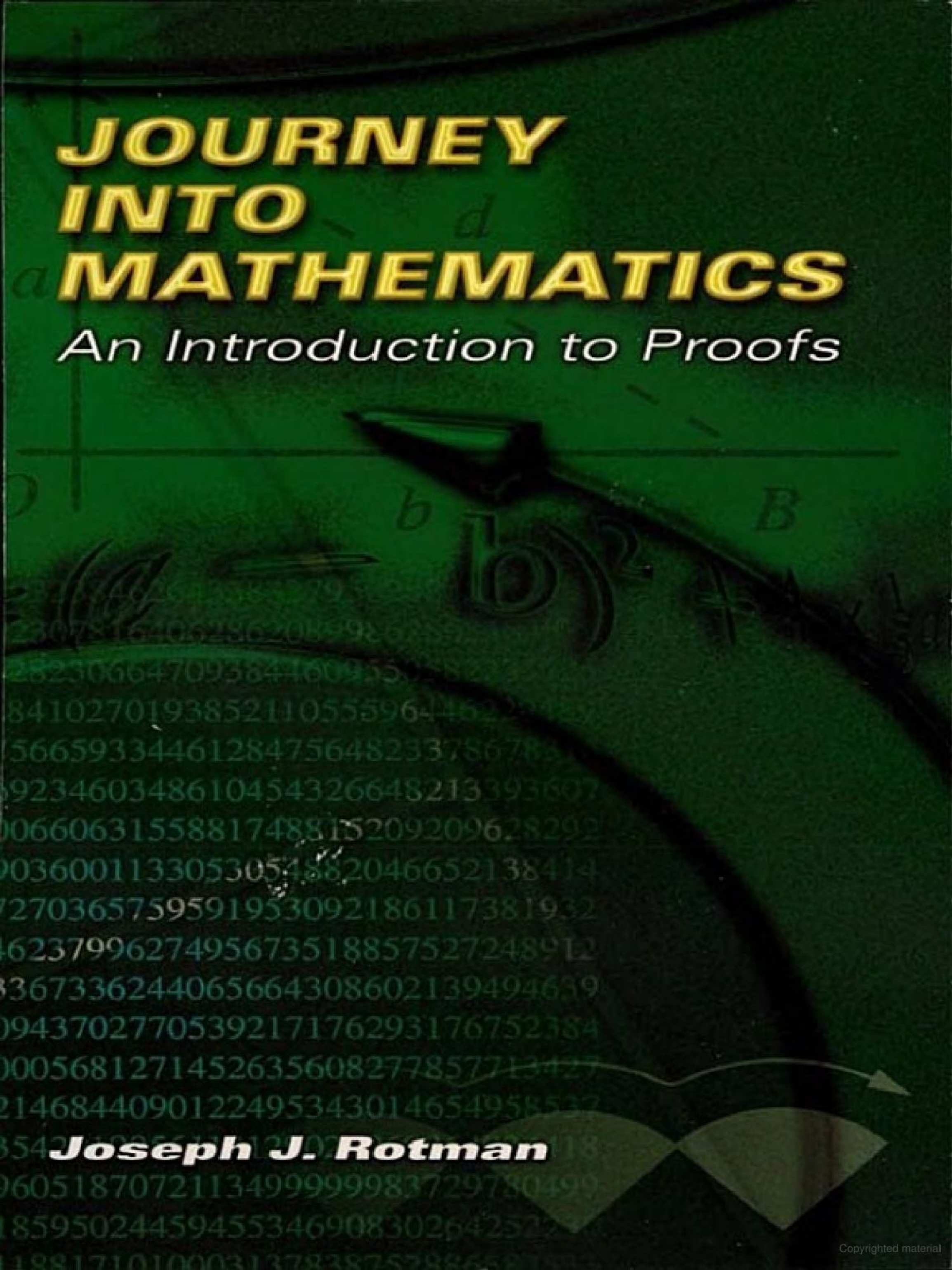


JOURNEY INTO MATHEMATICS

An Introduction to Proofs

Joseph J. Rotman

Copyright

Bibliographical Note

This Dover edition, first published in 2007, is a corrected republication of the first edition of the work originally published by Prentice Hall, Inc., Upper Saddle River, New Jersey, in 1998.

Readers of this book who would like to receive the solutions to the exercises may request them from the publisher at the following e-mail address: **editors@doverpublications.com**

9780486151687

Manufactured in the United States of America
Dover Publications, Inc., 31 East 2nd Street, Mineola, N.Y 11501

Table of Contents

Preface

Instructors have observed, when teaching junior level courses in abstract algebra, number theory, or real variables, that many students have difficulty out of proportion to the level of difficulty of the material. In an abstract algebra course introducing groups and rings, students' struggles are not affected by the changing of texts, instructors, or the order of presentation. Similarly, experimenting with courses in real variables (say, by treating only functions of a single variable instead of functions on euclidean n-space) offers little relief. The cause of this problem is plain when one considers the previous mathematics courses. The standard calculus sequence is presented, nowadays, to students from various disciplines who have different backgrounds, abilities, and goals, with the aim of teaching them how to differentiate, how to integrate, and how to use these techniques to solve problems. Theorems are stated but usually not proved; hypotheses of theorems are often not verified before applying the theorems (e.g., does one always check whether a given function is continuous?); definitions are given (e.g., limit and convergence) but not taken seriously. After two years of such "mathematics," is it any wonder that a junior-level student is woefully unprepared to read and do real mathematics?

There are two possible solutions to this problem. The obvious solution, revise the calculus sequence, is impractical. Many have tried; many are trying. I wish success to those still fighting the good fight, but I am pessimistic about there being a revolution in undergraduate mathematics, and I am even more pessimistic about there being such a revolution tomorrow. After all, scientists and engineers cannot afford the extra time before using calculus in their own domains, and so calculus courses are necessarily compromises between teaching the techniques of calculus and teaching an understanding of its principles.

My solution is a one semester intermediate course between calculus and the first courses in abstract algebra and real variables. This is not a new idea. There are many such “transition courses” designed to prepare students for junior-level courses, but they emphasize the elements of logic (from modus ponens and truth tables through quantifiers) and set theory (from Boolean operations through relations and functions). I find this material rather dull and uninspiring, and I imagine that this feeling is shared by most students. Of course, these things should be learned eventually; as Hermann Weyl wrote, “Logic is the hygiene that the mathematician practices to keep his ideas healthy and strong.” It is cruel, however, to inflict an entire course comprised of such things on defenseless students. And it doesn’t work; my unscientific observations indicate that those students who survive such tedious material do not fare any better in subsequent courses than do those who were spared. George Pólya wrote, “When introduced at the wrong time or place, good logic may be the worst enemy of good teaching.” I have attached an appendix, Glossary of Logic, at the end of the book, covering much of this material. Although this section is too brief an account to qualify as a text for a standard course, it is a place where readers can look to resolve the usual questions that tend to arise.

An introductory course should contain valuable material, it must be interesting, and it must give a fairly accurate picture of what mathematics really is and what mathematicians do. One learns how to read and write proofs by reading and writing them; merely reading *about* mathematics is not an adequate substitute for actually doing mathematics. This book begins with some very elementary mathematics – induction, binomial coefficients, and polygonal areas – because, when starting out, readers must be allowed to focus on the writing and reading of proofs without the distraction of absorbing unfamiliar ideas at the same time. From the outset, complete proofs are given to serve as models for the reader. The presentation is a coherent story, with historical and etymological asides, because it is more interesting and more natural to watch a subject grow and develop. The journey continues with elementary area problems, the irrationality of $\sqrt{2}$, the Pythagorean theorem, Pythagorean triples after Diophantus, and trigonometry. The Diophantine method of finding Pythagorean triples by parametrizing the circle with rational functions is extended to finding such parametrizations of other conic sections, and this leads to a glimpse of elliptic integrals. Next, one passes to disks, proving the area and circumference formulas (due to Eudoxus and Archimedes, respectively) essentially in the classical way. This early notion of approximation is subtle, but it is digestible because one can see areas of inscribed polygons approaching the area of the disk. One notes afterward, however, that this early notion has defects. Seeing how convergence remedies defects of the classical notion gives the reader a better understanding and appreciation of the modem definition of limit. We then see why $(-1) \times (-1) = +1$, discuss the quadratic formula, complex numbers, De Moivre’s theorem, the cubic formulas (Cardano’s version in terms of radicals as well as Viète’s trigonometric version), discriminants, and the quartic formula. The text ends with proofs of the irrationality of e, the irrationality of some specific values of sine and cosine, and the irrationality of π. Thus, geometry, algebra, number theory, and analysis are all intertwined. The journey travels a road from humble beginnings to a fairly sophisticated destination. I hope that students and instructors will enjoy this text, and that it will serve the several aims set forth for it.

I thank Paul Bateman, Richard Bishop, Peter Braunfeld, Everett

Dade, Heini Halberstam, Carl Jockusch, Daniel Saltz, Donald Sherbert, and Kenneth Stolarsky for their excellent suggestions. I give special thanks to Philippe Tondeur, whose notes on similar material were the starting point of this text, and I also give special thanks to J.-P Tignol, for permitting me to quote an excerpt from his wonderful book about Galois's theory of equations. I give extra special thanks to John Wetzel and Leon McCulloh who made many fine suggestions as they were teaching from a preliminary version of this text; they generously allowed me to use the ones I liked, and they did not complain in the rare cases when I did not use one. It also gives me great pleasure to thank my daughter, Ella Rose, for drawing and producing all the figures in the book.

I thank Joyce Woodworth for an excellent job of typing my manuscript in LaTeX, and I thank Adam Lewenberg for his expert help in the final stages of LaTeX typesetting. Lastly, I thank the reviewers for their good suggestions:

Linda A. Bolte, Eastern Washington University
Thomas G. Clarke, North Carolina A&T State University
Donald Nowlin, Eastern Washington University
Michael Stecher, Texas A&M University
David Walnut, George Mason University.

Joseph Rotman

TO THE READER

> Histories make men wise; poets, witty; the mathematics, subtile; natural philosophy, deep; moral, grave; logic and rhetoric, able to contend.
>
> *Francis Bacon*

One of the main purposes of this book is to help you learn how to read and write proofs. To further this aim, much of the early material is familiar (even at the beginning, however, there are new and interesting things) to allow you to focus on giving complete and clear proofs without distractions.

A proof is an explanation why something is true. There is a notion of formal proof, which is essentially an explanation to a machine, but we are concerned here with giving proofs to humans. Just as one does not give the same explanation to a ten-year old that one gives to an adult, one's proof, one's explanation, depends on whom one is speaking to. The audience for all of your proofs is not your instructor (who already knows the reasons!); your explanations are to be directed towards students in class, one of whom is yourself. Adequate reasons must be given to defend assertions against any possible objection; on the other hand, there is no need to explain why 3 = 3. Try your best to say enough to persuade, and try your best not to put others to sleep by belaboring the obvious. One role of the proofs in the text is to serve as models for your own proofs. Because one becomes more sophisticated as one learns, the proofs in the text also change;

certain points made explicit in the beginning are later left unsaid.

Some people think that a proof must be full of symbols, looking like ancient Egyptian hieroglyphics. Not so. Look in any mathematics book, and you will find words. Your proofs should be written in complete sentences. Of course, you may use symbols and pictures if necessary, but remember that a symbol is like a pronoun; it means nothing unless it is specified. Just as you wouldn't begin a story by saying, "He gave some of it to him there," you must not begin a proof by saying that $x = y^2$ without telling what x and y are (are they numbers? real? rational? integers? positive?).

Even though the context of this course is largely elementary, do not be lulled into thinking that it is an easy course with an inevitable grade of A at its conclusion. There are challenges within. If one wants the reward Bacon mentions, then there is no alternative but to do some mathematics. The journey may have some difficulties, but its goals are valuable. As Bacon says, the reward is understanding, subtlety and, we may add, pleasure.

TO THE INSTRUCTOR

There are several aims of this text:

to teach students how to read and write proofs;
to teach some valuable mathematics;
to show how attractive mathematics is.

Divide the course into three parts. Part I covers the first two chapters. Because students are learning the mechanics of writing proofs, one should proceed slowly. Students write proofs from the outset, using the proofs in the text, as well as the instructor's exposition, as models for their own proofs. Each student is assigned an exercise very early in the term that must be presented before the class. He or she has several days to prepare it, and prior discussion with other students and with me is encouraged. I assign no grade to the performance, the student is allowed to use any notes, and the class is allowed to heckle (I try not to make any comments until the end of the presentation). This exercise reinforces the notion that proofs are designed as explanations to the class. Regular homework assignments should be graded for presentation as well as for correctness.

Part II of the course covers Chapter 3. Some of the synthetic geometry in proving the area and circumference formulas can be done lightly, if desired. The basic idea of the chapter is to introduce convergence of sequences only after students have acquired some geometric experience with the simpler classical notion of limit.

Part III covers Chapter 4; the most important material is complex numbers and their application, via De Moivre’s theorem, to real numbers.

There will not be sufficient time in most courses to cover all the material in the text. The instructor should decide on what material to omit consistent with his or her goals for the course. I have found that students (and I) enjoy the historical and etymological asides, but I do not discuss them in class unless a question arises.

Chapter 1

Setting Out

INDUCTION

> So, naturalists observe, a flea
> Hath smaller fleas that on him prey;
> And these have smaller still to bite ’em;
> And so proceed ad infinitum.
>
> *Jonathan Swift*

There are many styles of proof, and mathematical induction is one of them. We begin by saying what mathematical induction is not. In the natural sciences, ***inductive reasoning*** is based on the principle that a frequently observed phenomenon will always occur. Thus, one says that the sun will rise tomorrow morning because, from the dawn of time, the sun has risen every morning. This is not a legitimate kind of proof in mathematics, for even though a phenomenon occurs frequently, it may not occur always.

Inductive reasoning is valuable in mathematics, because seeing patterns often helps in guessing what may be true. On the other hand, inductive reasoning is not adequate for proving theorems. Before we see examples, let us make sure we agree on the meaning of some standard terms.

Definition. An ***integer*** is one of 0, 1, – 1, 2, – 2, 3, – 3, . · · ·

Definition. An integer $p \geq 2$ is called a ***prime number***[1] if its only positive divisors are 1 and p. An integer $m \geq 2$ which is not prime is called ***composite***.

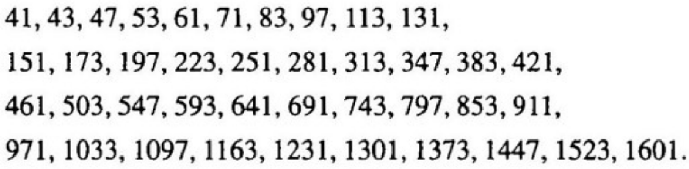
A positive integer m is composite if it has a factorization $m = ab$, where $a < m$ and $b < m$ are positive integers; the inequalities are present to eliminate the uninteresting factorization $m = m x$ 1. Notice that $a \geq 2$: since a is a positive integer, the only other option is a = 1, which implies $b = m$ (contradicting $b < m$); similarly, $b \geq 2$.

The first few primes are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41. That this sequence never ends is proved in Exercise 2.10.

Consider the statement:

$$f(n) = n^2 - n + 41 \text{ is a prime number for every } n \geq 1$$

(this is really a whole family of statements, one for each positive integer n). As we evaluate $f(n)$ for $n = 1, 2, 3, 4, \cdots, 40$, we obtain the following values:



41, 43, 47, 53, 61, 71, 83, 97, 113, 131,
151, 173, 197, 223, 251, 281, 313, 347, 383, 421,
461, 503, 547, 593, 641, 691, 743, 797, 853, 911,
971, 1033, 1097, 1163, 1231, 1301, 1373, 1447, 1523, 1601.

It is tedious but not difficult (see Exercise 1.7) to prove that every one of these numbers is prime. Can we now conclude that *all* the numbers of the form $f(n)$ are prime? For example, is the next number $f(41) = 1681$ prime? The answer is no: $f(41) = 412 – 41 + 41 = 41^2$, which obviously factors, and hence $f(41)$ is not prime.

Here is a more spectacular example (which I first saw in an article by W. Sierpinski). A ***perfect square*** is an integer of the form a^2 for some positive integer a; the first few perfect squares are: 1, 4, 9, 16, 25, 36, 49. Consider the statements S(n), one for each $n \geq 1$:

$$S(n) : 991n^2 + 1 \text{ is not a perfect square.}$$

It turns out that many of the statements $S(n)$ are true. In fact, the smallest number n for which $S(n)$ is false is

$$\begin{aligned} n &= 12,055,735,790,331,359,447,442,538,767 \\ &\approx 1.2 \times 10^{28}. \end{aligned}$$

The original problem is a special case of ***Pell's equation*** (given a prime *p,* when are there integers m and n with $m^2 = pn^2 + 1$), and there is a way of calculating all possible solutions of it. In fact, an even more spectacular example of Pell's equation involves the prime p = 1, 000, 099; the smallest n for which $1,000,099n^2 + 1$ is a perfect square has 1116 digits.) The latest scientific estimate of the age of the earth is 20 billion (20,000,000,000) years, or about 7.3×10^{12} days, a number very much smaller than 1.2×10^{28}, let alone 10^{1115}. If, starting on the very first day, mankind had verified statement $S(n)$ on the nth day, then there would be, today, as much evidence of the general truth of these statements as there is that the sun will rise tomorrow morning. And yet some statements *S*(*n*) are false!

As a final example, let us consider the following statement, known as ***Goldbach's Conjecture:*** Every even number $m \geq 4$ is a sum of two primes. For example,

$$\begin{aligned}
4 &= 2 + 2\\
6 &= 3 + 3\\
8 &= 3 + 5\\
10 &= 3 + 7 = 5 + 5\\
12 &= 5 + 7\\
14 &= 3 + 11 = 7 + 7\\
16 &= 3 + 13 = 5 + 11\\
18 &= 5 + 13 = 7 + 11\\
20 &= 3 + 17 = 7 + 13\\
22 &= 3 + 19 = 5 + 17 = 11 + 11.
\end{aligned}$$

It would be foolish to demand that all odd numbers be sums of two primes. For example, suppose that $27 = p + q$, where *p* and *q* are primes. If both *p* and *q* are odd, then their sum is even, contradicting 27 being odd. Since the only even prime is 2, we have $27 = 2 + q$, and so $q = 25$ is prime; this contradiction shows that 27 is not a sum of two primes.

No one has ever found a counterexample to Goldbach's conjecture, but neither has anyone ever proved it. At present, the conjecture has been verified for all even numbers $m \leq 10^{13}$ by H. J. J. te Riele and J.-M. Deshouillers. It has been proved by J.-R. Chen (with a simplification by P. M. Ross) that every sufficiently large even number *m* can be written as $p + q$, where *p* is prime and *q* is "almost" a prime; that is, *q* is either prime or a product of two primes. Even with all this positive evidence, however, no mathematician will say that Goldbach's conjecture must, therefore, be true for all even *m.*

We have seen what *mathematical induction* is not; let us now discuss what induction[2] is. Suppose one guesses that all the statements $S(n)$ of a certain sort are true (for example, suppose that $S(n)$ has been observed to be true for many values of n). Induction is a technique of proving that *all* the statements $S(n)$ are, indeed, true. For example, the reader may check that $2^n > n$ for many values of n, but is this inequality true for *every* value of n? We will prove below, using induction, that this is so.

The key idea is just this. Imagine a stairway to the sky; if its first step is white, and if the next step above a white step is also white, then all the steps of the stairway must be white. One can trace this idea back to Levi ben Gershon in 1321. There is an explicit description of induction (cited by Pascal) written by Francesco Maurolico in 1557.

Our discussion is based on the following property of positive integers (usually called the *Well Ordering Principle*).

Least Integer Axiom[3]. Every nonempty collection C of positive integers has a smallest number in it.

Saying that C is *nonempty* merely means that there is at least one integer in the collection C.

The Least Integer Axiom is certainly plausible. Given a nonempty collection C, check whether 1 is a number in C; if it is, then 1 is the smallest number in C. Otherwise, check whether 2 is a number in C; if it is, then 2 is the smallest number in C; if not, check whether 3 is a number in C. Continue this procedure; since there is some number in C, we will eventually bump into it.

The Least Integer Axiom can be restated in a more useful way.

Theorem 1.1 (Least Criminal). Let $S(n)$ be a family of statements, where n varies over some nonempty collection of positive integers. If some of these statements are false, then there is a first false statement.

Proof. Let C be the collection of all those positive integers n for which $S(n)$ is false; by hypothesis, C is nonempty. The Least Integer Axiom provides a smallest number m in C, and $S(m)$ is the first false statement.

Theorem 1.2. Every integer $n \geq 2$ is either a prime or a product of primes.

Proof. Were this not so, there would be a "least criminal" *m;* that is, $m \geq 2$, m is neither a prime nor a product of primes, and m is the smallest such integer. Since m is not a prime, it is composite: there is a factorization $m = ab$ with $a < m$ and $b < m$. Because m is the least criminal, both a and b are "honest"; i.e., $a = pp'p“$ for primes p, p’, p”, $\cdots$, and $b = qq'q'' \cdots$ for primes $q, q', q'',$ $. \cdots$ Therefore, $m = pp'\,p'' \cdots\ qq'q'' \cdots$ is a product of (at least two) primes, a contradiction.

Mathematical induction is a version of Least Criminal that is usually more convenient to use.

Theorem 1.3 (Mathematical Induction). Let $S(n)$ be a family of statements, one for each $n \geq 1$, and suppose that:

(i) $S(1)$ is true, and

(ii) if $S(n)$ is true, then $S(n + 1)$ is true.

Then $S(n)$ is true for every $n \geq 1$.

Proof. It suffices to show that there are no integers n for which $S(n)$ is false; that is, it suffices to show that the collection

$$C = \text{all positive integers } n \text{ for which } S(n) \text{ is false}$$

is empty.

If, on the contrary, C is nonempty, then there is a first false statement, say, $S(m)$. Since S(1) is true, by (i), we must have $m \geq 2$. This implies that $m - 1 \geq 1$, and so there is an $(m - 1)$st statement $S(m - 1)$ [there is no statement $S(0)$]. As m is the least criminal, $m - 1$ must be honest; that is, $S(m - 1)$ is true. But (ii) says that $S(m) = S([m - 1] + 1)$ is also true, and this is a contradiction. We conclude that C is empty and, hence, that all the statements are true.

Before we illustrate how to use mathematical induction, let us make sure we can manipulate inequalities. We recall that if two real numbers a and b are both positive, i.e., $a > 0$ and $b > 0$, then ab, $a + b$ and $1/a$ are also positive. On the other hand, the product of a positive number and a negative number is negative.

Definition. For any two real numbers c and d, define

$$d < c$$

to mean that $c - d$ is positive. We write $d \leq c$ to mean either $d < c$

or $d = c$.

Notice that if $a > b$ and $b >$ c, then $a > c$ [for $a - c = (a - b) + (b - c)$ is a sum of positive numbers and, hence, is itself positive]. One often abbreviates these two inequalities as $a > b > c$. The reader may check that if $a > b \geq c$, then $a > c$.

Theorem 1.4. Assume that $b < B$ are real numbers.

(i) If m is positive, then $mb < mB$, whereas if m is negative, then $mb > mB$.

(ii) For any number N, positive, negative, or zero, we have

$$N + b < N + B \text{ and } N - b > N - B.$$

(iii) Let c and d be positive numbers. If $d <$ c, then $1/d > 1/c$, and, conversely, if $1/c < 1/d$, then $c > d$.

Proof. (i) By hypothesis, $B - b > 0$. If $m > 0$, then the product of positive numbers being positive implies that m($B - b$) = $mB - mb$ is positive; that is, $mb < mB$. *If* $m < 0$, then the product m($B - b$) = $mB - mb$ is negative; that is, $mB < mb$.

(ii) The difference $(N + B) - (N + b)$ is positive, for it equals $B - b$. For the other inequality, $(N - b) - (N - B) = - b + B$ is positive, and, hence, $N - b > N - B$.

(iii) If $d < c$, then $c - d$ is positive. Hence, $1/d$-$1/c = (c - d)/cd$ is positive, being the product of the positive numbers $c - d$ and $1/cd$ (by hypothesis, both c and d are positive). Therefore, $1/d > 1/c$. Conversely, if $1/c < 1/d$, then part (i) gives $d = cd(1/c) < cd(1/d) =$

c; that is, $c > d$.

To illustrate, since 3 < 4, we have

$$9 \times 3 = 27 < 36 = 9 \times 4;$$

(i)

$$(-9) \times 3 = -27 > -36 = (-9) \times 4;$$
$$9 + 3 = 12 < 13 = 9 + 4;$$
$$9 - 3 = 6 > 5 = 9 - 4;$$

(ii)

$$\tfrac{1}{4} = 0.25 < 0.33 < \tfrac{1}{3}.$$

(iii)

It is always a good idea to see concrete examples of a theorem, for it makes the result more understandable by putting flesh on the bones of the statement. This is the first step in appreciating what a theorem means, and so it is an important habit to cultivate. Indeed, mathematics must be read with pencil and paper. If no example is given in a text, supply your own. There is an apocryphal story of a theorem so general that no particular case is known. Such a theorem would be bad mathematics.

Theorem 1.5. $2^n > n$ for all $n \geq 1$.

Proof. Regard this inequality as a sequence of statements, where the *n*th statement *S*(*n*) is:

$$S(n) : 2^n > n.$$

There are two steps required for mathematical induction.

Base step: The initial statement

$$S(1) : 2^1 > 1$$

is true, for $2^1 = 2 > 1$.

Inductive step: If *S*(*n*) is true, then *S*(*n* + 1) is also true; that is, if one uses the ***inductive hypothesis*** S(*n*) : “$2^n > n$ is a valid inequality,” then one can prove

$$S(n+1): 2^{n+1} > n + 1.$$

First, multiply both sides of $2^n > n$ by 2; Theorem 1.4(i) gives

$$2^{n+1} = 2 \times 2^n > 2\mathrm{n}.$$

Now $2^\mathrm{n} = n + n \geq n + 1$ (because $n \geq 1$); therefore, $2^{n+1} > 2n \geq n + 1$, and so $2^{n+1} > n + 1$, as desired.

Having verified both the base step and the inductive step, we conclude that $2^n > n$ for all $n \geq 1$.

Induction is plausible in the same sense that the Least Integer Axiom is plausible. Suppose that statements $S(1)$, $S(2)$, $S(3)$, $\cdots$ satisfy the hypotheses of mathematical induction. Since $S(1)$ is true, so is $S(2)$; since $S(2)$ is true, so is $S(3)$; since $S(3)$ is true, so is $S(4)$; and so forth. Induction replaces the phrase *and so forth* by the inductive step; this guarantees, for every n, that there is never an obstruction in the passage from a statement $S(n)$ to the next one, $S\ (n + 1)$.

Here are two comments before giving more applications of induction. First, one must verify both the base step and the inductive step; verification of only one of them is inadequate. For example, consider the statements $S(n)$: $n^2 = n$. The base step $S(1)$ is true, but one cannot prove the inductive step (of course, these statements are false for all $n > 1$). Another example is given by the statements $S(n)$: $n > n+1$. The next statement, $S(n+1)$, is: $n+1 > n+2$, and Theorem 1.4(ii) shows that the inductive step is true: if $n > n + 1$, then adding 1 to both sides gives $n + 1 > (n + 1) + 1$. But the base step is false (of course, all these statements are false).

Second, when first seeing induction, many people suspect that the inductive step is circular reasoning: one is using $S(n)$, and this is what one wants to prove! A closer analysis shows that this is not at all what is happening. The inductive step, by itself, does not prove that $S(\ n + 1)$ is true. Rather, it says that if $S(n)$ is true, then one can prove that $S(n+1)$ is also true. In other words, the inductive step proves that the ***implication*** “If $S(n)$ is true, then $S(n + 1)$ is true” is correct. The truth of this implication is not the same thing as the truth of its conclusion. For example, consider the two statements: “Your grade on every exam is 100%” and “Your grade in the course is A.” The implication: “If all your exams are perfect, then you will get the highest grade in the course” is true. Unfortunately, this does not say it is inevitable that your grade in the course will be A. The truth of an implication together with the truth of its hypothesis guarantee the truth of the conclusion; the truth of only the implication does not guarantee the conclusion. Our discussion above gives a mathematical example. The implication “If $n > n + 1$, then $n + 1 > n + 2$” is correct, but the conclusion ”$n + 1 > n + 2$” is false. (There is a discussion of implication, from the viewpoint of truth tables, given in the Glossary at the end of the book.)

This is an appropriate time to mention the ***converse*** of an implication. The converse of “If P is true, then Q is true” is the implication “If Q is true, then P is true.” It is possible that both an implication and its converse are true, in which case we say: “P is true ***if and only if Q*** is true.” On the other hand, it is possible that an implication is true but that its converse is false. For example, the converse of the implication: “If all your exams are perfect, then you will receive the highest grade in the course” is “If you received the highest grade in the course, then all your exams were perfect.” Fortunately, this converse is false. One need not be perfect to receive the grade A. According to my grading scheme, you receive the grade A in the course if and only if your exams average 90% or higher.

The next application of induction verifies a formula giving the sum of the first n integers.

Theorem 1.6. $1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$ for every $n \geq 1$.

Proof. The proof is by induction.

Base step: If $n = 1$, then the right side is $\frac{1}{2}1(1+1) = 1$ and the left side is 1, as desired.

Inductive step: It is always a good idea to write the $(n + 1)$ st statement S(n + 1) (so one can see what has to be proved). We must show that the sum of the first n + 1 integers is given by the formula:

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

Using the inductive hypothesis $S(n)$: $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$, we can rewrite the left side

$$[1 + 2 + \cdots + n] + (n + 1) = \tfrac{1}{2}n(n + 1) + (n + 1),$$

and high school algebra shows that $\frac{1}{2}n(n+1)+(n+1) = \frac{1}{2}(n+1)(n+2)$. We have verified the two steps necessary for induction, and so we can conclude that the formula is true for every $n \geq 1$.

Example. Here is an application of this last formula. How many pairs of positive integers (a, b) are there with $a < b < 12$? If $b = 2$, then $a = 1$; if $b = 3$, then $a = 1$ or 2; if $b = 4$, then $a = 1, 2,$ or 3; ⋯ ; if $b = 11$, then $a = 1, 2, \cdots$, or 10. The number of such pairs (a, b) is thus 1 + 2 + ⋯ + 10, and Theorem1.6 says that this sum is equal to $\frac{1}{2}10 \times 11 = 55$.

There is a story told about the great mathematician Gauss as a boy. One of his teachers asked the students in his class to add up all the numbers from 1 to 100, thereby hoping to get some time for himself (the story assumes that no one in the school knew Theorem 1.6). But Gauss quickly volunteered that the answer is 5050. Here is what he may have done (without induction). Let s denote the sum of all the numbers from 1 to 100: $s = 1 + 2 + \cdots + 99 + 100$. Of course, $s = 100 + 99 + \cdots + 2 + 1$. Arrange these nicely:

$$\begin{array}{rcrcrcccrcr} s & = & 1 & + & 2 & + & \cdots & + & 99 & + & 100 \\ s & = & 100 & + & 99 & + & \cdots & + & 2 & + & 1; \end{array}$$

now add the 100 columns:

$$\begin{array}{rcl} 2s & = & 101 + 101 + \cdots + 101 + 101 \\ & = & 101 \times 100 = 10,100, \end{array}$$

and s = 5050. The same argument works for any number in place of 100. Not only did Gauss give a different proof of Theorem 1.6, but he also discovered its formula. Induction is a technique of proof, but it is not a method of discovery. We displayed the formula for the sum of the first n integers in Theorem 1.6, and we used induction to prove it, but we did not say how the formula was found. The formula was not discovered by induction; it arose in some other way.

Example. Here is a problem using both inductive reasoning and mathematical induction. We seek a formula for the sum of the first n odd numbers: $1 + 3 + 5 + \cdots + (2n – 1)$. A list of the sums for n = 1, 2, 3, 4, 5 is 1, 4, 9, 16, 25. These are perfect squares; better, they are 1^2 2^2, 3^2, 4^2, 5^2. Inductive reasoning suggests the *guess*

$$S(n)\text{: } 1 + 3 + 5 + \cdots +(2n - 1)= n^2.$$

A formula has been discovered. We now use mathematical induction to prove that this guess is always true. The base step $S(1)$ has already been checked. For the inductive step, we must prove

$$S(n+1) : [1 + 3 + \cdots + (2n - 1)] + (2n + 1) = (n + 1)^2.$$

By the inductive hypothesis $S(n) : 1+3+5+\cdots+(2n-1) = n^2$; the bracketed term on the left side is n^2, and so the left side equals $n^2 + (2n + 1) = (n + 1)^2$. By induction, $S(n)$ is true for all $n \geq 1$.

Theorem 1.7. Assuming the product rule from calculus, one has

$$(x^n)' = nx^{n-1} \text{ for all } n \geq 1,$$

where′ denotes derivative.

Remark. Recall that the product rule says

$$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x).$$

Proof. We proceed by induction.

Base step. If $n = 1$, then we are asking whether $(x)' = x^0 = 1$. Now $f'(x) = \lim_{h\to 0}(1/h)[f(x + h) - f(x)]$. When $f(x) = x$, therefore, $(x)' = \lim_{h\to 0}(1/h)[x + h - x] = \lim_{h\to 0} hlh = 1$.

Inductive step. We must prove that $(x^{n+1})' = (n + 1)x^n$ using the nth statement $S(n) : (x^n)' = nx^{n-1}$. Since $x^{n+1} = xx^n$, the product rule and the base step give

$$\begin{aligned}(x^{n+1})' = (xx^n)' &= (x)'x^n + x(x^n)' \\ &= x^n + x(nx^{n-1}) = (n+1)x^n.\end{aligned}$$

We conclude that $(x^n)' = nx^{n-1}$ is true for all $n \geq 1$.

We have just seen that the base step of an inductive proof need not be a triviality; sometimes it is easier to prove than the inductive step, and sometimes it is not.

The base step of an induction may occur at an integer other than 1. For example, consider the statements

$$S(n) : 2^n > n^2.$$

Now $S(n)$ is not true for small n: if $n = 2$ or 4, then there is equality, not inequality; if $n = 3$, we have $2^3 = 8 < 9 = 3^2$, an inequality in the wrong direction. However, $S(5)$ is true, for $2^5 = 32 > 25 = 5^2$.

Theorem 1.8. $2^n > n^2$ is true for all $n \geq 5$.

Proof. We have just checked the base step $S(5)$. In proving the inductive step $2^{n+1} > (n + 1)^2$, we may assume not only the inductive hypothesis $2^n > n^2$ but also $n \geq 5$ (actually, we will need only $n \geq 3$). Multiplying $2^n > n^2$ by 2 gives

$$2^{n+1} = 2 \times 2^n > 2n^2 = n^2 + n^2 = n^2 + nn.$$

Since $n \geq 5$, we have $n \geq 3$ and so

$$nn \geq 3n = 2n + n \geq 2n + 1.$$

Therefore,

$$2^{n+1} > n^2 + 2n + 1 = (n+1)2. \cdot$$

We have seen that the base step of an induction can begin at $n = 1$ or at $n = 5$. Indeed, the base step of an induction can begin at any integer k; of course, the conclusion of such an induction is that the statements are true for all $n \geq k$. If there is a statement $S(0)$, one may also start an induction with base step at $n = 0$. On occasion, one finds an induction beginning at $n = -1$.[4]

One often begins an inductive proof with the phrase, “The proof is by induction on $n \geq k$.” This serves to alert the reader not only that an induction is coming up and that the base step will be at k, but also to indicate which of several variables in a statement will be relevant to the induction. For example, consider the statement, “$(m + 1)^n > mn$.” If I say that I will prove this by induction, do you expect that m is fixed and that the base step $n = 1$ says $m + 1 > m$, or do you expect that n is fixed and that the base step $m = 1$ says $2^n > n$? Stating that the proof is by induction on $n \geq 1$ makes one’s strategy clear.

So far, we have used induction to prove some minor results; let us now use it to prove something more substantial. Observe that if x and y are positive real numbers, then the identity

$$(x + y)^2 = (x - y)^2 + 4xy$$

gives

$$[\tfrac{1}{2}(x + y)]^2 = xy + [\tfrac{1}{2}(x - y)]^2.$$

It follows that

$$[\tfrac{1}{2}(x + y)]^2 \geq xy,$$

with the term $[\tfrac{1}{2}(x - y)]^2$ showing why, in general, the inequality is not an equality. If equality holds, then $[\tfrac{1}{2}(x - y)]^2 = 0$ and $x = y$; conversely, if $x = y$, then there is equality: $[\tfrac{1}{2}(x + x)]^2 = xx$, for $[\tfrac{1}{2}(x - x)]^2 = 0$.

Here is an application of this observation. The *hyperbolic cosine* is defined by

$$\cosh(x) = \tfrac{1}{2}(e^x + e^{-x}).$$

Since $e^x e^{-x} = 1$, it follows that

$$\cosh(x) \geq 1$$

for all x, with equality if and only if $e^x = e^{-x}$; that is, $\cosh(x) = 1$ if and only if $e^{2x} = 1$, so that $\cosh(x) = 1$ if and only if $x = 0$.

Given positive numbers $a_1, a_2, \cdots, a_n$, their ***arithmetic mean*** is defined to be their average: $A = (a_1 + a_2 + \cdots + a_n)/n$, and their ***geometric mean*** is defined to be $G = \sqrt[n]{a_1a_2\cdots a_n}$. Using these new words, we can rephrase what was shown above; the arithmetic mean of two positive numbers a_1 and a_2 is larger than their geometric mean, and equality holds precisely when $a_1 = a_2$. We are going to extend this result to many terms, but we begin with an elementary lemma followed by a normalized version of the inequality.

Lemma 1.9. If $0 < m < 1 < M$, then $m + M > 1 + mM$.

Proof. Since the product of positive numbers is positive,

$$(1 - m)(M - 1) = M - 1 - mM + m$$

is positive. Therefore, $M + m > 1 + mM$, as desired.

For example, there are inequalities $0 < \sin\theta < 1 < \sec\theta$ for any acute angle θ, and so the lemma gives the inequality

$$\sin\theta + \sec\theta > 1 + \sin\theta\sec\theta = 1 + \tan\theta.$$

Lemma 1.10 . If $k_1, k_2, \cdots, k_n$ are positive numbers with $k_1 k_2 \cdots k_n = 1$, then $k_1 + k_2 + \cdots + k_n \geq n$; moreover, equality holds if and only if $1 = k_1 = k_2 = \cdots = k_n$

Proof. Clearly, $k_1 + k_2 + \cdots + k_n = n$ if all $k_i = 1$. Therefore, to prove both statements, it suffices to show that if $k_1 k_2 \cdots k_n = 1$ and not all $k_i = 1$, then $k_1 + k_2 + \cdots + k_n > n$. We prove this by induction on $n \geq 2$.

Base step. Since $k_1 k_2 = 1$ and $k_1 \neq k_2$, we may assume that $0 < k_1 < 1 < k_2$ (if both are strictly larger than 1, then $k_1 k_2 > 1$; if both are strictly smaller than 1, then $k_1 k_2 < 1$). By the lemma, $k_1 + k_2 > 1 + k_1 k_2 = 2$.

Inductive step. Assume that $k_1 k_2 \cdots k_{n+1} = 1$, where $k_1, k_2, \cdots, k_{n+1}$ are positive numbers. We may further assume that some $k_i < 1$: otherwise, all $k_i \geq 1$, and the present assumption that not all $k_i = 1$ gives the contradiction $k_1 k_2 \cdots k_{n+1} > 1$. For notational convenience, let $k_1 < 1$. A similar argument, with all inequalities reversed, allows us to assume that $k_{n+1} > 1$. Define $a_1 = k_1 k_{n+1}$. Note that $a_1 k_2 \cdots k_n = k_1 k_2 \cdots k_{n+1} = 1$. By the lemma,

$$k_1 + k_{n+1} > 1 + k_1 k_{n+1} = 1 + a_1,$$

so that adding $k_2 + \cdots + k_n$ to both sides gives

$$k_1 + k_2 + \cdots + k_n + k_{n+1} > 1 + a_1 + k_2 + \cdots + k_n.$$

It remains to show that $1 + a_1 + k_2 + \cdots + k_n \geq n + 1$ (for we already have an earlier strict inequality). If $a_1 = 1 = k_2 = \cdots = k_n$, then $1 + a_1 + k_2 + \cdots + k_n = n + 1$, and we are done. Otherwise, the inductive hypothesis applies and gives $a_1 + k_2 + \cdots + k_n > n$, and hence $1 + a_1 + k_2 + \cdots + k_n > n + 1$.

Theorem 1.11 (Inequality of the Means). If $a_1, a_2, \cdots, a_n$ are positive numbers, then

$$(a_1 + a_2 + \cdots + a_n)/n \geq \sqrt[n]{a_1 a_2 \cdots a_n};$$

moreover, equality holds if and only if $a_1 = a_2 = \cdots = a_n$.

Proof. Define $G = \sqrt[n]{a_1 a_2 \cdots a_n}$, and define $k_i = a_i/G$ for all i. It follows that $k_1 k_2 \cdots k_n = a_1 a_2 \cdots a_n/G^n = 1$, and so the lemma gives $k_1 + k_2 + \cdots + k_n \geq n$; that is, $a_1 + a_2 + \cdots + a_n \geq nG$, or

$$(a_1 + a_2 + \cdots + a_n)/n \geq G = \sqrt[n]{a_1 a_2 \cdots a_n}.$$

Moreover, the lemma adds that there is equality if and only if all the $k_i = 1$; that is, if and only if all the a_i are equal (to G).

We shall give a geometric application of this inequality in Chapter 2: Of all the triangles with a given perimeter, the one with the greatest area must be equilateral.

The first proofs we have presented are straightforward, and it is easy to believe that we could have discovered them. The proof of the inequality of the means, however, is different; it is not so clear whether one could have discovered it without some pondering. Is there a more pedestrian proof? Probably not. Were all proofs in mathematics routine, there would be a machine that could solve any problem; press a button and wait until the machine presents its answer. But it can be proved that no such machine can ever exist. Your first reaction to this fact of life might be despair, but do not be discouraged. Mathematics is not the realm of a few “magicians”; you are not expected to compete with Archimedes, Gauss, Hilbert, and Poincaré. Each of us is inventive to some degree, and the more one learns, the more proficient one becomes. In music, we can listen and thrill to the beauty of Bach, Mozart, and Beethoven. Even though we cannot compose the sonatas for unaccompanied violin, Don Giovanni, or the late quartets, we can still sing.

The following proof of the inequality of the means is due to G. Pólya, who said that it came to him in a dream. We begin with a lemma from calculus.

Lemma 1.12. For all numbers x,

$$e^x \geq 1 + x,$$

with equality if and only if $x = 0$.

Proof. We consider the function $f(x) = (1 + x)/e^x = (1 + x)e^{-x}$ (notice that $f(x)$ is defined for all x, for the denominator e^x is never 0). Now

$$f'(x) = -(1 + x)e^{-x} + e^{-x} = -xe^{-x},$$

so that $f'(x) \geq 0$ for all $x \leq 0$, $f'(0) = 0$, and $f'(x) \leq 0$ for all $x \geq 0$. Hence, $f(x)$ is an increasing function for negative x, it has exactly one critical point, at $x = 0$, and it is a decreasing function for positive x. It follows that $f(x)$ has an absolute maximum at $x = 0$; that is,

$$(1 + x)/e^{x} = f(x) \leq f(0) = 1 \text{ for all } x.$$

Therefore, $1 + x \leq e^{x}$ for all x, and there is equality precisely when $x = 0$.

Here is Pólya's proof of the inequality of the means.

Theorem 1.13 (= Theorem 1.11). If $a_1, a_2, \cdots, a_n$ are positive numbers, then

$$[(a_1 + a_2 + \cdots + a_n)/n]^{n} \geq a_1a_2\cdots a_n;$$

moreover, equality holds if and only if $a_1 = a_2 = \cdots = a_n$.

Proof (Pólya). Let us write $A = (a_1 + a_2 + \cdots + a_n)/n$ and $G = \sqrt[n]{a_1a_2\cdots a_n}$, so that $G^{n} = a_1a_2 \cdots a_n$. We must show $A^{n} \geq G^{n}$, with equality if and only if $a_1=a_2= \cdots =a_n$.

For each i from 1 to n, define $x_i = -1 + \frac{1}{A}a_i$. By the lemma,

$$e^{x_i} = e^{-1+\frac{1}{A}a_i} \geq 1 + \left(-1 + \frac{1}{A}a_i\right) = \frac{1}{A}a_i,$$

(1)

and there is equality if and only if $x_i = -1 + \frac{1}{A}a_i = 0$; that is,

e^{xi} = 1 + x_i if and only if a_i = A.

We now prove the theorem. The law of exponents gives

(2)

$$\prod_{i=1}^{n} e^{-1+\frac{1}{A}a_i} = e^{-n+\Sigma_i \frac{1}{A}a_i}.$$

But $\sum_{i=1}^{n} \frac{1}{A}a_i = (a_1 + a_2 + \cdots + a_n)/[(a_1 + a_2 + \cdots + a_n)/n] = n$. Hence, the exponent $-n + \sum_i \frac{1}{A}a_i = -n + n = 0$, and so

$$\prod_{i=1}^{n} e^{-1+\frac{1}{A}a_i} = 1.$$

Therefore, Eq. (1) gives

$$1 = \prod_{i=1}^{n} e^{-1+\frac{1}{A}a_i} \geq \prod_{i=1}^{n} \left(\frac{1}{A}a_i\right) = a_1a_2\cdots a_n/A^n = G^n/A^n;$$

(3)

that is, $A^n \geq G^n$.

If all the a_i are equal, say, $a_i = a$ for all i, then $A = (a_1+\cdots+a_n)/n = na/n = a$; that is, $a_i = A$ for all *i*. Therefore, $G^n = a_1a_2 \cdots a_n = A^n$. Conversely, if $A^n = G^n$, then $1 = G^n/A^n$, and the inequality in Eq. (3) becomes

$$\prod_{i=1}^{n} e^{x_i} = \prod_{i=1}^{n}(1 + x_i).$$

By the lemma, $e^{x_i} \geq 1 + x_i$ for all *i*. If there is strict inequality $e^{xk} > 1 + x_k$ for some *k*, then there is strict inequality $\Pi\, e^{xi} > \Pi(1 + x_i)$. Therefore, $e^{xi} = 1 + x_i$ for all *i* ; by Eq. (2), this gives $a_i = A$ for all *i*.

This is an elegant proof, but we mere mortals must be content with more mundane ones.

There is another version of induction, usually called the *second form of induction,* that is often convenient to use.

Definition. The ***predecessors*** of an integer $n \geq 2$ are the positive integers *k* with $k < n$, namely, $1, 2, \cdots, n - 1$.

Theorem 1.14 (Second form of induction). Let S(n) be a family of statements, one for each $n \geq 1$, and suppose that

(i) S(1) is true, and

(ii) if S(k) is true for all predecessors k of n, then S(n) is itself true.

Then $S(n)$ is true for all $n \geq 1$.

Proof. It suffices to show that there are no integers n for which $S(n)$ is false; that is, the collection

$$C = \text{all positive integers } n \text{ for which } S(n) \text{ is false}$$

is empty.

If, on the contrary, C is nonempty, then there is a least criminal; that is, there is a first false statement S(m). Since $S(1)$ is true, by (i), we must have $m \geq 2$. As m is the *least* criminal, k must be honest for all $k < m$; that is, $S(k)$ is true for all the predecessors of m. By (ii), S(m) is true, and this is a contradiction. We conclude that C is empty and, hence, that all the statements $S(n)$ are true.

The second form of induction can be used to give a second proof of Theorem 1.2 (as with the first form, the base step need not occur at 1).

Theorem 1.15 (= Theorem 1.2). Every integer $n \geq 2$ is either a prime or a product of primes.

Proof.[5]. The base step $n = 2$ is true because 2 is a prime. Let $n >$ 2. If n is a prime, we are done. Otherwise, $n = ab$, where $2 \leq a < n$

and $2 \leq b < n$ (since a is an integer, $1 < a$ implies $2 \leq a$). As a and b are predecessors of n, each of them is either prime or a product of primes:

$$a = pp'p'' \cdots \text{ and } b = qq'q'' \cdots ,$$

and so $n = pp'p'' \cdots qq'q'' \cdots$ is a product of (at least two) primes.

The reason the second form of induction is more convenient here is that it is more natural to use S(a) and $S(b)$ than to use $S(n - 1)$; indeed, it is not at all clear how to use $S(n - 1)$.

Here is a notational remark. When using the second form of induction, we speak of n and its predecessors, not $n+1$ and its predecessors. If one wants to compare the two forms of induction, one could say that the first form uses $S(n - 1)$ to prove $S(n)$, whereas the second form uses any or all of the earlier statements $S(1)$, $S(2)$, $\cdots$, S(n – 1), to prove $S(n)$.

The next result says that one can always factor out a largest power of 2 from any integer.

Theorem 1.16. Every positive integer n has a factorization $n = 2^k m$, where $k \geq 0$ and $m \geq 1$ is odd.

Proof. We use the second form of induction on $n \geq 1$.

Base step: If $n = 1$, take $k = 0$ and $m = 1$.

Inductive step: If $n \geq 1$, then n is either odd or even. If n is odd, then take $k = 0$ and $m = n$. If n is even, then $n = 2b$. Because $b < n$, it is a predecessor of n, and so the inductive hypothesis allows us to assume $S(b) : b = 2^\ell m,$ where $\ell \geq 0$ and m is odd. The desired factorization is $n = 2b = 2^{\ell+1}m$.

Definition. The Fibonacci sequence $F_0, F_1, F_2, \cdots$ is defined as follows:

$$F_0 = 0,\ F_1 = 1, \text{ and } F_n = F_{n-1} + F_{n-2} \text{ for all } n \geq 2.$$

Thus, the sequence begins: 0, 1, 1, 2, 3, 5, 8, 13, .⋯

Theorem 1.17. If F_n denotes the nth term of the Fibonacci sequence, then

$$F_n = \frac{1}{\sqrt{5}}(\alpha^n - \beta^n)$$

for all $n \geq 0$, where $\alpha = \frac{1}{2}(1 + \sqrt{5})$ and $\beta = \frac{1}{2}(1 - \sqrt{5})$.

Remark. The number $\alpha = \frac{1}{2}(1 + \sqrt{5})$ is called the ***golden ratio.*** The ancient Greeks called a rectangular figure most pleasing if its edges *a* and *b* were in the proportion a : *b* = *b* : *a* + *b*. It follows that $b^2 = a\,(a + b)$, so that $b^2 - ab - a^2 = 0$, and the quadratic formula gives $b = \frac{1}{2}(a \pm \sqrt{a^2 + 4a^2}) = a\frac{1}{2}(1 \pm \sqrt{5})$. Therefore,

$$b/a = \alpha \text{ or } b/a = \beta.$$

Proof. We are going to use the second form of induction [the second form is the appropriate induction here, for the equation $F_n = F_{n-1} + F_{n-2}$ suggests that proving $S(n)$ will involve not only $S(n-1)$ but $S(n-2)$ as well].

Base step. The formula is true for $n = 0$: $\frac{1}{\sqrt{5}}(\alpha^0 - \beta^0) = 0 = F_0$ and

$$\begin{aligned}\frac{1}{\sqrt{5}}(\alpha^1 - \beta^1) &= \frac{1}{\sqrt{5}}(\alpha - \beta)\\ &= \frac{1}{\sqrt{5}}\left[\tfrac{1}{2}(1+\sqrt{5}) - \tfrac{1}{2}(1-\sqrt{5})\right]\\ &= \frac{1}{\sqrt{5}}(\sqrt{5}) = 1 = F_1\end{aligned}$$

(we have mentioned both $n = 0$ and $n = 1$ because the inductive step will use two predecessors).

Inductive step. If $n \geq 2$, then

$$\begin{aligned}F_n &= F_{n-1} + F_{n-2}\\ &= \frac{1}{\sqrt{5}}(\alpha^{n-1} - \beta^{n-1}) + \frac{1}{\sqrt{5}}(\alpha^{n-2} - \beta^{n-2})\\ &= \frac{1}{\sqrt{5}}\left[(\alpha^{n-1} + \alpha^{n-2}) - (\beta^{n-1} + \beta^{n-2})\right]\\ &= \frac{1}{\sqrt{5}}\left[\alpha^{n-2}(\alpha + 1) - \beta^{n-2}(\beta + 1)\right]\\ &= \frac{1}{\sqrt{5}}(\alpha^n - \beta^n),\end{aligned}$$

because the numbers α and β are the roots of the quadratic equation

$$x^2 = x + 1,$$

so that $\alpha + 1 = \alpha^2$ and $\beta + 1 = \beta^2$.

Remark. It is curious that the integers F_n are expressed in terms of the irrational number √5. An analogous phenomenon will be seen later: there are formulas that express real numbers in terms of complex numbers.

One can also use induction to give definitions. For example, we can define ***n factorial***, denoted by $n!$, by induction on $n \geq 0$. Define $0! = 1$, and if $n!$ is known, then define

$$(n + 1)! = n!(n + 1).$$

The reason for defining $0! = 1$ will be apparent in the next section.

Exercises

1.1. Find a formula for $1 + \sum_{j=1}^{n} j!j$, and use mathematical induction to prove that your formula is correct.

(*Remark.* This exercise illustrates the two types of induction described at the beginning of the chapter: your guess uses inductive reasoning, while its proof using base and inductive steps is mathematical induction.)

1.2. If $r \neq$, prove, for all $n \geq 1$, that

$$1 + r + r^2 + r^3 + \cdots + r^{n-1} = \frac{r^n - 1}{r - 1}.$$

13. Show, for all $n \geq 1$, that 10^n leaves remainder 1 after dividing by 9. (Hint: Prove $10^n = 9q_n + 1$ for some integer q_n.)

1.4. If $a \leq b$ are positive numbers, prove that $a^n \leq b^n$ for all integers $n \geq 0$.

1.5. (i) Prove that $1^2 + 2^2 + \cdots + n^2 = \frac{1}{6}n(n + 1)(2n + 1)$.

(ii) Prove that $1^3+2^3+\cdots+n^3 = (1+2+\cdots+n)^2$. (Hint: Use Theorem 1.6.)

(iii) Prove that $1^4 + 2^4 + \cdots + n^4 = \frac{1}{5}n^5 + \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n$

1.6. (i) Find a formula for $a_n = 1^3 + 3^3 + 5^3 + \cdots + (2n - 1)^3$, and then prove that your guess is correct using induction.

(ii) Give a second proof of part (i) based on Exercise 1.5(ii) and the following observation: If $b_m = 1^3 + 2^3 + 3^3 + \cdots + m^3$. then

$$\begin{aligned} b_{2n} &= a_n + [2^3 + 4^3 + \cdots + (2n)^3] \\ &= a_n + 8[1^3 + 2^3 + \cdots + n^3] = a_n + 8b_n. \end{aligned}$$

1.7. (i) Prove that if $n = ab$, where n, a, and b are positive integers, then either $a \le \sqrt{n}$ or $b \le \sqrt{n}$.

(ii) Prove that if n is composite, then it has a prime factor p with $p \le \sqrt{n}$. Conclude that if $n \ge 2$ has no prime factors $\le \sqrt{n}$, then n is a prime.

(iii) If $f(n) = n^2 - n + 41$, use part (ii) to show that $f(10)$, $f(20)$, $f(30)$, and $f(40)$ are prime. (If each student in a class checks that two or three values of $f(n)$ are prime, then the class will have shown that $f(n)$ is prime for all $n \le 40$.)

Remark. It is now simple to check that 991 is a prime, but checking that 1,000,099 is a prime is a longer enterprise, for its square root is a bit over 1000.

1.8. Prove, for all $n \ge 0$, that $(1+ x)^n \ge 1 + nx$ if $1 + x > 0$.

1.9. (i) Prove that $2^n > n^3$ for all $n \ge 10$.

(ii) Prove that $2^n > n^4$ for all $n \ge 17$.

1.10. Let $g_1(x), \ldots, g_n(x)$ be differentiable functions. If $f(x) = g_1(x) \ldots g_n(x)$, prove that its derivative is

$$f'(x) = \sum_{i=1}^{n} f(x)g_i'(x)/g_i(x).$$

1.11. Prove that every positive integer a has a factorization $a = 3^k m$, where $k \geq 0$ and m is not a multiple of 3. (Hint: Adapt the proof of Theorem 1.16.) *Remark*. This last exercise illustrates another reason for knowing proofs. The solution of Exercise 1.11 does not follow from the statement of Theorem 1.16, but a solution can be obtained by modifying the proof of that theorem.

1.12. Prove that $2^n < n!$ for all $n \geq 4$.

1.13. Prove that $\mathrm{F}_n < 2^n$ for all $n \geq 0$, where $F_0, F_1, F_2, \cdots$ is the Fibonacci sequence.

1.14. For every acute angle θ, i.e., $0° < \theta < 90°$, prove that

$$\sin\theta + \cot\theta + \sec\theta \geq 3.$$

(Hint: Use the inequality of the means or Lemma 1.10.)

1.15. Prove that if $a_1, a_2, \ldots, a_n$ are positive numbers, then

$$(a_1 + a_2 + \ldots + a_n)(1/a_1 + 1/a_2 + \ldots + 1/a_n) \geq n^2.$$

1.16. For every $n \geq 2$, prove that there are n consecutive composite numbers; that is, there is some integer b such that $b + 1, b + 2, \cdots, b + n$ are all composite. (Hint: If $2 \leq a \leq n + 1$, then a is a divisor of $(n + 1)! + a$.)

BINOMIAL COEFFICIENTS

> Let no one say that I have said nothing new ... the arrangement of the subject is new. When we play tennis, we both play with the same ball, but one of us places it better.
>
> *B. Pascal*

Consider the formulas for powers $(1 + x)^n$ of the binomial $1 + x$:

$$\begin{aligned}
(1+x)^0 &= 1 \\
(1+x)^1 &= 1+1x \\
(1+x)^2 &= 1+2x+1x^2 \\
(1+x)^3 &= 1+3x+3x^2+1x^3 \\
(1+x)^4 &= 1+4x+6x^2+4x^3+1x^4.
\end{aligned}$$

Is there a pattern in the coefficients in these formulas? Figure 1.1, called ***Pascal's triangle*** [after B. Pascal (1623 – 1662)], shows the first few coefficients.

```
                  1
               1     1
            1     2     1
         1     3     3     1
      1     4     6     4     1
   1     5    10    10     5     1
1     6    15    20    15     6     1
```

Figure 1.1

Figure 1.2 is a Chinese illustration made in the year 1303, which shows that Pascal’s triangle had been recognized long before Pascal.

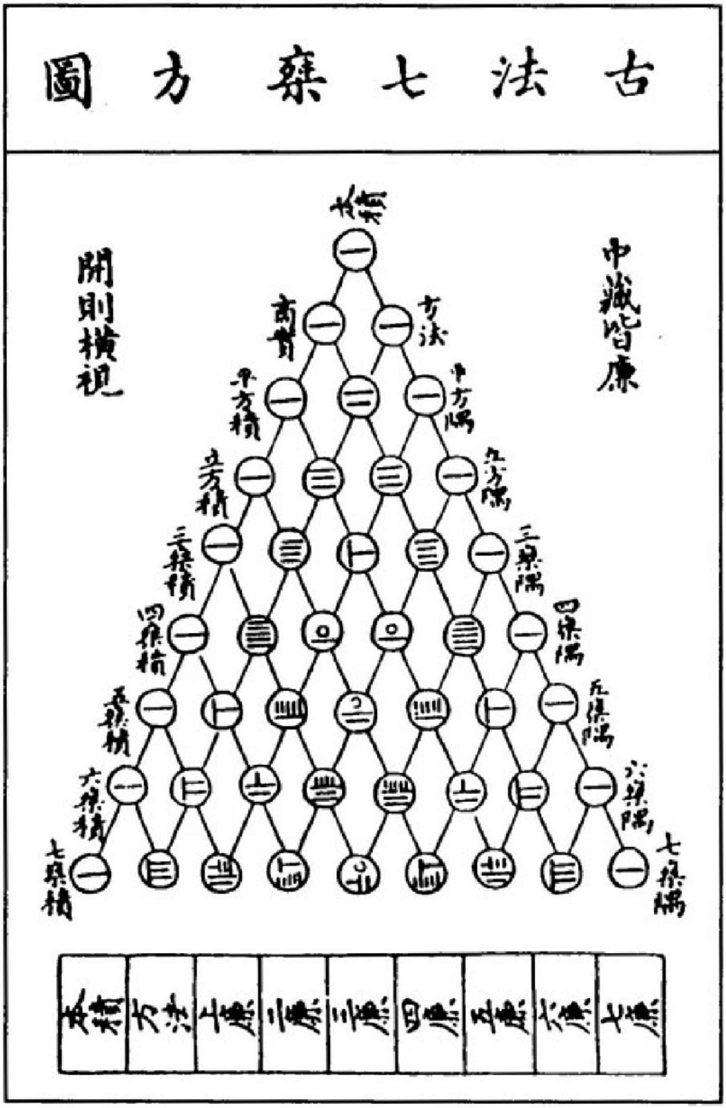


Figure 1.2

The expansion of $(1 + x)^n$ is an expression of the form

$$c_0 + c_1x + c_2x^2 + \ldots + c_nx^n,$$

where $c_0 = 1$ and $c_n = 1$. What are the “inside” coefficients $c_1, \ldots,$ c_{n-1}?

L. Euler (1707 – 1783) introduced the notation $\left(\frac{n}{r}\right)$, which lost its bar after fifty years, and this more simple form of it is now universally accepted:

$$\binom{n}{r} = \text{coefficient } c_r \text{ of } x^r \text{ in } (1+x)^n.$$

These numbers $\binom{n}{r}$ are called ***binomial coefficients;*** the number $\binom{n}{r}$ is pronounced “*n* choose *r*” because it arises in counting problems (as we shall see at the end of this section).

Observe, in Figure 1.1, that an inside number (i.e., not a 1 on the border) of the (n +1)th row can be computed by going up to the *n*th row and adding the two neighboring numbers above it:

$$\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}.$$

For example, the inside numbers in row 4 can be computed from row 3 as follows:

$$\begin{array}{ccccccccc} & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

1+3 = 4, 3+3 = 6, and 3+1 = 4. Let us prove that this observation always holds.

It is clear that $\binom{n}{0} = 1 = \binom{n}{n}$;that is, in the expansion of $(1+ x)^n$, both the constant term and the coefficient of x^n are equal to 1.

Theorem 1.18. For all r with $0 < r < n + 1$,

$$\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}.$$

Proof. We must show that if

$$(1 + x)^n = c_0 + c_1x + c_2x^2 \ldots + c_nx^n,$$

then the coefficient of x^r in $(1+ x)^{n+1}$ is $c_{r-1} + c_r$.

$$\begin{aligned}
(1+x)^{n+1} &= (1+x)(1+x)^n \\
&= (1+x)^n + x(1+x)^n \\
&= (c_0 + c_1x + c_2x^2 + \cdots + c_nx^n) + \\
&\qquad x(c_0 + c_1x + c_2x^2 + \cdots + c_nx^n) \\
&= (c_0 + c_1x + c_2x^2 + \cdots + c_nx^n) + \\
&\qquad c_0x + c_1x^2 + c_2x^3 + \cdots + c_nx^{n+1} \\
&= 1 + (c_0 + c_1)x + (c_1 + c_2)x^2 + (c_2 + c_3)x^3 + \cdots ,
\end{aligned}$$

because $c_0 = 1$. Thus, the coefficient of x^r in $(1 + x)^{n+1}$ is $c_{r-1} + c_r$.

We shall see how the formula in the next theorem arises when we discuss some counting problems.

Theorem 1.19 *(Pascal).* For all $n \geq 0$ and all r with $0 \leq r \leq n$,

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Proof. We prove the theorem by induction on $n \geq 0$.

The base step $n = 0$ is easy: by definition, $\binom{0}{0} = 1$, while $0!/0!0! = 1$ as well. To prove the inductive step, we must show

$$\binom{n+1}{r} = \frac{(n+1)!}{r!(n+1-r)!}.$$

By Theorem 1.18,

$$\begin{aligned}\binom{n+1}{r} &= \binom{n}{r-1}+\binom{n}{r} \\ &= \frac{n!}{(r-1)!(n-r+1)!}+\frac{n!}{r!(n-r)!} \\ &= \frac{n!}{(r-1)!(n-r)!}\cdot\left(\frac{1}{n-r+1}+\frac{1}{r}\right) \\ &= \frac{n!}{(r-1)!(n-r)!}\cdot\left(\frac{r+n-r+1}{r(n-r+1)}\right) \\ &= \frac{n!}{(r-1)!(n-r)!}\cdot\left(\frac{n+1}{r(n-r+1)}\right) \\ &= \frac{(n+1)!}{r!(n+1-r)!}. \quad \bullet\end{aligned}$$

One defines 0! = 1 to make formulas like this one more simple; without this convention, there would have to be an extra statement giving the formula in the special cases $r = 0$ and $r = n$. Moreover, the base step $n = 0$ is simpler than the base step $n = 1$.

Corollary 1.20. For any number x and for all $n \geq 0$,

$$(1+x)^n = \sum_{r=0}^{n}\binom{n}{r}x^r = \sum_{r=0}^{n}\frac{n!}{r!(n-r)!}x^r.$$

Proof. The binomial coefficients have been defined as the numbers c_r, where

$$(1 + x)^n = c_0 + c_1x + c_2x^2 + \ldots + c_nx^n.$$

Since $c_r = \binom{n}{r}$, we have

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{r}x^r + \cdots + \binom{n}{n}x^n,$$

and the formula in Pascal's theorem now gives the result.

Corollary 1.21 (Binomial Theorem). For all numbers *a* and *b* and for all integers $n \geq 0$,

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} b^r a^{n-r} = \sum_{i+j=n} \binom{n}{i} b^i a^j.$$

Proof. The result is trivially true when $a = 0$ (we agree upon the notation $0^0 = 1$). If $a \neq 0$, set $x = b / a$ in Corollary 1.20, and observe that

$$\left(1 + \frac{b}{a}\right)^n = \left(\frac{a+b}{a}\right)^n = \frac{(a+b)^n}{a^n}.$$

Thus, $(a + b)^n$ is obtained by multiplying the expression for $(1 + b / a)^n$ by a^n.

We have used a second convention: $0^0 = 1$, and we are using it for the same reason we use $0! = 1$: it simplifies the writing of formulas.

Here is a combinatorial interpretation of the binomial coefficients. Given a set X, an ***r*-subset** is a subset of X with exactly r elements. If X has n elements, denote the number of its r-subsets by

$$[n, r].$$

We compute $[n, r]$ by considering a related question. Given an “alphabet” with n (distinct) letters and a number r with $1 \leq r \leq n$, an r-letter *word* is a sequence of r of the letters with no repetitions (and with no regard to whether the “word” actually occurs in some dictionary). For example, the 2-letter words on the alphabet a, b, c are

$$ab, ba, ac, ca, bc, cb$$

(note that aa, bb, cc are not on this list). How many r-letter words are there on an alphabet with n letters? We count the number of words in two ways.

(I) There are n choices for the first letter; having chosen the first letter, there are now only $n - 1$ choices for the second letter, for no letter is repeated; having chosen the first two letters, there are only $n - 2$ choices for the third letter, and so forth. Thus, the number of r-letter words is

$$n(n - 1)(n - 2) \cdots (n - [r - 1]) = n(n - 1)(n - 2) \ldots (n - r + 1).$$

Note the special case $n = r$: the number of n-letter words on n letters is $n!$.

(II) Here is a second way to count these words. First choose an r-subset of the alphabet (consisting of r letters); there are $[n, r]$ ways to do this, for this is exactly what the symbol $[n, r]$ means. For each chosen r-subset, there are $r!$ ways to arrange the r letters in it (this is the special case of (I) when $n = r$). The number of r-letter words is thus

$$r![n, r].$$

We conclude that

$$r![n, r] = n(n - 1)(n - 2) \cdots (n - r + 1).$$

Therefore,

$$\begin{aligned} [n, r] &= \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} \\ &= \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} \cdot \frac{(n-r)!}{(n-r)!} \\ &= \frac{n!}{r!(n-r)!} = \binom{n}{r} \end{aligned}$$

(the last equation by Pascal's theorem). This fact is the reason one often pronounces the binomial coefficient $\binom{n}{r}$ as n choose r.

As an example, how many ways are there to choose 2 hats from a closet containing 14 different hats? (One of my friends does not like the phrasing of this question. After all, one can choose 2 hats with one's left hand, with one's right hand, with one's teeth, . . . ; but I continue the evil tradition.) The answer is $\binom{14}{2}$, and Pascal's formula allows us to compute this as 14 × 13/2 = 91.

Our first interpretation of the binomial coefficients $\binom{n}{r}$ was *algebraic;* that is, as Pascal's formula in terms of factorials; our second interpretation is *combinatorial;* that is, as n choose r. Quite often, each interpretation can be used to prove a desired result. For example, let us prove Theorem 1.18 combinatorially. Let X be a collection of n +1 balls, and let us color one ball red and the other n balls blue. Now $\binom{n+1}{r}$ is the number of r-subsets S of X. There are two possibilities for an r-subset S: either it contains the red ball or it does not. If S does contain the red ball, then S consists of the red ball and r – 1 blue balls, and so the number of such S is the same as the number of (r – 1)-subsets of the blue balls, namely, $\binom{n}{r-1}$. The other possibility is that S consists completely of blue balls; since there are n blue balls, there are $\binom{n}{r}$ such r-subsets. Therefore, $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$, as desired.

Exercises

1.17. Show that the binomial coefficients are "symmetric": if $0 \le r \le n$, then

$$\binom{n}{r} = \binom{n}{n-r}.$$

1.18. Show, for every *n*, that the sum of the binomial coefficients is 2^n:

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}=2^n.$$

1.19. (i) Show, for every $n > 0$, that the “alternating sum” of the binomial coefficients is zero:

$$\binom{n}{0}-\binom{n}{1}+\binom{n}{2}-\cdots\pm\binom{n}{n}=0.$$

(ii) Use part (i) to prove, for a given *n*, that the sum of all the binomial coefficients $\binom{n}{r}$ with r even is equal to the sum of all those $\binom{n}{r}$ with *r* odd.

1.20. What is the coefficient of x^{16} in $(1 + x)^{20}$?

1.21. How many ways are there to choose 4 colors from a palette containing 20 different paints?

1.22. Prove that a set *X* with *n* elements has exactly 2^n subsets. Can you give more than one proof of this?

1.23. A weekly lottery asks you to select 5 numbers between 1 and 45. At the week’s end, 5 such numbers are drawn at random, and you win the jackpot if all your numbers match, in some order, the drawn numbers. How many selections of 5 numbers are there?

Answer: 1, 221, 759.

1.24. Assume that *term-by-term differentiation* of power series is valid: if

$$f(x) = \sum_{k\geq 0} a_k x^k = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots,$$

then the power series for its derivative $f'(x)$ is

$$f'(x) = \sum_{k\geq 1} k a_k x^{k-1} = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} + \cdots.$$

(i) Prove that $f(0) = a_0$.

(ii) Prove, for all $n \geq 0$, that the nth derivative

$$f^{(n)}(x) = \sum_{k\geq n} k(k-1)(k-2)\cdots(k-n+1) a_k x^{k-n}.$$

[$f^{(0)}(x)$ is defined to be $f(x)$].[6] Conclude, for all $n \geq 0$, that

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

1.25. (***Leibniz***) A real-valued function $f(x)$ is called a ***C^{∞}-function*** if it has an nth derivative $f^{(n)}$ (x) for every $n \geq 0$ [$f^{(0)}(x)$ is defined to be $f(x)$]. Prove that if $f(x)$ and $g(x)$ are C^{∞}-functions, then

$$(f(x)g(x))^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(k)}(x) g^{(n-k)}(x) \text{ for all } n \geq 0.$$

1.26. Prove, for all $n \geq 1$ and for all $r > 1$, that

$$\binom{n}{0} + \binom{n+1}{1} + \binom{n+2}{2} + \cdots + \binom{n+r}{r} = \binom{n+r+1}{r}.$$

(Hint: Use induction on $r \geq 2$.)

Chapter 2

Things Pythagorean

AREA

> Let no one ignorant of geometry enter my door.
>
> *Plato*

One of the earliest ideas of geometry is the computation of the area of geometric figures. Let us start with the simplest figure; a rectangle with sides of lengths *a* and *b* has area *A = ba.*

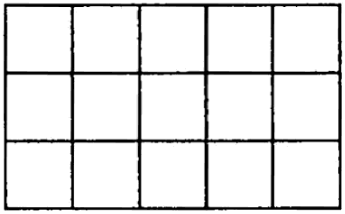


Figure 2.1

For instance, if the base is of length $b = 5$ and the height is $h = 3$, then the area is 15 units of area. If we cut the rectangle in half, we obtain two congruent right triangles:

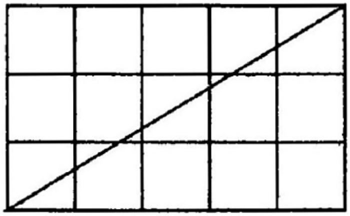


Figure 2.2

The area A of a right triangle is, thus, $\frac{1}{2}$ base × height.

The area A of an arbitrary triangle of base b and height h, as you recall, is also $A = \text{area} = \frac{1}{2}bh$. Observe that this formula is really three formulas in one, because each of the three sides of a triangle qualifies as a “base” of the triangle; the corresponding height is the length of the altitude to that side.

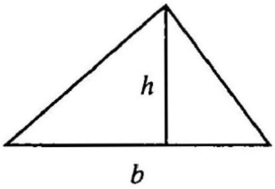


Figure 2.3

Let us prove the area formula by using the formula for a right triangle. If, as in Figure 2.4, the altitude divides the triangle into two right triangles of height h having bases b_1 and b_2, respectively (so that $b = b_1 + b_2$), then the area A is

$$A = \tfrac{1}{2}b_1h + \tfrac{1}{2}b_2h = \tfrac{1}{2}(b_1 + b_2)h = \tfrac{1}{2}bh.$$

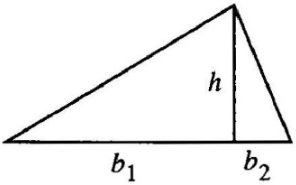


Figure 2.4

Suppose now that the altitude does not divide the triangle into two triangles, but that we have Figure 2.5; that is, the possibilities are either that *S* lies between *Q* and *R* or that it does not lie between them. Let *b′* be the length of *RS*, so that *b* + *b′* is the length of *QS*. Since Δ*QSP* and Δ*RSP* are right triangles, their areas are $\frac{1}{2}(b+b')h$ and $\frac{1}{2}b'h$, respectively. The area of Δ*PQR* is, thus, $\frac{1}{2}(b+b')h-\frac{1}{2}b'h=\frac{1}{2}(b+b'-b')h=\frac{1}{2}bh$. Having considered all the cases, we can now declare the result.

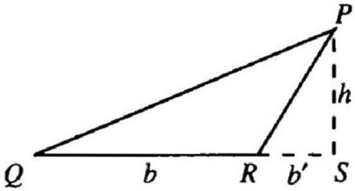


Figure 2.5

Theorem 2.1. The area *A* of any triangle with base b and height *h* is given by

$$A = \tfrac{1}{2}bh.$$

Let *PA* and *QR* be parallel, and consider the triangles ∆*PQR* and ∆*AQR* in Figure 2.6*a*. Since they have the same base and the same height, they have the same area, by Theorem 2.1. That is, if we hold the base fixed and slide the top along a horizontal line, then the area remains the same.

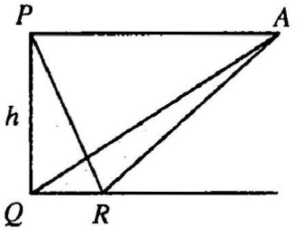


Figure 2.6a

This invariance of area after "sliding" was already recognized by Euclid, one of whose theorems is: "Triangles drawn to a point are to each other as their bases."

Euclid's theorem is illustrated in Figure 2.6*b*:

$$\frac{\text{area}(\triangle ABC)}{\text{area}(\triangle ADE)} = \frac{|BC|}{|DE|}.$$

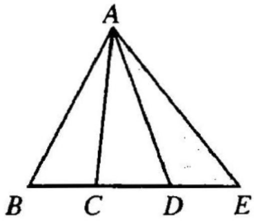


Figure 2.6b

Exercise 2.3 below shows that the two triangles in Figure 2.6*a* have horizontal cross sections of equal length. *Cavalieri's principle*, named after B. Cavalieri (ca. 1598 – 1647), extends the invariance of area after "sliding" to more general figures.

Cavalieri's Principle: Two figures with horizontal cross sections of equal length have the same area.

The next result shows that Cavalieri's principle applies to parallelograms: holding the base fixed and sliding the top does not change the area.

Theorem 2.2. The area A of a parallelogram with height h and base b is given by $A = hb$.

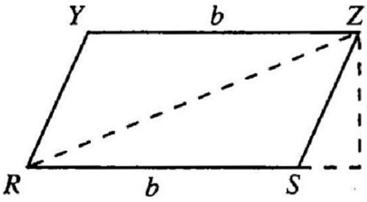


Figure 2.7

Proof. Divide the parallelogram with height *h* and base *b* into two triangles, as in Figure 2.7. Now Δ*ZRS* has base *RS* of length *b*, height *h*, and area $\frac{1}{2}hb;$ similarly, Δ*YZR* has base *YZ* of length *b*, height *h*, and area $\frac{1}{2}hb.$ We conclude that the parallelogram has area $\frac{1}{2}hb + \frac{1}{2}hb = hb.$

Another way to see that the area of a parallelogram is *hb* is indicated in Exercise 2.1. As the proof of Theorem 2.1, the proof suggested in this exercise also involves two cases, the second of which is often overlooked.

We sketch two proofs of Cavalieri's principle using calculus. The first proof involves the computation of the area of a region *R* by double integrals.

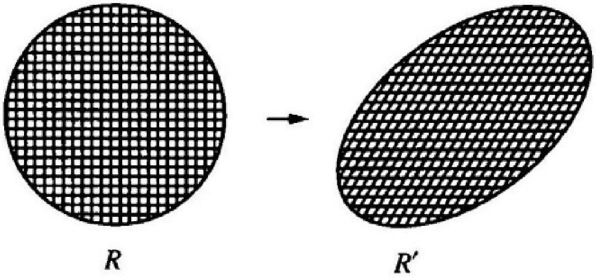


Figure 2.8

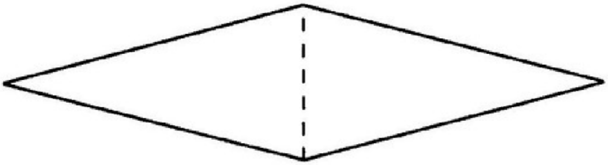


Figure 2.9

Here is an ancient problem, called ***doubling the square.*** Given a square of side length *a*, what is the side length *d* of the square having double its area, namely, $2a^2$? This problem is discussed in Plato's *Menon* in the form of a dialogue between Socrates and a slave. Their discussion leads to the geometric construction in Figure 2.10.

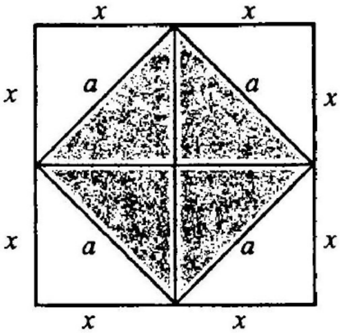


Figure 2.10

***Theorem* 2.3.** The length *d* of the diagonal of a square having side length *a* satisfies the equation

$$d^2 = 2a^2.$$

Proof. In Figure 2.10, the large region is a square with all sides of length d, where $d = 2x$. The shaded region is a rhombus with all sides of length a; it is a square because the base angles of the right triangles are each 45°, and hence the interior angles of the rhombus are each $45° + 45° = 90°$.

We compute the area of the shaded square in two ways. On the one hand, it has area a^2, for the length of a side is a. On the other hand, this square is divided into 4 right triangles, each of area $\frac{1}{2}x^2$. Therefore, $a^2 = 4 \times \frac{1}{2}x^2 =$ $2x^2$, and

$$2a^2 = 4x^2 = (2x)^2 = d^2.$$

We remark that there is a quick proof of the theorem using the Pythagorean theorem, for $d^2 = a^2 + a^2 = 2a^2$. Indeed, we have just proved the special case of the Pythagorean theorem for isosceles right triangles. We will prove the Pythagorean theorem in the next section.

Corollary 2.4. Let C be a circle, let S be a circumscribed square, and let T be an inscribed square. Then

$$\text{area}(S) = 2\ \text{area}(T).$$

Proof. Using the notation in Figure 2.10 (imagine a circle inscribed in the big square, circumscribing the shaded square), the diameter of C is equal to $d = 2x$. By the theorem,

$$\text{area}(S) = d^2 = 2a^2 = 2\ \text{area}(T).$$

Before continuing this discussion, we remind you of some familiar terminology.

Definition. A ***divisor*** of an integer *m* is an integer *d* for which

$$m = dq,$$

where *q* is an integer.

Recall that if *m* and *d* are positive integers, then long division gives integers *q* and *r* with

$$\frac{m}{d} = q + \frac{r}{d}$$

and $r/d < 1$. Clearing denominators, this equation involving fractions gives an equation of integers

$$m = dq + r,$$

where the ***quotient*** *q* is an integer ≥ 0 and the ***remainder*** *r* is one of 0, 1, 2, . . . , $d - 1$. Of course, *d* is a divisor of *m* precisely when the remainder $r = 0$.

For example, every integer can be written either as $2q$ or $2q + 1$, for some integer *q*, because the only possible remainders after dividing by 2 are 0 and 1.

Definition. A ***rational number*** *r* (also called a *fraction*) is a ratio of two integers; that is, *r* = *p*/*q*, where both *p* and *q* are integers and *q* ≠ 0. A real number that is not rational is called ***irrational.***

It turns out that there are plenty of irrational numbers, as we shall soon see. The terms *rational* and *irrational* come from "ratio;" in particular, this usage of irrational does not mean "unreasonable," which is the other contemporary usage of this word.

There are many ways to write a given rational number; for example, $\frac{1}{2} = \frac{2}{4} = \frac{3}{6} = \cdots$. Recall that

$$\frac{a}{b} = \frac{c}{d}$$

if and only if ***cross multiplication*** holds: *ad* = *bc*. Note that cross multiplying converts equations in rationals into equations in integers. Given a rational number *a*/*b*, let us show that we may assume, after suitable cancellation, that at least one of the numerator *a* and the denominator *b* is odd. Theorem 1.16 says that

$$a = 2^{\ell} m \quad \text{and} \quad b = 2^{k} n,$$

where ℓ, *k* ≥ 0 and *m* and n are odd. If ℓ ≥ *k*, then

$$\frac{a}{b} = \frac{2^{\ell} m}{2^{k} n} = \frac{2^{\ell - k} m}{n}.$$

Because ℓ ≥ *k*, the exponent ℓ – *k* ≥0 and $2^{\ell-k}m$ is an integer. We may thus replace *a* by $2^{\ell-k}m$ and *b* by *n*; that is, we may assume that the (new) denominator is odd. Similarly, if $\ell \leq k$, then

$$\frac{a}{b} = \frac{m}{2^{k-\ell}n},$$

and we may assume that the (new) numerator is odd. Hence, every rational number *a*/*b* has an expression of the form *p*/*q*, where at least one of *p* and *q* is odd. (More is true; *a*/*b* can always be put in *lowest terms*; that is, the numerator and denominator have no factors in common. This is proved using the notion of *greatest common divisor,* but we leave this discussion to another course.)

Theorem 2.5. $\sqrt{2}$ is an irrational number.

Proof. Assume, on the contrary, that $\sqrt{2}$ is rational; that is,

$$\sqrt{2} = \frac{p}{q},$$

where both *p* and *q* are positive integers. By our discussion above, we may assume that at least one of *p* and *q* is odd.

Squaring both sides, $2 = p^2/q^2$, and cross multiplying gives

$$2q^2 = p^2.$$

Were *p* odd, then p^2 would also be odd [by Exercise 2.4(ii): the product of odds is odd]. Since $p^2 = 2q^2$ is even, we conclude that p is even, and so we may write $p = 2r$ for some integer r. Substituting into $2q^2 = p^2$ gives $2q^2 = (2r)^2 = 4r^2$, so that

$$q^2 = 2r^2.$$

It follows, as above, that q is even (for $q^2 = 2r^2$ is even). This contradicts our assumption that at least one of p and q is odd.

This last result is significant in the history of mathematics. We accept the real number line without qualms; if two points on a line are chosen, one labeled 0 and the other 1, then every point on the line corresponds to a number. Not only was this not obvious to the ancient Greeks, they did not believe it. From its earliest days, about four to five thousand years ago in western Asia, mathematics was used for applications to practical problems. Numbers and geometry were studied using experience, and generalizations seem to have been made with the light of intuition, but not with proofs. Because of its utility, mathematics spread to China, India, Greece, and Egypt. In Greece, sometime after the time of Homer, the idea arose that logical reasoning was necessary to prove mathematical results. There are proofs attributed to Thales of Miletus (ca. 624 – 547 BC) (many historians describe him as the founding father of Greek mathematics). The ancient philosopher Pythagoras (ca. 570 – 500 BC), after whom the Pythagorean theorem is named, may have studied with Thales. Pythagoras founded a secret philosophical society, one of whose goals was to use integers and their ratios to explain all of nature's phenomena. (The word *mathematics*, meaning "that which is learned," is due to the Pythagoreans. Here is some more etymology. The word *geometry*, meaning "earth measure," probably goes back to its early applications, whereas the word *algebra* is a European version of the first word in the title of an influential book, *Al-jabr w'al muqabala*[7], by al-Khwarizmi, written in Arabic in 830.) For the Pythagoreans, numbers were defined to be positive integers, whereas other (positive) real numbers were not numbers at all; instead, they were viewed as pairs $a : b$ of line segments (which we can interpret as the number $|a|/|b|$, where $|a|$ is the length of a). There were geometric ways of viewing addition, subtraction, multiplication, and division of segments, but it was virtually impossible to do any algebra with them. For example, a sophisticated geometric argument (due to Eudoxus and given in Euclid's *Elements*) was needed to prove cross multiplication: if $a : b = c : d$, then $a : c = b : d$. Pythagoras dealt with rationals by assuming, given two segments x and y, that there is a segment z and integers m and n with $|x|= m|z|$ and $|y| = n|z|$; in modem notation: $|x| = \frac{m}{n}|y|$. He had hoped that such a relation would be true for any pair of segments x and y, but Theorems 2.3 and 2.5 showed him that this is not so when x is the diagonal of a square with side y.

Why did Pythagoras have such a constrictive view of numbers? We quote van der Waerden, *Science Awakening.*

> Nowadays we say that the length of the diagonal is the "irrational number" √2 and we feel superior to the poor Greeks who "did not know irrationals." But the Greeks knew irrational ratios very well. ... That they did not consider √2 as a number was not a result of ignorance, but of strict adherence to the definition of number. *Arithomos* means quantity, therefore whole number. Their logical rigor did not even allow them to admit fractions; they replaced them by ratios of integers.

> For the Babylonians, every segment and every area simply represented a number. ... When they could not determine a square root exactly, they calmly accepted an approximation. Engineers and natural scientists have always done this. But the Greeks were concerned with exact knowledge, with “the diagonal itself,” as Plato expresses it, not with an acceptable approximation.
>
> In the domain of numbers (positive integers), the equation $x^2 = 2$ cannot be solved, not even in that of ratios of numbers. But it is solvable in the domain of segments; indeed the diagonal of the unit square is a solution. Consequently, in order to obtain exact solutions of quadratic equations, we have to pass from the domain of numbers (positive integers) to that of geometric magnitudes. Geometric algebra is valid also for irrational segments and is nevertheless an exact science. It is therefore logical necessity, not the mere delight in the visible, which compelled the Pythagoreans to transmute their algebra into a geometric form.

Even though the Pythagorean definition of number is no longer popular, the Pythagorean dichotomy persists to this very day. For example, almost all American high schools teach one year of algebra and one year of geometry, instead of two years in which both subjects are developed together. The problem of defining *number* has arisen several times since the classical Greek era. In the 1500’s, mathematicians had to deal with negative numbers and with complex numbers (see our discussion of cubic polynomials in Chapter 4); the description of real numbers generally accepted today dates from the late 1800’s. There are echos of Pythagoras in our time. L. Kronecker (1823 – 1891) wrote, “Die ganzen Zahlen hat der liebe Gott gemacht; alles andere ist Menschenwerk" (The integers were created by God; all the rest is the work of Man), and even today, some logicians argue for a new definition of number.

Exercises

2.1. (i) In Figure 2.11*a*, *QPSR* is a rectangle and *YZSR* is a parallelogram. Show that Δ*QYR* and Δ*PZS* are congruent.

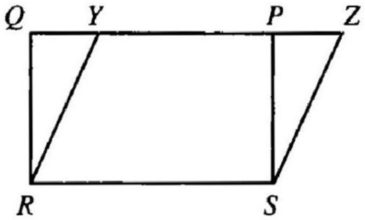


Figure 2.11a

(ii) Show that one can construct the parallelogram *YZSR* from the rectangle *QPSR* by cutting off Δ*QYR* and pasting it in position Δ*PZS*; conclude that the parallelogram has the same area as the rectangle if *Y* is between *Q* and *P*.

(iii) Prove that the parallelogram *YZSR* has the same area as the rectangle *QPSR* when *Y* is not between *Q* and *P* (see Figure 2.11b*b*.) [One must prove this in order to complete the argument that the areas of the parallelogram and the rectangle always agree. You may not use Theorem 2.2, for this exercise is to give an alternative proof of that theorem.]

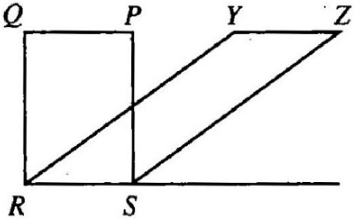


Figure 2.11b

2.2. Show that the trapezoid in Figure 2.12 has area $\frac{1}{2}(a+b)h$.

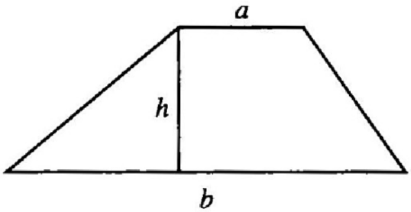


Figure 2.12

2.3. Assume, in Figure 2.13, that *PA*, *EH*, and *QR* are parallel. If *P* and *Q* are points, let us denote the length of the line segment *PQ* by |*PQ*| Prove that |*EF*| = |*GH*|, and then conclude that Cavalieri's Principle applies to Δ*PQR* and Δ*AQR*. (Hint: Let ℓ and ℓ′ be parallel lines, and let *t* and *t'* be transversals. If ℓ″ is parallel to ℓ (and to ℓ′), then ℓ″ divides the transversals proportionally. In Figure 2.13, |*PE*|/|*PQ*| = |*AH*|/|*AR*|.)

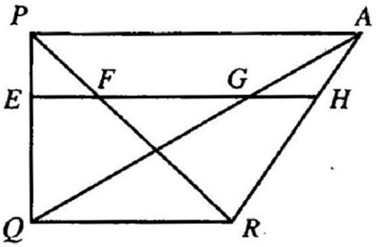


Figure 2.13

2.4. Let *a* and *b* be integers.

(i) If *a* is even, prove that *ab* is even for every integer *b*.

(ii) If both *a* and *b* are odd, prove that *ab* is odd while *a* + *b* is even.

(iii) If neither *a* nor *b* is a multiple of 3, prove that *ab* is not a multiple of 3.

2.5. If $r = p/q$ is a nonzero rational number, show that $r + \sqrt{2}$ and $r\sqrt{2}$ are irrational numbers. Conclude that there are infinitely many irrational numbers.

2.6. Use the Pythagorean theorem to prove that if *a* is the side length of a cube and $|AB|$ is the length of a diagonal joining opposite comers, then $|AB|^2 = 3a^2$.

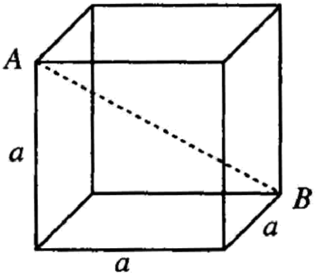


Figure 2.14

2.7. Prove that √3 is irrational. (Hint: Modify the proof of Theorem 2.5, replacing each occurrence of 2 by 3, “even” by “multiple of 3”, and ”odd“ by ”not divisible by 3.”)

Remark: Exercise 2.7 shows one reason why it is important to know proofs. One cannot use the statement of Theorem 2.5 to solve the problem. However, one can solve this problem by modifying the proof of that theorem.

2.8. (i) Prove that an integer $m \geq 2$ is a perfect square if and only if each of its prime factors occurs an even number of times. (Hint: Use the ***Fundamental Theorem of Arithmetic:*** If $p_1^{e_1} \cdots p_n^{e_n} = q_1^{f_1} \cdots q_t^{f_t}$, where $p_1 < p_2 < \cdots < p_n$ and $q_1 < q_2 < \cdots < q_t$ are primes and the e’s and f’s are positive integers, then $n = t$ and, for all i, $p_i = q_i$ and $e_i = f_i$. [This theorem is usually proved in the next course.])

(ii) Using part (i), prove that if m is a positive integer for which $\sqrt{m}$ is rational, then m is a perfect square.

Conclude that √2, √3, and √6 are irrational.

(iii) If n is a positive integer, show that $n^3 + n^2$ is a perfect square if and only if $n + 1$ is a perfect square. (Example: If $n = 8$ (so that $n + 1 = 9 = 3^2$), then $n^3 + n^2 = 512 + 56 = 576 = 24^2$. The next example occurs when $n = 15$.)

2.9. Let p be a prime number, and consider the number

$$N = N_p = 1 + (2 \times 3 \times 5 \times 7 \times 11 \times . . . \times p).$$

Prove that none of the prime numbers 2, 3, 5, 7, 11, ⋯ , p used in the definition of N is a divisor of N. (Hint: Dividing N by 3, for example, leaves remainder 1, for the quotient is $2 \times 5 \times 7 \times 11 \times . . . \times p$.)

2.10. Use Exercise 2.9 to prove that there are infinitely many prime numbers. (Hint: Assume, on the contrary, that there are only a finite number of primes, say, 2, 3, 5, 7, ⋯ , p; define N as in Exercise 2.9 and show that N does not satisfy Theorem 1.2.) (This is an argument given in the ninth book of Euclid's *Elements*.)

2.11. (i) If $p = 11$, the number N defined in Exercise 2.9 is 2311. Show that 2311 is prime.

(ii) If $p = 13$, the number N is 30031. Show that 30031 is not prime.

(iii) If p = 17, show that 19 is a divisor of N = 510511.

2.12. A mad architect has designed the symmetric building shown in Figure 2.15. Find the area of the building's front (not counting the two circular windows of radius 2 or the semicircular entrance way), given the dimensions in the figure.

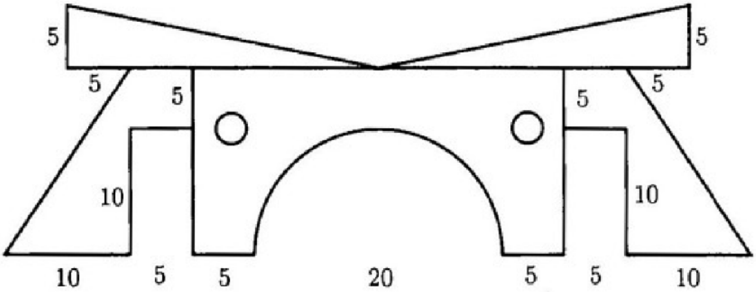


Figure 2.15

THE PYTHAGOREAN THEOREM

> Euclid alone has looked on beauty bare.
>
> *Edna St. Vincent Millay*

From earliest times, algebraic identities were verified by geometric figures. The simplest example is Figure 2.16, which is a geometric picture of the identity $(a + b)^2 = a^2 + 2ab + b^2$; the large square having sides of length $a + b$ is divided into two squares, with side lengths a and b, respectively, and two rectangles each of area ab.

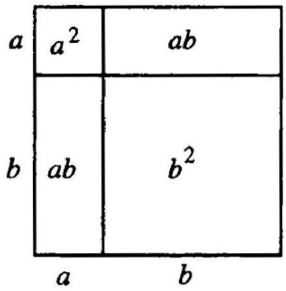


Figure 2.16

Recall that the *hypotenuse* (from the Greek word meaning "to stretch") of a right triangle is the longest of its three sides; the other sides are called its *legs*. In Theorem 2.3, we proved the special case of the Pythagorean theorem involving an isosceles right triangle.

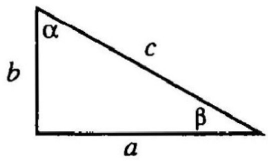


Figure 2.17

Theorem 2.6 (Pythagorean Theorem). In a right triangle with legs of lengths *a* and *b* and hypotenuse of length c, we have $c^2 = a^2 + b^2$.

Remark. Although the statement of this theorem was accepted centuries before him, Pythagoras was perhaps the first to prove it. The proof we give, due to Indian mathematicians around 400 AD, is based on Figure 2.18.

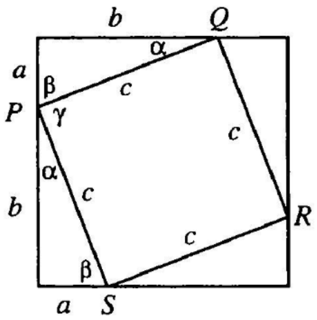


Figure 2.18

Proof. Figure 2.18 pictures the area of the big square in two ways: first, as a square with side lengths $a + b$; second, as dissected into a rhombus *PQRS* with side lengths c, and four congruent right triangles of area $\frac{1}{2}ab$. We claim that the rhombus is actually a square. Consider the interior angle γ at *P*, for example. Note that $\alpha + \gamma + \beta = 180°$. Inasmuch as $\alpha + \beta = 90°$, because *α* and *β* are the acute angles in a right triangle, we have $\gamma = 90°$. The Pythagorean theorem now follows from the algebraic identity

$$(a + b)^2 = c^2 + 4 \times \tfrac{1}{2}ab,$$

for the left side is $a^2 + 2ab + b^2$, while the right side is $c^2 + 2ab$.

The converse of the Pythagorean theorem is also true.

Theorem 2.7. A triangle having sides of lengths *a*, *b*, and *c* with $a^2 + b^2 = c^2$ must be a right triangle.

Proof. Take two perpendicular lines, as in Figure 2.19, and choose points A and B with $|AO|$ = a and $|BO| = b$. Now ΔOAB is a right triangle having side lengths a, b, and d. The Pythagorean theorem gives $a^2 + b^2 = d^2$, and so $d = c$. But ΔOAB and the given triangle are congruent, by "side-side-side," and so Δ is a right triangle.

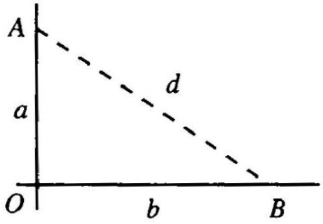


Figure 2.19

Figure 2.20 gives another proof, also in the Indian style, of the Pythagorean theorem.

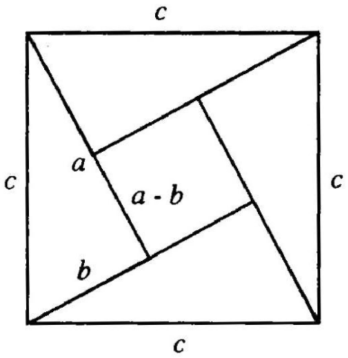


Figure 2.20

The square of side length c is partitioned, with a square of side length *a* – *b* in the center. This yields

$$c^2 = (a - b)^2 + 4 \times \tfrac{1}{2}ab$$

for the total area. The Pythagorean theorem follows. Query: What does Figure 2.20 look like when the right triangle is isosceles?

The proof of the Pythagorean theorem given in Euclid's *Elements* is based on Figure 2.21. Note that it is almost purely geometric (in contrast to the two proofs just given, which involve some algebra as well), and, thus, it is more complicated for us than the preceding proofs. On the other hand, notice that it actually displays the (geometric) squares sitting on the three sides of the right triangle. The big square is divided into two rectangles of areas pc and qc. It suffices to show that

$$a^2 = pc \text{ and } b^2 = qc.$$

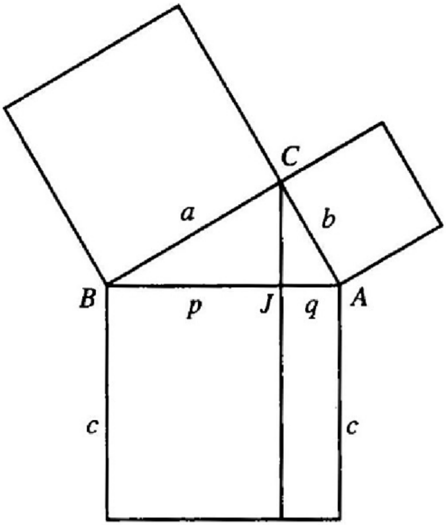


Figure 2.21

There is an algebraic proof of these equations. The altitude *CJ* forms similar right triangles Δ*JBC*, Δ*CBA*, and Δ*JCA*, as in Figure 2.22. The corresponding sides of these similar triangles can be seen in Figure 2.23. Thus, there are the proportions

$$\frac{p}{a} = \frac{a}{c} \quad \text{and} \quad \frac{q}{b} = \frac{b}{c},$$

and cross multiplication gives the desired equations.

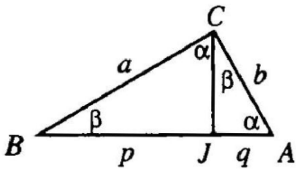


Figure 2.22

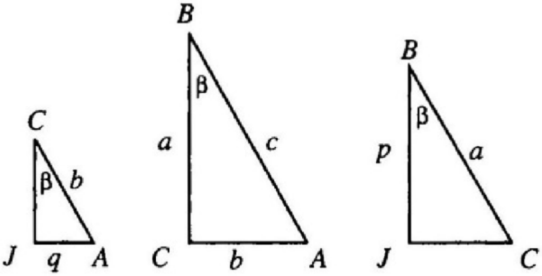


Figure 2.23

At this stage, however, Euclid did not want to use proportions. Although such algebraic manipulations are routine for us, $p : a = a : c$ in Euclid's time was not merely an equation $p/a = a/c$ (as we remarked earlier); rather, it was a relation between two pairs of line segments. Geometry and algebra were already living in separate worlds, and Euclid's treatment of proportions is very sophisticated. Thus, Euclid was not being a purist in avoiding proportions; his proof, which looks more complicated to us, was the most straightforward one that he knew.

Euclid's geometric argument is best explained by referring to Figure 2.24.

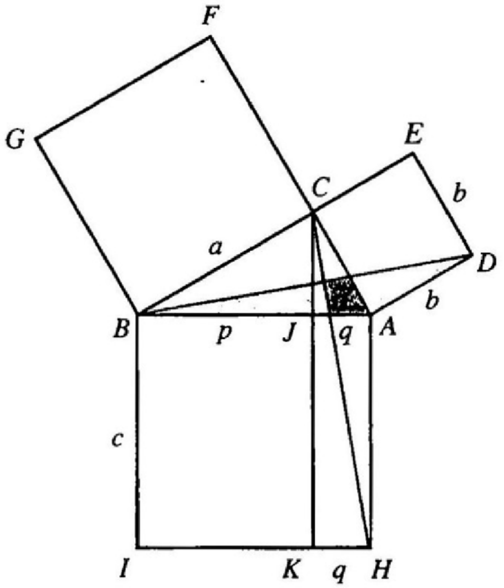


Figure 2.24

The area c^2 of the square *AHIB* is the sum of the areas of the rectangles *AHKJ* and *BJKI*. We want to show that

$$\text{area}(AHKJ) = b^2 \text{ and area}(BJKI) = a^2,$$

which will prove the desired result.

Now, by Theorem 2.1,

$$\text{area}(\triangle CAH) = \tfrac{1}{2}|AH||AJ|,$$

for *AH* is a base and *AJ* is the corresponding altitude. Thus,

$$\text{area}(\triangle CAH) = \tfrac{1}{2}\,\text{area}(\square AHKJ).$$

Next we compare ΔCAH with ΔABD. The angles $\angle CAH$ and $\angle DAB$ are equal, for each equals $\angle BAC + 90°$. By the side-angle-side theorem, it follows that ΔCAH and ΔABD are congruent. But

$$\text{area}(\triangle ABD) = \tfrac{1}{2}|AD||AC| = \tfrac{1}{2}b^2,$$

for *AD* is the base and *AC* is the altitude. It follows that

$$\text{area}(\quad AHKJ) = 2\ \text{area}(\Delta CAH) = 2\ \text{area}(\Delta ABD) = b^2.$$

A similar argument shows that

$$\text{area}(\quad BJKI) = a^2,$$

and the Pythagorean theorem follows.

We now give a geometric proof of the Pythagorean theorem that Euclid would have loved had he known it[8] (this proof can be found in a commentary in Heath’s 1926 translation of Euclid). Figure 2.25 displays a square of side length a+b dissected in two different ways. The left dissection is Figure 2.18; the right dissection is Figure 2.16 with each rectangle bisected by a diagonal.

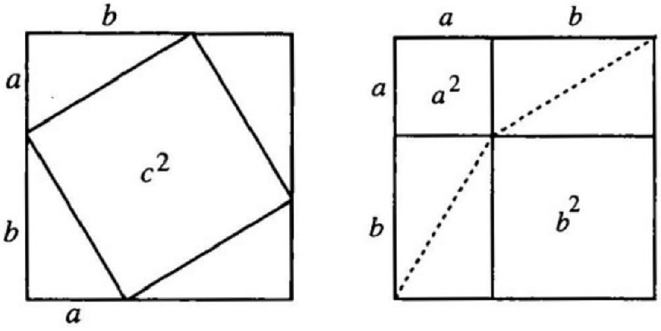


Figure 2.25

That both squares have the same area gives the equation

$$c^2 + 4 \times \tfrac{1}{2}ab = a^2 + b^2 + 4 \times \tfrac{1}{2}ab.$$

Here are some applications of the Pythagorean theorem. In a later section of this chapter, we will discuss a bit of trigonometry (which is also intimately related to the Pythagorean theorem).

Aristarchus (310 – 250 BC) used the Pythagorean theorem to draw conclusions about the distance from the Moon to the Earth. His idea was as follows. At halfmoon, the Sun *S*, the Moon *M*, and the Earth *E* form a right triangle, with right angle at *M*, as in Figure 2.26.

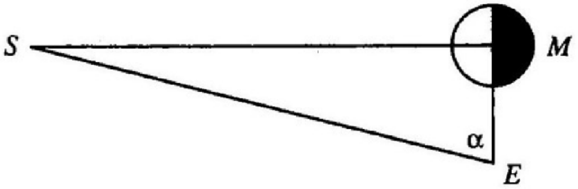


Figure 2.26

Let

$$|SE| = \text{distance from } S \text{ to } E,$$
$$|SM| = \text{distance from } S \text{ to } M,$$
$$|ME| = \text{distance from } M \text{ to } E.$$

The Pythagorean theorem gives

$$|SE|^2 = |SM|^2 + |ME|^2.$$

One conclusion Aristarchus drew from this is that the Earth is farther from the Sun than it is from the Moon. This is not at all obvious, for both the Sun and the Moon appear to be the same size (of course, having observed solar eclipses in which the Moon is seen to come between the Earth and the Sun, the Greeks would have known this fact without using the Pythagorean theorem). At sunset, if one is looking up at the (half) Moon, sunlight seems to be perpendicular to *EM,* the line of sight to the Moon; that is, the angle α seems very close to 90°. Now $\cos \alpha = |ME|/|SE|$. When α is close to 90°, cos α is close to zero. Aristarchus was, thus, able to conclude that the Earth is very much farther from the Sun than it is from the Moon, and so this fact follows without sophisticated instruments. [Modern measurements have $|ME| \approx 240,000$ miles and $|SE| \approx 93,000,000$ miles, so that $\cos \alpha = |ME|/|SE| \approx .0026$ and $\alpha \approx 89.85°$ (or $89°51'$).]

The following ***castle problem*** is from an old Chinese text.

> There is a circular castle, whose diameter is unknown; it is provided with four gates and two lengths out of the north gate there is a large tree, which is visible from a point six lengths east of the south gate. What is the length of the diameter?

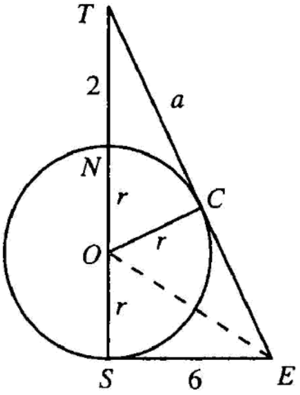


Figure 2.27

We compute the area of ΔTSE in two ways. On the one hand,

$$\text{area}(\Delta TSE) = \tfrac{1}{2}(2 + 2r)6 = 6 + 6r.$$

On the other hand, this area is the sum of the areas of the three smaller triangles ΔOSE, ΔOCE, and ΔOCT. Now the right triangles