Graduate Texts in
Mathematics

164

Additive Number Theory
The Classical Bases

Springer-Verlag

Melvyn B. Nathanson

Additive Number Theory

The Classical Bases

Springer

Melvyn B. Nathanson
Department of Mathematics
Lehman College of the
City University of New York
250 Bedford Park Boulevard West
Bronx, NY 10468-1589 USA

Mathematics Subject Classifications (1991): 11-01, 11P05, 11P32

Library of Congress Cataloging-in-Publication Data
Nathanson, Melvyn B. (Melvyn Bernard), 1944–
Additive number theory:the classical bases/Melvyn B. Nathanson.
p. cm. – (Graduate texts in mathematics;164)
Includes bibliographical references and index.
ISBN 0-387-94656-X (hardcover:alk. paper)
1. Number theory. I. Title. II. Series.
QA241.N347 1996
512′.72–dc20 96-11745

Printed on acid-free paper.

Production managed by Hal Henglein; manufacturing supervised by Jeffrey Taub.
Camera-ready copy prepared from the author's LaTeX files.
Printed and bound by R.R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-94656-X Springer-Verlag New York Berlin Heidelberg SPIN 10490794

Preface

> [Hilbert's] style has not the terseness of many of our modern authors in mathematics, which is based on the assumption that printer's labor and paper are costly but the reader's effort and time are not.
>
> H. Weyl [143]

The purpose of this book is to describe the classical problems in additive number theory and to introduce the circle method and the sieve method, which are the basic analytical and combinatorial tools used to attack these problems. This book is intended for students who want to learn additive number theory, not for experts who already know it. For this reason, proofs include many "unnecessary" and "obvious" steps; this is by design.

The archetypical theorem in additive number theory is due to Lagrange: Every nonnegative integer is the sum of four squares. In general, the set A of nonnegative integers is called an *additive basis of order* h if every nonnegative integer can be written as the sum of h not necessarily distinct elements of A. Lagrange's theorem is the statement that the squares are a basis of order four. The set A is called a *basis of finite order* if A is a basis of order h for some positive integer h. Additive number theory is in large part the study of bases of finite order. The classical bases are the squares, cubes, and higher powers; the polygonal numbers; and the prime numbers. The classical questions associated with these bases are Waring's problem and the Goldbach conjecture.

Waring's problem is to prove that, for every $k \geq 2$, the nonnegative kth powers form a basis of finite order. We prove several results connected with Waring's problem, including Hilbert's theorem that every nonnegative integer is the sum of

a bounded number of kth powers, and the Hardy–Littlewood asymptotic formula for the number of representations of an integer as the sum of s positive kth powers.

Goldbach conjectured that every even positive integer is the sum of at most two prime numbers. We prove three of the most important results on the Goldbach conjecture: Shnirel'man's theorem that the primes are a basis of finite order, Vinogradov's theorem that every sufficiently large odd number is the sum of three primes, and Chen's theorem that every sufficently large even integer is the sum of a prime and a number that is a product of at most two primes.

Many unsolved problems remain. The Goldbach conjecture has not been proved. There is no proof of the conjecture that every sufficiently large integer is the sum of four nonnegative cubes, nor can we obtain a good upper bound for the least number s of nonnegative kth powers such that every sufficiently large integer is the sum of s kth powers. It is possible that neither the circle method nor the sieve method is powerful enough to solve these problems and that completely new mathematical ideas will be necessary, but certainly there will be no progress without an understanding of the classical methods.

The prerequisites for this book are undergraduate courses in number theory and real analysis. The appendix contains some theorems about arithmetic functions that are not necessarily part of a first course in elementary number theory. In a few places (for example, Linnik's theorem on sums of seven cubes, Vinogradov's theorem on sums of three primes, and Chen's theorem on sums of a prime and an almost prime), we use results about the distribution of prime numbers in arithmetic progressions. These results can be found in Davenport's *Multiplicative Number Theory* [19].

Additive number theory is a deep and beautiful part of mathematics, but for too long it has been obscure and mysterious, the domain of a small number of specialists, who have often been specialists only in their own small part of additive number theory. This is the first of several books on additive number theory. I hope that these books will demonstrate the richness and coherence of the subject and that they will encourage renewed interest in the field.

I have taught additive number theory at Southern Illinois University at Carbondale, Rutgers University—New Brunswick, and the City University of New York Graduate Center, and I am grateful to the students and colleagues who participated in my graduate courses and seminars. I also wish to thank Henryk Iwaniec, from whom I learned the linear sieve and the proof of Chen's theorem.

This work was supported in part by grants from the PSC-CUNY Research Award Program and the National Security Agency Mathematical Sciences Program.

I would very much like to receive comments or corrections from readers of this book. My e-mail addresses are nathansn@alpha.lehman.cuny.edu and nathanson@worldnet.att.net. A list of errata will be available on my homepage at http://www.lehman.cuny.edu or http://math.lehman.cuny.edu/nathanson.

Melvyn B. Nathanson
Maplewood, New Jersey
May 1, 1996

Notation and conventions

Theorems, lemmas, and corollaries are numbered consecutively in each chapter and in the Appendix. For example, Lemma 2.1 is the first lemma in Chapter 2 and Theorem A.2 is the second theorem in the Appendix.

The lowercase letter p denotes a prime number.

We adhere to the usual convention that the *empty sum* (the sum containing no terms) is equal to zero and the *empty product* is equal to one.

Let f be any real or complex-valued function, and let g be a positive function. The functions f and g can be functions of a real variable x or arithmetic functions defined only on the positive integers. We write

$$f = O(g)$$

or

$$f \ll g$$

or

$$g \gg f$$

if there exists a constant $c > 0$ such that

$$|f(x)| \leq cg(x)$$

for all x in the domain of f. The constant c is called the *implied constant* . We write

$$f \ll_{a,b,\ldots} g$$

if there exists a constant $c > 0$ that depends on $a, b, \ldots$ such that

$$|f(x)| \leq cg(x)$$

for all x in the domain of f. We write

$$f = o(g)$$

if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0.$$

The function f *is asymptotic to* g, denoted

$$f \sim g,$$

if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

The real-valued function f is *increasing* on the interval I if $f(x_1) \le f(x_2)$ for all $x_1, x_2 \in I$ with $x_1 < x_2$. Similarly, the real-valued function f is *decreasing* on the interval I if $f(x_1) \ge f(x_2)$ for all $x_1, x_2 \in I$ with $x_1 < x_2$. The function f is *monotonic* on the interval I if it is either increasing on I or decreasing on I.

We use the following notation for exponential functions:

$$\exp(x) = e^x$$

and

$$e(x) = \exp(2\pi i x) = e^{2\pi i x}.$$

The following notation is standard:

$\mathbf{Z}$	the integers $0, \pm 1, \pm 2, \ldots$
$\mathbf{R}$	the real numbers
$\mathbf{R}^n$	n-dimensional Euclidean space
$\mathbf{Z}^n$	the integer lattice in $\mathbf{R}^n$
$\mathbf{C}$	the complex numbers
$\lvert z\rvert$	the absolute value of the complex number z
$\Re z$	the real part of the complex number z
$\Im z$	the imaginary part of the complex number z
$[x]$	the integer part of the real number x, that is, the integer uniquely determined by the inequality $[x] \le x < [x] + 1$.
$\{x\}$	the fractional part of the real number x, that is, $\{x\} = x - [x] \in [0, 1)$.
$\lVert x\rVert$	the distance from the real number x to the nearest integer, that is, $\lVert x\rVert = \min\{\lvert x - n\rvert : n \in \mathbf{Z}\} = \min(\{x\}, 1 - \{x\}) \in [0, 1/2]$.
$(a_1, \ldots, a_n)$	the greatest common divisor of the integers $a_1, \ldots, a_n$
$[a_1, \ldots, a_n]$	the least common multiple of the integers $a_1, \ldots, a_n$
$\lvert X\rvert$	the cardinality of the set X
hA	the h-fold sumset, consisting of all sums of h elements of A

Contents

Part I

Waring's problem

1

Sums of polygons

Imo propositionem pulcherrimam et maxime generalem nos primi deteximus: nempe omnem numerum vel esse triangulum vex ex duobus aut tribus triangulis compositum: esse quadratum vel ex duobus aut tribus aut quatuorquadratis compositum: esse pentagonum vel ex duobus, tribus, quatuor aut quinque pentagonis compositum; et sic deinceps in infinitum, in hexagonis, heptagonis polygonis quibuslibet, enuntianda videlicet pro numero angulorum generali et mirabili propostione. Ejus autem demonstrationem, quae ex multis variis et abstrusissimis numerorum mysteriis derivatur, hic apponere non licet. . . .[1]

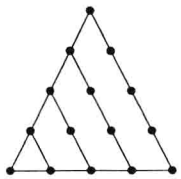
P. Fermat [39, page 303]

---

[1] I have discovered a most beautiful theorem of the greatest generality: Every number is a triangular number or the sum of two or three triangular numbers; every number is a square or the sum of two, three, or four squares; every number is a pentagonal number or the sum of two, three, four, or five pentagonal numbers; and so on for hexagonal numbers, heptagonal numbers, and all other polygonal numbers. The precise statement of this very beautiful and general theorem depends on the number of the angles. The theorem is based on the most diverse and abstruse mysteries of numbers, but I am not able to include the proof here. . . .

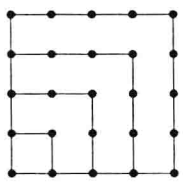
1.1 Polygonal numbers

Polygonal numbers are nonnegative integers constructed geometrically from the regular polygons. The triangular numbers, or triangles, count the number of points in the triangular array



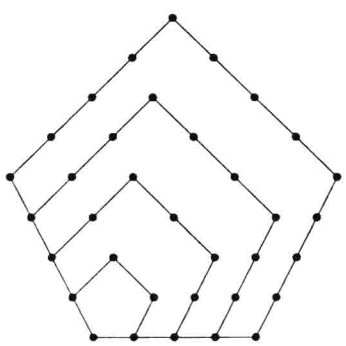
The sequence of triangles is 0, 1, 3, 6, 10, 15, ...

Similarly, the square numbers count the number of points in the square array



The sequence of squares is 0, 1, 4, 9, 16, 25, ...

The pentagonal numbers count the number of points in the pentagonal array



The sequence of pentagonal numbers is 0, 1, 5, 12, 22, 35, ... There is a similar sequence of m-gonal numbers corresponding to every regular polygon with m sides.

Algebraically, for every $m \geq 1$, the kth polygonal number of order $m+2$, denoted $p_m(k)$, is the sum of the first k terms of the arithmetic progression with initial value 1 and difference m, that is,

$$\begin{aligned} p_m(k) &= 1 + (m+1) + (2m+1) + \cdots + ((k-1)m+1) \\ &= \frac{mk(k-1)}{2} + k. \end{aligned}$$

This is a quadratic polynomial in k. The triangular numbers are the numbers

$$p_1(k) = \frac{k(k+1)}{2},$$

the squares are the numbers

$$p_2(k) = k^2,$$

the pentagonal numbers are the numbers

$$p_3(k) = \frac{k(3k-1)}{2},$$

and so on. This notation is awkward but traditional.

The epigraph to this chapter is one of the famous notes that Fermat wrote in the margin of his copy of Diophantus's *Arithmetica*. Fermat claims that, for every $m \geq 1$, every nonnegative integer can be written as the sum of $m+2$ polygonal numbers of order $m+2$. This was proved by Cauchy in 1813. The goal of this chapter is to prove Cauchy's polygonal number theorem. We shall also prove the related result of Legendre that, for every $m \geq 3$, every sufficiently large integer is the sum of five polygonal numbers of order $m+2$.

1.2 Lagrange's theorem

We first prove the polygonal number theorem for squares. This theorem of Lagrange is the most important result in additive number theory.

Theorem 1.1 (Lagrange) *Every nonnegative integer is the sum of four squares.*

Proof. It is easy to check the formal polynomial identity

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2, \tag{1.1}$$

where

$$\left.\begin{aligned} z_1 &= x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 \\ z_2 &= x_1y_2 - x_2y_1 - x_3y_4 + x_4y_3 \\ z_3 &= x_1y_3 - x_3y_1 + x_2y_4 - x_4y_2 \\ z_4 &= x_1y_4 - x_4y_1 - x_2y_3 + x_3y_2 \end{aligned}\right\} \tag{1.2}$$

This implies that if two numbers are both sums of four squares, then their product is also the sum of four squares. Every nonnegative integer is the product of primes, so it suffices to prove that every prime number is the sum of four squares. Since $2 = 1^2 + 1^2 + 0^2 + 0^2$, we consider only odd primes p.

The set of squares

$$\{a^2 \mid a = 0, 1, \ldots, (p-1)/2\}$$

represents $(p+1)/2$ distinct congruence classes modulo p. Similarly, the set of integers

$$\{-b^2 - 1 \mid b = 0, 1, \ldots, (p-1)/2\}$$

represents $(p+1)/2$ distinct congruence classes modulo p. Since there are only p different congruence classes modulo p, by the pigeonhole principle there must exist integers a and b such that $0 \le a, b \le (p-1)/2$ and

$$a^2 \equiv -b^2 - 1 \pmod{p},$$

that is,

$$a^2 + b^2 + 1 \equiv 0 \pmod{p}.$$

Let $a^2 + b^2 + 1 = np$. Then

$$p \le np = a^2 + b^2 + 1^2 + 0^2 \le 2\left(\frac{p-1}{2}\right)^2 + 1 < \frac{p^2}{2} + 1 < p^2,$$

and so

$$1 \le n < p.$$

Let m be the least positive integer such that mp is the sum of four squares. Then there exist integers x_1, x_2, x_3, x_4 such that

$$mp = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

and

$$1 \le m \le n < p.$$

We must show that $m = 1$.

Suppose not. Then $1 < m < p$. Choose integers y_i such that

$$y_i \equiv x_i \pmod{m}$$

and

$$-m/2 < y_i \le m/2$$

for $i = 1, \ldots, 4$. Then

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 \equiv x_1^2 + x_2^2 + x_3^2 + x_4^2 = mp \equiv 0 \pmod{m}$$

and

$$mr = y_1^2 + y_2^2 + y_3^2 + y_4^2$$

for some nonnegative integer r. If $r = 0$, then $y_i = 0$ for all i and each x_i^2 is divisible by m^2. It follows that mp is divisible by m^2, and so p is divisible by m. This is impossible, since p is prime and $1 < m < p$. Therefore, $r \ge 1$ and

$$mr = y_1^2 + y_2^2 + y_3^2 + y_4^2 \le 4(m/2)^2 = m^2.$$

Moreover, $r = m$ if and only if m is even and $y_i = m/2$ for all i. In this case, $x_i \equiv m/2 \pmod{m}$ for all i, and so $x_i^2 \equiv (m/2)^2 \pmod{m^2}$ and

$$mp = x_1^2 + x_2^2 + x_3^2 + x_4^2 \equiv 4(m/2)^2 = m^2 \equiv 0 \pmod{m^2}.$$

This implies that p is divisible by m, which is absurd. Therefore,

$$1 \leq r < m.$$

Applying the polynomial identity (1.1), we obtain

$$\begin{aligned} m^2rp &= (mp)(mr) \\ &= (x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) \\ &= z_1^2 + z_2^2 + z_3^2 + z_4^2, \end{aligned}$$

where the z_i are defined by equations (1.2). Since $x_i \equiv y_i \pmod{m}$, these equations imply that $z_i \equiv 0 \pmod{m}$ for $i = 1, \ldots, 4$. Let $w_i = z_i/m$. Then $w_1, \ldots, w_4$ are integers and

$$rp = w_1^2 + w_2^2 + w_3^2 + w_4^2,$$

which contradicts the minimality of m. Therefore, $m = 1$ and the prime p is the sum of four squares. This completes the proof of Lagrange's theorem.

A set of integers is called a *basis of order h* if every nonnegative integer can be written as the sum of h not necessarily distinct elements of the set. A set of integers is called a *basis of finite order* if the set is a basis of order h for some h. Lagrange's theorem states that the set of squares is a basis of order four. Since 7 cannot be written as the sum of three squares, it follows that the squares do not form a basis of order three. The central problem in additive number theory is to determine if a given set of integers is a basis of finite order. Lagrange's theorem gives the first example of a natural and important set of integers that is a basis. In this sense, it is the archetypical theorem in additive number theory. Everything in this book is a generalization of Lagrange's theorem. We shall prove that the polygonal numbers, the cubes and higher powers, and the primes are all bases of finite order. These are the classical bases in additive number theory.

1.3 Quadratic forms

Let $A = (a_{i,j})$ be an $m \times n$ matrix with integer coefficients. In this chapter, we shall only consider matrices with integer coefficients. Let A^T denote the transpose of the matrix A, that is, $A^T = \left(a_{i,j}^T\right)$ is the $n \times m$ matrix such that

$$a_{i,j}^T = a_{j,i}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Then $(A^T)^T = A$ for every $m \times n$ matrix A, and $(AB)^T = B^T A^T$ for any pair of matrices A and B such that the number of columns of A is equal to the number of rows of B.

Let $M_n(\mathbf{Z})$ be the ring of $n \times n$ matrices. A matrix $A \in M_n(\mathbf{Z})$ is *symmetric* if $A^T = A$. If A is a symmetric matrix and U is any matrix in $M_n(\mathbf{Z})$, then $U^T AU$ is also symmetric, since

$$(U^T AU)^T = U^T A^T (U^T)^T = U^T AU.$$