SUMMING IT UP

From One Plus One to Modern Number Theory

AVNER ASH AND
ROBERT GROSS

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Contents

Preface

Adding two whole numbers together is one of the first things we learn in mathematics. Addition is a rather simple thing to do, but it almost immediately raises all kinds of curious questions in the mind of an inquisitive person who has an inclination for numbers. Some of these questions are listed in the Introduction on page 3. The first purpose of this book is to explore in a leisurely way these and related questions and the theorems they give rise to. You may read a more detailed discussion of our subject matter in the Introduction.

It is in the nature of mathematics to be precise; a lack of precision can lead to confusion. For example, in Bloom's musings from *Ulysses*, quoted as one of our epigraphs, a failure to use exact language and provide full context makes it appear that his arithmetical claims are nonsense. They may be explained, however, as follows: "Two multiplied by two divided by half is twice one" means $\frac{2+2}{2} = 2 \cdot 1$, because Bloom meant "cut in half" when he used the ambiguous phrase "divided by half." Then "$1 + 2 + 6 = 7$" refers to musical intervals: If you add a unison, a second, and a sixth, the result is indeed a seventh. Bloom himself notes that this can appear to be "juggling" if you suppress a clear indication of what you are doing.

Bloom has the luxury of talking to himself. In this book, we strive to be clear and precise without being overly pedantic. The reader will decide to what extent we have succeeded. In addition to clarity and precision, rigorously logical proofs are characteristic of mathematics. All of the mathematical assertions in this book can

be proved, but the proofs often are too intricate for us to discuss in any detail. In a textbook or research monograph, all such proofs would be given, or reference made to places where they could be found. In a book such as this, the reader must trust us that all of our mathematical assertions have proofs that have been verified.

This book is the third in a series of books about number theory written for a general mathematically literate audience. (We address later exactly what we mean by "mathematically literate.") The first two books were *Fearless Symmetry* and *Elliptic Tales* (Ash and Gross, 2006; 2012). The first book discussed problems in Diophantine equations, such as Fermat's Last Theorem (FLT). The second discussed problems related to elliptic curves, such as the Birch–Swinnerton-Dyer Conjecture. In both of these books, we ended up mumbling something about modular forms, an advanced topic that plays a crucial role in both of these areas of number theory. By the time we reached the last chapters in these books, we had already introduced so many concepts that we could only allude to the theory of modular forms. One purpose of *Summing It Up* is to give in Part III a more leisurely and detailed explanation of modular forms, motivated by the kinds of problems we will discuss in Parts I and II.

Each of the three books in our trilogy may be read independently of the others. After reading the first two parts of *Summing It Up*, a very diligent person might gain from reading *Fearless Symmetry* or *Elliptic Tales* in tandem with the third part of *Summing It Up*, for they provide additional motivation for learning about modular forms, which are dealt with at length in Part III. Of course, this is not necessary—we believe that the number-theoretical problems studied in the first two parts by themselves lead naturally to a well-motivated study of modular forms.

The three parts of this book are designed for readers of varying degrees of mathematical background. Part I requires a knowledge of high school algebra and geometry. Only in a few dispensable sections is any deeper mathematical knowledge needed. Some of the exposition involves complicated and sometimes lengthy strings of algebraic manipulation, which can be skipped whenever you wish. To read Part II, you will need to have encountered much of

the content of the first year of a standard calculus course (mostly infinite series, differentiation, and Taylor series). You will also need to know about complex numbers, and we review them briefly. Part III does not require any additional mathematical knowledge, but it gets rather intricate. You may need a good dose of patience to read through all of the details.

The level of difficulty of the various chapters and sections sometimes fluctuates considerably. You are invited to browse them in any order. You can always refer back to a chapter or section you skipped, if necessary, to fill in the details. However, in Part III, things will probably make the most sense if you read the chapters in order.

It continues to amaze us what human beings have accomplished, starting with one plus one equals two, getting to two plus two equals four (the cliché example of a simple truth that we know for sure is true), and going far beyond into realms of number theory that even now are active areas of research. We hope you will enjoy our attempts to display some of these wonderful ideas in the pages that follow.

Chapter 1

PROEM

In the interest of allowing the reader to enjoy our book without constantly referring to many other references, we collect in this chapter many standard facts that we will often use in the remainder of the book. A reader familiar with elementary number theory can skip this chapter and refer back to it when necessary. We covered most of these topics in Ash and Gross (2006).

1. Greatest Common Divisors

If a is a positive integer and b is any integer, then long division tells us that we can always divide a into b and get an integer quotient q and integer remainder r. This means that $b = qa + r$, and the remainder r always satisfies the inequality $0 \leq r < a$. For example, if we take $a = 3$ and $b = 14$, then $14 = 4 \cdot 3 + 2$; the quotient $q = 4$ and the remainder $r = 2$. You may not be used to thinking about it, but you can do this with $b < 0$ also. Take $b = -14$ and $a = 3$, and $-14 = (-5) \cdot 3 + 1$; the quotient is $q = -5$, and the remainder is $r = 1$. Notice that if we divide by 2, the remainder will always be 0 or 1; if we divide by 3, the remainder will always be 0, 1, or 2; and so on.

If the result of the long division has $r = 0$, then we say that "a divides b." We write this sentence symbolically as $a \mid b$. Of course, one requirement for long division is that a cannot be 0, so whenever we write $a \mid b$, we implicitly assert that $a \neq 0$. If the remainder r is not zero, we say that "a does not divide b." We write that assertion symbolically as $a \nmid b$. For example, $3 \mid 6$, $3 \nmid 14$, and $3 \nmid (-14)$. Notice that if n is any integer (even 0), then $1 \mid n$. Also, if a is any positive integer, then $a \mid 0$. At the risk of giving too many

Chapter 2

SUMS OF TWO SQUARES

1. The Answer

You can amaze someone who enjoys playing around with numbers by having her choose a prime number and then telling her instantly whether it can be expressed as the sum of two squares. (In section 2 of the Introduction, we have precisely defined "the sum of two squares.") For example, suppose she chooses 97. You say immediately that 97 is the sum of two squares. Then she can experiment by adding pairs of numbers on the list

$$0, 1, 4, 9, 16, 25, 36, 49, 64, 81$$

(allowing a pair to contain the same square twice), and she will find that indeed 97 equals 16 plus 81. If she chooses 79, you say immediately that 79 is not the sum of two squares, and experiment shows that indeed you are correct. How do you do this trick?

THEOREM 2.1: *An odd prime number is the sum of two squares if and only if it leaves a remainder of* 1 *when divided by* 4.

We discuss the proof of this theorem later in the chapter. But even at this point, you can now do the trick yourself. It is as easy to answer the question for very large prime numbers as for small ones, because the remainder when a number in base 10 is divided by 4 depends only on the last two digits of the number. (PROOF: Take the number $n = ab \ldots stu$, where $a, b, \ldots, s, t, u$ are decimal digits. Then $n = 100(ab \ldots s) + tu$. Therefore, $n = 4(25)(ab \ldots s) + tu$, and so the remainder when you divide n by 4 is the same as the remainder

Chapter 3

SUMS OF THREE AND FOUR SQUARES

1. Three Squares

The question of which numbers are the sums of three squares is very much more difficult to answer:

THEOREM 3.1: *A positive integer n is the sum of three squares if and only if n is not equal to a power of 4 times a number of the form $8k + 7$.*

It's easy to see that numbers of the form $8k + 7$ cannot be written as sums of three squares. If $n = 8k + 7$, then $n \equiv 7 \pmod 8$. But a little bit of squaring will show you that every square integer is congruent to either 0, 1, or 4 (mod 8). So three squares could not add up to 7 (mod 8) and hence could not add up to n. Proving the whole theorem is quite difficult, and the proof is not given in such elementary textbooks as Davenport (2008) and Hardy and Wright (2008).

Why are three squares so much more difficult than two or four squares? One answer is that the product of a pair of sums of two squares is itself a sum of two squares, as we saw in formula (2.2). There is another such formula for a pair of sums of four squares, which we will give later in this chapter. There can be no such formula for sums of three squares. Indeed, $3 = 1^2 + 1^2 + 1^2$ and $5 = 0^2 + 1^2 + 2^2$ are each sums of three squares, but their product $15 = 8 \cdot 1 + 7$ is not.

Chapter 4

SUMS OF HIGHER POWERS: WARING'S PROBLEM

1. $g(k)$ and $G(k)$

We have just seen that every positive integer is the sum of four squares, where we count 0 as a square. Many generalizations immediately present themselves. Of these, one of the most interesting derives from an assertion of Edward Waring in 1770 that every number is the sum of 4 squares, of 9 cubes, of 19 biquadrates (fourth powers), and so on. The most interesting part of Waring's statement is "and so on": It implies that after picking a positive integer k, one can find some other integer N such that every positive integer is a sum of N nonnegative kth powers. It seems extremely unlikely that Waring could have had a proof in mind, given that the first proof was published by Hilbert in 1909.

It is customary to use the notation $g(k)$ to represent the smallest integer N such that every positive integer is the sum of N nonnegative kth powers. (To avoid unnecessary repetition, for the rest of this chapter, when we refer to kth powers, we mean nonnegative kth powers.) In this language, the equation $g(2) = 4$ means simultaneously that:

- Every integer is the sum of four squares.
- There are integers that cannot be written as the sum of three squares.

We showed in the previous chapter how to prove the first of these assertions. The second assertion follows by verifying from trial and error that 7 cannot be written as the sum of three squares.