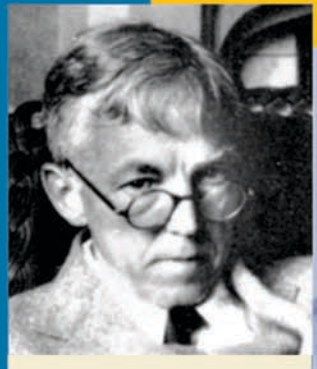
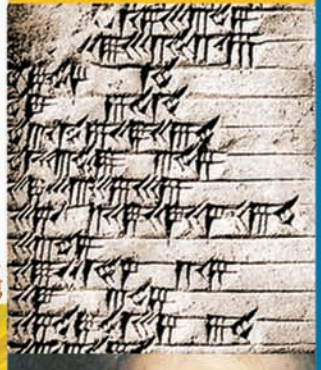
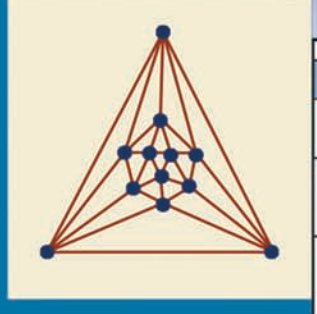
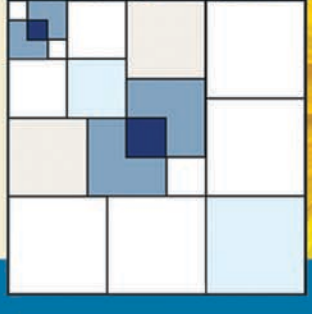
Number Theory

A Historical Approach



JOHN J. WATKINS

NUMBER THEORY

NUMBER THEORY

A Historical Approach

JOHN J. WATKINS

PRINCETON UNIVERSITY PRESS
Princeton and Oxford

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TW
press.prenceton.edu

Library of Congress Cataloging-in-Publication Data

Watkins, John J., author.
Number theory : a historical approach / John J. Watkins.
pages cm
Includes index.
Summary: “The natural numbers have been studied for thousands of years, yet most undergraduate textbooks present number theory as a long list of theorems with little mention of how these results were discovered or why they are important. This book emphasizes the historical development of number theory, describing methods, theorems, and proofs in the contexts in which they originated, and providing an accessible introduction to one of the most fascinating subjects in mathematics.Written in an informal style by an award-winning teacher, Number Theory covers prime numbers, Fibonacci numbers, and a host of other essential topics in number theory, while also telling the stories of the great mathematicians behind these developments, including Euclid, Carl Friedrich Gauss, and Sophie Germain. This one-of-a-kind introductory textbook features an extensive set of problems that enable students to actively reinforce and extend their understanding of the material, as well as fully worked solutions for many of these problems. It also includes helpful hints for when students are unsure of how to get started on a given problem. Uses a unique historical approach to teaching number theory Features numerous problems, helpful hints, and fully worked solutions Discusses fun topics like Pythagorean tuning in music, Sudoku puzzles, and arithmetic progressions of primes Includes an introduction to Sage, an easy-to-learn yet powerful open-source mathematics software package Ideal for undergraduate mathematics majors as well as non-math majors Digital solutions manual (available only to professors) ”– Provided by publisher.
ISBN 978-0-691-15940-9 (hardback)
1. Number theory. I. Title.
QA241.W328 2014
512.7–dc23
2013023273

British Library Cataloging-in-Publication Data is available

This book has been composed in ITC Stone Serif Std and ITC Stone Sons Std

Printed on acid-free paper. ∞

Typeset by S R Nova Pvt Ltd, Bangalore, India
Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

In Fond Memory

For David Roeder (1939–2011)

Contents

Preface

Many years ago, I was sitting in my second-grade classroom when I made what I thought was a remarkable discovery: *there is no largest number*. Whatever number I thought of, I realized I could just add one to it and get a larger number. This remains to this day one of my most vivid childhood memories. What I had "discovered" was the "dot, dot, dot" in the infinite collection of the numbers

$$1,\ 2,\ 3,\ 4,\ 5,\ \ldots\ ,$$

which we know as the *natural numbers*.

As simple as this collection of numbers may appear, humans have been studying these numbers for thousands of years, learning their properties, uncovering their secrets, finding one marvelous thing after another about them, and still we have only barely begun to tap this remarkable and ever-flowing current of ideas. These are the numbers we intend to study.

This book is an introduction to the study of the natural numbers; it evolved from courses I have taught at Colorado College, ranging from a general math course designed for nonmajors to a far more rigorous sophomore-level course required of all math majors. I hope to preserve several fundamental features of these courses in this book:

- Number theory is beautiful. It is fun. That's why people have done it for thousands of years and why people still do it today. Number theory is so naturally appealing that it provides a perfect introduction—either for math majors or for nonmajors—to the idea of doing mathematics for its own sake and for the pleasure we derive from it.
- Although number theory will always remain a part of pure mathematics (as opposed to applied mathematics), it has also in modern times become a spectacular instance of what the physicist Eugene Wigner called the "unreasonable effectiveness of mathematics" in that there are now important real-world applications of number theory. One of the most useful of these applications came along several centuries *after* the original concepts in number theory were developed and will be explored in the chapter on cryptography.
- Number theory is a subject with an extraordinarily long and rich history. Studying number theory with due attention to its history reminds us that this subject has always been an intensely

human activity. Many other mathematical subjects, calculus, for example, would have undoubtedly evolved much as they are today quite independent of the individual people involved in the actual development, but number theory has had a wonderfully quirky evolution that depended heavily upon the particular interests of the people who developed the subject over the years.

- Reading mathematics is very different from, say, reading a novel. It requires enormous patience to read mathematics. You cannot expect to digest new, and often complex, mathematical ideas in a single reading. It is frequently the case that multiple readings are needed. You will discover that individual sentences, paragraphs, and even whole chapters must be read carefully several times before the key ideas all fall into place.
- One of the primary goals of the book is to use the study of number theory as a context within which we learn to prove things. Proof plays a vital role in mathematics and is the way we bridge the gap between what our intuition tells us *might* be true and the certainty about what *is* true. You will encounter several quite different styles of proof as you read (and should feel free to skip any that you find either too difficult or simply not very interesting). In many cases, an informal argument or even a carefully examined example is sufficient to discover truth, but in other cases a far more rigorous and formal argument will be required to achieve certainty.

Another feature of our courses at Colorado College I hope to preserve in this book is the *interactive* nature of our classes. Learning mathematics requires active participation, and this book should be read with paper and pencil in hand, and a good calculator or computer nearby, checking details and working things through as you go. Sometimes, in order to understand an idea, it is best to go through a few examples by hand. Other times it is better to let a computer do the computations, and so an introduction to the computer software Sage has been provided at the back of the book. Sage is an extremely powerful aide to such computations and is a wonderful resource that can be used online or downloaded for free.

The problems at the end of each chapter are an important part of the text and you should try to do as many as you can. In this book problems are not merely exercises for you to do, but they also introduce definitions, explore new ideas, and prove additional results. Much of this material will be used later in the book, and so you should be sure to read *all* of the problems, even ones you make no attempt to solve. Problems that are either particularly important or explicitly referred to later in the book have the symbol $\star$ by them.

Solutions and answers are provided for many of these problems at the back of the book. There is also a separate section containing hints for you to consult to get an idea of how to start on a problem if you are stuck. Problems for which a hint is available have a letter H after them, and problems for which there is a solution or answer have a letter S after them. These solutions and hints can be used in a variety of ways: to check your answers; to compare your solutions with mine (there are often several ways to approach a given problem); to study how to write up a solution once you have figured out how to solve a problem; but, also, just to read as part of the text, since I occasionally make additional, and hopefully useful and interesting, comments about the material in these solutions.

Also at the back of the book are two useful tables. One is simply a short list of prime numbers. The other is a pronunciation guide to help you with the names of foreign mathematicians. Rather than using a phonetic alphabet, these pronunciations are given in a form that should make it easy for any speaker of standard (American) English to get reasonably close to an accurate pronunciation. So, for example, (THA bit) is used for the ninth-century Arab mathematician Thābit ibn Qurra, rather than the phonetically correct (Θa:bɪt).

It is probably obvious that covering all of the material in this book in a typical number theory course is not possible. I tend to think of Chapters 1–10 forming the core material and the topics covered in Chapters 11–15 being optional, perhaps to be done by students either individually or in groups as independent study. In the table of contents I have marked individual sections that I consider critical with a $\star$.

Many people have at various stages helped me write this book. The first and foremost was the long-time chair of our math department, Dave Roeder. It was Dave who put a number theory course at the very core of our math curriculum, and over the years it became my very favorite course to teach. I also owe a deep debt to my colleague Stefan Erickson who, unlike me, is a real number theorist and has used numerous drafts of this book in his own course on number theory. Stefan guided me with enormous patience through draft after draft. He also provided me with extraordinarily detailed student feedback from these courses. One result of this extensive "field-testing" is that there have been many students whose comments greatly improved this book. In particular, two of these students deserve special mention. Gautam Webb's careful reading of the latest draft uncovered more errors than I would have believed possible. Marina Gresham did the same sort of meticulous reading of several early drafts; more importantly, I relied almost exclusively on Marina's excellent judgment in deciding which problems needed to be provided with hints and solutions.

Finally, I would like to say that while this book is modeled upon specific courses I have taught at Colorado College, this book is nonetheless intended for a far more general audience; and so there is almost nothing in terms of prerequisites that a readers need to bring along with them except enthusiasm and curiosity. That is one of the fundamental charms of number theory. It really does begin with

$$1,\ 2,\ 3,\ 4,\ 5,\ \ldots,$$

and you can't be too young, or too old, to enjoy this amazing story.

John J. Watkins
Colorado Springs, Colorado

NUMBER THEORY

1

Number Theory Begins

Pierre de Fermat

Any good story can be told in a variety of ways. Often it is simply best to let events unfold in strict chronological order. But one must then sometimes pause to backtrack in the story every so often in order to explain one or two things that may not be clear to the audience. One can even tell a tale by beginning at the very end and spin the entire story out in a series of flashbacks, slowly and tantalizingly revealing everything one layer at a time.

I have chosen to tell the story of number theory by beginning with the first person who really thought about numbers in much the same way as we do today, and for this reason he is the first mathematician who could accurately be described as a *number theorist*. The man's name is Pierre de Fermat. The year is 1659 and Fermat has just written to his friend Christiaan Huygens bragging about having discovered a "most singular method" for proving mathematical propositions and mentioning as an example one of his most important early results:

> *There is no right-angled triangle in numbers whose area is a square.*

Let us begin our story of numbers here then, three hundred and fifty years ago with Fermat's proof of this proposition. His proof is actually quite short, but we will spend a great deal of time in this chapter developing his proof because I want to use this proposition as a way to introduce you informally to several basic ideas and topics in the theory of numbers. And so I intend to present the proposition in a series of flashbacks so that when we get to the actual proof you already know everything that Fermat knew when he discovered this proof. So try to keep in mind as you read this chapter that our ultimate goal is the proof of Fermat's proposition that no right-angled triangle has square area.

Pythagorean Triangles

The first thing to understand about this proposition is that Fermat is considering only whole numbers, what we now call *integers*—that is, an *integer* is a number in the set $\mathbf{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ (by

Figure 1.1 Pierre de Fermat, 1601–65.

the way, the letter Z is used for this set because the German word for numbers is *Zahlen*).

So when Fermat says "triangle in numbers" and "whose area is square" he means that the three sides of the triangle are all to have integer length and that the area is also an integer that is itself the square of another integer. In other words, all of the numbers in Fermat's proposition come from the set of *natural numbers*

$$\mathbf{N} = \{1, 2, 3, \ldots\}.$$

These days we call such right triangles *Pythagorean triangles*—a reference to the well-known Pythagorean theorem of high school geometry—and if three natural numbers a, b, c are such that $a^2 + b^2 = c^2$, then we call this set of numbers a *Pythagorean triple*. Furthermore, a Pythagorean triple $\{a, b, c\}$ is said to be *primitive* if the three numbers have no common positive factor other than 1.

Why are we interested in primitive Pythagorean triples? Because primitive Pythagorean triples represent fundamentally different triangles. For example, the two triples {3, 4, 5} and {6, 8, 10} correspond to two triangles that have exactly the same shape, so the only way these two triangles really differ is in terms of their size; one triangle is simply an expanded version of the other triangle. When trying to prove propositions such as the one of Fermat's about right triangles

not having square area, it is often much easier not to consider *all* right triangles, but just to consider those corresponding to primitive Pythagorean triples, such as {3, 4, 5}, {5, 12, 13}, and {8, 15, 17}.

Babylonian Mathematics

This brings us to a serious flashback because Pythagorean triangles have been known about for a very long time, even long before the time of Pythagoras, who himself lived in about the sixth century B.C. For instance, we happen to know that they were an important part of early Babylonian mathematics because the records of that Mesopotamian empire were kept by scribes on clay tablets written in a style known as *cuneiform* script (because of the distinctive wedge-shaped marks made in these clay tablets using a stylus). Many of these tablets survived to this day because of the dry climate of that region—Babylon, the capital of this empire, was located on the Euphrates about sixty miles south of present-day Baghdad. One of these ancient tablets somehow made its way into a private collection in Florida before finally becoming a permanent part of the Plimpton collection at Columbia University, where it was given the catalog number 322. This particular tablet is now quite famous and is called, simply, *Plimpton 322*.

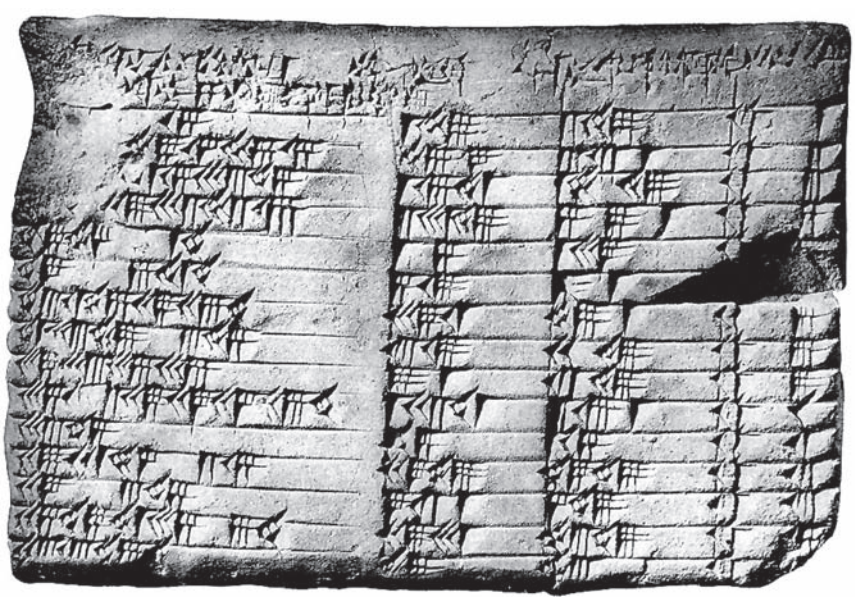


Figure 1.2 Plimpton 322.

Plimpton 322 has been described by mathematician and science historian Otto Neugebauer as "one of the most remarkable documents of Old-Babylonian mathematics." It contains fifteen rows and four columns (although there is some damage), the fourth column being just a numbering of the rows 1 through 15. Until Neugebauer deciphered this tablet it had been considered merely a "commercial account." What Neugebauer managed to figure out is that instead this tablet effectively contains a list of fifteen Pythagorean triples: the middle two columns are the hypotenuse and the shortest side of right triangles. For each of these fifteen triangles, if we call the hypotenuse c and the shortest side a, and then compute $\sqrt{c^2 - a^2}$, we get an integer; in other words, they knew the Pythagorean theorem in this part of the world twelve hundred years before Pythagoras!

The rows on the tablet begin in the first row with a right triangle {119, 120, 169} that is nearly isosceles—that is, the two legs are almost equal—and the triangles gradually change shape as you move down the tablet until you end at the bottom row with a right triangle {56, 90, 106} whose legs are not at all equal. It is worth noting that the largest triangle on the list is the fourth one {12 709, 13 500, 18 541} and remembering that this triangle was computed thirty-five hundred years before calculators. The first column turns out to contain the numbers $c^2/(c^2 - a^2)$, so this column records the square of the ratio of the hypotenuse to the "third" side—in modern terminology this would be represented as the square of the cosecant of the angle between the hypotenuse and the shortest side—and this ratio gradually diminishes as you go down the tablet.

At this point we could happily end this flashback into Babylonian mathematics satisfied that we have seen ample evidence of an awareness of the Pythagorean theorem from such a distant time in the human past, but Neugebauer discovered something even more interesting about Plimpton 322. He discovered why this tablet contains the fifteen Pythagorean triples that it does. To understand this, and thus continue our flashback, we need to talk about the way in which the Babylonians represented numbers.

Sexagesimal Numbers

In the 1997 film *Contact*, based on a novel of the same name by Carl Sagan, Jodie Foster plays the role of a brilliant astronomer who is the first human to receive a message from extraterrestrial beings. She has been monitoring an array of radio telescopes in New Mexico and knows the signal she is receiving from a distant star can only be coming from intelligent life because the signal is repeating a sequence of prime

numbers over and over again. What better way can there be to shout into the universe "We are here" than to send a message about numbers that can be universally understood by anyone who hears it?

Number theory is the study of inherent properties of numbers. For example, whether a number is odd or even is an inherent property of a number; it doesn't depend upon how the number is represented. The number 17 is odd whether we represent it, as we just did, in the familiar decimal system, or as XVII, in the Roman numeral system, or as 10001, in the binary system. Similarly, the fact that 17 is a prime number doesn't depend on how it is represented. That's why any curious, intelligent life form anywhere in the universe is eventually going to discover prime numbers. The fictional beings from that distant planet revolving around the star Vega who sent us a message in that movie had discovered prime numbers. And here on earth, thirty-five hundred years ago, the Babylonians had also discovered prime numbers, and had been fascinated by them.

The Babylonians used a *sexagesimal* number system much like our own decimal system, but based instead upon the number 60. They got the idea from the Sumerians and, in fact, we still use a version of their system today for some parts of our lives. We measure time in units of sixty: 60 minutes in an hour, and 60 seconds in a minute. We measure angles and navigate using degrees: 360 degrees in circle, 60 minutes of arc in a degree, 60 seconds of arc in minute of arc. The reason for choosing 60 as a base for a number system, and the reason we still use it for some purposes today, is that 60 has so many different factors. In particular, then, an hour, or a circle, can conveniently be broken up into 1, 2, 3, 4, 5, 6, or even more parts.

So, how does a sexagesimal system work? Well, when we write a number such as 3456 in our decimal system, what we mean is that

$$3456 \;=\; 3 \times 10^3 + 4 \times 10^2 + 5 \times 10^1 + 6 \times 10^0.$$

How do we write this same number in the Babylonian system? We need to write it in terms of powers of 60, that is, as something like

$$a \times 60^2 + b \times 60^1 + c \times 60^0,$$

where we just have to figure out what a, b, and c need to be. But $60^2 = 3600$, so it turns out we don't need the $a \times 60^2$ term at all for 3456. So divide 60 into 3456 and get 57.6, which means that we should let $b = 57$, and then $c = 3456 - 57 \times 60 = 36$. Therefore, in the sexagesimal system we would express 3456 as $57 \times 60^1 + 36 \times 60^0$. We write this more conveniently as $57,\ 36$ meaning 57 "sixties" plus 36 "ones."

Let's do one in the other direction. What number does the sexagesimal number $4, 37, 46, 40$ represent? It represents

$$4 \times 60^3 + 37 \times 60^2 + 46 \times 60^1 + 40,$$

which is equal to $4 \times 216\,000 + 37 \times 3600 + 46 \times 60 + 40 = 1\,000\,000$. Note that we are using the international system of marking off the digits in large numbers by groups of three using spaces rather than commas.

Now, of course, in the decimal system we have ten different symbols for our ten digits, but as you can imagine the Babylonians didn't have sixty different symbols to use. Instead, for the number 57 they would just make seven vertical marks next to five marks shaped like $<$ that each represented 10. You might barely be able to make out these two kinds of marks in Figure 1.2. Another way in which the Babylonians' system differed from our current decimal system is that they didn't have a symbol for zero, so they just left a blank space. This meant that two very different numbers such as 7232 and 432 032 would look the same since their sexagesimal representations are, respectively, $2, 0, 32$ and $2, 0, 0, 32$, and so, with only a space between 2 and 32 in each case, there would be no way to tell what power of 60 to use for the 2. This would have to be inferred from the context, which was not usually at all difficult in practice.

Regular Numbers

The Babylonian system differed from ours in still another way: it did not use a decimal point, or, rather, we should say a sexagesimal point. For us the numbers 3456 and 3.456 are very different. The latter number, of course, means

$$3 + 4 \times 10^{-1} + 5 \times 10^{-2} + 6 \times 10^{-3}.$$

The Babylonian system on the other hand was more flexible—and, again, context would be used to resolve any ambiguity. For example, $5, 30$ could represent 330, that is, $5 \times 60 + 30$; but it could also represent $5 + 30 \times \frac{1}{60}$, that is, $5\frac{1}{2}$.

As another example, we could represent the fraction $\frac{1}{3456}$ in the sexagesimal system as $1, 2, 30$ because

$$\frac{1}{3456} = \frac{1}{60^2} + \frac{2}{60^3} + \frac{30}{60^4}.$$

(Check this if you want.) This is really rather remarkable. A fraction such as $\frac{1}{3456}$, which in the decimal system becomes 0.000 289 351 85 . . . , and goes on forever, has in the sexagesimal system a *finite* representation.

There is a simple reason this happens, and it has to do with the factors of 60. Since $60 = 2^2 \cdot 3 \cdot 5$, the only prime factors of 60 are 2, 3, and 5. The Babylonians discovered that for any number that has no prime factors other than 2, 3, or 5—such as $3456 = 2^7 \cdot 3^3$, for example—the reciprocal of the number has a *finite* representation. But if you take any other number and try to write its reciprocal as a sexagesimal fraction, this fraction will go on forever. So, numbers that have no prime factors other than 2, 3, and 5 were very important to the Babylonians. Neugebauer called these numbers *regular* numbers.

Now, let's look again at Plimpton 322 and the fifteen triangles it contains. Here is a list of all fifteen of the "third" sides of these triangles, which we get by computing $\sqrt{c^2 - a^2}$: 120, 3456, 4800, 13500, 72, 360, 2700, 960, 600, 6480, 60, 2400, 240, 2700, 90. Notice that the number 3456 is on that list. These are not just any old integers, but they all share with 3456 the special property that was very important in Babylonian mathematics: the only prime factors of any of these numbers are 2, 3, and 5—that is, they are all regular numbers!

Square Numbers

I hope that during this flashback to Babylonian mathematics you haven't forgotten about Fermat's proposition that no Pythagorean triangle has a square area. As it happens, one of the earliest translations that was ever done of a Babylonian clay tablet was of a tablet that is nothing more than a table that lists the numbers from 49 to 60 and their squares. The property of a number being a square was something that was very important to the Babylonians.

The *square numbers* are the numbers 1, 4, 9, 16, 25, . . . and, as you can infer from their appearance on a Babylonian clay tablet, these particular numbers have fascinated people since ancient times. When you saw this list of square numbers just presented to you, you undoubtedly thought to yourself something like: of course I recognize 9 is a "square" number because $9 = 3^2$, and 16 is a "square" number because $16 = 4^2$. But twenty-five hundred years ago a young Greek student of mathematics in the city-state of Ionia would have thought something more like: of course 9 and 16 are "square" numbers because piles of nine stones and sixteen stones can each be arranged into square arrays

of stones on the ground. One of you thinks of the concept "square number" algebraically, and the other thinks of the same concept geometrically.

This will become a recurring theme as we continue our study of number theory. Just as we did with square numbers we will assign various traits to numbers and speak of there being *prime* numbers, *regular* numbers, *perfect* numbers, *triangular* numbers, *Fibonacci* numbers, *Mersenne* numbers—the list goes on and on, each term describing numbers that have a particular property that we find interesting. For each of these categories of numbers you will want to try to get a feeling for what makes that kind of number special. For square numbers we have lost in modern times that "feel" of the geometric quality that makes them special. Fermat undoubtedly had a much fuller appreciation of both the algebraic and geometric nature of square numbers than we do today. One of the great mathematicians of modern times, Paul Erdős, was legendary for the "feel" he had for numbers. At an international conference in Boca Raton, Florida, in 1994, Erdős expressed this great affection he had for numbers during one of his famous annual addresses to the conference in a typically humorous way by telling the audience that he suspected he was, at the age of eighty-one, "probably a square for the very last time."

As for Fermat's proposition about triangles, the reason we are able to restrict our attention to *primitive* Pythagorean triples is because of the intimate way area is linked to squaring. Suppose we have two similar triangles such as a $\{3, 4, 5\}$ triangle and a $\{6, 8, 10\}$ triangle. The larger triangle has sides that are *twice* as long as those in the smaller triangle, but its area is *four* times that of the smaller triangle. Area $= \frac{1}{2}$ (base $\times$ height), so the small triangle has area $\frac{1}{2} \cdot 3 \cdot 4 = 6$, and the larger triangle has area $\frac{1}{2} \cdot 6 \cdot 8 = 24$.

Similarly, the $\{9, 12, 15\}$ triangle has sides that are *three* times as long, but area that is *nine* times that of the smaller $\{3, 4, 5\}$ triangle, since its area is $\frac{1}{2} \cdot 9 \cdot 12 = 54$. In this same way, any Pythagorean triangle that is similar to the primitive triangle $\{3, 4, 5\}$ will have an area that is a *square* multiple of the area of this smaller triangle. Thus, since the $\{3, 4, 5\}$ triangle does not have square area, no Pythagorean triangle similar to it can have square area. (Of course, if the $\{3, 4, 5\}$ triangle *did* have square area, this same argument would mean that *any* Pythagorean triangle similar to it would also have square area.)

Note that this argument relies entirely on a basic fact about numbers that we will have to say more about later, namely, if you multiply a square number by a square number you get another square number, but if you multiply a non–square number by a square number you get a non–square number.

Primitive Pythagorean Triples

So, to prove his proposition, Fermat needed to know exactly which triples of numbers form primitive Pythagorean triples. Not only did Fermat know this, but this is something that had been known for a long time. Neugebauer even thinks the Babylonians knew this! And that brings us once again to Plimpton 322.

It turns out that the scribes made four errors on this tablet, and the nature of these errors makes it clear that the first column was not being computed directly from the middle two columns. In other words, they must have had some other method for producing the numbers in these three columns. Neugebauer believes that behind the scenes for each row of Plimpton 322 lay two small regular numbers s and t. For example, for the first row there would be the two regular numbers $s = 12$ and $t = 5$. Then, the numbers for the middle two columns would be computed by finding

$$c = s^2 + t^2; \quad a = s^2 - t^2.$$

For the first row, you would get $c = 12^2 + 5^2 = 169$, and $a = 12^2 - 5^2 = 119$. The number in the first column would be computed by finding $\left((s^2 + t^2)/(2st)\right)^2$, which in this case is $(169/120)^2$, and this is exactly the number 1.59, 0, 15 written sexagesimally that Neugebauer found on the tablet—or, at least he found the 15, the rest had been obliterated. The "third" side of this triangle, call it b, would be computed as $b = 2st$. Here, you would get $b = 120$.

When viewed in this light, Plimpton 322 starts to make a lot more sense. Take, for example, the largest triangle on the list, represented by the triple {12709, 13500, 18541}. At first glance this triangle seems quite out of place. But for this triple, the values of the two regular numbers s and t are $s = 125$ and $t = 54$, which again seems somewhat arbitrary until we look at their prime decomposition: $125 = 5^3$ and $54 = 2 \cdot 3^3$. So, in fact, these are very simple numbers built from the fundamental building blocks in the Babylonian system: 2, 3, and 5. This same pattern holds for each row on the tablet: one row would have $s = 2^5$ and $t = 3 \cdot 5$; another, $s = 2 \cdot 5^2$ and $t = 3^3$. There is one exception: the row for the triangle {45, 60, 75} is the only row where the numbers have a common factor. This triangle of course is similar to the triangle {3, 4, 5} (for which we can use $s = 2$ and $t = 1$) but is much more in the same scale as the other triangles on the list.

Row after row of the tablet, the values of s and t satisfy four properties: (i) $s > t$; (ii) one of s or t is even, and the other is odd; (iii) they are both regular numbers, that is, their prime decompositions use only the three primes 2, 3, and 5; and, finally, (iv) s and t are *relatively prime*,

that is, they have no common factor other than 1. These values of s and t never appear on the tablet of course; they are merely somewhere in the background. Perhaps the Babylonians were aware of these numbers and used them in their calculations, perhaps not. Perhaps they just used a simpler formula such as $b^2 = (c + a)(c - a)$—not that they would have been able to express it in this modern algebraic form. We just don't know.

Whether or not the Babylonians actually produced Plimpton 322 in the highly number theoretic way I have been suggesting, we are now ready to characterize primitive Pythagorean triples. This theorem, our first, was certainly known to Euclid, who gave a proof for this marvelous construction in the late fourth century B.C.

Theorem 1.1. *For any primitive Pythagorean triple* $\{x, y, z\}$ *where* $x^2 + y^2 = z^2$, *one of the numbers* x *or* y *must be even, and the other odd, so let* x *be the even number; then, there exist two positive integers* s *and* t, $s > t$, *one even and the other odd, with* s *and* t *having no common factor other than 1, such that*

$$x = 2st; \quad y = s^2 - t^2; \quad z = s^2 + t^2.$$

Moreover, if s *and* t *are any two such positive integers, then these formulas produce a primitive Pythagorean triple.*

Proof

First, we note that, for a primitive Pythagorean triple $\{x, y, z\}$, x and y cannot both be even, since then z would also be even and all three integers would have 2 as a factor. In order to show similarly that x and y cannot both be odd we will give an argument—much as Fermat would have done—based on an idea we will use often in this book having to do with the notion of remainders.

We know that a number can be either even or odd. We express this by saying that an even number can always be written in the form $2k$, where k is also an integer; and that an odd number can always be written in the form $2k+1$, where k is again an integer. So, for example, $26 = 2 \cdot 13$ and $57 = 2 \cdot 28 + 1$. Another way of saying this is that if we divide an integer by 2, there are only two possible remainders: 0 and 1.

Now, we want to show that x and y cannot both be odd, so let's see what would happen if they both *were* odd. That is, let us suppose that x and y are both odd. That means we can write $x = 2k + 1$ and $y = 2j + 1$. (We have to use a different letter j for y because we don't want to assume

that x and y are equal.) Now we can compute

$$x^2 + y^2 = (2k+1)^2 + (2j+1)^2 = 4k^2 + 4k + 1 + 4j^2 + 4j + 1$$
$$= 4(k^2 + k + j^2 + j) + 2.$$

We conclude that $x^2 + y^2$ is a number of the form $4i + 2$ (where in this case i is just the number $k^2 + k + j^2 + j$). Another way of saying this is that if we were to divide $x^2 + y^2$ by 4, we would get a remainder of 2. For example, since $70 = 4 \cdot 17 + 2$, we say that 70 is of the form $4i + 2$, or that if we divide 70 by 4 then the remainder is 2.

But recall that $x^2 + y^2 = z^2$, so z^2 must also be of the form $4i + 2$, and have a remainder 2 when divided by 4. However, as we shall see, this is impossible, because a square can *never* be of the form $4i + 2$ and have remainder 2 when divided by 4. Why not? Well, if z is an even number, then z can be written as $2k$, and then $z^2 = (2k)^2 = 4(k^2)$, and z^2 has a remainder 0 when divided by 4; on the other hand, if z is an odd number, then z can be written as $2k + 1$, and then $z^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 4(k^2 + k) + 1$, and z^2 has a remainder 1 when divided by 4. That is, squares have remainder 0 or 1 when divided by 4.

So, we found out exactly what happens if we assume that both x and y are odd. We end up with the conclusion that z^2 has a remainder 2 when divided by 4, but we know this is impossible. Therefore, our assumption had to be wrong, and we conclude that one of the numbers x or y must be even. And since we already decided they can't both be even, the other number is odd. We arbitrarily decide to let x be the even number and y the odd number. Note, then, that z will always be odd.

Since y and z are both odd, the numbers $z + y$ and $z - y$ are even, so we can write them as $z - y = 2u$ and $z + y = 2v$. (Note that it follows that $z = u + v$ and $y = v - u$.) Thus $x^2 = z^2 - y^2 = (z - y)(z + y) = 4uv$, and so $(\frac{x}{2})^2 = uv$. So uv is a square (note that $\frac{x}{2}$ is an integer since x is even). We claim that u and v are both squares.

To verify this claim we need to understand what makes a number a square in terms of its prime decomposition: each prime needs to occur an even number of times. So, $2^2 \cdot 5^6 \cdot 11^4$ will be a square, but $3^8 \cdot 7^3$ won't be a square. This means a product such as uv can be a square as long as each prime *collectively* shows up an even number of times in the prime decompositions of the two numbers u and v. But what if u and v don't have any primes in common? Then the only way uv can be a square is if u is a square and v is a square. That's the situation in our proof and what we need to show, namely, that u and v are relatively prime. But if u and v have a common positive factor d, then d is also a common factor for y and z, and hence a factor of x. Since $\{x, y, z\}$ is a *primitive* Pythagorean

triple, the only way this can happen is if $d = 1$. We conclude that u and v are both squares.

Therefore, we can write $u = t^2$ and $v = s^2$, and get

$$x^2 = 4uv = 4s^2t^2;\ y = v - u = s^2 - t^2;\ z = u + v = s^2 + t^2,$$

exactly as desired. Note that since u and v are relatively prime, so are s and t; this means that, in particular, one of s and t is even, and the other odd.

Finally, we still need to verify the converse, namely, that if s and t are any two positive integers with the given properties, then these formulas will produce a primitive Pythagorean triple. That the formulas produce a Pythagorean triple is straightforward algebra since $x^2 + y^2 = (2st)^2 + (s^2 - t^2)^2 = (s^2 + t^2)^2 = z^2$. If the triple is not primitive, then there is a prime number p that is a factor for all three numbers x, y, and z. Since one of s and t is even and one is odd and $z = s^2 + t^2$, we know that z is odd. So, in particular, we know that p can't be 2. Now, p is a factor of both y and z, so p is also a factor of their sum $y + z = (s^2 - t^2) + (s^2 + t^2) = 2s^2$. Therefore, since p is a prime number other than 2, p must be a factor of s. In the same way p also is a factor of the difference $z - y = (s^2 + t^2) - (s^2 - t^2) = 2t^2$, and so p is a factor of t as well. But this is a contradiction since s and t are supposed to have been relatively prime integers, so no prime could be a factor of both. Thus the triple must have been primitive after all. This completes the proof of the theorem. ■

The area of the triangle given by a primitive Pythagorean triple $\{x, y, z\}$ is $\frac{1}{2}xy$, and so the area of this triangle must by an integer because, by Theorem 1.1, one of x or y is even. Thus, the area of any Pythagorean triangle is an integer.

Infinite Descent

We are ready to go back to 1659 to see how Fermat proved his proposition. Fermat gave the barest outline of a proof of his proposition in his letter to Huygens saying only "if the area of such a triangle were a square, then there would also be a smaller one with the same property, and so on, which is impossible." Nevertheless, this single brief statement does capture beautifully the essence of the "most singular method" of proof of which Fermat was so justifiably proud, and which he called the *infinite descent*.

Fermat's method of infinite descent is based on a very simple idea. In order to prove that no Pythagorean triangle can have square area,

you assume that there is one that does have square area and show how to produce a smaller Pythagorean triangle that also has square area. If you can do that, you would be done, because you could then repeat the same process on the smaller triangle and get a still smaller triangle with square area, and again to get a still smaller triangle, and so on, forever. Why is this impossible? Because of the positive integers 1, 2, 3, ... If A_1 is the area of the first triangle, and A_2 is the area of the second triangle, and so on, then we have produced an infinite sequence of strictly decreasing *positive* integers

$$A_1 > A_2 > A_3 > A_4 > \cdots ,$$

which is an impossibility within the natural numbers. It is that simple.

What Fermat left out of his letter because it "would make his discourse too long" was any discussion at all about how to take one Pythagorean triangle with square area and produce from it a smaller Pythagorean triangle that also has square area. In other words, he left out his proof! He did, however, at least write down his proof in the margin of a book—of a very famous book—his copy of Bachet's translation of Diophantus, a book we discuss at some length in Chapter 4. Here is that proof.

Theorem 1.2. *No Pythagorean triangle has square area.*

Proof

This proof will use Fermat's method of infinite descent. Our strategy, therefore, will be to assume there *is* a Pythagorean triangle with square area, and then produce a smaller Pythagorean triangle with square area. Using infinite descent, that is all we have to do in order to prove the theorem. It is also worth recalling from our discussion in the section on square numbers that any Pythagorean triangle with square area will be similar to a primitive Pythagorean triangle that has square area, so we can focus on primitive Pythagorean triangles in this proof.

We start the proof—using the notation of Theorem 1.1—by assuming that $\{2st,\ s^2-t^2,\ s^2+t^2\}$ is a primitive Pythagorean triple that represents a triangle whose area is square. The area of this triangle is $A = \frac{1}{2}xy$, that is, $A = st(s+t)(s-t)$. Since A is a square and s and t are relatively prime, all four terms in this expression for A are also relatively prime to one another; therefore, all four terms are themselves squares, and we can write $s = a^2$, $t = b^2$, $s + t = u^2$, $s - t = v^2$. Note that u and v must both be odd, and are relatively prime.

Now, let's look at the three squares v^2, a^2, and u^2. We can compute

$$a^2 - v^2 = s - (s - t) = t = b^2 \quad \text{and} \quad u^2 - a^2 = (s + t) - s = t = b^2,$$

and so we see, first of all, that $v^2 < a^2 < u^2$.

Moreover, we see that the difference between v^2 and a^2 is b^2, and that the difference between a^2 and u^2 is also b^2. In other words, a^2 is in the exact middle between v^2 and u^2. (For example, here are three squares $49 < 169 < 289$ where 169 is in the exact middle.) We conclude that $u^2 - v^2 = 2b^2$, and we factor this to get $2b^2 = (u + v)(u - v)$.

Next, we observe that since u and v are odd, $u+v$ and $u-v$ will each be even. But one of them will be exactly divisible by 4, and the other one won't. (This is because these two numbers differ by $2v$ and v is an odd number—that is, they differ by $2v = 2(2k + 1) = 4k + 2$.) So, whichever one of them is exactly divisible by 4, we write that one as $4n^2$, and we write the other one as $2m^2$. Then $u = \frac{1}{2}((u + v) + (u - v)) = m^2 + 2n^2$, and $v = \frac{1}{2}((u + v) - (u - v)) = \pm(m^2 - 2n^2)$—the plus or minus depends on which number was exactly divisible by 4, and $2b^2 = (2m^2)(4n^2)$, so $b = 2mn$.

But we now have a smaller Pythagorean triangle. Taking m^2 and $2n^2$ as the two "legs" for this smaller triangle we compute

$$(m^2)^2 + (2n^2)^2 = m^4 + 4n^4 = \frac{1}{2}\big((m^2 + 2n^2)^2 + (m^2 - 2n^2)^2\big)$$

$$= \frac{1}{2}\big(u^2 + (\pm v)^2\big) = \frac{1}{2}(u^2 + v^2) = a^2,$$

and so m^2 and $2n^2$ form the legs of a new Pythagorean triangle whose hypotenuse is a, whereas the hypotenuse of the original triangle was $s^2 + t^2 = a^4 + b^4$. So this new triangle is definitely smaller. Also, the area of this new triangle is m^2n^2, so it has square area, namely, $(mn)^2$.

Thus we have accomplished what we set out to do: we produced a smaller Pythagorean triangle with square area. Hence, by infinite descent, we are done. This completes the proof of the theorem. ■

Arithmetic Progressions

The situation that occurred in the proof of Theorem 1.2, where there were three squares $v^2 < a^2 < u^2$ with a^2 in the exact middle, is worth another look. The example we gave there was $49 < 169 < 289$ where the square 169 is in the exact middle. In this case the *common difference* between 49 and 169, and between 169 and 289, is 120.

This is a specific case of what we call an *arithmetic progression*. An arithmetic progression is just a sequence of numbers—and the sequence can be either finite or infinite—where any successive pair of numbers in the sequence has a constant difference, and we call this constant difference the *common difference*. So, the infinite sequence 7, 12, 17, 22, 27, . . . , is an arithmetic sequence where the common difference is 5. Or, the finite sequence 49, 169, 289, 409, 529 is an arithmetic progression where the common difference is 120.

There is an interesting connection between Pythagorean triangles and squares in an arithmetic progression. Suppose three squares, a^2, b^2, and c^2, are in an arithmetic progression. This means that $b^2 - a^2 = c^2 - b^2$. Let $x = \frac{c+a}{2}$ and $y = \frac{c-a}{2}$; then, since b^2 is the middle term in the progression, we get

$$b^2 = \frac{a^2 + c^2}{2} = \frac{(c+a)^2 + (c-a)^2}{4} = x^2 + y^2.$$

(In this chain of three equalities, the first equality holds because b^2, being the middle term in an arithmetic progression, is the average of the two terms on either side; the second equality can be verified easily by expanding $(c + a)^2 + (c - a)^2$; and the last equality follows immediately from the definitions of x and y.) Thus we see that we have a Pythagorean triangle $\{x, y, b\}$, where the hypotenuse b comes from the middle square.

Moreover, the common difference, d, of this arithmetic progression is given by

$$d = \frac{c^2 - a^2}{2} = \frac{(c+a)(c-a)}{2} = 2xy.$$

(Again, in this chain of three equalities, the first equality holds because in the arithmetic progression a^2, b^2, c^2 the two terms c^2 and a^2 differ by $2d$, that is, by *two* of the common difference d; the second equality is obvious; and the last equality once again follows immediately from the definitions of x and y.)

But the area of this triangle is $\frac{1}{2}xy$, so we also see the remarkable fact that the common difference of the original arithmetic progression is *four* times the area of the Pythagorean triangle. What is even more remarkable is that this connection between Pythagorean triangles and three squares in an arithmetic progression has been known for more than a thousand years.

Let's look at the example we just mentioned during the proof of Theorem 1.2: $49 < 169 < 289$. In this case, $a = 7$, $b = 13$, $c = 17$,

so we get $x = \frac{17+7}{2} = 12$ and $y = \frac{17-7}{2} = 5$. The corresponding triangle is $\{12, 5, 13\}$, which is a Pythagorean triangle, and the area of this triangle is $\frac{1}{2}(12)(5) = 30$. The common difference of the three squares, 120, is in fact four times this area, 30.

We'll do another example of this remarkable connection by looking at a very old problem concerning the area of Pythagorean triangles from an eleventh-century Byzantine manuscript found in a library in Istanbul (formerly Constantinople): *find a Pythagorean triangle of area* $5m^2$. Note that this is similar to the proposition Fermat addressed in Theorem 1.2, except that the area is not a square, but a multiple of a square. The writer of this problem—writing almost a thousand years ago—begins his solution of this particular problem very casually by saying "we must take for m^2 a multiple of 6"; he then lets $m = 6$, which is certainly an easy way to make sure that m^2 is a multiple of 6.

Did you know that the area of a Pythagorean triangle is always divisible by 6? Apparently this was common knowledge in Constantinople a thousand years ago. Nevertheless, we had better check this fact. We'll do this by showing that the area is divisible by 2, and also that it is divisible by 3; hence it is divisible by 6. (Note that this line of reasoning works only because 2 and 3 are relatively prime—just because a number is divisible by 10 and 15 doesn't mean it is divisible by 150; for example, 30 is divisible by both 10 and 15 but not by 150.)

So, let $\{x, y, z\}$ be a Pythagorean triangle as in Theorem 1.1 (we can restrict our attention to primitive Pythagorean triangles because if the area of a triangle represented by any primitive Pythagorean triple is divisible by 6, then the area of any Pythagorean triangle will also be divisible by 6). Then the area of this triangle is given by $A = \frac{1}{2}xy$. But recall that $x = 2st$ where one of s and t is even, which means that x is in fact divisible by 4. Therefore, the area A is divisible by 2, as desired.

Next we show that A is also divisible by 3. If x is divisible by 3, we'll be done, and if either s or t is divisible by 3, then this is obvious, so let's suppose that *neither* s nor t is divisible by 3. This means that when we divide either s or t by 3 we get a remainder of 1 or 2. If s and t happen to have the same remainder, then $s - t$ will be divisible by 3, whereas if s and t happen to have different remainders, then $s + t$ will be divisible by 3 (simply because $1 + 2 = 3$). So, either way, $y = s^2 - t^2 = (s - t)(s + t)$ will be divisible by 3. In other words, if x isn't divisible by 3, then y will be. So, in a Pythagorean triangle, not only is one of the two legs divisible by 4, but one of the two legs is divisible by 3. (As we said, this fact has been known for at least a thousand years.) Hence the area is divisible by 6.

The other thing the writer of this problem knew about was the connection between Pythagorean triangles and squares in arithmetic

progression. So, since he had decided to let $m = 6$, he was looking for a triangle whose area is $5 \cdot 6^2 = 180$, and therefore he knew all he had to do was find three squares in an arithmetic progression where the common difference is four times that area; that is, the common difference should be 720.

The rest was easy for him because if the common difference d should be 720, then—using our previous notation where $d = 2xy$—we see that $xy = 360$ would work. Then he chose $x = 9$ and $y = 40$ as factors for 360, and got $9^2 + 40^2 = 41^2$. That's how he did it. So his answer for this problem is the triangle $\{9, 40, 41\}$, which does have an area of the desired form since $\frac{1}{2} \cdot 9 \cdot 40 = 180 = 5 \cdot 6^2$.

While we are at it, let's check the arithmetic progression. The middle square should be 41^2, and the common difference is supposed to be 720. Are $41^2 - 720$ and $41^2 + 720$ both squares as they should be? Well, $41^2 - 720 = 961 = 31^2$ and $41^2 + 720 = 2401 = 49^2$, and so 31^2, 41^2, 49^2 form an arithmetic progression.

Fibonacci's Approach

Fibonacci—who was born in Pisa around 1180, but grew up in North Africa and traveled extensively—could also solve this problem about finding a Pythagorean triangle of area $5m^2$ because he knew what we now know, namely, that the common difference for the related arithmetic progression of three squares would be given by

$$d = 2xy = 4st(s^2 - t^2),$$

where we are again using the notation of Theorem 1.1.

Hence 8 is going to divide d (because s or t is even), 3 is going to divide d (because, as we just saw, 3 divides x or y), and 5 is also going to divide d (because the area, $\frac{1}{2}xy$, is supposed to be $5m^2$). Thus d is a multiple of 120. Then Fibonacci just picked two convenient small values of s and t to make this happen, namely, $s = 5$ and $t = 4$, which yields the same value $d = 720$ as before. So Fibonacci gets the exact same triangle since $x = 2st = 2(5)(4) = 40$ and $y = s^2 - t^2 = 5^2 - 4^2 = 9$.

But Fibonacci noticed something else interesting about problems such as these. First of all, a square n^2 can be written as the sum of the first n *odd* integers. For example, $4^2 = 1 + 3 + 5 + 7$.

Why this is true is visually obvious if you just think about 16 stones arranged in a square array (see Figure 1.3). Remove 1 stone from, say, the top right-hand corner, then remove the next 3 stones in an L-shaped pattern from the top right, then the next 5 again in an L-shaped pattern, and then the final 7 remaining stones in this same pattern.

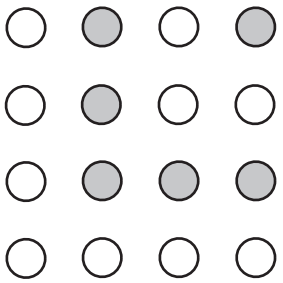


Figure 1.3 $1 + 3 + 5 + 7 = 16$.

Fibonacci used this basic fact in a clever way on the triangle problem. In this approach to the problem he concentrates on finding three squares in an arithmetic progression with common difference 720. He factors 720 into $8 \cdot 90$, and writes 720 as a sum of 8 consecutive odd integers whose average is 90 (that is, 90 is in the exact center of this sequence of odd numbers):

$$720 = 83 + 85 + 87 + 89 + 91 + 93 + 95 + 97.$$

Then he also factors 720 into $10 \cdot 72$, and writes 720 as a sum of 10 consecutive odd integers whose average is 72 (again, 72 is in the exact center):

$$720 = 63 + 65 + 67 + 69 + 71 + 73 + 75 + 77 + 79 + 81.$$

Note that, amazingly, these two sequences match up perfectly in that they could be combined into a single sequence beginning at 63 and ending at 97.

Now, it's just a matter of noticing that

$$1 + 3 + 5 + \cdots + 97 = 49^2,$$

$$1 + 3 + 5 + \cdots + 81 = 41^2,$$

$$1 + 3 + 5 + \cdots + 61 = 31^2.$$

Thus we know the three squares 31^2, 41^2, and 49^2 are in an arithmetic progression with common difference 720.

What made this idea work for Fibonacci is that the first sum has 8 consecutive odd integers centered on the number 90, and the second sum has 10 consecutive odd integers centered on the number 72, and furthermore, these consecutive sums fit together perfectly (since 83

is the next odd integer after 81). In general then, in order for this method of Fibonacci's to work, we need to be able to factor the common difference d in two ways, $d = \alpha\beta = \gamma\delta$ such that $\alpha + \gamma = \beta - \delta$. (By the way, we are using the four Greek letters $\alpha, \beta, \gamma, \delta$ here simply because they are four convenient letters to use; but any four available letters would be just as good, such as j, k, e, f.) In the problem above, for example, the fact that $8 + 10 = 90 - 72$ meant that there were 18 consecutive odd numbers with the first 10 centered on 72 and last 8 centered on 90.

Both Fibonacci and Fermat were well aware of the close connection between Pythagorean triangles and squares in arithmetic progression. In fact, another way to think of Theorem 1.2 is: *it is impossible to have three squares in an arithmetic progression whose common difference is a square*. Fibonacci had made this same assertion long before Fermat, but gave a completely inadequate argument to support his claim.

So we have begun our story of number theory with Fermat because modern number theory itself can be said to have begun with Fermat. Although numbers had engaged people in creative thought in many parts of the world for centuries and even millennia before the time of Fermat, he was the one who through his insight, curiosity, and vast correspondence set number theory on the path that we still follow today. We shall return to Fermat over and over again during the telling of our story in this book, but for now, in the next chapter, we again need to take a look much further back. Fermat did not invent number theory in a vacuum. The ultimate source of Fermat's ideas concerning numbers was the ancient Greeks, and these ideas came forward to him from them in a single book, *Arithmetica*, by Diophantus.

Problems

1.1 The Pythagorean triples $\{3, 4, 5\}$, $\{5, 12, 13\}$, and $\{7, 24, 25\}$ each represent right triangles in which the hypotenuse and one leg differ by only a single unit. Prove that there are infinitely many such Pythagorean triples by showing that for any odd number $2k + 1$, the triple $\{2k+1, 2k^2+2k, 2k^2+2k+1\}$ is a Pythagorean triple. Pythagoras knew of these triangles. Are they always primitive Pythagorean triples?

1.2 (H,S) Find all solutions in the positive integers to the equation

$$x^2 + y^2 = 1003.$$

1.3 (H,S) Find two primitive Pythagorean triples that represent triangles

having different hypotenuses but equal area. (Fermat proved that for any number n there are in fact n triangles with different hypotenuses and the same area.)

1.4 (H) Prove that for any integer $n \geq 3$ there is a Pythagorean triangle with one of its legs having length n.

For which integers n will there be a primitive Pythagorean triangle with n as the length of one of its legs?

1.5 (H,S) Prove that the radius of the inscribed circle of a Pythagorean triangle is always an integer.

1.6 (H,S) What is the longest possible hypotenuse a right triangle with integer sides can have if the radius of the inscribed circle is 12?

This problem appeared in the 2007 American Mathematics League Competition.

1.7 ⋆ (H,S) We say that a set of numbers is *pairwise relatively prime* if any two numbers in the set are relatively prime; in other words, for every pair of numbers from the set the only common factor of both numbers is 1. It is obvious that if a set of numbers is pairwise relatively prime, then the only common factor of all the numbers in the set is 1.

However, the converse of this statement is not true. Find a counterexample by finding a set of three numbers $\{a, b, c\}$ whose only common factor is 1, and yet no pair of these numbers is relatively prime.

Then determine whether a primitive Pythagorean triple is always pairwise relatively prime, and prove this one way or the other.

1.8 (H,S) Neugebauer called numbers that have no prime factors other than 2, 3, and 5 regular numbers. Regular numbers are therefore the numbers whose reciprocals can be expressed as finite sexagesimal fractions. For example, the reciprocal of 3 is an *infinite* decimal fraction ($\frac{1}{3} = 0.333\ldots$) but is *finite* as a sexagesimal fraction ($\frac{1}{3} = 0.20, 0, 0, \ldots$).

Express the reciprocal of 75 both as a decimal fraction and as a sexagesimal fraction. Then express the reciprocal of 7 both as a decimal fraction and as a sexagesimal fraction.

1.9 (S) Here are the values of s and t that correspond to the fifteen rows of Plimpton 322. (No values are given for row 11 because that particular row contains the triangle $\{45, 60, 75\}$ instead of the triangle $\{3, 4, 5\}$ for which $s = 2$ and $t = 1$).)

Row	1	2	3	4	5	6	7	8	9	10	12	13	14	15
s	12	64	75	125	9	20	54	32	25	81	48	15	50	9
t	5	27	32	54	4	9	25	15	12	40	25	8	27	5

Note that a few values seem to be missing such as $s = 5, t = 2$, which give triangle $\{21, 20, 29\}$. Perhaps this triangle was left out because $a = 21$ is not the shortest side—in other words, the angle of 43.60 degrees between a and the hypotenuse is less than 45 degrees. The values $s = 6, t = 5$ are also missing and would give triangle $\{11, 60, 61\}$. Maybe this triangle was left out because a is too short—that is, the angle 79.61 degrees is too big. The question remains: why does Plimpton 322 contain these fifteen triangles and no others?

One reasonable hypothesis to support Neugebauer's answer would be that Plimpton 322 is a list of all the triangles you would get for all regular s and t less than or equal to 125 (this number is chosen because $s = 125$ does occur behind the scenes in the fourth row) and assuming that the larger of the two acute angles is supposed to range from about 45 degrees in the triangle at the top of the list to almost 60 degrees at the bottom. Do you find this a plausible explanation for Plimpton 322? Support your answer.

1.10 ⋆ (S) Formulas such as $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ (which we "proved" geometrically in the text) can be proved algebraically using a method that is very much like Fermat's method of infinite descent except that it works in the other direction. The method is fundamentally simple: you prove the formula for a small value such as $n = 1$, and then you prove that whenever the formula is true for one value n it is also true for the next value $n + 1$. That's all there is to it. The formula is then true for all values of n; it is true for $n = 1$, so then it must be true for the next number $n = 2$, and for the next number $n = 3$, and the next number $n = 4$, and on, and on, forever.

This method could appropriately be called the method of *infinite ascent*, but it was given the name the method of *induction* by Augustus De Morgan in 1838, and first used by Blaise Pascal in 1654.

Use induction to prove that n^2 is the sum of the first n odd integers by assuming that

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2,$$

and then showing algebraically that the formula holds for $n+1$:

$$1+3+5+\cdots+(2n+1)=(n+1)^2.$$

Don't forget to show that the formula is true for $n=1$.

1.11 ⋆ (H,S) Let's look at Problem 1.10 again, and make sure we see exactly how the key inductive step is working. By the *inductive step* we mean the step in the proof where you show that if the formula is true for one value n, it is also true for the next value $n+1$. So, in this problem we are going to isolate the inductive step.

Assume that the formula in Problem 1.10 is true for $n=49$, that is, assume that

$$49^2=1+3+5+7+\cdots+93+95+97.$$

Then use this assumption to prove that the formula is also true for $n=50$, that is, prove that

$$50^2=1+3+5+7+\cdots+93+95+97+99.$$

1.12 ⋆ (H,S) (a) Use induction to prove the following formula for the sum of the first n squares:

$$1^2+2^2+3^2+\cdots+n^2 \;=\; \frac{n(n+1)(2n+1)}{6}.$$

We will give another proof of this formula in Chapter 2.

(b) It turns out that there is only one positive integer n (other than 1) such that the sum of the first n squares is itself a square. Use this formula to find that integer.

You can find a very nice geometric proof of this formula in "Counting Squares to Sum Squares" by Duane W. DeTemple, *The College Mathematics Journal* 41(2) (May 2010), 214–19. Here the idea is that the left side of this formula represents the total number of squares that can be found inside an $n\times n$ grid.

1.13 ⋆ (H,S) You should think about prime numbers the same way we think about atoms. They are the building blocks for the rest of the integers. And it is no accident that both the concept of prime number and the concept of the atom come down to us from the ancient Greeks. Modern theories of the atom can be traced back to the fifth-century B.C. Greek philosophers Democritus and Leucippus, who proposed that all matter is made up of very small *indivisible* particles.

In number theory this same notion can be traced back to the very same time and place. The Pythagoreans divided the natural numbers greater than 1 into two kinds: the indivisible numbers, called *prime* numbers, and the *composite* numbers, which are numbers that can be written as a product of two smaller numbers.

Prove that any integer greater than 1 can be written as a product of one or more primes. That is, prove that any integer greater than 1 has a *prime decomposition*.

1.14 ⋆ (H) In the proof of Theorem 1.1, we used the fact that if we divide a square by 4 the remainder will be either 0 or 1—that is, it will never be 2 or 3. And we express this fact by saying that a square n^2 must have the form $4k$ or $4k + 1$, and never the form $4k + 2$ or $4k + 3$.

Prove that if we divide a cube by 9, the remainder will be either 0, 1, or 8—that is, it will never be 2, 3, 4, 5, 6, or 7. In other words, show that a cube n^3 must have the form $9k$, $9k + 1$, or $9k + 8$.

1.15 ⋆ (S) In his letter of 1659 to Huygens, in which he reported having discovered his method of infinite descent, Fermat wrote: "At first I used it only to prove negative assertions such as: No number of the form $3n - 1$ can be written as $x^2 + 3y^2$." This is a somewhat puzzling statement because infinite descent is not needed at all to prove such an easy result. Prove this result by using the idea that any integer must have one of three forms: $3n$, $3n + 1$, or $3n + 2$. (Note that when Fermat talks about a number having the form $3n - 1$ that is equivalent to saying it has the form $3n + 2$.)

1.16 (H,S) The fact that a square must have the form $4n$ or $4n + 1$ immediately implies that no number of the form $4n + 3$ can be the sum of two squares. Still, it might be possible for such a number to be written as the sum of three squares. However, it is easy to check that the number 7 cannot be written as a sum of three squares, and that in fact 7 requires four squares: $7 = 4 + 1 + 1 + 1$. In 1638, Fermat wrote to Mersenne that no integer of the form $8n + 7$ can be written as the sum of three or fewer squares, and that this remains true even if you use rational squares. Mersenne passed this correspondence on to Descartes, who was quite disdainful that Fermat had announced such a trivial result.

Give a proof of this result for integers by first proving that any square n^2 must have the form $8k$, $8k + 1$, or $8k + 4$. Then prove that the result is also true for rational squares. A *rational* number is a number that can be written as a fraction $\frac{a}{b}$ where a and b are integers, and in this case we would call $(\frac{a}{b})^2$ a *rational square*.

1.17 (H,S) Show that if a, b, c are three positive integers such that

$$a^3 + b^3 = c^3,$$

then one of the three numbers must be divisible by 7.

1.18 ⋆ (H,S) Fermat would eventually prove that a cube can never be written as the sum of two cubes. This is one case of his famous conjecture known as *Fermat's last theorem*. However, it is possible for a cube to be the sum of three cubes. Find two different solutions in the positive integers to the equation

$$a^3 + b^3 + c^3 = d^3.$$

Such a solution $\{a, b, c, d\}$ is the cubic analog of a Pythagorean triple and is therefore called a *cubic quadruple*. One of these cubic quadruples should both surprise and delight you.

1.19 (H,S) Fibonacci was once challenged to find three squares in arithmetic progression with common difference 5, that is, to find three *rational* numbers a, b, and c such that $b^2 - a^2 = c^2 - b^2 = 5$. Solve this problem. A *rational* number is a number that can be written as a fraction $\frac{r}{m}$ where r and m are integers.

1.20 (S) Fibonacci picked the values $s = 5$ and $t = 4$ to find three squares in arithmetic progression with common difference 720. Try several other values of s and t such as $s = 5$ and $t = 2$, or even $s = 2$ and $t = 1$, to see what other arithmetic progressions with three squares that you come up with.

1.21 (H,S) Fibonacci used his method to find three squares in arithmetic progression with common differences other than 5. Use his method—as did Fibonacci himself—for the number 7 by taking $s = 16$ and $t = 9$. That is, find three squares in arithmetic progression with common difference 7. In particular, explicitly use two factorizations of the common difference d to find the three squares by expressing each occurrence of the common difference between the squares as a sum of consecutive odd numbers.

1.22 The Babylonians were not the only ones to do computations in base 60. The decimal system we use now only began to become known in Europe during the late Middle Ages. In particular, astronomers routinely used base 60 for their calculations. In 1483, a book of

astronomical data called the *Alphonsine tables* was published in Spain (named for Alfonso X of Castile, who commissioned the work).

The Alphonsine tables contain an amazingly accurate estimate of the length of a year, one that is within a few seconds of the current estimate. This estimate, written in base 60, was

$$365.14,\ 33,\ 9,\ 67$$

for the number of days in a year. Our current calendar is based on $365 + \frac{97}{400}$ as a fairly close estimate for the number of days in a year.

Show that the first two "digits" of the Alphonsine estimate—that is, $14,\ 33$—is exactly equal to the number $\frac{97}{400}$ that we use for our calendar today.

2

Euclid

Greek Mathematics

The source of modern number theory in terms of both its content and the manner in which we pursue it lies in the mathematics of the Greek-speaking people who lived throughout the eastern Mediterranean region in various independent city-states for more than a thousand years from roughly the sixth century B.C. onward.

One of the most well known of these ancient Greek mathematicians is the sixth-century B.C. philosopher Pythagoras, who was born on the island of Samos off the coast of present-day Turkey. Little is known of his life but he did travel widely and eventually settled in southern Italy with a group of followers we now call the "Pythagoreans." Pythagoras and his brotherhood are today perhaps best remembered for a philosophy that placed number at the very center of everything. Nicomachus, writing in about A.D. 100 in his *Introduction to Arithmetic*, captures this Pythagorean view of the universe:

> All that has by nature with systematic method been arranged in the universe seems both in part and as a whole to have been determined and ordered in accordance with number, by the forethought and the mind of him that created all things; for the pattern was fixed, like a preliminary sketch, by the domination of number preexistent in the mind of the world-creating God.

Pythagoras has been credited with discovering the simple relationship in music between harmony and number—that is, between the length of a string (or, as legend has it, the size of a blacksmith's anvil) and the pitch produced, so that a 2:1 ratio between the lengths of two strings will produce an octave difference in pitch, a 3:2 ratio will produce a "perfect fifth" (such as a G and a D, for example), and so on.

The Pythagoreans have also been credited with many discoveries in mathematics for which there is little evidence. The Pythagorean theorem is one famous example. Another is an often told story that it was the Pythagoreans who first discovered that $\sqrt{2}$ is an irrational number, which was extremely upsetting to them. A particularly dramatic version of this story, and one that is reminiscent of a famous scene in *The Godfather Part II*, has a group of irrate Pythagoreans row the unfortunate

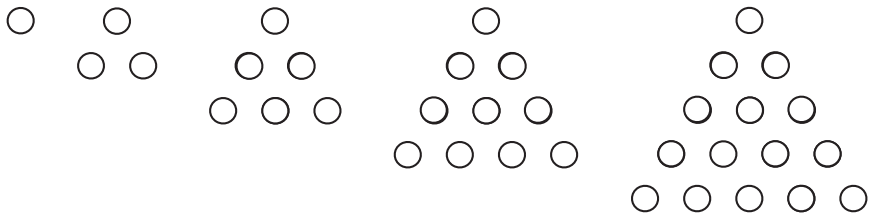


Figure 2.1 Triangular arrays.

fellow who made this discovery out to the middle of a lake and drown him in order to keep the discovery secret.

Triangular Numbers

The Pythagoreans did attach special properties to numbers. The number 2 represented man, 3 represented woman, and so 5 was marriage; 4 was justice, 1 was reason. Rather strangely, even numbers were considered feminine and odd numbers masculine.

More important for us, the Pythagoreans saw geometric properties in numbers. So, in addition to square numbers—after all, they were aware of the formula $n^2 = 1 + 3 + 5 + \cdots + (2n - 1)$ mentioned in the last chapter—that corresponded to square arrays of stones placed on the ground, they also studied *triangular numbers*, that is, numbers that represent numbers of stones that can be placed in triangular arrays on the ground. The triangular numbers, then, are 1, 3, 6, 10, 15, . . . , as shown in the *triangular arrays* in Figure 2.1.

The nth *triangular number* t_n is by definition the sum of the first n natural numbers, that is,

$$t_n = 1 + 2 + 3 + \cdots + n.$$

Furthermore, since we observe in Figure 2.1 that you can get from one triangular number to the next by adding a single row of stones, it is obvious that

$$t_n = t_{n-1} + n.$$

For example, $t_5 = t_4 + 5$; that is, $15 = 10 + 5$.

Another simple and extremely important formula for triangular numbers was known to the Pythagoreans. You can take two copies of the triangular array for a given triangular number t_n and place them together to form an $n \times (n + 1)$ rectangle. Let's do this for the triangular

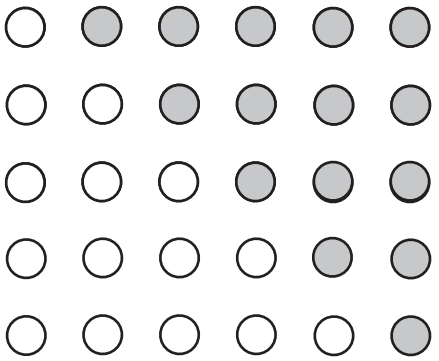


Figure 2.2 $t_5 = \frac{5(6)}{2}$.

number $t_5 = 15$ from Figure 2.1, and we see that we get the 5×6 rectangle shown in Figure 2.2, which of course has 30 stones. In general, the two triangles will form a single $n \times (n+1)$ rectangle with $n(n+1)$ stones. Therefore,

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

The great early nineteenth-century mathematician Carl Friedrich Gauss will be mentioned frequently in this book, and is even the main topic in Chapter 8, but one story about Gauss is worth telling now because it has to do with triangular numbers. Here is how the story goes. When Gauss was a young boy, about eight or so, his teacher one day gave his class the following problem in order to keep them occupied for awhile: add all the numbers from 1 to 100. Much to the teacher's surprise, Gauss did this in just a moment or two, presumably by noticing that $1 + 100 = 101$, $2 + 99 = 101$, $3 + 98 = 101$, and so on, and in this way he got fifty identical sums, each being 101, so the total sum is $50(101) = 5050$; that is, $1 + 2 + \cdots + 100 = \frac{100(101)}{2}$. In other words, as a boy, Gauss discovered the formula we gave above for triangular numbers! To be honest, there does seem to be some doubt as to whether this event actually took place. Nonetheless, this story of Gauss's early childhood genius, much like the famous but apocryphal story of the apple that fell on Isaac Newton's head, is now firmly rooted in mathematical folklore. By the way, in Problem 2.1 you will be asked to give a third proof of this extremely important formula for triangular numbers.

Another very nice fact about triangular numbers that was undoubtedly known to the Pythagoreans, but is usually attributed to Nicomachus, is that if you add any two *consecutive* triangular numbers, you get a square number. So, for example, $6 + 10 = 16$, and $15 + 21 = 36$.

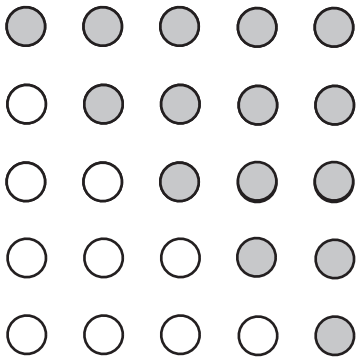


Figure 2.3 $t_4 + t_5 = 5^2$.

In other words,

$$t_{n-1} + t_n = n^2.$$

This fact can be made visually obvious simply by placing two triangular arrays together to form a square, as we see illustrated in Figure 2.3. You will be asked for an alternative proof in Problem 2.2.

A much less obvious fact about triangular numbers that was also known to the Pythagoreans, though attributed to Plutarch in about A.D. 100, is that *eight* times a triangular number plus one always yields a square. So, for example, $8 \cdot 6 + 1 = 49$, and $8 \cdot 15 + 1 = 121 = 11^2$. This fact is easy to prove algebraically since we can write

$$8 \cdot t_n + 1 = \frac{8n(n+1)}{2} + 1 = 4n^2 + 4n + 1 = (2n+1)^2.$$

You will be asked to provide a visual—and, we hope, quite beautiful—geometric proof of this fact in Problem 2.4 using ideas similar to those in Figures 2.2 and 2.3.

Tetrahedral and Pyramidal Numbers

The geometrical properties of numbers studied in ancient times were not limited to two dimensions. Even in modern times, as you travel around the world you see fruit and produce stacked in geometric patterns in markets or by the roadside. For centuries cannon balls have been stacked in similar geometric patterns, and today, golf balls are often set out in exactly the same way on driving ranges.

As we see in Figure 2.4, humans seem to have decided that there are two natural ways to stack things: either they start with a *triangular* base, or they start with a *square* base. Since we have been talking about

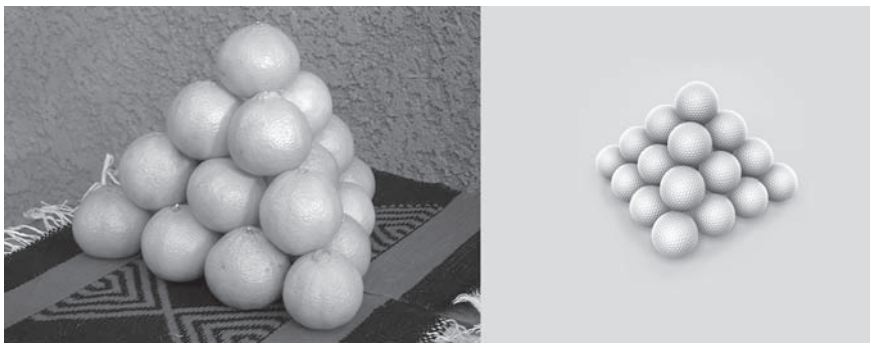


Figure 2.4 (a) A tetrahedron of oranges. (b) A pyramid of golf balls. Figure 2.4(b) is courtesy of Sgame/Shutterstock.

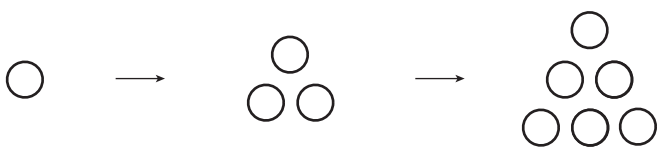


Figure 2.5 $T_3 = 1 + 3 + 6 = 10$.

triangular numbers, we will focus for the moment on the triangular base option. The oranges in Figure 2.4a form a tetrahedron (that is, a solid figure with four equal equilateral triangular faces). Note how perfectly the top orange nestles into the single space formed by the triangle of three oranges in the second layer of the tetrahedron, and how these three oranges similarly fit into the three spaces formed by the triangle of six oranges in the third layer, and so on. Figure 2.5 illustrates the idea that the number of oranges in each layer of such a tetrahedron is represented by a triangular number.

Since we could make larger and larger stacks of oranges by placing ever larger triangular layers at the bottom, we will define the nth *tetrahedral* number T_n to be

$$T_n = t_1 + t_2 + \cdots + t_n,$$

that is, the nth tetrahedral number is the sum of the first n triangular numbers (the layers of the tetrahedron).

For example, $T_3 = t_1 + t_2 + t_3 = 1 + 3 + 6$, as illustrated in Figure 2.5. And T_4 would be $1 + 3 + 6 + 10 = 20$ (so there are 20 oranges in the picture). The tetrahedral numbers form the sequence

$$1, 4, 10, 20, 35, 56, \ldots.$$

Note that each number in this sequence increases by the next triangular number, that is,

$$T_n = T_{n-1} + t_n.$$

A striking formula that was known in Egypt about 300 B.C., and was also found in India by Aryabhata around A.D. 500, involves the *sum* of the first n triangular numbers:

$$t_1 + t_2 + \cdots + t_n = \frac{n(n+1)(n+2)}{6}.$$

In other words, this is a formula for the nth tetrahedral number. For example, $T_6 = \frac{6(7)(8)}{6} = 56$. In Problem 2.6 you will be asked to prove this formula using induction.

There is also a very clever geometric proof of this formula, which we will describe using the tetrahedral number 20 as an example to illustrate the general idea. You might want to use actual objects such as oranges—balls of clay would be ideal—to help visualize what is going on in three dimensions.

We begin with the tetrahedral number 20, and represent it as the following triangle of numbers:

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 1 & & 2 & & \\ & 1 & & 2 & & 3 & \\ 1 & & 2 & & 3 & & 4 \end{array}$$

where the sum of each row is a triangular number; thus, $20 = 1 + 3 + 6 + 10$.

Now if we take three copies of this triangle of numbers (each triangle still representing the tetrahedral number $T_4 = 20$), and "add" these three triangles,

$$\begin{array}{ccccccc} & & 1 & & & & \\ & 1 & & 2 & & & \\ 1 & & 2 & & 3 & & \\ 1 & & 2 & & 3 & & 4 \end{array} \quad \begin{array}{ccccccc} & & 1 & & & & \\ & 2 & & 1 & & & \\ 3 & & 2 & & 1 & & \\ 4 & & 3 & & 2 & & 1 \end{array} \quad \begin{array}{ccccccc} & & 4 & & & & \\ & 3 & & 3 & & & \\ 2 & & 2 & & 2 & & \\ 1 & & 1 & & 1 & & 1 \end{array}$$

then we get the following triangle of numbers

$$\begin{array}{ccccccc} & & & 6 & & & \\ & & 6 & & 6 & & \\ & 6 & & 6 & & 6 & \\ 6 & & 6 & & 6 & & 6 \end{array}$$

and we conclude that $3 \cdot T_4 = 6 \cdot t_4$, that is $3 \cdot 20 = 6 \cdot 10$.

If we do this in general for the nth tetrahedral number we will get

$$3T_n = (n+2)t_n.$$

Solving for T_n gives us Aryabhata's formula for the nth tetrahedral number

$$T_n = \frac{n+2}{3}t_n = \frac{n+2}{3} \cdot \frac{n(n+1)}{2} = \frac{n(n+1)(n+2)}{6}.$$

Turning now to stacks formed by starting with a square base, as in the stack of 30 golf balls in Figure 2.4b, we define the nth *pyramidal* number P_n to be

$$P_n = 1^2 + 2^2 + \cdots + n^2,$$

that is, the nth *pyramidal* number is the sum of the first n square numbers (the layers of the pyramid).

For example, $P_4 = 1 + 4 + 9 + 16 = 30$, as illustrated by the stack of golf balls. The pyramidal numbers form the sequence given by

$$1, 5, 14, 30, 55, 91, \ldots .$$

Note that each number in this sequence increases by the next square number, that is,

$$P_n = P_{n-1} + n^2.$$

Since each layer of a pyramid is represented by a square k^2, we can use the formula $k^2 = t_{k-1} + t_k$ to find a formula for P_n,

as follows:

$$\begin{aligned} P_n &= 1^2 + 2^2 + 3^2 + 4^2 + \cdots + n^2 \\ &= t_1 + (t_1 + t_2) + (t_2 + t_3) + (t_3 + t_4) + \cdots + (t_{n-1} + t_n) \\ &= (t_1 + t_2 + t_3 + t_4 + \cdots + t_n) + (t_1 + t_2 + t_3 + \cdots + t_{n-1}) \\ &= T_n + T_{n-1} = \frac{n(n+1)(n+2)}{6} + \frac{(n-1)n(n+1)}{6} \\ &= \frac{n(n+1)(2n+1)}{6}. \end{aligned}$$

You proved this same formula by induction in Problem 1.12.

Let us now turn our attention from the specific content of the ancient Greek study of numbers to the manner in which they pursued mathematical truth.

The Axiomatic Method

During roughly the period from the sixth to the fourth century B.C., there developed within Greek mathematics the notion of proving things. This Greek notion of *proof* is the foundation upon which modern mathematics rests. The very way in which we go about doing mathematics today is something we inherited from the Greeks: the use of deductive reasoning to prove an ever-growing body of new facts from previously known facts, and the realization that you have to start with a set of simple "facts" called *axioms* that are taken as self-evidently true.

When Thomas Jefferson wrote the *Declaration of Independence of the Thirteen Colonies* in June of 1776 he adopted this very same *axiomatic method*:

> We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the pursuit of Happiness.

Thus, in this way and in this document Jefferson establishes individual liberty as the bedrock upon which his argument calling for colonies to break with England stands. Jefferson's *Declaration of Independence* owes its form to the ancient mathematics of the Greeks.

Euclid is justifiably one of the most famous Greek mathematicians of all—only Archimedes and Apollonius can compare. Little is known of Euclid's life. He lived most of it in Syracuse, in Sicily, but may have taught in Alexandria—the great center of learning on the north coast

of Egypt—around 300 B.C. However, Euclid wrote a text consisting of thirteen books that contained definitions, axioms, and theorems and brought together most of the mathematics that was known at that time. Euclid's *Elements* would go on to be not only the most wildly successful mathematics book ever written, but it is rivaled only by the Bible in terms of overall circulation and widespread influence. The *Elements* was one of the first books to be printed after the invention of the printing press. Since its first printing in Venice in 1482 more than a thousand editions of the *Elements* have been published. The first English edition appeared in 1570, and unfortunately its title page repeats a common error by confusing Euclid with an earlier philosopher, Eucleides of Megara (thus, in Latin, Euclid was often referred to as Euclidis Megarensis).

The very title of this great work tells us that it contains the "elements" from which all further mathematical truth can be produced. Euclid begins with definitions and self-evident truths (axioms) and then uses deductive reasoning to produce an ever-growing body of new results (theorems and propositions) that are also true—all results flow logically from the initial elementary starting point. This is the form that Thomas Jefferson adopted in his *Declaration of Independence*.

Most of the truths in the *Elements* deal with geometry—for example, Proposition 47 in Book I is the "Pythagorean Theorem"—but three of the books (VII, VIII, and IX) deal with number theory. For instance, Proposition 2 in Book IX is the statement that there are infinitely many prime numbers.

> **Proposition 20**. *Prime numbers are more than any assigned multitude of prime numbers.*

Euclid's proof of this fact is frequently cited as one of the most beautiful in all of mathematics. There is, however, no real evidence one way or the other that this proof can be attributed to Euclid (not that this detracts in the least from its beauty). The great early twentieth-century English mathematician G. H. Hardy called this proposition of Euclid's and the proof that $\sqrt{2}$ is irrational "theorems of the highest class," and wrote that "each is as fresh and significant as when it was discovered—two thousand years have not written a wrinkle on either of them."

Here is Euclid's proof. Euclid thought of numbers as representing line segments of various lengths. An explanatory remark has been added to his proof in brackets.

> Let A, B, C be the assigned prime numbers; I say there are more prime numbers than A, B, C. For let the least number measured by A, B, C be taken, and let it be DE; let the unit DF be added to

Figure 2.6 Title page of first English edition of Euclid's *Elements*, 1570.

DE. [When Euclid says "measured" we would today say "divisible," and so the number *DE* is the product of the three primes *A*, *B*, and *C* and then Euclid adds 1—that is, the length of *DF*—to *DE* to get the number *EF*.]

Then *EF* is either prime or not. First, let it be prime; then the prime numbers *A*, *B*, *C*, *EF* have been found which are more than *A*, *B*, *C*.

Next, let EF not be prime; therefore it is measured by some prime number. Let it be measured by the prime number G.

G ——————

I say that G is not the same with any of the numbers A, B, C. For, if possible, let it be so. Now A, B, C measure DE; therefore G will also measure DE. But it also measures EF. Therefore G, being a number, will measure the remainder, the unit DF, which is absurd. Therefore, G is not the same with any one of the numbers A, B, C. And by hypothesis it is prime. Therefore the prime numbers A, B, C, G have been found that are more than the assigned multitude of A, B, C. Q.E.D.

Paul Erdős, the great twentieth-century mathematician whom we mentioned in the first chapter, doubted the existence of God but believed deeply in something he called the Book. For him the Book was something that did exist and was even almost spiritual, for it "contains the best proofs of all mathematical theorems, proofs that are elegant and perfect." It is clear that Euclid's proof of the infinitude of the primes comes straight from the Book.

These days, we dress up Euclid's proof in contemporary clothing, but we can't pretend to improve it.

Theorem 2.1. *There are infinitely many prime numbers.*

Proof

Suppose this is not the case, and that $p_1, p_2, \ldots, p_n$ are all the primes. Then let

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

Now N cannot be prime because it is larger than all of the primes $p_1, p_2, \ldots, p_n$. So N is divisible by a prime p. However, none of the primes $p_1, p_2, \ldots, p_n$ can divide N since it would then also divide 1 (this is because $1 = N - p_1 p_2 p_3 \cdots p_n$). Hence we have a contradiction because p is a prime other than $p_1, p_2, \ldots, p_n$, which from the start were assumed to be all the primes. This completes the proof. ■

One change we made from Euclid in the proof of Theorem 2.1 is that instead of beginning with just three prime numbers A, B, C, we began in a style that is considered to be more "general" with an unknown number of primes $p_1, p_2, \ldots, p_n$. That's just the way we write proofs these days even though in this case the gain is minimal. It is clear in

Euclid's proof that the set of primes A, B, C represent "any assigned multitude of prime numbers" since their "assigned multitude" being 3 has nothing at all to do with anything that happens in the proof. There is even a disadvantage to the modern general style in that we are forced to use subscripts to list the prime numbers, and this has the effect of cluttering things up somewhat. Nonetheless, it is important to learn this modern style of proof because it has the huge advantage of greatly clarifying situations that are far more complicated than this particular proof is.

Proof by Contradiction

Another change we made in Euclid's proof is that we explicitly used "proof by contradiction." Euclid's proof has a different structure because he is proving a different statement—namely, that given any finite set of prime numbers, there is another prime number not in that set. Hence the set of *all* prime numbers cannot be finite. Euclid, however, does use contradiction—also known as *reductio ad absurdum*—within his proof to show that the prime G is not one of the numbers A, B, C.

It is worth pausing to see how proofs by contradiction work. The method of contradiction—which G. H. Hardy called "one of a mathematician's finest weapons"—can be traced back to the Eleatics, that is, to Parmenides of Elea and his followers. The Eleatics, contemporaries of the Pythagoreans in southern Italy in the fifth century B.C., emphasized pure reason and logic in their philosophy. Consider the way in which we start our proof of Theorem 2.1. What we say in effect is: *either* the number of primes is finite *or* the number of primes is *not* finite. Similarly, in his proof Euclid says in effect: *either* G is one of the numbers A, B, C, or *not*. The way contradiction works is to begin with a statement of pure reason, such as these, that clearly lays out two alternatives (only one of which is going to be true, the other therefore being false); next, you simply eliminate one of the two alternatives. What you are left with then must be true. What we did in our proof was eliminate the alternative that the number of primes is finite. What Euclid did in his proof was eliminate the alternative that G is one of the numbers A, B, C. The method is called contradiction, or *reductio ad absurdum*, because the way you eliminate an alternative is to *assume* that alternative, and then reach a contradiction—that is, a point of absurdity.

Sherlock Holmes was a firm advocate of the power of deductive reasoning—"that true cold reason which I place above all things"—and he frequently lectured his companion Watson on the virtues of an indirect proof. Here he is in 1890 speaking to Watson in the second

Sherlock Holmes novel, *The Sign of the Four* :

> How often have I said to you that when you have eliminated the impossible, whatever remains, *however improbable*, must be the truth.

You will be asked to heed this good advice of Sherlock Holmes in Problem 2.17 in order to prove that $\sqrt{2}$ is irrational (whether or not that sounds improbable to you).

A good illustration of the use of contradiction occurred in our proof of Theorem 1.1 in Chapter 1. Look back at that proof. In order to prove that one of the numbers x or y in a primitive Pythagorean triple $\{x, y, z\}$ must be even and the other odd, we eliminated the other two possibilities—namely, the possibility that both x and y might be even, *and* the possibility that both might be odd. Sir Arthur Conan Doyle and G. H. Hardy have it exactly right: contradiction is indeed "one of a mathematician's finest weapons."

Euclid, too, was a great fan of contradiction. He used it often in the *Elements*, and even used it as early as Proposition 6 in Book I to prove that if two angles in a triangle are equal, then the triangle is isosceles. He begins his proof by declaring that he will show that the two sides AB and AC subtending the two equal angles are themselves equal, and then goes straight into the contradiction mode by saying "for, if AB is unequal to AC, one of them is greater; let AB be greater."

Euclid's Self-Evident Truths

Euclid begins his study of numbers in Book VII with definitions: a "unit" is "that by virtue of which each of the things that exist is called one," a "number" then is "a multitude composed of units." There are the expected definitions: an "even number" is "that which is divisible into two equal parts", an "odd number" is "that which differs by a unit from an even number", and a "prime number" is "that which is measured by a unit alone"; but there are also a few unexpected definitions: an "even-times even number" is "that which is measured by an even number according to an even number," an "even-times odd number" is "that which is measured by an even number according to an odd number," and an "odd-times odd number" is "that which is measured by an odd number according to an odd number."

Then, as he had done in his books on geometry, and with these definitions as his starting point, Euclid goes on to arrange all of his propositions on number theory in Books VII–IX in a logical order so that each proposition can in turn be deduced from previous results. For

Figure 2.7 Euclid, shown in a detail of Raphael's *The School of Athens*, 1509/10.

example, in Book IX in order to prove Proposition 22, which says that the sum of an even number of odd numbers will always be even, Euclid uses Proposition 21, which guarantees that a sum of even numbers is always even. This is what we mean by the term "the axiomatic method" and Euclid's application of it in the *Elements* is the first, and still best, example of this method.

In the next chapter we will look at one of the most fundamental notions in number theory—divisibility—from precisely this axiomatic Euclidean point of view. You may find it somewhat surprising there that suddenly we shall be taking such painstaking care over what seem to be rather obvious ideas about the simple and familiar notion of division. But to forewarn you about some of the subtleties involved in such matters we will discuss a theorem that Euclid did *not* include in the *Elements*, even though it surely was known at the time. This is the important theorem that says every integer greater than 1 has a *prime decomposition* that you were asked to prove in Problem 1.13. For example, $60 = 2 \cdot 2 \cdot 3 \cdot 5$, and $210 = 2 \cdot 3 \cdot 5 \cdot 7$; and it should be clear to you that you could factor the number 123 456 789 into a product of primes if for some bizarre reason you decide to do so.

Theorem 2.2. *Every integer greater than 1 is either prime or can be written as a product of primes.*

Proof

This theorem might seem all but self-evident to you. In fact, before we give a rigorous proof of the theorem, let's try to see why the theorem is true. If n is an integer greater than 1, but not itself prime, then it is composite, and we can write $n = rs$ where both r and s are smaller than n. The same argument can now be repeated on each of the two factors r and s and, in turn, repeated again as many times as needed on any subsequent factors that arise. Since at each stage the factors always get smaller, this process eventually has to stop, and, when it does, each factor is a prime number, and n has in this way been written as a product of primes. You should recognize the basic idea here, because this has really been just an infinite descent argument.

This argument, while in some ways getting to the heart of the matter, also leaves much to be desired since it boils down to saying that you can't just keep factoring forever, which sounds way too simple. We can make this argument more rigorous in a couple of ways. Here is one way. Turn it into a proof by contradiction: suppose that there is a composite number that *cannot* be written as a product of primes. Then, there is a smallest such composite number, call this number n. But n is composite, so we can write $n = rs$ where both r and s are smaller than n but still greater than 1. Therefore, by the choice of n, both r and s *can* be written as a product of primes, hence so can n, which is a contradiction.

Another way to make this proof rigorous is to use a slightly different form of induction, sometimes called *strong induction*. The variation from the standard method of induction is a minor one: you prove the statement for a small value such as $n = 1$, and then you prove that whenever the statement is true for all values up to and including n, it is also true for the next value $n + 1$.

In this case, we first observe that the theorem is true for the smallest value, namely, $n = 2$, since 2 is prime. Now, assume that the theorem is true for all values up to and including some integer n where $n \geq 2$. We must prove that the theorem is true for the integer $n+1$. If $n+1$ is prime, then we are done. If, on the other hand, $n+1$ is composite, then we can write $n + 1 = rs$ where $1 < r, s < n + 1$. But then, $1 < r, s \leq n$ and so, by our assumption, r and s can each be written as a product of primes, hence so can $n + 1$. This completes the proof. ■

When reading the proof of this theorem, you might have paused momentarily at the point where we say that if there is a composite number that cannot be written as a product of primes, then there

must be a *smallest* such composite number. But the idea being used there is a fundamental one. All we are really saying is that any set of positive integers—in this case, it is the set of composite numbers that cannot be written as a product of primes—has to have a smallest element. This fact *is* self-evident, and even has a name: the well-ordering principle.

The *well-ordering principle* is the statement that every nonempty set of natural numbers contains a least element. This fundamental property of the natural numbers is frequently taken as an axiom, and then used to establish other properties of natural numbers and integers.

Unique Factorization

What is not at all obvious—and in fact requires careful proof, which we will be able to do in the next chapter—is that for a given number such as 107 207 100 there is only *one* way to write this number as a product of primes, namely, $107\,207\,100 = 2^2 \cdot 3^2 \cdot 5^2 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17$.

You may think this is obvious, but your only argument in support of this position at this point would be something along the lines of "well, how else could you possibly factor 107 207 100?" This is not an argument that would have impressed Euclid. In fact, in order to convince you how weak this argument is, we will now look at a number system where this argument fails completely. So, temporarily, we are going to step outside of the natural numbers and consider complex numbers of the form $a + b\sqrt{-6}$ where a and b are integers.

The first thing to say about this larger number system is that it is *closed* under addition and multiplication, which just means that if we *add* two numbers in the system, their sum will again be a number in the system

$$(2 + 3\sqrt{-6}) + (1 - 2\sqrt{-6}) = 3 + \sqrt{-6},$$

and if we *multiply* two numbers in the system, their product will again be a number in the system

$$(2 + 3\sqrt{-6}) \cdot (1 - 2\sqrt{-6}) = 2 - 4\sqrt{-6} + 3\sqrt{-6} - 6(-6) = 38 - \sqrt{-6}.$$

The next thing we need to do in this number system is have a way to decide whether a number $a + b\sqrt{-6}$ is prime. For example, we see above that the number $38 - \sqrt{-6}$ is composite because it factors into a product of two smaller numbers. But what about $10 + 3\sqrt{-6}$? Is it prime or composite? Well, we could try to factor it; after all that's what we do in the integers, but that is much more cumbersome here, and

besides, there are many more potential factors that we would have to try. A much better idea is to reduce this question to a different, and easier, question. Here is how we do this.

We will define a function on this number system specifically for this purpose, called the *norm* function. This norm function will map numbers from this number system to the integers, and will have the useful property that it reduces questions of whether numbers such as $10+3\sqrt{-6}$ are prime to vastly simpler questions of whether their images are prime in the integers.

So, for a number $a+b\sqrt{-6}$ in our number system, we define the *norm* of the number to be

$$N(a+b\sqrt{-6}) = a^2 + 6b^2.$$

Note that, as advertised, the norm of a number is an integer.

Now, the key property of the norm function is that it respects multiplication between these two number systems, that is, if $\alpha = a + b\sqrt{-6}$ and $\beta = c + d\sqrt{-6}$ are two numbers in our number system, then

$$N(\alpha\beta) = N(\alpha)N(\beta).$$

You will be asked to prove this fundamental property of the norm function in Problem 2.19.

Let's see how the norm can be used to show that $2 + \sqrt{-6}$ is prime. Suppose that $2 + \sqrt{-6} = \alpha\beta$, where $\alpha = a + b\sqrt{-6}$ and $\beta = c + d\sqrt{-6}$. Then, since $N(2 + \sqrt{-6}) = 2^2 + 6 \cdot 1^2 = 10$, when we apply the norm to both sides of $2 + \sqrt{-6} = \alpha\beta$, we get

$$10 = N(\alpha\beta) = N(\alpha)N(\beta),$$

where $N(\alpha) = a^2 + 6b^2$ and $N(\beta) = c^2 + 6d^2$. But now recall that this is all taking place inside the integers, and we know that the only factors of 10 are 1, 2, 5, and 10. At this point it is easy to see that neither $a^2 + 6b^2$ nor $c^2 + 6d^2$ can ever equal either 2 or 5, so the only option is that one of them equals 1, and the other equals 10. Thus we conclude that either $a = \pm 1$ or $c = \pm 1$; that is, either $\alpha = \pm 1$ or $\beta = \pm 1$. In other words, $2 + \sqrt{-6}$ is prime.

Not surprisingly we could use this same procedure with the norm to show that the number $2 - \sqrt{-6}$ is also prime in this number system. We can even use this procedure to show that the numbers 2 and 5 are prime in this system. (Note that numbers such as 2 and 5, which are prime as integers, still need to be shown to be prime in this new system

because there are simply more numbers present as potential divisors in this number system. Perhaps we should also point out the important detail that the numbers 2 and 5 are actually *in* this new number system since we can write $2 = 2 + 0 \cdot \sqrt{-6}$ and $5 = 5 + 0 \cdot \sqrt{-6}$.)

Now, we are finally prepared to produce two *different* prime factorizations for a single number:

$$10 = 2 \cdot 5 \qquad \text{and} \qquad 10 = (2 + \sqrt{-6}) \cdot (2 - \sqrt{-6}).$$

Note that the second factorization is a direct response to a question you never would have asked yourself because the answer would have seemed completely obvious (although wrong, as it turns out): "how else could you possibly factor 10?"

The moral of the story is that some number systems have unique factorization, and some don't. For those number systems that do, such as the integers, therefore, we have to prove that they possess this very desirable property, and that takes some very careful preparation, which we will begin in the next chapter.

But let's get back to $10 + 3\sqrt{-6}$ and decide whether it is prime. Suppose that $10 + 3\sqrt{-6} = \alpha\beta$, where $\alpha = a + b\sqrt{-6}$ and $\beta = c + d\sqrt{-6}$. Since $N(10 + 3\sqrt{-6}) = 10^2 + 6 \cdot 3^2 = 154$, and $N(\alpha) = a^2 + 6b^2$ and $N(\beta) = c^2 + 6d^2$, we need to consider the factorization of $154 = 2 \cdot 7 \cdot 11$. First, it is clear that neither $a^2 + 6b^2$ nor $c^2 + 6d^2$ can equal 2 or 11, so we might as well set $N(\alpha) = a^2 + 6b^2 = 7$ and $N(\beta) = c^2 + 6d^2 = 22$.

So we easily get $a = \pm 1$, $b = \pm 1$ as the only solutions for the first equation. Note that there are four solutions here, hence four possible values for α. The second equation is a little trickier, but a moment's reflection yields the only solutions: $c = \pm 4$, $d = \pm 1$, and again this means four possible values for β. And we quickly discover that in fact, by choosing $a = 1$, $b = 1$ and $c = 4$, $d = -1$, we get

$$(1 + \sqrt{-6}) \cdot (4 - \sqrt{-6}) = 10 + 3\sqrt{-6},$$

and so $10 + 3\sqrt{-6}$ is not prime.

Note that something quite interesting has emerged here. The question of whether $10 + 3\sqrt{-6}$ is prime in this number system seems to be connected to the kinds of questions involving sums of squares we dealt with in the first chapter, since it boiled down to deciding whether there were integer solutions to the equations $a^2 + 6b^2 = 7$ and $c^2 + 6d^2 = 22$. This turns out to be a familiar pattern in number theory, and in mathematics more generally, in that ideas that may at first appear unrelated often end up being connected in profound ways.