2

Strategies for Proofs

Rigour is to the mathematician what morality is to men.
– André Weil (1906–1998)

2.1 Mathematical Proofs—What They Are and Why We Need Them

Not all mathematics involves proofs. We learn a good bit of arithmetic in grade school long before we learn how to prove that the rules of arithmetic are correct. Mathematics originated in the ancient world, in various cultures, prior to the notion of proof. It was the contribution of the ancient Greeks (who, contrary to popular misconception, did not invent mathematics, nor even geometry) to bring the notion of proof into mathematics. The first use of proof is generally attributed to Thales of Miletus, who lived in the sixth century B.C.E. Euclid, who lived in Alexandria in the third century B.C.E., brought the notion of proofs based on axioms to its first peak of success. See [Hea21] for a discussion of ancient Greek mathematics.

Euclid used an axiomatic system—which is needed for proofs—in the field of geometry. Today, virtually all branches of pure mathematics are based on axiomatic systems, and work in pure mathematics involves the construction of rigorous proofs for new theorems. Much of the great mathematics of the past has been recast with a precision missing from its original treatment. Abstract algebra, for example, which received its modern form only in the last one hundred years, reconstructs the elementary algebra studied in high school in a rigorous, axiomatic fashion. A lot of applied mathematics today also has rigorous foundations (though the work of applied mathematicians, while no less challenging than pure mathematics, is not always oriented toward proofs).

Be the above as it may, the importance of proofs should be put in the proper perspective. Intuition, experimentation and even play are no less important in today's mathematical climate than rigor, because it is only by our intuition that we decide what new results to try to prove. The relation between intuition and formal rigor is not a trivial matter. Formal proofs and intuitive ideas essentially occupy different realms,

E.D. Bloch, *Proofs and Fundamentals: A First Course in Abstract Mathematics*,
Undergraduate Texts in Mathematics, DOI 10.1007/978-1-4419-7127-2_2,

and we cannot "prove" that an intuitive idea is true. Instead, there is essentially a dialectical relationship between intuition and rigor. We set up formal systems that mirror our intuition as closely as possible; we then use what we prove rigorously to further our intuitive understanding, which in turn points to new theorems requiring rigorous proofs, and so forth.

Mathematics has moved over time in the direction of ever greater rigor, though why that has happened is a question we leave to historians of mathematics to explain. We can, nonetheless, articulate a number of reasons why mathematicians today use proofs. The main reason, of course, is to be sure that something is true. Contrary to popular misconception, mathematics is not a formal game in which we derive theorems from arbitrarily chosen axioms. Rather, we discuss various types of mathematical objects, some geometric (for example, circles), some algebraic (for example, polynomials), some analytic (for example, derivatives) and the like. To understand these objects fully, we need to use both intuition and rigor. Our intuition tells us what is important, what we think might be true, what to try next and so forth. Unfortunately, mathematical objects are often so complicated or abstract that our intuition at times fails, even for the most experienced mathematicians. We use rigorous proofs to verify that a given statement that appears intuitively true is indeed true.

Another use of mathematical proofs is to explain why things are true, though not every proof does that. Some proofs tell us that certain statements are true, but shed no intuitive light on their subjects. Other proofs might help explain the ideas that underpin the result being proved; such proofs are preferable, though any proof, even if non-intuitive, is better than no proof at all. A third reason for having proofs in mathematics is pedagogical. A student (or experienced mathematician for that matter) might feel that she understands a new concept, but it is often only when attempting to construct a proof using the concept that a more thorough understanding emerges. Finally, a mathematical proof is a way of communicating to another person an idea that one person believes intuitively, but the other does not.

What does a rigorous proof consist of? The word "proof" has a different meaning in different intellectual pursuits. A "proof" in biology might consist of experimental data confirming a certain hypothesis; a "proof" in sociology or psychology might consist of the results of a survey. What is common to all forms of proof is that they are arguments that convince experienced practitioners of the given field. So too for mathematical proofs. Such proofs are, ultimately, convincing arguments that show that the desired conclusions follow logically from the given hypotheses.

There is no formal definition of proof that mathematicians use (except for mathematical logicians, when they develop formal theories of proofs, but these theories are distinct from the way mathematicians go about their daily business). Although we briefly discussed rules of inference and logical derivations in Section 1.4, what we are really interested in for the rest of this book is the way contemporary mathematicians do proofs, in order to prepare you for the kinds of proofs and basic mathematical concepts you will encounter in advanced mathematics courses.

Mathematicians who are not logicians virtually never write proofs as strings of logical symbols and rules of inference, for a number of reasons. First, and foremost, mathematical proofs are often much too long and complicated to be conve-

niently broken down into the two-column (statement-justification) format. Second, the mathematical ideas of the proof, not its logical underpinnings, are the main issue on which we want to focus, and so we do not even mention the rules of logical inference used, but rather mention only the mathematical justification of each step. Second, mathematicians who are not logicians, which means most mathematicians, find long strings of logical symbols not only unpleasant to look at, but in most cases rather difficult to follow. See [EFT94, pp. 70–71] for a fully worked out example of putting a standard mathematical proof in group theory into a two-column format using formal logic. The mathematical result proved in that example is given in Exercise 7.2.8; see Sections 7.2 and 7.3 for a brief introduction to groups. One look at the difference between the mathematicians' version of the proof and the logicians' version, in terms of both length and complexity, should suffice to convince the reader why mathematicians do things as they do.

To some extent mathematicians relate to proofs the way the general public often reacts to art—they know it when they see it. But a proof is not like a work of modern art, where self-expression and creativity are key, and all rules are to be broken, but rather like classical art that followed formal rules. (This analogy is not meant as an endorsement of the public's often negative reaction to serious modern art—classical art simply provides the analog we need here.) Also similarly to art, learning to recognize and construct rigorous mathematical proofs is accomplished not by discussing the philosophy of what constitutes a proof, but by learning the basic techniques, studying correct proofs, and, most importantly, doing lots of them. Just as art criticism is one thing and creating art is another, philosophizing about mathematics and doing mathematics are distinct activities (though of course it helps for the practitioner of each to know something about the other). For further discussion about the conceptual nature of proofs, see [Die92, Section 3.2] or [EC89, Chapter 5], and for more general discussion about mathematical activity see [Wil65] or [DHM95].

Ultimately, a mathematical proof is a convincing argument that starts from the premises, and logically deduces the desired conclusion. How someone may have thought of a proof is one thing, but the proof itself has to proceed logically from start to finish. The distinction between a valid mathematical proof itself and how it was thought of is something that is very important to keep in mind when you work on your own proofs. When solving a problem, you first try all sorts of approaches to find something that works, perhaps starting with the hypotheses and working forwards, or starting with the conclusion and working backwards, or some combination of the two. Whatever your explorations might be, a record of such exploration should never be mistaken for a final proof. Confusing the exploration with the proof is a very common mistake for students first learning advanced mathematics. We will see some examples of this distinction later on.

What is it that we prove in mathematics? We prove statements, which are usually called theorems, propositions, lemmas, corollaries and exercises. There is not much difference between these types of statements; all need proofs. Theorems tend to be important results; propositions are usually slightly less important than theorems; lemmas are statements that are used in the proofs of other results; corollaries are statements that follow easily from other results; exercises are statements that are

left to the reader to prove. When discussing proofs, we will generically refer to "theorems" when we mean any of theorems, propositions and the like.

Let us examine the statement of a very famous theorem.

Theorem 2.1.1 (Pythagorean Theorem). *Let $\triangle ABC$ be a right triangle, with sides of length a, b and c, where c is the length of the hypotenuse. Then $a^2 + b^2 = c^2$.*

When asked what the Pythagorean Theorem says, students often state "$a^2 + b^2 = c^2$." This expression alone is not the statement of the theorem—indeed, it is not a statement at all. Unless we know that a, b and c are the lengths of the sides of a right triangle, with c the length of the hypotenuse, we cannot conclude that $a^2 + b^2 = c^2$. (The formula $a^2 + b^2 = c^2$ is never true for the sides of a non-right triangle.) It is crucial to state theorems with all their hypotheses if we want to be able to prove them.

We will not give a proof of the Pythagorean Theorem; see [Loo40] for a variety of proofs. Rather, we want to consider its logical form. Although the words "if ... then" do not appear in the statement of the theorem, the statement is nonetheless a conditional statement (as discussed in Section 1.2). If we let $P =$ "a, b and c are the lengths of the sides of a right triangle, with c the length of the hypotenuse," and let $Q =$ "$a^2 + b^2 = c^2$," then the theorem has the form $P \to Q$. Many (if not all) statements of theorems are essentially conditional statements, or combinations of them, even though the words "if ... then" do not appear explicitly. A proof of a theorem is therefore an argument that shows that one thing implies another, or a combination of such arguments. It is usually much easier to formulate proofs for theorems when we recognize that they have the form $P \to Q$, even if they are not given to us in that form.

Theorems are not proved in a vacuum. To prove one theorem, we usually need to use various relevant definitions, and theorems that have already been proved. If we do not want to keep going backwards infinitely, we need to start with some objects that we use without definition, as well as some facts about these objects that are assumed without proof. Such facts are called axioms, and a body of knowledge that can be derived from a set of axioms is called an axiomatic system. In modern abstract mathematics, we take set theory as our basis for all arguments. In each branch of mathematics, we then give specific axioms for the objects being studied. For example, in abstract algebra, we study constructs such as groups, rings and fields, each of which is defined by a list of axioms; the axioms for groups are given in Section 7.2.

In Chapters 3–6 we will discuss sets, and various basic constructs using sets such as functions and relations, which together form the basis for much of modern mathematics. Our concern in the present chapter, by contrast, is not with the basis upon which we rely when we construct proofs, but rather the construction of proofs themselves. It may appear as if we are doing things backwards, in that we are not starting with what we say is the basis for modern mathematics, but we want to be able to give proofs about sets in Chapter 3, so we need to know how to write proofs before discussing set theory. As a basis for our work in the present chapter, we will make use of standard definitions and properties of the familiar number systems such as the integers, rational numbers and real numbers. We will assume that the reader is

informally familiar with these numbers. See the Appendix for a brief list of some of the standard properties of the real numbers.

We conclude this section with our first example of a proof. You are probably familiar with the statement "the sum of even numbers is even." This statement can be viewed in the form $P \to Q$ if we look at it properly, because it actually says "if n and m are even numbers, then $n+m$ is an even number." To construct a rigorous proof of our statement (as well as the corresponding result for odd numbers), we first need precise definitions of the terms involved.

Our theorem is concerned with the integers, that is, the numbers

$$\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots,$$

and so we need to assume that we know what the integers are, that we have the operations addition, subtraction, multiplication and division, and that these operations satisfy standard properties, for example the Distributive Law. Using only those standard facts about the integers, we can make the following definition, which is the basis for our theorem and its proof.

Definition 2.1.2. Let n be an integer. The number n is **even** if there is some integer k such that $n = 2k$. The number n is **odd** if there is some integer j such that $n = 2j+1$. △

As the reader knows intuitively, and as we will prove in Corollary 5.2.6, every integer is either even or odd, but not both.

We are now ready to state and prove our theorem. This result may seem rather trivial, but our point here is to see a properly done proof, not to learn an exciting new result about numbers.

Theorem 2.1.3. *Let n and m be integers.*

1. *If n and m are both even, then $n+m$ is even.*
2. *If n and m are both odd, then $n+m$ is even.*
3. *If n is even and m is odd, then $n+m$ is odd.*

Proof.

(1). Suppose that n and m are both even. Then there exist integers k and j such that $n = 2k$ and $m = 2j$. Then

$$n+m = 2k+2j = 2(k+j).$$

Because k and j are integers, so is $k+j$. Hence $m+n$ is even.

(2) & (3). These two parts are proved similarly to Part (1), and the details are left to the reader. □

There is a fourth possible case we did not state in Theorem 2.1.3, namely, the case when n is odd and m is even, because that case is really no different from Part (3) of the theorem, and hence it would not tell us anything new; it makes no difference whether we call the even number n and the odd number m, or vice versa.

The proof of Part (1) of Theorem 2.1.3 is quite simple, but there are a few features worth mentioning, because they are typical of what is found in virtually all our subsequent proofs (and in the proofs you will need to write). First, the proof relies completely on the definition of what it means to be an even or an odd integer. In a large number of proofs, going back to the formal definitions involved is the key step; forgetting to do so is a major source of error by students who are first learning about proofs.

Second, observe that the proof is written in grammatically correct English. Complete sentences are used, with proper punctuation. Each sentence begins with a capital letter, and ends with a period, even if the end of the sentence is in a displayed equation. Mathematical formulas and symbols are parts of sentences, and are treated no differently from other words. We will be writing all our proofs in this style; scratch work, by contrast, can be as careless as desired. The two-column method of writing proofs, which we used in our discussion of valid logical arguments in Section 1.4, and is often used in high school geometry, should be left behind at this point. Mathematics texts and research papers are all written in the style of Theorem 2.1.3. See Section 2.6 for more about writing mathematics.

An important consideration when writing a proof is recognizing what needs to be proved and what doesn't. There is no precise formula for such a determination, but the main factor is the context of the proof. In an advanced book on number theory, it would be unnecessary to prove the fact that the sum of two even integers is even; it would be safe to assume that the reader of such a book would either have seen the proof of this fact, or could prove it herself. For us, however, because we are just learning how to do such proofs, it is necessary to write out the proof of this fact in detail, even though we know from experience that the result is true. The reasons to prove facts that we already know are twofold: first, in order to gain practice writing proofs, we start with simple results, so that we can focus on the writing, and not on mathematical difficulties; second, there are cases where "facts" that seem obviously true turn out to be false, and the only way to be sure is to construct valid proofs.

Though mathematical proofs are logical arguments, observe that in the proof of Theorem 2.1.3 we did not use the logical symbols we discussed in Chapter 1. In general, it is not proper to use logical symbols in the writing of mathematical proofs. Logical symbols were used in Chapter 1 to help us become familiar with informal logic. When writing mathematical proofs, we make use of that informal logic, but we write using standard English (or whatever language is being used).

For the record, in the proof of Theorem 2.1.3 we did make use of some of the rules of inference discussed in Section 1.4, though as will always be the case, these rules are not mentioned explicitly in proofs to avoid unnecessary length and clutter. For instance, the hypothesis in Part (1) has the form $P \wedge Q$, where $P =$ "n is even" and $Q =$ "m is even." The proofs starts by assuming that $P \wedge Q$ is true. We then used Simplification to deduce that each of P and Q is true, so that we could apply

the definition of even numbers to each, to deduce that each of the statements "there exists an integer k such that $n = 2k$" and "there exists an integer j such that $m = 2j$" holds. We then applied Adjunction to deduce that the statement "$n = 2k$ and $m = 2j$" holds, so that we could do the calculation involving $n+m$. Finally, we made repeated use of Hypothetical Syllogism to put all the pieces of the proof together. Of course, even though mathematicians do not generally mention the rules of logical inference used in their proofs, care must be taken to ensure that the rules of inference are used correctly, even when not stated explicitly.

One final comment on writing proofs: neither thinking up proofs nor writing them properly is easy, especially as the material under consideration becomes more and more abstract. Mathematics is not a speed activity, and you should not expect to construct proofs rapidly. You will often need to do scratch work first, before writing up the actual proof. As part of the scratch work, it is very important to figure out the overall strategy for the problem being solved, prior to looking at the details. What type of proof is to be used? What definitions are involved? Not every choice of strategy ultimately works, of course, and so any approach needs to be understood as only one possible way to attempt to prove the theorem. If one approach fails, try another. Every mathematician has, in some situations, had to try many approaches to proving a theorem before finding one that works; the same is true for students of mathematics.

Exercises

Exercise 2.1.1. Reformulate each of the following theorems in the form $P \to Q$. (The statements of the theorems as given below are commonly used in mathematics courses; they are not necessarily the best possible ways to state these theorems.)

(1) The area of the region inside a circle of radius r is πr^2.

(2) Given a line l and a point P not on l, there is exactly one line m containing P that is parallel to l.

(3) Let $\triangle ABC$ be a triangle, with sides of length a, b and c. Then

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}.$$

(4) $e^{x+y} = e^x e^y$.

(5) (Fundamental Theorem of Calculus) Let f be a continuous function on $[a,b]$, and let F be any function for which $F'(x) = f(x)$. Then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

2.2 Direct Proofs

As mentioned in the previous section, the statement of virtually every theorem, when viewed appropriately, is of the form $P \to Q$, or some combination of such statements.

For example, each of the three parts of Theorem 2.1.3 is of the form $P \to Q$. To prove theorems, we therefore need to know how to prove statements of the form $P \to Q$.

The simplest form of proof, which we treat in this section, is the most obvious one: assume that P is true, and produce a series of steps, each one following from the previous ones, which eventually lead to Q. This type of proof is called a **direct proof**. That this sort of proof deserves a name is because there are other approaches that can be taken, as we will see in Section 2.3. An example of a direct proof is the proof of Theorem 2.1.3.

How do we construct direct proofs? There is no single answer to this question, but some useful strategies exist. To start, it is important to recognize that what is "direct" about a direct proof is the way the proof reads when you are done writing it. The completed proof starts at the beginning (the statement P) and ends at the end (the statement Q), and shows how to get logically from the former to the latter. How you think of the proof is another matter entirely. The way a proof looks when you are done constructing it often has little relation to how you went about thinking of it, especially for more difficult proofs. Similarly to writing a literature paper, for which you might take notes, make an outline, prepare a rough draft and revise it a number of times, so too with constructing a rigorous mathematical proof—the final version may be the result of a process involving a number of distinct steps, and much revision.

When constructing a proof, the first thing to do is specify what you are assuming, and what it is you are trying to prove. This comment may sound trivial, but the author has seen many students skip this important step in their rush to get to the details (which are usually more interesting). Then you pick a strategy for the proof; one such strategy is direct proof. The next stage is actually figuring out a proof, making use of your chosen strategy. If you cannot devise a proof using your chosen strategy, perhaps another strategy should be attempted. There is no fixed way of finding a proof; it requires experimentation, playing around and trying different things. Of course, with experience some standard ways of constructing proofs in certain familiar situations tend to suggest themselves.

Even when the chosen strategy is direct proof, there are a number of ways of trying to figure out the details of the proof. To find a direct proof of $P \to Q$, you might try assuming P, playing around with it, seeing where it leads. Or you might try looking at Q, determining what is needed to prove Q, and then what is needed to prove that, etc. Or you might do both of these, hoping to meet in the middle. However you go about working out the proof, once you understand it informally, you have only completed the "scratch work" stage of constructing the proof. Then comes the next stage, which is writing the proof in final form. No matter how convoluted a route you took in thinking up the proof, the final write-up should be direct and logical. In a direct proof, the write-up should start with P and go step by step until Q is reached. Therefore, this type of proof typically has the following form.

Proof. Suppose that P is true.

$\vdots$

(argumentation)

⋮

Then Q is true. □

We are now ready to give two simple examples of direct proof. We will put in more details here than one might normally include, in order to make each step as explicit as possible. We start with a definition concerning the integers.

Definition 2.2.1. Let a and b be integers. The number a **divides** the number b if there is some integer q such that $aq = b$. If a divides b, we write $a|b$, and we say that a is a **factor** of b, and that b is **divisible** by a. △

Before discussing the content of Definition 2.2.1, we need to make an important remark about its logical structure. The definition says that "the number a divides the number b if ...," where the ... describe a certain condition involving the numbers a and b. Strictly speaking, it would have been proper to write "if and only if" instead of just "if," because it is certainly meant to be the case that if the condition does not hold, then we do not say that a divides b. However, it is customary in definitions to write "if" rather than "if and only if," because it is taken as assumed that if the condition does not hold, then the term being defined cannot be applied. We will stick with the customary formulation of definitions, but it is important to think of definitions as meaning "if and only if."

To show the truth of a statement of the form "$a|b$," it is necessary to find an integer q such that $aq = b$. Therefore, a statement of the form "$a|b$" is an existence statement.

The expression "$a|b$" should not be confused with the fraction "a/b." The latter is a number, whereas the former is a shorthand way of writing the statement "the integer a divides the integer b." For example, even though it is not sensible to write the fraction $7/0$, it is perfectly reasonable to write the expression $7|0$, because 7 does in fact divide 0, because $7 \cdot 0 = 0$. Because of this potential confusion, and also to avoid ambiguous expressions such as $1/2+3$ (is that $\frac{1}{2} + 3$ or $\frac{1}{2+3}$?), we suggest writing all fractions as $\frac{a}{b}$ rather than a/b.

We now have two simple results about divisibility. The proof of each theorem is preceded by scratch work, to show how one might go about formulating such a proof.

Theorem 2.2.2. *Let a, b and c be integers. If $a|b$ and $b|c$, then $a|c$.*

Scratch Work. Our goal is to show that $a|c$, so that we need to find some integer k such that $ak = c$. We are free to choose any k that we can think of. Because $a|b$ and $b|c$, there are integers q and r such that $aq = b$ and $br = c$. Substituting the first equation into the second equation looks like a good idea to try, and we obtain $(aq)r = c$. By rearranging the left-hand side of this equation, we see that $k = qr$ is a good guess. ///

Proof. Suppose that $a|b$ and $b|c$. Hence there are integers q and r such that $aq = b$ and $br = c$. Define the integer k by $k = qr$. Then $ak = a(qr) = (aq)r = br = c$. Because $ak = c$, it follows that $a|c$. □

Compare the proof with the scratch work. The proof might not appear substantially better than the scratch work at first glance, and it might even seem a bit mysterious to someone who had not done the scratch work. Nonetheless, the proof is better than the scratch work, though in such a simple case the advantage might not be readily apparent. Unlike the scratch work, the proof starts with the hypotheses and proceeds logically to the conclusion, using the definition of divisibility precisely as stated. Later on we will see examples where the scratch work and the proof are more strikingly different.

Theorem 2.2.3. *Any integer divides zero.*

Scratch Work. In the statement of this theorem we are not given any particular choices of "variables," in contrast to the previous theorem (which was stated in terms of a, b and c). To prove something about any possible integer, we pick an arbitrary one, say n. Then we need to show that $n|0$. It would certainly not suffice to choose one particular number, say 5, and then show that 5 divides 0. Once we have chosen an arbitrary n, the rest of the details in this proof are extremely simple. ///

Proof. Let n be an integer. Observe that $n \cdot 0 = 0$. Hence $n|0$. □

The first step in proving a theorem often involves reformulating it in a more useful way, such as choosing n in the above proof.

The reader might be concerned that, in comparison to the scratch work for the above two theorems, the way we wrote the proofs involves "covering up our tracks." Although it might appear that way, the purpose of the proper writing of proofs is not at all to hide anything, but rather to make sure that what seemed like a good idea intuitively is indeed logical. The only way to check whether a proof is really valid is to write it up properly, and such a write-up does not include a description of everything that went through your mind when you were figuring out the details of the proof. The final proof must stand on its own, with no reference to what was written in the scratch work. For example, not all arguments are reversible, and an argument that worked backwards during scratch work might not work when written forwards, and it is only by writing the proof properly that we find out if the idea really works. Intuitive thinking that may have been useful in formulating the proof should be replaced with logical deduction in the final written proof.

In sum, there are two main steps to the process of producing a rigorous proof: formulating it and writing it. These two activities are quite distinct, though in some very simple and straightforward proofs you might formulate as you write. In most cases, you first formulate the proof (at least in outline form) prior to writing. For a difficult proof the relation between formulating and writing is essentially dialectical. You might formulate a tentative proof, try writing it up, discover some flaws, go back to the formulating stage and so on.

Exercises

Exercise 2.2.1. Outline the strategy for a direct proof of each of the following statements (do not prove them, because the terms are meaningless).

(1) Let n be an integer. If $7|n$, then n is bulbous.
(2) Every globular integer is even.
(3) If an integer is divisible by 13 and is greater than 100, then it is pesky.
(4) An integer is both tactile and filigreed whenever it is odd.

Exercise 2.2.2. Let n and m be integers.

(1) Prove that $1|n$.
(2) Prove that $n|n$.
(3) Prove that if $m|n$, then $m|(-n)$.

Exercise 2.2.3. Let n be an integer.

(1) Prove that if n is even, then $3n$ is even.
(2) Prove that if n is odd, then $3n$ is odd.

Exercise 2.2.4. **[Used in Theorem 2.3.5 and Theorem 2.4.1.]** Let n be an integer. Prove that if n is even then n^2 is even, and if n is odd then n^2 is odd.

Exercise 2.2.5. Let n and m be integers. Suppose that n and m are divisible by 3.

(1) Prove that $n+m$ is divisible by 3.
(2) Prove that nm is divisible by 3.

Exercise 2.2.6. Let a, b, c, m and n be integers. Prove that if $a|b$ and $a|c$, then $a|(bm+cn)$.

Exercise 2.2.7. Let a, b, c and d be integers. Prove that if $a|b$ and $c|d$, then $ac|bd$.

Exercise 2.2.8. Let a and b be integers. Prove that if $a|b$, then $a^n|b^n$ for all positive integers n. (There is no need for mathematical induction here.)

2.3 Proofs by Contrapositive and Contradiction

In this section we discuss two strategies for proving statements of the form $P \to Q$. Both these strategies are a bit more convoluted than direct proof, but in some situations they are nonetheless easier to work with. A less than perfect analogy might be when the straightest road between two cities leads up and down a mountain and through difficult terrain, whereas a curved road might at first seem to be going in the wrong direction, but in fact it bypasses the mountain and is ultimately easier and quicker than the straight road.

There is no foolproof method for knowing ahead of time whether a proof on which you are working should be a direct proof or a proof by one of these other methods. Experience often allows for an educated guess as to which strategy to try first. In any case, if one strategy does not appear to bear fruit, then another strategy should be attempted. It is only when the proof is completed that we know whether a given choice of strategy works.

Both strategies discussed in this section rely on ideas from our discussion of equivalence of statements in Section 1.3. For our first method, recall that the contrapositive of $P \to Q$, the statement $\neg Q \to \neg P$, is equivalent to $P \to Q$. Hence, in order

to prove $P \to Q$, we could just as well prove $\neg Q \to \neg P$, which we would do by the method of direct proof. We construct such a proof by assuming that Q is false, and then, in the final write-up, presenting a step-by-step argument going from $\neg Q$ to $\neg P$. A proof of this sort is called **proof by contrapositive**. This type of proof typically has the following form.

Proof. Suppose that Q is false.

$\vdots$

(argumentation)

$\vdots$

Then P is false. □

The following proof is a simple example of proof by contrapositive.

Theorem 2.3.1. *Let n be an integer. If n^2 is odd, then n is odd.*

Scratch Work. If we wanted to use a direct proof, we would have to start with the assumption that n^2 is odd. Then there would be some integer j such that $n^2 = 2j+1$. It is not clear, however, how to proceed from this point, so instead we try proof by contrapositive. Such a proof would involve assuming that n is not odd, which implies that it is even, and then deducing that n^2 is even, which implies that it is not odd. We start such a proof by observing that if n is even, then there is some integer k such that $n = 2k$, and we then compute n^2 in terms of k, leading to the desired result. ///

Proof. Suppose that n is even. Then there is some integer k such that $n = 2k$. Hence $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$. Because $2k^2$ is an integer, it follows that n^2 is even. By contrapositive, we see that if n^2 is odd then n is odd. □

In the above proof we mentioned that we used proof by contrapositive. In general, it is often helpful to the reader to have the method of proof stated explicitly.

Another method of proof for theorems with statements of the form $P \to Q$, which looks similar to proof by contrapositive but is actually distinct from it, is **proof by contradiction**.

Logicians use the term "proof by contradiction" to mean the proof of a statement A by assuming $\neg A$, then reaching a contradiction, and then deducing that A must be true. For our purposes, we are interested in proof by contradiction for the special case where the statement A has the form $P \to Q$, because that is how mathematical theorems are formulated. We now take a closer look at this particular type of proof by contradiction.

Recall from Section 1.3 that $\neg(P \to Q)$ is equivalent to $P \wedge \neg Q$. Suppose that we could prove that $P \wedge \neg Q$ is false. It would follow that $\neg(P \to Q)$ is false, and hence that $\neg(\neg(P \to Q))$ is true. Then, using Double Negation (Fact 1.3.2 (1)), we could conclude that $P \to Q$ is true.

The method of proof by contradiction is to show that $P \to Q$ is true by assuming that $P \wedge \neg Q$ is true, and then deriving a logical contradiction, by which we mean, as discussed in Section 1.2, a statement that cannot be true under any circumstances;

often such statements have the form $B \wedge \neg B$ for some statement B. Once we reach a contradiction, we conclude that $P \wedge \neg Q$ is false, and then as above we deduce that $P \rightarrow Q$ is true.

Another way to think of proof by contradiction is to observe from the truth table for $P \rightarrow Q$ that the only way for this statement to be false is if P is true and Q is false, that is, if P is true and $\neg Q$ is true. Hence, if we assume both of these, and then derive a contradiction, we would know that $P \rightarrow Q$ cannot be false; hence $P \rightarrow Q$ must be true.

A proof by contradiction typically has the following form.

Proof. We prove the result by contradiction. Suppose that P is true and that Q is false.

$\vdots$

(argumentation)

$\vdots$

We have therefore reached a contradiction. Therefore P implies Q. □

We now turn to a simple example of proof by contradiction. It is a good idea to start such a proof by stating that you are using this strategy.

Theorem 2.3.2. *The only consecutive non-negative integers a, b and c that satisfy* $a^2 + b^2 = c^2$ *are* 3, 4 *and* 5.

Scratch Work. The statement of this theorem has the form $P \rightarrow Q$, because it can be restated as "if a, b and c are consecutive non-negative integers such that $a^2 + b^2 = c^2$, then a, b and c are 3, 4 and 5." It is hard to prove the result directly, because we are trying to prove that something does not exist. Rather, we will assume that consecutive integers a, b and c, other than 3, 4 and 5, exist and satisfy $a^2 + b^2 = c^2$, and we will then derive a contradiction. Also, we observe that if a, b and c are consecutive integers, then $b = a + 1$ and $c = a + 2$. ///

Proof. We prove the result by contradiction. Suppose that a, b and c are non-negative consecutive integers other than 3, 4 and 5, and that $a^2 + b^2 = c^2$. Because a, b and c are not 3, 4 and 5, we know that $a \neq 3$, and because the three numbers are consecutive, we know that $b = a + 1$ and $c = a + 2$. From $a^2 + b^2 = c^2$ we deduce that $a^2 + (a+1)^2 = (a+2)^2$. After expanding and rearranging we obtain $a^2 - 2a - 3 = 0$. This equation factors as $(a-3)(a+1) = 0$. Hence $a = 3$ or $a = -1$. We have already remarked that $a \neq 3$, and we know a is non-negative. Therefore we have a contradiction, and the theorem is proved. □

Our next two theorems are both famous results that have well-known proofs by contradiction. These clever proofs are much more difficult than what we have seen so far, and are more than would be expected of a student to figure out on her own at this point.

Our first result involves irrational numbers, which we will shortly define. Irrational numbers are a type of real number, and so we need to assume informal knowledge of the real numbers, just as we assumed informal knowledge of the integers

in Sections 2.1 and 2.2. The real numbers are the collection of all the numbers that are generally used in elementary mathematics (not including the complex numbers), and they have operations addition, subtraction, multiplication and division, and these operations satisfy standard properties such as the Commutative Law for addition and multiplication. See the Appendix for a brief summary of some of the standard properties of real numbers. We now turn to the matter at hand.

Definition 2.3.3. Let x be a real number. The number x is a **rational number** if there exist integers n and m such that $m \neq 0$ and $x = \frac{n}{m}$. If x is not a rational number, it is an **irrational number**. △

Observe that if x is a rational number, then there are many different fractions of the form $\frac{n}{m}$ such that $x = \frac{n}{m}$. Given any fraction $\frac{n}{m}$ such that $n \neq 0$, we can always reduce it to "lowest terms," by which we mean that the numerator and denominator have no common factors other than 1 and -1. See the Appendix for a reference, where this fact about rational numbers is stated as Theorem A.6.

Are there any irrational numbers? Though it is not at all obvious, there are in fact infinitely many of them, and in a certain sense there are more irrational numbers than rational ones, as will be made precise in Section 6.7.

At this point, however, we will have to be satisfied with verifying that irrational numbers exist. In particular, we will prove that $\sqrt{2}$ is an irrational number. To us this fact may seem rather innocuous, though when first discovered it was something of a shock. The result was discovered by someone in the Pythagorean school in ancient Greece (possibly the sixth century B.C.E.). This school, centered around the figure of Pythagoras, was dedicated to mathematics as well as various mystical beliefs. Among other things, the Pythagoreans believed in the importance of whole numbers, and held that anything meaningful in the universe could be related to whole numbers or to ratios of whole numbers. The ancient Greeks tended to think of numbers geometrically, and they probably did not think of $\sqrt{2}$ as an algebraically defined object, as we do today. However, by using the Pythagorean Theorem, we see that if a square has sides of length 1, then the diagonal of the square will have length $\sqrt{2}$. Hence $\sqrt{2}$ would be a geometrically meaningful number to the Pythagoreans, and therefore they were very disturbed to discover that this number was not expressible as a ratio of whole numbers. One legend has it that the discoverer of this fact, in despair, threw himself overboard from a ship.

Before we state and prove our theorem about $\sqrt{2}$, we need a proper definition for this number.

Definition 2.3.4. Let p be a positive real number. The **square root** of p, denoted $\sqrt{p}$, is a positive real number x such that $x^2 = p$. △

Our goal is to prove that $\sqrt{2}$ is an irrational number, but there is a more fundamental question about $\sqrt{2}$ that needs to be addressed first, which is whether it exists. Definition 2.3.4 states that if there is a number denoted $\sqrt{2}$, it would be a positive real number x such that $x^2 = 2$, but nothing in the definition guarantees that such a number x exists. Clearly, if there is no such real number x, it would make no sense to try to prove that such a number is irrational. In fact, as expected, it is indeed true

that there is a positive real number x such that $x^2 = 2$ (and there is only one such number), but unfortunately it is beyond the scope of this book to give a proof of that fact. The proof requires tools from real analysis; see [Blo11, Theorem 2.6.9] for a proof.

Assuming that $\sqrt{2}$ exists, however, we can prove here that this number is irrational. Observe that the following theorem is self-contained, and does not rely upon a proof that $\sqrt{2}$ exists; it only says that if $\sqrt{2}$ exists, then it is irrational.

Theorem 2.3.5. *There is no rational number x such that $x^2 = 2$.*

Preliminary Analysis. The statement of our theorem says that something does not exist, which is hard to prove directly. However, we can easily reformulate the statement to avoid that problem, because to say that there is no rational number with a certain property means that if a real number has that property, that number cannot be rational. That is, we can reformulate our theorem as "if x is a real number and $x^2 = 2$, then x is irrational," which has the familiar form $P \to Q$. We then use proof by contradiction, which we start by assuming that x is a real number such that $x^2 = 2$, and also that x is not irrational (and hence it is rational). ///

Proof. Let x be a real number. Suppose that $x^2 = 2$, and that x is rational. We will derive a contradiction. Because x is rational, there are integers n and m such that $x = \frac{n}{m}$. Observe that $n \neq 0$. If $\frac{n}{m}$ is not in lowest terms, then we could cancel any common factors, bringing it to lowest terms. There is no problem assuming that this has been done already, and so we may assume that n and m have no common factors other than 1 and -1.

Because $x^2 = 2$, then $\left(\frac{n}{m}\right)^2 = 2$. It follows that $\frac{n^2}{m^2} = 2$, and hence $n^2 = 2m^2$. We now ask whether n is even or odd. If n were odd, then using Exercise 2.2.4 we would see that n^2 would be odd. This last statement is not possible, because $n^2 = 2m^2$, and $2m^2$ must be even, because it is divisible by 2. It follows that n cannot be odd; hence n must be even. Therefore there is some integer k such that $n = 2k$. Then $(2k)^2 = 2m^2$, so that $4k^2 = 2m^2$, and therefore $2k^2 = m^2$. By an argument similar to the one used above, we see that m is even. We therefore conclude that both n and m are even. We have therefore reached a contradiction, because any two even numbers have 2 as a common factor, and yet we assumed that n and m have no common factors other than 1 and -1. Hence x is not rational. □

The proof of Theorem 2.3.5 is mentioned (without details) in Aristotle's "Prior Analytics" (I.23), and is presumed to be of earlier origin; perhaps it is the proof used by the Pythagoreans (though they would not have formulated it as we do).

Our second famous result involves prime numbers, and has a proof by contradiction for a subpart of a proof by contradiction. We will make use of the definition of divisibility given in Section 2.2.

Definition 2.3.6. Let p be an integer greater than 1. The number p is a **prime number** if the only positive integers that divide p are 1 and p. The number p is a **composite number** if it is not a prime number. △

The first few prime numbers are $2, 3, 5, 7, 11, \ldots$. The study of prime numbers is quite old and very extensive; see any book on elementary number theory, for example [Ros05], for details.

The number 1 is not considered to be either prime or composite. On the one hand, the only positive integers that divide 1 are 1 and itself, which would make it seem as if 1 were a prime number. However, the prime numbers are always defined as being 2 or larger to avoid special cases and awkward statements of theorems. For example, if 1 were a prime number, then the factorization of integers into prime numbers would not be unique, and uniqueness would hold only for "factorization into prime numbers other than 1," which is cumbersome to state. On the other hand, the number 1 is not considered composite, because there are no positive integers other than 1 and itself that divide it.

Whereas we restrict our attention to the integers greater than 1 when we discuss prime numbers and composite numbers, some authors consider negative numbers such as $-2, -3, -5, \ldots$ to be prime numbers, and similarly for composite numbers. Moreover, the term "prime" is used in the more general context of rings, a structure that is studied in abstract algebra, and that includes the integers as a special case; see any introductory abstract algebra text, for example [Fra03], for details.

Observe that a composite number n can always be written as $n = ab$ for some positive integers a and b such that $1 < a < n$ and $1 < b < n$.

How many prime numbers are there? In particular, are there only finitely many prime numbers, or infinitely many? The following theorem answers this question. The proof we give is very commonly used, and goes back to Euclid; see [Rib96, Chapter 1] for further discussion, as well as some other nice proofs of this theorem.

Theorem 2.3.7. *There are infinitely many prime numbers.*

Preliminary Analysis. We have not yet seen a rigorous treatment of what it means for there to be infinitely many of something, and so for now we need to use this concept in an intuitive fashion. A thorough discussion of finite vs. infinite is found in Chapter 6. The essential idea discussed in that chapter is that if a collection of objects can be listed in the form $a_1, a_2, \ldots, a_n$ for some positive integer n, then the collection of objects is finite; if the collection of objects cannot be described by any such list, then it is infinite. In Chapter 6 we will see a rigorous formulation of this idea in terms of sets and functions, but this intuitive explanation of finite vs. infinite completely captures the rigorous definition.

To say that there are infinitely many prime numbers means that there is no list of the form $P_1, P_2, \ldots, P_n$, for any positive integer n, that contains all prime numbers. It is easier to prove this statement if we reformulate it as "if n is a positive integer, and $P_1, P_2, \ldots, P_n$ are prime numbers, then $P_1, P_2, \ldots, P_n$ does not include all prime numbers." The proof of this last statement is by contradiction. ///

Proof. Let n be a positive integer, and let $P_1, P_2, \ldots, P_n$ be a collection of prime numbers. Suppose that $P_1, P_2, \ldots, P_n$ contains all prime numbers.

Let $Q = (P_1 \times P_2 \times \cdots \times P_n) + 1$. We will show that Q is a prime number. Because Q is clearly larger than any of the numbers $P_1, P_2, \ldots, P_n$, it will follow that Q is a prime number that is not in the collection $P_1, P_2, \ldots, P_n$, and we will therefore know that the collection $P_1, P_2, \ldots, P_n$ does not contain all prime numbers, which is a contradiction. It will then follow that if n is a positive integer, and $P_1, P_2, \ldots, P_n$ are prime numbers, then $P_1, P_2, \ldots, P_n$ does not include all prime numbers, and we will conclude that there are infinitely many prime numbers.

To show that Q is a prime number, we use proof by contradiction. Suppose that Q is not a prime number. Therefore Q is a composite number. By Theorem 6.3.10 we deduce that Q has a factor that is a prime number. (Though this theorem comes later in the text, because it needs some tools we have not yet developed, it does not use the result we are now proving, and so it is safe to use.) The only prime numbers are $P_1, P_2, \ldots, P_n$, and therefore one of these numbers must be a factor of Q. Suppose that P_k is a factor of Q, for some integer k such that $1 \leq k \leq n$. Therefore there is some integer R such that $P_k R = Q$. Hence

$$P_k R = (P_1 \times P_2 \times \cdots \times P_n) + 1,$$

and therefore

$$P_k[R - (P_1 \times \cdots \times P_{k-1} \times P_{k+1} \times \cdots \times P_n)] = 1.$$

It follows that P_k divides 1. However, the only integers that divide 1 are 1 and -1. (We will not provide a proof of this last fact; it is stated as Theorem A.4 in the Appendix.) Because P_k is a prime number it cannot possibly equal 1 or -1, which is a contradiction. We deduce that Q is not a composite number, and hence it is a prime number. □

The proof of Theorem 2.3.7 actually yields more than just what the statement of the theorem says; it in fact gives an explicit procedure for producing arbitrarily many prime numbers. We start by letting $P_1 = 2$, which is the smallest prime number. We then let $P_2 = P_1 + 1 = 3$, and then $P_3 = (P_1 \times P_2) + 1 = 7$, and then $P_4 = (P_1 \times P_2 \times P_3) + 1 = 43$, and so on. We could continue this process indefinitely, producing as many prime numbers as we liked. This process is not entirely satisfying, however, both because it does not yield a simple explicit formula for P_n as a function of n, and also because this process skips over many prime numbers. In fact, no one has yet found a simple procedure to produce all prime numbers.

We conclude this section with the observation that proof by contradiction implicitly uses Double Negation, which ultimately relies upon the Law of the Excluded Middle, which says that any statement is either true or false. (See Section 1.2 for more discussion of this issue.) Any mathematician who does not believe in the Law of the Excluded Middle would therefore object to proof by contradiction. There are such mathematicians, though the majority of mathematicians, including the author of this book, are quite comfortable with the Law of the Excluded Middle, and hence with proof by contradiction.

Exercises

Exercise 2.3.1. For each of the statements in Exercise 2.2.1, outline the strategy for a proof by contrapositive, and the strategy for a proof by contradiction (do not prove the statements, because the terms are meaningless).

Exercise 2.3.2. Let n be an integer. Prove that if n^2 is even, then n is even.

Exercise 2.3.3. Let a, b and c be integers. Prove that if a does not divide bc, then a does not divide b.

Exercise 2.3.4. **[Used in Theorem 6.7.4.]** Prove that the product of a non-zero rational number and an irrational number is irrational.

Exercise 2.3.5. Let a, b and c be integers. Suppose that there is an integer d such that $d|a$ and $d|b$, but that d does not divide c. Prove that the equation $ax+by=c$ has no solution such that x and y are integers.

Exercise 2.3.6. Let c be an integer. Suppose that $c \geq 2$, and that c is not a prime number. Prove that there is an integer b such that $b \geq 2$, that $b|c$ and that $b \leq \sqrt{c}$.

Exercise 2.3.7. Let q be an integer. Suppose that $q \geq 2$, and that for any integers a and b, if $q|ab$ then $q|a$ or $q|b$. Prove that $\sqrt{q}$ is irrational.

Exercise 2.3.8. Let q be an integer. Suppose that $q \geq 2$, and that for any integers a and b, if $q|ab$ then $q|a$ or $q|b$. Prove that q is a prime number. (The converse to this statement is also true, though it is harder to prove; see [Dea66, Section 3.6] for details, though note that his use of the term "prime," while keeping with the standard usage in ring theory, is not the same as ours.)

2.4 Cases, and If and Only If

The notion of equivalence of statements, as discussed in Section 1.3, has already been seen to be useful in proving theorems, for example in proof by contrapositive. In this section we will make use of some other equivalences of statements to prove certain types of theorems.

One commonly used method for proving a statement of the form $P \to Q$ is by breaking up the proof into a number of cases (and possibly subcases, subsubcases and so on). Formally, we use proof by cases when the premise P can be written in the form $A \vee B$. We then use Exercise 1.3.2 (6) to see that $(A \vee B) \to Q$ is equivalent to $(A \to Q) \wedge (B \to Q)$. Hence, in order to prove that a statement of the form $(A \vee B) \to Q$ is true, it is sufficient to prove that each of the statements $A \to Q$ and $B \to Q$ is true. The use of this strategy often occurs when proving a statement involving a quantifier of the form "for all x in U," and where no single proof can be found for all such x, but where U can be divided up into two or more parts, and where a proof can be found for each part.

For the following simple example of proof by cases, recall the definition of even and odd integers in Section 2.1.

Theorem 2.4.1. *Let n be an integer. Then n^2+n is even.*

Preliminary Analysis. Because we know about sums and products of even numbers and odd numbers, it seems like a good idea to try breaking up the proof into two cases, one case where n is even and one case where n is odd. Formally, let $A =$ "n is an even integer," let $B =$ "n is an odd integer" and let $Q =$ "n^2+n is even." Then the theorem has the form $(A \vee B) \to Q$. We will prove the theorem by proving that $(A \to Q)$ and $(B \to Q)$ are both true; each of these statements will be proved as a separate case. The proof of this theorem could be done either by making use of Theorem 2.1.3 and Exercise 2.2.4, or from scratch; because the latter is simple enough, we will do that. ///

Proof. Case 1: Suppose that n is even. By definition we know that there is some integer k such that $n = 2k$. Hence

$$n^2+n = (2k)^2+2k = 4k^2+2k = 2(2k^2+k).$$

Because k is an integer, so is $2k^2+k$. Therefore n^2+n is even.

Case 2: Suppose that n is odd. By definition we know that there is some integer j such that $n = 2j+1$. Hence

$$\begin{aligned} n^2+n &= (2j+1)^2+(2j+1) = (4j^2+4j+1)+(2j+1) \\ &= 4j^2+6j+2 = 2(2j^2+3j+1). \end{aligned}$$

Because j is an integer so is $2j^2+3j+1$. Therefore n^2+n is even. □

It is not really necessary to define A and B explicitly as we did in the scratch work for Theorem 2.4.1, and we will not do so in the future, but it was worthwhile doing it once, just to see how the equivalence of statements is being used.

In the proof of Theorem 2.4.1 we had two cases, which together covered all possibilities, and which were exclusive of each other. It is certainly possible to have more than two cases, and it is also possible to have non-exclusive cases; all that is needed is that all the cases combined cover all possibilities. The proof of Theorem 2.4.4 below has two non-exclusive cases.

We now turn to theorems that have statements of the form $P \to (A \vee B)$. Such theorems are less common than the previously discussed type, but do occur, and it is worth being familiar with the standard proof strategies for such theorems. There are two commonly used strategies, each one being advantageous in certain situations. One approach would be to use the contrapositive together with De Morgan's Law (Fact 1.3.2 (13)), which together imply that $P \to (A \vee B)$ is equivalent to $(\neg A \wedge \neg B) \to \neg P$. The other would be to use Exercise 1.3.2 (5), which says that $P \to (A \vee B)$ is equivalent to $(P \wedge \neg A) \to B$. The roles of A and B could also be interchanged in this last statement. The second approach is more commonly used, and so we use it in the following proof, although in this particular case the first approach would work quite easily, as the reader should verify.

Theorem 2.4.2. *Let x and y be real numbers. If xy is irrational, then x or y is irrational.*

Preliminary Analysis. The statement of this theorem has the form $P \to (A \vee B)$. We will prove $(P \wedge \neg A) \to B$, which we do by assuming that xy is irrational and that x is rational, and deducing that y is irrational. ///

Proof. Suppose that xy is irrational and that x is rational. Hence $x = \frac{a}{b}$ for some integers a and b such that $b \neq 0$. We will show that y is irrational, by using proof by contradiction. Suppose that y is rational. It follows that $y = \frac{m}{n}$ for some integers m and n such that $n \neq 0$. Hence $xy = \frac{am}{bn}$, and $bn \neq 0$, contradicting the fact that xy is irrational. Hence y is irrational. □

Having discussed the appearance of $\vee$ in the statements of theorems, we could also consider the appearance of $\wedge$, though these occurrences are more straightforward. As expected, a theorem with statement of the form $(A \wedge B) \to Q$ is proved by assuming A and B, and using both of these statements to derive Q. To prove a theorem with statement of the form $P \to (A \wedge B)$, we can use Exercise 1.3.2 (4), which states that $P \to (A \wedge B)$ is equivalent to $(P \to A) \wedge (P \to B)$. Hence, to prove a theorem with statement of the form $P \to (A \wedge B)$, we simply prove each of $P \to A$ and $P \to B$, again as expected.

Not only are there a variety of ways to structure proofs, but there are also variants in the logical form of the statements of theorems. Whereas the most common logical form of the statement of a theorem is $P \to Q$, as we have discussed so far, another common form is $P \leftrightarrow Q$. We refer to such theorems as "if and only if" theorems (often abbreviated "iff" theorems). To prove such a theorem, we make use of the fact that $P \leftrightarrow Q$ is equivalent to $(P \to Q) \wedge (Q \to P)$, as was shown in Fact 1.3.2 (11). Hence, to prove a single statement of the form $P \leftrightarrow Q$, it is sufficient to prove the two statements $P \to Q$ and $Q \to P$, each of which can be proved using any of the methods we have seen so far. We now give a typical example of such a proof; it is sufficiently straightforward so that we dispense with the scratch work. Recall the definition of divisibility of integers in Section 2.2.

Theorem 2.4.3. *Let a and b be non-zero integers. Then $a|b$ and $b|a$ if and only if $a = b$ or $a = -b$.*

Proof.

$\Rightarrow$**.** Suppose that $a|b$ and $b|a$. Because $a|b$, there is some integer m such that $am = b$, and because $b|a$, there is some integer k such that $bk = a$. Substituting this last equation into the previous one, we obtain $(bk)m = b$, and hence $b(km) = b$. Because $b \neq 0$, it follows that $km = 1$. Because k and m are integers, then either $k = 1$ and $m = 1$, or $k = -1$ and $m = -1$. (We will not provide a proof of this last fact; it is stated as Theorem A.4 in the Appendix.) In the former case $a = b$, and in the latter case $a = -b$.

$\Leftarrow$**.** Suppose that $a = b$ or $a = -b$. First, suppose that $a = b$. Then $a \cdot 1 = b$, so $a|b$, and $b \cdot 1 = a$, so $b|a$. Similarly, suppose that $a = -b$. Then $a \cdot (-1) = b$, so $a|b$, and $b \cdot (-1) = a$, so $b|a$. □

Our next example of an if and only if theorem combines a number of the methods we have discussed so far.

Theorem 2.4.4. *Let m and n be integers. Then mn is odd if and only if both m and n are odd.*

Scratch Work. The "$\Leftarrow$" part of this theorem, which is the "if" part, says that if m and n are both odd, then mn is odd. This implication will be straightforward to prove, using the definition of odd integers.

The "$\Rightarrow$" part of this theorem, which is the "only if" part, says that if mn is odd, then both m and n are odd. A direct proof of this part of the theorem would start with the assumption that mn is odd, which would mean that $mn = 2p+1$ for some integer p, but it is not clear how to go from there to the desired conclusion. It is easier to make assumptions about m and n and proceed from there, so we will prove this part of the theorem by contrapositive, in which case we assume that m and n are not both odd, and deduce that mn is not odd. When we assume that m and n are not both odd, we will have two (overlapping) cases to consider, namely, when m is even or when n is even. Alternatively, it would be possible to make use of three non-overlapping cases, which are when m is even and n is odd, when m is odd and n is even, and when m and n are both even; however, the proof is no simpler as a result of the non-overlapping cases, and in fact the proof would be longer with these three cases rather than the two overlapping ones as originally proposed, and so we will stick with the latter. ///

Proof.

$\Leftarrow$**.** Suppose that m and n are both odd. Hence there is an integer j such that $m = 2j+1$, and there is an integer k such that $n = 2k+1$. Therefore

$$mn = (2j+1)(2k+1) = 4jk+2j+2k+1 = 2(2jk+j+k)+1.$$

Because k and j are integers, so is $2jk+j+k$. Therefore mn is odd.

$\Rightarrow$**.** Suppose that m and n are not both odd. We will deduce that mn is not odd, and the desired result will follow by contrapositive. If m and n are not both odd, then at least one of them is even. Suppose first that m is even. Then there is an integer p such that $m = 2p$. Hence $mn = (2p)n = 2(pn)$. Because p and n are integers, so is pn. Therefore mn is even. Next assume that n is even. The proof in this case is similar to the previous case, and we omit the details. □

A slightly more built-up version of an if and only if theorem is a theorem that states that three or more statements are all mutually equivalent. Such theorems often include the phrase "the following are equivalent," sometimes abbreviated "TFAE." The following theorem, which involves 2×2 matrices, is an example of this type of result. For the reader who is not familiar with matrices, we summarize the relevant notation. A 2×2 matrix is a square array of numbers of the form $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, for some real numbers a, b, c and d. The determinant of such a matrix is defined by $\det M = ad - bc$, and the trace of the matrix is defined by $\mathrm{tr}\, M = a + d$. An upper

triangular 2×2 matrix has the form $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, for some real numbers a, b and d. See any introductory text on linear algebra, for example [AR05, Chapters 1 and 2], for the relevant information about matrices.

Theorem 2.4.5. *Let $M = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ be an upper triangular 2×2 matrix. Suppose that a, b and d are integers. The following are equivalent.*

a. $\det M = 1$.
b. $a = d = \pm 1$.
c. $\operatorname{tr} M = \pm 2$ *and* $a = d$.

What Theorem 2.4.5 says is that (a) if and only if (b), that (a) if and only if (c), and that (b) if and only if (c). Hence, to prove these three if and only if statements we would in principle need to prove that (a) $\Rightarrow$ (b), that (b) $\Rightarrow$ (a), that (a) $\Rightarrow$ (c), that (c) $\Rightarrow$ (a), that (b) $\Rightarrow$ (c), and that (c) $\Rightarrow$ (b). In practice we do not always need to prove six separate statements. The idea is to use the transitivity of logical implication, which follows from Fact 1.3.1 (12). For example, suppose that we could prove that (a) $\Rightarrow$ (b), that (b) $\Rightarrow$ (c), and that (c) $\Rightarrow$ (a); the other three implications would then hold automatically. We could just as well prove that (a) $\Rightarrow$ (c), that (c) $\Rightarrow$ (b), and that (b) $\Rightarrow$ (a), if that were easier. Another way to prove the theorem would be to prove that (a) $\Rightarrow$ (b), that (b) $\Rightarrow$ (a), that (a) $\Rightarrow$ (c), and that (c) $\Rightarrow$ (a). It is sufficient to prove any collection of logical implications from which the remaining logical implications can be deduced using transitivity; the choice of what to prove and what to deduce depends upon the particular theorem being proved. Similar reasoning holds when more than three statements are being proved equivalent.

Proof of Theorem 2.4.5. We will prove that (a) $\Rightarrow$ (b), that (b) $\Rightarrow$ (c), and that (c) $\Rightarrow$ (a).

(a) $\Rightarrow$ (b). Suppose that $\det M = 1$. Hence $ad - b \cdot 0 = 1$, and therefore $ad = 1$. Because both a and d are integers, it must be the case that either $a = 1$ and $d = 1$, or $a = -1$ and $d = -1$, using Theorem A.4.

(b) $\Rightarrow$ (c). Suppose that $a = d = \pm 1$. First, suppose that $a = d = 1$. Then $\operatorname{tr} M = a + d = 2$. Second, suppose that $a = d = -1$. Then $\operatorname{tr} M = a + d = -2$. Hence $\operatorname{tr} M = \pm 2$ and $a = d$.

(c) $\Rightarrow$ (a). Suppose that $\operatorname{tr} M = \pm 2$ and $a = d$. We can rewrite $\operatorname{tr} M = \pm 2$ as $a + d = \pm 2$. Hence $4 = (a+d)^2 = a^2 + 2ad + d^2$. Because $a = d$, then $a^2 = ad = d^2$, and therefore $4 = 4ad$. It follows that $ad = 1$. Because $\det M = ad - b \cdot 0 = ad$, we deduce that $\det M = 1$. □

Exercises

Exercise 2.4.1. Outline the strategy for a proof of each of the following statements (do not prove them, because the terms are meaningless).

(1) If an integer is combustible then it is even or prime.
(2) A 2×2 matrix is collapsible if and only if its determinant is greater than 3.

(3) For an integer to be putrid, it is necessary and sufficient that it is both odd and divisible by 50.
(4) Let n be an integer. The following are equivalent: (a) the integer n is composite and greater than 8; (b) the integer n is suggestive; (c) the integer n is indifferent or fragile.

Exercise 2.4.2. Let a, b and c be integers. Suppose that $c \neq 0$. Prove that $a|b$ if and only if $ac|bc$.

Exercise 2.4.3. **[Used in Exercise 4.4.8, Exercise 6.7.9 and Section 8.8.]** Let a and b be integers. The numbers a and b are **relatively prime** if the following condition holds: if n is an integer such that $n|a$ and $n|b$, then $n = \pm 1$. See Section 8.2 for further discussion and references.

(1) Find two integers p and q that are relatively prime. Find two integers c and d that are not relatively prime.
(2) Prove that the following are equivalent.
a. a and b are relatively prime.
b. a and $-b$ are relatively prime.
c. $a+b$ and b are relatively prime.
d. $a-b$ and b are relatively prime.

Exercise 2.4.4. Let n be an integer. Prove that one of the two numbers n and $n+1$ is even, and the other is odd. (You may use the fact that every integer is even or odd.)

Exercise 2.4.5. It follows from Corollary 5.2.5, using $n = 3$, that if a is an integer, then precisely one of the following holds: either $a = 3k$ for some integer k, or $a = 3k+1$ for some integer k, or $a = 3k+2$ for some integer k.

Let n and m be integers.

(1) Suppose that 3 divides n, and that 3 does not divide m. Prove that 3 does not divide $n+m$.
(2) Prove that 3 divides mn if and only if 3 divides m or 3 divides n.

Exercise 2.4.6. Are there any integers p such that $p > 1$, and such that all three numbers p, $p+2$ and $p+4$ are prime numbers? If there are such triples, prove that you have all of them; if there are no such triples, prove why not. Use the discussion at the start of Exercise 2.4.5.

Exercise 2.4.7. Let n be an integer. Using only the fact that every integer is even or odd, and without using Corollary 5.2.5, prove that precisely one of the following holds: either $n = 4k$ for some integer k, or $n = 4k+1$ for some integer k, or $n = 4k+2$ for some integer k, or $n = 4k+3$ for some integer k.

Exercise 2.4.8. Let n be an integer. Suppose that n is odd. Prove that there is an integer k such that $n^2 = 8k+1$.

Exercise 2.4.9. Let x be a real number. Define the **absolute value** of x, denoted $|x|$, by

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

Let a and b be real numbers. Prove the following statements.

(1) $|-a| = |a|$.
(2) $|a|^2 = a^2$.
(3) $|a-b| = |b-a|$.
(4) $|ab| = |a||b|$.

Exercise 2.4.10. Let x and y be real numbers. Let $x \frown y$ and $x \smile y$ be defined by

$$x \frown y = \begin{cases} x, & \text{if } x \ge y \\ y, & \text{if } x \le y, \end{cases} \qquad \text{and} \qquad x \smile y = \begin{cases} y, & \text{if } x \ge y \\ x, & \text{if } x \le y. \end{cases}$$

(Observe that $x \frown y$ is simply the maximum of x and y, and $x \smile y$ is the minimum, though our notation is more convenient for the present exercise than writing $\max\{x,y\}$ and similarly for the minimum.)

Let a, b and c be real numbers. Prove the following statements. The definition of absolute value is given in Exercise 2.4.9.

(1) $(a \frown b) + (a \smile b) = a + b$.
(2) $(a \smile b) + c = (a+c) \smile (b+c)$ and $(a \frown b) + c = (a+c) \frown (b+c)$.
(3) $(a \frown b) \frown c = a \frown (b \frown c)$ and $(a \smile b) \smile c = a \smile (b \smile c)$.
(4) $(a \frown b) - (a \smile b) = |a-b|$.
(5) $a \frown b = \frac{1}{2}(a+b+|a-b|)$ and $a \smile b = \frac{1}{2}(a+b-|a-b|)$.

2.5 Quantifiers in Theorems

A close look at the theorems we have already seen, and those we will be seeing, shows that quantifiers (as discussed in Section 1.5) appear in the statements of many theorems—implicitly if not explicitly. The presence of quantifiers, and especially multiple quantifiers, in the statements of theorems is a major source of error in the construction of valid proofs by beginners. So, extra care should be taken with the material in this section; mastering it now will save much difficulty later on. Before proceeding, it is worth reviewing the material in Section 1.5. Though we will not usually invoke them by name, to avoid distraction, the rules of inference for quantifiers discussed in Section 1.5 are at the heart of much of what we do with quantifiers in theorems.

We start by considering statements with a single universal quantifier, that is, statements of the form "$(\forall x \text{ in } U)P(x)$." Many of the theorems we have already seen have this form, even though the expression "for all" might not appear in their statements. For example, Theorem 2.3.1 says "Let n be an integer. If n^2 is odd, then n is odd." This statement implicitly involves a universal quantifier, and it can be rephrased as "For all integers n, if n^2 is odd, then n is odd." In order to prove that something is true for all integers, we picked an arbitrary integer that we labeled n (any other symbol would do), and proved the result for this arbitrarily chosen integer n. It was crucial

that we picked an arbitrary integer n, rather than a specific integer, for example 7. It is true that $7^2 = 49$ is odd, and that 7 is odd, but checking this one particular case does not tell us anything about what happens in all the other cases where n is an integer with n^2 odd.

More generally, suppose that we want to prove a theorem with statement of the form $(\forall x \text{ in } U)P(x)$. The key observation is that the statement "$(\forall x \text{ in } U)P(x)$" is equivalent to "if x is in U, then $P(x)$ is true." This latter statement has the form $A \rightarrow B$, and it can be proved by any of the methods discussed previously. A direct proof for $(\forall x \text{ in } U)P(x)$ would therefore proceed by choosing some arbitrary x_0 in U, and then deducing that $P(x_0)$ holds. Phrases such as "let x_0 be in U" are often used at the start of an argument to indicate an arbitrary choice of x_0. This type of proof typically has the following form.

Proof. Let x_0 be in U.

$\vdots$

(argumentation)

$\vdots$

Then $P(x_0)$ is true. □

Again, we stress that it is crucial in this type of proof that an arbitrary x_0 in U is picked, not some particularly convenient value. It is not possible to prove that something is true for all values in U by looking at only one (or more) particular cases. In terms of rules of inference, look closely at the discussion of the variable in the Universal Generalization rule of inference in Section 1.5.

For example, a well-known function due to Leonhard Euler is defined by the formula $f(n) = n^2 + n + 41$ for all integers n. If you substitute the numbers $n = 0, 1, 2, \ldots, 39$ into this function, you obtain the numbers 41, 43, 47, …, 1601, all of which are prime numbers. It therefore might appear that substituting in every positive integer into this function would result in a prime number (which would be a very nice property), but it turns out that $f(40) = 1681 = 41^2$, which is not prime. See [Rib96, p. 199] for more discussion of this, and related, functions. The point is that if you want to prove that a statement is true for all x in U, it does not suffice to try only some of the possible values of x.

Statements of the form $(\forall x \text{ in } U)P(x)$ can be proved by strategies other than direct proof. For example, the proof of such a statement using proof by contradiction typically has the following form.

Proof. We use proof by contradiction. Let y_0 be in U. Suppose that $P(y_0)$ is false.

$\vdots$

(argumentation)

$\vdots$

Then we arrive at a contradiction. □

We will not show here any examples of proofs of statements of the form $(\forall x \text{ in } U)P(x)$, because we have already seen a number of such proofs in the previous sections of this chapter.

We now consider statements with a single existential quantifier, that is, statements of the form "$(\exists x \text{ in } U)P(x)$." Using the Existential Generalization rule of inference in Section 1.5, we see that to prove a theorem of the form $(\exists x)P(x)$ means that we need to find some z_0 in U such that $P(z_0)$ holds. It does not matter if there are actually many x in U such that $P(x)$ holds; we need to produce only one of them to prove existence. A proof of "$(\exists x \text{ in } U)P(x)$" can also be viewed as involving a statement of the form $A \to B$. After we produce the desired object z_0 in U, we then prove the statement "if $x = z_0$, then $P(x)$ is true." Such a proof typically has the following form.

Proof. Let $z_0 = \ldots$.

$\vdots$

(argumentation)

$\vdots$

Then z_0 is in U.

$\vdots$

(argumentation)

$\vdots$

Then $P(z_0)$ is true. □

How we find the element z_0 in the above type of proof is often of great interest, and sometimes is the bulk of the effort we spend in figuring out the proof, but it is not part of the actual proof itself. We do not need to explain how we found z_0 in the final write-up of the proof. The proof consists only of defining z_0, and showing that z_0 is in U, and that $P(z_0)$ is true. It is often the case that we find z_0 by going backwards, that is, assuming that $P(z_0)$ is true, and seeing what z_0 has to be. However, this backwards work is not the same as the actual proof, because, as we shall see, not all mathematical arguments can be reversed—what works backwards does not necessarily work forwards.

We now turn to a simple example of a proof involving an existential quantifier. Recall the definitions concerning 2×2 matrices prior to Theorem 2.4.5. We say that a 2×2 matrix $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has integer entries if a, b, c and d are integers.

Proposition 2.5.1. *There exists a* 2×2 *matrix* A *with integer entries such that* $\det A = 4$ *and* $\operatorname{tr} A = 7$.

Scratch Work. Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. The condition $\det A = 4$ means that $ad - bc = 4$; the condition $\operatorname{tr} A = 7$ means that $a + d = 7$. We have two equations with four unknowns. Substituting $d = 7 - a$ into the first equation and rearranging, we obtain $a^2 - 7a + (bc + 4) = 0$. Applying the quadratic equation yields

$$a = \frac{7 \pm \sqrt{33 - 4bc}}{2}.$$

Because we want a, b, c and d to be integers, we need to find integer values of b and c such that $33-4bc$ is the square of an odd integer. Trial and error shows that $b=2$ and $c=3$ yield either $a=5$ and $d=2$, or $a=2$ and $d=5$. (There are other possible solutions, for example $b=-2$ and $c=2$, but we do not need them). ///

Proof. Let $A=\left(\begin{smallmatrix}5 & 2\\ 3 & 2\end{smallmatrix}\right)$. Then $\det A=5\cdot 2-2\cdot 3=4$, and $\operatorname{tr} A=5+2=7$. □

The difference between the scratch work and the actual proof for the above proposition is quite striking, as often occurs in proofs of theorems involving existential quantifiers. In the scratch work we went backwards, by which we mean that we started with the desired conclusion, in this case the assumption that there is some matrix A as desired, and proceeded to find out what criteria would then be imposed on a, b, c, d. We then found a, b, c, d that satisfy these criteria. Such a procedure was helpful, but it could not be our final proof, because we needed to show that the matrix A existed; we were not asked to show what could be said about A if it existed, which is what we did in the scratch work. To show that the desired matrix A existed, we simply had to produce it, and then show that it satisfied the requisite properties regarding its determinant and trace. This is what we did in the proof. How we produced A is irrelevant to the final proof (though not to our understanding of matrices). It is important that the actual proof reads "forwards," not backwards. Moreover, because we were asked to show only that A existed, and not describe how many possible matrices A there were, we needed to exhibit only one value of A in the actual proof, even though we knew that there was more than one possibility from our scratch work. Not everything we learn in the scratch work is necessarily needed in the final proof.

Backwards proofs are so common, especially in elementary mathematics, that unfortunately they are often unnoticed by students, and rarely criticized by instructors. Whereas backwards proofs might not produce any real harm in elementary mathematics, it is crucial to avoid them in advanced mathematics, where questions of logical implication are often much trickier.

Let us examine two simple examples of backwards proofs. First, suppose that we are asked to solve the equation $7x+6=21+4x$. A typical solution submitted by a high school student might look like

$$\begin{aligned} 7x+6&=21+4x\\ 3x-15&=0\\ 3x&=15\\ x&=5. \end{aligned} \tag{2.5.1}$$

There is nothing wrong with the algebra here, and indeed $x=5$ is the correct solution. For computational purposes such a write-up is fine, but logically it is backwards. We were asked to find the solutions to the original equation. A solution to an equation is a number that can be plugged into the equation to obtain a true statement. To solve an equation in the variable x, we simply have to produce a collection of numbers, which we then plug into the equation one at a time, verifying that each one makes the equation a true statement when plugged in. How these solutions are found is

logically irrelevant (though, of course, of great pedagogical interest). A logically correct "forwards" write-up of the solution to $7x+6=21+4x$ would be as follows.

"Let $x=5$. Plugging $x=5$ into the left-hand side of the equation yields $7x+6=7\cdot 5+6=41$, and plugging it into the right-hand side of the equation yields $21+4x=21+4\cdot 5=41$. Therefore $x=5$ is a solution. Because the equation is linear, it has at most one solution. Hence $x=5$ is the only solution."

Such a write-up seems ridiculously long and overly pedantic, given the simplicity of the original equation, and in practice no one would (or should) write such a solution. Logically, however, it is the correct form for the solution to the problem as stated. The backwards approach in Equation 2.5.1 did happen to produce the correct solution to our problem, because all steps in this particular case are reversible. Not all computations are reversible, however, as we now see.

Suppose that we are asked to solve the equation

$$\sqrt{x^2-5}=\sqrt{x+1},$$

where, as is common in high school, we consider only real number solutions. A typical (and backwards) write-up might look like

$$\begin{aligned}\sqrt{x^2-5}&=\sqrt{x+1}\\ x^2-5&=x+1\\ x^2-x-6&=0\\ (x-3)(x+2)&=0\\ x=3 \quad &\text{or} \quad x=-2.\end{aligned}$$

The above write-up is definitely not correct, because $x=-2$ is not a solution to the original equation. In fact, it is not even possible to substitute $x=-2$ into either side of the original equation, because we cannot take the square root of negative numbers. The source of the error in the write-up is that not every step in it is reversible; it is left to the reader to figure out which step cannot be reversed. In an elementary course such as high school algebra or calculus, it would suffice to write up the above computation, and then observe that $x=-2$ should be dropped. In more rigorous proofs, however, it is best to stick to logically correct writing, in order to avoid errors that might otherwise be hard to spot. In your scratch work you can go forwards, backwards, sideways or any combination of these; in the final write-up, however, a proof should always go forwards, starting with the hypothesis and ending up with the desired conclusion.

Returning to our discussion of existence results, one variant on such results concerns theorems that involve existence and uniqueness, of which the following theorem is an example. This theorem concerns 2×2 matrices, as discussed prior to Theorem 2.4.5. This time we need some additional aspects of matrices, namely, the 2×2 identity matrix $I=\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, and matrix multiplication. It would take us too far afield to define matrix multiplication here; we assume that the reader is familiar

with such multiplication. See any introductory text on linear algebra, for example [AR05, Chapter 1], for information about matrix multiplication. It is easy to verify that $AI = A = IA$ for any 2×2 matrix A. It can also be verified (by a slightly tedious computation) that $(AB)C = A(BC)$ for any three 2×2 matrices A, B and C.

The following theorem concerns inverse matrices. Given a 2×2 matrix A, an inverse matrix for A is a 2×2 matrix B such that $AB = I = BA$. Does every 2×2 matrix have an inverse matrix? The answer is no. For example, the matrix $\left(\begin{smallmatrix} 3 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ has no inverse matrix, as the reader may verify (by supposing it has an inverse matrix, and seeing what happens). The following theorem gives a very useful criterion for the existence of inverse matrices. In fact, the criterion is both necessary and sufficient for the existence of inverse matrices, and its analog holds for square matrices of any size, but we will not prove these stronger results.

Theorem 2.5.2. *Let A be a 2×2 matrix such that $\det A \neq 0$. Then A has a unique inverse matrix.*

The phrase "A has a unique inverse matrix" means that an inverse matrix for A exists, and that only one such inverse matrix exists. The logical notation for such a statement is $(\exists!x)P(x)$, where "$\exists!x$" means "there exists unique x." To prove such a statement, we need to prove two things, namely, existence and uniqueness, and it is usually best to prove each of these two things separately. It makes no difference which part is proved first. To prove existence, we proceed as before, and produce an example of the desired object. To prove uniqueness, the standard strategy is to assume that there are two objects of the sort we are looking for, and then show that they are the same. (It is also possible to assume that there are two different objects of the sort we are looking for, and then arrive at a contradiction by showing that the two object are actually the same, but there is rarely any advantage to using this alternative strategy.)

Scratch Work for Theorem 2.5.2. We start with the uniqueness part of the proof, to show that it really is independent of the existence part of the proof. To prove uniqueness, we assume that A has two inverse matrices, say B and C, and then use the properties of matrices cited above, together with the definition of inverse matrices, to show that $B = C$. The proof of existence is rather different. A backwards calculation to try to find an inverse matrix for A would be as follows. Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Suppose that $B = \left(\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right)$ is an inverse matrix of A. Then $BA = I$ and $AB = I$. The latter equality says

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which yields

$$\begin{pmatrix} ax+bz & ay+bw \\ cx+dz & cy+dw \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This matrix equation yields the four equations

$$\begin{aligned} ax+bz &= 1 \\ ay+bw &= 0 \end{aligned}$$

$$cx + dz = 0$$
$$cy + dw = 1,$$

where x, y, z and w are to be thought of as the variables and a, b, c and d are to be thought of as constants. We then solve for x, y, z and w in terms of a, b, c and d. The solution to these four equations turns out to be $x = \frac{d}{ad-bc}$, and $y = \frac{-b}{ad-bc}$, and $z = \frac{-c}{ad-bc}$ and $w = \frac{a}{ad-bc}$. Because $\det A = ad - bc$, we see why the hypothesis that $\det A \neq 0$ is necessary. ///

Proof of Theorem 2.5.2. Uniqueness: Suppose that A has two inverse matrices, say B and C. Then $AB = I = BA$ and $AC = I = CA$. Using standard properties of matrix multiplication, we then compute

$$B = BI = B(AC) = (BA)C = IC = C.$$

Because $B = C$, we deduce that A has a unique inverse.

Existence: Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. The condition $\det A \neq 0$ means that $ad - bc \neq 0$. Let B be the 2×2 matrix defined by

$$B = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} = \begin{pmatrix} \frac{ad}{ad-bc} + \frac{-bc}{ad-bc} & \frac{-ab}{ad-bc} + \frac{ab}{ad-bc} \\ \frac{cd}{ad-bc} + \frac{-cd}{ad-bc} & \frac{-bc}{ad-bc} + \frac{ad}{ad-bc} \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

A similar calculation shows that $BA = I$. Hence B is an inverse matrix of A. □

An understanding of quantifiers is also useful when we want to prove that a given statement is false. Suppose that we want to prove that a statement of the form "$(\forall x \text{ in } U)P(x)$" is false. We saw in Section 1.5 that $\neg[(\forall x \text{ in } U)Q(x)]$ is equivalent to $(\exists x \text{ in } U)(\neg Q(x))$. To prove that the original statement is false, it is sufficient to prove that $(\exists x \text{ in } U)(\neg Q(x))$ is true. Such a proof would work exactly the same as any other proof of a statement with an existential quantifier, that is, by finding some x_0 in U such that $\neg Q(x_0)$ is true, which means that $Q(x_0)$ is false. The element x_0 is called a "counterexample" to the original statement $(\forall x \text{ in } U)P(x)$.

For example, suppose that we want to prove that the statement "all prime numbers are odd" is false. The statement has the form $(\forall x)Q(x)$, where x has values in the integers, and where $Q(x) =$ "if x is prime, then it is odd." Using the reasoning above, it is sufficient to prove that $(\exists x)(\neg Q(x))$ is true. Using Fact 1.3.2 (14), we see that $\neg Q(x)$ is equivalent to "x is prime, and it is not odd." Hence, we need to find some integer x_0 such that x_0 is prime, and it is not odd, which would be a counterexample to the original statement. The number $x_0 = 2$ is just such a number (and in fact it is the only even prime number, though we do not need that fact). This example is so

simple that it may seem unnecessary to go through a lengthy discussion of it, but our point is to illustrate the general approach.

Similar considerations can be used to prove that a statement of the form $(\exists y)R(y)$ is false. It is often very hard to show directly that something does not exist, because one would have to examine all possible cases, and show that none of them have the desired property. Rather, we use the fact that $\neg[(\exists y)R(y)]$ is equivalent to $(\forall y)(\neg R(y))$, and we prove this last statement by our usual methods.

Finally, we look at theorems with statements that involve more than one quantifier. Such theorems might typically have the form $(\forall y)(\exists x)P(x,y)$ or $(\exists a)(\forall b)Q(a,b)$. We saw in Section 1.5 that there are eight possible ways of forming statements with two quantifiers, and clearly with more than two quantifiers there are many more possibilities. There is no point in giving detailed instructions on how to proceed for each different combination of quantifiers, both because there would be too many cases to consider, and because one single strategy works in all cases: take one quantifier at a time, from the outside in. The following two simple results are typical examples of this strategy.

Proposition 2.5.3. *For every real number a, there exists a real number b such that* $a^2 - b^2 + 4 = 0$.

Scratch Work. This proposition has the form $(\forall a)(\exists b)(a^2 - b^2 + 4 = 0)$, where a and b are real numbers. To prove this proposition, we start with the outside quantifier, which is $\forall a$. We can rewrite the statement to be proved as $(\forall a)Q(a)$, where $Q(a) =$ "$(\exists b)(a^2 - b^2 + 4 = 0)$." To prove the statement $(\forall a)Q(a)$, which is a statement with a single universal quantifier, we proceed as before, namely, by picking an arbitrary real number a_0, and then showing that $Q(a_0)$ holds. Therefore we need to show that $(\exists b)((a_0)^2 - b^2 + 4 = 0)$ is true for the given a_0. Again, we have a statement with one quantifier, this time an existential quantifier, and we do a backwards computation to solve for b, which yields $b = \pm\sqrt{(a_0)^2 + 4}$, though we need only one of these solutions. As always, we now write the proof forwards, to make sure that everything is correct. ///

Proof. Let a_0 be a real number. Let $b_0 = \sqrt{(a_0)^2 + 4}$. Then

$$(a_0)^2 - (b_0)^2 + 4 = (a_0)^2 - (\sqrt{(a_0)^2 + 4})^2 + 4 = 0.$$

Hence, for each real number a_0, we found a real number b_0 such that $(a_0)^2 - (b_0)^2 + 4 = 0$. □

Proposition 2.5.4. *There exists a real number x such that* $(3 - x)(y^2 + 1) > 0$ *for all real numbers y.*

Scratch Work. This proposition has the form $(\exists x)(\forall y)((3 - x)(y^2 + 1) > 0)$, where x and y are real numbers. Again, we start with the outside quantifier, which is $\exists x$. We rewrite the statement to be proved as $(\exists x)R(x)$, where $R(x) =$ "$(\forall y)((3 - x)(y^2 + 1) > 0)$." We prove the statement $(\exists x)R(x)$ by producing a single real number x_0 for which $R(x_0)$ holds. That is, we need to find a real number x_0 such that $(\forall y)((3 - x_0)(y^2 +$

$1) > 0)$ is true, and hence we need to find a real number x_0, such that if we pick an arbitrary real number y_0, then $(3-x_0)((y_0)^2+1) > 0$ will hold. Again we do our scratch work backwards. Observe that $(y_0)^2+1 > 0$ for all real numbers y_0, and that $3-x_0 > 0$ for all $x_0 < 3$. We need to pick a single value of x_0 that works, and we randomly pick $x_0 = 2$. ///

Proof. Let $x_0 = 2$. Let y_0 be a real number. Observe that $(y_0)^2+1 > 0$. Then

$$(3-x_0)((y_0)^2+1) = (3-2)((y_0)^2+1) > 0.$$

Hence, we have found a real number x_0 such that $(3-x_0)((y_0)^2+1) > 0$ for all real numbers y_0. □

As discussed in Section 1.5, the order of the quantifiers in the statement of a theorem often matters. The statement of Proposition 2.5.3 is "For every real number a, there exists a real number b such that $a^2-b^2+4=0$," which is $(\forall a)(\exists b)(a^2-b^2+4=0)$. If we were to reverse the quantifiers, we would obtain $(\exists b)(\forall a)(a^2-b^2+4=0)$, which in English would read "there is a real number b such that $a^2-b^2+4=0$ for all real numbers a." This last statement is not true, which we can demonstrate by showing that its negation is true. Using Fact 1.5.1 (2), it follows that $\neg[(\exists b)(\forall a)(a^2-b^2+4=0)]$ is equivalent to $(\forall b)(\exists a)(a^2-b^2+4\neq 0)$. To prove this latter statement, let b_0 be an arbitrary real number. We then choose $a_0 = b_0$, in which case $(a_0)^2-(b_0)^2+4 = 4 \neq 0$. Hence the negation of the statement is true, so the statement is false. We therefore see that the order of the quantifiers in Proposition 2.5.3 does matter. On the other hand, changing the order of the quantifiers in the statement of Proposition 2.5.4, while changing the meaning of the statement, does not make it become false, as the reader may verify.

Exercises

Exercise 2.5.1. Convert the following statements, which do not have their quantifiers explicitly written, into statements with explicit quantifiers (do not prove them, because the terms are meaningless).

(1) If a 5×5 matrix has positive determinant then it is bouncy.
(2) There is a crusty integer that is greater than 7.
(3) For each integer k, there is an opulent integer w such that $k|w$.
(4) There is a fibrous 2×2 matrix P such that $\det P > m$, for each ribbed integer m.
(5) Some 2×2 matrix M has the property that every subtle integer divides $\mathrm{tr}\, M$.

Exercise 2.5.2. A problem that might be given in a high school mathematics class is "Prove that the equation $e^x = 5$ has a unique solution." We could rewrite the problem as "Prove that there exists a unique real number x such that $e^x = 5$." First, write up a solution to the problem as would be typically found in a high school class. Second, write up a proper solution to the problem, using the ideas discussed in this section. Write up the uniqueness first, without making use of the existence part of

the proof; avoid a backwards proof when showing existence. Do not use a calculator (the number x does not have to be given explicitly in decimal expansion).

Exercise 2.5.3. Prove or give a counterexample to each of the following statements.

(1) For each non-negative number s, there exists a non-negative number t such that $s \geq t$.
(2) There exists a non-negative number t such that for all non-negative numbers s, the inequality $s \geq t$ holds.
(3) For each non-negative number t, there exists a non-negative number s such that $s \geq t$.
(4) There exists a non-negative number s such that for all non-negative numbers t, the inequality $s \geq t$ holds.

Exercise 2.5.4. Prove or give a counterexample to each of the following statements.

(1) For each integer a, there exists an integer b such that $a|b$.
(2) There exists an integer b such that for all integers a, the relation $a|b$ holds.
(3) For each integer b, there exists an integer a such that $a|b$.
(4) There exists an integer a such that for all integers b, the relation $a|b$ holds.

Exercise 2.5.5. Prove or give a counterexample to each of the following statements.

(1) For each real number x, there exists a real number y such that $e^x - y > 0$.
(2) There exists a real number y such that for all real numbers x, the inequality $e^x - y > 0$ holds.
(3) For each real number y, there exists a real number x such that $e^x - y > 0$.
(4) There exists a real number x such that for all real numbers y, the inequality $e^x - y > 0$ holds.

Exercise 2.5.6. Prove or give a counterexample to the following statement. For each positive integer a, there exists a positive integer b such that

$$\frac{1}{2b^2+b} < \frac{1}{ab^2}.$$

Exercise 2.5.7. Prove or give a counterexample to the following statement. For every real number y, there is a real number x such that $e^{3x} + y = y^2 - 1$.

Exercise 2.5.8. Prove or give a counterexample to the following statement. For each real number p, there exist real numbers q and r such that $q \sin\left(\frac{r}{5}\right) = p$.

Exercise 2.5.9. Prove or give a counterexample to the following statement. For each integer x, and for each integer y, there exists an integer z such that $z^2 + 2xz - y^2 = 0$.

Exercise 2.5.10. Let $P(x,y)$ be a statement with free variables x and y that are real numbers. Let a and b be real numbers. The real number u is called the least P-number for a and b if two conditions hold: (1) the statements $P(a,u)$ and $P(b,u)$ are both true; and (2) if w is a real number such that $P(a,w)$ and $P(b,w)$ are both true, then $u \leq w$. Suppose that c and d are real numbers, and that there is a least P-number for c and d. Prove that this least P-number is unique.

Exercise 2.5.11. A student is asked to show that the equation $x(x-1) = 2(x+2)$ has a solution. In the context of writing rigorous proofs, what is wrong with the following solution she handed in?

"Proof:

$$\begin{aligned} x(x-1) &= 2(x+2) \\ x^2 - x &= 2x+4 \\ x^2 - 3x - 4 &= 0 \\ (x-4)(x+1) &= 0 \\ x = 4 \quad &\text{or} \quad x = -1. \end{aligned}$$

Therefore there are two solutions."

Exercise 2.5.12. Look through mathematics textbooks that you have previously used (in either high school or college), and find an example of a backwards proof.

2.6 Writing Mathematics

In mathematics—as in any other field—careful writing is of great importance for both the writer and the reader. Careful writing is clearly necessary if the writer's proofs are to be understood by the reader. For the writer's own benefit, putting a mathematical idea into written form forces her to pay attention to all the details of an argument. Often an idea that seemed to make sense in one's head is found to be insufficient when put on paper. Any experienced mathematician knows that until an idea has been written up carefully, its correctness cannot be assumed, no matter how good the idea seemed at first.

Mathematical correctness is certainly the ultimate test of the validity of a proof, but to allow us to judge mathematical correctness, however, a number of important factors in the proper writing of mathematics are needed. Some of these ideas are described below. See [Gil87], [Hig98], [KLR89] and [SHSD73] for further discussion of writing mathematics.

1. A Written Proof Should Stand on Its Own

The first rule of writing proofs actually applies to all forms of writing, not just mathematical writing: The written text should stand on its own, without any need for clarification by the writer. Unlike writing of a more personal nature such as poetry and fiction, a written proof is not an expression of the writer's feelings, but rather a document that should work according to objective standards. When writing a proof, state everything you are doing as explicitly and clearly as possible. DO NOT ASSUME THE READER IS A MIND READER. Err on the side of too much explanation.

2. Write Precisely and Carefully

There is no room in mathematics for ambiguity. The most minute matters of phraseology in mathematics may make a difference. For example, compare the statement "If the given integer n is prime then it is not less than 2, and it is a perfect number" with "If the given integer n is prime, then it is not less than 2 and it is a perfect number." Something as seemingly insignificant as the change of the location of a comma can change the meaning of a statement. MAKE SURE WHAT YOU WRITE IS WHAT YOU MEAN.

As in non-mathematical writing, revision is often the key to achieving precision and clarity. Do not confuse the rough drafts of a proof with the final written version. You should revise your proofs just as you should revise all writing, which is by trying to read what you wrote as if someone else (whose thoughts you do not know) had written it.

Write mathematics in simple, straightforward, plodding prose. Leave your imagination to the mathematical content of your writing, but keep it out of your writing style, so that your writing does not get in the way of communicating your mathematical ideas. Serious mathematics is hard enough as it is, without having unnecessary verbiage or convoluted sentences making it even less clear.

Particular care should be taken with the use of mathematical terminology, where common words are sometimes given technical meanings different from their colloquial meanings (for example, the word "or"). Precision should not be overlooked in the statement of what is being proved. Mathematics is often read by skipping back and forth, and so it is important that the statements of theorems, lemmas, propositions and the like contain all their hypotheses, rather than having the hypotheses in some earlier paragraphs. Better a bit of redundancy than a confused reader.

3. Prove What Is Appropriate

A good proof should have just the right amount of detail—neither too little nor too much. The question of what needs to be included in a proof, and what can be taken as known by the reader, is often a matter of judgment. A good guideline is to assume that the reader is at the exact same level of knowledge as you are, but does not know the proof you are writing. It is certainly safe to assume that the reader knows elementary mathematics at the high school level (for example, the quadratic formula). In general, do not assume that the reader knows anything beyond what has been covered in your mathematics courses. When in doubt—prove.

4. Be Careful with Saying Things Are "Obvious"

It is very tempting to skip over some details in a proof by saying that they are "obvious" or are "similar to what has already been shown." Such statements are legitimate if true, but are often used as a cover for uncertainty or laziness. "Obvious" is in the eye of the beholder; what may seem obvious to the writer after spending hours (or

days) on a problem might not be so obvious to the reader. That something is obvious should mean that another person at your level of mathematical knowledge could figure it out in very little time and with little effort. If it does not conform to this criterion, it is not "obvious." As an insightful colleague once pointed out, if something is truly obvious, then there is probably no need to remind the reader of that fact.

The words "trivial" and "obvious" mean different things when used by mathematicians. Something is trivial if, after some amount of thought, a logically very simple proof is found. Something is obvious if, relative to a given amount of mathematical knowledge, a proof can be thought of very quickly by anyone at the given level. According to an old joke, a professor tells students during a lecture that a certain theorem is trivial; when challenged by one student, the professor thinks and thinks, steps out of the room to think some more, comes back an hour later, and announces to the class that the student was right, and that the result really is trivial. The joke hinges on the fact that something can be trivial without being obvious.

5. Use Full Sentences and Correct Grammar

The use of correct grammar (such as complete sentences and correct punctuation) is crucial if the reader is to follow what is written. Mathematical writing should be no less grammatically correct than literary prose. Mathematics is not written in a language different from the language we use for general speech. In this text all mathematics is written in English.

A distinguishing feature of mathematical writing is the use of symbols. It is very important to understand that mathematical symbols are nothing but shorthand for expressions that could just as well be written out in words. For example, the phrase "$x = z^2$" could be written as "the variable x equals the square of the variable z." Mathematical symbols are therefore subject to the rules of grammar just as words are. Mathematical symbols floating freely on a page are neither understandable nor acceptable. All symbols, even those displayed between lines, should be embedded in sentences and paragraphs.

A proof is an explanation of why something is true. A well-written proof is an explanation that someone else can understand. Proper grammar helps the reader follow the logical flow of the proof. Connective words such as "therefore," "hence" and "it follows that" help guide the logical flow, and should be used liberally. Look through this entire book, and you will see that we always use complete sentences and paragraphs, as well as correct grammar and the frequent use of connective words (except, of course, for some instances of typographical errors). Though it may at times seem cumbersome when you are writing a proof, and would like to get it done as quickly as possible, sticking with correct grammar and a readable style will pay off in the long run.

The following two examples of poor writing, both of which contain all the mathematical ideas of the proof of Theorem 2.3.5, are written without regard to proper grammar and style, and are modeled on homework assignments the author has received from students. Compare these versions of the proof with the proof as originally given in Section 2.3.

The first version is genuinely awful, though for reasons the author does not understand, some students seem to be given the impression in high school that this sort of writing is acceptable.

> $x^2 = 2$ and x rational
> $\therefore x = \frac{n}{m}$
> n and m have no common factors
> $\left(\frac{n}{m}\right)^2 = 2 \Rightarrow \frac{n^2}{m^2} = 2 \Rightarrow n^2 = 2m^2$ which is even
> if n odd, n^2 odd (Exercise 2.2.4) contradiction
> $\therefore n$ even
> $n = 2k \Rightarrow (2k)^2 = 2m^2 \Rightarrow 4k^2 = 2m^2 \Rightarrow 2k^2 = m^2$
> m even (as before)
> $\therefore n$ and m both even—impossible (no common factors)
> $\therefore x$ is not rational.

This second version is slightly better, being in paragraph form and with a few more words, but it is still far from desirable.

> $x^2 = 2$, x is rational. so $x = \frac{n}{m}$; n and m have no common factors. $\left(\frac{n}{m}\right)^2 = 2$, $\frac{n^2}{m^2} = 2$, $n^2 = 2m^2$. If n were odd, then n^2 would be odd by Exercise 2.2.4 a contradiction because $2m^2$ is even because it is divisible by 2. n not odd and hence is even. $n = 2k$ $(2k)^2 = 2m^2$, $4k^2 = 2m^2$, $2k^2 = m^2$. m is even as before both n and m even—impossible because any two even numbers have 2 as a factor, but n and m have no common factors. x is not rational.

Mathematicians do not write papers and books this way; please do not write this way yourself!

6. Use "=" Signs Properly

One of the hallmarks of poor mathematical writing is the improper use of "=" signs. It is common for beginners in mathematics to write "=" when it is not appropriate, and to drop "=" signs when they are needed. Both these mistakes should be studiously avoided. For example, suppose that a student is asked to take the derivative of the function defined by $f(x) = x^2$ for all real numbers x. The first type of mistake occurs when someone writes something such as "$f(x) = x^2 = 2x = f'(x)$." What is meant is correct, but what is actually written is false (because this function does not equal its derivative), and it is therefore extremely confusing to anyone other than the writer of the statement. THE READER SHOULD NOT HAVE TO GUESS WHAT THE WRITER INTENDED.

The second type of mistake occurs when someone writes "$f(x) = x^2$, and so $2x$." Here again the reader has to guess what is meant by $2x$. If it is meant that $f'(x) = 2x$, then why not write that?

Both of these examples of the improper use of "=" signs may seem far-fetched, but the author has seen these and similar mistakes quite regularly on homework assignments and tests in calculus courses. A proper write-up could be either "$f(x) = x^2$ for all real numbers x, so $f'(x) = 2x$ for all x," or simply "$(x^2)' = 2x$."

Another common type of error involving "=" signs involves lengthier calculations. Suppose that a student is asked to show that

$$(x^2+2x)(x^2-4)(x^2-2x) = (x^3-4x)^2.$$

An incorrect way of writing the calculation, which the author has seen very regularly on homework assignments, would be

$$\begin{aligned}(x^2+2x)(x^2-4)(x^2-2x) &= (x^3-4x)^2\\ x(x+2)(x-2)(x+2)x(x-2) &= (x^3-4x)^2\\ x^2(x+2)^2(x-2)^2 &= (x^3-4x)^2\\ [x(x-2)(x+2)]^2 &= (x^3-4x)^2\\ (x^3-4x)^2 &= (x^3-4x)^2.\end{aligned}$$

The problem here is that this calculation as written is a backwards proof, as discussed in Section 2.5. The calculation starts by stating the equation that we are trying to prove, and deducing from it an equation that is clearly true. A correct proof should start from what we know to be true, and deduce that which we are trying to prove. In principle, if the writer of such a backwards proof were to verify that every step is reversible, and indicate this fact after the above write-up, then the calculation would be correct. However, no one ever does that, and doing so would be more complicated than doing the proof correctly to begin with.

Another incorrect way of writing this same calculation, and also one that the author has seen regularly, is

$$\begin{gathered}(x^2+2x)(x^2-4)(x^2-2x)\\ x(x+2)(x-2)(x+2)x(x-2)\\ x^2(x+2)^2(x-2)^2\\ [x(x-2)(x+2)]^2\\ (x^3-4x)^2.\end{gathered}$$

The problem here is with what is not written, namely, the "=" signs. What is written is a collections of formulas, without any explicit indication of what equals what. The reader can often deduce what the writer of such a collection of formulas meant, but why risk confusion? Written mathematics should strive for clarity, and should therefore state exactly what the writer means.

A helpful way to think about this second type of error is via the need for correct grammar. The statement "$(x^2+2x)(x^2-4)(x^2-2x) = (x^3-4x)^2$" is a complete sentence, with subject "$(x^2+2x)(x^2-4)(x^2-2x)$," with verb "=" and with object "$(x^3-4x)^2$." To drop the = sign is to drop the verb in this sentence. Few students would ever turn in a literature paper with missing verbs. And yet, unfortunately, many students do the equivalent in mathematics homework assignments—not because of any ill intention, but because, sadly, improper ways of writing lengthy calculations are actually taught to many students in high school. These errors should be discarded.

There are a number of correct ways of writing the above calculation, for example

$$\begin{aligned}(x^2+2x)(x^2-4)(x^2-2x) &= x(x+2)(x-2)(x+2)x(x-2)\\ &= x^2(x+2)^2(x-2)^2\\ &= [x(x-2)(x+2)]^2\\ &= (x^3-4x)^2,\end{aligned}$$

and

$$\begin{aligned}(x^2+2x)(x^2-4)(x^2-2x) &= x(x+2)(x-2)(x+2)x(x-2)\\ &= x^2(x+2)^2(x-2)^2 = [x(x-2)(x+2)]^2 = (x^3-4x)^2.\end{aligned}$$

The differences between these correctly written calculations and the incorrect ones may seem extremely minor and overly picky, but mathematics is a difficult subject, and every little detail that makes something easier to follow (not to mention logically correct) is worthwhile. A lack of attention to fundamentals such as writing "=" signs correctly can often be a symptom of a general lack of attention to logical thoroughness. A good place to start building logical thinking is with the basics.

7. Define All Symbols and Terms You Make Up

Any mathematical symbols used as variables, even simple ones such as x or n, need to be defined before they are used. Such a definition might be as simple as "let x be a real number." (If you are familiar with programming languages such as C++ or Java, think of having to declare all variables before they are used.) For example, it is not acceptable to write "$x+y$" without somewhere stating that x and y are real numbers (or whatever else they might be); the symbol $+$ needs no definition, because it is not a variable, and its meaning is well-known. The same need for definition holds when the variable is a set, function, relation or anything else. Just because a letter such as n is often used to denote an integer, or the letter f is often used to denote a function, one cannot rely upon such conventions, because these same letters can be used to mean other things as well. If you want to use n to denote an integer, you must say so explicitly, and similarly for f denoting a function.

The need to define variables can get a bit tricky when quantifiers are involved. It is important to understand the scope of any quantifier being used. Suppose that somewhere in a proof you have the statement "for each positive integer n, there is an integer p such that ...." The variables n and p are bound variables, and are defined only inside that statement. They cannot be used subsequently, unless they are redefined. If you subsequently want to use a positive integer, you cannot assume that the symbol n has already been defined as such. You would need to define it for the current use, by saying, as usual, something such as "let n be a positive integer."

Finally, it is tempting in the course of a complicated proof to make up new words and symbols, and to use all sorts of exotic alphabets. For the sake of readability, avoid

this temptation as much as possible. Do not use more symbols than absolutely necessary, and avoid exotic letters and complications (such as subscripts of subscripts) where feasible. Try to stick to standard notation. If you do make up some notation, make sure you define it explicitly.

8. Break Up a Long Proof into Steps

If a proof is long and difficult to follow, it is often wise to break it up into steps, or to isolate preliminary parts of the proof as lemmas (which are simply smaller theorems used to prove bigger theorems). If you use lemmas, be sure to state them precisely. Prior to going into the details of a long proof, it is often useful to give a sentence or two outlining the strategy of the proof. All lemmas and their proofs should be placed before they are used in the main theorem. Do not put a lemma inside the proof of the main theorem—doing so can be very confusing to the reader.

9. Distinguish Formal vs. Informal Writing

Writing mathematics involves both formal and informal writing. Formal writing is used for definitions, statements of theorems, proofs and examples; informal writing is for motivation, intuitive explanations, descriptions of the mathematical literature, etc. When writing up the solution to an exercise for a mathematics course, the writing should be a formal proof. A lengthier exposition (such as a thesis or a book) will make use of both kinds of writing—formal writing to make sure that mathematical rigor is maintained, and informal writing to make the text understandable and interesting. Do not confuse the two types of writing, or each will fail to do what it is supposed to do. Intuitive aids such as drawings, graphs, Venn diagrams and the like are extremely helpful when writing up a proof, though such aids should be in addition to the proof, not instead of it.

10. Miscellaneous Writing Tips

Most of the following items are from [KLR89] and [OZ96, pp. 109–118], which have many other valuable suggestions not included here for the sake of brevity. All the examples of poor writing given below are based on what the author has seen in homework assignments and tests.

(A) Do not put a mathematical symbol directly following punctuation. As a corollary, do not start a sentence with a symbol. The only exception to this rule is when the punctuation is part of the mathematical notation, for example (x,y). It is important to avoid ambiguities that might arise from using punctuation without proper care. For example, does the expression "$0 < x, y < 1$" mean that both x and y are between 0 and 1, or does it mean that $0 < x$ and $y < 1$?

Bad: For all $x > 3$, $x^2 > 9$. $y \leq 0$, so $xy < 0$.

Good: For all $x > 3$, it follows that $x^2 > 9$. Moreover, because $y \leq 0$, then $xy < 0$.

(B) In the final write-up of a proof, do not use logical symbols, such as $\wedge$, $\vee$, $\exists$, $\forall$ and $\Rightarrow$, as abbreviations for words. Unless you are writing about logic, where logical symbols are necessary, the use of logical symbols makes proofs harder for others to read. Of course, you may use any symbols you want in your scratch work.

Bad: $\forall$ distinct real numbers $x \wedge y$, if $x < y \Rightarrow \exists$ rational q such that $x < q < y$.

Good: For all distinct real numbers x and y, if $x < y$ then there exists a rational number q such that $x < q < y$.

(C) Use equal signs only in equations (and only then when the two sides are equal!). Do not use equal signs when you mean "implies," "the next step is" or "denotes." Do not use equal signs instead of punctuation, or as a substitute for something properly expressed in words.

Bad: $n =$ odd $= 2k+1$.

Good: Let n be an odd number. Then $n = 2k+1$ for some integer k.

Bad: For the next step, let $i = i+1$.

Good: For the next step, replace i with $i+1$.

Bad: Let $P =$ the # of people in the room.

Good: Let P denote the number of people in the room.

(D) Use consistent notation throughout a proof. For example, if you start a proof using uppercase letters for matrices and lowercase letters for numbers, stick with that notation for the duration of the proof. Do not use the same notation to mean two different things, except when it is unavoidable due to standard mathematical usage—for example, the multiple uses of the notation "(a,b)."

(E) Display long formulas, as well as short ones that are important, on their own lines. Recall, however, that such displayed formulas are still parts of sentences, and require normal punctuation. In particular, if a sentence ends with a displayed formula, do not forget the period at the end of the formula. Also, do not put an unnecessary colon in front of a displayed formula that does not require it.

Bad: From our previous calculations, we see that:

$$x^5 - r\cos\theta = \sqrt{y^2+3}$$

Good: From our previous calculations, we see that

$$x^5 - r\cos\theta = \sqrt{y^2+3}.$$

(F) Colons are very rarely needed. They are usually either unnecessary, as in the bad example in Item (E), or meant as substitutes for words in situations where words would be much more clear. In mathematical writing, colons should normally be used only in headings or at the starts of lists, and in certain mathematical symbols. Do not use a colon in mathematical writing in a place where you would not use one in non-mathematical writing.

Bad: $x^2 + 10x + 3 = 0$ has two real solutions: $10^2 - 4 \cdot 1 \cdot 3 > 0$.

Good: The equation $x^2 + 10x + 3 = 0$ has two real solutions because $10^2 - 4 \cdot 1 \cdot 3 > 0$.

(G) Capitalize names such as "Theorem 2.3" and "Lemma 17." No capitalization is needed in phrases such as "by the previous theorem."

Exercises

Exercise 2.6.1. State what is wrong with each of the following write-ups; some have more than one error.

(1) We make use of the fact about the real numbers that if $x > 0$, $x^2 > 0$.

(2) To solve $x^2 + 6x = 16$:

$$x^2 + 6x = 16$$
$$x^2 + 6x - 16 = 0$$
$$(x-2)(x+8) = 0$$

and $x = 2$, $x = -8$.

(3) In order to solve $x^2 + 6x = 16$, then $x^2 + 6x - 16 = 0$, $(x-2)(x+8) = 0$, and therefore $x = 2$, $x = -8$.

(4) We want to solve the equation $x^2 - 2x = x + 10$. then $x^2 - 3x - 10$, so $(x-5)(x+2)$, so 5 and -2.

(5) We want to multiply the two polynomials $(7+2y)$ and (y^2+5y-6), which we do by computing

$$(7+2y)(y^2+5y-6)$$
$$7y^2 + 35y - 42 + 2y^3 + 10y^2 - 12y$$
$$2y^3 + 17y^2 + 23y - 42$$

the answer is $2y^3 + 17y^2 + 23y - 42$.

(6) A real number x is gloppy if there is some integer n such that $x^2 - n$ is sloppy. Suppose that x is gloppy. Because n is an integer, then its square is an integer, ... (The terms here are meaningless.)

(7) Let x be a real number. Then $x^2 \geq 0$ for all real numbers x, ...

(8) It is known that $\sqrt{a} < a$ for all $a > 1$. Hence $\sqrt{a} + 3 < a + 3$. Hence $(\sqrt{a}+3)^2 < (a+3)^2$.

http://www.springer.com/978-1-4419-7126-5

Proofs and Fundamentals
A First Course in Abstract Mathematics
Bloch, E.D.
2011, XXVI, 358 p., Hardcover
ISBN: 978-1-4419-7126-5