Introduction

WHAT THIS BOOK IS ABOUT

1. Plus

Counting—one, two, three, four or *uno*, *dos*, *tres*, *cuatro* (or in whatever language); or I, II, III, IV or 1, 2, 3, 4, or in whatever symbols—is probably the first theoretical mathematical activity of human beings. It is theoretical because it is detached from the objects, whatever they might be, that are being counted. The shepherd who first piled up pebbles, one for each sheep let out to graze, and then tossed them one by one as the sheep came back to the fold, was performing a practical mathematical act—creating a one-to-one correspondence. But this act was merely practical, without any theory to go with it.

This book is concerned with what may have been the next mathematical activity to be discovered (or invented): addition. We may think that addition is primitive or easy, because we teach it to young children. A moment's reflection will convince you that it must have taken an enormous intellectual effort to conceive of an abstract theory of addition. There cannot be addition of two numbers before there are numbers, and the formation of pure numbers is sophisticated because it involves abstraction.

You may want to plunge into the mathematics of addition and skip the rest of this section. Or you may want to muse a little on philosophical problems connected with the concepts and practice of mathematics, as exemplified in the subject matter of this book. Those who have a taste for these philosophical questions may enjoy the following extremely brief survey.

Perhaps this abstraction of pure numbers was formed through the experience of counting. Once you had a series of number words, you could count axes one day, sheep the next, and apples the third day. After a while, you might just recite those words without any particular thing being counted, and then you might stumble onto the concept of a pure number. It is more probable that arithmetic and the abstract concept of number were developed together.[1]

One way to see the difficulty of the concepts of number, counting, and addition is to look at the philosophy of mathematics, which to this day has not been able to decide on a universally acceptable definition of "number." The ancient Greek philosophers didn't even consider that the number one was a number, since in their opinion numbers were what we counted, and no one would bother to count "one, period."

We won't say more about the very difficult philosophy of mathematics, except to mention one type of issue. Immanuel Kant and his followers were very concerned with addition and how its operations could be justified philosophically. Kant claimed that there were "synthetic truths *a priori*." These were statements that were true, and could be known by us to be true prior to any possible experience, but whose truth was not dependent on the mere meaning of words. For example, the statement "A bachelor doesn't have a wife" is true without needing any experience to vouch for its truth, because it is part of the definition of "bachelor" that he doesn't have a wife. Such a truth was called "analytic *a priori*." Kant claimed that "five plus seven equals twelve" was indubitably true, not needing any experience to validate its truth, but it was "synthetic" because (Kant claimed) the concept of "twelve" was not logically bound up and implied by the concepts of "five," "seven," and "plus." In this way, Kant could point to arithmetic to show that synthetic truths *a priori* existed, and then could go on to consider other such truths that came up later in his philosophy.

In contrast, other philosophers, such as Bertrand Russell, have thought that mathematical truths are all analytic.

[1] Of course, in practice, two apples are easily added to two apples to give four apples. But a theoretical approach that could lead to the development of number theory is more difficult. There was a period in the mathematical thought of the ancient Greek world when "pure" numbers were not clearly distinguished from "object" numbers. For an account of the early history of arithmetic and algebra, we recommend Klein (1992).

These philosophers often think logic is prior to mathematics. And then there is the view that mathematical truths are "*a posteriori*," meaning they are dependent on experience. This seems to have been Ludwig Wittgenstein's opinion. It was also apparently the opinion of the rulers in George Orwell's novel *1984*, who were able to get the hero, in defeat, to believe firmly that two plus two equals five.

The philosophy of mathematics is exceedingly complicated, technical, and difficult. During the twentieth century, things grew ever more vexed. W.V.O. Quine questioned the analytic–synthetic distinction entirely. The concept of truth (which has always been a hard one to corner) became more and more problematic. Today, philosophers do not agree on much, if anything, concerning the philosophical foundations of numbers and their properties. Luckily, we do not need to decide on these philosophical matters to enjoy some of the beautiful theories about numbers that have been developed by mathematicians. We all have some intuitive grasp of what a number is, and that grasp seems to be enough to develop concepts that are both free from contradictions and yield significant theorems about numbers. We can test these theorems doing arithmetic by hand or by computer. By verifying particular number facts, we have the satisfaction of seeing that the theorems "work."

2. Sums of Interest

This book is divided into three parts. The first part will require you to know college algebra and Cartesian coordinates but, except in a few places, nothing substantially beyond that. In this part, we will ask such questions as:

- Is there a short formula for the sum $1 + 2 + 3 + \cdots + k$?
- How about the sum $1^2 + 2^2 + 3^2 + \cdots + k^2$?
- We can be even more ambitious. Let n be an arbitrary integer, and ask for a short formula for $1^n + 2^n + 3^n + \cdots + k^n$.

- How about the sum $1 + a + a^2 + \cdots + a^k$?
- Can a given integer N be written as a sum of perfect squares? Cubes? nth powers? Triangular numbers? Pentagonal numbers?
- Obviously, an integer larger than 1 can be written as a sum of smaller positive integers. We can ask: In how many different ways can this be done?
- If a number can be written as a sum of k squares, in how many different ways can this be done?

Why do we ask these questions? Because it is fun and historically motivated, and the answers lead to beautiful methods of inquiry and amazing proofs.

In the second part of this book, you will need to know some calculus. We will look at "infinite series." These are infinitely long sums that can only be defined using the concept of *limit*. For example,

$$1 + 2 + 3 + \cdots = ?$$

Here the dots mean that we intend the summing to go on "forever." It seems pretty clear that there can't be any answer to this sum, because the partial totals just keep getting bigger and bigger. If we want, we can define the sum to be "infinity," but this is just a shorter way of saying what we said in the previous sentence.

$$1 + 1 + 1 + \cdots = ?$$

This, too, is clearly infinity.

How should we evaluate

$$1 - 1 + 1 - 1 + 1 - 1 + 1 - \cdots = ?$$

Now you might hesitate. Euler[2] said this sum added up to $\frac{1}{2}$.

$$1 + a + a^2 + \cdots = ?$$

[2] One way to justify Euler's answer is to use the formula for the sum of an infinite geometric series. In chapter 7, section 5, we have the formula

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \cdots ,$$

valid as long as $|z| < 1$. If we dare to go outside the region of validity and substitute $z = -1$, we see why Euler interpreted the sum as he did.

We will see that this problem has a nice answer if a is a real number strictly between -1 and 1. You may have learned this answer when you studied "geometric series." We will expand our algebra so that we can use a complex number for a.

Then we can ask about

$$1^n + 2^n + 3^n + \cdots = ?$$

where n is any complex number. This answer (for some values of n) gives a function of n called the ζ-function.

Going back one step, we can add in coefficients:

$$b_0 + b_1 a + b_2 a^2 + \cdots = ?$$

This is the setting in which we introduce the concept of *generating function*, where a itself is a variable.

We can also add coefficients to the ζ-function series and ask about series like

$$c_1 1^n + c_2 2^n + c_3 3^n + \cdots = ?$$

which are called *Dirichlet series*.

These questions and their answers put us in position in the third part of this book to define and discuss *modular forms*. The surprising thing will be how the modular forms tie together the subject matter of the first two parts. This third part will require a little bit of group theory and some geometry, and will be on a somewhat higher level of sophistication than the preceding parts.

One motivation for this book is to explain modular forms, which have become indispensable in modern number theory. In both of our previous two books, modular forms appeared briefly but critically toward the end. In this book, we want to take our time and explain some things about them, although we will only be scratching the surface of a very broad and deep subject. At the end of the book, we will review how modular forms were used in Ash and Gross (2006) to tie up Galois representations and prove Fermat's Last Theorem, and in Ash and Gross (2012) to be able to phrase the fascinating Birch–Swinnerton-Dyer Conjecture about solutions to cubic equations.

As a leitmotif for this book, we will take the problem of "sums of squares," because it is a very old and very pretty problem whose solution is best understood through the theory of modular forms. We can describe the problem a little now.

Consider a whole number n. We say n is a *square* if it equals m^2, where m is also a whole number. For example, 64 is a square, because it is 8 times 8, but 63 is not a square. Notice that we define $0 = 0^2$ as a square, and similarly $1 = 1^2$. We easily list all squares by starting with the list 0, 1, $2, \ldots$ and squaring each number in turn. (Because a negative number squared is the same as its absolute value squared, we only have to use the nonnegative integers.) We obtain the list of squares

$$0, 1, 4, 9, 16, 25, 36, 49, 64, 81, 100, \ldots,$$

in which, as you can see, the squares get farther and farther apart as you go along. (PROOF: The distance between successive squares is $(m+1)^2 - m^2 = m^2 + 2m + 1 - m^2 = 2m + 1$, so the distance gets larger as m gets larger. Notice that this argument gives us more precise information: The list of differences between successive squares is in fact the list of positive odd numbers in increasing order.) We can be very pedantic, if a bit ungrammatical, and say that the list of squares is the list of numbers that are "the sum of one square."

This raises a more interesting question: What is the list of numbers that are the sum of two squares? You could write a computer program that would output this list up to some limit N, and your computer program could generate the list in at least two different ways. First, list all the squares up to N. Then:

METHOD 1: Add pairs of squares on your list together in all possible ways. Then arrange the answers in ascending order.

METHOD 2: Form a loop with n going from 0 to N. For each n, add up all pairs of squares less than or equal to n to see if you get n. If you do, put n on the list and go on to $n + 1$. If you don't, leave n off the list, and go on to $n + 1$.

NOTE: We have defined 0 as a square, so any square number is also a sum of two squares. For example, $81 = 0^2 + 9^2$. Also, we allow a square to be reused, so twice any square number is a sum of two squares. For example, $162 = 9^2 + 9^2$.

Run your program, or add squares by hand. Either way, you get a list of the sums of two squares that starts out like this:

$$0, 1, 2, 4, 5, 8, 9, 10, 13, \ldots.$$

As you can see, not every number is on the list, and it is not immediately clear how to predict if a given number will be a sum of two squares or not. For instance, is there a way to tell if 12345678987654321 is on the list without simply running your computer program? Nowadays, your program would probably only take a fraction of a second to add up all squares up to 12345678987654321, but we can easily write down a number large enough to slow down the computer. More importantly, we would like a theoretical answer to our question, whose proof would give us some insight into which numbers are on the list and which ones are not.

Pierre de Fermat asked this question in the seventeenth century[3] and must have made such a list. There were no computers in the seventeenth century, so his list could not have been all that long, but he was able to guess the correct answer as to which numbers are sums of two squares.[4] In chapter 2, we will supply the answer and discuss the proof in a sketchy way. Because this book is not a textbook, we don't want to give complete proofs. We prefer to tell a story that may be more easily read. If you wish, you can refer to our references and find the complete proof.

Once you get interested in this kind of problem (as did Fermat, who gave a huge impulse to the study of number theory), then it is easy to create more of them. Which numbers are sums of three squares? Of four squares? Of five squares? This particular list of puzzles will stop because, 0 being a square, any sum of four squares

[3] Apparently, another mathematician, Albert Girard, had asked the question and guessed the answer before Fermat, but Fermat publicized the problem.

[4] Fermat announced the answer in a letter to another mathematician, Marin Mersenne, without giving a proof. The first printed proof was written by Leonhard Euler.

will also be a sum of five, six, or any larger amount of squares, and we will see that in fact every positive integer is a sum of four squares.

You could also ask: Which numbers are sums of two cubes, three cubes, four cubes, and so on? You could then substitute higher powers for cubes.

You could ask, as did Euler: Any number is a sum of four squares. A geometrical square has four sides. Is every number a sum of three triangular numbers, five pentagonal numbers, and so on? Cauchy proved that the answer is "yes."

At some point in mathematical history, something very creative happened. Mathematicians started to ask an apparently harder question. Rather than wonder if n can be written as a sum of 24 squares (for example), we ask: *In how many different ways* can n be written as the sum of 24 squares? If the number of ways is 0, then n is not a sum of 24 squares. But if n is a sum of 24 squares, we get more information than just a yes/no answer. It turned out that this harder question led to the discovery of powerful tools that have a beauty and importance that transcends the puzzle of sums of powers, namely tools in the theory of generating functions and modular forms. And that's another thing that this book is about.