AN INTRODUCTION TO THE THEORY OF NUMBERS

IVAN NIVEN
University of Oregon

HERBERT S. ZUCKERMAN
University of Washington

NEW YORK · LONDON
John Wiley & Sons, Inc.

Library of Congress Catalog Card Number: 60-10322

Printed in the United States of America

AN INTRODUCTION TO THE THEORY OF NUMBERS

PREFACE

Our purpose is to present a reasonably complete introduction to the theory of numbers within the compass of a single volume. The basic concepts are presented in the first part of the book, followed by more specialized material in the final three chapters. Paralleling this progress from general topics to more particular discussions, we have attempted to begin the book at a more leisurely pace than we have followed later. Thus the later parts of the book are set forth in a more compact and sophisticated presentation than are the earlier parts.

The book is intended for seniors and beginning graduate students in American and Canadian universities. It contains at least enough material for a full year course; a short course can be built by the use of Sections 1.1 to 1.3, 2.1 to 2.4, 3.1, 3.2, 4.1, 5.1 to 5.3, 5.5, 6.1, and 6.2. Various other arrangements are possible because the chapters beyond the fourth are, apart from a very few exceptions, independent of one another. The final three chapters are entirely independent of each other.

To enable the student to deepen his understanding of the subject, we have provided a considerable number of problems. The variety of these exercises is extensive, ranging from simple numerical problems to additional developments of the theory. The beginner at number theory should take warning that the subject is noted for the difficulty of its problems. Many an innocent looking problem gives, by the very simplicity of its statement, very little notion of the considerable ingenuity or depth of insight required for its solution. As might be expected, the more difficult problems are placed toward the ends of the sets. In many instances three or four consecutive problems constitute a related series in which the last ones can be solved more readily by use of information from the first ones. As a matter of principle we

have made the text itself entirely independent of the problems. In no place does the proof of a theorem depend on the results of any problem.

In choosing methods of proof, we have tried to include as many methods as possible. We have tried to state the proofs accurately, avoiding statements that could be misleading and also avoiding unduly long discussions of unimportant details. As the reader progresses he will become familiar with more and more methods, and he should be able to construct accurate proofs by patterning them after our proofs.

The reader interested in further exploration of the subject will find the bibliography at the end of the book of considerable use. In particular, anyone interested in the history of the subject should consult O. Ore, *Number Theory and Its History*, and, for more specific information, L. E. Dickson, *History of the Theory of Numbers*. Our approach is analytical, not historical, and we make no attempt to attribute various theorems and proofs to their original discoverers. However, we do wish to point out that we followed the suggestion of Peter Scherk that we use F. J. Dyson's formulation of the proof of Mann's $\alpha\beta$ Theorem. Our proof is based on notes graciously placed at our disposal by Peter Scherk. For permission to use several problems from the *American Mathematical Monthly*, we are indebted to the editors. We also appreciate the careful reading of the manuscript by Margaret Maxfield, whose efforts resulted in numerous improvements. Finally, we would like to record our deep appreciation of and our great debt to the mathematicians whose lectures were vital to our introduction to the theory of numbers: L. E. Dickson, R. D. James, D. N. Lehmer, and Hans Rademacher.

IVAN NIVEN
HERBERT S. ZUCKERMAN

June 1960

CONTENTS

CHAPTER 1

DIVISIBILITY

1.1 Introduction

The theory of numbers is primarily concerned with the properties of the natural numbers, 1, 2, 3, 4, $\cdots$, also called the positive integers. However, the theory is not strictly confined to just the natural numbers or even to the set of all integers: 0, ± 1, ± 2, ± 3, $\cdots$. In fact, some theorems of number theory are most easily proved by making use of the properties of real or complex numbers even though the statement of the theorems may involve only natural numbers. Also, there are theorems concerning real numbers that depend so heavily on the properties of integers that they are properly included in the theory of numbers.

An integer n greater than 1 is called a prime if it has no divisor d such that $1 < d < n$. The fact that for every given positive integer m there is a prime greater than m is stated in terms of integers, and it can be proved from the properties of the natural numbers alone. The fact that every natural number can be expressed as a sum of, at most, fifty-four fifth powers of integers is also stated in terms of natural numbers, but any known proof depends on properties of complex numbers. Finally, the question as to how many primes there are that do not exceed x clearly belongs to the theory of numbers but its answer involves the function $\log x$ and is well outside of the realm of the natural numbers. The last two examples are beyond the scope of this book. However, we do not restrict ourselves to the integers but will use real and complex numbers when it is convenient. The questions discussed in this book are not numerical computations or numerical curiosities, except insofar as these are relevant to general propositions. Nor do we discuss the foundations

of the number system; it is assumed that the reader is familiar not only with the integers, but also with the rational and real numbers. However, a rigorous logical analysis of the real-number system is not prerequisite to the study of number theory.

The theory of numbers relies for proofs on a great many ideas and methods. Of these, there are two basic principles to which we draw especial attention. The first is that any set of positive integers has a smallest element if it contains any members at all. In other words, if a set S of positive integers is not empty, then it contains an integer s such that for any member a of S, the relation $s \leqq a$ holds. The second principle, mathematical induction, is a logical consequence of the first.* It can be stated as follows: if a set S of positive integers contains the integer 1, and contains $n + 1$ whenever it contains n, then S consists of all the positive integers.

It may be well to point out that a negative assertion such as, for example, "Not every positive integer can be expressed as a sum of the squares of three integers," requires only that we produce a single example—the number 7 cannot be so expressed. On the other hand, a positive assertion such as "Every positive integer can be expressed as a sum of the squares of four integers," cannot be proved by producing examples, however numerous. This result is Theorem 5.6 in Chapter 5, where a proof is supplied.

Finally, it is presumed that the reader is familiar with the usual formulation of mathematical propositions. In particular, if A denotes some assertion or collection of assertions, and B likewise, the following statements are logically equivalent—they are just different ways of saying the same thing.

A implies B.

If A is true, then B is true.

In order that A be true it is necessary that B be true.

B is a necessary condition for A.

A is a sufficient condition for B.

If A implies B and B implies A, then one can say that B is a necessary and sufficient condition for A to hold.

* Compare G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, Macmillan, revised edition, 1953, pp. 10–13.

In general, we shall use roman letters a, b, c, $\cdots$, m, n, $\cdots$, x, y, z to designate integers unless otherwise specified.

1.2 Divisibility

Definition 1.1 *An integer b is divisible by an integer a, not zero, if there is an integer x such that* $b = ax$, *and we write* $a|b$. *In case b is not divisible by a we write* $a\nmid b$.

Other language for the divisibility property $a|b$ is that a divides b, that a is a divisor of b, and that b is a multiple of a. If $a|b$ and $0 < a < b$ then a is called a proper divisor of b. It is understood that we never use 0 as the left member of the pair of integers in $a|b$. On the other hand, not only may 0 occur as the right member of the pair, but also in such instances we always have divisibility. Thus $a|0$ for every integer a not zero. The notation $a^K \| b$ is sometimes used to indicate that $a^K|b$ but $a^{K+1}\nmid b$.

Theorem 1.1

(1) $a|b$ *implies* $a|bc$ *for any integer* c;
(2) $a|b$ *and* $b|c$ *imply* $a|c$;
(3) $a|b$ *and* $a|c$ *imply* $a|(bx + cy)$ *for any integers* x *and* y;
(4) $a|b$ *and* $b|a$ *imply* $a = \pm b$;
(5) $a|b$, $a > 0$, $b > 0$, *imply* $a \leqq b$.

Proof. The proofs of these results follow at once from the definition of divisibility. Property 3 admits an obvious extension to any finite set, thus:

$$a|b_1,\ a|b_2,\ \cdots,\ a|b_n \text{ imply } a \left| \sum_{j=1}^{n} b_j x_j \right. \text{ for any integers } x_j.$$

Property 2 can be extended similarly.

Theorem 1.2 *The division algorithm. Given any integers a and b, with* $a > 0$, *there exist integers q and r such that* $b = qa + r$, $0 \leqq r < a$. *If* $a\nmid b$, *then r satisfies the stronger inequalities* $0 < r < a$.

Proof. Consider the arithmetic progression

$$\cdots,\ b - 3a,\ b - 2a,\ b - a,\ b,\ b + a,\ b + 2a,\ b + 3a,\ \cdots$$

extending indefinitely in both directions. In this sequence, select the

smallest non-negative member and denote it by r. Thus by definition r satisfies the inequalities of the theorem. But also r, being in the sequence, is of the form $b - qa$, and thus q is defined in terms of r, and the proof is complete.

We have stated the theorem with the assumption $a > 0$. However this hypothesis is not necessary, and we may formulate the theorem without it: given any integers a and b, with $a \neq 0$, there exist integers q and r such that $b = qa + r$, $0 \leqq r < |a|$.

Theorem 1.2 is called the division algorithm. An algorithm is a mathematical procedure or method to obtain a result. We have stated Theorem 1.2 in the form "there exist integers q and r," and this wording suggests that we have a so-called existence theorem rather than an algorithm. However, it may be observed that the proof does give a method for obtaining the integers q and r, because the infinite arithmetic progression $\cdots, b - a, b, b + a, \cdots$ need be examined only in part to yield the smallest positive member r.

In actual practice the quotient q and the remainder r are obtained by the arithmetic division of a into b.

Definition 1.2 *The integer a is a* common divisor *of b and c in case $a|b$ and $a|c$. Since there is only a finite number of divisors of any non-zero integer, there is only a finite number of common divisors of b and c, except in the case $b = c = 0$. If at least one of b and c is not 0, the greatest among their common divisors is called the* greatest common divisor *of b and c, and is denoted by (b, c). Similarly we denote the greatest common divisor g of the integers $b_1, b_2, \cdots, b_n$, not all zero, by $(b_1, b_2, \cdots, b_n)$.*

Thus the greatest common divisor (b, c) is defined for every pair of integers b, c except $b = 0$, $c = 0$, and we note that $(b, c) \geqq 1$.

Theorem 1.3 *If g is the greatest common divisor of b and c, then there exist integers x_0 and y_0 such that $g = (b, c) = bx_0 + cy_0$.*

Proof. Consider the linear combinations $bx + cy$, where x and y range over all integers. This set of integers $\{bx + cy\}$ includes positive and negative values, and also 0 by the choice $x = y = 0$. Choose x_0 and y_0 so that $bx_0 + cy_0$ is the least positive integer l in the set; thus $l = bx_0 + cy_0$.

Next we prove that $l|b$ and $l|c$. We establish the first of these, and the second follows by analogy. We give an indirect proof that $l|b$, that is, we assume $l \nmid b$ and obtain a contradiction. From $l \nmid b$ it follows that

there exist integers q and r, by Theorem 1.2, such that $b = lq + r$ with $0 < r < l$. Hence we have $r = b - lq = b - q(bx_0 + cy_0) = b(1 - qx_0) + c(-qy_0)$, and thus r is in the set $\{bx + cy\}$. This contradicts the fact that l is the least positive integer in the set $\{bx + cy\}$.

Now since g is the greatest common divisor of b and c, we may write $b = gB$, $c = gC$, and $l = bx_0 + cy_0 = g(Bx_0 + Cy_0)$. Thus $g|l$, and so by part 5 of Theorem 1.1, we conclude that $g \leqq l$. Now $g < l$ is impossible, since g is the *greatest* common divisor, and so $g = l = bx_0 + cy_0$.

Theorem 1.4 *The greatest common divisor g of b and c can be characterized in the following two ways*: (1) *it is the least positive value of $bx + cy$ where x and y range over all integers*; (2) *it is the positive common divisor of b and c which is divisible by every common divisor.*

Proof. Part 1 follows from the proof of Theorem 1.3. To prove part 2, we observe that if d is any common divisor of b and c, then $d|g$ by part 3 of Theorem 1.1. Moreover, there cannot be two distinct integers with property 2, because of Theorem 1.1, part 5.

Theorem 1.5 *Given any integers $b_1, b_2, \cdots, b_n$ not all zero, with greatest common divisor g, there exist integers $x_1, x_2, \cdots, x_n$ such that*

$$g = (b_1, b_2, \cdots, b_n) = \sum_{j=1}^{n} b_j x_j.$$

Furthermore g is the least positive value of the linear form $\sum_{j=1}^{n} b_j y_j$ where the y_j range over all integers; also g is the positive common divisor of $b_1, b_2, \cdots, b_n$ which is divisible by every common divisor.

Proof. This result is a straightforward generalization of the preceding two theorems, and the proof is analogous without any complications arising in the passage from two integers to n integers.

Theorem 1.6 *For any positive integer m,*

$$(ma, mb) = m(a, b).$$

Proof. By Theorem 1.4 we have

$$\begin{aligned}(ma, mb) &= \text{least positive value of } max + mby \\ &= m \cdot \{\text{least positive value of } ax + by\} \\ &= m(a, b).\end{aligned}$$

Theorem 1.7 *If $d|a$ and $d|b$ and $d > 0$ then*

$$\left(\frac{a}{d}, \frac{b}{d}\right) = \frac{1}{d}(a, b).$$

If $(a, b) = g$, then

$$\left(\frac{a}{g}, \frac{b}{g}\right) = 1.$$

Proof. The second assertion is the special case of the first obtained by using the greatest common divisor g of a and b in the role of d. The first assertion in turn is a direct consequence of Theorem 1.6 obtained by replacing m, a, b in that theorem by d, (a/d), (b/d) respectively.

Theorem 1.8 *If $(a, m) = (b, m) = 1$, then $(ab, m) = 1$.*

Proof. By Theorem 1.3 there exist integers x_0, y_0, x_1, y_1 such that $1 = ax_0 + my_0 = bx_1 + my_1$. Thus we may write $(ax_0)(bx_1) = (1 - my_0)(1 - my_1) = 1 - my_2$ where y_2 is defined by the equation $y_2 = y_0 + y_1 - my_0y_1$. From the equation $abx_0x_1 + my_2 = 1$ we note, by part 3 of Theorem 1.1, that any common divisor of ab and m is a divisor of 1, and hence $(ab, m) = 1$.

Definition 1.3 *We say that a and b are* relatively prime *in case $(a, b) = 1$, and that $a_1, a_2, \cdots, a_n$ are* relatively prime *in case $(a_1, a_2, \cdots, a_n) = 1$. We say that $a_1, a_2, \cdots, a_n$ are* relatively prime in pairs *in case $(a_i, a_j) = 1$ for all $i = 1, 2 \cdots, n$ and $j = 1, 2, \cdots, n$ with $i \neq j$.*

The fact that $(a, b) = 1$ is sometimes expressed by saying that a and b are coprime.

Theorem 1.9 *For any x, $(a, b) = (b, a) = (a, -b) = (a, b + ax)$.*

Proof. Denote (a, b) by d and $(a, b + ax)$ by g. It is clear that $(b, a) = (a, -b) = d$.

By application of Theorem 1.1, parts 3 and 4, we obtain $d|g$, $g|d$ and $d = g$.

Theorem 1.10 *If $c|ab$ and $(b, c) = 1$, then $c|a$.*

Proof. By Theorem 1.6, $(ab, ac) = a(b, c) = a$. But $c|ab$ and $c|ac$, and so $c|a$ by Theorem 1.4.

Given two integers b and c, how can the greatest common divisor g be found? Definition 1.2 gives no answer to this question, and neither

does Theorem 1.3 which merely asserts the existence of a pair of integers x_0 and y_0 such that $g = ax_0 + by_0$. If b and c are small, values of g, x_0, and y_0 can be found by inspection. For example, if $b = 10$ and $c = 6$, then it is obvious that $g = 2$, and one pair of values for x_0, y_0 is $2, -3$. We now state an algorithm which gives a general method for finding the value of g and also values of x_0 and y_0. By Theorem 1.9, $(b, c) = (b, -c)$, and hence we may presume c positive, because the case $c = 0$ is very special: $(b, 0) = |b|$.

Theorem 1.11 *The Euclidean algorithm. Given integers b and $c > 0$, we make a repeated application of the division algorithm, Theorem* 1.2, *to obtain a series of equations*

$$\begin{aligned} b &= cq_1 + r_1, & 0 < r_1 < c, \\ c &= r_1q_2 + r_2, & 0 < r_2 < r_1, \\ r_1 &= r_2q_3 + r_3, & 0 < r_3 < r_2, \\ &\cdots & \cdots \\ r_{j-2} &= r_{j-1}q_j + r_j, & 0 < r_j < r_{j-1}, \\ r_{j-1} &= r_jq_{j+1}. \end{aligned}$$

The greatest common divisor (b, c) of b and c is r_j, the last non-zero remainder in the division process. Values of x_0 and y_0 in $(b, c) = bx_0 + cy_0$ can be obtained by eliminating $r_1, r_2, \cdots, r_{j-1}$ from the set of equations.

Example. $b = 963$, $c = 657$.

$$\begin{aligned} 963 &= 657 \cdot 1 + 306 \\ 657 &= 306 \cdot 2 + 45 \\ 306 &= 45 \cdot 6 + 36 \\ 45 &= 36 \cdot 1 + 9 \\ 36 &= 9 \cdot 4 \end{aligned}$$

Thus $(963, 657) = 9$, and we can express 9 as a linear combination of 963 and 657 by eliminating the remainders 36, 45, and 306 as follows:

$$\begin{aligned} 9 &= 45 - 36 \\ &= 45 - (306 - 45 \cdot 6) \\ &= -306 + 7 \cdot 45 \\ &= -306 + 7(657 - 306 \cdot 2) \\ &= 7 \cdot 657 - 15 \cdot 306 \\ &= 7 \cdot 657 - 15(963 - 657) \\ &= 22 \cdot 657 - 15 \cdot 963. \end{aligned}$$

Proof. The chain of equations is obtained by dividing c into b, r_1 into c, r_2 into $r_1, \cdots, r_j$ into r_{j-1}. The process stops when the division is

exact, that is when the remainder is zero. Thus in our application of Theorem 1.2 we have written the inequalities for the remainder without an equality sign. Thus, for example, $0 < r_1 < c$ in place of $0 \leqq r_1 < c$, because if r_1 were equal to zero, the chain would stop at the first equation $b = cq_1$, in which case the greatest common divisor of b and c would be c.

We now prove that r_j is the greatest common divisor g of b and c. Since $g|b$ and $g|c$, we see that $g|r_1$ by the first equation of the chain. Since $g|c$ and $g|r_1$ we see that $g|r_2$ by the second equation. Continuing by mathematical induction we find that $g|r_j$. On the other hand, the final equation implies that $r_j|r_{j-1}$. This, together with the next to last equation, implies $r_j|r_{j-2}$. Continuing by mathematical induction, we conclude that $r_j|b$ and $r_j|c$. By Theorem 1.4, $r_j|g$. Hence $g = r_j$ by Theorem 1.1.

To see that r_j is expressible as a linear combination of b and c, we need merely eliminate r_1 from the first two equations of the chain, then eliminate r_2 from this and the third equation. Proceeding by successive eliminations of $r_3, r_4, \cdots, r_{j-1}$, we obtain r_j in the form $bx_0 + cy_0$.

Definition 1.4 *The integers $a_1, a_2, \cdots, a_n$, all different from zero, have a* common multiple *b if $a_i|b$ for $i = 1, 2, \cdots, n$. (Note that common multiples do exist, for example the product $a_1a_2 \cdots a_n$ is one.) The least of the positive common multiples is called the least common multiple, and it is denoted by $[a_1, a_2, \cdots, a_n]$.*

Theorem 1.12 *If b is any common multiple of $a_1, a_2, \cdots, a_n$, then $[a_1, a_2, \cdots, a_n]|b$. This is the same as saying that if h denotes $[a_1, a_2, \cdots, a_n]$, then $0, \pm h, \pm 2h, \pm 3h, \cdots$ comprise all the common multiples of $a_1, a_2, \cdots, a_n$.*

Proof. Let m be any common multiple and divide m by h. By Theorem 1.2 there is a quotient q and a remainder r such that $m = qh + r, 0 \leqq r < h$. We must prove that $r = 0$. If $r \neq 0$ we argue as follows. For each $i = 1, 2, \cdots, n$ we know that $a_i|h$ and $a_i|m$, so that $a_i|r$. Thus r is a positive common multiple of $a_1, a_2, \cdots, a_n$ contrary to the fact that h is the least positive of all the common multiples.

Theorem 1.13 *If $m > 0$, $[ma, mb] = m[a, b]$. Also $[a, b] \cdot (a, b) = |ab|$.*

Proof. Since $[ma, mb]$ is a multiple of ma, it is *a fortiori* a multiple of m, and so can be written in the form mh_1. Denoting $[a, b]$ by h_2, we note that $a|h_2$, $b|h_2$, $am|mh_2$, $bm|mh_2$, and so $mh_1|mh_2$ by Theorem 1.12. Thus $h_1|h_2$. On the other hand, $am|mh_1$, $bm|mh_1$, $a|h_1$, $b|h_1$ and so,

$h_2|h_1$. We conclude that $h_1 = h_2$ and thus the first part of the theorem is established.

It will suffice to prove the second part for positive integers a and b, since $[a, -b] = [a, b]$. We begin with the special case where $(a, b) = 1$. Now $[a, b]$ is a multiple of a, say ma. Then $b|ma$ and $(a, b) = 1$, so by Theorem 1.10 we conclude that $b|m$. Hence $b \leqq m$, $ba \leqq ma$. But ba, being a positive common multiple of b and a cannot be less than the least common multiple, and so $ba = ma = [a, b]$.

Turning to the general case where $(a, b) = g > 1$, we have $((a/g), (b/g)) = 1$ by Theorem 1.7. Applying the result of the preceding paragraph, we obtain

$$\left[\frac{a}{g}, \frac{b}{g}\right]\left(\frac{a}{g}, \frac{b}{g}\right) = \frac{a}{g}\frac{b}{g}.$$

Multiplying by g^2 and using Theorem 1.6 as well as the first part of the present theorem, we get $[a, b]\,(a, b) = ab$.

PROBLEMS

1. By using the Euclidean algorithm find the greatest common divisor (g.c.d.) of
   (*a*) 7469 and 2464; (*b*) 2689 and 4001;
   (*c*) 2947 and 3997; (*d*) 1109 and 4999.
2. Find the greatest common divisor g of the numbers 1819 and 3587, and then find integers x and y to satisfy
   $1819x + 3587y = g$.
3. Find values of x and y to satisfy
   (*a*) $243x + 198y = 9$
   (*b*) $71x - 50y = 1$
   (*c*) $43x + 64y = 1$
   (*d*) $93x - 81y = 3$
   (*e*) $6x + 10y + 15z = 1$.
4. Find the least common multiple (l.c.m.) of (*a*) 482 and 1687; (*b*) 60 and 61.
5. Prove that the product of three consecutive integers is divisible by 6; of four consecutive integers by 24.
6. Exhibit three integers that are relatively prime but not relatively prime in pairs.
7. Two integers are said to be of the same parity if they are both even or both odd; if one is even and the other odd, they are said to be of opposite parity, or of different parity. Given any two integers, prove that their sum and their difference are of the same parity.