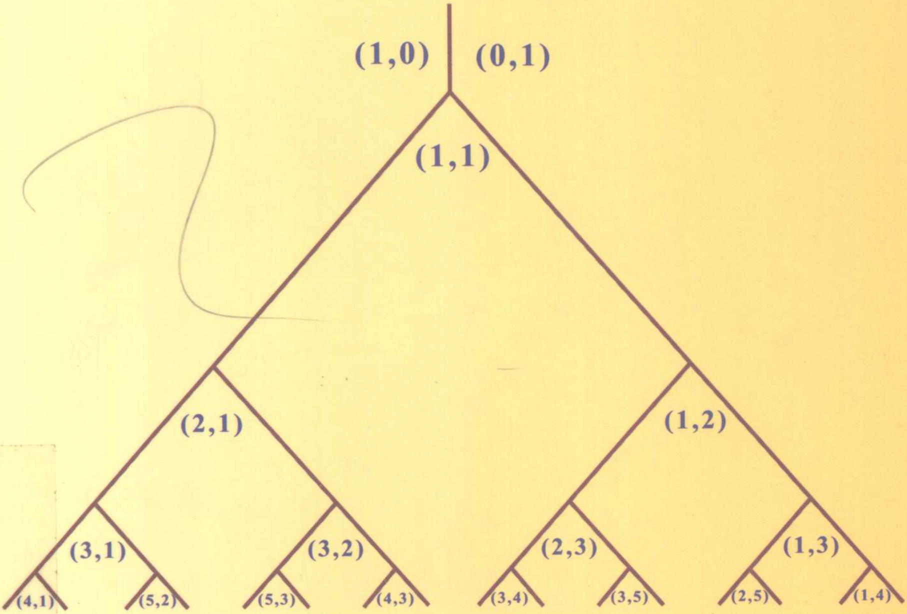
John Stillwell

ELEMENTS OF NUMBER THEORY

数论基础



Springer

世界图书出版公司
www.wpcbj.com.cn

John Stillwell

Elements of Number Theory

With 35 figures

Springer

John Stillwell
Mathematics Department
University of San Francisco
2130 Fulton Street
San Francisco, CA 94117-1080
USA
stillwell@usfca.edu

Mathematics Subject Classification (2000): 11-01

Library of Congress Cataloging-in-Publication Data
Stillwell, John.
Elements of number theory / John Stillwell.
p. cm.—(Undergraduate texts in mathematics)
Includes bibliographical references and index.
ISBN 0-387-95587-9. I. Title. II. Series.
QA241 .S813 2002
512'.7—dc21 2002030569

ISBN 0-387-95587-9

springeronline.com

图书在版编目（CIP）数据

数论基础＝Elements of Number Theory：英文/（美）史迪威（Stillwell，J.）著.—北京：世界图书出版公司北京公司，2009.5
ISBN 978-7-5100-0467-4

Ⅰ.数… Ⅱ.史… Ⅲ.数论—高等学校—教材—英文
Ⅳ.O156

中国版本图书馆CIP数据核字（2009）第066890号

书　　名：Elements of Number Theory
作　　者：John Stillwell

中 译 名：数论基础
责任编辑：高蓉　刘慧

出 版 者：世界图书出版公司北京公司
印 刷 者：三河国英印务有限公司
发　　行：世界图书出版公司北京公司（北京朝内大街137号　100010）
联系电话：010-64021602，010-64015659
电子信箱：kjb@wpcbj.com.cn

开　　本：24开
印　　张：11.5
版　　次：2009年05月
版权登记：图字：01-2009-1079

书　　号：978-7-5100-0467-4/O·682　　　定　　价：35.00元

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics

Abbott: Understanding Analysis.
Anglin: Mathematics: A Concise History and Philosophy.
Readings in Mathematics.
Anglin/Lambek: The Heritage of Thales.
Readings in Mathematics.
Apostol: Introduction to Analytic Number Theory. Second edition.
Armstrong: Basic Topology.
Armstrong: Groups and Symmetry.
Axler: Linear Algebra Done Right. Second edition.
Beardon: Limits: A New Approach to Real Analysis.
Bak/Newman: Complex Analysis. Second edition.
Banchoff/Wermer: Linear Algebra Through Geometry. Second edition.
Berberian: A First Course in Real Analysis.
Bix: Conics and Cubics: A Concrete Introduction to Algebraic Curves.
Brémaud: An Introduction to Probabilistic Modeling.
Bressoud: Factorization and Primality Testing.
Bressoud: Second Year Calculus.
Readings in Mathematics.
Brickman: Mathematical Introduction to Linear Programming and Game Theory.
Browder: Mathematical Analysis: An Introduction.
Buchmann: Introduction to Cryptography.
Buskes/van Rooij: Topological Spaces: From Distance to Neighborhood.
Callahan: The Geometry of Spacetime: An Introduction to Special and General Relavitity.
Carter/van Brunt: The Lebesgue–Stieltjes Integral: A Practical Introduction.
Cederberg: A Course in Modern Geometries. Second edition.
Childs: A Concrete Introduction to Higher Algebra. Second edition.
Chung: Elementary Probability Theory with Stochastic Processes. Third edition.
Cox/Little/O'Shea: Ideals, Varieties, and Algorithms. Second edition.
Croom: Basic Concepts of Algebraic Topology.
Curtis: Linear Algebra: An Introductory Approach. Fourth edition.
Devlin: The Joy of Sets: Fundamentals of Contemporary Set Theory. Second edition.
Dixmier: General Topology.
Driver: Why Math?
Ebbinghaus/Flum/Thomas: Mathematical Logic. Second edition.
Edgar: Measure, Topology, and Fractal Geometry.
Elaydi: An Introduction to Difference Equations. Second edition.
Erdős/Surányi: Topics in the Theory of Numbers.
Estep: Practical Analysis in One Variable.
Exner: An Accompaniment to Higher Mathematics.
Exner: Inside Calculus.
Fine/Rosenberger: The Fundamental Theory of Algebra.
Fischer: Intermediate Real Analysis.
Flanigan/Kazdan: Calculus Two: Linear and Nonlinear Functions. Second edition.
Fleming: Functions of Several Variables. Second edition.
Foulds: Combinatorial Optimization for Undergraduates.
Foulds: Optimization Techniques: An Introduction.
Franklin: Methods of Mathematical Economics.
Frazier: An Introduction to Wavelets Through Linear Algebra.

(continued after index)

To Elaine

Preface

This book is intended to complement my *Elements of Algebra*, and it is similarly motivated by the problem of solving polynomial equations. However, it is independent of the algebra book, and probably easier. In *Elements of Algebra* we sought *solution by radicals*, and this led to the concepts of *fields* and *groups* and their fusion in the celebrated theory of Galois. In the present book we seek *integer solutions*, and this leads to the concepts of *rings* and *ideals* which merge in the equally celebrated *theory of ideals* due to Kummer and Dedekind.

Solving equations in integers is the central problem of number theory, so this book is truly a number theory book, with most of the results found in standard number theory courses. However, numbers are best understood through their algebraic structure, and the necessary algebraic concepts—rings and ideals—have no better motivation than number theory.

The first nontrivial examples of rings appear in the number theory of Euler and Gauss. The concept of ideal—today as routine in ring theory as the concept of normal subgroup is in group theory—also emerged from number theory, and in quite heroic fashion. Faced with failure of unique prime factorization in the arithmetic of certain generalized "integers", Kummer created in the 1840s a new kind of number to overcome the difficulty. He called them "ideal numbers" because he did not know exactly what they were, though he knew how they behaved. Dedekind in 1871 found that these "ideal numbers" could be realized as *sets* of actual numbers, and he called these sets *ideals*.

Dedekind found that ideals could be defined quite simply; so much so that a student meeting the concept today might wonder what all the fuss is about. It is only in their role as "ideal numbers", where they realize Kummer's impossible dream, that ideals can be appreciated as a genuinely brilliant idea.

Thus solution in integers—like solution by radicals—is a superb setting in which to show algebra at its best. It is the right place to introduce rings and ideals and the right place first to apply them. It even gives an opportunity to introduce some exotic rings, such as the quaternions, which we use to prove Lagrange's theorem that every natural number is the sum of four squares.

The book is based on two short courses (about 20 lectures each) given at Monash University in recent years; one on elementary number theory and one on ring theory with applications to algebraic number theory. Thus the amount of material is suitable for a one-semester course, with some variation possible through omission of the optional starred sections. A slower-paced course could stop at the end of Chapter 9, at which point most of the standard results have been covered, from Euclid's theorem that there are infinitely many primes to quadratic reciprocity.

It should be stressed, however, that this is not meant to be a standard number theory course. I have tried to avoid the ad hoc proofs that once gave number theory a bad name, in favor of unifying ideas that work in many situations. These include algebraic structures but also ideas from elementary number theory, such as the Euclidean algorithm and unique prime factorization. In particular, I use the Euclidean algorithm as a bridge to Conway's visual theory of quadratic forms, which offers a new approach to the Pell equation.

There are exercises at the end of almost every section, so that each new idea or proof receives immediate reinforcement. Some of them focus on specific ideas, while others recapitulate the general line of argument (in easy steps) to prove a similar result. The purpose of each exercise should be clear from the accompanying commentary, so instructors and independent readers alike will be able to find an enjoyable path through the book.

My thanks go to the Monash students who took the courses on which the book is based. Their reactions have helped improve the presentation in many ways. I am particularly grateful to Ley Wilson, who showed that it is possible to master the book by independent study.

Special thanks go to my wife Elaine, who proofread the first version of the book, and to John Miller and Abe Shenitzer, who carefully read the revised version and saved me from many mathematical and stylistic errors.

JOHN STILLWELL
South Melbourne, July 2002

Contents

1

Natural numbers and integers

PREVIEW

Counting is presumably the origin of mathematical thought, and it is certainly the origin of difficult mathematical problems. As the great Hungarian problem-solver Paul Erdős liked to point out, if you can think of an open problem that is more than 200 years old, then it is probably a problem in number theory.

In recent decades, difficulties in number theory have actually become a virtue. *Public key encryption*, whose security depends on the difficulty of factoring large numbers, has become one of the commonest applications of mathematics in daily life.

At any rate, problems are the life blood of number theory, and the subject advances by building theories to make them understandable. In the present chapter we introduce some (not so difficult) problems that have played an important role in the development of number theory because they lead to basic methods and concepts.

- Counting leads to *induction*, the key to all facts about numbers, from banalities such as $a+b=b+a$ to the astonishing result of Euclid that there are infinitely many primes.
- Division (with remainder) is the key computational tool in Euclid's proof and elsewhere in the study of primes.
- Binary notation, which also results from division with remainder, leads in turn to a method of "fast exponentiation" used in public key encryption.
- The Pythagorean equation $x^2+y^2=z^2$ from geometry is equally important in number theory because it has integer solutions.

In this chapter we are content to show these ideas at work in few interesting but seemingly random situations. Later chapters will develop the ideas in more depth, showing how they unify and explain a great many astonishing properties of numbers.

1.1 Natural numbers

Number theory starts with the *natural numbers*

$$1,2,3,4,5,6,7,8,9,\ldots,$$

generated from 1 by successively adding 1. We denote the set of natural numbers by $\mathbb{N}$. On $\mathbb{N}$ we have the operations $+$ and $\times$, which are simple in themselves but lead to more sophisticated concepts.

For example, we say that *a divides n* if $n = ab$ for some natural numbers a and b. A natural number p is called *prime* if the only natural numbers dividing p are 1 and p itself.

Divisibility and primes are behind many of the interesting questions in mathematics, and also behind the recent applications of number theory (in cryptography, internet security, electronic money transfers etc.).

The sequence of prime numbers begins with

$$2,3,5,7,11,13,17,19,23,29,31,37,\ldots$$

and continues in a seemingly random manner. There is so little pattern in the sequence that one cannot even see clearly whether it continues forever. However, Euclid (around 300 BCE) proved that *there are infinitely many primes*, essentially as follows.

Infinitude of primes. *Given any primes $p_1, p_2, p_3, \ldots, p_k$, we can always find another prime p.*

Proof. Form the number

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

Then none of the given primes $p_1, p_2, p_3, \ldots, p_k$ divides N because they all leave remainder 1. On the other hand, *some* prime p divides N. If N itself is prime we can take $p = N$, otherwise $N = ab$ for some smaller numbers a and b. Likewise, if either a or b is prime we take it to be p, otherwise split a and b into smaller factors, and so on. Eventually we must reach a prime p dividing N because natural numbers cannot decrease forever. □

Exercises

Not only is the sequence of primes without apparent pattern, there is not even a known simple formula that produces only primes. There are, however, some interesting "near misses".

1.1.1 Check that the quadratic function n^2+n+41 is prime for all small values of n (say, for n up to 30).

1.1.2 Show nevertheless that n^2+n+41 is not prime for certain values of n.

1.1.3 Which is the smallest such value?

1.2 Induction

The method just used to find the prime divisors of N is sometimes called *descent*, and it is an instance of a general method called *induction.*

The "descent" style of induction argument relies on the fact that any process producing smaller and smaller natural numbers must eventually halt. The process of repeatedly adding 1 reaches any natural number n in a finite number of steps, hence there are only finitely many steps *downward* from n. There is also an "ascent" style of induction that imitates the construction of the natural numbers themselves—starting at some number and repeatedly adding 1.

An "ascent" induction proof is carried out in two steps: the *base step* (getting started) and the *induction step* (going from n to $n+1$). Here is an example: proving that *any number of the form* k^3+2k *is divisible by* 3.

Base step. The claim is true for $k=1$ because $1^3+2\times 1=3$, which is certainly divisible by 3.

Induction step. Suppose that the claim is true for $k=n$, that is, 3 divides n^3+2n. We want to deduce that it is true for $k=n+1$, that is, that 3 divides $(n+1)^3+2(n+1)$. Well,

$$\begin{aligned}&(n+1)^3+2(n+1)\\&=n^3+3n^2+3n+1+2n+2\\&=n^3+2n+3n^2+3n+3\\&=n^3+2n+3(n^2+n+1)\end{aligned}$$

And the right-hand side is the sum of n^3+2n, which we are supposing to be divisible by 3, and $3(n^2+n+1)$, which is obviously divisible by 3. Therefore $(n+1)^3+2(n+1)$ is divisible by 3, as required. □

Induction is fundamental not only for proofs of theorems about $\mathbb{N}$ but also for defining the basic functions on $\mathbb{N}$. Only one function needs to be assumed, namely the *successor function* $s(n) = n+1$; then $+$ and $\times$ can be defined by induction. In this book we are not trying to build everything up from bedrock, so we shall assume $+$ and $\times$ and their basic properties, but it is worth mentioning their inductive definitions, since they are so simple.

For any natural number m we define $m+1$ by

$$m+1 = s(m).$$

Then, given the definition of $m+n$ for all m, we define $m+s(n)$ by

$$m+s(n) = s(m+n).$$

It then follows, by induction on n, that $m+n$ is defined for all natural numbers m and n. The definition of $m\times n$ is similarly based on the successor function and the $+$ function just defined:

$$\begin{aligned} m\times 1 &= m \\ m\times s(n) &= m\times n+m. \end{aligned}$$

From these inductive definitions one can give inductive *proofs* of the basic properties of $+$ and $\times$, for example $m+n=n+m$ and $l(m+n)=lm+ln$. Such proofs were first given by Grassmann (1861) (in a book intended for high school students!) but they went unnoticed. They were rediscovered, together with an analysis of the successor function itself, by Dedekind (1888). For more on this see Stillwell (1998), Chapter 1.

Exercises

An interesting process of descent may be seen in the algorithm for the so-called *Egyptian fractions* introduced by Fibonacci (1202). The goal of the algorithm is to represent any fraction $\frac{b}{a}$ with $0<b<a$ as sum of distinct terms $\frac{1}{n}$, called *unit fractions*. (The ancient Egyptians represented fractions in this way.)

Fibonacci's algorithm, in a nutshell, is to *repeatedly subtract the largest possible unit fraction*. Applied to the fraction $\frac{11}{12}$, for example, it yields

$$\frac{11}{12}-\frac{1}{2}=\frac{5}{12}, \quad \text{subtracting the largest unit fraction, } \tfrac{1}{2}\text{, less than } \tfrac{11}{12},$$

$$\frac{5}{12}-\frac{1}{3}=\frac{1}{12}, \quad \text{subtracting the largest unit fraction, } \tfrac{1}{3}\text{, less than } \tfrac{5}{12},$$

$$\text{hence } \frac{11}{12}=\frac{1}{2}+\frac{1}{3}+\frac{1}{12}.$$