A)$, we will describe a method to show that it is not onto. That is, for any such function $f$, we must find some subset $S_f$ of $A$ that gets “missed” by the function $f$. One such set is

$$S_f = \{a \in A | a \notin f(a)\}.$$

Since for every $a$, $a \in S_f \iff a \notin f(a)$, we see that $S_f \neq f(a)$; that is, $S_f$ is missed by the function $f$. Since we have not assumed anything special about $f$, this shows that no function $f$ can be onto. Therefore $|A| \neq |\mathcal{P}(A)|$.

\hfill \square

Exercises

1. Prove Theorem 2.10.

2. Show that the set of irrational numbers is uncountable.

3. Which of the following are countable, which are uncountable?

   (a) The set of all functions from $\mathbb{N}$ to $\mathbb{N}$.
   (b) The set of all one-to-one functions from $\mathbb{N}$ to $\mathbb{N}$.
   (c) The set of all functions from $\mathbb{N}$ to $\mathbb{N}$ that are eventually 0.
   (d) The set of all functions from $\mathbb{N}$ to $\mathbb{N}$ that are eventually constant.