CHAPTER

2

Mathematical Induction and Recursive Definitions

2.1 | PROOFS

A *proof* of a statement is essentially just a convincing argument that the statement is true. Ideally, however, a proof not only convinces but explains why the statement is true, and also how it relates to other statements and how it fits into the overall theory. A typical step in a proof is to derive some statement from (1) assumptions or hypotheses, (2) statements that have already been derived, and (3) other generally accepted facts, using general principles of logical reasoning. In a very careful, detailed proof, we might allow no "generally accepted facts" other than certain axioms that we specify initially, and we might restrict ourselves to certain specific rules of logical inference, by which each step must be justified. Being this careful, however, may not be feasible or worthwhile. We may take shortcuts ("It is obvious that ..." or "It is easy to show that ...") and concentrate on the main steps in the proof, assuming that a conscientious or curious reader could fill in the low-level details.

Usually what we are trying to prove involves a statement of the form $p \rightarrow q$. A *direct* proof assumes that the statement p is true and uses this to show q is true.

The Product of Two Odd Integers Is Odd — EXAMPLE 2.1

To prove: For any integers a and b, if a and b are odd, then ab is odd.

■ Proof

We start by saying more precisely what our assumption means. An integer n is odd if there exists an integer x so that $n = 2x + 1$. Now let a and b be any odd integers. Then according to this definition, there is an integer x so that $a = 2x + 1$, and there is an integer y so that

$b = 2y + 1$. We wish to show that there is an integer z so that $ab = 2z + 1$. Let us therefore calculate ab:

$$\begin{aligned} ab &= (2x + 1)(2y + 1) \\ &= 4xy + 2x + 2y + 1 \\ &= 2(2xy + x + y) + 1 \end{aligned}$$

Since we have shown that there is a z, namely, $2xy + x + y$, so that $ab = 2z + 1$, the proof is complete.

This is an example of a *constructive* proof. We proved the statement "There exists z such that . . ." by constructing a specific value for z that works. A nonconstructive proof shows that such a z must exist without providing any information about its value. Such a proof would not explain, it would only convince. Although in some situations this is the best we can do, people normally prefer a constructive proof if one is possible. In some cases, the method of construction is interesting in its own right. In these cases, the proof is even more valuable because it provides an algorithm as well as an explanation.

Since the statement we proved in Example 2.1 is the quantified statement "For any integers a and $b, \ldots,$" it is important to understand that it is *not* sufficient to give an example of a and b for which the statement is true. If we say "Let $a = 45$ and $b = 11$; then $a = 2(22) + 1$ and $b = 2(5) + 2$; therefore, $ab = (2 * 22 + 1)(2 * 5 + 1) = \ldots = 2 * 247 + 1$," we have proved nothing except that $45 * 11$ is odd. Finding a value of x so that the statement $P(x)$ is true is a proof of the statement "There exists x such that $P(x)$." Finding a value of x for which $P(x)$ is false *disproves* the statement "For every x, $P(x)$" (or, if you prefer, proves the statement "It is not the case that for every x, $P(x)$"); this is called a *proof by counterexample*. To prove "For every x, $P(x)$," however, requires that we give an argument in which there are no restrictions on x. (Let us return briefly to the example with 45 and 11. It is not totally unreasonable to claim that the argument beginning "Let $a = 45$ and $b = 11$" is a proof of the quantified statement—after all, the algebraic steps involved are the same as the ones we presented in our official proof. The crucial point, however, is that there is nothing special about 45 and 11. Someone who offers this as a proof should at least point out that the same argument would work in general. For an argument this simple, such an observation may be convincing; even more convincing is an argument involving a and b like the one we gave originally.)

The alternative to a direct proof is an *indirect* proof, and the simplest form of indirect proof is a *proof by contrapositive*, using the logical equivalence of $p \rightarrow q$ and $\neg q \rightarrow \neg p$.

EXAMPLE 2.2 A Proof by Contrapositive

To prove: For any positive integers i, j, and n, if $i * j = n$, then either $i \leq \sqrt{n}$ or $j \leq \sqrt{n}$.

Proof

The statement we wish to prove is of the general form "For every x, if $p(x)$, then $q(x)$." For each x, the statement "If $p(x)$ then $q(x)$" is logically equivalent to "If not $q(x)$ then not $p(x)$," and therefore (by a general principle of logical reasoning) the statement we want to prove is equivalent to this: For any positive integers i, j, and n, if it is not the case that $i \leq \sqrt{n}$ or $j \leq \sqrt{n}$, then $i * j \neq n$.

If it is not true that $i \leq \sqrt{n}$ or $j \leq \sqrt{n}$, then $i > \sqrt{n}$ and $j > \sqrt{n}$. A generally accepted fact from mathematics is that if a and b are numbers with $a > b$, and c is a number > 0, then $ac > bc$. Applying this to the inequality $i > \sqrt{n}$ with $c = j$, we obtain $i * j > \sqrt{n} * j$. Since $n > 0$, we know that $\sqrt{n} > 0$, and we may apply the same fact again to the inequality $j > \sqrt{n}$, this time letting $c = \sqrt{n}$, to obtain $j\sqrt{n} > \sqrt{n}\sqrt{n} = n$. We now have $i * j > j\sqrt{n} > n$, and it follows that $i * j \neq n$.

The second paragraph in this proof illustrates the fact that a complete proof, with no details left out, is usually not feasible. Even though the statement we are proving here is relatively simple, and our proof includes more detail than might normally be included, there is still a lot left out. Here are some of the details that were ignored:

1. $\neg(p \vee q)$ is logically equivalent to $\neg p \wedge \neg q$. Therefore, if it is not true that $i \leq \sqrt{n}$ or $j \leq \sqrt{n}$, then $i \not\leq \sqrt{n}$ and $j \not\leq \sqrt{n}$.
2. For any two real numbers a and b, exactly one of the conditions $a < b$, $a > b$, and $a = b$ holds. (This is a generally accepted fact from mathematics.) Therefore, if $i \not\leq \sqrt{n}$, then $i > \sqrt{n}$, and similarly for j.
3. For any two real numbers a and b, $a * b = b * a$. Therefore, $\sqrt{n} * j = j\sqrt{n}$.
4. The $>$ relation on the set of real numbers is transitive. Therefore, from the fact that $i * j > j\sqrt{n}$ and $j\sqrt{n} > n$ it follows that $i * j > n$.

Even if we include all these details, we have not stated explicitly the rules of inference we have used to arrive at the final conclusion, and we have used a number of facts about real numbers that could themselves be proved from more fundamental axioms. In presenting a proof, one usually tries to strike a balance: enough left out to avoid having the minor details obscure the main points and put the reader to sleep, and enough left in so that the reader will be convinced.

A variation of proof by contrapositive is *proof by contradiction*. In its most general form, proving a statement p by contradiction means showing that if it is not true, some contradiction results. Formally, this means showing that the statement $\neg p \rightarrow \mathit{false}$ is true. It follows that the contrapositive statement $\mathit{true} \rightarrow p$ is true, and this statement is logically equivalent to p. If we wish to prove the statement $p \rightarrow q$ by contradiction, we assume that $p \rightarrow q$ is false. Because of the logical equivalence of $p \rightarrow q$ and $\neg p \vee q$, this means assuming that $\neg(\neg p \vee q)$, or $p \wedge \neg q$, is true. From this assumption we try to derive some statement that contradicts some statement we know to be true—possibly p, or possibly some other statement.

EXAMPLE 2.3 $\sqrt{2}$ Is Irrational

A real number x is *rational* if there are two integers m and n so that $x = m/n$. We present one of the most famous examples of proof by contradiction: the proof, known to the ancient Greeks, that $\sqrt{2}$ is irrational.

■ Proof

Suppose for the sake of contradiction that $\sqrt{2}$ is rational. Then there are integers m' and n' with $\sqrt{2} = m'/n'$. By dividing both m' and n' by all the factors that are common to both, we obtain $\sqrt{2} = m/n$, for some integers m and n having no common factors. Since $m/n = \sqrt{2}$, $m = n\sqrt{2}$. Squaring both sides of this equation, we obtain $m^2 = 2n^2$, and therefore m^2 is even (divisible by 2). The result proved in Example 2.1 is that for any integers a and b, if a and b are odd, then ab is odd. Since a conditional statement is logically equivalent to its contrapositive, we may conclude that for any a and b, if ab is not odd, then either a is not odd or b is not odd. However, an integer is not odd if and only if it is even (Exercise 2.21), and so for any a and b, if ab is even, then a or b is even. If we apply this when $a = b = m$, we conclude that since m^2 is even, m must be even. This means that for some k, $m = 2k$. Therefore, $(2k)^2 = 2n^2$. Simplifying this and canceling 2 from both sides, we obtain $2k^2 = n^2$. Therefore, n^2 is even. The same argument that we have already used shows that n must be even, and so $n = 2j$ for some j. We have shown that m and n are both divisible by 2. This contradicts the previous statement that m and n have no common factor. The assumption that $\sqrt{2}$ is rational therefore leads to a contradiction, and the conclusion is that $\sqrt{2}$ is irrational.

EXAMPLE 2.4 Another Proof by Contradiction

To prove: For any sets A, B, and C, if $A \cap B = \emptyset$ and $C \subseteq B$, then $A \cap C = \emptyset$.

■ Proof

Again we try a proof by contradiction. Suppose that A, B, and C are sets for which the conditional statement is false. Then $A \cap B = \emptyset$, $C \subseteq B$, and $A \cap C \neq \emptyset$. Therefore, there exists x with $x \in A \cap C$, so that $x \in A$ and $x \in C$. Since $C \subseteq B$ and $x \in C$, it follows that $x \in B$. Therefore, $x \in A \cap B$, which contradicts the assumption that $A \cap B = \emptyset$. Since the assumption that the conditional statement is false leads to a contradiction, the statement is proved.

There is not always a clear line between a proof by contrapositive and one by contradiction. Any proof by contrapositive that $p \rightarrow q$ is true can easily be reformulated as a proof by contradiction. Instead of assuming that $\neg q$ is true and trying to show $\neg p$, assume that p and $\neg q$ are true and derive $\neg p$; then the contradiction is that p and $\neg p$ are both true. In the last example it seemed slightly easier to argue by contradiction, since we wanted to use the assumption that $C \subseteq B$. A proof by contrapositive would assume that $A \cap C \neq \emptyset$ and would try to show that

$$\neg((A \cap B = \emptyset) \wedge (C \subseteq B))$$

This approach seems a little more complicated, just because the formula we are trying to obtain is more complicated.

It is often convenient (or necessary) to use several different proof techniques within a single proof. Although the overall proof in the following example is not a proof by contradiction, this technique is used twice within the proof.

There Must Be a Prime Between n and $n!$ — EXAMPLE 2.5

For a positive integer n the number $n!$ is defined to be the product $n * (n-1) * \cdots * 2 * 1$ of all the positive integers less than or equal to n. *To prove:* For any integer $n > 2$, there is a prime p satisfying $n < p < n!$.

■ Proof

Since $n > 2$, two distinct factors in $n!$ are n and 2. Therefore, $n! \geq 2n = n + n > n + 1$, and thus $n! - 1 > n$. The number $n! - 1$ must have a factor p that is a prime. (See Example 1.2 for the definition of a prime. The fact that every integer greater than 1 has a prime factor is a basic fact about positive integers, which we will prove in Example 2.11.) Since p is a divisor of $n! - 1$, $p \leq n! - 1 < n!$. This gives us one of the inequalities we need. To show the other one, suppose for the sake of contradiction that $p \leq n$. Then since p is one of the positive integers less than or equal to n, p is a factor of $n!$. However, p cannot be a factor of both $n!$ and $n! - 1$; if it were, it would be a factor of 1, their difference, and this is impossible. Therefore, the assumption that $p \leq n$ leads to a contradiction, and we may conclude that $n < p < n!$.

Another useful technique is to divide the proof into separate cases; this is illustrated by the next example.

Strings of Length 4 Contain Substrings yy — EXAMPLE 2.6

To prove: Every string x in $\{0, 1\}^*$ of length 4 contains a nonnull substring of the form yy.

■ Proof

We can show the result by considering two separate cases. If x contains two consecutive 0's or two consecutive 1's, then the statement is true for a string y of length 1. In the other case, any symbol that follows a 0 must be a 1, and vice versa, so that x must be either 0101 or 1010. The statement is therefore true for a string y of length 2.

Even though the argument is simple, let us state more explicitly the logic on which it depends. We want to show that some proposition P is true. The statement P is logically equivalent to $\mathit{true} \rightarrow P$. If we denote by p the statement that x contains two consecutive 0's or two consecutive 1's, then $p \vee \neg p$ is true. This means $\mathit{true} \rightarrow P$ is logically equivalent to

$$(p \vee \neg p) \rightarrow P$$

which in turn is logically equivalent to

$$(p \rightarrow P) \wedge (\neg p \rightarrow P)$$

This last statement is what we actually prove, by showing that each of the two separate conditional statements is true.

In this proof, there was some choice as to which cases to consider. A less efficient approach would have been to divide our two cases into four subcases: (i) x contains two consecutive 0's; (and so forth). An even more laborious proof would be to consider the 16 strings of length 4 individually, and to show that the result is true in each case. Any of these approaches is valid, as long as our cases cover all the possibilities and we can complete the proof in each case.

The examples in this section provide only a very brief introduction to proofs. Learning to read proofs takes a lot of practice, and creating your own is even harder. One thing that does help is to develop a critical attitude. Be skeptical. When you read a step in a proof, ask yourself, "Am I convinced by this?" When you have written a proof, read it over as if someone else had written it (it is best to read aloud if circumstances permit), and as you read each step ask yourself the same question.

2.2 THE PRINCIPLE OF MATHEMATICAL INDUCTION

Very often, we wish to prove that some statement involving a natural number n is true for every sufficiently large value of n. The statement might be a numerical equality:

$$\sum_{i=1}^{n} i = n(n+1)/2$$

The number of subsets of $\{1, 2, \ldots, n\}$ is 2^n.

It might be an inequality:

$$n! > 2^n$$

It might be some other assertion about n, or about a set with n elements, or a string of length n:

There exist positive integers j and k so that $n = 3j + 7k$.

Every language with exactly n elements is regular.

If $x \in \{0, 1\}^*$, $|x| = n$, and $x = 0y1$, then x contains the substring 01.

(The term *regular* is defined in Chapter 3.) In this section, we discuss a common approach to proving statements of this type.

In both the last two examples, it might seem as though the explicit mention of n makes the statement slightly more awkward. It would be simpler to say, "Every finite language is regular," and this statement is true; it would also be correct to let the last statement begin, "For any x and y in $\{0, 1\}^*$, if $x = 0y1, \ldots$." However, in both cases the simpler statement is equivalent to the assertion that the original statement is true for every nonnegative value of n, and formulating the statement so that it involves n will allow us to apply the proof technique we are about to discuss.

The Sum of the First n Positive Integers **EXAMPLE 2.7**

We begin with the first example above, expressed without the summation notation:

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

This formula is supposed to hold for every $n \geq 1$; however, it makes sense to consider it for $n=0$ as well if we interpret the left side in that case to be the empty sum, which by definition is 0. Let us therefore try to prove that the statement is true for every $n \geq 0$.

How do we start? Unless we have any better ideas, we might very well begin by writing out the formula for the first few values of n, to see if we can spot a pattern.

$$\begin{aligned}
n=0: &\quad 0=0(0+1)/2\\
n=1: &\quad 0+1=1(1+1)/2\\
n=2: &\quad 0+1+2=2(2+1)/2\\
n=3: &\quad 0+1+2+3=3(3+1)/2\\
n=4: &\quad 0+1+2+3+4=4(4+1)/2
\end{aligned}$$

As we are verifying these formulas, we probably realize after a few lines that in checking a specific case, say $n=4$, it is not necessary to do all the arithmetic on the left side: $0+1+2+3+4$. We can take the left side of the previous formula, which we have already calculated, and add 4. When we calculated $0+1+2+3$, we obtained $3(3+1)/2$. So our answer for $n=4$ is

$$3(3+1)/2+4=4(3/2+1)=4(3+2)/2=4(4+1)/2$$

which is the one we wanted. Now that we have done this step, we can take care of $n=5$ the same way, by taking the sum we just obtained for $n=4$ and adding 5:

$$4(4+1)/2+5=5(4/2+1)=5(4+2)/2=5(5+1)/2$$

These two calculations are similar—in fact, this is the pattern we were looking for, and we can probably see at this point that it will continue. Are we ready to write our proof?

■ Example 2.7. Proof Number 1

To show

$$0+1+2+\cdots+n=n(n+1)/2 \qquad \text{for every } n \geq 0$$

$n=0$: $0=0(0+1)/2$

$n=1$:
$$\begin{aligned}
0+1 &= 0(0+1)/2+1 \quad \text{(by using the result for } n=0)\\
&= 1(0/2+1)\\
&= 1(0+2)/2\\
&= 1(1+1)/2
\end{aligned}$$

$n=2$:
$$\begin{aligned}
0+1+2 &= 1(1+1)/2+2 \quad \text{(by using the result for } n=1)\\
&= 2(1/2+1)\\
&= 2(1+2)/2\\
&= 2(2+1)/2
\end{aligned}$$

$$
\begin{aligned}
n = 3: \quad 0+1+2+3 &= 2(2+1)/2+3 \quad \text{(by using the result for } n=2\text{)} \\
&= 3(2/2+1) \\
&= 3(2+2)/2 \\
&= 3(3+1)/2
\end{aligned}
$$

Since this pattern continues indefinitely, the formula is true for every $n \geq 0$.

Now let us criticize this proof. The conclusion, "the formula is true for every $n \geq 0$," is supposed to follow from the fact that "this pattern continues indefinitely." The phrase "this pattern" refers to the calculation that we have done three times, to derive the formula for $n = 1$ from $n = 0$, for $n = 2$ from $n = 1$, and for $n = 3$ from $n = 2$. There are at least two clear deficiencies in the proof. One is that we have not said explicitly what "this pattern" is. The second, which is more serious, is that we have not made any attempt to justify the assertion that it continues indefinitely. In this example, the pattern is obvious enough that people might accept the assertion without much argument. However, it would be fair to say that the most important statement in the proof is the one for which no reasons are given!

Our second version of the proof tries to correct both these problems at once: to describe the pattern precisely by doing the calculation, not just for three particular values of n but for an *arbitrary* value of n, and in the process, to demonstrate that the pattern does not depend on the value of n and therefore *does* continue indefinitely.

■ Example 2.7. Proof Number 2

To show

$$0+1+2+\cdots+n = n(n+1)/2 \qquad \text{for every } n \geq 0$$

$$
\begin{aligned}
n = 0: \quad 0 &= 0(0+1)/2 \\
n = 1: \quad 0+1 &= 0(0+1)/2+1 \quad \text{(by using the result for } n=0\text{)} \\
&= 1(0/2+1) \\
&= 1(0+2)/2 \\
&= 1(1+1)/2 \\
n = 2: \quad 0+1+2 &= 1(1+1)/2+2 \quad \text{(by using the result for } n=1\text{)} \\
&= 2(1/2+1) \\
&= 2(1+2)/2 \\
&= 2(2+1)/2 \\
n = 3: \quad 0+1+2+3 &= 2(2+1)/2+3 \quad \text{(by using the result for } n=2\text{)} \\
&= 3(2/2+1) \\
&= 3(2+2)/2 \\
&= 3(3+1)/2
\end{aligned}
$$

In general, for any value of $k \geq 0$, the formula for $n = k+1$ can be derived from the one for $n = k$ as follows:

$$\begin{aligned} 0+1+2+\cdots+(k+1) &= (0+1+\cdots+k)+(k+1) \\ &= k(k+1)/2+(k+1) \quad \text{(from the result for } n=k\text{)} \\ &= (k+1)(k/2+1) \\ &= (k+1)(k+2)/2 \\ &= (k+1)((k+1)+1)/2 \end{aligned}$$

Therefore, the formula holds for every $n \geq 0$.

We might now say that the proof has more than it needs. Presenting the calculations for three specific values of n originally made it easier for the reader to spot the pattern; now, however, the pattern has been stated explicitly. To the extent that the argument for these three specific cases is taken to be part of the proof, it obscures the two *essential* parts of the proof: (1) checking the formula for the initial value of n, $n = 0$, and (2) showing in general that once we have obtained the formula for one value of n ($n = k$), we can derive it for the next value ($n = k + 1$). These two facts together are what allow us to conclude that the formula holds for every $n \geq 0$. Neither by itself would be enough. (On one hand, the formula for $n = 0$, or even for the first million values of n, might be true just by accident. On the other hand, it would not help to know that we can always derive the formula for the case $n = k+1$ from the one for the case $n = k$, if we could never get off the ground by showing that it *is* actually true for some starting value of k.)

The principle that we have used in this example can now be formulated in general.

The Principle of Mathematical Induction

Suppose $P(n)$ is a statement involving an integer n. Then to prove that $P(n)$ is true for every $n \geq n_0$, it is sufficient to show these two things:

1. $P(n_0)$ is true.
2. For any $k \geq n_0$, if $P(k)$ is true, then $P(k+1)$ is true.

A *proof by induction* is an application of this principle. The two parts of such a proof are called the *basis step* and the *induction* step. In the induction step, we *assume* that k is a number $\geq n_0$ and that the statement $P(n)$ is true in the case $n = k$; we call this assumption the *induction hypothesis*. Let us return to our example one last time in order to illustrate the format of a proof by induction.

■ Example 2.7. Proof Number 3 (by induction)

Let $P(n)$ be the statement

$$1+2+3+\cdots+n = n(n+1)/2$$

To show that $P(n)$ is true for every $n \geq 0$.

Basis step. We must show that $P(0)$ is true. $P(0)$ is the statement $0 = 0(0+1)/2$, and this is obviously true.

Induction hypothesis.

$$k \geq 0 \quad \text{and} \quad 1+2+3+\cdots+k = k(k+1)/2$$

Statement to be shown in induction step.

$$1+2+3+\cdots+(k+1) = (k+1)((k+1)+1)/2$$

Proof of induction step.

$$\begin{aligned} 1+2+3+\cdots+(k+1) &= (1+2+\cdots+k)+(k+1) \\ &= k(k+1)/2+(k+1) \quad \text{(by the induction hypothesis)} \\ &= (k+1)(k/2+1) \\ &= (k+1)(k+2)/2 \\ &= (k+1)((k+1)+1)/2 \end{aligned}$$

Whether or not you follow this format exactly, it is advisable always to include in your proof explicit statements of the following:

- The general statement involving n that is to be proved.
- The statement to which it reduces in the basis step (the general statement, but with n_0 substituted for n).
- The induction hypothesis (the general statement, with k substituted for n, and preceded by "$k \geq n_0$, and").
- The statement to be shown in the induction step (with $k+1$ substituted for n).
- The point during the induction step at which the induction hypothesis is used.

The advantage of formulating a general principle of induction is that it supplies a general framework for proofs of this type. If you read in a journal article the phrase "It can be shown by induction that . . . ," even if the details are missing, you can supply them. Although including these five items explicitly may seem laborious at first, the advantage is that it can help you to clarify for yourself exactly what you are trying to do in the proof. Very often, once you have gotten to this point, filling in the remaining details is a straightforward process.

EXAMPLE 2.8 Strings of the Form 0*y*1 Must Contain the Substring 01

Let us prove the following statement: For any $x \in \{0, 1\}^*$, if x begins with 0 and ends with 1 (i.e., $x = 0y1$ for some string y), then x must contain the substring 01.

You may wonder whether this statement requires an induction proof; let us begin with an argument that does not involve induction, at least explicitly. If $x = 0y1$ for some string $y \in \{0, 1\}^*$, then x must contain at least one 1. The *first* 1 in x cannot occur at the beginning, since x starts with 0; therefore, the first 1 must be immediately preceded by a 0, which means that x contains the substring 01. It would be hard to imagine a proof much simpler than this, and it seems convincing. It is interesting to observe, however, that this proof uses a fact about natural numbers (every nonempty subset has a smallest element) that is equivalent to

the principle of mathematical induction. We will return to this statement later, when we have a slightly modified version of the induction principle. See Example 2.12 and the discussion before that example.

In any case, we are interested in illustrating the principle of induction at least as much as in the result itself. Let us try to construct an induction proof. Our initial problem is that mathematical induction is a way of proving statements of the form "For every $n \geq n_0, \ldots,$" and our statement is not of this form. This is easy to fix, and the solution was suggested at the beginning of this section. Consider the statement $P(n)$: If $|x| = n$ and $x = 0y1$ for some string $y \in \{0, 1\}^*$, then x contains the substring 01. In other words, we are introducing an integer n into our statement, specifically in order to use induction. If we can prove that $P(n)$ is true for every $n \geq 2$, it will follow that the original statement is true. (The integer we choose is the length of the string, and we could describe the method of proof as *induction on the length of the string*. There are other possible choices; see Exercise 2.6.)

In the basis step, we wish to prove the statement "If $|x| = 2$ and $x = 0y1$ for some string $y \in \{0, 1\}^*$, then x contains the substring 01." This statement is true, because if $|x| = 2$ and $x = 0y1$, then y must be the null string Λ, and we may conclude that $x = 01$. Our induction hypothesis will be the statement: $k \geq 2$, and if $|x| = k$ and $x = 0y1$ for some string $y \in \{0, 1\}^*$, then x contains the substring 01. In the induction step, we must show: if $|x| = k + 1$ and $x = 0y1$ for some $y \in \{0, 1\}^*$, then x contains the substring 01. (These three statements are obtained from the original statement $P(n)$ very simply: first, by substituting 0 for n; second, by substituting k for n, and adding the phrase "$k \geq 2$, and" at the beginning; third, by substituting $k + 1$ for n. These three steps are always the same, and the basis step is often as easy to prove as it is here. Now the mechanical part is over, and we must actually think about how to continue the proof!)

We have a string x of length $k + 1$, about which we want to prove something. We have an induction hypothesis that tells us something about certain strings of length k, the ones that begin with 0 and end with 1. In order to apply the induction hypothesis, we need a string of length k to apply it to. We can get a string of length k from x by leaving out one symbol. Let us try deleting the initial 0. (See Exercise 2.5.) The remainder, $y1$, is certainly a string of length k, and we know that it ends in 1, but it may not begin with 0—and we can apply the induction hypothesis only to strings that do. However, if $y1$ does not begin with 0, it must begin with 1, and in this case x starts with the substring 01! If $y1$ does begin with 0, then the induction hypothesis tells us that it must contain the substring 01, so that $x = 0y1$ must contain the substring too.

Now that we have figured out the crucial steps, we can afford to be a little more concise in our official proof. We are trying to prove that for every $n \geq 2$, $P(n)$ is true, where $P(n)$ is the statement: If $|x| = n$ and $x = 0y1$ for some string $y \in \{0, 1\}^*$, then x contains the substring 01.

Basis step. We must show that the statement $P(2)$ is true. $P(2)$ says that if $|x| = 2$ and $x = 0y$ for some $y \in \{0, 1\}^*$, then x contains the substring 01. $P(2)$ is true, because if $|x| = 2$ and $x = 0y1$ for some y, then $x = 01$.

Induction hypothesis. $k \geq 2$ and $P(k)$; in other words, if $|x| = k$ and $x = 0y1$ for some $y \in \{0, 1\}^*$, then x contains the substring 01.

Statement to be shown in induction step. $P(k + 1)$; that is, if $|x| = k + 1$ and $x = 0y1$ for some $y \in \{0, 1\}^*$, then x contains the substring 01.

Proof of induction step. Since $|x| = k + 1$ and $x = 0y1$, $|y1| = k$. If y begins with 1, then x begins with the substring 01. If y begins with 0, then $y1$ begins with 0 and ends with 1; by the induction hypothesis, y contains the substring 01, and therefore x does also.

EXAMPLE 2.9 Verifying a Portion of a Program

The program fragment below is written in pseudocode. Lowercase letters represent constants, uppercase letters represent variables, and the constant n is assumed to be nonnegative:

```
Y = 1;
for I = 1 to n
   Y = Y * x;
write(Y);
```

We would like to show that when this code is executed, the value printed out is x^n. We do this in a slightly roundabout way, by introducing a new integer j, the number of iterations of the loop that have been performed. Let $P(j)$ be the statement that the value of Y after j iterations is x^j. The result we want will follow from the fact that $P(j)$ is true for any $j \geq 0$, and the fact that "For $I = 1$ to n" results in n iterations of the loop.

Basis step. $P(0)$ is the statement that after 0 iterations of the loop, Y has the value x^0. This is true because Y receives the initial value 1 and after 0 iterations of the loop its value is unchanged.

Inductive hypothesis. $k \geq 0$, and after k iterations of the loop the value of Y is x^k.

Statement to be proved in induction step. After $k + 1$ iterations of the loop, the value of Y is x^{k+1}.

Proof of induction step. The effect of the assignment statement $Y = Y * x$ is to replace the old value of Y by that value times x; therefore, the value of Y after any iteration is x times the value before that iteration. Since $x * x^k = x^{k+1}$, the proof is complete.

Although the program fragment in this example is very simple, the example should suggest that the principle of mathematical induction can be a useful technique for verifying the correctness of programs. For another example, see Exercise 2.56.

You may occasionally find the principle of mathematical induction in a disguised form, which we could call the *minimal counterexample principle*. The last example in this section illustrates this.

EXAMPLE 2.10 A Proof Using the Minimal Counterexample Principle

To show: For every integer $n \geq 0$, $5^n - 2^n$ is divisible by 3.

Just as in an ordinary induction proof, we begin by checking that $P(n)$ is true for the starting value of n. This is true here, since $5^0 - 2^0 = 1 - 1 = 0$, and 0 is divisible by 3. Now if it is *not* true that $P(n)$ is true for every $n \geq 0$, then there are values of n greater than or equal to 0 for which $P(n)$ is false, and therefore there must be a smallest such value, say $n = k$.

(See Example 2.12.) Since we have verified $P(0)$, k must be at least 1. Therefore, $k-1$ is at least 0, and since k is the smallest value for which P fails, $P(k-1)$ is true. This means that $5^{k-1}-2^{k-1}$ is a multiple of 3, say $3j$. Then, however,

$$5^k - 2^k = 5 * 5^{k-1} - 2 * 2^{k-1} = 3 * 5^{k-1} + 2 * (5^{k-1} - 2^{k-1}) = 3 * 5^{k-1} + 2 * 3j$$

This expression is divisible by 3. We have derived a contradiction, which allows us to conclude that our original assumption is false. Therefore, $P(n)$ is true for every $n \geq 0$.

You can probably see the similarity between this proof and one that uses the principle of mathematical induction. Although an induction proof has the advantage that it does not involve proof by contradiction, both approaches are equally valid.

Not every statement involving an integer n is appropriate for mathematical induction. Using this technique on the statement

$$(2^n + 1)(2^n - 1) = 2^{2n} - 1$$

would be silly because the proof of the induction step would not require the induction hypothesis at all. The formula for $n = k+1$, or for any other value, can be obtained immediately by expanding the left side of the formula and using laws of exponents. The proof would not be a *real* induction proof, and it would be misleading to classify it as one.

A general rule of thumb is that if you are tempted to use a phrase like "Repeat this process for each n," or "Since this pattern continues indefinitely" in a proof, there is a good chance that the proof can be made more precise by using mathematical induction. When you encounter one of these phrases while reading a proof, it is very likely a substitute for an induction argument. In this case, supplying the details of the induction may help you to understand the proof better.

2.3 THE STRONG PRINCIPLE OF MATHEMATICAL INDUCTION

Sometimes, as in our first example, a proof by mathematical induction is called for, but the induction principle in Section 2.2 is not the most convenient tool.

EXAMPLE 2.11 Integers Bigger Than 2 Have Prime Factorizations

Recall that a *prime* is a positive integer, 2 or bigger, that has no positive integer divisors except itself and 1. Part of the fundamental theorem of arithmetic is that every integer can be factored into primes. More precisely, let $P(n)$ be the statement that n is either prime or the product of two or more primes; we will try to prove that $P(n)$ is true for every $n \geq 2$.

The basis step does not present any problems. $P(2)$ is true, since 2 is a prime. If we proceed as usual, then we take as the induction hypothesis the statement that $k \geq 2$ and k is either prime or the product of two or more primes. We would like to show that $k+1$ is either prime or the product of primes. If $k+1$ happens to be prime, there is nothing left to prove. Otherwise, by the definition of prime, $k+1$ has some positive integer divisor other than itself

and 1. This means $k + 1 = r * s$ for some positive integers r and s, neither of which is 1 or $k + 1$. It follows that r and s must both be greater than 1 and less than $k + 1$.

In order to finish the induction step, we would like to show that r and s are both either primes or products of primes; it would then follow, since $k + 1$ is the product of r and s, that $k + 1$ is a product of two or more primes. Unfortunately, the only information our induction hypothesis gives us is that k is a prime or a product of primes, and this tells us nothing about r or s.

Consider, however, the following intuitive argument, in which we set about verifying the statement $P(n)$ one value of n at a time:

2 is a prime.

3 is a prime.

$4 = 2 * 2$, which is a product of primes since $P(2)$ is known to be true.

5 is a prime.

$6 = 2 * 3$, which is a product of primes since $P(2)$ and $P(3)$ are known to be true.

7 is a prime.

$8 = 2 * 4$, which is a product of primes since $P(2)$ and $P(4)$ are known to be true.

$9 = 3 * 3$, which is a product of primes since $P(3)$ is known to be true.

$10 = 2 * 5$, which is a product of primes since $P(2)$ and $P(5)$ are known to be true.

11 is a prime.

$12 = 2 * 6$, which is a product of primes since $P(2)$ and $P(6)$ are known to be true.

...

This seems as convincing as the intuitive argument given at the start of Example 2.7. Furthermore, we can describe explicitly the pattern illustrated by the first 11 steps: For each $k \geq 2$, either $k + 1$ is prime or it is the product of two numbers r and s for which the proposition P has already been shown to hold.

The difference between the pattern appearing here and the one we saw in Example 2.7 is this: At each step in the earlier example we were able to obtain the truth of $P(k + 1)$ by knowing that $P(k)$ was true, and here we need to know that P holds, not only for k but also for all the values up to k. The following modified version of the induction principle will allow our proof to proceed.

The Strong Principle of Mathematical Induction

Suppose $P(n)$ is a statement involving an integer n. Then to prove that $P(n)$ is true for every $n \geq n_0$, it is sufficient to show these two things:

1. $P(n_0)$ is true.
2. For any $k \geq n_0$, if $P(n)$ is true for every n satisfying $n_0 \leq n \leq k$, then $P(k + 1)$ is true.

To use this principle in a proof, we follow the same steps as before except for the way we state the induction hypothesis. The statement here is that k is some integer $\geq n_0$ and that *all* the statements $P(n_0)$, $P(n_0 + 1), \ldots, P(k)$ are true. With this change, we can finish the proof we began earlier.

■ Example 2.11. Proof by induction.

To show: $P(n)$ is true for every $n \geq 2$, where $P(n)$ is the statement: n is either a prime or a product of two or more primes.

Basis step. $P(2)$ is the statement that 2 is either a prime or a product of two or more primes. This is true because 2 is a prime.

Induction hypothesis. $k \geq 2$, and for every n with $2 \leq n \leq k$, n is either prime or a product of two or more primes.

Statement to be shown in induction step. $k + 1$ is either prime or a product of two or more primes.

Proof of induction step. We consider two cases. If $k + 1$ is prime, the statement $P(k + 1)$ is true. Otherwise, by definition of a prime, $k + 1 = r * s$, for some positive integers r and s, neither of which is 1 or $k + 1$. It follows that $2 \leq r \leq k$ and $2 \leq s \leq k$. Therefore, by the induction hypothesis, both r and s are either prime or the product of two or more primes. Therefore, their product $k + 1$ is the product of two or more primes, and $P(k + 1)$ is true.

The strong principle of induction is also referred to as the principle of *complete* induction, or *course-of-values* induction. The first example suggests that it is as plausible intuitively as the ordinary induction principle, and in fact the two are equivalent. As to whether they are *true*, the answer may seem a little surprising. Neither can be proved using other standard properties of the natural numbers. (Neither can be disproved, either!) This means, in effect, that in order to use the induction principle, we must adopt it as an axiom. A well-known set of axioms for the natural numbers, the *Peano* axioms, includes one similar to the induction principle.

Twice in Section 2.2 we had occasion to use the *well-ordering* principle for the natural numbers, which says that every nonempty subset of $\mathcal{N}$ has a smallest element. As obvious as this statement probably seems, it is also impossible to prove without using induction or something comparable. In the next example, we show that it follows from the strong principle of induction. (It can be shown to be equivalent.)

The Well-ordering Principle for the Natural Numbers — EXAMPLE 2.12

To prove: Every nonempty subset of $\mathcal{N}$, the set of natural numbers, has a smallest element. (What we are actually proving is that if the strong principle of mathematical induction is true, then every nonempty subset of $\mathcal{N}$ has a smallest element.)

First we need to find a way to express the result in the form "For every $n \geq n_0$, $P(n)$." Every nonempty subset A of $\mathcal{N}$ contains a natural number, say n. If every subset of $\mathcal{N}$

containing n has a smallest element, then A does. With this in mind, we let $P(n)$ be the statement "Every subset of $\mathcal{N}$ containing n has a smallest element." We prove that $P(n)$ is true for every $n \geq 0$. (See Exercise 2.7.)

Basis step. $P(0)$ is the statement that every subset of $\mathcal{N}$ containing 0 has a smallest element. This is true because 0 is the smallest natural number and therefore the smallest element of the subset.

Induction hypothesis. $k \geq 0$, and for every n with $0 \leq n \leq k$, every subset of $\mathcal{N}$ containing n has a smallest element. (Put more simply, $k \geq 0$ and every subset of $\mathcal{N}$ containing an integer less than or equal to k has a smallest element.)

Statement to be shown in induction step. Every subset of $\mathcal{N}$ containing $k+1$ has a smallest element.

Proof of induction step. Let A be any subset of $\mathcal{N}$ containing $k+1$. We consider two cases. If A contains no natural number less than $k+1$, then $k+1$ is the smallest element of A. Otherwise, A contains some natural number n with $n \leq k$. In this case, by the induction hypothesis, A contains a smallest element.

The strong principle of mathematical induction is more appropriate here, since when we come up with an n to which we want to apply the induction hypothesis, all we know about n is that $n \leq k$. We do *not* know that $n = k$. It may not be obvious at the beginning of an induction proof whether the strong induction principle is required or whether you can get by with the original version. You can avoid worrying about this by *always* using the strong version. It allows you to adopt a stronger induction hypothesis, and so if an induction proof is possible at all, it will certainly be possible with the strong version. In any case, you can put off the decision until you reach the point where you have to prove $P(k+1)$. If you can do this with only the assumption that $P(k)$ is true, then the original principle of induction is sufficient. If you need information about earlier values of n as well, the strong version is needed.

We will see more examples of how the strong principle of mathematical induction is applied once we have discussed recursive definitions and the close relationship between them and mathematical induction.

2.4 | RECURSIVE DEFINITIONS

2.4.1 Recursive Definitions of Functions with Domain $\mathcal{N}$

The chances are that in a programming course you have seen a translation into some high-level programming language of the following definition:

$$n! = \begin{cases} 1 & \text{if } n = 0 \\ n * (n-1)! & \text{if } n > 0 \end{cases}$$

This is one of the simplest examples of a recursive, or inductive, definition. It defines the factorial function on the set of natural numbers, first by defining the value at 0,