die Summe ändert also ihren Wert nicht, wenn man sie mit der von eins verschiedenen Zahl $e^{4\pi ik/p}$ multipliziert, und damit muß sie verschwinden. Somit ist

$$|\widehat{r}(m)|^2 = \frac{1}{p}e^0 \cdot p = 1$$

für alle m, wir haben also die gewünschte Diffusionseigenschaft.

Die obige Abbildung zeigt den Querschnitt über ein solches Phasengitter, hier für $p = 23$. Entsprechende SCHROEDER-Reflektoren zu den verschiedensten Primzahlen gibt es in vielen Konzertsälen und Opernhäusern, oft allerdings verborgen hinter schalldurchlässigem Material.

MANFRED ROBERT SCHROEDER wurde 1926 in Deutschland geboren. Er studierte Physik an der Universität Göttingen, wo er 1952 promovierte. Danach arbeitete er bei den AT & T Bell Laboratories in Murray Hill, New Jersey auf dem Gebiet der Akustik; diese Arbeit führte unter anderem zu 45 Patenten. 1969 wechselte er als Professor für Akustik an die Universität Göttingen, wo er bis zu seiner Emeritierung lehrte. Er schrieb mehrere Bücher, unter anderem *Number theory in Science and Communication* und *Fractals, Chaos, Power Laws.* Der Inhalt dieses Abschnitts ist kurz im ersten dieser Bücher dargestellt sowie ausführlich in M.R. SCHROEDER: Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion, *J. Acoust. Soc. Am.* **65** (4), 1979

`www.physik3.gwdg.de/~mrs`

Kapitel 7
Primzahlen

§1: Das Sieb des Eratosthenes

Das klassische Verfahren zur Bestimmung aller Primzahlen unterhalb einer bestimmten Schranke geht zurück auf ERATOSTHENES im dritten vorchristlichen Jahrhundert. Es funktioniert folgendermaßen:

Wenn man alle Primzahlen kleiner oder gleich einer Zahl N finden möchte, schreibt man zunächst die Zahlen von Eins bis N in eine Reihe.

Eins ist nach Definition keine Primzahl – für griechische Mathematiker wie EUKLID war die Eins nicht einmal eine Zahl. Also streichen wir die Eins durch.

Ansonsten ist eine Primzahl eine Zahl, die außer der Eins und sich selbst keine Teiler hat. Damit muß zwei eine Primzahl sein.

Die echten Vielfachen von zwei sind natürlich keine Primzahlen; also streichen wir sie durch. Dazu müssen wir nicht von jeder Zahl nachprüfen, ob sie durch zwei teilbar ist, sondern wir streichen einfach nach der Zwei jede zweite Zahl aus der Liste durch.

Die erste nichtdurchgestrichene Zahl der Liste ist dann die Drei. Sie muß eine Primzahl sein, denn hätte sie einen von eins verschiedenen kleineren Teiler, könnte das nur die Zwei sein, und alle Vielfachen von zwei (außer der Zwei selbst) sind bereits durchgestrichen.

Auch die echten Vielfachen der Drei sind keine Primzahlen, werden also durchgestrichen. Auch dazu streichen wir wieder einfach jede dritte Zahl aus der Liste durch, unabhängig davon, ob sie bereits durchgestrichen ist oder nicht. (Alle durch sechs teilbaren Zahlen sind offensichtlich schon durchgestrichen.)

Genauso geht es weiter mit der Fünf *usw.;* nach jedem Durchgang durch die Liste muß offenbar die erste noch nicht durchgestrichene Zahl eine Primzahl sein, denn alle Vielfache von kleineren Primzahlen sind bereits durchgestrichen, und wenn eine Zahl überhaupt einen echten Teiler hat, dann hat sie natürlich auch eine Primzahl als echten Teiler.

Wie lange müssen wir dieses Verfahren durchführen? Wenn eine Zahl x Produkt zweier echt kleinerer Faktoren u, v ist, können u und v nicht beide größer sein als $\sqrt{x}$: Sonst wäre schließlich $x = uv$ größer als x. Also ist einer der beiden Teiler u, v kleiner oder gleich $\sqrt{x}$, so daß x mindestens einen Teiler hat, dessen Quadrat kleiner oder gleich x ist. Damit ist eine zusammengesetzte Zahl x durch mindestens eine Primzahl p teilbar mit $p^2 \leq x$.

Für das Sieb des ERATOSTHENES, angewandt auf die Zahlen von Eins bis N heißt das, daß wir aufhören können, sobald die erste nichtdurchgestrichene Zahl p ein Quadrat $p^2 > N$ hat; denn dann können wir sicher sein, daß jede zusammengesetzte Zahl $x \leq N$ bereits einen kleineren Primteiler als p hat und somit bereits durchgestrichen ist. Die noch nicht durchgestrichenen Zahlen in der Liste sind also Primzahlen.

ERATOSTHENES (Ερατοσθένες) wurde 276 v.Chr. in Cyrene im heutigen Lybien geboren, wo er zunächst von Schülern des Stoikers ZENO ausgebildet wurde. Danach studierte er noch einige Jahre in Athen, bis ihn 245 der Pharao PTOLEMAIOS III als Tutor seines Sohns nach Alexandrien holte. 240 wurde er dort Bibliothekar der berühmten Bibliothek im Museion.
Heute ist er außer durch sein Sieb vor allem durch seine Bestimmung des Erdumfangs bekannt. Er berechnete aber auch die Abstände der Erde von Sonne und Mond und entwickelte einen Kalender, der Schaltjahre enthielt. 194 starb er in Alexandrien, nach einigen Überlieferungen, indem er sich, nachdem er blind geworden war, zu Tode hungerte.

Damit lassen sich leicht von Hand alle Primzahlen bis hundert finden, mit etwas Fleiß auch die bis Tausend, aber sicher nicht die hundertstelligen.

Trotzdem kann uns ERATOSTHENES helfen, zumindest zu zeigen, daß gewissen Zahlen nicht prim sind: Wenn wir Primzahlen in einem Intervall $[a, b]$ suchen, d.h. also Primzahlen p mit

$$a \leq p \leq b\,,$$

so können wir ERATOSTHENES auf dieses Intervall fast genauso anwenden wie gerade eben auf das Intervall $[1, N]$:

Wir gehen aus von einer Liste $p_1, \dots, p_r$ der ersten Primzahlen; dabei wählen wir r so, daß die Chancen auf nicht durch p_r teilbare Zahlen im Intervall $[a, b]$ noch einigermaßen realistisch sind, d.h. wir gehen bis zu einer Primzahl p_r, die ungefähr in der Größenordnung der Intervallänge $b - a$ liegt.

Nun können wir mit jeder der Primzahlen p_i sieben wie im klassischen Fall; wir müssen nur wissen, wo wir anfangen sollen.

Dazu berechnen wir für jedes p_i den Divisionsrest $r_i = a \bmod p_i$. Dann ist $a - r_i$ durch p_i teilbar, liegt allerdings nicht ims Intervall $[a, b]$. Die erste Zahl, die wir streichen müssen, ist also $a - r_i + p_i$, und von da an streichen wir einfach, ohne noch einmal dividieren zu müssen, wie gehabt jede p_i-te Zahl durch.

Was nach r Durchgängen noch übrigbleibt, sind genau die Zahlen aus $[a, b]$, die durch keine der Primzahlen p_i teilbar sind. Sie können zwar noch größere Primteiler haben, aber wichtig ist, daß wir mit minimalem Aufwand für den Großteil aller Zahlen aus $[a, b]$ gesehen haben, daß sie keine Primzahlen sind. Für den Rest brauchen wir andere Verfahren, aber die sind allesamt erheblich aufwendiger als ERATOSTHENES, so daß sich diese erste Reduktion auf jeden Fall lohnt.

§2: Der Fermat-Test

Nach dem kleinen Satz von FERMAT gilt für jede Primzahl p und jede nicht durch p teilbare Zahl a die Formel $a^{p-1} \equiv 1 \bmod p$. Im Umkehrschluß folgt sofort:

Falls für eine natürliche Zahl $1 \leq a \leq p - 1$ gilt $a^{p-1} \not\equiv 1 \bmod p$, kann p keine Primzahl sein.

Beispiel: Ist $F_{20} = 2^{2^{20}} + 1$ eine Primzahl? Falls ja, ist nach dem kleinen Satz von FERMAT insbesondere

$$3^{F_{20}-1} = 1 \bmod F_{20} .$$

Nachrechnen zeigt, daß

$$3^{(F_{20}-1)/2} \not\equiv \pm 1 \bmod F_{20} ,$$

die Zahl ist also nicht prim. (Das „Nachrechnen“ ist bei dieser 315 653-stelligen Zahl natürlich keine Übungsaufgabe für Taschenrechner: 1988 brauchte eine Cray X-MP dazu 82 Stunden, eine Cray-2 immerhin noch zehn; siehe *Math. Comp.* **50** (1988), 261–263. Die anscheinend etwas weltabgewandt lebenden Autoren meinen, dies sei die teuerste bislang produzierte 1-Bit-Information.)

Ein anderes Beispiel, daß sich leicht mit einem Computeralgebrasystem nachrechnen läßt, wäre $M_{67} = 2^{67} - 1$. Hier ist

$$13^{M_{67}-1} \equiv 81\,868\,480\,399\,682\,966\,751 \bmod M_{67} ,$$

also ist auch M_{67} keine Primzahl.

Zur Not auch mit dem Taschenrechner läßt sich

$$F_5 = 2^{2^5} + 1 = 2^{32} + 1 = 2\,863\,311\,531$$

überprüfen: Hier ist zum Beispiel

$$3^{F_5-1} = 3^{2^{32}} \equiv 2\,863\,311\,531 \bmod F_5 ,$$

wie man durch 32-faches Quadrieren modulo F_5 feststellt.

(Allgemein bezeichnet man $F_n = 2^{2^n} + 1$ als die n-te FERMAT-Zahl, da FERMAT 1650 behauptet hatte, er könne beweisen, daß alle diese Zahlen prim seien. $F_0 = 3, F_1 = 5, F_2 = 17, F_3 = 257$ und $F_4 = 65\,537$ sind in der Tat prim, aber 1732 zeigte EULER, daß F_5 durch 641 teilbar ist. Inzwischen ist bekannt, daß F_n für $5 \le n \le 32$ sowie viele andere Werte von n zusammengesetzt ist; FERMATsche Primzahlen mit $n > 4$ sind nicht bekannt.

So einfach es ist, auf diese Weise eine Zahl als zusammengesetzt zu erkennen, so unmöglich ist es, umgekehrt so zu beweisen, daß sie prim ist. So ist beispielsweise

$$18^{322} \equiv 1 \bmod 323 \quad \text{aber} \quad 323 = 17 \cdot 19 .$$

Immerhin gibt es nicht viele $a \le 323$ mit $a^{322} \equiv 1 \bmod 323$: Die einzigen Möglichkeiten sind $a = \pm 1$ und $a = \pm 18$.

Zumindest theoretisch läßt sich der FERMAT-Test allerdings auch ausbauen zu einem echten Primzahlbeweis:

Satz: Ist für zwei natürliche Zahlen p, a zwar $a^{p-1} \equiv 1 \bmod p$, aber für jeden Primteiler q von $p-1$ $a^{(p-1)/q} \not\equiv 1 \bmod p$, so ist p eine Primzahl.

Beweis: Offensichtlich muß dann die Ordnung von a in $(\mathbb{Z}/p\mathbb{Z})^\times$ gleich $p-1$ sein. Wie wir aus Kapitel 1, §7 wissen, hat $(\mathbb{Z}/p\mathbb{Z})^\times$ die Ordnung $\varphi(p)$, und für jede zusammengesetzte Zahl folgt aus der dort angegebenen Formel leicht, daß $\varphi(p) < p-1$ ist. Also muß p prim sein. ∎

Beispiel: Ist $p = 2^{16} + 1$ prim? Hier hat $p - 1 = 2^{16}$ nur die Zwei als Primteiler; nach dem Satz ist p also prim, falls wir eine Zahl a finden können, so daß $a^{p-1} \equiv 1 \bmod p$, aber $a^{(p-1)/2} \not\equiv 1 \bmod p$. Für $a = 2$ sind beide Potenzen eins, für $a = 3$ aber ist die zweite gleich -1. Somit ist p eine Primzahl.

Für FERMAT-Zahlen läßt sich also recht einfach entscheiden, ob sie prim sind oder nicht; bei sonstigen Zahlen hat man das Problem, daß zunächst $p-1$ faktorisiert werden muß. Falls p Primfaktor eines sicheren RSA-Moduls werden soll, im Idealfall also über dreihundert Dezimalstellen haben sollte, ist dies jedoch unrealistisch; hier braucht man alternative Verfahren.

Es kann nicht vorkommen, daß für eine zusammengesetzte Zahl n und alle $1 \le a \le n$ gilt $a^{n-1} \equiv 1 \bmod n$, denn ist p ein Primteiler von n, so ist für jedes Vielfache a von p natürlich auch a^{n-1} durch p teilbar, kann also nicht kongruent eins modulo des Vielfachen n von p sein. Zumindest für die a mit $\mathrm{ggT}(a, n) > 1$ kann die Gleichung also nicht erfüllt sein.

Bei großen Zahlen n mit nur wenigen Primfaktoren ist die Chance, ein solches a zu erwischen, allerdings recht klein; wenn dies die einzigen Gegenbeispiele sind, wird uns der FERMAT-Test also fast immer in die Irre führen.

Definition: Eine natürliche Zahl n heißt CARMICHAEL-Zahl, wenn sie keine Primzahl ist, aber trotzdem für jede natürliche Zahl a mit $\mathrm{ggT}(a,n)=1$ gilt: $a^{n-1}\equiv 1 \bmod n$.

ROBERT DANIEL CARMICHAEL (1879–1967) war ein amerikanischer Mathematiker, der unter anderem Bücher über die Relativitätstheorie, über Zahlentheorie, über Analysis und über Gruppentheorie veröffentlichte. Ab 1915 lehrte er an der University of Illinois.

Lemma: Eine CARMICHAEL-Zahl n ist ein Produkt von mindestens drei paarweise verschiedenen Primzahlen p_i; für jede davon ist p_i-1 ein Teiler von $n-1$.

Beweis: Die Primzerlegung von n sei $\prod_{i=1}^{r} p_i^{e_i}$. Falls $a^{n-1}\equiv 1 \bmod n$ ist, haben wir erst recht $a^{n-1}\equiv 1 \bmod p_i^{e_i}$ für jedes i. Andererseits wissen wir, daß die Ordnung von a in $\mathbb{Z}/p_i^{e_i}\mathbb{Z}$ die Gruppenordnung $\varphi(p_i^{e_i})=p_i^{e_i-1}(p_i-1)$ teilt, außerdem muß sie $n-1$ teilen. Damit kann sie auf keinen Fall ein Teiler von $p_i^{e_i-1}$ sein, denn das ist ein Teiler von n. Also muß sie ein Teiler von p_i-1 sein.

Für $e_i>1$ kann dies aber nicht für alle zu p_i primen a der Fall sein, denn beispielsweise ist $(p_i+1)^{p_i}\equiv 1 \bmod p_i^2$, so daß nicht gleichzeitig $(p_i+1)^{p_i-1}\equiv 1 \bmod p_i^2$ sein kann, denn die beiden Potenzen unterscheiden sich um den Faktor $p_i+1\not\equiv 1 \bmod p_i^2$. Daher muß $e_i=1$ sein, und da es in $(\mathbb{Z}/p_i\mathbb{Z})^\times$ Elemente der Ordnung p_i-1 gibt, ist p_i-1 ein Teiler von $n-1$.

Schließlich müssen wir uns noch überlegen, daß $r\geq 3$ ist. Wäre $n=pq$ nur das Produkt zweier Primzahlen, müßte

$$n-1=pq-1=(p-1)q+(q-1)$$

sowohl durch $p-1$ als auch durch $q-1$ teilbar sein, also müßten $p-1$ und $q-1$ durcheinander teilbar sein, d.h. $p=q$, was wir bereits augeschlossen haben. ■

Als Beispiel können wir ein Produkt $n=(6t+1)(12t+1)(18t+1)$ mit drei primen Faktoren betrachten, z.B.

$$1729=7\times 13\times 19 \quad \text{für } t=1 \quad \text{oder} \quad 294409=37\times 73\times 109$$

für $t=6$. Hier ist $n-1=1296t^3+396t^2+36t=36t\cdot(36t^2+11t+1)$ offensichtlich durch $6t$, $12t$ und $18t$ teilbar, n ist also eine CARMICHAEL-Zahl.

Die kleinste CARMICHAEL-Zahl ist $561=3\cdot 11\cdot 17$; wie man inzwischen weiß, gibt es unendlich viele CARMICHAEL-Zahlen, auch wenn sie ziemlich selten sind.

Für große Zahlen p wird es zunehmend unwahrscheinlich, daß sie auch nur für ein a den FERMAT-Test bestehen, ohne Primzahl zu sein. Rechnungen von

SU HEE KIM, CARL POMERANCE: The probability that a Random Probable Prime is Composite, *Math. Comp.* **53** (1989), 721–741

geben folgende obere Schranke für die Fehlerwahrscheinlichkeit ε:

$p\approx 10^{60}$	10^{70}	10^{80}	10^{90}	10^{100}
$\varepsilon\leq 7{,}16\cdot 10^{-2}$	$2{,}87\cdot 10^{-3}$	$8{,}46\cdot 10^{-5}$	$1{,}70\cdot 10^{-6}$	$2{,}77\cdot 10^{-8}$

$p\approx 10^{120}$	10^{140}	10^{160}	10^{180}	10^{200}
$\varepsilon\leq 5{,}28\cdot 10^{-12}$	$1{,}08\cdot 10^{-15}$	$1{,}81\cdot 10^{-19}$	$2{,}76\cdot 10^{-23}$	$3{,}85\cdot 10^{-27}$

(Sie geben natürlich auch eine allgemeine Formel an, jedoch ist diese zu grausam zum Abtippen.)

Selbst wenn wir noch mit 1024-Bit-Moduln arbeiten und somit etwa 155-stellige Primzahlen brauchen, liegt also die Fehlerwahrscheinlichkeit bei nur etwa 10^{-15}; falls man sie erniedrigen möchte, testet man einfach mit mehreren zufällig gewählten Basen und hat dann etwa bei zwei verschiedenen Basen eine Wahrscheinlichkeit von höchsten etwa 10^{-30}, daß beide Tests das falsche Ergebnis liefern. Dies sollte für die meisten Anwendungen genügen: Die Bundesnetzagentur empfiehlt bei probabilistischen Primzahltests eine Irrtumswahrscheinlichkeit von höchsten $2^{-80}\approx 8{,}27\cdot 10^{-25}$ zuzulassen, die hier deutlich unterschritten wäre. Ab etwa zweihundertstelligen Primzahlen reicht sogar bereits ein einziger FERMAT-Test.

Einige Leute reden bei Zahlen, die einen FERMAT-Test bestanden haben, von „wahrscheinlichen Primzahlen“. Das ist natürlich Unsinn: Eine Zahl ist entweder *sicher* prim oder *sicher* zusammengesetzt; für Wahrscheinlichkeiten gibt es hier keinen Spielraum. Besser ist der ebenfalls gelegent zu höherende Ausdruck „industrial grade primes", also „Industrieprimzahlen", der ausdrücken soll, daß wir zwar nicht *bewiesen* haben, daß die Zahl wirklich prim ist, daß sie aber für industrielle Anwendungen gut genug ist.

§3: Der Test von Miller und Rabin

Der Test von MILLER und RABIN ist eine etwas strengere Version des Tests von FERMAT: Um zu testen, ob p eine Primzahl sein kann, schreiben wir $p-1$ zunächst als Produkt $2^n u$ einer Zweierpotenz und einer ungeraden Zahl; sodann berechnen wir $a^u \bmod p$. Falls wir das Ergebnis eins erhalten, ist erst recht $a^{p-1} \equiv 1 \bmod p$, und wir können nicht folgern, daß p zusammengesetzt ist.

Andernfalls quadrieren wir das Ergebnis bis zu n-mal modulo p. Falls dabei nie eine Eins erscheint, folgt nach FERMAT, daß p zusammengesetzt ist. Falls vor der ersten Eins eine von -1 (*bzw.* $p-1$) verschiedene Zahl erscheint, folgt das auch, denn im Körper $\mathbb{F}_p$ hat die Eins nur die beiden Quadratwurzeln ± 1. In allen anderen Fällen erfahren wir nicht mehr als bei FERMAT.

Algorithmisch funktioniert der Test also folgendermaßen:

Schritt 0: Wähle ein zufälliges a, schreibe $p-1 = 2^n u$ mit einer ungeraden Zahl u und berechne $b = a^u \bmod p$. Falls dies gleich Eins ist, endet der Algorithmus und wir konnten nicht zeigen, daß p eine zusammengesetzte Zahl ist; sie kann prim sein.

Schritt $i, 1 \le i \le n$: Falls $b \equiv -1 \bmod p$, endet der Algorithmus und wir können nicht ausschließen, daß p prim ist. Falls $b = 1$ ist (was frühestens im zweiten Schritt der Fall sein kann), ist p zusammengesetzt und der Algorithmus endet. Andernfalls wird b durch $b^2 \bmod p$ ersetzt und es geht weiter mit Schritt $i+1$.

Schritt $n+1$: Der Algorithmus endet mit dem Ergebnis, daß p zusammengesetzt ist.

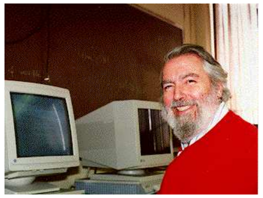
Beispiel: Ist 247 eine Primzahl? Wir wählen $a = 77$, und da $77^{246} \bmod 247 = 1$ ist, können wir mit FERMAT nicht ausschließen, daß 247 prim ist. Da aber $77^{123} \bmod 247 = 77$ ist, sagt uns der Algorithmus von MILLER und RABIN im zweiten Schritt, wenn wir $77^2 \equiv 1 \bmod 247$ betrachten, daß die Zahl zusammengesetzt sein muß.

Hätten wir allerdings mit $a = 87$ gearbeitet, hätten wir im nullten Schritt $87^{123} \equiv 1 \bmod 247$ berechnet und hätten $247 = 13 \cdot 19$ nicht als zusammengesetzt erkannt.

GARY L. MILLER entwickelte diesen Test 1975 im Rahmen seiner Dissertation (in Informatik) an der Universität von Berkeley. Dabei ging es ihm nicht um einen probabilistischen Test, sondern um einen Test, der immer die richtige Antwort liefert. Er konnte zeigen, daß dies hier beim Test von hinreichend vielen geeigneten Basen der Fall ist **vorausgesetzt** die bis heute immer noch offene verallgemeinerte RIEMANN-Vermutung ist richtig. Er lehrte später zunächst einige Jahre an der University of Waterloo, inzwischen an der Carnegie Mellon University. Seine späteren Arbeiten stammen hauptsächlich aus dem Gebiet der rechnerischen Geometrie. `www.cs.icmu.edu/~glmiller`

MICHAEL O. RABIN wurde 1931 in Breslau geboren. Die Familie wanderte nach Israel aus, wo er an der hebräischen Universität von Jerusalem Mathematik studierte. Nach seinem Diplom 1953 ging er nach Princeton, wo er 1957 promovierte. Seit 1958 lehrt er an der hebräischen Universität, wo er unter anderem auch Dekan der mathematischen Fakultät und Rektor war. Seit 1983 ist er zusätzlich Inhaber des THOMAS J. WATSON-Lehrstuhls für Informatik an der Harvard University. Seine Forschungen, für die er u.a. 1976 den TURING-Preis erhielt, beschäftigen sich mit der Komplexität mathematischer Operationen und der Sicherheit von Informationssystemen. Seine home page in Harvard ist zu finden unter `www.seas.harvard.edu/ourfaculty/profile/Michael_Rabin` .

Anscheindend wurde der Test von MILLER und RABIN bereits 1974, also vor MILLERs Veröffentlichung, von SELFRIDGE verwendet; daher sieht man gelegentlich auch die korrektete Bezeichnung *Test von* MILLER, RABIN *und* SELFRIDGE.



Der amerikanische Mathematiker JOHN L. SELFRIDGE promivierte 1958 an der University of California in Los Angelos. Bis zu seiner Emeritierung lehrte er an der Northern Illinois University. Seine Arbeiten befassen sich vor allem mit der analytischen sowie der konstruktiven Zahlentheorie. Vierzehn davon schrieb er mit PAUL ERDŐS. `math.niu.edu/faculty/index.php?cmd=detail&id=91`

§4: Der Test von Agrawal, Kayal und Saxena

Im August 2002 stellten MANINDRA AGRAWAL, NEERAJ KAYAL und NITIN SAXENA, zwei Bachelor-Studenten am Indian Institute of Technology in Kanpur und ihr Professor, einen Primzahltest vor, der ebenfalls auf dem kleinen Satz von FERMAT beruht, aber (natürlich auf Kosten eines erheblich größeren Aufwands) immer die richtige Antwort liefert; er ist inzwischen erschienen in

> MANINDRA AGRAWAL, NEERAJ KAYAL, NITIN SAXENA: PRIMES is in P, *Annals of Mathematics* **160** (2004), 781-793.

Selbstverständlich war dies nicht der erste Primzahltest, der deutlich schneller als Probedivisionen zeigt, ob eine gegebene Zahl prim ist oder nicht; es ist auch bei weitem nicht der schnellste solche Test. Er hat aber gegenüber anderen solchen Tests zwei Besonderheiten:

1. Zu seinem Verständnis ist – nach einigen in der letzten Zeit gefundenen Vereinfachungen –nur elementare Zahlentheorie notwendig.
2. Es ist der bislang einzige Test, von dem man beweisen kann, daß seine Laufzeit für n-stellige Zahlen durch ein Polynom in n begrenzt werden kann.

Für uns ist vor allem der erste Punkt wichtig; der zweite ist zwar ein für Komplexitätstheoretiker sehr interessantes Ergebnis, hat aber keinerlei praktische Bedeutung: Im Buch

> VICTOR SHOUP: A computational Introduction to Number Theory and Algebra, *Cambridge University Press,* 2005,

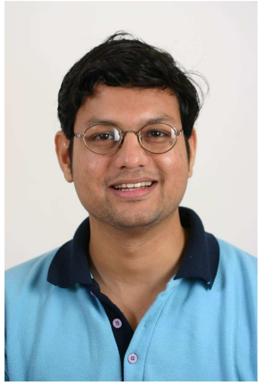
dem dieser Paragraph im wesentlichen folgt, argumentiert SHOUP, daß alternative Algorithmen, so man sich auf Zahlen von weniger als 2^{256} Bit

MANINDRA AGRAWAL erhielt 1986 seinen BTech und 1991 seinen PhD in Informatik am Indian Institute of Technology in Kanpur, wo er –abgesehen von Gastaufenthalten in Madras, Ulm, Princeton und Singapur – seither als Professor lehrt. Seine Arbeiten befassen sich hauptsächlich mit der Komplexität von Schaltungen und von Algorithmen. Für die Arbeit mit KAYAL und SAXENA erhielt er gemeinsam mit diesen unter anderem den GÖDEL-Preis 2006 für die besten Zeitschriftenveröffentlichung auf dem Gebiet der Theoretischen Informatik.

`http://www.cse.iitk.ac.in/users/manindra/`

NEERAJ KAYAL wurde 1979 geboren. Er erhielt 2002 seinen BTech und 2006 seinen PhD bei MANINDRA AGRAWAL am Indian Institute of Technology in Kanpur. Derzeit arbeitet er am Institute for Advances Study in Princeton, wo er bereits im akademischen Jahr 2003/2004 als *visiting student research collaborator* war. Neuere Arbeiten beschäftigen sich mit der Komplexität des Isomorphieproblems bei endlichen Ringen sowie der Lösbarkeit von bivariaten Polynomgleichungen über endlichen Körpern.

`http://www.math.ias.edu/~kayaln/`



NITIN SAXENA wurde 1981 geboren. Er erhielt 2002 seinen Bachelor of Technology und 2006 seinen PhD bei MANINDRA AGRAWAL am Indian Institute of Technology in Kanpur. Während der Arbeit an seiner Dissertation über die Anwendung von Ringhomomorphismen auf Fragen der Komplexitätstheorie besuchte er jeweils ein Jahr lang die Universitäten Princeton und Singapur. Derzeit arbeitet er als Postdoc in der Gruppe *Quantum Computing and Advanced Systems Research* am Centrum voor Wiskunde en Informatica in Amsterdam. Sein Interesse gilt für algorithmischen Verfahren der Algebra und Zahlentheorie sowie Fragen der Komplexitätstheorie.

`http://homepages.cwi.nl/~ns/`

beschränkt, durch eine vergleichbare Schranke abgeschätzt werden können, und natürlich sind die Zahlen, mit denen wir es üblicherweise zu tun haben, deutlich kleiner. In der Praxis sind die alternativen Algorithmen deutlich schneller.

(2^{256} liegt knapp über 10^{77}; derzeitige Schätzungen für die Anzahl der Nukleonen im Universum liegen bei etwa 10^{80}. Damit ist klar, daß kein Computer, der mit irgendeiner Art von heute üblicher Technologie arbeitet, je eine solche Zahl speichern kann, geschweige denn damit rechnen.)

Im folgenden wird es daher nur um eine mathematische Betrachtung des Algorithmus von AGRAWAL, KAYAL und SAXENA gehen; für einen (kurzen und elementaren) Beweis der Komplexitätsaussage sei beispielsweise auf das zitierte Buch von SHOUP verwiesen.

Die Grundidee des Algorithmus steckt im folgenden

Satz: $n > 1$ sei eine natürliche Zahl und $a \in \mathbb{N}$ sei dazu teilerfremd. n ist genau dann prim, wenn im Polynomring über $\mathbb{Z}/N$ gilt:

$$(X + a)^n = X^n + a\,.$$

Beweis: Nach dem binomischen Lehrsatz ist

$$(X + a)^n = X^n + a^n + \sum_{i=1}^{n-1} \binom{n}{i} a^i X^{n-i}\,.$$

Für eine Primzahl n gilt nach dem kleinen Satz von FERMAT in $\mathbb{Z}/n$ die Gleichung $a^n = a$. Außerdem ist für $1 \le i \le n-1$ der Binomialkoeffizient

$$\binom{n}{i} = \frac{n(n-1)\cdots(n-i+1)}{i!}$$

durch n teilbar, da n Faktor des Zählers, nicht aber des Nenners ist. Somit verschwinden in $\mathbb{Z}/n$ alle diese Binomialkoeffizienten, und die Gleichung aus dem Satz ist bewiesen.

Umgekehrt sei n eine zusammengesetzte Zahl und p ein Primteiler von n. Genauer sei $n = p^k m$ mit einer zu p teilerfremden Zahl m. Dann ist der Zähler von $\binom{n}{p}$ genau durch p^k teilbar, denn die Faktoren $(n-1), \dots, (n-p-1)$ sind allesamt teilerfremd zu p, und der Nenner ist genau durch p teilbar. Somit ist $\binom{n}{p}$ zwar durch p^{k-1} teilbar, nicht aber durch p^k und damit erst recht nicht durch n. Wenn wir $(X + a)^n$ über $\mathbb{Z}/n$ ausmultiplizieren, kann daher der Summand $\binom{n}{p} a^p X^{n-p}$ nicht verschwinden, und damit kann die Gleichung aus dem Satz nicht gelten. ■

In dieser Form führt der Satz allerdings noch nicht zu einem praktikablen Primzahltest: Das Ausmultiplizieren von $(X+a)^n$ führt schließlich auf $n+1$ Summanden, der Aufwand ist also proportional zu n und damit vergleichbar damit, daß wir für jede natürliche Zahl $1 < m < n$ nachprüfen, ob n ohne Rest durch m teilbar ist. Die wesentliche neue Idee von AGRAWAL, KAYAL und SAXENA besteht darin zu zeigen, daß es bereits reicht, Gleichungen der im Satz genannten Art modulo einem geeigneten Polynom $X^r - 1$ mit einem relativ kleinen Grad r nachzuprüfen.

Konkret geht ihr Algorithmus folgendermaßen vor:

n sei die zu testende natürliche Zahl und $\ell(n) = [\log_2 n] + 1$ die Anzahl ihrer Binärziffern.

1. Schritt: Stelle sicher, daß n keine Potenz einer anderen natürlichen Zahl ist.

Das läßt sich beispielsweise dadurch bewerkstelligen, daß man die Quadratwurzel, Kubikwurzel *usw.* von n soweit ausrechnet bis man erkennt, daß es sich um keine natürliche Zahl handelt. Der ungünstigste Fall ist offenbar der, daß n eine Zweierpotenz sein könnte; man muß also bis zur $[\log_2 n]$-ten Wurzel gehen.

2. Schritt: Finde die kleinste natürliche Zahl $r > 1$ mit der Eigenschaft, daß entweder $\mathrm{ggT}(n, r) > 1$ ist oder aber $\mathrm{ggT}(n, r) = 1$ ist und $n \bmod r$ in $(\mathbb{Z}/r)^\times$ eine größere Ordnung als $4\ell(n)^2$ hat.

Dies geschieht einfach dadurch, daß man die Zahlen $r = 2, 3, \dots$ allesamt durchprobiert, bis zum ersten mal eine der beiden Bedingungen erfüllt ist. Die Bedingung über die Ordnung der Restklasse von n in $(\mathbb{Z}/r)^\times$ prüft man nach, indem man nacheinander ihre Potenzen ausrechnet, bis man entweder eine Eins gefunden hat oder aber der Exponent größer als $4\ell(n)^2$ ist.

3. Schritt: Falls $r = n$, ist n prim und der Algorithmus endet.

In der Tat: Dann haben wir für alle $r < n$ überprüft, daß $\mathrm{ggT}(n, r) = 1$ ist. Wenn der Algorithmus etwas taugt, darf er natürlich höchstens für sehr kleine Werte von n mit diesem Schritt enden.

4. Schritt: Falls im zweiten Schritt ein r gefunden wurde, für das der ggT von n und r größer als eins ist, muß n zusammengesetzt sein und der Algorithmus endet.

Denn dann haben wir einen Teiler von n gefunden.

5. Schritt: Teste für $j = 1, \dots, \ell \underset{\text{def}}{=} 2\ell(n)[\sqrt{r}] + 1$, ob über $\mathbb{Z}/n$

$$(X + j)^n \equiv X^n + j \bmod (X^r - 1)\,.$$

Sobald ein j gefunden wird, für das dies nicht erfüllt ist, endet der Algorithmus mit dem Ergebnis n *ist zusammengesetzt.*

Falls nämlich n eine Primzahl ist, stimmen $(X + j)^n$ und $X^n + j$ als Polynome mit Koeffizienten aus $\mathbb{Z}/n$ nach obigem Satz überein, sind also erst recht auch gleich modulo $(X^r - 1)$.

6. Schritt: Wenn alle Tests im fünften Schritt bestanden sind, ist n eine Primzahl.

Dies zu beweisen ist die Hauptarbeit dieses Paragraphen.

Nach den Kommentaren zu den einzelnen Schritten ist klar, daß der Algorithmus für eine Primzahl n stets das richtige Ergebnis liefert; wir müssen zeigen, daß er auch zusammengesetzte Zahlen stets erkennt.

Sei also n eine zusammengesetzte Zahl. Falls n Potenz einer anderen natürlichen Zahl ist, wird dies im ersten Schritt erkannt; wir können und werden im folgenden daher annehmen, daß dies nicht der Fall ist.

Das r aus dem zweiten Schritt ist auf jeden Fall echt kleiner als n, denn als zusammengesetzte Zahl hat n insbesondere einen Teiler $r < n$. Der Algorithmus kann daher nicht im dritten Schritt mit der Antwort „n ist prim“ enden. Falls er im vierten Schritt endet, lieferte der zweite Schritt einen Teiler von n, und wir erhalten die richtige Antwort „n ist zusammengesetzt“.

Für den Rest des Paragraphen können wir somit annehmen, daß der zweite Schritt auf ein r führte, für das $\mathrm{ggT}(n, r) = 1$ ist. Wir müssen zeigen, daß einer der Tests im fünften Schritt scheitert, daß es also eine natürliche Zahl j gibt mit

$$1 \le j \le \ell \quad \text{und} \quad (X + j)^n \not\equiv X^n + j \bmod (X^r - 1) \text{ in } \mathbb{Z}/n[X]\,.$$

Wir nehmen an, das sei nicht der Fall, und betrachten einen Primteiler p von n. Dieser muß größer als r sein, denn sonst hätte der Algorithmus bereits mit dem vierten Schritt spätestens bei $r = p$ geendet.

Jede Kongruenz modulo n ist erst recht eine Kongruenz modulo p; wir können daher davon ausgehen, daß für alle j mit $1 \le j \le \ell$ gilt

$$(X + j)^n \equiv X^n + j \bmod (X^r - 1) \text{ in } \mathbb{F}_p[X]\,.$$

Wenn wir zum Faktorring $R = \mathbb{F}_p[X]/(X^r - 1)$ übergehen, ist dort also

$$(X + j)^n = X^n + j \quad \text{falls} \quad 1 \le j \le \ell\,.$$

Um diese seltsame Relation genauer zu untersuchen, betrachten wir für jede zu r teilerfremde natürliche Zahl k die Abbildung

$$\widehat{\sigma}_k \colon \begin{cases} \mathbb{F}_p[X] \to R \\ \qquad g \mapsto g(X^k) \bmod (X^r - 1) \end{cases},$$

die in jedem Polynom g die Variable X überall durch X^k ersetzt.

Lemma: $\widehat{\sigma}_k$ ist surjektiv und sein Kern besteht genau aus den Vielfachen des Polynoms $X^r - 1$.

Beweis: Wir betrachten $\widehat{\sigma}_k$ nur für Indizes k, die zu r teilerfremd sind. Zu jedem solchen Index gibt es daher ein k', so daß $kk' \equiv 1 \bmod r$ ist, und modulo $X^r - 1$ ist damit $X^{kk'} \equiv X$. Für ein beliebiges Polynom $g \in \mathbb{F}_p[X]$ und $h(X) = g(X^{k'})$ ist daher in R

$$\widehat{\sigma}_k(h) = h(X^k) = g(X^{kk'}) = g(X) = g\,,$$

die Abbildung ist also surjektiv.

Was ihren Kern betrifft, so enthält er auf jeden Fall $X^r - 1$ und alle seine Vielfachen, denn

$$\widehat{\sigma}_k(X^r - 1) = (X^{kr} - 1) \bmod (X^r - 1) = 1^k - 1 = 0\,,$$

da $X^r \equiv 1 \bmod (X^r - 1)$.

Umgekehrt sei g irgendein Polynom aus dem Kern von $\widehat{\sigma}_k$. Dann ist das Polynom $h(X) = g(X^k)$ modulo $X^r - 1$ gleich dem Nullpolynom, ist also ein Vielfaches von $X^r - 1$. Konkret sei $h = (X^r - 1)f$. Im Faktorring R ist dann

$$g(X) = g(X^{kk'}) = h(X^{k'}) = (X^{k'r} - 1)f(X^{k'}) = 0\,,$$

denn wegen $X^r = 1$ in R ist dort $X^{k'r} - 1 = 0$.

In $\mathbb{F}_p[X]$ muß $g(X)$ daher ein Vielfaches von $X^r - 1$ sein, und genau das war die Behauptung über den Kern von $\widehat{\sigma}_k$.

■

Da alle Vielfachen von $X^r - 1$ im Kern von $\widehat{\sigma}_k$ liegen, definiert $\widehat{\sigma}_k$ eine Abbildung σ_k von R nach R, die jedem Polynom $g \bmod (X^r - 1)$ aus R das Element $\widehat{\sigma}_k(g)$ zuordnet; nach dem gerade bewiesenen Lemma hängt dieses wirklich nur von der Restklasse $g \bmod (X^r - 1)$ ab. Außerdem zeigt das Lemma, daß σ_k sowohl surjektiv als auch injektiv ist, denn der Kern von $\widehat{\sigma}_k$ ist gleich dem Kern der Restklassenabbildung von $\mathbb{F}_p[X]$ nach R. Damit ist σ_k ein bijektiver Homomorphismus von R nach R, ein sogenannter *Automorphismus* von R. Wir haben damit für jede zu r teilerfremde natürliche Zahl k einen Automorphismus $\sigma_k\colon R \to R$, der jedem Polynom in X das entsprechende Polynom in X^k zuordnet. Da wir in R rechnen, werden natürlich alle Polynome modulo $X^r - 1$ betrachtet.

Unmittelbar aus der Definition folgt, daß die verschiedenen Automorphismen σ_k miteinander kommutieren; genauer ist

$$\sigma_k \circ \sigma_{k'} = \sigma_{k'} \circ \sigma_k = \sigma_{kk'}\,,$$

denn in allen drei Fällen wird im Endeffekt X durch $X^{kk'}$ ersetzt.

Speziell für das Element $X + j$ aus R ist $\sigma_k(X + j) = X^k + j$. Für $j = 1, \ldots, \ell$ ist andererseits auch

$$\sigma_k(X + j) = (X + j)^k\,,$$

denn für diese j wurde ja nach unserer Annahme der Test im fünften Schritt bestanden.

Wir wollen genauer untersuchen, wann die Gleichung $\sigma_k(f) = f$ erfüllt ist. Dazu definieren zwei Arten von Mengen:

$$C(f) = \left\{k \in (\mathbb{Z}/r)^\times \;\middle|\; \sigma_k(f) = f^k\right\} \quad \text{für alle } f \in R \qquad \text{und}$$
$$D(k) = \left\{f \in R \;\middle|\; \sigma_k(f) = f^k\right\} \quad \text{für alle } k \in (\mathbb{Z}/r)^\times$$

Beide Mengen enthalten mit zwei Elementen auch deren Produkt, denn für zwei Elemente $k, k' \in C(f)$ ist

$$\sigma_{kk'}(f) = \sigma_k(f)\sigma_{k'}(f) = \sigma_k\big(\sigma_{k'}(f)\big) = \sigma_k(f^{k'}) = \sigma_k(f)^{k'} = f^{kk'}\,,$$

und für $f, g \in D(k)$ ist

$$\sigma_k(fg) = \sigma_k(f)\sigma_k(g) = f^k g^k = (fg)^k\,.$$

Der Rest des Beweises besteht darin, daß wir die „Größe“ der Menge $D(n)$ auf zwei verschiedene Weisen abschätzen und daraus einen Widerspruch herleiten zur Annahme, daß n zusammengesetzt ist, aber trotzdem vom Algorithmus als Primzahl klassifiziert wird. Wir definieren zunächst zwei neue Zahlen:

- s sei die Ordnung der Restklasse von p in $(\mathbb{Z}/r)^\times$. Dann ist r ein Teiler von $p^s - 1$, denn $p^s \equiv 1 \bmod r$.
- t sei die Ordnung der von den Restklassen von p und n erzeugten Untergruppe von $(\mathbb{Z}/r)^\times$, d.h also die Ordnung der kleinsten Untergruppe, die beide Restklassen enthält. Da diese Untergruppe insbesondere die Restklasse von p und deren Potenzen enthält, ist t ein Vielfaches von s.

Als nächstes betrachten wir einen Körper K mit p^s Elementen. Einen solchen Körper kann man konstruieren, indem man den Vektorraum $\mathbb{F}_p^s$ identifiziert mit dem Vektorraum aller Polynome vom Grad kleiner s mit Koeffizienten aus $\mathbb{F}_p$ und dort eine Multiplikation einführt, die zwei Polynomen deren Produkt modulo einem festen irreduziblen Polynom vom Grad s über $\mathbb{F}_p$ zuordnet. Man kann zeigen (siehe Algebra-Vorlesung oder entsprechendes Lehrbuch), daß es für jedes s ein solches Polynom gibt, und daß zwei verschiedene irreduzible Polynome vom Grad s zu isomorphen Körpern führen.

Aus Kapitel 1 wissen wir, daß die multiplikative Gruppe jedes endlichen Körpers zyklisch ist; $K^\times$ ist also eine zyklische Gruppe der Ordnung $p^s - 1$. Diese Zahl ist, wie wir gerade gesehen haben, ein Vielfaches von r; somit gibt es in $K^\times$ (mindestens) ein Element ζ der Ordnung r. Für irgendein solches Element definieren wir einen Homomorphismus

$$\widehat{\tau}\colon \begin{cases} \mathbb{F}_p[X] \to K \\ \qquad g \mapsto g(\zeta) \end{cases} .$$

Da $\widehat{\tau}(X^r - 1) = \zeta^r - 1$ verschwindet, induziert $\widehat{\tau}$ einen Ringhomomorphismus $\tau\colon R \to K$. Die angekündigten Abschätzungen der „Größe" von $D(n)$ beziehen sich auf die Mächtigkeit der Menge $S = \tau\bigl(D(n)\bigr)$:

Lemma: $S = \tau\bigl(D(n)\bigr)$ hat höchstens $n^{2[\sqrt{t}]}$ Elemente.

Beweis: Wir gehen davon aus, daß n weder eine Primzahl noch eine Primzahlpotenz ist; daher gibt es außer dem Primteiler p noch mindestens einen weiteren Primteiler q. Wenn wir (in $\mathbb{N}$) Potenzen der Form $n^u p^v$ und $n^{u'} p^{v'}$ mit $u, u', v, v' \in \mathbb{N}_0$ betrachten, sind diese daher genau dann gleich, wenn $(u, v) = (u', v')$ ist: Ist nämlich $u \neq u'$, so tritt q in der Primzerlegung der beiden Elemente mit verschiedenen Exponenten auf, und ist $u = u'$, aber $v \neq v'$, so gilt entsprechendes für p. Daher hat die Menge

$$I = \bigl\{ n^u p^v \bigm| 0 \le u, v \le [\sqrt{t}] \bigr\}$$

mindestens $\bigl([\sqrt{t}]+1\bigr)^2$ Elemente, und diese Zahl ist offensichtlich größer als t.

Nun war aber t definiert als die Ordnung der Untergruppe von $(\mathbb{Z}/r)^\times$, die von den Restklassen von n und von p erzeugt wird; daher muß es mindestens zwei Elemente

$$k = n^u p^v \quad \text{und} \quad k' = n^{u'} p^{v'}$$

aus I geben, die dieselbe Restklasse in $(\mathbb{Z}/r)^\times$ definieren, für die also gilt: $k \equiv k' \bmod r$. Da die Exponenten u, u', v, v' höchstens gleich $[\sqrt{t}]$ sind und p ein Teiler von n ist, können wir $n^{2[\sqrt{t}]}$ als (sehr grobe) obere Schranke für k und k' nehmen.

Nun sei $f \in R$ ein Element von $D(n)$. Nach Definition der Mengen $C(f)$ und $D(n)$ ist dann auch n ein Element von $C(f)$. Außerdem enthält $C(f)$ stets die Eins und nach dem kleinen Satz von FERMAT auch die Primzahl p, denn Potenzieren mit p ist über $\mathbb{F}_p$ ein Homomorphismus. Da mit zwei Elementen stets auch deren Produkt in $C(f)$ liegt, liegen daher die Restklassen modulo r aller Elemente von I in $C(f)$. Insbesondere sind daher k und k' Elemente von $C(f)$, d.h.

$$\sigma_k(f) = f^k \quad \text{und} \quad \sigma_{k'}(f) = f^{k'} .$$

Wegen $k \equiv k' \bmod r$ ist aber σ_k dieselbe Abbildung wie $\sigma_{k'}$; daher ist $f^k = f^{k'}$ für jedes $f \in D(n)$. Somit sind die Bilder $\tau(f)$ aller $f \in R$ Nullstellen des Polynomes $X^k - X^{k'}$. Dessen Grad ist das Maximum von k und k', und da $\tau(f)$ im Körper K liegt, gibt es höchstens so viele Nullstellen, wie der Grad angibt. Aufgrund der obigen Abschätzung für k und k' hat das Polynom daher höchstens $n^{2[\sqrt{t}]}$ Nullstellen, und damit kann auch S nicht mehr Elemente enthalten. ■

Als untere Grenze für die Elementanzahl von S erhalten wir

Lemma: S enthält mindestens $2^{\min(t,\ell)} - 1$ Elemente.

Beweis: Wegen der bestandenen Tests in Schritt 5 liegt $\tau(X + j)$ in $D(n)$ für $j = 1, \dots, \ell$. Da $p > r > t \ge m$ ist, sind die Zahlen von 1 bis m auch modulo p paarweise verschieden. Die Teilmenge

$$P = \Bigl\{ \prod_{j=1}^{m} (X + j)^{e_j} \Bigm| e_j \in \{0, 1\} \text{ und } \sum_{j=1}^{m} e_j < m \Bigr\}$$

von $\mathbb{F}_p[X]$ enthält daher $2^m - 1$ Polynome.

Aus diesen Polynomen können wir Elemente von R *bzw.* K machen, indem wir für die Variable X die Restklasse $\eta = X \bmod (X^r - 1)$ *bzw.* das oben gewählte Element ζ der Ordnung r einsetzen; wir erhalten Teilmengen

$$P(\eta) = \bigl\{ f(\eta) \bigm| f \in P \bigr\} \subseteq R \quad \text{und} \quad P(\zeta) = \bigl\{ f(\zeta) \bigm| f \in P \bigr\} \subseteq K .$$

Da sowohl n als auch p in $D(n)$ liegen und mit zwei Elementen auch deren Produkt, liegt $P(\eta)$ in $D(n)$ und damit $\tau\bigl(P(\eta)\bigr) = P(\zeta)$ in S.

Das Lemma ist daher bewiesen, sobald wir gezeigt haben, daß $P(\zeta)$ mindestens $2^m - 1$ Elemente enthält.

Falls dies nicht der Fall wäre, müßte es in P zwei verschiedene Polynome g und h geben, für die $g(\zeta) = h(\zeta)$ wäre. Wir müssen also zeigen, daß $g(\zeta) = h(\zeta)$ nur dann gelten kann, wenn $g = h$ ist.

Wie im vorigen Lemma folgt, da $1, p$ und n alle drei sowohl in $C\big(g(\eta)\big)$ als auch in $C\big(h(\eta)\big)$ liegen, daß alle natürlichen Zahlen k der Form $k = n^u p^v$ in diesen beiden Mengen liegen.

Da $g(\zeta) = h(\zeta)$, gilt für jedes solche k

$$\begin{aligned} 0 = g(\zeta)^k - h(\zeta)^k &= \tau\big(g(\eta)\big)^k - \tau\big(h(\eta)\big)^k = \tau\big(g(\eta)^k\big) - \tau\big(h(\eta)^k\big) \\ &= \tau\big(g(\eta^k)\big) - \tau\big(h(\eta^k)\big) = g(\zeta^k) - h(\zeta^k)\,. \end{aligned}$$

Da ζ in K die Ordnung r hat, hängt ζ^k nur von k mod r ab; die Anzahl verschiedener Restklassen der Form $n^u p^v$ modulo r hatten wir oben mit t bezeichnet. Somit hat die Differenz $g - h$ mindestens t Nullstellen. Andererseits sind aber g und h und damit auch ihre Differenz Polynome vom Grad höchstens $t - 1$, also muß $g - h$ das Nullpolynom sein, d.h. $g = h$. Somit enthält S mindestens $2^m - 1$ Elemente, wie behauptet. ■

Zum Abschluß des Beweises, daß der Test von AGRAWAL, KAYAL und SAXENA stets die richtige Antwort liefert, müssen wir nun nur noch zeigen, daß die Schranken aus den beiden letzten Lemmata, die ja unter der Voraussetzung bewiesen wurde, daß eine zusammengesetzte Zahl als prim erkannt wird, einander widersprechen, daß also die untere Schranke größer ist als die obere:

Lemma: $2^{\min(t,\ell)} - 1 > n^{2[\sqrt{t}]}$.

Beweis: Da $\ell(n) > \log_2 n$, genügt es zu zeigen, daß

$$2^{\min(t,\ell)} - 1 > 2^{2\ell(n)[\sqrt{t}]}\,.$$

Da beide Exponenten natürliche Zahlen sind, genügt dazu wiederum, daß $\min(t,\ell) > 2\ell(n)[\sqrt{t}]$ ist, denn wenn sich die Exponenten um mindestens eins unterscheiden, ist die Differenz zwischen den Potenzen mindestens zwei. Wir müssen daher zeigen, daß sowohl t als auch ℓ größer sind als $2\ell(n)[\sqrt{t}]$.

Für $\ell = 2\ell(n)[\sqrt{r}] + 1$ ist das klar, da t die Ordnung einer Untergrupe von $(\mathbb{Z}/r)^\times$ bezeichnet und damit auf jeden Fall kleiner als r ist.

Die Ungleichung $t > 2\ell(n)[\sqrt{t}]$ ist sicherlich dann erfüllt, wenn sogar $t > 2\ell(n)\sqrt{t}$ ist, und dies wiederum ist äquivalent zur Ungleichung $t > 4\ell(n)^2$. Nun ist aber t die Ordnung jener Untergruppe von $(\mathbb{Z}/r)^\times$, die von den Restklassen von n und p erzeugt wird. Da wir im zweiten Schritt des Algorithmus sichergestellt haben, daß dort allein die Ordnung der Restklasse von n schon größer ist als $4\ell(n)^2$, ist auch die Ungleichung für t trivial. ■

Damit ist die Korrektheit des Algorithmus vollständig bewiesen.

§5: Die Verteilung der Primzahlen

Wenn wir Primzahlen einer vorgegebenen Größenordnung suchen (z.B. für einen RSA-Schlüssel), sollten wir zumindest ungefähr wissen, wie die Primzahlen verteilt sind. Damit können wird dann beispielsweise abschätzen, wie groß ein Intervall sein muß, damit wir eine einigermaßen gute Chance haben, dort mindestens eine Primzahl zu finden.

Natürlich sind die Abstände zwischen aufeinanderfolgenden Primzahlen sehr ungleichmäßig verteilt: Der kleinstmögliche Abstand zwischen zwei verschiedenen Primzahlen ist offensichtlich eins, der Abstand zwischen zwei und drei. Er kommt nur an dieser einen Stelle vor, denn außer der Zwei sind schließlich alle Primzahlen ungerade.

Der Abstand zwei ist schon deutlich häufiger: Zwei ist beispielsweise der Abstand zwischen fünf und drei, aber auch der zwischen den Primzahlen $10^{50} + 18307$ und $10^{50} + 18309$. Seit langer Zeit wird vermutet, daß es unendlich viele solcher Primzahlzwillinge gibt; experimentelle Untersuchungen deuten sogar darauf hin, daß ihre Dichte für Zahlen der Größenordnung n bei ungefähr $1 : (\log n)^2$ liegen sollte, aber bislang konnte noch niemand beweisen, daß es wirklich unendlich viele gibt.

Eine obere Grenze für den Abstand zwischen zwei aufeinanderfolgenden Primzahlen gibt es nicht: Ist $n \geq 2$ und $2 \leq i \leq n$, so ist die Zahl $n! + i$ durch i teilbar und somit keine Primzahl. Der Abstand zwischen der größten Primzahl kleiner oder gleich $n! + 1$ und ihrem Nachfolger ist somit mindestens n.

Um einen ersten Eindruck von der Verteilung der Primzahlen zu bekommen, betrachten wir den Graphen der Funktion

$$\pi\colon \begin{cases} \mathbb{R}_{>0} \to \mathbb{N}_0 \\ x \mapsto \text{Anzahl der Primzahlen} \leq x \end{cases}.$$

Die Abbildungen auf der folgenden Seite zeigen ihn für die Intervalle von null bis 10^i für $i = 1, \ldots, 5$. Wie man sieht, werden die Graphen immer glatter, und bei den beiden letzten Bilder könnte man glauben, es handle sich um den Graphen einer differenzerbaren Funktion; auf den ersten Blick sieht sie sogar fast linear aus.

Sieht man sich allerdings die Zahlenwerte genauer an, so sieht man schnell, daß $\pi(x)$ etwas langsamer wächst als eine lineare Funktion; die Funktion $x/\log x$ ist eine deutlich bessere Approximation.

In der Tat können wir auch mit unseren sehr elementaren Mitteln eine entsprechende Aussage beweisen:

Satz: Es gibt Konstanten $c_1, c_2 > 0$, so daß gilt:

$$c_1 \frac{x}{\log x} < \pi(x) < c_2 \frac{x}{\log x}.$$

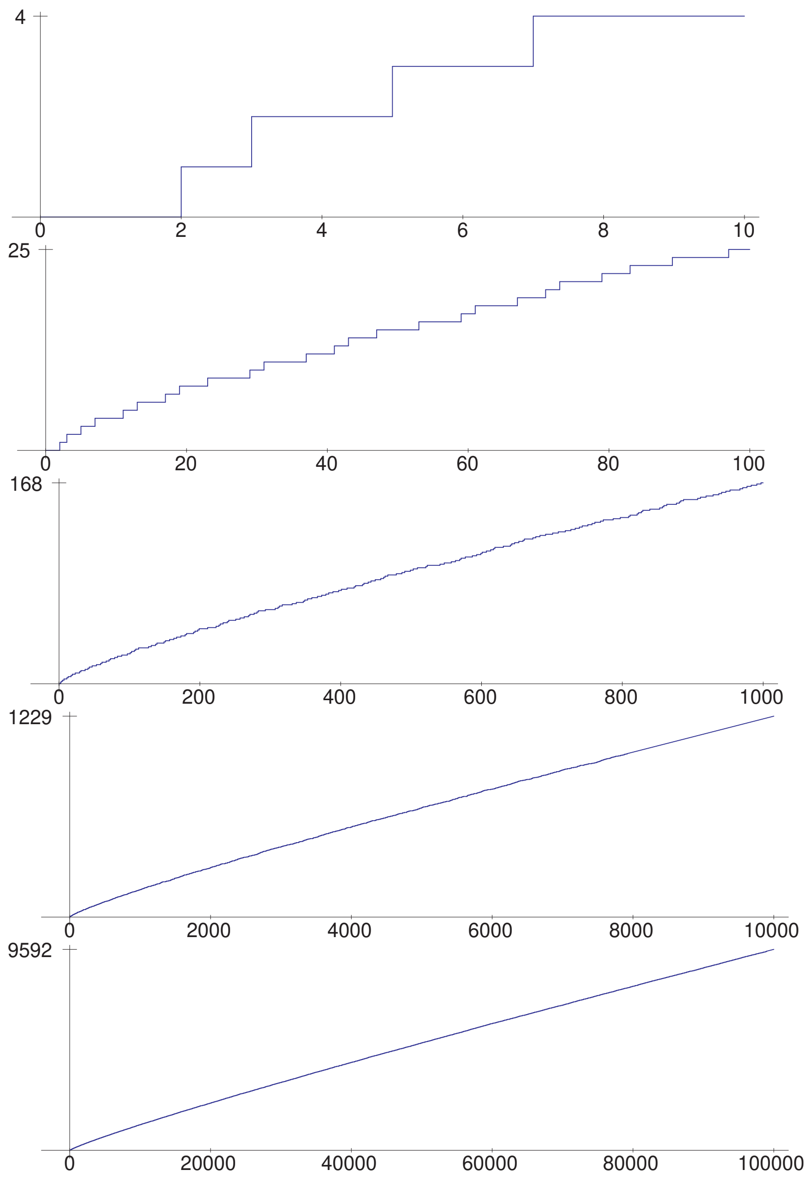
Beweis: Wir betrachten die neue Funktion

$$\vartheta(x) = \sum_{p \leq x} \log p\,,$$

wobei ein Summationsindex p hier wie stets in diesem Beweis bedeuten soll, daß wir über alle *Primzahlen* mit der jeweils angegebenen Eigenschaft summieren.

Dann ist einerseits

$$\pi(x) = \sum_{p \leq x} \frac{\log p}{\log p} \geq \sum_{p \leq x} \frac{\log p}{\log x} = \frac{\vartheta(x)}{\log x}\,,$$



andererseits ist

$$\vartheta(x) = \sum_{p\le x} \log p \ge \sum \sqrt{x} < p \le x \log p > \log\left(\sqrt{x}\right)\left(\pi(x) - \pi\left(\sqrt{x}\right)\right)$$
$$= \frac{1}{2}\log(x)\left(\pi(x) - \pi\left(\sqrt{x}\right)\right)$$

und damit auch

$$\pi(x) < \frac{2\vartheta(x)}{\log x} + \pi\left(\sqrt{x}\right) < \frac{2\vartheta(x)}{\log x} + \sqrt{x}\,.$$

Wenn wir also zeigen können

1. Es gibt Konstanten $c_1, c_3 > 0$, so daß $c_1 x < \vartheta(x) < c_3 x$
2. $\displaystyle\sum_{p\le x} \frac{\log p}{p} = \log x + O(1)$,

dann folgt die Behauptung des Satzes.

Zum Beweis der ersten Aussage betrachten wir die Primzerlegung

$$n! = \prod_{p\le n} p^{e_p}$$

von $n!$. Unter den natürlichen Zahlen bis n sind $\left[\frac{n}{p}\right]$ durch p teilbar, $\left[\frac{n}{p^2}\right]$ durch p^2, *usw.;* daher ist

$$e_p = \sum_{k\ge 1}\left[\frac{n}{p^k}\right] \quad \text{und} \quad \log n! = \sum_{p\le n} e_p \log p = \sum_{p\le n}\sum_{k\ge 1}\left[\frac{n}{p^k}\right]\log p\,.$$

Die Summanden mit $k > 1$ liefern dabei nur einen kleinen Beitrag:

$$\sum_{p\le n}\sum_{k\ge 2}\left[\frac{n}{p^k}\right]\log p \le \sum_{p\le n}\left(\log p \cdot \sum_{k\ge 2}\frac{n}{p^k}\right) = n\sum_{p\le n}\frac{\log p}{p(p-1)}$$

nach der Summenformel für die geometrische Reihe:

$$\sum_{k\ge 2}\frac{1}{p^k} = \frac{1}{p^2}\cdot\frac{1}{1-\frac{1}{p}} = \frac{1}{p^2-p} = \frac{1}{p(p-1)}\,.$$

Zur weiteren Abschätzung ersetzen wir die Summe über alle Primzahlen kleiner oder gleich n durch die Summe aller natürlicher Zahlen bis n und beachten, daß für alle reellen Zahlen $x \ge 2$ gilt

$$\frac{\log x}{x(x-1)} < \frac{\sqrt{x}}{x^2} = \frac{1}{x^{3/2}}:$$

$$\sum_{p\le n}\frac{\log p}{p(p-1)} = \sum_{i=1}^{n}\frac{\log i}{i(i-1)} = \sum_{i=2}^{n}\frac{\log i}{i(i-1)} \le \sum_{i=2}^{n}\frac{1}{i^{3/2}}\,.$$

Da $\sum_{i=1}^{\infty}\frac{1}{i^s}$ für alle $s > 1$ konvergiert, konvergiert die rechts stehende Summe für $n \to \infty$ gegen einen endlichen Wert (ungefähr 1,612375), ist also $O(1)$, und damit ist

$$\sum_{k\ge 2}\frac{1}{p^k} = O(n)\,.$$

Setzen wir dies in die Formel für $\log n!$ ein, erhalten wir nach allen bislang bewiesenen Abschätzungen, daß

$$\log n! = \sum_{p\le n}\left[\frac{n}{p}\right]\log p + O(n)\,.$$

Dies können wir vergleichen mit der STIRLINGschen Formel

$$\log n! = n\log n - n + O(\log n)\,,$$

deren Beweis für Leser, die sie noch nicht kennen, im Anhang zu diesem Paragraphen skizziert ist. Kombinieren wir dies mit der gerade bewiesenen Formel, ist also

$$\sum_{p\le n}\left[\frac{n}{p}\right]\log p = n\log n + O(n)\,.$$

Damit ist

$$\sum_{p\le 2n}\left(\left[\frac{2n}{p}\right] - 2\left[\frac{n}{p}\right]\right) = 2n\log 2n - 2n\log n + O(2n)$$
$$= 2n\log 2 + O(n) = O(n)\,.$$

Hier ist $\left[\frac{2n}{p}\right] - 2\left[\frac{n}{p}\right]$ stets entweder null oder eins; speziell für die Primzahlen p mit $n < p < 2n$ ist $\left[\frac{n}{p}\right] = 0$ und $\left[\frac{2n}{p}\right] = 1$. Somit ist

$$\vartheta(2n) - \vartheta(n) = \sum_{n<p<2n} \log p \le \sum_{p\le 2n} \left(\left[\frac{2n}{p}\right] - 2\left[\frac{n}{p}\right]\right) \log p = O(n)\,.$$

Die Formel $\vartheta(2n) - \vartheta(n) = O(n)$ bleibt gültig, wenn wir n durch eine reelle Zahl x ersetzen; somit ist

$$\vartheta(x) = \sum_{i=1}^{\infty} \left(\vartheta\left(\frac{x}{2^{i+1}}\right) - \vartheta\left(\frac{x}{2^i}\right)\right) = O\left(\sum_{i=1}^{\infty} \frac{x}{2^i}\right) = O(x)\,,$$

womit die obere Schranke für $\vartheta(x)$ bewiesen wäre.

Bevor wir uns der unteren Schranke zuwenden, beweisen wir zunächst die zweite Aussage.

Natürlich ist $\frac{n}{p} = \left[\frac{n}{p}\right] + O(1)$, also ist

$$\begin{aligned}\sum_{p\le n} \frac{n}{p} \log p &= \sum_{p\le n} \left[\frac{n}{p}\right] \log p + O\left(\sum_{p\le n} \log p\right)\\ &= n\log n + O(n) + O\big(\vartheta(n)\big) = n\log n + O(n)\,,\end{aligned}$$

denn wie wir gerade gesehen haben ist $\vartheta(n) = O(n)$. Kürzen wir die obige Formel durch n, erhalten wir die gewünschte Aussage

$$\sum_{p\le n} \frac{\log p}{p} = \log n + O(1)\,,$$

die natürlich auch dann gilt, wenn wir n durch eine reelle Zahl x ersetzen: Der Term $O(1)$ schluckt alle dabei auftretenden zusätzlichen Fehler.

Für $0 < \alpha < 1$ ist daher

$$\sum_{\alpha x<p\le x} \frac{\log p}{p} = \log x - \log \alpha x + O(1) = \log \frac{1}{\alpha} + O(1)\,,$$

wobei der Fehlerterm $O(1)$ nicht von α abhängt.

Da $\log \frac{1}{\alpha}$ für $\alpha \to 0$ gegen ∞ geht, ist für hinreichend kleine Werte von α und $x > c/\alpha$ für irgendein $c > 2$ beispielsweise

$$\sum_{\alpha x<p\le x} \frac{\log p}{p} > 10\,,$$

und für solche Werte von α und c ist dann

$$10 < \sum_{\alpha x<p\le x} \frac{\log p}{p} \le \frac{1}{\alpha x} \sum_{\alpha x<p\le \alpha} \log p \le \frac{\vartheta(x)}{\alpha x}\,.$$

Somit ist $10\alpha x < \vartheta(x)$, womit auch die untere Schranke aus der ersten Behauptung bewiesen wäre und damit der gesamte Satz. ■

Der bewiesene Satz ist nur ein schwacher Abglanz dessen, was über die Funktion $\pi(x)$ bekannt ist. Zum Abschluß des Kapitels seien kurz einige der wichtigsten bekannten und vermuteten Eigenschaften von $\pi(x)$ zusammengestellt. Diese knappe Übersicht folgt im wesentlichen dem Artikel *Primzahlsatz* aus

> DAVID WELLS: Prime Numbers – The Most Mysterious Figurs in Math, *Wiley,* 2005,

einer Zusammenstellung im Lexikonformat von interessanten Tatsachen und auch bloßen Kuriosa aus dem Umkreis der Primzahlen.

GAUSS kam 1792, im Alter von 15 Jahren also, durch seine Experimente zur Vermutung, daß $\pi(x)$ ungefähr gleich dem sogenannten *Integrallogarithmus* von x sein sollte:

$$\pi(x) \approx \mathrm{Li}(x) \underset{\text{def}}{=} \int_2^x \frac{d\xi}{\log \xi}\,.$$

Auch LEGENDRE versuchte, $\pi(x)$ anhand experimenteller Daten anzunähern. Er stellte dazu eine Liste aller Primzahlen bis 400 000 zusammen, das sind immerhin 33 860 Stück, und suchte eine glatte Kurve, die den Graphen von π möglichst gut annähert. In seinem 1798 erschienenen Buch *Essai sur la théorie des nombres* gab er sein Ergebnis an als

$$\pi(x) \approx \frac{x}{\log x - 1{,}08366}\,.$$

Über ein halbes Jahrhundert später gab es den ersten Beweis einer Aussage: PAFNUTIJ L'VOVIČ ČEBYŠËV (1821–1894) zeigte 1851: *Falls*

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x}$$

existiert, dann muß er den Wert eins haben.

1852 zeigte er dann ein deutlich schärferes Resultat als den oben bewiesenen Satz: Für *hinreichend große* Werte von x ist

$$c_1 \cdot \frac{x}{\log x} < \pi(x) < c_2 \cdot \frac{x}{\log x} \quad \text{mit} \quad c_1 \approx 0{,}92 \;\text{ und }\; c_2 \approx 1{,}105\,.$$

1896 schließlich zeigten der französische Mathematiker JACQUES SALOMON HADAMARD (1865–1963) und sein belgischer Kollege CHARLES JEAN GUSTAVE NICOLAS BARON DE LA VALLÉE POUSSIN (1866–1962) unabhängig voneinander die Aussage, die heute als **Primzahlsatz** bekannt ist:

$$\pi(x) \sim \frac{x}{\log x}\,.$$

Dies bedeutet nun freilich nicht, daß damit die Formeln von GAUSS und von LEGENDRE überflüssig wären: Die Tatsache, daß der Quotient zweier Funktionen gleich eins ist, erlaubt schließlich immer noch beträchtliche Unterschiede zwischen den beiden Funktionen: Nur der *relative* Fehler muß gegen null gehen.

Offensichtlich ist für jedes $a \in \mathbb{R}$

$$\lim_{x\to\infty} \frac{x/\log x}{x/(\log x - a)} = \lim_{x\to\infty} \frac{\log x - a}{\log x} = 1 - \lim_{x\to\infty} \frac{a}{\log x} = 1\,,$$

und es ist auch nicht schwer zu zeigen, daß

$$\lim_{x\to\infty} \frac{x/\log x}{\operatorname{Li}(x)} = 1$$

ist. Nach dem Primzahlsatz ist daher auch für jedes $a \in \mathbb{R}$

$$\pi(x) \sim \frac{x}{\log x - a} \quad \text{und} \quad \pi(x) \sim \operatorname{Li}(x)\,.$$

Wie DE LA VALLÉE POUSSIN zeigte, liefert der Wert $a = 1$ unter allen reellen Zahlen a die beste Approximation an $\pi(x)$, aber $\operatorname{Li}(x)$ liefert eine

noch bessere Approximation. Für kleine Werte von x sieht man das auch in der folgenden Tabelle, in der alle reellen Zahlen zur nächsten ganzen Zahl gerundet sind. Wie kaum anders zu erwarten, liefert LEGENDREs Formel für 10^4 und 10^5 die besten Werte:

n	$\pi(n)$	$\frac{n}{\log n}$	$\frac{n}{\log n - 1}$	$\frac{n}{\log n - 1{,}08366}$	$\operatorname{Li}(n)$
10^3	168	145	169	172	178
10^4	1 229	1 086	1 218	1 231	1 246
10^5	9 592	8 686	9 512	9 588	9 630
10^6	78 489	72 382	78 030	78 534	78 628
10^7	664 579	620 420	661 459	665 138	664 918
10^8	5 761 455	5 428 681	5 740 304	5 769 341	5 762 209
10^9	50 847 478	48 254 942	50 701 542	50 917 519	50 849 235

Wenn wir genaue Aussagen über $\pi(x)$ machen wollen, sollten wir also etwas über die Differenz $\operatorname{Li}(x) - \pi(x)$ wissen. Hier kommen wir in das Reich der offenen Fragen, und nach derzeitigem Verständnis hängt alles ab von der RIEMANNschen Zetafunktion

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}\,.$$

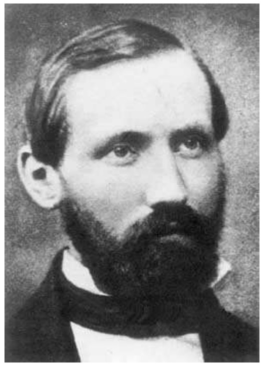
Es ist im wesentlichen eine Analysis I Übungsaufgabe zu zeigen, daß diese Summe für reelle $s > 1$ konvergiert; wer mit komplexen Zahlen umgehen kann, folgert daraus dann leicht, daß sie auch für alle komplexen s mit Realteil größer eins konvergiert.

Etwas trickreicher, aber durchaus noch im Rahmen einer Vorlesung *Funktionentheorie I* durchführbar, ist der Beweis, daß $\zeta(s)$ zu einer analytischen Funktion auf $\mathbb{C} \setminus \{1\}$ forgesetzt werden kann. (Für $s = 1$ haben wir eine harmonische Reihe, und die divergiert bekanntlich, so daß der von rechts kommende Limes von $\zeta(s)$ für $s \to 1$ unendlich sein muß.) Wie RIEMANN erkannte, hängt die Primzahlverteilung eng mit der Frage zusammen, welche Nullstellen $\zeta(s)$ für jene Argumente s hat, deren Realteil zwischen null und eins liegt.

Nach einer berühmten Vermutung von RIEMANN haben alle diese Nullstellen den Realteil ein halb. Falls dies stimmt, ist

$$\pi(x) = \operatorname{Li}(x) + O(\sqrt{x}\log x)\,.$$

Die RIEMANNsche Vermutung ist eines der wichtigsten ungelösten Probleme der heutigen Mathematik; sie war 1900 eines der HILBERTschen Probleme und ist auch eines der sieben *Millennium problems,* für deren Lösung das CLAY Mathematics Institute in Cambridge, Mass. einen Preis von jeweils einer Million Dollar ausgesetzt hat.



GEORG FRIEDRICH BERNHARD RIEMANNN (1826-1866) war Sohn eines lutherischen Pastors und schrieb sich 1946 auf Anraten seines Vaters an der Universität Göttingen für das Studium der Theologie ein. Schon bald wechselte an die Philosophische Fakultät, um dort unter anderem bei GAUSS Mathematikvorlesungen zu hören. Nach Promotion 1851 und Habilitation 1854 erhielt er dort 1857 einen Lehrstuhl. Trotz seines frühen Todes initiierte er grundlegende auch noch heute fundamentale Entwicklungen in der Geometrie, der Zahlentheorie und über abelsche Funktionen. Wie sein Nachlaß zeigte, stützte er seine 1859 aufgestellte Vermutung über die Nullstellen der ζ-Funktion auf umfangreiche Rechnungen.

Anhang: Die Eulersche Summenformel und die Stirlingsche Formel

Die EULERsche Summenformel erlaubt es, eine endliche Summe auf ein Integral zurückzuführen und dadurch in vielen Fällen erst rechnerisch handhabbar zu machen. Wir betrachten eine reellwertige differenzierbare Funktion f, deren Definitionsbereich das Intervall $[1,\ n]$ enthält.

Für eine reelle Zahl x bezeichnen wir weiterhin mit $[x]$ die größte ganze Zahl kleiner oder gleich x; außerdem führen wir noch die Bezeichnung $\{x\} \underset{\text{def}}{=} x - [x]$ ein für den gebrochenen Anteil von x. Für eine ganze Zahl k ist somit $\{x\} = x - k$ für alle x aus dem Intervall $[k,\ k+1)$.

Partielle Integration führt auf die Gleichung

$$\int_k^{k+1} \left(\{x\} - \tfrac{1}{2}\right) f'(x)\,dx = \left(x - k - \tfrac{1}{2}\right) f(x)\,\Big|_k^{k+1} - \int_k^{k+1} f(x)\,dx$$

$$= \frac{f(k+1) + f(k)}{2} - \int_k^{k+1} f(x)\,dx\,.$$

Addition aller solcher Gleichungen von $k = 1$ bis $k = n - 1$ liefert

$$\int_1^n \left(\{x\} - \tfrac{1}{2}\right) f'(x)\,dx = \frac{f(1)}{2} + \sum_{k=2}^{n-1} f(k) + \frac{f(n)}{2} - \int_1^n f(x)\,dx\,,$$

womit man die Summe der $f(k)$ berechnen kann:

Satz (EULERsche Summenformel): Für eine differenzierbare Funktion $f\colon D \to \mathbb{R}$, deren Definitionsbereich das Intervall $[1,\ n]$ umfaßt, ist

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)\,dx + \frac{f(1) + f(n)}{2} + \int_1^n \left(\{x\} - \tfrac{1}{2}\right) f'(x)\,dx\,.$$

■

Für die Abschätzung von $n!$ interessiert uns speziell der Fall, daß $f(x) = \log x$ der natürliche Logarithmus ist; hier wird die EULERsche Summenformel zu

$$\begin{aligned} \log n! &= \int_1^n \log x\,dx + \frac{\log n}{2} + \int_1^n \frac{\{x\} - \frac{1}{2}}{x}\,dx \\ &= x(\log x - 1)\,\Big|_1^n + \frac{\log n}{2} + \int_1^n \frac{\{x\} - \frac{1}{2}}{x}\,dx \\ &= n(\log n - 1) + 1 + \frac{\log n}{2} + \int_1^n \frac{\{x\} - \frac{1}{2}}{x}\,dx\,. \end{aligned}$$

In dieser Formel stört noch das rechte Integral; dieses können wir wie folgt abschätzen: Für eine natürliche Zahl k ist

$$\int_k^{k+1} \frac{\{x\} - \frac{1}{2}}{x}\,dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{x}{k + \frac{1}{2} + x}\,dx$$

$$= \int\limits_0^{\frac12} \left(\frac{x}{k+\frac12+x} - \frac{x}{k+\frac12-x} \right) dx = \int\limits_0^{\frac12} \frac{-2x^2}{\left(k+\frac12\right)^2 - x^2}\, dx\,.$$

Im Intervall von 0 bis $\frac12$ ist der Integrand monoton fallend, d.h.

$$0 \ge \frac{-2x^2}{\left(k+\frac12\right)^2 - x^2} \ge \frac{-\frac12}{\left(k+\frac12\right)^2 - \frac14} = \frac{-2}{(2k+1)^2-1} \ge -\frac{1}{2k^2}\,,$$

und damit ist

$$0 \ge \int\limits_k^{k+1} \frac{\{x\}-\frac12}{x}\, dx = \int\limits_0^{\frac12} \frac{-2x^2}{\left(k+\frac12\right)^2 - x^2}\, dx \ge -\frac{1}{4k^2}\,,$$

denn wir können das Integral abschätzen durch das Produkt aus der Länge des Integrationsintervalls und dem Minimum des Integranden. Summation von $k=1$ bis $n-1$ schließlich gibt die Abschätzung

$$0 \ge \int\limits_1^n \frac{\{x\}-\frac12}{x}\, dx \ge -\sum_{k=1}^{n-1} \frac{1}{4k^2}$$

für das störende Integral aus der obigen Formel.

Wie wohl jeder schon einmal in einer Analysis I Übungsaufgabe zeigen mußte, konvergiert die rechtsstehende Summe (egal ob mit oder ohne vier im Nenner) für $n \to \infty$; wer mit FOURIER-Reihen vertraut ist, weiß wahrscheinlich auch, daß der Grenzwert $\pi^2/24$ ist. Auf jeden Fall können wir folgern, daß das uneigentliche Integral

$$\int\limits_1^\infty \frac{\{x\}-\frac12}{x}\, dx$$

konvergiert; den Grenzwert wollen wir mit I bezeichnen. Dann ist

$$\log n! = n(\log n - 1) + \frac{\log n}{2} + C + o(1) \quad \text{mit} \quad C = I + 1\,,$$

also folgt insbesondere die Abschätzung

$$\log n! = n \log n + O(n)\,,$$

die wir im Beweis des Satzes über $\pi(n)$ verwendet haben.

Kapitel 8
Faktorisierungsverfahren

Wie wir in §2 des letzten Kapitels gesehen haben, ist $M_{67} = 2^{67} - 1$ keine Primzahl, denn

$$13^{M_{67}-1} \equiv 81\,868\,480\,399\,682\,966\,751 \bmod M_{67} \not\equiv 1 \bmod M_{67}\,.$$

Somit ist M_{67} ein Produkt von mindestens zwei nichttrivialen Faktoren. Welche sind das?

FRANK NELSON COLE gab das Ergebnis am 31. Oktober 1903 auf einer Sitzung der American Mathematical Society bekannt: Er schrieb die Zahl

$$2^{67} - 1 = 147\,573\,952\,589\,676\,412\,927$$

auf eine der beiden Tafeln und

$$193\,707\,721 \times 761\,838\,257\,287$$

auf die andere. Dieses Produkt rechnete er wortlos aus (nach der üblichen Schulmethode zur schriftlichen Multiplikation), und als er dieselbe Zahl erhielt, die auf der anderen Tafel stand, schrieb er ein Gleichheitszeichen zwischen die beiden Zahlen und setzte sich wieder. Das Ergebnis, d.h. die Faktorisierung von M_{67}, findet ein Computeralgebrasystem heute in weniger als einer Sekunden; für die damalige Zeit war sie eine Sensation! COLE gab später zu, daß er drei Jahre lang jeden Sonntag nachmittag daran gearbeitet hatte. Er versuchte M_{67} in der Form x^2-y^2 darzustellen, wobei er mit Hilfe quadratischer Reste Kongruenzbedingungen für x modulo verschiedener relativ kleiner Primzahlen aufstellte und auch verwendete, daß jeder Teiler von M_{67} kongruent eins modulo 67 und kongruent ± 1 modulo acht sein muß. Dies führte zu einer ganzen Reihe von Kongruenzen für x, die er in

$$x \equiv 1\,160\,932\,384 \bmod 1\,323\,536\,760$$

zusammenfassen konnte. Untersuchung quadratischer Reste zeigt, daß

$$x_k = 1\,323\,536\,760\,k + 1\,160\,932\,384$$

frühestens für $k = 287$ in Frage kommt, und mit $x = x_{287}$ ist tatsächlich

$$\begin{aligned} M_{67} &= 381\,015\,982\,504^2 - 380\,822\,274\,783^2 \\ &= 193\,707\,721 \times 761\,838\,257\,287\,. \end{aligned}$$

Für Einzelheiten siehe

F. N. COLE: On the factoring of large numbers, *Bull. Am. Math. Soc.* **10** (1903), 134–137

FRANK NELSON COLE (1861–1926) wurde in Massachusetts geboren; 1878 ging er dort an die Harvard University, wo er 1882 seinen Bachelor erhielt. Mit einem Stipendium konnte er dann drei Jahre lang nach Deutschland gehen, wo er bei FELIX KLEIN in Leipzig studierte. Mit einer von KLEIN betreuten Arbeit über Gleichungen sechsten Grades wurde er 1886 in Harvard promoviert. Nach verschiedenen Positionen in Harvard und Michigan ging er 1895 als Professor an die Columbia University in New York, wo er bis zu seinem Tod lehrte. Seine Arbeiten befassen sich hauptsächlich mit Primzahlen und mit Gruppentheorie.

Der Auftritt von COLE schlug selbst außerhalb der Mathematik so große Wellen, daß seine Faktorisierung noch fast ein Jahrhundert später vorkommt in einer New Yorker (off-Broadway) Show von RINNE GROFF mit dem Titel *The five hysterical girls theorem.* Dort bringt sich ein junger Mathematiker um, weil er in einem Beweis von der *Primzahl* $2^{67} - 1$ ausgeht und die Tochter des Professors die obige Faktorisierung an die Tafel schreibt. Einzelheiten kann man, so man unbedingt möchte, unter `http://www.playscripts.com/play.php3?playid=551` nachlesen.

In diesem Kapitel soll es um zumindest einige der Verfahren gehen, mit denen man heute das Problem der Faktorisierung von Zahlen wie $2^{67} - 1$ und auch erheblich größeren Zahlen behandelt.

Es gibt kein „bestes“ Faktorisierungsverfahren; für Zahlen verschiedener Größenordnungen haben jeweils andere Verfahren ihre Stärken. Auch Vorwissen über die zu faktorisierende Zahl kann bei der Wahl eines geeigneten Verfahrens helfen: Bei einem RSA-Modul, der das Produkt zweier Primzahlen ähnlicher Größenordung ist, wird man anders vorgehen als etwa bei einer Zahl der Form $a^n \pm 1$. Mehr noch als bei Primzahltests gilt, daß asymptotische Komplexitätsaussagen als Auswahlkriterium nutzlos sind: Das für die Faktorsierung 150-stelliger RSA-Moduln heute optimale Verfahren, das Zahlkörpersieb, wird beim Versuch eine sechsstellige Zahl zu faktorisieren, oft nicht in der Lage sein die Faktoren zu trennen, und selbst in den Fällen, in denen es erfolgreich ist, braucht es erheblich länger als einfache Probedivisionen. Im folgenden sollen einige der einfachsten gebräuchlichen Verfahren vorgestellt werden.

§ 1: Die ersten Schritte

a) Test auf Primzahl

Der schlimmste Fall für praktisch jedes Faktorisierungsverfahren tritt dann ein, wenn die zu faktorisierende Zahl eine Primzahl ist: Gerade bei den fortgeschrittenen Verfahren gibt es oft kein anderes Abbruchkriterium als das Auffinden eines Faktors. Daher sollte (außer eventuell bei ganz kleinen Zahlen) zu Beginn einer Faktorisierung immer ein Primzahltest stehen. Da auch das Testen auf Potenzen relativ einfach ist, läßt sich eventuell auch das noch durchführen – es sei denn, daß von der Situation her (beispielsweise bei RSA-Moduln) nicht mit einer Potenz zu rechnen ist.

b) Abdividieren kleiner Primteiler

Falls eine Zahl n zusammengesetzt ist, hat sie mindestens einen Primteiler $p \leq [\sqrt{n}]$. Bei kleinen Zahlen n besteht die effizienteste Art der Faktorisierung im allgemeinen darin, einfach alle diese Primzahlen durchzuprobieren, indem man sie der Reihe nach so lange abdividiert, wie es geht.

Für große Werte von n ist $[\sqrt{n}]$ zu groß; trotzdem sollte man auch da zumindest alle Primteiler bis zu einer gewissen Schranke N eliminieren. Ein typischer Wert für PCs wäre etwa $N = 2^{15}$ oder $N = 2^{16}$. Die Vorgehensweise ist folgende:

1. Schritt: Bestimme nach ERATOSTHENES die Folge $p_1, \ldots, p_r$ aller Primzahlen $p_i \leq N$ und setze $m = n$ sowie $e_1 = \cdots = e_r = 0$.

2. Schritt: Führe für $i = 1, \ldots, r$ die folgenden Anweisungen aus:
Falls m nicht durch p_i teilbar, geht es weiter mit dem nächsten i; andernfalls wird so lange m durch m/p_i und e_i durch $e_i + 1$ ersetzt, bis p_i kein Teiler von m mehr ist. Falls $m = 1$, geht es weiter zu Schritt drei, andernfalls geht es weiter mit dem nächsten i.

3. Schritt: Falls $m = 1$, ist $n = \prod_{i=1}^{r} p_i^{e_i}$ faktorisiert; andernfalls ist $n = \prod_{i=1}^{r} p_i^{e_i} \cdot m$ mit einer Zahl m, die keinen Primteiler $p \leq N$ hat. Falls $m < N^2$, ist m eine Primzahl, und n ist ebenfalls komplett faktorisiert. Andernfalls teste man, ob m nicht eventuell doch prim ist, womit die Faktorisierung ebenfalls beendet ist. Im Falle eines zusammengesetzten m muß dieses mit einem anderen Verfahren weiter untersucht werden.

§2: Die Verfahren von Pollard und ihre Varianten

a) Die Monte-Carlo-Methode

Als etwas weniger systematische Alternative zum Abdividieren könnte man auch eine Folge von Zufallszahlen x_i erzeugen und jeweils den ggT von x_i mit der zu faktorisierenden Zahl N bilden. Dies hat zwar den Nachteil, daß ein EUKLIDischer Algorithmus aufwendiger ist als eine bloße Division mit Rest und daß der ggT möglicherweise eine zusammengesetzte Zahl ist, daür testet man aber in vielen Schritten mehrere Primzahlen auf einmal, und selbst ein zusammengesetzter Faktor ist nützlich, denn je kleiner eine Zahl ist, desto einfacher ist sie zu faktorisieren. Eine weitere Optimierung wird dadurch erreicht, daß wir mehrere x_i modulo N miteinander multiplizieren können und dann erst den ggT des Produkts modulo N mit N bilden. Offensichtlich ist dieser ggT genau dann durch eine Primzahl p teilbar, wenn diese Teiler von N

und von mindestens einem der Faktoren ist. Die Anzahl der Faktoren darf natürlich nicht zu groß sein, denn sonst besteht die Gefahr, daß der ggT einfach gleich N ist. Wenn man aber die kleinen Primzahlen bereits durch Abdividieren eliminiert hat, kann man i.a. relativ gefahrlos mit der Zusammenfassung von etwa hundert Zufallszahlen arbeiten.

Ist p ein Primteiler von N, so sollte bei echten Zufallszahlen etwa jede p-te durch p teilbar sein; ist also p der kleinste Primteiler von N, so kann man erwarten, daß nach p Versuchen ein nichttrivialer Faktor gefunden wird, der p enthält. Dies ist kein Problem für vierstellige Faktoren (die wir allerdings mindestens genauso schnell auch durch Abdividieren bestimmen können), ist aber schon für achtstellige Faktoren viel zu aufwendig.

POLLARDs Idee zur Beschleunigung beruht auf dem Geburtstagsparadoxon: Die Wahrscheinlichkeit dafür, daß eine gegebene Zufallszahl durch p teilbar ist, liegt zwar nur bei $1 : p$, aber die Wahrscheinlichkeit, daß zwei der x_i modulo p gleich sind, steigt in der Nähe von etwa $\sqrt{p}$ Folgegliedern ziemlich steil von nahe null zu nahe eins. Wenn wir also anstelle der größten gemeinsamen Teiler von N mit den x_i die mit den Differenzen $x_i - x_j$ berechnen, haben wir bereits bei einer Folge der Länge um $\sqrt{p}$ gute Chancen, einen nichttrivialen ggT zu finden.

In dieser Form ist das Verfahren allerdings noch nicht praktikabel: Wenn wir ein neues x_i mit $i \approx \sqrt{p}$ erzeugt haben, müssen wir für alle $j < i$ den ggT von $x_i - x_j$ berechnen, was noch einmal rund $\sqrt{p}$ Schritte sind, so daß der Gesamtaufwand nicht proportional zu $\sqrt{p}$ ist, sondern eher zu

$$\int_0^{\sqrt{p}} x\,dx = \frac{p}{2}\,,$$

was keine große Ersparnis ist. Dazu kommt, daß alle bereits berechneten Folgeglieder gespeichert werden müssen, der Algorithmus hat also auch einen Platzbedarf in der Größenordnung $\sqrt{p}$.

Dieses Problem können wir umgehen, indem wir keine echten Zufallszahlen verwenden, sondern algorithmisch eine Folge sogenannter Pseudozufallszahlen erzeugen. Typischerweise verwendet man dazu eine

Rekursionsvorschrift der Form $x_{i+1} = Q(x_i) \bmod N$ mit einem quadratischen Polynom Q. (Die bei Simulationen sehr beliebten Pseudozufallsgeneratoren nach der linearen Kongruenzmethode sind für die Monte-Carlo-Methode der Faktorisierung nicht geeignet.)

Wegen der speziellen Form der Rekursion hängt die Restklasse $x_{i+1} \bmod p$ nur ab von $x_i \bmod p$; insbesondere ist also $x_{i+1} \equiv x_{j+1} \bmod p$, falls $x_i \equiv x_j \bmod p$, und entsprechend stimmen auch für jedes $r \geq 0$ die Zahlen x_{i+r} und x_{j+r} modulo p überein, d.h. die Folge wird modulo p periodisch mit einer Periode π, die $|i-j|$ teilt.

Das Problem, Periodizität in einer Folge zu entdecken, tritt nicht nur in der Zahlentheorie auf, sondern beispielsweise auch in der Zeitreihenanalyse und anderen Anwendungen. Ein möglicher Algorithmus zu seiner Lösung, auch als Hase und Schildkröte Algorithmus bekannt, stammt von FLOYD (1967) und beruht auf folgender Beobachtung:

Wird eine Folge (y_i) irgendwann periodisch, so gibt es Indizes k derart, daß $y_k = y_{2k}$ ist.

In der Tat, ist $y_{i+\pi} = y_i$ für alle $i \geq r$, so können wir für k jedes Vielfache $\ell\pi$ der Periode nehmen, das mindestens gleich r ist.

ROBERT W. FLOYD (1936–2001) beendete seine Schulausbildung bereits im Alter von 14 Jahren, um dann mit einem Stipendium an der Universität von Chicago zu studieren, wo er mit 17 einen Bachelor in *liberal arts* bekam. Danach finanzierte er sich durch Arbeit ein zweites Bachelorstudium in Physik, das er 1958 abschloß. Damit war seine akademische Ausbildung beendet; er arbeitete als Operator in einem Rechenzentrum, brachte sich selbst Programmieren bei und begann einige Jahre später mit der Publikation wissenschaftlicher Arbeiten auf dem Gebiet der Informatik. Mit 27 wurde er Assistenzprofessor in Carnegie Mellon, fünf Jahre später erhielt er einen Lehrstuhl in Stanford. Zu den vielen Entwicklungen, die er initiierte, gehört die semantische Verifikation von Programmen, Design und Analyse von Algorithmen, Refactoring, dazu kommen Arbeiten über Graphentheorie und das FLOYD-STEINBERG dithering in der Computergraphik. 1978 erhielt er den TURING-Preis, die höchste Auszeichnung der Informatik. Stanfords Nachruf auf FLOYD ist zu finden unter `news-service.stanford.edu/news/2001/november7/floydobit-117.html`.

Damit sieht der Grobablauf der Monte-Carlo-Faktorisierung einer natürlichen Zahl N folgendermaßen aus:

Schritt 0: Man wähle ein quadratisches Polynom Q und einen Startwert x_0. Setze $x = y = x_0$.
Schritt $i, i > 0$**:** Ersetze x durch $Q(x)$ und y durch $Q\big(Q(y)\big)$; berechne dann $\mathrm{ggT}(x-y, N)$. Falls dieser weder eins noch N ist, wurde ein Faktor gefunden.

Man beachte, daß hier im i-ten Schritt $x = x_i$ und $y = x_{2i}$ ist; wir erzeugen also die Folge der x_i und die der x_{2i} simultan, ohne Zwischenergebnisse zu speichern.

Natürlich kann man auch bei dieser Form des Algorithmus mehrere ggT-Berechnungen zusammenfassen: Sollen etwa jeweils s Berechnungen zusammengefaßt weden, so führt man eine neue Variable P ein mit Anfangswert ein ersetzt P im i-ten Schritt durch $P \cdot (x-y) \bmod N$. Nur falls i durch r teilbar ist, wird anschließend der ggT von N und P berechnet; andernfalls geht es gleich weiter mit dem $(i+1)$-ten Schritt.

Die Monte-Carlo-Methode wird auch als ρ-Methode bezeichnet, da die Folge der X_i nicht von Anfang an periodisch sein muß. Sie muß aber, da es nur p Restklassen modulo p gibt, schließlich periodisch werden, d.h. sie beginnt auf dem unteren Ast des ρ und mündet irgendwann in den Kreis. Erfahrungsgemäs ist diese Methode sehr erfolgreich im Auffinden sechs- bis achtstelliger Faktoren; danach wird sie recht langsam, und kleine Faktoren kann sie oft nicht trennen.

JOHN M. POLLARD ist ein britischer Mathematiker, der hauptsächlich bei British Telecom arbeitete. Er publizierte rund zwanzig mathematische Arbeiten, größtenteils auf dem Gebiet der algorithmischen Zahlentheorie. Bekannt sind auch seine Beiträge zur Kryptographie, für die er 1999 den RSA Award erhielt. Außer den hier vorgestellten Faktorisierungsalgorithmen entwickelte er unter anderem auch das Zahlkörpersieb, eine Variante des weiter hinten vorgestellten quadratischen Siebs, dessen Weiterentwicklungen derzeit die schnellten Faktorisierungsalgorithmen für große Zahlen sind.

b) Die $(p-1)$-Methode

POLLARDs zweite Methode beruht auf dem kleinen Satz von FERMAT: Ist p ein Primteiler von N und r ein Vielfaches von $p-1$, so ist $a^r \equiv 1 \bmod$

p für jedes zu p teilerfremde a; der ggT von a^r mod N und N ist also durch p teilbar.

Natürlich ist $p-1$ nicht bekannt, wir können aber hoffen, daß $p-1$ nur durch vergleichsweise kleine Primzahlen teilbar ist. Sei etwa B eine Schranke mit der Eigenschaft, daß $p-1$ durch keine Primzahlpotenz größer B teilbar ist. Dann ist das Produkt r *aller* Primzahlpotenzen q^e, die höchstens gleich B sind, sicherlich ein Vielfaches von $p-1$, wenn auch ein extrem großes, das sich kaum mit realistischem Aufwand berechnen läßt. Für jedes konkrete a kann a^r mod N jedoch verhältnismäßig einfach berechnet werden: Man potenziert einfach nacheinander für jede Primzahl $q \le B$ modulo N mit deren größter Potenz, die immer noch kleiner oder gleich B ist; mit dem Algorithmus zur modularen Exponentiation aus Kapitel 1 geht das auch für sechs- bis siebenstellige Werte von B noch recht flott.

Insgesamt funktioniert POLLARDs $(p-1)$-Methode zur Faktorisierung einer natürlichen<Zahl N also folgendermaßen:

Schritt 0: Wähle eine Schranke B und eine Basis a zwischen 1 und N.

Schritt 1: Erstelle (z.B. nach ERATOSTHENES) eine Liste aller Primzahlen $q \le B$.

Schritt 2: Berechne für jede dieser Primzahlen q den größten Exponenten e derart, daß auch noch $q^e \le B$ ist, d.h. $e = [\log B / \log q]$. Ersetze dann den aktuellen Wert von a durch a^{q^e} mod N.

Schritt 3: Berechne $\mathrm{ggT}(a, N)$. Falls ein Wert ungleich eins oder N gefunden wird, war das Verfahren erfolgreich, ansonsten nicht.

Es ist klar, daß der Erfolg dieses Verfahrens wesentlich davon abhängt, daß N einen Primteiler p hat mit der Eigenschaft, daß alle Primfaktoren von $p-1$ relativ klein sind. Ob dies der Fall ist, läßt sich im Voraus nicht sagen; die $(p-1)$-Methode liefert daher gelegentlich ziemlich schnell sogar 20- oder 30-stellige Faktoren, während sie andererseits deutlich kleinere Faktoren oft nicht findet.

Als Beispiel betrachten wir noch einmal $M_{67} = 2^{67}-1$. Wenn wir mit der Basis $a = 17$ und der Schranke $B = 3\,000$ arbeiten, wird a modulo M_{67}

potenziert zum neuen

$$a = 111\,153\,665\,932\,902\,146\,348 \text{ mit } \mathrm{ggT}(a-1, M_{67}) = 193\,707\,721\,.$$

Damit ist eine nichttriviale Faktorisierung gefunden, und ein Primzahltest zeigt, daß sowohl der gefundene Faktor als auch sein Komplement prim sind.

c) Varianten

Falls $p-1$ nicht nur relativ kleine Primfaktoren hat, führt die $(p-1)$-Methode nicht zum Erfolg. In solchen Fällen hat dann aber vielleicht $p+1$ oder irgendeine andere Zahl in der Nähe von p nur kleine Primfaktoren. In solchen Fällen können Varianten der $(p-1)$-Methode zum Erfolg führen.

Um diese Varianten zu definieren, empfiehlt es sich, zunächst die $(p-1)$-Methode etwas abstrakter unter gruppentheoretischen Gesichtspunkten zu betrachten.

Wir rechnen in der primen Restklassengruppe $(\mathbb{Z}/N)^\times$ und damit implizit auch in $(\mathbb{Z}/p)^\times$ für jeden Primteiler p von N – egal ob wir ihn kennen, oder nicht. In $(\mathbb{Z}/p)^\times$ ist für jedes Element a die $(p-1)$-te Potenz gleich dem Einselement; genau dasselbe gibt für jede r-te Potenz, für die der Exponent r ein Vielfaches von $(p-1)$ ist. Bei der $(p-1)$-Methode wird ein r berechnet, das durch alle Primzahlpotenzen bis zu einer gewissen Schranke teilbar ist; falls in der Primzerlegung von $p-1$ keine Primzahlpotenz oberhalb der Schranke liegt, ist r ein Vielfaches von $p-1$.

Allgemeiner können wir statt in $(\mathbb{Z}/N)^\times$ und $(\mathbb{Z}/p)^\times$ auch in einem anderen Paar von Gruppen rechnen: Wir gehen aus von einer endlichen Gruppe G_n, deren Elemente sich in irgendeiner Weise als r-tupel über $(\mathbb{Z}/N)$ auffassen lassen; außerdem nehmen wir an, daß sich die Gruppenmultiplikation für zwei so dargestellte Elemente auf Grundrechenarten über $\mathbb{Z}/N$ zurückführen läßt. Dann können wir die Elemente von G_n zu Tupeln über $\mathbb{Z}/p$ reduzieren und die Menge aller so erhaltenen Tupel bildet eine Gruppe G_p. Wieder ist jede Rechnung in G_n implizit auch eine Rechnung in G_p.

Die Elementanzahl von G_p sei $N(p)$.

Wir wählen irgendein Element von G_n und potenzieren es mit demselben Exponenten r, mit dem wir bei der $p-1$-Methode die Zahl a modulo N potenziert haben. Falls r ein Vielfaches von $N(p)$ ist, erhalten wir ein Element $b \in G_n$, dessen Reduktion modulo p das Einselement von G_p ist. Ist daher b_i die i-te Koordinate von b und e_i die von e, so muß die Differenz $b_i - e_i$ durch p teilbar sein, und mit etwas Glück können wir p als ggT von n und $b_i - e_i$ bestimmen.

Bleibt nur noch das Problem, geeignete Gruppen zu finden. Bei der $(p-1)$-Methode ist $G_n = (\mathbb{Z}/n)^\times$ und $N(p) = p-1$. Ein anderer Vorschlag von POLLARD war $G_p = \mathbb{F}_{p^2}^\times/\mathbb{F}_p^\times$; hier ist

$$N(p) = \frac{p^2-1}{p-1} = p+1\,,$$

daher der Name $(p+1)$-Methode. Zur Konstruktion vom G_n brauchen wir zunächst eine Gruppe, die modulo p auf $\mathbb{F}_p^\times$ reduziert; dazu können wir eine geeignete quadratische Erweiterung von $(\mathbb{Z}/n)^\times$ nehmen. G_n ist dann die Faktorgruppe dieser Gruppe nach $(\mathbb{Z}/n)^\times$.

Derzeit am populärsten ist aber eine andere Wahl von G_n und G_p: Wir nehmen für G_n eine elliptische Kurve über $\mathbb{Z}/n$. Dabei handelt es sich um die Menge aller Punkte $(x,y) \in (\mathbb{Z}/n)^2$, die einer vorgegebenen Gleichung

$$y^2 = x^3 - ax - b$$

genügen, wobei a, b Elemente von $\mathbb{Z}_n$ sind, für die $\Delta = 4a^3 - 27b^2$ teilerfremd zu n ist; dazu kommt ein weiterer Punkt O, den wir formal als $(0,\infty)$ schreiben. G_p ist dann die entsprechende Punktmenge in $\mathbb{F}_p^2$ zusammen mit O. Nach einem Satz von HELMUT HASSE (1898–1979) ist

$$p+1-2\sqrt{p} < N(p) < p+1+2\sqrt{p}\,,$$

und wie man inzwischen weiß, kann man auch für jeden Wert, der diese Ungleichung erfüllt, Parameterwerte a und b finden, so daß $N(p)$ gleich diesem Wert ist. Wenn man mit hinreichend vielen verschiedenen Kurven arbeitet, ist daher die Chance recht groß, daß der Exponent r wenigstens für eine davon ein Vielfaches von $N(p)$ ist.

Die Multiplikation ist folgendermaßen definiert: Durch zwei Punkte (x_1, y_1) und (x_2, y_2) auf der Kurve geht genau eine Gerade; setzt man deren Gleichung $y = mx + c$ ein die Kurvengleichung ein, erhält man ein Polynom dritten Grades in x. Dieses hat natürlich die beiden Nullstellen x_1, x_2, und daneben noch eine dritte Nullstelle x_3. Der dritte Schnittpunkt der Geraden mit der Kurve ist somit $(x_3, mx_3 + c)$; als Summe der beiden Punkte definiert man aber

$$(x_1, y_1) \oplus (x_2, y_2) = \bigl(x_3, -(mx_3 + c)\bigr)\,.$$

Man kann zeigen, daß dies die Menge der Kurvenpunkte zu einer Gruppe mit Neutralelement O macht, in der man genauso vorgehen kann wie bei der klassischen $(p-1)$-Methode.

§3: Das Verfahren von Fermat und seine Varianten

Die bisher betrachteten Verfahren funktionieren vor allem dann gut, wenn die zu faktorisierende Zahl mindestens einen relativ kleinen Primteiler hat. Das hier beschriebene Verfahren von FERMAT führt genau dann schnell ans Ziel, wenn sie sich als Produkt zweier fast gleich großer Faktoren schreiben läßt. In seiner einfachsten Form beruht es auf der

$$x^2 - y^2 = (x+y)(x-y)\,.$$

Ist $N = pq$ Produkt zweier ungerader Primzahlen, so ist

$$N = (x+y)(x-y) \quad \text{mit} \quad x = \frac{p+q}{2} \quad \text{und} \quad y = \frac{p-q}{2}\,;$$

zusammen mit obiger Formel folgt $N + y^2 = x^2$.

FERMAT berechnet für $y = 0, 1, 2, \ldots$ die Zahlen $N + y^2$; falls er auf ein Quadrat x^2 stößt, hat er zwei Faktoren $x \pm y$ gefunden.

Anstelle der Zahlen $N + y^2$ kann man auch für ein festes k die Zahlen $kN + y^2$ betrachten. Falls dies eine Quadratzahl x^2 ist, gilt entsprechend

$$kN = x^2 - y^2 = (x+y)(x-y)\,,$$

und wenn man Glück hat, sind $\mathrm{ggT}(x \pm y, N)$ echte Faktoren von N. Wenn man Pech hat, sind dies die beiden Zahlen eins und N, so daß dies auf den ersten Blick keine Vorteile gegenüber dem klassischen FERMAT-Verfahren hat, insbesondere da es keine offensichtliche Wahl für k gibt.

Es gibt allerdings eine ganze Reihe von Algorithmen, die ohne Rücksicht auf einen konkreten Wert von k einfach Zahlen $x, y \in \mathbb{Z}$ suchen, für die

$$x^2 \equiv y^2 \bmod N$$

ist, und unter diesen Verfahren sind die besten derzeit bekannten zur Faktorisierung großer Zahlen ohne kleine Primteiler.

Auch hier sind mit etwas Glück $\mathrm{ggT}(x \pm y, N)$ echte Faktoren von N, wenn man Pech hat sind es einfach wieder eins und N. In der Praxis wird man daher von vornherein gleich mehrere Paare (x, y) mit $x^2 \equiv y^2 \bmod N$ suchen um die Chance zu erhöhen, daß zumindest ein Paar auf eine nichttriviale Faktorisierung führt.

Hier soll nur kurz der Grundalgorithmus, das sogenannte quadratische Sieb, beschrieben werden; die wirklich für Rekordfaktorisierungen benutzen Modifikationen sind teilweise mathematisch recht anspruchsvolle Varianten davon.

Im einfachsten Fall arbeiten wir ausschließlich mit dem Polynom

$$f(x) = \left(x + \left\lceil\sqrt{N}\right\rceil\right)^2 - N .$$

Für jedes x ist dann $f(x) \equiv \left(x + \left\lceil\sqrt{N}\right\rceil\right)^2 \bmod N$, wobei links und rechts verschiedene Zahlen stehen. Insbesondere steht links im allgemeinen keine Quadratzahl.

Falls wir allerdings Werte $x_1, x_2, \ldots, x_r$ finden können, für die das Produkt der $f(x_i)$ eine Quadratzahl ist, dann ist

$$\prod_{i=1}^{r} f(x_i) \equiv \prod_{i=1}^{r} \left(x + \left\lceil\sqrt{N}\right\rceil\right)^2 \bmod N$$

eine Relation der gesuchten Art.

Um die x_i zu finden, betrachten wir eine Menge $\mathcal{B}$ von Primzahlen, die sogenannte Faktorbasis. Typischerweise enthält $\mathcal{B}$ für die Faktorisierung einer etwa hundertstelligen Zahl etwa 100–120 Tausend Primzahlen, deren größte somit, wie die folgende Tabelle zeigt, im einstelligen Millionenbereich liegt.

n	n-te Primzahl	n	n-te Primzahl
100 000	1 299 709	600 000	8 960 453
200 000	2 750 159	700 000	10 570 841
300 000	4 256 233	800 000	12 195 257
400 000	5 800 079	900 000	13 834 103
500 000	7 368 787	1 000 000	15 485 863

Beim quadratischen Sieb interessieren nur x-Werte, für die $f(x)$ als Produkt von Primzahlen aus $\mathcal{B}$ (und eventuell auch Potenzen davon) darstellbar ist. Ist

$$f(x_i) = \prod_{p\in\mathcal{B}} p^{e_{ip}} ,$$

so ist

$$\prod_{i=1}^{r} f(x_i)^{\varepsilon_i} = \prod_{p\in\mathcal{B}} p^{\sum_{i=1}^{r} \varepsilon_i e_{ip}}$$

genau dann ein Quadrat, wenn

$$\sum_{i=1}^{r} \varepsilon_i e_{ip}$$

für alle $p \in \mathcal{B}$ gerade ist. Dies hängt natürlich nur ab von den $\varepsilon_i \bmod 2$ und den $e_{ip} \bmod 2$; wir können ε_i und e_{ip} daher als Elemente des Körpers mit zwei Elementen auffassen und bekommen dann über $\mathbb{F}_2$ die Bedingungen

$$\sum_{i=1}^{r} \varepsilon_i e_{ip} = 0 \quad \text{für alle } p \in \mathcal{B} .$$

Betrachten wir die ε_i als Variablen, ist dies ein homogenes lineares Gleichungssystem in r Variablen mit soviel Gleichungen, wie es Primzahlen in der Faktorbasis gibt. Dieses Gleichungssystem hat nichttriviale Lösungen, falls die Anzahl der Variablen die der Gleichungen übersteigt, falls es also mehr Zahlen x_i gibt, für die $f(x_i)$ über der Faktorbasis faktorisiert werden kann, als Primzahlen in der Faktorbasis.

Für jede nichttriviale Lösung ist

$$\prod_{i=1}^{r} f(x_i)^{\varepsilon_i} = \prod_{i=1}^{r} \left(x + \left[\sqrt{N}\right]\right)^{2\varepsilon_i} \bmod N$$

eine Relation der Form $x^2 \equiv y^2 \bmod N$, die mit einer Wahrscheinlichkeit von etwa ein halb zu einer Faktorisierung von N führt. Falls wir zehn linear unabhängige Lösungen des Gleichungssystems betrachten, führt also mit einer Wahrscheinlichkeit von etwa 99,9% mindestens eine davon zu einer Faktorisierung.

Da ε_i nur die Werte 0 und 1 annimmt, stehen in obigem Produkt natürlich keine echten Potenzen: Man multipliziert einfach nur die Faktoren miteinander, für die $\varepsilon_i = 1$ ist. Außerdem interessieren nicht die links- und rechtsstehenden Quadrate, sondern deren Quadratwurzeln; tatsächlich also berechnet man (hier natürlich in $\mathbb{N}_0$)

$$x = \prod_{p\in\mathcal{B}} p^{\frac{1}{2}\sum_{i=1}^{r} \varepsilon_i e_{ip}} \bmod N \quad \text{und} \quad y = \prod_{i=1}^{r} \left(x + \left[\sqrt{N}\right]\right)^{\varepsilon_i} \bmod N\,.$$

Zum besseren Verständnis des Verfahrens wollen wir versuchen, damit die Zahl 15 zu faktorisieren. Dies ist zwar eine sehr untypische Anwendung, da das quadratische Sieb üblicherweise erst für mindestens etwa vierzigstellige Zahlen angewandt wird, aber zumindestens das Prinzip sollte auch damit klarwerden.

Als Faktorbasis verwenden wir die Menge

$$\mathcal{B} = \{2, 3, 7, 11\}\,;$$

die Primzahl fünf fehlt, da $3 \cdot 5 = 15$ ist und daher bei einer Faktorbasis, die sowohl drei als auch fünf enthält, die Gefahr zu groß ist, daß die linke wie auch die rechte Seite der Kongruenz durch fünfzehn teilbar ist. Bei realistischen Anwendungen muß man auf solche Überlegungen keine Rücksicht nehmen, denn dann sind die Elemente der Faktorbasis höchstens siebenstellig und somit erheblich kleiner als die gesuchten Faktoren.

Wir berechnen $f(x)$ für $x = 1, 2, \ldots$, bis wir einige Funktionswerte haben, die über der Faktorbasis faktorisiert werden können. Die faktorisierbaren Werte sind in folgender Tabelle zusammengestellt:

x	$x + \left[\sqrt{N}\right]$	$f(x)$	*Faktorisierung*
1	4	1	
3	6	21	$3 \cdot 7$
5	8	49	7^2
6	9	66	$2 \cdot 3 \cdot 11$
10	13	154	$2 \cdot 7 \cdot 11$
54	57	3234	$2 \cdot 3 \cdot 7^2 \cdot 11$

Die erste und die dritte Zeile sind selbst schon Relationen der gesuchten Art, nämlich

$$4^2 \equiv 1 \bmod 15 \quad \text{und} \quad 8^2 \equiv 7^2 \bmod 15\,.$$

Die zweite Relation ist nutzlos, denn $8 - 7 = 1$ und $8 + 7 = 15$. Die erste dagegen führt zur Faktorisierung, denn

$$\mathrm{ggT}(4 + 1, 15) = 5 \quad \text{und} \quad \mathrm{ggT}(4 - 1, 15) = 3\,.$$

Da dies aber ein Zufall ist, der bei großen Werten von N so gut wie nie vorkommt, wollen wir das ignorieren und mit den Relationen zu $x = 3, 6, 10$ und 51 arbeiten:

$$\begin{aligned} 6^2 &\equiv 3 \cdot 7 && \bmod 15 \\ 9^2 &\equiv 2 \cdot 3 \cdot 11 && \bmod 15 \\ 13^2 &\equiv 2 \cdot 7 \cdot 11 && \bmod 15 \\ 54^2 &\equiv 2 \cdot 3 \cdot 7^2 \cdot 11 && \bmod 15 \end{aligned}$$

Multipliziert man die ersten drei dieser Relationen miteinander, folgt

$$(6 \cdot 9 \cdot 13)^2 \equiv (2 \cdot 3 \cdot 7 \cdot 11)^2 \bmod 15$$

oder $702^2 \equiv 462^2 \bmod 15$. Da

$$\mathrm{ggT}(702 - 462, 15) = \mathrm{ggT}(240, 15) = 15$$

ist, bringt das leider nichts.

Wir erhalten auch dann rechts ein Quadrat, wenn wir das Produkt der ersten, dritten und vierten Relation bilden; dies führt auf

$$(6 \cdot 13 \cdot 57)^2 \equiv (2 \cdot 3 \cdot 7^2 \cdot 11)^2 \bmod 15$$

oder $4446^2 \equiv 3234^2 \bmod 15$. Hier ist

$$\operatorname{ggT}(4446 - 3234, 15) = \operatorname{ggT}(1212, 15) = 3\,,$$

womit wir die Zahl 15 faktorisiert haben – wenn auch nicht unbedingt auf die einfachstmögliche Weise.

Bei realistischen Beispielen sind die Funktionswerte $f(x)$ deutlich größer als die Primzahlen aus der Faktorbasis; außerdem liegen die vollständig faktorisierbaren Zahlen viel dünner als hier: Bei der Faktorisierung einer hundertstelligen Zahl etwa muß man davon ausgehen, daß nur etwa jeder 10^9-te Funktionswert über der Faktorbasis zerfällt.

Daher ist es wichtig, ein Verfahren zu finden, mit dem diese wenigen Funktionswerte schnell und einfach bestimmt werden können. Das ist zum Glück möglich:

Der Funktionswert $f(x)$ ist genau dann durch p teilbar, wenn

$$f(x) \equiv 0 \bmod p$$

ist. Für ein Polynom f mit ganzzahligen Koeffizienten ist offensichtlich $f(x) \equiv f(y) \bmod p$, falls $x \equiv y \bmod p$ ist. Daher ist für ein x mit $f(x) \equiv 0 \bmod p$ auch

$$f(x+kp) \equiv 0 \bmod p \quad \text{für alle } k \in \mathbb{Z}\,.$$

Es genügt daher, im Bereich $0 \le x < p-1$ nach Werten zu suchen, für die $f(x)$ durch p teilbar ist.

Dazu kann man f auch als Polynom über dem Körper mit p Elementen betrachten und nach Nullstellen in diesem Körper suchen. Für Polynome großen Grades und große Werte von p kann dies recht aufwendig sein; hier, bei einem quadratischen Polynom, müssen wir natürlich einfach eine quadratische Gleichung lösen: In $\mathbb{F}_p$ wie in jedem anderen Körper auch gilt

$$f(x) = \left(x - \left[\sqrt{N}\right]\right)^2 - N = 0 \iff \left(x - \left[\sqrt{N}\right]\right)^2 = N\,,$$

und diese Gleichung ist genau dann lösbar, wenn es ein Element $w \in \mathbb{F}_p$ gibt mit Quadrat N, wenn also in $\mathbb{Z}$

$$w^2 \equiv N \bmod p$$

ist. Für $p > 2$ hat $f(x) = 0$ in $\mathbb{F}_p$ dann die beiden Nullstellen

$$x = \left[\sqrt{N}\right] \pm w\,;$$

andernfalls gibt es keine Lösung.

Insbesondere kann also $f(x)$ nur dann durch p teilbar sein, wenn N modulo p ein Quadrat ist; dies ist für etwa die Hälfte aller Primzahlen der Fall. Offensichtlich sind alle anderen Primzahlen nutzlos, die Faktorbasis sollte also nur Primzahlen enthalten, für die N modulo p ein Quadrat ist. Mit Hilfe des quadratischen Reziprozitätsgesetzes läßt sich leicht bestimmen, für welche Primzahlen dies der Fall ist. Für solche p kann man dann (z.B. mit dem Algorithmus von SHANKS) die beiden Lösungen der Gleichung $f(x) = 0$ in $\mathbb{F}_p$ berechnen.

Das eigentliche Sieben zum Auffinden der komplett über der Faktorbasis zerlegbaren Funktionswerte $f(x)$ geht dann folgendermaßen vor sich: Man legt ein Siebintervall $x = 0, 1, \ldots, M$ fest und speichert in einem Feld der Länge $M+1$ für jedes x eine ganzzahlige Approximation von $\log_2 f(x)$.

Für jede Primzahl p aus der Faktorbasis berechnet man dann die beiden Nullstellen $x_{1/2}$ von f modulo p im Intervall von 0 bis $p-1$ und subtrahiert von jedem Feldelement mit Index der Form $x_1 + kp$ oder $x_2 + kp$ eine ganzzahlige Approximation von $\log_2 p$.

Falls $f(x)$ über der Faktorbasis komplett faktorisierbar ist, sollte dann am Ende der entsprechende Feldeintrag bis auf Rundungsfehler gleich null sein; um keine Fehler zu machen, untersucht man daher für alle Feldelemente, die unterhalb einer gewissen Grenze liegen, durch Abdividieren, ob sie wirklich komplett faktorisieren, und man bestimmt auf diese Weise auch *wie* sie faktorisieren. Damit läßt sich dann das oben erwähnte Gleichungssystem über $\mathbb{F}_2$ aufstellen und, falls genügend viele Relationen gefunden sind, nichttrivial so lösen, daß eine der daraus resultierenden Gleichungen $x^2 \equiv y^2 \bmod p$ zu einer nichttrivialen Faktorisierung von N führt.