Modern Birkhäuser Classics

Hans Riesel

Prime Numbers and Computer Methods for Factorization

Second Edition

Birkhäuser

Modern Birkhäuser Classics

Many of the original research and survey monographs in pure and applied mathematics, as well as textbooks, published by Birkhäuser in recent decades have been groundbreaking and have come to be regarded as foundational to the subject. Through the MBC Series, a select number of these modern classics, entirely uncorrected, are being re-released in paperback (and as eBooks) to ensure that these treasures remain accessible to new generations of students, scholars, and researchers.

Prime Numbers and Computer Methods for Factorization

Second Edition

Hans Riesel

Reprint of the 1994 Edition

Birkhäuser

Hans Riesel
Department of Mathematics
The Royal Institute of Technology
S-100 44 Stockholm
Sweden
riesel@nada.kth.se

Originally published as Volume 126 in the series *Progress in Mathematics*

ISBN 978-0-8176-8297-2 e-ISBN 978-0-8176-8298-9
DOI 10.1007/978-0-8176-8298-9
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011943337

Springer is part of Springer Science+Business Media
(www.birkhauser-science.com)

Hans Riesel

Prime Numbers and Computer Methods for Factorization

Second Edition

Birkhäuser
Boston • Basel • Berlin

Hans Riesel
Department of Mathematics
The Royal Institute of Technology
S-100 44 Stockholm
Sweden

Library of Congress Cataloging-in-Publication Data

Riesel, Hans, 1929-
Prime numbers and computer methods for factorization / Hans Riesel. -- 2nd ed.
p. cm. -- (Progress in mathematics ; v. 126)
Includes bibliographical references and index.
Contents: v. 1. Fundamental algorithms
ISBN 0-8176-3743-5
1. Numbers, Prime--Data processing. 2. Factorization (Mathematics)--Data processing. I. Title. II. Series : Progress in mathematics (Boston, Mass.) ; vol. 126.
II. Title. III. Series.
QA246.R54 1994 94-27688
512'.72--dc20 CIP

Printed on acid-free paper

Birkhäuser ®

ISBN 0-8176-3743-5
ISBN 3-7643-3743-5
Typeset by the author.
Printed and bound by Quinn-Woodbine, Woodbine, NJ.
Printed in the U.S.A.

9 8 7 6 5 4 3 2 1

PREFACE

In this book the author treats four fundamental and apparently simple problems. They are: the number of primes below a given limit, the approximate number of primes, the recognition of prime numbers and the factorization of large numbers. A chapter on the details of the distribution of the primes is included as well as a short description of a recent application of prime numbers, the so-called RSA public-key cryptosystem. The author is also giving explicit algorithms and computer programs. Whilst not claiming completeness, the author has tried to give all important results known, including the latest discoveries. The use of computers has in this area promoted a development which has enormously enlarged the wealth of results known and that has made many older works and tables obsolete.

As is often the case in number theory, the problems posed are easy to understand but the solutions are theoretically advanced. Since this text is aimed at the mathematically inclined layman, as well as at the more advanced student, not all of the proofs of the results given in this book are shown. Bibliographical references in these cases serve those readers who wish to probe deeper. References to recent original works are also given for those who wish to pursue some topic further.

Since number theory is seldom taught in basic mathematics courses, the author has appended six sections containing all the algebra and number theory required for the main body of the book. There are also two sections on multiple precision computations and, finally, one section on Stieltjes integrals. This organization of the subject-matter has been chosen in order not to disrupt the reader's line of thought by long digressions into other areas. It is also the author's hope that the text, organized in this way, becomes more readable for specialists in the field. Any reader who gets stuck in the main text should first consult the appropriate appendix, and then try to continue.

The six chapters of the main text are quite independent of each other, and need not be read in order.

For those readers who have a computer (or even a programmable calculator) available, computer programs have been provided for many of the methods described. In order to achieve a wide understanding, these programs are written in the high-level programming language PASCAL. With this choice the author hopes that most readers will be able to translate the programs into the language of the computer they use with reasonable effort.

PREFACE

At the end of the book a large amount of results are collected in the form of tables, originating partly from the author's own work in this field. All tables have been verified and up-dated as far as possible. Also in connection with the tables, bibliographical references are given to recent or to more extensive work in the corresponding area.

The text is an up-dated version of an earlier book by the same author: "En bok om primtal," existing in Swedish only.

The author is very much indebted to Richard Brent, Arne Fransén, Gunnar Hellström, Hans Karlgren, D. H. Lehmer, Thorkil Naur, Andrzej Schinzel, Bob Vaughan and many others for their reading of the manuscript and for suggesting many improvements. Thanks are also due to Beatrice Frock for revising the English, and to the late Ken Clements for reading and correcting one of the last versions of the manuscript.

The author wishes you a pleasant reading!

Stockholm, February 1985

PREFACE TO THE SECOND EDITION

During the last ten years the science of computational number theory has seen great advances. Several important new methods have been introduced to recognize prime numbers and to factor composite integers. These advances have stimulated the author to add subsections on the applications of the elliptic curve method and on the number field sieve to the main text as well as two new appendices covering the basics of these new subjects.—Also, very many minor updatings and corrections of the text have been made.

The author wishes to express his thanks to the many readers who have written during the years and proposed improvements of the text. Special thanks are going to Richard Brent, François Morain, Peter Montgomery and in particular to Harvey Dubner and Wilfrid Keller, who all helped during various stages of the preparation of the new manuscript.

Stockholm, June 1994

CONTENTS

Chapter 1. The Number of Primes Below a Given Limit

Chapter 2. The Primes Viewed at Large

Chapter 3. Subtleties in the Distribution of Primes

Chapter 4. The Recognition of Primes

Chapter 5. Classical Methods of Factorization

Chapter 6. Modern Factorization Methods

Appendix 2. Basic Concepts in Higher Arithmetic

Appendix 3. Quadratic Residues

Appendix 4. The Arithmetic of Quadratic Fields

NOTATIONS

Symbol	Meaning
$f(x) \approx g(x)$	$f(x)$ is approximately equal to $g(x)$
$f(x) \sim g(x)$	$f(x)$ is asymptotically equal to $g(x)$, meaning that $\lim f(x)/g(x) = 1$, usually as $x \to \infty$
$[a, b]$	closed interval: all x in $a \leq x \leq b$
$\mathbf{C}$	the complex numbers
$\overset{c}{=}$, $\overset{c}{\sim}$, $\overset{c}{<}$	conjectured relation
$\overline{a}$	conjugate number: if $a = p + q\sqrt{D}$, then $\overline{a} = p - q\sqrt{D}$
$a \equiv b \bmod n$	a is congruent to b modulus n
$a \not\equiv b \bmod n$	a is not congruent to b modulus n
$a + \frac{b\|}{\|c} + \frac{d\|}{\|e}$	continued fraction $= a + \cfrac{b}{c + \cfrac{d}{e}}$
$a \mid b$	a divides b
$a \nmid b$	a does not divide b
$p^{\alpha} \| n$	highest power of p dividing n, i.e. $p^{\alpha} \mid n$, but $p^{\alpha+1} \nmid n$
$\varphi(n)$	Euler's totient function $= \prod p_i^{\alpha_i - 1}(p_i - 1)$,if $n = \prod p_i^{\alpha_i}$
$\Phi_n(x)$	the nth cyclotomic polynomial, having degree $\varphi(n)$
γ	Euler's constant $= 0.57721566\,49015328\ldots$
GCD(a, b)	greatest common divisor of a and b
M_N	group of primitive residue classes mod$N$
$\lfloor x \rfloor$	integer part of x : $\lfloor \pi \rfloor = 3$, $\lfloor -\pi \rfloor = -4$
$[a, b]$	the interval $a \leq x \leq b$
$\left(\frac{a}{N}\right)$	Jacobi's symbol, defined if GCD$(a, N) = 1$
$\lambda(n)$	Carmichael's function $=$ LCM$[\lambda(p_i^{\alpha_i})]_i$, if $n = \prod p_i^{\alpha_i}$
LCM $[a, b]$	least common multiple of a and b
LHS	left hand side
$\left(\frac{a}{p}\right)$	Legendre's symbol, defined if p is an odd prime

Symbol	Meaning
$\operatorname{li} x$	the logarithmic integral of x
$\ln x$	$\log_e x$ $(e = 2.71828182\,84590452\ldots)$
$\log x$	logarithm to an unspecified base
$\mu(n)$	Möbius' function
$a >> b$	a is much larger than b
$a << b$	a is much smaller than b
#	"number sign": number of elements in a set
$\omega(N)$	the number of different prime factors of N
$\Omega(N)$	the total number of prime factors of N
$\Omega(f(x))$	"big omega": greater than $Cf(x)$ for some constant $C > 0$ and an infinitude of x, usually as $x \to \infty$
$O(f(x))$	"big ordo": less than $Cf(x)$ for some constant $C > 0$, usually as $x \to \infty$
$o(f(x))$	"little ordo": less than $\epsilon f(x)$, where $\epsilon \to 0$, usually as $x \to \infty$
p	a prime number
$P(N)$	the largest prime factor of N
$P_k(N)$	the kth largest prime factor of N
$\pi(x)$	number of primes $\leq x$
$\prod$	product symbol: $\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdot a_3 \cdots a_n$
Q	the rational numbers
R	the real numbers
$\Re(z)$	real part x of complex number $z = x + iy$
RHS	right hand side
$\sum$	summation symbol: $\sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + \cdots + a_n$
$\zeta(s)$	zeta-function of Riemann, $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$
$\chi(A)$	group character
Z	the rational integers
10(5)100	the numbers 10, 15, 20, ..., 100

CHAPTER 1

THE NUMBER OF PRIMES BELOW A GIVEN LIMIT

What Is a Prime Number?

Consider the positive integers $1, 2, 3, 4, \ldots$ Among them there are *composite numbers* and *primes*. A composite number is a product of several factors $\neq 1$, such as $15 = 3 \cdot 5$; or $16 = 2 \cdot 8$. A prime p is characterized by the fact that its only possible factorization apart from the order of the factors is $p = 1 \cdot p$. Every composite number can be written as a product of *primes*, such as $16 = 2 \cdot 2 \cdot 2 \cdot 2$.—Now, what can we say about the integer 1? Is it a prime or a composite? Since 1 has the only possible factorization $1 \cdot 1$ we could agree that it is a prime. We might also consider the product $1 \cdot p$ as a product of two primes; somewhat awkward for a *prime number* p.—The dilemma is solved if we adopt the convention of classifying the number 1 as neither prime nor composite. We shall call the number 1 a *unit*. The positive integers may thus be divided into:

1. The unit 1.
2. The prime numbers $2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots$
3. The composite numbers $4, 6, 8, 9, 10, 12, 14, 15, 16, \ldots$

Frequently, it is of interest to study not only the *positive* integers, but *all* integers:

$$\ldots -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots$$

The numbers n and $-n$ are called *associated* numbers. The number 0, which is divisible by any integer without remainder (the quotient always being 0), is of a special kind. The integers are thus categorized as:

1. The number 0.
2. The units -1 and 1.
3. The primes $\ldots -7, -5, -3, -2, 2, 3, 5, 7, \ldots$
4. The composite numbers $\ldots -9, -8, -6, -4, 4, 6, 8, 9, \ldots$

Generally, when only the *factorization* of numbers is of interest, associated numbers may be considered as equivalent, therefore in this case the two different classifications of the integers given above can be considered equivalent (if we neglect the number 0).—We shall often find that the numbers 0 and 1 have special properties in the theory of numbers which will necessitate some additional explanation, just as in the above case.

H. Riesel, *Prime Numbers and Computer Methods for Factorization: Second Edition*,
Modern Birkhäuser Classics, DOI 10.1007/978-0-8176-8298-9_1,

The Fundamental Theorem of Arithmetic

When we were taught at school how to find the *least common denominator,* LCM, of several fractions, we were actually using *the fundamental theorem of arithmetic.* It is a theorem which well illustrates the fundamental role of the prime numbers. The theorem states that every positive integer n can be written as a product of primes, and in one way only:

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s} = \prod_{i=1}^{s} p_i^{\alpha_i}. \tag{1.1}$$

In order that this decomposition be unique, we have to consider, as identical, decompositions that differ only in the order of the prime factors, and we must also refrain from using associated factors.—Having done much arithmetic, we have become so used to this theorem that we regard it as self-evident, and not necessary to prove. This is, however, not at all the case, and the reader will find a proof on p. 261 in Appendix 2. In Appendix 3 an example is given which shows why the proof is a logical necessity. And on p. 295 an arithmetic system is constructed which resembles the ordinary integers in many ways, *but in which the fundamental theorem of arithmetic does not hold.*—As a matter of fact, the logical necessity for a proof of the fundamental theorem was recognized by Euclid, who gave the proof of the almost equivalent Theorem A2.1 on p. 261.

Which Numbers Are Primes? The Sieve of Eratosthenes

All the primes in a given interval can be found by a sieve method invented by Eratosthenes. This method deletes all the composite numbers in the interval, leaving the primes. How can we find all the composites in an interval? To check if a number n is a multiple of p, divide n by p and see whether the division leaves no remainder. This so-called trial division method for finding primes is much too laborious for the construction of prime tables when applied to each number individually, but it turns out that *only one division by each prime p suffices* for the whole interval. To see why this is so, suppose that the first number in the interval is m, and that $m - 1 = p \cdot q + r$, with $0 \leq r < p$. Then, obviously, the numbers

$$m - 1 - r + p,\ m - 1 - r + 2p,\ m - 1 - r + 3p,\ \dots$$

are precisely those multiples of p which are $\geq m$. Thus all multiples of p can be identified by one single division by p and the number of divisions performed for each number examined will be reduced accordingly. This saves much labor, particularly if the interval is long.

Which values of p need to be tested as factors of any given number n? It obviously suffices to test all values of $p \leq \sqrt{n}$, since a composite number n cannot

have two factors, both $> \sqrt{n}$, because the product of these factors would then exceed n. Thus, if all composite numbers in an interval $m \leq x \leq n$ are to be sieved, it will suffice to cross out the multiples of all primes $\leq \sqrt{n}$ (with the possible exception of some primes in the beginning of the interval, a case which may occur if $m \leq \sqrt{n}$). Using these principles, we shall show how to construct a table of primes between 100 and 200. Commence by enumerating all integers from 100 to 200: 100, 101, ..., 199, 200. The multiples to be crossed out are the multiples of all primes $\leq \sqrt{200}$, i.e., of the primes $p = 2, 3, 5, 7, 11$ and 13. First, strike out the multiples of 2, all even numbers, leaving the odd numbers:

$$101,\ 103,\ 105,\ \ldots,\ 195,\ 197,\ 199.$$

Next delete all multiples of 3. Since $99 = 3 \cdot 33 + 0$, the smallest multiple of 3 in the interval is $99 + 3 = 102$. By counting 3 steps at a time, starting from 102, we find all multiples of 3 in the interval: 102, 105, 108, 111,... When we are working with paper and pencil, we might strike out the multiples of 5 in the same round, since these are easy to recognize in our decimal number system. After this step the following numbers remain:

$$101,\ 103,\ 107,\ 109,\ 113,\ 119,\ 121,\ 127,\ 131,\ 133,$$
$$137,\ 139,\ 143,\ 149,\ 151,\ 157,\ 161,\ 163,\ 167,\ 169,$$
$$173,\ 179,\ 181,\ 187,\ 191,\ 193,\ 197 \text{ and } 199.$$

Next, we must locate the multiples of 7, 11 and 13. These are 119, 133, 161 (the remaining multiples of 7), 121, 143, 187 (the remaining multiples of 11) and 169 (the remaining multiple of 13). The 21 numbers now remaining,

$$101,\ 103,\ 107,\ 109,\ 113,\ 127,\ 131,\ 137,\ 139,\ 149,\ 151,$$
$$157,\ 163,\ 167,\ 173,\ 179,\ 181,\ 191,\ 193,\ 197 \text{ and } 199,$$

are the primes between 100 and 200. This example demonstrates the fact that it requires approximately the same amount of work to construct a prime table as it would take to devise a table for the smallest factor of each number for a given interval. After having made some simple modifications increasing the efficiency of the sieve of Eratosthenes described above, D. N. Lehmer at the beginning of this century compiled and in 1909 published the largest factor table [1] and the largest prime table [2] ever published as books. These cover the integers up to 10,017,000. It is unlikely that more extensive factor tables or prime tables will ever be published as books since, with a computer, the factorization of a small number is much faster than looking it up in a table. Before the era of computers a considerable amount of work was invested in the construction of prime and factor tables. Most famous of these is probably Kulik's manuscript, containing a table of the smallest factor of each number up to 10^8. The manuscript, of which parts have

been lost, was never published. Before computers became readily available, a lot of effort was spent on big prime table projects. For example, in 1956–57 the author of this book compiled a prime table for the interval 10,000,000–70,000,000. The computations were performed on the vacuum-tube computer BESK, which had an addition time of 56 micro-seconds. To sieve an interval containing 48,000 integers took 3 minutes of computing time. The output from the BESK was a teletype paper tape which had to be re-read and printed by special equipment, consisting of a paper tape reader and an electric typewriter.

Finally, in 1959, C. L. Baker and F. J. Gruenberger published a micro-card edition of a table containing the first six million primes [3]. This table comprises all the primes below 104,395,289. In addition, there exist printed tables for the primes in certain intervals, such as the 1000:th million $[999 \cdot 10^6,\ 10^9]$, see [4], and the intervals $[10^n,\ 10^n + 150{,}000]$ for $n = 8, 9, \ldots, 15$, see [5].

General Remarks Concerning Computer Programs

For all numerical work on primes it is essential to have prime tables accessible in your computer. This may be achieved by the computer generating the primes by means of the sieve of Eratosthenes, and then either immediately performing the computations or statistics required on the primes, or compressing the prime table generated in a way suitable for storing in the computer and then performing the necessary computation or statistics.

We shall provide a number of *algorithms* in this book (i.e., schemes for calculating desired quantities). These will be presented in one or more of the following forms:

1. Verbal descriptions.
2. Mathematical formulas.
3. Flow-chart diagrams.
4. Computer programs in PASCAL.

The choice of the high-level programming language PASCAL has been made because this language is simple, reasonably well-known, and close to the mathematical formulas which are to be transformed into computer programs. There are, however, some minor differences between the various versions of PASCAL in use, in particular concerning the input and output routines; hopefully this will not cause any great obstacle. The version of PASCAL used in this book is Standard PASCAL. It is the author's hope that readers trained in computer programming will be able to transform the PASCAL programs presented in this book into the language of their choice without undue difficulty.

All the programs in this book have been collected in a package, which can be obtained by `anonymous ftp` from the site `ftp.nada.kth.se` in the directory `Num`. The package is named `riesel-comp.tar`.

A Sieve Program

For a sieve to operate on the interval $[m, n]$, we require all the primes up to p_s, the largest prime $\leq \lfloor\sqrt{n}\rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x. We assume that these have already been computed and stored in a one-dimensional array, `Prime[1:s]`, with `Prime[1]:=2, Prime[2]:=3, Prime[3]:=5,...,` `Prime[s]:=` the sth prime. Now, in order to simplify the computer program, we shall work only with the *odd* integers. Without loss of generality, let both m and n be *odd* (if this is not the case to begin with, we may change m and n during the input part of the program, if we really want the program to work also for *even* values of m and/or n). Next we give each of the $(n - m + 2)/2$ *odd* integers between m and n, inclusive, a corresponding storage location in the fast access memory of the computer. This may be a computer word, a byte, or even a bit, depending on the level on which you are able to handle data in your computer. We shall call the array formed by these storage locations `Table[1:(n-m+2)/2]`, where $(n - m + 2)/2$ is the number of elements in the array (its dimension). Suppose that these storage locations are filled with ZEROs to begin with. Each time we find a multiple of some prime $\leq \sqrt{m}$, we shall put the number 1 into the corresponding storage location. Here is a PASCAL program to do this for an interval $[m, n]$ below 1000:

```
PROGRAM Eratosthenes
{Prints a prime table between odd m and n < 1000}
(Input,Output);
LABEL 1;
CONST imax=11; jmax=500;
VAR Prime : ARRAY[1..imax] OF INTEGER;
    Table : ARRAY[1..jmax] OF INTEGER;
  m,n,p,p2,q,i,j,start,stop : INTEGER;

BEGIN
Prime[1]:=2;  Prime[2]:=3;   Prime[3]:=5;  Prime[4]:=7;
Prime[5]:=11; Prime[6]:=13;  Prime[7]:=17; Prime[8]:=19;
Prime[9]:=23; Prime[10]:=29; Prime[11]:=31;
write('Input m and n: '); read(m,n);
stop:=(n-m+2) DIV 2;
FOR i:=2 TO stop DO
  BEGIN p:=Prime[i]; p2:=p*p;
    IF p2 < m THEN
      BEGIN q:=2*p; start:=(m DIV q)*q+p;
        {start is the odd multiple of p
          which is closest to m}
        IF start < m THEN start:=start+q
      END
```

```
        ELSE start:=p2;
    IF p2 > n THEN GOTO 1 ELSE
        BEGIN j:=(start-m) DIV 2 + 1;
          WHILE j <= stop DO
            BEGIN {Here the odd multiples of p are
              marked:} Table[j]:=1; j:=j+p
            END
        END
  END;
1: {When arriving at this point, the elements of Table
     corresponding to the primes have the value 0, the
     others have the value 1}
  FOR i:=1 TO stop DO
    IF Table[i]=0 THEN write(m+2*i-2:4)
    {Here the prime table generated is printed out}
END.
```

The table of small primes required by this program is built up by the statements at the beginning: `Prime[1]:=2;...Prime[11]:=31;` When dealing with a larger interval it is practical also to produce the table of small primes required by a sieving program rather than defining them arithmetically. The only necessary augmentation of the program in that case involves starting at the beginning of the generated table, searching for the next entry = 0, and replacing this ZERO with the corresponding integer, which is the next prime.—Since it is convenient to have computations of general interest, such as sieving with the primes, readily available as *computer procedures*, we also give below a modified version of the above sieve program, writen in the form of a PASCAL procedure `primes`, which generates the odd primes in the interval [3, n], n odd. Since it is the first time we show a PASCAL procedure we provide a complete program `Primegenerator` containing this procedure to enable the reader to understand the context in which a PASCAL procedure should be written. The following program reads an odd integer $m < 1000$, and generates and prints all odd primes below m:

```
PROGRAM Primegenerator
{Generates and prints the primes up to m < 1000, m odd}
(Input,Output);
CONST n=500;
TYPE vector=ARRAY[0..n] OF INTEGER;
VAR Table : vector; i,j,m : INTEGER;

PROCEDURE primes(n : INTEGER; VAR Prime : vector);
LABEL  1;
VAR i,j,k,p,p2,stop : INTEGER;
```

```
BEGIN
  stop:=(n-1) DIV 2; j:=1;
1:  FOR k:=j TO stop DO
      IF Prime[k]=0 THEN {The next prime for sieving
          has been obtained}
        BEGIN p:=2*k+1; j:=k+1; p2:=p*p;
          IF p2 <= n THEN BEGIN
            i:=(p2-1) DIV 2; WHILE i <= stop DO
              BEGIN Prime[i]:=1; i:=i+p END;
            GOTO 1 END END;
    {When arriving here all elements of the array Prime
      corresponding to the primes below m have the value
      0, the others have the value 1}
END {primes};

BEGIN
  write('Input m: '); read(m); primes(m,Table);
  j:=(m-1) DIV 2;
  FOR i:=1 TO j DO IF Table[i]=0 THEN write(2*i+1:4);
END.
```

With computer programs of the type shown above, it is possible to investigate properties of the series of primes, depending on all the individual primes as high as $3 \cdot 10^{11}$, see for instance [6] or [6′].

Exercise 1.1. Primes in intervals. Write a program for your computer, that performs the following: Read two (odd) integers m and $n \geq m$. Generate an array containing the odd primes below $\sqrt{n}$. Sieve out all composites between m and n by the sieve of Eratosthenes, utilizing the array of primes below $\sqrt{n}$. Print the primes between m and n and their numbers, equalling $\pi(n) - \pi(m-1)$, where the function $\pi(x)$ is defined on p. 10 below. Suitable test values are: $(m, n) = (99, 201)$, $(12113, 12553)$ (check the computer's answer against the last column of Table 1 on p. 377), $(9553, 9585)$ (an interval entirely without primes). More test values can be picked from Tables 2 and 3.

Compact Prime Tables

If the prime table generated by the computer needs to be kept for later use or perhaps for output, the information can be given in a rather compact form. The simplest method is to store the sequence of ZEROs and ONEs representing the status (composite or prime) of all the odd numbers. Please note that this time we denote the *primes* by ONEs! A table of primes below 107 in this representation will appear as

01110 11011 01001 10010 11010 01001 10010 11001 01001 00010 1101

Here the first five entries in the table, viz. 0, 1, 1, 1, 0, correspond to the integers 1, 3 (=prime), 5 (=prime), 7 (=prime) and 9. The final four entries show that 101, 103 and 107 are primes, while 105 is a composite number.

Since most computers are binary, it is convenient to store this sequence of ZEROs and ONEs (bits) in computer words, where the number of bits per word depends upon the computer's word-size. Thus, in a 36-bit computer we could store a table of the primes among 36 consecutive odd numbers in one computer word. (It might actually be much easier to use only 35 of the 36 bits available, at least when programming in a high-level language.)

This method of storing primes is rather simple. We can, by eliminating also the multiples of 3 and 5 in our example, compress the prime tables even more, but at the cost of working with a more complicated pattern. We are then left with all numbers of the forms

$$30k+1,\ 30k+7,\ 30k+11,\ 30k+13,\ 30k+17,\ 30k+19,\ 30k+23,\ \text{and}\ 30k+29$$

i.e., with only 8 numbers out of 30. Thus, we need to store only one bit for 8 *odd* numbers out of 15, or just about half of them. By removing also multiples of 7, 11 and 13, we can further reduce the storage required by a factor of $\frac{6}{7} \cdot \frac{10}{11} \cdot \frac{12}{13} = 0.72$. If we do so, however, the pattern of the remaining integers is quite complicated and repeats periodically after as many as 5760 steps, leading to some tricky computer programming as well as to a time-consuming table look-up function.—Thus, optimal efficiency requires balancing how large the prime table to be stored against how much computational work required each time a table look-up is requested. A reasonable compromise is to exclude only multiples of 3 saving 1/3 of the storage otherwise needed, i.e., to store one bit of information for each of the numbers $6k \pm 1$ only. The sequence of primes between 5 and 107 in this representation will be

11111 11011 01111 01011 01110 11010 01111

which is the sequence previously shown with each third bit removed, the removal starting with the second bit, corresponding to the number 3. The leading zero, corresponding to the number 1, has also been removed. The final 5 bits in the string shown represent the character (prime or composite) of the numbers 95, 97, 101, 103 and 107.

In a 36-bit computer, the primes in an interval from $105k$ to $105(k+1)$ can thus be stored in 35 of the 36 bits of a computer word. This string of bits may be reversed and printed out as an integer $< 2^{35}$. A prime table up to $105 \times 100 = 10500$ looks rather strange when printed out in this way (see next page). The reader should compare this with the prime table up to 12553 provided at the end of this book. The table printed there contains slightly more information than the print-out on next page, as the print-out here comprises all the primes up to 10499 only. On a 3.5 inch magneto-optical disk, having a storage capacity of 128 Mbytes, there is enough room to store the primes up to about 3,000,000,000 in this way.

Compact prime table up to 10500 (to be read horizontally)

32596917119	19221276355	32294916984	27056746064	13260585324
19153906256	11044217692	10628959443	23930632312	27274595010
12929300524	09758853778	21477751664	18735703058	06820532604
01946775235	27961930040	10687629457	28253630548	10613958227
25803963148	26662686226	21859162944	17449165506	06723734372
27325713986	26053724260	09204998354	06548095832	00563096657
30104951352	26603773969	32489161484	18886509697	30344308812
10336150736	06524317784	08657858241	13223332700	17248824849
02496475956	09263457856	27997067788	17534371331	12956543856
18014274691	30375432704	09021230674	17224718372	01007756482
06518293316	01562192512	10744524916	10695737491	00048255592
27610139283	21483511888	27182047424	17255969348	26401637587
02460821844	00154165328	02423076900	10092832833	25850294296
00746875411	02489932592	01755325507	28234041380	26190874626
21544575008	27592828609	00369648472	01896680659	27988934752
00294462083	21785479944	18472241811	32321401632	01688470608
04297614176	10219687506	11113138744	26578275025	08600449332
09084889235	27959757100	26452510928	15137323024	02040873601
06782734684	18550637056	17558145328	18936318161	10776021528
10336339456	06816596588	09070971010	04330713676	18395700226

Hexadecimal Compact Prime Tables

A related method, proposed by Weintraub, is to store the primes by using one hexadecimal symbol to denote the pattern of primes in each interval of the form $[10k, 10k + 9]$, $k = 1, 2, 3, \ldots$ Thus the primes between 20 and 29 are stored as $0101 = 5$, since 21 is composite (0), 23 is prime (1), 27 is composite (0) and 29 is prime (1). The primes between 10 and 99 are in this way stored as F5AE5AD52. (The hexadecimal symbols have the values A=10, B=11, C=12, D=13, E=14 and F=15.) The storage capacity required is 4 bits out of each 10 numbers.

Difference Between Consecutive Primes

The method demonstrated above is not the only means of storing the primes in compact form. An alternative is to store $(p_{i+1} - p_i)/2$. The table on p. 80 shows how large these values can become. This is convenient if we wish to run sequentially through the list of primes (e.g. for trial division by all small primes). The number of bits needed for each prime stored in this manner is 6, 8, or 9, for n up to 10^6, 10^9, and 10^{12}, respectively. Taking the number of primes up to these limits into account, this leads to storage requirements of 59 kbytes, 51 Mbytes, and 42 Gbytes, respectively.

The Number of Primes Below x

The number of primes $\leq x$, $\pi(x)$, is an important number-theoretic function. Thus $\pi(2) = 1$ because we count 2 as the first prime, $\pi(10) = 4$ and $\pi(\sqrt{1000}) = 11$, since there are 11 primes ≤ 31.6. To calculate arithmetically the exact number of primes below a given limit is an extremely complicated and labour-consuming task, as we shall see, and the early computations of $\pi(x)$ were carried out simply by counting the number of primes in existing prime tables. But since those early prime tables were not free from errors, the results of the computations of $\pi(x)$ by this method were somewhat unreliable.

As mentioned above, the existing formulas for $\pi(x)$ are quite complicated. The simplest (but unfortunately also the most labour-consuming) was found by Legendre, and reads

$$1 + \pi(x) = \pi(\sqrt{x}) + \lfloor x \rfloor - \sum_{p_i \leq \sqrt{x}} \left\lfloor \frac{x}{p_i} \right\rfloor + \sum_{p_i < p_j \leq \sqrt{x}} \left\lfloor \frac{x}{p_i p_j} \right\rfloor -$$

$$- \sum_{p_i < p_j < p_k \leq \sqrt{x}} \left\lfloor \frac{x}{p_i p_j p_k} \right\rfloor + \cdots \tag{1.2}$$

where $\lfloor z \rfloor$ denotes the integer part of z. Legendre's formula is almost self-evident. It uses the idea that

$$1 + \text{the number of primes} = \text{the number of all integers} -$$
$$- \text{the number of composites in the interval } [1, x].$$

The term $\lfloor x/p \rfloor$ enumerates the integers divisible by p in this interval, since $\lfloor x/p \rfloor = n$ if and only if $np \leq x < (n+1)p$. Since all composite numbers in the interval $[1, x]$ have *some* prime factor $\leq \sqrt{x}$, there are obviously $\sum_{p_i \leq \sqrt{x}} \lfloor x/p_i \rfloor$ multiples of primes $\leq \sqrt{x}$ in this interval. However, we must not count the multiples $1 \cdot p_i$ as composites; that is why the term $\pi(\sqrt{x})$ has been added. This reasoning accounts for the first three terms in the right-hand-side of the formula. Where do the rest come from? Some of the composites in $[1, x]$ are divisible by *two* of the primes $\leq \sqrt{x}$, p_j as well as p_i. These composites will be counted *twice* when computing the sum $\sum \lfloor x/p_i \rfloor$; since any integer of the form $ap_i p_j$ will count as a multiple of p_i as well as a multiple of p_j. That is why the total has to be corrected by again adding the number of integers having this particular form, which is carried out by the next term in the formula, $\sum \lfloor x/p_i p_j \rfloor$. As a result of this new term, all those integers in $[1, x]$ that happen to be divisible by *three* different primes, all $\leq \sqrt{x}$, now will not be subtracted at all. This omission has to be corrected, which explains the next term in (1.2), and so forth.

In order to understand the formula better, let us, as an example, compute $\pi(100)$ and $\pi(200)$ and check with the number of primes between 100 and 200, $\pi(200) - \pi(100) = 21$ which we have found on p. 3 by the sieve of Eratosthenes.

Legendre's formula gives

$$\pi(100) = \pi(10) + 100 - \left\lfloor \frac{100}{2} \right\rfloor - \left\lfloor \frac{100}{3} \right\rfloor - \left\lfloor \frac{100}{5} \right\rfloor - \left\lfloor \frac{100}{7} \right\rfloor + \left\lfloor \frac{100}{2 \cdot 3} \right\rfloor +$$

$$+ \left\lfloor \frac{100}{2 \cdot 5} \right\rfloor + \left\lfloor \frac{100}{2 \cdot 7} \right\rfloor + \left\lfloor \frac{100}{3 \cdot 5} \right\rfloor + \left\lfloor \frac{100}{3 \cdot 7} \right\rfloor + \left\lfloor \frac{100}{5 \cdot 7} \right\rfloor - \left\lfloor \frac{100}{2 \cdot 3 \cdot 5} \right\rfloor -$$

$$- \left\lfloor \frac{100}{2 \cdot 3 \cdot 7} \right\rfloor - \left\lfloor \frac{100}{2 \cdot 5 \cdot 7} \right\rfloor - \left\lfloor \frac{100}{3 \cdot 5 \cdot 7} \right\rfloor + \left\lfloor \frac{100}{2 \cdot 3 \cdot 5 \cdot 7} \right\rfloor - 1 =$$

$$= 4 + 100 - 50 - 33 - 20 - 14 + 16 + 10 + 7 + 6 + 4 + 2 -$$
$$- 3 - 2 - 1 - 0 + 0 - 1 = 25.$$

When $x = 200$, the primes 11 and 13 also come into play. This gives 66 different terms in all, of which, however, quite a few are equal to 0:

$$\pi(200) = \pi(14) + 200 - 1 - 100 - 66 - 40 - 28 - 18 - 15 + 33 +$$
$$+ 20 + 14 + 9 + 7 + 13 + 9 + 6 + 5 + 5 + 3 + 3 + 2 + 2 +$$
$$+ 1 - 6 - 4 - 3 - 2 - 2 - 1 - 1 - 1 - 1 - 1 - 1 - 1 = 46,$$

finally yielding

$$\pi(200) - \pi(100) = 46 - 25 = 21.$$

Note that this computation agrees with the value found earlier for the number of primes between 100 and 200.

As is obvious from these examples, Legendre's formula as it stands is impracticable for the computation of $\pi(x)$ for large values of x. The large number of terms makes the computation difficult to organize. By various tricks, however, it is possible to collect some of the terms into partial sums and thus arrive at a more feasible calculus. The first efforts in this direction were already made by Legendre himself, who managed to compute $\pi(1,000,000)$. The value he found was not completely correct which tells us something about the difficulties involved in these kinds of computations. Thus Legendre [7] found $\pi(10^6)$ to be 78,526 instead of 78,498. The erroneous prime tables that existed at his time displayed 78,492 primes.

The next important progress in the counting of primes was made by Meissel who, with an efficient modification of Legendre's formula, computed $\pi(x)$ for, among other values, $x = 10^7$, 10^8 and 10^9. However, Meissel's computations are not free from errors. For instance, in 1885 Meissel published the value 50,847,478 for $\pi(10^9)$ and this is probably the most often quoted faulty value in the literature on primes. This error passed unnoticed for a long time and was revealed by D. H. Lehmer only in 1958, the correct value of $\pi(10^9)$ being 50,847,534.