Wolfgang K. Seiler

Algebra

Vorlesung an der Universität Mannheim
im Herbstsemester 2020

Dieses Skriptum entsteht parallel zur Vorlesung und soll mit möglichst geringer Verzögerung erscheinen. Es ist daher in seiner Qualität auf keinen Fall mit einem Lehrbuch zu vergleichen; insbesondere sind Fehler bei dieser Entstehensweise nicht nur möglich, sondern **sicher**. Dabei handelt es sich wohl leider nicht immer nur um harmlose Tippfehler, sondern auch um Fehler bei den mathematischen Aussagen. Da mehrere Teile aus anderen Skripten für Hörerkreise der verschiedensten Niveaus übernommen sind, ist die Präsentation auch teilweise ziemlich inhomogen.

Das Skriptum sollte daher mit Sorgfalt und einem gewissen Mißtrauen gegen seinen Inhalt gelesen werden. Falls Sie Fehler finden, teilen Sie mir dies bitte persönlich oder per e-mail (seiler@math.uni-mannheim.de) mit. Auch wenn Sie Teile des Skriptums unverständlich finden, bin ich für entsprechende Hinweise dankbar.

Falls genügend viele Hinweise eingehen, werde ich von Zeit zu Zeit Listen mit Berichtigungen und Verbesserungen zusammenstellen. In der online Version werden natürlich alle bekannten Fehler korrigiert.

Biographische Angaben von Mathematikern beruhen größtenteils auf den entsprechenden Artikeln im *MacTutor History of Mathematics archive* (`www-history.mcs.st-andrews.ac.uk/history/`), von wo auch die meisten abgedruckten Bilder stammen. Bei noch lebenden Mathematikern bezog ich mich, soweit möglich, auf deren eigenen Internetauftritt.

KAPITEL 0:WAS IST ALGEBRA? 1

KAPITEL I: KLASSISCHE LÖSUNGSFORMELN 3
§1: Lineare Gleichungen 3
§2: Lösung quadratischer Gleichungen 3
§3: Der Wurzelsatz von VIÈTE 6
§4: Kubische Gleichungen 13
§5: Biquadratische Gleichungen 29
§6: Gleichungen höheren Grades 31
§7: Symmetrische Polynome 32
§8: Die Diskriminante eines Polynoms 35
§9: Der casus irreducibilis bei kubischen Gleichungen 36

KAPITEL II: RECHNEN MIT GANZEN ZAHLEN 39
§1: Der EUKLIDische Algorithmus 41
§2: Die multiplikative Struktur der ganzen Zahlen 50
§3: Die Verteilung der Primzahlen 53
Anhang: Die EULERsche Summenformel und die STIRLINGsche Formel 66
§4: Das Sieb des ERATOSTHENES 68
§5: Kongruenzenrechnung 70
§6: Der kleine Satz von FERMAT 76
§7: Anwendungen in der Kryptographie 79

KAPITEL 3: GRUNDLEGENDE ALGEBRAISCHE STRUKTUREN 97
§1: Halbgruppen und Monoide 97
§2: Gruppen 100
§3: Ringe 119

KAPITEL IV: NULLSTELLEN UND KÖRPERERWEITERUNGEN 157
§1: Zerfällungskörper und der Fundamentalsatz der Algebra 157
§2: Automorphismen von Körpererweiterungen 168
§3: Lösbarkeit von Gleichungen durch Radikale 184
§4: Konstruktionen mit Zirkel und Lineal 191
§5: Transzendente Zahlen und die Quadratur des Kreises 205
§6: Endliche Körper 218
§7: Mehr über Einheitswurzeln 218

KAPITEL V: DIE FERMAT-VERMUTUNG FÜR ZAHLEN
UND FÜR POLYNOME 287
§1: Zahlen und Funktionen 237
§2: Der Satz von MASON 239
§3: Die abc-Vermutung 242
§4: Die FREY-Kurve 247

Kapitel 0
Was ist Algebra?

Die Zahlentheorie beschäftigt sich mit den (ganzen) Zahlen, die Geometrie von γεωμετρία mit dem Messen der Erde, aber woher kommt das Wort Algebra?

Um 830 legte der arabische Gelehrte ABU DSCHA'FAR MUḤAMMAD IBN MŪSĀ AL-CHWĀRIZMĪ sein zweites Buch *Al-Kitāb al-muchtasar fi hisab al-dschabr wa-'l-muqābala* oder kurz *Kitāb al-dschabr wa-'l-muqābala* vor; *al-dschabr* gab der Algebra ihren Namen, und der Autorenname AL-CHWĀRIZMĪ führte zum Wort Algorithmus. In deutscher Übersetzung heißt der volle Titel etwa *Kurzgefaßtes Buch über das Rechnen durch Ergänzen und Ausgleichen. Al-dschabr,* das Ergänzen oder Vervollständigen, besteht darin, negative Terme in einer Gleichung auf die andere Seite zu bringen; in einem Beispiel aus dem Buch wird etwa aus (in moderner Schreibweise) $x^2 = 40x - 4x^2$ durch *al-dschabr* die Gleichung $5x^2 = 40x$. *Al-muqābala,* das Ausgleichen, besteht darin, von zwei positiven Termen auf den beiden Seiten der Gleichung den einen auf Null zu reduzieren; aus $x^2+3x+5 = 7x+2$ wird also zunächst $x^2 + 5 = 4x + 2$ und dann $x^2 + 3 = 4x$.

ABU DSCHA'FAR MUHAMMAD IBN MŪSĀ AL-CHWĀRIZMĪ wurde um 780 geboren und arbeitete die meiste Zeit seines Lebens in Bagdad, insbesondere auch im *Haus der Weisheit,* das AL-MA'MŪM, der siebte Kalif, als wissenschaftliches Zentrum seines Reichs gegründet hatte. Eine der Aufgaben dieses Zentrums bestand darin, Texte klassischer griechischer Wissenschaftler ins Arabische zu übersetzen; viele Texte sind heute nur noch über diese Übersetzungen bekannt. Arbeitsgebiete am *Haus der Weisheit* waren vor allem Mathematik und Astronomie. Außer einem weiteren mathematischen Buch, das sich mit den indischen Ziffern befaßte, schrieb AL-CHWĀRIZMĪ auch Bücher über Geographie und Kartographie.

Aus heutiger Sicht besteht kaum ein Unterschied zwischen *al-dschabr*

und *al-muqābala;* wir sagen einfach, daß wir einen Term auf die andere Seite bringen. Demnach ist Algebra also die Lehre vom auf die andere Seite bringen. Im neunten Jahrhundert waren die beiden Methoden noch grundverschieden, denn negative Zahlen begannen außerhalb Indiens erst im 16. Jahrhundert langsam in der Mathematik aufzutauchen. Auch die Null fing gerade erst an verwendet zu werden; davon handelt AL-CHWĀRIZMĪs erstes Buch, in dem er die indische Zahlenschrift in die arabische Welt brachte. Die Null wurde aber nicht als *Zahl* eingeführt, sondern nur als *Ziffer.* Dieses Wort kommt vom arabischen Wort für Null *ṣifr,* was von *ṣafira = leer sein* kommt, und das wiederum kommt vom Sanskrit-Wort *śūnya,* das Nichts oder die Leere.

In heutiger Sprechweise ist der Gegenstand des Buchs von AL-CHWĀRIZMĪ die Lösung linearer und quadratischer Gleichungen. Quadrate der Unbekannten kommen in seinen Gleichungen entweder gar nicht oder ohne Koeffizient (d.h. mit Koeffizient eins) vor; lineare und konstante Terme können auf beiden Seiten mit positiven Koeffizienten stehen, können aber auch fehlen. Da es die Null als Zahl noch nicht gab, konnte sie auch auf keiner der beiden Seiten stehen.

Unter diesen Randbedingungen entstehen Gleichungen, die sich durch *al-dschabr* und *al-muqābala* auf eine der folgenden sechs Normalformen bringen lassen:

$$x^2 = px,\quad x^2 = q,\quad px = q,\quad x^2+px = q,\quad x^2+q = px \quad \text{und} \quad x^2 = px+q\,,$$

wobei p und q natürlich positive Zahlen sein müssen. Vom Lösen dieser sechs Typen von Gleichungen handelt das Buch. AL-CHWĀRIZMĪ benutzt dabei häufig eine geometrische Sprechweise und veranschaulicht seine Vorgehensweise auch geometrisch.

Die nach *al-dschabr* benannte Algebra befaßte sich somit traditionell mit dem Lösen von Gleichungen. Erst im neunzehnten Jahrhundert begann man sich auch für in diesem Zusammenhang auftretende strukturelle Fragen zu interessieren. Was wir heute als *abstrakte Algebra* bezeichnen, geht größtenteils erst auf den Beginn des zwanzigsten Jahrhunderts zurück.

Kapitel 1
Klassische Lösungsformeln

Wir beginnen mit dem klassischen Grundproblem der Algebra, dem Lösen von Polynomgleichungen in einer Variablen, d.h. Gleichungen der Form

$$a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 = 0 \,.$$

Über den Zahlbereich, in dem die Koeffizienten liegen, wollen wir uns dabei im Augenblick noch keine großen Gedanken machen. In den meisten Beispielen werden die Koeffizienten ganze, rationale, reelle oder komplexe Zahlen sein; der Zahlbereich könnte aber auch einfach irgendein Körper oder Ring sein. Um Lösungen zu finden, müssen wir oft auch in einem größeren Zahlbereich suchen: Die Gleichung $2x - 3 = 0$ hat beispielsweise ganzzahlige Koeffizienten, aber keine ganzzahligen Lösungen, sondern nur die rationale Lösung $x = \frac{2}{3}$.

Wir werden stets annehmen, daß a_d nicht verschwindet und bezeichnen dann d als den *Grad* der Gleichung.

§1: Lineare Gleichungen

Gleichungen vom Grad eins oder lineare Gleichungen sind problemlos zu lösen: Da gemäß unserer Annahme in $ax + b = 0$ der Koeffizient a von x nicht verschwindet, können wir (eventuell erst nach Übergang zu einem größeren Zahlbereich) b auf die andere Seite bringen (je nach Vorzeichen von b ist das *al-dschabr* oder *al-muqābala*) und dann beide Seiten durch a dividieren, um die Lösung $x = -\frac{b}{a}$ zu erhalten.

§2: Quadratische Gleichungen

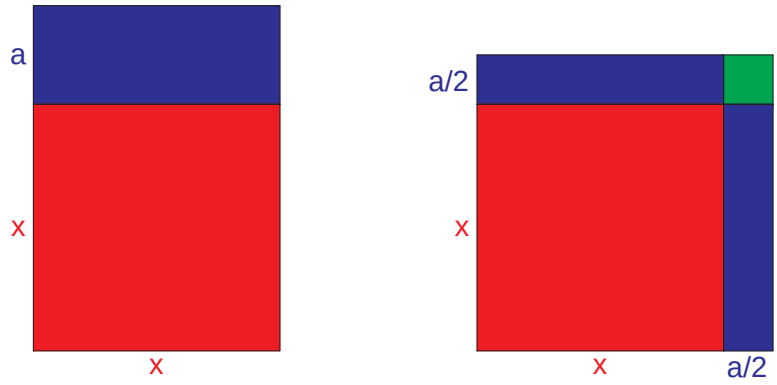
Gleichungen vom Grad zwei werden üblicherweise als quadratische

Gleichungen bezeichnet; Verfahren zur ihrer Lösung waren in allen frühen Hochkulturen bekannt. Die ältesten erhaltenen Hinweise deuten darauf hin, daß die Babylonier schon vor rund vier Jahrtausenden damit vertraut waren.

Der Ansatz zur Lösung der Gleichung $x^2 + ax = b$ läßt sich am einfachsten geometrisch verstehen: Wir suchen nach einem Quadrat mit unbekannter Seitenlänge x derart, daß die Fläche des Quadrats zusammen mit der des Rechtecks mit Seiten x und a gleich b ist.

Die linke unter den beiden folgenden Zeichnungen zeigt dieses Quadrat und darüber das Rechteck; auf der rechten Seite ist die Hälfte des Rechtecks neben das Quadrat gewandert, so daß abgesehen von dem kleinen Quadrat rechts oben nun ein Quadrat mit Seitenlänge $x + \frac{a}{2}$ entstanden ist. Die Größe des kleinen Quadrats ist bekannt: Seine Seitenlänge ist $\frac{a}{2}$. Wir suchen somit eine Zahl x derart, daß das Quadrat mit Seitenlänge $x+\frac{a}{2}$ die Fläche $b+\frac{a^2}{4}$ hat; das Problem ist also zurückgeführt auf das Ziehen einer Quadratwurzel:

$$x = -\frac{a}{2} \pm \sqrt{b + \frac{a^2}{4}}\,.$$



(Eine ähnliche Zeichnung befindet sich übrigens auch im Buch von AL-CHWĀRIZMĪ; er teilt das Rechteck mit Seiten a und x allerdings auf in vier Rechtecke mit Seiten $a/4$ und x und setzt diese an die vier Seiten des Quadrats. Das gibt eine etwas schönere Zeichnung, dafür muß er vier Quadrate mit Seitenlänge $a/4$ hinzufügen, um auf ein Quadrat mit Seitenlänge $x + a/2$ zu kommen.)

Wie die Babylonier auf diese Lösungsformel kamen, ist nicht bekannt; in den überlieferten Schriften wird nur der fertige Lösungsweg anhand von Beispielen präsentiert. Sie wußten aber auf jeden Fall, daß die Summe der Lösungen der Gleichung $x^2 - ax + b = 0$ gleich a ist und ihr Produkt gleich b – einen Beweis in einem allgemeineren Zusammenhang werden wir in Kürze kennen lernen. Damit ist das Lösen der Gleichung $x^2 - ax + b$ äquivalent dazu, zwei Zahlen x_1 und x_2 zu finden mit

$$x_1 + x_2 = a \quad \text{und} \quad x_1 x_2 = b\,.$$

Die führt zu einer alternativen Herleitung der Lösungsformel: Wir machen den Ansatz

$$x_{1/2} = \frac{a}{2} \pm u$$

mit einer neuen Unbekannten u; damit ist die erste Gleichung automatisch erfüllt. Für die zweite erhalten wir nach der den Babyloniern bekannten dritten binomischen Formel

$$b = x_1 x_2 = \left(\frac{a}{2} + u\right)\left(\frac{a}{2} - u\right) = \frac{a^2}{4} - u^2\,, \quad \text{also} \quad u = \sqrt{\frac{a^2}{4} - b}\,.$$

Somit ist $x_{1/2} = \dfrac{a}{2} \pm \sqrt{\dfrac{a^2}{4} - b}$.

Falls $\frac{a^2}{4} - b > 0$ ist, liefert uns das zwei reelle Lösungen; falls der Radikand Null ist, fallen beide zusammen. Für die Babylonier, die ihre Mathematik benutzten, um Größen aus der realen Welt zu berechnen, ging es nur um reelle Lösungen; heute interessieren wir uns auch für Gleichungen, bei denen unter der Wurzel eine negative oder sogar eine komplexe Zahl steht. Im Falle einer negativen Zahl ist die Wurzel rein imaginär und somit problemlos; für eine komplexe Zahl allerdings stellt sich die Frage, wie wir die Wurzel aus $c + di$ mit $c, d \in \mathbb{R}$ und $d \neq 0$ in der Form $u + iv$ mit $u, v \in \mathbb{R}$ darstellen können. Die Gleichung

$$(u + iv)^2 = u^2 - v^2 + 2iuv = c + id$$

führt auf die beiden reellen Gleichungen

$$u^2 - v^2 = c \quad \text{und} \quad 2uv = d\,.$$

Wegen $d \neq 0$ können auch u und v nicht verschwinden; daher können wir die zweite Gleichung umformen zu $v = d/2u$ und das in die erste Gleichung einsetzen:

$$u^2 - \frac{d^2}{4u^2} = c\,.$$

Multiplikation mit $4u^2$ macht daraus

$$4u^4 - d^2 = 4cu^2 \quad \text{oder} \quad u^4 - cu^2 - \frac{d^2}{4} = 0\,.$$

Dies ist eine quadratische Gleichung für u^2 mit den beiden Lösungen

$$u^2 = \frac{c}{2} \pm \frac{1}{2}\sqrt{c^2 + d^2}\,.$$

Als Quadrat einer reellen Zahl muß $u^2 \geq 0$ sein; die Lösung mit dem Minuszeichen kommt daher nicht in Frage: Für negatives $c \in \mathbb{R}$ ist sie offensichtlich negativ, und für positives c auch, denn wegen $d \neq 0$ ist $\sqrt{c^2 + d^2}$ größer als der Betrag von c. Somit sind

$$u = \pm\sqrt{\frac{c}{2} + \frac{1}{2}\sqrt{c^2 + d^2}} \quad \text{und} \quad v = \frac{d}{2u}$$

problemlos berechenbar.

§3: Der Wurzelsatz von Viète

Während Lösungsverfahren für quadratische Gleichungen seit Jahrtausenden bekannt sind, tauchte die erste allgemeine Formel zur Lösung einer kubischen Gleichung erst vor gut 500 Jahren auf. Das lag nicht daran, daß sich vorher niemand dafür interessierte: Die klassische griechische Mathematik etwa kannte eine ganze Reihe von Problemen, die auf Gleichungen dritten Grades führten, und sie kannte auch geometrische Lösungsverfahren dafür. Diese Verfahren kamen allerdings nicht mit Zirkel und Lineal aus, so daß sie als weniger „rein“ und damit auch weniger interessant galten. Von einer allgemeinen Lösungsformel war man weit entfernt.

In speziellen Fällen lassen sich aber gelegentlich leicht Lösungen auch von Gleichungen sehr hohen Grades finden. Ausgangspunkt dazu ist eine Umkehrung des Problems: Wir fragen uns nicht, wie wir aus den

Koeffizienten der Gleichung die Lösungen bestimmen können, sondern wie wir aus den Lösungen die Koeffizienten erhalten.

Dazu erinnern wir uns zunächst an die den meisten wohl aus der Schule oder aus Anfängervorlesungen bekannte Polynomdivision mit Rest: Ein Polynom ist bekanntlich eine formale Summe

$$f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0$$

mit Koeffizienten $a_i \in k$ und einem „Symbol“ X. Falls alle a_i verschwinden, reden wir vom *Nullpolynom;* ansonsten nehmen wir, wie bei den Gleichungen, an, daß a_d nicht verschwindet und bezeichnen $d = \deg f$ als den *Grad* und a_d als den führenden Koeffizienten von f. Das Nullpolynom hat keinen Grad.

Oft ist es üblich, den Buchstaben x sowohl als Bezeichnung für eine Variable als auch für eine konkrete Lösung einer Gleichung zu verwenden. Die folgenden Überlegungen werden aber wohl klarer, wenn wir zwischen den beiden Bedeutungen unterscheiden: Große Buchstaben stehen für Variablen und kleine für Zahlen. Der Wert des obigen Polynoms f an der Stelle x ist somit die Zahl

$$f(x) = x^d + a_{d-1} x^{d-1} + a_{d-2} x^{d-2} + \cdots + a_2 x^2 + a_1 x + a_0 \,.$$

Nun seien f und g Polynome mit Koeffizienten aus einem Körper k mit $\deg f = d$ und $\deg g = e$. Der Divisionsalgorithmus konstruiert dazu Polynome q und r mit Koeffizienten aus k für die $f = qg + r$ gilt, wobei r entweder das Nullpolynom ist oder einen kleineren Grad als g hat. Wir bezeichnen q als den *Quotienten* und r als den *Divisionsrest.* Sie werden wie folgt bestimmt:

Schritt 0: Setze $r = f$ und $q = 0$. b_e sei der führende Koeffizient von g.

Schritt i**,** $i \geq 1$**:** Falls $r = 0$ ist oder $\deg r < \deg g$, endet der Algorithmus. Andernfalls sei a der führende Koeffizient von r. Wir eliminieren den führenden Term von r, indem wir r ersetzen durch $r - \frac{a}{b_e} X^{\deg r - e} g$. Gleichzeitig ersetzen wir q durch $q + \frac{a}{b_e} X^{\deg r - e}$.

Dieser Algorithmus endet nach endlich vielen Schritten, denn in jedem Schritt ab dem ersten wird der Summand von r mit der höchsten X-Potenz eliminiert, so daß der Grad von r um mindestens eins kleiner wird

oder r sogar zum Nullpolynom wird. Nach endlich vielen Schritten ist daher entweder $r = 0$ oder $\deg r < \deg g$, so daß der Algorithmus endet.

Nach Schritt 0 ist $qg + r = 0 \cdot g + f = f$, und wenn diese Gleichung $f = qg + r$ vor Beginn des i-ten Schritts gilt, gilt sie auch danach, denn q wird ersetzt durch $q + \frac{a}{b_e} X^{\deg r - e}$ und r durch $r - \frac{a}{b_e} X^{\deg r - e} g$, und

$$\left(q + \frac{a}{b_e} X^{\deg r - e} \right) q + \left(r - \frac{a}{b_e} X^{\deg r - e} g \right) = qg + r = f \, .$$

Nach Beendigung des Algorithmus ist außerdem noch $r = 0$ oder $\deg r < \deg e$, so daß der Algorithmus das gewünschte Ergebnis liefert.

Da wir wiederholt durch b_e dividieren, wobei die Werte von a von Schritt zu Schritt variieren können, mußten wir annehmen, daß die Koeffizienten in einem Körper liegen. Es gibt allerdings eine Ausnahme: Falls der führende Koeffizient von g gleich eins ist, sind keine Divisionen notwendig, und der Algorithmus funktioniert auch bei Koeffizienten aus einem (kommutativen) Ring wie beispielsweise den ganzen Zahlen.

Das wollen wir anwenden auf die Nullstellen eines Polynoms, die im betrachteten Koeffizientenbereich liegen. x sei also eine solche Nullstelle des Polynoms

$$f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0 \, ,$$

das heißt

$$f(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 = 0 \, .$$

Wir wenden den Divisionsalgorithmus an auf f und $g = X - x$. Er liefert Polynome q, r derart, daß $f = qg + r$ ist mit $r = 0$ oder $\deg r < \deg g = 1$. Der Divisionsrest r ist also in jedem Fall eine Konstante. Wenn wir x einsetzen, ergibt sich $0 = f(x) = q(x)g(x) + r = r$, da $g(x) = x - x = 0$ ist. Somit gibt es ein Polynom q derart, daß $f = q \cdot (X - x)$ ist. Falls dabei auch $q(x) = 0$ ist, gibt es ein weiteres Polynom q_2, so daß $q = q_2 \cdot (X - x)$ ist und damit $f = q_2 \cdot (X - x)^2$. Wenn auch $q_2(x)$ verschwindet, können wir weitermachen, bis wir schließlich eine Darstellung $f = q_n \cdot (X - x)^n$ erhalten mit $q_n(x) \neq 0$. Wir sagen dann, x sei eine n-fache Nullstelle oder die Vielfachheit der Nullstelle x sei n.

Falls wir eine weitere Nullstelle $x' \neq x$ von f kennen, muß $q_n(x')$ verschwinden, denn $X - x$ verschwindet natürlich nicht an der Stelle x'. Wenn wir ℓ verschiedene Nullstellen $x_1, \ldots, x_\ell$ kennen mit Vielfachheiten $n_1, \ldots, n_\ell$, erhalten wir somit eine Darstellung

$$f = \widetilde{q} \cdot (X - x_1)^{n_1} \cdots (X - x_\ell)^{n_\ell} .$$

Ist d der Grad von f, so ist $d = \deg \widetilde{q} + n_1 + \cdots n_\ell$; damit folgt

Lemma: Die Anzahl der Nullstellen eines Polynoms vom Grad d ist, auch mit Vielfachheiten gezählt, höchstens gleich dem Grad. ∎

Wir werden später sehen, daß es stets einen Körper gibt, in dem die Anzahl der Nullstellen des Polynoms mit Vielfachheiten gezählt gleich dem Grad ist. Für Polynome mit reellen Koeffizienten ist das beispielsweise der Körper $\mathbb{C}$ der komplexen Zahlen, aber es gibt stets auch noch deutlich kleinere Körper mit dieser Eigenschaft.

Die Tatsache, daß $X - x$ für eine Nullstelle x eines Polynoms f ein Teiler von f ist, läßt sich für Polynome mit ganzzahligen Koeffizienten zum Erraten zumindest der ganzzahligen Nullstellen verwenden:

Lemma: $f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0$ sei ein Polynom mit ganzzahligen Koeffizienten. Falls $f(x)$ für eine ganze Zahl $x \neq 0$ verschwindet, ist x ein Teiler von a_0. ($f(0) = 0$ ist natürlich äquivalent zu $a_0 = 0$.)

Beweis: Wie wir oben gesehen haben, gibt es ein Polynom q derart, daß $f = q \cdot (X - x)$ ist. Da $X - x$ den höchsten Koeffizienten eins hat und $x \in \mathbb{Z}$, entstehen in jedem Schritt des Divisionsalgorithmus wieder Polynome mit ganzzahligen Koeffizienten; daher hat auch der Quotient q ganzzahlige Koeffizienten. Ist b_0 der konstante Koeffizient von q, so ist a_0 das Produkt von b_0 mit dem konstanten Koeffizienten $-x$ von $X - x$, d.h. $a_0 = -b_0 x$, d.h. a_0 ist ein Vielfaches von x und x damit ein Teiler von a_0. ∎

Als Beispiel betrachten wir die kubische Gleichung $x^3 - 7x + 6 = 0$. Das Polynom $f = X^3 - 7X + 6$ hat den konstanten Koeffizienten 6;

falls es ganzzahlige Nullstellen gibt, müssen diese also unter den Zahlen $\pm 1, \pm 2, \pm 3$ und ± 6 sein. Einsetzen zeigt, daß $f(1) = f(2) = 0$, $f(-1) = f(-2) = f(3) = 12$ und $f(-3) = 0$ ist. Da es nicht mehr als drei Nullstellen geben kann, hat die kubische Gleichung daher die drei Lösungen $x_1 = 1$, $x_2 = 2$ und $x_3 = -3$.

Bei einer Gleichung ohne ganzzahlige Lösungen führt dieser Ansatz natürlich nicht zum Ziel, aber da wir im Voraus nicht wissen, ob es ganzzahlige Lösungen gibt, ist er doch oft einen Versuch wert. Selbst wenn wir damit nur einen Teil der Nullstellen finden, können wir die zugehörigen Linearfaktoren abdividieren und erhalten für die restlichen Nullstellen eine Gleichung kleineren Grades.

Tatsächlich läßt sich die Methode noch ausbauen. Wir nehmen dazu an, wir hätten ein Polynom f, das über einem hinreichend großen Körper in Linearfaktoren zerfällt, d.h.

$$\begin{aligned} f &= a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0 \\ &= a_d (X - x_1)(X - x_2) \cdots (X - x_d)\,, \end{aligned}$$

wobei die x_i natürlich nicht alle verschieden sein müssen. Ausmultiplizieren und Koeffizientenvergleich führt auf die Gleichungen

$$\begin{aligned} a_{d-1} &= -a_d \sigma_1(x_1, \ldots, x_d) \quad \text{mit} \quad \sigma_1(x_1, \ldots, x_d) = x_1 + \cdots + x_n \\ a_{d-2} &= a_d \sigma_2(x_1, \ldots, x_d) \quad \text{mit} \quad \sigma_2(x_1, \ldots, x_d) = \sum_{i<j} x_i x_j \\ a_{d-3} &= -a_d \sigma_3(x_1, \ldots, x_d) \quad \text{mit} \quad \sigma_3(x_1, \ldots, x_d) = \sum_{i<j<k} x_i x_j x_k \\ &\vdots \qquad\qquad \vdots \\ a_0 &= (-1)^d a_d \sigma_d(x_1, \ldots, x_d) \quad \text{mit} \quad \sigma_d(x_1, \ldots, x_d) = x_1 \cdots x_d\,. \end{aligned}$$

Allgemein ist a_{d-r} bis aufs Vorzeichen gleich der Summe alle Produkte aus r Werten x_i mit verschiedenem Index. Diese Summen bezeichnet man als die *elementarsymmetrischen Funktionen* $\sigma_r(x_1, \ldots, x_d)$ und die obigen Gleichungen als den Wurzelsatz von VIÈTE.

FRANÇOIS VIÈTE (1540–1603) studierte Jura an der Universität Poitiers, danach arbeitete er als Hauslehrer. 1573, ein Jahr nach dem Massaker an den Hugenotten, berief ihn CHARLES IX (obwohl VIÈTE Hugenotte war) in die Regierung der Bretagne; unter HENRI III wurde er geheimer Staatsrat. 1584 wurde er auf Druck der katholischen Liga vom Hofe verbannt und beschäftigte sich fünf Jahre lang nur mit Mathematik. Unter HENRI IV arbeitete er wieder am Hof und knackte u.a. verschlüsselte Botschaften an den spanischen König PHILIP II. In seinem Buch *In artem analyticam isagoge* rechnete er als erster systematisch mit symbolischen Größen, führte also die „Buchstabenrechnung“ ein. Auch die mathematische Formelschreibweise geht auf ihn zurück, insbesondere auch die Zeichen „+“ und „-“ für Addition und Subtraktion.

Für eine quadratische Gleichung $x^2 + px + q = 0$ besagt der Satz von VIÈTE einfach, daß die Summe der Lösungen gleich $-p$ und das Produkt gleich q ist, was die Babylonier schon vor rund vier Jahrtausenden die wußten.

Diese elementarsymmetrischen Funktionen sind für r-Werte im mittleren Bereich recht umfangreiche Summen, die beiden Fälle $r = 0$ und $r = d - 1$ können aber gelegentlich sehr nützlich sein, um Lösungen zu erraten, vor allem wenn $a_d = 1$ ist:

Falls wir aus irgendeinem Grund erwarten, daß alle Nullstellen eine Polynoms mit ganzzahligen Koeffizienten ganzzahlig sind, ist ihr Produkt gleich $(-1)^d a_0/a_d$ ist und ihre Summe gleich $-a_{d-1}/a_d$.

Bei der oben betrachteten Gleichung $f(x) = x^3 - 7x + 6 = 0$ etwa ist das Produkt aller Nullstellen gleich -6 und ihre Summe verschwindet. Aus den Zahlen $\pm 1, \pm 2, \pm 3$ und ± 6 müssen wir also drei (nicht notwendigerweise verschiedene) finden mit Summe Null und Produkt -6. Das geht offensichtlich nur mit $1, 2$ und -3; Einsetzen zeigt, daß dies auch tatsächlich Nullstellen sind. Damit mußten wir nur drei Zahlen einsetzen, statt wie oben sechs, um alle Lösungen zu finden.

Man beachte, daß dieses Einsetzen unbedingt notwendig ist: Bei der Gleichung $g(x) = x^3 - 6x + 6 = 0$ hätten wir genauso vorgehen können und wären auf dieselben drei Kandidaten gekommen, aber $g(1) = 1, g(2) = 2$ und $g(-3) = -3$. (Daß die Lösungsmenge

nicht$\{1, 2, -3\}$ sein kann, erkennt man auch daran, daß

$$\sigma_2(1, 2, -3) = 1 \cdot 2 + 1 \cdot (-3) + 2 \cdot (-3) = -7$$

nicht gleich dem Koeffizienten Null von x^2 ist.)

Auch die Gleichung $x^3 - 21x - 20 = 0$ läßt sich leicht nach VIÈTE lösen: Hier ist das Produkt aller Nullstellen gleich 20; *falls* sie alle ganzzahlig sind, kommen also nur $\pm 1, \pm 2, \pm 4, \pm 5, \pm 10$ und ± 20 in Frage. Aus diesen zwölf Zahlen müssen wir drei (nicht notwendigerweise verschiedene) auswählen mit Produkt 20 und Summe null. Das geht offensichtlich nur mit $-1, -4$ und 5, und wieder zeigt Einsetzen, daß dies auch tatsächlich Nullstellen sind.

Betrachten wir als nächstes Beispiel das Polynom

$$f = X^4 + 14X^3 - 52X^2 - 14X + 51$$

mit $a_0 = 51 = 3 \cdot 17$. Da das Produkt aller Nullstellen diesen Wert haben muß, kommen – *falls* alle Nullstellen ganzzahlig sind – für diese nur die Werte $\pm 1, \pm 3, \pm 17$ und ± 51 in Frage. Wäre eine der Nullstellen ± 51, müßten alle anderen den Betrag eins haben und die Summe könnte nicht gleich -14 sein. Daher muß eine Nullstelle Betrag drei und eine Betrag 17 haben, die beiden anderen Betrag eins. Produkt 51 und Summe -14 erzwingt dabei offensichtlich, daß sowohl +1 als auch -1 Nullstellen sind, außerdem -17 und +3. Einsetzen zeigt, daß alle vier auch tatsächlich Nullstellen sind.

Beim Polynom $X^3 - 3X - 2$ ist das Produkt aller Nullstellen -2 und die Summe verschwindet. Die Teiler von -2 sind ± 1 und ± 2; Einsetzen zeigt, daß nur -1 und 2 Nullstellen sind. Sowohl aus dem Verschwinden der Summe als auch aus dem Produkt -2 folgt, daß -2 eine doppelte Nullstelle sein muß, d.h. $X^3 - 3X - 2 = (X + 1)^2(X - 2)$.

Beim Polynom

$$f = X^6 + 27X^5 - 318X^4 - 5400X^3 - 10176X^2 + 27648X + 32768$$

ist $a_0 = 32768 = 2^{15}$; hier wissen wir also nur, daß – sofern alle Nullstellen ganzzahlig sind – jede Nullstelle die Form $\pm 2^i$ haben muß, wobei

die Summe aller Exponenten gleich 15 sein muß und die Anzahl der negativen Vorzeichen gerade. Einsetzen zeigt, daß

$$-1, \quad 2, \quad -4, \quad -8, \quad 16, \quad -32$$

die Nullstellen sind.

Gelegentlich lassen sich auch nicht ganzzahlige Nullstellen mit Hilfe des Satzes von VIÈTE erraten: Bei Polynom $X^4+X^3-7X^2-5X+10$ ist das Produkt der Nullstellen zehn. Wie Einsetzen zeigt, sind 1 und -2 Nullstellen. Ihre Summe ist -1 und ihr Produkt -2. Die beiden restlichen Nullstellen haben somit die Summe Null und das Produkt -5. Wir suchen also Zahlen x_3 und x_4 mit $x_4 = -x_3$ und $x_3x_4 = -x_3^2 = -5$. Daher müssen x_3 und x_4 gleich $\pm\sqrt{5}$ sein.

Man beachte, daß die Anwendung des Satzes von VIÈTE nur deshalb so gut funktionierte, weil die betrachteten Polynome höchsten Koeffizient eins hatten. Ist das nicht der Fall, ist das Produkt der Nullstellen gleich dem Quotienten aus konstantem Koeffizienten und führendem Koeffizienten mal $(-1)^{\text{Grad}}$, und wenn das keine ganze Zahl ist, können wir nicht mehr mit Teilbarkeit argumentieren, sondern müssen uns auf der Suche nach rationalen Lösungen auch mit den möglichen Nennern beschäftigen

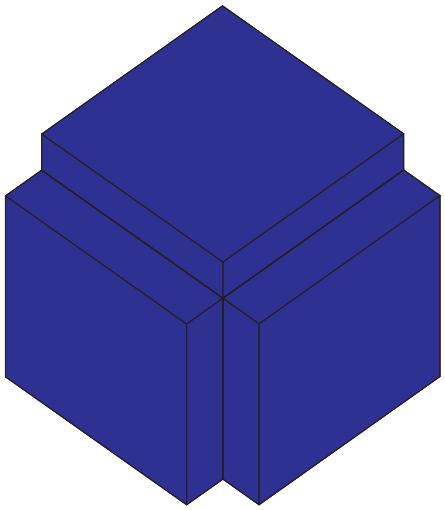
§4: Kubische Gleichungen

Wie bereits zu Beginn des vorigen Paragraphen erwähnt, gibt es seit rund 500 Jahren, genauer seit 1515, auch Ansätze zur Lösung allgemeiner kubischer Gleichungen.

Wenn wir versuchen, für die Gleichungen $x^3 + ax^2 = b$ eine ähnlich Strategie zu finden wie im Fall der Gleichung $x^2 + ax = b$, müssen wir ins Dreidimensionale gehen und auf den Würfel mit Kantenlänge x eine quadratische Säule mit Basisquadrat der Seitenlänge x und Höhe a stellen. Um sie so zu verteilen, daß wir möglichst nahe an einen neuen Würfel kommen, müssen wir jeweils ein Drittel davon auf drei der Seitenflächen des Würfels platzieren.

Leider fehlt hier nun nicht nur ein Würfel der Kantenlänge $\frac{a}{3}$, sondern auch noch drei quadratische Säulen der Höhe x auf Grundflächen mit

Seitenlänge $\frac{a}{3}$. Wir können das Volumen des Würfels mit Seitenlänge $x + \frac{a}{3}$ also nicht einfach durch die bekannten Größen a, b ausdrücken, sondern haben auch noch einen Term mit der Unbekannten x.



Trotzdem ist diese Idee nützlich, sogar für Gleichungen höheren Grades. Die allgemeine Gleichung d-ten Grades hat die Form

$$a_d x^n + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 = 0\,,$$

wobei wir natürlich wie immer voraussetzen, daß a_d nicht verschwindet. Falls wir über einem Körper arbeiten, können wir durch a_d dividieren und erhalten die neue Gleichung

$$x^d + c_{d-1} x^{d-1} + \cdots + c_1 x + c_0 = 0 \quad \text{mit} \quad c_i = \frac{a_i}{a_d}\,,$$

deren höchster Koeffizient eins ist.

Geometrisch betrachtet wollen wir einen d-dimensionalen Hyperwürfel bekommen, dessen Seitenlänge $x + \frac{c_{d-1}}{d}$ sein sollte; rechnerisch bedeutet dies, daß wir die Zahl $y = x + \frac{c_{d-1}}{d}$ betrachten und überall in der Gleichung x durch $y - \frac{c_{d-1}}{d}$ ersetzen:

$$\begin{aligned}
&x^d + c_{d-1} x^{d-1} + c_{d-2} x^{d-2} + \cdots + c_1 x + c_0 \\
&= \left(y - \frac{c_{d-1}}{d}\right)^d + c_{d-1}\left(y - \frac{c_{d-1}}{d}\right)^{d-2} + \cdots + c_1\left(y - \frac{d_{n-1}}{d}\right) + c_0 \\
&= \left(y^d - c_{d-1} y^{d-1} + d c_{d-1}^2 y^{d-2} + \cdots\right) \\
&\quad + c_{d-1}\left(y^{d-1} - \frac{(d-1)c_{d-1}}{d} y^{d-2} + \cdots\right)
\end{aligned}$$

$$
\begin{aligned}
&+ c_{d-2}\left(y^{d-2} - \frac{(d-2)c_{d-1}}{d}y^{d-3} + \cdots\right)\\
&\qquad + \cdots\\
&= y^d + \left(dc_{d-1}^2 - \frac{(d-1)c_{d-1}^2}{d} + c_{d-2}\right)y_{d-2} + \cdots\ .
\end{aligned}
$$

Wir kommen also auf eine Gleichung d-ten Grades in y, die keinen Term mit y^{d-1} hat.

Im Falle $d = 2$ hat diese Gleichung die Form $y^2 + p = 0$; wir können ihre Nullstellen also einfach durch Wurzelziehen ermitteln. Für $d > 2$ haben wir immerhin einen Term weniger als in der allgemeinen Gleichung d-ten Grades und müssen sehen, ob uns das bei der Lösung helfen kann.

Im Falle der kubischen Gleichungen reicht es somit, die etwas speziellere Gleichung

$$y^3 + py + q = 0$$

zu lösen. Auch wenn die Griechen geometrische Konstruktionen (jenseits von Zirkel und Lineal) kannten, mit denen sie Lösungen kubischer Gleichungen konstruieren konnten, dauerte es bekanntlich bis ins 16. Jahrhundert, bevor eine explizite Lösungsformel gefunden war – ein Zeichen dafür, daß der Lösungsansatz nicht gerade offensichtlich ist.

Der Trick, der schließlich zum Erfolg führte, ist folgender: Wir schreiben y als Summe zweier neuer Zahlen u und v und machen dadurch das Problem auf den ersten Blick nur schwieriger. Andererseits ist diese Summendarstellung natürlich alles andere als eindeutig; wir können daher hoffen, daß es auch dann noch Lösungen gibt, wenn wir an u und v zusätzliche Forderungen stellen und dadurch das Problem vielleicht vereinfachen.

Einsetzen von $y = u + v$ führt auf die Bedingung

$$(u+v)^3 + p(u+v) + q = u^3 + 3u^2v + 3uv^2 + v^3 + p(u+v) + q = 0\,.$$

Dies können wir auch anders zusammenfassen als

$$(u^3 + v^3 + q) + (3uv + p)(u+v) = 0\,,$$

und natürlich verschwindet diese Summe insbesondere dann, wenn beide Summanden einzeln verschwinden. Falls es uns also gelingt, zwei Zahlen u, v zu finden mit

$$u^3 + v^3 = -q \quad \text{und} \quad 3uv = -p\,,$$

haben wir eine Lösung gefunden.

Zwei solche Zahlen u, v erfüllen erst recht die schwächere Bedingung

$$u^3 + v^3 = -q \quad \text{und} \quad u^3 \cdot v^3 = -\frac{p^3}{27}\,;$$

wir kennen also die Summe und das Produkt ihrer dritten Potenzen. Damit kennen wir aber, wie wir bereits sowohl in §2 als auch §3 gesehen haben, auch u^3 und v^3 als Lösungen der quadratischen Gleichung $x^2 + qx - \frac{1}{27}p^3$. Somit ist

$$u^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}$$
$$\text{und} \qquad v^3 = -\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}\,,$$

wobei es auf die Reihenfolge natürlich nicht ankommt.

Damit kennen wir u^3 und v^3. Für u und v selbst gibt es dann jeweils drei Möglichkeiten, allerdings führen nicht alle neun Kombinationen dieser Möglichkeiten zu Lösungen, denn für eine Lösung muß ja die Bedingung $3uv = -p$ erfüllt sein, nicht nur $u^3 \cdot v^3 = -\frac{1}{27}p^3$.

Dies läßt sich am besten dadurch gewährleisten, daß wir für u *irgendeine* der drei Kubikwurzeln von u^3 nehmen und dann $v = -p/3u$ setzen. Die drei Lösungen der kubischen Gleichung $y^3 + py + q = 0$ sind also

$$y = u - \frac{p}{3u} \quad \text{mit} \quad u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}\,,$$

wobei für u nacheinander jede der drei Kubikwurzeln eingesetzt werden muß. (Es spielt keine Rolle, welche der beiden Quadratwurzeln wir nehmen, denn ersetzen wir die eine durch die andere, vertauschen wir dadurch einfach u und v.)

Da selbst von den drei Kubikwurzeln einer reellen Zahl nur eine reell ist, müssen wir zur Bestimmung aller drei Lösungen einer kubischen Gleichung mit reellen Koeffizienten *immer* auch mit komplexen Zahlen rechnen, selbst wenn alle Lösungen reell sind.

Wenn wir eine Kubikwurzel w_0 einer komplexen Zahl kennen, lassen sich die beiden anderen leicht bestimmen: Ist w eine von ihnen, so ist $(w/w_0)^3 = 1$, d.h. w/w_0 ist eine Nullstelle des Polynoms $X^3 - 1$. Dieses hat die Eins als Nullstelle; wenn wir durch $X - 1$ dividieren erhalten wir den Quotienten $X^2 + X + 1$, der nach der Lösungsformel für quadratische Gleichungen die beiden Nullstellen

$$\rho = -\frac{1}{2} + \frac{i}{2}\sqrt{3} \quad \text{und} \quad \bar{\rho} = -\frac{1}{2} - \frac{i}{2}\sqrt{3}$$

hat. Die beiden anderen Kubikwurzeln sind also $w_0\rho$ und $w_0\bar{\rho}$. Ihr Produkt $\rho\bar{\rho}$ ist $|\rho|^2 = 1$, d.h. die beiden sind invers zueinander, so daß die Division durch eine der beiden Wurzeln äquivalent ist zur Multiplikation mit der anderen. Ist in der Lösungsformel für die kubische Gleichung daher u irgendeine feste dritte Wurzel, so sind die drei Lösungen gleich

$$u - \frac{p}{3u}, \quad u\rho - \frac{p}{3u\rho} = u\rho - \frac{p}{3u}\bar{\rho} \quad \text{und} \quad u\bar{\rho} - \frac{p}{3u\bar{\rho}} = u\bar{\rho} - \frac{p}{3u}\rho\,.$$

Im sechzehnten Jahrhundert wurde das natürlich nicht so formuliert: Die mathematische Formelschreibweise führte schließlich erst VIÈTE einige Jahrzehnte später ein. TARTAGLIA, der die Lösungsmethode für die Gleichung $x^3 + px = q$ für positive Werte p, q fand, arbeite im übrigen auch nicht mit den Größen u und v, sondern mit deren dritten Potenzen. Er beschrieb seine Methode gegenüber CARDANO in einem Gedicht:

> Quando chel cubo con le cose appresso
> Se agguaglia à qualche numero discreto
> Trouan dui altri differenti in esso.
>
> Depoi terrai questo per consueto
> Che'llor produtto sempre sta eguale
> Al terzo cubo delle cose neto.
>
> El residuo poi suo generale
> Delli lor lati cubi ben sottratto
> Varra la tua cosa principale.

Die erste Lösung einer kubischen Gleichung geht wohl aus SCIPIONE DEL FERRO (1465–1526) zurück, der von 1496 bis zu seinem Tod an der Universität Bologna lehrte. 1515 fand er eine Methode, um die Nullstellen von $x^3 + px = q$ für *positive* Werte von p und q zu bestimmen (Negative Zahlen waren damals in Europa noch nicht im Gebrauch). Er veröffentlichte diese jedoch nie, so daß NICCOLO FONTANA (1499–1557, oberes Bild), genannt TARTAGLIA (der Stotterer), dieselbe Methode 1535 noch einmal entdeckte und gleichzeitig auch noch eine Modifikation, um einen leicht verschiedenen Typ kubischer Gleichungen zu lösen. TARTAGLIA war mathematischer Autodidakt, war aber schnell als Fachmann anerkannt und konnte seinen Lebensunterhalt als Mathematiklehrer in Verona und Venedig verdienen.

Die Lösung allgemeiner kubischer Gleichungen geht auf den Mathematiker, Arzt und Naturforscher GIROLAMO CARDANO (1501–1576, unteres Bild) zurück, dem TARTAGLIA nach langem Drängen und unter dem Siegel der Verschwiegenheit seine Methode mitgeteilt hatte. LODOVICO FERRARI (1522–1565) kam 14-jährig als Diener zu CARDANO; als dieser merkte, daß FERRARI schreiben konnte, machte er ihn zu seinem Sekretär. 1540 fand FERRARI die Lösungsmethode für biquadratische Gleichungen; 1545 veröffentlichte CARDANO trotz seines Schweigeversprechens gegenüber TARTAGLIA die Lösungsmethoden für kubische und biquadratische Gleichungen in seinem Buch *Ars magna.*

Frei übersetzt: Wenn der Kubus zusammen mit dem Produkt mit einer Sache eine gewisse Zahl ergibt, drücke diese aus als eine Differenz zweier anderen. Danach stelle sicher, daß das Produkt dieser beiden immer gleich dem Kubus eines Drittels der Sache ist. Die Lösung ist dann die Differenz der Kubikwurzeln der beiden.

In heutiger mathematischer Sprechweise: Zur Lösung der Gleichung $x^3 + px = q$ schreibe q als eine Differenz $q = U - V$. Stelle sicher, daß $UV = \left(\frac{p}{3}\right)^3$ ist. Dann ist $x = \sqrt[3]{U} - \sqrt[3]{V}$.

Betrachten wir als einfaches Beispiel die Gleichung

$$(x-1)(x-2)(x-3) = x^3 - 6x^2 + 11x - 6 = 0\,;$$

sie hat nach Konstruktion die drei Lösungen $1, 2$ und 3.

Falls wir das nicht wüßten, würden wir als erstes durch die Substitution $y = x - 2$ den quadratischen Term eliminieren. Einsetzen von $x = y + 2$ liefert

$$(y+2)^3 - 6(y+2)^2 + 11(y+2) - 6$$
$$= y^3 + 6y^2 + 12y + 8 - 6y^2 - 24y - 24 + 11y + 22 - 6 = y^3 - y\,,$$

wir müssen also zunächst die Gleichung $y^3 - y = 0$ lösen. Hierzu brauchen wir selbstverständlich keine Lösungstheorie kubischer Gleichungen: Ausklammern von y und die dritte binomische Formel zeigen sofort, daß

$$y^3 - y = y(y^2 - 1) = y(y+1)(y-1)$$

genau an den Stellen $y = -1, 0, 1$ verschwindet, und da $x = y + 2$ ist, hat die Ausgangsgleichung die Lösungen $x = 1, 2, 3$.

Wenden wir trotzdem unsere Lösungsformel an: Bei dieser Gleichung ist $p = -1$ und $q = 0$, also

$$u_1 = \sqrt[3]{\sqrt{\frac{-1}{27}}} = \sqrt[6]{\frac{-1}{27}} = \sqrt{\frac{-1}{3}}$$

für die rein imaginäre Kubikwurzel. Das zugehörige v_1 muß die Gleichung $u_1 v_1 = \frac{1}{3}$ erfüllen, also ist $v_1 = -u_1$, und wir erhalten als erste Lösung $y_1 = u_1 + v_1 = 0$.

Die beiden anderen Kubikwurzeln erhalten wir, indem wir die bekannte Kubikwurzel mit einer der beiden komplexen dritten Einheitswurzeln multiplizieren, d.h. also mit ρ und mit $\bar{\rho}$.

$$u_2 = \sqrt{\frac{-1}{3}}\,\rho = \frac{\sqrt{3}}{3} i \left(-\frac{1}{2} + \frac{\sqrt{3}\,i}{2} \right) = -\frac{1}{2} - \frac{\sqrt{3}}{6} i$$

und

$$v_2 = \frac{1}{3u_2} = \frac{-2}{3 + \sqrt{3}\,i} = \frac{-2(3 - \sqrt{3}\,i)}{3^2 + (\sqrt{3})^2} = -\frac{1}{2} + \frac{\sqrt{3}}{6} i\,;$$

wir erhalten somit die Lösung $y_2 = u_2 + v_2 = -1$.

Die dritte Kubikwurzel

$$u_3 = \sqrt{\frac{-1}{3}}\,\bar{\rho} = \frac{\sqrt{3}}{3} i \left(-\frac{1}{2} - \frac{\sqrt{3}\,i}{2} \right) = \frac{1}{2} - \frac{\sqrt{3}}{6} i$$

schließlich führt auf

$$v_3 = \frac{1}{3u_3} = \frac{2}{3-\sqrt{3}\,i} = \frac{2(3+\sqrt{3}\,i)}{3^2+(\sqrt{3})^2} = \frac{1}{2} + \frac{\sqrt{3}}{6}i$$

und liefert so die Lösung $y_3 = u_3 + v_3 = 1$.

Etwas komplizierter wird es bei der aus §3 bekannten Gleichung

$$x^3 - 7x + 6 = 0\,,$$

deren Lösungen wir dort über den Satz von VIÈTE so leicht erraten konnten. Da sie keinen x^2-Term hat, können wir gleich $p = -7$ und $q = 6$ in die Formel einsetzen und erhalten

$$u = \sqrt[3]{-3+\sqrt{\frac{6^2}{4}-\frac{7^3}{27}}} = \sqrt[3]{-3+\sqrt{-\frac{400}{4\cdot 27}}} = \sqrt[3]{-3+\frac{10}{9}\sqrt{3}\,i}\,.$$

Was nun? Wenn wir einen Ansatz der Form $u = r + is$ machen, kommen wir auf ein System von zwei kubischen Gleichungen in zwei Unbekannten, also ein schwierigeres Problem als unsere Ausgangsgleichung.

Eine Alternative ist die Polarkoordinatendarstellung komplexer Zahlen: Eine komplexe Zahl $z = x + iy$ läßt sich bekanntlich auch darstellen als $z = re^{i\varphi}$ mit $r = |z| = \sqrt{x^2+y^2}$ und $x = r\cos\varphi$, $y = r\sin\varphi$. Da $e^{i\varphi}\cdot e^{i\psi} = e^{i(\varphi+\psi)}$ ist, werden bei der Multiplikation zweier komplexer Zahlen in Polarkoordinatendarstellung die Beträge miteinander multipliziert und die Winkel addiert. Daher ist $\sqrt[3]{|z|}\big(\cos\frac{\varphi}{3} + i\sin\frac{\varphi}{3}\big)$ eine dritte Wurzel von z. Leider gibt es aber keine einfache Formel, die Sinus und Kosinus von $\frac{\varphi}{3}$ durch $\cos\varphi$ und $\sin\varphi$ ausdrückt. Aus den Additionstheoremen können wir uns natürlich leicht Formeln für $\cos 3\varphi$ verschaffen; wir erhalten

$$\cos 3\varphi = 4\cos^3\varphi - 3\cos\varphi\,.$$

Um $x = \cos\frac{\varphi}{3}$ zu berechnen, müssen wir also die kubische Gleichung $4x^3 - 3x = \cos\varphi$ lösen, was uns wiederum auf die Berechnung einer Kubikwurzel führt, *usw.*

Trotzdem ist die obige Darstellung der Lösung nicht völlig nutzlos: Sie gibt uns immerhin Formeln für den Real- und den Imaginärteil der Lösung, und diese Formeln können wir numerisch auswerten.

Für den hier interessierenden Radikanden $z = -3 + \frac{10}{9}\sqrt{3}\,i$ ist

$$|z| = \sqrt{(-3)^2 + \left(\frac{10}{9}\sqrt{3}\right)^2} = \sqrt{\frac{9+300}{81}} = \sqrt{\frac{9\cdot 27+100}{27}}$$
$$= \sqrt{\frac{343}{27}} = \sqrt{\frac{7^3}{3^3}} = \frac{7}{9}\sqrt{21}\,.$$

Somit ist

$$\cos\varphi = \frac{x}{|z|} = -\frac{9}{49}\sqrt{21} \approx -0{,}84169975767$$

und

$$\sin\varphi = \frac{y}{|z|} = \frac{10}{49}\sqrt{7} \approx 0{,}5399492473\,.$$

Der Arkuskosinus des ersten Werts ist ungefähr 2,571215844, der Arkussinus des zweiten 0,5703768102. Wenn wir φ im Intervall $(-\pi,\ \pi]$ suchen, folgt aus der Negativität von $\cos\varphi$, daß $|\varphi| > \frac{\pi}{2}$ sein muß; daher ist φ ungefähr gleich dem ersten der beiden Werte. (Der zweite ist natürlich $\pi - \varphi$, was den gleichen Sinus hat.)

Damit können wir eine dritte Wurzel näherungsweise bestimmen; wir erhalten

$$u_1 = \sqrt[3]{|z|}\left(\cos\frac{\varphi}{3} + i\sin\frac{\varphi}{3}\right) \approx 0{,}9999999994 + 1{,}154700538i$$

(Je nach Taschenrechner oder Programm kann das Ergebnis auch leicht verschieden sein.) und damit als erste Lösung

$$x_1 = u_1 + \frac{7}{3u_1} \approx 1{,}9999999999 - 10^{-9}i\,.$$

Wie jedes numerische Ergebnis stimmt diese Zahl natürlich nur näherungsweise und hängt im übrigen auch sowohl von der Stellenzahl als auch der Rundung ab. Zumindest in diesem Fall ist die Hypothese, daß es sich hier um eine durch Rundungsfehler verfälschte Zwei handeln könnte, eine Überlegung wert. Einsetzen zeigt, daß die Zwei tatsächlich eine Lösung ist. Für die beiden anderen dividieren am besten das Polynom durch $X-2$ und lösen dann die Quotientengleichung $x^2+2x-3 = 0$.

Obwohl die drei Lösungen $1, 2$ und -3 unserer Gleichung allesamt ganzzahlig sind, konnten wir dies also durch bloßes Einsetzen in unsere Formel nicht erkennen und konnten insbesondere die Kubikwurzel nur durch Erraten und Nachprüfen in einer einfachen Form darstellen.

Bei der ebenfalls in §3 betrachteten Gleichung $x^3+6x+6 = 0$, bei der uns VIÈTE nicht weiterhelfen konnte, ist die Anwendung der Lösungsformel übrigens deutlich einfacher: Einsetzen der Parameter $p = -6$ und $q = 6$ in die Lösungsformel führt zunächst auf

$$u_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{-3 + \sqrt{9-8}} = \sqrt[3]{-2} = -\sqrt[3]{2}$$

für die reelle Wurzel; die erste Lösung ist also

$$x_1 = u_1 - \frac{p}{3u_1} = -\sqrt[3]{2} - \frac{2}{\sqrt[3]{2}} = -\sqrt[3]{2} - \sqrt[3]{4}\,.$$

Für die zweite und dritte Lösung müssen wir mit $u_2 = u_1\rho$ *bzw.* $u_3 = u_1\bar\rho$ anstelle von u_1 arbeiten und erhalten

$$x_2 = -\sqrt[3]{2}\,\rho - \frac{2}{\sqrt[3]{2}\,\rho} = -\sqrt[3]{2}\,\rho - \sqrt[3]{4}\,\bar\rho \qquad \text{und}$$

$$x_3 = -\sqrt[3]{2}\,\bar\rho - \frac{2}{\sqrt[3]{2}\,\bar\rho} = -\sqrt[3]{2}\,\bar\rho - \sqrt[3]{4}\,\rho\,,$$

was nach Einsetzen von $\rho = -\frac{1}{2} + \frac{1}{2}\sqrt{3}i$ und $\bar\rho = -\frac{1}{2} - \frac{1}{2}\sqrt{3}i$ auf die beiden komplexen Lösungen

$$\begin{aligned} x_{2/3} &= -\sqrt[3]{2}\left(-\frac{1}{2} \pm \frac{\sqrt{3}}{2}i\right) - \sqrt[3]{4}\left(-\frac{1}{2} \mp \frac{\sqrt{3}}{2}i\right) \\ &= \frac{\sqrt[3]{2} + 3\sqrt[3]{4}}{2} \pm \frac{\sqrt{3}\left(\sqrt[3]{2} - \sqrt[3]{4}\right)}{2}i \end{aligned}$$

führt. Natürlich erfüllen auch diese Zahlen den Satz von VIÈTE, jedoch hilft uns dieser nicht, sie zu erraten.

Wenn wir eine reelle Kubikwurzel finden können, ist die Situation auch nicht unbedingt viel besser. Betrachten wir etwa die Gleichung

$$x^3 - 3x^2 + 9x + 13 = 0\,.$$

Hier setzen wir $x = y + 1$ und erhalten die neue Gleichung

$$\begin{aligned}&(y+1)^3 - 3(y+1)^2 + 9(y+1) + 13\\ &= y^3 + 3y^2 + 3y + 1 - 3(y^2 + 2y + 1) + 9y + 9 + 13\\ &= y^3 + 6y + 20 = 0\end{aligned}$$

mit $p = 6$ und $q = 20$. Damit ist $\frac{p}{3} = 2$ und $\frac{q}{2} = 10$, also

$$u = \sqrt[3]{-10 + \sqrt{100 + 8}} = \sqrt[3]{-10 + \sqrt{108}} = \sqrt[3]{-10 + 6\sqrt{3}}$$

Da 108 größer ist als $(-10)^2 = 100$, gibt es eine positive reelle Wurzel u_1; wir rechnen zunächst mit dieser und erhalten als erste Lösung

$$y_1 = u_1 - \frac{p}{3u_1} = \sqrt[3]{-10 + 6\sqrt{3}} - \frac{2}{\sqrt[3]{-10 + 6\sqrt{3}}}\,.$$

Damit haben wir im Prinzip eine Lösung gefunden. Wenn wir sie allerdings numerisch auswerten, erhalten wir etwas wie -1,99999998, und damit drängt sich natürlich der Verdacht auf, daß dies gleich -2 sein könnte. Einsetzen von $y = -2$ in unsere kubische Gleichung zeigt in der Tat, daß

$$(-2)^3 + 6 \cdot (-2) + 20 = -8 - 12 + 20 = 0$$

ist. Aber warum ist

$$\sqrt[3]{-10 + 6\sqrt{3}} - \frac{2}{\sqrt[3]{-10 + 6\sqrt{3}}} = -2\,,$$

und wie, vor allem, kann man das der linken Seite ansehen?

Wie die Erfahrung der Computeralgebra zeigt, kann es extrem schwierig sein, auch nur zu entscheiden, ob zwei Wurzelausdrücke gleich sind; direkte allgemeine Verfahren dazu gibt es nicht. Unsere Formel gibt uns daher zwar immer drei Wurzelausdrücke, die Lösungen der gegebenen Gleichung sind, aber diese können für Zahlen stehen, die sich auch sehr viel einfacher ausdrücken lassen.

Im vorliegenden Fall, wo die numerische Berechnung eine Vermutung nahelegt, können wir wieder versuchen, diese zu beweisen: Aus der vermuteten Gleichung

$$u_1 - \frac{2}{u_1} = -2 \quad \text{folgt} \quad u_1^2 - 2 = -2u_1\,.$$

Quadratische Ergänzung macht daraus $(u_1+1)^2 = 3$, also ist u_1 eine der beiden Zahlen $-1 \pm \sqrt{3}$. Die dritte Potenz davon ist

$$(-1 \pm \sqrt{3})^3 = -1 \pm 3\sqrt{3} - 3 \cdot 3 \pm 3\sqrt{3} = -10 \pm 6\sqrt{3}\,,$$

also ist tatsächlich $u_1 = -1+\sqrt{3}$ und

$$\begin{aligned} y_1 &= -1+\sqrt{3} - \frac{2}{-1+\sqrt{3}} = -1+\sqrt{3} - \frac{2(-1-\sqrt{3})}{(-1+\sqrt{3})(-1-\sqrt{3})} \\ &= -1+\sqrt{3} + \frac{2+2\sqrt{3}}{-2} = -2\,. \end{aligned}$$

Nachdem wir u_1 in einfacher Form ausgedrückt haben, lassen sich auch die anderen beiden Lösungen berechnen:

$$u_2 = u_1\rho = (-1+\sqrt{3}) \cdot \frac{-1+\sqrt{3}\,i}{2} = \frac{(1-\sqrt{3})+(3-\sqrt{3})i}{2}$$

und

$$u_3 = u_1\bar{\rho} = (-1+\sqrt{3}) \cdot \frac{-1-\sqrt{3}\,i}{2} = \frac{(1-\sqrt{3})-(3-\sqrt{3})i}{2}$$

Damit ist

$$\begin{aligned} \frac{2}{u_2} &= \frac{4\big((1-\sqrt{3})-(3-\sqrt{3})i\big)}{(1-\sqrt{3})^2+(3-\sqrt{3})^2} = \frac{4\big((1-\sqrt{3})-(3-\sqrt{3})i\big)}{16-8\sqrt{3}} \\ &= \frac{\big((1-\sqrt{3})-(3-\sqrt{3})i\big)(2+\sqrt{3})}{2(2-\sqrt{3})(2+\sqrt{3})} = \frac{-(1+\sqrt{3})-(3+\sqrt{3})i}{2}\,, \end{aligned}$$

also

$$y_2 = u_2 - \frac{2}{u_2} = \frac{(1-\sqrt{3})+(3-\sqrt{3})i}{2} + \frac{(1+\sqrt{3})+(3+\sqrt{3})i}{2} = 1+3i\,.$$

Entsprechend folgt $y_3 = u_3 - \dfrac{2}{u_3} = 1-3i$.

Die Mathematiker des sechzehnten Jahrhunderts, auf die die Lösungsformel für kubische Gleichungen zurückgeht, hatten natürlich weder Computer noch Taschenrechner noch komplexe Zahlen; auch Dezimalbrüche in ihrer heutigen Form kamen erst im siebzehnten Jahrhundert auf, als die ersten Tafeln trigonometrischer Funktionen und kurz später

auch Logarithmen veröffentlicht wurden. Doch auch ohne diese Hilfsmittel konnten sie erstaunlich gut mit der Lösungsformel umgehen. In §3.2 des Buchs

TEO MORA: Solving Polynomial Equation Systems I: The Kronecker-Duval Philosophy, *Cambridge University Press,* 2003

sind zwei Beispiele für ihre Vorgehensweise zu finden:

Bei der Gleichung $x^3 + 3x - 14 = 0$ ist $p = 3$ und $q = -14$, also

$$u = \sqrt[3]{7 + \sqrt{7^2 + 1^3}} = \sqrt[3]{7 + 5\sqrt{2}}\,.$$

Beim vorigen Beispiel hatten wir gesehen, daß

$$\sqrt[3]{-10 + 6\sqrt{3}} = -1 + \sqrt{3}$$

ist; eine Zahl der Form $a + b\sqrt{3}$ hat in diesem speziellen Fall also eine Kubikwurzel derselben Form $c+d\sqrt{3}$. Das gilt natürlich nicht allgemein; die Kubikwurzel aus $\sqrt{3}$, die sechste Wurzel von drei also, läßt sich sicher nicht in der Form $c + d\sqrt{3}$ mit ganzen Zahlen c und d schreiben. Trotzdem können wir unser Glück versuchen.

Für den Radikanden $7 + 5\sqrt{2}$ machen wir natürlich einen Ansatz der Form $c + d\sqrt{2}$. Wir wollen, daß

$$(c + d\sqrt{2})^3 = c^3 + 3c^2d\sqrt{2} + 6cd^2 + 2d^3\sqrt{2} = 7 + 5\sqrt{2}$$

ist mit $c, d \in \mathbb{Z}$, also

$$c^3 + 6cd^2 = 7 \quad \text{und} \quad 3c^2d + 2d^3 = 5\,.$$

Damit haben wir, wie schon oben erwähnt, ein System von *zwei* kubischen Gleichungen anstelle von einer, jetzt allerdings suchen wir nur nach ganzzahligen Lösungen. Aus der ersten Gleichung können wir c ausklammern und erhalten $c(c^2 + 6d^2) = 7$. Somit muß c ein Teiler von sieben sein, d.h. $c = \pm 1$ oder $c = \pm 7$. Die negativen Zahlen scheiden aus, da die Klammer nicht negativ werden kann, und auch $c = 7$ ist nicht möglich, denn dann wäre die linke Seite mindestens gleich 7^3. Wenn es eine ganzzahlige Lösung gibt, muß daher $c = 1$ sein; durch Einsetzen folgt, daß dann mit $d = \pm 1$ die erste Gleichung in der Tat erfüllt ist. Die

zweite Gleichung $d(3c^2 + 2d^2) = 5$ zeigt, daß auch d positiv sein muß und $c = d = 1$ beide Gleichungen erfüllt. Somit ist

$$u = \sqrt[3]{7 + 5\sqrt{2}} = 1 + \sqrt{2}$$

für die reelle unter den drei Kubikwurzeln. Da wir eine Gleichung mit reellen Koeffizienten haben, muß auch das zugehörige v reell sein und kann genauso wie u bestimmt werden:

$$v = \sqrt[3]{7 - 5\sqrt{2}} = 1 - \sqrt{2} \quad \text{und} \quad x = u + v = 2\,.$$

Damit war die Gleichung für die Zwecke des sechzehnten Jahrhunderts gelöst, denn da es noch keine komplexen Zahlen gab, suchte auch niemand nach komplexen Lösungen.

Wir interessieren uns allerdings für komplexe Lösungen; die beiden noch fehlenden Lösungen können wir entweder berechnen als $u\rho + v\bar{\rho}$ und $u\bar{\rho} + v\rho$, oder aber wir dividieren das Polynom $X^3 + 3X - 14$ durch $X - 2$ und erhalten das quadratische Polynom $X^2 + 2X + 7$ mit den Nullstellen $-1 \pm \sqrt{6}\,i$.

Bei Gleichungen mit drei reellen Nullstellen führt die Lösungsformel, wie wir in §9 sehen werden, *immer* übers Komplexe, aber auch damit wurden CARDANO und seine Zeitgenossen fertig. MORA betrachtet als Beispiel dafür die Gleichung $x^3 - 21x - 20 = 0$. Hier ist

$$u = \sqrt[3]{10 + \sqrt{10^2 - 7^3}} = \sqrt[3]{10 + \sqrt{-243}} = \sqrt[3]{10 + 9\sqrt{-3}}\,.$$

$\sqrt{-3}$ war für CARDANO im Gegensatz zu $\sqrt{2}$ keine Zahl; trotzdem rechnete er damit als mit einem abstrakten Symbol gemäß der Regel $\sqrt{-3} \cdot \sqrt{-3} = -3$.

Wenn wir wieder auf unser Glück vertrauen und einen Ansatz der Form $u = c + d\sqrt{-3}$ machen, kommen wir auf das Gleichungssystem

$$c^3 - 9cd^2 = 10 \quad \text{und} \quad 3c^2d - 3d^3 = 9\,.$$

Ausklammern von c *bzw.* d und Kürzen der zweiten Gleichung durch drei führt auf

$$c(c^2 - 9d^2) = 10 \quad \text{und} \quad d(c^2 - d^2) = 3\,.$$

Wenn es ganzzahlige Lösungen gibt, muß wegen der zweiten Gleichung $d = \pm 1$ oder $d = \pm 3$ sein. $d = \pm 1$ führt auf $c^2 - 1 = \pm 3$, also $d = 1$ und $c = \pm 2$; für $d = \pm 3$ läßt sich kein ganzzahliges c finden. Einsetzen in die erste Gleichung zeigt, daß $c = -2, d = 1$ das System löst, also ist $u_1 = -2 + \sqrt{-3}$ eine der drei Wurzeln. Die erste Lösung der kubischen Gleichung ist also

$$\begin{aligned} x_1 &= -2 + \sqrt{-3} + \frac{7}{-2 + \sqrt{-3}} \\ &= -2 + \sqrt{-3} + \frac{7(-2 + \sqrt{-3})}{(-2 + \sqrt{-3})(-2 - \sqrt{-3})} \\ &= -2 + \sqrt{-3} + \frac{-14 + 7\sqrt{-3}}{7} = -4\,. \end{aligned}$$

Zur Bestimmung der beiden anderen Lösungen haben nun viele Möglichkeiten: Wir könnten das Polynom $X^3 - 21X - 20$ durch $X - x_1$, also $X + 4$, dividieren und damit das Problem auf die Lösung einer quadratischen Gleichung reduzieren, oder wir könnten die beiden weiteren Werte für u als $u_2 = u_1\rho$ und $u_3 = u_1\bar{\rho}$ berechnen, oder wir könnten den Satz von VIÈTE anwenden, der uns hier problemlos alle drei Lösungen gibt. Zur Zeit CARDANOs gab es keine dieser Möglichkeiten: Mit Polynomen rechnete man erst im achtzehnten Jahrhundert, und die komplexen Zahlen wurden sogar erst im neunzehnten Jahrhundert eingeführt. CARDANO rechnete zwar mit „Symbolen“ wie $\sqrt{-3}$, aber die einzige Lösung der Gleichung $x^3 = 1$ war für ihn – wie für alle seiner Zeitgenossen – die Eins. FRANÇOIS VIÈTE schließlich war 1545, als CARDANOs *Ars magna* erschien, gerade fünf Jahre alt.

Wie die obige Rechnung zeigte, sind $c = -2$ und $d = 1$ die einzigen ganzzahligen Lösungen der Gleichung $(c + d\sqrt{-3})^3 = 10 + 9\sqrt{-3}$. Vielleicht gibt es aber weitere Lösungen, wenn wir für c und d auch rationale Zahlen zulassen. Nun haben wir allerdings bei der Suche nach ganzzahligen Lösungen viel mit Teilbarkeit argumentiert, und im Körper der rationalen Zahlen ist jedes Element durch jedes andere außer der Null teilbar. Wir müssen uns daher auf Zahlen mit einem festen Nenner beschränken; dann kommen wir wieder zu Beziehungen zwischen ganzen Zahlen und können versuchen, wie oben vorzugehen.

Der kleinstmögliche Nenner ist zwei; versuchen wir also unser Glück mit dem Ansatz

$$\left(\frac{c}{2}+\frac{d}{2}\sqrt{-3}\right)^3 = 10+9\sqrt{-3}\,,$$

wobei c und d wieder ganze Zahlen sein sollen. Ausmultiplizieren, Multiplikation mit acht und Ausklammern führt auf die Gleichungen

$$c(c^2-9d^2) = 80 \quad \text{und} \quad d(c^2-d^2) = 24\,,$$

wobei mindestens eine der Zahlen c und d ungerade sein muß, da wir ansonsten wieder eine Wurzel mit ganzzahligem Real- und Imaginärteil bekommen, also die bereits bekannte. Da rechts jeweils gerade Zahlen stehen, sieht man leicht, daß dann beide Zahlen ungerade sein müssen; damit bleiben also für c als Teiler von achtzig nur die Möglichkeiten ±1 und ±5. Für d als Teiler von 24 sind $d=\pm1$ und $d=\pm3$ möglich. Einsetzen zeigt, daß $c=-1, d=-3$ und $c=5, d=1$ die einzigen Lösungen sind. Die beiden verbleibenden Kubikwurzeln von $-10+9\sqrt{-3}$ sind somit

$$u_2 = \frac{5}{2}+\frac{1}{2}\sqrt{-3} \quad \text{und} \quad u_3 = -\frac{1}{2}-\frac{3}{2}\sqrt{-3}\,.$$

Damit lassen sich nun leicht

$$x_2 = u_2+\frac{7}{u_2} = -1 \quad \text{und} \quad x_3 = u_3+\frac{7}{u_3} = 5$$

berechnen. Die Gleichung $x^3-21x-20=0$ hat also die drei ganzzahligen Lösungen $-1, -4$ und 5.

Wie die Beispiele in diesem Paragraphen zeigen, haben wir es beim exakten Lösen kubischer Gleichungen mit der hier betrachteten Formel oft mit komplizierten Ausdrücken zu tun, von denen sich nachher (nach teilweise recht trickreichen Ansätzen) herausstellt, daß sie sich tatsächlich sehr viel einfacher darstellen lassen. Dies ist ein allgemeines Problem der Computeralgebra, zu dem es leider keine allgemeine Lösung gibt: Wie D. RICHARDSON 1968 gezeigt hat, kann es keinen Algorithmus geben, der von zwei beliebigen reellen Ausdrücken entscheidet, ob sie gleich sind oder nicht. Dabei reicht es schon, wenn wir nur Ausdrücke betrachten, die aus ganzen Zahlen, den Grundrechenarten, der Sinus- und der Betragsfunktion sowie der Zahl π aufgebaut

werden können. Jedenfalls sollte klar geworden sein, daß das Lösung von kubischen Gleichung deutlich aufwendiger ist als das von quadratischen und daß die Lösungsformel hier keine problemlos anwendbare Mitternachtsformel ist. Kein Wunder, daß CARDANO seinem Buch, in dem er die Lösung kubischer und biquadratischer Gleichungen behandelte, den Titel *Ars magna,* die „große Kunst“ gab.

§5: Biquadratische Gleichungen

1840 fand CARDANOs ehemaliger Diener und späterer Sekretär LODOVICO FERRARI eine Lösungsmethode für Gleichungen vom Grad vier. Hier wird zunächst der kubische Term von

$$x^4 + ax^3 + bx^2 + cx + d = 0$$

eliminiert durch die Substitution $x = y - \frac{a}{4}$; dies führt auf eine Gleichung der Form

$$y^4 + py^2 + qy + r = 0\,.$$

Zu deren Lösung benutzen wir (nach FERRARI) einen anderen Trick als im kubischen Fall: Wir versuchen, die Gleichung so zu modifizieren, daß wir ihre Lösungen als die Lösungen zweier quadratischer Gleichungen berechnen können.

Dazu nehmen wir an, wir hätten eine Lösung y der Gleichung und betrachten dazu für eine zunächst noch beliebige Zahl u die Zahl $y^2 + u$. Da $y^4 + py^2 + qy + r$ verschwindet, ist $y^4 = -py^2 - qy - r$, also

$$(y^2 + u)^2 = y^4 + 2uy^2 + u^2 = (2u - p)y^2 - qy + u^2 - r\,.$$

Falls rechts das Quadrat eines linearen Polynoms $sy + t$ steht, ist

$$(y^2 + u)^2 = (sy + t)^2 \Longrightarrow y^2 + u = \pm(sy + t)\,,$$

wir müssen also nur die beiden quadratischen Gleichungen

$$y^2 \mp (sy + t) + u = 0$$

lösen, um die Lösungen der biquadratischen Gleichung zu finden.

Natürlich läßt sich die rechte Seite $(2u-p)y^2 - qy + u^2 - r$ im allgemeinen nicht als ein Quadrat $(sy+t)^2$ schreiben; wir können aber hoffen, daß sie zumindest für gewisse spezielle Werte der bislang noch willkürlichen Konstanten u eines ist.

Ein quadratisches Polynom $\alpha Y^2 + \beta Y + \gamma$ ist genau dann Quadrat eines linearen, wenn die beiden Nullstellen der quadratischen Gleichung $\alpha y^2 + \beta y + \gamma = 0$ übereinstimmen. Diese Nullstellen können wir nach der Formel aus §2 berechnen:

$$y^2+\frac{\beta}{\alpha}y+\frac{\gamma}{\alpha} = 0 \Longrightarrow y = -\frac{\beta}{2\alpha} \pm \sqrt{\frac{\beta^2}{4\alpha^2} - \frac{\gamma}{\alpha}} = -\frac{\beta}{2\alpha} \pm \frac{1}{2\alpha}\sqrt{\beta^2 - 4\alpha\gamma}\,.$$

Die beiden Lösungen fallen somit genau dann zusammen, wenn $\beta^2 - 4\alpha\gamma$ verschwindet. In unserem Fall ist $\alpha = (2u - p)$, $\beta = -q$ und $\gamma = u^2 - r$; wir erhalten also die Bedingung

$$q^2 - 4(2u - p)(y^2 - r) = -8u^3 + 4pu^2 + 8ru + q^2 - 4pr = 0\,.$$

Dies ist eine kubische Gleichung für u, die wir mit der Methode aus dem vorigen Abschnitt (vielleicht) lösen können. Ist u_0 eine der Lösungen, so steht in der Gleichung

$$(y^2 + u_0)^2 = (2u_0 - p)y^2 - qy + u_0^2 - r$$

rechts das Quadrat eines linearen Polynoms $sy + t$, das wir – da wir alle Koeffizienten kennen – problemlos hinschreiben können. Dies führt dann nach Wurzelziehen zu zwei quadratischen Gleichungen für y, deren Wurzeln die Nullstellen von $y^4 + py^2 + qy + r = 0$ sind.

Es wäre nicht schwer, mit Hilfe der Lösungsformel für kubische Gleichungen, eine explizite Formel für die vier Lösungen hinzuschreiben; sie ist allerdings erstens deutlich länger und zweitens für die praktische Berechnung reeller Nullstellen mindestens genauso problematisch wie die für kubische Gleichungen. Auf Beispiele zur Lösung biquadratischer Gleichungen verzichte ist, denn schon in einfachen Fällen ist die kubische Gleichung für u selbst dann sehr kompliziert, wenn es eine reelle Lösung gibt, die in der Lösungsformel in rein reeller Form auftaucht.

Für numerische Berechnungen sind übrigen sowohl die Lösungsmethode für kubische Gleichungen als auch die für biquadratische eher nicht geeignet, da es beim Einsetzen in die Formeln oft vorkommen kann, daß zwei ungefähr gleich große Zahlen voneinander subtrahiert werden, so daß die Anzahl der geltenden Ziffern dramatisch kleiner wird. Die klassischen numerischen Verfahren zur Nullstellenbestimmung liefern genauere Ergebnisse und sind einfacher anzuwenden.

§6: Gleichungen höheren Grades

Nach der (mehr oder weniger) erfolgreichen Auflösung der kubischen und biquadratischen Gleichungen in der ersten Hälfte des sechzehnten Jahrhunderts beschäftigten sich natürlich viele Mathematiker mit dem nächsten Fall, der Gleichung fünften Grades. Hier gab es jedoch über 250 Jahre lang keinerlei Fortschritt, bis zu Beginn des neunzehnten Jahrhunderts ABEL glaubte, eine Lösung gefunden zu haben. Er entdeckte dann aber recht schnell seinen Fehler und bewies stattdessen 1824, daß es keine allgemeinen Lösungsformel für Gleichungen fünften (oder höheren) Grades geben kann, die nur mit Grundrechenarten und Wurzeln auskommt.

Die Grundidee seines Beweises liegt in der Betrachtung von Symmetrien innerhalb der Lösungsmenge: Man betrachtet die Menge aller Permutationen der Nullstellenmenge, die durch Abbildungen $\varphi\colon \mathbb{C} \to \mathbb{C}$ erreicht werden können, wobei φ sowohl mit der Addition als auch der Multiplikation verträglich sein muß. ABEL zeigt, daß diese Permutationen für allgemeine Gleichungen vom Grad größer vier eine (in heutiger Terminologie) *nichtauflösbare* Gruppe bilden und daß es aus diesem Grund keine Lösungsformel geben kann, in der nur Grundrechenarten und Wurzeln vorkommen. Ein großer Teil dieser Vorlesung wird sich damit beschäftigen, dies genauer zu verstehen.

Der norwegische Mathematiker NILS HENRIK ABEL (1802–1829) ist trotz seinen frühen Todes (an Tuberkulose) Initiator vieler Entwicklungen der Mathematik des neunzehnten Jahrhunderts; Begriffe wie abelsche Gruppen, abelsche Integrale, abelsche Funktionen, abelsche Varietäten, die auch in der heutigen Mathematik noch allgegenwärtig sind, verdeutlichen seinen Einfluß. Zu seinem 200. Geburtstag stiftete die norwegische Regierung einen ABEL-Preises für Mathematik mit gleicher Ausstattung und Vergabebedingungen wie die Nobelpreise; erster Preisträger war 2003 JEAN-PIERRE SERRE (*1926) vom Collège de France für seine Arbeiten über algebraische Geometrie, Topologie und Zahlentheorie.

Der ABELsche Satz besagt selbstverständlich nicht, daß Gleichungen höheren als vierten Grades *unlösbar* seien; er sagt nur, daß es *im allgemeinen* nicht möglich ist, die Lösungen durch Wurzelausdrücke in

den Koeffizienten darzustellen: Für eine allgemeine Lösungsformel muß man also außer Wurzeln und Grundrechenarten noch weitere Funktionen zulassen. Beispielsweise fanden sowohl HERMITE als auch KRONECKER 1858 Lösungsformeln für Gleichungen fünften Grades mit sogenannten elliptischen Modulfunktionen; 1870 löste JORDAN damit Gleichungen beliebigen Grades.

§7: Symmetrische Polynome

In §3 haben wir gesehen, daß sich die Koeffizienten eines Polynoms als Polynome in den Nullstellen darstellen lassen. Die Darstellung der Nullstellen als Polynome in den Koeffizienten ist nicht möglich; schon bei quadratischen Polynomen brauchen wir auch Wurzeln, und ab Grad fünf geht es nach dem Satz von ABEL nicht einmal mit diesen. Wir können uns aber fragen, ob es nicht zumindest für manche Polynome in den Nullstellen möglich ist, sie auch als Polynome in den Koeffizienten auszudrücken. Im nächsten Paragraphen wollen wir das Ergebnis dann anwenden auf die Frage, wann ein Polynom mehrfache Nullstellen hat. Über die Nullstellen läßt sich das natürlich einfach entscheiden, und wir werden sehen, daß wir das Kriterium mit den Methoden aus diesem Paragraphen so umschreiben können, daß wir ein Polynom in den Koeffizienten erhalten, das genau dann verschwindet, wenn es eine mehrfache Nullstelle gibt.

Wir kennen bereits Polynome in den Nullstellen, die sich auch als Polynome in den Koeffizienten schreiben lassen, nämlich die elementarsymmetrischen Polynome, die ja bei führendem Koeffizienten eins bis eventuell aufs Vorzeichen gerade die Koeffizienten sind. Sie sind Polynome, die sich nicht verändern, wenn ihre Variablen in irgendeiner Weise permutiert werden. Solche Polynome bezeichnen wir als symmetrisch:

Definition: *a)* Die *symmetrische Gruppe* $\mathfrak{S}_n$ ist die Menge aller bijektiver Abbildungen der Menge $\{1, \ldots, n\}$ auf sich selbst; ihre Elemente heißen *Permutationen.*
b) Ein Polynom f in den n Variablen $X_1, \ldots, X_n$ heißt *symmetrisch,* wenn für jede Permutation $\pi \in \mathfrak{S}_n$ gilt:

$$f(X_{\pi(1)}, \ldots, X_{\pi(n)}) = f(X_1, \ldots, X_n) .$$

Der folgende Satz besagt, daß sich jedes symmetrische Polynom durch die elementarsymmetrischen ausdrücken läßt:

Satz: f sei ein symmetrisches Polynom in $X_1, \ldots, X_n$. Dann gibt es ein Polynom g in n neuen Variablen $Y_1, \ldots, Y_n$, so daß gilt

$$f(X_1, \ldots, X_n) = g\big(\sigma_1(X_1, \ldots, X_n), \ldots, \sigma_n(X_1, \ldots, X_n)\big) .$$

Beweis: Zur Konstruktion von g ordnen wir die Monome $X_1^{e_1} \cdots X_n^{e_n}$ von f graduiert lexikographisch an, d.h. das Monom $X_1^{d_1} \cdots X_n^{d_n}$ heißt größer als $X_1^{e_1} \cdots X_n^{e_n}$, wenn sein Grad $d_1 + \cdots + d_n$ größer ist als $e_1 + \cdots + e_n$, oder wenn die Grade gleich sind und die erste von Null verschiedene Differenz $d_i - e_i$ positiv ist. Für ein symmetrisches Polynom, muß im ersten, dem sogenannten *führenden* Monom, $e_1 \geq e_2 \geq \cdots \geq e_n$ sein, denn wegen der Symmetrie muß mit $X_1^{e_1} \cdots X_n^{e_n}$ auch jedes Monom $X_1^{e_{\pi(1)}} \cdots X_n^{e_{\pi(n)}}$ vorkommen.

Um die Formeln übersichtlich zu halten, schreiben wir im folgenden kurz σ_i an Stelle von $\sigma_i(X_1, \ldots, X_n)$. Die Beweisstrategie besteht darin, nacheinander alle Terme von f zu eliminieren durch Subtraktion eines Produkts elementarsymmetrischer Polynome, wobei in jedem Schritt das graduiert lexikographisch größte unter den noch verbliebenen Monomen eliminiert wird.

Zur Konstruktion eines Produkt elementarsymmetrischer Polynome mit führendem Monom $X_1^{e_1} \cdots X_n^{e_n}$ beachten wir, daß das führende Monom von σ_i gleich $X_1 \cdots X_i$ ist. Wir betrachten nur Monome, deren Exponenten die Ungleichungen $e_1 \geq \cdots \geq e_n$ erfüllen. Dann sind $\delta_i = e_i - e_{i+1}$ für $i = 1, \ldots, n-1$ und $\delta_n = e_n$ größer oder gleich Null, und $\sigma_1^{\delta_1} \cdots \sigma_n^{\delta_n}$ hat das führende Monom $X_1^{e_1} \cdots X_n^{e_n}$, da e_i gleich der Summe $\delta_i + \cdots + \delta_n$ ist.

Ist daher $aX_1^{e_1} \cdots X_n^{e_n}$ der führende Term von f, so heben sich in der Differenz $f_1 = f - a\sigma_1^{\delta_1} \cdots \sigma_n^{\delta_n}$ die führenden Terme weg, und f_1 hat ein führendes Monom, das kleiner ist als $X_1^{e_1} \cdots X_n^{e_n}$. Als Differenz zweier symmetrischer Polynome ist auch f_1 wieder symmetrisch. Wir schreiben $f = a\sigma_1^{\delta_1} \cdots \sigma_n^{\delta_n} + f_1$, wenden die gleiche Konstruktion an auf f_1, und so weiter. Da die führenden Monome dabei immer kleiner werden und

es nur endlich viele Monome gibt, die graduiert lexikographisch kleiner sind als ein gegebenes Monom, endet diese Konstruktion nach endlich vielen Schritten und drückt f aus als Linearkombination von Monomen in den σ_i. Das gesuchte Polynom g ist nun einfach die entsprechende Linearkombination mit Monomen in den Y_i an Stelle der σ_i. ∎

Als Beispiel betrachten wir das symmetrische Polynom $f = X^3Y+XY^3$. Der führende Term ist X^3Y, wir haben also $e_1 = 3$ und $e_2 = 1$. Damit ist $\delta_1 = 2$ und $\delta_2 = 1$; wir subtrahieren daher im ersten Schritt

$$\sigma_1^{\delta_1}\sigma_2^{\delta_2} = (X+Y)^2(XY) = X^3Y + 2X^2Y^2 + XY^3 \,.$$

Übrig bleibt

$$f_1 = f - (X^3Y + 2X^2Y^2 + XY^3) = -2X^2Y^2 \,.$$

Da es nur ein Monom gibt, ist dieses führend; wir haben $e_1 = e_2 = 2$, also $\delta_1 = 0$ und $\delta_2 = 2$. Daher subtrahieren wir $-2\sigma_2^2$ und erhalten

$$f_2 = f_1 + 2(XY)^2 = 0 \,.$$

Somit ist

$$f = \sigma_1^2\sigma_2 + f_1 = \sigma_1^2\sigma_2 - 2\sigma_2^2 + f_2 = \sigma_1^2\sigma_2 - 2\sigma_2^2 \,.$$

Kombinieren wir den gerade bewiesenen Satz mit dem Wurzelsatz von VIÈTE, besagt er, daß sich jedes symmetrische Polynom in den Nullstellen eines Polynoms als Polynom in den Koeffizienten schreiben läßt:

Satz: P sei ein symmetrisches Polynom in $z_1, \ldots, z_n$, und für jedes n-tupel $z = (z_1, \ldots, z_n)$ sei

$$f^{(z)} = (X - z_1)\cdots(X - z_n) = X^n + a_{n-1}^{(z)}X^{n-1} + \cdots + a_1^{(z)}X + a_0^{(z)} \,.$$

Dann gibt es ein Polynom Q in neuen Variablen $A_0, \ldots, A_{n-1}$ mit der Eigenschaft, daß für alle n-tupel $z = (z_1, \ldots, z_n)$ gilt:

$$P(z) = Q\left(a_0^{(z)}, \ldots, a_{n-1}^{(z)}\right) \,.$$

∎

§8: Die Diskriminante eines Polynoms

Die Lösungsformel

$$x_{1/2} = -\frac{p}{2} \pm \frac{\sqrt{p^2 - 4q}}{2}$$

für eine quadratische Gleichung führt genau dann auf zwei verschiedene Lösungen, wenn der Ausdruck unter der Wurzel nicht verschwindet. Man bezeichnet $\Delta = p^2 - 4q$ als die *Diskriminante* der Gleichung. Wir wollen in diesem Paragraphen auch für Gleichungen höheren Grades eine entsprechende Größe einführen; sie soll genau dann verschwinden, wenn die Gleichung mindestens eine mehrfache Nullstelle hat.

Für ein Polynom, das bereits als Produkt von Linearfaktoren vorliegt, läßt sich so eine Diskriminante leicht konstruieren:

Definition: Die Diskriminante des Polynoms $(X - x_1) \cdots (X - x_n)$ ist

$$\Delta \underset{\text{def}}{=} \prod_{i<j} (x_i - x_j)^2 \, .$$

Für die quadratische Gleichung führt dies auf

$$(x_1 - x_2)^2 = \left(-\frac{p}{2} + \frac{\sqrt{p^2 - 4q}}{2} - \left(-\frac{p}{2} - \frac{\sqrt{p^2 - 4q}}{2} \right) \right)^2 = p^2 - 4q$$

wie oben.

Dadurch, daß die Differenzen der Nullstellen in obiger Definition quadriert werden, ist die Diskriminante unabhängig von der Reihenfolge der Nullstellen – was sie natürlich sinnvollerweise ohnehin sein muß. Sie ist daher eine symmetrische Funktion der Nullstellen und läßt sich als Polynom in den elementarsymmetrischen Funktionen und damit in den Koeffizienten des Polynoms schreiben.

Für das kubische Polynom $(X - x_1)(X - x_2)(X - x_3)$ ist

$$\begin{aligned} \Delta &= (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2 \\ &= x_1^4 x_2^2 - 2x_1^4 x_2 x_3 + x_1^4 x_3^2 - 2x_1^3 x_2^2 + 2x_1^3 x_2^2 x_3 + 2x_1^3 x_2 x_3^2 - 2x_1^3 x_3^3 \\ &\quad + x_1^2 x_2^4 + 2x_1^2 x_2^3 x_3 - 6x_1^2 x_2^2 x_3^2 + 2x_1^2 x_2 x_3^3 + x_1^2 x_3^4 \\ &\quad - 2x_1 x_2^4 x_3 + 2x_1 x_2^3 x_3^2 + 2x_1 x_2^2 x_3^3 - 2x_1 x_2 x_3^4 + x_2^4 x_3^2 - 2x_2^3 x_3^3 + x_2^2 x_3^4 \end{aligned}$$

fast zu grausam, um damit weiter zu rechnen. Da wir immer von der Gleichung X^3+pX+q ausgehen, müssen wir das zum Glück auch nicht: Nach dem Satz von VIÈTE ist die Summe der drei Nullstellen gleich dem negativen Koeffizienten von X^2, und da wir keinen X^2-Term haben, ist diese Summe Null, d.h. $x_3 = -x_1 - x_2$. Damit ist

$$\Delta = (x_1 - x_2)^2(2x_1 + x_2)^2(2x_2 + x_1)^2$$
$$= 4x_1^6 + 12x_1^5x_2 - 3x_1^4x_2^2 - 26x_1^3x_2^3 - 3x_1^2x_2^4 + 12x_1x_2^5 + 4x_2^6$$

Nach dem Wurzelsatz von VIÈTE ist

$$p = x_1x_2 + x_1x_3 + x_2x_3 = x_1x_2 - x_1(x_1 + x_2) - x_2(x_1 + x_2)$$
$$= x_1x_2 - (x_1 + x_2)^2 = -x_1^2 - x_1x_2 - x_2^2$$

und $q = -x_1x_2x_3 = x_1x_2(x_1 + x_2) = x_1^2x_2 + x_1x_2^2$.

Wir gehen nun vor wie im Beweis des Hauptsatzes über symmetrische Funktionen: Um im Ausdruck für die Diskriminante den führenden Term $4x_1^6$ zum Verschwinden zu bringen, addieren wir $4p^3$ und erhalten (nach mühsamer Rechnung) das Ergebnis

$$\Delta + 4p^3 = -27x_1^4x_2^2 - 54x_1^3x_2^3 - 27x_1^2x_2^4 ;$$

addieren wir dazu noch $27q^2$, wird $\Delta+4p^3+27q^2 = 0$. Die Diskriminante des Polynoms $X^3 + pX + q$ ist also $\Delta = -(4p^3 + 27q^2)$. Das kubische Polynom $X^3 + pX + q$ hat somit genau dann eine mehrfache Nullstelle, wenn $\Delta = -(4p^3 + 27q^2)$ verschwindet.

§9: Der casus irreducibilis bei kubischen Gleichungen

Falls alle drei Nullstellen des kubischen Polynoms X^3+pX+q mit reellen Koeffizienten p, q reell und verschieden sind, ist die Diskriminante als Produkt von Quadraten reeller Zahlen positiv. In

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} ,$$

ist daher

$$\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = \frac{q^2}{4} + \frac{p^3}{27} = \frac{27q^2 + 4p^3}{108} = \frac{-\Delta}{108}$$

negativ. In diesem Fall steht somit unter der Quadratwurzel eine negative Zahl, so daß u^3 einen nichtverschwindenden Imaginärteil hat. *Falls alle drei Nullstellen reell und verschieden sind, muß also u eine nichtreelle komplexe Zahl sein.* Man bezeichnet diesen Fall aus diesem Grund als den *casus irreducibilis.* Mit der Irreduzibilität von Polynomen, die wir bald betrachten werden, hat dies nichts zu tun.

Wir wollen uns überlegen, daß umgekehrt im Falle einer positiven Diskriminanten bei reellen Koeffizienten auch alle drei Lösungen reell sein müssen. Dazu machen wir einen trigonometrischen Ansatz, mit dem wir sie ohne Umweg über die komplexen Zahlen rein reell bestimmen können.

Wir schreiben $x = r\cos\varphi$ mit einer nichtnegativen Zahl r und einem Winkel φ zwischen 0 und π. (Da der Kosinus eine gerade Funktion ist, können wir uns auf nichtnegative Winkel beschränken.) Die kubische Gleichung wird dann zu

$$x^3 + px + q = r^3\cos^3\varphi + pr\cos\varphi + q = 0\,.$$

Nach den EULERschen Formeln ist

$$(\cos\varphi + i\sin\varphi)^3 = (e^{i\varphi})^3 = e^{3i\varphi} = \cos 3\varphi + i\sin 3\varphi\,.$$

Andererseits ist nach dem binomischen Lehrsatz

$$(\cos\varphi + i\sin\varphi)^3 = \cos^3\varphi + 3i\cos^2\varphi\sin\varphi - 3\cos\varphi\sin^2\varphi - i\sin^3\varphi\,;$$

durch Vergleich der Realteile sehen wir, daß

$$\begin{aligned}\cos 3\varphi &= \cos^3\varphi - 3\cos\varphi\sin^2\varphi = \cos^3\varphi - 3\cos\varphi\cdot(1-\cos^2\varphi)\\ &= 4\cos^3\varphi - 3\cos\varphi\end{aligned}$$

ist und damit $\cos^3\varphi = \dfrac{\cos 3\varphi + 3\cos\varphi}{4}$.

Die Gleichung wird damit zu

$$\begin{aligned}&\frac{r^3}{4}(\cos 3\varphi + 3\cos\varphi) + pr\cos\varphi + q\\ =\ &\frac{r^3}{4}\cos 3\varphi + r\cdot\left(\frac{3}{4}r^2 + p\right)\cos\varphi + q = 0\,.\end{aligned}$$

Wir können sie vereinfachen, wenn wir r so wählen, daß die Klammer verschwindet: Mit $r = \sqrt{-4p/3}$ erhalten wir

$$\frac{r^3}{4}\cos 3\varphi + q = 0$$

und damit $\cos 3\varphi = -\frac{4q}{r^3} = -4q\sqrt{-\frac{27}{4^3p^3}} = -\frac{q}{2}\sqrt{\frac{-27}{p^3}}$.

r ist eine reelle Zahl, denn da $-\Delta = 4p^3 + 27q^2 < 0$ ist und $q^2 \geq 0$, muß p negativ sein. Außerdem hat der Ausdruck für $\cos 3\varphi$ höchstens den Betrag eins, denn sein Quadrat ist

$$\frac{q^2}{4} \cdot \frac{-27}{p^3} = \frac{27q^2}{-4p^3} ,$$

wobei hier im Nenner eine positive und im Zähler zumindest eine nichtnegative Zahl steht. Wegen $\Delta < 0$ ist $27q^2 < -4p^3$, also hat der Bruch einen Wert zwischen 0 und 1, und obiger Ausdruck liegt zwischen -1 und 1. Somit können wir einen Winkel φ aus $[0, \pi]$ finden mit $\cos 3\varphi = -q/2\sqrt{-27/p^3}$.

Tatsächlich finden wir dann nicht nur einen Winkel, sondern drei, denn $\cos 3\varphi$ nimmt im Intervall $[0, \pi]$ jeden Wert aus $[-1, 1]$ dreimal an: Ist $\varphi_1 \in [0, \frac{1}{3}\pi]$ der kleinste solche Winkel, sind $\varphi_{2/3} = \frac{2}{3}\pi \pm \varphi_1$ die beiden anderen. Die drei Lösungen der gegebenen Gleichung sind also

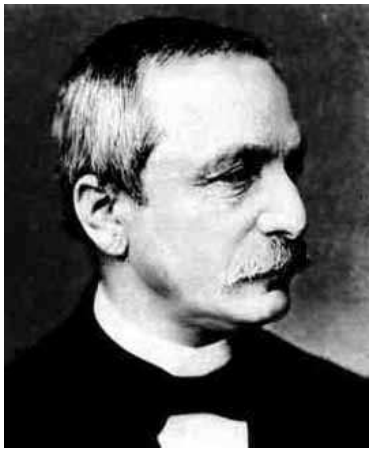
$$x_i = r\cos\varphi_i = \sqrt{\frac{-4p}{3}}\cos\varphi_i \quad \text{mit} \quad \cos 3\varphi_i = -\frac{q}{2}\sqrt{\frac{-27}{p^3}} ,$$

und sie sind allesamt reell.

Man beachte, daß wir hier nicht nur Grundrechenarten und Wurzeln verwenden, um die x_i auszudrücken: Um die Winkel φ_i zu bestimmen, brauchen wir den Arkuskosinus und zur Berechnung der x_i zusätzlich noch den Kosinus.

Kapitel 2
Rechnen mit ganzen Zahlen

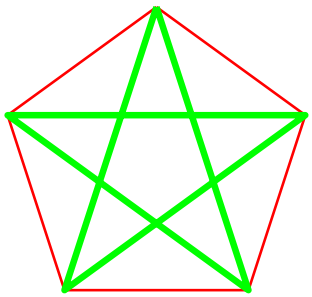
In einen Vortrag bei der Berliner Naturforscher-Versammlung sagte LEOPOLD KRONECKER 1886: „Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.“ In seinem gesamten Werk versuchte er immer wieder, alles auf arithmetische Eigenschaften ganzer Zahlen zurückzuführen.



LEOPOLD KRONECKER (1823–1891) ist heute zwar Vielen nur im Zusammenhang mit dem KRONECKER-δ bekannt, er war aber einer der bedeutendsten deutschen Mathematiker seiner Zeit. Seine Arbeiten befaßten sich mit Algebra, Zahlentheorie und Analysis, wobei er insbesondere die Verbindungen zwischen der Analysis und den beiden anderen Gebieten erforschte. Bekannt ist auch seine Ablehnung jeglicher mathematischer Methoden, die, wie die Mengenlehre oder Teile der Analysis, unendliche Konstruktionen verwenden. Er war deshalb mit vielen anderen bedeutenden Mathematikern seiner Zeit verfeindet, z.B. mit CANTOR und mit WEIERSTRASS.

Auch in der frühen griechischen Mathematik spielten die natürlichen Zahlen eine herausragende Rolle. Zwar ging es dort vor allem um Geometrie, und eine Strecke läßt sich beispielsweise beliebig oft halbieren, was bei natürlichen Zahlen bekanntlich nicht der Fall ist. Schon PYTHAGORAS ($\sim$570–$\sim$510) und seine Schüler versuchten aber stets, zu zwei Strecken ein gemeinsames „Maß“ zu finden, d.h. eine dritte Strecke, von der beide Ausgangsstrecken ganzzahlige Vielfache sind. Einige Gelehrte spekulieren sogar, daß PLATON (428/427–348/347) in seiner (nicht überlieferten) ungeschriebenen Lehre die gesamte Welt der Ideen auf (natürliche) Zahlen zurückführen wollte.

Das Wahrzeichen der Pythagoräer war das Pentagramm, d.h. die Figur aus allen Diagonalen eines regelmäßigen Fünfecks (Pentagons). Um 450 vor Christus erkannte der Pythagoräer HIPPASSOS VON METAPONT, daß es für die Diagonale und die Seite eines Pentagons kein solchen gemeinsames Maß geben konnte, daß es also auch das gibt, was wir heute als irrationale Zahlen bezeichnen. Hier stehen die beiden Streckenlängen im Verhältnis des *goldenen Schnitts,* in Zahlen ausgedrückt $\frac{1}{2}(1+\sqrt{5})$.



Die Mitglieder von PLATONs Akademie interessierten sich nicht für die Lösung von Gleichungen. Praktische Anwendungen der Arithmetik waren für sie etwas Minderwertiges, das nur für Handwerker, Händler und andere Leute, die ihr Geld durch Arbeit verdienen mußten, taugte, und denen war es egal, ob der Weinmenge, den sie verkauften, ein ganzzahliges Vielfaches einer Maßeinheit war oder nicht.

Erst DIOPHANTOS von Alexandrien beschäftigte sich in seinem Buch *Arithmetika* systematisch mit ganzzahligen Lösungen von Gleichungen mit ganzzahligen Koeffizienten; man bezeichnet diese deshalb heute nach ihm als *diophantische Gleichungen.*

Über das Leben von DIOPHANTOS ist praktisch nichts bekannt. Anhand der Daten anderer Autoren, die ihn zitierten *bzw.* die von ihm zitiert wurden, läßt sich *mit Sicherheit* nur sagen, daß sein Werk später als 150 vor Christus und früher als 350 nach Christus entstanden sein muß. Es ist nur teilweise überliefert. Bemerkenswert ist auch, daß er als erster ein eigenes Symbol für eine unbekannte Zahl benutzte. Dieses Symbol war allerdings kein Buchstabe, sondern eine Neuschöpfung.

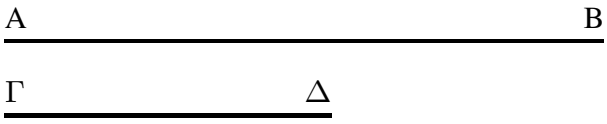
In diesem Kapitel wollen wir zumindest lineare diophantische Gleichungen betrachten sowie einige andere Probleme im Umgang mit ganzen Zahlen. Ausgangspunkt für die meisten Anwendungen ist der aus der Linearen Algebra bekannte EUKLIDische Algorithmus, den wir deshalb kurz wiederholen wollen.

§1: Der Euklidische Algorithmus

Bei EUKLID, in Proposition 2 des siebten Buchs seiner *Elemente,* wird er (in der Übersetzung von CLEMENS THAER in Oswalds Klassikern der exakten Wissenschaft) so beschrieben:

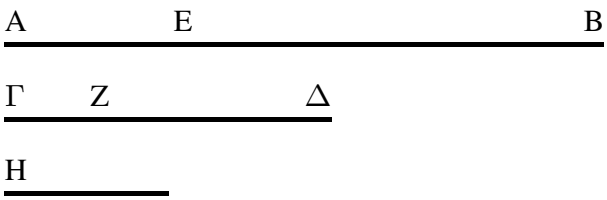
Zu zwei gegebenen Zahlen, die nicht prim gegeneinander sind, ihr größtes gemeinsames Maß zu finden.

Die zwei gegebenen Zahlen, die nicht prim, gegeneinander sind, seien AB, ΓΔ. Man soll das größte gemeinsame Maß von AB, ΓΔ finden.



Wenn ΓΔ hier AB mißt – sich selbst mißt es auch – dann ist ΓΔ gemeinsames Maß von ΓΔ, AB. Und es ist klar, daß es auch das größte ist, denn keine Zahl größer ΓΔ kann ΓΔ messen.

Wenn ΓΔ aber AB nicht mißt, und man nimmt bei AB, ΓΔ abwechselnd immer das kleinere vom größeren weg, dann muß (schließlich) eine Zahl übrig bleiben, die die vorangehende mißt. Die Einheit kann nämlich nicht übrig bleiben; sonst müßten AB, ΓΔ gegeneinander prim sein, gegen die Voraussetzung. Also muß eine Zahl übrig bleiben, die die vorangehende mißt. ΓΔ lasse, indem es BE mißt, EA, kleiner als sich selbst übrig; und EA lasse, indem es ΔZ mißt, ZΓ, kleiner als sich selbst übrig; und ΓZ messe AE.



Da Γ*Z* AE mißt und AE ΔZ, muß ΓZ auch ΔZ messen; es mißt aber auch sich selbst, muß also auch das Ganze ΓΔ messen. ΓΔ mißt aber BE; also mißt ΓZ auch BE; es mißt aber auch EA, muß also auch das Ganze BA messen. Und es mißt auch ΓΔ; ΓZ mißt also AB und ΓΔ; also ist ΓZ gemeinsames Maß von AB, ΓΔ. Ich behaupte, daß es auch das größte ist. Wäre nämlich ΓZ nicht das größte gemeinsame Maß von AB, ΓΔ, so müßte irgendeine Zahl größer ΓZ die Zahlen AB und ΓΔ messen. Dies geschehe; die Zahl sei H. Da H dann ΓΔ mäße und ΓΔ BE mißt, mäße H auch BE; es soll aber auch das Ganze BA messen, müßte also auch den Rest AE messen. AE mißt aber ΔZ; also müßte H auch ΔZ messen; es soll aber auch das Ganze ΔΓ messen, müßte also auch den Rest ΓZ messen, als größere Zahl die kleinere; dies ist unmöglich. Also kann keine Zahl größer ΓZ die Zahlen AB und ΓΔ messen; ΓZ ist also das größte gemeinsame Maß von AB, ΓΔ; dies hatte man beweisen sollen.

Aus heutiger Sicht erscheint die Voraussetzung, daß die betrachteten Größen nicht teilerfremd sein dürfen, seltsam. Sie erklärt sich daraus, daß in der griechischen Philosophie und Mathematik die Einheit eine Sonderrolle einnahm und nicht als Zahl angesehen wurde: Die Zahlen begannen erst mit der Zwei. Dementsprechend führt EUKLID in Proposition 1 des siebten Buchs fast wörtlich dieselbe Konstruktion durch für den Fall von teilerfremden Größen. Schon wenig später wurde die Eins auch in Griechenland als Zahl anerkannt, und für uns heute ist die Unterscheidung ohnehin bedeutungslos. Wir können die Bedingung, daß der ggT ungleich eins sein soll, also einfach ignorieren.

Das dem EUKLIDischen Algorithmus zugrunde liegende Prinzip der *Wechselwegnahme* oder wechselseitigen Subtraktion war in der griechischen Mathematik spätestens gegen Ende des fünften vorchristlichen Jahrhunderts bekannt unter dem Namen Antanairesis (ἀνταναιρεσις) oder auch Anthyphairesis (ἀνθυφαιρεσις); damit bewies bereits HIPPASSOS, daß das Längenverhältnis zwischen der Diagonale und der Seite eines regelmäßigen Fünfecks keine rationale Zahl ist. Auch der Algorithmus selbst geht mit ziemlicher Sicherheit, wie so vieles in den Elementen, *nicht* erst auf EUKLID zurück: Seine *Elemente* waren das wohl mindestens vierte Buchprojekt dieses Namens, und alles spricht dafür, daß er vieles von seinen Vorgängern übernommen hat. Seine Elemente waren dann aber mit Abstand die erfolgreichsten, so daß die anderen in Vergessenheit gerieten und verloren gingen; EUKLID wurde schließlich als *der* Stoichist bekannt nach dem griechischen Titel στοιχεῖα der Elemente.

Es ist nicht ganz sicher, ob EUKLID (Εὐκλειδης) wirklich gelebt hat; es ist möglich, wenn auch sehr unwahrscheinlich, daß EUKLID nur ein Pseudonym für eine Autorengruppe ist. (Das nebenstehende Bild aus dem 18. Jahrhundert ist reine Phantasie.) EUKLID ist vor allem bekannt als Autor der *Elemente,* in denen er die Geometrie seiner Zeit systematisch darstellte und (in gewisser Weise) auf wenige Definitionen sowie die berühmten fünf Postulate zurückführte. Sie entstanden um 300 v. Chr.. EUKLID arbeitete wohl am Museion in Alexandrien. Außer den Elementen schrieb er ein Buch über Optik und weitere, teilweise verschollene Bücher.

Wenn wir nicht mit Zirkel und Lineal arbeiten, sondern rechnen, können wir die mehrfache „Wegnahme“ einer Strecke von einer anderen einfacher beschreiben durch eine Division mit Rest: Sind a und b die (als natürliche Zahlen vorausgesetzten) Längen der beiden Strecken und ist $a : b = q$ Rest r, so kann man q mal die Strecke b von a wegnehmen; was übrig bleibt ist eine Strecke der Länge $r < b$.

EUKLIDs Konstruktion wird dann zu folgendem Algorithmus für zwei natürliche Zahlen a, b:

Schritt 0: Setze $r_0 = a$ und $r_1 = b$.

Schritt i, $i \geq 1$**:** Falls r_i verschwindet, endet der Algorithmus mit $\mathrm{ggT}(a, b) = r_{i-1}$; andernfalls sei r_{i+1} der Rest bei der Division von r_{i-1} durch r_i.

EUKLID behauptet, daß dieser Algorithmus stets endet und daß das Ergebnis der größte gemeinsame Teiler der Ausgangszahlen a, b ist, d.h. die größte natürliche Zahl, die sowohl a als auch b teilt.

Da der Divisionsrest r_{i+1} stets echt kleiner ist als sein Vorgänger r_i und eine Folge immer kleiner werdender nichtnegativer ganzer Zahlen notwendigerweise nach endlich vielen Schritten die Null erreicht, muß der Algorithmus in der Tat stets enden. Daß er mit dem richtigen Ergebnis endet, ist ebenfalls leicht zu sehen, denn im i-ten Schritt ist

$$r_{i-1} = q_i r_i + r_{i+1} \quad \text{oder} \quad r_{i+1} = r_{i-1} - q_i r_i \,,$$

so daß jeder gemeinsame Teiler von r_i und r_{i+1} auch ein Teiler von r_{i-1} ist und umgekehrt jeder gemeinsame Teiler von r_{i-1} und r_i auch r_{i+1} teilt. Somit haben r_i und r_{i-1} dieselben gemeinsamen Teiler wie r_i und r_{i+1}, insbesondere haben sie denselben größten gemeinsamen Teiler. Durch Induktion folgt, daß in jedem Schritt $\mathrm{ggT}(r_i, r_{i-1}) = \mathrm{ggT}(a, b)$ ist. Im letzten Schritt ist $r_i = 0$; da jede natürliche Zahl Teiler der Null ist, ist dann $r_{i-1} = \mathrm{ggT}(r_i, r_{i-1}) = \mathrm{ggT}(a, b)$, wie behauptet.

Mehr als zwei Tausend Jahre nach der Entdeckung von Anthyphairesis und EUKLIDischem Algorithmus, 1624 in Bourg-en-Bresse, modifizierte BACHET DE MÉZIRIAC in der zweiten Auflage seines Buchs *Problèmes*

plaisants et délectables qui se fonts par les nombres den Algorithmus so, daß er zu zwei teilerfremden natürlichen Zahlen a, b zwei weitere natürliche Zahlen x, y konstruiert, für die $ax - by = 1$ ist. Bei ihm heißt das in seiner Proposition XVIII: *Deux nombres premiers entre eux estant donnéz, treuuer le moindre multiple de chascun d'iceux, surpassant de l'unité un multiple de l'autre.* (Für zwei gegebene teilerfremde Zahlen das kleinste Vielfache von jeder der beiden zu finden, das um eins größer ist als ein Vielfaches der anderen.) Er sucht also nicht nur irgendwelche natürlichen Zahlen x, y, sondern verlangt auch noch, daß x minimal ist. (1993 brachte der Verlag Blanchard eine vereinfachte fünfte Auflage heraus, in der so „komplizierte" Dinge wie der Beweis dieser Proposition leider fehlen.)

CLAUDE GASPAR BACHET SIEUR DE MÉZIRIAC (1581-1638) verbrachte den größten Teil seines Lebens in seinem Geburtsort Bourg-en-Bresse. Er studierte bei den Jesuiten in Lyon und Milano und trat 1601 in den Orden ein, trat aber bereits 1602 wegen Krankheit wieder aus und kehrte nach Bourg zurück. Die erste Auflage seines Buchs erschien 1612. Am bekanntesten ist BACHET für seine lateinische Übersetzung der *Arithmetika* von DIOPHANTOS. In einem Exemplar davon schrieb FERMAT seine Vermutung an den Rand. Auch Gedichte von BACHET sind erhalten. 1635 wurde er Mitglied der französischen Akademie der Wissenschaften.

Seine Methode läßt sich leicht verallgemeinern auf den Fall, daß a, b nicht teilerfremd sind: Man muß einfach beide durch ihren ggT teilen und das Ergebnis wieder mit diesem multiplizieren.

Das Verfahren beruht darauf, daß wir bei der Division mit Rest den Divisionsrest als Dividend minus Quotient mal Divisor darstellen; im EUKLIDischen Algorithmus ist also jedes r_i eine ganzzahlige Linearkombination von r_{i-1} und r_{i-2}; indem wir diese Linearkombinationen ineinander einsetzen, erhalten wir den ggT als letzten von null verschiedenen Divisionsrest als ganzzahlige Linearkombination der beiden Ausgangszahlen. Obwohl dies bei EUKLID nicht zu finden ist, redet man heute vom *erweiterten* EUKLID*ischen Algorithmus* oder von der *Identität von* BÉZOUT, benannt nach einem Mathematiker, der das Verfahren

142 Jahre nach BACHET beschrieb und auf Polynome in einer Variablen verallgemeinerte.

ETIENNE BÉZOUT (1730-1783) wurde in Nemours in der Île-de-France geboren, wo seine Vorfahren Magistrate waren. Er ging stattdessen an die Akademie der Wissenschaften und schrieb mehrere Lehrbücher für die Militärausbildung. Im 1766 erschienenen dritten Band (von vier) seines *Cours de Mathématiques à l'usage des Gardes du Pavillon et de la Marine* ist die Identität von BÉZOUT dargestellt. Seine Bücher waren so erfolgreich, daß sie ins Englische übersetzt und als Lehrbücher z.B. in Harvard benutzt wurden. Heute ist er vor allem auch bekannt durch seinen Beweis, daß sich zwei Kurven der Grade n und m ohne gemeinsame Komponenten in höchstens nm Punkten schneiden können.

Formal sieht der erweiterte EUKLIDische Algorithmus folgendermaßen aus:

Schritt 0: Setze $r_0 = a$, $r_1 = b$, $\alpha_0 = \beta_1 = 1$ und $\alpha_1 = \beta_0 = 0$. Für $i = 1$ ist dann

$$r_{i-1} = \alpha_{i-1}a + \beta_{i-1}b \quad \text{und} \quad r_i = \alpha_i a + \beta_i b\,.$$

Im i-ten Schritt werden neue Zahlen berechnet derart, daß diese Gleichungen auch für $i + 1$ gelten:

Schritt i, $i \geq 1$**:** Falls r_i verschwindet, endet der Algorithmus mit

$$\text{ggT}(a, b) = r_{i-1} = \alpha_{i-1}a + \beta_{i-1}b\,.$$

Andernfalls dividiere man r_{i-1} durch r_i; der Divisionsrest sei r_{i+1}. Dann ist

$$\begin{aligned} r_{i+1} = r_{i-1} - q_i r_i &= (\alpha_{i-1}a + \beta_{i-1}b) - q_i(\alpha_i a + \beta_i b) \\ &= (\alpha_{i-1} - q_i\alpha_i)a + (\beta_{i-1} - q_i\beta_i)b\,; \end{aligned}$$

die gewünschten Gleichungen gelten also für

$$\alpha_{i+1} = \alpha_{i-1} - q_i\alpha_i \quad \text{und} \quad \beta_{i+1} = \beta_{i-1} - q_i\beta_i\,.$$

Genau wie oben folgt, daß der Algorithmus für alle natürlichen Zahlen a und b endet und daß am Ende der richtige ggT berechnet wird; außerdem sind die α_i und β_i so definiert, daß in jedem Schritt $r_i = \alpha_i a + \beta_i b$ ist,

insbesondere wird also im letzten Schritt der ggT als Linearkombination der Ausgangszahlen dargestellt.

Dieser Algorithmus liefert sofort ein Verfahren, mit dem wir diophantische Gleichungen der Form $ax + by = c$ mit $a, b, c \in \mathbb{Z}$ für zwei Unbekannte $x, y \in \mathbb{Z}$ lösen können:

Der größte gemeinsame Teiler $d = \text{ggT}(a, b)$ von a und b teilt offensichtlich jeden Ausdruck der Form $ax + by$ mit $x, y \in \mathbb{Z}$; falls d kein Teiler von c ist, kann es also keine ganzzahlige Lösung geben.

Ist aber $c = rd$ ein Vielfaches von d und ist $d = \alpha a + \beta b$ die lineare Darstellung des ggT nach dem erweiterten EUKLIDischen Algorithmus, so haben wir mit $x = r\alpha$ und $y = r\beta$ offensichtlich eine Lösung gefunden.

Ist (x', y') eine weitere Lösung, so ist

$$a(x - x') + b(y - y') = c - c = 0 \quad \text{oder} \quad a(x - x') = b(y' - y)\,.$$

$v = a(x - x') = b(y' - y)$ ist also ein gemeinsames Vielfaches von a und b und damit auch ein Vielfaches des kleinsten gemeinsamen Vielfachen von a und b. Dieses kleinste gemeinsame Vielfache ist ab/d, es muß also eine ganze Zahl m geben mit

$$x - x' = m \cdot \frac{b}{d} \quad \text{und} \quad y' - y = m \cdot \frac{a}{d}\,.$$

Die allgemeine Lösung der obigen Gleichung ist somit

$$x = r\alpha - m \cdot \frac{b}{d} \;\text{ und }\; y = r\beta + m \cdot \frac{a}{d} \quad \text{mit} \quad m \in \mathbb{Z}\,.$$

Als Beispiel betrachten wir eines der Probleme aus dem Buch von BACHET:

Il y a 41 personnes en un banquet tant hommes que femmes et enfants qui en tout dépensent 40 sous, mais chaque homme paye 4 sous, chaque femme 3 sous, chaque enfant 4 deniers. Je demande combien il y a d'hommes, combien de femmes, combien d'enfants.

(Bei einem Bankett sind 41 Personen, Männer, Frauen und Kinder, die zusammen vierzig Sous ausgeben, aber jeder Mann zahlt vier Sous, jede

Frau drei Sous und jedes Kind 4 Deniers. Ich frage, wie viele Männer, wie viele Frauen und wie viele Kinder es sind.)

Sobald man weiß, daß zwölf Deniers ein Sou sind (und zwanzig Sous ein Pfund), kann man dies in ein lineares Gleichungssystem übersetzen: Ist x die Zahl der Männer, y die der Frauen und z die der Kinder, so muß gelten $x + y + z = 41$ und $4x + 3y + \frac{1}{3}z = 40$.

Im Gegensatz zum Fall der in Schule und Linearer Algebra betrachteten linearen Gleichungssystemen kommen hier natürlich nur nichtnegative ganze Zahlen als Lösungen in Frage.

Zur Lösung kann man zunächst die erste Gleichung nach z auflösen und in die zweite Gleichung einsetzen; dies führt auf die Gleichung

$$\frac{11}{3}x + \frac{8}{3}y = \frac{79}{3} \quad \text{oder} \quad 11x + 8y = 79\,.$$

Da elf und acht teilerfremd sind, teilt ihr ggT die rechte Seite; das Problem hat also ganzzahlige Lösungen. Um diese zu finden, müssen wir zunächst den ggT von 11 und 8 als Linearkombination dieser Zahlen darstellen.

Elf durch acht ist eins Rest drei, also ist $3 = 1 \cdot 11 - 1 \cdot 8$.

Im nächsten Schritt dividieren wir acht durch drei mit dem Ergebnis zwei Rest zwei, also ist $2 = 1 \cdot 8 - 2 \cdot 3 = 1 \cdot 8 - 2 \cdot (1 \cdot 11 - 1 \cdot 8) = -2 \cdot 11 + 3 \cdot 8$.

Im letzten Schritt wird daher drei durch zwei dividiert und wir sehen erstens, daß der ggT gleich eins ist (was hier keine Überraschung ist), und zweitens, daß gilt $1 = 3 - 2 = (1 \cdot 11 - 1 \cdot 8) - (-2 \cdot 11 + 3 \cdot 8) = 3 \cdot 11 - 4 \cdot 8$.

Damit haben wir auch eine Darstellung von 79 als Linearkombination von elf und acht:

$$79 = 79 \cdot (3 \cdot 11 - 4 \cdot 8) = 237 \cdot 11 - 316 \cdot 8\,.$$

Dies ist allerdings nicht die gesuchte Lösung: BACHET dachte sicherlich nicht an 237 Männer, -316 Frauen und 119 Kinder.

Nun ist aber die obige Gleichung $1 = 3 \cdot 11 - 4 \cdot 8$ nicht die einzige Möglichkeit zur Darstellung der Eins als Linearkombination von acht

und elf: Da $8 \cdot 11 - 11 \cdot 8$ verschwindet, können wir ein beliebiges Vielfaches dieser Gleichung dazu addieren und bekommen die allgemeinere Lösung

$$(3 + 8k) \cdot 11 - (4 + 11k) \cdot 8 = 1 \, .$$

Entsprechend können wir auch ein beliebiges Vielfaches dieser Gleichung zur Darstellung von 79 addieren:

$$79 = (237 + 8k) \cdot 11 - (316 + 11k) \cdot 8 \, .$$

Wir müssen k so wählen, daß sowohl die Anzahl $237 + 8k$ der Männer als auch die Anzahl $-(316 + 11k)$ der Frauen positiv oder zumindest nicht negativ wird, d.h. $-\frac{237}{8} \leq k \leq -\frac{316}{11}$. Da k ganzzahlig sein muß, kommt nur $k = -29$ in Frage; es waren also fünf Männer, drei Frauen und dazu noch $41 - 5 - 3 = 33$ Kinder. Ihre Gesamtausgaben belaufen sich in der Tat auf $5 \cdot 4 + 3 \cdot 3 + 33 \cdot \frac{1}{3} = 40$ Sous.

Im Beweis, daß der EUKLIDische Algorithmus stets nach endlich vielen Schritten abbricht, hatten wir argumentiert, daß der Divisionsrest stets kleiner ist als der Divisor, so daß er irgendwann einmal null werden muß; dann endet der Algorithmus.

Damit haben wir auch eine obere Schranke für den Rechenaufwand zur Berechnung von $\mathrm{ggT}(a, b)$: Wir müssen höchstens b Divisionen durchführen.

Das erscheint zwar auf den ersten Blick als ein recht gutes Ergebnis; wenn man aber bedenkt, daß der EUKLIDische Algorithmus heute in der Kryptographie auf rund tausendstellige Zahlen angewendet wird, verliert diese Schranke schnell ihre Nützlichkeit: Da unser Universum ein geschätztes Alter von zehn Milliarden Jahren, also ungefähr $3 \cdot 10^{18}$ Sekunden hat, ist klar, daß auch der schnellste heutige Computer, selbst wenn er zu Beginn des Universum zu rechnen begonnen hätte, bis heute nur einen verschwindend kleinen Bruchteil von 10^{1000} Divisionen ausgeführt hätte. Wäre 10^{1000} eine realistische Aufwandsabschätzung, könnten wir an eine Anwendung des EUKLIDischen Algorithmus auf tausendstellige Zahlen nicht einmal denken. Zum Glück fand GABRIEL LAMÉ 1844 eine viel schärfere Schranke, für deren Beweis ich auf

Lehrbücher der Zahlentheorie oder auf das Skriptum meiner Zahlentheorievorlesung verweisen möchte:

Satz von Lamé: Die kleinsten natürlichen Zahlen a, b, für die beim EUKLIDischen Algorithmus $n \geq 2$ Divisionen benötigt werden, sind $a = F_{n+2}$ und $b = F_{n+1}$. Dabei sind die sogenannten FIBONACCI-Zahlen F_n rekursiv definiert durch

$$F_0 = F_1 = 1 \quad \text{und} \quad F_{n+1} = F_n + F_{n-1} \quad \text{für } n \geq 1 \,.$$

(Für $n = 1$ gilt der Satz nur, wenn wir zusätzlich voraussetzen, daß $a \neq b$ ist; für $n \geq 2$ ist dies automatisch erfüllt.)

GABRIEL LAMÉ (1795–1870) studierte von 1813 bis 1817 Mathematik an der Ecole Polytechnique, danach bis 1820 Ingenieurwissenschaften an der Ecole des Mines. Auf Einladung Alexanders I. kam er 1820 nach Rußland, wo er in St. Petersburg als Professor und Ingenieur unter anderem Vorlesungen über Analysis, Physik, Chemie und Ingenieurwissenschaften hielt. 1832 erhielt er einen Lehrstuhl für Physik an der Ecole Polytechnique in Paris, 1852 einen für mathematische Physik und Wahrscheinlichkeitstheorie an der Sorbonne. 1836/37 war er wesentlich am Bau der Eisenbahnlinien Paris-Versailles und Paris–St. Germain beteiligt.

Man kann auch eine geschlossene Formel für die F_n finden; setzt man diese ein, erhält man für $b = F_{n+1}$ mit dem Satz von LAMÉ die Abschätzung

$$\begin{aligned} n &\approx \log_\phi \sqrt{5}\, b - 1 = \log_\phi b + \log_\phi \sqrt{5} - 1 = \frac{\ln b}{\ln \phi} + \frac{\ln \sqrt{5}}{\ln \phi} - 1 \\ &\approx 2{,}078 \ln b + 0{,}672 \,. \end{aligned}$$

Für beliebige Zahlen $a > b$ können nicht mehr Divisionen notwendig sein als für die auf b folgenden nächstgrößeren FIBONACCI-Zahlen, also gibt obige Formel für jedes b eine obere Grenze. Die Anzahl der Divisionen wächst daher nicht (wie oben bei der naiven Abschätzung) wie b, sondern höchstens wie $\log b$. Für tausendstellige Zahlen a, b müssen wir daher nicht mit 10^{1000} Divisionen rechnen, sondern mit weniger als fünf Tausend, was auch mit weniger leistungsfähigen Computern problemlos und schnell möglich ist.

Tatsächlich gibt natürlich auch die hier berechnete Schranke nur selten den tatsächlichen Aufwand wieder; fast immer werden wir mit erheblich weniger auskommen. Im übrigen ist auch alles andere als klar, ob wir den ggT auf andere Weise nicht möglicherweise schneller berechnen können. Da wir aber für Zahlen der Größenordnung, die in heutigen Anwendungen interessieren, selbst mit der Schranke für den schlimmsten Fall ganz gut leben können, sei hier auf diese Fragen nicht weiter eingegangen. Interessenten finden mehr dazu z.B. in den Abschnitten 4.5.2+3 des Buchs
DONALD E. KNUTH: The Art of Computer Programming, *vol. 2:* Seminumerical Algorithms, *Addison-Wesley,* ^{3}1997
Eine deutsche Übersetzung des relevanten Kapitels erschien 2001 bei Springer unter dem Titel *Arithmetik.*

§2: Die multiplikative Struktur der ganzen Zahlen

Eine Primzahl ist bekanntlich eine natürliche Zahl p, die genau zwei Teiler hat, nämlich die Eins und sich selbst; insbesondere ist also $p \neq 1$. Der erweiterte EUKLIDische Algorithmus zeigt eine wichtige Folgerung aus dieser Definition:

Lemma: Wenn eine Primzahl das Produkt $\prod_{i=1}^{n} a_i$ von n natürlichen Zahlen a_i teilt, teilt sie mindestens einen der Faktoren.

Beweis: Für $n = 1$ gibt es nichts zu beweisen.

Für $n = 2$ setzen wir kurz $a_1 = a$ und $a_2 = b$. Falls p ein Teiler von a ist, stimmt die Behauptung. Andernfalls muß der ggT von a und p gleich eins sein, denn er ist ein Teiler von p und ungleich p. Es gibt daher eine Darstellung

$$1 = \alpha a + \beta p \quad \text{mit} \quad \alpha, \beta \in \mathbb{Z}\,.$$

Dann ist $b = \alpha ab + \beta pb$ durch p teilbar, denn sowohl ab also auch pb sind Vielfache von p, was die Behauptung beweist.

Für $n > 2$ beweisen wir die Behauptung induktiv: Angenommen, sie gilt für $n - 1$ und p teilt $\prod_{i=1}^{n} a_i$. Falls p ein Teiler von a_n ist, stimmt die Behauptung; andernfalls muß p wegen des bereits bewiesenen Falls

$n = 2$ ein Teiler von $\prod_{i=1}^{n-1} a_i$ sein und teilt nach Induktionsannahme einen der Faktoren.

■

Eine wichtige Folgerung aus diesem Lemma ist der sogenannte *Hauptsatz der elementaren Zahlentheorie:*

Satz: Jede natürliche Zahl läßt sich bis auf Reihenfolge eindeutig als ein Produkt von Primzahlpotenzen schreiben.

Beweis: Wir zeigen zunächst, daß sich jede natürliche Zahl überhaupt als Produkt von Primzahlpotenzen schreiben läßt. Falls dies nicht der Fall wäre, gäbe es ein minimales Gegenbeispiel M. Dies kann nicht die Eins sein, denn die ist ja das leere Produkt, und es kann auch keine Primzahl sein, denn die ist ja das Produkt mit sich selbst als einzigem Faktor. Somit hat M einen echten Teiler N, d.h. $1 < N < M$.

Da M das minimale Gegenbeispiel war, lassen sich N und $\frac{M}{N}$ als Produkte von Primzahlpotenzen schreiben, also auch $M = N \cdot \frac{M}{N}$.

Bleibt noch zu zeigen, daß die Produktdarstellung bis auf die Reihenfolge der Faktoren eindeutig ist. Auch hier gäbe es andernfalls wieder ein minimales Gegenbeispiel M, das somit mindestens zwei verschiedene Darstellungen

$$M = \prod_{i=1}^{r} p_i^{e_i} = \prod_{j=1}^{s} q_j^{f_j}$$

hätte. Da die Eins durch kein Produkt dargestellt werden kann, in dem wirklich eine Primzahl vorkommt, ist $M > 1$, und somit steht in jedem der beiden Produkte mindestens eine Primzahl.

Da p_1 Teiler von M ist, teilt es auch das rechtsstehende Produkt, also nach dem gerade bewiesenen Lemma mindestens einen der Faktoren, d.h. mindestens ein q_j. Da q_j eine Primzahl ist, muß dann $p_1 = q_j$ sein. Da M als minimales Gegenbeispiel vorausgesetzt war, unterscheiden sich die beiden Produkte, aus denen dieser gemeinsame Faktor gestrichen wurde, höchstens durch die Reihenfolge der Faktoren, und damit gilt dasselbe für die beiden Darstellungen von M.

■

Als erste Anwendung dieses Satzes wollen wir zeigen

Satz: Die reelle Zahl x erfülle die Gleichung

$$x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0 = 0 \quad \text{mit} \quad a_i \in \mathbb{Z}\,.$$

Dann ist x entweder ganzzahlig oder irrational.

Beweis: Jede rationale Zahl x kann als Quotient $x = p/q$ zweier zueinander teilerfremder ganzer Zahlen p und q geschrieben werden. Multiplizieren wir die Gleichung

$$\left(\frac{p}{q}\right)^d + a_{d-1}\left(\frac{p}{q}\right)^{d-1} + \cdots + a_1\left(\frac{p}{q}\right) + a_0 = 0$$

mit q^d, erhalten wir die nennerlose Gleichung

$$p^d + a_{d-1}p^{d-1}q + \cdots + a_1 pq^{d-1} + a_0 q^d = 0\,.$$

Auflösen nach p^d führt auf

$$\begin{aligned} p^d &= -a_{d-1}p^{d-1}q - \cdots - a_1 pq^{d-1} - a_0 q^d \\ &= q\cdot(-a_{n-1}p^{n-1} - \cdots - a_1 pq^{n-2} - a_0 q^{n-1})\,, \end{aligned}$$

d.h. q muß ein Teiler von p^d sein, was wegen der Eindeutigkeit der Primfaktorzerlegung von p^d sowie der vorausgesetzten Teilerfremdheit von p und q nur für $q = \pm 1$ der Fall sein kann. Somit ist x eine ganze Zahl, wie behauptet. ∎

Insbesondere ist also eine n-te Wurzel einer natürlichen Zahl entweder ganzzahlig oder irrational.

Als Übungsaufgabe kann man mit der Beweismethode des obigen Satzes zusammen mit unserem früheren Ergebnis, wonach ganzzahlige Nullstellen ganzzahliger Polynome den konstanten Koeffizienten teilen müssen, leicht ein Verfahren zur Bestimmung rationaler Lösungen durch Probieren herleiten: Ist eine rationale Lösung der Gleichung

$$a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0 = 0 \quad \text{mit} \quad a_i \in \mathbb{Z}\,,$$

so ist bei einer Darstellung von x als gekürztem Bruch der Nenner ein Teiler von a_d und der Zähler teilt a_0.

§3: Die Verteilung der Primzahlen

Nachdem wir wissen, daß jede natürliche Zahl als Produkt von Primzahlpotenzen darstellbar ist, stellt sich als nächstes die Frage, wie viele Primzahlen es gibt. Die Antwort finden wir schon in EUKLIDs Elementen; der dort gegebene Beweis dürfte immer noch der einfachste sein: Es gibt unendlich viele Primzahlen, denn gäbe es nur endlich viele Primzahlen $p_1, \ldots, p_n$, so könnten wir deren Produkt P bilden und die Primzerlegung von $P+1$ betrachten. Da P durch alle p_i teilbar ist, kann $P+1$ durch kein p_i teilbar sein. Andererseits ist natürlich auch $P+1$ als Produkt von Primzahlpotenzen darstellbar; also muß es außer $p_1, \ldots, p_n$ noch weitere Primzahlen geben, im Widerspruch zur Annahme.

Um nicht ganz auf dem Stand von vor rund zweieinhalb Jahrtausenden stehen zu bleiben, wollen wir uns noch einen zweiten, auf EULER zurückgehenden Beweis ansehen.

Dazu betrachten wir für eine reelle Zahl $s > 1$ die unendliche Reihe

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

Als erstes müssen wir uns überlegen, daß diese Reihe konvergiert. Da alle Summanden positiv sind, müssen wir dafür nur zeigen, daß es eine gemeinsame obere Schranke für alle Teilsummen gibt. Da die Funktion $x \mapsto 1/x^s$ für $x > 0$ monoton fallend ist, haben wir für $n-1 \le x \le n$ die Abschätzung $1/n^s \le 1/x^s$, d.h.

$$\sum_{n=1}^{N} \frac{1}{n^s} = 1 + \sum_{n=2}^{N} \frac{1}{n^s} \le 1 + \int_1^N \frac{dx}{x^s}$$

$$< 1 + \int_1^{\infty} \frac{dx}{x^s} = 1 + \frac{1}{s-1} = \frac{s}{s-1} \, .$$

Somit ist $\zeta(s)$ für alle $s > 1$ wohldefiniert.

Einen Zusammenhang mit Primzahlen liefert der folgende

Satz: *a)* Für $s > 1$ ist $\zeta(s) = \prod_{p \text{ prim}} \frac{1}{1 - \frac{1}{p^s}}$.

b) Für alle $N \in \mathbb{N}$ und alle reellen $s > 0$ ist $\sum_{n=1}^{N} \frac{1}{n^s} \leq \prod_{\substack{p \leq N \\ p \text{ prim}}} \frac{1}{1 - \frac{1}{p^s}}$.

Beweis: Wir beginnen mit *b).* Für $N = 1$ steht hier die triviale Formel $1 \leq 1$; sei also $N \geq 2$, und seien $p_1, \ldots, p_r$ die sämtlichen Primzahlen kleiner oder gleich N. Nach der Summenformel für die geometrische Reihe ist

$$\frac{1}{1 - \frac{1}{p_k^s}} = \sum_{\ell=0}^{\infty} \frac{1}{p_k^{\ell s}} ,$$

und das Produkt der rechtsstehenden Reihen über $k = 1$ bis r ist wegen der Eindeutigkeit der Primzerlegung die Summe über alle jene $1/n^s$, für die n keinen Primteiler größer N hat. Darunter sind insbesondere alle $n \leq N$, womit *b)* bewiesen wäre.

Die Differenz zwischen $\zeta(s)$ und dem Produkt auf der rechten Seite von *b)* ist gleich der Summe über alle $1/n^s$, für die n mindestens einen Primteiler größer N haben. Diese Summe ist natürlich höchstens gleich der Summe aller $1/n^s$ mit $n > N$, und die geht wegen der Konvergenz von $\zeta(s)$ gegen null für $N \to \infty$. Damit ist auch *a)* bewiesen. ∎

Auch daraus folgt, daß es unendlich viele Primzahlen gibt: Gäbe es nämlich nur endlich viele, so stünde auf der rechten Seite von *b)* für jedes hinreichend große N das Produkt über die *sämtlichen* Primzahlen. Da es nur endlich viele Faktoren hat, wäre es auch für $s = 1$ endlich, und damit müßte

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

kleiner oder gleich dieser Zahl sein, im Widerspruch zur Divergenz der harmonischen Reihe.

Verglichen mit dem Beweis aus EUKLIDs Elementen ist EULERs Methode erheblich komplizierter. Um trotzdem ihre Existenzberechtigung

zu haben, sollte sie uns daher auch mehr Informationen liefern. In welchem Maße sie dies tatsächlich leistet, geht wahrscheinlich sogar noch deutlich über alles hinaus, was EULER seinerzeit träumen konnte.

Zunächst einmal können wir Teil *b)* für $s = 1$ zu einer quantitativen Abschätzung bezüglich der Anzahl $\pi(N)$ der Primzahlen kleiner oder gleich N umformulieren: Wie oben im Konvergenzbeweis für $\zeta(s)$ können wir aus der Monotonie der Funktion $x \mapsto 1/x$ folgern, daß für alle $N \in \mathbb{N}$ gilt

$$\log(N+1) = \int_1^{N+1} \frac{dx}{x} < \sum_{n=1}^{N} \frac{1}{n} < 1 + \int_1^{N} \frac{dx}{x} = 1 + \log N \, .$$

Zur Abschätzung der linken Seite beachten wir einfach, daß der Faktoren $1/(1-1/p)$ für $p = 2$ gleich zwei ist, ansonsten aber kleiner. Somit ist

$$\log(N+1) < \sum_{n=1}^{N} \frac{1}{n} \leq \prod_{\substack{p \leq N \\ p \text{ prim}}} \frac{1}{1-\frac{1}{p}} \leq 2^{\pi(N)}$$

und damit

$$\pi(N) \geq \frac{\log\log(N+1)}{\log 2} \, .$$

Wie wir bald sehen werden, ist das allerdings eine sehr schwache Abschätzung.

EULERs Methode erlaubt uns auch, die Dichte der Primzahlen zu vergleichen mit der Dichte beispielsweise der Quadratzahlen: Wie wir oben gesehen haben, konvergiert $\zeta(s)$ für alle $s > 1$, insbesondere also konvergiert die Summe $\zeta(2)$ der inversen Quadratzahlen. EULER konnte mit seiner Methode zeigen, daß die Summe der inversen Primzahlen *divergiert,* so daß die Primzahlen zumindest in diesem Sinne dichter liegen als die Quadratzahlen und alle anderen Potenzen mit (reellem) Exponenten $s > 1$.

Zum Beweis fehlt uns nur noch eine Analysis I Übungsaufgabe: Wir wollen uns überlegen, daß für alle $0 \leq x \leq \frac{1}{2}$ gilt $(1-x) \geq 4^{-x}$. An den Intervallenden stimmen beide Funktionen überein, und $1-x$ ist eine lineare Funktion. Es reicht daher, wenn wir zeigen, daß 4^{-x} eine

konvexe Funktion ist, daß also ihre zweite Ableitung überall im Intervall positiv ist. Das ist aber klar, denn die ist einfach $\log(4)^2 \cdot 4^{-x}$. Für jede Primzahl p ist daher

$$1 - \frac{1}{p} \geq 4^{-1/p} \quad \text{und} \quad \frac{1}{1 - \frac{1}{p}} \leq 4^{1/p} \, .$$

Zusammen mit der vorigen Abschätzung folgt

$$\log(N+1) < \prod_{\substack{p \leq N \\ p \text{ prim}}} \frac{1}{1 - \frac{1}{p}} \leq \prod_{\substack{p \leq N \\ p \text{ prim}}} 4^{1/p} = 4^{\sum \frac{1}{p}} \, ,$$

wobei die Summe im Exponenten über alle Primzahlen $p \leq N$ geht. Da $\log(N+1)$ für $N \to \infty$ gegen unendlich geht, muß somit auch die Summe der inversen Primzahlen divergieren.

Mit diesen Bemerkungen fängt allerdings die Nützlichkeit der Funktion $\zeta(s)$ für das Verständnis der Funktion $\pi(N)$ gerade erst an: Ein Jahrhundert nach EULER erkannte RIEMANN, daß die Funktion $\zeta(s)$ ihre wahre Nützlichkeit für das Studium von $\pi(N)$ erst zeigt, wenn man sie auch für komplexe Argumente s betrachtet. Jeder, der sich ein bißchen mit Funktionen einer komplexer Veränderlichen auskennt, kann leicht zeigen, daß $\zeta(s)$ auch für komplexe Zahlen mit Realteil größer ein konvergiert: Der Imaginärteil des Exponenten führt schließlich nur zu einem Faktor vom Betrag eins.

GEORG FRIEDRICH BERNHARD RIEMANN (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.

RIEMANNs wesentliche Erkenntnis war, daß sich $\zeta(s)$ fortsetzen läßt zu einer analytischen Funktion auf der gesamten Menge der komplexen Zahlen mit Ausnahme der Eins (wo die ζ-Funktion wegen der Divergenz der harmonischen Reihe keinen endlichen Wert haben kann).

Für Leser, die nicht mit dem Konzept der analytischen Fortsetzung vertraut sind, möchte ich ausdrücklich darauf hinweisen, daß dies selbstverständlich nicht bedeutet, daß die definierende Summe der ζ-Funktion für reelle Zahlen kleiner eins oder komplexe Zahlen mit Realteil kleiner oder gleich eins konvergiert: Analytische Fortsetzung besteht darin, daß eine differenzierbare Funktion (die im Komplexen automatisch beliebig oft differenzierbar ist und um jeden Punkt in eine TAYLOR-Reihe entwickelt werden kann) via TAYLOR-Reihen über ihren eigentlichen Definitionsbereich hinweg ausgedehnt wird. Man kann beispielsweise zeigen, das $\zeta(-1) = -\frac{1}{12}$ ist. Setzt man $s = -1$ in die für $s > 1$ gültige Reihe ein, erhält man die Summe aller natürlicher Zahlen, die selbstverständlich nicht gleich $-\frac{1}{12}$ ist, sondern divergiert. Entsprechend hat $\zeta(s)$ Nullstellen bei allen geraden negativen Zahlen, obwohl auch hier die entsprechenden Reihen divergieren. Diese Nullstellen bezeichnet man als die sogenannten *trivialen* Nullstellen der ζ-Funktion, da sie sich sofort aus einer bei der Konstruktion der analytischen Fortsetzung zu beweisenden Funktionalgleichung ablesen lassen. Für die Primzahlverteilung spielen vor allem die übrigen, die sogenannten nichttrivialen Nullstellen, eine große Rolle.

Wie wir gerade gesehen haben, liegen die Primzahlen zumindest in einem gewissen Sinne dichter als die Quadratzahlen. Zur Einstimmung auf das Problem der Primzahlverteilung wollen wir uns kurz mit der (deutlich einfacheren) Verteilung der Quadratzahlen beschäftigen.

Die Folge der Abstände zwischen zwei aufeinanderfolgenden Quadratzahlen ist einfach die Folge der ungeraden Zahlen, denn

$$(n+1)^2 - n^2 = 2n + 1 \,.$$

Zwei aufeinanderfolgende Quadratzahlen $Q < Q'$ haben daher die Differenz $Q' - Q = 2\sqrt{Q} + 1$.

Bei den Primzahlen ist die Situation leider sehr viel unübersichtlicher: EULER meinte sogar, die Verteilung der Primzahlen sei ein Geheimnis,

das der menschliche Verstand nie erfassen werde. 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 drei und fünf, aber auch der zwischen den Primzahlen $10^{100} + 35737$ und $10^{100} + 35739$. Das größte derzeit bekannte Beispiel bilden die im September 2016 im Rahmen von PrimeGrid (www.primegrid.com) gefundenen beiden Zahlen

$$2\,996\,863\,034\,895 \cdot 2^{1290000} \pm 1 \,.$$

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 auch nur beweisen, daß es unendlich viele gibt.

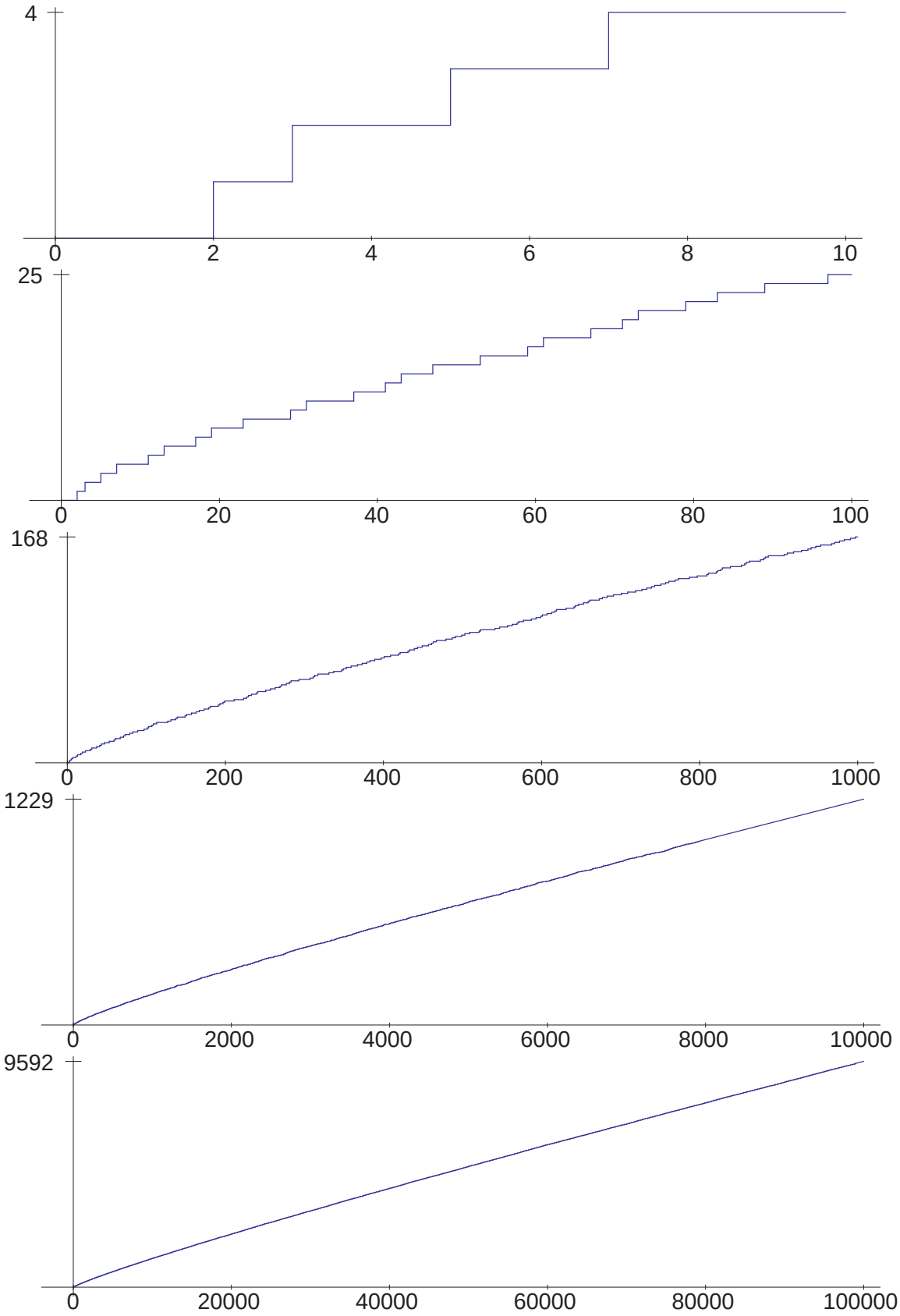
Eine obere Grenze für den Abstand zwischen zwei aufeinanderfolgenden Primzahlen gibt es genauso wenig wie bei den Quadratzahlen: Ist $n \geq 2$ und $2 \leq i \leq n$, so ist die Zahl $n! + i$ durch i teilbar und somit keine Primzahl. Der Abstand zwischen der größten Primzahl kleiner oder gleich $n! +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

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

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

Die Abbildungen auf der vorigen 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 differenzierbaren Funktion; daher auch die Schreibweise $\pi(x)$ statt – wie bisher – $\pi(N)$.

Auf den ersten Blick sieht diese Funktion fast linear aus.; sieht man sich allerdings die Zahlenwerte genauer an, so sieht man schnell, daß


4
0
2
4
6
8
10
25
0
20
40
60
80
100
168
0
200
400
600
800
1000
1229
0
2000
4000
6000
8000
10000
9592
0
20000
40000
60000
80000
100000

$\pi(x)$ 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

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

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

$$\pi(x) = \sum_{p \le x} \frac{\log p}{\log p} \ge \sum_{p \le x} \frac{\log p}{\log x} = \frac{\vartheta(x)}{\log x} \, ,$$

andererseits ist

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

und damit auch $\pi(x) < \dfrac{2\vartheta(x)}{\log x} + \pi\left(\sqrt{x}\right) < \dfrac{2\vartheta(x)}{\log x} + \sqrt{x}$. 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

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

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

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

nach der Summenformel für die geometrische Reihe:

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

Zur weiteren Abschätzung ersetzen wir die Summe über alle Primzahlen kleiner oder gleich n durch die Summe über alle Zahlen bis n und beachten, daß für reellen $x \geq 2$ gilt $\log x < \sqrt{x}$, also

$$\frac{\log x}{x(x-1)} < \frac{\sqrt{x}}{x^2} = \frac{1}{x^{3/2}} \quad \text{und damit folgt}$$

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

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 $\sum_{k\geq 2} \frac{1}{p^k} = O(n)$. Setzen wir dies in die Formel für $\log n!$ ein, erhalten wir nach allen bislang bewiesenen Abschätzungen, daß

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

Dies können wir vergleichen mit der STIRLINGschen Formel

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

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

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

Damit ist

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

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

$$\vartheta(x) = \sum_{i=0}^{\infty} \left(\vartheta\Big(\frac{x}{2^i}\Big) - \vartheta\Big(\frac{x}{2^{i+1}}\Big)\right) = O\left(\sum_{i=0}^{\infty} \frac{x}{2^i}\right) = O(x)\,,$$

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 nach $(*)$

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

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

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

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 Figures 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 ČEBYŠĒV (1821–1894), in der Numerik meist bekannt in der Schreibweise TSCHEBYTSCHEFF, zeigte 1851: *Falls* der Grenzwert

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

existiert, muß er den Wert eins haben.

1852 bewies er dann ein deutlich schärferes Resultat: 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 asymptotisch 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 $\lim\limits_{x\to\infty} \dfrac{x/\log x}{\mathrm{Li}(x)} = 1$ ist. Nach dem Primzahlsatz ist daher auch für jedes $a \in \mathbb{R}$

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

Wie DE LA VALLÉE POUSSIN zeigte, liefert der Wert $a = 1$ unter allen reellen Zahlen a die beste Approximation an $\pi(x)$, aber $\mathrm{Li}(x)$ 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	$\pi(n)$	$\frac{n}{\log n}$	$\frac{n}{\log n-1}$	$\frac{n}{\log n-1{,}08366}$	$\mathrm{Li}(n)$
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 $\mathrm{Li}(x) - \pi(x)$ wissen. Hier kommen wir in das Reich der offenen Fragen, und nach derzeitigem Verständnis hängt alles ab von der oben erwähnten RIEMANNschen Zetafunktion. Nach einer berühmten Vermutung von RIEMANN haben alle nichttrivialen Nullstellen von $\zeta(s)$ den Realteil ein halb. Falls dies stimmt, ist

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

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. 2000 einen Preis von jeweils einer Million Dollar ausgesetzt hat; für Einzelheiten siehe `http://www.claymath.org/millennium/` .

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

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

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

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

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

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

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

■

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

$$\log n! = \int_1^n \log x\,dx + \frac{\log n}{2} + \int_1^n \frac{\{x\}-\frac{1}{2}}{x}\,dx$$

$$= x(\log x - 1)\Big|_1^n + \frac{\log n}{2} + \int_1^n \frac{\{x\}-\frac{1}{2}}{x}\,dx$$

$$= n(\log n - 1) + 1 + \frac{\log n}{2} + \int_1^n \frac{\{x\}-\frac{1}{2}}{x}\,dx\,.$$

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

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

$$= \int_0^{\frac{1}{2}} \left(\frac{x}{k+\frac{1}{2}+x} - \frac{x}{k+\frac{1}{2}-x}\right) dx = \int_0^{\frac{1}{2}} \frac{-2x^2}{\left(k+\frac{1}{2}\right)^2 - x^2}\,dx\,.$$

Im Intervall von 0 bis $\frac{1}{2}$ ist der Integrand monoton fallend, d.h.

$$0 \geq \frac{-2x^2}{\left(k+\frac{1}{2}\right)^2 - x^2} \geq \frac{-\frac{1}{2}}{\left(k+\frac{1}{2}\right)^2 - \frac{1}{4}} = \frac{-2}{(2k+1)^2 - 1} \geq -\frac{1}{2k^2}\,,$$

und damit ist

$$0 \geq \int_k^{k+1} \frac{\{x\}-\frac{1}{2}}{x}\,dx = \int_0^{\frac{1}{2}} \frac{-2x^2}{\left(k+\frac{1}{2}\right)^2 - x^2}\,dx \geq -\frac{1}{4k^2}\,,$$

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

$$0 \geq \int_1^n \frac{\{x\}-\frac{1}{2}}{x}\,dx \geq -\sum_{k=1}^{n-1} \frac{1}{4k^2} > -\frac{1}{4}\sum_{k=1}^{\infty} \frac{1}{k^2}$$

für das störende Integral aus der obigen Formel. Da die Summe rechts konvergiert, konvergiert auch das Integral für $n \to \infty$ gegen einen Grenzwert I. Somit ist

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

also folgt insbesondere die Abschätzung

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

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

§4: Das Sieb des Eratosthenes

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

Um alle Primzahlen kleiner oder gleich einer Zahl N zu finden, schreibe man zunächst die Zahlen von eins bis N in eine Reihe.

Eins ist nach Definition keine Primzahl – für klassische griechische Mathematiker wie EUKLID war die Eins schließlich nicht einmal eine Zahl. Also streichen wir die Eins durch. Die Zwei ist prim, aber ihre echten Vielfachen sind natürlich keine Primzahlen, werden also durchgestrichen. 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 ist sie natürlich auch durch eine echt kleinere Primzahl teilbar.

ERATOSTHENES (Ερατοσθένες) wurde 276 v.Chr. in Cyrene im heutigen Libyen 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.

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; 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.

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 $p_1, \ldots, p_r$ 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 im 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 übrig bleibt, 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.

§5: Kongruenzenrechnung

Zwei ganze Zahlen lassen sich im allgemeinen nicht durcheinander dividieren. Trotzdem – oder gerade deshalb – spielen Teilbarkeitsfragen in der Zahlentheorie eine große Rolle. Das technische Werkzeug zu ihrer Behandlung ist die Kongruenzenrechnung.

Definition: Wir sagen, zwei ganze Zahlen $x, y \in \mathbb{Z}$ seien kongruent modulo m für eine natürliche Zahl m, in Zeichen $x \equiv y \bmod m$, wenn $x - y$ durch m teilbar ist.

Die Kongruenz modulo m definiert offensichtlich eine Äquivalenzrelation auf $\mathbb{Z}$: Jede ganze Zahl ist kongruent zu sich selbst, denn $x - x = 0$

ist durch jede natürliche Zahl teilbar. Wenn $x - y$ durch m teilbar ist, so auch $y - x = -(x - y)$, und ist schließlich $x \equiv y \bmod m$ und $y \equiv z \bmod m$, so sind $x - y$ und $y - z$ durch m teilbar, also auch ihre Summe $x - z$, und damit ist auch $x \equiv z \bmod m$.

Zwei Zahlen $x, y \in \mathbb{Z}$ liegen genau dann in derselben Äquivalenzklasse, wenn sie bei der Division durch m denselben Divisionsrest haben; es gibt somit m Äquivalenzklassen, die den m möglichen Divisionsresten $0, 1, \ldots, m-1$ entsprechen.

Lemma: Ist $x \equiv x' \bmod m$ und $y \equiv y' \bmod m$, so ist auch

$$x \pm y \equiv x' \pm y' \bmod m \quad \text{und} \quad x'y' \equiv xy \bmod m\,.$$

Beweis: Sind $x - x'$ und $y - y'$ durch m teilbar, so auch

$$(x \pm y) - (x' \pm y') = (x - x') \pm (y - y') \qquad \text{und}$$
$$xy - x'y' = x(y - y') + y'(x - x')$$

■

Im folgenden wollen wir das Symbol „mod“ nicht nur in Kongruenzen wie $x \equiv y \bmod m$ benutzen, sondern auch – wie in einigen Programmiersprachen üblich – als Rechenoperation:

Definition: Für eine ganze Zahl x und eine natürliche Zahl m bezeichnet $x \bmod m$ jene ganze Zahl $0 \le r < m$ mit $x \equiv r \bmod m$.

$x \bmod m$ ist also einfach der Divisionsrest bei der Division von x durch m.

Beim Rechnen modulo einer Zahl m ersetzt man alle Rechenergebnisse durch ihren Wert modulo m; sie liegen also stets zwischen null und $m - 1$. Anwendungen findet dies beispielsweise in der Computeralgebra: Da man auch für Polynome in einer Veränderlichen über einem Körper eine Division mit Rest hat, kann man auch hier größte gemeinsame Teiler mit dem EUKLIDischen Algorithmus berechnen. Wenn die führenden Koeffizienten nicht eins sind, bekommt man dabei selbst bei Polynomen mit ganzzahligen Koeffizienten oft sehr schnell gigantische Nenner. Wenn man allerdings weiß, daß der ggT ein Polynom mit ganzzahligen Koeffizienten ist und auch eine Schranke M für den Betrag

der Koeffizienten kennt, genügt es, wenn man den ggT modulo einer Zahl $m \geq 2M + 1$ berechnen kann. Tatsächlich genügt es sogar, wenn man ihn modulo hinreichend vieler kleinerer Zahlen kennt, denn der folgende Satz zeigt uns, wie man diese Ergebnisse kombinieren kann zu einer Kongruenz modulo einer größeren Zahl.

Der Legende nach zählten chinesische Generäle ihre Truppen, indem sie diese mehrfach antreten ließen in Reihen verschiedener Breiten $m_1, \ldots, m_r$ und jedesmal nur die Anzahl a_r der Soldaten in der letzten Reihe zählten. Aus den r Relationen

$$x \equiv a_1 \bmod m_1, \quad \ldots, \quad x \equiv a_r \bmod m_r$$

bestimmten sie dann die Gesamtzahl x der Soldaten.

Ob es im alten China wirklich Generäle gab, die soviel Mathematik konnten, sei dahingestellt. Beispiele zu diesem Satz finden sich jedenfalls bereits 1247 in den chinesischen *Mathematischen Abhandlungen in neun Bänden* von CH'IN CHIU-SHAO (1202–1261), allerdings geht es dort nicht um Soldaten, sondern um Reis.

CH'IN CHIU-SHAO oder QIN JIUSHAO wurde 1202 in der Provinz Sichuan geboren. Auf eine wilde Jugend mit vielen Affairen folgte ein wildes und alles andere als gesetzestreues Berufsleben in Armee, öffentlicher Verwaltung und illegalem Salzhandel. Als Jugendlicher studierte er an der Akademie von Hang-chou Astronomie, später brachte ihm ein unbekannter Lehrer Mathematik bei. Insbesondere studierte er die in vorchristlicher Zeit entstandenen *Neun Bücher der Rechenkunst,* nach deren Vorbild er 1247 seine deutlich anspruchsvolleren *Mathematischen Abhandlungen in neun Bänden* publizierte, die ihn als einen der bedeutendsten Mathematiker nicht nur Chinas der damaligen Zeit ausweisen. Zum chinesischen Restesatz schreibt er, daß er ihn von den Kalendermachern gelernt habe, diese ihn jedoch nur rein mechanisch anwendeten ohne ihn zu verstehen. CH'IN CHIU-SHAO starb 1261 in Meixian, wohin er nach einer seiner vielen Entlassungen aus einer Haft wegen krimineller Machenschaften geschickt worden war.

Wir wollen uns zunächst überlegen, unter welchen Bedingungen ein solches Verfahren überhaupt funktionieren kann. Offensichtlich können die obigen r Relationen eine natürliche Zahl nicht eindeutig festlegen, denn ist x eine Lösung und M irgendein gemeinsames Vielfaches der sämtlichen m_i, so ist auch $x+M$ eine Lösung – M ist schließlich modulo aller m_i kongruent zur Null.

Außerdem gibt es Relationen obiger Form, die unlösbar sind, beispielsweise das System

$$x \equiv 2 \bmod 4 \quad \text{und} \quad x \equiv 3 \bmod 6\,,$$

dessen erste Gleichung nur gerade Lösungen hat, während die zweite nur ungerade hat. Das Problem hier besteht darin, daß zwei ein gemeinsamer Teiler von vier und sechs ist, so daß jede der beiden Kongruenzen auch etwas über $x \bmod 2$ aussagt: Nach der ersten ist x gerade, nach der zweiten aber ungerade.

Dieses Problem können wir dadurch umgehen, daß wir nur Moduln m_i zulassen, die paarweise teilerfremd sind. Dies hat auch den Vorteil, daß jedes gemeinsame Vielfache der m_i Vielfaches des Produkts aller m_i sein muß, so daß wir x modulo einer vergleichsweise großen Zahl kennen.

Chinesischer Restesatz: Das System von Kongruenzen

$$x \equiv a_1 \bmod m_1, \quad \ldots, \quad x \equiv a_r \bmod m_r$$

hat für paarweise teilerfremde Moduln m_i genau eine Lösung x mit $0 \le x < m_1 \cdots m_r$. Jede andere Lösung $y \in \mathbb{Z}$ läßt sich in der Form $x + km_1 \cdots m_r$ schreiben mit $k \in \mathbb{Z}$.

Beweis: Wir beginnen mit dem Fall zweier Kongruenzen

$$x \equiv a \bmod m \quad \text{und} \quad y \equiv b \bmod n$$

mit zueinander teilerfremden Zahlen m und n. Ihr ggT eins läßt sich nach dem erweiterten EUKLIDischen Algorithmus als $1 = \alpha m + \beta n$ schreiben. Somit ist

$$1 - \alpha m = \beta n \equiv \begin{cases} 1 & \bmod m \\ 0 & \bmod n \end{cases} \quad \text{und} \quad 1 - \beta n = \alpha m \equiv \begin{cases} 0 & \bmod m \\ 1 & \bmod n \end{cases},$$

also löst

$$x = \beta n a + \alpha m b \equiv \begin{cases} a & \bmod m \\ b & \bmod n \end{cases}$$

das Problem. Für jede weitere Lösung y ist $x - y \equiv 0$ sowohl modulo n als auch modulo m, also ist $x - y$ durch nm teilbar und hat somit die

Form $y = x + kmn$ mit einem $k \in \mathbb{Z}$. Umgekehrt löst natürlich auch jedes solche y die Kongruenz; die allgemeine Lösung ist daher

$$x = \beta na + \alpha mb + kmn$$

mit einem beliebigen $k \in \mathbb{Z}$.

Der allgemeine Satz folgt nun leicht durch vollständige Induktion nach r: Für $r = 1$ ist die Aussage trivial; sei also $r \geq 2$. Nach Induktionsannahme gibt es ein $y \in \mathbb{Z}$ mit $0 \leq y < m_1 \cdots m_{r-1}$, so daß

$$y \equiv a_1 \bmod m_1, \quad \ldots, \quad y \equiv a_{r-1} \bmod m_{r-1}\,,$$

und die sämtlichen Lösungen sind genau die Zahlen $y + km_1 \cdots m_{r-1}$ mit $k \in \mathbb{Z}$. Sei $m = m_1 \cdots m_{r-1}$; nach dem bereits bewiesenen Fall von nur zwei Kongruenzen gibt es ein $x \in \mathbb{Z}$ mit $0 \leq x < mm_r$. so daß

$$x \equiv y \bmod m \quad \text{und} \quad x \equiv a_r \bmod m_r\,,$$

und x ist modulo $mm_r = m_1 \cdots m_r$ eindeutig. Damit ist der Satz vollständig bewiesen.

■

Als Beispiel betrachten wir die beiden Kongruenzen

$$x \equiv 1 \bmod 17 \quad \text{und} \quad x \equiv 5 \bmod 19\,.$$

Zunächst wenden wir den erweiterten EUKLIDischen Algorithmus an auf die beiden Moduln 17 und 19:

$$19 : 17 = 1 \text{ Rest } 2 \Longrightarrow 2 = 19 - 17$$

$$17 : 2 = 8 \text{ Rest } 1 \Longrightarrow 1 = 17 - 8 \cdot 2 = 9 \cdot 17 - 8 \cdot 19$$

Also ist $9 \cdot 17 = 153 \equiv 0 \bmod 17$ und $\equiv 1 \bmod 19$; außerdem ist $-8 \cdot 19 = -152$ durch 19 teilbar und $\equiv 1 \bmod 17$. Die Zahl

$$x = -152 \cdot 1 + 153 \cdot 5 = 613$$

löst somit das Problem. Da 613 größer ist als $17 \cdot 19 = 323$, ist allerdings nicht 613 die kleinste positive Lösung, sondern $613 - 323 = 290$.

Zur Lösung des Systems

$$x \equiv 5 \bmod 10, \quad x \equiv 9 \bmod 11, \quad x \equiv 6 \bmod 13$$

lösen wir zunächst nur das System

$$x \equiv 5 \bmod 10 \quad \text{und} \quad x \equiv 9 \bmod 11 \,.$$

Da $1 = 11 - 10$, ist $11 \equiv 0 \bmod 11$ und $11 \equiv 1 \bmod 10$; entsprechend ist $-10 \equiv 0 \bmod 10$ und $-10 \equiv 1 \bmod 11$. Also ist

$$x = 5 \cdot 11 - 9 \cdot 10 = -35$$

eine Lösung; die allgemeine Lösung ist $-35 + 110k$ mit $k \in \mathbb{Z}$. Die kleinste positive Lösung ist $-35 + 110 = 75$.

Unser Ausgangssystem ist somit äquivalent zu den beiden Kongruenzen

$$x \equiv 75 \bmod 110 \quad \text{und} \quad x \equiv 6 \bmod 13 \,.$$

Um es zu lösen, müssen wir zunächst die Eins als Linearkombination von 110 und 13 darstellen. Hier bietet sich keine offensichtliche Lösung an, also verwenden wir den erweiterten EUKLIDischen Algorithmus:

$$110 : 13 = 8 \text{ Rest } 6 \Longrightarrow 6 = 110 - 8 \cdot 13$$

$$13 : 6 = 2 \text{ Rest } 1 \Longrightarrow 1 = 13 - 2 \cdot (110 - 8 \cdot 12) = 17 \cdot 13 - 2 \cdot 110$$

Also ist $17 \cdot 13 = 221 \equiv 1 \bmod 110$ und $\equiv 0 \bmod 13$; genauso ist $-2 \cdot 110 = 220 \equiv 1 \bmod 13$ und $\equiv 9 \bmod 110$. Eine ganzzahlige Lösung unseres Problems ist somit

$$75 \cdot 221 - 6 \cdot 220 = 15\,255 \,.$$

Die allgemeine Lösung ist

$$15\,255 + k \cdot 110 \cdot 13 = 15\,255 + 1\,430k \quad \text{mit} \quad k \in \mathbb{Z} \,.$$

Da $15\,255 : 1\,430 = 10$ Rest 955 ist, erhalten wir 955 als kleinste Lösung.

Alternativ läßt sich die Lösung eines Systems aus r Kongruenzen auch in einer geschlossenen Form darstellen, allerdings um den Preis einer n-maligen statt $(n-1)$-maligen Anwendung des EUKLIDischen Algorithmus und größeren Zahlen schon von Beginn an: Um das System

$$x \equiv a_i \bmod m_i \quad \text{für} \quad i = 1, \ldots, r$$

zu lösen, berechnen wir zunächst für jedes i das Produkt

$$\widehat{m}_i = \prod_{j \neq i} m_j$$

der sämtlichen übrigen m_j und bestimmen dazu ganze Zahlen α_i, β_i, für die gilt $\alpha_i m_i + \beta_i \widehat{m}_i = 1$. Dann ist

$$x = \sum_{j=1}^{n} \beta_j \widehat{m}_j a_j \equiv \beta_i \widehat{m}_i a_i = (1 - \alpha_i m_i) a_i \equiv a_i \bmod m_i \, .$$

Natürlich wird x hier – wie auch bei der obigen Formel – oft größer sein als das Produkt der m_i; um die kleinste Lösung zu finden, müssen wir also noch modulo diesem Produkt reduzieren.

Im obigen Beispiel wäre

$$\begin{array}{llll} m_1 = 10 & \widehat{m}_1 = 11 \cdot 13 = 143 & 1 = 43 \cdot 10 - 3 \cdot 143 \\ m_2 = 11 & \widehat{m}_2 = 10 \cdot 13 = 130 & 1 = -59 \cdot 11 + 5 \cdot 130 \\ m_3 = 13 & \widehat{m}_3 = 10 \cdot 11 = 110 & 1 = 17 \cdot 13 - 2 \cdot 110 \, , \end{array}$$

also

$$x = -3 \cdot 143 \cdot 5 + 5 \cdot 130 \cdot 9 - 2 \cdot 110 \cdot 6 = -2145 + 5850 - 1320 = 2385 \, .$$

Modulo $10 \cdot 11 \cdot 13$ erhalten wir natürlich auch hier wieder 955.

§6: Der kleine Satz von Fermat

Die meisten Mathematiker kennen FERMAT vor allem wegen seiner 1994 von ANDREW WILES bewiesenen Vermutung über Summen n-ter Potenzen; im englischen Sprachraum wurde sie auch schon lange vor diesem Beweis als FERMAT's *big theorem* bezeichnet. In Analogie zu diesem „großen" Satz von FERMAT gibt es auch einen kleinen; diesen hat er wirklich bewiesen.

Kleiner Satz von Fermat: Für jedes $a \in \mathbb{Z}$ und jede Primzahl p ist

$$a^p \equiv a \bmod p \, ;$$

ist a nicht durch p teilbar, gilt auch $a^{p-1} \equiv 1 \bmod p$.

Beweis: Wir betrachten zunächst nur nichtnegative Werte von a und beweisen die erste Aussage dafür durch vollständige Induktion:

Für $a = 0$ ist $0^p = 0$, also erst recht kongruent Null modulo p; genauso ist für $a = 1$ auch $a^p = 1$.

Für $a > 1$ schreiben wir

$$a^p = \big((a-1)+1\big)^p = \sum_{i=0}^{p} \binom{p}{i} (a-1)^i \quad \text{mit} \quad \binom{p}{i} = \frac{p!}{i!\,(p-i)!}\,.$$

Falls $1 \le i \le p-1$, ist der Nenner von $\binom{p}{i}$ nicht durch p teilbar, wohl aber der Zähler. Somit ist auch $\binom{p}{i}$ durch p teilbar, also kongruent Null modulo p. Damit ist

$$a^p \equiv \binom{p}{0}(a-1)^0 + \binom{p}{p}(a-1)^p = 1 + (a-1) = a \bmod p$$

nach Induktionsannahme.

Dies beweist die erste Aussage für $a \ge 0$. Für $a < 0$ ist im Falle $p = 2$ sowohl $-a \equiv a \bmod 2$ als auch $a^p = (-a)^p$; für ungerades p ist $(-a)^p = -(a^p)$, so daß die Behauptung in beiden Fällen folgt.

Zum Beweis der zweiten Behauptung beachten wir, daß

$$a^p - a = a\,(a^{p-1} - 1)\,,$$

wie wir gerade bewiesen haben, durch p teilbar ist. Falls a nicht durch p teilbar ist, muß also $a^{p-1} - 1$ durch p teilbar sein, und genau das ist die Behauptung.

■

Der französische Mathematiker PIERRE DE FERMAT (1601–1665) wurde in Beaumont-de-Lomagne im Departement Tarn et Garonne geboren. Bekannt ist er heutzutage vor allem für seine 1994 von ANDREW WILES bewiesene Vermutung, wonach die Gleichung $x^n + y^n = z^n$ für $n \ge 3$ keine ganzzahlige Lösung mit $xyz \neq 0$ hat. Dieser „große“ Satzes von FERMAT, von dem FERMAT lediglich in einer Randnotiz behauptete, daß er ihn beweisen könne, erklärt den Namen der obigen Aussage. Obwohl FERMAT sich sein Leben lang sehr mit Mathematik beschäftigte und wesentliche Beiträge zur Zahlentheorie, Wahrscheinlichkeitstheorie und Analysis lieferte, war er hauptberuflich Jurist.

Wenn wir entscheiden wollen, ob eine gegebene Zahl p prim ist, kann die folgende Umkehrung des kleinen Satzes von FERMAT nützlich sein:

Lemma: Sind a, p zwei zueinander teilerfremde natürliche Zahlen, und ist $a^{p-1} \not\equiv 1 \bmod p$, so ist p keine Primzahl. ∎

Beispiel: Ist $F_{20} = 2^{2^{20}} + 1$ eine Primzahl? Falls ja, ist nach dem kleinen Satz von FERMAT insbesondere $3^{F_{20}-1} \equiv 1 \bmod F_{20}$. 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 315653-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 meinten, das sei die teuerste bislang produzierte 1-Bit-Information.)

Die Umkehrung gilt leider nicht: Es gibt unendlich viele Nichtprimzahlen n, die sogenannten CARMICHAEL-Zahlen, für die $a^{n-1} \equiv 1 \bmod n$ ist für jedes zu n teilerfremde a. Trotzdem wird es für große Zahlen zunehmend unwahrscheinlich, daß eine Zahl p für auch nur ein a den obigen Test besteht, ohne Primzahl zu sein. In der Arbeit

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

sind u.a. die folgenden Schranken ε zu finden für die Wahrscheinlichkeit, daß für eine Nichtprimzahl p und ein zufällig gewähltes a die Kongruenz $a^{p-1} \equiv 1 \bmod p$ gilt:

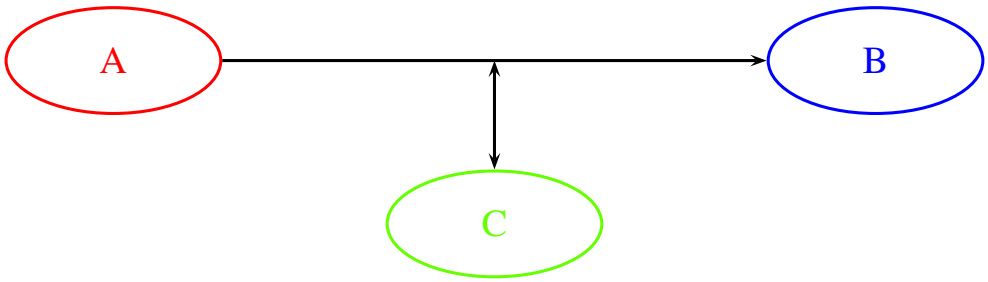
$p \approx 10^{60}$	10^{70}	10^{80}	10^{90}	10^{100}
$\varepsilon \le 7{,}16 \cdot 10^{-2}$	$2{,}87 \cdot 10^{-3}$	$8{,}46 \cdot 10^{-5}$	$1{,}70 \cdot 10^{-6}$	$2{,}77 \cdot 10^{-8}$
$p \approx 10^{120}$	10^{140}	10^{160}	10^{180}	10^{200}
$\varepsilon \le 5{,}28 \cdot 10^{-12}$	$1{,}08 \cdot 10^{-15}$	$1{,}81 \cdot 10^{-19}$	$2{,}76 \cdot 10^{-23}$	$3{,}85 \cdot 10^{-27}$
$p \approx 10^{300}$	10^{400}	10^{500}	10^{600}	10^{700}
$\varepsilon \le 5{,}8 \cdot 10^{-29}$	$5{,}7 \cdot 10^{-42}$	$2{,}3 \cdot 10^{-55}$	$1{,}7 \cdot 10^{-68}$	$1{,}8 \cdot 10^{-82}$
$p \approx 10^{800}$	10^{900}	10^{1000}	10^{2000}	10^{3000}
$\varepsilon \le 5{,}4 \cdot 10^{-96}$	$1{,}0 \cdot 10^{-109}$	$1{,}2 \cdot 10^{-123}$	$8{,}6 \cdot 10^{-262}$	$3{,}8 \cdot 10^{-397}$

Die Arbeit beruht im wesentlichen auf der von POMERANCE betreuten Masterarbeit des ersten Autors und enthält natürlich auch eine Formel für ε; diese ist aber zu grausam zum Abdrucken.

Natürlich kennt die Zahlentheorie auch effiziente Tests, mit denen sich *beweisen* läßt, daß eine gegebene Zahl p prim ist. Diese sind allerdings deutlich aufwendiger als der FERMAT-Test, so daß man sie erst anwendet, wenn die Zahl bereits den FERMAT-Test bestanden hat.

§7: Anwendungen in der Kryptographie

Das Grundproblem der Kryptographie ist das folgende:



A möchte eine Nachricht m an B übermitteln, jedoch besteht die Gefahr, daß alles, was er an B schickt, auf dem Weg dorthin von C gelesen und vielleicht auch verändert wird; außerdem könnte C eventuell versuchen, sich gegenüber B als A auszugeben oder umgekehrt.

Die Kryptographie versucht, dies zu verhindern, indem A anstelle von m eine verschlüsselte Nachricht c schickt, aus der zwar B, nicht aber C die Nachricht m und gegebenenfalls weitere Informationen rekonstruieren kann. Bei sogenannten *Blockchiffren* teilt man die Nachricht auf in Blöcke aus einer endlichen Menge M, im einfachsten Fall die Menge der Buchstaben des Alphabets, und verschlüsselt durch eine bijektive Abbildung von M nach M.

Eine bekannte (und sehr schlechte) Form der Verschlüsselung geht auf CAESAR zurück. Der römische Schriftsteller SUETON schreibt in Kapitel 56 des ersten Buchs DIVUS IULIUS *(der göttliche Julius)* seines Werks DE VITA CAESARUM:

> extant et ad ciceronem, item ad familiares domesticis de rebus, in quibus, si qua occultius perferenda erant, per notas scripsit, id est sic structo litterarum ordine, ut nullum verbum effici posset: quae si qui investigare et persequi velit, quartam elementorum litteram, id est d pro a et perinde reliquas commutet.

> Erhalten sind auch seine Briefe an CICERO, ebenso an seine engeren Freunde über private Angelegenheiten, in denen er, was etwa geheim zu überbringen war, in verschlüsselter Form schrieb, nämlich in einer solchen Anordnung der Buchstaben, daß kein einziges Wort herauskam. Falls hier jemand nachforschen und der Sache nachgehen will, möge er den vierten Buchstaben des Alphabets, d.h. D für A und so fort setzen.

(aus SUETON: Kaiserbiographien*, Akademie Verlag Berlin,* 1993)

CAESAR verschob das Alphabet also einfach zyklisch nach dem Schema

$$A \to D, \quad B \to E, \ldots, W \to Z, \quad X \to A, \quad Y \to B, \quad Z \to C\,.$$

Gerade für jemand, der seine Verbündete so oft wechselte wie CAESAR hat ein solches Verfahren den großen Nachteil, daß jeder, der es einmal benutzt hat, auch künftig alle damit verschlüsselten Nachrichten lesen kann.

Wie AUGUSTE KERCKHOFFS 1883 in seiner grundlegenden Arbeit *La cryptographie militaire* feststellte, muß man bei jedem in größerem Umfang eingesetzten Verfahren davon ausgehen, daß es sich nicht über einen längeren Zeitraum hinweg geheimhalten läßt. Anstelle einer einfachen Verschlüsselungsfunktion f, die jeder Nachricht m einen Chiffretext $c = f(m)$ zuordnet, soll man eine Funktion benutzen, die außer von m auch noch von einem zweiten Parameter s abhängt, dem *Schlüssel.* Somit ist also $c = f(m, s)$. Die Sicherheit des Verfahrens darf laut KERCKHOFFS nur von der Geheimhaltung des (häufig zu wechselnden) Schlüssels s abhängen, nicht von der der Funktion f.

Viele heutige Kryptoverfahren sind oder beruhen auf Blockchiffren. Dazu wird die zu übermittelnde Nachricht aufgespalten in eine Folge von Blöcken einer vorgegebenen Länge. Oft sind das 128 Bit, also 16 Byte; bei den hier betrachteten Verfahren werden die Blöcke allerdings deutlich länger sein.

Bezeichnet $\mathcal{B}$ die Menge aller möglicher Blöcke und $\mathcal{S}$ die Menge aller möglicher Schlüssel, so ist eine Verschlüsselungsfunktion also von der

Form

$$f\colon \begin{cases} \mathcal{B} \times \mathcal{S} \to \mathcal{B} \\ (m, s) \mapsto f(m, s) \end{cases},$$

und die Entschlüsselungsfunktion

$$g\colon \begin{cases} \mathcal{B} \times \mathcal{S} \to \mathcal{B} \\ (m, s) \mapsto g(m, s) \end{cases}$$

ist so definert, daß $g\big(f(m, s), s\big) = m$ ist für alle $m \in \mathcal{B}$.

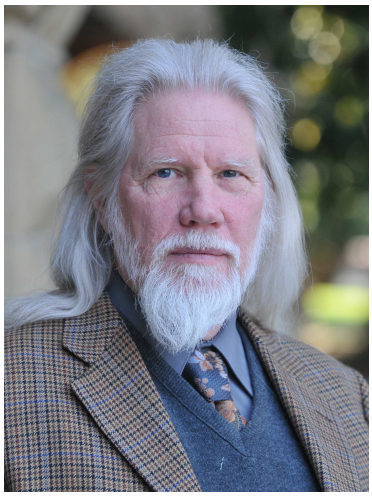
JEAN-GUILLAUME-HUBERT-VICTOR-FRANÇOIS-ALEXANDRE-AUGUSTE KERCKHOFFS VON NIEUWENHOF (1835–1903) wurde in der heute niederländischen Ortschaft Nuth geboren. Er studierte an der Universität Liège, wo er mit dem Doktor der Literaturwissenschaften abschloß. Nachdem er mehrere Stellen als Lehrer in den Niederlanden und in Frankreich bekleidet hatte, wurde er schließlich Professor für Deutsch an der Ecole des Hautes Etudes Commerciales in Paris. Außer für seine Arbeit zur Militärkryptographie ist er vor allem auch noch für linguistische Studien bekannt, insbesondere auch zur heute weithin vergessenen Kunstsprache Volapük.

Wenn ein Schlüssel häufig gewechselt wird, müssen sich die beteiligen Partner jeweils miteinander verständigen, was entweder ein Treffen oder vertrauenswürdige Boten voraussetzt – beides ist mit großem Aufwand verbunden.

1976 publizierten MARTIN HELLMAN, damals Assistenzprofessor an der Stanford University, und sein Forschungsassistent WHITFIELD DIFFIE eine Arbeit mit dem Titel *New directions in cryptography* (IEEE Trans. Inform. Theory **22**, 644–654; inzwischen auch im Netz zu finden), in der sie vorschlugen, den Vorgang der *Ver*schlüsselung und den der *Ent*schlüsselung völlig voneinander zu trennen: Es sei schließlich nicht notwendig, daß der Sender einer verschlüsselten Nachricht auch in der Lage sei, diese zu *ent*schlüsseln.

Der Vorteil eines solchen Verfahrens wäre, daß jeder potentielle Empfänger nur einen einzigen Schlüssel bräuchte und dennoch sicher sein

könnte, daß nur er selbst seine Post entschlüsseln kann. Der Schlüssel müßte nicht einmal geheimgehalten werden, da es ja nicht schadet, wenn jedermann Nachrichten *ver*schlüsseln kann. In einem Netzwerk mit n Teilnehmern bräuchte man also nur n Schlüssel, um jedem Teilnehmer zu erlauben, mit jeden anderen zu kommunizieren, und diese Schlüssel könnten sogar in einem öffentlichen Verzeichnis stehen. Bei einem symmetrischen Kryptosystem wäre der gleiche Zweck nur erreichbar mit $\frac{1}{2}n(n-1)$ Schlüsseln, die zudem noch durch ein sicheres Verfahren wie etwa ein persönliches Treffen oder durch vertrauenswürdige Boten ausgetauscht werden müßten.



BAILEY WHITFIELD DIFFIE wurde 1944 geboren. Erst im Alter von zehn Jahren lernte er lesen; im gleichen Jahr hielt eine Lehrerin an seiner New Yorker Grundschule einen Vortrag über Chiffren. Er ließ sich von seinem Vater alle verfügbare Literatur darüber besorgen, entschied sich dann 1961 aber doch für ein Mathematikstudium am MIT. Um einer Einberufung zu entgehen, arbeitete er nach seinem Bachelor bei Mitre; später, nachdem sein Interesse an der Kryptographie wieder erwacht war, kam er zu Martin Hellman nach Stanford, der ihn als Forschungsassistent einstellte. 1991–2009 arbeitete er als *chief security officer* bei Sun Microsystems; heute ist er *consulting professor* in Stanford.
`http://cisac.stanford.edu/people/whitfield_diffie/`

MARTIN HELLMAN wurde 1945 in New York geboren. Er studierte Elektrotechnik zunächst bis zum Bachelor an der dortigen Universität; für das Studium zum Master und zur Promotion ging er nach Stanford. Nach kurzen Zwischenaufenthalten am Watson Research Center der IBM und am MIT wurde er 1971 Professor an der Stanford University. Nach seiner Emeritierung 1996 gab er noch lange Kurse, mit denen er Schüler für mathematische Probleme interessieren wollte. 2015 erhielt er den Turing Award.
`http://www-ee.stanford.edu/~hellman/`

DIFFIE und HELLMAN machten nur sehr vage Andeutungen, wie ein System mit öffentlichen Schritten aussehen könnte. Es ist zunächst einmal klar, daß ein solches System keine beweisbare absolute Sicherheit

bieten kann, denn die Verschlüsselungsfunktion ist eine bijektive Abbildung zwischen endlichen Mengen, und jeder, der die Funktion kennt, kann zumindest im Prinzip auch ihre Umkehrfunktion berechnen.

Nun läßt sich aber nicht jede theoretisch mögliche Berechnung auch praktisch durchführen; es reicht, wenn wir sicher sein können, daß derzeit niemand die Umkehrfunktion wirklich berechnen kann. Nur darauf beruht die Sicherheit eines Kryptosystems mit öffentlichen Schlüsseln, und leider sind wir uns nie ganz sicher, sondern können nur mit unseren Erfahrungen argumentieren, die umso verläßlicher sind, je länger sich Wissenschaftler im öffentlichen Bereich mit dem Problem beschäftigt haben. Ideal sind also alte, klassische Probleme der Mathematik, an denen schon Generationen von Mathematikern gearbeitet haben.

DIFFIE und HELLMAN bezeichnen eine Funktion, deren Umkehrfunktion nicht mit vertretbarem Aufwand berechnet werden kann, als *Einwegfunktion* und schlagen als Verschlüsselungsfunktion eine solche Einwegfunktion vor. Damit hat man aber noch kein praktikables Kryptosystem, denn bei einer echten Einwegfunktion ist es auch für den legitimen Empfänger nicht möglich, seinen Posteingang zu entschlüsseln. DIFFIE und HELLMAN schlagen deshalb eine Einwegfunktion mit *Falltür* vor, wobei der legitime Empfänger zusätzlich zu seinem öffentlichen Schlüssel noch über einen geheimen Schlüssel verfügt, mit dem er (und nur er) diese Falltür öffnen kann.

Natürlich hängt alles davon ab, ob es solche Einwegfunktionen mit Falltür wirklich gibt. DIFFIE und HELLMAN gaben keine an, und unter den Experten gab es durchaus einige Skepsis bezüglich der Möglichkeit, solche Funktionen zu finden.

Tatsächlich existierten aber bereits damals Systeme, die auf solchen Funktionen beruhten, auch wenn sie nicht in der offenen Literatur dokumentiert waren: Die britische *Communications-Electronics Security Group* (CESG) hatte bereits Ende der sechziger Jahre damit begonnen, nach entsprechenden Verfahren zu suchen, um die Probleme des Militärs mit dem Schlüsselmanagement zu lösen, aufbauend auf (impraktikablen) Ansätzen von AT&T zur Sprachverschlüsselung während des zweiten Weltkriegs. Die Briten sprachen nicht von Kryp-

tographie mit öffentlichen Schlüsseln, sondern von *nichtgeheimer Verschlüsselung,* aber das Prinzip war das gleiche.

Erste Ideen dazu sind in einer auf Januar 1970 datierten Arbeit von JAMES H. ELLIS zu finden, ein praktikables System in einer auf den 20. November 1973 datierten Arbeit von CLIFF C. COCKS. Wie im Milieu üblich, gelangte nichts über diese Arbeiten an die Öffentlichkeit; erst 1997 veröffentlichten die *Goverment Communications Headquarters* (GCHQ), zu denen CESG gehört, einige Arbeiten aus der damaligen Zeit; eine Zeitlang waren sie auch auf dem Server `http://www.cesg.gov.uk/` zu finden, wo sie allerdings inzwischen anscheinend wieder verschwunden sind.

In der offenen Literatur erschien ein Jahr nach der Arbeit von DIFFIE und HELLMAN das erste Kryptosystem mit öffentlichen Schlüsseln: RON RIVEST, ADI SHAMIR und LEN ADLEMAN, damals alle drei am Massachussetts Institute of Technology, fanden nach rund vierzig erfolglosen Ansätzen 1977 schließlich jenes System, das heute nach ihren Anfangsbuchstaben mit RSA bezeichnet wird:

Das System wurde 1983 von der eigens dafür gegründeten Firma RSA Computer Security Inc. patentiert und mit großem kommerziellem Erfolg vermarktet. Das Patent lief zwar im September 2000 aus, die Firma ist aber weiterhin erfolgreich im Kryptobereich tätig.

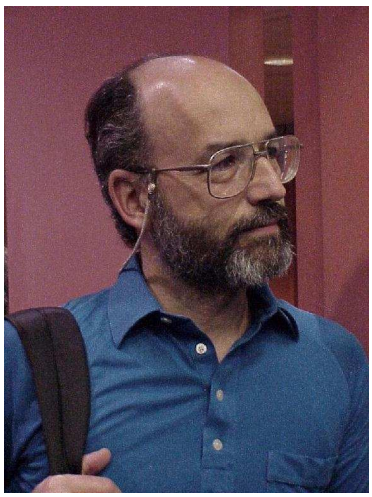
RSA ist übrigens identisch mit dem von laut GCHQ von COCKS vorgeschlagenen System. Die Beschreibung durch RIVEST, SHAMIR und ADLEMAN erschien 1978 unter dem Titel *A method for obtaining digital signatures and public-key cryptosystems* in Comm. ACM **21**, 120–126.

Ausgangspunkt ist die folgende Kombination aus kleinem Satz von FERMAT mit dem erweiterten EUKLIDischen Algorithmus:

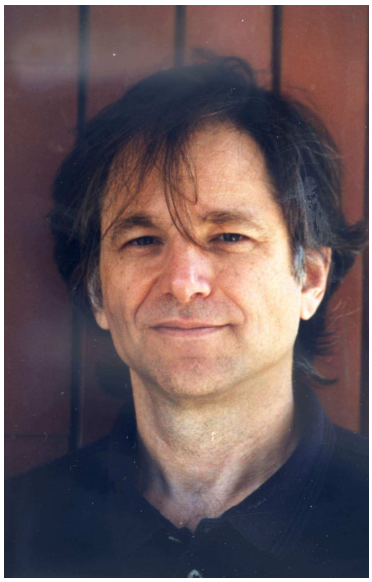
Lemma: Ist p eine Primzahl und $e \in \mathbb{N}$ teilerfremd zu $p-1$, so ist die Abbildung

$$\begin{cases} \{0,\ldots,p-1\} \to \{0,\ldots,p-1\} \\ x \mapsto x^e \bmod p \end{cases}$$

RONALD LINN RIVEST wurde 1947 in Schenectady im US-Bundesstaat New York geboren. Er studierte zunächst Mathematik an der Yale University, wo er 1969 seinen Bachelor bekam; danach studierte er in Stanford Informatik. Nach seiner Promotion 1974 wurde er Assistenzprofessor am Massachussetts Institute of Technology, wo er heute einen Lehrstuhl hat. Er arbeitet immer noch auf dem Gebiet der Kryptographie und entwickelte eine ganze Reihe weiterer Verfahren, auch symmetrische Verschlüsselungsalgorithmen und Hashverfahren. Er ist Koautor eines Lehrbuchs über Algorithmen. Im laufenden Herbstsemester hält er u.a. eine Vorlesung über COVID-19. Seine home page ist `//http://theory.lcs.mit.edu/~rivest/` .



ADI SHAMIR wurde 1952 in Tel Aviv geboren. Er studierte zunächst Mathematik an der dortigen Universität; nach seinem Bachelor wechselte er ans Weizmann Institut, wo er 1975 seinen Master und 1977 die Promotion in Informatik erhielt. Nach einem Jahr als Postdoc an der Universität Warwick und drei Jahren am MIT kehrte er ans Weizmann Institut zurück, wo er bis heute Professor ist. Außer für RSA ist er bekannt sowohl für die Entwicklung weiterer Kryptoverfahren als auch für erfolgreiche Angriffe gegen Kryptoverfahren. Er schlug auch einen optischen Spezialrechner zur Faktorisierung großer Zahlen vor. `http://www.wisdom.weizmann.ac.il/profile/scientists/shamir-profile.html`



LEONARD ADLEMAN wurde 1945 in San Francisco geboren. Er studierte in Berkeley, wo er 1968 einen BS in Mathematik und 1976 einen PhD in Informatik erhielt. Thema seiner Dissertation waren zahlentheoretische Algorithmen und ihre Komplexität. Von 1976 bis 1980 war er an der mathematischen Fakultät des MIT; seit 1980 arbeitet er an der University of Southern California in Los Angeles. Seine Arbeiten beschäftigen sich mit Zahlentheorie, Kryptographie und Molekularbiologie. Er führte nicht nur 1994 die erste Berechnung mit einem „DNS-Computer“ durch, sondern arbeitete auch auf dem Gebiet der Aidsforschung. Heute hat er einen Lehrstuhl für Informatik und Molekularbiologie. `https://adelman.usc.edu`

bijektiv, und ihre Umkehrabbildung ist von der Form

$$\begin{cases} \{0,\dots,p-1\} \to \{0,\dots,p-1\} \\ x \mapsto x^d \bmod p \end{cases}$$

mit einem $d \in \mathbb{N}$.

Beweis: Nach dem erweiterten EUKLIDischen Algorithmus gibt es natürliche Zahlen d und k derart, daß

$$ed - k(p-1) = \mathrm{ggT}(e, p-1) = 1$$

ist und damit $ed = 1 + k(p-1)$. Für jedes zu p teilerfremde $x \in \mathbb{Z}$ ist dann

$$\left(x^e\right)^d = x^{ed} = x^{1+k(p-1)} = x \cdot \left(x^{p-1}\right)^k \equiv x \bmod p$$

nach dem kleinen Satz von FERMAT. Ist x nicht teilerfremd zu p, also ein Vielfaches von p, ist auch x^{ed} eines, so daß $x^{ed} \equiv x \equiv 0 \bmod p$ ist. Damit ist $x^{ed} \equiv x \bmod p$ für alle $x \in \mathbb{Z}$. ∎

Hier ist $\mathcal{B}$ die Menge aller ganzer Zahlen zwischen Null und $p-1$, und $\mathcal{S}$ besteht aus allen Paaren (p, e) aus einer Primzahl p und einer zu $p-1$ teilerfremden natürlichen Zahl $e > 1$. Die Verschlüsselungsfunktion ist

$$f\colon \begin{cases} \mathcal{B} \times \mathcal{S} \to \mathcal{B} \\ \big(m, (p,e)\big) \mapsto m^e \bmod p \end{cases},$$

und die Entschlüsselungsfunktion

$$g\colon \begin{cases} \mathcal{B} \times \mathcal{S} \to \mathcal{B} \\ \big(m, (p,d)\big) \mapsto m^d \bmod p \end{cases}$$

verwendet einen anderen Schlüssel als f. Nach dem gerade bewiesenen Lemma gilt für alle $m \in \mathcal{B}$ und $(p, e) \in \mathcal{S}$ die Gleichung

$$g\Big(f\big(m,(p,e)\big),(p,d)\Big) = m\,,$$

wenn d wie oben aus p und e berechnet wird.

Damit haben wir aber leider noch kein Kryptosystem Problem: Jeder, der außer p und e auch den erweiterten EUKLIDischen Algorithmus kennt, kann d berechnen.

Wenn wir die Primzahl p aber ersetzen durch das Produkt $N = pq$ zweier verschiedener Primzahlen p, q und verlangen, daß e teilerfremd sowohl zu $p-1$ als auch zu $q-1$ ist, ändert sich die Situation ganz entscheidend: Natürlich ist weiterhin

$$x^{1+\ell(p-1)} \equiv x \bmod p \quad \text{und} \quad x^{1+m(q-1)} \equiv x \bmod q$$

für alle $\ell, m \in \mathbb{N}_0$. Wenn wir für $\ell = k(q-1)$ ein Vielfaches von $q-1$ wählen und für $m = k(p-1)$ das entsprechende Vielfache von $p-1$, folgt, daß $x^{1+k(p-1)(q-1)} \equiv x$ ist sowohl modulo p als auch modulo q, also auch modulo $N = pq$ für alle $k \in \mathbb{N}_0$.

Da e teilerfremd zu $(p-1)(q-1)$ vorausgesetzt war, liefert uns EUKLID natürliche Zahlen d, k, so daß

$$ed - k(p-1)(q-1) = 1 \quad \text{und} \quad ed = 1 + k(p-1)(q-1)$$

ist. Somit ist für alle $x \in \{0, 1, \ldots N-1\}$

$$\left(x^e\right)^d = x^{ed} = x^{1+k(p-1)(q-1)} \equiv x \bmod N\,.$$

Für die Menge $\mathcal{B}$ aller ganzer Zahlen von Null bis $N-1$ und die Menge $\mathcal{S}$ aller Paare (N, e), wobei $N = pq$ das Produkt zweier verschiedener Primzahen ist und die natürliche Zahl e teilerfremd zu $p-1$ und zu $q-1$ sein muß, erfüllen

$$f\colon \begin{cases} \mathcal{B} \times \mathcal{S} \to \mathcal{B} \\ \big(m, (N, e)\big) \mapsto m^e \bmod N \end{cases}$$

und

$$g\colon \begin{cases} \mathcal{B} \times \mathcal{S} \to \mathcal{B} \\ \big(m, (N, d)\big) \mapsto m^d \bmod N \end{cases},$$

somit die Gleichung

$$g\Big(f\big(m, (N, e)\big), (N, d)\Big) = m$$

für alle $m \in \mathcal{B}$ und alle $(N, e) \in \mathcal{S}$, sofern d zu (N, e) wie oben berechnet wurde. Insbesondere sind beide Funktionen für festgehaltene Schlüssel bijektive Abbildung von $\mathcal{B}$ nach $\mathcal{B}$.

Zum Verschlüsseln muß man nur N und e kennen, zum Entschlüsseln N und d. Um d aus N und e zu berechnen, muß man aber noch $(p-1)(q-1)$ kennen, und das ist äquivalent zur Kenntnis von p und q:

$$(p-1)(q-1) = pq - (p+q) + 1 = N - (p+q) + 1 \; ;$$

wer N und $(p-1)(q-1)$ kennt, kennt also sowohl pq als auch $p+q$, und kann damit p und q berechnen.

Wegen der Eindeutigkeit der Primzerlegung sind p und q natürlich schon durch N eindeutig bestimmt, und für kleine N lassen sie sich auch leicht finden. Für Zahlen mit mehreren hundert Dezimalstellen wird das Problem allerdings ungleich schwieriger; der derzeitige Rekord in der offenen Literatur für Zahlen, die ein gut gewähltes Produkt zweier ungefähr gleich großer Primzahlen sind, ist die Faktorisierung einer 250-stelligen Zahl (829 Bit) im Februar 2020 durch drei Teams bestehend aus FABRICE BOUDOT, der an der Universität von Limoges, aber auch für das französische Verteidigungsministerium arbeitet, PIERRICK GAUDRY, AURORE GUILLEVIC, EMMANUEL THOME und PAUL ZIMMERMANN aus Nancy und NADIA HENINGER von der University of California in San Diego. Sie geben den Rechenaufwand mit 2700 CPU-Jahren an, bezogen auf einen mit 2,1 GHz getakteten Prozessor.

Selbstverständlich gibt es Geheimdienste mit erheblich besserer Rechnerausstattung als der an Universitäten und Forschungseinrichtungen; insbesondere dürften diese auch Spezialhardware haben, während die hier zitierten Forscher mit (vielen) Standard-PCs arbeiteten. Auf der algorithmischen Seite allerdings ist es unwahrscheinlich, daß Geheimdienste wesentlich mehr wissen als die Experten an Universitäten und Forschungsinstituten; daher kann man hoffen, daß Geheimdienste gut gewählte Produkte zweier Primzahlen mit wesentlich mehr als 1000 Bit nicht faktorisieren können.

Um vor Überraschungen geschützt zu sein, sollte man allerdings darauf noch einen erheblichen Sicherheitszuschlag geben. In der Europäischen Union ist die *Senior Officials Group Information Security* (SOGIS) für entsprechende Empfehlungen zuständig; ihr aktuellstes Dokument stammt vom Januar 2020 und unterscheidet zwischen *legacy* und *recommended mechanisms. Legacy* bedeutet *überliefert, hergebracht*

oder in diesem Zusammenhang auch *Altlast,* d.h. übergangsweise noch toleriert; bis zum 31. Dezember 2025 werden RSA-Moduld mit mindestens 1900 Bit noch toleriert. Empfohlen sind aber heute schon welche mit mindestens drei Tausend Bit.

Ansonsten sollten die Primzahlen ungefähr gleiche Größenordnung haben, denn wenn eine davon sehr klein ist, kann man sie natürlich schnell finden. Sie dürfen allerdings auch nicht zu nahe beieinander liegen, denn sonst führt ein Verfahren von FERMAT zur Faktorisierung: Dieser berechnet die Zahlen $N + x^2$ für $x = 1, 2, 3, \ldots$ so lange, bis eine Quadratzahl $N + x^2 = y^2$ gefunden ist. Dann liefert die dritte binomische Formel die Faktorisierung

$$N = y^2 - x^2 = (y + x)(y - x) = p \cdot q \,.$$

Da $p - q = 2x$ ist, sind hier $\frac{1}{2}\,|p - q|$ Schritte notwendig.

Es gibt allerdings eine Modifikation, mit der es schneller geht: Für große Zahlen N ist es besser, nicht die Zahlen $N + x^2$ für $x \in \mathbb{N}_0$ darauf zu testen, ob $N + x^2 = y^2$ ein Quadrat ist, denn offensichtlich ist $y \geq \sqrt{N}$, und der Abstand zwischen zwei Quadraten dieser Größenordnung ist recht groß: $(y + 1)^2 = y^2 + 2y + 1 > 2\sqrt{N}$. Dagegen sind die Abstände zwischen den Zahlen $N + x^2$ zumindest am Anfang sehr klein. Es ist daher effizienter, wenn man nacheinander die Zahlen y^2 mit $y \geq \sqrt{N}$ darauf testet, ob $y^2 - N$ das Quadrat einer ganzen Zahl x ist. Da x zumindest am Anfang relativ klein ist, geht dieser Test auch schneller als bei der klassischen Vorgehensweise.

Wenn wir davon ausgehen, daß N kein Quadrat ist (was bei RSA selbstverständlich gilt), ist $y = [\sqrt{N}] + 1$ die kleinste ganze Zahl nach $\sqrt{N}$. Der modifizierte Algorithmus verläuft somit wie folgt:

1. Schritt: Setze $y = [\sqrt{N}] + 1$ und $D = y^2 - N$.

2. Schritt: Teste, ob $D = x^2$ Quadrat einer natürlichen Zahl x ist; falls ja, endet der Algorithmus und $N = y^2 - x^2 = (y - x)(y + x)$.

3. Schritt: Ersetze D durch $D + 2y + 1$ und y durch $y + 1$ und gehe zurück zu Schritt 2.

Man beachte, daß nach dem dritten Schritt wieder $D = y^2 - N$ ist, da $(y+1)^2 = y^2 + 2y + 1$ ist. Die Addition von $2y + 1$ zu D geht allerdings, gerade für große Zahlen, deutlich schneller, als die Berechnung des Quadrats $(y + 1)^2$.

Betrachten wir als Beispiel die Zahl $N = 159\,212\,357$. Ihre Quadratwurzel ist ungefähr 12 617,938, also beginnen wir mit $y = 12\,618$ und setzen $D = y^2 - N = 1\,567$. Das ist keine Quadratzahl; daher erhöhen wir D auf $D + 2y + 1 = 26\,804$ und y auf $y + 1 = 12\,619$. Auch die Quadratwurzel des neuen D ist nicht ganzzahlig, also wird im nächsten Durchgang

$$D = 28\,804 + 2 \cdot 12\,619 + 1 = 52\,043 \quad \text{und} \quad y = 12\,619 + 1 = 12\,620\,.$$

Wieder ist D kein Quadrat. Im nächsten Durchgang wird

$$D = 52\,043 + 2 \cdot 12\,620 + 1 = 77\,284 = 278^2$$

und $y = 12\,620 + 1 = 12\,621$. Mit $x = 278$ ist daher

$$N = (y - x)(y + x) = 12343 \cdot 12899\,.$$

Damit hatten wir im vierten Anlauf Erfolg; nach FERMATs klassischer Vorgehensweise wäre dies erst beim 278. Versuch der Fall gewesen.

Für interessierte Leser sei (auch wenn dies nicht wirklich zum Stoff der Vorlesung gehört) gezeigt, wie man den Aufwand nach der modifizierten Methode abschätzen kann:

Angenommen, $N = pq$ mit $p < \sqrt{N} < q$. Dann ist

$$N = pq = \frac{(p+q)^2 - (p-q)^2}{4} = \left(\frac{p+q}{2}\right)^2 - \left(\frac{p-q}{2}\right)^2\,,$$

der Algorithmus endet also mit $y = \frac{1}{2}(p + q)$, und die Anzahl der getesteten y-Werte ist $\frac{1}{2}(p + q) - [\sqrt{N}]$. Um zu sehen, wie viele das sind, brauchen wir eine Abschätzung für $p + q$:

Lemma: $N = pq$ sei das Produkt zweier verschiedener Zahlen p und q, und $\Delta = p - q > 0$. Dann ist

$$0 < p + q - 2\sqrt{N} < \frac{\Delta^2}{4\sqrt{N}}\,.$$

Beweis: Nach den drei binomischen Formeln ist

$$\Delta^2 = (p-q)^2 = p^2 - 2pq + q^2 = (p+q)^2 - 4pq = (p+q)^2 - 4N$$
$$= (p+q-2\sqrt{N})(p+q+2\sqrt{N})\,.$$

Da Δ^2 und $p+q+2\sqrt{N}$ positiv sind, muß auch $p+q-2\sqrt{N}$ positiv sein, d.h. $p+q > 2\sqrt{N}$. (Dies ist, für unseren speziellen Fall, der allgemeine Satz, wonach das geometrische Mittel $\sqrt{xy}$ zweier positiver reeller Zahlen größer oder gleich dem arithmetischen ist mit Gleichheit nur im Fall $x = y$.) Die obige Gleichung für Δ^2 zeigt daher, daß

$$0 < p+q-2\sqrt{N} < \frac{\Delta^2}{p+q+2\sqrt{N}} < \frac{\Delta^2}{4\sqrt{N}}$$

ist. ∎

Wir interessieren uns für die Zahl $\frac{1}{2}(p+q) - [\sqrt{N}]$; diese ist ungefähr gleich

$$\frac{1}{2}(p+q) - \sqrt{N} < \frac{\Delta^2}{8\sqrt{N}}\,.$$

Ist also $\Delta \leq c\sqrt[4]{N}$, so brauchen wir höchstens $c^2/8$ Versuche. Bei den Zahlen, um die es bei RSA-Moduln geht, nimmt jeder einzelne Versuch nur wenig Zeit in Anspruch; auch für ein c in der Größenordnung von mehreren Tausend ist die Faktorisierung daher leicht durchführbar. Damit ist klar, daß die Differenz der beiden Primfaktoren eines RSA-Moduls deutlich größer als die vierte Wurzel von N sein muß. Für ein N mit 2000 Bit heißt dies, daß sich die beiden Faktoren schon deutlich vor dem 1500. Bit in mindestens einem Bit unterscheiden müssen.

Empfohlen wird, daß die beiden Primfaktoren p, q zufällig und unabhängig voneinander erzeugt werden und aus einem Bereich stammen, in dem

$$\varepsilon_1 < |\log_2 p - \log_2 q| < \varepsilon_2$$

gilt. Als *Anhaltspunkte* werden dabei die Werte $\varepsilon_1 = 0{,}1$ und $\varepsilon_2 = 30$ vorgeschlagen; ist p die kleinere der beiden Primzahlen, soll also

$$10^{-3}p \approx 2^{-10}p < q < 2^{30}p \approx 10^9 p$$

gelten.

Für den Exponenten e wurde lange Zeit der kleinstmögliche Wert drei verwendet; das ist nicht nur problematisch, wenn die zu verschlüsselnde Nachricht x so klein ist, daß $c = x^3 < N$ ist, so daß sich x einfach als $\sqrt[3]{c}$ berechnen läßt. Probleme gibt es auch, wenn die gleiche Nachricht als Rundbrief an mehrere Adressaten verschickt wird: Falls ein Angreifer die Nachrichten an drei Adressaten mit RSA-Moduln N_1, N_2 und N_3 abfängt, kennt er x^3 mod N_i für $i = 1, 2, 3$ und kann nach dem chinesischen Restesatz x^3 mod $N_1N_2N_3$ berechnen. Da x kleiner ist als jedes N_i, ist $x^3 < N_1N_2N_3$, und er kann x als dritte Wurzel in $\mathbb{Z}$ berechnen.

Das Bundesamt für Sicherheit in der Informationstechnik empfiehlt daher, daß e nicht zu klein sein darf, sondern die Ungleichungen $2^{16} + 1 \leq e < 2^{256}$ erfüllen sollte. (Bei zu großen werden von e besteht die Gefahr, daß d klein wird, was zu gefährlichen Angriffsmöglichkeiten führen kann.)

Auch bei Beachtung aller dieser Vorschriften und Empfehlungen ist das Verfahren so wie beschrieben immer noch unbrauchbar: Nehmen wir an, wir verschicken einfach eine (streng geheime) Antwort *Ja* oder *Nein.* Ein Gegner muß dann nur die beiden Nachrichten *Ja* und *Nein* verschlüsseln und sehen, welche zum abgefangenen Chiffretext führt. Um dieses Problem zu umgehen, „opfert“ man einen Teil der möglichen Bits und setzt die höchsten mindestens 128 Bit der Nachricht auf Zufallswerte. Dann kann jeder Block auf 2^{128} verschiedene Weisen verschlüsselt werden, und derzeit geht man davon aus, daß 2^{128} Versuche jenseits der Möglichkeiten eines jeden Gegners liegen. Die Nachricht selbst wird natürlich einfach in irgendeiner Weise binär kodiert, meist im ASCII-Code, und die gesamte Bitfolge einschließlich der Zufallsbits wird dann als natürliche Zahl interpretiert.

Man kann übrigens oft kleinere öffentliche Exponenten d zu gebenen öffentlichen Schlüsseln (N, e) bekommen, wenn man den erweiterten EUKLIDischen Algorithmus nicht auf e und $(p-1)(q-1)$ anwendet, sondern auf e und ein kleineres gemeinsames Vielfaches von $p-1$ und $q-1$. Man überzeugt sich leicht davon, daß alle obigen Beweise für beliebige gemeinsame Vielfache von $p-1$ und $q-1$ funktionieren.

Für die praktische Anwendung des RSA-Verfahrens müssen wir uns noch überlegen, wie man die Potenzen x^e mod N *bzw.* x^d mod N effizient berechnen kann. Nehmen wir der Einfachheit halber an, wir rechnen mit einem Modul N von 2 048 Bit, was heute ja gerade noch toleriert wird.

Damit haben auch die zu übermittelnde Nachrichtenblöcke eine Länge von mindestens 2 048 Bit, also 256 Byte, und die e-te Potenz der entsprechenden Zahlen hat die e-fache Länge. Für die Verschlüsselung können wir einen kleinen Exponenten e wählen, für die Entschlüsselung allerdings wird der Exponent d mit an Sicherheit grenzender Wahrscheinlichkeit in der Größenordnung von N liegen, so daß m^d eine Bitlänge von etwa $(2\,048)^2$ Bit hat, also ein halbes Megabyte.

Dafür hat ein heutiger Computer natürlich mehr als genug Speicherplatz, aber er muß die Zahlen auch berechnen, und zumindest wenn man das in der dümmstmöglichen Weise durchführt, indem man sukzessive die Potenzen $m, m^2, m^3, \ldots$ berechnet, überfordern auch deutlich kleinere Exponenten selbst die besten heutigen Supercomputer um Größenordnungen.

Tatsächlich gibt es aber keinen Grund, die natürliche Zahl m^d wirklich zu berechnen: Wir brauchen schließlich nur m^d mod N. Außerdem käme hoffentlich auch kein Leser auf die dumme Idee, die Zahl 3^{32} durch 31-fache Multiplikation mit Drei zu berechnen: Da $32 = 2^5$ ist, läßt sich das Ergebnis viel schneller als

$$3^{32} = \left(\left(\left(\left(3^2\right)^2\right)^2\right)^2\right)^2$$

durch nur fünfmaliges Quadrieren berechnen.

Entsprechend können wir für jede gerade Zahl $n = 2m$ die Potenz x^n als Quadrat von x^m berechnen. Für einen ungeraden Exponenten e ist $e-1$ gerade, wenn wir also m^e als Produkt von m und m^{e-1} darstellen, können wir zumindest im nächsten Schritt wieder die Formel für gerade Exponenten verwenden. Da uns das Ergebnis nur modulo N interessiert, können wir zudem nach jeder Multiplikation und jeder Quadrierung

das Ergebnis modulo N reduzieren; auf diese Weise entsteht nie ein Zwischenergebnis, das größer ist als N^2.

Dies führt auf folgenden rekursiven Algorithmus zur Berechnung von $m^e \bmod N$:

Falls $e = 2f$ gerade ist, berechne man zunächst $m^f \bmod N$ nach diesem Algorithmus und quadriere das Ergebnis modulo N; andernfalls gibt es im Falle $e = 1$ nichts zu tun, und für $e > 1$ berechne man zunächst $m^{e-1} \bmod N$ und multipliziere das Ergebnis modulo N mit m.

Falls e eine Zahl mit r Bit ist, erfordert dieser Algorithmus $r - 1$ Quadrierungen und höchstens r, im Mittel rund $r/2$ Multiplikationen mit m. Für einen Exponenten mit 2 048 Bit brauchen wir also im Mittel rund 3 072 Multiplikationen, auf keinen Fall aber mehr als 4 096, und damit wird ein heutiger Computer problemlos fertig.

Bleibt noch die Frage: Wie multiplizieren wir zwei Zahlen mit einer Länge von mehreren Tausend Bit?

Für Taschenrechner wie auch für die 32- oder 64-Bit-Register eines Computers sind sie natürlich viel zu groß. Trotzdem ist die vielleicht erstaunliche Antwort auf obige Frage, daß wir genau so vorgehen können, wie wir es in der Schule gelernt haben: Zwar gibt es Multiplikationsalgorithmen, die asymptotisch schneller sind als die Schulmethode, aber tatsächlich schneller werden sie erst, wenn die Zahlen eine Bitlänge haben, die eher bei Millionen liegt als bei bloßen Tausenden.

Einen Unterschied zur Schule sollten wir freilich machen: Während uns in der Grundschule das kleinen Einmaleins eingepaukt wird, also die Produkte der Zahlen von Eins bis Zehn untereinander, sind in den CPUs unserer Computer Algorithmen implementiert, die zwei 32-Bit-Zahlen zu einer 64-Bit-Zahl multiplizieren oder, falls der Computer hinreichend neu ist, zwei 64-Bit-Zahlen zu einer 128-Bit-Zahl. Wir sollten die Zahlen also nicht im Zehnersystem betrachten, sondern im Ziffernsystem mit Basis 2^{32} oder 2^{64}.

Nach jeder Multiplikation muß das Ergebnis modulo N reduziert werden; wir müssen also durch N dividieren. Auch dazu können wir *im Prinzip* genauso vorgehen wie in der Schule, haben dabei allerdings das

Problem, daß das in der Schule gelehrte Divisionsverfahren kein Algorithmus ist: Wir müssen schließlich in jedem Schritt die nächste Ziffer des Quotienten *erraten* und sehen erst nach Multiplikation mit dem Divisor oder sogar erst nach Subtraktion dieses Produkts vom Dividenden, ob wir das korrekte Ergebnis haben.

Zum Glück läßt sich dieses „Erraten“ selbst für beliebige Basen des Ziffernsystems zumindest insoweit algorithmisch machen, daß das Ergebnis nie um mehr als zwei danebenliegt, und auch ein Fehler von zwei nur mit verschwindend geringer Wahrscheinlichkeit auftritt. Da es inzwischen viele Unterprogrammpakete und auch Programme gibt, in denen Algorithmen zum Rechnen mit Langzahlen implementiert sind, sei hier nicht auf Einzelheiten eingegangen; Interessenten finden diese zusammen mit allen Beweisen z.B. in Abschnitt 4.3 des bereits im Zusammenhang mit der Aufwandsabschätzung für den EUKLIDischen Algorithmus zitierten Buchs

> DONALD E. KNUTH: The Art of Computer Programming, *vol. 2:* Seminumerical Algorithms, *Addison Wesley,*, ^{3}1997

Ein großer Vorteil eines Verfahrens mit öffentlichen Schlüsseln gegenüber einem klassischen Kryptoverfahren ist die Möglichkeit elektronischer Unterschriften. Angenommen, der Besitzer des privaten Exponenten d zum öffentlichen Schlüsseln (N, e) möchte eine Nachricht x unterschreiben. Dann kann er dazu einfach $u = x^d \bmod N$ berechnen. Da niemand außer ihm d kennt, kann nur er diese Zahl bestimmen. Jeder, der den öffentlichen Schlüssel kennt, kann aber $u^e \equiv x^{de} \equiv x \bmod N$ berechnen und sich davon überzeugen, daß er wirklich das Ergebnis x erhält. Solche elektronischen Unterschriften sind innerhalb der Europäischen Union rechtsverbindlich sofern sie den jeweils geltenden Vorschriften genügen. Derzeit ist das die sogenannte eIDAS-Verordnung *(**e**lectronic **ID**entification, **A**uthentication and Trust **S**ervices)* vom 23. Juli 2014.

Der Wert einer elektronischen Unterschrift steht und fällt damit, daß der korrekte öffentliche Schlüssel des Unterschreibenden bekannt ist: Falls es jemandem gelingt, einem anderen einen falschen Schlüssel für eine

Person zu unterschieben, kann er beliebig viele Unterschriften in deren Namen leisten. Zu elektronischen Unterschriften (und allgemein zur Kryptographie mit öffentlichen Schlüsseln) gehört daher eine *public key infrastructure* mit zertifizierten Schlüsseln. An der Spitze stehen einige wenige Zertifizierungsagenturen, deren öffentliche Schlüssel weithin bekannt sind (insbesondere sind sie in den gängigen Browsern gespeichert), und diese unterschreiben für Ihre Kunden Zertifikate, die Namen, öffentlichen Schlüssel, *usw.* enthalten. Oft ist das Verfahren mehrstufig, d.h. der Inhaber eines Zertifikats kann auf dessen Grundlage selbst Zertifikate ausstellen. Erst mit zertifizierten Unterschriften wird ein sicherer Handel über das Internet möglich: Da RSA zur Verschlüsselung langer Texte zu aufwendig ist, verwendet man dazu symmetrische Verfahren, typischerweise den sogenannten *Advanced Encryption Standard* AES, der mit eine Schlüssellänge von 128 Bit arbeitet. Wenn der Kunde ein Zertifikat mit dem öffentlichen Schlüssel des Händlers bekommt, kann er ihm einen damit verschlüsselten AES-Schlüssel schicken, den die beiden dann zur weiteren Kommunikation verwenden. (Das tatsächlich verwendete Verfahren TLS/SSL ist aufwendiger; hier wirken beide Seiten an der Erstellung des Schlüssels mit.)

Zum Schluß dieses Kapitels möchte ich noch darauf hinweisen, daß sichere Kryptographie keineswegs nur Mathematik ist; auch das beste Kryptoverfahren wird wertlos, wenn sich ein Gegner Zugriff auf Ihren Computer verschafft oder die Routine zur Erzeugung einer „zufälligen“ Zahl als Ausgangspunkt zur Primzahlsuche manipuliert oder . . .
Nicht nur NSA hat viele Möglichkeiten.

Kapitel 3
Grundlegende algebraische Strukturen

Bisher sind wir so mit Zahlen und mit Gleichungen umgegangen, wie es bereits vor mehreren hundert oder sogar seit über zwei Tausend Jahren üblich war. Zu Beginn des zwanzigsten Jahrhunderts erhielt das Wort *Algebra* jedoch langsam eine andere Bedeutung: Im Mittelpunkt standen nicht mehr Gleichungen, sondern Strukturen.

Rechengesetze wie etwa das Kommutativgesetz der Addition oder Multiplikation waren natürlich schon lange bekannt; der neue Gesichtspunkt war, daß man völlig von der Art der Verknüpfung und den zu verknüpfenden Objekten abstrahierte und nur von den Rechenregeln ausging. Das führt zwar zu abstrakten und eher unanschaulichen Strukturen, hat aber den Vorteil, daß ein Satz, der nur unter Voraussetzung gewisser Regeln bewiesen wurde, in allen Zahl- und sonstigen Bereichen gilt, für die man diese Regeln nachweisen kann.

Wir beginnen mit dem Fall nur einer Verknüpfung, denn es gibt ja auch Gemeinsamkeiten zwischen beispielsweise Addition und Multiplikation, die auf diese Weise zusammen betrachtet werden können.

§1: Halbgruppen und Monoide

Fast alle Verknüpfungen, mit denen man in der Mathematik arbeitet, erfüllen das Assoziativgesetz; eine der wenigen Ausnahmen ist das Kreuzprodukt im $\mathbb{R}^3$: Hier ist beispielsweise

$$\begin{pmatrix}1\\0\\0\end{pmatrix} \times \left(\begin{pmatrix}1\\0\\0\end{pmatrix} \times \begin{pmatrix}0\\0\\1\end{pmatrix} \right) = \begin{pmatrix}1\\0\\0\end{pmatrix} \times \begin{pmatrix}0\\-1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\-1\end{pmatrix} ,$$

aber

$$\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\times\begin{pmatrix}1\\0\\0\end{pmatrix}\right)\times\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}\times\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}\ .$$

Es bietet sich daher an, als einfachste Struktur eine zu definieren, bei der die Verknüpfung nur das Assoziativgesetz erfüllen muß:

Definition: Eine *Halbgruppe* ist eine Menge H zusammen mit einer Verknüpfung $*\colon H\times H\to H$, für die gilt:

$$x*(y*z)=(x*y)*z\quad\text{für alle } x,y,z\in H\ .$$

Beispiele von Halbgruppen sind die natürlichen Zahlen, sowohl bezüglich der Addition als auch bezüglich der Multiplikation, genauso natürlich die ganzen, rationalen und reellen Zahlen, $n\times m$-Matrizen bezüglich der Addition, $n\times n$-Matrizen bezüglich der Multiplikation und viele mehr. Vor allem in der Informatik populär sind sogenannte Worthalbgruppen; hier geht man aus von einem Alphabet A und betrachtet die Menge H aller nichtleerer Folgen von Elementen aus A; Verknüpfung ist das Hintereinanderschreiben:

$$\text{hintereinander}*\text{schreiben}=\text{hintereinanderschreiben}\ .$$

Wenn man in einer Halbgruppe ein Produkt von n Elementen $x_1,\dots,x_n$ berechnen will, muß man durch Klammerung die Reihenfolge der Rechenoperationen festlegen, bei vier Elementen etwa

$$x_1*\big(x_2*(x_3*x_4)\big),\quad\big((x_1*(x_2*x_3)\big)*x_4,\quad(x_1*x_2)*(x_3*x_4)$$

oder auf verschiedene andere Weisen. Wir definieren das Produkt

$$p_n=\prod_{i=1}^{n}x_i$$

von n Elementen $x_1,\dots,x_n$ für $n\geq 1$ rekursiv durch

$$\prod_{i=1}^{1}x_i=x_1\quad\text{und}\quad\prod_{i=1}^{n+1}x_i=\left(\prod_{i=1}^{n}x_i\right)*x_{n+1}\ .$$

Lemma: Das Produkt von n Elementen $x_1, x_2, \ldots, x_n$ (in dieser Reihenfolge) ist unabhängig von der Klammerung stets gleich $\prod_{i=1}^{n} x_i$.

Den *Beweis* führen wir durch vollständige Induktion: Für $n = 1$ und $n = 2$ sind alle Klammern überflüssig und können daher weggelassen werden; somit gibt es hier nichts zu beweisen. Sei also $n > 2$. Wir gehen aus von einem irgendwie geklammerten Produkt und betrachten die Operation, die als letzte ausgeführt wird. Diese verknüpft für ein $m < n$ ein irgendwie geklammertes Produkt der Elemente $x_1, \ldots, x_m$ mit einem irgendwie geklammerten Produkt der Elemente $x_{m+1}, \ldots, x_n$. Nach Induktionsannahme ist das erste Produkt gleich $p_m \underset{\text{def}}{=} \prod_{i=1}^{m} x_i$, das zweite ist $q_{nm} \underset{\text{def}}{=} \prod_{i=m+1}^{n} x_i$. Nach Definition ist $q_{nm} = q_{n-1,m} * x_n$, also ist $p_m * q_{nm} = p_m * (q_{n-1,m} * x_n) = (p_m * q_{n-1,m}) * x_n$. Da $p_m * q_{n-1,m}$ ein Produkt von $n - 1$ Faktoren ist, zeigt die Induktionsannahme, daß es den Wert p_{n-1} hat; somit ist $p_m * q_{nm} = p_{n-1} * x_n = p_n$. Damit ist das Lemma bewiesen. ∎

In einer Halbgruppe muß es kein Element geben, das sich bei Verknüpfung mit jedem anderen Element „neutral“ verhält wie etwa die Null bei der Addition. Wenn es so ein Element gibt, reden wir von einem *Monoid:*

Definition: Eine Halbgruppe H mit Verknüpfung $*$ heißt *Monoid,* wenn es ein Element $e \in H$ gibt, so daß $e * x = x * e = x$ ist für alle $x \in H$.

So sind beispielsweise die natürlichen Zahlen bezüglich der Addition *kein* Monoid, da $0 \notin \mathbb{N}$, sie sind aber ein Monoid bezüglich der Multiplikation, da $1 \in \mathbb{N}$. Die Menge $\mathbb{N}_0$ ist sowohl bezüglich der Addition als auch bezüglich der Multiplikation ein Monoid, da sie sowohl die Null als auch die Eins enthält. Eine Worthalbgruppe wird zum Monoid, wenn wir das leere Wort zulassen; es ist dadurch charakterisiert, daß es keinen einzigen Buchstaben enthält, also eine leere Zeichenkette ist.

In der Definition eines Monoids wurde nur gefordert, daß es *mindestens* ein Element e gibt, für das $x * e = e * x = x$ ist; tatsächlich ist e, wenn es existiert, durch die Verknüpfung eindeutig festgelegt:

Lemma: Erfüllt das Element $n \in H$ die Gleichung $n * x = x * n = x$ für alle $x \in H$, so ist $n = e$.

Beweis: Setzen wir speziell $h = e$, folgt, daß $n*e = e$ ist. Nach Definition eines Monoids ist aber $n * e = n$, so daß $n = e$ sein muß. ∎

§2: Gruppen

Außer neutralen Elementen haben wir oft auch noch inverse Elemente; ein Monoid, in dem es Inverse gibt, bezeichnen wir als eine *Gruppe.* Gruppen sind erheblich wichtiger als Halbgruppen und Monoide; sie spielen in vielen Teilen der Mathematik (auch außerhalb der Algebra) eine entscheidende Rolle. Ich gehe davon aus, daß alle Hörer dieser Vorlesung bereits mit dem Begriff der Gruppe vertraut sind; schließlich wird er auch in der Linearen Algebra vielfach benötigt, beispielsweise bei der Berechnung von Determinanten. Zur Fixierung der Begriffe seien die wesentlichen Definitionen trotzdem noch einmal wiederholt:

Definition: *a)* Eine *Gruppe* ist eine Menge G zusammen mit einer Verknüpfung $*$: $G * G \to G$, für die gilt

1.) $(x * y) * z = x * (y * z)$ für alle $x, y, z \in G$.

2.) Es gibt ein Element $e \in G$, das Neutralelement, so daß für alle $x \in G$ gilt $e * x = x * e = x$.

3.) Zu jedem $x \in G$ gibt es ein $x' \in G$, so daß $x * x' = x' * x = e$ ist.

b) Ist G eine endliche Menge, bezeichnen wir deren Elementanzahl als die *Gruppenordnung* $|G|$ von G.

c) Die Gruppe heißt *kommutativ* oder *abelsch,* wenn zusätzlich noch folgende Bedingung erfüllt ist:

4.) $x * y = y * x$ für alle $x, y \in G$.

d) Eine Teilmenge $U \subseteq G$ heißt *Untergruppe* von G, in Zeichen $U \leq G$, wenn sie bezüglich der Operation $*$ selbst eine Gruppe ist.

Unter den bekannten Zahlbereichen sind $\mathbb{N}$ und $\mathbb{N}_0$ weder bezüglich der Addition noch bezüglich der Multiplikation Gruppen, $\mathbb{Z}$ ist eine Gruppe bezüglich der Addition, nicht aber der Multiplikation. Die Körper $\mathbb{Q}, \mathbb{R}$ und $\mathbb{C}$ (und auch alle anderen) sind ebenfalls additive Gruppen; wenn man die Null entfernt, werden sie zu multiplikativen Gruppen.

Bezüglich der Addition ist $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$, bezüglich der Multiplikation $\mathbb{Q} \setminus \{0\} \leq \mathbb{R} \setminus \{0\} \leq \mathbb{C} \setminus \{0\}$.

Als weiteres Beispiel für Untergruppen können wir die sämtlichen Untergruppen der (additiven) Gruppe $\mathbb{Z}$ bestimmen:

Lemma: Die Untergruppen von $\mathbb{Z}$ sind genau die Mengen

$$m\mathbb{Z} \underset{\text{def}}{=} \{mz \mid z \in \mathbb{Z}\} \quad \text{mit} \quad m \in \mathbb{N}_0 \,.$$

Beweis: Natürlich sind alle diese Mengen Untergruppen, denn

$$mz_1 + mz_2 = m(z_1 + z_2) \quad \text{und} \quad (-mz) + mz = 0 \,.$$

Umgekehrt sei $U \leq \mathbb{Z}$ eine Untergruppe von $\mathbb{Z}$. Falls U nur aus der Null besteht, ist $U = \{0\} = 0\mathbb{Z}$. Andernfalls enthält U eine kleinste positive Zahl m, denn zu jedem negativen x muß U auch dessen Inverses $-x$ enthalten.

Nun sei x ein beliebiges Element von U. Nach dem erweiterten EUKLIDischen Algorithmus gibt es $\alpha, \beta \in \mathbb{Z}$, so daß ggT$(m, x) = \alpha m + \beta x$ ist. Da m und x in U liegen, ist auch der ggT ein Element von U. Er ist einerseits ein positiver Teiler von m, andererseits ist m die kleinste positive Zahl in U. Also muß er gleich m sein, d.h. x ist ein Vielfaches von m und $U = \mathbb{Z}m$. ∎

Da eine Gruppe insbesondere auch ein Monoid ist, gibt es auch in einer Gruppe genau ein Neutralelement. Auch die inversen Elemente sind eindeutig bestimmt, denn ist $x' * x = x * x' = e$ und $x'' * x = x * x'' = e$, so ist $x' = x' * e = x' * (x * x'') = (x' * x) * x'' = e * x'' = x''$.

Am Rande sei noch vermerkt, daß man sich bei der Definition einer Gruppe auch darauf beschränken könnte, nur zu fordern. daß es ein Element $e \in G$ gibt mit $e * x = x$ für alle $x \in G$, und zu jedem $x \in G$ ein $x' \in G$ mit $x' * x = e$. Zu diesem x' existiert dann auch ein $x'' \in G$ mit $x'' * x' = e$, und

$$\begin{aligned} x * x' &= e * (x * x') = (x'' * x') * (x * x') = x'' * (x' * x) * x' \\ &= x'' * e * x' = x'' * x' = e \,. \end{aligned}$$

Außerdem ist $x * e = x * (x' * x) = (x * x') * x = e * x = x$, so daß in G auch die beiden nicht vorausgesetzten Gleichungen erfüllt sind.

Man beachte, daß man sich bei der Definition eines Monoids *nicht* darauf beschränken kann, nur zu fordern, daß es ein Element e gibt mit $e * x = x$ für alle x: Beispielsweise wird die Menge $H = \{1, 2\}$ mit der Verknüpfung $x * y = y$ zu einer Halbgruppe, denn $x * (y * z) = y * z = z$ und auch $(x * y) * z = z$. Weiterhin ist $1 * x = x$ für alle $x \in H$, aber $2 * 1 = 1 \neq 2$.

Im folgenden werden wir, wenn keine Verwechslung zu befürchten ist, die Gruppenoperation meist weglassen und statt $x * y$ einfach xy schreiben. Wenn die Gruppenoperation als Addition oder Multiplikation aufgefaßt werden kann, schreiben wir $x+y$ *bzw.* $x \cdot y$, wobei der Malpunkt ebenfalls meist weggelassen wird. Das eindeutig bestimmte Inverse bezeichnen wir meist mit x^{-1}; wenn die Gruppe additiv geschrieben wird, schreiben wir $-x$. In additiven Gruppen wird das Neutralelement meist als 0 geschrieben, in multiplikativen als 1.

Damit eine Teilmenge $U \subseteq G$ eine Untergruppe ist, reicht es offensichtlich, daß das Produkt zweier Elemente $x, y \in U$ wieder in U liegt, außerdem auch das Neutralelement und zu jedem $x \in U$ das Inverse, denn das Assoziativgesetz gilt für alle Elemente von G, also erst recht für alle Elemente der Teilmenge U. Auf die Bedingung $e \in U$ können wir verzichten, falls U nicht leer ist, denn dann liegt zu $x \in U$ auch x^{-1} in U und damit auch $e = xx^{-1}$.

Die obigen Bedingungen können wir etwas kompakter formulieren mit Hilfe des sogenannten Komplexprodukts:

Definition: Sind $A, B \subseteq G$ zwei Teilmengen einer Gruppe G, so bezeichnen wir

$$AB = \{ab \mid a \in A \wedge b \in B\}$$

als das *Komplexprodukt* von A und B. Besteht $A = \{a\}$ nur aus einem Element a, schreiben wir statt $\{a\}B$ auch kurz aB, entsprechend auch Ab, falls $B = \{b\}$. Außerdem definieren wir

$$A^{-1} = \{a^{-1} \mid a \in A\}$$

als die Menge aller inverser Elemente zu den Elementen von A.

Damit werden die beiden ersten Bedingungen für eine Untergruppe zu $UU \subseteq U$ und $U^{-1} \subseteq U$; als dritte Bedingung können wir entweder $e \in U$ oder $U \neq \emptyset$ fordern.

Wenn wir uns für die Lösung von Gleichungen interessieren, unterscheiden sich Gruppen von Halbgruppen und Monoiden durch die folgende

Kürzungsregel: Gilt für drei Elemente a, x, y einer Gruppe eine der Gleichungen $ax = ay$ oder $xa = ya$, so ist $x = y$.

Beweis: Wenn wir die Gleichung $ax = ay$ von links mit a^{-1} multiplizieren, erhalten wir $a^{-1}ax = a^{-1}ay$, also $x = y$. Bei $xa = ya$ müssen wir entsprechend von rechts mit a^{-1} multiplizieren. ∎

Alternativ folgt diese Regel aus dem etwa allgemeineren

Lemma: Für jede Gruppe G und jedes Element $a \in G$ sind die Abbildungen $x \mapsto ax$ und $x \mapsto xa$ bijektiv.

Beweis: Offensichtlich sind die Abbildungen $y \mapsto a^{-1}y$ und $y \mapsto ya^{-1}$ Umkehrabbildungen. ∎

Korollar: In einer Gruppe G haben die Gleichungen $ax = b$ und $ya = b$ für jedes Paar von Elementen $a, b \in G$ genau eine Lösung. ∎

Diese Lösungen lassen sich natürlich auch konkret angeben: $x = a^{-1}b$ und $y = ba^{-1}$.

Umgekehrt läßt sich nun als Übungsaufgabe leicht zeigen, daß eine nichtleere Menge G mit einer assoziativen Verknüpfung genau dann eine Gruppe ist, wenn für alle $a, b \in G$ die Gleichungen $ax = b$ und $xa = b$ lösbar sind. Dies zeigt, daß auch der abstrakte Gruppenbegriff eng mit dem Grundproblem der klassischen Algebra, dem Lösen von Gleichungen, verbunden ist.

Im vorigen Kapitel haben wir verschiedentlich Rechnungen in den ganzen Zahlen modulo einer oder mehrerer natürlicher Zahlen ausgeführt.

Um auch solche Zusammenhänge abstrakt beschreiben zu können, brauchen wir nicht nur Strukturen wie Gruppen, Monoide und Halbgruppen, sondern auch Abbildungen zwischen solchen Strukturen, die mit den Verknüpfungen kompatibel sind. Diese werden als *Homomorphismen* bezeichnet. Die folgende Definition könnte wörtlich übernommen werden für Halbgruppen und Monoide; da wir es aber in dieser Vorlesung kaum je mit Halbgruppen zu tun haben werden, die keine Gruppen sind, beschränke ich mich aber auf diese:

Definition: *a)* Eine Abbildung $\varphi\colon G \to H$ zwischen zwei Gruppen G und H mit Verknüpfungen $*$ und $\times$ heißt (Gruppen-)*Homomorphismus,* falls für alle $x, y \in G$ gilt: $\varphi(x * y) = \varphi(x) \times \varphi(y)$.
b) Ein $\left\{\begin{array}{c} \textit{Monomorphismus} \\ \textit{Epimorphismus} \\ \textit{Isomorphismus} \end{array}\right\}$ ist ein $\left\{\begin{array}{l} \textit{injektiver} \\ \textit{surjektiver} \\ \textit{bijektiver} \end{array}\right\}$ Homomorphismus.
Zwei Gruppen G und H heißen *isomorph,* in Zeichen $G \cong H$, wenn es einen Isomorphismus $\varphi\colon G \to H$ gibt.
c) Ist $G = H$, bezeichnen wir einen Homomorphismus von G nach G auch als *Endomorphismus* und einen Isomorphismus als *Automorphismus.*
d) Das *Bild* eines Homomorphismus $\varphi\colon G \to H$ ist

$$\operatorname{Bild}\varphi \underset{\text{def}}{=} \varphi(G) = \{\varphi(x) \mid x \in G\} ;$$

sein *Kern* ist

$$\operatorname{Kern}\varphi \underset{\text{def}}{=} \{x \in G \mid \varphi(x) = e'\} ,$$

wobei e' das Neutralelement von H bezeichnet.

Lemma: Für jeden Homomorphismus $\varphi\colon G \to H$ gilt:
a) Für die Neutralelemente $e \in G$ und $e' \in H$ ist $\varphi(e) = e'$.
b) Für alle $x \in G$ ist $\varphi(x)^{-1} = \varphi\big(x^{-1}\big)$
c) $\operatorname{Kern}\varphi$ ist eine Untergruppe von G und $\operatorname{Bild}\varphi$ ist eine Untergruppe von H.

Beweis: Nach Definition eines Homomorphismus ist

$$\varphi(e)\varphi(e) = \varphi(ee) = \varphi(e) ;$$

außerdem ist natürlich auch $e'\varphi(e) = \varphi(e)$. Da $x\varphi(e) = \varphi(e)$ in einer Gruppe genau eine Lösung hat, muß $\varphi(e) = e'$ sein. Genauso muß $\varphi\big(x^{-1}\big) = \varphi(x)^{-1}$ sein, denn $\varphi\big(x^{-1}\big)\varphi(x) = \varphi\big(x^{-1}x\big) = \varphi(e) = e'$ und auch $\varphi(x)^{-1}\varphi(x) = e'$.

Für zwei Elemente $x, y \in \text{Kern}\,\varphi$ ist $\varphi(xy) = \varphi(x)\varphi(y) = e'e' = e'$ und $\varphi\big(x^{-1}\big) = \varphi(x)^{-1} = e'^{-1} = e'$, so daß auch Produkte und Inverse wieder im Kern liegen. Da $\varphi(e) = e'$, liegt auch e im Kern, also bildet dieser eine Untergruppe. Auch beim Bild liegen Produkte und Inverse wegen $\varphi(x)\varphi(y) = \varphi(xy)$ und $\varphi(x)^{-1} = \varphi\big(x^{-1}\big)$ wieder im Bild, und da G als Gruppe nicht leer ist, ist auch $\text{Bild}\,\varphi \neq \emptyset$.

■

Die Abbildungen $x \mapsto ax$ und $x \mapsto xa$ sind für $a \neq e$ natürlich keine Homomorphismen; ein Homomorphismus bildet schließlich das Neutralelement ab auf das Neutralelement. Um dies zu erreichen, können wir die Multiplikation von rechts mit a kombinieren mit der Multiplikation von links mit a^{-1}; wir betrachten also die Abbildung $x \mapsto a^{-1}xa$. Dann wird natürlich e auf e abgebildet, und wir haben auch einen Homomorphismus, denn für alle $x, y \in G$ ist

$$a^{-1}(xy)a = a^{-1}\big(x(aa^{-1})y\big)a = \big(a^{-1}xa\big)\big(a^{-1}ya\big)\,.$$

Die Abbildung ist auch bijektiv, also ein Automorphismus von G, denn sowohl die Linksmultiplikation mit a^{-1} als auch die Rechtsmultiplikation mit a sind bijektiv, also auch ihre Hintereinanderausführung.

Definition: Die Abbildung von G nach G, die jedem $x \in G$ das Element $x^a \underset{\text{def}}{=} a^{-1}xa$ zuordnet, heißt *Konjugation* mit $a \in G$. Ein Automorphismus $\varphi\colon G \to G$ heißt *innerer Automorphismus,* wenn es ein $a \in G$ gibt, so daß $\varphi(x) = x^a$ für alle $x \in G$.

Man beachte, daß die Konjugation genau die gleiche Gestalt hat, wie die aus der Linearen Algebra bekannte Konjugation von Matrizen: Auch dort betrachtet man zu einer Matrix M und einer invertierbaren Matrix B, der Matrix des Basiswechsels, die Matrix $B^{-1}MB$. Falls auch M invertierbar ist, stimmt dies mit der Konjugation in der Gruppe der invertierbaren $n \times n$-Matrizen überein.

Zu einem Vektorraum V betrachtet man in der Linearen Algebra auch seine Untervektorräume U und die Faktorräume V/U; wir wollen etwas analoges auch für Gruppen definieren:

Definition: Die Linksnebenklassen einer Untergruppe $U \leq G$ sind die Mengen $aU = \{au \mid u \in U\}$, die Rechtsnebenklassen entsprechend $Ua = \{ua \mid u \in U\}$ für ein $a \in G$

Für $a \in U$ ist wegen der Bijektivität der Multiplikation mit a natürlich $aU = Ua = U$.

Wenn zwei Nebenklassen aU und bU nichtleeren Durchschnitt haben, muß es zwei Elemente $u, v \in U$ geben mit $au = bv$. Das ist äquivalent zu $a = bvu^{-1}$, also ist $aU = bvu^{-1}U = bU$, da $vu^{-1} \in U$. Für $aU \neq bU$ ist somit $aU \cap bU = \emptyset$. Genauso ist auch $Ua \cap Ub = \emptyset$, falls $Ua \neq Ub$.

Nun nehmen wir an, die Gruppe G sei endlich. Dann ist natürlich erst recht jede Untergruppe $U \leq G$ endlich, und es gibt nur endlich viele Nebenklassen. Da die Multiplikation mit einem Gruppenelement eine injektive Abbildung ist, hat sowohl aU als auch Ua für jedes $a \in G$ dieselbe Elementanzahl, nämlich die von U. Insbesondere ist also die Anzahl der Linksnebenklassen gleich der der Rechtsnebenklassen.

Dies gilt auch, wenn G (und eventuell U) unendlich sind, denn da jedes Element einer Gruppe ein eindeutig bestimmtes Inverses hat, ist auch die Abbildung, die jedem Element von G dessen Inverses zuordnet, bijektiv, und sie ordnet der Linksnebenklasse aU die Rechtsnebenklasse Ua^{-1} zu und umgekehrt. Damit gibt es auch hier eine Bijektion zwischen der Menge der Linksnebenklassen und der der Rechtsnebenklassen: Sind a_iU mit i aus irgendeiner Indexmenge I die sämtlichen Linksnebenklassen, sind die Ua_i^{-1} die sämtlichen Rechtsnebenklassen. Wenn wir von der Anzahl der Nebenklassen sprechen, ist es also egal, ob wir Links- oder Rechtsnebenklassen meinen.

Definition: Falls die Untergruppe $U \leq G$ der Gruppe G nur eine endliche Anzahl n von Nebenklassen hat, sagen wir, U habe den Index n, in Zeichen $[G : U] = n$. Falls es unendlich viele Nebenklassen gibt, sagen wir, U habe unendlichen Index.

Im Falle einer endlichen Gruppe G ist auch die Anzahl $[G : U]$ der Nebenklassen endlich. Jede dieser Nebenklassen hat genau so viele Elemente wie U und jedes Element von G liegt in genau einer Nebenklasse; damit folgt der

Satz von Lagrange: Für eine endliche Gruppe G und eine Untergruppe $U \leq G$ ist $|G| = [G : U] \cdot |U|$. Insbesondere sind die Ordnung und der Index von U Teiler der Ordnung von G.

■

JOSEPH-LOUIS LAGRANGE (1736–1813) wurde als GIUSEPPE LODOVICO LAGRANGIA in Turin geboren und studierte dort zunächst Latein. Erst eine alte Arbeit von HALLEY über algebraische Methoden in der Optik weckte sein Interesse an der Mathematik, woraus ein ausgedehnter Briefwechsel mit EULER entstand. In einem Brief vom 12. August 1755 berichtete er diesem unter anderem über seine Methode zur Berechnung von Maxima und Minima; 1756 wurde er, auf EULERs Vorschlag, Mitglied der Berliner Akademie; zehn Jahre später zog er nach Berlin und wurde dort EULERs Nachfolger als mathematischer Direktor der Akademie. 1787 wechselte er an die Pariser Académie des Sciences, wo er bis zu seinem Tod blieb und unter anderem an der Einführung des metrischen Systems beteiligt war. Seine Arbeiten umspannen weite Teile der Analysis, Algebra und Geometrie.

Der Satz von LAGRANGE liefert auch einen neuen Beweis für den kleinen Satz von FERMAT: Für eine Primzahl p ist die Menge $\{1, \ldots, p-1\}$ bezüglich der Multiplikation modulo p eine abelsche Gruppe: Kommutativ- und Assoziativgesetz sind klar, eins ist das Neutralelement, und für jedes x aus der Menge ist ggT$(x, p) = 1$, so daß uns der erweiterte EUKLIDische Algorithmus eine Darstellung $ux + vp = 1$ liefert, wobei wir erreichen können, daß $1 \leq u < p$ ist. Dann ist u das inverse Element zu x. Für jedes nicht durch p teilbare $a \in \mathbb{Z}$ ist a mod p ein Element dieser Gruppe, und die Potenzen von a bilden eine Untergruppe: $1 = a^0$, und ist r die kleinste natürliche Zahl, für die $a^r \equiv 1 \bmod p$ ist, so gibt es genau r verschiedene Potenzen $a^0, \ldots, a^{r-1}$, und das inverse Element zu a^i ist a^{r-i}. Also teilt r die Gruppenordnung $p - 1$. Wegen $a^r \equiv 1 \bmod p$ muß daher auch $a^{p-1} \equiv 1 \bmod p$ gelten.

Ist V ein Vektorraum und $U \leq V$ ein Untervektorraum, so können wir die Menge aller Nebenklassen (bei denen wir hier wegen der Kommutativität der Vektoraddition nicht zwischen links und rechts unterscheiden müssen) wieder zu einem Vektorraum machen mit der Addition $(v + U) + (w + U) = (v + w) + U$ Wir wollen uns überlegen, ob wir auch die Menge aller Links- oder Rechtsnebenklassen einer Untergruppe zu einer Gruppe machen können. Da beide Fälle völlig analog zueinander sind, beschränken wir uns auf Linksnebenklassen.

Falls diese eine Gruppe bilden bezüglich des Komplexprodukts, muß es zu $a, b \in G$ ein $c \in G$ geben, so daß $(aU)(bU) = cU$ ist. Da a in aU liegt und b in bU, liegt ab in cU, d.h. $cU = (ab)U$. Somit ist $(aU)(bU) = (ab)U$. Zu beliebigen Elementen $u, v \in U$ muß es also ein $w \in U$ geben, so daß $(au)(bv) = (ab)w$ ist. Speziell für $v = 1$ folgt, daß es zu jedem $u \in U$ ein $w \in U$ geben muß, so daß $aub = abw$ ist. Multiplizieren wir dies von links mit a^{-1}, folgt daß $ub = bw$ sein muß. Für jedes $u \in U$ muß es also ein $w \in U$ geben, so daß $ub = bw$ ist, d.h. die Linksnebenklasse von b muß gleich der Rechtsnebenklasse von b sein.

Wir können die Gleichung $ub = bw$ durch Linksmultiplikation mit b^{-1} auch umschreiben zu $b^{-1}ub = w$; damit haben wir die Bedingung, daß jedes Element von U durch Konjugation mit einem beliebigen $b \in G$ wieder auch ein Element von U abgebildet wird, daß U also unter der Konjugation mit b auf sich selbst abgebildet wird.

Dies gilt aber nicht für jede Untergruppe: Nehmen wir etwa für G die Gruppe $\mathfrak{S}_4$ aller bijektiver Abbildungen der Menge $\{1, 2, 3, 4\}$ auf sich selbst und für U die Untergruppe, die 4 festläßt, also im wesentlichen die Gruppe $\mathfrak{S}_3$ der Abbildungen von $\{1, 2, 3\}$ auf sich selbst, so liegt der Dreierzyklus (1 2 3) natürlich in der Untergruppe, aber sein Konjugiertes unter der Transposition (1 4), die Permutation

$$(1\ 4)^{-1}(1\ 2\ 3)(1\ 4) = (1\ 4)(1\ 2\ 3)(1\ 4) = (1\ 4\ 2\ 3)(1\ 4) = (2\ 3\ 4)$$

liegt nicht in U.

Die Menge der Links- oder Rechtsnebenklassen von $U \leq G$ kann also höchstens dann mit der offensichtlichen Verknüpfung zu einer Gruppe

gemacht werden, wenn die Linksnebenklasse eines jeden Elements gleich seiner Rechtsnebenklasse ist oder, äquivalent, wenn jeder innere Automorphismus von G die Untergruppe U auf sich selbst abbildet.

In diesem Fall bilden die Nebenklassen auch tatsächlich eine Gruppe bezüglich des Komplexprodukts, denn

$$(aU)(bU) = (aU)(Ub) = a\big((UU)b\big) = a(Ub) = a(bU) = (ab)U\ .$$

Definition: *a)* Eine Untergruppe $N \leq G$ einer Gruppe G heißt *Normalteiler* von G, in Zeichen $N \trianglelefteq G$, wenn für jedes $n \in N$ und jedes $g \in G$ das konjugierte Element $n^g = g^{-1}ng$ in N liegt.
b) Die von den Nebenklassen $gN = Ng$ gebildete Gruppe heißt *Faktorgruppe* und wird mit G/N bezeichnet.

In einer abelschen Gruppe ist stets $g^{-1}ng = g^{-1}gn = n$, so daß alle Untergruppen Normalteiler sind. Insbesondere sind daher alle Untergruppen der additiven Gruppe $\mathbb{Z}$ der ganzen Zahlen Normalteiler. Wie wir oben gesehen haben, ist jede Untergruppe von der Form $m\mathbb{Z}$ mit einem $m \in \mathbb{N}_0$; mit den Faktorgruppen $\mathbb{Z}/m\mathbb{Z}$ werden wir es noch häufiger zu tun haben. Wir vereinbaren deshalb die Abkürzung

$$\mathbb{Z}/m \underset{\text{def}}{=} \mathbb{Z}/m\mathbb{Z}\,.$$

In einigen Büchern wird auch einfach $\mathbb{Z}_m$ geschrieben; das möchte ich vermeiden, da es im Primzahlfall zu Verwechslungen mit den sogenannten p-adischen Zahlen führen kann.

In der nichtabelschen Gruppe $\mathfrak{S}_4$ ist die Untergruppe U bestehend aus allen Permutationen, die die Vier fest lassen, kein Normalteiler. Die alternierende Gruppe $\mathfrak{A}_n$ dagegen ist für jedes n ein Normalteiler von $\mathfrak{S}_n$: Bekanntlich läßt sich jede Permutation als Produkt von Transpositionen schreiben. Diese Darstellung ist zwar nicht eindeutig, aber für jede Darstellung einer festen Permutation ist die Anzahl der Faktoren entweder immer gerade oder immer ungerade. Die Permutation heißt im ersten Fall gerade, im zweiten ungerade, und die alternierende Gruppe $\mathfrak{A}_n$ besteht aus allen geraden Permutationen.

Für eine Permutation $\omega \in \mathfrak{A}_n$ und eine beliebige Permutation $\pi \in \mathfrak{S}_n$ läßt sich ω (nur) als ein Produkt aus einer geraden Anzahl von Transpositionen schreiben, und ist $\pi = \tau_1 \cdots \tau_r$ eine Darstellung von π als

Produkt von Transpositionen, so ist $\pi^{-1} = \tau_r \cdots \tau_1$, das heißt $\pi^{-1}\omega\pi$ hat eine Darstellung als Produkt einer geraden Anzahl von Transpositionen, so daß $\pi^{-1}\omega\pi$ in $\mathfrak{A}_n$ liegt.

Lemma: Der Kern eines jeden Homomorphismus $\varphi\colon G \to H$ ist ein Normalteiler von G, und umgekehrt gibt es auch zu jedem Normalteiler $N \trianglelefteq G$ einen Homomorphismus $\varphi\colon G \to H$ mit Kern N.

Beweis: Ist $n \in \operatorname{Kern}\varphi$ und $g \in G$, so ist

$$\begin{aligned}\varphi\big(n^g\big) = \varphi\big(g^{-1}ng\big) &= \varphi\big(g^{-1}\big)\varphi(n)\varphi(g) = \varphi\big(g^{-1}\big)e\varphi(g)\\ &= \varphi\big(g^{-1}\big)\varphi(g) = \varphi\big(g^{-1}g\big) = \varphi(e') = e\,,\end{aligned}$$

wobei $e' \in G$ das Neutralelement von G bezeichnet. Somit liegt n^g in $\operatorname{Kern}\varphi$, was die Normalität beweist.

Umgekehrt ist jeder Normalteiler $N \trianglelefteq G$ beispielsweise Kern der Restklassenabbildung $\varphi\colon G \to G/N$, die jedem Element $x \in G$ seine Nebenklasse xN zuordnet. ∎

Die Normalteiler einer Gruppe G sind somit genau die Kerne der von G ausgehenden Homomorphismen. Damit läßt sich etwas einfacher zeigen, daß $\mathfrak{A}_n$ ein Normalteiler von $\mathfrak{S}_n$ ist, denn $\mathfrak{A}_n$ ist der Kern der Signaturabbildung $\varepsilon\colon \mathfrak{S}_n \to \{+1, -1\}$, die jede gerade Permutation auf +1 und jede ungerade Permutation aus -1 abbildet.

Wie für Vektorräume gilt auch für Gruppen ein

Homomorphiesatz: Für jedem Homomorphismus $\varphi\colon G \to H$ ist

$$G/\operatorname{Kern}\varphi \cong \operatorname{Bild}\varphi\,.$$

Beweis: Sei $N = \operatorname{Kern}\varphi$. Wir definieren eine Abbildung $\overline{\varphi}$ von G/N nach H, die die Nebenklasse Nx auf $\varphi(x)$ abbildet. Sie ist wohldefiniert, denn ist $Nx = Ny$, so liegt y in Nx, läßt sich also in der Form $y = nx$ schreiben mit $n \in N = \operatorname{Kern}\varphi$. Somit ist

$$\varphi(y) = \varphi(nx) = \varphi(n)\varphi(x) = e'\varphi(x) = \varphi(x)\,.$$

Da φ ein Homomorphismus ist, ist auch $\overline{\varphi}$ einer, und $\overline{\varphi}$ ist injektiv, denn ist $\overline{\varphi}(Nx) = \overline{\varphi}(Ny)$, so ist $\varphi(x) = \varphi(y)$, also $\varphi(xy^{-1}) = e'$. Also ist

$xy^{-1} \in N$ und damit $x \in Ny$, d.h. $Nx = Ny$. Das Bild von $\overline{\varphi}$ ist natürlich gleich dem von φ; schränken wir $\overline{\varphi}$ ein zu einer Abbildung von G/N nur nach Bild φ statt nach ganz H, haben wir daher einen bijektiven Homomorphismus, d.h. einen Isomorphismus. ∎

Nach diesen vielen Begriffen, Lemmata und Sätzen wird es Zeit, endlich mehr Beispiele von Gruppen zu betrachten. Am einfachsten sind Gruppen, die von einem Element erzeugt werden:

Definition: Eine Gruppe G heißt *zyklisch,* wenn es ein $g \in G$ gibt, so daß sich jedes Element $x \in G$ als $x = g^n$ schreiben läßt mit einem $n \in \mathbb{Z}$.

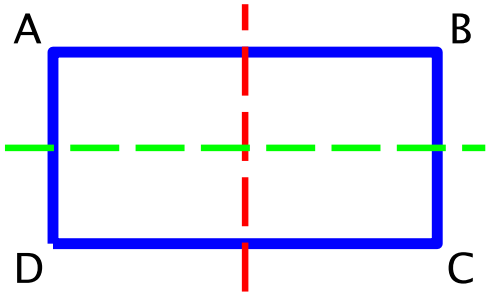
Lemma: Eine zyklische Gruppe G ist entweder isomorph zu $\mathbb{Z}$ oder es gibt ein $m \in \mathbb{N}$, so daß $G \cong \mathbb{Z}/m$.

Beweis: Wir betrachten die Abbildung

$$\varphi: \begin{cases} \mathbb{Z} \to G \\ n \mapsto g^n \end{cases}$$

Nach Definition einer zyklischen Gruppe ist sie surjektiv, und auf Grund des allgemeinen Assoziativgesetzes ist sie auch ein Homomorphismus. Falls sie auch injektiv ist, folgt $G \cong \mathbb{Z}$; andernfalls ist ihr Kern als Untergruppe von $\mathbb{Z}$ von der Form Kern $\varphi = m\mathbb{Z}$ mit einem $m \in \mathbb{N}$. Nach dem Homomorphiesatz ist dann $G \cong \mathbb{Z}/m\mathbb{Z} = \mathbb{Z}/m$. ∎

Interessante Beispiele von Gruppen liefern auch die Symmetrien geometrischer Objekte. Betrachten wir etwa ein Rechteck mit Ecken (im Uhrzeigersinn) A, B, C, D.



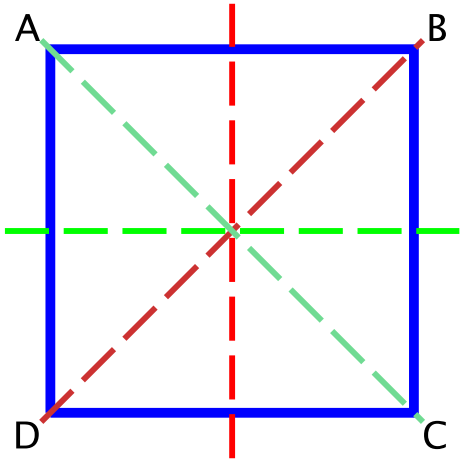
Abgesehen von der identischen Abbildung id gibt es drei Symmetrien: Die Spiegelung σ an der gemeinsamen Mittelsenkrechten der Strecken $\overline{AB}$ und $\overline{DC}$, die A mit B und C mit D vertauscht, die Spiegelung τ an der gemeinsamen Mittelsenkrechten von $\overline{AD}$ und $\overline{BC}$, die A mit D und B mit C vertauscht, und die Punktspiegelung ρ am Mittelpunkt des Rechtecks, die die gegenüberliegenden Ecken A, C sowie B, D miteinander vertauscht.

Offensichtlich ist $\sigma^2 = \tau^2 = \rho^2 = \mathrm{id}$; die Gruppe ist also nicht zyklisch, da es kein Element mit vier verschiedenen Potenzen gibt. Indem man den Effekt auf die Ecken betrachtet, rechnet man auch leicht nach, daß $\sigma\tau = \tau\sigma = \rho$ ist, und daraus folgen die Gleichungen $\rho\sigma = \sigma\rho = \tau$ und $\rho\tau = \tau\rho = \sigma$. Die Verknüpfung in dieser Gruppe ist also durch die folgende Tafel gegeben:

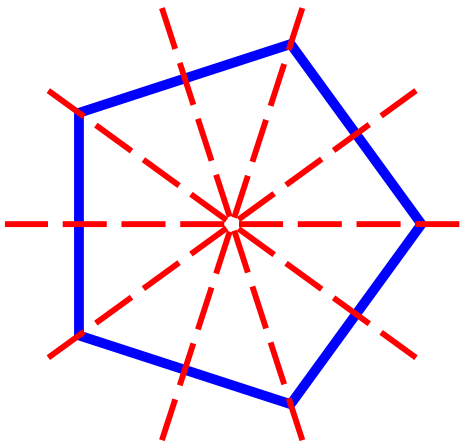
	id	σ	τ	ρ
id	id	σ	τ	ρ
σ	σ	id	ρ	τ
τ	τ	ρ	id	σ
ρ	ρ	τ	σ	id

Die Gruppe heißt KLEINsche Vierergruppe V_4 nach dem deutschen Mathematiker FELIX KLEIN (1849–1925), der sich intensiv mit diskreten Symmetriegruppen beschäftigt hatte. Die Mengen $\{\sigma, \mathrm{id}\}$, $\{\tau, \mathrm{id}\}$ und $\{\rho, \mathrm{id}\}$ sind Untergruppen und, da die Gruppe abelsch ist, auch Normalteiler von V_4. Weitere Untergruppen, abgesehen von den trivialen Untergruppen $\{\mathrm{id}\}$ und V_4 selbst, gibt es nicht, denn jede solche Untergruppe muß nach LAGRANGE die Ordnung zwei haben, besteht also aus der Identität und einem weiteren Element, dessen Quadrat die Identität ist.

Falls das Rechteck ein Quadrat ist, gibt es noch weitere Symmetrien, zum Beispiel die Drehung δ um 90° und ihre Potenzen. δ^2 ist die Punktspiegelung ρ, und $\delta^3 = \delta^{-1}$ ist die Drehung um -90°. Außerdem gibt es noch die Spiegelungen an den beiden Diagonalen, so daß die Gruppe acht Elemente enthält.



Allgemein wird die Symmetriegruppe des regelmäßigen n-Ecks als *Diedergruppe* D_n bezeichnet; die Symmetriegruppe des Quadrats ist also die Diedergruppe D_4. (Ein Polyeder ist es Körper, der von ebenen Polygonen begrenzt wird, der Würfel als Hexaeder etwa von sechs Quadraten und das Oktaeder von acht Dreiecken. Bei nur zwei Flächen entsteht etwas ebenes, ein *Dieder*, gesprochen Di-eder.) Für gerade n sind die Spiegelungsachsen Mittelsenkrechte von Kanten oder Verbindungsgeraden gegenüberliegender Ecken; für ungerade n sind es Geraden durch eine Ecke und den Mittelpunkt der gegenüberliegenden Kante.



Zur Untersuchung der allgemeinen Diedergruppen, identifizieren wir die reelle Ebene mit der komplexen Zahlenebene und betrachten das regelmäßige n-Eck mit Ecken

$$E_k = e^{2k\pi i/n}, \quad k = 0, \ldots, n-1 .$$

Die Drehung δ um den Winkel $360^\circ/n$ können wir dann beschreiben durch die Multiplikation mit $\zeta = e^{2\pi i/n}$, seine Potenzen durch die mit $\zeta^k = e^{2k\pi i/n}$. Wenn n gerade ist, gehört dazu für $k = n/2$ auch die Multiplikation mit $e^{\pi i} = -1$, also die Punktspiegelung am Nullpunkt. Ansonsten haben für gerade $n = 2m$ noch die Spiegelungen an den Geraden durch zwei gegenüberliegende Ecken E_k und E_{k+m}, $k = 0, \ldots, m-1$, und die Spiegelungen an den Geraden durch die Mittelpunkte zweier gegenüberliegender Seiten. Die Steigungswinkel der Geraden durch zwei gegenüberliegende Ecken sind $k \cdot 360^\circ/n$, die durch zwei Kantenmittelpunkte liegen jeweils dazwischen und haben daher die Steigungen $(k + \frac{1}{2}) \cdot 360^\circ/n$. Insgesamt haben wir somit Spiegelungen σ_k an den n Geraden mit Steigungswinkeln $k \cdot 180^\circ/n$ für $k = 0, \ldots, n-1$. Für ungerade n haben wir keine Punktsymmetrie, aber auch Spiegelungen an den Geraden der gerade aufgelisteten Steigungswinkel; der einzige Unterschied zum geraden Fall besteht darin, daß nun jede dieser Geraden durch eine Ecke und den gegenüberliegenden Kantenmittelpunkt geht. Insgesamt gibt es also in beiden Fällen $2n$ Symmetrieoperationen.

Die Potenzen der Drehung δ bilden eine zyklische Untergruppe der Ordnung n; diese ist ein Normalteiler. Dies kann man leicht direkt nachrechnen durch Konjugation einer Drehung mit einer Spiegelung; es geht aber auch einfacher ganz allgemein und abstrakt:

Lemma: Ist G eine (nicht notwendigerweise endliche) Gruppe und U eine Untergruppe vom Index zwei, so ist U ein Normalteiler.

Beweis: Da $[G : U] = 2$ ist, hat U genau zwei Nebenklassen. Eine davon ist natürlich U selbst, und die andere besteht aus den übrigen Elementen von G. Für jedes $x \notin U$ ist daher $Ux = xU = G \setminus U$, und damit ist U Normalteiler. ∎

Daraus folgt auch noch einmal, daß die alternierende Gruppe $\mathfrak{A}_n$ ein Normalteiler von $\mathfrak{S}_n$ ist.

Da jede Spiegelungen zu sich selbst invers ist, bildet jede von ihnen zusammen mit der Identität eine Untergruppe der Ordnung zwei. Diese Untergruppen sind keine Normalteiler: Konjugieren wir etwa σ_0 mit δ,

so drehen wir zunächst um den Winkel $360^\circ/n$, ersetzen also einen Punkt $z \in \mathbb{C}$ durch ζz. Danach wird an der reellen Achse gespiegelt; dies entspricht der komplexen Konjugation, und dann wird für die Drehung δ^{-1} mit ζ^{-1} multipliziert. Insgesamt wird also z abgebildet auf $\zeta^{-1} \cdot \overline{\zeta z} = \zeta^{-1}\overline{\zeta}\overline{z}$. Wegen $\zeta\overline{\zeta} = |\zeta|^2 = 1$ ist $\overline{\zeta} = \zeta^{-1}$, also geht z insgesamt auf $\zeta^{-2}\overline{z}$. Wegen der komplexen Konjugation kann dies keine Drehung sein, also ist es eine der Spiegelungen σ_k. Da bei einer Spiegelung genau die Punkte auf der Achse fest bleiben, können wir die Achse leicht bestimmen: Für ein z vom Betrag eins ist $\overline{z} = z^{-1}$, also $\zeta^{-2}\overline{z} = z$ genau dann, wenn $\zeta^{-2} = z^2$ oder $z = \pm\zeta^{-1}$ ist. Die Fixgerade ist also die Gerade durch E_{n-1} und den gegenüberliegenden Punkt, d.h. $\delta^{-1}\sigma_0\delta = \sigma_{n-2} \neq \sigma_0$.

Da eine Symmetrieoperation auf einem n-Eck durch die Bilder der Ecken eindeutig bestimmt ist, gibt es eine natürlichen Monomorphismus φ von D_n in die symmetrische Gruppe $\mathfrak{S}_n$: Falls $\sigma \in D_n$ die Ecke E_k abbildet auf $E_{\pi(k)}$, ist $\varphi(\sigma) = \pi$. Für $n = 3$ ist φ ein Isomorphismus: Die drei Spiegelungen σ_k vertauschen jeweils zwei Ecken und lassen eine fest, werden also auf die drei Transpositionen aus $\mathfrak{S}_n$ abgebildet, und die Drehungen um $\pm 120^\circ$ gehen auf die beiden Dreierzykel; die Identität geht natürlich auf die Identität.

Für $n = 4$ haben wir keinen Isomorphismus mehr, denn $\mathfrak{S}_4$ hat 24 Elemente, D_4 aber nur acht. In der Tat gibt es keine Symmetrie eines Quadrats, die zwei benachbarte Ecken eines Quadrats vertauscht, die beiden anderen aber festläßt; von den $\binom{4}{2} = 6$ Transpositionen liegen also nur die beiden im Bild, die zwei gegenüberliegende Ecken vertauschen und den Rest fest lassen. Sie sind die Bilder der Spiegelungen an den beiden Diagonalen; wenn wir die Ecken A, B, C, D mit den Zahlen $1, 2, 3, 4$ identifizieren, sind das die Transpositionen (1 3) und (2 4). Auch die acht Dreierzykel lassen sich nicht als Symmetrieoperationen eines Quadrats realisieren, denn sie lassen genau eine Ecke fest, während Spiegelungen entweder keine oder zwei Ecken festlassen und Drehungen keine. Von den drei Produkten zweier elementfremder Transpositionen entsprechen (1 2)(3 4) und (1 4)(2 3) den beiden Spiegelungen an den Mittelsenkrechten, und (1 3)(2 4) der Drehung um 180°, die den gleichen Effekt hat wie die Punktspiegelung am Mittelpunkt. Die

Drehungen um 90° und 270° werden auf die beiden (zueinander inversen) Vierzykeln (1 2 3 4) und (1 4 3 2) abgebildet.

Es ist keine spezielle Eigenschaft der Gruppe D_n, daß sie in eine symmetrische Gruppe eingebettet werden kann:

Lemma: Für jede endliche Gruppe G gibt es einen Monomorphismus φ von G in eine symmetrische Gruppe $\mathfrak{S}_n$.

Beweis: Wir nummerieren die Elemente von G in irgendeiner Weise; für $|G| = n$ sei also $G = \{g_1, \ldots, g_n\}$. Für jedes Element $x \in G$ ist die Multiplikation mit x bijektiv, es gibt also eine Permutation $\pi_x \in \mathfrak{S}_n$, so daß $xg_i = g_{\pi_x(i)}$ ist. Die Abbildung $\varphi\colon G \to \mathfrak{S}_n$, die jedes $x \in G$ auf $\pi_x \in \mathfrak{S}_n$ abbildet, ist natürlich injektiv. Sie ist auch ein Homomorphismus, denn

$$g_{\pi_{xy}(i)} = (xy)g_i = x(yg_i) = xg_{\pi_y(i)} = g_{\pi_x(\pi_y(i))} = g_{\pi_x \circ \pi_y(i)}\,,$$

d.h. $\pi_{xy} = \pi_x \circ \pi_y$, und damit ist $\varphi(xy) = \varphi(x) \circ \varphi(y) = \varphi(x)\varphi(y)$. ∎

Wie das Beispiel der Diedergruppen zeigt, kann man zumindest gelegentlich auch in eine $\mathfrak{S}_n$ einbetten, bei der n kleiner ist als die Gruppenordnung; das Extrembeispiel ist natürlich die symmetrische Gruppe $\mathfrak{S}_n$ selbst, die wir nicht erst in $\mathfrak{S}_{n!}$ einbetten müssen.

Die Einbettbarkeit einer beliebigen endlichen Gruppe in eine symmetrische Gruppe ist nicht nur theoretisch interessant: In der symmetrischen Gruppe können wir explizit rechnen nach einfachen Regeln, die wir auch einem Computer beibringen können. Sobald wir eine Gruppe also in eine $\mathfrak{S}_n$ eingebettet haben, können wir dort beliebige Rechnungen per Computer ausführen. Die Einbettungen einer endlichen Gruppe G in symmetrische Gruppen bezeichnet man als ihre *Permutationsdarstellungen;* Computeralgebrasysteme wie Maple benutzen unter anderen diese, um konkret mit Gruppen zu rechnen.

Wenn wir bei den regelmäßigen n-Ecken n gegen unendlich gehen lassen, bekommen wir einen Kreis. Diesen können wir zunächst einmal selbst als Gruppe betrachten: Wenn wir ihn in die komplexe Zahlenebene einbetten als

$$\mathbb{S}^1 = \left\{ z \in \mathbb{C} \;\middle|\; |z| = 1 \right\},$$

ist er offensichtlich eine Gruppe bezüglich der Multiplikation (und hat für jedes n die Menge der Ecken E_k des regelmäßigen n-Ecks mit der obigen Einbettung als Untergruppe). Diese Gruppe operiert durch Multiplikation auf der *Menge* $\mathbb{S}^1$ dadurch, daß das Element z der Gruppe das Element w der Menge auf zw abbildet. Für $z = e^{i\varphi}$ und $w = e^{i\psi}$ bedeutet das geometrisch, daß der Kreispunkt mit Winkel ψ bezüglich der x-Achse um den Winkel φ gedreht wird. Diese Drehungen sind aber nicht die einzigen Symmetrieoperationen auf der Kreislinie: Genau wie bei regelmäßigen n-Eck haben wir auch noch Spiegelungen, und zwar hier an jeder Geraden durch den Nullpunkt. Wir haben also unendlich viele Untergruppen der Ordnung zwei, die genau wie im Fall der Diedergruppen keine Normalteiler sind. Die Gruppe aller Drehungen ist auch hier ein Normalteiler, da sie Index zwei hat: Eine orientierungserhaltende Bewegung, die den Kreis invariant läßt, muß eine Drehung sein, (Die Punktspiegelung am Mittelpunkt ist die Drehung um 180°.) Ist also σ eine beliebige Drehung und τ eine Symmetrieoperation des Kreises, die die Orientierung nicht erhält, so ist $\sigma\tau$ orientierungserhaltend, also eine Drehung. Damit hat auch hier die Gruppe $\mathbb{S}^1$ der Drehungen den Index zwei, ist also Normalteiler. In Analogie zu den Gruppen D_n wird die Symmetriegruppe des Kreises mit D_∞ bezeichnet.

Die Lineare Algebra liefert eine ganze Reihe von Beispielen für Gruppen, vor allem als Teilmengen der Menge $k^{n\times n}$ der $n \times n$-Matrizen über einem Körper k. Am bekanntesten sind die allgemeine *(general)* lineare Gruppe

$$\mathrm{GL}_n(k) \underset{\text{def}}{=} \{A \in k^{n\times n} \mid \det A \neq 0\}$$

und die spezielle lineare Gruppe

$$\mathrm{SL}_n(k) \underset{\text{def}}{=} \{A \in k^{n\times n} \mid \det A = 1\}\,.$$

Natürlich ist $\mathrm{SL}_n(k)$ eine Untergruppe von $\mathrm{GL}_n(k)$; als Kern der Determinantenabbildung ist sie sogar ein Normalteiler, denn nach dem Multiplikationssatz für Determinanten ist diese Abbildung ein Homomorphismus.

Ist G eine endliche Gruppe, so gibt es auch einen Monomorphismus von G in eine Gruppe $\mathrm{GL}_n(k)$: Im einfachsten Fall können wir n gleich

der Gruppenordnung von G nehmen, die Gruppenelemente wieder irgendwie als $g_1, \ldots, g_n$ durchnummerieren und g_i abbilden auf die Matrix der folgenden linearen Abbildung φ_i: Ist $g_i g_j = g_\ell$, so soll $\varphi_i(e_j) = e_\ell$ sein, wobei $e_1, \ldots, e_n$ die Standardbasis von k^n ist. Man beachte, daß diese Matrix eine Permutationsmatrix ist; wenn wir $\mathfrak{S}_n$ identifizieren mit der Untergruppe aller Permutationsmatrizen in $\mathrm{GL}_n(k)$ haben wir also wieder eine der oben betrachteten Permutationsdarstellungen. Allgemein bezeichnet man Homomorphismen $G \to \mathrm{GL}_n(k)$ als *lineare Darstellungen* der Gruppe G; die meisten dieser Darstellungen lassen sich nicht als Permutationsdarstellungen interpretieren.

Die sogenannte *Darstellungstheorie* beschäftigt sich mit der systematischen Untersuchung der linearen Darstellungen einer Gruppe. Sie ist ein wichtiges Teilgebiet der Gruppentheorie; mit ihr kann man Strukturaussagen über Gruppen beweisen, die sich mit anderen Auch zum Rechnen in Gruppen sind lineare Darstellungen nützlich. Methoden nur schwerer oder überhaupt nicht beweisen lassen. Oft reichen bereits die einfacher zu untersuchenden *Charaktere* der linearen Darstellungen, d.h. man betrachtet an Stelle der Darstellungsmatrizen nur deren Spuren.

Fast alle hier betrachteten Gruppen wurden so definiert, daß ihre Elemente als Operationen auf einer Menge aufgefaßt werden können: bei der D_n auf der EUKLIDischen Ebenen, bei $\mathrm{GL}_n(k)$ und $\mathrm{SL}_n(k)$ auf k^n. Wir definieren allgemein:

Definition: *a)* Eine *Operation* einer Gruppe G auf einer Menge M ist eine Abbildung

$$\begin{cases} G \times M \to M \\ (g, m) \mapsto g(m) \end{cases},$$

für die gilt: $g\big(h(m)\big) = (gh)(m)$ für alle $g, h \in G$ und $m \in M$ und $e(m) = m$ für das Neutralelement $e \in G$ und alle $m \in M$.
b) Die *Bahn* eines Element $m \in M$ ist die Menge

$$O(m) \underset{\text{def}}{=} \{g(m) \mid g \in G\}\,.$$

c) Der *Stabilisator* eines Elements $m \in M$ ist

$$\mathrm{Stab}(m) \underset{\text{def}}{=} \{g \in G \mid g(m) = m\}\,.$$

Für *Operation* sind auch die Synonyme *Aktion* oder *Wirkung* gebräuchlich; statt *Bahn* sagt man auch *Orbit.*

Der Stabilisator eines Element $m \in M$ ist eine Untergruppe von g, denn für zwei Elemente $g, h \in \text{Stab}(m)$ ist $g\big(h(m)\big) = g(m) = m$ und

$$g^{-1}(m) = g^{-1}\big(g(m)\big) = (gg^{-1})(m) = e(m) = m\,.$$

Für zwei Elemente $g, h \in G$ ist $g(m) = h(m)$ genau dann, wenn $(h^{-1}g)(m) = m$ ist, wenn also $h^{-1}g$ im Stabilisator von m liegt. Dies ist genau dann der Fall, wenn h in der Nebenklasse $g\text{Stab}(m)$ liegt, wenn also die beiden Nebenklassen $g\text{Stab}(m)$ und $h\text{Stab}(m)$ übereinstimmen. Bezeichnen wir mit $G/\text{Stab}(m)$ die Menge aller Linksnebenklassen von G modulo $\text{Stab}(m)$, gilt daher:

Lemma: Für jedes $m \in M$ gibt es eine bijektive Abbildung

$$\begin{cases} G/\text{Stab}(m) \to O(m) \\ \quad g\text{Stab}(m) \mapsto g(m) \end{cases}.$$

Im Falle einer endlichen Gruppe G gilt somit die *Bahnbilanzgleichung* $|O(m)| = |G| \,/\, |\text{Stab}(m)|$.

■

§3: Ringe

Ein Ring ist eine algebraische Struktur mit zwei Rechenoperationen, von denen die eine als Addition und die andere als Multiplikation aufgefaßt wird. Wir fordern, daß wir bezüglich der Addition eine abelsche Gruppe haben und bezüglich der Multiplikation ein Monoid; außerdem sollen die üblichen Distributivgesetze gelten. Ausführlich aufgeschrieben:

Definition: Ein *Ring* ist eine Menge R zusammen mit zwei Verknüpfungen $+, \cdot : R \times R \to R$, für die gilt

1.) Bezüglich + ist R eine abelsche Gruppe.

2.) $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ für alle $x, y, z \in R$.

3.) Es gibt ein Element $1 \in R$, so daß $1 \cdot x = x \cdot 1 = x$ für alle $x \in R$.

4.) $x \cdot (y+z) = x \cdot y + x \cdot z$ und $(x+y) \cdot z = x \cdot z + y \cdot z$ für alle $x, y, z \in R$.

b) Der Ring heißt *kommutativ,* wenn zusätzlich noch gilt

5.) $x \cdot y = y \cdot x$ für alle $x, y \in R$.

c) Ein kommutativer Ring heißt *Körper,* wenn $R \setminus \{0\}$ bezüglich der Multiplikation eine Gruppe bildet, wenn also zusätzlich gilt

6.) Zu jedem $x \neq 0$ aus R gibt es ein $x' \in R$ gibt, so daß $x \cdot x' = 1$ ist.

d) Eine Abbildung $\varphi\colon R \to S$ zwischen zwei Ringen heißt (Ring-) *Homomorphismus,* wenn für alle $x, y \in R$ gilt

$$\varphi(x + y) = \varphi(x) + \varphi(y) \quad \text{und} \quad \varphi(x \cdot y) = \varphi(x) \cdot \varphi(y)\,,$$

wobei + und $\cdot$ auf der linken Seite jeweils die Operationen von R bezeichnen und rechts die von S.

e) Ein $\left\{\begin{array}{l}\textit{Monomorphismus}\\ \textit{Epimorphismus}\\ \textit{Isomorphismus}\end{array}\right\}$ ist ein $\left\{\begin{array}{l}\textit{injektiver}\\ \textit{surjektiver}\\ \textit{bijektiver}\end{array}\right\}$ Homomorphismus.

Zwei Ringe R und S heißen *isomorph,* in Zeichen $R \cong S$, wenn es einen Isomorphismus $\varphi\colon R \to S$ gibt.

f) Ist $R = S$, bezeichnen wir einen Homomorphismus von R nach R auch als *Endomorphismus* und einen Isomorphismus als *Automorphismus.*

g) Das *Bild* eines Homomorphismus $\varphi\colon R \to S$ ist

$$\operatorname{Bild}\varphi \underset{\text{def}}{=} \varphi(R) = \{\varphi(x) \mid x \in R\}\,;$$

sein *Kern* ist

$$\operatorname{Kern}\varphi \underset{\text{def}}{=} \{x \in R \mid \varphi(x) = 0\}\,.$$

Das bekannteste Beispiel eines Rings ist der Ring $\mathbb{Z}$ der ganzen Zahlen; er ist kommutativ. Ein Beispiel eines nichtkommutativen Rings bilden die $n \times n$-Matrizen über einem Körper (oder einem kommutativen Ring) k für $n \geq 2$.

Auch die additiven Gruppen $\mathbb{Z}/m$ können zu Ringen gemacht werden, indem wir als Multiplikation die Multiplikation ganzer Zahlen modulo m nehmen. Wir bezeichnen auch diesen Ring mit $\mathbb{Z}/m$.

Die Gleichung $0 \cdot x = 0$ für alle $x \in R$ gilt in jedem Ring, denn

$$x = 1 \cdot x = (1 + 0) \cdot x = 1 \cdot x + 0 \cdot x = x + 0 \cdot x \Longrightarrow 0 \cdot x = x - x = 0\,.$$

Genauso folgt auch $x \cdot 0 = 0$.

Falls $1 = 0$ sein sollte, besteht der ganze Ring nur aus der Null, denn für jedes x ist dann $x = 1 \cdot x = 0 \cdot x = 0$. In einem Körper muß $1 \neq 0$ sein, da sonst $R \setminus \{0\}$ die leere Menge wäre, und die ist keine Gruppe.

In Ringen muß es keine multiplikativen Inverse geben, eine Gleichung $ax = b$ mit $a, b \in R$ muß also keine Lösung haben. In $\mathbb{Z}$ ist sie genau dann lösbar, wenn a ein Teiler von b ist; dieses Konzept wollen wir auch auf andere Ringe verallgemeinern. Wenn wir eindeutige Lösungen wollen, können wir allerdings keine beliebigen Ringe zulassen: In $\mathbb{Z}/10$ hat etwa die Gleichung $2x \equiv 4 \bmod 10$ sowohl $x = 2$ als auch $x = 7$ als Lösungen. Der Grund liegt darin, daß $2 \cdot (7 - 2) = 2 \cdot 5 \equiv 0 \bmod 10$ ist, daß es also Elemente $a, b \neq 0$ gibt, deren Produkt verschwindet. Solche Elemente a, b bezeichnet man als *Nullteiler;* für eine Teilbarkeitstheorie, die dem entspricht, was wir von $\mathbb{Z}$ gewohnt sind, müssen wir die ausschließen.

Definition: Ein Ring heißt *nullteilerfrei* wenn gilt: Ist $x \cdot y = 0$, so muß mindestens einer der beiden Faktoren x, y verschwinden. Ein nullteilerfreier kommutativer Ring heißt *Integritätsbereich* (englisch *domain*).

In diesem Sinne ist also $\mathbb{Z}$ ein Integritätsbereich, erst recht natürlich auch jeder Körper. In einem Integritätsbereich hat die Gleichung $ax = b$ für $a \neq 0$ höchstens eine Lösung, denn ist $ax = ay$, so ist $a(y - x) = 0$, also $y - x = 0$ und $y = x$.

Der Kern eines Ringhomomorphismus ist im Allgemeinen kein Ring: Ansonsten müßte er insbesondere die Eins enthalten, und für jedes Element $x \in R$ ist dann $\varphi(x) = \varphi(1 \cdot x) = \varphi(1) \cdot \varphi(x) = 0 \cdot \varphi(x) = 0$, so daß φ die Nullabbildung sein muß. (Vor allem in der älteren Literatur verzichtet man aus diesem Grund bei der Definition eines Rings häufig auf die Forderung, daß es ein Neutralelement für die Multiplikation geben muß; dann bilden beispielsweise auch die geraden Zahlen einen Unterring von $\mathbb{Z}$. Da praktisch alle interessanten Beispiele von Ringen eine Eins enthalten, betrachten wir nur Ringe mit Eins.)

Liegen zwei Elemente x, y im Kern eines Homomorphismus $\varphi\colon R \to S$, so ist

$$\varphi(x+y) = \varphi(x)+\varphi(y) = 0+0 = 0 \quad \text{und} \quad \varphi(x \cdot y) = \varphi(x) \cdot \varphi(y) = 0 \cdot 0 = 0\,,$$

also liegen auch Summe und Produkt im Kern. Für das Produkt hätte es aber offensichtlich gereicht, wenn nur einer der beiden Faktoren im Kern liegt. Der Kern ist daher ein Ideal im Sinne der folgenden Definition:

Definition: Eine Teilmenge I eines Rings R heißt *Ideal* von R, in Zeichen $I \lhd R$, wenn I eine additive Untergruppe von R ist und wenn für alle $r \in R$ und $x \in I$ gilt: rx und xr liegen in I.

Für kommutative Ringe reicht natürlich eine der beiden Forderungen $rx \in I$ oder $xr \in I$. Bei nichtkommutativen Ringen betrachtet man auch Linksideale, bei denen nur rx in I liegen muß und Rechtsideale, bei denen di›es nur für xr der Fall sein muß; wenn – wie in obiger Definition gefordert – beides gilt, spricht man von einem beidseitigen Ideal.

Der Name *Ideal* geht auf KUMMER zurück, der für einen Beweis der FERMAT-Vermutung eindeutige Primzerlegung in Einheitswurzelringen benötigte. Da dies im allgemeinen nicht gilt, aber mit Idealen erreichbar ist, bezeichnete er diese als *ideale Zahlen.*

Da jedes Ideal eine additive Untergruppe ist, kommen in $\mathbb{Z}$ als Ideale nur die Mengen $m\mathbb{Z}$ mit $m \in \mathbb{N}_0$ in Frage, und die sind offensichtlich auch alle Ideale. Wenn wir sie als Ideale betrachten, schreiben wir meist (m) an Stelle von $m\mathbb{Z}$ gemäß der folgenden Konvention:

Definition: *a)* Für eine Teilmenge M eines Rings R bezeichnet (M) das kleinste Ideal von R, das M enthält. Für eine endliche Menge $M = \{a_1, \ldots, a_r\}$ schreiben wir $(M) = (a_1, \ldots, a_r)$.
b) Ein Ideal $I \trianglelefteq R$ eines Rings R heißt *Hauptideal,* wenn es ein $a \in R$ gibt, so daß $I = (a)$ ist.
c) Ein Integritätsbereich R heißt *Hauptidealring,* wenn jedes Ideal von R ein Hauptideal ist.

In diesem Sinne ist also $\mathbb{Z}$ ein Hauptidealring. Sind $a_1, \ldots, a_r$ ganze Zahlen, so wird das Ideal $(a_1, \ldots, a_r)$ erzeugt vom größten gemeinsamen Teiler der a_i, der in diesem Ideal liegt, weil er sich nach dem erweiterten EUKLIDischen Algorithmus als Linearkombination der a_i schreiben läßt. Dies legt nahe, daß der erweiterte EUKLIDische Algorithmus etwas mit Hauptidealringen zu tun haben könnte.

Der EUKLIDische Algorithmus beruht auf der Division mit Rest; wir definieren daher

Definition: Ein EUKLIDischer Ring ist ein Integritätsbereich R zusammen mit einer Abbildung $\nu\colon R \setminus \{0\} \to \mathbb{N}_0$, so daß gilt: Ist $x|y$, so ist

$\nu(x) \leq \nu(y)$, und zu je zwei Elementen $x, y \in R$ gibt es Elemente $q, r \in R$ mit

$$x = qy + r \quad \text{und} \quad r = 0 \quad \text{oder} \quad \nu(r) < \nu(y)\,.$$

Wir schreiben auch $x : y = q$ Rest r und bezeichnen r als Divisionsrest bei der Division von x durch y.

Das Standardbeispiel ist natürlich $R = \mathbb{Z}$ mit $\nu(z) = |z|$.

Erwartungsgemäß gilt

Lemma: Jeder EUKLIDische Ring ist ein Hauptidealring.

Beweis: $I \neq (0)$ sei ein Ideal von R, und M sei die Menge aller $\nu(f)$ für $f \in I \setminus \{0\}$. Das ist eine Teilmenge von $\mathbb{N}_0$; sie hat also ein kleinstes Element. Dieses sei gleich $\nu(f)$. Wir wollen uns überlegen, daß $I = (f)$ ist: Für ein beliebiges Element $g \in I$ können wir g mit Rest durch f dividieren, es also als $g = qf + r$ schreiben, wobei entweder $r = 0$ ist oder $\nu(r) < \nu(f)$. Letzteres ist wegen der Minimalität von $\nu(f)$ nicht möglich; also liegt $g = qf$ in (f). ∎

Die Umkehrung dieses Lemmas gilt nicht, allerdings sind Gegenbeispiele nicht einfach zu konstruieren, da die Funktion ν aus der Definition eines EUKLIDischen Rings a priori völlig beliebig sein kann und außerdem der einfachste Beweis, daß ein Ring Hauptidealring ist, meist darin besteht, zu zeigen, daß er EUKLIDisch ist. THEODORE MOTZKIN (1908–1970) gab 1949 den Ring $\mathbb{Z} \oplus \mathbb{Z}\omega$ mit $\omega = \frac{1}{2}(1 + \sqrt{-19})$ als Gegenbeispiel an in

> T. MOTZKIN: The Euclidian Algorithm, *Bulletin of the American Mathematical Society* **55** (1949), 1142–1146;

einen ausführlichen Beweis dafür, daß dies ein Hauptidealring ist, aber kein EUKLIDischer Ring, findet man in

> JACK C. WILSON: A principal ideal ring that is not a Euclidean ring, *Mathematics Magazine* **46** (1973), 34–38

Das Standardbeispiel eines EUKLIDischen Rings ist natürlich der Ring $\mathbb{Z}$ der ganzen Zahlen mit $\nu(x) = |x|$. Aus der Schule bekannt ist aber auch

die Division mit Rest bei Polynomen; hier definieren wir $\nu(g)$ für ein von Null verschiedenes Polynom als den Grad von g.

Polynome können wir wir nicht nur über den reellen Zahlen betrachten, sondern über beliebigen Ringen; hier wollen wir uns allerdings mit kommutativen Ringen begnügen:

Definition: R sei ein kommutativer Ring. Der *Polynomring* $R[X]$ ist die Menge aller (formaler) Summen

$$a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0$$

mit $d \in \mathbb{N}_0$ und $a_i \in R$ für alle i. Ist $a_d \neq 0$, bezeichnen wir $d = \deg f$ als den *Grad* von f. Addition und Multiplikation sind durch die üblichen Regeln definiert.

Offensichtlich ist auch $R[X]$ ein kommutativer Ring; wir können daher auch den Polynomring $R[X][Y]$ über $R[X]$ betrachten, den wir kurz als $R[X, Y]$ bezeichnen, und so weiter. Die Elemente des Polynomrings $R[X_1, \ldots, X_n]$ in n Variablen lassen sich schreiben als endliche Linearkombinationen sogenannter *Monome* $X_1^{e_1} \cdots X_n^{e_n}$ mit $e_1, \ldots, e_n \in \mathbb{N}_0$. Der Grad eines solchen Monoms ist die Summe $e_1 + \ldots + e_n$ der Exponenten; der Grad eines Polynoms ungleich dem Nullpolynom ist der größte Grad eines darin vorkommenden Monoms.

Wir können nicht erwarten, daß jeder Polynomring EUKLIDisch ist; im allgemeinen muß er schließlich nicht einmal nullteilerfrei sein. Immerhin gilt

Lemma: Ist R ein Integritätsbereich, so auch der Polynomring $R[X]$.

Beweis: Wir betrachten zwei Polynome

$$f = \sum_{i=0}^{d} a_i X^i \quad \text{und} \quad g = \sum_{j=0}^{e} b_j X^j \,,$$

die beide von Null verschieden sind. Wir können annehmen, daß d und e so gewählt sind, daß a_d und b_e beide nicht verschwinden. Da R

Integritätsbereich ist, kann dann auch das Produkt $a_d b_e$ nicht verschwinden, also ist der führende Term $a_d b_e X^{d+e}$ von fg von Null verschieden und damit auch fg selbst.

∎

Tatsächlich beweist dies sogar etwas mehr als die Nullteilerfreiheit, denn wir wissen nun, daß sich bei der Multiplikation zweier Polynome über einem Integritätsbereich die Grade addieren.

Auch der Polynomring über einem Integritätsbereich muß nicht EUKLIDisch sein; wie wir sehen werden, ist weder $\mathbb{Z}[X]$ noch ein Polynomring in mehr als einer Variablen EUKLIDisch. Es gilt aber

Lemma: Der Polynomring $k[X]$ über einem Körper k ist EUKLIDisch und damit auch ein Hauptidealring.

Beweis: Wir definieren $\nu(f)$ für ein Polynom $f \neq 0$ als den Grad von f; dann zeigt der Algorithmus zur Polynomdivision, daß es zu je zwei Polynomen $f, g \in k[X]$ mit $g \neq 0$ Polynome $q, r \in k[X]$ gibt mit $r = 0$ oder $\nu(r) < \nu(g)$ derart, daß $f = qg + r$ ist.

∎

Der EUKLIDische Algorithmus wird dazu verwendet, größte gemeinsame Teiler zu berechnen und als Linearkombination darzustellen; bevor wir das genauer untersuchen können, müssen wir erst definieren, was Teiler in einem beliebigen Ring sein sollen. Eine sinnvolle Theorie erhalten wir nur für Integritätsbereiche; daher wollen wir uns darauf beschränken.

Definition: R sei ein Integritätsbereich.
a) $u \in R$ heißt *Teiler* von $x \in R$, in Zeichen $u|x$, wenn es ein $q \in R$ gibt, so daß $x = q \cdot u$.
b) $u \in R$ heißt *größter gemeinsamer Teiler* von x und y, wenn u Teiler von x und von y ist und wenn für jeden anderen gemeinsamen Teiler v von x und y gilt: $v|u$.
c) Ein Element $e \in R$ heißt *Einheit,* falls es ein $e' \in R$ gibt, so daß $e \cdot e' = 1$ ist. Die Menge aller Einheiten von R bezeichnen wir mit $R^\times$.
d) Zwei Elemente $x, y \in R$ heißen *assoziiert,* wenn es eine Einheit $e \in R$ gibt mit $y = e \cdot x$.

Mit Idealen ausgedrückt ist $u|x$ äquivalent zu $(x) \subseteq (u)$, und x, y sind genau dann assoziiert, wenn sie das gleiche Hauptideal erzeugen.

Für $R = \mathbb{Z}$ ist wie gewohnt fünf ein Teiler von zehn, aber auch minus fünf ist einer. Entsprechend sind sowohl vier als auch minus vier größte gemeinsame Teiler von zwölf und zwanzig. Die Einheiten sind eins und minus eins; zwei ganze Zahlen sind genau dann assoziiert, wenn sie sich höchstens im Vorzeichen unterscheiden.

Ist $R = k$ ein Körper, so ist jedes $x \in R \setminus \{0\}$ eine Einheit und teilt jedes Element von k. Alle von Null verschiedene Elemente sind assoziiert.

Für jeden Ring R gilt

Lemma: *a)* Die Menge $R^\times$ aller Einheiten eines Rings R bildet eine Gruppe bezüglich der Multiplikation.
b) Ein kommutativer Ring R ist genau dann ein Integritätsbereich, wenn die folgende ***Kürzungsregel*** erfüllt ist: Gilt für $x, y, z \in R$ mit $z \neq 0$ die Gleichung $xz = yz$, so ist $x = y$.
c) Zwei Elemente x, y eines Integritätsbereich R sind genau dann assoziiert, wenn $x|y$ und $y|x$.
d) Ein größter gemeinsamer Teiler, so er existiert, ist bis auf Assoziiertheit eindeutig bestimmt.

Beweis: a) Sind $e, f \in R$ Einheiten, so gibt es inverse Elemente e', f'. Mit denen ist $(ef)(f'e') = e(ff')e' = ee' = 1$, d.h. auch ef ist eine Einheit. Außerdem ist jede Einheit nach Definition invertierbar, und 1 ist eine Einheit.
b) Ist R ein Integritätsbereich und $xz = yz$, so ist $(x - y)z = 0$; da $z \neq 0$ vorausgesetzt war, folgt $x - y = 0$, also $x = y$. Folgt umgekehrt aus $xz = yz$ und $z \neq 0$ stets $x = y$, so ist R nullteilerfrei, denn ist $xz = 0$ und $z \neq 0$, so ist $xz = 0z$, also $x = 0$.
c) Ist $y = ex$, so ist entweder $x = y = 0$, oder beide sind von Null verschieden und x teilt y. Da Einheiten invertierbar sind, ist auch $x = e^{-1}y$, d.h. $y|x$.
Gilt umgekehrt $x|y$ und $y|x$, so gibt es Elemente q, r mit $x = qy$ und $y = rx$. Damit ist $1x = x = qy = q(rx) = (qr)x$, also $qr = 1$, da $x \neq 0$. Somit ist q eine Einheit.

d) Sind u, v zwei größte gemeinsame Teiler von x, y, so ist nach Definition u Teiler von v und v Teiler von u, also sind u und v assoziiert. ∎

Schauen wir uns an, was das für einen Polynomring bedeutet!

Lemma: Die Einheiten im Polynomring $R[X]$ über einem Integritätsbereich R sind genau die Einheiten von R.

Beweis: Ist $f \in R[X]$ eine Einheit, so gibt es ein $g \in R[X]$ mit $fg = 1$; da das konstante Polynom 1 den Grad Null hat, muß dasselbe auch für f und g gelten, d.h. $f, g \in R$ und damit in $R^\times$. ∎

Für Polynome über einem Körper bedeutet dies, daß zwei Polynome genau dann assoziiert sind, wenn sie sich durch eine multiplikative Konstante ungleich Null unterscheiden; wenn es einen größten gemeinsamen Teiler gibt, ist er also nur bis auf eine solche Konstante bestimmt. Das nächste Lemma zeigt, daß es ihn wirklich gibt:

Lemma: In einem EUKLIDischen Ring R gibt es zu je zwei Elementen $x, y \in R$ einen ggT. Dieser kann nach dem EUKLIDischen Algorithmus berechnet werden und läßt sich als Linearkombination mit Koeffizienten aus R von x und y darstellen

Beweis: Eigentlich könnte man hier den entsprechenden Beweis aus dem letzten Kapitel fast wörtlich wiederholen, wobei man nur statt von ganzen oder natürlichen Zahlen von Elementen von R reden müßte, und statt $a < b$ müßte man $\nu(a) < \nu(b)$ sagen. Da man in der Mathematik alles auf viele verschiedene Weisen ausdrücken kann, möchte ich den Beweis stattdessen hier in einer weniger formalen Weise präsentieren.

In jedem Integritätsbereich folgt aus der Gleichung $x = qy + r$ mit $x, y, q, r \in R$, daß die gemeinsamen Teiler von x und y gleich denen von y und r sind. Speziell in einem EUKLIDischen Ring können wir dabei r als Divisionsrest wählen und, wie beim klassischen EUKLIDischen Algorithmus, danach y durch r dividieren *usw.,* wobei wir eine Folge (r_i) von Divisionsresten erhalten mit der Eigenschaft, daß in jedem Schritt die gemeinsamen Teiler von x und y gleich denen von r_{i-1}

und r_i sind. Außerdem ist stets entweder $r_i = 0$ oder $\nu(r_i) < \nu(r_{i-1})$, so daß die Folge nach endlich vielen Schritten mit einem $r_n = 0$ abbrechen muß. Auch hier sind die gemeinsamen Teiler von r_{n-1} und $r_n = 0$ genau die gemeinsamen Teiler von x und y. Da jede Zahl Teiler der Null ist, sind die gemeinsamen Teiler von r_{n-1} und Null aber genau die Teiler von r_{n-1}, und unter diesen gibt es natürlich einen größten, nämlich r_{n-1} selbst. Somit haben auch x und y einen größten gemeinsamen Teiler, nämlich den nach dem EUKLIDischen Algorithmus berechneten letzten von Null verschiedenen Divisionsrest r_{n-1}.

Auch die lineare Kombinierbarkeit folgt wie im klassischen Fall: Bei jeder Division mit Rest ist der Divisionsrest als Linearkombination von Dividend und Divisor darstellbar; beim EUKLIDischen Algorithmus beginnen wir mit der Division von x durch y, deren Rest somit als Linearkombinationen von x und y darstellbar ist. x und y selbst sind natürlich trivialerweise als Linearkombinationen von x und y darstellbar.

Wir wollen induktiv zeigen, daß auch bei allen weiteren Divisionen alle beteiligten Größen eine solche Darstellung haben.

Bei jeder dieser Divisionen wird der Divisor der vorigen Division durch deren Rest dividiert; nach Induktionsannahme haben beide eine Darstellung als Linearkombination von x und y. Der Divisionsrest ist eine Linearkombinationen von Dividend und Divisor, also auch eine von x und y. Insbesondere ist der ggT als letzter nichtverschwindender Divisionsrest Linearkombination von x und y, und die Koeffizienten können wie im vorigen Kapitel für $R = \mathbb{Z}$ mit dem erweiterten EUKLIDischen Algorithmus berechnet werden.

■

Im vorigen Kapitel hatten wir unter anderem mit Hilfe des erweiterten EUKLIDischen Algorithmus die eindeutige Primzerlegung gezeigt. Für ein ähnliches Resultat in allgemeineren Ringen definieren wir

Definition: *a)* Ein Element x eines Integritätsbereichs R heißt *irreduzibel*, falls gilt: x ist keine Einheit, und wenn sich x als Produkt $x = yz$ zweier Elemente aus R schreiben läßt, muß y oder z eine Einheit sein.
b) Ein Integritätsbereich R heißt *faktoriell* oder *ZPE*-Ring, wenn gilt: Jedes Element $x \in R$ läßt sich bis auf Reihenfolge und Assoziiertheit ein-

deutig schreiben als Produkt $x = u \prod_{i=1}^{r} p_i^{e_i}$ mit einer Einheit $u \in R^\times$, irreduziblen Elementen $p_i \in R$ und natürlichen Zahlen e_i.
(*ZPE* steht für **Z**erlegung in **P**rimfaktoren **E**indeutig.)

Lemma: In einem faktoriellen Ring gibt es zu je zwei Elementen x, y einen größten gemeinsamen Teiler.

Beweis: Wir wählen zunächst aus jeder Klasse assoziierter irreduzibler Elemente einen Vertreter; für die Zerlegung eines Elements in ein Produkt irreduzibler Elemente reicht es dann, wenn wir nur irreduzible Elemente betrachten, die Vertreter ihrer Klasse sind.

Sind $x = u \prod_{i=1}^{r} p_i^{e_i}$ und $y = v \prod_{j=1}^{s} q_j^{f_j}$ mit $u, v \in R^\times$ und p_i, q_j jeweils Vertreter ein Klasse irreduzibler Elemente die Zerlegungen von x und y in Primfaktoren, so können wir, indem wir nötigenfalls Exponenten Null einführen, o.B.d.A. annehmen, daß $r = s$ ist und $p_i = q_i$ für alle i. Dann ist offenbar $\prod_{i=1}^{r} p_i^{\min(e_i, f_i)}$ ein ggT von x und y, denn $z = \prod_{i=1}^{r} p_i^{g_i}$ ist genau dann Teiler von x, wenn $g_i \leq e_i$ für alle i, und Teiler von y, wenn $g_i \leq f_i$.

∎

Satz: Jeder Hauptidealring ist faktoriell.

Beweis: Wir müssen zeigen, daß jedes Element $x \neq 0$ bis auf Reihenfolge und Assoziiertheit eindeutig als Produkt aus einer Einheit und geeigneten Potenzen irreduzibler Elemente geschrieben werden kann. Wir beginnen damit, daß sich x überhaupt in dieser Weise darstellen läßt. Wie beim Hauptsatz der elementaren Zahlentheorie beweisen wir dies durch Widerspruch.

Wir nehmen also an, es gäbe Elemente $x \neq 0$, die sich nicht als Produkte von irreduziblen Elementen und Einheiten darstellen lassen und betrachten in der Menge M aller dieser Elemente ein bezüglich der Teilbarkeit minimales, d.h. ein Element $x \in M$ derart, daß jedes $y \in M$, das x teilt, zu y assoziiert sein muß. Wir müssen uns zunächst uberlegen, daß es so ein Element überhaupt gibt.

Tatsächlich gibt es sogar in *jeder* Teilmenge M eines Hauptidealrings ein in diesem Sinne minimales Element x, denn andernfalls hätten wir

eine unendliche Folge von Elementen $x_1, x_2, \ldots$ aus M derart, daß stets x_{i+1} ein echter Teiler von x_i wäre. Für die Hauptideale (x_i) bedeutete dies, daß (x_i) stets echt in (x_{i+1}) enthalten wäre:

$$(x_1) \subset (x_2) \subset (x_3) \subset \cdots .$$

Die Vereinigung I der sämtlichen Hauptideale (x_i) wäre wieder ein Ideal, und da wir in einem Hauptidealring sind, wäre $I = (x)$ ein Hauptideal. Da I die Vereinigung alle (x_i) ist, müßte x in einem dieser Ideale (x_i) liegen. Für $j > i$ wäre dann

$$(x) \subseteq (x_i) \subset (x_j) \subset I = (x) ,$$

im Widerspruch zur Annahme, daß (x_i) echt in (x_j) enthalten ist. Also gibt es ein bezüglich der Teilbarkeit minimales Element $x \in M$.

x kann nicht irreduzibel sein, denn sonst wäre $x = x$ eine Darstellung als Produkt irreduzibler Elemente. Daher läßt sich x als Produkt $x = yz$ zweier Elemente y, z schreiben, die beide keine Einheiten sind. Als echte Teiler von x können y und z nicht in M liegen, lassen sich also als Produkt einer Einheit mit einem Produkt irreduzibler Elemente darstellen. Multiplizieren wir die beiden Darstellungen miteinander und fassen die beiden Einheiten zusammen zu deren Produkt, erhalten wir eine entsprechende Darstellung für x, im Widerspruch zur Annahme $x \in M$. Also kann es keine Gegenbeispiele geben.

Als nächstes müssen wir uns überlegen, daß diese Darstellung bis auf Reihenfolge und Einheiten eindeutig ist. Das wesentliche Hilfsmittel hierzu ist wieder die folgende Zwischenbehauptung:

Falls ein irreduzibles Element p ein Produkt xy teilt, teilt es mindestens einen der beiden Faktoren.

Zum *Beweis* nehmen wir an, p sei kein Teiler von x. Dann liegt x nicht im Ideal (p), das Ideal $(p) \subset (p, x)$ ist also echt größer als (p). Da wir den Ring als Hauptidealring vorausgesetzt haben, gibt es ein Element q, so daß $(p, x) = (q)$ ist, d.h. q ist ein Teiler von p. Da p irreduzibel ist, sind alle Teiler entweder assoziiert zu p oder Einheiten. Im Falle der Assoziiertheit wäre $(q) = (p)$, was hier nicht der Fall ist; somit muß q

eine Einheit sein. Dann enthält (q) auch $q^{-1}q = 1$, ist also der ganze Ring. Da $(q) = (p, x)$ ist, gibt es daher eine Darstellung

$$1 = \alpha p + \beta x$$

mit zwei Ringelementen α, β. Multiplikation dieser Gleichung mit y führt auf $y = \alpha px + \beta xy$, und hier sind beide Summanden auf der rechten Seite durch p teilbar: Bei αpx ist das klar, und bei βxy folgt es daraus, daß nach Voraussetzung p ein Teiler von xy ist. Also ist p Teiler von y, und die Zwischenbehauptung ist bewiesen.

Induktiv folgt sofort:

Falls ein irreduzibles Element p ein Produkt $\prod_{i=1}^{r} x_i$ teilt, teilt es mindestens einen der Faktoren x_i.

Um den Beweis des Satzes zu beenden, müssen wir noch zeigen, daß die Zerlegung bis auf Reihenfolge und Einheiten eindeutig ist. Wieder nehmen wir an, dies sei nicht der Fall und wählen in der Menge aller Gegenbeispiele ein bezüglich der Teilbarkeit minimales Element x. Dieses hat somit mindestens zwei Zerlegungen

$$x = u \prod_{i=1}^{r} p_i^{e_i} = v \prod_{j=1}^{s} q_j^{f_j} \ ,$$

wobei wir annehmen können, daß alle $e_i, f_j \geq 1$ sind. Dann ist p_1 trivialerweise Teiler des ersten Produkts, also auch des zweiten. Wegen der Zwischenbehauptung teilt p_1 daher mindestens eines der Elemente q_j, d.h. $p_1 = wq_j$ ist bis auf eine Einheit w gleich q_j. Da $x/p_i = x/(wq_j)$ ein echter Teiler von x ist, liegt dieses Element nicht in M, hat also eine bis auf Reihenfolge und Einheiten eindeutige Zerlegung in irreduzible Elemente. Damit hat auch x eine solche Zerlegung.

■

Da der Polynomring in einer Veränderlichen über einem Körper Hauptidealring ist, läßt sich dort somit jedes Polynom in irreduzible Faktoren zerlegen. Wir wollen uns überlegen, daß dies auch für Polynome in mehreren Veränderlichen gilt, sogar dann, wenn wir den Körper ersetzen durch einen beliebigen faktoriellen Ring. Der entsprechende Satz geht zurück auf GAUSS, der dazu einen beliebigen solchen Polynomring einbettet in einen Polynomring in einer Variablen über einem Körper.

Als ersten Schritt konstruieren wir zu einem Integritätsbereich R einen Körper, der R enthält. Vorbild ist die Konstruktion der rationalen Zahlen aus den ganzen.

Wir betrachten auf der Menge aller Paare (f, g) mit $f, g \in R$ und $g \neq 0$ die Äquivalenzrelation

$$(f, g) \sim (r, s) \iff fs = gr\,;$$

die Äquivalenzklasse von (f, g) bezeichnen wir als den *Bruch* $\dfrac{f}{g}$.

Verknüpfungen zwischen diesen Brüchen werden nach den üblichen Regeln der Bruchrechnung definiert:

$$\frac{f}{g} + \frac{r}{s} = \frac{fs + rg}{gs} \quad \text{und} \quad \frac{f}{g} \cdot \frac{r}{s} = \frac{fr}{gs}\,.$$

Dies ist wohldefiniert, denn sind $(f, g) \sim (\tilde{f}, \tilde{g})$ und $(r, s) \sim (\tilde{r}, \tilde{s})$, so ist $f\tilde{g} = \tilde{f}g$ und $r\tilde{s} = \tilde{r}s$, und

$$\frac{\tilde{f}}{\tilde{g}} + \frac{\tilde{r}}{\tilde{s}} = \frac{\tilde{f}\tilde{s} + \tilde{r}\tilde{g}}{\tilde{g}\tilde{s}} \quad \text{und} \quad \frac{\tilde{f}}{\tilde{g}} \cdot \frac{\tilde{r}}{\tilde{s}} = \frac{\tilde{f}\tilde{r}}{\tilde{g}\tilde{s}}\,.$$

Damit ist $(fs + rg)\tilde{g}\tilde{s} = f\tilde{g}s\tilde{s} + r\tilde{s}g\tilde{g} = \tilde{f}gs\tilde{s} + \tilde{r}sg\tilde{g} = (\tilde{f}\tilde{s} + \tilde{r}\tilde{g})gs$ und $(fr)(\tilde{g}\tilde{s}) = f\tilde{g}r\tilde{s} = \tilde{f}g\tilde{r}s = (\tilde{f}\tilde{r})(gs)$, d.h. auch die Ergebnisse sind äquivalent.

Man rechnet leicht nach (wie bei $\mathbb{Q}$), daß diese Äquivalenzklassen einen Ring bilden mit $\frac{0}{1}$ als Null und $\frac{1}{1}$ als Eins; er ist sogar ein Körper, denn für $f, g \neq 0$ ist $\frac{g}{f}$ ein multiplikatives Inverses zu $\frac{f}{g}$, da $(fg, fg) \sim (1, 1)$. Identifizieren wir schließlich ein Element $f \in R$ mit dem Bruch $\frac{f}{1}$, so können wir R in den Körper K einbetten.

Definition: Der so konstruierte Körper K heißt Quotientenkörper von R, in Zeichen $K = \operatorname{Quot} R$.

Das Standardbeispiel ist natürlich $\mathbb{Q} = \operatorname{Quot} \mathbb{Z}$, aber auch der Quotientenkörper $k(X) \underset{\text{def}}{=} \operatorname{Quot} k[X]$ eines Polynomrings über einem Körper k ist wichtig: $k(X)$ heißt rationaler Funktionenkörper in einer Veränderlichen über k. Seine Elemente sind rationale Funktionen in X, d.h. Quotienten von Polynomen in X, wobei der Nenner natürlich nicht das Nullpolynom sein darf.

Für Polynome, die statt über einem Körper nur über einem faktoriellen Ring definiert sind, sind die beiden folgenden Begriffe sehr wesentlich:

Definition: *a)* Der *Inhalt* eines Polynoms $f = a_d X^d + \cdots + a_0 \in R[X]$ über einem faktoriellen Ring R ist der größte gemeinsame Teiler

$$I(f) = \mathrm{ggT}(a_0, \ldots, a_d)$$

seiner Koeffizienten a_i.
b) f heißt *primitiv,* wenn die a_i zueinander teilerfremd sind.

Indem wir alle Koeffizienten eines Polynoms durch ihren gemeinsamen ggT dividieren sehen wir, daß sich jedes Polynom aus $R[X]$ als Produkt seines Inhalts mit einem primitiven Polynom schreiben läßt. Diese Zerlegung bleibt bei der Multiplikation zweier Polynome erhalten:

Lemma: R sei ein faktorieller Ring. Für zwei Polynome

$$f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0 \quad \text{und}$$
$$g = b_e X^e + b_{e-1} X^{e-1} + \cdots + b_1 X + b_0$$

aus $R[X]$ ist (bis auf Assoziiertheit) $I(fg) = I(f) \cdot I(g)$. Insbesondere ist das Produkt zweier primitiver Polynome wieder primitiv.

Beweis: Wir schreiben $f = I(f) \cdot f^*$ und $g = I(g) \cdot g^*$ mit primitiven Polynomen f^* und g^*; dann ist $fg = I(f) \cdot I(g) \cdot (f^* g^*)$. Falls $f^* g^*$ wieder ein primitives Polynom ist, folgt, daß $I(fg) = I(f) \cdot I(g)$ sein muß.

Es genügt daher, zu zeigen, daß das Produkt zweier primitiver Polynome f und g wieder primitiv ist.

Das Produkt sei $fg = c_{d+e} X^{d+e} + c_{d+e-1} X^{d+e-1} + \cdots + c_1 X + c_0$; dabei ist $c_r = \sum_{i,j \text{ mit } i+j=r} a_i b_j$.

Angenommen, diese Koeffizienten c_r haben einen gemeinsamen Teiler, der keine Einheit ist. Wegen der Faktorialität von R gibt es dann auch ein irreduzibles Element p, das alle Koeffizienten c_r teilt.

Insbesondere ist p ein Teiler von $c_0 = a_0 b_0$; da p irreduzibel ist, muß mindestens einer der beiden Faktoren a_0, b_0 durch p teilbar sein. Da es

im Lemma nicht auf die Reihenfolge von f und g ankommt, können wir o.B.d.A. annehmen, daß a_0 Vielfaches von p ist.

Da f ein primitives Polynom ist, kann nicht jeder Koeffizient a_i durch p teilbar sein; ν sei der kleinste Index, so daß a_ν kein Vielfaches von p ist. Genauso gibt es auch einen kleinsten Index $\mu \geq 0$, für den b_μ nicht durch p teilbar ist. In

$$c_{\mu+\nu} = \sum_{i,j \text{ mit } i+j=\mu+\nu} a_i b_j$$

ist dann der Summand $a_\nu b_\mu$ nicht durch p teilbar, aber für jeden anderen Summanden $a_i b_j$ ist entweder $i < \nu$ oder $j < \mu$, so daß mindestens einer der Faktoren und damit auch das Produkt durch p teilbar ist. Insgesamt ist daher $c_{\mu+\nu}$ nicht durch p teilbar, im Widerspruch zur Annahme. Somit muß fg ein primitives Polynom sein.

■

Satz von Gauß: R sei ein faktorieller Ring und $K = \operatorname{Quot} R$. Falls sich ein Polynom $f \in R[X]$ in $K[X]$ als Produkt zweier Polynome $g, h \in K[X]$ schreiben läßt, gibt es ein $\lambda \in K$, so daß $\widetilde{g} = \lambda g$ und $\widetilde{h} = \lambda^{-1} h$ in $R[X]$ liegen und $f = \widetilde{g} \cdot \widetilde{h}$.

Beweis: Durch Multiplikation mit einem gemeinsamen Vielfache der Nenner aller Koeffizienten können wir aus einem Polynom mit Koeffizienten aus K eines mit Koeffizienten aus R machen. Dieses wiederum ist gleich seinem Inhalt mal einem primitiven Polynom. Somit läßt sich jedes Polynom aus $K[X]$ schreiben als Produkt eines Elements von K mit einem primitiven Polynom aus $R[X]$. Für g und h seien dies die Zerlegungen $g = cg^*$ und $h = dh^*$. Dann ist $f = (cd)g^*h^*$, und nach dem gerade bewiesenen Lemma ist g^*h^* ein primitives Polynom. Daher liegt $cd = I(f)$ in R, und wir können beispielsweise $\widetilde{g} = I(f)g^*$ und $\widetilde{h} = h^*$ setzen.

■

Korollar: Ein primitives Polynom $f \in R[X]$ ist genau dann irreduzibel in $R[X]$, wenn es in $K[X]$ irreduzibel ist.

■

Für nichtprimitive Polynome gilt diese Aussage natürlich nicht: Das Polynom $2X + 2$ ist zwar irreduzibel in $\mathbb{Q}[X]$, hat aber in $\mathbb{Z}[X]$ die beiden irreduziblen Faktoren 2 und $X + 1$.

CARL FRIEDRICH GAUSS (1777–1855) leistete wesentliche Beiträge zur Zahlentheorie, zur nichteuklidischen Geometrie, zur Differentialgeometrie und Kartographie, zur Fehlerrechnung und Statistik, zur Astronomie und Geophysik *usw.* Als Direktor der Göttinger Sternwarte baute er zusammen mit dem Physiker Weber den ersten Telegraphen. Er leitete die erste Vermessung und Kartierung des Königreichs Hannover, was sowohl seine Methode der kleinsten Quadrate als auch sein *Theorema egregium* motivierte. Zeitweise leitete er auch den Witwenfond der Universität Göttingen. Seine hierbei gewonnene Erfahrung nutzte er für erfolgreiche Aktienspekulationen.

Aus dem Satz von GAUSS folgt induktiv sofort, daß seine Aussage auch für Produkte von mehr als zwei Polynomen gilt, und daraus folgt

Satz: Der Polynomring über einem faktoriellen Ring R ist faktoriell.

Beweis: Wir müssen zeigen, daß sich jedes $f \in R[X]$ bis auf Reihenfolge und Einheiten eindeutig als Produkt von Potenzen irreduzibler Elemente aus $R[X]$ und einer Einheit schreiben läßt. Dazu schreiben wir $f = I(f) \cdot f^*$ mit einem primitiven Polynom $f^* \in R[X]$ und zerlegen zunächst den Inhalt $I(f)$ in R. Da R nach Voraussetzung faktoriell ist, ist diese Zerlegung eindeutig bis auf Reihenfolge und Einheiten in R, und wie wir bereits wissen, sind die Einheiten von $R[X]$ gleich denen von R.

Als nächstes zerlegen wir das primitive Polynom f^* über dem Quotientenkörper K von R; dies ist möglich, da $K[X]$ als EUKLIDischer Ring faktoriell ist. Jedes der irreduziblen Polynome q_i, die in dieser Zerlegung vorkommen, läßt sich schreiben als $q_i = \lambda_i p_i$ mit einem $\lambda_i \in K^\times$ und einem primitiven Polynom $p_i \in R[X]$. Wir können daher annehmen, daß in der Zerlegung von f nur primitive Polynome aus $R[X]$ auftreten sowie eine Einheit aus K. Diese muß, da f^* Koeffizienten aus R hat und ein Produkt primitiver Polynome primitiv ist, in R liegen; da auch f^* primitiv ist, muß sie dort sogar eine Einheit sein.

Kombinieren wir diese Primzerlegung von f^* mit der Primzerlegung des Inhalts, haben wir eine Primzerlegung von f gefunden; sie ist (bis

auf Reihenfolge und Einheiten) eindeutig, da entsprechendes für die Zerlegung des Inhalts, die Zerlegung von f^* sowie die Zerlegung eines Polynoms in Inhalt und primitiven Anteil gilt.

■

Da wir einen Polynomring $R[X_1, \ldots, X_n]$ in n Veränderlichen als Polynomring $R[X_1, \ldots, X_{n-1}][X_n]$ in einer Veränderlichen über dem Polynomring $R[X_1, \ldots, X_{n-1}]$ in $n-1$ Veränderlichen auffassen können, folgt induktiv sofort:

Satz: Der Polynomring $R[X_1, \ldots, X_n]$ in n Veränderlichen über einem faktoriellen Ring R ist faktoriell. Insbesondere sind $\mathbb{Z}[X_1, \ldots, X_n]$ sowie $k[X_1, \ldots, X_n]$ für jeden Körper k faktoriell.

■

Damit wissen wir also, daß auch Polynome in mehreren Veränderlichen über $\mathbb{Z}$ oder über einem Körper in Produkte irreduzibler Polynome zerlegt werden können; insbesondere existieren daher auch in dieser Ringen größte gemeinsame Teiler.

Der Beweis des obigen Satzes ist allerdings nicht konstruktiv; die Computeralgebra kennt zwar Algorithmen, mit denen man die Faktorisierung für Polynome, auch in mehreren Veränderlichen, über den ganzen oder rationalen Zahlen (und einigen anderen) konstruktiv durchführen kann, sie benutzen aber ganz andere Methoden als der obige Beweis.

Nachdem wir nun wissen, daß auch beispielsweise die Ringe $\mathbb{Z}[X]$ und $\mathbb{R}[X, Y]$ faktoriell sind, wissen wir, daß auch dort größte gemeinsame Teiler existieren. Offensichtlich sind in $\mathbb{Z}[X]$ sowohl die Zwei als auch X irreduzible Elemente; ihr größter gemeinsamer Teiler ist also eins. Es gibt aber natürlich keine Darstellung $1 = 2\alpha + X\beta$ mit ganzzahligen Polynomen $\alpha, \beta \in \mathbb{Z}[X]$, denn der konstante Term eines Polynom der Form $2\alpha + X\beta$ ist immer gerade. Genauso ist in $\mathbb{R}[X, Y]$ der ggT von X und Y gleich eins, aber eine Darstellung in der Form $1 = X\alpha + Y\beta$ ist nicht möglich, da $X\alpha + Y\beta$ keinen konstanten Term hat. Beide Ringe sind daher zwar faktoriell, aber nicht EUKLIDisch. Sie sind auch keine Hauptidealringe, denn auch in einem Hauptidealring ist der ggT linear kombinierbar: Ist nämlich (x, y) das von zwei Elementen x, y erzeugte Ideal, so ist dieses ein Hauptideal (u). Da (u) sowohl x als auch y enthält

und damit auch die Hauptideale (x) und (y), ist u sowohl Teiler von x als auch von y. Ist umgekehrt t ein gemeinsamer Teiler von x und y, so liegen x und y in s, also auch $(x, y) = (u)$. Also liegt (u) in (s), d.h. s ist ein Teiler von u. Somit ist u ein größter gemeinsamer Teiler von x und y; als Element von (x, y) hat er natürlich eine Darstellung der Form $u = \alpha x + \beta y$.

Zu Beginn dieses Paragraphen haben wir Ideale eingeführt als Teilmengen deren Eigenschaften gerade die der Kerne von Ringhomomorphismen sind. Von daher sollte es möglich sein, Faktorringe modulo einem Ideal zu bilden.

Wir beschränken uns auf kommutative Ringe R und betrachten ein Ideal $I \lhd R$. Da R insbesondere eine (additive) Gruppe ist und I eine Untergruppe, also wegen der Kommutativität der Addition automatisch ein Normalteiler ist, können wir auf jeden Fall die additive Gruppe R/I bilden. Um sie zu einem Ring zu machen, brauchen wir noch eine Multiplikation. Es bietet sich an, diese durch die Vorschrift

$$(x + I)(y + I) = xy + I$$

zu definieren – falls dies wohldefiniert ist.

Ist $x + I = x' + I$ und $y + I = y' + I$, so ist

$$x'y' = \big(x + (x' - x)\big)\big(y + (y' - y)\big)$$
$$= xy + x(y' - y) + (x' - x)y + (x' - x)(y' - y)\,.$$

$x' - x$ und $y' - y$ liegen in I; nach Definition eines Ideals liegen daher auch alle Produkte einer dieser Differenzen mit einem beliebigen Ringelement in I. Somit unterscheiden sich xy und $x'y'$ nur durch ein Element von I, d.h. $xy + I = x'y' + I$. Dies zeigt die Wohldefiniertheit der Multiplikation; Assoziativ- und Distributivgesetz folgen daraus, daß sie in R gelten.

Definition: R/I mit den beiden Rechenoperationen

$$(x + I) + (y + I) = (x + y) + I \quad \text{und} \quad (x + I)(y + I) = xy + I$$

heißt *Faktorring* von R modulo dem Ideal I.

Wie bei Gruppen haben wir auch bei Ringen einen

Homomorphiesatz: Ist $\varphi\colon R \to S$ ein Homomorphismus von kommutativen Ringen, so ist

$$R/\operatorname{Kern}\varphi \cong \operatorname{Bild}\varphi\,.$$

Beweis: Zwei Elemente $x, y \in R$ haben genau dann das gleiche Bild $\varphi(x) = \varphi(y)$, wenn $y - x$ im Kern liegt. Daher werden alle Elemente einer Nebenklasse $x + \operatorname{Kern}\varphi$ auf dasselbe Element von S abgebildet, so daß

$$\widetilde{\varphi}\colon \begin{cases} R/\operatorname{Kern}\varphi \to S \\ x + \operatorname{Kern}\varphi \mapsto \varphi(x) \end{cases}$$

eine wohldefinierte Abbildung ist. Da verschiedene Nebenklassen verschiedene Bilder haben, ist sie injektiv, und sie ist ein Homomorphismus, denn

$$\begin{aligned}\widetilde{\varphi}\big((x + \operatorname{Kern}\varphi) + (y + \operatorname{Kern}\varphi)\big) &= \widetilde{\varphi}\big((x+y) + I\big) = \varphi(x+y) \\ &= \varphi(x) + \varphi(y) = \widetilde{\varphi}(x + \operatorname{Kern}\varphi) + \widetilde{\varphi}(y + \operatorname{Kern}\varphi)\end{aligned}$$

und

$$\begin{aligned}\widetilde{\varphi}\big((x + \operatorname{Kern}\varphi)\,(y + \operatorname{Kern}\varphi)\big) &= \widetilde{\varphi}(xy + I) = \varphi(xy) = \varphi(x)\varphi(y) \\ &= \widetilde{\varphi}(x + \operatorname{Kern}\varphi)\;\widetilde{\varphi}(y + \operatorname{Kern}\varphi)\,.\end{aligned}$$

Wenn wir sie einschränken zu einer Abbildung von $R/\operatorname{Kern}\varphi$ nach Bild φ, ist sie auch surjektiv, also ein Isomorphismus. ∎

Betrachten wir als Beispiel den Homomorphismus $\varphi\colon \mathbb{Q}[X] \to \mathbb{R}$, der jedes Polynom abbildet auf seinen Wert an der Stelle $x_0 = \sqrt{2}$. Für $f = a_d X^d + \cdots + a_0 \in \mathbb{Q}[X]$ ist

$$f\big(\sqrt{2}\big) = \sum_{i=0}^{d} a_i \big(\sqrt{2}\big)^i = \sum_{\substack{i=0 \\ i \text{ gerade}}}^{d} a_i 2^{i/2} + \sum_{\substack{i=0 \\ i \text{ ungerade}}}^{d} a_i 2^{(i-1)/2}\sqrt{2}\,,$$

$\varphi(f)$ ist also von der Form $a + b\sqrt{2}$ mit $a, b \in \mathbb{Q}$. Umgekehrt liegt auch jede solche Zahl im Bild, denn $\varphi(bX + a) = a + b\sqrt{2}$. Somit ist $\operatorname{Bild}\varphi = \mathbb{Q} \oplus \mathbb{Q}\sqrt{2}$.

Für ein Polynom f aus dem Kern ist $f\left(\sqrt{2}\right) = 0$; da $\sqrt{2}$ keine rationale Zahl ist, müssen daher in obiger Zerlegung *beide* Summanden verschwinden. Damit ist auch

$$f\left(-\sqrt{2}\right) = \sum_{i=0}^{d} a_i \left(-\sqrt{2}\right)^i = \sum_{\substack{i=0 \\ i \text{ gerade}}}^{d} a_i 2^{i/2} - \sum_{\substack{i=0 \\ i \text{ ungerade}}}^{d} a_i 2^{(i-1)/2}\sqrt{2} = 0\,,$$

so daß f sowohl $\sqrt{2}$ als auch $-\sqrt{2}$ als Nullstellen hat. Somit ist f durch $\left(X - \sqrt{2}\right)\left(X + \sqrt{2}\right) = X^2 - 2$ teilbar.

Daraus folgt insbesondere, daß $X^2 - 2$ bis auf Assoziiertheit das einzige irreduzible Polynom aus $\mathbb{Q}[X]$ ist, das $\sqrt{2}$ als Nullstelle hat. Damit ist Kern φ das von $X^2 - 2$ erzeugte Hauptideal, und der Homomorphiesatz wird für dieses Beispiel zur Formel

$$\mathbb{Q}[X]/(X^2 - 2) \cong \mathbb{Q} \oplus \mathbb{Q}\sqrt{2}\,.$$

Natürlich ist $\varphi(X) = \sqrt{2}$, was wir auch so interpretieren können, daß wir völlig unabhängig von reellen Zahlen und Wurzeln mit $\mathbb{Q}[X]/(X^2 - 2)$ einen Ring konstruiert haben, in dem die über $\mathbb{Q}$ unlösbare Gleichung $x^2 = 2$ eine Lösung hat, denn natürlich haben X^2 und die Zwei modulo dem Ideal $(X^2 - 2)$ die gleiche Nebenklasse.

Auf den ersten Blick mag dies wie ein überflüssiger Taschenspielertrick erscheinen; wenn wir uns aber daran erinnern, wie die komplexen Zahlen aus den reellen konstruiert wurden, dann entspricht das genau der *Definition* von $\mathbb{C}$ als $\mathbb{R}[X]/(X^2+1)$, wobei die Nebenklasse von X mit i bezeichnet wird.

Betrachten wir allgemein einen Körper k und ein irreduzibles Polynom $f \in k[X]$ vom Grad mindestens zwei. Dann hat f in k keine Nullstelle, denn wäre $z \in k$ eine Nullstelle, so wäre $(X-z)$ ein Teiler von f in $k[X]$, was für ein irreduzibles Polynom vom Grad mindestens zwei nicht der Fall sein kann. Im Faktorring $k[X]/(f)$ allerdings ist die Nebenklasse x von X natürlich eine Lösung, denn $f(x) = f + (f) = (f)$ ist die Null des Faktorrings.

Dieser Faktorring $k[X]/(f)$ ist tatsächlich sogar ein Körper, denn wegen der Irreduzibilität von f ist jedes Polynom $g \in k[X]$, das nicht in (f)

liegt, teilerfremd zu f. Da $k[X]$ ein EUKLIDischer Ring ist, gibt es daher Polynome $a, b \in k[X]$ mit $ag + bf = 1$. Somit liegen af und 1 in der gleichen Nebenklasse modulo (f), d.h. in $k[X]/(f)$ ist die Nebenklasse von a invers zu der von g. Damit hat in $k[X]/(f)$ jedes von der Null verschiedene Element ein Inverses; der Ring ist also ein Körper.

Ist K irgendein Körper, der k enthält und in dem f eine Nullstelle z hat, können wir den Homomorphismus

$$\varphi\colon \begin{cases} k[X] \to K \\ \quad f \mapsto f(z) \end{cases}$$

betrachten. Wegen der Irreduzibilität von f ist jedes Polynom g, das in z verschwindet, ein Vielfaches von f, denn auch ggT(f, g) als Linearkombination von f und g verschwindet in z, hat also positiven Grad und muß deshalb gleich f sein. Somit ist Kern $\varphi = (f)$ und wir erhalten eine Einbettung des Körpers $k[X]/(f)$ in K als Bild φ.

Wir haben damit zu einem irreduziblen Polynom $f \in k[X]$ einen Erweiterungskörper gefunden, in dem das Polynom eine Nullstelle hat. Wenn wir eine numerische Lösung suchen, sind wir damit noch nicht viel weiter; wir brauchen dann zusätzlich eine Einbettung des Körpers $k[X]/(f)$ in einen Körper wie $\mathbb{R}$ oder $\mathbb{C}$. Im nächsten Kapitel werden wir aber sehen, daß uns der Körper $k[X]/(f)$ trotzdem eine große Hilfe ist bei der Untersuchung der Nullstellen des Polynoms f und der Möglichkeit, sie durch Grundrechenarten und Wurzeln auszudrücken.

Der Homomorphiesatz führt uns auch zu einer Verallgemeinerung des chinesischen Restesatzes auf Ringe und Ideale. Um auch auf Ringniveau mit simultanen Kongruenzen umgehen zu können, führen wir den Begriff der direkten Summe von Gruppen und Ringen ein:

Definition: *a)* G und H seien Gruppen. Die direkte Summe $G \oplus H$ ist die Produktmenge $G \times H$ mit der Komposition

$$(g, h) * (g', h') = (g * g', h * h') .$$

Im Falle von $n > 2$ Gruppen $G_1, \ldots, G_n$ ist $G_1 \oplus \cdots \oplus G_n$ entsprechend das Produkt $G_1 \times \cdots \times G_n$ mit

$$(x_1, \ldots, x_n) * (y_1, \ldots, y_n) = (x_1 * y_1, \ldots, x_n * y_n) .$$

b) R und S seien Ringe. Die direkte Summe $R \oplus S$ ist die Produktmenge $R \times S$ mit den Rechenoperationen

$$(r, s) + (r', s') = (r + r', s + s') \quad \text{und} \quad (r, s) \cdot (r', s') = (r \cdot r', s \cdot s') \,.$$

Im Falle von $n > 2$ Ringen $R_1, \ldots, R_n$ ist $R_1 \oplus \cdots \oplus R_n$ entsprechend das Produkt $R_1 \times \cdots \times R_n$ mit

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

und $(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = (x_1 \cdot y_1, \ldots, x_n \cdot y_n)$.

Das Neutralelement von $G \oplus H$ ist natürlich das Paar aus den beiden Netralelementen von G und H, und bei Ringen ist die Null in $R \oplus S$ das Paar $(0, 0)$, und die Eins ist $(1, 1)$. Entsprechendes gilt für direkte Summen von mehr als zwei Gruppen oder Ringen. Man beachte, daß der Ring $R \oplus S$ auch im Falle zweier Integritätsbereiche R und S stets Nullteiler enthält: Beispielsweise ist $(1, 0) \cdot (0, 1) = (0, 0)$.

Für n Ideale $I_1, \ldots, I_n$ eines Rings R können wir die Abbildung

$$\begin{cases} R \to R/I_1 \oplus \ldots \oplus R/I_n \\ x \mapsto (x + I_1, \ldots, x + I_n) \end{cases}$$

betrachten. Beim klassischen chinesischen Restesatz ist $R = \mathbb{Z}$, die Ideale I_ν werden von natürlichen Zahlen m_ν erzeugt, und es geht darum zu entscheiden, wann ein n-tupel

$$\big(x_1 + (m_1), \ldots, x_n + (m_n)\big) \in \mathbb{Z}/m_1 \oplus \cdots \oplus \mathbb{Z}/m_n$$

ein Urbild $x \in \mathbb{Z}$ hat und modulo welcher Zahl dieses gegebenenfalls eindeutig bestimmt ist. Diese Frage wollen wir nun für beliebige Ringe untersuchen.

Beginnen wir mit dem Fall von nur zwei Idealen:

Lemma: R sei ein kommutativer Ring, und I, J seien Ideale von R. Dann gibt es einen Monomorphismus von Ringen

$$\varphi\colon \begin{cases} R/(I \cap J) \to R/I \oplus R/J \\ \qquad\quad x \mapsto (x + I, x + J) \end{cases} .$$

Beweis: Es gibt natürlich einen (Ring-)Homomorphismus

$$\widetilde{\varphi}\colon \begin{cases} R \to R/I \oplus R/J \\ x \mapsto (x+I, x+J) \end{cases}$$

Ein Element $x \in R$ liegt genau dann im Kern von $\widetilde{\varphi}$, wenn sowohl $x+I$ das Nullelement von R/I ist als auch $x+J$ das von R/J. Dann liegt x sowohl in I als auch in J, also in $I \cap J$, und Kern $\widetilde{\varphi} = I \cap J$. Nach dem Homomorphiesatz ist daher $R/(I \cap J) \cong \text{Bild}\,\widetilde{\varphi}$, und da Bild $\widetilde{\varphi}$ eine Teilmenge von $R/I \oplus R/J$ ist, folgt die Behauptung. ∎

Wir können allerdings nicht erwarten, daß dieser Monomorphismus stets ein Isomorphismus ist: Wenn wir $\mathbb{Z}$ nach $\mathbb{Z}/4 \oplus \mathbb{Z}/10$ abbilden, kann etwa das Element $\big(1+(4), 2+(10)\big)$ unmöglich ein Urbild haben, denn dieses müßte ja sowohl gerade als auch ungerade sein. Da 4 und 10 beide gerade sind, kann $\big(y+(4), z+(10)\big)$ höchstens dann im Bild von $\widetilde{\varphi}$ liegen, wenn $y \equiv z \bmod 2$.

Bevor wir uns überlegen, was wir im allgemeinen Fall sagen können, zunächst eine

Vorbemerkung: Sind $I' \subset I \lhd R$ zwei Ideale eines Rings R, so ist die Projektion $\pi\colon R \to R/I$ die Hintereinanderausführung der Projektion $\pi'\colon R \to R/I'$ und des Homomorphismus

$$\varphi\colon \begin{cases} R/I' \to R/I \\ x+I' \mapsto x+I \end{cases} .$$

Beweis: φ ist wohldefiniert, da I' ganz in I liegt, und für jedes $x \in R$ ist $(\varphi \circ \pi')(x) = \varphi(x+I') = x+I = \pi(x)$. ∎

Um dies auf zwei Ideale $I, J \lhd R$ eines Rings anwenden zu können, brauchen wir ein Ideal, das beide enthält. Ein solches Ideal ist die Menge

$$I+J \underset{\text{def}}{=} \{x+y \mid x \in I \text{ und } y \in J\} \ :$$

I und J sind Teilmengen, da die Null sowohl in J als auch in I liegt, und für $x_1, x_2 \in I$ und $y_1, y_2 \in R$ liegt wegen der Kommutativität und Assoziativität der Addition auch $(x_1+y_1)+(x_2+y_2) = (x_1+x_2)+(y_1+y_2)$

in $I + J$. Für $r \in R$ ist wegen des Distributivgesetzes schließlich auch $r(x_1 + y_1) = rx_1 + ry_1 \in I + J$.

Nach der Vorbemerkung haben wir zu den drei Projektionen

$$R \to R/I\,, \quad R \to R/J \quad \text{und} \quad \pi\colon R \to R/(I + J)$$

Homomorphismen

$$R/(I \cap J) \to R/I\,, \quad R/(I \cap J) \to R/J\,, \quad R/(I \cap J) \to R/(I + J)\,,$$
$$\pi_1\colon R/I \to R/(I + J) \quad \text{und} \quad \pi_2\colon R/J \to R/(I + J)\,.$$

In der Abbildungsfolge

$$\begin{array}{ccccc} & & & \nearrow & R/I & \searrow & \\ R & \to & R/(I \cap J) & & \longrightarrow & & R/(I + J) \\ & & & \searrow & R/J & \nearrow & \end{array}$$

ist es offensichtlich gleichgültig, auf welchem Weg wir von R nach $R/(I + J)$ gehen. Wenn es zu zwei Nebenklassen $y + I \in R/I$ und $z + J \in R/J$ ein $x \in R$ gibt mit $x + I = y + I$ und $x + J = z + J$, muß daher gelten

$$\pi_1(y + I) = \pi_2(z + J) = \pi(x)\,.$$

Ist umgekehrt $\pi_1(y + I) = \pi_2(z + J)$, so ist $y + (I + J) = z + (I + J)$, also $y - z \in I + J$. Es gibt daher Elemente $i \in I$ und $j \in J$, so daß $y - z = i + j$ ist und damit $y - i = z + j$. Da $y - i$ in $y + I$ liegt und $z + j$ in $z + J$, liegt $x = y - i = z + j$ im Durchschnitt von $y + I$ und $z + J$, und es gilt $x + I = y + I$ und $x + J = z + J$. Somit ist $x + (I \cap J) \in R/(I \cap J)$ ein gemeinsames Urbild von $y + I \in R/I$ und $z + J \in R/J$.

Zusammenfassend können wir also sagen, daß es genau dann ein $x \in R$ gibt mit $x + I = y + I$ und $x + J = z + J$, wenn die beiden Elemente $\pi_1(y + I)$ und $\pi_2(z + J)$ aus $R/(I + J)$ übereinstimmen, wenn also $y + (I + J) = z + (I + J)$ ist. In diesem Fall ist die Nebenklasse $x + (I \cap J)$ durch y und z eindeutig bestimmt.

Mit Ringen und Homomorphismen ausgedrückt heißt das:

Lemma: Die Abbildung

$$\varphi\colon \begin{cases} R/(I\cap J) \to R/I \oplus R/J \\ \qquad\qquad x \mapsto (x+I, x+J) \end{cases}$$

induziert einen Isomorphismus von $R/(I\cap J)$ auf

$$\big\{(y+I, z+J) \in R/I \oplus R/J \;\big|\; y+(I+J) = z+(I+J)\big\}\,.$$

∎

Im Falle $I+J=R$ ist $y+(I+J) = z+(I+J)$ für alle $y, z \in R$; in diesem Fall ist $R/(I\cap J)$ daher isomorph zu $R/I \oplus R/J$. Durch vollständige Induktion folgt:

Chinesischer Restesatz für Ringe: R sei ein kommutativer Ring und $I_1, \ldots, I_n$ seien Ideale von R derart, daß $I_\mu + I_\nu = R$ für alle $\mu \neq \nu$. Dann gibt es einen Isomorphismus von Ringen

$$\varphi\colon R/(I_1\cap\cdots\cap I_n) \to R/I_1 \oplus \cdots \oplus R/I_n\,.$$

Beweis: Für $n = 1$ gibt es nichts zu beweisen; für $n = 2$ haben wir den Satz gerade bewiesen.

Für $n > 2$ betrachten wir die Ideale $I = I_1\cap\cdots\cap I_{n-1}$ und $J = I_n$. Für diese ist $I + J = R$, denn nach Voraussetzung ist $I_\nu + I_n = R$ für alle $\nu < n$; es gibt also Elemente $x_\nu \in I_\nu$ und $y_\nu \in I_n$ mit $x_\nu + y_\nu = 1$ für $\nu = 1, \ldots, n-1$. Dann liegt $x = x_1 \cdots x_{n-1}$ in $I = I_1\cap\cdots\cap I_{n-1}$, und

$$x = \prod_{\nu=1}^{n-1} x_\nu = \prod_{\nu=1}^{n-1}(1-y_\nu) = 1 - y \quad \text{mit} \quad y = -\sum_{\emptyset\neq M\subseteq\{1,\ldots,n-1\}} \prod_{\nu\in M}(-y_\nu)\,.$$

Da hier jedes Produkt mindestens ein y_ν enthält und dieses in I_n liegt, liegt auch y in $I_n = J$, und $x + y = 1$.

Nach dem gerade bewiesenen Lemma ist daher $R/(I\cap J) \cong R/I\oplus R/J$, das heißt $R/(I_1\cap\cdots\cap I_n) \cong R/(I_1\cap\cdots\cap I_{n-1}) \oplus R/I_n$. Nach Induktionsvoraussetzung ist der erste Summand

$$R/(I_1\cap\cdots\cap I_{n-1}) \cong R/I_1 \oplus \cdots \oplus R/I_{n-1}\,,$$

also ist $R/(I_1\cap\cdots\cap I_n) \cong R/I_1 \oplus \cdots \oplus R/I_n$, wie behauptet. ∎

Speziell für $R = \mathbb{Z}$ und Ideale $I_\nu = (m_\nu)$ erhalten wir den klassischen chinesischen Restesatz in der Form

Satz: Für paarweise teilerfremde natürliche Zahlen $m_1, \ldots, m_n$ ist die Abbildung

$$\varphi\colon \begin{cases} \mathbb{Z}/m_1 \cdots m_n \to \mathbb{Z}/m_1 \oplus \cdots \oplus \mathbb{Z}/m_n \\ x \bmod m_1 \cdots m_n \mapsto (x \bmod m_1, \ldots, x \bmod m_n) \end{cases}$$

ein Isomorphismus von Ringen.

∎

Als nächstes wollen wir uns überlegen, was dieser Isomorphismus für die Einheitengruppen bedeutet. Offensichtlich induziert ein Isomorphismus $\varphi\colon R \to S$ von Ringen einen Isomorphismus $R^\times \to S^\times$ zwischen den Einheitengruppen, denn ist $x \in R^\times$, so gibt es ein $y \in R$ mit $xy = 1$, also ist auch $\varphi(x)\varphi(y) = 1$. Ist umgekehrt $\varphi(x)$ eine Einheit von S, so gibt es in S ein multiplikatives Inverses, das sich wegen der Surjektivität von φ als $\varphi(y)$ mit einem $y \in R$ schreiben läßt. Da $\varphi(xy) = \varphi(x)\varphi(y) = 1 = \varphi(1)$ ist, muß $xy = 1$ sein wegen der Injektivität von φ.

Damit können wir zeigen

Korollar: Sind $m_1, \ldots, m_n$ paarweise teilerfremd, so gibt es einen Isomorphismus multiplikativer Gruppen

$$\left(\mathbb{Z}/m_1 \cdots m_n\right)^\times \cong \left(\mathbb{Z}/m_1\right)^\times \oplus \cdots \oplus \left(\mathbb{Z}/m_n\right)^\times .$$

Beweis: Wie wir uns gerade überlegt haben, folgt aus dem obigen Satz die Isomorphie der Einheitengruppen der Ringe $\mathbb{Z}/(m_1 \cdots m_n)$ und $\mathbb{Z}/m_1 \oplus \cdots \oplus \mathbb{Z}/m_n$. Da die Multiplikation sowohl in einer direkten Summe von Ringen als auch in einer direkten Summe von Gruppen komponentenweise definiert ist, ist letztere Gruppe isomorph zur direkten Summe der Einheitengruppen $\left(\mathbb{Z}/m_\mu\right)^\times$, womit das Korollar bewiesen ist.

∎

Definition: Für eine natürliche Zahl $m \geq 2$ bezeichnen wir $\left(\mathbb{Z}/m\right)^\times$ als die *prime Restklassengruppe modulo* m; ihre Gruppenordnung wird mit $\varphi(m)$ bezeichnet. Die Funktion $\varphi\colon \mathbb{N} \setminus \{1\} \to \mathbb{N}$ heißt EULERsche *φ-Funktion.*

LEONHARD EULER (1707–1783) wurde in Basel geboren und ging auch dort zur Schule und, im Alter von 14 Jahren, zur Universität. Dort legte er zwei Jahre später die Magisterprüfung in Philosophie ab und begann mit dem Studium der Theologie; daneben hatte er sich seit Beginn seines Studium unter Anleitung von JOHANN BERNOULLI mit Mathematik beschäftigt. 1726 beendete er sein Studium in Basel und bekam eine Stelle an der Petersburger Akademie der Wissenschaften, die er 1727 antrat. Auf Einladung FRIEDRICHS DES GROSSEN wechselte er 1741 an die preußische Akademie der Wissenschaften; nachdem sich das Verhältnis zwischen den beiden dramatisch verschlechtert hatte, kehrte er 1766 nach St. Petersburg zurück. Im gleichen Jahr erblindete er vollständig; trotzdem schrieb er rund die Hälfte seiner zahlreichen Arbeiten (Seine gesammelten Abhandlungen umfassen 73 Bände) danach. Sie enthalten bedeutende Beiträge zu zahlreichen Teilgebieten der Mathematik, Physik, Astronomie und Kartographie.

Das gerade bewiesene Korollar zeigt, daß die EULERsche φ-Funktion in manchen Fällen multiplikativ ist, genauer:

Definition: Eine Funktion φ von einer Teilmenge der natürlichen Zahlen nach $\mathbb{N}$ heißt *schwach multiplikativ,* wenn für zwei teilerfremde Zahlen m, n gilt: $\varphi(mn) = \varphi(m)\varphi(n)$.

Lemma: Die EULERsche φ-Funktion ist schwach multiplikativ.

Beweis: Sind m und n teilerfremd, so ist nach obigem Korollar $\big(\mathbb{Z}/mn\big)^\times \cong \big(\mathbb{Z}/m\big)^\times \oplus \big(\mathbb{Z}/n\big)^\times$. Die Gruppe links hat $\varphi(mn)$ Elemente; rechts steht, mengentheoretisch gesehen, das kartesische Produkt zweier Mengen mit $\varphi(m)$ und $\varphi(n)$ Elementen. Dies beweist die Behauptung. ∎

Die Voraussetzung, daß m und n teilerfremd sind, ist notwendig: Beispielsweise enthält $\big(\mathbb{Z}/4\big)^\times$ die Nebenklassen der Eins und der Drei, so daß $\varphi(4) = 2$ ist. Im Ring $\mathbb{Z}/2$ ist nur die Eins eine Einheit, d.h. $\varphi(2) = 1$ und $\varphi(2 \cdot 2) \neq \varphi(2) \cdot \varphi(2)$.

Wir können die Elemente der primen Restklassengruppe auch bestimmen, ohne daß wir zu jedem ein Inverses finden müssen:

Lemma: Die Nebenklasse x+(m) in $\mathbb{Z}/m$ liegt genau dann in $\left(\mathbb{Z}/m\right)^\times$, wenn ggT$(x, m)$ = 1 ist.

Beweis: Sind x und m teilerfremd, so liefert uns der erweiterte EUKLIDische Algorithmus ganze Zahlen y, n, für die $xy + mn = 1$ ist, also $xy = 1 - mn \equiv 1 \bmod m$. Somit ist x eine Einheit.

Umgekehrt gibt es zu jeder Einheit x ein y, so daß $xy \equiv 1 \bmod m$ ist. Es gibt also ein $n \in \mathbb{Z}$, so daß $xy = 1 + mn$ oder $xy - mn = 1$ ist. Daher muß jeder gemeinsame Teiler von m und x auch die Eins teilen; die beiden Zahlen sind also teilerfremd. ∎

Daraus folgt zu EULERs Verallgemeinerung des kleines Satzes von FERMAT:

Satz: Sind a, m zueinander teilerfremde natürliche Zahlen, so ist

$$a^{\varphi(m)} \equiv 1 \bmod m\,.$$

Beweis: Da a teilerfremd zu m ist, liegt die Restklasse von a modulo m in $(\mathbb{Z}/m)^\times$; die Ordnung von a teilt also nach LAGRANGE die Gruppenordnung $\varphi(m)$ der primen Restklassengruppe. ∎

Ist $m = p_1^{e_1} \cdots p_r^{e_r}$ die Primzerlegung einer natürlichen Zahl $m \geq 2$, so können wir $\varphi(m)$ wegen der schwachen Multiplikativität der EULERschen φ-Funktion berechnen, sobald wie die Werte $\varphi(p_i^{e_i})$ kennen. Eine ganze Zahl ist genau dann teilerfremd zu $p_i^{e_i}$, wenn sie nicht durch p_i teilbar ist. Unter den Zahlen von 0 bis $p_i^{e_i} - 1$ ist jede p_i-te durch p_i teilbar, also $p_i^{e_i-1}$ Stück, so daß $\varphi(p_i^{e_i}) = p_i^{e_i} - p_i^{e_i-1} = p_i^{e_i-1}(p_i - 1)$ dieser Zahlen nicht durch p_i teilbar sind. Damit folgt:

Satz: Für $m = p_1^{e_1} \cdots p_r^{e_r}$ ist $\varphi(m) = p_1^{e_1-1} \cdots p_r^{e_r-1}(p_1-1) \cdots (p_r-1)$. ∎

Für ein Produkt $m = pq$ zweier verschiedener Primzahlen ist also $\varphi(m) = (p - 1)(q - 1)$; dies zeigt noch einmal, warum diese Zahl beim RSA-Verfahren eine so wichtige Rolle spielt.

Auch wenn die additive Gruppe $\mathbb{Z}/m$ stets zyklisch ist, gibt es natürlich keinen Grund, daß auch die prime Restklassengruppe $(\mathbb{Z}/m)^\times$ zyklisch sein müßte: $(\mathbb{Z}/12)^\times$ etwa besteht aus den vier Nebenklassen $1, 5, 7$ und 11, die allesamt das Quadrat eins haben, ist also isomorph zur KLEINschen Vierergruppe. Ist allerdings $m = p$ eine Primzahl, so ist $\mathbb{Z}/p$ ein Körper, und in diesem Fall ist $(\mathbb{Z}/p)^\times$ zyklisch nach dem folgenden

Satz: Die multiplikative Gruppe eines endlichen Körpers ist zyklisch.

Beweis: Da die multiplikative Gruppe eines Körpers mit q Elementen aus allen Körperelementen außer der Null besteht, hat sie die Ordnung $q-1$, d.h. nach LAGRANGE ist die Ordnung eines jeden Elements ein Teiler von $q-1$. Wir müssen zeigen, daß es mindestens ein Element g gibt, dessen Ordnung *genau* $q-1$ ist.

Für jeden Primteiler p_i von $q-1$ hat die Polynomgleichung

$$x^{(q-1)/p_i} = 1$$

höchstens $(q-1)/p_i$ Lösungen im Körper; es gibt also zu jedem p_i ein Körperelement a_i mit $a_i^{(q-1)/p_i} \neq 1$.

q_i sei die größte Potenz von p_i, die $q-1$ teilt, und $g_i = a_i^{(q-1)/q_i}$ die $(q-1)/q_i$-te Potenz von a_i. Dann ist

$$g_i^{q_i} = a_i^{q-1} = 1 \quad \text{und} \quad g_i^{q_i/p_i} = a_i^{(q-1)/p_i} \neq 1\,;$$

g_i hat also die Ordnung q_i.

Da q_i eine Potenz der Primzahl p_i ist, haben alle Elemente der von g_i erzeugten Untergruppe eine p_i-Potenz als Ordnung. Wegen $p_i \neq p_j$ für $i \neq j$ haben zwei solche Untergruppen daher außer der Eins kein gemeinsames Element. Für das Produkt g aller g_i bedeutet dies, daß eine Potenz g^n von g nur dann gleich eins sein kann, wenn alle $g_i^n = 1$ sind. Somit muß n durch alle q_i teilbar sein und damit durch deren Produkt $q-1$. Damit ist die multiplikative Gruppe des Körpers die von g erzeugte zyklische Gruppe.

∎

Definition: Ein Element g eines endlichen Körpers k heißt *primitive Wurzel,* wenn es die zyklische Gruppe $k^\times$ erzeugt.

Selbst im Fall der Körper $\mathbb{F}_p = \mathbb{Z}/p$ gibt es keine Formel, mit der man eine solche primitive Wurzel explizit in Abhängigkeit von p angeben kann. Üblicherweise wählt man zufällig ein Element aus und testet, ob es die Ordnung $p-1$ hat. Die Wahrscheinlichkeit dafür ist offenbar $\varphi(p-1) : (p-1)$, was für die meisten Werte von p recht gut ist. Der Test, ob die Ordnung gleich $p-1$ ist, läßt sich allerdings nur dann effizient durchführen, wenn die Primteiler p_i von $p-1$ bekannt sind, denn dann kann man einfach testen, ob alle Potenzen mit den Exponenten $(p-1)/p_i$ von eins verschieden sind. Für große Werte von p, wie sie in der Kryptographie benötigt werden, kann dies ein Problem sein, so daß man hier im allgemeinen von faktorisierten Zahlen r ausgeht und dann testet, ob $r + 1$ prim ist.

Ist p eine Primzahl und a eine primitive Wurzel modulo p, so ist die Abbildung

$$\begin{cases} \mathbb{Z}/(p-1) \to (\mathbb{Z}/p)^\times \\ \qquad\quad x \mapsto a^x \end{cases}$$

ein Gruppenisomorphismus. Auch für große Primzahlen p ist es mit relativ geringem Aufwand möglich, das Bild eines Elements x zu berechnen; wir können dabei genauso vorgehen, wie bei der RSA-Verschlüsselung. Für die Berechnung der Umkehrfunktion, den sogenannten diskreten Logarithmus modulo p zur Basis a, sind allerdings keine effizienten Algorithmen bekannt, und daher lassen sich auch mit dieser Funktion Kryptoverfahren konzipieren.

Das älteste und einfachste stammt von DIFFIE und HELLMAN; damit können zwei Personen über eine unsichere Leitung einen Schlüssel vereinbaren, ohne daß sie vorher irgendwelche geheime Information vereinbart haben; auch öffentliche Schlüssel sind nicht notwendig.

Die beiden Teilnehmer einigen sich zunächst (über die unsichere Leitung) auf eine Primzahl p und eine natürliche Zahl a derart, daß die Potenzfunktion $x \mapsto a^x$ möglichst viele Werte annimmt. Als nächstes wählt Teilnehmer A eine Zufallszahl $x < p$ und B entsprechend ein $y < p$. A schickt $u = a^x \bmod p$ an B und erhält dafür $y = a^y \bmod p$ von diesem. Sodann berechnet A die Zahl

$$v^x \bmod p = \left(a^y\right)^x \bmod p = a^{xy} \bmod p$$

und B entsprechend $u^y \bmod p = \left(a^x\right)^y \bmod p = a^{xy} \bmod p$. Beide haben also auf verschiedene Weise dieselbe Zahl berechnet, die sie nun zum Beispiel verwenden können, um daraus einen Schlüssel für ein symmetrisches Kryptosystem zu bestimmen. Verfahren dazu gibt es mehr als genug: Sie könnten etwa die letzten oder sonst irgendwelche Bits dieser Zahl verwenden, aber auch einen irgendwie definierten Hashwert.

Ein Gegner, der den Datenaustausch abgehört hat, kennt die Zahlen p, a, u und v; er kann also problemlos alle möglichen Zahlen modulo p der Art $a^{\alpha x+\beta y} = u^\alpha \cdot v^\beta$ berechnen. Es fällt aber schwer, sich eine Art und Weise vorzustellen, wie er $a^{xy} \bmod p$ finden kann, ohne den diskreten Logarithmus von u oder v zu berechnen. (Bewiesen ist hier, wie üblich, natürlich nichts.)

In der Praxis wird dieses Verfahren nur selten verwendet wegen der folgenden Angriffsmöglichkeit: Nehmen wir an, der Gegner habe eine gewisse Kontrolle über das Netz, in dem der Datenaustausch stattfindet – beispielsweise, weil er Systemverwalter eines für die betreffende Verbindung unbedingt notwendigen Knotenrechners ist. Dann kann er eine sogenannte *man in the middle attack* durchführen: Er fängt alle Datenpakete zwischen A und B ab und ersetzt sie durch selbstfabrizierte eigene Pakete.

Damit kann er sich gegenüber A als B auszugeben und umgekehrt: Alles, was A an B zu schicken glaubt, geht tatsächlich an den Gegner C, und alles was B von A zu erhalten glaubt, kommt tatsächlich von C. In Gegenrichtung funktioniert das natürlich entsprechend.

Im einzelnen läuft der Angriff folgendermaßen ab: Falls die Zahlen a und p nicht ohnehin Konstanten eines Verbunds sind, dem A und B angehören, läßt C die Kommunikation, die zu deren Vereinbarung führt, ungehindert zu: In diesem Stadium beschränkt er sich auf reines Abhören.

Als nächstes wählen A und B ihre Zufallszahlen $x < p$ und $y < p$; gleichzeitig wählt C eine Zufallszahl $z < p$ oder vielleicht auch zwei verschiedene solche Zahlen z_A und z_B für die beiden Teilnehmer.

Wenn A die Zahl $u = a^x \bmod p$ an B schickt, fängt C diese Nachricht ab

und ersetzt sie durch $w_B = a^{z_B} \bmod p$; entsprechend fängt er Bs Nachricht $y = a^y \bmod p$ ab und schickt stattdessen $w_A = a^{z_A}$ an A. Dies führt dazu, daß am Ende A und C einen gemeinsamen Schlüssel s_A haben und B und C einen gemeinsamen Schlüssel s_B. Sowohl A als auch B glauben, der ihnen bekannte Schlüssel s_A *bzw.* s_B sei aus $a^{xy} \bmod p$ abgeleitet und senden nun damit verschlüsselte Nachrichten an ihren Partner. Diese Nachrichten fängt C ab, entschlüsselt sie mit dem Schlüssel, den er mit dem Absender gemeinsam hat, und verschlüsselt sie anschließend, gegebenenfalls nach einer seinen Interessen entsprechenden Modifikation, mit dem Schlüssel, den er mit dem Empfänger gemeinsam hat. Auf diese Weise hat er die gesamte Konversation unter Kontrolle, ohne daß A und B etwas merken.

Diese Attacke funktioniert, weil sich A und B nicht sicher sein können, den jeweils anderen am anderen Ende der Leitung zu haben. Die kryptographisch einwandfreie Modifikation, die das Verfahren gegen diese Art von Angriff sicher macht, bestünde beispielsweise darin, daß A und B ihre Nachrichten x und y vor dem Versenden unterschreiben – aber dann verschwindet auch wieder der Vorteil, daß sie ohne Kenntnis irgendeines Schlüssels miteinander kommunizieren können: Zur Verifikation einer Unterschrift braucht man schließlich den öffentlichen Schlüssel des Unterschreibenden.

Falls sich A und B hinreichend gut kennen, um die Stimme des jeweils anderen am Telephon einigermaßen sicher zu erkennen, können sie diese Art von Attacke auch dadurch erschweren, daß sie nach dem Austausch von u und v per Telephon über diese Zahlen (z.B. die 317. bis 320. Ziffer) und gegebenenfalls auch über Persönliches reden. Bei Videokonferenzen könnte man auch die Zahlen langsam über den Bildschirm des jeweils anderen laufen lassen. Die volle Sicherheit einer Schlüsselvereinbarung via RSA wird aber nicht erreicht, und da oft zumindest einer der Teilnehmer ein Unternehmen ist, das sich einen zertifizierten RSA-Schlüssel leisten kann, werden Schlüssel für symmetrische Kryptoverfahren in der Praxis sehr viel häufiger via RSA vereinbart als via DIFFIE-HELLMAN.

Zwischen RSA und den Verfahren mit diskreten Logarithmen gibt es einen ganz wesentlichen Unterschied: Wer die Faktorisierung des RSA-Moduls N kennt, kann die sonst schwer zugängliche Umkehrfunktion

von $x \mapsto x^e \bmod N$ leicht berechnen, so daß Potenzieren mit e direkt als Verschlüsselung benutzt werden kann.

Bei der modularen „Exponentialfunktion“ $x \mapsto a^x \bmod p$ sind keine speziellen Wahlen von a und p bekannt, die vermöge einer geheimen Information zu einer einfachen Umkehrfunktion führen – diskrete Logarithmen sind für alle gleich schwer zu berechnen.

Die geheime Information bei einem asymmetrischen Verfahren auf der Basis diskreter Logarithmen kann daher nur in der Kenntnis *einzelner* diskreter Logarithmen bestehen: Wer für einen speziellen Wert x die Potenz $u = a^x \bmod p$ berechnet hat, weiß anschließend, daß x der diskrete Logarithmus von u modulo p zur Basis a ist.

Bei diesen sehr viel spezielleren „Geheimnissen“ ist klar, daß Kryptoverfahren auf der Basis von diskreten Logarithmen anders aussehen müssen als RSA.

Im Prinzip könnte man die Schlüsselvereinbarung nach DIFFIE und HELLMAN direkt zu einem Verschlüsselungsverfahren erweitern: Nachdem das gemeinsame Geheimnis $\gamma = a^{xy} \bmod p$ vereinbart ist, können Nachrichtenblöcke m_i mit $0 \leq m_i < p - 1$ in beide Richtungen verschlüsselt werden als $c_i = \gamma m_i \bmod p$. Da beide Partner den Wert von γ kennen, können sie leicht nach dem erweiterten EUKLIDischen Algorithmus ein δ berechnen, so daß $\gamma\delta \equiv 1 \bmod p$, und die verschlüsselte Information kann einfach entschlüsselt werden als $m_i = \delta c_i \bmod p$.

Solange nur ein einzelner Block m übertragen werden soll, ist dagegen nichts einzuwenden. Sobald aber mehrere Blöcke zu übertragen sind, wird dieses Verfahren verwundbar gegen Angriffe mit bekanntem Klartext: Falls ein Gegner für einen einzigen Chiffreblock c_i den Klartextblock m_i kennt (oder errät), kann er $\delta = m_i/c_i \bmod p$ berechnen und damit den gesamten Klartext entschlüsseln. Um das Verfahren sicher zu machen, müßte man daher für jeden Block ein eigenes γ vereinbaren und dazu jedes Mal das gesamte DIFFIE-HELLMAN-Protokoll durchlaufen, was sehr aufwendig wäre.

Das Verfahren von ELGAMAL umgeht dieses Problem, indem es exakt dieselbe Mathematik mit einem leicht modifizierten Protokoll zu einem asymmetrischen Kryptoverfahren macht:

Die Parameter a und p sind entweder allgemein bekannte Systemparameter, oder jeder Teilnehmer A wählt sie selbst als Teil seines öffentlichen Schlüssels. Zusätzlich wählt er sich eine geheime Zufallszahl x und veröffentlicht $u = a^x \bmod p$.

Wer immer eine Nachricht $m_1, \ldots, m_r$ an A schicken möchte, erzeugt für jeden Block m_i eine Zufallszahl y_i berechnet daraus $v_i = a^{y_i} \bmod p$ und $c_i = u^{y_i} m_i$. Dann schickt er die Folge der Paare (v_i, c_i) an A. Der Chiffretext ist damit doppelt so lang wie der Klartext, was das Verfahren insbesondere für lange Texte nicht sonderlich attraktiv macht.

A muß zur Entschlüsselung den Multiplikator u^{y_i} kennen; dann kann er m_i als $c_i u^{-y_i}$ berechnen. Da $u^{y_i} \equiv a^{xy_i} \equiv \left(a^{y_i}\right)^x \equiv {v_i}^x \bmod p$ ist, hat er damit keine Probleme.

TAHER ELGAMAL wurde 1955 in Ägypten geboren. Er studierte zunächst Elektrotechnik an der Universität Kairo; nachdem er dort seinen BSc bekommen hatte, setzte er seine Studien fort an den Information Systems Laboratories der Stanford University. In seiner Masterarbeit ging es hauptsächlich um Systemtheorie, jedoch hörte er parallel auch freiwillig viele Mathematikvorlesungen und kam auf diesem Weg zur Kryptographie, die zum Thema seiner Doktorarbeit wurde. Nach dem Studium arbeitete er für eine ganze Reihe von Unternehmen, beispielsweise war er von 1995–1998 als Chefwissenschaftler von Netscape maßgeblich an der Entwicklung von SSL beteiligt. Zeitweise arbeitete er auch in selbst gegründeten Firmen. 2006 wurde er Chief Technology Officer der Tumbleweed Communications Corporation; seitdem diese 2008 von Axway übernommen wurde, ist er deren Chief Security Officer sowie Berater einer Reihe weiterer Unternehmen. Seit 2013 ist er Chief Technical Officer for Security des Cloud-Anbieters salesforce.com. Sein Name wird in der Literatur oft auch EL GAMAL oder ELGAMAL geschrieben; die obige Schreibweise ist die, die er selbst im Englischen benutzt. Eine mögliche Transkription der arabischen Schreibweise seines Namens ins Deutsche wäre TAHIR AL-DSCHAMAL; „al“ ist der bestimmte Artikel im Arabischen.

Der offensichtliche Angriff eines Gegners besteht darin, aus u und a den diskreten Logarithmus x zu ermitteln, was nach derzeitigem Stand der Dinge schwierig erscheint. Ob andere Angriffe zum Erfolg führen könnten, ist (wie üblich) unbekannt – hoffentlich auch unseren Gegnern.

Der Nachteil des Verfahrens von ELGAMAL und anderer Verfahren auf der Basis diskreter Logarithmen ist, daß man für jeden Nachrichtenblock zwei Blöcke übertragen muß. Daher werden solche Verfahren nur selten zur Verschlüsselung eingesetzt; sie liefern aber nützliche und viel verwendete Ansätze für elektronische Unterschriften.

Eine RSA-Unterschrift sollte nach derzeitigem Sicherheitsstandard eine Länge von mindestens drei Tausend Bit haben. Was damit unterschrieben wird, ist meist ein Hashwert einer Länge von etwa 256 Bit.

Verglichen mit dieser Länge erscheint eine drei Tausend Bit lange Unterschrift weit übertrieben. Andererseits wäre eine Unterschrift, die auf diskreten Logarithmen in einem Körper mir nur etwa 2^{256} Elementen beruht, ohne großen Aufwand fälschbar.

Der *Digital Signature Algorithm* DSA bietet einen Ausweg aus diesem Dilemma, indem er zwar in einer großen Gruppe rechnet, dabei aber kurze Unterschriften aus einer deutlich kleineren Untergruppe liefert. Dieser Algorithmus wurde im *Digital Signature Standard* DSS der USA spezifiziert und zählt neben RSA auch zu den vom Bundesamt für Sicherheit in der Informationstechnik und auf EU-Ebene von SOGIS empfohlenen Verfahren.

Als Ordnung der Untergruppe wählt man eine Primzahl q einer Länge von mindestens 256 Bit.

Die Sicherheit wird gewährleistet (soweit dies möglich ist) durch eine zweite Primzahl p, die so gewählt wird, daß $p \equiv 1 \bmod q$ ist; für ihre Größe sind mindestens drei Tausend Bit empfohlen.

Primzahlen $p \equiv 1 \bmod q$ sind nicht schwerer zu finden als beliebige Primzahlen: Falls man bei der Primzahlsuche wirklich auf Nummer sicher geht und Zufallszahlen auf Primalität testet, nimmt man hier einfach Zufallszahlen k und testet $kq + 1$ auf Primalität. Falls man mit ERATOSTHENES arbeitet, kann man das Sieben leicht so modifizieren, daß nur Zahlen der Form $kq + 1$ gesiebt werden. An den Erfolgschancen ändert dies in beiden Fällen nichts: Nach einem Satz von DIRICHLET über Primzahlen in arithmetischen Folgen ist die Dichte der Primzahlen der Form $kq + i$ für jedes i mit $0 < i < q$ dieselbe; in der Größenordnung n ist also weiterhin im Mittel jede $\ln n$-te solche Zahl eine Primzahl. (Tatsächlich sind es sogar geringfügig mehr, denn außer q selbst gibt es natürlich keine Primzahl der Form $p = kq$. Bei den Größenordnungen von q mit denen wir arbeiten, geht aber der Unterschied zwischen q und $q - 1$ definitiv im „Rauschen“ der im Kleinen sehr unregelmäßigen Primzahlverteilung unter.)

Als nächstes muß ein Element g gefunden werden, dessen Potenzen im Körper $\mathbb{F}_p$ eine Gruppe der Ordnung q bilden. Auch das ist einfach: Man starte mit irgendeinem Element $g_0 \in \mathbb{F}_p \setminus \{0\}$ und berechne seine $(p-1)/q$-te Potenz. Falls diese ungleich eins ist, muß sie wegen $g_0^{p-1} = 1$ die Ordnung q haben; andernfalls muß ein neues g_0 betrachtet werden.

Die so bestimmten Zahlen q, p und g werden veröffentlicht und können auch in einem ganzen Netzwerk global eingesetzt werden. Geheimer Schlüssel jedes Teilnehmers ist eine Zahl x zwischen eins und $q-1$; der zugehörige öffentliche Schlüssel ist $u = g^x \bmod p$.

Unterschreiben lassen sich mit diesem Verfahren Nachrichtenblöcke m mit $0 \le m < q$; im allgemeinen wird es sich dabei um Hashwerte der eigentlich zu unterschreibenden Nachricht handeln. Dazu wählt man für jede Nachricht eine Zufallszahl k mit $0 < k < q$ und berechnet

$$r = (g^k \bmod p) \bmod q\,.$$

Man beachte, daß es in dieser Formel nicht um Restklassen geht, sondern um Zahlen aus $\mathbb{N}_0$: Aus $x \equiv y \bmod p$ folgt selbstverständlich nicht, daß auch $x \equiv y \bmod q$ sein muß. Der Operator mod in dieser Gleichung bezeichnet den (nichtnegativen) Divisionsrest, also eine ganze Zahl zwischen 0 und $p-1$ *bzw.* $q-1$.

Da q eine Primzahl ist, hat k ein multiplikatives Inverses modulo q; man kann also modulo q durch k dividieren und somit eine Zahl s berechnen, für die gilt

$$sk \equiv m + xr \bmod q$$

Die Unterschrift unter die Nachricht m besteht dann aus den beiden Zahlen r und $s = k^{-1}(m + xr) \bmod q$, die beide zwischen 0 und $q-1$ liegen. Sie kann nur erzeugt werden von jemanden, der den geheimen Schlüssel x kennt.

Überprüfen kann die Unterschrift allerdings jeder: Ist t das multiplikative Inverse zu s modulo q, so ist $k \equiv tsk \equiv tm + xtr \bmod q$, also, da g die Ordnung q hat, $g^k \bmod p = g^{tm} g^{xtr} \bmod p = g^{tm} u^{tr} \bmod p$. Modulo q ist die linke Seite gleich r, und auf der rechten Seite können sowohl g^{tm} als auch u^{tr} aus öffentlicher Information und der Unterschrift berechnet

werden. Modulo q kann diese Gleichung somit überprüft werden; die Unterschrift wird anerkannt, wenn

$$r \equiv (g^{tm} u^{tr} \bmod p) \bmod q$$

ist. (Die beiden Potenzen und ihr Produkt müssen natürlich auch hier zunächst modulo p berechnet werden: Zwei modulo p kongruente Zahlen sind praktisch nie auch kongruent modulo q.)

Ein Angreifer müßte sich nach allem was wir wissen x aus u verschaffen, müßte also ein diskretes Logarithmenproblem modulo der großen Primzahl p lösen, so daß der Sicherheitsstandard dem des diskreten Logarithmenproblems modulo p entsprechen sollte, obwohl die Unterschriften deutlich kürzer sind.

Diskrete Logarithmen lassen sich nicht nur für die prime Restklassengruppe definieren; wir können grundsätzlich für jede Gruppe G und jedes Element $a \in G$ der Ordnung r die „Exponentialfunktion“

$$\varphi\colon \begin{cases} \mathbb{Z}/r \to G \\ \quad x \mapsto a^x \end{cases}$$

betrachten und ihre Umkehrfunktion Bild $\varphi \to G$ als diskreten Logarithmus bezeichnen. Für manche Gruppen ist dieser recht einfach zu berechnen, etwa wenn G eine additive zyklische Gruppe ist; für andere kann die Berechnung sogar noch deutlich aufwendiger sein als im Fall der primen Restklassengruppe. Ein in der kryptographischen Praxis viel verwendetes Beispiel sind diskrete Logarithmen für elliptische Kurven. Dabei handelt es sich um ebene Kurven vom Grad drei (ohne Doppelpunkte), also Kurven mit Gleichungen wie $y^2 = x^3 + 2x + 5$. Man kann auf diesen Kurven eine Addition von Punkten erklären, mit einem „unendlich fernen“ zusätzlichen Punkt O als Nullelement. Mit diskreten Logarithmen auf solchen Kurven arbeitet beispielsweise die Unterschriftsfunktion der deutschen Bundespersonalausweise.

Kapitel 4
Nullstellen und Körpererweiterungen

§1: Zerfällungskörper und der Fundamentalsatz der Algebra

Ist k ein Körper und $f \in k[X]$ ein irreduzibles Polynom vom Grad mindestens zwei, so hat f in k keine Nullstelle. Wir kennen aber bereits viele Fälle, in denen es einen größeren Körper K gibt, in dem f eine oder mehrere Nullstellen hat. Solche Körper lassen sich auf verschiedene Weisen konstruieren: Ist etwa $k = \mathbb{Q}$ und $f = X^2 - 2$, so können wir bekanntlich mit dem Verfahren von HENON durch die Iteration

$$x_0 = 1, \qquad x_n = \frac{1}{2}\left(x_{n-1} + \frac{2}{x_{n-1}}\right) \quad \text{für alle } n \in \mathbb{N}$$

immer bessere Näherungslösungen konstruieren, und wenn wir die rationalen Zahlen durch Hinzunahme aller Grenzwerte von CAUCHY-Folgen (oder Intervallschachtelungen) zu den reellen Zahlen erweitern, ist dort der Grenzwert

$$x = \lim_{n\to\infty} x_n$$

dieser Folge eine Lösung.

Für die Nullstellen des Polynoms $X^2 + 1$ ist ein solcher Ansatz nicht möglich; hier müssen wir die „imaginäre Einheit“ i einführen als „Symbol“ mit dem wir rechnen. Ähnlich hatten wir uns bereits gegen Ende des vorigen Kapitels überlegt, daß wir zu jedem Körper k und jedem irreduziblen Polynom über f einen größeren Körper finden können, in dem f eine Nullstelle hat. Diesen Ansatz wollen wir nun systematisch ausbauen.

Beginnen wir mit Körpererweiterungen:

Definition: Sind $k \subseteq K$ zwei Körper, so bezeichnen wir k als *Teilkörper* von K und K als *Erweiterungskörper* von k. Wir sagen auch, K/k, gesprochen K über k, sei eine *Körpererweiterung*.

Ist K/k eine Körpererweiterung, so ist K ein k-Vektorraum, denn K ist bezüglich seiner Addition eine abelsche Gruppe, und die Einschränkung der Multiplikation in K auf $k \times K$ ist die Multiplikation der „Skalare“ aus k mit den „Vektoren“ aus K. Klassisches Beispiel ist die Betrachtung des Körpers $\mathbb{C}$ der komplexen Zahlen als zweidimensionalen Vektorraum $\mathbb{R}^2$.

Im allgemeinen muß dieser Vektorraum nicht endlichdimensional sein: $\mathbb{R}$ kann beispielsweise unmöglich ein endlichdimensionaler $\mathbb{Q}$-Vektorraum sein, denn genau wie $\mathbb{Q}$ ist auch jeder Vektorraum $\mathbb{Q}^n$ abzählbar, aber $\mathbb{R}$ ist überabzählbar.

Definition: Ist K ein endlichdimensionaler k-Vektorraum, sagen wir, die Körpererweiterung sei endlich, und wir bezeichnen die Dimension des k-Vektorraums K als deren Grad $[K : k]$. Andernfalls sagen wir, sie sei unendlich und schreiben $[K : k] = \infty$.

Als Beispiel betrachten wir ein irreduzibles Polynom $f \in k[X]$ vom Grad d und den Faktorring $K = k[X]/(f)$. Wie wir gegen Ende des vorigen Kapitels gesehen haben, ist er ein Körper, und als Vektorraum hat er beispielsweise die Basis $1, x, \ldots, x^{d-1}$, wobei $x = X + (f)$ die Restklasse von X bezeichnet. Ist $f = a_d X^d + \cdots + a_0$ mit $a_d \neq 0$, so ist

$$X^d \equiv -\frac{a_{d-1}X^{d-1} + \cdots + a_1 X + a_0}{a_d} \bmod (f)$$

und damit

$$x^d = -\frac{a_{d-1}x^{d-1} + \cdots + a_1 x + a_0}{a_d} ,$$

so daß x^d und die höheren Potenzen nicht zur Erzeugung gebraucht werden. Die genannten Elemente sind auch linear unabhängig über k, denn falls es Elemente $\lambda_i \in k$ gibt, so daß

$$\lambda_0 \cdot 1 + \lambda_1 \cdot x + \cdots + \lambda_{d-1} \cdot x^{d-1} = 0$$

ist, so muß das Polynom $\lambda_0 + \lambda_1 \cdot X + \cdots + \lambda_{d-1} \cdot X^{d-1}$ in $k[X]$ durch f teilbar sein. Da sein Grad höchstens gleich $d-1$ sein kann, geht das nur, wenn es das Nullpolynom ist, wenn also alle λ_i verschwinden.

Wir können dieses Ergebnis und die Diskussion im vorigen Kapitel zusammenfassen zum

Lemma: Ist $f \in k[X]$ ein irreduzibles Polynom vom Grad $d \geq 1$, so ist $K = k[X]/(f)$ ein Erweiterungskörper vom Grad d, in dem f mindestens eine Nullstelle hat. ∎

Sind L/K und K/k zwei Körpererweiterungen, so ist auch L/k eine; hier gilt:

Lemma: *a)* Sind L/K und K/k zwei endliche Körpererweiterungen, so ist auch L/k eine endliche Körpererweiterung und

$$[L : k] = [L : K] \cdot [K : k] .$$

b) Ist L/k eine endliche Körpererweiterung und ist $k \subseteq K \subseteq L$, so sind sowohl L/K als auch K/k endliche Körpererweiterungen. Ist $[L : k] = [K : k]$, so ist $K = L$.

Beweis: a) $b_1, \ldots, b_r$ seine eine Basis von K als k-Vektorraum, und $c_1, \ldots, c_s$ sei eine Basis von L als K-Vektorraum. Dann können wir in L die rs-Produkte $b_i c_j$ bilden, und wollen uns überlegen, daß diese eine Basis des k-Vektorraums L bilden.

Zunächst erzeugen sie diesen Vektorraum, denn jedes $v \in L$ läßt sich als Linearkombination

$$v = \lambda_1 c_1 + \cdots + \lambda_s c_s \quad \text{mit} \quad \lambda_j \in K$$

schreiben, und jedes λ_j läßt sich schreiben als

$$\lambda_j = \mu_{1j} b_1 + \cdots + \mu_{rj} b_r \quad \text{mit} \quad \mu_{ij} \in k .$$

Setzt man dies in die darüberliegende Formelzeile ein, erhält man v als Summe aller $\mu_{ij} b_i c_j$.

Zum Beweis der linearen Unabhängigkeit nehmen wir an,

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}b_ic_j = \sum_{j=1}^{s}\left(\sum_{i=1}^{r}\mu_{ij}b_i\right)c_j = 0$$

für irgendwelche Elemente $\mu_{ij} \in k$. Die Summen in der Klammer sind Elemente von K; wegen der linearen Unabhängigkeit der c_j über K müssen sie also alle verschwinden. Dann müssen aber auch alle μ_{ij} verschwinden, denn die b_i sind linear unabhängig über k.

Somit ist $[L : k] = rs = [K : k] \cdot [L : K]$, wie behauptet.

b) Betrachten wir K und L als Vektorräume über k, so ist K ein Untervektorraum von L, und natürlich sind Untervektorräume endlichdimensionaler Vektorräume selbst endlichdimensional. Wenn beide die gleiche Dimension haben, müssen sie sogar gleich sein. Als K-Vektorraum ist L endlichdimensional, da eine k-Basis von L insbesondere ein Erzeugendensystem von L über dem größeren Körper K ist.

■

Am einfachsten findet man die Nullstellen eines Polynoms, wenn das Polynom bereits als Produkt von Linearfaktoren gegeben ist. Wir wollen uns überlegen, daß es für jedes Polynom einen Körper gibt, über dem es so zerlegt werden kann:

Definition: k sei ein Körper und $f \in k[X]$ sei ein Polynom. Ein Körper K mit $k \subseteq K$ heißt *Zerfällungskörper* von f über k, wenn gilt:

1.) Es gibt Elemente $z_1, \ldots, z_d \in K$ und $a \in k$, so daß im Polynomring $K[X]$ gilt $f = a(X - z_1) \cdots (X - z_n)$.
2.) Ist $k \subseteq L \subseteq K$ und gibt es eine solche Zerlegung auch über L, so ist $L = K$.

In einem Zerfällungskörper *zerfällt* das Polynom also in ein Produkt von Linearfaktoren, und es gibt keinen echt kleineren Teilkörper, über dem dies bereits der Fall ist.

Für das Polynom $X^2 - 2 \in \mathbb{Q}[X]$ ist somit $\mathbb{Q} \oplus \mathbb{Q}\sqrt{2}$ ein Zerfällungskörper, denn $X^2 - 2 = (X + \sqrt{2})(X - \sqrt{2})$. Auch $\mathbb{Q}[X]/(X^2 - 2)$ ist ein Zerfällungskörper, denn bezeichnet x die Restklasse von X, so ist auch $(X + x)(X - x) = X^2 - x^2 = X^2 - 2$.

Satz: k sei ein Körper und $f \in k[X]$ ein Polynom. Dann gibt es einen Zerfällungskörper K von f über k.

Beweis durch Induktion nach $d = \deg f$: Für Polynome vom Grad Null gibt es nichts zu beweisen, für das Polynom $aX + b$ mit $a \neq 0$ ist k selbst der Zerfällungskörper, denn

$$aX + b = a\left(X - \frac{(-b)}{a}\right) .$$

Sei nun $d > 1$ und $f \in k[X]$ ein Polynom vom Grad d. Weiter sei g ein irreduzibler Faktor von f; für irreduzible f setzen wir natürlich $g = f$. Wie wir aus dem Lemma zu Beginn dieses Paragraphen wissen, ist $k[X]/(g)$ ein Körper, in dem g (mindestens) eine Nullstelle z_1 hat. Da es hierbei auf den Namen der Variablen nicht ankommt und wir X im folgenden noch als Variable brauchen, betrachten wir stattdessen den Körper $k_1 = k[Y]/\big(g(Y)\big)$, wobei $g(Y)$ aus g entsteht, indem wir Y für die Variable X einsetzen. Bezeichnet z_1 die Restklasse von Y in k_1, ist also $g(z_1) = 0$ und damit auch $f(z_1) = 0$.

Nun betrachten wir f und g als Elemente des Polynomring $k_1[X]$; da k ein Teilkörper von k_1 ist, geht das ohne Probleme. Da $f(z_1)$ verschwindet, ist f dort ein Vielfaches von $(X - z_1)$. Sei etwa $f = (X - z_1) \cdot f_1$ mit einem Polynom $f_1 \in k_1[X]$ vom Grad $d - 1$. Nach Induktionsannahme gibt es einen Zerfällungskörper K von f_1 über k_1. In diesem Körper läßt sich f_1 als Produkt von Linearfaktoren und einer Konstanten schreiben, also gibt es auch für $f = (X - z_1)f_1$ eine solche Darstellung

$$f = a(X - z_1)(X - z_2) \cdots (X - z_d) \quad \text{mit} \quad z_i \in K .$$

Der kleinste Teilkörper von K, der k und alle z_i enthält, ist somit ein Zerfällungskörper von f über k.

∎

Für das Polynom $X^3 - 2$ über $\mathbb{Q}$ etwa konstruieren wir zunächst den Körper $k_1 = \mathbb{Q}[Y]/(Y^3 - 2)$; die Nebenklasse von Y in k_1 bezeichnen wir als z_1. In k_1 ist dann $z_1^3 = 2$.

Nun dividieren wir $X^3 - 2 = X^3 - z_1^3$ in $k_1[X]$ durch $X - z_1$ und erhalten den Quotienten $f_1 = X^2 + z_1 X + z_1^2$. Mit einer neuen Variablen Z bilden wir den neuen Faktorring $k_2 = k_1[Z]/(Z^2 + z_1 Z + z_1^2)$; die Restklasse

von Z modulo (f_1) sei z_2. In $k_2[X]$ ist f_1 durch $X - z_2$ teilbar, und da f_2 den Grad zwei hat, ist der Quotient linear. Somit liegen beide Nullstellen von f_2 in k_2; das Polynom $X^3 - 2$ zerfällt also über k_2 in Linearfaktoren.

k_2 ist ein zweidimensionaler k_1-Vektorraum mit Basis $1, z_2$, und k_1 ist ein dreidimensionaler k_2-Vektorraum mit Basis $1, z_1, z_1^2$. Als k-Vektorraum hat k_2 somit die Dimension sechs und die Basis $1, z_1, z_1^2, z_2, z_1 z_2, z_1^2 z_2$.

Wir können k_1 in $\mathbb{R}$ einbetten, indem wir z_1 auf $\sqrt[3]{2}$ abbilden. Dann haben wir für z_2 über $\mathbb{R}$ die quadratische Gleichung $z_2^2 + \sqrt[3]{2}\, z_2 \oplus \sqrt[3]{4} = 0$ mit Lösungen

$$\left(-\frac{1}{2} + \frac{1}{2}\sqrt{-3}\right) \sqrt[3]{2} \quad \text{und} \quad \left(-\frac{1}{2} - \frac{1}{2}\sqrt{-3}\right) \sqrt[3]{2}$$

in $\mathbb{C}$, wie erwartet. Wir hätten aber aber natürlich k_1 auch in $\mathbb{C}$ einbetten können, indem wir z_1 auf $\left(-\frac{1}{2} + \frac{1}{2}\sqrt{3}\right) \sqrt[3]{2}$ abbilden und hätten dann eine quadratische Gleichung mit komplexen Koeffizienten bekommen, die die konjugiert komplexe Zahl sowie $\sqrt[3]{2}$ als Lösungen hätte.

Es ist kein Wunder, daß die Gleichung in $\mathbb{C}$ drei Nullstellen hat; der sogenannte *Fundamentalsatz der Algebra* besagt, daß jedes Polynom mit komplexen Koeffizienten über $\mathbb{C}$ in Linearfaktoren zerfällt. Für diesen Satz gibt es mehrere Beweise, unter anderem über die Funktionentheorie oder mit Hilfe der algebraischen Topologie. Der folgende Beweis stammt aus dem Buch *Théorie algébrique des nombres* von PIERRE SAMUEL (Hermann, Paris, ^{2}1971), und geht nach Angaben des Autors „im wesentlichen“ zurück auf LAGRANGE. Er verwendet nur elementare, aus der Analysisvorlesung bekannte Eigenschaften der reellen und komplexen Zahlen. Der wesentliche Beweisschritt ist der folgende

Satz: Jedes nichtkonstante Polynom $f \in \mathbb{R}[X]$ hat mindestens eine komplexe Nullstelle.

Beweis: Wir schreiben den Grad d eines Polynoms in der Form $d = 2^n \cdot u$ mit $n \in \mathbb{N}_0$ und einer ungeraden Zahl u und beweisen den Satz durch Induktion nach n.

Für den Induktionsanfang $n = 0$ müssen wir somit beweisen, daß jedes reelle Polynom ungeraden Grades mindestens eine komplexe Nullstelle hat. Da wir aus der Analysis wissen, daß es sogar eine reelle Nullstelle hat, ist das klar.

Nun sei $n > 0$; wir nehmen an, daß die Behauptung für alle Grade d, in deren Zerlegung ein kleineres n auftaucht, bereits bewiesen sei, und betrachten ein Polynom $f \in \mathbb{R}[X]$ vom Grad $d = 2^n u$ mit irgendeinem ungeraden u. Wie wir wissen, gibt es einen Zerfällungskörper $K/\mathbb{R}$, über dem das Polynom in Linearfaktoren zerfällt. Die d (nicht notwendigerweise verschiedenen) Nullstellen seien $z_1, \ldots, z_d$.

Mit diesen Nullstellen konstruieren wir nun neue Polynome, deren Grad zwar größer als d ist, aber nur durch 2^{n-1} teilbar ist, so daß wir die Induktionsannahme anwenden können.

Zu jedem $\lambda \in \mathbb{R}$ betrachten wir für alle Paare (i, j) mit $1 \leq i < j \leq d$ die Elemente

$$w_{ij}(\lambda) = z_i + z_j + \lambda z_i z_j \in K$$

sowie das Polynom

$$g_\lambda = \prod_{\substack{(i,j) \\ 1 \leq i < j \leq d}} \big(X - w_{ij}(\lambda)\big) \in K[X]\,.$$

Tatsächlich liegt g_λ sogar in $\mathbb{R}[X]$, denn seine Koeffizienten sind nach dem Satz von VIÈTE (Kap. 1, §6) bis aufs Vorzeichen die elementarsymmetrischen Funktionen in den $w_{ij}(\lambda)$. Damit sind sie auch symmetrische Funktionen in den z_i, denn jede Permutation $z_i \mapsto z_{\pi(i)}$ führt zu einer Permutation $w_{ij}(\lambda) \mapsto w_{\pi(i)\pi(j)}(\lambda)$. Nach dem Hauptsatz über symmetrische Funktionen (Kap. 1, §7) lassen sie sich daher als Polynome in den elementarsymmetrischen Funktionen der z_i schreiben, also, wieder nach VIÈTE, als Polynome in den Koeffizienten von f. Diese Koeffizienten sind reelle Zahlen; also sind auch die Koeffizienten aller g_λ reelle Zahlen, d.h. $g_\lambda \in \mathbb{R}[X]$ für alle λ.

Da es $\frac{1}{2}d(d-1)$ Paare (i, j) gibt, hat g_λ den Grad

$$\frac{d(d-1)}{2} = \frac{2^n u(d-1)}{2} = 2^{n-1}u(d-1)\,.$$

Wegen $n \geq 1$ ist $d-1$ ungerade, also auch $u(d-1)$; die Grade der g_λ sind daher nur durch 2^{n-1} teilbar, nicht aber durch 2^n. Somit können wir die Induktionsvoraussetzung anwenden und folgern, daß jedes der Polynome g_λ mindestens eine komplexe Nullstelle hat.

Die Nullstellen von g_λ sind die $w_{ij}(\lambda) \in K$; damit wissen wir, daß es zu jedem $\lambda \in \mathbb{R}$ ein Paar (i,j) gibt derart, daß $w_{ij}(\lambda)$ in $\mathbb{C}$ liegt.

Nun verwenden wir ein klassisches Beweisprinzip der Mathematik, das DIRICHLETsche Schubfachprinzip: Hat man n Schubfächer und verteilt mehr als n Objekte darauf, so müssen in mindestens einem dieser Schubfächer mindestens zwei Objekte liegen.

JOHANN PETER GUSTAV LEJEUNE DIRICHLET (1805 – 1859) wurde in der damals zu Frankreich gehörenden Stadt Düren geboren; er lehrte an den Universitäten Breslau, Berlin und Göttingen. 1828 gab er den ersten strengen Beweis für die Konvergenz von FOURIERreihen und untersuchte die Darstellbarkeit beliebiger Funktionen durch solche Reihen. Auch unser heutiger Funktionsbegriff geht auf DIRICHLET zurück. Sein wohl bekanntester Satz besagt, daß eine arithmetische Progression, deren Glieder keinen gemeinsamen Teiler haben, unendlich viele Primzahlen enthält.

Unsere Objekte sind die reellen Zahlen λ, die Schubfächer sind die Paare (i,j). Es gibt $\frac{1}{2}d(d-1)$ Schubfächer, aber unendlich viele reelle Zahlen, also muß es zwei Werte $\lambda \neq \lambda'$ und ein Paar (i,j) geben, so daß sowohl $w_{ij}(\lambda)$ als auch $w_{ij}(\lambda')$ in $\mathbb{C}$ liegen.

Gehen wir zurück zur Definition der w_{ij}, sehen wir, daß

$$z_i + z_j + \lambda z_i z_j = w_{ij}(\lambda) \quad \text{und} \quad z_i + z_j + \lambda' z_i z_j = w_{ij}(\lambda')$$

beides komplexe Zahlen sind, also auch

$$z_i z_j = \frac{w_{ij}(\lambda') - w_{ij}(\lambda)}{\lambda' - \lambda} \quad \text{und} \quad z_i + z_j = w_{ij}(\lambda) - \lambda z_i z_j \,.$$

Aus §2 von Kapitel 1 wissen wir, daß wir zwei Zahlen, deren Produkt P und Summe S wir kennen, als Lösungen der quadratischen Gleichung $x^2 - Sx + P = 0$ bestimmen können, und dort haben wir auch gesehen,

daß wir zu jeder komplexen Zahl eine komplexe Quadratwurzel finden können, so daß sich die Lösungsformel für quadratische Gleichungen auch im Komplexen anwenden läßt und komplexe Lösungen liefert. Somit sind z_i und z_j komplex, f hat also in der Tat mindestens eine komplexe Nullstelle.

∎

Korollar: Jedes nichtkonstante Polynom $f \in \mathbb{C}[X]$ hat mindestens eine komplexe Nullstelle.

Beweis: Wir betrachten zu $f = a_d X^d + \cdots + a_0$ das Polynom

$$\overline{f} = \overline{a_d} X^d + \cdots + \overline{a_0}$$

mit den konjugiert komplexen Koeffizienten und multiplizieren die beiden miteinander. Der Koeffizient von X^r in $f\overline{f}$ ist die Summe aller Produkte $a_i\overline{a_j}$ mit $i + j = r$. Ist $i \neq j$, kommt also in der Summe außer dem Summanden $a_i\overline{a_j}$ auch noch $a_j\overline{a_i}$ vor. Diese beiden Zahlen sind konjugiert komplex, so daß ihre Summe reell ist. Für gerade r gibt es noch einen Term der Form $a_i\overline{a_i} = |a_i|^2$; auch der ist reell. Also ist die gesamte Summe reell, und damit ist $f\overline{f} \in \mathbb{R}[X]$. Nach dem gerade bewiesenen Satz hat $f\overline{f}$ mindestens eine komplexe Nullstelle z; es gibt also ein $z \in \mathbb{C}$, so daß $f(z)\overline{f}(z) = 0$ ist. Dann ist entweder $f(z) = 0$, und wir sind fertig, oder $\overline{f}(z) = 0$. In diesem Fall ist auch $\overline{\overline{f}(z)} = f(\overline{z}) = 0$, also ist $\overline{z}$ eine komplexe Nullstelle von f.

∎

Induktiv folgt sofort der

Fundamentalsatz der Algebra: Jedes Polynom $f \in \mathbb{C}[X]$ vom Grad $d \geq 1$ zerfällt vollständig in Linearfaktoren, läßt sich also schreiben als

$$f = a(X - z_1) \cdots (X - z_d) \quad \text{mit} \quad a, z_1, \ldots, z_d \in \mathbb{C}\,.$$

∎

Ist k ein Teilkörper von $\mathbb{C}$ und $f \in k[X]$ ein irreduzibles Polynom über k, so zerfällt f über $\mathbb{C}$ in Linearfaktoren:

$$f = a(X - z_1) \cdots (X - z_d) \quad \text{mit} \quad a \in k, \;\; z_1, \ldots, z_d \in \mathbb{C}\,.$$

Der kleinste Teilkörper von $\mathbb{C}$, der sowohl k als auch die Elemente $z_1, \ldots, z_d$ enthält, ist daher ein Zerfällungskörper von f über k.

Definition: Ist K/k eine Körpererweiterung und sind $z_1, \ldots, z_r$ Elemente von K, so bezeichnen wir mit $k(z_1, \ldots, z_r)$ den kleinsten Teilkörper von K, der sowohl k als auch die Elemente $z_1, \ldots, z_r$ enthält.

Dieser Körper existiert; er ist einfach der Durchschnitt aller Teilkörper von K, die sowohl k als auch die sämtlichen z_i enthalten. Wir sagen, $k(z_1, \ldots, z_r)$ entstehe aus k durch *Adjunktion* der Elemente $z_1, \ldots, z_r$.

Beispielsweise ist $\mathbb{Q}\big(\sqrt{2}\big) = \mathbb{Q} \oplus \mathbb{Q}\sqrt{2}$ und $\mathbb{Q}\big(\sqrt[3]{2}\big) = \mathbb{Q} \oplus \mathbb{Q}\sqrt[3]{2} \oplus \mathbb{Q}\sqrt[3]{4}$ in $\mathbb{R}/\mathbb{Q}$; dabei bezeichnet $\oplus$ die direkte Summe von $\mathbb{Q}$-Vektorräumen.

Für ein irreduzibles Polynom über $\mathbb{Q}$ oder einem anderen Teilkörper von $\mathbb{C}$ haben wir somit zwei wesentlich verschiedene Zugänge zum Zerfällungskörper: Einmal durch Adjunktion der komplexen Nullstellen (wie immer wir die bekommen) oder rein formal durch die Restklassenkonstruktion, mit der wir oben die Existenz des Zerfällungskörpers allgemein bewiesen haben. Natürlich sollten wir uns fragen, was diese beiden Zerfällungskörper miteinander zu tun haben.

Lemma: $\varphi\colon k \to k'$ sei ein Isomorphismus von Körpern,

$$f = a_d X^d + \cdots + a_1 X + a_0 \in k[X]$$

sei ein Polynom über k und

$$f' = \varphi(a_d) X^d + \cdots + \varphi(a_1) X + \varphi(a_0)$$

das entsprechende Polynom aus $k'[X]$. Weiter seien K/k und K'/k' Zerfällungskörper von f *bzw.* f'. Dann gibt es einen Isomorphismus $\Phi\colon K \to K'$, der φ fortsetzt.

Beweis: In $K[X]$ läßt sich das Polynom f als Produkt von Linearfaktoren schreiben: $f = a_d(X - z_1) \cdots (X - z_d)$ mit $z_i \in K$. Wir beweisen den Satz durch Induktion nach der Anzahl r jener z_i, die *nicht* in k liegen.

Im Fall $r = 0$ zerfällt das Polynom bereits über k in Linearfaktoren, d.h. $K = k$. Außerdem ist dann auch

$$f' = \varphi(a_d)\big(X - \varphi(z_1)\big) \cdots \big(X - \varphi(z_d)\big) \quad \text{mit} \quad \varphi(z_i) \in k' \,,$$

so daß auch $K' = k'$ ist und wir einfach $\Phi = \varphi$ setzen können.

Der Fall $r = 1$ tritt nicht auf, denn liegen etwa $z_2, \ldots, z_d$ in k, so muß auch z_1 in k liegen, denn nach dem Wurzelsatz von VIÈTE ist

$$z_1 = -\frac{a_{d-1}}{a_d} - z_2 - \cdots - z_d \,.$$

Sei nun $r > 1$. Dann hat f mindestens einen irreduziblen Faktor g vom Grad größer eins. Durch Umnummerieren der Nullstellen können wir erreichen, daß z_1 eine Nullstelle von g ist. Nun betrachten wir das Polynom $g' \in k'[X]$, das aus g entsteht, indem wir alle Koeffizienten durch ihr Bild unter φ ersetzen. Offensichtlich ist g' ein irreduzibler Faktor von f'; das Element $z_1' \in K'$ sei eine Nullstelle von g'.

Im Körper $k(z_1)$ hat g mindestens eine Nullstelle, nämlich z_1, und in $k'(z_1')$ hat g' mindestens eine Nullstelle, nämlich z_1'. Außerdem ist

$$k(z_1) \cong k[X]/(g) \cong k'[X]/(g') \cong k'(z_1') \,.$$

Indem wir $\widetilde{\varphi}(z_1) = z_1'$ setzen, können wir φ daher fortsetzen zu einen Isomorphismus $\widetilde{\varphi}\colon k(z_1) \to k'(z_1')$.

Da $k(z_1)$ in K liegt und $k'(z_1')$ in K', können wir das zu beweisende Lemma auch für die Zerfällungskörper $K/k(z_1)$ und $K'/k'(z_1')$ und den Morphismus $\widetilde{\varphi}$ formulieren. Da r dann um mindestens eins kleiner ist, gilt es hier nach Induktionsannahme, und damit gilt es auch für K/k und K'/k'.

∎

Korollar: Je zwei Zerfällungskörper K, K' eines Polynoms $f \in k[X]$ sind isomorph.

Beweis: Wir müssen nur das gerade bewiesene Lemma auf den Fall anwenden, daß $k' = k$ ist und φ die Identität.

∎

Was uns wirklich interessiert ist natürlich dieses Korollar. Die allgemeinere Formulierung im Lemma war notwendig, da selbst im Fall $k = k'$ die Körper $k(z_1)$ und $k(z_1')$ im allgemeinen nicht gleich sind, sondern nur isomorph. Der Induktionsbeweis funktionierte daher nur, weil wir in der Formulierung des Lemmas von der allgemeineren Situation isomorpher Körper ausgegangen sind.

§2: Automorphismen von Körpererweiterungen

Um die Nullstellenmenge eines Polynoms $f \in k[X]$ zu untersuchen, können wir den Zerfällungskörper K von f betrachten. Wenn wir den konstruieren können als eine Folge von Körpererweiterungen, die jeweils durch Adjunktion der Wurzel eines Elements entstehen, können wir alle Elemente von K und damit insbesondere auch die Nullstellen von f ausgehend von k durch Wurzelausdrücke beschreiben.

Wenn wir etwa eine kubische Gleichung $x^3 + px + q = 0$ für zwei rationale Zahlen p, q lösen wollen, betrachten wir nach der Lösungsformel zunächst die Zahl

$$U = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3},$$

sodann die dritten Wurzeln u_1, u_2 und u_3 von U, und erhalten schließlich die Lösungen

$$x_1 = u_1 - \frac{p}{3u_1}, \quad x_2 = u_2 - \frac{p}{3u_2} \quad \text{und} \quad x_3 = u_3 - \frac{p}{3u_3}.$$

Wir berechnen also zunächst in $\mathbb{Q}$ die Zahl $\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3$; falls sie in $\mathbb{Q}$ ein Quadrat ist, können wir U als rationale Zahl berechnen. Im allgemeinen wird Δ kein Quadrat sein; dann gehen wir über zum Körper $k_1 = \mathbb{Q}(\sqrt{\Delta})$ mit $[k_1 : \mathbb{Q}] = 2$. Dort können wir das Element U berechnen und müssen schauen, ob alle dritten Wurzeln von U in k_1 liegen. Auch das wird im allgemeinen nicht der Fall sein; dann gehen wir weiter zum Zerfällungskörper k_2 des Polynoms $X^3 - U$. Dort liegen die drei Kubikwurzeln u_1, u_2, u_3 von U und damit auch die drei Lösungen x_1, x_2, x_3 der Gleichung. Wie wir im ersten Kapitel gesehen haben, bedeutet das aber nicht unbedingt, daß $k_2/\mathbb{Q}$ der Zerfällungskörper des Polynoms $X^3 + pX + q$ sein muß: Selbst im Falle dreier ganzzahliger Lösungen sind für die Berechnung von $\sqrt{\Delta}$ und der dritten Wurzeln von U Körpererweiterungen notwendig. Immerhin wissen wir, daß der Zerfällungskörper ein Teilkörper von k_2 ist.

In diesem Paragraphen wollen wir versuchen, die Struktur einer Körpererweiterung über ihre Zwischenkörper zu verstehen; diese Zwischenkörper wiederum wollen wir mit Hilfe von Automorphismen in den Griff bekommen.

Homomorphismen, Isomorphismen, Automorphismen, *usw.* von Körpern sind natürlich einfach Ringhomomorphismen, -isomorphismen, -automorphismen, *usw.;* bei Körpern sind diese allerdings automatisch injektiv:

Lemma: Ist k ein Körper und R ein Ring, so ist jeder Ringhomomorphismus $\varphi\colon k \to R$ injektiv.

Beweis: Kern φ ist ein Ideal von k; falls es das Nullideal ist, sind wir fertig. Andernfalls gibt es mindestens ein Element $x \neq 0$, und wegen der Idealeigenschaft liegen auch alle Vielfachen von x im Kern. Da in einem Körper alle Elemente außer der Null invertierbar sind, ist jedes $y \in k$ ein Vielfaches von x: $y = x \cdot (x^{-1}y)$, so daß Kern $\varphi = k$ wäre. Das ist aber nicht möglich, denn zumindest die Eins muß bei einem Ringhomomorphismus auf 1 abgebildet werden. ∎

Die folgende Diskussion des Zusammenhangs zwischen Automorphismen und Zwischenkörpern folgt im wesentlichen der besonders kompakten und einfachen Darstellung aus

EMIL ARTIN: Galoissche Theorie, *Leipzig 1959* (u.a.)

Lemma: $\sigma_1, \ldots, \sigma_r\colon k \to K$ seien paarweise verschiedene Homomorphismen des Körpers k in einen Körper K. Dann sind $\sigma_1, \ldots, \sigma_r$ linear unabhängig im folgenden Sinne: Ist $a_1\sigma_1(x) + \cdots + a_r\sigma_r(x) = 0$ für alle $x \in k$ mit irgendwelchen Koeffizienten $a_i \in K$, so müssen alle a_i verschwinden.

Beweis durch Induktion nach r. Im Falle $r = 1$ können wir einfach $x = 1$ einsetzen und erhalten $a_1 = a_1\sigma(1) = 0$; die Behauptung ist also richtig.

Nun sei $r > 1$ und $a_1\sigma_1(x) + \cdots + a_r\sigma_r(x) = 0$ für alle $x \in k$. Für jedes $y \in k$ gilt dann auch

$$a_1\sigma_1(xy) + \cdots + a_r\sigma_r(xy)$$
$$= a_1\sigma_1(x)\sigma_1(y) + \cdots + a_r\sigma_r(x)\sigma_r(y) = 0\,.$$

Wenn wir die ursprüngliche Gleichung mit $\sigma_r(y)$ multiplizieren, erhalten wir die weitere Gleichung

$$a_1\sigma_1(x)\sigma_r(y) + \cdots + a_r\sigma_r(x)\sigma_r(y) = 0\,.$$

Subtraktion der letzten beiden Gleichungen voneinander liefert die neue Gleichung

$$a_1\big(\sigma_1(y)-\sigma_r(y)\big)\sigma_1(x)+\cdots+a_{r-1}\big(\sigma_{r-1}(y)-\sigma_r(y)\big)\sigma_{r-1}(x)=0$$

für alle $x \in k$. Da diese nur $r-1$ Summanden enthält, verschwinden nach Induktionsannahme alle Koeffizienten, insbesondere der Koeffizient $a_1\big(\sigma_1(y)-\sigma_r(y)\big)$ von $\sigma_1(x)$. Da σ_1 und σ_r verschiedenen Homomorphismen sind, können wir ein $y \in k$ finden, für das $\sigma_1(y) \neq \sigma_r(y)$ ist, und wenn wir mit diesem y arbeiten, sehen wir, daß der Koeffizient a_1 verschwinden muß. Unsere Beziehung ist daher von der Form $a_2\sigma_2(x)+\cdots+a_r\sigma_r(x)=0$, und da auch hier nur $r-1$ Homomorphismen vorkommen, zeigt eine nochmalige Anwendung der Induktionsannahme das Verschwinden der restlichen Koeffizienten $a_2,\ldots,a_r$. ∎

Definition: *a)* Ist K/k eine Körpererweiterung, so bezeichnen wir mit $\mathrm{Aut}(K/k)$ die Menge aller (Körper)-Automorphismen $\sigma\colon K \to K$, die auf k die Identität sind, d.h. $\sigma(x)=x$ für alle $x \in k$.
b) Ist G eine Menge von Automorphismen des Körpers K, so bezeichnen wir

$$K^G = \{x \in K \mid \sigma(x)=x \ \text{ für alle } \sigma \in G\}$$

als den Fixkörper von G.

Es ist klar, daß $\mathrm{Aut}(K/k)$ eine Gruppe ist und K^G ein Teilkörper von K. Im Falle einer endlichen Gruppe $G=\{\sigma_1,\ldots,\sigma_n\}$ haben wir zwei Abbildungen von K nach K^G, gegeben durch

$$S(x)=\sigma_1(x)+\cdots+\sigma_n(x) \quad \text{und} \quad N(x)=\sigma_1(x)\cdot\cdots\cdot\sigma_n(x)\,.$$

$S(x)$ heißt die *Spur* von x, $N(x)$ die *Norm.* Beide liegen in K^G, denn für jedes $\sigma \in G$ ist

$$\sigma\big(S(x)\big)=\sigma\big(\sigma_1(x)+\cdots+\sigma_n(x)\big)=\sigma\circ\sigma_1(x)+\cdots+\sigma\circ\sigma_n(x)=S(x)$$

und

$$\sigma\big(N(x)\big)=\sigma\big(\sigma_1(x)\cdot\cdots\cdot\sigma_n(x)\big)=\sigma\circ\sigma_1(x)\cdot\cdots\cdot\sigma\circ\sigma_n(x)=N(x)\,,$$

denn da die Multiplikation mit σ eine bijektive Abbildung von G nach G definiert, ist auch die Menge aller $\sigma\circ\sigma_i$ gleich G. Natürlich ist weder

die Spur noch die Norm ein Ringhomomorphismus; immerhin ist die Spur ein Homomorphismus der additiven Gruppen und die Norm einer der multiplikativen Gruppen. Die Spurabbildung kann nicht gleich der Nullabbildung sein, denn wäre $\sigma_1(x) + \cdots + \sigma_n(x) = 0$ für alle $x \in K$, wären die Automorphismen $\sigma_1, \ldots, \sigma_n$ linear abhängig, im Widerspruch zum obigen Lemma.

Lemma: Ist G eine endliche Menge von Automorphismen eines Körpers K, so ist $[K : K^G] \geq |G|$.

Wir beweisen dieses Lemma im Hinblick auf eine spätere Anwendung gleich etwas allgemeiner als

Lemma: Ist G eine endliche Menge von Homomorphismen des Körpers K in einen Körper L und ist

$$k = \{x \in K \mid \sigma(x) = \tau(x) \text{ für alle } \sigma, \tau \in G\} ,$$

so ist $[K : k] \geq |G|$.

Daraus folgt das vorige Lemma, denn für $K = L$ und $G' = G \cup \{\mathrm{id}_K\}$ ist $k = K^{G'} = K^G$, und nach dem Lemma ist $[K : k] \geq |G'| \geq |G|$.

Beweis des zweiten Lemmas: Konkret sei $G = \{\sigma_1, \ldots, \sigma_n\}$. Wir nehmen an, daß der Grad $[K : k] = r < n$ sei, und wollen daraus einen Widerspruch ableiten.

Da $[K : k] = r$ ist, gibt es r Elemente $b_1, \ldots, b_r \in K$, die eine k-Basis von K bilden. Wir betrachten über K das homogene lineare Gleichungssystem aus den r Gleichungen

$$\sigma_1(b_i)x_1 + \sigma_2(b_i)x_2 + \cdots + \sigma_n(b_i)x_n = 0$$

für $i = 1, \ldots, r$. Da wir mehr Variablen als Gleichungen haben, muß es nichttriviale Lösungen geben; eine davon sei $(x_1, \ldots, x_n)$. Weiter sei x ein beliebiges Element von K; wir schreiben es als k-Linearkombinationen $x = a_1b_1 + \cdots + a_rb_r$ der Basiselemente. Da die a_i in k liegen, ist $\sigma_j(a_i) = \sigma_1(a_i)$ und $\sigma_j(a_ib_i) = \sigma_1(a_i)\sigma_j(b_i)$ für alle i, j. Multiplizieren wir die i-te Gleichung des obigen Systems mit $\sigma_1(a_i)$, können wir das Ergebnis daher auch schreiben als

$$\sigma_1(a_ib_i)x_1 + \sigma_2(a_ib_i)x_2 + \cdots + \sigma_n(a_ib_i)x_n = 0 .$$

Addieren wir diese Gleichungen für $i = 1, \ldots, r$, hat x_j in der Summe den Koeffizienten

$$\begin{aligned}&\sigma_j(a_1b_1) + \sigma_j(a_2b_2) + \cdots + \sigma_j(a_rb_r)\\ &= \sigma_j(a_1b_1 + a_2b_2 + \cdots + a_rb_r) = \sigma_j(x)\,.\end{aligned}$$

Die Summe der r Gleichungen ist daher

$$x_1\sigma_1(x) + x_2\sigma_2(x) + \cdots + x_n\sigma_n(x) = 0\,.$$

Da x als beliebiges Element von K vorausgesetzt war, gilt dies für alle $x \in K$ und widerspricht somit der oben gezeigten linearen Unabhängigkeit von Körperhomomorphismen. Also kann r nicht kleiner als n sein, was das Lemma beweist. ∎

Für eine *Gruppe* G von Automorphismen können wir das erste Lemma verschärfen zu

Satz: Ist G eine endliche Gruppe von Automorphismen eines Körpers K, so ist $[K : K^G] = |G|$.

Beweis: Sei wieder $G = \{\sigma_1, \ldots, \sigma_n\}$. Da wir bereits wissen, daß $[K : K^G] \geq |G|$ ist, muß nur noch gezeigt werden, daß je $n + 1$ Elemente $b_1, \ldots, b_{n+1}$ von K linear abhängig über K^G sind. Auch dazu betrachten wir ein homogenes lineares Gleichungssystem mit weniger Gleichungen als Variablen; die i-te der n Gleichungen sei

$$\sigma_i^{-1}(b_1)x_1 + \sigma_i^{-1}(b_2)x_2 + \cdots + \sigma_i^{-1}(b_{n+1})x_{n+1} = 0\,.$$

Wieder muß es eine nichttriviale Lösung $(x_1, \ldots, x_{n+1})$ geben; durch Umindizieren der b_i können wir erreichen, daß $x_1 \neq 0$ ist. Für jedes $\lambda \neq 0$ aus K ist auch $(\lambda x_1, \ldots, \lambda x_{n+1})$ eine nichttriviale Lösung; durch geeignete Wahl von λ können wir also x_1 zu einem beliebigen Element von $K^\times$ machen. Wie wir wissen, ist die Spurabbildung nicht gleich der Nullabbildung; wir wählen x_1 so, daß $S(x_1) \neq 0$ ist.

Als nächstes wenden wir im obigen System auf die i-te Gleichung den Automorphismus σ_i an und erhalten

$$b_1\sigma_i(x_1) + b_2\sigma_i(x_2) + \cdots + b_{n+1}\sigma_i(x_{n+1}) = 0\,.$$

Die Summe aller dieser Gleichungen ist

$$b_1 S(x_1) + b_2 S(x_2) + \cdots + b_{n+1} S(x_{n+1}) = 0\,,$$

wobei die Spuren $S(x_i)$ im Fixkörper K^G liegen und nicht alle verschwinden, da zumindest $S(x_1) \neq 0$ ist. Somit sind $b_1, \ldots, b_{n+1}$ linear abhängig über K^G.

■

Als erstes Resultat über den Zusammenhang zwischen Automorphismengruppen und Teilkörpern erhalten sofort das folgende

Korollar: Sind G und H zwei verschiedene Gruppen von Automorphismen eines Körpers K, so sind K^G und K^H verschiedene Teilkörper.

Beweis: Von zwei verschiedenen Gruppen enthält mindestens eine ein Element, das nicht in der anderen enthalten ist. Nehmen wir an, H enthalte einen Automorphismus σ von K, der nicht in G liegt. Wäre $K^G = K^H$, so müßte σ den Körper K^G punktweise festlassen, K^G wäre also auch der Fixkörper der Menge $G \cup \{\sigma\}$. Nach dem Lemma vor dem gerade bewiesenen Satz wäre damit $[K : K^G] \geq |G| + 1$, aber nach dem Satz ist $[K : K^G] = |G|$.

■

Leider ist nicht jeder Teilkörper Fixkörper einer Automorphismengruppe: Für $\mathbb{Q}(\sqrt{2}/\mathbb{Q}$ haben wir zwar eine Automorphismengruppe der Ordnung zwei, bestehend aus der Identität und der Abbildung, die der Zahl $a + b\sqrt{2}$ deren konjugiertes Element $a - b\sqrt{2}$ zuordnet, doch schon für $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ gibt es aber nichts entsprechendes mehr: Jeder Automorphismus $\sigma \in \mathrm{Aut}\big(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}\big)$ muß $\sqrt[3]{2}$ auf eine Zahl x mit $x^3 = 2$ abbilden, und da wir in einem Teilkörper der reellen Zahlen sind, läßt das nur die Möglichkeit $\sigma(\sqrt[3]{2}) = \sqrt[3]{2}$ zu. Also wird auch das Quadrat von $\sqrt[3]{2}$ auf sich selbst abgebildet und damit ganz $\mathbb{Q}(\sqrt[3]{2})$; es gibt also keinen Automorphismus außer der Identität.

Selbst $\mathrm{Aut}(\mathbb{R}/\mathbb{Q})$ besteht nur aus der Identität: Wir betrachten einen beliebigen Automorphismus $\varphi\colon \mathbb{R} \to \mathbb{R}$ der auf $\mathbb{Q}$ die Identität ist. (Wie man sich leicht überlegt, gilt das für jeden Automorphismus von $\mathbb{R}$ automatisch.) Da eine reelle Zahl x genau dann größer oder gleich Null ist, wenn es ein $w \in \mathbb{R}$ gibt mit $w^2 = x$, muß φ nichtnegative

Zahlen auf nichtnegative Zahlen abbilden, denn $\varphi(w^2) = \varphi(w)^2$. Wegen $\varphi(x) - \varphi(y) = \varphi(x - y)$ muß dann auch für alle $x \leq y$ gelten, daß $\varphi(x) \leq \varphi(y)$ ist. Nun kann man für jede reelle Zahl x eine rationale Intervallschachtelung $\big([a_n, b_n]\big)_{n\in\mathbb{N}}$ angeben, d.h. eine Folge von Intervallen mit $a_n, b_n \in \mathbb{Q}$ derart, daß $x \in [a_n, b_n]$ für alle n und $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ für alle n und $\lim_{n\to\infty}(b_n - a_n) = 0$. Da $\varphi(a_n) = a_n$ und $\varphi_n(b_n) = b_n$ ist, liegt auch $\varphi(x)$ in allen diesen Intervallen; also muß, wegen der letzten Bedingung, $\varphi(x) = x$ sein. Um solche Beispiele zumindest vorläufig auszuschließen definieren wir

Definition: Eine endliche Körpererweiterung K/k heißt GALOISsch, wenn es eine Gruppe G von Automorphismen von K gibt, für die $k = K^G$ ist.

Natürlich muß dann nach obigem Korollar $G = \mathrm{Aut}(K/k)$ sein; wir bezeichnen diese Gruppe als die GALOIS-Gruppe von K/k.

ÉVARISTE GALOIS (1811 – 1832) wurde in Bourg La Reine in der Nähe von Paris geboren. Obwohl die französische Revolution zu seiner Zeit schon Jahrzehnte zurücklag, war er stark von ihr geprägt und überzeugter Republikaner, der deshalb immer wieder ins Gefängnis kam. In seiner Jugend wurde er nur von seiner Mutter unterrichtet; erst 1823 ging er auf eine Schule, und 1827 besuchte er erstmalig eine Mathematikklasse. Die Mathematik begeisterte ihn so sehr, daß er darüber alle anderen Fächer vernachlässigte. Trotzdem schaffte er 1828 nicht die Aufnahmeprüfung zur École polytechnique. 1829 veröffentlichte er seine erste mathematische Arbeit; sie handelte von Kettenbrüchen. 1830 folgte eine Arbeit über die Lösung algebraischer Gleichungen. Nachdem er eine posthum veröffentlichte Arbeit von ABEL über dieses Thema gelesen hatte, schrieb er, auf CAUCHYs Rat hin, eine Arbeit, die dessen Ergebnisse mit seinen kombinierte. Er reichte sie 1830 bei FOURIER, dem damaligen Sekretär der Akademie der Wissenschaften ein; nachdem dieser kurz darauf starb, ist diese Arbeit bis heute verschollen. Kurz vor einem Duell, dessen Hintergrund nicht ganz klar ist, schrieb er seine Resultate noch einmal kurz auf; am Tag nach dem Duell starb er an dessen Folgen.

Bevor wir uns überlegen, wie wir einer Körpererweiterung ansehen können, ob sie GALOISsch ist oder nicht, und ob diese Erweiterungen für uns nützlich sind, wollen wir uns zunächst überlegen, daß wir für

GALOISsche Erweiterungen in der Tat alles über die Zwischenkörper aus der Automorphismengruppe ablesen können.

Eine wesentliche Eigenschaft GALOISscher Erweiterungen ist die zunächst eher überflüssig erscheinende Bedingung der Separabilität:

Definition: Ein Polynom $f \in k[X]$ über einem Körper k heißt *separabel,* wenn keiner seiner irreduziblen Faktoren im Zerfällungskörper von f eine mehrfache Nullstelle hat. Für eine Körpererweiterung K/k heißt ein Element $x \in K$ separabel, falls es Nullstelle eines separablen Polynoms aus $k[X]$ ist. K/k heißt separabel, wenn jedes Element $x \in K$ separabel über x ist.

Im Falle $k = \mathbb{R}$ ist offensichtlich jedes Polynom separabel: Hat nämlich ein irreduzibles Polynom $f \in \mathbb{R}[X]$ eine mehrfache Nullstelle, so ist diese auch eine Nullstelle der Ableitung f'. Damit ist $\mathrm{ggT}(f, f') \neq 1$. Andererseits ist aber $\mathrm{ggT}(f, f')$ ein Teiler von f, also entweder assoziiert zu eins oder zu f. Da f' kleineren Grad als f hat und im Falle eines Polynoms mit einer mehrfachen Nullstelle nicht konstant sein kann, ist auch das unmöglich. Also gibt es in $\mathbb{R}[X]$ keine nichtseparablen Polynome.

Für Polynome über einen beliebigen Körper k können wir natürlich nicht von einer über Grenzwerte definierten Ableitung reden. Wir können sie aber trotzdem formal definieren:

Definition: k sei ein Körper. Die (formale) Ableitung eines Polynoms $f = \sum_{i=0}^{d} a_i X^i \in k[X]$ ist das Polynom $f' = \sum_{i=1}^{d} i a_i X^{i-1}$.

Für $k = \mathbb{R}$ und $k = \mathbb{C}$ ist das natürlich die übliche Ableitung.

Aus der Definition folgt sofort, daß

$$\begin{cases} k[X] \to k[X] \\ \quad f \mapsto f' \end{cases}$$

eine k-lineare Abbildung ist. Um zu sehen, daß die formale Ableitung auch die Produktregel erfüllt, vergleichen wir die beiden Abbildungen

$$\begin{cases} k[X] \to k[X] \\ \quad f \mapsto (fg)' \end{cases} \quad \text{und} \quad \begin{cases} k[X] \to k[X] \\ \quad f \mapsto fg' + f'g \end{cases}.$$

Beide sind linear, und für $g = \sum_{j=0}^{e} b_j X^j$ wird das Basiselement X^i für $i \geq 0$ von der ersten abgebildet auf $\sum_{j=0}^{e}(i+j)b_j X^{i+j-1}$, und von der zweiten auf

$$\sum_{j=1}^{e} j b_j X^{i+j-1} + \sum_{j=0}^{e} i b_j X^{i+j-1} = \sum_{j=0}^{e}(i+j)b_j X^{i+j-1} .$$

Somit stimmen beide Abbildungen überein, und die Produktregel ist bewiesen.

Wie in der Analysis gilt

Lemma: Die Nullstelle $z \in k$ des Polynoms $f \in k[X]$ hat genau dann eine Vielfachheit größer eins, wenn sie auch Nullstelle von f' ist.

Beweis: Da z Nullstelle von f ist, gibt es ein Polynom $g \in k[X]$ derart, daß $f = (X - z)g$ ist. Nach der Produktregel ist $f' = (X - z)g' + g$, d.h. $f'(z)$ verschwindet genau dann, wenn $g(z)$ verschwindet, und das ist genau dann der Fall, wenn z eine mehrfache Nullstelle von f ist. ∎

Für ein Polynom aus $\mathbb{R}[X]$ (und für jede differenzierbare Funktion $f\colon \mathbb{R} \to \mathbb{R}$) ist f' genau dann identisch Null, wenn f konstant ist. Dies gilt nicht für Polynome über einem beliebigen Körper: Sei etwa $\mathbb{F}_p = \mathbb{Z}/p$ der Körper mit p Elementen. (Wenn wir $\mathbb{Z}/p$ als Körper betrachten, schreiben wir meist $\mathbb{F}_p$; das $\mathbb{F}$ steht für *finit,* also endlich.) Dann hat das Polynom $f = X^p$ die Ableitung $f' = pX^{p-1} = 0$, denn in $\mathbb{F}_p$ ist $p = 0$. Trotzdem ist f ein nichtkonstantes Polynom.

Um zu sehen, wann so etwas passieren kann, betrachten wir zu einem beliebigen Körper k den einzig möglichen Ringhomomorphismus $\chi\colon \mathbb{Z} \to k$. Er bildet die Eins auf die Eins ab, und jede ganze Zahl n auf das n-fache der Eins von k. Sein Kern ist ein Ideal von $\mathbb{Z}$, also ein Hauptideal (m).

Definition: Die *Charakteristik* eines Körpers k ist jene Zahl $m \in \mathbb{N}_0$, für die Kern $\chi = (m)$ ist. Wir schreiben $m = \operatorname{char} k$.

Lemma: Die Charakteristik eines Körpers ist entweder Null oder eine Primzahl.

Beweis: Wäre $\operatorname{char} k = ab$ eine zusammengesetzte Zahl mit $a, b > 1$, so wäre $\chi(a) \cdot \chi(b) = \chi(ab) = 0$, aber $\chi(a)$ und $\chi(b)$ beide von Null verschieden. Da Körper nullteilerfrei sind, ist das nicht möglich. ∎

Natürlich ist $\operatorname{char} \mathbb{Q} = \operatorname{char} \mathbb{R} = 0$ und $\operatorname{char} \mathbb{F}_p = p$ für jede Primzahl p. Über einem Körper der Charakteristik Null ist die Ableitung iX^{i-1} von X^i für $i \geq 1$ stets von Null verschieden; hier ist also $f' = 0$ genau dann, wenn f konstant ist. Über einem Körper der Charakteristik $p > 0$ ist dagegen die Ableitung von X^{np} für jedes $n \in \mathbb{N}_0$ gleich Null; die übrigen X-Potenzen haben nichtverschwindende Ableitungen. Die Ableitung eines Polynoms verschwindet daher genau dann, wenn alle Exponenten durch p teilbar sind.

Lemma: Über einem Körper der Charakteristik Null sind alle Polynome separabel.

Beweis: Wir können genau so vorgehen, wie im Fall $k = \mathbb{R}$: Hat ein irreduzibles Polynom f eine mehrfache Nullstelle, so ist diese auch eine Nullstelle von f' und damit von $\operatorname{ggT}(f, f')$. Da f eine Nullstelle hat, ist f nicht konstant, so daß f' nicht das Nullpolynom sein kann, sondern einen echt kleineren Grad als f hat. Damit ist $\operatorname{ggT}(f, f')$ ein Faktor positiven Grades von f mit echt kleinerem Grad als f. Dies widerspricht der Irreduzibilität von f. ∎

Auch über manchen Körpern positiver Charakteristik sind alle Polynome separabel, zum Beispiel über den Körpern $\mathbb{F}_p$. Um dies zu zeigen, betrachten wir zunächst einen speziellen Homomorphismus für Körper positiver Charakteristik und die Polynomringe darüber:

Lemma: Ist R ein Körper mit Charakteristik $p > 0$ oder ein Polynomring über einem solchen Körper, so ist die Abbildung

$$\begin{cases} R \to R \\ x \mapsto x^p \end{cases}$$

ein Ringhomomorphismus.

Beweis: Da R ein kommutativer Ring ist, muß natürlich $(xy)^p = x^p y^p$ sein. Bezüglich der Addition ist

$$(x+y)^p = x^p + \binom{p}{1} x^{p-1}y + \cdots + \binom{p}{p-1} xy^{p-1} + y^p \,,$$

und für alle i mit $1 \leq i \leq p-1$ ist $\binom{p}{i} = \frac{p\cdots(p-(i-1))}{i!} \equiv 0 \bmod p$, da zwar der Zähler, nicht aber der Nenner durch p teilbar ist. Somit ist $(x+y)^p = x^p + y^p$.

∎

Definition: Der Homomorphismus $x \mapsto x^p$ heißt FROBENIUS-Homomorphismus.

Damit können wir zeigen

Lemma: Jedes Polynom aus $\mathbb{F}_p[X]$ ist separabel.

Beweis: Wieder sei $f \in \mathbb{F}_p[X]$ ein irreduzibles Polynom. Falls f' nicht das Nullpolynom ist, können wir genauso vorgehen wie beim Beweis des entsprechenden Lemmas für Körper der Charakteristik Null. Andernfalls hat f, wie wir oben gesehen haben, die Form $f = \sum_{i=0}^{d} a_i X^{ip}$ mit einem $d > 0$, da f nicht konstant sein kann. Nach dem kleinen Satz von FERMAT ist in $\mathbb{F}_p$ jedes Element a_i gleich seiner p-ten Potenz, und außerdem ist nach dem vorigen Lemma für zwei Polynome g und h stets $(g+h)^p = g^p + h^p$. Somit ist

$$\left(\sum_{i=0}^{d} a_i X^i\right)^p = \sum_{i=0}^{d} a_i^p X^{ip} = \sum_{i=0}^{d} a_i X^{ip} = f \,,$$

im Widerspruch zur Irreduzibilität von f.

∎

Betrachten wir aber den Körper $k = \mathbb{F}_p(T)$ aller rationaler Funktionen über $\mathbb{F}_p$, also den Quotientenkörper des Polynomrings $\mathbb{F}_p[T]$, so können wir inseparable Polynome finden, etwa das Polynom $f = X^p - T$ aus $k[X]$. Es ist irreduzibel in $\mathbb{F}_p[T][X] = \mathbb{F}_p[T, X]$, da es linear in T ist, und nach dem Satz von GAUSS ist es somit auch irreduzibel in $k[X] = \mathbb{F}_p(T)[X]$.

Wie für jedes Polynom über einem Körper können wir dazu einen Körper K/k finden, in dem f eine Nullstelle s hat. In $K[X]$ ist nach dem vorletzten Lemma $(X-s)^p = X^p - s^p = X^p - T$. Die Zerlegung von f in irreduzible Faktoren im Zerfällungskörper ist also $(X-s)^p$; es gibt daher nur eine einzige Nullstelle, und die hat Vielfachheit p, so daß f nicht separabel ist.

In einer GALOISschen Erweiterung kann so etwas nicht passieren; hier gilt

Satz: Jede GALOISsche Erweiterung K/k ist separabel. Für ein Element $z \in K$ sei $M_z = \{\sigma(z) \mid \sigma \in \mathrm{Aut}(K/k)\}$; dann ist z eine Nullstelle des über k irreduziblen Polynoms

$$f = \prod_{w \in M_z} (X - w)\,.$$

Beweis: Da auch die Identität in $\mathrm{Aut}(K/k)$ liegt, ist $z \in M_z$ eine Nullstelle von f. Auch die Separabilität von f ist klar, denn die Elemente von M_z sind paarweise verschieden. Zu zeigen bleibt, daß f in $k[X]$ liegt und irreduzibel ist. Die Koeffizienten von f sind bis aufs Vorzeichen die elementarsymmetrischen Funktionen in den Nullstellen $w \in M_z$. Daher sind sie invariant unter $\mathrm{Aut}(K/k)$, liegen also, da K/k eine GALOISsche Erweiterung ist, in k. Ist $f = gh$ eine Zerlegung von f mit $g, h \in k[X]$, so muß mindestens einer der beiden Faktoren bei z verschwinden; sei etwa $g(z) = 0$. Dann ist für jedes $\sigma \in \mathrm{Aut}(K/k)$ auch

$$g\big(\sigma(z)\big) = \sigma\big(g(z)\big) = \sigma(0) = 0\,,$$

so daß alle $w \in M_z$ Nullstellen von g sind. Somit ist g assoziiert zu f und h eine Einheit, f also irreduzibel. Da alle Elemente von M_z verschieden sind, ist f auch separabel. ∎

Korollar: Ist K/k eine GALOISsche Erweiterung und $f \in k[X]$ ein irreduzibles Polynom, das in K eine Nullstelle z hat, so zerfällt f in $K[X]$ in Linearfaktoren.

Beweis: Wie wir am Ende des obigen Beweises gesehen haben, verschwindet ein Polynom mit Koeffizienten aus k, das bei einem $z \in K$

verschwindet, auch in allen $\sigma(z)$ mit $\sigma \in \mathrm{Aut}(K/k)$. Daher ist f teilbar durch das im Satz angegebene Polynom zu z. Wegen der Irreduzibilität von f unterscheiden sich die beiden höchstens um eine Einheit, also zerfällt auch f in Linearfaktoren. ∎

Die Bedingung, daß K/k eine GALOISsche Erweiterung sein soll, ist hier wesentlich: Beispielsweise zerfällt das Polynom $X^3 - 2$ über $\mathbb{Q}\big(\sqrt[3]{2}\big)$ nicht in Linearfaktoren, sondern nur in ein Produkt von $(X - \sqrt[3]{2})$ mit einem quadratischen Polynom.

Satz: L/k sei eine GALOISsche Erweiterung, und K sei ein Zwischenkörper. Dann ist auch L/K GALOISsch.

Beweis: $\mathrm{Aut}(L/K)$ ist die Untergruppe jener Automorphismen aus $\mathrm{Aut}(L/k)$, die jedes Element von K festlassen; ihre Gruppenordnung sei r und ihr Fixkörper sei K'. Natürlich ist $K \leq K'$; wir müssen zeigen, daß die beiden Körper gleich sind. Da $[L : K'] = r$ ist, genügt dazu, daß auch $[L : K] = r$ ist.

Ein Automorphismus $\sigma \in \mathrm{Aut}(L/k)$ bildet K ab auf einen Teilkörper $\sigma(K) \leq L$. Ein weiterer Automorphismus $\tau \in \mathrm{Aut}(L/k)$ stimmt genau dann auf K mit σ überein, wenn $\sigma^{-1}\tau$ die Identität auf K ist, wenn also $\sigma^{-1}\tau$ in $\mathrm{Aut}(L/K)$ liegt. Dies wiederum ist äquivalent dazu, daß die beiden Nebenklassen $\sigma\,\mathrm{Aut}(L/K)$ und $\tau\,\mathrm{Aut}(L/K)$ übereinstimmen.

Zwei Automorphismen $\sigma, \tau\colon L \to L$ definieren bei Einschränkung auf K somit genau dann die gleiche Abbildung, wenn sie in der gleichen Nebenklasse von $\mathrm{Aut}(L/k)$ modulo $\mathrm{Aut}(L/K)$ liegen. Die Anzahl dieser Nebenklassen ist der Index s von $\mathrm{Aut}(L/K)$ in $\mathrm{Aut}(L/k)$. Nehmen wir aus jeder Nebenklasse einen Vertreter, erhalten wir somit s verschiedene Homomorphismen von K nach L. Da sie in $\mathrm{Aut}(L/k)$ liegen, induzieren sie allesamt die Identität auf k. Nach einem der oben bewiesenen Lemmata ist daher $[K : k] \geq s$. Außerdem wissen wir, daß $[L : K] \geq r$ ist, also ist $[L : k] = [L : K][K : k] \geq rs$. Da L/k GALOISsch ist, ist $[L : k]$ die Ordnung von $\mathrm{Aut}(L/k)$. Die Ordnung dieser Gruppe ist nach LAGRANGE das Produkt der Ordnung der Untergruppe $\mathrm{Aut}(L/K)$ mit dem Index von $\mathrm{Aut}(L/K)$ in $\mathrm{Aut}(L/k)$, also rs. Damit ist einerseits

$[L : K][K : k] = rs$, andererseits $[L : K] \geq r$ und $[K : k] \geq s$. Das ist nur möglich, wenn $[L : K] = r$ und $[K : k] = s$ ist, und $[L : K] = r$ ist genau das, was wir beweisen wollten.

■

Damit können wir den Hauptsatz der GALOIS-Theorie beweisen:

Satz: L/k sei eine GALOISsche Erweiterung und $G = \mathrm{Aut}(L/k)$. Dann gibt es eine Bijektion zwischen der Menge aller Untergruppen von G und der Menge aller Körper K mit $k \leq K \leq L$, die jeder Untergruppe $H \leq G$ deren Fixkörper L^H zuordnet und jedem Zwischenkörper K die Gruppe $\mathrm{Aut}(L/K)$. Ist $H \leq H' \leq G$, so ist $K^H \geq K^{H'}$. Außerdem ist $[L : L^H] = |H|$ und $[L^H : k] = [G : H]$.

Beweis: $\mathcal{U}$ sei die Menge aller Untergruppen von G und $\mathcal{Z}$ die Menge aller Zwischenkörper K mit $k \leq K \leq L$. Wir müssen zeigen, daß die beiden Abbildungen

$$\begin{cases} \mathcal{U} \to \mathcal{Z} \\ H \mapsto L^H \end{cases} \quad \text{und} \quad \begin{cases} \mathcal{Z} \to \mathcal{U} \\ K \mapsto \mathrm{Aut}(L/K) \end{cases}$$

zueinander invers sind. Ausgehend von einer Untergruppe $H \leq G$ müssen wir also zeigen, daß $\mathrm{Aut}(L/L^H) = H$ ist: H ist auf jeden Fall eine Untergruppe von $\mathrm{Aut}(L/L^H)$; wären die beiden verschieden, hätten sie verschiedene Fixkörper, da der Grad eines Körpers über seinem Fixkörper nach einem früheren Lemma gleich der Ordnung der Automorphismengruppe ist. Der Fixkörper von $\mathrm{Aut}(L/L^H)$ enthält den Körper L^H, und $[L : L^H] = |H|$; daher müssen die beiden Körper und somit auch die beiden Untergruppen übereinstimmen.

Umgekehrt sei K ein Zwischenkörper; wir müssen zeigen, daß K der Fixkörper von $\mathrm{Aut}(L/K)$ ist. Da L/K nach dem vorigen Satz GALOISsch ist, gilt dies in der Tat. Die restlichen Behauptungen sind klar.

■

Der Zwischenkörper $K = L^H$ ist genau dann GALOISsch über k, wenn k der Fixkörper einer Gruppe von Automorphismen des Körpers K ist. Da $[L : k] = |G|$ und $[L : K] = |H|$ ist, folgt $[K : k] = [G : H]$, die Gruppe hat also $[G : H]$ Elemente. Aus dem Beweis des vorletzten Satzes wissen

wir, daß die Automorphismen von L/k genau $[G : H]$ verschiedene Isomorphismen von K auf Teilkörper von L induzieren, die k festlassen. Mehr solche Isomorphismen kann es nicht geben, denn ist G' eine Menge von Körperhomomorphismen $K \to L$, die k festlassen und zu denen auch die Identität gehört, so wissen wir, daß der Körper

$$k' = \{x \in K \mid \sigma(x) = \tau(x) \text{ für alle } \sigma, \tau \in G'\},$$

einerseits den Körper k enthält, andererseits aber ist nach einem der früheren Lemmata in diesem Paragraphen $[K : k'] \geq |G'|$. Gäbe es also mehr als $[G : H]$ Isomorphismen von K auf Teilkörper von L, so wäre

$$[G : H] = [K : k] \geq [K : k'] \geq |G'| > [G : H]\,,$$

was nicht sein kann. Nun ist aber K/k genau dann GALOISsch, wenn es eine Gruppe von $[G : H]$ Automorphismen von K/k gibt, die k als Fixkörper hat. Jeder solche Automorphismus ist natürlich insbesondere ein Isomorphismus von K auf einen Teilkörper von L; da es insgesamt nur $[G : H]$ solche Isomorphismen gibt heißt dies, daß jeder dieser Isomorphismen ein Automorphismus von K sein muß, d.h. $\sigma(K) = K$ für jeden Automorphismus σ von L.

Für einen beliebigen Automorphismus σ von L und einen beliebigen Teilkörper $K = L^H$ ist $\sigma(K)$ ein Teilkörper von L. Ein weiterer Automorphismus $\tau\colon L \to L$ läßt $\sigma(K)$ genau dann punktweise fest, wenn für jedes $x \in K$ gilt $\tau\big(\sigma(x)\big) = \sigma(x)$ oder $\sigma^{-1}\tau\sigma(x) = x$. Somit muß $\sigma^{-1}\tau\sigma$ in H liegen und τ in $\sigma H \sigma^{-1}$, d.h. $\sigma(K)$ ist der Zwischenkörper zur Untergruppe $\sigma H \sigma^{-1}$. Wenn $\sigma(K)$ gleich K sein soll, muß dies gleich H sein; daher ist $\sigma(K) = K$ für alle $\sigma \in \mathrm{Aut}(L/k)$ genau dann, wenn H ein Normalteiler ist. In diesem Fall bilden die Nebenklassen von H eine Gruppe, die Faktorgruppe G/H. Also gilt:

Satz: L/k sei eine GALOISsche Erweiterung. Für einen Zwischenkörper K ist K/k genau dann GALOISsch, wenn $\mathrm{Aut}(L/K)$ ein Normalteiler von $\mathrm{Aut}(L/k)$ ist, und $\mathrm{Aut}(K/k)$ ist dann isomorph zur Faktorgruppe. ∎

Damit können wir die Zwischenkörper einer GALOISschen Erweiterung vollständig beschreiben durch die Untergruppen ihrer GALOIS-Gruppe. Das nützt uns allerdings nur dann etwas, wenn es interessante GALOISsche Erweiterungen gibt.

Satz: Eine endliche Körpererweiterung K/k ist genau dann GALOISsch, wenn K Zerfällungskörper eines über k separablen Polynoms ist.

Beweis: Sei zunächst K/k GALOISsch, und $b_1, \ldots, b_n$ sei eine Basis des k-Vektorraums K. Da dieser endlichdimensional ist, sind die verschiedenen Potenzen eines jeden b_i linear abhängig. Es gibt daher zu jedem b_i ein Polynom aus $k[X]$, das dort verschwindet; f_i sei ein irreduzible Faktor davon, der b_i als Nullstelle hat. Wie wir bereits gesehen haben, ist f_i separabel und zerfällt über K in Linearfaktoren. Damit ist auch das Produkt f aller f_i separabel, und K ist der Zerfällungskörper von f über k.

Umgekehrt sei f ein separables Polynom, und K/k sei der Zerfällungskörper von f über k. Wir müssen zeigen, daß K der Fixkörper von $G = \mathrm{Aut}(K/k)$ ist. Wir beweisen dies durch Induktion nach der Anzahl r jener Nullstellen von f, die *nicht* in k liegen. Im Falle $r = 0$ ist $K = k$, und die Behauptung ist trivialerweise richtig.

Für $r > 0$ betrachten wir eine nicht in k liegende Nullstelle z von f. Dann ist K auch Zerfällungskörper von f über $k(z)$, und da die Nullstelle z in $k(z)$ liegt, ist die Anzahl der nicht in $k(z)$ liegenden Nullstellen von f kleiner als r. Nach Induktionsannahme ist daher $K/k(z)$ GALOISsch, und $k(z)$ ist der Fixkörper von $\mathrm{Aut}\big(K/k(z)\big)$.

Als Nullstelle von f ist z auch Nullstelle eines irreduziblen Faktors g von f, und mit f ist auch g separabel. Bezeichnet d den Grad von g, hat g daher d verschiedene Nullstellen $z_1, \ldots, z_d$. Für jedes i ist $k(z_i) \cong k[X]/(g)$, also gibt es Isomorphismen $\sigma_i \colon k(z) \to k(z_i)$. Beim Beweis der Tatsache, daß Zerfällungskörper isomorpher Körper isomorph sind, haben wir gesehen, daß sich jeder solche Isomorphismus fortsetzen läßt zu einem Isomorphismus der Zerfällungskörper. Da der Zerfällungskörper von f über jedem der Körper $k(z_i)$ gleich K ist, gibt es also d Automorphismen $\tau_i \colon K \to K$, die auf $k(z)$ mit σ_i übereinstimmen.

Nun sei x ein beliebiges Element des Fixkörpers von $\mathrm{Aut}(K/k)$. Da x dann insbesondere von allen Automorphismen von $K/k(z)$ festgelassen wird, ist auf jeden Fall $x \in k(z)$. Die Potenzen $1, z, \ldots, z^{d-1}$ bilden

eine Basis des Vektorraums $k(z)$ über k; daher können wir x schreiben als

$$x = c_0 + c_1 z + \cdots + c_{d-1} z^{d-1} \quad \text{mit} \quad c_i \in k\,.$$

Die Automorphismen τ_i lassen k punktweise fest, und da x im Fixkörper von $\mathrm{Aut}(K/k)$ liegt, ist auch $\tau_i(x) = x$. Daher ist

$$\begin{aligned} x = \tau_i(x) &= c_0 + c_1 \tau_i(z) + \cdots + c_{d-1} \tau_i(z)^{d-1} \\ &= c_0 + c_1 z_i + \cdots c_{d-1} z_i^{d-1}\,. \end{aligned}$$

Das Polynom

$$c_{d-1} X^{d-1} + \cdots + c_1 X + (c_0 - x) \in K[X]$$

hat somit mindestens die d verschiedenen Nullstellen $z_1, \ldots, z_d$. Da es höchstens den Grad $d-1$ hat, muß es gleich dem Nullpolynom sein. Insbesondere ist $c_0 - x = 0$, d.h. $x = c_0$ liegt in k. Damit ist die Behauptung bewiesen.

∎

Korollar: Ist K/k eine separable endliche Körpererweiterung, so gibt es eine Körpererweiterung L/K derart, daß L/k GALOISsch ist.

Beweis: Nach Voraussetzung gibt es endich viele über k separable Elemente $z_1, \ldots, z_r \in K$ derart, daß $K = k(z_1, \ldots, z_r)$ ist. Für jedes z_i sei $f_i \in k[X]$ ein irreduzibles Polynom, das z_i als Nullstelle hat. Dann ist der Zerfällungskörper des Produkts aller f_i eine GALOISsche Erweiterung von k, die K enthält.

∎

§3: Lösbarkeit von Gleichungen durch Radikale

In diesem Paragraphen wollen wird uns überlegen, daß Polynomgleichungen vom Grad mindestens fünf *im allgemeinen* nicht durch Wurzelausdrücke lösbar sind. Dazu betrachten wir zunächst die Körpererweiterungen, die durch Adjunktion einer Wurzel entstehen. Wie wir vom Beispiel der dritten Wurzel aus zwei über $\mathbb{Q}$ wissen, sind diese im allgemeinen nicht GALOISsch; sie werden aber GALOISsch, wenn wir über einem Körper arbeiten, der genügend viele Einheitswurzeln enthält:

Definition: Ein Element ζ eines Körpers k heißt n-te *Einheitswurzel,* wenn $\zeta^n = 1$ ist. Wenn es kein $m < n$ gibt, für das bereits $\zeta^m = 1$ ist, bezeichnen wir ζ als eine *primitive* n-te Einheitswurzel.

Satz: Der Körper k enthalte eine primitive n-te Einheitswurzel ζ, und $a \in k$ sei ein beliebiges Element von k. Dann ist $k(\sqrt[n]{a})/k$ eine GALOISsche Erweiterung mit einer zyklischen GALOIS-Gruppe.

Dabei bezeichnet $\sqrt[n]{a}$ ein festes Element w eines Erweiterungskörpes von k mit $w^n = a$. Beispielsweise können wir, falls das Polynom $W^n - a \in k[W]$ irreduzibel ist, die Restklasse von W in $k[W]/(W^n - a)$ nehmen oder, falls k ein Teilkörper von $\mathbb{C}$ ist, eine feste komplexe Zahl w mit $w^n = a$.

Beweis: Das Polynom $X^n - a \in k[X]$ hat in $k(\sqrt[n]{a})$ die n verschiedenen Nullstellen $\zeta^i \sqrt[n]{a}$ für $i = 0, \ldots, n-1$, ist also separabel. Außerdem ist $k(\sqrt[n]{a})/k$ Zerfällungskörper von $X^n - a$ über k; die Körpererweiterung ist also GALOISsch. Jeder Automorphismus von $k(\sqrt[n]{a})/k$ muß $\sqrt[n]{a}$ abbilden auf ein Element x mit $x^n = a$, also auf eines der Elemente $\zeta^i \sqrt[n]{a}$. Durch dieses Element ist der Automorphismus eindeutig bestimmt, denn jedes Element von $k(\sqrt[n]{a})$ läßt sich als Polynom in $\sqrt[n]{a}$ mit Koeffizienten aus k schreiben. Da die Potenzen von ζ eine zyklische Gruppe der Ordnung n bilden, ist $\mathrm{Aut}\big(k(\sqrt[n]{a})/k\big)$ isomorph zu einer Untergruppe von $\mathbb{Z}/n$, also auch zyklisch. ∎

Auf der Suche nach einer allgemeinen Lösungsformel für Gleichungen d-ten Grades

$$a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0$$

über einem Körper k können wir uns beschränken auf den Fall $a_d = 1$, denn da a_d nicht verschwindet, können wir die Gleichung durch a_d dividieren. Wie wir in Kapitel 1 gesehen haben, könnten wir uns zusätzlich noch beschränken auf den Fall $a_{d-1} = 0$; da dies für die allgemeinen Betrachtungen hier keinen Vorteil bringt, wollen wir aber darauf verzichten. Wir betrachten die Koeffizienten $a_0, \ldots, a_{d-1}$ als Unbestimmte, d.h. wir arbeiten über dem Körper

$$K = k(a_0, \ldots, a_{d-1}) = \mathrm{Quot}\, k[a_0, \ldots, a_{d-1}]\,,$$

dessen Elemente rationale Funktionen (Quotienten von Polynomen) in den d Variablen $a_0, \ldots, a_{d-1}$ sind. Über diesem Körper betrachten wir den Zerfällungskörper L des Polynoms

$$f = X^d + a_{d-1}X^{d-1} + \cdots + a_1 X + a_0 \in K[X]\,.$$

In $L[X]$ ist $f = (X - z_1)\cdots(X - z_d)$ mit $z_1, \ldots, z_d \in L$, und natürlich ist $L = K(z_1, \ldots, z_d)$. Tatsächlich ist sogar $L = k(z_1, \ldots, z_d)$, denn nach dem Wurzelsatz von VIÈTE ist $a_j = (-1)^{d-j}\varphi_{d-j}(z_1, \ldots, z_d)$, wobei φ_j das j-te elementarsymmetrische Polynom in d Variablen bezeichnet, so daß alle a_j und damit auch K in $k(z_1, \ldots, z_d)$ liegen.

Die Gruppe $\mathfrak{S}_d$ aller Permutationen der Menge $\{1, \ldots, d\}$ operiert auf L, indem z_i abgebildet wird auf $z_{\pi(i)}$ für jedes i und jede Permutation $\pi \in \mathfrak{S}_d$. Die Koeffizienten a_j als (bis aufs Vorzeichen) elementarsymmetrische Funktionen in den z_i bleiben bei dieser Operation fix, also bleibt ganz K fix. Wir wollen uns überlegen, daß *nur* K fix bleibt, daß K also der Fixkörper unter dieser Operation ist.

Da L der Zerfällungskörper von f über K ist, läßt sich jedes Element $x \in L$ schreiben als Polynom in $z_1, \ldots, z_d$ mit Koeffizienten aus K. Da diese unter der Operation von $\mathfrak{S}_d$ fix bleiben, bleibt $x \in L$ genau dann fix, wenn es sich über K als ein symmetrisches Polynom in den z_i schreiben läßt. Nach dem Hauptsatz über symmetrische Polynome läßt sich jedes symmetrische Polynom schreiben als Polynom in den elementarsymmetrischen Polynomen, also läßt sich x schreiben als Polynom in den a_i und liegt somit in K. Somit ist K der Fixkörper von L unter der Operation der symmetrischen Gruppe $\mathfrak{S}_d$. Insbesondere ist L/K eine GALOISsche Erweiterung mit $\mathrm{Aut}(L/K) \cong \mathfrak{S}_d$.

Wir sagen, die allgemeine Gleichung vom Grad d lasse sich durch Radikale auflösen, wenn sich die z_i schreiben lassen als Ausdrücke in den a_j, die nur Grundrechenarten und Wurzeln enthalten. Für $d = 2$ etwa haben wir im Falle eines Grundkörpers k der Charakteristik 0 mit den Lösungsformeln

$$z_{1/2} = -\frac{a_1}{2} \pm \sqrt{\left(\frac{a_1}{2}\right)^2 - a_0}$$

eine solche Darstellung; tatsächlich gilt dies sogar über jedem Körper k mit $\mathrm{char}\, k \neq 2$.

Falls sich die allgemeine Gleichung d-ten Grades durch Radikale auflösen läßt, gibt es zur obigen Körpererweiterung L/K eine Folge von Zwischenkörpern

$$K = K_0 < K_1 < \cdots < K_r = L$$

derart, daß jeder Körper K_{i+1} aus K_i durch Adjunktion einer Wurzel entsteht. Dazu gibt es nach dem Hauptsatz der GALOIS-Theorie eine Folge von Gruppen

$$G_r = \{\mathrm{id}\} < G_{r-1} < \cdots G_1 < G_0 = \mathfrak{S}_d$$

derart, daß $K_i = L^{G_i}$ der Fixkörper von G_i ist.

Für das Folgende wollen wir annehmen, daß der Grundkörper k (und damit erst recht jeder Körper K_i) eine primitive $d!$-te Einheitswurzel enthält. Da der Grad jeder Körpererweiterung K_i/K_{i-1} ein Teiler von $[L : K] = d!$ ist und K_i aus K_{i-1} durch Adjunktion einer n-ten Wurzel entsteht, wobei $n \leq d!$ sein muß, folgt aus dem zu Beginn dieses Paragraphen bewiesenen Satz, daß K_{i+1}/K_i GALOISsch ist mit einer zyklischen GALOIS-Gruppe. Wir bezeichnen ihre Ordnung mit n_i.

Da L/K GALOISsch ist, sind auch alle Erweiterungen L/K_i GALOISsch und $\mathrm{Aut}(L/K_i) = G_i$. Der Körper K_{i+1} ist ein Zwischenkörper dieser Erweiterung, und da er GALOISsch über K_i ist, muß G_{i+1} ein Normalteiler von G_i sein mit einer zyklischen Faktorgruppe $G_i/G_{i+1} \cong \mathbb{Z}/n_i$.

Für $d = 3$ etwa haben wir die Folge $\{\mathrm{id}\} < \mathfrak{A}_3 < \mathfrak{S}_3$. Die Körpererweiterung L/K hat als Grad die Gruppenordnung sechs der $\mathfrak{S}_3$, und der Fixkörper K_1 von $\mathfrak{A}_3$ ist eine quadratische Erweiterung von K. Sie kommt zustande durch das Ziehen der Quadratwurzel aus der Diskriminante in der Lösungsformel. L/K_1 hat Grad drei; dies entspricht dem Ziehen der Kubikwurzel. Da K nach Voraussetzung eine primitive dritte Einheitswurzel enthält, enthält diese Erweiterung alle dritten Wurzeln des Radikanden, ist also GALOISsch.

Mit der Frage, wann wir in $\mathfrak{S}_d$ eine solche Folge von Untergruppen finden können, wollen wir uns als nächstes beschäftigen.

Im Hinblick auf die Auflösbarkeit von Gleichungen definieren wir:

Definition: Eine endliche Gruppe G heißt *auflösbar,* wenn es eine Folge

$$G_r = \{\mathrm{id}\} \lhd G_{r-1} \lhd \cdots G_1 \lhd G_0 = G$$

von Untergruppen gibt derart, daß G_{i+1} stets ein Normalteiler von G_i mit zyklischer Faktorgruppe G_i/G_{i+1} ist.

Falls sich die allgemeine Gleichung d-ten Grades durch Radikale lösen läßt, muß die Permutationsgruppe $\mathfrak{S}_d$ also auflösbar sein. Wir wollen uns überlegen, daß dies für $d \geq 5$ nicht der Fall ist. Dazu führen wir zunächst einen weiteren Begriff ein:

Definition: Eine endliche Gruppe G heißt *einfach,* wenn sie nicht nur aus dem Neutralelement e besteht und keine Normalteiler außer sich selbst und $\{e\}$ hat.

Die einfachsten Beispiele einfacher Gruppen sind die zyklischen Gruppen von Primzahlordnung, die ja überhaupt keine echten Untergruppen haben. Eine zyklische Gruppe $\mathbb{Z}/n$ mit zusammengesetztem n ist nicht einfach, denn ist r ein echter Teiler von n und g ein Erzeugendes der Gruppe, so erzeugt $g^{n/r}$ eine Untergruppe der Ordnung r, die natürlich wie jede Untergruppe einer abelschen Gruppe ein Normalteiler ist. Auch die symmetrische Gruppe $\mathfrak{S}_d$ ist für $d > 2$ nicht einfach, da sie die alternierende Gruppe $\mathfrak{A}_d$ als Normalteiler enthält. Wir wollen uns aber überlegen, daß $\mathfrak{A}_d$ für $d \geq 5$ einfach ist. Daraus wird dann insbesondere folgen, daß $\mathfrak{S}_d$ für $d \geq 5$ nicht auflösbar ist und es somit keine allgemeine Lösungsformel für Gleichungen vom Grad größer vier geben kann, die mit Grundrechenarten und Wurzeln auskommt:

Lemma: Ist $\mathfrak{A}_d$ einfach und nicht zyklisch, so ist $\mathfrak{S}_d$ nicht auflösbar.

Beweis: Falls $\mathfrak{A}_d$ einfach ist, kann es in $\mathfrak{S}_d$ außer $\mathfrak{A}_d$ keinen nichttrivialen Normalteiler N geben, denn für den wäre $N' = N \cap \mathfrak{A}_d$ ein Normalteiler von $\mathfrak{A}_d$, müßte also entweder gleich $\mathfrak{A}_d$ sein oder nur aus der Identität bestehen. $N' = \mathfrak{A}_d$ ist äquivalent zu $N = \mathfrak{S}_d$ oder $N = \mathfrak{A}_\mathfrak{d}$, und wenn N' nur aus der Identität besteht, kann N außer der Identität keine gerade Permutation enthalten. Da das Produkt zweier ungerader Permutationen gerade ist, besteht daher N entweder nur aus der Identität oder ist die von einer ungeraden Permutation ω der Ordnung zwei erzeugte zyklische

Untergruppe der Ordnung zwei. Letztere kann aber für $d \geq 3$ kein Normalteiler sein, da es immer eine Permutation $\pi \in \mathfrak{S}_d$ gibt, für die $\pi^{-1}\omega\pi$ eine von ω verschiedene ungerade Permutation ist. (Für $d \leq 2$ besteht $\mathfrak{A}_d$ nur aus der Identität, ist also nicht einfach.)

Falls $\mathfrak{S}_d$ auflösbar wäre, gäbe es für die Untergruppenfolge aus der obigen Definition daher nur die beiden Möglichkeiten $\{e\} \lhd \mathfrak{S}_d$ und $\{e\} \lhd \mathfrak{A}_d \lhd \mathfrak{S}_d$. Im ersten Fall müßte $\mathfrak{S}_d$ zyklisch sein; dies ist nur für $d = 2$ erfüllt, und dann ist $\mathfrak{A}_d$ nicht einfach. Im zweiten Fall müßte $\mathfrak{A}_d$ zyklisch sein, was wir ausgeschlossen haben (und was im übrigen genau für $d = 3$ erfüllt ist, wo diese Folge von Inklusionen die Auflösbarkeit der $\mathfrak{S}_3$ zeigt). Somit kann $\mathfrak{S}_d$ im Falle einer nichtzyklischen einfachen alternierenden Gruppe $\mathfrak{A}_d$ nicht auflösbar sein.

Satz: Für $d \geq 5$ ist die alternierende Gruppe $\mathfrak{A}_d$ einfach.

Wir führen den *Beweis* in fünf Schritten:

1. Schritt: $\mathfrak{A}_d$ wird von den Dreierzykeln erzeugt.

Beweis: Jedes Element von $\mathfrak{A}_d$ ist ein Produkt einer geraden Anzahl von Transpositionen. Falls zwei aufeinanderfolgende Transpositionen ein gemeinsames Element haben, ist $(a\ b)(b\ c) = (a\ b\ c)$ ein Dreierzyklus; andernfalls ist

$$(a\ b)(c\ d) = (a\ b)(b\ c)(b\ c)(c\ d) = (a\ b\ c)(b\ c\ d)$$

Produkt zweier Dreierzykeln. Somit läßt sich jedes Element von $\mathfrak{A}_d$ als Produkt von Dreierzykeln schreiben.

2. Schritt: Alle Dreierzykeln in $\mathfrak{S}_d$ sind zueinander konjugiert.

Beweis: $(a\ b\ c)$ und $(a'\ b'\ c')$ seien zwei Dreierzykeln und π eine Permutation, die a' auf a, b' auf b und c' auf c abbildet. Dann bildet $\pi^{-1}(a\ b\ c)\pi$ das Element a' zunächst ab auf a, dann via $(a\ b\ c)$ auf b, und weiter via π^{-1} auf b'. Genauso überlegt man sich, daß b' auf c' und c' auf a' abgebildet wird.

Ein $x \notin \{a', b', c'\}$ wird von π abgebildet auf ein $\pi(x) \notin \{a, b, c\}$, so daß $\pi(x)$ von $(a\ b\ c)$ festgelassen wird. Durch π^{-1} wird es wieder zurück auf x abgebildet. Somit ist $\pi^{-1}(a\ b\ c\ )\pi = (a'\ b'\ c')$.

3. Schritt: Für $d \geq 5$ sind je zwei Dreierzykel sogar bereits in $\mathfrak{A}_d$ zueinander konjugiert.

Beweis: $(a\ b\ c)$ und $(a'\ b'\ c')$ seien zwei Dreierzykeln und $\pi \in \mathfrak{S}_d$ eine Permutation, für die $\pi^{-1}(a\ b\ c)\pi = (a'\ b'\ c')$ ist. Falls π in $\mathfrak{A}_d$ liegt, sind wir fertig. Andernfalls existiert wegen $d \geq 5$ eine Transposition τ, die jedes der drei Elemente a, b, c festläßt. Somit ist $\tau^{-1}(a\ b\ c)\tau = (a\ b\ c)$ und damit $(\tau\pi)^{-1}(a\ b\ c)(\tau\pi) = \pi^{-1}(a\ b\ c\ )\pi = (a'\ b'\ c')$. Da π eine ungerade Permutation ist, muß $\tau\pi$ gerade sein, also in $\mathfrak{A}_d$ liegen.

4. Schritt: Für $d \geq 5$ ist jeder Normalteiler von $\mathfrak{A}_d$, der einen Dreierzyklus enthält, gleich $\mathfrak{A}_d$.

Beweis: Falls er einen Dreierzyklus enthält, muß er nach dem vorigen Schritt *alle* Dreierzykeln enthalten, ist also nach dem ersten Schritt gleich $\mathfrak{A}_d$.

Der Satz folgt nun aus

5. Schritt: Für $d \geq 5$ enthält jeder nichttriviale Normalteiler $N \trianglelefteq \mathfrak{A}_d$ einen Dreierzyklus.

Beweis: Ist π irgendein Element von N, so ist auch $\pi^{-1} \in N$, und für jedes $\omega \in N$ liegt auch $\omega^{-1}\pi^{-1}\omega$ in N, und damit auch $\pi\omega^{-1}\pi^{-1}\omega$.

Jedes Element $\pi \in N$ läßt sich schreiben als Produkt elementfremder Zykeln. Wir betrachten ein Element, das einen Zyklus der maximal vorkommenden Länge enthält.

Ist diese Länge mindestens gleich vier, ist π Produkt eines Zykels $(a\ b\ c\ d\ \ldots)$ der Länge mindestens vier und eventuell weiterer Zykeln. Wir setzen $x = \pi^{-1}(a)$ und betrachten $\omega = (c\ a\ b)$. Das Produkt $\pi\omega^{-1}\pi^{-1}\omega$ bildet a über die Zwischenergebnisse $a \mapsto c \mapsto b \mapsto c \mapsto d$ ab auf d, welches via $d \mapsto d \mapsto c \mapsto a \mapsto b$ auf b geht. Dieses wiederum wird via $b \mapsto a \mapsto x \mapsto x \mapsto a$ auf a abgebildet. Bei den übrigen Elementen überzeugt man sich leicht, daß sie auf sich selbst abgebildet werden; somit ist $\pi\omega^{-1}\pi^{-1}\omega = (a\ d\ b)$ ein Dreierzyklus aus N.

Falls die maximale Länge gleich drei ist und wir keinen Dreierzyklus haben, gibt es eine Permutation $\pi \in N$, die das Produkt eines Dreierzyklus mit Dreier- und Zweierzykeln ist. Der Dreierzyklus sei

$(a\ b\ c)$, der nächste nichttriviale Faktor sei entweder ein Dreierzyklus $(d\ e\ f)$ oder eine Transposition $(d\ e)$. Wir betrachten wieder die Permutation $\pi\omega^{-1}\pi^{-1}\omega$, dieses Mal für $\omega = (d\ b\ a)$. Nun geht a via $a \mapsto d \mapsto x \mapsto x \mapsto d$ auf d, welches via $d \mapsto b \mapsto a \mapsto b \mapsto c$ auf c geht, welches wiederum via $c \mapsto c \mapsto b \mapsto d \mapsto e$ auf e geht. Damit enthält $\pi\omega^{-1}\pi^{-1}\omega$ einen Zyklus der Länge mindestens vier, im Widerspruch zu unserer Annahme, der längste Zyklus eines jeden Elements sei höchstens ein Dreierzyklus.

Bleibt noch der Fall, daß sich jedes Element von N als Produkt elementfremder Transpositionen schreiben läßt. Deren Anzahl muß gerade sein, es gibt also ein Element, das einen Faktor $(a\ b)(c\ d)$ enthält mit vier verschiedenen Zahlen a, b, c, d, und es gibt noch mindestens eine weitere Zahl e, die von diesen vier Zahlen verschieden ist. Wir setzen $x = \pi^{-1}(e)$ und betrachten $\pi\omega^{-1}\pi^{-1}\omega$ für $\omega = (e\ c\ a)$. Hier zeigen die Abbildungsketten $a \mapsto e \mapsto x \mapsto x \mapsto e$, $e \mapsto c \mapsto d \mapsto d \mapsto c$ und $c \mapsto a \mapsto b \mapsto b \mapsto a$, daß $\pi\omega^{-1}\pi^{-1}\omega$ den Dreierzyklus $(a\ e\ c)$ enthält, im Widerspruch zur Annahme, daß alle Elemente von N Produkte elementfremder Transpositionen sind. Also tritt auch dieser Fall nicht auf, und wir haben gezeigt, daß N auf jeden Fall einen Dreierzyklus enthalten muß. ∎

Da die Gruppe $\mathfrak{A}_d$ für $d \geq 4$ nicht abelsch ist, ist sie erst recht nicht zyklisch; daher zeigt der gerade bewiesenen Satz zusammen mit dem davor bewiesenen Lemma, daß die Gruppe $\mathfrak{S}_d$ für $d \geq 5$ nicht auflösbar ist. Als Korollar folgt sofort der

Satz von Abel: Für $d \geq 5$ ist die allgemeine Gleichung d-ten Grades nicht durch Radikale auflösbar. ∎

§4: Konstruktionen mit Zirkel und Lineal

In der klassischen EUKLIDischen Geometrie geht man aus von einer Menge $\{P_0, \ldots, P_r\}$ von Punkten der Ebene und konstruiert daraus „mit Zirkel und Lineal“ weitere Punkte. Dabei sind folgende Operationen erlaubt:

- Durch zwei der vorhandenen Punkte wird eine Gerade gezeichnet (mit dem Lineal)
- Um einen der vorhandenen Punkte wird eine Kreislinie gezeichnet, die einen anderen der vorhandenen Punkte enthält (mit dem Zirkel)
- Schnittpunkte der gezeichneten Geraden und/oder Kreise werden zu den vorhandenen Punkten dazugenommen.

Diese Operationen können beliebig oft wiederholt werden.

Um solche Konstruktionen mit Körpererweiterungen in Verbindung zu bringen, wählen wir ein kartesisches Koordinatensystem und haben nun zwei Möglichkeiten: Entweder wir adjungieren für jeden Punkt $P_j = (x_j, y_j)$ die Koordinaten $x_j, y_j \in \mathbb{R}$ an $\mathbb{Q}$ und erhalten so einen Körper $k_0 < \mathbb{R}$, oder wir adjungieren stattdessen die komplexe Zahl $x_j + iy_j$ an $\mathbb{Q}$ und erhalten einen Körper $k_0' < \mathbb{C}$. Da wir in der Geometrie eher an reelle Koordinaten gewohnt sind, ist der erste Zuganng anschaulicher, so daß wir zunächst diesen benutzen werden. Bei der Frage nach der Konstruierbarkeit regelmäßiger n-Ecke werden sich allerdings die komplexen Zahlen als nützlicher erweisen.

Die Gerade durch zwei Punkte $P_i = (x_i, y_i)$ und $P_j = (x_j, y_j)$ ist die Lösungsmenge der Gleichung

$$(x - x_i)(y_j - y_i) + (y - y_i)(x_j - x_i) = 0 \,,$$

deren sämtliche Koeffizienten in k_0 liegen. Die Kreislinie um P_i, auf der P_j liegt, wird entsprechend beschrieben durch die quadratische Gleichung

$$(x - x_i)^2 + (y - y_i)^2 = (x_j - x_i)^2 + (y_j - y_i)^2 \,,$$

deren Koeffizienten ebenfalls in k_0 liegen.

Zur Berechnung des Schnittpunkts zweier verschiedener Geraden müssen wir ein lineares Gleichungssystem mit Koeffizienten aus k_0 lösen; falls es eine Lösung gibt, d.h. wenn die Geraden nicht parallel sind, ist diese ein Punkt mit Koordinaten in k_0.

Beim Schnitt einer Geraden $ax + by = c$ mit einem Kreis beachten wir zunächst, daß a und b nicht beide verschwinden können. Wir können die Gleichung also nach mindestens einer der beiden Variablen auflösen.

Das Ergebnis setzen wir ein in die Kreisgleichung und erhalten eine quadratische Gleichung in der anderen Variablen. Wenn es Schnittpunkte gibt, hat diese reelle Lösungen, die entweder in k_0 liegen oder in einem Körper k_1/k_0, der aus k_0 entsteht durch Adjunktion der Quadratwurzel eines Elements von k_0. Im letzteren Fall ist k_1/k eine Erweiterung vom Grad zwei.

Ähnlich ist die Situation beim Schnitt von zwei Kreisen: Die Differenz der beiden Gleichungen

$$(x-a)^2 + (y-b)^2 = r^2 \quad \text{und} \quad (x-c)^2 + (y-d)^2 = s^2$$

ist eine lineare Gleichung in x und y (es sei denn, die beiden Kreise wären konzentrisch), definiert also eine Gerade, und die Schnittmenge der beiden Kreislinien ist gleich der Schnittmenge dieser Geraden mit einer der beiden Kreislinien.

Falls wir eine Konstruktion mit Zirkel und Lineal durchführen können, liegen die Koordinaten aller konstruierter Punkte somit in einem Körper k, der aus k_0 durch schrittweise Körpererweiterungen vom Grad zwei entsteht:

$$k_0 < k_1 < \cdots < k_n = k \quad \text{und} \quad [k_i : k_{i-1}] = 2 \quad \forall i = 1, \ldots, n\,.$$

Insbesondere ist $[k : k_0] = 2^n$ eine Zweierpotenz.

Betrachten wir einige klassische mathematische Konstruktionsprobleme unter diesem Gesichtspunkt! Am einfachsten geht das beim sogenannten Delischen Problem: Der Legende nach fragten die Einwohner der griechischen Insel Delos (eine der kleinsten der Kykladen im Ägäischen Meer) anläßlich einer Pestepedemie ihr Orakel um Rat. Dieses verlangte, daß sie den würfelförmigen Altar im Tempel des Apollon durch einen Würfel mit doppeltem Volumen ersetzen sollten. Natürlich mußte dessen Kantenlänge aus der des alten Würfels mit Zirkel und Lineal konstruiert werden.

Wir haben also zwei Ausgangspunkte P_0 und P_1 derart, daß die Strecke $\overline{P_0P_1}$ der Kantenlänge des alten Würfels entspricht. Da wir das Koordinatensystem und die Einheit frei wählen können, sei etwa $P_0 = (0,0)$ und $P_1 = (1,0)$. Wir müssen zwei Punkte P_i, P_j konstruieren, deren

Verbindungsstrecke die Länge $\sqrt[3]{2}$ hat. Wenn wir das können, können wir diese Strecke von P_1 aus auf der x-Achse abtragen und erhalten den Punkt $(\sqrt[3]{2}, 0)$; der Körper k, in dem nach Ende der Konstruktion alle Koordinaten liegen, muß also die dritte Wurzel aus zwei enthalten und hat somit $\mathbb{Q}(\sqrt[3]{2})$ als Teilkörper. Damit muß $[k : \mathbb{Q}]$ durch drei teilbar sein, ist also keine Zweierpotenz. Daher ist das Delische Problem nicht mit Zirkel und Lineal lösbar.

Als nächstes betrachten wir das Problem der Konstruktion des regelmäßigen n-Ecks mit Zirkel und Lineal. Die griechischen Mathematiker konnten natürlich gleichseitige Dreiecke und Quadrate konstruieren, ebenso das regelmäßige Fünfeck, das Fünfzehneck, und über die Halbierung des Innenwinkeld damit auch jedes n-Eck, dessen Eckenanzahl eine der genannten Zahlen mal einer Zweierpotenz ist. Erst rund zwei Tausend Jahre später gelang 1796 dem damals 19-jährigen GAUSS die Konstruktion eines weiteren regelmäßigen n-Ecks, des Siebzehnecks. In seinem 1798 geschriebenen und 1801 erschienenen Buch *Disquisitiones Arithmeticae* bewies er allgemein, welche regelmäßigen n-Ecke sich mit Zirkel und Lineal konstruieren lassen und welche nicht.

Um sein Ergebnis zu verstehen, empfiehlt es sich, die Ebene mit der komplexen Zahlenebene zu identifizieren und das Problem so in Algebra zu übersetzen, daß wir nach der Konstruktion eines Punktes P mit Koordinaten (x, y) die komplexe Zahl $x + iy$ adjungieren.

Wenn wir ausgehen vom Mittelpunkt $P_0 = (0, 0)$ des regelmäßigen n-Ecks und einer Ecke $P_1 = (1, 0)$, haben die weiteren Ecken die Koordinaten $(\cos \frac{2\pi j}{n}, \sin \frac{2\pi j}{n})$ für $j = 1, \ldots, n-1$. Diese Punkte werden identifiziert mit den komplexen Zahlen

$$\cos \frac{2\pi j}{n} + i \sin \frac{2\pi j}{n} = e^{2\pi i j/n} = \left(e^{2\pi i/n}\right)^j ;$$

falls das regelmäßige n-Eck mit Zirkel und Lineal konstruierbar ist, muß also die primitive n-te Einheitswurzel $\zeta = e^{2\pi i/n}$ in einem Erweiterungskörper von Zweipotenzordnung über $\mathbb{Q}$ liegen.

Der Körper $\mathbb{Q}(\zeta)$ enthält natürlich auch alle Potenzen von ζ, ist also ein Zerfällungskörper des Polynoms $X^n - 1$. Dieses ist aber nicht irreduzibel, beispielsweise ist $X - 1$ ein Teiler, da die Eins eine Nullstelle ist.

Allgemeiner: Ist $n = mq$, so ist $X-1$ auch ein Teiler von X^q-1; ersetzen wir hier X durch X^m, folgt, daß X^m-1 Teiler von $X^{mq}-1 = X^n-1$ ist. Bezeichnet also f den irreduziblen Faktor von X^n-1, der ζ als Nullstelle hat, so hat f nur primitive n-te Einheitswurzeln als Nullstellen, denn ist x bereits eine m-te Einheitswurzel für einen echten Teiler m von n, so ist x Nullstelle von X^m-1. Wäre sie auch eine Nullstelle von f, so wäre x eine mehrfache Nullstelle des Polynoms X^n-1. Da dessen Ableitung nX^{n-1} nur bei der Null verschwindet, ist das nicht möglich.

Die primitiven n-ten Einheitswurzel allerdings müssen allesamt Nullstellen von f sein nach dem folgenden Argument von DEDEKIND: Da die primitiven n-ten Einheitswurzeln genau die Potenzen ζ^j sind, für die j teilerfremd zu n sind, läßt sich j schreiben als Produkt von Primzahlen, die keine Teiler von n sind. Daher reicht es zu zeigen, daß für jede Nullstelle ξ von f und jede Primzahl p, die kein Teiler von n ist, auch ξ^p eine Nullstelle von f ist. Falls dies für irgendein ξ und irgendein p nicht der Fall ist, muß ξ^p Nullstelle eines weiteren irreduziblen Polynoms g sein. Da ξ^p eine primitive n-te Einheitswurzel ist, muß auch g ein Teiler von X^n-1 sein. Betrachten wir das Polynom $G = g(X^p) \in \mathbb{Q}[X]$. Da $g(\xi^p)$ verschwindet, ist ξ eine Nullstelle von G.

Nach dem Lemma von GAUSS können wir bei der Zerlegung des Polynoms X^n-1 in $\mathbb{Q}[X]$ annehmen, daß alle Faktoren ganzzahlige Koeffizienten haben, daß also f, g und damit auch G in $\mathbb{Z}[X]$ liegen. Wenn wir alle Koeffizienten modulo p reduzieren, erhalten wir Polynome $\overline{f}, \overline{g}$ und $\overline{G}$ aus $\mathbb{F}_p[X]$, wobei $\overline{f}$ und $\overline{g}$ zwei verschiedene (nicht notwendigerweise irreduzible) Faktoren von X^n-1 in $\mathbb{F}_p[X]$ sind. Da in $\mathbb{F}_p$ jedes Element gleich seiner p-ten Potenz ist, ist $\overline{G} = g^p$. Somit sind alle Nullstellen von $\overline{G}$ auch Nullstellen von $\overline{g}$. Die Polynome f und G aus $\mathbb{Z}[X]$ haben in $\mathbb{Q}(\zeta)$ die gemeinsame Nullstelle ξ, also hat der ggT h der beiden Polynome positiven Grad. Betrachten wir ihn modulo p, erhalten wir ein Polynom $\overline{h}$, das sowohl $\overline{f}$ als auch $\overline{G}$ teilt. Wegen $\overline{G} = \overline{g}^p$ folgt, daß auch der ggT von $\overline{f}$ und $\overline{g}$ positiven Grad hat. Somit hat das Polynom $X^n-1 \in \mathbb{F}_p[X]$ in seinem Zerfällungskörper mindestens eine mehrfache Nullstelle. Das ist aber nicht möglich, denn seine (formale) Ableitung nX^{n-1} ist nicht das Nullpolynom, da

p kein Teiler von n ist, und es hat nur die Null als Nullstelle, die aber keine Nullstelle von $X^n - 1$ ist. Daher hat $X^n - 1$ auch als Polynom über $\mathbb{F}_p$ keine mehrfache Nullstelle, so daß die Annahme $f(\xi^p) \neq 0$ zu einem Widerspruch führt. Dies zeigt, daß f genau die primitiven n-ten Einheitswurzeln als Nullstellen hat.

Somit hat das irreduzible Polynom $f \in \mathbb{Q}[X]$ mit $f(\zeta) = 0$ den Grad $\varphi(n)$, wobei φ die aus dem zweiten Kapitel bekannte EULERsche φ-Funktion ist, die die Anzahl der primen Restklassen modulo n angibt. Da alle Nullstellen von f Potenzen von ζ sind, ist $\mathbb{Q}(\zeta)$ ein Zerfällungskörper von f über $\mathbb{Q}$; insbesondere ist die Körpererweiterung $\mathbb{Q}(\zeta)/\mathbb{Q}$ GALOISsch und hat den Grad $\varphi(n)$.

Dies zeigt, daß das regelmäßige n-Eck nur dann mit Zirkel und Lineal konstruierbar sein kann, wenn $\varphi(n)$ eine Zweierpotenz ist. Wie wir uns in Kapitel zwei überlegt haben, läßt sich $\varphi(n)$ anhand der Primzerlegung von n bestimmen: Für

$$n = \prod_{i=1}^{r} p_i^{e_i} \qquad \text{ist} \quad \varphi(n) = \prod_{i=1}^{r} \left(p_i^{e_i - 1} \cdot (p_i - 1) \right) .$$

Somit müssen alle Faktoren in diesem Produkt Zweierpotenzen sein.

Im Falle $p_i = 2$ ist das automatisch erfüllt; alle anderen möglichen p_i sind ungerade, so daß $e_i = 1$ sein muß und $p_i - 1$ eine Zweierpotenz.

Primzahlen der Form $2^r + 1$ heißen FERMATsche Primzahlen. $2^r + 1$ kann nur dann prim sein, wenn r eine Zweierpotenz ist, denn ist r ungerade, so ist $2^r + 1 \equiv (-1)^r + 1 = 0 \bmod 3$ durch drei teilbar, und ist $r = 2^s u$ mit einer ungeraden Zahl $u > 1$, so ist $2^r + 1 \equiv (-1)^u + 1 \equiv 0 \bmod 2^s + 1$ durch $2^s + 1$ teilbar.

Definition: Die m-te FERMAT-Zahl ist $F_m = 2^{2^m} + 1$; falls F_m prim ist, heißt F_m eine FERMATsche Primzahl.

FERMAT vermutete, daß alle F_m prim seien; das ist eine der sehr wenigen seiner Vermutungen, die sich als falsch herausstellten. $F_0 = 3$, $F_1 = 5$, $F_2 = 17$, $F_3 = 257$ und $F_4 = 65\,537$ sind in der Tat allesamt prim (was auch FERMAT wußte), aber wie EULER 1732 zeigte, ist

$$F_5 = 4\,294\,967\,297 = 641 \cdot 6\,700\,417 .$$

Auch alle anderen F_m mit $m \geq 5$, die bislang getestet wurden, sind keine Primzahlen; es ist also nicht bekannt, ob es ein $m \geq 5$ gibt, für das F_m prim ist.

Für die Konstruierbarkeit des regelmäßigen n-Ecks folgt:

Satz: Falls das regelmäßige n-Eck mit Zirkel und Lineal konstruierbar ist, läßt sich n als Produkt einer Zweierpotenz (die auch eins sein kann) mit verschiedenen FERMATschen Primzahlen schreiben. ∎

Als Korollar zu diesem Satz können wir sofort auch die Unlösbarkeit eines anderen klassischen geometrischen Problems zeigen:

Korollar: Es ist nicht möglich, einen beliebigen Winkel mit Zirkel und Lineal in drei gleiche Teile zu zerlegen.

Beweis: Falls es ein solches Verfahren gäbe, könnte man insbesondere den Innenwinkel eines gleichseitigen Dreiecks dreiteilen. Damit wäre der Innenwinkel des regelmäßigen Neunecks und damit dieses selbst mit Zirkel und Lineal konstruierbar, im Widerspruch zum gerade gezeigten Resultat von GAUSS. ∎

GAUSS bewies auch die Umkehrung des obigen Satzes; bevor wir uns damit beschäftigen, wollen wir uns aber davon überzeugen, daß zumindest in vielen Fällen klassische Konstruktionsverfahren ausreichen. Beginnen wir mit den einzelnen Faktoren:

Die Konstruierbarkeit des regelmäßigen Dreiecks ist aus der Schule bekannt, das 2^m-Eck für $m \geq 2$ kann aus dem Quadrat durch Winkelhalbierungen konstruiert werde. Die Konstruktion des regelmäßigen Fünfecks wird in der Schule üblicherweise nicht behandelt, war aber bereits den Pythagoräern bekannt: Diese brauchten sie für ihr Symbol, den fünfzackigen Stern, bestehend aus den sämtlichen Diagonalen eines regelmäßigen Fünfecks. Die Konstruktion des regelmäßigen Siebzehnecks geht, wie erwähnt, zurück auf GAUSS, der auch den obigen Satz (einschließlich seiner Umkehrung) bewies. Das grundsätzliche Verfahren, wie er aus der Struktur der GALOIS-Gruppe eine Konstruktion des

Siebzehnecks herleitete, führte später auch zur Konstruktion des 257-Ecks durch

MAGNUS GEORG PAUCKER: Geometrische Verzeichnung des regelmäßigen Siebzehn-Ecks und des regelmäßige Zweyhundersiebenundfunfzig-Ecks in den Kreis, *Jahresverhandlungen der Kurländischen Gesellschaft für Literatur und Kunst* **2**, 1822, S. 160–219

(die Konstruktion des 257-Ecks beginnt Seite 188) und

FRIEDRICH JULIUS RICHELOT: De resolutione algebraica aequationis $x^{257} = 1$, sive de divisione circuli per bisectionem anguli septies repetitam in partes 257 inter se aequales commentatio coronata, *Journal für die reine und angewandte Mathematik* **9**, 1832, S. 1–26, 146–161, 209–230, 337–358.

Das regelmäßige 65 537-Eck konstruierte JOHANN GUSTAV HERMES in über zehnjähriger Arbeit; er hinterlegte das aus mehr als zweihundert großformatigen Seiten bestehende Manuskript 1889 in einem Handkoffer im mathematischen Institut der Universität Göttingen, wo es immer noch zu finden ist. 1894 veröffentlichte er eine siebzehnseitige Zusammenfassung

J. HERMES: Ueber die Teilung des Kreises in 65537 gleiche Teile, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse,* 1894, S. 170–186.

Da er als Königsberger kein Mitglied der Göttinger Gesellschaft der Wissenschaften war, wurde das Manuskript dort von FELIX KLEIN vorgelegt.

Falls es ein $m \geq 5$ geben sollte, für das F_m prim ist, folgt aus dem Satz von GAUSS, daß auch das regelmäßige F_m-Eck mit Zirkel und Lineal konstruierbar ist; eine entsprechende Konstruktion konnte natürlich bislang noch niemand vorlegen, und auch in Zukunft wird das nur schwer möglich sein: Die kleinste FERMAT-Zahl, von der nicht bekannt ist, ob sie prim ist oder nicht, ist F_{33}, und diese Zahl hat über fünf Milliarden Dezimalstellen.

Beschäftigen wir uns als nächstes mit den Produkten aus Zweierpotenzen und verschiedenen FERMATschen Primzahlen. Wir müssen zeigen:

Sind n und m zwei zueinander teilerfremde Zahlen derart, daß das regelmäßige n-Eck und das regelmäßige m-Eck beide mit Zirkel und Lineal konstruierbar sind, so läßt sich auch das regelmäßige nm-Eck konstruieren. Allgemein läßt sich das regelmäßige r-Eck genau dann mit Zirkel und Lineal konstruieren, wenn der Winkel beim Mittelpunkt zwischen zwei benachbarten Ecken konstruierbar ist. Beim n-Eck und beim m-Eck ist er $2\pi/n$ *bzw.* $2\pi/m$; wir müssen zeigen, daß sich daraus der Winkel $2\pi/nm$ konstruieren läßt. Da n und m teilerfremd sind, gibt es ganze Zahlen a, b, so daß $am + bn = 1$ ist. Multiplikation mit $2\pi/nm$ macht daraus

$$a \cdot \frac{2\pi}{n} + b \cdot \frac{2\pi}{m} = \frac{2\pi}{nm} \, .$$

Ganzzahlige Vielfache eines Winkels und Summen und Differenzen von Winkeln lassen sich problemlos mit Zirkel und Lineal konstruieren; somit ist auch der Winkel $2\pi/nm$ und damit das regelmäßige nm-Eck mit Zirkel und Lineal konstruierbar.

Für einen vollständigen Beweis der Umkehrung können wir beispielsweise zeigen, daß jeder Punkt mit Koordinaten in einer GALOISschen Körpererweiterung vom Grad 2^m mit Zirkel und Lineal konstruiert werden kann. Wir beginnen mit den Punkten, deren Koordinaten im Grundkörper selbst liegen.

Lemma: $P_0, \ldots, P_r$ seien Punkte der Ebene $\mathbb{R}^2$ mit $P_0 = (0,0)$ und $P_1 = (1,0)$, und $K/\mathbb{Q}$ entstehe aus $\mathbb{Q}$ durch Adjunktion der Koordinaten der P_i. Dann kann jeder Punkt P mit Koordinaten in K aus den Punkten P_i mit Zirkel und Lineal konstruiert werden.

Beweis: P habe die Koordinaten (x, y) mit $x, y \in K$, und die Koordinaten der P_i seien (x_i, y_i). Dann ist

$$K = \mathbb{Q}(x_0, \ldots, x_r, y_0, \ldots, y_r) \, ,$$

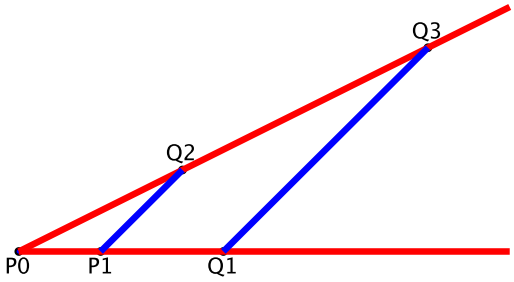
und x, y lassen sich als rationale Funktionen in den x_i und y_i schreiben. Da wir P konstruieren können, sobald wir Strecken der Längen x und y konstruiert haben, reicht es, daß wir aus gegebenen Streckenlängen auch alle Streckenlängen konstruieren können, die sich als rationale Funktionen in den gegebenen ausdrücken lassen. Dies folgt induktiv, sobald

wir wissen, daß wir aus gegebenen Längen auch deren Summen, Differenzen, Produkte und Quotienten konstruieren lassen. Für Summen und Differenzen ist dies trivial: Wir können die beiden Strecken einfach mit dem Zirkel auf einer festen Geraden abtragen.

Für Produkte und Quotienten können wir o.B.d.A. annehmen, daß beide Strecken positive Längen haben. Für das Produkt tragen wir auf dem von P_0 ausgehenden Strahl durch P_1 eine Strecke $\overline{P_0Q_1}$ der Länge a ab, auf einem anderen Strahl durch P_0 eine Strecke $\overline{P_0Q_2}$ der Länge b. Sodann konstruieren wir die Parallele zur Geraden P_1Q_2 durch Q_1; sie schneide die Gerade P_0Q_2 im Punkt Q_3. Nach dem Strahlensatz ist dann

$$\left|\overline{P_0Q_3}\right| : \left|\overline{P_0Q_2}\right| = \left|\overline{P_0Q_1}\right| : \left|\overline{P_0P_1}\right| = a : 1 \,.$$

Da die Strecke $\overline{P_0Q_2}$ die Länge b hat, muß daher $\overline{P_0Q_3}$ die Länge ab haben.



Um für den Quotienten a/b die gleiche Zeichnung verwenden zu können, tragen wir auf einen Strahl durch P_0 die Strecke $\overline{P_0Q_3}$ der Länge a ab und auf dem Strahl von P_0 durch P_1 die Strecke $\overline{P_0Q_1}$ der Länge b. Q_2 sei der Schnittpunkt der Parallelen zu Q_1Q_3 durch P_1 mit dem Strahl durch Q_3. Nach dem Strahlensatz ist dann

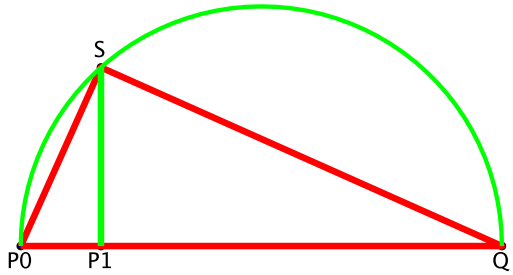
$$\left|\overline{P_0Q_3}\right| : \left|\overline{P_0Q_2}\right| = \left|\overline{P_0Q_1}\right| : \left|\overline{P_0P_1}\right| = a : 1 \,,$$

und da $\overline{P_0Q_3}$ die Länge b hat, erfüllt die Länge x von $\overline{P_0Q_2}$ die Relation $b : x = a : 1$, d.h. $x = b/a$.

■

Lemma: $P_0, \ldots, P_r$ seien Punkte der Ebene $\mathbb{R}^2$ mit $P_0 = (0,0)$ und $P_1 = (1,0)$, und $K/\mathbb{Q}$ entstehe aus $\mathbb{Q}$ durch Adjunktion der Koordinaten der P_i. Dann kann jeder Punkt P mit Koordinaten in einem Erweiterungskörper L/K vom Grad zwei aus den Punkten P_i mit Zirkel und Lineal konstruiert werden.

Beweis: Jedes Element von L liegt entweder bereits in K oder ist Lösung einer quadratischen Gleichung mit Koeffizienten in K. Wir müssen uns also zusätzlich zum bereits im Beweis des vorigen Lemmas gezeigten überlegen, daß sich jede quadratische Gleichung mit Koeffizienten aus K mit Zirkel und Lineal lösen läßt. Da wir alle Grundrechenarten ausführen können, müssen wir dazu nur noch zeigen, daß wir zu einer Strecke der Länge $a > 0$ auch eine Strecke der Länge $\sqrt{a}$ konstruieren können.



Dazu können wir beispielsweise auf der Geraden P_0P_1 die Strecke a von P_1 aus als $\overline{P_1Q}$ in die von P_0 abgewandte Richtung abtragen. Über der Strecke $\overline{P_0Q}$ (mit Länge $a+1$) konstruieren wir den THALES-Kreis, d.h. wir konstruieren zunächst die Mittelsenkrechte zu $\overline{P_0Q}$, die diese im Punkt M schneide; dann zeichnen wir den Kreis um M durch Q (und damit auch P_0). Mit diesem Kreis schneiden wir die Senkrechte zur Strecke $\overline{P_0Q}$ durch den Punkt P_1; einer der beiden Schnittpunkte sei S. Nach dem Satz des THALES ist ΔP_0QS ein rechtwinkliges Dreieck. Seine Hypothenuse $\overline{P_0Q}$ wird durch den Fußpunkt P_1 der Höhe $\overline{P_1S}$ in die Teilstrecken $\overline{P_0P_1}$ der Länge eins und $\overline{P_1Q}$ der Länge a aufgeteilt. Nach dem Höhensatz ist das Quadrat der Höhe $h = \left|\overline{P_1S}\right|$ gleich dem Produkt $1 \cdot a$, d.h. $h = \sqrt{a}$.

■

Korollar: $P_0, \ldots, P_r$ seien Punkte der Ebene $\mathbb{R}^2$ mit $P_0 = (0,0)$ und $P_1 = (0,1)$, und $K/\mathbb{Q}$ entstehe aus $\mathbb{Q}$ durch Adjunktion der Koordinaten der P_i. Zum Körper L/K gebe es eine Folge von Zwischenkörpern

$$K = L_0 < L_1 < \cdots < L_r = L$$

derart, daß $[L_i : L_{i-1}] = 2$ ist für $i = 1, \ldots, r$. Dann kann jeder Punkt P mit Koordinaten in L aus den Punkten P_i mit Zirkel und Lineal konstruiert werden.

Beweis durch Induktion nach r: Für $r = 1$ ist das gerade das obige Lemma, und für $r > 1$ wissen wir nach der Induktionsvoraussetzung, daß sich jeder Punkt mit Koordinaten aus L_{r-1} aus den P_i konstruieren läßt. Da L_r/L_{r-1} eine quadratische Erweiterung ist, gibt es ein $w \in L_{r-1}$, so daß sich jedes Element von L_r in der Form $a + b\sqrt{w}$ mit $a, b \in L_{r-1}$ schreiben läßt. Da wir Strecken der Längen a, b und w nach Induktionsannahme konstruieren können und nach obigem Lemma zu w auch $\sqrt{w}$, folgt die Behauptung. ∎

Die Voraussetzung dieses Korollars an der Körper L ist schwer nachzuweisen; wir wollen uns überlegen, daß wir sie ersetzen können durch die einfachere Voraussetzung, daß L/K GALOISsch ist und $[L : K]$ eine Zweierpotenz. $\mathrm{Aut}(L/K)$ ist dann eine Gruppe von Zweipotenzordnung; als erstes soll gezeigt werden, daß jede solche Gruppe auflösbar ist. Für die vorbereitenden Lemmata brauchen wir die Voraussetzung über die Gruppenordnung noch nicht und können sie daher etwas allgemeiner formulieren:

Lemma: Die Gruppe G habe einen Normalteiler N derart, daß sowohl N als auch G/N auflösbar sind. Dann ist auch G auflösbar.

Beweis: Wegen der Auflösbarkeit von N gibt es eine Folge von Untergruppen $N_0, \ldots, N_r$ von N mit

$$\{1\} = N_r \trianglelefteq N_{r-1} \trianglelefteq \cdots \trianglelefteq N_1 \trianglelefteq N_0 = N$$

derart, daß alle Faktorgruppen N_{j-1}/N_j zyklisch sind. Entsprechend gibt es eine Folge von Untergruppen

$$\{1\} = N/N = \overline{G}_s \trianglelefteq \overline{G}_{s-1} \trianglelefteq \cdots \trianglelefteq \overline{G}_1 \trianglelefteq \overline{G}_0 = G/N$$

derart, daß alle Faktorgruppen $\overline{G}_{j-1}/\overline{G}_j$ zyklisch sind. Die Abbildung $\varphi\colon G \to G/N$, die jedes $g \in G$ auf seine Restklasse module N abbildet, ist ein Homomorphismus; daher sind die Urbilder $G_i = \varphi^{-1}(\overline{G}_i)$ der $\overline{G}_i$ Untergruppen von G und

$$N = G_s \trianglelefteq G_{s-1} \trianglelefteq \cdots \trianglelefteq G_1 \trianglelefteq G_0 = G\,.$$

Da N ein Normalteiler sowohl von G_{j-1} als auch von G_j ist, können wir G_{j-1}/N abbilden nach G_{j-1}/G_j, indem wir die Nebenklasse gN

abbilden auf gG_j. Diese Abbildung ist offensichtlich surjektiv und hat den Kern G_j/N; nach dem Homomorphiesatz ist also

$$\overline{G}_{j-1}/\overline{G}_j = (G_{j-1}/N)/(G_j/N) \cong G_{j-1}/G_j$$

für alle j. Daher sind auch die Faktorgruppen G_{j-1}/G_j zyklisch. Setzen wir die beiden Reihen hintereinander, sehen wir, daß G auflösbar ist. ∎

Lemma: Jede endliche abelsche Gruppe G ist auflösbar. Ist ihre Ordnung m das Produkt der Primzahlpotenzen $p_i^{e_i}$ und e die Summe der Exponenten e_i, so gibt es Untergruppen $G_0, \ldots, G_e$ von G derart daß

$$\{1\} = G_e \trianglelefteq G_{e-1} \trianglelefteq \cdots G_1 \trianglelefteq G_0 = G$$

ist und jeweils e_i der Faktorgruppen G_{j-1}/G_j isomorph zu $\mathbb{Z}/p_i$ sind.

Beweis durch Induktion nach e: Für $e = 1$ ist G eine zyklische Gruppe von Primzahlordnung; mit $G_1 = \{1\}$ und $G_0 = G$ ist die Behauptung trivialerweise erfüllt.

Nun sei $e > 1$ und $g \neq 1$ ein Element von g. Seine Ordnung sei r, und p sei ein Primteiler von r. Dann ist $h = g^{r/p}$ ein Element der Ordnung p, erzeugt also eine zyklische Untergruppe N der Ordnung p in G. Wegen der Kommutativität von G ist N ein Normalteiler, und G/N ist eine abelsche Gruppe der Ordnung $|G|/p$. Somit ist die Exponentensumme e für G/N um eins kleiner als für G; nach Induktionsvoraussetzung gilt die Behauptung also für G/N, und sie gilt natürlich auch für $N \cong \mathbb{Z}/p$. Nach dem vorigen Lemma ist G daher auflösbar, und da die Faktorgruppen sowohl für N als auch für G/N zyklisch von Primzahlordnung sind, gilt dies auch für die von G.

∎

Satz: Jede Gruppe G von Primzahlpotenzordnung p^n ist auflösbar. Es gibt eine Folge von Untergruppen $G_0, \ldots, G_n$ von G derart, daß

$$\{1\} = G_n \trianglelefteq G_{n-1} \trianglelefteq \cdots \trianglelefteq G_1 \trianglelefteq G_0 = G$$

ist und $G_{i-1}/G_i \cong \mathbb{Z}/p$ für $i = 1, \ldots n$.

Beweis durch Induktion nach n: Für $n = 0$ gibt es nichts zu beweisen; im Falle $n = 1$ ist G isomorph zu $\mathbb{Z}/p$, denn jedes Element außer dem

Neutralelement muß nach LAGRANGE die Ordnung p haben. Sei also $n \geq 2$.

Wir definieren das *Zentrum* einer Gruppe G als

$$Z(G) = \{x \in G \mid xg = gx \text{ für alle } g \in G\} .$$

es ist eine Untergruppe, denn es enthält offensichtlich das Neutralelement, und für zwei Elemente $x, y \in Z(G)$ ist

$$(xy)g = x(yg) = x(gy) = (xg)y = (gx)y = g(xy)$$

und

$$x^{-1}g = (g^{-1}x)^{-1} = (xg^{-1})^{-1} = gx^{-1}$$

für alle $g \in G$. Tatsächlich ist $Z(G)$ sogar ein Normalteiler von G, denn durch Multiplikation von links mit g^{-1} wird die Gleichung $xg = gx$ zu $g^{-1}xg = g^{-1}gx = x$; ein Element $x \in G$ liegt also genau dann im Zentrum von G, wenn $x^g = x$ ist für alle $g \in G$.

Wir wollen uns als nächstes überlegen, daß $Z(G)$ für eine Gruppe der Ordnung p^n mit $n \geq 1$ nicht nur aus dem Neutralelement bestehen kann: Dazu lassen wir G durch Konjugation auf sich selbst operieren, betrachten also die Operation

$$\begin{cases} G \times G \to G \\ \quad (g, x) \mapsto x^g = g^{-1}xg \end{cases} .$$

Die Bahn eines Elements $x \in G$ besteht aus allen Konjugierten von x. Genau dann, wenn x im Zentrum liegt, sind alle Konjugierten von x gleich x, die Bahn von x besteht also nur aus x selbst. Für alle anderen Elemente von G enthält sie mindestens ein weiteres Element. Nach der Bahnbilanzgleichung ist ihre Elementanzahl der Quotient aus der Gruppenordnung durch die Ordnung des Stabilisators; für eine Gruppe von p-Potenzordnung ist das eine p-Potenz ungleich eins, also eine durch p teilbare Zahl.

Somit enthält die Bahn eines Elements aus dem Zentrum genau ein Element; jede andere Bahn enthält eine durch p teilbare Anzahl von Elementen. Da jedes Element von G in genau einer Bahn liegt, ist die Summe der Mächtigkeiten gleich der Gruppenordnung p^n. Diese

Summe ist gleich der Ordnung des Zentrums, die wegen $e \in Z(G)$ mindestens eins sein muß, plus einer durch p teilbaren Zahl. Somit muß auch die Ordnung des Zentrums durch p teilbar und damit mindestens p sein.

Haben wir also eine Gruppe G der Ordnung p^n mit $n > 1$, so hat $Z(G)$ mindestens die Ordnung p. Das Zentrum ist Normalteiler und als abelsche Gruppe auch auflösbar; nach einem der obigen Lemmata sind alle Faktorgruppen zyklisch von Primzahlordnung, also isomorph zu $\mathbb{Z}/p$, da die Ordnung von $Z(G)$ eine p-Potenz ist. Die Faktorgruppe $G/Z(G)$ hat höchstens Ordnung p^{n-1}, ist also nach Induktionsvoraussetzung auflösbar mit allen Faktorgruppen isomorph zu $\mathbb{Z}/p$, und damit ist auch G aufllösbar mit allen Faktorgruppen isomorph zu $\mathbb{Z}/p$. ∎

Uns interessiert hier vor allem der Fall $p = 2$; hier sind also alle Faktorgruppen isomorph zu $\mathbb{Z}/2$, d.h. für eine GALOISsche Erweiterung L/K, deren Grad eine Zweierpotenz ist, gibt es eine Folge von ZwKn, von denen jeder eine quadratische Erweiterung seines Vorgängers ist. Daraus folgt insbesondere, daß das regelmäßige n-Eck mit Zirkel und Lineal konstruiert werden kann, wenn n das Produkt einer Zweierpotenz mit verschiedenen FERMATschen Primzahlen ist.

§5: Transzendente Zahlen und die Quadratur des Kreises

Eines der berühmtesten Probleme der klassischen Geometrie, das es sogar in die Umgangssprache geschafft hat, ist die Quadratur des Kreises, d.h. die Konstruktion eines Quadrats, das den gleichen Flächeninhalt hat wie ein vorgegebener Kreis. Wie wir in diesem Paragraphen sehen werden, kann diese Konstruktion nicht mit Zirkel und Lineal ausgeführt werden, da beispielsweise für den Kreis mit Radius eins die Seitenlänge des Quadrats in keiner endlichen Körpererweiterung von $\mathbb{Q}$ liegt.

Definition: $K/\mathbb{Q}$ sei eine Körpererweiterung. Ein Element $x \in K$ heißt *algebraisch,* wenn es ein Polynom $f \in \mathbb{Z}[X]$ gibt, das an der Stelle x verschwindet. Andernfalls heißt x *transzendent.*

Natürlich ist jede rationale Zahl algebraisch, denn der Bruch p/q ist eine Nullstelle des linearen Polynoms $qX - p$. Auch Wurzeln, egal

ob reell oder nicht aus ganzen Zahlen, sind algebraisch als Nullstellen von Polynomen $X^n - a$. Allgemeiner ist sogar jeder Ausdruck, der nur rationale Zahlen, Grundrechenarten und Wurzeln enthält, algebraisch, denn die so konstruierte Zahl liegt in einer Körpererweiterung endlichen Grades $K/\mathbb{Q}$, und jedes Element x eines solchen Körpers ist algebraisch: Da K als $\mathbb{Q}$-Vektorraum endliche Dimension hat, können die Potenzen x^n nicht alle linear unabhängig sein; wir erhalten also eine lineare Abhängigkeit, d.h. ein Polynom aus $\mathbb{Q}[X]$, das für x verschwindet. Multiplikation mit dem Hauptnenner der Koeffizienten macht daraus ein Polynom aus $\mathbb{Z}[X]$, das bei x verschwindet.

Die Idee, daß es nichtalgebraische Zahlen geben könne, kam erst im Laufe des achtzehnten Jahrhunderts auf, unter anderem bei GOTTFRIED WILHELM LEIBNIZ (1646–1716), der von ihnen sagte: Omnem rationem trancendunt. (Sie übersteigen jede Vernunft.) 1844 konnte JOSEPH LIOUVILLE (1809–1882) als erster von einer reellen Zahl beweisen, daß sie transzendent ist; es handelte sich um die ansonsten völlig uninteressante Zahl $\sum 10^{-i!}$, wobei über alle $i \in \mathbb{N}$ summiert wird. 1874 zeigte GEORG CANTOR (1845–1918) durch eine Variante seines ersten Diagonalverfahrens, daß es nur abzählbar viele algebraische (reelle oder komplexe) Zahlen gibt, während er mit seinem zweiten Diagonalverfahren die Überabzählbarkeit von $\mathbb{R}$ und $\mathbb{C}$ bewies und daraus folgerte, daß es überabzählbar viele transzendente Zahlen gibt. Trotzdem ist es im Einzelfall meist sehr schwer, die Transzendenz einer Zahl zu beweisen; es gibt hier noch viele offene Probleme.

Wenn wir ausgehen von Punkten mit rationalen Koordinaten, hat jeder Punkt, den wir daraus mit Zirkel und Lineal konstruieren können, algebraische Zahlen als Koordinaten, und auch die Länge jeder konstruierten Strecke ist algebraisch. Zur Quadratur eines Kreises mit gegebenem (und daher algebraischem) Radius r müssen wir ein Quadrat mit Seitenlänge $r\sqrt{\pi}$ konstruieren. Falls dies mit Zirkel und Lineal möglich wäre, müßte $r\sqrt{\pi}$ und damit auch $\sqrt{\pi}$ und schließlich auch π selbst algebraisch sein. Aus der Transzendenz von π folgt somit die Unmöglichkeit der Quadratur des Kreises mit Zirkel und Lineal.

Ein erstes Problem beim Nachweis der Transzendenz von π ist eine geeignete Definition von π. Klassisch war π definiert als das Verhältnis des Kreisumfangs zu seinem Durchmesser, aber es gibt bislang keinen Transzendenzbeweis, der damit auskommt. Alle bekannten Beweise gehen aus von der Beziehung $e^{\pi i} = -1$.

1873 bewies CHARLES HERMITE die Transzendenz von e, danach erst

1882 CARL LOUIS FERDINAND VON LINDEMANN die von π. Für dessen Beweis wird die Transzendenz von e nicht wirklich benötigt, allerdings verwendete und erweiterte er HERMITEs Methoden, so daß sein Beweis besser verständlich wird, wenn wir zunächst den von HERMITE betrachten.

CHARLES HERMITE (1822–1901) war einer der bedeutendsten Mathematiker des neunzehnten Jahrhunderts. Zu seinen Resultaten zählen eine Vereinfachung des ABELschen Beweises, daß Gleichungen fünften Grades im allgemeinen nicht durch Wurzelausdrücke gelöst werden können, die explizite Lösung solcher Gleichungen durch elliptische Funktionen, die er sodann auf zahlentheoretische Probleme anwendete, der Nachweis, daß e eine transzendente Zahl ist, eine Interpolationsformel und vieles mehr. HERMITE galt als ein sehr guter akademischer Lehrer; er unterrichtete an der École Polytechnique, dem Collège de France, der École Normale Supérieure und der Sorbonne.

CARL LOUIS FERDINAND VON LINDEMANN (1852–1939) wurde in Hannover geboren und studierte in Göttingen, Erlangen, wo er bei FELIX KLEIN promovierte, und München. Danach besuchte er Universitäten in Oxford, Cambridge, London und Paris, wo er unter anderem mit HERMITE über dessen Transzendenzbeweis für e und seine Versuche, diesen auf π auszudehnen diskutierte. Während HERMITE damit erfolglos blieb, fand LINDEMANN 1882 den Trick, der HERMITE noch gefehlt hatte und konnte so mit dessen Methoden die Transzendenz von π beweisen. Zunächst aber habilitierte er sich 1877 in Würzburg und wurde daraufhin Professor zunächst in Freiburg, dann 1883 in Königsberg und schließlich 1893 in München. Er hatte über sechzig Doktoranden, darunter als wohl berühmtesten DAVID HILBERT (1862–1943). Auch mathematische Seminare in ihrer heutigen Form gehen im wesentlichen auf ihn zurück.

Die folgende Darstellung ist eine Vereinfachung der Beweise von HERMITE und LINDEMANN; sie folgt im wesentlichen dem Anhang zu Kapitel 1 aus dem Buch

ANTOINE CHAMBERT-LOIR: Algèbre corporelle, *Les éditions de l'École Polytechnique, Palaiseau,* 2005.

Ausgangspunkt ist, für ein zunächst beliebiges Polynom $f \in \mathbb{R}[X]$ und eine komplexe Zahl z, das Integral

$$I(f,z) = \int_0^1 z e^{z(1-u)} f(zu)\, du\,.$$

Sein Wert kann über Funktionswerte von f und seinen Ableitungen $f^{(j)}$ berechnet werden:

Lemma: Für ein Polynom $f \in \mathbb{R}[X]$ vom Grad d ist

$$I(f,z) = e^z \sum_{j=0}^{d} f^{(j)}(0) - \sum_{j=0}^{d} f^{(j)}(z)\,.$$

Beweis durch Induktion nach d: Für $d = 0$ ist f konstant; es gibt also ein $c \in \mathbb{R}$ mit $f(z) = c$ für alle $z \in \mathbb{R}$. Dann ist

$$I(f,z) = \int_0^1 cze^{z(1-u)}\, du = ce^z \int_0^1 ze^{-zu}\, du = ce^z \left(-e^{-zu}\Big|_0^1 \right)$$
$$= ce^z(-e^{-z}+1) = -c + ce^z = e^z f(0) - f(z)\,,$$

wie behauptet.

Für $d > 0$ führen wir das Integral durch partielle Integration auf $I(f',z)$ zurück: Der Integrand ist das Produkt von $f(zu)$ mit $ze^{z(1-u)}$, und wie wir beim Fall $d = 0$ gesehen haben, ist $ze^{z(1-u)}$ die Ableitung von $-e^{z(1-u)}$ nach u. Somit ist

$$I(f,z) = -e^{z(1-u)} f(zu)\Big|_0^1 + \int_0^1 ze^{z(1-u)} \cdot f'(zu)\, du$$
$$= -f(z) + e^z f(0) + I(f',z)\,.$$

f' ist ein Polynom vom Grad $d-1$; nach Induktionsannahme ist daher

$$I(f',z) = e^z \sum_{j=0}^{d-1} f^{(j+1)}(0) - \sum_{j=0}^{d-1} f^{(j+1)}(z) = e^z \sum_{j=1}^{d} f^{(j)}(0) - \sum_{j=1}^{d} f^{(j)}(z)\,.$$

Um $I(f,z)$ zu bekommen, müssen wir nach der vorangehenden Formel noch $-f(z)+e^z f(0)$ addieren; dadurch bekommen die beiden Summen rechts noch ihren Term zum Index $j = 0$, was die behauptete Formel beweist.

∎

Die grundsätzliche Strategie bei den Transzendenzbeweisen für e und π besteht darin, daß uns die Annahme, eine Zahl sei algebraisch, für geeignete Polynome f und geeignete Zahlen z untere Schranken für den gerade hergeleiteten Ausdruck liefert. Diese können wir dann vergleichen mit oberen Schranken für geeignete Summen solcher Integrale und dabei auf einen Widerspruch hoffen.

Für die Konstruktion unterer Schranken im Falle der Ableitungen in obiger Formel ist vor allem die folgende elementare Aussage nützlich:

Lemma: Für ein Polynom $f \in \mathbb{Z}[X]$ sind alle Koeffizienten der r-ten Ableitung durch $r!$ teilbar. Insbesondere sind damit auch alle Werte von $f^{(r)}(x)$ an ganzzahligen Stellen x durch $r!$ teilbar.

Beweis: Durch r-malige Ableitung wird der Summand $a_j X^j$ von f im Falle $r > 0$ zu Null, ansonsten zu

$$\big(j-(r-1)\big)\big(j-(r-2)\big)\cdots(j-1)j\cdot a_j X^{j-r} = r!\binom{j}{r}\cdot a_j X^{j-r}\ .$$

Da sowohl die Binomialkoeffizienten als auch alle a_j ganze Zahlen sind, sind alle Koeffizienten von $f^{(r)}$ durch $r!$ teilbar.

∎

Eine obere Schranke für $I(f,z)$ ist leicht zu finden: Der Betrag des Faktors $ze^{z(1-u)}$ ist für alle $u \in [0,\ 1]$ kleiner oder gleich $|z|\cdot e^{|z|}$, also ist der Betrag des Integranden höchstens gleich dieser Zahl mal dem Supremum von $|f(zu)|$ im Intervall $[0,\ 1]$. Um eine Schranke für das Integral zu bekommen, müssen wir nur noch mit der Länge eins des Integrationsintervalls multiplizieren und erhalten

Lemma: $|I(f,z)| \le |z|\cdot e^{|z|}\cdot \sup_{u\in[0,\,1]}|f(zu)|\ .$

∎

Damit haben wir im wesentlichen alles beisammen, um die Transzendenz von e zu beweisen:

Satz von HERMITE**:** e ist transzendent.

Beweis: Andernfalls müßte e Nullstelle eines Polynoms $f \in \mathbb{Z}[X]$ sein, also etwa

$$f(e) = a_d e^d + a_{d-1} e^{d-1} + \cdots + a_1 e + a_0 = 0 \quad \text{mit} \quad a_j \in \mathbb{Z}\,.$$

Dabei können wir annehmen, daß a_0 nicht verschwindet, denn andernfalls können wir so lange durch e dividieren, bis wir ein Polynom mit $a_0 \neq 0$ erhalten. Wir wollen uns überlegen, daß die Annahme $f(e) = 0$ auf einen Widerspruch führt.

Dazu betrachten wir für jede Primzahl p das Polynom

$$f_p = X^{p-1}(X-1)^p \cdots (X-d)^p$$

vom Grad $(d+1)p - 1$ und die Zahl

$$J_p = a_0 I(f_p, 0) + a_1 I(f_p, 1) + \cdots + a_d I(f_p, d)\,.$$

Wie wir oben nachgerechnet haben, ist für $z = k$

$$I(f_p, k) = e^k \sum_{j=0}^{(d+1)p-1} f_p^{(j)}(0) - \sum_{j=0}^{(d+1)p-1} f_p^{(j)}(k)\,.$$

Die erste Summe ist unabhängig von k; wenn wir also über k summieren, um J_p zu bestimmen, können wir diese Summe ausklammern und erhalten

$$\begin{aligned} J_p &= \sum_{j=0}^{(d+1)p-1} f_p^{(j)}(0) \cdot \sum_{k=0}^{d} a_k e^k - \sum_{k=0}^{d} a_k \sum_{j=0}^{(d+1)p-1} f_p^{(j)}(k) \\ &= -\sum_{k=0}^{d} a_k \sum_{j=0}^{(d+1)p-1} f_p^{(j)}(k)\,, \end{aligned}$$

denn nach unserer Annahme verschwindet $f(e) = \sum_{k=0}^{d} a_k e^k$.

Nach Konstruktion von f_p verschwinden auch alle $f_p^{(j)}(k)$ für $1 \leq k \leq d$ und $j = 0, \ldots, p-1$, denn k ist ja eine p-fache Nullstelle von f_p. Für $j \geq p$ wissen wir immerhin nach dem obigen Lemma, daß alle Funktionswerte von $f^{(j)}$ an ganzzahligen Stellen Vielfache von $j!$ und damit insbesondere auch von $p!$ sind.

Da die Null nur eine $(p-1)$-fache Nullstelle von f_p ist, verschwinden nur die Werte $f_p^{(j)}(0)$ mit $j < p-1$, und nach obigem Argument sind die mit $j \geq p$ durch $p!$ teilbar. Über $f_p^{(p-1)}(0)$ wissen wir noch nichts. Hier hilft uns die Verallgemeinerung der LEIBNIZ-Regel auf höhere Ableitungen, die man entweder aus der Analysis kennt oder durch Induktion mit Hilfe der klassischen Produktregel beweist: Für $r \geq 1$ und zwei mindestens r mal differenzierbare Funktionen u, v ist

$$(uv)^{(r)} = \sum_{j=0}^{r} \binom{r}{j} u^{(r-j)} v^{(j)} .$$

Dies wenden wir an auf $u = X^{p-1}$ und $v = (X-1)^p \cdots (X-d)^p$ und $r = p-1$. Alle Ableitungen von u außer der $(p-1)$-ten verschwinden an der Stelle Null; daher ist

$$f_p^{(p-1)}(0) = (uv)^{(p-1)}(0) = \sum_{j=0}^{p-1} \binom{p-1}{j} u^{(p-1-j)}(0) v^{(j)}(0)$$

$$= u^{(p-1)}(0) v(0) = (p-1)!(-1)^p \cdots (-d)^p = (p-1)! \cdot (-1)^{dp} d!^p ,$$

und das ist für $p > d$ eine nicht durch p teilbare Zahl. In der Summe

$$J_p = -\sum_{k=0}^{d} a_k \sum_{j=0}^{(d+1)p-1} f_p^{(j)}(k)$$

ist dann also jeder nichtverschwindende Summand durch $p!$ teilbar mit der einzigen möglichen Ausnahme von $a_0 f_p^{(p-1)}(0)$, der genau dann durch $p!$ teilbar ist, wenn a_0 Vielfaches von p ist. Für $p > |a_0|$ ist das nicht der Fall; daher ist für alle hinreichend großen Primzahlen p

$$J_p \equiv (-1)^{dp+1} a_0 (p-1)! d!^p \not\equiv 0 \bmod p! .$$

Kürzen durch $(p-1)!$, was nach obiger Rechnung auch $f_p^{(p-1)}(0)$ teilt, macht daraus

$$\frac{J_p}{(p-1)!} \equiv (-1)^{dp-1} a_0 d!^p \not\equiv 0 \bmod p .$$

Insbesondere ist $J_p/(p-1)!$ damit auch selbst ungleich Null. Da es eine ganze Zahl ist, muß es mindestens Betrag eins haben, d.h. $(p-1)! \leq \left|J_p\right|$.

Um dies zu einem Widerspruch zu führen, verwenden wir das Lemma vom Beginn des Paragraphen über eine obere Abschätzung für die Integrale $I(f, z)$: Danach ist

$$\left|I(f_p, k)\right| \le k \cdot e^k \cdot \sup_{u \in [0,\,1]} \left|f_p(ku)\right| = k \cdot e^k \cdot \sup_{x \in [0,\,k]} \left|f_p(x)\right| \ .$$

Die Schranke auf der rechten Seite ist offensichtlich monoton wachsend in k; daher ist

$$\left|I(f_p, k)\right| \le d \cdot e^d \cdot \sup_{x \in [0,\,d]} \left|f_p(x)\right|$$

für $k = 1, \ldots, d$ und

$$|J|_p \le \|f\|_1 \cdot d \cdot e^d \cdot \sup_{x \in [0,\,d]} \left|f_p(x)\right| \ ,$$

wobei $\|f\|_1 = |a_0| + \cdots + |a_d|$ die L^1-Norm jenes hypothetischen Polynoms f ist, das in e verschwindet.

Bleibt noch die Abschätzung der rechten Seite durch eine Funktion von p. Dazu zerlegen wir f_p wieder in ein Produkt

$$f_p = X^{p-1} \cdot g^p \quad \text{mit} \quad g = (X-1)(X-2)\cdots(X-d) \, .$$

Im Intervall $[0,\, d]$ ist $|x| \le d$, und $|g(x)|$ nimmt irgendwo sein Maximum M an. Somit ist

$$\sup_{x \in [0,\,d]} \left|f_p(x)\right| \le d^{p-1} M^p$$

für alle p und

$$\left|J_p\right| \le \|f\|_1 \cdot d \cdot e^d \cdot d^{p-1} M^p = \|f\|_1 e^d (dM)^p < \left(\|f\|_1 d e^d M\right)^p$$

für alle Primzahlen p. Zusammen mit der obigen Abschätzung zeigt dies, daß für $c = \|f\|_1 d e^d M$ und alle hinreichend großen Primzahlen p gilt

$$(p-1)! \le \left|J_p\right| \le c^p \, .$$

Das kann aber nicht sein, da $(p-1)!$ schneller wächst als jede p-te Potenz. Genauer ist nach der STIRLINGschen Formel

$$(p-1)! \sim \left(\frac{p-1}{e}\right)^{p-1} \sqrt{2\pi} \, ;$$

sobald $(p-1)/e$ hinreichend viel größer als c ist, kann die Ungleichung also unmöglich richtig sein. Damit führt die Annahme, e sei Nullstelle eines Polynoms mit ganzen Koeffizienten, zu einem Widerspruch, und die Transzendenz von e ist bewiesen. ∎

Der Beweis für die Transzendenz von π ist aufwendiger. Wie man heute weiß, bilden Funktionen wie die Exponentialfunktion algebraische Zahlen, abgesehen von offensichtlichen Ausnahmefällen wie $e^0 = 1$, nie auf algebraische Zahlen ab. Da $e^{\pi i} = -1$ algebraisch ist, kann daher πi und damit auch π nicht algebraisch sein. Der ursprüngliche Beweis von LINDEMANN konnte nicht auf solche allgemeinen Resultate zurückgreifen, verwendete aber auch die Beziehung $e^{\pi i} = -1$.

Wie HERMITE arbeitete er mit Integralen der Form $I(f, z)$, aber er definierte die Summen J_p nicht wie HERMITE, indem er die Koeffizienten eines hypothetischen Polynoms mit Nullstelle e *bzw.* π verwendete, sondern er arbeitete mit den sämtlichen Nullstellen eines solchen Polynoms. Um Summen über die Nullstellen eines Polynoms typographisch einfach darstellen zu können, definieren wir

Definition: $f \in \mathbb{C}[X]$ sei ein Polynom vom Grad d mit den (nicht notwendigerweise verschiedenen) Nullstellen $z_1, \ldots, z_d$, und $g\colon \mathbb{C} \to \mathbb{C}$ sei eine beliebige Funktion. Dann setzen wir

$$\sum_{f(z)=0} g(z) \underset{\text{def}}{=} \sum_{j=1}^{d} g(z_j) \quad \text{und} \quad \prod_{f(z)=0} g(z) \underset{\text{def}}{=} \prod_{j=1}^{d} g(z_j)\,.$$

Wenn f ein Polynom mit ganzzahligen Koeffizienten ist, sind die Nullstellen z_j algebraische Zahlen, und die Werte $g(z_j)$ werden im allgemeinen keine ganze Zahlen sein. Trotzdem gilt

Lemma: Für $f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0$ mit $a_i \in \mathbb{Z}$ und ein weiteres Polynom $g \in \mathbb{Z}[X]$ vom Grad n ist $a_d^n \sum_{f(z)=0} g(z)$ eine ganze Zahl.

Beweis: Unter einer Permutation der Nullstellen von f ändert sich in der Summe nur die Reihenfolge der Summanden; $\sum_{f(z)=0} g(z)$ ist also ein

symmetrisches Polynom in den Nullstellen von f. Nach dem Hauptsatz über symmetrische Funktionen läßt sich dieses schreiben als Polynom mit ganzzahligen Koeffizienten in den elementarsymmetrischen Funktionen der Nullstellen und damit nach dem Wurzelsatz von VIÈTE als Polynom mit ganzzahligen Koeffizienten in den Koeffizienten des Polynoms f/a_d. (Der Wurzelsatz von VIÈTE gilt nur für Polynome mit höchstem Koeffizient eins.) Da $\sum_{f(z)=0} g(z)$ als Polynom in den Nullstellen z_j von f höchstens den Grad n hat, hat auch dieses Polynom in den Koeffizienten a_i/a_d von f/a_d höchstens den Grad n. Durch Multiplikation mit a_d^n erhalten wir daher ein Polynom mit ganzzahligen Koeffizienten in $a_0, \ldots, a_d$ und damit eine ganze Zahl. ∎

Der erste Schritt im Beweis der Transzendenz von π zeigt, daß die entsprechende Aussage für eine Summe von Exponentialfunktionen höchstens dann gelten kann, wenn die Summe verschwindet:

Lemma: Für ein Polynom $f \in \mathbb{Z}[X]$ mit $f(0) \neq 0$ kann

$$\sum_{f(z)=0} e^z$$

keine von Null verschiedene ganze Zahl sein.

Beweis: Angenommen, die Summe wäre $N \in \mathbb{Z} \setminus \{0\}$. Wir wählen eine (beliebige) Primzahl p und betrachten das Polynom $g = X^{p-1} f^p$ aus $\mathbb{Z}[X]$. Wenn wir den Grad von f mit d bezeichnen, hat es den Grad $m = p - 1 + pd = p(d+1) - 1$. Wir verwenden wieder eine Summe gewisser der von HERMITE eingeführten Integrale, hier

$$J_p \underset{\text{def}}{=} \sum_{f(z)=0} I(g, z) \, .$$

Nach der oben bewiesenen Abschätzung ist

$$|I(g,z)| \leq |z| \, e^{|z|} \sup_{u \in [0,1]} |g(zu)| \, .$$

Da f ein Polynom ist, ist $|f(zu)|$ eine stetige Funktion von u und nimmt daher auf dem kompakten Intervall [0, 1] ein Maximum M_1 an. Der Betrag von uz ist dort kleiner oder gleich $|z|$. M_2 bezeichne das

Maximum der Beträge sämtlicher Nullstellen von f, falls dieses größer oder gleich eins ist, und eins sonst. Entsprechend sei M_3 das Maximum der $e^{|z|}$, mindestens aber eins. Dann ist

$$|I(g,z)| \leq M_2 \cdot M_3 \cdot M_2^{p-1} \cdot M_1^p \leq (M_1 M_2 M_3)^p$$

und $\left|J_p\right| \leq d(M_1 M_2 M_3)^p \leq (dM_1 M_2 M_3)^p$. Es gibt daher eine reelle Zahl M, für die gilt

$$\left|J_p\right| \leq M^p \,.$$

Um auch eine untere Schranke zu bekommen, beginnen wir wieder mit der expliziten Formel für den Wert von $I(g,z)$:

$$I(g,z) = e^z \sum_{j=0}^{m} g^{(j)}(0) - \sum_{j=0}^{m} g^{(j)}(z) \,.$$

Summation über alle Nullstellen von f führt auf

$$\begin{aligned} J_p &= \sum_{f(z)=0} I(g,z) = \left(\sum_{f(z)=0} e^z \right) \sum_{j=0}^{m} g^{(j)}(0) - \sum_{f(z)=0} \sum_{j=0}^{m} g^{(j)}(z) \\ &= N \sum_{j=0}^{m} g^{(j)}(0) - \sum_{j=0}^{m} \sum_{f(z)=0} g^{(j)}(z) \,, \end{aligned}$$

wobei N nach unserer Annahme eine von Null verschiedene ganze Zahl ist. Da Null eine $(p-1)$-fache Nullstelle von g ist, verschwindet $g^{(j)}(0)$ für $j < p-1$. Für $j = p-1$ wenden wir wieder die verallgemeinerte LEIBNIZ-Regel

$$(uv)^{(r)} = \sum_{k=0}^{r} \binom{r}{k} u^{(k)} v^{(r-k)}$$

an; hier für $u = X^{p-1}, v = f^p$ und $uv = g$. Da $u^{(j)}(0)$ für $j < p-1$ verschwindet und $u^{(p-1)}(0) = (p-1)!$ ist, folgt

$$g^{(p-1)}(0) = (p-1)! f(0)^p \,.$$

Für $j \geq p$ schließlich wissen wir, daß $g^{(j)}(0)$ durch $j!$ teilbar ist, also insbesondere durch $p!$. Es gibt daher eine ganze Zahl A_p, so daß gilt

$$\sum_{j=0}^{m} g^{(j)}(0) = (p-1)! f(0)^p + p! A_p \,.$$

Zur Untersuchung der Doppelsumme im Ausdruck für J_p beachten wir, daß jede Nullstelle z von f eine p-fache Nullstelle von g ist. Daher verschwindet $g^{(j)}(z)$ für $j < p$. Für $j \geq p$ sind alle Koeffizienten von $g^{(j)}$ durch $j!$ teilbar, also insbesondere durch $p!$. Es gibt daher Polynome $g_j \in \mathbb{Z}[X]$, so daß $g^{(j)} = p!g_j$ ist. Wie $g^{(j)}$ hat auch g_j dem Grad $m-j$. Nach dem vorigen Lemma ist daher $a_d^{m-j} \sum_{f(z)=0} g_j(z)$ eine ganze Zahl, und

$$a_d^{m-j} \sum_{f(z)=0} g^{(j)}(z) = a_d^{m-j} p! \sum_{f(z)=0} g_j(z)$$

ist eine durch $p!$ teilbare ganze Zahl. Für $j \geq p$ ist a_d^{m-j} ein Teiler von a_d^{m-p}; daher ist $a_d^{m-p} \sum_{f(z)=0} g^{(j)}(z)$ für alle $j \geq p$ eine durch $p!$ teilbare ganze Zahl. Es gibt somit eine ganze Zahl B_p derart, daß

$$-a_d^{m-p} \sum_{j=0}^{m} \sum_{f(z)=0} g^{(j)}(z) = -a_d^{m-p} \sum_{j=p}^{m} \sum_{f(z)=0} g^{(j)}(z) = p!B_p$$

ist. Insgesamt erhalten wir so die Formel

$$\begin{aligned} J_p &= N\big((p-1)!f(0)^p + p!A_p\big) + p!\frac{B_p}{a_d^{m-p}} \\ &= N(p-1)!f(0)^p + p!\left(NA_p + \frac{B_p}{a_d^{m-p}}\right) . \end{aligned}$$

Multiplikation mit a_d^{m-p} und Division durch $(p-1)!$ führt auf

$$\frac{a_d^{m-p}}{(p-1)!} J_p = N a_d^{m-p} f(0) + p(N a_d^{m-p} A_p + B_p) ,$$

und das ist eine ganze Zahl. Der Grad m von g ist $(d+1)p-1$, also ist $m-p = dp-1$

Nach unseren Annahmen sind N, a_d und $f(0)$ allesamt von Null verschieden; es gibt daher höchstens endlich viele Primzahlen, die eine dieser drei Zahlen teilen. Für jede andere Primzahl p ist diese ganze Zahl nicht durch p teilbar und damit insbesondere von Null verschieden. Sie hat daher mindestens den Betrag eins, was auf die Abschätzung

$$\big|J_p\big| \geq \frac{(p-1)!}{a_d^{m-p}} = \frac{(p-1)!}{a_d^{dp-1}}$$

führt, denn der Grad m von g ist $(p+1)d-1$. Wie wir bereits gezeigt haben, ist andererseits $\left|J_p\right| \leq M^p$ für eine geeignete Konstante M; daher ist

$$\frac{(p-1)!}{a_d^{dp-1}} \leq M^p \quad \text{und} \quad (p-1)! \leq \frac{(a_d M)^p}{a_d}$$

für jede hinreichend große Primzahl p. Da $(p-1)!$ schneller wächst als jede p-te Potenz, führt das auf den gesuchten Widerspruch. ∎

Nach diesen Vorbereitungen folgt nun recht schnell der Satz von LINDEMANN:

Satz: π ist transzendent.

Beweis: Wäre π algebraisch, so wäre auch πi algebraisch; es gäbe also ein irreduzibles Polynom $f \in \mathbb{Z}[X]$, das πi als Nullstelle hätte. Sein Grad sei d, und die (komplexen) Nullstellen von f seien $z_1, \ldots, z_d$. Da πi zu diesen Nullstellen gehört, ist

$$\prod_{f(z)=0} (1+e^z) = \prod_{j=1}^{d} (1+e^{z_j}) = 0\,.$$

Ausmultipliziert wird das zu

$$\sum_{\varepsilon \in \{0,1\}^d} e^{\sum_{j=1}^{d} \varepsilon_j z_j} = 0\,.$$

Die 2^d Summen $\sum \varepsilon_j z_j$ sind die Nullstellen des Polynoms

$$P_0 = \prod_{\varepsilon \in \{0,1\}^d} \left(X - \sum_{j=1}^{d} \varepsilon_j z_j \right)\,,$$

dessen Koeffizienten nach VIÈTE die elementarsymmetrischen Funktionen in diesen Summen sind.

Eine Permutation der Nullstellen $z_1, \ldots, z_d$ von f permutiert auch die Nullstellen $\sum \varepsilon_j z_j$ von P_0; daher lassen sich die Koeffizienten von P_0 nach dem Hauptsatz über symmetrische Polynome als Polynome in den

elementarsymmetrischen Funktionen in den z_j schreiben, also als Polynome in den Koeffizienten jenes normierten Polynoms, das die z_j als Nullstellen hat. Da f nicht als normiert vorausgesetzt war, sind dies nicht die Koeffizienten von f, sondern die von f/a_d, wobei a_d den führenden Koeffizienten von f bezeichnet. Die Koeffizienten von P_0 sind somit rationale Funktionen der Koeffizienten von f, und da letztere ganzzahlig sind, folgt, daß alle Koeffizienten von P_0 in $\mathbb{Q}$ liegen. Multiplizieren wir P_0 mit deren Hauptnenner und dividieren wir durch die größtmögliche X-Potenz X^q, erhalten wir ein Polynom $P \in \mathbb{Z}[X]$ mit $P(0) \neq 0$, das alle nichtverschwindenden Summen $\sum \varepsilon_j z_j$ als Nullstellen hat.

Nun ist einerseits

$$\sum_{\varepsilon \in \{0,1\}^d} e^{\sum_{j=1}^d \varepsilon_j z_j} = \prod_{f(z)=0} (1 + e^z) = 0 \,,$$

andererseits ist

$$\sum_{\varepsilon \in \{0,1\}^d} e^{\sum_{j=1}^d \varepsilon_j z_j} = \sum_{P_0(z)=0} e^z = qe^0 + \sum_{P(z)=0} e^z = q + \sum_{P(z)=0} e^z \,.$$

Wenn alle ε_j verschwinden, ist die Summe der $\varepsilon_j z_j$ gleich Null; daher ist $q \geq 1$. Zusammen mit den beiden letzten Formelzeilen folgt

$$\sum_{P(z)=0} e^z = -q \in \mathbb{Z} \setminus \{0\} \,,$$

was nach dem vorigen Lemma unmöglich ist. Die Annahme, π sei algebraisch, führt daher zu einem Widerspruch, d.h. π ist transzendent. ∎

§6: Endliche Körper

Da jeder Körper der Charakteristik Null den Körper der rationalen Zahlen enthält, haben endliche Körper k notwendigerweise positive Charakteristik. Ist $\operatorname{char} k = p$, so enthält k einen zu $\mathbb{F}_p$ isomorphen Teilkörper, das Bild des kononischen Homomorphismus $\mathbb{Z} \to k$, und ist somit isomorph zu einem $\mathbb{F}_p$-Vektorraum. Die Elementanzahl eines endlichen Körpers ist somit stets eine Primzahlpotenz.

Wie wir in §2 gesehen haben, ist die Abbildung

$$F\colon \begin{cases} k \to k \\ x \mapsto x^p \end{cases}$$

für jeden Körper der Charakteristik p ein Homomorphismus, der sogenannte FROBENIUS-Homomorphismus. Wenn wir ihn r mal hintereinander ausführen, erhalten wir die Abbildung, die jedes Element $x \in k$ auf x^{p^r} abbildet; auch sie ist natürlich ein Homomorphismus, den wir mit F^r bezeichnen.

Ein Homomorphismus eines Körpers in einen Körper ist stets injektiv. Das gilt natürlich auch für die Homomorphismen $F^r\colon k \to k$. Im Falle eines endlichen Körpers k folgt aus der Injektivität die Surjektivität, in diesem Fall ist F^r also ein Automorphismus von k. Für die Körper $\mathbb{F}_p$ können wir den kleinen Satz von FERMAT auch so formulieren, daß F (und damit auch jedes F^r) die identische Abbildung ist.

Aus §3 des vorigen Kapitels wissen wir, daß die multiplikative Gruppe eines endlichen Körpers stets zyklisch ist. In einem Körper k mit p^n Elementen hat sie die Ordnung $p^n - 1$, so daß es ein Element x der Ordnung $p^n - 1$ geben muß. Dieses ist natürlich, genau wie alle seine Potenzen, eine Nullstelle des Polynoms $X^{p^n-1} - 1$; da dessen Grad gleich der Elementanzahl der multiplikativen Gruppe von k ist, besteht diese somit genau aus den Nullstellen dieses Polynoms. Insbesondere ist k ein Zerfällungskörper von $X^{p^n-1} - 1$, und da alle Zerfällungskörper eines festen Polynoms zueinander isomorph sind, sind auch alle Körper mit p^n Elementen zueinander isomorph.

Da das Polynom $X^{p^n-1} - 1$ alle Elemente außer der Null als Nullstellen hat, hat $X^{p^n} - X$ *alle* Elemente eines Körpers mit p^n Elementen als Nullstelle; somit ist die n-te Potenz F^n des FROBENIUS-Automorphismus gleich der Identität auf k.

Mit dieser Bemerkung können wir auch leicht einsehen, daß es zu jeder Primzahlpotenz p^n einen Körper mit p^n Elementen gibt: Wir bilden zunächst über $\mathbb{F}_p$ den Zerfällungskörper des Polynoms $X^{p^n-1} - 1$; er enthält alle Nullstellen dieses Polynoms und natürlich auch die Null. Diese Elemente bilden zusammen einen Teilkörper, nämlich den

Fixkörper von F^n. Da der Zerfällungskörper der kleinste Körper ist, der alle Nullstellen enthält, ist er gleich dieser Menge aus p^n Elementen.

Zusammenfassend können wir festhalten

Satz: Für jede Primzahlpotenz p^n gibt es Körper mit p^n Elementen; sie sind alle zueinander isomorph. Für jedes Element x eines solchen Körpers ist $x^{p^n} = x$, die n-te Potenz des FROBENIUS-Automorphismus ist also die Identität. ∎

Wir bezeichnen „den“ Körper mit p^n Elementen mit $\mathbb{F}_{p^n}$.

Falls der Körper $\mathbb{F}_{p^n}$ einen der Körper $\mathbb{F}_{p^m}$ enthält, ist er ein $\mathbb{F}_{p^m}$-Vektorraum; daher ist p^n eine Potenz von p^m, d.h. m muß ein Teiler von n sein. Ist umgekehrt m ein Teiler von n, so folgt aus $x^m = 1$, daß auch $x^n = 1$ ist, d.h. jede Nullstelle von $X^m - 1$ (im Zerfällungskörper $\mathbb{F}_{p^m}$) ist auch eine Nullstelle von $X^n - 1$, so daß $X^m - 1$ ein Teiler von $X^n - 1$ ist und der Zerfällungskörper von $X^n - 1$ einen Zerfällungskörper von $X^m - 1$ enthält, d.h. $\mathbb{F}_{p^m}$ ist in $\mathbb{F}_{p^n}$ enthalten.

$\mathbb{F}_{p^m}$ ist der Fixkörper unter F^m; daher ist die Erweiterung $\mathbb{F}_{p^n}/\mathbb{F}_{p^m}$ GALOISsch. Die GALOIS-Gruppe wird erzeugt von F^m; da F^n die Identität auf $\mathbb{F}_{p^n}$ ist, ist diese eine zyklische Gruppe mit n/m Elementen.

Zum expliziten Rechnen in einem Körper $\mathbb{F}_{p^n}$ muß dieser zunächst irgendwie konkret als Vektorraum über $\mathbb{F}_p$ dargestellt werden; in $\mathbb{F}_p$ können wir schließlich rechnen. Wir wissen, daß $\mathbb{F}_{p^n}$ aus $\mathbb{F}_p$ entsteht durch Adjunktion eines erzeugenden Elements x der Gruppe $\mathbb{F}_{p^n}^\times$; da die Körpererweiterung den Grad n hat, ist x Nullstelle eines irreduziblen Polynoms vom Grad n über $\mathbb{F}_p$. Ist umgekehrt f ein irreduzibles Polynom vom Grad n, so ist $\mathbb{F}_p[X]/(f)$ über $\mathbb{F}_p$ eine Körpererweiterung vom Grad n, hat also p^n Elemente und ist somit isomorph zu $\mathbb{F}_{p^n}$.

Wenn wir also ein irreduzibles Polynom $f \in \mathbb{F}_p[X]$ vom Grad n gefunden haben, können wir $\mathbb{F}_{p^n}$ identifizieren mit $\mathbb{F}_p[X]/(f)$, und dort können wir die Elemente identifizieren mit den Polynomen aus $\mathbb{F}_p[X]$ vom Grad kleiner n. Die Addition und Subtraktion sind problemlos, bei der Multiplikation erhalten wir im allgemeinen allerdings ein Polynom

vom Grad n oder größer. Dieses muß dann ersetzt werden durch seinen Rest bei der Division durch f. Multiplikative Inverse schließlich lassen sich mit Hilfe des erweiterten EUKLIDischen Algorithmus bestimmen: Ist das Element $x \neq 0$ aus $\mathbb{F}_{p^n}$ gegeben durch das Polynom $g \in \mathbb{F}_p[X]$ vom Grad kleiner n, so sind f und g teilerfremd, da $g \neq 0$ und f irreduzibel ist. Daher gibt es Polynome g^*, f^* mit $\deg g^* < \deg f = n$ und $\deg f^* < \deg g$, so daß $gg^* + ff^* = 1$ ist. Modulo f ist somit $gg^* = 1$.

Die Computeralgebra kennt, insbesondere für Polynome aus $\mathbb{F}_p[X]$, effiziente Faktorisierungsverfahren; man kann sich daher irreduzible Polynome vom Grad n über $\mathbb{F}_p$ verschaffen, indem man das Polynom $X^{p^n-1} - 1 \in \mathbb{F}_p[X]$ in seine irreduziblen Faktoren zerlegt und einen der Faktoren vom Grad n auswählt. Wegen der Existenz des Körpers $\mathbb{F}_{p^n}$ muß es mindestens einen solchen Faktor geben, oft gibt es aber mehrere, die aber alle zu zueinander isomorphen Körpern führen.

Im Falle des Körpers $\mathbb{F}_{256}$, der sowohl für den *Advanced Encryption Standard* AES als auch für die Fehlerkorrektur auf CDs und DVDs verwendet wird, lassen sich die Faktoren von $X^{255} - 1$ über $\mathbb{F}_2$ per Computer leicht bestimmen. Wie sich zeigt, haben dreißig davon den Grad acht, den wir für einen Körper mit $256 = 2^8$ Elementen brauchen. Zweckmäßigerweise sollten wir einen wählen, der das Rechnen modulo diesem Polynom möglichst einfach macht; insbesondere sollte das verwendete Polynom möglichst wenige Terme habe.

Dreizehn der dreißig Polynome haben sieben nichtverschwindende Terme, die restlichen siebzehn nur fünf. Wir wählen natürlich eines der letzteren. Alle diese Polynome haben, wie jedes Polynom vom Grad acht über $\mathbb{F}_2$, den führenden Term X^8; danach folgen vier weitere Terme. Bei der Reduktion modulo einem solchen Polynom $P = X^8 + \mathit{Rest}$ benutzt man, daß dann

$$X^8 \equiv \mathit{Rest}, \qquad X^9 \equiv X \cdot \mathit{Rest}, \quad \ldots$$

ist; dies wird umso häufiger mehrfach angewandt werden müssen, je höheren Grad das Polynom *Rest* hat. Am effizientesten kann man also rechnen, wenn das Polynom *Rest* den kleinstmöglichen Grad hat. Bei unseren siebzehn Kandidaten ist dies der Grad vier; er kommt zweimal

vor, nämlich bei

$$X^8 + X^4 + X^3 + X + 1 \quad \text{und} \quad X^8 + X^4 + X^3 + X^2 + 1\,.$$

Das erste dieser Polynome wird für AES verwendet, das zweite bei der Fehlerkorrektur auf CDs.

Die genaue Festlegung für das Rechnen in $\mathbb{F}_{256} = \mathbb{F}_2^8$ für die Zwecke von AES ist folgende: Wir schreiben ein Byte als $(a_7, a_6, \ldots, a_0)$ und identifizieren es mit dem Polynom

$$a_7X^7 + a_6X^6 + \cdots + a_1X + a_0\,;$$

das Byte 0000 0010 entspricht also X.

Der Einfachheit halber schreiben wir Bytes meist als zweiziffrige Hexadezimalzahlen: Im betrachteten Beispiel wäre das 02_{hex}, und das Byte $\text{A5}_{\text{hex}} = 1010\,0101$ entspricht dem Polynom $X^7 + X^5 + X^2 + 1$.

Man beachte, daß trotz dieser Schreibweise die Addition und Multiplikation in $\mathbb{F}_{256}$ natürlich nichts mit der Addition und Multiplikation von Hexadezimalzahlen zu tun haben. Zwar ist $01_{\text{hex}} + 02_{\text{hex}} = 03_{\text{hex}}$, aber $01_{\text{hex}} + 01_{\text{hex}} = 00_{\text{hex}}$ und $05_{\text{hex}} + 04_{\text{hex}} = 01_{\text{hex}}$.

Zur Berechnung von $\text{A5}_{\text{hex}} \cdot 01_{\text{hex}}$ müssen wir das Polynom

$$(X^7 + X^5 + X^2 + 1) \cdot X = X^8 + X^6 + X^3 + X$$

berechnen und modulo $m(X) = X^8 + X^4 + X^3 + X + 1$ reduzieren. Da

$$X^8 \bmod m(X) = X^4 + X^3 + X^2 + 1$$

ist (in $\mathbb{F}_2$ ist $-1 = 1$), ist dies

$$(X^4 + X^3 + X^2 + 1) + (X^6 + X^3 + X) = X^6 + X^4 + X^2 + X + 1\,.$$

Somit ist $\text{A5}_{\text{hex}} \cdot 01_{\text{hex}} = 56_{\text{hex}}$.

Trotz der Wahl des optimalen Polynoms ist die Multiplikation also immer noch erheblich aufwendiger als die Addition.

§7: Mehr über Einheitswurzeln

Die n-ten Einheitswurzeln spielen eine wesentliche Rolle bei der Konstruktion des regelmäßigen n-Ecks mit Zirkel und Lineal, und wir wir gerade gesehen haben, sind auch alle Elemente außer der Null im Körper mit p^n Elementen $(p^n - 1)$-te Einheitswurzeln. Es liegt daher nahe, Einheitswurzeln etwas genauer zu betrachten.

Definition: Für einen Körper k bezeichnen wir

$$\mu_n(k) = \left\{ x \in k \;\middle|\; x^n = 1 \right\}$$

als die Gruppe der n-ten Einheitswurzeln von k. Der Zerfällungskörper des Polynoms $X^n - 1$ über k heißt der n-te *Kreisteilungskörper* über k und wird mit k_n bezeichnet.

$\mu_n(k)$ ist in der Tat eine Gruppe, denn natürlich enthält sie die Eins, und ist $x^n = y^n = 1$, so ist auch $(xy)^n = x^n y^n = 1$. Auch das inverse Element x^{-1} liegt in $\mu_n(k)$, denn $\left(x^{-1}\right)^n = x^{-n} = \left(x^n\right)^{-1} = 1$.

Für $k = \mathbb{Q}$ und $k = \mathbb{R}$ beispielsweise haben wir

$$\mu_n(\mathbb{Q}) = \mu_n(\mathbb{R}) = \begin{cases} \{1\} & \text{für ungerade } n \\ \{1, -1\} & \text{für gerade } n \end{cases} .$$

Deutlich größer ist

$$\mu_n(\mathbb{C}) = \left\{ e^{2\pi i j/n} \;\middle|\; j = 0, \ldots, n-1 \right\} .$$

Für endliche Körper ist $\mu_{p-1}(\mathbb{F}_p) = \mathbb{F}_p^\times$ nach dem kleinen Satz von FERMAT. Da die multiplikative Gruppe eines endlichen Körpers zyklisch ist, gilt allgemeiner auch $\mu_{p^n-1}(\mathbb{F}_{p^n}) = \mathbb{F}_{p^n}^\times$, aber $\mu_{p^m}(\mathbb{F}_{p^n}) = \{1\}$ für alle m, denn $x^{p^n} = x$ für alle x, so daß für $x \neq 1$ auch x^{p^m} nicht gleich eins sein kann.

Lemma: Falls die Charakteristik von k kein Teiler von n ist, ist $\mu_n(k_n)$ eine Gruppe der Ordnung n. Ist $n = p^r m$ mit einer zu $p = \operatorname{char} k$ teilerfremden Zahl m, so ist $\mu_n(k_n) = \mu_m(k_n)$ eine Gruppe der Ordnung m.

Beweis: Falls n kein Vielfaches der Charakteristik ist, ist die Ableitung nX^{n-1} des Polynoms $X^n - 1$ nicht das Nullpolynom. Da nX^{n-1} nur für $x = 0$ verschwindet, $X^n - 1$ dort aber den Wert -1 annimmt, gibt es keine gemeinsame Nullstelle des Polynoms mit seiner Ableitung, so daß alle Nullstellen von $X^n - 1$ einfach sind. Da k_n der Zerfällungskörper dieses Polynoms ist, hat es dort also n verschiedene Nullstellen.

Ist $n = p^r m$ und $\operatorname{char} k = p$, so ist $(X^m - 1)^{p^r} = X^{mp^r} - 1^{p^m} = X^n - 1$, und $X^m - 1$ hat, da p kein Teiler von m ist, m verschiedene Nullstellen. Jede dieser Nullstellen ist eine p^r-fache Nullstelle von $X^n - 1$. ∎

Als nächstes wollen wir uns überlegen, daß $\mu_n(k)$ stets eine zyklische Gruppe ist. Wir betrachten zunächst Körper k der Charakteristik Null. Da diese $\mathbb{Q}$ als Teilkörper enthalten, ist auch $\mathbb{Q}_n$ ein Teilkörper von k_n. Da sowohl $\mu(\mathbb{Q}_n)$ als auch $\mu(k_n)$ die Ordnung n hat, ist $\mu_n(k_n) = \mu_n(\mathbb{Q}_n)$. Somit ist $\mu_n(k_n)$ genau dann zyklisch, wenn $\mu_n(\mathbb{Q}_n)$ zyklisch ist. Da $\mathbb{C}$ algebraisch abgeschlossen ist und $\mathbb{Q}_n$ enthält, ist auch $\mu_n(\mathbb{Q}_n) = \mu_n(\mathbb{C})$, und letzteres ist eine zyklische Gruppe, die von $e^{2\pi i/n}$ erzeugt wird. Somit ist $\mu_n(k_n)$ in Charakteristik Null stets zyklisch. $\mu_n(k)$ als Untergruppe davon ist ebenfalls zyklisch nach dem folgenden

Lemma: Jede Untergruppe einer zyklischen Gruppe ist zyklisch.

Beweis: Die zyklische Gruppe G sei erzeugt vom Element $g \in G$, und U sei eine Untergruppe. Falls U nur aus dem Neutralelement besteht, ist U zyklisch. Andernfalls enthält U Potenzen g^n mit $n > 0$; der kleinste vorkommende positive Exponent sei s. Für jedes Element g^m von U können wir m mit Rest durch s dividieren und erhalten eine Gleichung $m = qs + r$ mit $0 \leq r < s$. Mit g^m und $g^{qs} = \left(g^s\right)^q$ liegt auch $g^r = g^m g^{-qs}$ in U; wegen der Minimalität von s muß daher $r = 0$ sein. Somit ist g^m eine Potenz von g^s, d.h. g^s erzeugt die Untergruppe U. ∎

Ist $\operatorname{char} k = p > 0$, so enthält k den Körper $\mathbb{F}_p$ als Teilkörper und k_n damit den Körper $\left(\mathbb{F}_p\right)_n$. Dieser ist ein endlicher Körper, und wie wir aus Kapitel 3, §3, wissen, ist die multiplikative Gruppe eines endlichen Körpers stets zyklisch. Damit ist auch $\mu_n\left(\left(\mathbb{F}_p\right)_n\right)$ als Untergruppe dieser multiplikativen Gruppe zyklisch, und da $\mu_n(k_n)$ nach dem Lemma die gleiche Ordnung wie $\mu_n\left(\left(\mathbb{F}_p\right)_n\right)$ hat, ist auch $\mu_n(k_n)$ zyklisch. $\mu_n(k)$ schließlich ist zyklisch als Untergruppe davon.

Definition: Eine n-te Einheitswurzel $\zeta \in \mu_n(k)$ heißt *primitive n-te Einheitswurzel,* wenn es keine natürliche Zahl $d < n$ gibt, für die bereits $\zeta^d = 1$ ist. Die Menge der primitiven n-ten Einheitswurzeln wird mit $\mu_n^*(k)$ bezeichnet. Falls $\mu_n^*(k_n)$ nicht leer ist, bezeichnen wir

$$\Phi_n = \prod_{\zeta \in \mu_n^*(k_n)} (X - \zeta)$$

als das n-te *Kreisteilungspolynom* über k.

Beispielsweise ist also $\mu_4(\mathbb{R}) = \emptyset$, aber $\mu_4^*(\mathbb{C}) = \{i, -i\}$. Da $\mathbb{C}$ der Zerfällungskörper von $X^4 - 1$ über $\mathbb{R}$ ist, ist $(X + i)(X - i) = X^2 + 1$ das vierte Kreisteilungspolynom sowohl über $\mathbb{R}$ als auch über $\mathbb{C}$.

Zumindest für $k = \mathbb{Q}$ sind wir dem n-ten Kreisteilungspolynom in einem anderen Zusammenhang bereits begegnet: Für den Satz von GAUSS, wonach das regelmäßige n-Eck genau dann mit Zirkel und Lineal konstruiert werden kann, wenn $\varphi(n)$ eine Zweierpotenz ist, betrachteten wir einen irreduziblen Faktor des Polynoms $X^n - 1$, der eine primitive n-te Einheitswurzel als Nullstelle hat, und zeigten dann mit einem Argument von DEDEKIND, daß dieser Faktor genau die primitiven n-ten Einheitswurzeln als Nullstellen hat. Somit ist dieser Faktor, wenn wir ihn auf führenden Koeffizienten eins normieren, gerade das n-te Kreisteilungspolynom Φ_n. Insbesonder folgt:

Satz: Das n-te Kreisteilungspolynom Φ_n über $\mathbb{Q}$ ist irreduzibel. ∎

Sobald wir eine primitive n-te Einheitswurzel kennen, können wir leicht auch alle anderen bestimmen nach dem folgenden

Lemma: Ist G eine zyklische Gruppe der Ordnung n und ist g ein Erzeugendes von G, so ist g^r genau dann ebenfalls ein Erzeugendes, wenn r teilerfremd zu n ist. Ist also ζ eine primitive n-te Einheitswurzel, so besteht $\mu_n^*(k)$ genau aus den Elementen ζ^r mit $0 \leq r < n$ und $\mathrm{ggT}(n, r) = 1$.

Beweis: Ist r teilerfremd zu n, liefert uns der erweiterte EUKLIDische Algorithmus ganze Zahlen α, β, für die $\alpha r + \beta n = 1$ ist. Damit ist

$$g = g^{\alpha r + \beta n} = \left(g^r\right)^{\alpha} \left(g^n\right)^{\beta} = \left(g^r\right)^{\alpha}$$

eine Potenz von g^r, so daß sich jede Potenz von g auch als eine Potenz von g^r schreiben läßt. Ist dagegen d ein positiver gemeinsamer Teiler von r und n, so ist $\alpha r \bmod n$ für jedes $\alpha \in \mathbb{Z}$ durch d teilbar, so daß das Erzeugnis von g^r keine Potenzen g^e enthalten kann, für die e kein Vielfaches von d ist; insbesondere ist g keine Potenz von g^r. ∎

Korollar: Falls char k kein Teiler von n ist, enthält $\mu_n^*(k_n)$ genau $\varphi(n)$ Elemente.

Beweis: Nach Kapitel 2 haben genau die zu n teilerfremden Elemente von $\mathbb{Z}/n$ ein multiplikatives Inverses, und die EULERsche φ-Funktion gibt die Elementanzahl der primen Restklassengruppe $\left(\mathbb{Z}/n\right)^\times$ an. ∎

Lemma: Ist char $k = 0$, so ist Φ_n ein Polynom mit ganzzahligen Koeffizienten vom Grad $\varphi(n)$.

Beweis durch Induktion nach n.

Für $n = 1$ hat $\Phi_1 = X - 1$ offensichtlich ganzzahlige Koeffizienten.

Nun sei $n > 1$; wir nehmen an, daß Φ_d für $d < n$ in $\mathbb{Z}[X]$ liegt. Ist $\zeta \in \mu_n(k_n)$ keine primitive n-te Einheitswurzel, so gibt es einen echten Teiler d von n derart, daß ζ eine primitive d-te Einheitswurzel ist. Deshalb ist

$$X^n - 1 = \prod_{d|n} \Phi_d \,.$$

Nach Induktionsvoraussetzung sind die Φ_d mit $d < n$ Polynome mit ganzzahligen Koeffizienten; daher liegt auch

$$\Psi_n = \prod_{\substack{d|n \\ d<n}} \Phi_d$$

in $\mathbb{Z}[X]$. In $\mathbb{Q}[X]$ haben wir eine Polynomdivision mit Rest; dort dividieren wir $X^n - 1$ durch Ψ_n. Der Quotient sei Q, und der Rest R ist entweder das Nullpolynom oder ein Polynom, dessen Grad kleiner als der von Ψ_n ist. Da der führende Koeffizient eines Kreisteilungspolynoms nach Konstruktion stets eins ist, hat auch Ψ_n den führenden Koeffizienten eins, so daß bei der Polynomdivision keine Nenner entstehen. Daher liegen sowohl Q als auch R in $\mathbb{Z}[X]$, und sie sind die einzigen Polynome aus $\mathbb{Q}[X]$ mit $\deg R < \deg \Psi_n$, für die $X^n - 1 = Q\Psi_n + R$ ist. In $\mathbb{Q}[X]$ ist aber auch $X^n - 1 = \Phi_n \Psi_n$, also muß $R = 0$ sein und $\Phi_n = Q$ liegt in $\mathbb{Z}[X]$. ∎

Da Φ_d den Grad $\varphi(d)$ hat, folgt aus der Formel $X^n - 1 = \prod_{d|n} \Phi_d$ durch Gradvergleich, daß

$$n = \sum_{d|n} \varphi(d)$$

sein muß, beispielsweise ist also

$$6 = \varphi(6) + \varphi(3) + \varphi(2) + \varphi(1) = 2 + 2 + 1 + 1 \, .$$

Die Formel erlaubt auch die rekursive Berechnung der Polynome Φ_n:

$$\Phi_1 = X - 1$$

$$\Phi_2 = \frac{X^2 - 1}{\Phi_1} = \frac{X^2 - 1}{X - 1} = X + 1$$

$$\Phi_3 = \frac{X^3 - 1}{\Phi_1} = \frac{X^3 - 1}{X - 1} = X^2 + X + 1$$

$$\Phi_4 = \frac{X^4 - 1}{\Phi_1\Phi_2} = \frac{X^4 - 1}{(X - 1)(X + 1)} = X^2 + 1$$

$$\Phi_5 = \frac{X^5 - 1}{\Phi_1} = \frac{X^5 - 1}{X - 1} = X^4 + X^3 + X^2 + X + 1$$

$$\Phi_6 = \frac{X^6 - 1}{\Phi_1\Phi_2\Phi_3} = \frac{X^6 - 1}{(X - 1)(X + 1)(X^2 + X + 1)} = X^2 - X + 1$$

$$\Phi_7 = \frac{X^7 - 1}{\Phi_1} = \frac{X^7 - 1}{X - 1} = X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$$

$$\Phi_8 = \frac{X^8 - 1}{\Phi_1\Phi_2\Phi_4} = \frac{X^8 - 1}{(X - 1)(X + 1)(X^2 + 1)} = X^4 + 1 \, ,$$

und so weiter. Nach dem obigen Satz sind sie allesamt irreduzibel in $\mathbb{Q}[X]$ und damit wegen ihrer Primitivität auch in $\mathbb{Z}[X]$. Wenn wir diese Polynome allerdings über einen Körper positiver Charakteristik betrachten, müssen sie dort nicht irreduzibel sei: Über $\mathbb{F}_3$ ist beispielsweise $(X + 2)^2 = X^2 + 4X + 1 = X^2 + X + 1$ eine Zerlegung von Φ_3, und über $\mathbb{F}_7$ ist $\Phi_3 = (X + 3)(X + 5)$.

Auch über Erweiterungskörpern von $\mathbb{Q}$ müssen die Kreisteilungspolynome natürlich nicht irreduzibel sei; über $\mathbb{C}$ etwa zerfallen sie trivialer-

weise in ein Produkt von Linearfaktoren. Die Irreduzibilität von Φ_n über einem Körper k hängt eng mit der GALOIS-Gruppe von k_n/k zusammen:

Lemma: Falls n kein Vielfaches von $\operatorname{char} k$ ist, ist k_n/k eine GALOISsche Erweiterung von k, und ihre GALOIS-Gruppe ist isomorph zu einer Untergruppe von $\left(\mathbb{Z}/n\right)^\times$. Sie ist genau dann gleich $\left(\mathbb{Z}/n\right)^\times$, wenn Φ_n in $k[X]$ irreduzibel ist.

Beweis: Falls n kein Vielfaches von $\operatorname{char} k$ ist, können wir wie im Fall $k = \mathbb{Q}$ argumentieren: Die Ableitung nX^{n-1} von $X^n - 1$ ist dann nicht das Nullpolynom und hat keine gemeinsame Nullstelle mit $X^n - 1$; daher ist $X^n - 1$ separabel und k_n/k als Zerfällungskörper eines separablen Polynoms GALOISsch. Für jede primitive n-te Einheitswurzel ζ ist $k_n = k(\zeta)$, und jeder Automorphismus von k_n/k muß ζ wieder auf eine primitive n-te Einheitswurzel abbilden. Da sich diese als Potenz von ζ schreiben läßt mit einem zu n teilerfremden Exponenten, muß $\operatorname{Aut}(k_n/k)$ eine Untergruppe von $\left(\mathbb{Z}/n\right)^\times$ sein. Genau dann, wenn Φ_n in $k[X]$ irreduzibel ist, ist $k_n \cong k[X]/(\Phi_n)$ und hat somit den Grad $\varphi(n)$ über k, so daß die GALOIS-Gruppe in diesem Fall $\varphi(n)$ Elemente hat und damit ganz $\left(\mathbb{Z}/n\right)^\times$ sein muß.

∎

Speziell für einen endlichen Körper $k = \mathbb{F}_q$ können wir die GALOIS-Gruppe explizit angeben:

Lemma: Ist $k = \mathbb{F}_q$ mit $q = p^r$, und ist p kein Teiler von n, so ist $\operatorname{Aut}(k_n/k)$ isomorph zur von q mod n in $\left(\mathbb{Z}/n\right)^\times$ erzeugten zyklischen Untergruppe.

Beweis: Nach dem vorigen Lemma ist die Erweiterung GALOISsch. Die GALOIS-Gruppe einer endlichen Erweiterung von $\mathbb{F}_q$ ist zyklisch und wird erzeugt von der r-ten Potenz des FROBENIUS-Automorphismus F. Für jede primitive n-te Einheitswurzel ζ ist $k_n = \mathbb{F}_q(\zeta)$, und ist dies der Körper mit p^s Elementen, so ist $F^s(\zeta) = \zeta^{p^s} = \zeta$ für jede primitive n-te Einheitswurzel ζ. Also ist $\zeta^{p^s-1} = 1$, so daß n ein Teiler von $p^s - 1$ sein muß. Ist $s = ar$, so ist $p^s - 1 = q^a - 1$, d.h. in $\left(\mathbb{Z}/n\right)^\times$ hat q die Ordnung a. Da $\mathbb{F}_{p^s}$ als $\mathbb{F}_{p^r}$-Vektorraum die Dimension a hat, ist

dies auch die Ordnung der GALOIS-Gruppe, die somit isomorph ist zur von q mod n erzeugten Untergruppe der primen Restklassengruppe. ∎

Um zu sehen, wie wir anhand der GALOIS-Gruppe die Zwischenkörper von k_n/k explizit bestimmen können, brauchen wir zunächst ein allgemeines Lemma über GALOISsche Erweiterungen. Ziemlich am Anfang von §2 hatten wir für eine GALOISsche Erweiterung K/k die Spur $S(x)$ eines Elements $x \in K$ definiert als die Summe aller Bilder von x unter den Elementen der GALOIS-Gruppe. Jetzt definieren wir etwas allgemeiner

Definition: Für eine Untergruppe H der GALOIS-Gruppe G der Körpererweiterung K/k setzen wir

$$S_H(x) \underset{\text{def}}{=} \sum_{\sigma \in H} \sigma(x) \, .$$

Für $H = G$ ist das gerade die Spur, und wie wir wissen, liegt $S(x)$ stets in k. Völlig analog können wir zeigen, daß $S_H(x)$ im Fixkörper K^H liegen muß, denn für jedes Element $\tau \in H$ ist

$$\tau\big(S_H(x)\big) = \tau\left(\sum_{\sigma \in H} \sigma(x)\right) = \sum_{\sigma \in H} (\tau \circ \sigma)(x) = S_H(x) \, ,$$

da die Multiplikation mit τ eine bijektive Abbildung von H nach H ist. Offensichtlich ist $S_H \colon K \to K^H$ die Spurabbildung der GALOISschen Erweiterung K/K^H. Insbesondere ist daher, wie wir in §2 gleich nach der Definition der Spur gesehen haben, S_H nicht die Nullabbildung; es gibt also Elemente $x \in K$, für die $S_H(x)$ nicht verschwindet. Außerdem ist S_H eine K^H-lineare Abbildung, denn für $x \in K$ und $y \in K^H$ ist

$$S_H(xy) = \sum_{\sigma \in H} \sigma(xy) = \sum_{\sigma \in H} \sigma(x)\sigma(y) = \sum_{\sigma \in H} \sigma(x)y = yS_H(x) \, ,$$

und als Spurabbildung ist S_H natürlich ein Homomorphismus bezüglich der Addition. Schließlich ist S_H surjektiv, denn ist $x \in K$ ein Element mit $S_H(x) \neq 0$, so ist für jedes $y \in K^H$

$$S_H\left(\frac{y}{S_H(x)}x\right) = \frac{y}{S_H(x)} S_H(x) = y \, .$$

Daraus wiederum folgt

Lemma: K/k sei eine GALOISsche Erweiterung mit GALOIS-Gruppe G, und H sei eine Untergruppe von G. Weiter sei $\{x_1, \ldots, x_n\}$ ein Erzeugendensystem von K als k-Vektorraum. Dann erzeugen die Elemente $S_H(x_i)$ den Fixkörper K^H als k-Vektorraum.

Beweis: Wegen der Surjektivität von S_H gibt es zu jedem $y \in K^H$ ein $x \in K$ mit $S_H(x) = y$. Als Element des k-Vektorraums K läßt sich x schreiben als $x = \sum_{i=1}^n c_i x_i$ mit $c_i \in k$. Als K^H-lineare Abbildung ist S_H erst recht k-linear; somit ist

$$y = S_H(x) = S_H\left(\sum_{i=1}^n c_i x_i\right) = \sum_{i=1}^n c_i S_H(x_i)\,.$$

Damit läßt sich jedes Element von K^H als k-Linearkombination der Elemente $S_H(x_i)$ darstellen, was die Behauptung beweist. ∎

Für die Körpererweiterung k_n/k bilden die n-ten Einheitswurzeln ein solches Erzeugendensystem, aber für $n \neq 1$ natürlich keine Basis, denn es gibt ja n verschiedene n-te Einheitswurzeln, aber der Grad von k_n/k ist höchstens gleich $\varphi(n)$. In der Tat ist für jede n-te Einheitswurzel $\zeta \neq 1$ beispielsweise

$$1 + \zeta + \zeta^2 + \cdots + \zeta^{n-1} = 0\,,$$

denn multipliziert man diese Summe mit ζ, so erhält man bis auf die Reihenfolge der Summanden die gleiche Summe, d.h. das Produkt dieser Summe mit $1 - \zeta$ ist Null. Im Falle $\zeta \neq 1$ folgt daraus, daß die Summe verschwinden muß. Weitere lineare Abhängigkeiten können sich etwa auch dadurch ergeben, daß außer der Eins noch weitere Einheitswurzeln bereits in k liegen.

Die GALOIS-Gruppe von k_n/k ist eine Untergruppe von $\left(\mathbb{Z}/n\right)^\times$. Ist etwa $k = \mathbb{Q}$ und $n = 5$, so ist $\mathbb{Z}/5 = \mathbb{F}_5$ ein Körper; die prime Restklassengruppe ist also zyklisch von der Ordnung vier. Sie wird erzeugt von der Zwei, denn $2^2 = 4$, $2^3 = 3$ und $2^4 = 1$. Die GALOIS-Gruppe ist wegen der Irreduzibilität von Φ_5 über $\mathbb{Q}$ die gesamte Gruppe $\left(\mathbb{Z}/5\right)^\times$. Schreiben wir $k_5 = \mathbb{Q}(\zeta)$ mit einer primitiven fünften Einheitswurzel ζ, entspricht das Element $j \in \left(\mathbb{Z}/5\right)^\times$ dem Automorphismus von $k_5/\mathbb{Q}$,

der ζ auf ζ^j abbildet. Die einzige echte Untergruppe H der GALOIS-Gruppe hat die Ordnung zwei und wird erzeugt vom Quadrat dieses Automorphismus, also der Abbildung, die ζ auf $\zeta^4 = \zeta^{-1}$ abbildet. Sie bildet auch jede Potenz ζ^r ab auf ζ^{-r}; der Zwischenkörper L vom Grad zwei wird also erzeugt von den Elementen

$$S_H(\zeta^r) = \zeta^r + \zeta^{-r} \quad \text{für } r = 0, \ldots, 4 \,.$$

Man beachte, daß diese Zahlen allesamt reell sind, denn für jede komplexe Zahl z vom Betrag eins ist $z\bar{z} = |z|^2 = 1$. Somit ist $z + z^{-1} = z + \bar{z} = 2\,\mathfrak{Re}\, z$. Tatsächlich erhalten wir nur drei verschiedene Werte, denn wegen $\zeta^5 = 1$ ist $S_H(\zeta^3) = S_H(\zeta^2)$ und $S_H(\zeta^4) = S_H(\zeta)$. Somit wird L als $\mathbb{Q}$-Vektorraum erzeugt von den drei Zahlen $1, \zeta + \zeta^{-1}$ und $\zeta^2 + \zeta^{-2}$. Da $L/\mathbb{Q}$ nur den Grad zwei hat, können diese allerdings nicht linear unabhängig über $\mathbb{Q}$ sein, und in der Tat folgt aus dem Verschwinden von $1 + \zeta + \zeta^2 + \zeta^3 + \zeta^4$ nach Division durch ζ^2, daß

$$\zeta^{-2} + \zeta^{-1} + 1 + \zeta + \zeta^2 = 1 + \left(\zeta + \zeta^{-1}\right) + \left(\zeta^2 + \zeta^{-2}\right) = 0$$

ist. Somit hat L als $\mathbb{Q}$-Vektorraum beispielsweise die Basis bestehend aus der Eins und aus $\zeta + \zeta^{-1}$.

Da $L/\mathbb{Q}$ eine quadratische Körpererweiterung ist, muß $\zeta + \zeta^{-1}$ einer quadratischen Gleichung über $\mathbb{Q}$ genügen. $\left(\zeta + \zeta^{-1}\right)^2 = \zeta^2 + \zeta^{-2} + 2$, und wie wir gerade gesehen haben, ist $\left(\zeta^2 + \zeta^{-2}\right) + 1 = -\left(\zeta + \zeta^{-1}\right)$. Somit ist

$$(\zeta + \zeta^{-1})^2 + (\zeta + \zeta^{-1}) = 1 \,.$$

$\zeta + \zeta^{-1}$ ist also eine Nullstelle des Polynoms $X^2 + X - 1$, d.h. einer der beiden Werte $-\frac{1}{2} \pm \frac{1}{2}\sqrt{5}$, Für $\zeta = e^{2\pi i/5}$ haben ζ und $\zeta^{-1} = \bar{\zeta}$ positiven Realteil, so daß $\zeta + \zeta^{-1}$ gleich $\frac{1}{2}\left(\sqrt{5} - 1\right)$ sein muß; für $\zeta = e^{2\cdot 2\pi i/5}$ etwa wäre $\zeta + \zeta^{-1} = -\frac{1}{2}\left(\sqrt{5} + 1\right)$. In beiden Fällen ist $L = \mathbb{Q}\left(\zeta + \zeta^{-1}\right) = \mathbb{Q}\left(\sqrt{5}\right)$.

k_5/L ist ebenfalls eine quadratische Erweiterung. Wenn wir von der Einheitswurzel $\zeta = e^{2\pi i/5}$ ausgehen, ist sie wegen $\zeta + \zeta^{-1} = \frac{1}{2}\left(\sqrt{5} - 1\right)$ gegeben durch die quadratische Gleichung $\zeta^2 - \frac{1}{2}\left(\sqrt{5} - 1\right)\zeta + 1 = 0$. Deren Lösungen sind

$$\frac{1 - \sqrt{5}}{4} \pm i\frac{\sqrt{2}\,\sqrt{5 + \sqrt{5}}}{4} \,,$$

wobei ζ die Lösung mit positivem Imaginärteil ist und ζ^{-1} die mit dem negativen. Der Körper k_5 entsteht also aus L durch Adjunktion von $\sqrt{-2}\cdot\sqrt{5+\sqrt{5}}$.

Für $\zeta = e^{4\pi i/5}$ erhalten wir ζ und ζ^{-1} entsprechend als Lösungen der quadratischen Gleichung $\zeta^2 + \frac{1}{2}\big(\sqrt{5}+1\big)\zeta + 1 = 0$, also

$$\frac{1+\sqrt{5}}{4} \pm i\frac{\sqrt{2}\,\sqrt{5-\sqrt{5}}}{4}\,.$$

k_5 ist also auch gleich $L\Big(\sqrt{-2}\cdot\sqrt{5-\sqrt{5}}\Big)$.

Zusammenfassend können wir sagen, daß wir für alle primitiven fünften Einheitswurzeln sowohl den Realteil als auch den Imaginärteil explizit durch Wurzelausdrücke dargestellt haben. Da wir mit Quadratwurzeln ausgekommen sind, können wir an diesen Ausdrücken auch leicht eine Vorschrift zur Konstruktion des regelmäßigen Fünfecks ablesen.

Für $n = 7$ wird die Situation schon deutlich komplexer: Die GALOIS-Gruppe von $k_7/k = \mathbb{Q}$ ist wieder wegen der Irreduzibilität von Φ_7 über $\mathbb{Q}$ die gesamte prime Restklassengruppe. $\big(\mathbb{Z}/7\big)^\times$ wird aber nicht von der Zwei erzeugt, denn $2^3 \equiv 1 \bmod 7$. Die Drei ist allerdings eine primitive Wurzel modulo sieben, denn in $\big(\mathbb{Z}/7\big)^\times$ ist $3^2 = 2$, $3^3 = 6$, $3^4 = 4$, $3^5 = 5$ und $3^6 = 1$. Die GALOIS-Gruppe wird somit erzeugt vom Automorphismus τ, der jede siebte Einheitswurzel ζ auf ζ^3 abbildet, und sie ist isomorph zur zyklischen Gruppe $\mathbb{Z}/6$. Damit gibt es genau zwei nichttriviale Untergruppen, eine der Ordnung drei und eine der Ordnung zwei.

Die Untergruppe H der Ordnung drei wird erzeugt von τ^2 und enthält auch τ^4. Wegen $\tau^2(\zeta) = \zeta^9 = \zeta^2$ und $\tau^4(\zeta) = \tau^2\big(\tau^2(\zeta)\big) = \tau^2(\zeta^2) = \zeta^4$ ist $S_H(\zeta) = \zeta + \zeta^2 + \zeta^4$. Dieses Element liegt im Fixkörper $L = K^H$ von H, aber nicht in $\mathbb{Q}$, denn wäre es gleich einer rationalen Zahl q, so wäre ζ Nullstelle des Polynoms $X^4 + X^2 + X - q \in \mathbb{Q}[X]$, im Widerspruch zur Irreduzibilität von Φ_7 über $\mathbb{Q}$. Daher bildet $\zeta + \zeta^2 + \zeta^4$ zusammen mit der Eins eine $\mathbb{Q}$-Basis von L und muß Nullstelle eines quadratischen Polynoms aus $\mathbb{Q}[X]$ sein. Dieses Polynom hat auch $\tau\big(\zeta + \zeta^2 + \zeta^4\big)$ als

Lösung, ist also nach VIÈTE

$$\left(X-\zeta-\zeta^2-\zeta^4\right)\left(X-\zeta^3-\zeta^5-\zeta^6\right)=X^2-aX+b$$

mit $a=(\zeta+\zeta^2+\zeta^4)+(\zeta^3+\zeta^5+\zeta^6)=-1$ und

$$b=(\zeta+\zeta^2+\zeta^4)(\zeta^3+\zeta^5+\zeta^6)=\zeta^4+\zeta^6+1+\zeta^5+1+\zeta+1+\zeta^2+\zeta^3=2\,.$$

Somit ist $\zeta+\zeta^2+\zeta^4$ eine Nullstelle des Polynoms X^2+X+2, also eine der beiden Zahlen $-\frac{1}{2}\pm\frac{1}{2}\sqrt{-7}$. Insbesondere ist $L=K^H=\mathbb{Q}\left(\sqrt{-7}\right)$,

Für jede primitive siebte Einheitswurzel ζ ist

$$\zeta^2+\left(\zeta^2\right)^2+\left(\zeta^2\right)^4=\zeta^2+\zeta^4+\zeta$$

und

$$\zeta^4+\left(\zeta^4\right)^2+\left(\zeta^4\right)^4=\zeta^4+\zeta+\zeta^2\,,$$

so daß $S_H(\zeta)=S_H(\zeta^2)=S_H(\zeta^4)$ ist. Dieser gemeinsame Wert ist eine der beiden Zahlen $-\frac{1}{2}\pm\frac{1}{2}\sqrt{-7}$, die andere ist gleich $S_H(\zeta^3)$, $S_H(\zeta^5)$ und $S_H(\zeta^6)$. Man beachte, daß ζ^3,ζ^5 und ζ^6 die Inversen von ζ^4,ζ^2 und ζ sind und damit konjugiert komplex zu diesen Zahlen.

Für $\zeta=e^{2\pi i/7}$ ist der Imaginärteil von $\zeta+\zeta^2+\zeta^4$ gleich

$$\sin(2\pi/7)+\sin(4\pi/7)+\sin(8\pi/7)\approx 1{,}322875656\approx\tfrac{1}{2}\sqrt{7}\,,$$

so daß $S_H(\zeta)=-\frac{1}{2}+\frac{1}{2}\sqrt{-7}$ ist. ζ ist daher eine Nullstelle des Polynoms $X^4+X^2+X+\frac{1}{2}-\frac{1}{2}\sqrt{-7}$ über $\mathbb{Q}\left(\sqrt{-7}\right)$. Dieses Polynom kann allerdings unmöglich irreduzibel über $\mathbb{Q}\left(\sqrt{-7}\right)$ sein, denn $\left[\mathbb{Q}\left(\sqrt{-7}\right):\mathbb{Q}\right]=2$ und $\left[k_7:\mathbb{Q}\right]=6$, so daß $\left[k_7:\mathbb{Q}\left(\sqrt{-7}\right)\right]=3$ ist.

Die Computeralgebra kennt Algorithmen, mit denen man auch Polynome über endlichen Erweiterungen von $\mathbb{Q}$ in ihre irreduziblen Faktoren zerlegen kann; für obiges Polynom liefern sie die beiden Faktoren

$$X-\frac{1}{2}+\frac{\sqrt{-7}}{2}\quad\text{und}\quad X^3+\frac{1-\sqrt{-7}}{2}X^2-\frac{1+\sqrt{-7}}{2}X-1\,.$$

Der erste hat die Nullstelle $\frac{1}{2}-\frac{1}{2}\sqrt{-7}$, der zweite hat ζ,ζ^2 und ζ^4 als Nullstellen. $\mathbb{Q}(\zeta)$ ist somit der Zerfällungskörper dieses kubischen

Polynoms über $\mathbb{Q}\big(\sqrt{-7}\big)$. Anwendung der Formel von CARDANO führt beispielsweise auf die Nullstelle

$$\frac{\sqrt[3]{56+12\sqrt{21}-4\sqrt{-7}}}{6}+\frac{2\sqrt{-7}}{3\sqrt[3]{56+12\sqrt{21}-4\sqrt{-7}}}-\frac{1-\sqrt{-7}}{6}\,,$$

von der die numerische Auswertung (nach Regeln von Maple für die Werte von Kubikwurzeln komplexer Zahlen) zeigt, daß sie gleich $e^{2\pi i/7}$ ist.

Real- und Imaginärteil kann man aus dieser Darstellung nur schlecht ablesen, höchstens über die allgemeinen Formeln

$$\mathfrak{Re}\,z=\frac{z+\bar z}{2}\quad\text{und}\quad\mathfrak{Im}\,z=\frac{z-\bar z}{2i}\,.$$

In einem beliebigen Erweiterungskörper $K/\mathbb{Q}$ müssen allerdings zu einem Element $z\in K$ weder Realteil noch Imaginärteil in K liegen, denn weder muß $\bar z$ in K liegen noch i. In unserem Fall, für $K=k_7$, liegt aber zu jedem Element z auch $\bar z$ in K, denn τ^3 bildet ζ aus auf $\zeta^{27}=\zeta^{-1}$. Damit wird jede siebte Einheitswurzel auf ihr Inverses abgebildet, das komplex konjugiert zu ihr ist. Da die siebten Einheitswurzeln (ohne die Eins) eine $\mathbb{Q}$-Vektorraumbasis von k_7 bilden und τ^3 auf dieser Basis mit der komplexen Konjugation übereinstimmt, ist τ^3 auf k_7 die komplexe Konjugation. Damit liegt für jedes $z\in k_7$ auch der Realteil in k_7. Ein nichtverschwinder Imaginärteil kann aber nicht in k_7 liegen, denn sonst wäre $i\in\mathbb{Q}(\zeta)$, und $\mathbb{Q}(i)$ wäre ein Teilkörper vom Grad zwei. Da die GALOIS-Gruppe nur eine Untergruppe der Ordnung drei hat, gibt es aber nur einen Teilkörper vom Grad zwei, und das ist $L=\mathbb{Q}\big(\sqrt{-7}\big)$.

Die Gleichung

$$\zeta^4+\zeta^2+\zeta+\frac12\pm\frac12\sqrt{-7}=0$$

zeigt, daß für $x=\mathfrak{Re}\,\zeta=\cos\alpha$ gilt

$$\cos 4\alpha+\cos 2\alpha+\cos\alpha+\tfrac12=0\,.$$

Da ζ eine siebte Einheitswurzel ist, sind ζ^4 und ζ^3 invers zueinander, also komplex konjugiert, und haben daher denselben Realteil. Somit ist $\cos 4\alpha=\cos 3\alpha$ und damit auch

$$\cos 3\alpha+\cos 2\alpha+\cos\alpha+\tfrac12=0\,.$$

Aus der Gleichung

$$(\cos\alpha + i\sin\alpha)^2 = \left(e^{i\alpha}\right)^2 = e^{i\cdot 2\alpha} = \cos 2\alpha + i\sin 2\alpha$$

folgt, daß

$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha = \cos^2\alpha - (1-\cos^2\alpha) = 2\cos^2\alpha - 1$$

ist. Entsprechend führt die Gleichung

$$(\cos\alpha + i\sin\alpha)^3 = \left(e^{i\alpha}\right)^3 = e^{i\cdot 3\alpha} = \cos 3\alpha + i\sin 3\alpha$$

auf die Formel

$$\begin{aligned}\cos 3\alpha &= \cos^3\alpha - 3\cos\alpha\,\sin^2\alpha = \cos^3\alpha - 3\cos\alpha\,(1-\cos^2\alpha)\\ &= 4\cos^3\alpha - 3\cos\alpha\,.\end{aligned}$$

Speziell für $\alpha = 2\pi/7$ genügt $x = \cos\alpha$ somit der Gleichung

$$(4x^3 - 3x) + (2x^2 - 1) + x + \frac{1}{2} = 4x^3 + 2x^2 - 2x - \frac{1}{2} = 0\,.$$

Diese kubische Gleichung hat die drei verschiedenen Lösungen $\cos\alpha$, $\cos 2\alpha$ und $\cos 3\alpha$; wir sind also im *casus irreducibilis* mit drei verschiedenen reellen Lösungen, die bei Anwendung der Formel von CARDANO aber als Ausdrücke mit Wurzeln aus nichtreellen Zahlen erscheinen. Eine der Lösungen ist beispielsweise

$$x = \frac{\sqrt[3]{28 + 84\sqrt{-3}}}{28} + \frac{7}{3\sqrt[3]{28 + 84\sqrt{-3}}} - \frac{1}{6}\,,$$

was Maple numerisch mit $\cos\alpha$ identifiziert.

Es gibt auch einen Zwischenkörper L' vom Grad drei über $\mathbb{Q}$; er ist der Fixkörper von k_7 bezüglich der Untergruppe $H = \{1, \tau^3\}$. Wie wir gerade gesehen haben, ist der Automorphismus τ^3 die komplexe Konjugation, bildet also ζ ab auf $\bar{\zeta} = \zeta^{-1}$. Somit ist $S_{H'}(\zeta) = \zeta + \zeta^{-1}$.

Der Zwischenkörper L' wird als $\mathbb{Q}$-Vektorraum nach dem obigen Lemma erzeugt von den Elementen $S_{H'}(\zeta^j)$, also von den drei Zahlen $S_{H'}(\zeta) = \zeta + \zeta^{-1} = \zeta + \zeta^6$, $S_{H'}(\zeta^2) = \zeta^2 + \zeta^{-2} = \zeta^2 + \zeta^5$ und

$S_{H'}(\zeta^3) = \zeta^3 + \zeta^{-2} = \zeta^3 + \zeta^4$. Das Polynom $X^3 + aX^2 + bX + c$ mit diesen drei Nullstellen hat nach dem Satz von VIÈTE die Koeffizieten

$$\begin{aligned}
a &= -(\zeta + \zeta^6 + \zeta^2 + \zeta^5 + \zeta^3 + \zeta^4) = 1 \\
b &= (\zeta + \zeta^6)(\zeta^2 + \zeta^5) + (\zeta + \zeta^6)(\zeta^3 + \zeta^4) + (\zeta^2 + \zeta^5)(\zeta^3 + \zeta^4) \\
&= \zeta^{11} + \zeta^{10} + 2\zeta^9 + 2\zeta^8 + 2\zeta^6 + 2\zeta^5 + \zeta^4 + \zeta^3 \\
&= 2\zeta^6 + 2\zeta^5 + 2\zeta^4 + 2\zeta^3 + 2\zeta^2 + 2\zeta = -2 \\
c &= -(\zeta + \zeta^6)(\zeta^2 + \zeta^5)(\zeta^3 + \zeta^4) \\
&= -(\zeta^{15} + \zeta^{14} + \zeta^{12} + \zeta^{11} + \zeta^{10} + \zeta^9 + \zeta^7 + \zeta^6) \\
&= -(\zeta^6 + \zeta^5 + \zeta^4 + \zeta^3 + \zeta^2 + \zeta + 2) = -1\,.
\end{aligned}$$

$\zeta+\zeta^{-1}$ ist also eine Nullstelle des kubischen Polynoms X^3+X^2-2X-1. Auch hier haben wir drei verschiedene reelle Nullstellen, denn

$$\zeta^j + \zeta^{-j} = \zeta^j + \overline{\zeta^j} = 2\,\mathfrak{Re}\,\zeta^j = 2\cos j\alpha\,.$$

Wieder sind wir also im *casus irreducibilis* mit drei reellen Lösungen, die aber in der Lösungsformel von CARDANO als Kombinationen komplizierter komplexer Wurzelausdrücke dargestellt werden. In der Tat haben wir fast die gleiche kubische Gleichung wie eben, denn ist x eine Nullstelle von $X^3 + X^2 - 2X - 1$, so ist

$$4\left(\frac{x}{2}\right)^3 + 2\left(\frac{x}{2}\right)^2 - 2\frac{x}{2} - \frac{1}{2} = \frac{x^3 + x^2 - 2x - 1}{2} = 0\,,$$

d.h. $\frac{x}{2}$ ist eine Nullstelle von $4X^3 + 2X^2 - 2X - \frac{1}{2}$. Der Körper L' ist somit der Zerfällungskörper über $\mathbb{Q}$ für jedes der beiden Polynome $X^3 + X^2 - 2X - 1$ und $4X^3 + 2X^2 - 2X - \frac{1}{2}$.

$\mathbb{Q}(\zeta)/L'$ ist eine quadratische Erweiterung; da

$$\zeta + \zeta^{-1} = u$$

eine Nullstelle des Polynoms $X^3 + X^2 - 2X - 1$ ist, können wir wie beim Fall der fünften Einheitswurzeln vorgehen und sehen nach Multiplikation obiger Gleichung mit ζ, daß $\mathbb{Q}(\zeta)$ der Zerfällungskörper des Polynoms $X^2 - uX + 1$ über L' ist. Damit sind alle Zwischenkörper mehr oder weniger explizit bestimmt.

Kapitel 5
Die Fermat-Vermutung für Zahlen und für Polynome

§1: Zahlen und Funktionen

Wir haben im Verlauf dieser Vorlesung zweimal einen Satz über eindeutige Primzerlegung bewiesen: Zuerst für den Ring $\mathbb{Z}$ der ganzen Zahlen, und später für Polynomringe über Körpern (oder allgemeiner über faktoriellen Ringen). Speziell im Falle von Polynomringen in einer Variablen über einem Körper war der Beweis praktisch identisch zu dem für die ganzen Zahlen; beide Male ging es darum, daß der Ring EUKLIDisch ist.

Die Rolle der Primzahlen spielten im Polynomring die irreduziblen Polynome. Im Falle eines algebraisch abgeschlossenen Körpers k sind das gerade die linearen Polynome, und die bilden im Gegensatz zu den Primzahlen eine sehr übersichtliche Menge. Da es auf konstante Faktoren nicht ankommt, kann man sich auf Polynome der Form $X - a$ mit $a \in k$ beschränken. Die Menge aller dieser Polynome wiederum kann identifiziert werden mit der Menge aller $a \in k$, deren Elemente man mit den Punkten einer Geraden identifizieren kann, so daß in einigen Anwendungen auch geometrische Argumente möglich sind.

Natürlich gibt es – zum Teil beträchtliche – Unterschiede zwischen $\mathbb{Z}$ und dem Polynomring über einem Körper, aber gerade das macht die Analogie so interessant: Da es für jeden der beiden Ringe ein eigenes Instrumentarium gibt, kann man versuchen die damit bewiesenen Resultate auf den jeweils anderen Fall zu übertragen, was idealerweise zu neuen Sätzen und sonst zumindest zu interessanten Vermutungen führt.

Als Beispiel für Parallelen und Unterschiede zwischen den beiden Situationen wollen wir die FERMAT-Vermutung betrachten. FERMAT schrieb bekanntlich um 1637 an den Rand seiner Arithmetik des DIOPHANTOS von Alexandrien, daß die Gleichung

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

für $n \geq 3$ keine Lösung in ganzen Zahlen habe – außer natürlich den trivialen Lösungen, bei denen eine der drei Zahlen verschwindet. (Die französische Übersetzung der Arithmetik, die er dabei benutzte, stammt übrigens von BACHET DE MÉZIRIAC, denn wir als Entdecker des erweiterten EUKLIDischen Algorithmus kennen. Bekannt wurde FERMATs Randbemerkung erst, als sein Sohn CLÉMENT-SAMUEL DE FERMAT 1670 die Arithmetik mit den Randbemerkungen seines fünf Jahre zuvor gestorbenen Vaters veröffentlichte.)

Es ist nicht möglich, einen Kubus in zwei Kuben oder ein Biquadrat in zwei Biquadrate und ganz allgemein irgendeine der unendlich vielen Potenzen jenseits des Quadrats in zwei ebensolche zu teilen. Ich habe einen wunderbaren Beweis hierfür gefunden, aber der Rand ist zu schmal, um ihn zu fassen.

Die direkte Verallgemeinerung auf Polynomringe ist sicherlich falsch: Die Gleichung $f^n + g^n = h^n$ ist zumindest für *konstante* Polynome über einem algebraisch abgeschlossenen Körper immer lösbar: Für beliebig vorgegebene Konstanten $f, g \in k$ muß man einfach $h = \sqrt[n]{f^n + g^n}$ setzen. Das sind allerdings, wenn wir uns wirklich für Polynome interessieren, uninteressante Lösungen, vergleichbar den Lösungen $x^n + 0^n = x^n$ der klassischen FERMAT-Gleichung.

Auch wenn wir verlangen, daß die Grade aller beteiligter Polynome positiv sein sollen, gibt es triviale Lösungen: Ist f irgendein beliebiges Polynom und sind $a, b, c \in k$ so, daß gilt $a^n + b^n = c^n$, ist natürlich auch $(af)^n + (bf)^n = (cf)^n$. Was wir höchstens erwarten können ist also das folgende Analogon zur klassischen FERMAT-Vermutung:

Für $n \geq 3$ gibt es keine teilerfremden Polynome f, g, h mit positivem Grad, so daß $f^n + g^n = h^n$ ist.

(Normalerweise unterscheiden wir sorgfältig zwischen paarweise teilerfremden Polynomen und solchen, die nur insgesamt keinen gemeinsamen Teiler haben. Hier sind beide Begriffe äquivalent, da jeder gemeinsame Teiler zweier der Polynome f, g, h wegen $f^n + g^n = h^n$ auch das dritte teilen muß.)

Für Körper positiver Charakteristik ist selbst das noch falsch: Über einen Körper der Charakteristik p ist schließlich $f^p+g^p = (f+g)^p$ für beliebige Polynome f und g, und dasselbe gilt auch wenn man den Exponenten p durch eine seiner Potenzen ersetzt. Wir können also höchstens für Körper der Charakteristik null erwarten, daß diese Vermutung für alle Exponenten $n \geq 3$ richtig ist, und genau das werden wir im nächsten Paragraphen zumindest für den Körper der komplexen Zahlen und damit auch jeden Teilkörper davon beweisen.

§2: Der Satz von Mason

Wir wollen zeigen, daß es für $n \geq 3$ keine zueinander teilerfremden Polynome positiven Grades $f, g, h \in \mathbb{C}[X]$ gibt mit $f^n + g^n = h^n$.

Der *Beweis* beruht darauf, daß die Polynome f^n und g^n dieselben Nullstellen wie f und g haben, aber mit n-facher Vielfachheit. Ist $f^n + g^n = h^n$, so hat auch die Summe dieser beiden Potenzen im Vergleich zum Grad relativ wenige Nullstellen, diese aber mit mindestens n-facher Vielfachheit. Nach einem 1983 von R.C. MASON bewiesenen Satz können in einer solchen Situation aber f^n, g^n und h^n nicht zu wenige verschiedene Nullstellen haben:

Satz: Bezeichnet $n_0(f)$ die Anzahl verschiedener (komplexer) Nullstellen eines Polynoms f, so gilt für drei nichtkonstante, teilerfremde Polynome $f, g, h \in \mathbb{C}[X]$ mit $f + g = h$

$$n_0(fgh) \geq \max(\deg f, \deg g, \deg h) + 1 \, .$$

Bevor wir diesen Satz beweisen, wollen wir uns zunächst überlegen, daß daraus wirklich das Analogon der FERMAT-Vermutung für Polynome folgt:

Für drei nichtkonstante teilerfremde Polynome f, g, h mit $f^n + g^n = h^n$ ist nach dem Satz von MASON

$$\begin{aligned} n_0\big(f^n g^n h^n\big) &\geq \max\big(\deg f^n, \deg g^n, \deg h^n\big) + 1 \\ &= n \max\big(\deg f, \deg g, \deg h\big) + 1 \, . \end{aligned}$$

Andererseits ist aber

$$\begin{aligned} n_0\big(f^n g^n h^n\big) = n_0\big(fgh\big) &\le \deg f + \deg g + \deg h \\ &\le 3\max(\deg f, \deg g, \deg h)\,, \end{aligned}$$

denn die Anzahl *verschiedener* Nullstellen einer Potenz eines Polynoms ist gleich der Anzahl verschiedener Nullstellen des Polynoms selbst, und die Nullstellenanzahl eines Polynom kann nicht größer sein als der Grad.

Damit haben wir insgesamt die Ungleichung

$$\begin{aligned} 3\max\big(\deg f, \deg g, \deg h\big) &\ge n_0\big(f^n g^n h^n\big) \\ &\ge n\max\big(\deg f, \deg g, \deg h\big) + 1\,, \end{aligned}$$

die nur für $n \le 2$ gelten kann. Somit gibt es für $n \ge 3$ keine nichtkonstanten teilerfremden Polynome, für die $f^n + g^n = h^n$ ist.

Zu einem vollständigen Beweis der FERMAT-Vermutung für Polynome fehlt nun nur noch der Beweis des Satzes von MASON. Die Idee dazu ist folgende: Ist $f + g = h$, so betrachten wir den Quotienten g/f im rationalen Funktionenkörper $\mathbb{C}(X)$. Da f und g teilerfremd sind, ist das ein gekürzter Bruch. Falls wir diesen auch in der Form $g/f = G/F$ schreiben können mit Polynomen F, G vom Grad höchstens $n_0(fgh)-1$, haben auch Zähler und Nenner f und g des gekürzten Bruchs höchstens den Grad $n_0(fgh) - 1$. Wegen $f + g = h$ gilt dasselbe auch für h, und damit ist $n_0(fgh) - 1 \le \max(\deg f, \deg g, \deg h)$, was zur Aussage des Satzes äquivalent ist.

Um g/f als Quotienten zweier neuer Polynome auszudrücken, schreiben wir zunächst

$$\frac{g}{f} = \frac{S}{R} \quad \text{mit} \quad R = \frac{f}{h} \;\text{ und }\; S = \frac{g}{h}\,.$$

Wegen $f + g = h$ ist $R + S = 1$, die Summe $R' + S'$ der Ableitungen verschwindet also. Dies können wir etwas umschreiben

$$R' + S' = \frac{R'}{R}R + \frac{S'}{S}S = 0 \Longrightarrow \frac{R'}{R}R = -\frac{S'}{S}S \Longrightarrow \frac{R}{S} = -\frac{R'}{R}\Big/\frac{S'}{S}\,,$$

und damit erhalten wir die neue Darstellung

$$\frac{g}{f} = \frac{S}{R} = -\frac{R'/R}{S'/S}\,.$$

Rechts stehen die logarithmischen Ableitungen von R und S im Zähler und Nenner, und damit lassen sich gut die Nullstellen von f, g und h ins Spiel bringen: Nach der LEIBNIZ-Regel ist bekanntlich

$$(uv)' = u'v + uv', \quad \text{also} \quad \frac{(uv)'}{uv} = \frac{u'}{u} + \frac{v'}{v}\,,$$

die logarithmische Ableitung eines Produkts ist also einfach die Summe der logarithmischen Ableitungen der Faktoren. Daraus folgt sofort, daß die logarithmische Ableitung eines Quotienten gleich der Differenz aus logarithmischer Ableitung des Zählers und logarithmischer Ableitung des Nenners ist. Schreiben wir

$$f = f_0 \prod_{i=1}^{r}(x-a_i)^{n_i}, \quad g = g_0 \prod_{j=1}^{s}(x-b_j)^{m_j} \quad \text{und} \quad h = h_0 \prod_{k=1}^{t}(x-c_k)^{p_k}$$

mit $f_0, g_0, h_0 \in \mathbb{C}^\times$, ist also wegen $R = \frac{f}{h}$ und $S = \frac{g}{h}$

$$\frac{R'}{R} = \frac{f'}{f} - \frac{h'}{h} = \sum_{i=1}^{r} \frac{n_i}{x-a_i} - \sum_{k=1}^{t} \frac{p_k}{x-c_k}\,,$$

$$\frac{S'}{S} = \frac{g'}{g} - \frac{h'}{h} = \sum_{j=1}^{s} \frac{m_j}{x-b_j} - \sum_{k=1}^{t} \frac{p_k}{x-c_k}$$

$$\text{und} \quad \frac{g}{f} = -\frac{R'/R}{S'/S} = -\frac{\displaystyle\sum_{i=1}^{r} \frac{n_i}{x-a_i} - \sum_{k=1}^{t} \frac{p_k}{x-c_k}}{\displaystyle\sum_{j=1}^{s} \frac{m_j}{x-b_j} - \sum_{k=1}^{t} \frac{p_k}{x-c_k}}\,.$$

Erweitern wir Zähler und Nenner mit dem Hauptnenner aller Summanden, d.h. mit dem Polynom vom Grad $r + s + t = n_0(fgh)$

$$H = \prod_{i=1}^{r}(x-a_i) \cdot \prod_{j=1}^{s}(x-b_j) \cdot \prod_{k=1}^{t}(x-c_k)\,,$$

so erhalten wir im Zähler wie auch im Nenner Summen von Polynomen vom Grad $n_0(fgh)-1$, also Polynome vom Grad höchstens $n_0(fgh)-1$, wie gewünscht. Damit sind sowohl der Satz von MASON als auch das Analogon der FERMATschen Vermutung für Polynome bewiesen.

§3: Die abc-Vermutung

Der Erfolg des Satzes von MASON beim Beweis der FERMAT-Vermutung für Polynome legt es nahe, etwas Ähnliches auch im klassischen Fall zu versuchen.

Da natürliche Zahlen weder Grade noch Nullstellen haben, müssen wir dazu den Satz von MASON zunächst einmal so umformulieren, daß wir eine Aussage bekommen, die ein sinnvolles Analogon für natürliche Zahlen hat.

Dazu ordnen wir einem Polynom f anstelle der Anzahl $n_0(f)$ seiner (verschiedenen) Nullstellen ein *Polynom* $N_0(f)$ dazu, das genau diese Nullstellen mit jeweils der Vielfachheit eins haben soll: Für

$$f = f_0 \prod_{i=1}^{r} (x-a_i)^{n_i} \quad \text{sei} \quad N_0(f) \underset{\text{def}}{=} \prod_{i=1}^{r} (x-a_i)\,,$$

so daß der Grad von $N_0(f)$ gerade die im vorigen Paragraphen definierte Zahl $n_0(f)$ ist.

Der Vorteil des Polynoms $N_0(f)$ besteht darin, daß wir eine analoge Definition leicht auch für natürliche Zahlen hinschreiben können: Für

$$n = \prod_{i=1}^{r} p_i^{e_i} \quad \text{setzen wir} \quad N_0(n) \underset{\text{def}}{=} \prod_{i=1}^{r} p_i\,.$$

Mit Hilfe der Polynome $N_0(f)$ läßt sich der Satz von MASON folgendermaßen umformulieren:

Gilt für drei teilerfremde Polynome f, g und h die Gleichung $f+g=h$, so hat jedes der drei Polynome einen kleineren Grad als das Polynom $N_0(fgh)$.

In dieser Formulierung kommt immer noch der Grad vor, für den wir bei natürlichen Zahlen keine Verwendung haben. Betrachten wir aber den

Grad (wie bei der Polynomdivision mit Rest) lediglich als eine Methode, einem Polynom eine Zahl aus $\mathbb{N}_0$ zuzuordnen, so können wir, wenn wir bereits natürliche Zahlen haben, einfach ganz auf ihn verzichten; falls wir ganze Zahlen betrachten, liegt es nahe, ihn durch den Betrag zu ersetzen.

Gemäß dieser Philosophie können wir nun probeweise die folgende Aussage formulieren:

A1: *Ist a+b = c für drei zueinander teilerfremde natürliche Zahlen a, b, c, so ist jede der drei Zahlen kleiner als $N_0(abc)$.*

Damit haben wir eine sinnvolle Aussage über natürliche Zahlen gefunden, die – falls sie korrekt ist – sofort die FERMAT-Vermutung impliziert: Gäbe es nämlich drei natürliche Zahlen x, y, z mit der Eigenschaft, daß $x^n + y^n = z^n$ für ein $n \geq 3$, so gäbe es auch drei zueinander teilerfremde Zahlen x, y, z mit dieser Eigenschaft: Wir müssen einfach die drei Zahlen durch ihren größten gemeinsamen Teiler kürzen. Alsdann müßte, falls obige Aussage richtig wäre, jede der drei Potenzen x^n, y^n, z^n kleiner sein als $N_0(x^n y^n z^n)$. Nun ist aber

$$N_0(x^n y^n z^n) = N_0(xyz) \leq xyz \,,$$

d.h. nach **A1** wäre jede der drei Zahlen x^n, y^n, z^n kleiner als xyz. Damit wäre

$$(xyz)^n = x^n y^n z^n < (xyz)^3 \,,$$

was für $n \geq 3$ offensichtlich nicht möglich ist.

Angesichts der Komplexität des WILESschen Beweises fällt es schwer, an einen so einfachen Beweis zu glauben, und in der Tat ist die Aussage **A1** falsch:

Betrachten wir etwa die Gleichung $8 + 1 = 9$. Offensichtlich sind die drei Summanden teilerfremd zueinander, aber sowohl 8 als auch 9 sind größer als $N_0(8 \cdot 1 \cdot 9) = 2 \cdot 3 = 6$. Ganz so einfach geht es also nicht.

Da der Grad eines Polynoms nicht durch konstante Faktoren beeinflußt wird, könnte man versuchen, als „richtiges“ Analogon zum Satz von MASON eine abgeschwächte Aussage zu nehmen, die nur eine Abschätzung bis auf einen konstanten Faktor enthält, etwa

A2: *Ist a+b = c für drei zueinander teilerfremde natürliche Zahlen a, b, c, so gibt es eine Konstante K derart, daß jede der drei Zahlen kleiner ist als $K \cdot N_0(abc)$.*

Diese Aussage ist trivialerweise richtig: Wir müssen nur eine Konstante K wählen, die größer ist als das Maximum von a, b und c. Leider ist sie auch völlig nutzlos, denn solange die Konstante von a, b und c abhängen darf, haben wir keine Chance, damit die FERMAT-Vermutung zu beweisen.

Wir müssen die Aussage also noch einmal umformulieren:

A3: *Es gibt eine Konstante K, für die gilt: Ist a+b = c für drei zueinander teilerfremde natürliche Zahlen a, b, c, so ist jede der drei Zahlen kleiner als $K \cdot N_0(abc)$.*

Auch daraus würde die FERMAT-Vermutung zumindest für alle hinreichend großen Exponenten n folgen, allerdings ist die Aussage, so wie sie dasteht, leider immer noch falsch:

Betrachten wir die Gleichung

$$a_n + b_n = c_n \quad \text{mit} \quad a_n = 3^{2^n} - 1, \quad b_n = 1 \quad \text{und} \quad c_n = 3^{2^n}. \quad (*)$$

Wäre sie richtig, müßte für jedes n gelten:

$$3^{2^n} \le K N_0\big((3^{2^n} - 1) \cdot 3^{2^n}\big) = K \cdot 3 \cdot N_0(3^{2^n} - 1)\,.$$

Um $N_0(3^{2^n} - 1)$ abzuschätzen, beachten wir, daß gilt

$$3^{2^n} = \big(3^{2^{n-1}}\big)^2 \quad \text{und} \quad 3^{2^n} - 1 = \big(3^{2^{n-1}} + 1\big)\big(3^{2^{n-1}} - 1\big)$$

nach der dritten binomischen Formel. Wenden wir dies mehrfach an, erhalten wir

$$\begin{aligned} 3^{2^n} - 1 &= \big(3^{2^{n-1}} + 1\big)\big(3^{2^{n-1}} - 1\big) \\ &= \big(3^{2^{n-1}} + 1\big)\big(3^{2^{n-2}} + 1\big)\big(3^{2^{n-2}} - 1\big) \\ &= \big(3^{2^{n-1}} + 1\big)\big(3^{2^{n-2}} + 1\big)\big(3^{2^{n-3}} + 1\big)\big(3^{2^{n-3}} - 1\big) \\ &= \cdots \\ &= \big(3^{2^{n-1}} + 1\big)\big(3^{2^{n-2}} + 1\big) \ldots \big(3^2 + 1\big)\big(3^1 + 1\big)\big(3^1 - 1\big)\,. \end{aligned}$$

In der letzten Zeile steht ein Produkt aus $n+1$ geraden Zahlen; somit ist $3^{2^n}-1$ durch 2^{n+1} teilbar. Das Produkt $N_0(3^{2^n}-1)$ aller *verschiedener* Primteiler von $3^{2^n}-1$ erfüllt daher die Ungleichung

$$N_0(3^{2^n}-1) \le 2 \cdot \frac{3^{2^n}-1}{2^{n+1}} = \frac{3^{2^n}-1}{2^n} ,$$

denn das Produkt aller ungerader Primteiler kann höchstens gleich $(3^{2^n}-1)/2^n$ sein.

Falls **A3** korrekt wäre, müßte nach Gleichung $(*)$ also gelten

$$3^{2^n} \le \frac{3K}{2^n}(3^{2^n}-1) \quad \text{für alle } n .$$

Das kann aber unmöglich der Fall sein, denn für hinreichend große n ist der Faktor $\frac{3K}{2^n}$ kleiner als eins, so daß 3^{2^n} echt kleiner als sich selbst sein müßte.

Auf der Suche nach einem Analogon für den Satz von MASON müssen wir daher noch weiter abschwächen. *Eine* Möglichkeit dazu ist die 1986 aufgestellte

abc-Vermutung von MASSER und OESTERLÉ: Zu jedem $\varepsilon > 0$ gibt es eine Konstante $K(\varepsilon)$, so daß für alle teilerfremden natürlichen Zahlen a, b, c mit $a+b=c$ gilt: Jede der drei Zahlen a, b, c ist kleiner oder gleich $K(\varepsilon) \cdot N_0(abc)^{1+\varepsilon}$.

Wir wollen uns überlegen, daß sie zumindest für große Exponenten n die FERMAT-Vermutung impliziert.

Dazu betrachten wir eine Lösung $x^n + y^n = z^n$ mit o.B.d.A. teilerfremden natürlichen Zahlen x, y, z und wählen uns irgendein $\varepsilon > 0$. Nach der abc-Vermutung gibt es dazu eine Konstante $K(\varepsilon)$, so daß x^n, y^n und z^n allesamt höchstens gleich

$$K(\varepsilon)N_0(x^n y^n z^n)^{1+\varepsilon} = K(\varepsilon)N_0(xyz)^{1+\varepsilon} \le K(\varepsilon)(xyz)^{1+\varepsilon}$$

sind. Für ihr Produkt gilt daher

$$x^n y^n z^n \le K(\varepsilon)^3 (xyz)^{3(1+\varepsilon)} \quad \text{oder} \quad (xyz)^{n-3-3\varepsilon} \le K(\varepsilon)^3 .$$

$K(\varepsilon)^3$ ist eine feste Zahl; es gibt daher einen Exponenten m derart, daß $2^m > K(\varepsilon)^3$ ist. Da das Produkt xyz auf jeden Fall nicht kleiner als zwei

sein kann, ist $2^{n-3-3\varepsilon} < 2^m$, also $n - 3 - 3\varepsilon < m$ und $n < m + 3 + 3\varepsilon$. Für $n \geq m + 3 + 3\varepsilon$ ist daher

$$(xyz)^{n-3-3\varepsilon} > K(\varepsilon)^3 \,,$$

und damit kann es keine zueinander teilerfremden natürlichen Zahlen x, y, z geben mit $x^n + y^n = z^n$.

Ob und gegebenenfalls welche konkreten Schranken für n man damit erreichen kann, hängt natürlich davon ab, wie $K(\varepsilon)$ von ε abhängt. Nach einigen Spekulationen könnte aus der abc-Vermutung die FERMAT-Vermutung für alle $n \geq 6$ folgen, und für $n = 3, 4, 5$ ist der Satz schon lange bekannt: Für $n = 4$ und möglicherweise auch $n = 3$ hatte FERMAT selbst bereits spätestens um 1640 einen (grob skizzierten) Beweis; den für $n = 4$ arbeitete BERNARD FRÉNICLE DE BESSY aus und veröffentlichte ihn 1676. EULER fand 1753 einen Beweis für den Fall $n = 3$, den er 1770 veröffentlichte. Er arbeitete dazu mit den dritten Einheitswurzeln. Ebenfalls mit Einheitswurzeln bewies ERNST EDUARD KUMMER die FERMAT-Vermutung für alle Exponenten, die durch eine sogenannte reguläre Primzahl teilbar sind, d.h. durch eine Primzahl, für die es im Ring der ganzen Zahlen im Körper $\mathbb{Q}(\zeta_p)$ eine eindeutige Primzerlegung gibt. (Eine Zahl aus einen Körper $K/\mathbb{Q}$ heißt ganz, wenn sie Nullstelle eines *normierten* Polynoms mit ganzzahligen Koeffizienten ist. Im Gegensatz zur Definition einer algebraischen Zahl wird hier also noch verlangt, daß der führende Koeffizient des Polynoms eins ist.) Zu diesen regulären Primzahlen gehört insbesondere auch die Zahl fünf.

Der derzeitige Stand der abc-Vermutung ist innerhalb der Mathematik umstritten. 2012 kündigte SHINICHI MOCHIZUKI vom Research Institute for Mathematical Sciences (RIMS) der Universität Kyoto einen Beweis an, der nach langer kontroverser Diskussion im Februar 2020 von den *Publications of the RIMS* zur Veröffentlichung angenommen wurde. Da MOCHIZUKI in seiner rund sechshundert Seiten langen Arbeit mit vielen, von ihm selbst entwickelten neuen und unkonventionellen Methoden arbeitet, wird der Beweis allerdings zumindest außerhalb Japans von den meisten Experten nicht akzeptiert.

Für weitere Informationen zu §2 und §3 sei auf einen Vortrag verwiesen, den SERGE LANG (1927 – 2005) im Jahr 1992 an der ETH Zürich vor

einem „allgemeinen“ Publikum hielt und dem ich hier im wesentlichen gefolgt bin:

SERGE LANG: Die abc-Vermutung, *Elemente der Mathematik* **48** (1993), 89-99

Der Artikel ist (wie die gesamte Zeitschrift *Elemente der Mathematik)* unter `http://www.bibliothek.uni-regensburg.de/ezeit/?2135837` frei zugänglich.

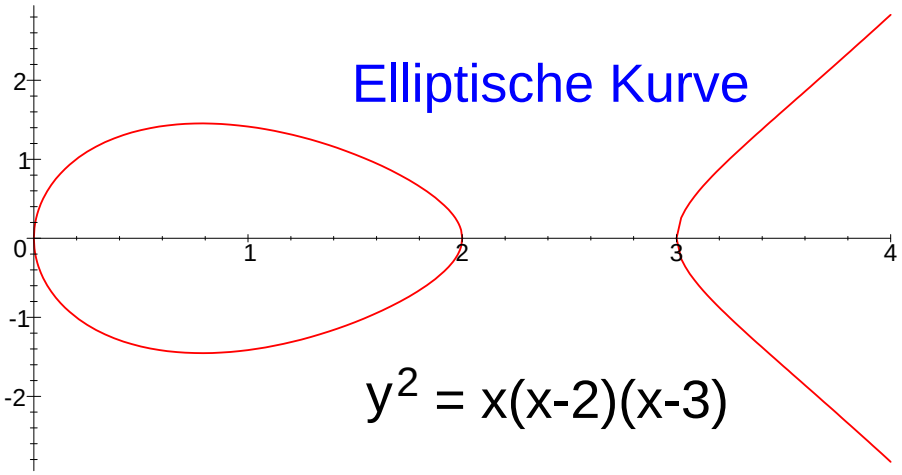
§4: Die Frey-Kurve

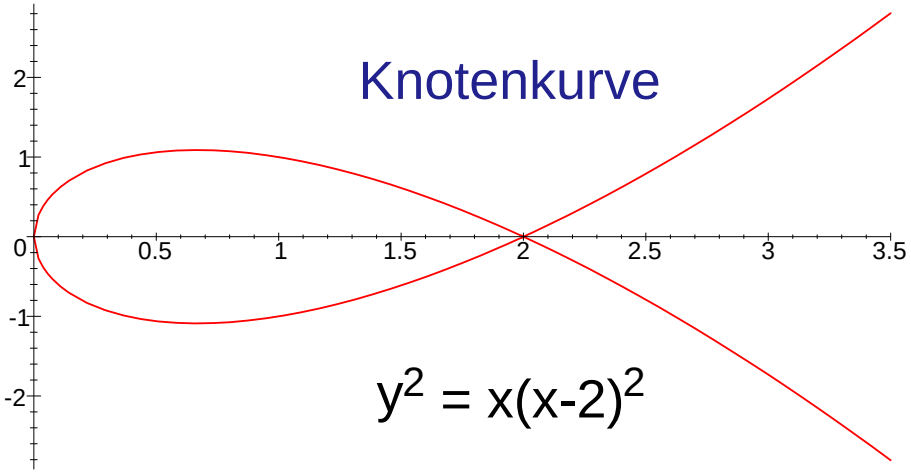
Da die FERMAT-Vermutung seit 1994 bewiesen ist, die abc-Vermutung aber immer noch offen, mußte der Beweis der FERMAT-Vermutung natürlich andere Wege gehen. Die meisten dieser Wege führen in Gebiete, die weit jenseits dessen liegen, was selbst ein guter auf Zahlentheorie spezialisierter Mathematiker im Laufe seines Studiums lernen kann. Zumindest die Grundidee der abc-Vermutung, daß man nämlich Summenbeziehungen zwischen großen Zahlen nicht ohne ein gewisses Minimum an verschiedenen Primfaktoren realisieren kann, spielt in modifizierter Weise in der Tat eine große Rolle.

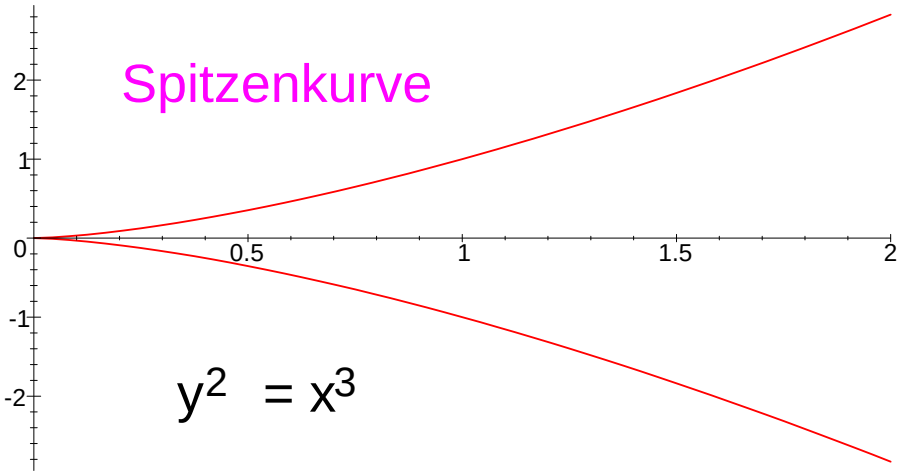
Der Anstoß kam 1984 von GERHARD FREY, damals Professor an der Universität Saarbrücken, wo er auf dem Gebiet der Arithmetik elliptischer Kurven arbeitete. (Von 1990 bis zu seiner Pensionierung im Jahr 2009 leitete er die Arbeitsgruppe Zahlentheorie am Institut für experimentelle Mathematik der (inzwischen mit Duisburg vereinigten) Universität Essen und beschäftigte sich unter anderem mit der Anwendung elliptischer Kurven in der Kryptologie.)

Elliptische Kurven sind ebene Kurven, die durch eine Gleichung der Form $y^2 = f_3(x)$ beschrieben werden mit einem Polynom $f_3(x)$ vom Grad drei mit drei verschiedenen Nullstellen. Da das Quadrat einer reellen Zahl nicht negativ sein kann, gibt es im Reellen nur Punkte mit x-Koordinaten, für die $f_3(x) \geq 0$ ist. Im Falle $f_3(x) > 0$ erfüllt mit y auch $-y$ die obige Gleichung, die Kurve ist also symmetrisch zur x-Achse.

Falls $f_3(x)$ nur zwei verschiedene Nullstellen hat, muß eine der Nullstellen doppelt sein, und bei diesem x-Wert überkreuzt sich die Kurve;


Elliptische Kurve
2
1
0
-1
-2
1
2
3
4
$y^2 = x(x-2)(x-3)$


Knotenkurve
2
1
0
-1
-2
0.5
1
1.5
2
2.5
3
3.5
$y^2 = x(x-2)^2$


Spitzenkurve
2
1
0
-1
-2
0.5
1
1.5
2
$y^2 = x^3$

wir reden dann von einer Knotenkurve.

Hat schließlich $f_3(x)$ nur eine, dafür aber dreifache Nullstelle, entsteht eine Spitzenkurve.

FREY betrachtete eine (hypothetische) Lösung

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

der FERMAT-Gleichung mit teilerfremden natürlichen Zahlen x, y, z und $n \geq 5$. (Den Fall $n = 4$ hat bereits FERMAT selbst gelöst, den Fall $n = 3$ wenig später EULER.) Wenn es eine solche Lösung gibt, dann gibt es auch eine Lösung für einen Primzahlexponenten ℓ, denn ist ℓ ein Primteiler von n und $n = \ell m$, so ist

$$a^\ell + b^\ell = c^\ell \quad \text{mit} \quad a = x^m, \quad b = y^m \quad \text{und} \quad c = z^m \,,$$

und auch a, b, c sind teilerfremd. Auch für ℓ genügt es, den Fall $\ell \geq 5$ zu betrachten, denn wenn wir für ℓ den größten Primteiler von n nehmen, bedeutet $\ell = 2$, daß n eine Zweierpotenz sein muß, was für $n = 2$ kein Widerspruch zur FERMAT-Vermutung ist und für $n = 4$ und damit auch jede höhere Zweierpotenz nach FERMATs Beweis ausgeschlossen ist. Für den Fall $\ell = 3$ kann wieder auf EULER verwiesen werden.

Zur obigen Lösung betrachtete FREY die elliptische Kurve

$$y^2 = x(x - a^\ell)(x + b^\ell) \,,$$

die er aber nicht nur über den reellen oder komplexen Zahlen betrachtet, sondern auch über den Körpern $\mathbb{F}_p$.

FREYs Gleichung definiert genau dann eine elliptische Kurve, wenn alle drei Nullstellen verschieden sind, wenn also

$$a^\ell b^\ell (a^\ell + b^\ell) = a^\ell b^\ell c^\ell = (abc)^\ell$$

nicht verschwindet. Über $\mathbb{F}_p$ sind die drei Nullstellen genau dann verschieden, wenn p kein Teiler dieser Zahl ist, wenn p also keine der drei Zahlen a, b, c teilt.

Da $(abc)^\ell$ verglichen mit a, b, c ziemlich groß ist, heißt das, daß es im Verhältnis zur Größe der Koeffizienten erstaunlich wenige Primzahlen gibt, modulo derer wir *keine* elliptische Kurve erhalten; wir sind

also wieder einer ähnlichen Situation wie bei der abc-Vermutung. Die FREYsche Kurve sieht damit so aus, als sei sie fast zu schön, um wirklich zu existieren.

Einen Anhaltspunkt zum Beweis dieser Nichtexistenz liefert eine Vermutung, die auf um 1955 durchgeführte Rechnungen und Spekulationen des japanischen Mathematikers TANIYAMA zurückgeht und heute je nach Autor mit irgendeiner Kombination der drei Namen TANIYAMA, SHIMURA und WEIL bezeichnet wird. Danach sollte es zu einer elliptischen Kurve E mit ganzzahligen Koeffizienten eine surjektive Abbildung $X_0(N) \to E$ von einer sogenannten Modulkurve $X_0(N)$ auf E geben, wobei N im wesentlichen das Produkt aller Primzahlen p ist, modulo derer E keine elliptische Kurve mehr ist. Wie FREYs Rechnungen zeigen, hat seine Kurve vor diesem Hintergrund sehr seltsame Eigenschaften.

Als er damals hier in Mannheim über seine Resultate vortrug, meinte er noch, er glaube nicht, daß die FERMAT-Vermutung so bewiesen werde; er veröffentlichte sein Ergebnis auch nicht in einer der großen internationalen Fachzeitschriften, sondern als Band 1, Heft 1 einer gerade neu gestarteten Schriftenreihe der Universität Saarbrücken, in einfachster Aufmachung xerographiert mit einem nur schwarz-weiß gestalteten Karton als Umschlag:

GERHARD FREY: Links between stable elliptic curves and certain diophantine equations, *Annales Universitatis Saraviensis, Series Mathematicae,* **1** *(1), 1986*

1987 verschärfte der französische Mathematiker JEAN-PIERRE SERRE die TANIYAMA-Vermutung, und aus dieser stärkeren Vermutung folgt in der Tat, daß die FREY-Kurve nicht existieren kann. Leider ist die SERREsche Vermutung bis heute noch nicht bewiesen.

SERRE erhielt übrigens 2002 den ersten der vom norwegischen Parlament gestifteten ABEL-Preise, die seither zur Erinnerung an den norwegischen Mathematiker NIELS HENRIK ABEL (1802–1829) jedes Jahr in gleicher Weise und gleicher Ausstattung wie die Nobel-Preise für hervorragende Leistungen auf dem Gebiet der Mathematik vergeben werden.

SERRE stellte jedoch noch zusätzlich seine sogenannte ε-Vermutung auf, und auch aus der TANIYAMA-Vermutung zusammen mit der ε-Vermutung

folgt die Nichtexistenz der FREY-Kurve und damit die FERMAT-Vermutung. Diese ε-Vermutung bewies KENNETH RIBET von der Universität Berkeley 1990. Die Grundidee seines Beweises läßt sich interpretieren als eine Art zweidimensionale Version eines Beweises von ERNST EDUARD KUMMER (1810–1893), der die FERMAT-Vermutung 1846 für sogenannte reguläre Primzahlen als Exponenten bewies. (Eine Primzahl p heißt regulär, wenn es für eine primitive p-te Einheitswurzel so etwas wie eine eindeutige Primzerlegung für die hier nicht definierten gazen Elemente von $\mathbb{Q}(\zeta)$ gibt). Der Beweis von RIBET ist allerdings erheblich aufwendiger.

Damit war also die FERMAT-Vermutung zurückgeführt auf die TANIYAMA-Vermutung. Diese Vermutung schließlich (die für die mathematische Forschung erheblich wichtiger ist als die FERMAT-Vermutung) bewies WILES 1994.