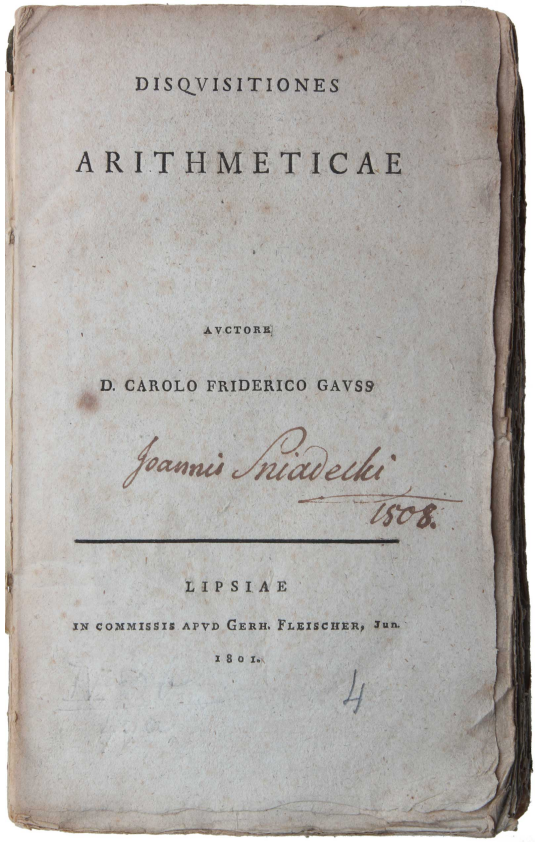
Von der reinen Theorie zur Anwendung: Kodieren, suchen und verschlüsseln

Meinolf Geck

IAZ - Lehrstuhl für Algebra

Tag der Wissenschaft, Juli 2014



Carl Friedrich Gauß, 30.4.1777 – 23.2.1855

Begründung der modernen Zahlentheorie

(geschrieben 1798, veröffentlicht 1801)

Die "5." Grundrechenart (außer $+$, $-$, $\cdot$, $/$)

$$14 \bmod 3 = 2, \quad 6 \bmod 2 = 0, \quad -9 \bmod 4 = 3.$$

Weit verbreitet und gut bekannt in Programmiersprachen:

$14\%3$ (C, Python, Java, ...), $\mathrm{mod}(14, 3)$ (Fortran, ...)

- $14 \bmod 3$ ("14 *modulo* 3") ist das Ergebnis folgender Operation: "*Ziehe solange* 3 *von* 14 *ab, bis noch etwas Positives (oder* 0*) übrig bleibt; dieses Überbleibsel ist das Ergebnis*". Also:

  $14 - 3 = 11, \ 11 - 3 = 8, \ 8 - 3 = 5, \ 5 - 3 = 2$ STOP!

- Genauso: $6 - 2 = 4, \ 4 - 2 = 2, \ 2 - 2 = 0$ STOP!

- "*Bei einer negativen Zahl addiere solange, bis etwas Positives (oder* 0*) herauskommt*"; also:

  $-9 + 4 = -5, \ -5 + 4 = -1, \ -1 + 4 = 3$ STOP!

Modulare Arithmetik

Halte Modulus (z.B. 6) fest und rechne dann stets *modulo* 6:

$$10 +_6 3 = 1 \quad \text{kurz für} \quad (10+3) \bmod 6 = 13 \bmod 6 = 1.$$

$$10 \cdot_6 3 = 0 \quad \text{kurz für} \quad (10 \cdot 3) \bmod 6 = 30 \bmod 6 = 0.$$

$\rightsquigarrow$ erhalte neue "*Grundrechenarten modulo* 6".

Vereinfachungsregel

In einer Summe $17 +_6 14 = (17+14) \bmod 6 = 31 \bmod 6 = 1$

darf man auch zuerst "mod" berechnen und dann addieren; also:

$$17 +_6 14 = (17 \bmod 6) +_6 (14 \bmod 6) = 5 +_6 2 = 7 \bmod 6 = 1.$$

Genauso beim Produkt $17 \cdot_6 14 = (17 \cdot 14) \bmod 6 = ?$

$$17 \cdot_6 14 = (17 \bmod 6) \cdot_6 (14 \bmod 6) = 5 \cdot_6 2 = 10 \bmod 6 = 4.$$

$\rightsquigarrow$ selbst bei längeren Rechnungen bleiben Zahlen "klein".

Ein Satz der Zahlentheorie. Gegeben seien

- eine Primzahl p (also eine Zahl > 1, die nur durch 1 und sich selbst teilbar ist, z.B., $2, 3, 5, 7, 11, \ldots$)
- und eine beliebige Zahl n, die nicht durch p teilbar ist.

Dann gilt: $\boxed{n^{p-1} \bmod p = 1}$ "Kleiner Satz von Fermat".

Zum Beispiel $p = 3$, $n = 8$ $\rightsquigarrow$ $8^2 = 64 \bmod 3 = 1$.

Pierre de Fermat, 17.8.1601 – 12.1.1665

Eine Anwendung: Prüfziffern (Beispiel IBAN)

Deutsches Bankkonto
Konto-Nr. 0356843503
BLZ 37010050

$\rightsquigarrow$ DE12 $\underbrace{37010050}_{\text{BLZ}}$ $\underbrace{0356843503}_{\text{Konto-Nr.}}$

Prüfziffer 12 wird nach folgendem Verfahren berechnet:

- Schreibe BLZ, gefolgt von Konto-Nr., Land und 00: 370100500356843503DE00.
- Wandle Buchstaben in Zahlen um:

A	B	C	D	$\ldots$	Z
10	11	12	13	$\ldots$	35

- Berechne $370100500356843503131400 \bmod 97 = 86$.
- Ziehe Rest von 98 ab: Ergebnis 12.

  (Falls Ergebnis einstellig, ergänze führende Null.)

Umgekehrt: Validierung einer IBAN

DE12370100500356843503 $\rightsquigarrow$

370100500356843503DE12 $\rightsquigarrow$

$370100500356843503131412 \mod 97 = 1$

(Kommt als Ergebnis nicht 1 heraus, wird die IBAN nicht akzeptiert.)

Britische IBAN: GB52TSBS $\underbrace{873410}_{\text{sort code}}\underbrace{81152768}_{\text{account nr.}}$

TSBS873410811527 68GB52 $\rightsquigarrow$ (T=29, S=28, B=11, G=16)

$292811287341081152768161152 \mod 97 = 1$.

Eine weitere Anwendung: Verschlüsseln von Nachrichten

Alice möchte die Ergebnisse der Fußballspiele Deutschlands bei der WM 2014 an Bob übermitteln; für das Spiel gegen Brasilien ist die Nachricht also 7:1. Der Einfachheit halber will sie also $N = 71$ an Bob übermitteln. (Analog für andere Spiele.)

> *Aber Bob möchte nicht, dass allzu viele Leute erfahren, was für Nachrichten Alice ihm schickt!*

Bob wählt zwei Primzahlen p, q und bildet $n = pq$; dabei soll $n > N$ sein. Also z.B. $p = 101$, $q = 103$, $n = 10403$.

Außerdem bildet Bob $(p-1)(q-1) = 10200$ und wählt zwei Zahlen e, d mit $(ed) \bmod 10200 = 1$; dabei soll e groß sein. (So etwas geht immer.) Hier z.B. $e = 8743$, $d = 7$.

Bob behält $d = 7$ für sich, und veröffentlicht die beiden Zahlen

$$n = 10403, \quad e = 8743.$$

Verschlüsselung: Anstelle von N, übermittele $r = N^e \bmod n$.

Also: Anstelle der ursprünglichen Nachricht $N = 71$ schickt Alice die verschlüsselte Nachricht $r = 71^{8743} \bmod 10403 = 5512$ an Bob.

Was kann Bob (oder jemand anders!) mit $r = 5512$ anfangen?

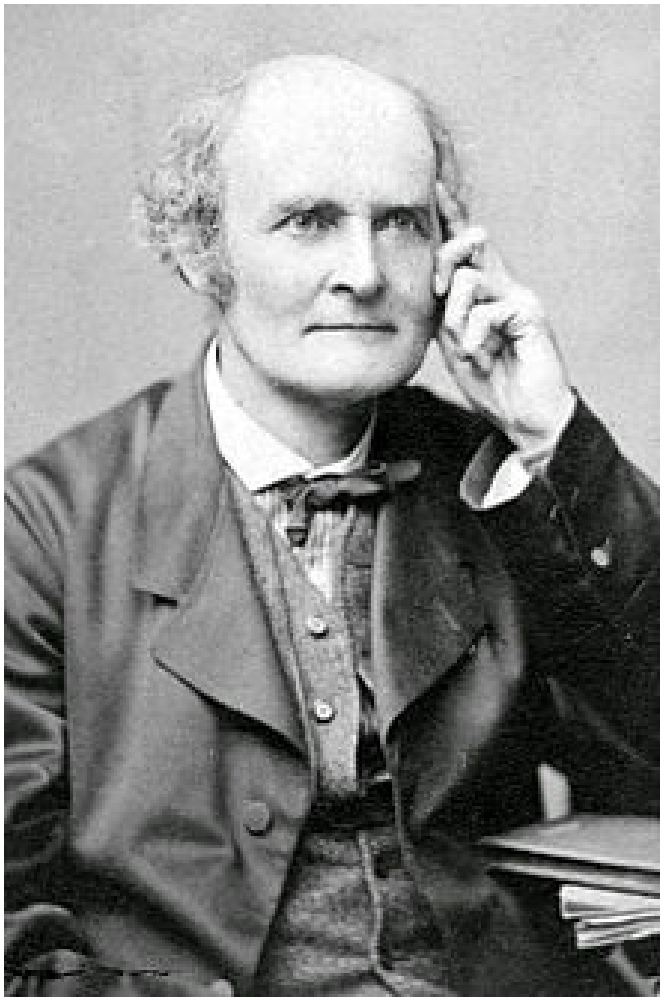
- Kleiner Satz von Fermat $\rightsquigarrow$ Es gilt $N = r^d \bmod n$. Da Bob d kennt, kann er $N = 5512^7 \bmod 10403 = 71$ berechnen.
- Wer d nicht kennt, hat zwei Möglichkeiten:
  - Probieren: Teste $a = 1, 2, 3 \ldots$, bis $a^{8743} \bmod 10403 = 5512$ gilt. Dann ist $a = N$ die ursprüngliche Nachricht. Sind n, e groß, so dauert dies zu lange (selbst mit den besten Computern).

- Zweite Möglichkeit:
  - Versuche p, q zu finden mit $n = pq$. Denn dann kann man (so wie Bob) auch d mit $(ed) \bmod (p-1)(q-1) = 1$ finden.
    Aber: Faktorisierung $n = pq$ sehr schwierig für große n.
- Fazit: Die Sicherheit des Verfahrens (RSA) beruht darauf, dass
  - es entweder zu lange dauert, durch Probieren von $a = 1, 2, 3, \ldots, n$ die ursprüngliche Nachricht zu finden,
  - oder es zu schwierig ist, n in Primzahlen zu zerlegen.

(Beachte: Unser $n = 10403$ ist 5stellig. Tatsächlich werden n mit mehreren Hundert Stellen verwendet.)

Mein Tipp für das Endspiel Deutschland – Argentinien: $r = 3288$.

(Wie oben verschlüsselt mit $n = 10403$ und $e = 8743$.)



II. *A Memoir on the Theory of Matrices.* *By* Arthur Cayley, *Esq.*, *F.R.S.*

Received December 10, 1857,—Read January 14, 1858.

The term matrix might be used in a more general sense, but in the present memoir I consider only square and rectangular matrices, and the term matrix used without qualification is to be understood as meaning a square matrix; in this restricted sense, a set of quantities arranged in the form of a square, *e. g.*

$$\begin{pmatrix} a, & b, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{pmatrix}$$

is said to be a matrix. The notion of such a matrix arises naturally from an abbreviated notation for a set of linear equations, viz. the equations

$$\begin{aligned} X &= ax + by + cz, \\ Y &= a'x + b'y + c'z, \\ Z &= a''x + b''y + c''z, \end{aligned}$$

may be more simply represented by

$$(X, Y, Z) = \begin{pmatrix} a, & b, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{pmatrix} (x, y, z),$$

and the consideration of such a system of equations leads to most of the fundamental notions in the theory of matrices. It will be seen that matrices (attending only to those of the same order) comport themselves as single quantities; they may be added, multiplied or compounded together, &c.: the law of the addition of matrices is precisely similar to that for the addition of ordinary algebraical quantities; as regards their multiplication (or composition), there is the peculiarity that matrices are not in general convertible; it is nevertheless possible to form the powers (positive or negative, integral or fractional) of a matrix, and thence to arrive at the notion of a rational and integral function, or generally of any algebraical function, of a matrix. I obtain the remarkable theorem that any matrix whatever satisfies an algebraical equation of its own order, the coefficient of the highest power being unity, and those of the other powers functions of the terms of the matrix, the last coefficient being in fact the determinant; the rule for the formation of this equation may be stated in the following condensed form, which will be intelligible after a perusal of the memoir, viz. the determi-

Arthur Cayley 16.8.1821 – 26.1.1895, Gesammelte Schriften: 967 Arbeiten

Lineare Gleichungssysteme

$$\begin{array}{rcrcrcr} 2x & - & y & + & z & = & 1 \\ -x & + & 3y & & & = & 0 \\ 5x & - & y & + & 2z & = & -2 \end{array}$$

Matrix-Schreibweise:

$$\underbrace{\begin{pmatrix} 2 & -1 & 1 \\ -1 & 3 & 0 \\ 5 & -1 & 2 \end{pmatrix}}_{=A} \cdot \underbrace{\begin{pmatrix} x \\ y \\ z \end{pmatrix}}_{=v} = \underbrace{\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}}_{=b}$$

Gleichungssystem wird zu einer Gleichung:

$$A \cdot v = b$$

(Analog für m lineare Gleichungen in n Variablen, mit m, n beliebig.)

Eine Anwendung: Verkaufsprognosen

Ein Unternehmen verkauft zwei Waschmittel A und B und möchte ein drittes C einführen. Kapazität: Insgesamt 100000 Stück pro Monat. Aufgrund von Kundenbefragungen ist zu erwarten, dass:

- von den Kunden, die A gekauft haben, 50% dabei bleiben, im nächsten Monat aber 10% zu B und 40% zu C wechseln würden;
- Käufer von B zu 60% auch bei B bleiben, 30% zu C und 10% zu A wechseln würden;
- von denen, die C gekauft haben, auch 90% bei C bleiben, keiner zu A und nur 10% zu B wechseln würden.

Fragestellung: Wieviel soll von jedem Waschmittel konkret pro Monat hergestellt werden, damit die Kapazität optimal ausgenutzt wird?

Mathematische Modellierung:

Seien x, y, z die Verkaufszahlen von A, B, C in einem Monat und x', y', z' die Verkaufszahlen im nächsten Monat. Dann gilt:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 0.5 & 0.1 & 0 \\ 0.1 & 0.6 & 0.1 \\ 0.4 & 0.3 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Optimal: Die Verkaufszahlen ändern sich nicht, d.h., es gilt $x = x'$, $y = y'$, $z = z'$. Ist A die obige 3×3-Matrix, so ist also gesucht:

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{mit} \quad x + y + z = 100000 \quad \text{und} \quad A \cdot v = v.$$

Ein solches v heißt dann **stationärer Vektor**.

Obige Matrix $A = \begin{pmatrix} 0.5 & 0.1 & 0 \\ 0.1 & 0.6 & 0.1 \\ 0.4 & 0.3 & 0.9 \end{pmatrix}$ hat folgende Eigenschaft:

Die Summe der Einträge in jeder Spalte ist genau 1.

Eine solche Matrix A heißt deshalb **spalten-stochastisch**.

Ein Satz der Matrix-Theorie (1. Semester Lineare Algebra):

Sei A eine spalten-stochastische Matrix beliebiger Größe. Dann gibt es stets einen stationären Vektor v (der nicht nur aus Nullen besteht).

Es gibt Berechnungsverfahren, um solche stationären Vektoren auch zu finden (Gauß-Algorithmus).

Lösung in unserem Beispiel: Es sollten pro Monat 4000 Stück von A, 20000 Stück von B und 76000 Stück von C hergestellt werden.

Eine weitere Anwendung: Suchen im Internet

Tippt man einen Suchbegriff in Google ein, so wird im Internet nach allen Webseiten gesucht, die diesen Begriff enthalten: Dies können mehrere Millionen sein, also viel zu viele für eine einfache Anzeige.

Problem: Wie kann man automatisch die gefundenen Seiten so sortieren, dass nach Möglichkeit zuerst die interessanteren und erst danach die weniger interessanten auf dem Bildschirm erscheinen?

- Definiere Maß $W(P)$ für die Wichtigkeit einer Webseite P.
- Bestimmung von $W(P)$ muss nach einem formalen Verfahren durch einfache Rechnungen möglich sein (und nicht dadurch, dass jemand die Seite P liest und inhaltlich bewertet).
- Das Ganze muss sehr flexibel sein: es kommen ja ständig neue Webseiten und Links hinzu oder werden gelöscht.

Prinzipien zur Bestimmung von $W(P)$

- $W(P)$ steigt, je mehr andere (wichtige) Seiten auf P verweisen.
- Verweist eine andere Webseite P' auf P, so ergibt sich ein Beitrag von $W(P')$ zu $W(P)$, umgekehrt proportional zur Gesamtzahl der Seiten, auf die P' verweist.

Mathematisches Modell

Seien $P_1, \ldots, P_N$ die insgesamt verfügbaren Webseiten. Schreibe:

$P_j \to P_i$, wenn P_j einen Verweis auf P_i enthält;

ℓ_j = Anzahl aller Verweise von P_j auf andere Webseiten.

$$W(P_i) = \sum_{\text{alle } j \text{ mit } P_j \to P_i} \frac{W(P_j)}{\ell_j}$$

(D.h.: Verweist P_j auf P_i, so gibt P_j den ℓ_j-ten Bruchteil seiner eigenen Wichtigkeit an P_i weiter.)

- $W(P_i) = \sum_{j \text{ mit } (\ldots)} \dfrac{W(P_j)}{\ell_j}$ ist keine Formel, um die $W(P_i)$ einzeln auszurechnen. Es handelt sich um ein Gleichungssystem, durch das sich die $W(P_i)$ gegenseitig bestimmen.
- Sei A "Internet-Matrix": N Zeilen und N Spalten mit

  Eintrag in i-ter Zeile und j-ter Spalte

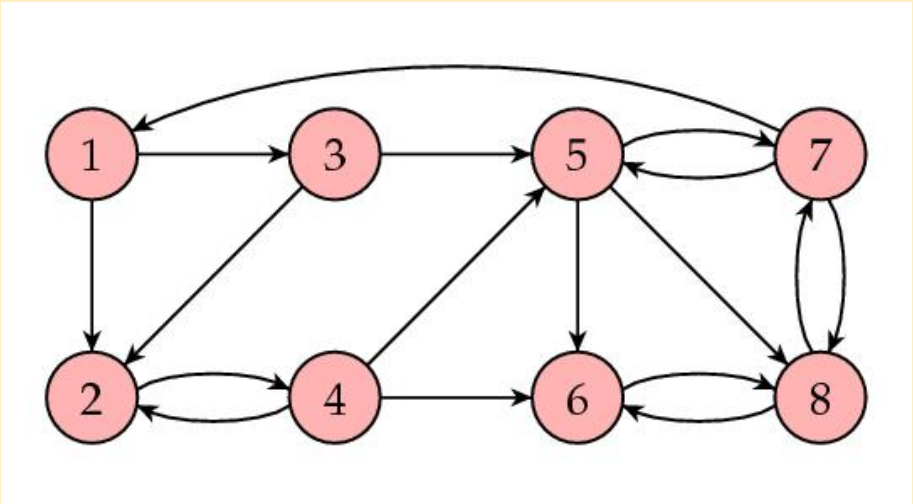
$$= \begin{cases} 1/\ell_j & \text{wenn } P_j \to P_i, \\ 0 & \text{sonst.} \end{cases}$$

  Dann ist A spalten-stochastisch und

$$A \cdot v = v \qquad \text{mit} \qquad v = \begin{pmatrix} W(P_1) \\ \vdots \\ W(P_N) \end{pmatrix}$$

$\rightsquigarrow$ v stationärer Vektor für A $\rightsquigarrow$ Ansatz zur Berechnung der $W(P_i)$.

Beispiel: Modell-Internet mit 8 Webseiten und Verweisen wie folgt:



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{pmatrix} \qquad v \approx \begin{pmatrix} 0.0600 \\ 0.0675 \\ 0.0300 \\ 0.0675 \\ 0.0975 \\ 0.2025 \\ 0.1800 \\ 0.2950 \end{pmatrix}$$

Also ist P_8 die wichtigste Seite! (Danach P_6, P_7, P_5, P_4, P_2, P_1, P_3)

(`http://www.ams.org/samplings/feature-column/fcarc-pagerank`)