Wolfgang K. Seiler

Computeralgebra

Stichworte zur Vorlesung
im Wintersemester 2005/2006
an der Universität Mannheim

Kapitel 1
Einführung

§ 1: Was ist Computeralgebra

Sobald kurz nach dem zweiten Weltkrieg die ersten Computer an Universitäten auftauchten, wurden sie von Mathematikern nicht nur zum numerischen Rechnen eingesetzt, sondern auch für alle anderen Arten mathematischer Routinearbeiten, genau wie auch schon früher alle zur Verfügung stehenden Mittel benutzt wurden: Beispielsweise konstruierte D.H. LEHMER bereits vor rund achtzig Jahren, lange vor den ersten Computern, mit Fahrradketten Maschinen, die (große) natürliche Zahlen in ihre Primfaktoren zerlegen konnten.

Computer manipulieren Bitfolgen; von den meisten Anwendern wurden diese zur Zeit der ersten Computer zwar als Zahlen interpretiert, aber wie wenig später selbst die Buchhalter bemerkten, können sie natürlich auch Informationen ganz anderer Art darstellen. Deshalb wurden bereits auf den ersten Computern (deren Leistungsfähigkeit nach heutigen Standards nicht einmal der eines programmierbaren Taschenrechners entspricht) algebraische, zahlentheoretische und andere abstrakt mathematische Berechnungen durchgeführt wurden. Programmiert wurde meist in Assembler, da die gängigen höhere Programmiersprachen der damaligen Zeit (FORTRAN, ALGOL 60, COBOL, . . .) vor allem mit Blick auf numerische *bzw.*, im Fall von COBOL, betriebswirtschaftliche Anwendungen konzipiert worden waren.

Eine Ausnahme bildete die 1958 von JOHN MCCARTHY entwickelte Programmiersprache LISP, die speziell für symbolische Manipulation entwickelt wurde, vor allem solche im Bereich der künstlichen Intelligenz. In dieser Sprache wurden Ende der Sechzigerjahre die ersten Computeralgebrasysteme geschrieben: MACSYMA ab 1968 ebenfalls am M.I.T. zunächst vor allem für alle Arten von symbolischen Rechnungen in Forschungsprojekten des M.I.T., REDUCE ungefähr gleichzeitig von ANTHONY C. HEARN vor allem für Berechnungen in der Hochenergiephysik.

Beide Systeme verbreiteten sich schnell an den Universitäten und wurden bald auch schon für eine Vielzahl anderer Anwendungen benutzt; dies wiederum führte zur Weiterentwicklung der Systeme sowohl durch die ursprünglichen Autoren als auch durch Benutzer, die neue Pakete hinzufügten, und es führte auch dazu, daß anderswo neue Computeralgebrasysteme entwickelt wurden, wie beispielsweise Maple an der University of Waterloo (einer der Partneruniversitäten von Mannheim). Mit der zunehmenden Nachfrage lohnte es sich auch, deutlich mehr Arbeit in die Entwicklung der Systeme zu stecken, so daß die neuen Systeme oft nicht mehr in LISP geschrieben waren, sondern in klassischen Programmiersprachen wie MODULA oder C *bzw.* später C++, die zwar für das symbolische Rechnen einen erheblich höheren Programmieraufwand erfordern als LISP, die dafür aber auch zu deutlich schnelleren Programmen führen.

Eine gewisse Zäsur bedeutete das Auftreten von *Mathematica* im Jahr 1988. Dies ist das erste System, das von Anfang an rein kommerziell entwickelt wurde. Der Firmengründer und Initiator STEVE WOLFRAM kommt zwar aus dem Universitätsbereich (bevor er seine Firma gründete, forschte er am *Institute for Advanced Studies* in Princeton über zelluläre Automaten), aber *Mathematica* war von Anfang an gedacht als ein Produkt, das an Naturwissenschaftler, Ingenieure und Mathematiker *verkauft* werden sollte. Ein wesentlicher Aspekt, der aus Sicht dieser Zielgruppe den Kauf von *Mathematica* attraktiv machte, obwohl zumindest damals noch eine ganze Reihe anderer Systeme frei oder gegen nominale Gebühr erhältlich waren, bestand in der Möglichkeit, auf einfache Weise Graphiken zu erzeugen. Bei den ersten Systemen hatte dies nie eine Rolle gespielt, da Graphik damals nur über teure Plotter und (zumindest in Universitätsrechenzentrum) mit Wartezeiten von rund einem Tag erstellt werden konnte. 1988 gab es bereits PCs mit (damals

noch sehr schwachen) graphikfähigen Bildschirmen, und Visualisierung spielte plötzlich in allen Wissenschaften eine erheblich größere Rolle als zuvor.

Der Nachteil der ersten *Mathematica*-Versionen war eine im Vergleich zur Konkurenz ziemlich hohe Fehlerquote bei den mathematischen Berechnungen. (Perfekt ist in diesem Punkt auch heute noch kein Computeralgebrasystem.) Der große Vorteil der einfachen Erzeugung von Graphiken sowie das sehr gute Begleitbuch von STEVE WOLFRAM, das deutlich über dem Qualitätsniveau auch heute üblicher Softwaredokumentation liegt, bescherte *Mathematica* einen großen Erfolg. Da auch Systeme wie MACSYMA und MAPLE mittlerweile in selbständige Unternehmen ausgegliedert worden waren, führte die Konkurenz am Markt schnell dazu, daß Graphik auch ein wesentlicher Bestandteil anderer Computeralgebrasysteme wurde und daß *Mathematica* etwas vorsichtiger mit den Regeln der Mathematik umging; heute unterscheiden sich die beiden kommerziell dominanten Systeme Maple und *Mathematica* nicht mehr wesentlich in ihren Graphikfähigkeiten und ihrer (geringen, aber bemerkbaren) Häufigkeit mathematischer Fehler. Hinzu kam der Markt der Schüler und Studenten, so daß ein am Markt erfolgreiches Computeralgebrasystem auch in der Lage sein muß, die Grundaufgaben der Schulmathematik und der Mathematikausbildung zumindest der ersten Semester der gefragtesten Studiengänge zu lösen.

Da die meisten, die mit dem Begriff *Computeralgebra* überhaupt etwas anfangen können, an Computeralgebrasysteme denken, hat sich dadurch auf die Bedeutung des Worts *Computeralgebra* verändert: Gemeinhin versteht man darunter nicht mehr nur ein Programm, das symbolische Berechnungen ermöglicht, sondern eines, das über ernstzunehmende Graphikfähigkeiten verfügt und viele gängige Aufgabentypen lösen kann, ohne daß der Benutzer notwendigerweise versteht, wie man solche Aufgaben löst.

Hier in der Vorlesung wird es in erster Linie um die Algorithmen gehen, die hinter solche System stehen, insbesondere denen, die sich mit der klassischen Aufgabe des symbolischen Rechnens befassen. In den Übungen wird es allerdings zumindest auch teilweise darum gehen,

Computeralgebrasysteme effizient einzusetzen auch zur Visualisierung mathematischer Sachverhalte.

§ 2: Numerisches Rechnen, exaktes Rechnen und symbolisches Rechnen

Numerisches Rechnen gilt gemeinhin als das Rechnen mit reellen Zahlen. Kurzes Nachdenken zeigt, daß wirkliches Rechnen mit reellen Zahlen weder mit Papier und Bleistift noch per Computer wirklich möglich ist: Die Menge $\mathbb{R}$ der reellen Zahlen ist schließlich überabzählbar, aber sowohl unsere Gehirne als auch unsere Computer sind endlich. Der Datentyp **real** oder **float** oder auch **double** kann daher unmöglich das Rechnen mit reellen Zahlen exakt wiedergeben.

Tatsächlich genügt das Rechnen mit reellen Zahlen per Computer völlig anderen Regeln als denen, die wir vom Körper der reellen Zahlen gewohnt sind. Zunächst einmal müssen wir uns notgedrungen auf eine endliche Teilmenge von $\mathbb{R}$ beschränken; in der Numerik sind dies traditionellerweise die sogenannten Gleichtkommazahlen.

Eine Gleitkommazahl wird dargestellt in der Form $x = \pm m \cdot b^{\pm e}$, wobei die *Mantisse* m zwischen 0 und 1 liegt und der *Exponent* e eine ganze Zahl aus einem gewissen vorgegebenen Bereich ist. Die Basis b ist in heutigen Computern gleich zwei, in einigen alten Mainframe Computern sowie in vielen Taschenrechnern wird auch $b = 10$ verwendet.

Praktisch alle heute gebräuchliche CPUs für Computer richten sich beim Format für m und e nach dem IEEE-Standard 754 von 1985. Hier ist $b = 2$, und einfach genaue Zahlen werden in einem Wort aus 32 Bit gespeichert. Das erste dieser Bits steht für das Vorzeichen, null für positive, eins für negative Zahlen. Danach folgen acht Bit für den Exponenenten e und 23 Bit für die Mantisse m.

Die acht Exponentenbit können interpretiert werden als eine ganze Zahl n zwischen 0 und 255; wenn n keinen der beiden Extremwerte 0 und 255 annimmt, wird das Bitmuster interpretiert als die Gleitkommazahl (Mantisse im Zweiersystem)

$$\pm 1{,}m_1 \ldots m_{23} \times 2^{n-127} .$$

Die Zahlen, die in obiger Form dargestellt werden können, liegen somit zwischen $2^{-126} \approx 1{,}175 \cdot 10^{-37}$ und $(2 - 2^{-23}) \cdot 2^{127} \approx 3{,}403 \cdot 10^{38}$. Das führende Bit der Mantisse ist stets gleich eins (sogenannte normalisierte Darstellung) und wird deshalb gleich gar nicht erst abgespeichert. Der Grund liegt natürlich darin, daß man ein führendes Bit null durch Erniedrigung des Exponenten zum Verschwinden bringen kann – es sei denn, man hat bereits den niedrigstmöglichen Exponenten $n = 0$, entsprechend $e = -127$.

Für $n = 0$ gilt daher eine andere Konvention: Jetzt wird die Zahl interpretiert als

$$\pm 0{,}m_1 \ldots m_{23} \times 2^{-126};$$

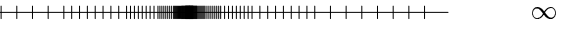
man hat somit einen (unter Numerikern nicht unumstrittenen) *Unterlaufbereich* aus sogenannten *subnormalen* Zahlen, in dem mit immer weniger geltenden Ziffern Zahlen auch noch positive Werte bis hinunter zu $2^{-23} \times 2^{-126} = 2^{-149} \approx 1{,}401 \cdot 10^{-44}$ dargestellt werden können, außerdem natürlich die Null, bei der sämtliche 32 Bit gleich null sind.

Auch der andere Extremwert $n = 255$ hat eine Sonderbedeutung: Falls alle 23 Mantissenbit gleich null sind, steht dies je nach Vorzeichenbit für $\pm\infty$, andernfalls für NAN *(not a number)*, d.h das Ergebnis einer illegalen Rechenoperation wie $\sqrt{-1}$ oder $0/0$. Das Ergebnis von $1/0$ dagegen ist nicht NAN, sondern $+\infty$, und $-1/0 = -\infty$.

Doppeltgenaue Gleitkommazahlen werden entsprechend dargestellt; hier stehen insgesamt 64 Bit zur Verfügung, eines für das Vorzeichen, elf für den Exponenten und 52 für die Mantisse. Durch die elf Exponentenbit können ganze Zahlen zwischen null und 2047 dargestellt werden; abgesehen von den beiden Extremfällen entspricht dies dem Exponenten $e = n - 1023$.

Der Exponent e sorgt dafür, daß Zahlen aus einem relativ großen Bereich dargestellt werden können, er hat aber auch zur Folge, daß die Dichte der darstellbaren Zahlen in den verschiedenen Größenordnung stark variiert: Am dichtesten liegen die Zahlen in der Umgebung der Null, und mit steigendem Betrag werden die Abstände benachbarter Zahlen immer größer.

Um dies anschaulich zu sehen, betrachten wir ein IEEE-ähnliches Gleitkommasystem mit nur sieben Bit, einem für das Vorzeichen und je drei für Exponent und Mantisse. Das folgende Bild zeigt die Verteilung der so darstellbaren Zahlen (mit Ausnahme von NAN):



Um ein Gefühl dafür zu bekommen, was dies für das praktische Rechnen mit Gleitkommazahlen bedeutet, betrachten wir ein analoges System mit der uns besser vertrauten Dezimaldarstellung von Zahlen (für die es einen eigenen IEEE-Standard 854 von 1987 gibt), und zwar nehmen wir an, daß wir eine dreistellige dezimale Mantisse haben und Exponenten zwischen -3 und 3. Da es bei einer von zwei verschiedenen Basis keine Möglichkeit gibt, bei einer normalisierten Mantisse die erste Ziffer einzusparen, schreiben wir die Zahlen in der Form $\pm 0{,}m_1m_2m_3 \cdot 10^e$.

Zunächst einmal ist klar, daß die Summe zweier Gleitkommazahlen aus diesem System nicht immer als Gleitkommazahl im selben System darstellbar ist: Ein einfaches Gegenbeispiel wäre die Addition der größten darstellbaren Zahl $0{,}999 \cdot 10^3 = 999$ zu $5 = 0{,}5 \cdot 10^1$: Natürlich ist das Ergebnis 1004 nicht mehr im System darstellbar. Der IEEE-Standard sieht vor, daß in so einem Fall eine *overflow*-Bedingung gesetzt wird und das Ergebnis gleich $+\infty$ wird. Wenn man (wie es die meisten Compiler standardmäßig tun) die *overflow*-Bedinung ignoriert und mit dem Ergebnis $+\infty$ weiterrechnet, kann dies zu akzeptablen Ergebnissen führen: Beispielsweise wäre die Rundung von $1/(999 + 5)$ auf die Null für viele Anwendungen kein gar zu großer Fehler, auch wenn es dafür in unserem System die sehr viel genauere Darstellung $0{,}996 \cdot 10^{-3}$ gibt. Spätestens wenn man das Ergebnis mit 999 multipliziert, um den Wert von $999/(999 + 5)$ zu berechnen, sind die Konsequenzen aber katastrophal: Nun bekommen wir eine Null anstelle von $0{,}996 \cdot 10^0$. Ähnlich sieht es auch aus, wenn wir anschließend 500 subtrahieren: $\infty - 500 = \infty$, aber $(999 + 5) - 500 = 504$ ist eine Zahl, die sich in unserem System sogar exakt darstellen ließe!

Auch ohne Bereichsüberschreitung kann es Probleme geben: Beispiels-

weise ist

$$123 + 0{,}0456 = 0{,}123 \cdot 10^3 + 0{,}456 \cdot 10^{-1} = 123{,}0456$$

mit einer nur dreistelligen Mantisse nicht exakt darstellbar. Hier sieht der Standard vor, daß das Ergebnis zu einer darstellbaren Zahl gerundet wird, wobei mehrere Rundungsvorschriften zur Auswahl stehen. Voreingestellt ist üblicherweise eine Rundung zur nächsten Maschinenzahl; wer etwas anderes möchte, kann dies durch spezielle Bits in einem Prozessorstatusregister spezifizieren. Im Beispiel würde man also $123+0{,}0456 = 123$ oder (bei Rundung nach oben) 124 setzen und dabei zwangsläufig einen Rundungsfehler machen.

Wegen solcher unvermeidlicher Rundungsfehler gilt das Assoziativgesetz selbst dann nicht, wenn es keine Bereichsüberschreitung gibt: Bei Rundung zur nächsten Maschinenzahl ist beispielsweise

$$(0{,}456\cdot 10^0 + 0{,}3\cdot 10^{-3}) + 0{,}4\cdot 10^{-3} = 0{,}456\cdot 10^0 + 0{,}4\cdot 10^{-3} = 0{,}456\cdot 10^0,$$

aber

$$0{,}456\cdot 10^0 + (0{,}3\cdot 10^{-3} + 0{,}4\cdot 10^{-3}) = 0{,}456\cdot 10^0 + 0{,}7\cdot 10^{-3} = 0{,}457\cdot 10^0.$$

Ein mathematischer Algorithmus, dessen Korrektheit unter Voraussetzung der Körperaxiome für $\mathbb{R}$ bewiesen wurde, muß daher bei Gleitkommarechnung kein korrektes oder auch nur annähernd korrektes Ergebnis mehr liefern – ein Problem, das keinesfalls nur theoretische Bedeutung hat.

In der numerischen Mathematik ist dieses Problem natürlich schon seit Jahrzehnten bekannt; das erste Buch, das sich ausschließlich damit beschäftigte, war

J.H. WILKINSON: Rounding errors in algebraic processes, *Prentice Hall*, 1963; Nachdruck bei *Dover*, 1994.

Heute enthält fast jedes Lehrbuch der Numerischen Mathematik entsprechende Abschnitte; zwei neuere Bücher in denen es speziell um diese Probleme, ihr theoretisches Verständnis und praktische Algorithmen geht, sind

FRANÇOISE CHAITIN-CHATELIN, VALÉRIE FRAYSSÉ: Lectures on finite precision computations, SIAM, 1996

sowie das sehr ausführlichen Buch

NICHOLAS J. HIGHAM: Accuracy and stability of numerical algorithms, SIAM, 1996.

Eine ausführliche und elementare Darstellung der IEEE-Arithmetik und des Umgangs damit findet man in

MICHAEL L. OVERTON: Numerical Computing with IEEE Floating Point Arithmetic – *Including One Theorem, One Rule of Thumb and One Hundred and One Exercises*, SIAM, 2001.

§3: Unentscheidbarkeitsprobleme

Ein auch nur moderat koplizierter symbolischer Ausdruck läßt sich praktisch immer auf eine Vielzahl von Arten darstellen, die teils offensichtlich gleich sind, teils aber auch auf den ersten Blick nichts miteinander zu tun haben. Einige Beispiele:

$$\frac{10}{15} = \frac{2}{3}, \quad \sqrt{8} = 2\sqrt{2}, \quad \sqrt{4+2\sqrt{3}} = 1+\sqrt{3}$$

$$(a+b)^2 = a^2 + 2ab + b^2, \quad \frac{x^5-1}{x-1} = 1 + x + x^2 + x^3 + x^4,$$

$$X^5 - 15X^4 + 85X^3 - 225X^2 + 274X - 120$$
$$= (X-1)(X-2)(X-3)(X-4)(X-5),$$

$$\sin x \cos x = \frac{\sin 2x}{2}, \quad 1 + \tan^2 x = \frac{1}{\cos^2 x}$$

Nur in wenigen dieser Fälle ist eine der beiden Darstellung für alle Arten von Anwendungen der anderen vorzuziehen; meist hat mal die eine, mal die andere Form ihre Vorteile.

Andererseits gehört es zu den Grundaufgaben jeglicher Art des Rechnens, daß man entscheiden muß, ob zwei Ausdrücke gleich sind. Dies ist dann am einfachsten, wenn jeder Ausdruck intern durch eine eindeutig bestimmte kanonische Form dargestellt wird. In einem System, daß alle Ergebnisse auf eine solche kanonische Form bringt, lassen sich zwei Ausdrücke einfach dadurch auf Gleichheit testen, daß man ihre

Differenz berechnet; die Ausdrücke sind genau dann gleich, wenn das Ergebnis die kanonische Darstellung der Null ist.

Gegen eine solche Darstellung sprechen sowohl theoretische als auch praktische Gründe: Wenn beispielsweise Polynome stets in ausmultiplizierter Form dargestellt werden, läuft man Gefahr, ein als Produkt von Linearfaktoren gegebenes Polynom zunächst auszumultiplizieren, um dann anschließend mit großer Mühe seine Nullstellen zu bestimmen. Stellt man Polynome dagegen in faktorisierter Form da, so kann es passieren, daß ein als Summe von Potenzen gegebenes Polynom zunächst mit großem Aufwand faktorisiert wird, und wir anschließend beispielsweise eine Stammfunktion suchen, wofür diese Faktorisierung wieder rückgängig gemacht werden muß. Das Ergebnis müßte dann wieder faktorisert werden, wobei je nach Wahl der Integrationskonstanten sehr verschiedene Ergebnisse entstehen können.

In älteren Computeralgebrasystemen wie REDUCE war es üblich, alles auszumultiplizieren; in den heute gebräuchlichen Systemen wie MAPLE und MATHEMATICA werden Umformungen nur noch durchgeführt, wenn es entweder für die jeweilige Rechnung notwendig ist (Zur Berechnung der Stammfunktion eines Polynoms muß dieses in ausmultiplizierter Form vorliegen) oder wenn es der Anwender explizit verlangt. Lediglich in einigen offensichtlichen Fällen bemühen sich auch diese Systeme um Normalisierung: Beispielsweise werden Brüche stets in gekürzter Form dargestellt und bei Summen werden gleichartige Terme zusammengefaßt.

Das theoretische Argument gegen kanonische Darstellungen ist, daß es solche Darstellungen nur für sehr eingeschränkte Klassen von Zahlen und Funktionen gibt: Wie wir gleich sehen werden, ist selbst für reelle Zahlen im allgemeinen unentscheidbar, wann zwei auf unterschiedliche Weise dargestellte Zahlen gleich sind.

§4: Das zehnte Hilbertsche Problem und der Satz von Richardson

Auf dem Internationalen Mathematikerkongreß 1900 stellt DAVID HILBERT 23 Probleme vor, von denen er glaubte, daß sie für die Mathematik

des 20. Jahrhunderts wichtig sein sollten. Die Probleme kamen aus allen Teilgebieten der Mathematik und hatten auch sehr unterschiedlichen Schwierigkeitsgrad: Einige wurden schon sehr bald gelößt, andere sind auch ein Jahrhundert später noch ungelöst. Das zehnte Problem lautete:

§5: Wichtige Datenstrukturen der Computeralgebra

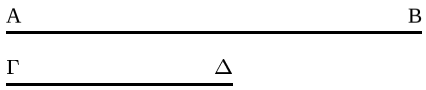
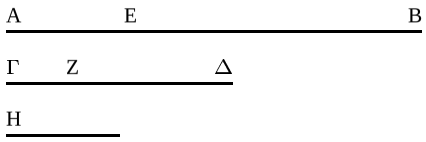
§6: Gängige Computeralgebrasysteme

Kapitel 2
Grundalgorithmen

§ 1: Der Euklidische Algorithmus für ganze Zahlen

a) Der klassische Euklidische Algorithmus

Bei EUKLID, in Proposition 2 des siebten Buchs seiner *Elemente*, wird er 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 ΓZ 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.

Was hier als erstes überrascht, ist die Beschränkung auf nicht zueinander teilerfremde Zahlen. Der Grund dafür liegt darin, daß die klassische griechische Philosophie und Mathematik die Eins nicht als Zahl betrachtete: Zahlen begannen erst bei zwei, und auch Mengen mußten mindestens zwei Elemente haben. Auch bei den Aristotelischen Syllogismen musste sich ein Prädikat auf mindestens zweielementige Klassen beziehen: Die oft als klassischer Syllogismus zitierte Schlußweise

> Alle Menschen sind sterblich
> SOKRATES ist ein Mensch
> Also ist SOKRATES sterblich

wäre von ARISTOTELES nicht anerkannt worden, denn es gab schießlich nur einen SOKRATES. Erst bei seinen Nachfolgern, den Peripatetikern, setzte sich langsam auch die Eins als Zahl durch. EUKLID macht noch brav eine Fallunterscheidung: In Proposition 1, unmittelbar vor der abgedruckten Proposition 2, führt er praktisch dieselbe Konstruktion durch für teilerfremde Zahlen. Außerdem fällt auf, daß EUKLID seine Konstruktion rein geometrisch durchführt; wenn er von einer Strecke eine andere Strecke abträgt solange es geht, ist das natürlich in unserer heutigen arithmetischen Sprache gerade die Konstruktion des Divisionsrests bei der Division der beiden Streckenlängen durcheinander.

In dieser Sprechweise wird der EUKLIDische Algorithmus für heutige Leser wohl auch klar: Wir suchen den größten gemeinsamen Teiler zweier natürlicher Zahlen a und b, d.h. die größte natürliche Zahl d, die

sowohl a als auch b teilt. Wir schreiben kurz

$$d = \mathrm{ggT}(a, b)\,.$$

Grundidee des EUKLIDischen Algorithmus ist die Anwendung der Division mit Rest: Für je zwei natürliche Zahlen x und y gibt es nichtnegative ganze Zahlen q und r, so daß

$$x = qy + r \quad \text{und} \quad 0 \le r < y$$

ist. Alsdann ist

$$\mathrm{ggT}(x, y) = \mathrm{ggT}(y, r)\,,$$

denn wegen der beiden Gleichungen

$$x = qy + r \quad \text{und} \quad r = x - qy$$

teilt jeder gemeinsame Teiler von x und y auch r, und jeder gemeinsame Teiler von y und r teilt auch x.

Der EUKLIDische Algorithmus nutzt dies aus, um die Zahlen, deren ggT bestimmt werden muß, sukzessive zu verkleinern, bis der ggT zweier Zahlen berechnet werden muß, von denen die eine Teiler der anderen ist; in diesem Fall ist natürlich die kleinere der beiden Zahlen gleich dem ggT.

Formal sieht der Algorithmus demnach folgendermaßen aus:

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

Schritt i, $i \ge 1$**:** Falls $r_i = 0$ ist, endet der Algorithmus mit dem Ergebnis $\mathrm{ggT}(a, b) = r_{i-1}$; andernfalls dividiere man r_{i-1} mit Rest durch r_i und bezeichne den Divisionsrest mit r_{i+1}.

(Bei einer tatsächlichen Implementierung bieten sich natürlich einige offensichtliche Optimierungen an.)

Der Algorithmus muß nach endlich vielen Schritten enden, denn bei der Division mit Rest ist stets $0 \le r_{i+1} < r_i$, so daß r_i mit jedem Schritt kleiner wird, was bei natürlichen Zahlen nicht unbegrenzt möglich ist. Da außerdem in jedem Schritt

$$\mathrm{ggT}(r_{i-1}, r_i) = \mathrm{ggT}(r_i, r_{i+1})$$

ist und im letzten Schritt, wenn r_{i-1} den vorigen Wert r_{i-2} teilt,

$$\mathrm{ggT}(r_{i-1}, r_{i-2}) = r_{i-1}$$

ist, folgt induktiv

$$\mathrm{ggT}(a, b) = r_{i-1}\,,$$

so daß der Algorithmus das richtige Ergebnis liefert.

Es ist nicht ganz sicher, ob EUKLID wirklich gelebt hat; das nebenstehende Bild aus dem 18. Jahrhundert ist mit Sicherheit reine Phantasie. EUKLID ist vor allem bekannt als Autor der *Elemente*, in denen er u.a. die Geometrie seiner Zeit systematisch darstellte und (in gewisser Weise) auf wenige Definitionen sowie die berühmten fünf Postulate zurückführte. Diese Elemente entstanden um 300 v. Chr. und waren zwar nicht der erste, aber doch der erfolgreichste Versuch einer solchen Zusammenfassung. EUKLID arbeitete wohl am Museion in Alexandrien; außer den Elementen schrieb er noch ein Buch über Optik und weitere, teilweise verschollene Bücher.

b) Abschätzung des Rechenaufwands

Der obige Beweis, daß der EUKLIDische Algorithmus nach endlich vielen Schritten zum Ziele führt, nutzt aus, daß der Divisionsrest in jedem Schritt kleiner ist als im Schritt zuvor; die Anzahl der Divisionen ist als beschränkt durch das Minimum der beiden Zahlen, auf die wir den Algorithmus anwenden. In der Kryptographie gibt es Anwendungen, bei denen diese Zahlen etwa 600-stellig sind, und natürlich ist es undenkbar, 10^{600} Rechenoperationen auszuführen. Zum Glück ist der tatsächliche Aufwand deutlich geringer.

Um zu einer realistischeren Abschätzung zu kommen, suchen wir die kleinsten natürlichen Zahlen a, b, für die n Divisionen notwendig sind. Im Falle $n = 1$ sind dies offensichtlich $a = b = 1$; im Fall $a = b$ kommt man immer mit genau einer Division aus.

Dies ist allerdings ein eher untypischer Fall, der sich insbesondere nicht rekursiv verallgemeinern läßt, denn ab dem zweiten Schritt des EUKLIDischen Algorithmus ist der Divisor stets kleiner als der Dividend: Ersterer ist schließlich der Rest bei der vorangegangenen Division und

letzterer der Divisor. Die kleinsten natürlichen Zahlen $a \neq b$, für die man mit nur einer Division auskommt, sind $a = 2$ und $b = 1$.

Als nächstes Suchen wir die kleinsten Zahlen a, b, für die zwei Divisionen notwendig sind. Ist r der Rest bei der ersten Division, so ist $b : r$ die zweite Division. Für diese muß $r \geq 1$ und $b \geq 2$ sein, und $a = qb + r$, wobei q der Quotient bei der ersten Division ist. Dieser ist mindestens eins, die kleinstmöglichen Werte sind damit

$$r = 1, \quad b = 2 \quad \text{und} \quad a = b + r = 3\,.$$

Allgemeiner seien a_n und b_n mit $a_n > b_n$ die kleinsten Zahlen, für die n Divisionen notwendig sind, und r_n sei der Rest bei der Division von a_n durch b_n. Im nächsten Schritt des EUKLIDischen Algorithmus wird dann b_n durch r_n dividiert; da a_n und b_n die kleinstmöglichen Zahlen sind, muß dabei der Quotient gleich eins sein und $b_n = a_{n-1}$ sowie $r_n = b_{n-1}$. Also ist

$$a_n = b_n + r_n = a_{n-1} + b_{n-1} = a_{n-1} + a_{n-2} \quad \text{und} \quad b_n = a_{n-1}\,.$$

Da wir $a_1 = 2$ und $b_1 = 1$ kennen, können wir daraus alle a_n und b_n berechnen; was wir erhalten, sind die sogenannten FIBONACCI-Zahlen.

Sie sind durch folgende Rekursionsformel definiert:

$$F_0 = 0, \qquad F_1 = 1 \qquad \text{und} \qquad F_i = F_{i-1} + F_{i-2} \quad \text{für } i \geq 2\,.$$

Somit ist $a_1 = F_3$ und $b_1 = F_2$, und es folgt rekursiv, daß $a_n = F_{n+2}$ und $b_n = F_{n+1}$ ist.

Damit folgt

Satz von Lamé (1844)**:** Die kleinsten natürlichen Zahlen a, b, für die bei EUKLIDischen Algorithmus $n \geq 2$ Divisionen notwendig sind, sind $a = F_{n+2}$ und $b = F_{n+1}$. ∎

(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 Zar Alexanders I. ging er 1820 nach Rußland, wo er Vorlesungen über Analysis, Physik, Chemie und Ingenierswissenschaften hielt. Nach seiner Rückkehr 1832 erhielt er einen Lehrstuhl für Physik an der Ecole Polytechnique, 1852 einen für mathematische Physik und Wahrscheinlichkeitstheorie an der Sorbonne. 1836/37 war er war auch wesentlich am Bau der Eisenbahnlinien Paris-Versailles und Paris-St. Germain beteiligt.

Um zu einer Aufwandsabschätzung zu kommen, müssen wir uns die FIBONACCI-Zahlen etwas genauer ansehen. FIBONACCI führte sie ein, um die Vermehrung einer Karnickelpopulation durch ein einfaches Modell zu berechnen. In seinem 1202 erschienenen Buch *Liber abaci* schreibt er:

Ein Mann bringt ein Paar Karnickel auf einen Platz, der von allen Seiten durch eine Mauer umgeben ist. Wie viele Paare können von diesem Paar innerhalb eines Jahres produziert werden, wenn man annimmt, daß jedes Paar jeden Monat ein neues Paar liefert, das vom zweiten Monat nach seiner Geburt an produktiv ist?

LEONARDO PISANO (1170–1250) ist heute vor allem unter seinem Spitznamen FIBONACCI bekannt; gelegentlich nannte er sich auch BIGOLLO, auf Deutsch *Tunichtgut* oder *Reisender.* Seine Bücher waren mit die ersten, die die indisch-arabischen Ziffern in Europa einführten. Er behandelt darin nicht nur Rechenaufgaben für Kaufleute, sondern auch zahlentheoretische Fragen, beispielsweise daß man die Quadratzahlen durch Aufaddieren der ungeraden Zahlen erhält. Auch betrachtet er Beispiele nichtlinearer Gleichungen, die er approximativ löst, und erinnert an viele in Vergessenheit geratene Ergebnisse der antiken Mathematik.

Um die Zahlen F_i durch eine geschlossene Formel darzustellen, betrachten wir die (formale) Potenzreihe

$$X(z) = \sum_{i=0}^{\infty} F_i z^i\,.$$

Auf Grund der Rekursionsformel $F_i = F_{i-1} + F_{i-2}$ für $i \geq 2$ ist

$$\sum_{i=2}^{\infty} F_i z^i = \sum_{i=2}^{\infty} F_{i-1} z^i + \sum_{i=2}^{\infty} F_{i-1} z^i = z \sum_{i=1}^{\infty} F_i z^i + z^2 \sum_{i=0}^{\infty} F_{i-1} z^i ,$$

was wir wegen $F_0 = 0$ und $F_1 = 1$ auch in der Form

$$X(z) - z = zX(z) + z^2 X(z)$$

schreiben können. Auflösen nach $X(z)$ führt auch

$$X(z) = \frac{z}{1 - z - z^2} .$$

Um die rechte Seite als Potenzreihe in z zu schreiben, versuchen wir, sie durch Terme der Form $\frac{1}{1-q}$ darzustellen, die wir als Summen geometrischer Reihen $\sum_{i=0}^{\infty} q^i$ schreiben können.

Da $z^2 + z - 1 = \left(z + \frac{1}{2}\right)^2 - \frac{5}{4}$ ist, verschwindet der Nenner für die beiden Werte

$$z = z_{1/2} = -\frac{1}{2} \pm \sqrt{\frac{5}{4}} = -\frac{1 \mp \sqrt{5}}{2} .$$

Nach dem Satz von VIÈTE ist $z_1 z_2 = z_1 + z_2 = -1$, also

$$1 - z - z^2 = -(z - z_1)(z - z_2) = \frac{(z - z_1)(z - z_2}{z_1 z_2}$$

$$= \left(1 - \frac{z}{z_1}\right)\left(1 - \frac{z}{z_2}\right) = (1 + z_2 z)(1 + z_1 z) .$$

Da wir die Summenformel der geometrischen Reihe besser anwenden können, wenn wir Terme der Form $(1 - q)$ haben, definieren wir die beiden neuen Zahlen

$$\phi = \frac{1 + \sqrt{5}}{2} \quad \text{und} \quad \overline{\phi} = \frac{1 - \sqrt{5}}{2} ;$$

dann ist

$$1 - z - z^2 = (1 - \phi z)(1 - \overline{\phi} z) .$$

Bemerkung: ϕ und $\overline{\phi}$ erfüllen die Gleichung $\phi^2 - \phi - 1 = 0$ oder $\phi^2 = \phi + 1$, d.h. ϕ ist das Verhältnis des *goldenen Schnitts:* Zwei Größen $a > b$ stehen bekanntlich in diesem Verhältnis, wenn sich $a + b$ zu a verhält wie a zu b. Für $\phi = a/b$ ist dies die Bedingung

$$1 + \phi^{-1} = 1 + \frac{b}{a} = \frac{a + b}{a} = \frac{a}{b} = \phi ,$$

die nach Multiplikation mit ϕ zu $\phi + 1 = \phi^2$ wird.

Nach diesen Vorbereitungen können wir mit der Partialbruchzerlegung von $X(z)$ beginnen: Nach der allgemeinen Theorie machen wir den Ansatz

$$X(z) = \frac{z}{1 - z - z^2} = \frac{\alpha}{1 - \phi z} + \frac{\beta}{1 - \overline{\phi} z} = \frac{(\alpha + \beta) - (\alpha\overline{\phi} + \beta\phi)z}{1 - z - z^2} ,$$

der auf die beiden Gleichungen

$$\alpha + \beta = 0 \quad \text{und} \quad \alpha\overline{\phi} + \beta\phi = -1$$

führt. Einsetzen von $\beta = -\alpha$ in die zweite Gleichung zeigt, daß

$$\alpha(\overline{\phi} - \phi) = -\alpha\sqrt{5} = -1 \quad \text{oder} \quad \alpha = \frac{1}{\sqrt{5}}$$

ist. Also ist

$$X(z) = \frac{1}{\sqrt{5}} \left(\frac{1}{1 - \phi z} - \frac{1}{1 - \overline{\phi} z} \right) .$$

Diese beiden Summanden können wir nun als Summen geometrischer Reihen interpretieren und erhalten

$$X(z) = \frac{1}{\sqrt{5}} \left(\sum_{i=0}^{\infty} \phi^i z^i + \sum_{i=0}^{\infty} \overline{\phi}^i z^i \right) = \frac{1}{\sqrt{5}} \sum_{i=1}^{\infty} (\phi^i + \overline{\phi}^i) .$$

Koeffizientenvergleich zeigt, daß

$$F_i = \frac{\phi^i + \overline{\phi}^i}{\sqrt{5}}$$

ist, womit wir die gesuchte explizite Formel gefunden hätten.

In Zahlen ist $\phi = \frac{1 + \sqrt{5}}{2} \approx 1{,}618034$, $\overline{\phi} = 1 - \phi \approx -0{,}618034$ und $\sqrt{5} \approx 2{,}236068$; der Quotient $\overline{\phi}^i / \sqrt{5}$ ist also für jedes i kleiner als $1/2$.

Daher können wir F_i auch einfacher berechnen als nächste ganze Zahl zu $\phi^i/\sqrt{5}$. Insbesondere folgt, daß F_i exponentiell mit i wächst.

Für $a = F_{n+2}$ und $b = F_{n+1}$, die beiden kleinsten Zahlen, für die beim EUKLIDischen Algorithmus n Divisionen notwendig sind, ist also

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

Für beliebige Zahlen $a > b$ können nicht mehr Divisionen notwendig sein für die auf b folgenden nächstgrößeren FIBONACCI-Zahlen, also gibt obige Formel für jedes b eine obere Grenze. Die Anzahl der Divisionen wächst also nicht, wie bei der naiven Abschätzung im vorigen Abschnitt, mit b, sondern nur mit $\ln b$. Für sechshundertstellige Zahlen a, b müssen wir also nicht mit 10^{600} Divisionen rechnen, sondern mit weniger als drei Tausend, was auch für weniger leistungsfähige Computer problemlos und schnell möglich ist.

Tatsächlich ist natürlich auch noch 3 000 für die meisten sechshundertstelligen Zahlen eine gewaltige Überschätzung des tatsächlichen Aufwands, denn hier handelt es sich ja um eine obere Grenze, von der wir nur im Falle der FIBONACCI-Zahlen wissen, daß sie wirklich angenommen wird. Für zufällig gewählte Zahlen ist der Aufwand im Durchschnitt erheblich geringer, siehe dazu DONALD E. KNUTH: *The Art of Computer Programming 2: Seminumerical Algorithms,* Addison Wesley, viele Auflagen.

c) Der erweiterte Euklidische Algorithmus

Die Grundform des EUKLIDische Algorithmus reicht uns nicht aus; für viele Zwecke (nicht nur) der Computeralgebra ist mindestens genauso wichtig, den ggT als ganzzahlige Linearkombination der Ausgangsdaten darzustellen wie ihn zu berechnen. Daß eine solche Darstellung tatsächlich möglich ist, zeigt der erweiterte EUKLIDische Algorithmus, der diese Darstellung auch explizit liefert:

Ausgangspunkt ist wieder die Division mit Rest; die zugehörige Gleichung

$$r_{i-1} = q_i r_i + r_{i+1}$$

läßt sich umschreiben als

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

so daß r_{i+1} eine ganzzahlige Linearkombination von r_i und r_{i-1} ist. Da entsprechend auch r_i Linearkombination von r_{i-1} und r_{i-2} ist, folgt induktiv, daß der ggT von a und b als ganzzahlige Linearkombination von $r_0 = a$ und $r_1 = b$ dargestellt werden kann.

Algorithmisch sieht dies folgendermaßen aus:

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

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

Diese Relationen werden in jedem der folgenden Schritte erhalten:

Schritt i, $i \geq 1$**:** Falls $r_i = 0$ ist, endet der Algorithmus mit

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

Andernfalls dividiere man r_{i-1} mit Rest durch r_i mit dem Ergebnis

$$r_{i-1} = q_i r_i + r_{i+1}\,.$$

Dann ist

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

man setze also

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

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 ist also im letzten Schritt der ggT als Linearkombination der Ausgangszahlen dargestellt.

Als Beispiel wollen wir den ggT von 200 und 148 als Linearkombination darstellen. Im nullten Schritt haben wir 200 und 148 als die trivialen Linearkombinationen

$$200 = 1 \cdot 200 + 0 \cdot 148 \quad \text{und} \quad 148 = 0 \cdot 200 + 1 \cdot 148\,.$$

Im ersten Schritt dividieren wir, da 148 nicht verschwindet, 200 mit Rest durch 148:

$$200 = 1 \cdot 148 + 52 \Longrightarrow 52 = 1 \cdot 200 - 1 \cdot 148$$

Da auch $52 \neq 0$, dividieren wir im zweiten Schritt 148 durch 52 mit Ergebnis $148 = 2 \cdot 52 + 44$, d.h.

$$44 = 148 - 2 \cdot (1 \cdot 200 - 1 \cdot 148) = 3 \cdot 148 - 2 \cdot 200$$

Auch $44 \neq 0$, wir dividieren also weiter: $52 = 1 \cdot 44 + 8$ und

$$\begin{aligned} 8 &= 52 - 44 = (1 \cdot 200 - 1 \cdot 148) - (3 \cdot 148 - 2 \cdot 200) \\ &= 3 \cdot 200 - 4 \cdot 148\,. \end{aligned}$$

Im nächsten Schritt erhalten wir $44 = 5 \cdot 8 + 4$ und

$$\begin{aligned} 4 &= 44 - 5 \cdot 8 = (3 \cdot 148 - 2 \cdot 200) - 5 \cdot (3 \cdot 200 - 4 \cdot 148) \\ &= 23 \cdot 148 - 17 \cdot 200\,. \end{aligned}$$

Bei der Division von acht durch vier schließlich erhalten wir Divisionsrest Null; damit ist vier der ggT von 148 und 200 und kann in der angegebenen Weise linear kombiniert werden.

Der erweiterte EUKLIDische Algorithmus kann auch zur Lösung linearer diophantischer Gleichungen verwendet werden: Angenommen wir suchen ganzzahlige Lösungen (x, y) der linearen Gleichung

$$ax + by = c \quad \text{mit} \quad a, b, c \in \mathbb{Z}\,.$$

Da die linke Seite für alle x, y ein Vielfaches des ggT von a und b ist, kann es offensichtlich nur dann Lösungen geben, wenn $\mathrm{ggT}(a, b)$ ein Teiler von c ist. Falls dies gilt, können wir aus der linearen Darstellung

$$\mathrm{ggT}(a, b) = \alpha a + \beta b$$

durch Multiplikation mit $c/ggT(a, b)$ eine lineare Darstellung

$$c = xa + yb$$

konstruieren, also eine Lösung der Gleichung.

Dies ist allerdings nicht die einzige Lösung: Wegen $ba - ab = 0$ ist offensichtlich auch $(x+b, y-a)$ eine. Allgemeiner gilt $au+bv = 0$ auch für $u = b/gT(a, b)$ und $v = -a/ggT(a, b)$, und die allgemeine Lösung der Gleichung ist daher

$$\left(x + \frac{kb}{\mathrm{ggT}(a, b)},\ y - \frac{ka}{\mathrm{ggT}(a, b)}\right) \quad \text{mit} \quad k \in \mathbb{Z}\,.$$

Lineare diophantische Gleichungen mit mehr als zwei Unbekannten haben die Form

$$a_1 x_1 + \cdots + a_n x_n = c$$

mit ganzen Zahlen a_i, c; gesucht sind ganzzahlige Lösungen x_i. Auch eine solche Gleichung ist offensichtlich unlösbar, wenn der ggT der Koeffizienten a_i die rechte Seite nicht teilt. Wenn er sie teilt, können wir im wesentlichen wie im Fall zweier Veränderlichen vorgehen, indem wir zunächst den ggT als Linearkombination der a_i ausdrücken: Dazu berechnen wir zunächst den ggT d_2 von a_1 und a_2 und stellen diesen als Linearkombination von a_1 und a_2 dar. Sodann berechnen wir den ggT von d_2 und a_3; das ist gleichzeitig der ggT von a_1, a_2 und a_3. Wir stellen ihn als Linearkombination dieser Zahlen dar, indem wir ihn zunächst als Linearkombination von d_2 und a_3 schreiben und dann für d_2 die im vorigen Schritt berechnete Darstellung als Linearkombination von a_1 und a_2 einsetzen, und so weiter. Durch Multiplikation läßt sich aus dem Ergebnis eine Lösung der obigen Gleichung finden; weitere Lösungen erhält man durch Addition von Lösungen der homogenen Gleichung.

Löungen von Systemen linearer diophantischer Gleichungen findet man, indem man den GAUSS-Algorithmus ohne Divisionen anwendet: Möchte man aus einer Gleichung

$$a_{i1} x_1 + \cdots + a_{in} x_n = b_i \quad \text{mit} \quad a_{i1} \neq 0$$

die Variable x_1 eliminieren mittels der Gleichung

$$a_{j1} x_1 + \cdots + a_{jn} x_n = b_j \quad \text{mit} \quad a_{j1} \neq 0\,,$$

so subtrahiert man beim klassischen GAUSS-Algorithmus das a_{i1}/a_{j1}-fache dieser Gleichung von der Ausgangsgleichung, wodurch im allgemeinen Brüche ins Spiel kommen. Beim GAUSS-Algorithmus ohne

Divisionen bildet man stattdessen die Linearkombination a_{i1} mal zweite Gleichung minus a_{j1} mal erste Gleichung, in der x_1 ebenfalls nicht mehr vorkommt.

...

d) Der chinesische Restesatz

Hier geht es darum, eine ganze Zahl x zu finden derart, die modulo vorgegebener Zahlen $m_1, \ldots, m_r$ kongruent ebenfalls vorgegebener Zahlen $a_1, \ldots, a_r$ sind. Damit dieses Problem stets lösbar ist, werden die Zahlen $m_1, \ldots, m_r$ als paarweise teilerfremd vorausgesetzt.

Betrachten wir zunächst den Fall $r = 2$: Hier geht es darum, ein x zu finden mit

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

für zwei zueinander teilerfremde Zahlen m und n.

Da m und n teilerfremd sind, haben sie den ggT eins, der sich nach dem erweiterten EUKLIDischen Algorithmus als

$$1 = \alpha m + \beta n$$

schreiben läßt. 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.

Es ist natürlich nicht die einzige Lösung; wenn wir ein gemeinsames Vielfaches von m und n addieren, ändert sich nichts an den Kongruenzen. Da wir von teilerfremden Zahlen ausgegangen sind, ist das Produkt das kleinste gemeinsame Vielfache; die allgemeine Lösung ist somit

$$x + (\beta n + \lambda b)a + (\alpha m - \lambda a)b\,;$$

insbesondere ist die Lösung eindeutig modulo nm.

Bei mehr als zwei Kongruenzen geht man rekursiv vor: Man löst die ersten beiden Kongruenzen $x \equiv a_1 \bmod m_1$ und $x \equiv a_2 \bmod m_2$ wie

gerade besprochen; das Ergebnis ist eindeutig modulo $m_1 m_2$. Ist c_2 eine feste Lösung, so läßt sich die Lösung schreiben als Kongruenz

$$x \equiv c_2 \bmod m_1 m_2\,,$$

und da die m_i paarweise teilerfremd sind, ist auch $m_1 m_2$ teilerfremd zu m_3. Mit EUKLID können wir daher das System

$$x \equiv c_2 \bmod m_1 m_2 \quad \text{und} \quad x \equiv a_3 \bmod m_3$$

lösen und die Lösung schreiben als

$$x \equiv c_3 \bmod m_1 m_2 m_3$$

und so weiter, bis wir schließlich x modulo dem Produkt aller m_i kennen und somit das Problem gelöst haben.

Alternativ läßt sich die Lösung auch in einer geschlossenen Formel 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, berechnet man zunächst für jedes i das Produkt

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

der sämtlichen übrigen m_j und bestimmt 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 den obigen Formel – oft größer sein als das Produkt der m_i; um die kleinste Lösung zu finden, muß man also noch modulo diesem Produkt reduzieren.

Der chinesische Restesatz hat seinen Namen daher, daß angeblich chinesische Generäle ihre Truppen in Zweier-, Dreier-, Fünfer-, Siebenerreihen *usw.* antreten ließen und jeweils nur die (i.a. unvollständige) letzte Reihe abzählten. Aus den Ergebnissen lies sich die Gesamtzahl der Soldaten berechnen, wenn das Produkt der verschiedenen Reihenlängen größer war als diese Anzahl.

Es ist fraglich, ob die chinesischen Generäle wirklich soviel Mathematik konnten; Beispiele zu diesem Satz finden sich jedenfalls bereits 1247 in den *Mathematischen Abhandlungen in neun Bänden* von CH'IN CHIU-SHAO (1202–1261), allerdings geht es dort nicht um Soldaten, sondern um Reis.

§2: Der allgemeine Euklidische Algorithmus

Wenn wir uns den vorigen Paragraphen genau anschauen, sehen wir, daß von den vielen Eigenschaften der ganzen Zahlen vor allem eine wesentlich war: Die Division mit Rest. Division mit Rest gibt es beispielsweise auch für Polynome in einer Veränderlichen über einem Körper, so daß es möglich sein sollte, dieselben Algorithmen dafür zu formulieren. In der Tat spielt der EUKLIDische Algorithmus für Polynome in der Computeralgebra eine noch größere Rolle als der für Zahlen und wird uns im Rahmen dieser Vorlesung noch vielfach begegnen. Allerdings sind ganze Zahlen und Polynome erstens nicht die einzigen Beispiele, für die es einen EUKLIDischen Algorithmus gibt, und zweitens lassen sich viele Fragen simultan für beide Fälle klären, wenn wir den EUKLIDischen Algorithmus nur etwas abstrakter formulieren.

a) Grundbegriffe der Ringtheorie

Definition: *a)* Ein Ring ist eine Menge R zusammen mit zwei Rechenoperationen „+“ und „·“, so daß gilt:

1.) R bildet bezüglich „+“ eine abelsche Gruppe

2.) Die Verknüpfung „·“: $R \times R \to R$ erfüllt das Assoziativgesetz $x(yz) = (xy)z$ und es gibt ein Element $1 \in R$, so daß $1x = x1 = x$.

3.) „+“ und „·“ erfüllen die Distributivgesetze $x(y + z) = xy + xz$ und $(x + y)z = xz + yz$..

b) Ein Ring heißt *kommutativ*, falls zusätzlich noch das Kommutativgesetz $xy = yx$ der Multiplikation erfüllt ist.

c) Ein Ring heißt *nullteilerfrei* wenn gilt: Falls ein Produkt $xy = 0$ verschwindet, muß mindestens einer der beiden Faktoren x, y gleich Null sein. Ein nullteilerfreier kommutativer Ring heißt *Integritätsbereich.*

d) Wir sagen, ein Element u eines Integritätsbereichs R sei *Teiler* von $x \in R$, in Zeichen $u|x$, wenn es ein $q \in R$ gibt, so daß $x = qu$.

e) $u \in R$ 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: $v|u$.

f) Ein Element $e \in R$ heißt *Einheit*, falls es ein $e' \in R$ gibt mit $ee' = 1$. Die Menge aller Einheiten von R bezeichnen wir mit $R^\times$.

g) Zwei Elemente $x, y \in R$ heißen assoziiert, wenn es eine Einheit $e \in R$ gibt, so daß $y = ex$.

Der Prototyp eines kommutativen Rings ist der Ring $\mathbb{Z}$ der ganzen Zahlen; er ist offensichtlich ein Integritätsbereich und seine einzigen Einheiten sind ± 1.

Der Menge aller $n \times n$-Matrizen über einem Körper ist ein Beispiel eines nichtkommutativen Rings. Er ist nicht nullteilerfrei, und seine Einheiten sind genau die invertierbaren Matrizen.

Auch die Polynome über einem Körper k bilden einen Ring, den Polynomring $k[X]$. Allgemeiner gilt sogar:

Lemma: Ist R ein Integritätsbereich, so auch der Polynomring

$$R[X] = \left\{ \sum_{i=0}^{n} a_i X^i \;\middle|\; b \in \mathbb{N}_0, a_i \in R \right\} .$$

Seine Einheiten sind genau die Einheiten von R.

Beweis: Wenn wir Addition und Multiplikation nach den üblichen Regeln definieren, ist klar, daß $R[X]$ alle Ringaxiome erfüllt.

Um zu zeigen, daß $R[X]$ nullteilerfrei ist, betrachten wir zwei Polynome

$$f = \sum_{i=0}^{n} a_i X^i \quad \text{und} \quad g = \sum_{j=0}^{m} b_j X^j ,$$

die beide von Null verschieden sind. Wir können etwa annehmen, daß n und m so gewählt sind, daß a_n und b_m beide nicht verschwinden.

Da R Integritätsbereich ist, kann dann auch das Produkt $a_n b_m$ nicht verschwinden, also ist der führende Term $a_n b_m X^{n+m}$ 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 die Grade addieren.

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 $R^\times$. ∎

Allgemein gilt:

Lemma: *a)* Die Menge $R^\times$ aller Einheiten von R ist eine abelsche Gruppe bezüglich der Multiplikation.

b) Ein kommutativer Ring R ist genau dann ein Integritätsbereich, wenn gilt: Ist $xz = yz$ für ein Element $z \neq 0$ und zwei beliebige Elemente x, y, so ist $x = y$.

c) Zwei Elemente x, y eines Integritätsbereich R sind genau dann assoziert, 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 Elemente e', f' mit $ee' = ff' = 1$. Damit ist $(ef)(f'e') = e(ff')e' = ee' = 1$, d.h. auch ef ist eine Einheit. Außerdem ist jede Einheit invertierbar, denn offensichtlich ist e' ein multiplikatives Inverses zu e.

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 $xy = 0$ und $y \neq 0$, so ist $xy = 0y$, also $x = 0$.

c) Ist $y = ex$, so ist x ein Teiler von y. Da Einheiten invertierbar sind, ist auch $x = e^{-1}y$, d.h. $y|x$.

Ist umgekehrt $x|y$ und $y|x$, so gibt es Elemente q, r mit $x = qy$ und $y = rx$. Damit ist $1x = x = (qr)x$, also $qr = 1$. 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. ∎

In Integritätsbereichen können wir somit einen Teilbarkeitsbegriff einführen, der den üblichen, von $\mathbb{Z}$ her gewohnten Regeln genügt. Manchmal können wir auch, wie in $\mathbb{Z}$, von einer eindeutigen Primzerlegung reden:

Definition: *a)* Ein Element x eines Integritätsbereichs R heißt *irreduzibel*, falls gilt: x ist keine Einheit, und ist $x = yz$ das Produkt zweier Elemente aus R, so 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 eindeutig schreiben als Produkt $x = e \prod_{i=1}^{r} p_i^{s_i}$ mit einer Einheit $e \in R^\times$, irreduziblen Elementen $p_i \in R$ und natürlichen Zahlen s_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: 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 irreduzibel 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 $g_i \leq f_i$. ∎

b) Euklidische Ringe

EUKLIDische Ringe sind die Ringe, in denen es einen EUKLIDischen Algorithmus gibt. Wie wir gesehen haben, ist dazu die Division mit Rest das wichtigste Werkzeug, und wie sich in diesem Abschnitt herausstellen wird, brauchen wir kein weiteres. 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 der Ring $\mathbb{Z}$ der ganzen Zahlen mit $\nu(x) = |x|$. Ein anderes Beispiel ist der Polynomring $k[X]$ über einem Körper k: Hier können wir $\nu(f)$ für ein Polynom $f \neq 0$ als den Grad von f definieren; dann erfüllt auch die Polynomdivision mit Rest die Forderung an einen EUKLIDischen Ring.

Als letztes Beispiel schließlich kann sich der Leser überlegen, daß auch der Ring $\mathbb{Z}[i]$ der komplexen Zahlen mit ganzzahligen Real- und Imaginärteilen ein EUKLIDischer Ring ist; hier kann man $\nu(x+iy) = x^2 + y^2$ setzen.

Wie angekündigt, gilt

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: 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 Dividend x und Divisor y, die natürlich beide als Linearkombinationen von x und y darstellbar sind, und induktiv folgt, daß auch alle folgenden Dividenden und Divisoren sind als Reste einer vorangegangenen Division Linearkombinationen von x und y sind, also ist es auch ihr Divisionsrest. Insbesondere ist auch der ggT als letzter nichtverschwindender Divisionsrest Linearkombination von x und y, und die Koeffizienten können wie im vorigen Paragraphen mit dem erweiterten EUKLIDischen Algorithmus berechnet werden. ∎

c) Eindeutige Primzerlegung in Euklidischen Ringen

Satz: Jeder EUKLIDische Ring ist faktoriell.

Beweis: Wir müssen zeigen, daß jedes Element $x \neq 0$ aus R 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.

Dazu benutzen wir die Betragsfunktion $\nu\colon R \setminus \{0\} \to \mathbb{N}_0$ des EUKLIDischen Rings R und beweisen induktiv, daß für $n \in \mathbb{N}_0$ alle $x \neq 0$ mit $\nu(x) \leq n$ in der gewünschten Weise darstellbar sind.

Ist $\nu(x) = 0$, so ist x eine Einheit: Bei der Division $1 : x = q$ Rest r ist nämlich entweder $r = 0$ oder aber $\nu(r) < \nu(x) = 0$. Letzteres ist nicht möglich, also ist $qx = 1$ und x eine Einheit. Diese kann als sich selbst mal dem leeren Produkt von Potenzen irreduzibler Elemente geschrieben werden.

Für $n > 1$ unterscheiden wir zwei Fälle: Ist x irreduzibel, so ist $x = x$ eine Darstellung der gewünschten Form, und wir sind fertig.

Andernfalls läßt sich $x = yz$ als Produkt zweier Elemente schreiben, die beide keine Einheiten sind. Da y und z Teiler von x sind, sind $\nu(y), \nu(z) \leq \nu(x)$. Wir wollen uns überlegen, daß sie tatsächlich sogar echt kleiner sind.

Dazu dividieren wir y mit Rest durch x; das Ergebnis sei q Rest r, d.h. $y = qx+r$ mit $r = 0$ oder $\nu(r) < \nu(x)$. Wäre $r = 0$, wäre y ein Vielfaches von x, es gäbe also ein $u \in R$ mit $y = ux = u(yz) = (uz)y$. Damit wäre $uz = 1$, also z eine Einheit, im Widerspruch zur Annahme. Somit ist $\nu(r) < \nu(x)$.

Weiter ist y ein Teiler von $r = y - qx = y(1 - qz)$, also folgt $\nu(y) \leq \nu(r) < \nu(x)$. Genauso folgt, daß auch $\nu(z) < \nu(x)$ ist.

Nach Induktionsvoraussetzung lassen sich daher y und z als Produkte von Einheiten und Potenzen irreduzibler Elemente schreiben, und damit läßt sich auch $x = yz$ so darstellen.

Als nächstes müssen wir uns überlegen, daß diese Darstellung bis auf Reihenfolge und Einheiten eindeutig ist. Das wesentliche Hilfsmittel hierzu ist die folgende Zwischenbehauptung:

Falls ein irreduzibles Element p ein Produkt xy teilt, teilt es mindestens einen der beiden Faktoren.

Zum *Beweis* betrachten wir den ggT von x und p. Dieser ist insbesondere ein Teiler von p, also bis auf Assoziiertheit entweder p oder 1. Im ersten Fall ist p Teiler von x und wir sind fertig; andernfalls können wir

$$1 = \alpha p + \beta x$$

als Linearkombination von p und x schreiben. Multiplikation mit y macht daraus $y = \alpha px + \beta xy$, 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.

Um den Beweis des Satzes zu beenden, zeigen wir induktiv, daß für jedes $n \in \mathbb{N}_0$ alle Elemente mit $\nu(x) \leq n$ eine bis auf Reihenfolge und Einheiten eindeutige Primfaktorzerlegung haben.

Für $n = 0$ haben wir oben gesehen, daß x eine Einheit sein muß, und hier ist die Zerlegung $x = x$ eindeutig.

Seien nun

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

zwei Zerlegungen eines Elements $x \in R$, 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 also mindestens eines der Elemente q_j, d.h. $p_1 = wq_j$ ist bis auf eine Einheit w gleich q_j. Da p_i keine Einheit ist, ist $\nu(x/p_i) < \nu(x)$; nach Induktionsannahme hat also $x/p_i = x/(wq_j)$ eine bis auf Reihenfolge und Einheiten eindeutige Zerlegung in irreduzible Elemente. Damit hat auch x diese Eigenschaft. ∎

Bemerkung: Die Umkehrung dieses Satzes gilt nicht: Wir werden gleich sehen, daß $\mathbb{Z}[X]$ oder auch Polynomringe in mehr als einer Veränderlichen über einem Körper faktoriell sind, aber keiner dieser Ringe ist EUKLIDisch, da sich weder der ggT 1 von 2 und X in $\mathbb{Z}[X]$ noch der ggT 1 von X und Y in $k[X, Y]$ als Linearkombination der Ausgangselemente schreiben läßt.

§3: Der Euklidische Algorithmus für Polynome

Aus dem vorigen Paragraphen wissen wir, daß der Polynomring über einem Körper EUKLIDisch ist, so daß auch dort größte gemeinsame Teiler existieren und nach dem EUKLIDischen Algorithmus berechnet werden können. Wir wollen uns zunächst überlegen, daß größte gemeinsame Teiler auch in Polynomringen in mehreren Veränderlichen über $\mathbb{Z}$ oder einem Körper existieren und dann sehen, wie man diese berechnen kann.

a) Der Satz von Gauß

Für einen beliebigen Integritätsbereich R ist der Polynomring $R[X]$ im allgemeinen nicht EUKLIDisch. Falls wir allerdings R in einen Körper K einbetten können, sind wir in einem EUKLIDischen Ring $K[X]$ und können dort den EUKLIDischen Algorithmus anwenden. Der Satz von GAUSS sagt uns, wie Faktorzerlegungen in $R[X]$ und in $K[X]$ miteinander zusammenhängen.

Zunächst brauchen wir einen geeigneten Kandidaten für einen Körper K, in den wir R einbetten können; dies ist der sogenannte *Quotientenkörper.* Seine Konstruktion ist völlig analog zur Konstruktion der rationalen Zahlen aus den ganzen Zahlen:

Wir betrachten für einen Integritätsbereich R auf der Menge aller Paare (r, s) mit $r, s \in R$ und $s \neq 0$ die Äquivalenzrelation

$$(r, s) \sim (u, v) \Longleftrightarrow rv = us\,;$$

die Äquivalenzklasse von (r, s) bezeichnen wir als den *Bruch* $\frac{r}{s}$.

Verknüpfungen zwischen diesen Brüchen werden nach den üblichen Regeln der Bruchrechnung definiert:

$$\frac{r}{s} + \frac{u}{v} = \frac{rv + us}{sv} \quad \text{und} \quad \frac{r}{s} \cdot \frac{u}{v} = \frac{ru}{sv}\,.$$

Dies ist wohldefiniert, denn sind $(r, s) \sim (r', s')$ und $(u, v) \sim (u', v')$, so ist

$$\frac{r'}{s'} + \frac{u'}{v'} = \frac{r'v' + u's'}{s'v'} \quad \text{und} \quad \frac{r'}{s'} \cdot \frac{u'}{v'} = \frac{r'u'}{s'v'}$$

und $rs' = r's$ sowie $uv' = u'v$. Damit ist auch

$$\begin{aligned}(r'v' + u's') \cdot sv &= r'v'sv + u's'sv = r'svv' + u'vss' \\ &= rs'vv' + uv'ss' = (rv + uy)s'v'\end{aligned}$$

und $(r'u')(sv) = r'su'v = rs'uv' = (ru)(s'v')$, 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 $r, s \neq 0$ ist $\frac{s}{r}$ ein multiplikatives Inverses zu $\frac{r}{s}$, da $(rs, rs) \sim (1, 1)$.

Identifizieren wir schließlich ein Element $r \in R$ mit dem Bruch $\frac{r}{1}$, so können wir R in den Körper K einbetten.

Definition: Der so konstruierte Körper K heißt Quotientenkörper von R; in Zeichen $K = \text{Quot}\, R$.

Das Standardbeispiel ist natürlich $\mathbb{Q} = \text{Quot}\, \mathbb{Z}$, aber auch der Quotientenkörper $k(X) \underset{\text{def}}{=} \text{Quot}\, k[X]$ 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 Integritätsbereich definiert sind, sind die beiden folgenden Begriffe sehr wesentlich:

Definition: *a)* Der *Inhalt* eines Polynoms $f = a_n X^n + \cdots + a_0 \in R[X]$ ist der ggT $I(f)$ der Koeffizienten a_i.
b) f heißt *primitiv,* wenn die a_i zueinander teilerfremd sind.

Indem wir die sämtlichen Koeffizienten eines Polynoms durch deren 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 Integritätsbereich. Für zwei Polynome

$$f = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0$$

und

$$g = b_m X^m + b_{m-1} X^{m-1} + \cdots + b_1 X + b_0$$

aus $R[X]$ ist $I(fg) = I(f) \cdot I(g)$. 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 wieder primitiv ist. Sei

$$fg = c_{n+m}X^{n+m} + c_{n+m-1}X^{n+m-1} + \ldots + c_1X + c_0\,;$$

dann ist $c_r = \sum_{i,j \text{ mit } i+j=r} a_i b_j$.

Angenommen, diese Koeffizienten c_r haben einen gemeinsamen Teiler vom Betrag größer eins. Dann gibt es auch ein irreduzibles Element p, das alle Koeffizienten c_r teilt.

Insbesondere ist p ein Teiler von $c_0 = a_0b_0$; da p irreduzibel ist, muß somit 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, ist nicht jeder Koeffizient a_i durch p teilbar; ν sei der kleinste Index, so daß a_ν kein Vielfaches von p ist. Da auch g primitiv ist, gibt es auch einen kleinsten Index μ, 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; für jeden anderen Summanden a_ib_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. ∎

Korollar: Ein primitives Polynom $f \in R[X]$ ist genau dann irreduzibel in $R[X]$, wenn es in $K[X]$ irreduzibel ist. ∎

Satz von Gauß: R sei ein Integritätsbereich 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 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

$$g = cg^* \quad \text{und} \quad h = dh^* .$$

Dann ist $f = (cd)g^*h^*$, und nach dem Lemma ist g^*h^* ein primitives Polynom. Daher liegt $cd = I(f)$ in R, und wir können beispielsweise $\widetilde{g} = I(P)g^*$ und $\widetilde{h} = h^*$ setzen. ∎

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, und zeitweise auch den Witwenfond der Universität Göttingen. Seine hierbei gewonnene Erfahrung nutzte er für erfolgreiche Spekulationen mit Aktien.

Aus diesem Satz folgt induktiv sofort, daß dieselbe Aussage auf für Produkte von mehr als zwei Polynomen gilt. Damit folgt insbesondere

Satz: Der Polynomring über einem faktoriellen Ring R ist selbst 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 aus §2*a)* 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 selbst 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 für Polynome in mehreren Veränderlichen über $\mathbb{Z}$ oder über einem Körper größte gemeinsame Teiler existieren. Der Rest des Paragraphen wird sich damit beschäftigen, wie wir diese effizient berechnen können.

Eine mögliche Strategie folgt aus den obigen Sätzen:

Ist R ein Integritätsbereich und sind $f, g \in R[X]$ zwei Polynome, so schreiben wir $f = I(f) \cdot f^*$ und $g = I(g) \cdot g^*$. Wegen der eindeutigen Primzerlegung in $R[X]$ ist

$$\mathrm{ggT}(f, g) = \mathrm{ggT}\big(I(f), I(g)\big) \cdot \mathrm{ggT}(f^*, g^*) .$$

Für den ersten Faktor müssen wir wissen, wie man den ggT in R ausrechnet; den zweiten Faktor können wir in $K[X]$ berechnen und dann durch seinen primitiven Anteil ersetzen, denn der ggT zweier primitiver Polynome ist als Teiler dieser Polynome insbesondere selbst primitiv.

b) Ein erstes Beispiel

Betrachten wir die beiden Polynome

$$f = X^8 + X^6 - 3X^4 - 3X^3 + 8X^2 + 2X - 5$$

und

$$g = 3X^6 + 5X^4 - 4X^2 - 9X + 21$$

aus $\mathbb{Z}[X]$. Offensichtlich sind beide primitiv, also ist auch ihr ggT primitiv. Wir berechnen ihn zunächst in $\mathbb{Q}[X]$, wo uns der EUKLIDische Algorithmus zur Verfügung steht:

Division von f durch g führt auf den Quotienten $X^2/3 - 2/9$ und Divisionsrest

$$r_2 = -\frac{5}{9}X^4 + \frac{1}{9}X^2 - \frac{1}{3} .$$

Division von g durch r_2 ergibt den Divisionsrest

$$r_3 = -\frac{117}{25}X^2 - 9X + \frac{441}{25} ,$$

bei der Division von r_2 durch r_3 bleibt Rest

$$r_4 = \frac{233150}{6591}X - \frac{102500}{2197} ,$$

und bei der letzten Division verbleibt als Rest der ggT

$$r_5 = \frac{1288744821}{543589225} .$$

Der primitive Anteil dieses „Polynoms“ ist Eins, also folgt

$$\mathrm{ggT}(f, g) = 1 .$$

Leider folgte dies aber nur auf dem Umweg über sehr große Zwischenergebnisse. Wenn wir auch bei größeren Problemen noch effizient rechnen wollen, müssen wir einen Weg finden, um diese zu vermeiden.

Eine wichtige Beobachtung dazu ist folgende: Falls wir den EUKLIDischen Algorithmus statt über $\mathbb{Q}$ über einem endlichen Körper ausführen, kann es diese Explosion von Zwischenergebnissen nicht geben: Im Körper mit p Elementen wird jedes Element reprässentiert durch eine Zahl zwischen Null und $p-1$, und das gilt selbstverständlich auch für alle Zwischenergebnisse.

Hätten wir im obigen Beispiel etwa im Körper mit elf Elementen gerechnet, so wäre dort

$$f = X^8 + X^6 + 8X^4 + 8X^3 + 8X^2 + 2X + 6$$

und

$$g = 3X^6 + 5X^4 + 7X^2 + 2X + 10\,.$$

Division von f durch g führt auf den Quotienten $4X^2 + 1$ und Divisionsrest

$$r_2 = 8X^4 + 5X^2 + 7\,.$$

Division von g durch r_2 ergibt den Divisionsrest

$$r_3 = 5X^2 + 2X + 4\,,$$

bei der Division von r_2 durch r_3 bleibt Rest

$$r_4 = 10X + 10\,,$$

und bei der letzten Division verbleibt als Rest der ggT

$$r_5 = 7\,,$$

der in $\mathbb{F}_{11}$ natürlich eine Einheit ist, so daß die beiden Polynome auch dort teilerfremd sind.

Eine der beiden gebräuchlichen Strategien zur Berechnung des größten gemeinsamen Teilers zweier Polynome mit ganzzahligen Koeffizienten führt über den EUKLIDischen Algorithmus über endlichen Körpern, die andere (für die wir im Rahmen dieser Vorlesung keine Zeit haben) arbeitet mit sogenannten Subresultanten. Beide Methoden habe dieselbe asymptotische Laufzeit; die tatsächliche Laufzeit ist allerdings meist bei der modularen Methode deutlich besser als bei den Subresultantenalgorithmen.

c) Allgemeines zum Rechnen mit homomorphen Bildern

Modulare Methoden sind ein Spezialfall des Rechnens mit homomorphen Bildern, das wir in einem etwas anderen Zusammenhang auch später zur Bestimmung des ggTs zweier Polynome in mehreren Veränderlichen verwenden werden. Es lohnt sich daher, das Problem gleich abstrakt algebraisch zu formulieren:

Definition: *a)* Ein Homomorphismus $\varphi\colon R \to S$ von Ringen ist eine Abbildung, für die gilt

$$\varphi(r+s) = \varphi(r) + \varphi(s) \quad \text{und} \quad \varphi(r \cdot s) = \varphi(r) \cdot \varphi(s)$$

für alle $r, s \in \mathbb{R}$.
b) Der *Kern* eines Homomorphismus $\varphi\colon R \to S$ von Ringen ist die Menge aller $r \in R$ mit $\varphi(r) = 0$.
c) Eine Teilmenge $I \subseteq R$ heißt *Ideal* von R, in Zeichen $I \lhd R$, wenn gilt:
1.) I ist eine additive Untergruppe von R
2.) Für $r \in R$ und $s \in I$ ist auch rs ein Element von I.
d) Die Menge $(a) = \{ra | r \in R\}$ aller Vielfachen eines Elements $a \in R$ heißt das von a erzeugte *Hauptideal*.

Bei der modularen Berechnung des ggT verwenden wir für geeignete Primzahlen p den Homomorphismus

$$\varphi_p\colon \begin{cases} \mathbb{Z} \to \mathbb{F}_p \\ x \mapsto x \bmod p \end{cases};$$

sein Kern ist offensichtlich das von p erzeugte Hauptideal. Später werden wir oft auch *Einsetzungsmorphismen* betrachten, die ein Polynom an einer vorgegebenen Stelle auswerten: Für ein festes Element $a \in R$ eines Integritätsbereichs R betrachten wir hier den Homomorphismus

$$\varphi_a\colon \begin{cases} R[X] \to R \\ f \mapsto f(a) \end{cases}.$$

Sein Kern besteht aus allen Polynomen, die an der Stelle a verschwinden; das ist offenbar gerade das von $X - a$ erzeugte Hauptideal.

Es ist kein Zufall, daß es sich in beiden Fällen um Hauptideale handelt, denn allgemein gilt

Lemma: Jedes Ideal eines EUKLIDischen Rings ist ein Hauptideal.

Beweis: Das Nullideal ist ein Hauptideal, und für jedes andere Ideal I betrachten wir die Menge

$$M = \left\{ \nu(r) \;\middle|\; r \in I \setminus \{0\} \right\}.$$

Diese hat als Teilmenge von $\mathbb{N}_0$ ein minimales Element m; wir wählen ein $a \in I$ mit $\nu(a) = m$.

Für jedes Element $x \in I$ können wir x mit Rest durch a dividieren; das Ergebnis sei q Rest r. Falls $r = 0$ liegt $x = qa$ in (a); andernfalls ist $\nu(r) < \nu(a)$. Letzteres ist aber nicht möglich, denn $r = x - qa$ liegt in I, so daß $\nu(r) \geq \nu(a)$ sein müßte. Somit ist $I = (a)$. ∎

Ideale spielen bei Ringen genau dieselbe Rolle wie Normalteiler bei Gruppen, d.h. es gilt:

Lemma: Zu einem Ring R und einer Teilmenge $I \subseteq R$ gibt es genau dann einen Homomorphismus $\varphi\colon R \to S$ mit Kern I, wenn I ein Ideal von R ist.

Beweis: Ist I Kern des Homomorphismus $\varphi\colon R \to S$, so ist I natürlich eine additive Untergruppe von R, da φ insbesondere auch ein Gruppenhomomorphismus ist. Für $r \in R$ und $s \in I$ ist $\varphi(s) = 0$, also auch $\varphi(rs) = \varphi(r)\varphi(s) = 0$. Somit ist I ein Ideal.

Ist umgekehrt I ein Ideal von R, so können wir auf R eine Äquivalenzrelation definieren durch $r \sim s$ genau dann, wenn $r - s \in I$. Die Äquivalenzklasse von r bezeichnen wir mit $\overline{r}$, die Menge aller Äquivalenzklassen mit $\overline{R} = R/I$.

Für $r, r', s, s' \in R$ mit $r \sim r'$ und $s \sim s'$ liegt mit $r - r'$ und $s - s'$ auch $r + s - r' - s'$ in I, d.h. $\overline{r+s} = \overline{r} + \overline{s}$. Genauso ist auch $\overline{r \cdot s} = \overline{r} \cdot \overline{s}$, denn ist $r' = r + i$ und $s' = s + j$ mit $i, j \in I$, so ist

$$r' \cdot s' = (r+i)(s+j) = rs + is + rj + ij\,,$$

und wegen der Idealeigenschaft von i liegen rs, is und rj allesamt in I. Somit ist $\overline{R}$ ein Ring, und die Abbildung $\varphi\colon R \to \overline{R}$, die jedes $r \in R$ auf seine Äquivalenzklasse $\overline{r}$ in $\overline{R}$ abbildet, ist ein Homomorphismus, dessen Kern natürlich I ist. ∎

Die Ideale haben ihren Namen von KUMMER, der sie als *ideale Zahlen* betrachtete: KUMMER glaubte zunächst, er habe einen Beweis der FERMAT-Vermutung gefunden, allerdings war er davon ausgegangen, daß der Ring $\mathbb{Z}[\zeta_p]$, wobei p eine primitive p-te Einheitswurzel bezeichnet, faktoriell ist. Dies ist zwar für unendlich viele Primzahlen p der Fall, aber eben nicht für alle. KUMMER konnte aber zeigen, daß es auf dem Niveau der Ideale eine eindeutige Primzerlegung gibt – nur reichte das leider nicht aus, um seinen Beweis auch für die Primzahlen zu retten für die $\mathbb{Z}[\zeta_p]$ nicht faktoriell ist.

Natürlich definiert jeder Homomorphismus $\varphi\colon R \to S$ einen Homomorphismus

$$\varphi\colon \begin{cases} R[X] \to S[X] \\ a_n X^n + \cdots + a_0 \mapsto \varphi(a_n)X^n + \cdots + \varphi(a_0) \end{cases}$$

zwischen den Polynomringen darüber; die Grundidee beim Rechnen mit homomorphen Bildern besteht darin, ein Problem für Polynome über S auf das entsprechende Problem über S zurückzuführen.

d) Zusammenhang zwischen ggT und modularem ggT

Wir gehen aus von zwei Polynomen

$$f = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0$$

und

$$g = b_m X^m + b_{m-1} X^{m-1} + \cdots + b_1 X + b_0$$

mit Koeffizienten aus R und wollen die Berechnung des ggT von f und g zurückführen auf die des ggT von $\varphi(f)$ und $\varphi(g)$. Um zu sehen, was dabei zu beachten ist, betrachten wir einen Teiler

$$h = c_r X^r + c_{r-1} X^{r-1} + \cdots + c_1 X + c_0$$

von f und g. Dazu gibt es Kofaktoren $F, G \in R[X]$, für die $f = hF$ und $g = hG$ ist; da φ auch auf dem Niveau der Polynomringe ein Homomorphismus ist, folgt

$$\varphi(f) = \varphi(h)\varphi(F) \quad \text{und} \quad \varphi(g) = \varphi(h)\varphi(G)\,.$$

Damit ist also auch $\varphi(h)$ ein gemeinsamer Teiler von $\varphi(f)$ und $\varphi(g)$.

Falls der führende Koeffizient von h im Kern von φ liegt (und nur dann) ist allerdings der Grad von $\varphi(h)$ kleiner als der von h. Als Beispiel dazu können wir die beiden Polynome $p = 3X^6 + 3X^5 + X + 1$ und $q = 3X^6 - 3X^5 + X - 1$ betrachten. Hier ist

$$p : q = 1 \quad \text{Rest } 6X^5 + 2$$

und

$$q : (6X^5 + 2) = \frac{X}{2} - \frac{1}{2} \quad \text{Rest } 0\,,$$

also ist das primitive Polynom $h = 3X^5 + 1$ zu $6X^5 + 2$ ein ggT von p und q in $\mathbb{Z}[X]$. Die Polynome $p \bmod 3 = X + 1$ und $q \bmod 3 = X - 1$ sind aber teilerfremd, und natürlich ist auch $h \bmod 3 = 1$.

Andererseits kann es auch gemeinsame Teiler von $\varphi(f)$ und $\varphi(g)$ geben, die nicht von einem gemeinsamen Teiler von f und g kommen. Für die im vorigen Paragraphen betrachteten Polynome f und g passiert dies beispielsweise modulo sieben: Dort ist

$$f = X^8 + X^6 + 4X^4 + 4X^3 + X^2 + 2X + 2$$

und

$$g = 3X^6 + 5X^4 + 3X^2 + 5X\,.$$

Division von f durch g in $\mathbb{F}_7[X]$ führt auf den Divisionsrest

$$r_2 = X^4 + 4X^2 + 2\,,$$

Division von g durch r_2 ergibt den Divisionsrest

$$r_3 = 4X^2 + 5X\,,$$

und bei der Division von r_2 durch r_3 bleibt Rest

$$r_4 = 3X + 2\,,$$

was über $\mathbb{F}_7$ ein Teiler von r_3 ist. Somit ist hier der ggT gleich dem linearen Polynom $3X + 2$, während f und g in $\mathbb{Z}[X]$ teilerfremd sind.

Das erste Problem, daß der ggT von $\varphi(f)$ und $\varphi(g)$ kleineren Grad hat als der von f und g, läßt sich leicht vermeiden: Liegt nämlich der

führende Koeffizient von $h = \text{ggT}(f, g)$ im Kern von φ, so müssen auch die führenden Koeffizienten von f und g dort liegen; denn der führende Koeffizient eines Vielfachen von h muß ein Vielfaches des führenden Koeffizienten von h sein. Um dieses Problem zu vermeiden, müssen wir somit einfach alle Homomorphismen ausschließen, die die führenden Koeffizienten von f und g auf Null abbilden. Falls wir als Homomorphismen die Abbildungen $\varphi_p\colon \mathbb{Z} \to \mathbb{F}_p$ verwenden, müssen wir also alle Primzahlen vermeiden, die beide führende Koeffizienten teilen.

Schwieriger ist es mit dem Problem, daß der ggT von $\varphi(f)$ und $\varphi(g)$ größeren Grad haben kann als der von f und g. Um dieses Problem in den Griff zu bekommen, müssen wir zunächst untersuchen, wann zwei Polynome über einem Körper überhaupt einen (nichtkonstanten) gemeinsamen Teiler haben.

e) Die Resultante

Wir gehen aus von einem faktoriellen Ring R mit Quotientenkörper K und zwei Polynomen

$$f = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0$$

und

$$g = b_m X^m + b_{m-1} X^{m-1} + \cdots + b_1 X + b_0$$

aus $R[X]$. Ist $h \in R[X]$ ein Polynom positiven Grades, das sowohl f als auch g teilt, ist

$$\frac{fg}{h} = \frac{f}{h} \cdot g = \frac{q}{h} \cdot f$$

ein gemeinsames Vielfaches von f und q, dessen Grad

$$\deg f + \deg g - \deg h = n + m - \deg h$$

echt kleiner ist als $\deg f + \deg g = n + m$.

Haben umgekehrt f und g ein gemeinsames Vielfaches, dessen Grad kleiner ist als $n + m$, so hat auch ihr kleinstes gemeinsames Vielfaches p einen kleineren Grad als $n + m$. (Ein kleinstes gemeinsames Vielfaches existiert, da wir in Abschnitt *a)* gesehen haben, daß mit R auch $R[X]$ faktoriell ist.)

Zu p gibt es einerseits Polynome $u, v \in R[X]$, für die $p = uf = vg$ ist, andererseits ist p als *kleinstes* gemeinsame Vielfache von f und g Teiler von fg, es gibt also ein Polynom $h \in R[X]$ mit $fg = ph$. Für dieses ist

$$hv = \frac{fg}{p} \cdot v = f \cdot \frac{vg}{p} = f \quad \text{und} \quad hu = \frac{fg}{p} \cdot u = g \cdot \frac{uf}{p} = g\,,$$

es teilt also sowohl f als auch g und sein Grad $n + m - \deg p$ ist positiv. Damit ist gezeigt:

Lemma: Zwei Polynome $f, g \in R[X]$ haben genau dann einen gemeinsamen Teiler positiven Grades, wenn es nichtverschwindende Polynome $u, v \in R[X]$ gibt mit $\deg u < \deg g$ und $\deg v < \deg f$, so daß $uf = vg$ ist. ∎

Diese Bedingung schreiben wir um in ein lineares Gleichungssystem für die Koeffizienten von u und v: Da $\deg u < \deg g = m$ ist und $\deg v < \deg f = n$, lassen sich die beiden Polynome schreiben als

$$u = u_{m-1}X^{m-1} + u_{m-2}X^{m-2} + \ldots + u_1X + u_0$$

und

$$v = v_{n-1}X^{n-1} + v_{n-2}X^{n-2} + \ldots + v_1X + v_0\,,$$

und im Polynom uf hat die Potenz X^r den Koeffizienten

$$\sum_{i,j \text{ mit } i+j=r} a_i u_j\,,$$

während sie im Polynom vg den Koeffizienten

$$\sum_{i,j \text{ mit } i+j=r} b_i v_j$$

hat. f und g haben daher genau dann einen gemeinsamen Teiler positiven Grades, wenn es nicht allesamt verschwindende Körperelemente $u_0, \ldots, u_{m-1}$ und $v_0, \ldots, v_{n-1}$ gibt, so daß

$$\sum_{i,j \text{ mit } i+j=r} a_i u_j - \sum_{i,j \text{ mit } i+j=r} b_i v_j = 0 \qquad \text{für } r = 0, \ldots, n+m-1$$

ist. Dies ist ein homogenes lineares Gleichungssystem aus $n + m$ Gleichungen für die $n + m$ Unbekannten $u_0, \ldots, u_{m-1}$ und $v_0, \ldots, v_{n-1}$; es

hat genau dann eine nichttriviale Lösung, falls seine Matrix singulär ist, falls also deren Determinante verschwindet.

Ausgeschrieben wird dieses Gleichungssystem, wenn wir mit dem Koeffizienten von X^{m+n-1} anfangen, zu

$$\begin{gathered}
a_n u_{m-1} - b_m v_{n-1} = 0 \\
a_{n-1}u_{m-1} + a_n u_{m-2} - b_{m-1}v_{n-1} - b_m v_{n-2} = 0 \\
a_{n-2}u_{m-1} + a_{n-1}u_{m-2} + a_n u_{m-3} - b_{m-2}v_{n-1} - b_{m-1}v_{n-2} b_m v_{n-3} = 0 \\
\ldots \\
a_0u_2 + a_1u_1 + a_2u_0 - b_0v_2 - b_1v_1 - b_2v_0 = 0 \\
a_0u_1 + a_1u_0 - b_0v_1 - b_1v_0 = 0 \\
a_0u_0 - b_0v_0 = 0
\end{gathered}$$

Natürlich ändert sich nichts an der nichttrivialen Lösbarkeit oder Unlösbarkeit dieses Gleichungssystems, wenn wir anstelle der Variablen v_j die Variablen $-v_j$ betrachten, womit alle Minuszeichen im obigen Gleichungssystem zu Pluszeichen werden; außerdem hat es sich – der größeren Übersichtlichkeit wegen – eingebürgert, die Transponierte der Matrix des Gleichungssystems zu betrachten. Dies führt auf die Determinante

$$\begin{vmatrix}
a_n & a_{n-1} & a_{n-2} & \ldots & a_1 & a_0 & 0 & 0 & \ldots & 0 \\
0 & a_n & a_{n-1} & \ldots & a_2 & a_1 & a_0 & 0 & \ldots & 0 \\
0 & 0 & a_n & \ldots & a_3 & a_2 & a_1 & a_0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & a_n & a_{n-1} & a_{n-2} & a_{n-3} & \ldots & a_0 \\
b_m & b_{m-1} & b_{m-2} & \ldots & b_2 & b_1 & b_0 & 0 & \ldots & 0 \\
0 & b_m & b_{m-1} & \ldots & b_3 & b_2 & b_1 & b_0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0 & b_m & b_{m-1} & b_{m-2} & \ldots & b_0
\end{vmatrix}$$

Die Matrix dazu bezeichnet man als SYLVESTER-Matrix, ihre Determinante als *Resultante* $\operatorname{Res}(f, g)$ der beiden Polynome f und g. Falls man, etwa bei späteren Anwendungen auf Polynome mehrerer Veränderlicher,

auf die Variable X hinweisen möchte, schreibt man auch $\mathrm{Res}_X(f,g)$. Die Resultante ist offensichtlich ein Polynom in den a_i und den b_j, das genau dann verschwindet, wenn f und g einen gemeinsamen Faktor haben.

JAMES JOSEPH SYLVESTER (1814–1897) wurde geboren als JAMES JOSEPH; erst als sein Bruder nach USA auswanderte und dazu einen dreiteiligen Namen brauchte, erweiterte er aus Solidarität auch seinem Namen. 1837 bestand er das berüchtigte Tripos-Examen der Universität Cambridge als Zweitbester, bekam aber keinen akademischen Abschluß, da er als Jude den dazu vorgeschriebenen Eid auf die 39 Glaubensartikel der Church of England nicht leisten konnte. Trotzdem wurde er Professor am University College in London; seine akademischen Grade bekam er erst 1841 aus Dublin, wo die Vorschriften gerade mit Rücksichtauf die Katholiken geändert worden waren. Während seiner weiteren Tätigkeit an sowohl amerikanischen als auch englischen Universitäten beschäftigte er sich mit Matrizen, fand die Diskriminante kubischer Gleichungen und entwickelte auch die allgemeine Theorie der Diskriminanten. In seiner Zeit an der Johns Hopkins University in Baltimore gründete er das American Journal of Mathematics, das auch heute noch mit die wichtigste mathematische Zeitschrift Amerikas ist.

f) Die Landau-Mignotte-Schranke

$f \in \mathbb{Z}[X]$ sei ein bekanntes Polynom mit ganzzahligen Koeffizienten, und $g \in \mathbb{Z}[X]$ sei ein (im allgemeinen noch unbekannter) Teiler von f. Wir wollen eine obere Schranke für die Koeffizienten von g finden.

Dazu ordnen wir zunächst jedem Polynom

$$f = \sum_{k=0}^{d} a_k X^k$$

mit komplexen Koeffizienten a_k eine Reihe von Maßzahlen für die Größe der Koeffizienten zu: Am wichtigsten ist natürlich

$$H(f) = \max_{k=0}^{d} |a_k| \ ,$$

die sogenannte *Höhe* des Polynoms. Unser Ziel ist es, für ein gegebenes Polynom $f \in \mathbb{Z}[X]$ die Höhe seiner Teiler abzuschätzen. Auf dem Weg

zu dieser Abschätzung wird es sich als nützlich erweisen, zunächst Polynome mit beliebigen *komplexen* Koeffizienten zu betrachtet; für diese können wir genau wie oben ihre Höhe definieren. Zusätzlich werden werden uns noch eine Reihe anderer Größen nützlich sein, darunter die L^1- und die L^2-Norm

$$\|f\|_1 = \sum_{k=0}^{d} |a_k| \quad \text{und} \quad \|f\|_2 = \sqrt{\sum_{k=0}^{d} a_k \overline{a_k}} = \sqrt{\sum_{k=0}^{d} |a_k|^2} \ .$$

Für die drei bislang definierten Größen gilt

Lemma 1: $H(f) \le \|f\|_2 \le \|f\|_1 \le \sqrt{d+1}\ \|f\|_2 \le (d+1)H(f)$

Beweis: Ist a_ν der betragsgrößte Koeffizient von f, so ist $H(f) = |a_\nu| = \sqrt{|a_\nu|^2}$ offensichtlich kleiner oder gleich $\|f\|_2$. Dies wiederum ist nach der Dreiecksungleichung kleiner oder gleich $\|f\|_1$, denn schreiben wir in $\mathbb{C}^{d+1}$ den Koeffizientenvektor von f als Summe von Vielfachen der Basisvektoren, d.h.

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{pmatrix} = \begin{pmatrix} a_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ a_d \end{pmatrix} \ ,$$

so steht links ein Vektor der Länge $\|f\|_2$, und rechts stehen Vektoren, deren Längen sich zu $\|f\|_1$ summieren.

Das nächste Ungleichheitszeichen ist die CAUCHY-SCHWARZsche Ungleichung, angewandt auf die Vektoren

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \ ,$$

und das letzte schließlich ist klar, denn

$$\|f\|_2 = \sqrt{\sum_{j=0}^{d} |a_j|^2} \le \sqrt{\sum_{j=0}^{d} |a_\nu|^2} = \sqrt{d+1}\,|a_\nu| = \sqrt{d+1}\,H(f) \ .$$

■

Es ist alles andere als offensichtlich, wie sich die drei bislang definierten Maßzahlen für einen Teiler eines Polynoms durch die entsprechende Größe für das Polynom selbst abschätzen lassen, denn über die Koeffizienten eines Teilers können wir leider nur sehr wenig sagen. Über seine Nullstellen allerdings schon: Die Nullstellen eines Teilers bilden natürlich eine Teilmenge der Nullstellen des Polynoms. Also sollten wir versuchen, auch die Nullstellen ins Spiel zu bringen; den Zusammenhang zwischen Nullstellen und Koeffizienten liefern die sogenannten *elementarsymmetrische* Funktionen $\psi_k(z_1,\ldots,z_n)$, die definiert sind als Summe Summe aller möglicher Produkte von k Werten z_j mit verschiedenen Indizes.

Wurzelsatz von Viète: Hat das Polynom $\sum_{k=0}^{d} a_k X^k$ die Nullstellen $z_1,\ldots,z_d$, so ist $a_k = (-1)^{d-k} a_d \psi_k(z_1,\ldots,z_n)$, wobei $\psi_k(z_1,\ldots,z_n)$.

(Die Funktionen ψ_k bezeichnet man in der Algebra als *elementarsymmetrische Funktionen;* man kann zeigen, daß sich jedes in n Variablen symmetrische Polynom als Polynom in diesen elementarsymmetrischen Funktionen schreiben läßt.)

Der *Beweis* des Satzes von VIEÈTE ist fast trivial: Man muß einfach $\prod_{j=1}^{d}(X - z_j)$ ausmultiplizieren. Dabei entstehen 2^d Summanden, die jeweils Produkte von d Faktoren sind: Aus jedem der Faktoren $(X - z_j)$ wird entweder das X genommen oder $(-z_j)$. Die Summanden, in denen X in der k-ten Potenz steht, haben somit $n - k$ Faktoren der Form $(-z_j)$, d.h. insgesamt steht im Produkt vor X^k der Term $(-1)^{n-k}\psi_{n-k}(z_1,\ldots,z_n)$. ∎

(Ohne Beweis sei an den *Fundamentalsatz der Algebra* erinnert, wonach sich jedes Polynom über den komplexen Zahlen als Produkt von Linearfaktoren schreiben läßt.)

FRANÇOIS VIÈTE (1540–1603) . . .

Um die Koeffizienten eines Polynoms durch die Nullstellen abschätzen zu können, brauchen wir obere Schranken für die Beträge der Produkte aus k Nullstellen. Natürlich ist jedes solche Produkt ein Teilprodukt des Produkts $z_1 \cdots z_d$ aller Nullstellen, aber das führt zu keiner Abschätzung, da unter den fehlenden Nullstellen auch welche sein können, deren Betrag kleiner als eins ist. Um eine obere Schranke für den Betrag zu kommen, müssen wir diese Nullstellen im Produkt $z_1 \cdots z_d$ durch Einsen ersetzen; dann können wir sicher sein, daß kein Produkt von k Nullstellen einen größeren Betrag hat als das so modifizierte Produkt.

Diese Überlegungen führen auf die

Definition: Das *Maß* $\mu(f)$ eines nichtkonstanten Polynoms

$$f = a_d \prod_{j=1}^{d}(X - z_j)$$

ist das Produkt der Beträge aller Nullstellen von Betrag größer eins mal dem Betrag des führenden Koeffizienten a_d von f:

$$\mu(f) = |a_d| \prod_{j=1}^{d} \max\bigl((1, |z_j|\bigr) .$$

Dieses Maß ist im allgemeinen nur schwer explizit berechenbar, da man dazu die sämtlichen Nullstellen des Polynoms explizit kennen muß. Es

hat aber den großen Vorteil, daß für zwei Polynome f und g trivialerweise gilt

$$\mu(f \cdot g) = \mu(f) \cdot \mu(g)\,.$$

Auch können wir es nach dem Wurzelsatz von VIÈTE leicht für eine Abschätzung der Koeffizienten verwenden:

$$a_k = (-1)^k a_d\, \psi_k(z_1, \ldots, z_d)$$

ist eine Summe von Termen, von denen jeder einzelne höchstens den Betrag $\mu(f)$ hat. Die Anzahl dieser Terme ist die Anzahl von Möglichkeiten, aus d Indizes eine k-elementige Teilmenge auszuwählen, also $\binom{d}{k}$. Damit folgt

Lemma 2: Für ein nichtkonstantes Polynom $f = \sum_{k=0}^{d} a_d X^d$ ist

$$|a_k| \le \binom{d}{k} \mu(f)\,.$$

∎

Der größte unter den Binomialkoeffizienten $\binom{d}{k}$ ist bekanntlich der mittlere *bzw.* sind die beiden mittleren, und die Summe aller Binomialkoeffizienten $\binom{d}{k}$ ist, wie die binomische Formel für $(1+1)^d$ zeigt, gleich 2^d. Damit folgt

Korollar: Für ein nichtkonstantes Polynom $f \in \mathbb{C}[X]$ ist

$$H(f) \le \binom{d}{[d/2]} \mu(f) \quad \text{und} \quad H(f) \le \|f\|_1 \le 2^d \mu(d)\,.$$

∎

Um umgekehrt das Maß durch eine Norm abschätzen zu können, zeigen wir zunächst

Lemma 3: Für jedes Polynom $f \in \mathbb{C}[X]$ und jede komplexe Zahl z ist

$$\|(X - z)f\|_2 = \|(\overline{z}X - 1)f\|_2\,.$$

Beweis durch explizite Berechnung der beiden Seiten: Sei $f = \sum_{k=0}^{d} X^d$.

Dann ist $(X - z)f = a_d X^{d+1} + \sum_{k=1}^{d}(za_k - a_{k-1})X^k - a_0 z$ und

$\|(X - z)f\|_2^2$ als Summe aller Koeffizientenquadrate errechnet sich zu

$$a_d\overline{a_d} + \sum_{k=1}^{d}(za_k - a_{k-1})\,\overline{(za_k - a_{k-1})} + a_0\overline{a_0}$$

$$= |a_d|^2 + \sum_{k=1}^{d}\left(|a_k|^2\,|z|^2 - 2\,\mathfrak{Re}(za_k\overline{a_{k-1}}) + \left|a_{k-1}\right|^2\right) + |a_0|^2$$

$$= (1 + |z|^2)\sum_{k=0}^{d}|a_k|^2 - 2\sum_{k=1}^{d}\mathfrak{Re}(za_k\overline{a}_{k-1})\,.$$

Entsprechend ist $(\overline{z}X - 1)f = a_d\overline{z}X^{d+1} + \sum_{k=1}^{d}(\overline{z}a_{k-1} - a_k)X^k - a_0$ und $\|(\overline{z}X - 1)f\|_2^2$ wird zu

$$a_d\overline{z}\cdot\overline{a_d}z + \sum_{k=1}^{d}(\overline{z}a_{k-1} - a_k)(z\overline{a_{k-1}} - \overline{a_k}) + a_0\overline{a_0}$$

$$= |za_d|^2 + \sum_{k=1}^{d}\left(\left|za_{k-1}\right|^2 - 2\,\mathfrak{Re}(za_k\overline{a_{k-1}}) + |a_k|^2\right) + |a_0|^2$$

$$= (1 + |z|^2)\sum_{k=0}^{d}|a_k|^2 - 2\sum_{k=1}^{d}\mathfrak{Re}(za_k\overline{a}_{k-1})\,.$$

Also stimmen die beiden Normen überein. ∎

Für das Polynom $f = a_d\prod_{j=1}^{d}(X - z_j)$ bedeutet dies, daß wir den Faktor $(X - z_j)$ durch $(\overline{z}_j X - 1)$ ersetzen können, ohne daß sich die L^2-Norm ändert. Wenden wir dies an auf alle Faktoren $(X - z_j)$, für die $\left|z_j\right| > 1$ ist, erhalten wir ein Polynom, dessen sämtliche Nullstellen Betrag kleiner oder gleich eins haben, denn $\overline{z}_j X - 1$ verschwindet für $X = 1/\overline{z}_j$, was für $|z|_j > 1$ einen Betrag kleiner Eins hat. Das Maß des modifizierten Polynoms ist also gleich dem Betrag des führenden Koeffizienten, und dieser wiederum ist natürlich kleiner oder gleich der L^2-Norm. Andererseits ist das Maß des modifizierten Polynoms

gleich dem des ursprünglichen, denn für jeden Faktor $(X - z_j)$ wird der führende Koeffizient bei der Modifikation mit $\overline{z}_j$ multipliziert, was denselben Betrag hat wie z_j. Damit folgt:

Lemma 4: Für ein nichtkonstantes Polynom $f \in \mathbb{C}[X]$ ist $\mu(f) \leq \|f\|_2$.

■

Nach diesen Vorbereitungen können wir uns an die Abschätzung der Koeffizienten eines Teilers machen. Sei dazu

$$g = \sum_{j=0}^{e} b_j X^j \quad \text{Teiler von} \quad f = \sum_{i=0}^{d} a_i X^i .$$

Da jede Nullstelle von g auch Nullstelle von f ist, lassen sich die Maße der beiden Polynome leicht vergleichen:

$$\mu(g) \leq \left|\frac{b_e}{a_d}\right| \cdot \mu(f) .$$

Kombinieren wir dies mit dem Korollar zu Lemma 2 und mit Lemma 4, erhalten wir die LANDAU-MIGNOTTE-Schranke:

$$H(g) \leq \binom{e}{[e/2]} \left|\frac{b_e}{a_d}\right| \|f\|_2 \quad \text{und} \quad \|g\|_1 \leq 2^e \left|\frac{b_e}{a_d}\right| \|f\|_2 .$$

Der ggT zweier Polynome f und g muß diese Abschätzung für beide Polynome erfüllen, allerdings kennen wir *a priori* weder den Grad noch den führenden Koeffizienten des ggT. Falls wir Polynome mit ganzzahligen Koeffizienten betrachten und einen ggT in $\mathbb{Z}[X]$ suchen, wissen wir nur, daß sein führender Koeffizient die führenden Koeffizienten sowohl von f als auch von g teilen muß, und daß sein Grad natürlich weden den von f noch den von g übersteigen kann. Damit erhalten wir die LANDAU-MIGNOTTE-Schranke für den ggT zweier Polynome: Schreiben wir f und g wie oben, so ist für $f, g \in \mathbb{Z}[X]$

$$H\big(\mathrm{ggT}(f,g)\big) \leq \|\mathrm{ggT}(f,g)\|_1 \leq \mathrm{LM}(f,g) \underset{\text{def}}{=} 2^{\min(d,e)} \mathrm{ggT}(a_d, b_e) \min\left(\frac{\|f\|_2}{|a_d|}, \frac{\|g\|_2}{|b_e|}\right) .$$

EDMUND GEORG HERMANN LANDAU (1877–1938) wurde in Berlin geboren und studierte an der dortigen Universität, wo er auch von 1899 bis 1909 lehrte. Dann bekam er einen Ruf an die damals führende deutsche Mathematikfakultät in Göttingen. 1933 verlor er seinen dortigen Lehrstuhl, denn die Studenten boykotierten seine Vorlesungen, da sie meinten, sie könnten Mathematik nur von einem Professor ihrer eigenen Rasse lernen. LANDAUS zahlreiche Publikationen beschäftigen sich vor allem mit der Zahlentheorie, über die er auch ein bedeutendes Lehrbuch schrieb. Sehr bekannt sind insbesondere seine Arbeiten über Primzahlverteilung.

MAURICE MIGNOTTE arbeitet am Institut de Recherche Mathématique Avancée der Universität Straßburg; sein Hauptforschungsgebiet sind diophantische Gleichungen. Er ist Autor mehrerer Lehrbücher, unter anderem aus dem Gebiet der Computeralgebra.

g) Die modulare Berechnung des ggT

Der Algorithmus zur modularen Berechnung des ggT zweier Polynome $f, g \in \mathbb{Z}[X]$ mit ganzzahligen Koeffizienten geht nun folgendermaßen:

1. Schritt: Berechne die LANDAU-MIGNOTTE-Schranke $\mathrm{LM}(f,g)$ und setze $M = 2\,\mathrm{LM}(f,g) + 1$. Außerdem wird $\mathcal{P} = \emptyset$ gesetzt, d.h. die Menge der bereits betrachteten Primzahlen ist noch leer.

Da der Betrag eines jeden Koeffizienten des ggT höchstens gleich $\mathrm{LM}(f,g)$ ist, kennen wie die Koeffizienten in $\mathbb{Z}$, sobald wir sie modulo M kennen.

2. Schritt: Wähle eine zufällige Primzahl $p \notin \mathcal{P}$, die weder den führenden Koeffizienten von f noch den von g teilt, und berechne in $\mathbb{F}_p[X]$ den ggT von f mod p und g mod p. Falls dieser gleich eins ist, endet der Algorithmus und $\mathrm{ggT}(f,g) = 1$. Andernfalls wird $N = p$ gesetzt und $h \in \mathbb{Z}[X]$ sei ein Polynom, dessen Reduktion modulo p gleich dem in $\mathbb{F}_p[X]$ berechneten ggT ist. Außerdem ersetze man $\mathcal{P}$ durch $\mathcal{P} \cup \{p\}$.

(Die Zahl N soll jeweils so gewählt werden, daß wir in jeder Phase des Algorithmus die Koeffizienten des ggT modulo N kennen oder zumindest zu kennen glauben. $N = 1$ bedeutet, daß wir nichts darüber wissen und uns dessen auch bewußt sind.)

3. Schritt: Falls $N \geq M$ ist, ändere man die Koeffizienten von h modulo N nötigenfalls so ab, daß ihre Beträge höchstens gleich LM(f, g) sind. Falls das nicht möglich ist, zurück zum zweiten Schritt. Andernfalls wird überprüft, ob h sowohl f also auch g teilt; falls ja ist h der gesuchte ggT und der Algorithmus endet; andernfalls geht es ebenfalls zurück zum zweiten Schritt.

4. Schritt: Im Fall $N < M$ wähle man eine zufällige Primzahl $p \notin \mathcal{P}$, die weder den führenden Koeffizienten von f noch den von g teilt, ersetze $\mathcal{P}$ durch $\mathcal{P} \cup \{p\}$ und berechne in $\mathbb{F}_p[X]$ den ggT von f mod p und g mod p. Falls dieser gleich eins ist, endet der Algorithmus und ggT$(f, g) = 1$. Falls sein Grad größer als der von h ist, wiederhole man den vierten Schritt, falls er kleiner ist, setze man $N = p$ und $\mathcal{P} = \{p\}$; alsdann geht es weiter mit dem dritten Schritt. Sind schließlich die beiden Grade gleich, so konstruiere man nach dem chinesischen Restesatz ein Polynom, das modulo N gleich h ist und modulo p gleich dem gerade berechneten ggT. N wird ersetzt durch Np, und es geht weiter mit dem dritten Schritt.

Der Algorithmus muß enden, da es nur endlich viele schlechte Primzahlen p gibt, für die der in $\mathbb{F}_p[X]$ berechnete ggT nicht einfach die Reduktion von ggT(f, g) modulo p ist, und nach endlich vielen Durchläufen sind genügend viele davon zusammengekommen, daß ihr Produkt die Zahl M übersteigt. Da der ggT in $\mathbb{F}_p[X]$ für Primzahlen, die keinen der führenden Koeffizienten teilen, höchstens höheren Grad als ggT(f, g) haben kann, ist auch klar, daß der Algorithmus mit einem korrekten Ergebnis abbricht.

h) Die Berechnung der Resultante

Angenommen, wir wollen die Resultante zweier Polynome der Grade dreißig und vierzig bestimmen. Das ist die Determinante einer 70×70-Matrix, und eine solche Determinante hat nach LAGRANGE $70!$ Summanden; das sind geringfügig mehr als 10^{100}. So viele Rechenoperationen sind weit jenseits der Möglichkeiten selbst der besten heutigen Supercomputer.

Tatsächlich verwendet natürlich niemand den Entwicklungssatz von LAGRANGE um eine Determinante zu berechnen – außer vielleicht bei

einigen kleineren Spielzeugdeterminanten in Mathematikklausuren. In allen anderen Fällen wird man die Matrix durch Zeilen- und/oder Spaltenoperationen auf Dreiecksform bringen und dann die Determinante einfach als Produkt der Diagonaleinträge berechnen. Das dauert für die SYLVESTER-Matrix zweier Polynome der Grade dreißig und vierzig auf heutigen Computern weniger als eine halbe Minute.

Stellt man allerdings keine Matrix auf, sondern verlangt von einem Computeralgebrasystem einfach, daß es die Resultante der beiden Polynome berechnen soll, hat man das Ergebnis nach weniger als einem Zehntel der Zeit. Einer der Schlüssel dazu ist wieder der EUKLIDische Algorithmus.

Angenommen, wir haben zwei Polynome f, g in einer Variablen X über einem Körper k:

$$f = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0 \quad \text{und}$$
$$g = b_m X^m + b_{m-1} X^{m-1} + \cdots + b_1 X + b_0 \quad \text{mit} \quad n \leq m\,.$$

Falls $f = a_0$ konstant ist, also $n = 0$, gibt es in der SYLVESTER-Matrix null Zeilen aus Koeffizienten von g und m Zeilen aus Koeffizienten von f; die Matrix ist also einfach a_0 mal der $m \times m$-Einheitsmatrix und die Resultante als ihre Determinante ist a_0^m.

Andernfalls dividieren wir g durch f und erhalten einen Rest h:

$$g : f = q \text{ Rest } h \quad \text{oder} \quad h = g - qf\,.$$

Der zentrale Punkt beim EUKLIDischen Algorithmus ist, daß die gemeinsamen Teiler von f und g genau dieselben sind wie die von f und h. Insbesondere haben also f und g genau dann einen gemeinsamen Teiler von positivem Grad, wenn f und h einen haben, d.h. $\mathrm{Res}_x(f, g)$ verschwindet genau dann, wenn $\mathrm{Res}_x(f, h)$ verschwindet. Damit sollte es also einen Zusammenhang zwischen den beiden Resultanten geben, und den können wir zur Berechnung von $\mathrm{Res}_x(f, g)$ ausnützen, denn natürlich ist $\mathrm{Res}_x(f, h)$ kleiner und einfacher als $\mathrm{Res}_x(f, g)$.

Überlegen wir uns, was bei der Polynomdivision mit den Koeffizienten passiert.

Wir berechnen eine Folge von Polynomen $g_0 = g, g_1, \ldots, g_r = h$, wobei g_i aus seinem Vorgänger dadurch entsteht, daß wir ein Vielfaches von $X^j f$ subtrahieren, wobei $j = \deg g_i - \deg f$ ist. Der maximale Wert, den j annehmen kann, ist offenbar $\deg g - \deg f = m - n$.

Die Zeilen der SYLVESTER-Matrix sind Vektoren in k^{n+m}; die ersten m sind die Koeffizientenvektoren von $X^{m-1}f, \ldots, Xf, f$, danach folgen die von $X^{n-1}g, \ldots, Xg, g$.

Im ersten Divisionschritt subtrahieren wir von g ein Vielfaches $\lambda X^j f$ mit $j = m - n$; damit subtrahieren wir auch von jeder Potenz $X^i g$ das Polynom $\lambda X^{i+j} f$. Für $0 \le i < n$ und $0 \le j \le m+n$ ist $0 \le i+j < m$, was wir subtrahieren entspricht auf dem Niveau der Koeffizientenvektoren also stets einem Vielfachen einer Zeile der SYLVESTER-Matrix. Damit ändert sich nichts am Wert der Determinanten, wenn wir den Koeffizientenvektor von g nacheinander durch den von $g_1, \ldots, g_r = h$ ersetzen.

Die Resultante ändert sich also nicht, wenn man in der SYLVESTER-Matrix entsteht jede Zeile mit Koeffizienten von g ersetzt durch die entsprechende Zeilen mit Koeffizientenvon h, wobei h als ein Polynom vom Grad m behandelt wird, dessen führende Koeffizienten verschwinden. Ist $h = c_s x^s + \cdots + c_0$, so ist also $\mathrm{Res}_x(f, g)$ gleich

$$\begin{vmatrix}
a_n & a_{n-1} & a_{n-2} & \ldots & a_1 & a_0 & 0 & 0 & \ldots & 0 \\
0 & a_n & a_{n-1} & \ldots & a_2 & a_1 & a_0 & 0 & \ldots & 0 \\
0 & 0 & a_n & \ldots & a_3 & a_2 & a_1 & a_0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & a_n & a_{n-1} & a_{n-2} & a_{n-3} & \ldots & a_0 \\
c_m & c_{m-1} & c_{m-2} & \ldots & c_2 & c_1 & c_0 & 0 & \ldots & 0 \\
0 & c_m & c_{m-1} & \ldots & c_3 & c_2 & c_1 & c_0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0 & c_m & c_{m-1} & c_{m-2} & \ldots & c_0
\end{vmatrix},$$

wobei die Koeffizienten $c_m, \ldots, c_{s+1}$ alle verschwinden.

Somit beginnt im unteren Teil der Matrix jede Zeile mit $m - s$ Nullen.

In den ersten $m-s$ Spalten der Matrix stehen daher nur noch Koeffizienten von f: In der ersten ist dies ausschließlich der führende Koeffizient a_n von f in der ersten Zeile. Entwickeln wir nach der ersten Zeile, können wir also einfach die erste Zeile und die erste Zeile streichen; die Determinante ist dann a_n mal der Determinante der übrigbleibenden Matrix. Diese hat (falls $m > s+1$) wieder dieselbe Gestalt, wir können also wieder einen Faktor a_n ausklammern und bekommen eine Determinante mit einer Zeile und einer Spalte weniger *uws.;* das Ganze funktioniert $m - s$ mal, dann ist der führende Koeffizient von h in die erste Spalte gerutscht und die übriggebliebene Matrix ist die Sylvestermatrix von f und h – falls etwas übrigbleibt.

Offensichtlich bleibt genau dann nichts übrig, wenn h das Nullpolynom ist: Dann sind die unteren m Zeilen Null, d.h. die Resultante verschwindet.

Andernfalls ist

$$\mathrm{Res}_x(f, g) = a_n^{m-s} \mathrm{Res}_x(f, h),$$

und da diese Formel auch für $h = 0$ gilt, haben wir gezeigt

Lemma: Hat f keinen größeren Grad als g und ist h der Divisionsrest von g durch f, der den Grad s habe, so ist $\mathrm{Res}(f, g) = a_n^{m-s} \mathrm{Res}(f, h)$. ∎

Dies läßt sich nun nach Art des EUKLIDischen Algorithmus iterieren: Berechnen wir wie dort die Folge der Reste $r_1 = h$ der Division von g durch f und dann (mit $r_0 = g$) weiter r_{i+1} gleich dem Rest bei der Division von r_i durch r_{i-1}, so können wie die Berechnung von $\mathrm{Res}_x(f, g)$ durch Multiplikation mit Potenzen der führenden Koeffizienten der Divisoren zurückführen auf die viel kleineren Resultanten $\mathrm{Res}_x(r_i, r_{i+1})$. Sobald r_{i+1} eine Konstante ist, egal ob Null oder nicht, haben wir eine explizite Formel und der Algorithmus endet. Für den Fall, daß f größeren Grad als g hat brauchen wir noch

Lemma: Für ein Polynom, f vom Grad n und ein Polynom g vom Grad n ist $\mathrm{Res}(f, g) = (-1)^{nm} \mathrm{Res}(g, f)$.

Beweis: Wir müssen in der SYLVESTER-Matrix m Zeilen zu f mit den n Zeilen zu g vertauschen. Dies kann beispielsweise so realisiert werden, daß wir die unterste f-Zeile nacheinander mit jeder der g-Zeilen vertauschen, bis sie nach n Vertauschungen schließlich unten steht. Dies müssen wir wiederholen, bis alle f-Zeilen unten stehen, wir haben also insgesamt nm Zeilenvertauschungen. Somit ändert sich das Vorzeichen der Determinante um den Faktor $(-1)^{nm}$. ∎

i) Resultanten und homomorphe Bilder

Ein aufmerksamer Leser muß sich an dieser Stelle (mindestens) zwei Fragen stellen:

1.) Die Resultante sagt uns, ob zwei Polynome einen gemeinsamen Teiler haben oder nicht. Um dies zu entscheiden, berechnen wir den größten gemeinsamen Teiler und je nachdem, welchen Grad dieser hat, setzen wir die Resultante auf Null oder einen anderen Wert. Wozu brauchen wir, wenn wir ohnehin den ggT berechnen, dann überhaupt eine Resultante?

2.) Da der Algorithmus à la EUKLID die Folge der sukkzessiven Reste berechnet, bekommen wir genau dieselben Probleme mit explodierenden Zwischenergebnissen wie bei EUKLIDischen Algorithmus. Dort hielten wir diese für inakzeptabel; warum sollten wir sie hier tolerieren?

Tatsächlich ist kaum eine Situation vorstellbar, in der es sonderlich sinnvoll wäre, die Resultante zweier konkret gegebener Polynome aus $\mathbb{Q}[X]$ zu berechnen: Wenn wir wissen wollen, ob sie gemeinsame Nullstellen haben, berechnen wir ihren ggT.

Die wahre Nützlichkeit von Resultanten kommt von Situationen wie der, für die wir Resultanten bereits angewandt haben: Wir haben Resultanten für Polynome mit ganzzahligen Koeffizienten berechnet und daraus geschlossen, modulo welcher Primzahlen zwei Polynome einen gemeinsamen Teiler haben. Dort ging es in erster Linie um den Beweis, daß das ggT-Problem nur modulo endlich vieler Primzahlen schlechte Reduktion hat.

Dort ging es zwar nur um einen abstrakten Beweis, aber entsprechende Situationen lassen sich auch für algorithmische Anwendungen nutzen.

Das Grundprinzip, das uns auf beide Fragen eine Antwort geben wird, ist das bereits beim modularen EUKLIDischen Algorithmus betrachtete Rechnen modulo homomorpher Bilder.

Zunächst einige Begriffe aus der Algebra:

Definition: *a)* Ein Homomorphismus $\varphi\colon R \to S$ von Ringen ist eine Abbildung, für die gilt

$$\varphi(r+s) = \varphi(r) + \varphi(s) \quad \text{und} \quad \varphi(r \cdot s) = \varphi(r) \cdot \varphi(s)$$

für alle $r, s \in \mathbb{R}$.
b) Der *Kern* eines Homomorphismus $\varphi\colon R \to S$ von Ringen ist die Menge aller $r \in R$ mit $\varphi(r) = 0$.
c) Eine Teilmenge $I \subseteq R$ heißt *Ideal* von R, in Zeichen $I \lhd R$, wenn gilt:

1.) I ist eine additive Untergruppe von R
2.) Für $r \in R$ und $s \in I$ ist auch rs ein Element von I.

Ideale spielen bei Ringen genau dieselbe Rolle wie Normalteiler bei Gruppen, d.h. es gilt:

Lemma: Zu einem Ring R und einer Teilmenge $I \subseteq R$ gibt es genau dann einen Homomorphismus $\varphi\colon R \to S$ mit Kern I, wenn I ein Ideal von R ist.

Beweis: Ist I Kern des Homomorphismus $\varphi\colon R \to S$, so ist I natürlich eine additive Untergruppe von R, da φ insbesondere auch ein Gruppenhomomorphismus ist. Für $r \in R$ und $s \in I$ ist $\varphi(s) = 0$, also auch $\varphi(rs) = \varphi(r)\varphi(s) = 0$. Somit ist I ein Ideal.

Ist umgekehrt I ein Ideal von R, so können wir auf R eine Äquivalenzrelation definieren durch $r \sim s$ genau dann, wenn $r - s \in I$. Die Äquivalenzklasse von r bezeichnen wir mit $\overline{r}$, die Menge aller Äquivalenzklassen mit $\overline{R} = R/I$.

Für $r, r', s, s' \in R$ mit $r \sim r'$ und $s \sim s'$ liegt mit $r - r'$ und $s - s'$ auch $r + s - r' - s'$ in I, d.h. $\overline{r+s} = \overline{r} + \overline{s}$. Genauso ist auch $\overline{r \cdot s} = \overline{r} \cdot \overline{s}$, denn ist $r' = r + i$ und $s' = s + j$ mit $i, j \in I$, so ist

$$r' \cdot s' = (r+i)(s+j) = rs + is + rj + ij\,,$$

und wegen der Idealeigenschaft von i liegen rs, is und rj allesamt in I. Somit ist $\overline{R}$ ein Ring, und die Abbildung $\varphi\colon R \to \overline{R}$, die jedes $r \in R$ auf seine Äquivalenzklasse $\overline{r}$ in $\overline{R}$ abbildet, ist ein Homomorphismus, dessen Kern natürlich I ist. ∎

Die Ideale haben ihren Namen von KUMMER, der sie als *ideale Zahlen* betrachtete: KUMMER glaubte zunächst, er habe einen Beweis der FERMAT-Vermutung gefunden, allerdings war er davon ausgegangen, daß der Ring $\mathbb{Z}[\zeta_p]$, wobei p eine primitive p-te Einheitswurzel bezeichnet, faktoriell ist. Dies ist zwar für unendlich viele Primzahlen p der Fall, aber eben nicht für alle. KUMMER konnte aber zeigen, daß es auf dem Niveau der Ideale eine eindeutige Primzerlegung gibt – leider reichte das aber nicht aus, um seinen Beweis auch für die Primzahlen zu retten für die $\mathbb{Z}[\zeta_p]$ nicht faktoriell ist.

Natürlich definiert jeder Homomorphismus $\varphi\colon R \to S$ einen Homomorphismus

$$\varphi\colon \begin{cases} R[X] \to S[X] \\ a_n X^n + \cdots + a_0 \mapsto \varphi(a_n)X^n + \cdots + \varphi(a_0) \end{cases}$$

zwischen den Polynomringen darüber, und da die Resultante zweier Polynome als Summe von Produkten von Koeffizienten der beiden Polynome dargestellt werden kann, ist

$$\operatorname{Res}_X\big(\varphi(f), \varphi(g)\big) = \varphi\big(\operatorname{Res}_X(f,g)\big)\,.$$

Diese Formel hatten wir bereits im Fall $R = \mathbb{Z}$ und $S = \mathbb{F}_p$ angewandt; für praktische Anwendungen interessanter sind aber Fälle wie die Abbildungen

$$\varphi_a\colon \begin{cases} R[X] \to R \\ X \mapsto a \end{cases} \quad \text{für ein Element } a \in R\,.$$

...

Kapitel 3
Faktorisierung von Polynomen

Die Lösung sowohl einzelner Polynomgleichungen als auch von Systemen solcher Gleichungen wird mit steigendem Grad der Polynome sehr schnell sehr viel schwieriger; falls man einzelne der Polynome in Faktoren zerlegen kann, ist es meist effizienten, mit diesen zu arbeiten – obwohl die Anzahl der betrachteten Fälle bei Systemen von hinreichend vielen Polynomgleichungen auch da ziemlich groß werden kann.

Wie aus der Analysis I bekannt, läßt sich jedes Polynom über den reellen Zahlen in ein Produkt von Potenzen linearer und quadratischer Faktoren zerlegen; über den komplexen Zahlen reichen sogar lineare. Diese Art von Faktorisierung ist allerdings algorithmisch selbst bei Polynomen mit ganzzahligen Koeffizienten extrem aufwendigund lohnt nur in seltenen Fällen. Sinnvoll ist dagegen die Faktorisierung über Körpern wie $\mathbb{Q}$ oder endlichen Erweiterungen davon.

Wirklich effiziente direkte Algorithmen zur Faktorisierung sind allerdings nur über endlichen Körpern bekannt; deshalb wird auch hier unsere Strategie sein, daß wir wie beim ggT den Umweg über endliche Körper machen. Zunächst aber wollen wir Polynome über „beliebigen“ Körpern in einem ersten und billigen Schritt in quadratfreie Faktoren zerlegen, d.h. in Faktoren, in deren Primfaktorzerlegung kein irreduzibles Polynom mit einem Exponenten größer eins vorkommt. Alle folgenden Algorithmen werden sich nur mit quadratfreien Polynomen beschäftigen

Das Wort *beliebig* im vorigen Abschnitt ist in Anführungszeichen gesetzt, da wir natürlich nur über solchen Körpern arbeiten können, in denen wir die Grundrechenarten und den Test auf Gleichheit algorithmisch beschreiben können. Nach dem Satz von RICHARDSON sind damit beispielsweise die reellen und die komplexen Zahlen ausgeschlossen.

§1: Quadratfreie Zerlegung

Wir betrachten ein Polynom f über einem Körper k. Da der Polynomring $k[X]$ faktoriell ist, zerfällt f dort in ein Produkt aus einer Einheit $e \in k^\times$ und Potenzen irreduzibler Polynome aus $k[X]$:

$$f = e \prod_{i=1}^{r} f_i^{n_i} .$$

Falls alle $n_i = 1$ und kein zwei f_i zueinander assoziiert sind, bezeichnen wir f als quadratfrei. Ziel der quadratfreien Zerlegung ist es, ein beliebiges Polynom f in der Form

$$f = \prod_{j=1}^{s} g_j^{m_j}$$

zu schreiben, wobei die g_j zueinander teilerfremde quadratfreie Polynome sind. Vergleichen wir mit der obigen Darstellung und vernachlässigen wir für den Augenblick die Einheit u, so folgt, daß g_j das Produkt aller f_i mit $n_i = j$ ist.

a) Quadratfreie Zerlegung über den reellen Zahlen

Wenn ein Polynom $f \in \mathbb{R}[X]$ eine mehrfache Nullstelle hat, verschwindet dort auch die Ableitung f'. Allgemeiner gilt, daß für ein Polynom $h \in \mathbb{R}[X]$, dessen e-te Potenz f teilt, zumindest h^{e-1} auch die Ableitung f' teilen muß, denn ist $f = h^e g$, so ist

$$f' = eh^{e-1}h'g + h^e g' = h^{e-1}(eh'g + hg') .$$

Falls f *genau* durch h^e teilbar ist, ist auch f' genau durch h^{e-1} teilbar, denn wäre es sogar durch h^e teilbar, so wäre auch $eh^{e-1}h'g$ durch h^e teilbar, so daß h ein Teiler von g wäre.

Damit ist $\mathrm{ggT}(f, f') = \prod_{i=1}^{r} f_i^{e_i-1}$ und

$$q_1 = \frac{f}{\mathrm{ggT}(f, f')} = \prod_{i=1}^{r} f_i$$

ist das Produkt aller irreduzibler Faktoren von f. Alle irreduziblen Faktoren von f, die dort mindestens in der zweiten Potenz vorkommen, sind auch Teiler von f', also ist

$$g_1 = \frac{q_1}{\mathrm{ggT}(q_1, f')}$$

das Produkt aller irreduzibler Faktoren von f, die dort genau in der ersten Potenz vorkommen.

In $f_1 = f/q_1$ kommen alle irreduziblen Faktoren von f mit einem um eins verminderten Exponenten vor; insbesondere sind also die mit $e_i = 1$ verschwunden. Wenden wir darauf dieselbe Konstruktion an, erhalten wir die Zerlegung $\mathrm{ggT}(f_1, f_1') = \prod_{i=1}^{r} f_i^{\min(e_i-2,0)}$, und

$$q_2 = \frac{f_1}{\mathrm{ggT}(f_1, f_1')} = \prod_{i=1}^{r} f_i$$

ist das Produkt aller irreduzibler Faktoren von f_1, also das Produkt aller Faktoren von f, die mit einem Exponenten mindestens zwei vorkommen. Damit ist

$$g_2 = \frac{q_2}{\mathrm{ggT}(q_2, f_1')}$$

das Produkt aller Faktoren, die in f mit Multiplizität genau zwei vorkommen. Entsprechend lassen sich auch alle folgenden g_i konstruieren.

b) Ableitungen über einem beliebigen Körper

Auch wenn Ableitungen ursprünglich über Grenzwerte definiert sind, ist doch die Ableitung eines Polynoms rechnerisch gesehen eine rein algebraische Operation, die sich im Prinzip über jedem beliebigen Körper oder sogar Ring erklären läßt. Wir beschränken uns hier auf Polynome über einem Körper k und definieren die Ableitung eines Polynoms

$$f = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0 \in k[X]$$

als das Polynom

$$f' = n a_n X^{n-1} + (n-1) a_{n-1} X^{n-2} + \cdots + a_1 \in k[X]\,.$$

Es ist klar, daß die so definierte Abbildung $k[X] \to k[X]$, die jedem Polynom $f \in k[X]$ seine Ableitung f' zuordnet, k-linear ist. Auch

die LEIBNIZsche Produktregel $(fg)' = fg' + fg'$ ist erfüllt: Wegen der Linearität der Ableitung und der Linearität beider Seiten der Formel sowohl in f als auch in g genügt es, dies für x-Potenzen nachzurechnen, und für $f = x^n, g = x^m$ ist $(fg)' = (n+m)x^{n+m-1}$ gleich

$$fg' + f'g = x^n m x^{m-1} + n x^{n-1} x^m = (m+n) x^{n+m-1}\,.$$

Damit gelten die üblichen Ableitungsregeln auch für die formale Ableitung von Polynomen aus $k[X]$.

Falls ein Polynom f durch das Quadrat q^2 eines anderen teilbar ist, gibt es ein Polynom $g \in k[X]$ mit $f = q^2 g$, und nach der Produktregel ist $f' = 2qg + q^2 g' = q(2g + qg')$, d.h. q teilt auch f' und damit den ggT von f und f'.

Ist umgekehrt ein irreduzibles Polynom $q \in k[X]$ Teiler von f, etwa $f = qh$, so ist $f' = q'h + qh'$ genau dann durch q teilbar, wenn $q'h$ durch q teilbar ist. Da q irreduzibel ist, muß dann entweder q' oder h durch q teilbar sein; da ersteres nicht möglich ist, folgt:

Lemma: Ein irreduzibles Polynom q ist genau dann ein mindestens quadratischer Faktor von f, wenn es den ggT von f und f' teilt. ∎

Genauer: Wenn q in der Primfaktorzerlegung von f in der Potenz q^e auftritt, d.h. $f = q^e g$ mit $q \nmid g$, so ist $f' = eq^{e-1} g + q^e g'$.

Über $\mathbb{R}$ würde daraus folgen, daß q^{e-1} die höchste q-Potenz ist, die f' teilt. Da wir aber über einem beliebigen Körper arbeiten, könnte es sein, daß der erste Faktor verschwindet: Dies passiert genau dann, wenn der Exponent e durch die Charakteristik p des Grundkörpers teilbar ist. In diesem Fall ist $f' = q^e g$ mindestens durch q^e teilbar. Da f genau durch q^e teilbar ist, folgt

Lemma: Ist $f = a \prod q_i^{e_i}$ mit $a \in k^\times$ die Zerlegung eines Polynoms $f \in k[X]$ in irreduzible Faktoren, so ist der ggT von f und f' gleich $\prod q_i^{d_i}$ mit $d_i = \begin{cases} e_i - 1 & \text{falls } p \nmid e_i \\ e_i & \text{falls } p \mid e_i \end{cases}$. ∎

Nach dem Lemma ist zumindest klar, daß $h_1 = f/\operatorname{ggT}(f, f')$ ein quadratfreies Polynom ist, nämlich das Produkt aller jener Primfaktoren von f, deren Exponent nicht durch p teilbar ist. In Charakteristik Null ist also $f/\operatorname{ggT}(f, f')$ einfach das Produkt der sämtlichen irreduziblen Faktoren von f. Diejenigen Faktoren, die mindestens quadratisch vorkommen, sind gleichzeitig Teiler des ggT; das Produkt g_1 der Faktoren, die genau in der ersten Potenz vorkommen, ist also $h_1/\operatorname{ggT}\big(h_1, \operatorname{ggT}(f, f')\big)$.

Falls $\operatorname{ggT}(f, f')$ kleineren Grad als f hat, können wir rekursiv weitermachen und nach derselben Methode das Produkt aller Faktoren bilden, die in $f_1 = \operatorname{ggT}(f, f')$ genau mit Exponent eins vorkommen; in f selbst sind das quadratische Faktoren. Weiter geht es mit $f_2 = \operatorname{ggT}(f_2, f_2')$, dessen Faktoren mit Exponent eins kubisch in f auftreten, *usw.*

Über einem Körper der Charakteristik Null liefert diese Vorgehensweise die gesamte quadratfreie Zerlegung; in positiver Charakteristik kann es allerdings vorkommen, daß $\operatorname{ggT}(f, f') = f$ ist. Da $\deg f' < \deg f$, ist dies genau dann der Fall, wenn $f' = 0$ ist. Dies ist in Charakteristik Null genau dann der Fall, wenn f konstant ist; in Charakteristik p verschwindet aber auch die Ableitung ex^{e-1} einer jeden x-Potenz, deren Exponent ein Vielfaches von p ist. Somit ist hier $f' = 0$ genau dann, wenn alle in f vorkommenden x-Potenzen einen durch p teilbaren Exponenten haben. Dann ist für $f \in \mathbb{F}_p[X]$

$$\begin{aligned} f &= a_{np}X^{np} + a_{(n-1)p}X^{(n-1)p} + \cdots + a_pX^p + a_0 \\ &= \left(a_{np}X^p + a_{(n-1)p}X^{(n-1)} + \cdots + a_pX + a_0\right)^p , \end{aligned}$$

da nach dem kleinen Satz von FERMAT in $\mathbb{F}_p$ jedes Element gleich seiner p-ten Potenz ist. f ist dann also die p-te Potenz eines anderen Polynoms, und wir können den Algorithmus auf dieses anwenden. Im Endergebnis müssen dann natürlich alle hier gefundenen Faktoren in die p-te Potenz gehoben werden.

In anderen Körpern der Charakteristik p ist die Situation etwas komplizierter: Dort müssen wir zunächst Elemente b_i finden mit $b_i^p = a_{ip}$; dann ist

$$f = \left(b_nX^n + b_{n-1}X^{n-1} + \cdots + b_1X + b_0\right)^p .$$

Solche Elemente müssen nicht existieren, es gibt aber eine große Klasse von Körpern, in denen sie stets existieren:

Definition: Ein Körper k der Charakteristik $p > 0$ heißt vollkommen, wenn die Abbildung $k \to k$; $x \mapsto x^p$ surjektiv ist.

Man kann zeigen, daß jeder endliche Körper vollkommen ist: Im Körper mit p^n Elementen ist $x^{p^n} = x$ für alle $x \in \mathbb{F}_{p^n}$, und damit ist x die p-te Potenz von $y = x^{p^{n-1}}$. Ein Beispiel eines nicht vollkommenen Körpers wäre $\mathbb{F}_p(X)$, wo X offensichtlich nicht als p-te Potenz eines anderen Körperelements geschrieben werden kann.

Über einem vollkommenen Körper der Charakteristik p kann man also jedes Polynom, dessen Ableitung das Nullpolynom ist, als p-te Potenz eines anderen Polynoms schreiben und so, falls man die p-ten Wurzeln auch effektiv berechnen kann, den Algorithmus zur quadratfreien Zerlegung durchführen.

§ 2: Der Berlekamp-Algorithmus

Wir gehen aus von einem *quadratfreien* Polynom über dem Körper $\mathbb{F}_p$ mit p Elementen, d.h. $f \in \mathbb{F}_p[X]$ ist ein Produkt von *verschiedenen* irreduziblen Polynomen $f_1, \ldots, f_r$. Durch quadratfreie Zerlegung läßt sich jedes Faktorisierungsproblem in $\mathbb{F}_p[X]$ auf diesen Fall zurückführen.

a) Ein erster Ansatz

Um zu sehen, wie wir die f_i bestimmen können, nehmen wir zunächst an, sie seien bereits bekannt. Wir wählen uns dann irgendwelche Zahlen $s_1, \ldots, s_r \in \mathbb{F}_p$ und suchen ein Polynom $v \in \mathbb{F}_p[X]$ mit

$$v \equiv s_i \bmod f_i \quad \text{für alle } i = 1, \ldots, r .$$

Da die f_i als verschiedene irreduzible Polynome vorausgesetzt waren, sind sie insbesondere paarweise teilerfremd; es gibt daher nach dem chinesischen Restesatz

Für ein quadratfreies Polynom $f \in \mathbb{F}_p[X]$ mit irreduziblen Faktoren $f_1, \ldots, f_r$ ist somit der Vektorraum V aller Polynome $v \in \mathbb{F}_p[X]$ mit

$v^p \equiv v \bmod f$ gleich dem Vektorraum aller Polynome, zu denen es Elemente $s_1, \ldots, s_r \in \mathbb{F}_p$ gibt, so daß $v \equiv s_i \bmod f_i$. Für $v \in V$ ist daher ggT$(v - \lambda, f)$ gleich dem Produkt aller f_i mit $s_i = \lambda$. Falls alle s_i verschieden sind, bekommen wir also alle Faktoren, indem wir ggT$(v-\lambda, f)$ für alle $\lambda \in \mathbb{F}_p$ berechnen; wenn sie nicht alle verschieden sind (etwa weil es mehr Faktoren gibt als Elemente von $\mathbb{F}_p$), haben wir zumindest eine teilweise Zerlegung und können mit einem neuen Polynom v zu anderen Werten s_i weitermachen.

Der einzige Nachteil an dieser Vorgehensweise besteht darin, daß wir f erst konstruieren können, *nachdem* wir die Faktoren f_i von f bereits kennen – genau diese Faktoren suchen wir aber gerade. Die Idee des BERLEKAMP-Algorithmus besteht darin, den Vektorraum V auf eine andere Weise zu charakterisieren, die ohne Kenntnis der f_i auskommt. Für diese Charakterisierung brauchen wir zunächst einige algebraische Vorbereitungen:

b) Der kleine Satz von Fermat

Zu jedem Ring R gibt es genau einen Homomorphismus $\varphi\colon \mathbb{Z} \to R$ von den ganzen Zahlen nach R, denn ein Homomorphismus muß die Eins auf die Eins abbilden und jede natürliche Zahl n entsprechend auf die Summe von n Einsen. Der Kern dieses Homomorphismus ist – wie jeder Kern eines Homomorphismus von Ringen – ein Ideal in $\mathbb{Z}$, also ein Hauptideal (p) mit $p \in \mathbb{N}_0$.

Definition: Die Charakteristik eines Körpers k ist jenes $p \in \mathbb{N}_0$, für das (p) der Kern des Homomorphismus $\mathbb{Z} \to k$ ist. Wir schreiben $p = \operatorname{char} k$.

Lemma: Die Charakteristik eines Körpers ist entweder Null oder eine Primzahl.

Beweis: Angenommen, das ist nicht der Fall. Dann ist p eine zusammengesetzte Zahl, es gibt also zwei Zahlen $a, b < p$ in $\mathbb{N}$, so daß $p = ab$ ist. In k ist dann $a \cdot 1 \neq 0$ und $b \cdot 1 \neq 0$, aber das Produkt dieser beiden Zahlen ist $ab \cdot 1 = p \cdot 1 = 0$. Wegen der Nullteilerfreiheit eines Körpers ist das nicht möglich. ∎

Die Körper $\mathbb{Q}, \mathbb{R}$ und $\mathbb{C}$ haben natürlich allesamt Charakteristik Null, denn dort ist $\mathbb{Z}$ ein Teilring. Der Körper $\mathbb{F}_p$ ist schon konstruiert als Faktorring von $\mathbb{Z}$ mit (p) als Kern der Projektion $\mathbb{Z} \to \mathbb{F}_p$; somit ist $\operatorname{char} \mathbb{F}_p = p$.

Eines der wichtigsten Hilfsmittel der Algebra über Körpern positiver Charakteristik ist der FROBENIUS-Homomorphismus:

Lemma: Für einen Körper k der Charakteristik $p > 0$ ist die Abbildung $k \to k;\ \ x \mapsto x^p$ ein Homomorphismus.

Beweis: Natürlich ist in jedem Körper $(xy)^p = x^p y^p$; wir müssen uns überlegen, daß auch $(x+y)^p = x^p + y^p$ ist. Nach der binomischen Formel gilt

$$(x+y)^p = \binom{p}{0}x^p + \binom{p}{1}x^{p-1}y + \binom{p}{2}x^{p-2}y^2 + \ldots + \binom{p}{1}xy^{p-1} + \binom{p}{p}y^p .$$

Dabei hat

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

für $i \geq 1$ einen durch p teilbaren Zähler, während der Nenner für $i \leq p-1$ nicht durch p teilbar ist, denn eine Primzahl p kann natürlich keine natürliche Zahl echt kleiner p teilen. Somit ist $\binom{p}{i}$ für $1 \leq i \leq p-1$ durch p teilbar; der Term $\binom{p}{i}x^i y^{p-1}$ verschwindet also in $\mathbb{F}_p$ für alle i außer $i = 0$ und $i = p$. In diesen beiden Fällen ist der Binomialkoeffizient gleich eins, also ist $(x + y)^p = x^p + y^p$, wie behauptet. ∎

Der Homomorphismums $x \mapsto x^p$ heißt FROBENIUS-Homomorphismus; in den endlichen Körpern, die uns am meisten interessieren, sagt uns der folgende Satz, daß er trivial ist – was, wie wir im folgenden sehen werden, allerdings keinesfalls bedeutet, daß er uninteressant wäre:

Kleiner Satz von Fermat: *a)* Für eine ganze Zahl $x \in \mathbb{Z}$ und eine Primzahl p ist $x^p \equiv x \bmod p$. Ist p kein Teiler von x, so ist auch $x^{p-1} \equiv 1 \bmod p$.
b) Ist p eine Primzahl, so ist $x^p = x$ für alle $x \in \mathbb{F}_p$ und $x^{p-1} = 1$ für alle $x \neq 0$ aus $\mathbb{F}_p$.

Beweis: Natürlich ist $1^p = 1 \equiv 1 \bmod p$. Da nach dem vorigen Lemma $(x+1)^p \equiv x^p + 1^p \bmod p$, folgt daraus induktiv, daß $x^p \equiv x \bmod p$ für alle natürlichen Zahlen x. Für diese ist dann auch $(-x)^p = (-1)^p x^p \equiv -x \bmod p$ (man überzeuge sich davon, daß dies auch für $p = 2$ gilt!), und natürlich ist $0^p = 0$. Somit ist $x^p \equiv x \bmod p$ für alle $x \in \mathbb{Z}$. Da jeses $x \in \mathbb{F}_p$ einen Reprässentanten in $\mathbb{Z}$ hat, ist damit auch $x^p = x$ für alle $x \in \mathbb{F}_p$. Für $x \in \mathbb{F}_p^\times = \mathbb{F}_p \setminus \{0\}$ können wir durch x dividieren und erhalten $x^{p-1} = 1$. Für eine nicht durch p teilbare ganze Zahl x folgt, daß ihre Restklasse modulo p die Eins als $(p-1)$-te Potenz hat, also ist $x^{p-1} \equiv 1 \bmod p$.

∎

Korollar: Über einem Körper k der Charakteristik $p > 0$ ist

$$X^p - X = \prod_{i=0}^{p-1}(X - i) \quad \text{und} \quad X^{p-1} - 1 = \prod_{i=1}^{p-1}(X - i)\,.$$

Beweis: Nach dem kleinen Satz von FERMAT sind die betreffenden Elemente Nullstellen, und da ein von Null verschiedenes Polynom vom Grad n nicht mehr als n Nullstellen haben kann, gibt es keine weiteren. Die Gleichheit beider Seiten folgt somit daraus, daß die führenden Koeffizienten beider Polynome Eins sind.

∎

c) Anwendung auf den Berlekamp-Algorithmus

Sei nun wieder $f = f_1 \cdots f_r \in \mathbb{F}_p[X]$ ein quadratfreies Polynom vom Grad n mit irreduziblen Faktoren f_i. Wir betrachteten Polynome $v \in \mathbb{F}_p[X]$ vom Grad höchstens $n-1$ mit der Eigenschaft $v \equiv s_i \bmod f_i$ für irgendwelche Elemente $s_i \in \mathbb{F}_p$. Für ein solches Polynom

$$v = v_{n-1}X^{n-1} + v_{n-2}X^{n-2} + \cdots + v_1 X + v_0$$

...

d) Durchführung des Berlekamp-Algorithmus

Die zweite Charakterisierung von V zeigt, daß die Dimension von V gleich der Anzahl r der irreduziblen Faktoren von f ist; insbesondere ist also V eindimensional genau dann, wenn f irreduzibel ist.

Andernfalls wählen wir irgendein Element $v \in V$ und berechnen die Polynome $\mathrm{ggT}(v - \lambda, f)$ für alle $\lambda \in \mathbb{F}_p$. Falls wir dabei r mal ein nichtkonstantes Polynom bekommen, haben wir f faktorisiert. Wenn wir zu wenige Faktoren bekommen, waren für das betrachtete Polynom v einige der Werte s_i gleich; wir bilden eine Liste der gefundenen (und zumindest noch nicht in allen Fällen irreduziblen) Faktoren und wählen wir ein von v linear unabhängiges neues Polynom $w \in V$ und machen damit dasselbe. Indem wir für jedes nichtkonstante Polynom $\mathrm{ggT}(w - \lambda, f)$ den ggT mit den in der Liste stehenden Faktoren bilden, können wir die Listenelemente weiter zerlegen. Bei jeder gefundenen Zerlegung ersetzen wir das zerlegte Element durch seine Faktoren. Sobald die Liste r Faktoren enthält, sind wir fertig.

Falls die sämtlichen $\mathrm{ggT}(w - \lambda, f)$ immer noch nicht ausreichen, um r Faktoren zu produzieren, müssen wir ein neues, von v und w linear unabhängiges Element von V wählen und damit weitermachen, *usw.*

Das Verfahren muß spätestens mit dem r-ten Polynom v enden, denn dann haben wir eine Basis $v_1, \ldots, v_n$ von V durchprobiert. Hätten wir dann noch nicht alle r Faktoren isoliert, müßte es (mindestens) zwei Faktoren f_i und f_j geben, so daß $v_\ell \bmod f_i = v_\ell \bmod f_j$ ist für alle Basiselemente v_ℓ und damit auch $v \bmod f_i = v \bmod f_j$ für alle $v \in V$. Das ist aber nicht möglich, denn nach dem chinesischen Restesatz enthält V beispielsweise auch ein Element v mit $v \bmod f_i = 0$ und $v \bmod f_j = 1$.

§3: Faktorisierung über den ganzen Zahlen und über endlichen Körpern

Wie bei der Berechnung des ggT zweier Polynome wollen wir auch bei der Faktorisierung den Umweg über endliche Körper benutzen, um Probleme für Polynome über $\mathbb{Z}$ zu lösen. Allerdings kann es hier häufiger passieren, daß sich Ergebnisse über $\mathbb{F}_p$ deutlich unterscheiden von denen über Z.

Betrachten wir dazu als erstes Beispiel das Polynom $X^2 + 1$ aus $\mathbb{Z}[X]$. Es ist irreduzibel, da eine Zerlegung die Form $(X - a)(X + a)$ haben müßte mit $a \in \mathbb{Z}$, und in Z gibt es kein Element a mit $a^2 = -1$.

Auch über dem Körper $\mathbb{F}_p$ muß eine eventuelle Faktorisierung die Form $(X-a)(X+a)$ haben mit $a^2=-1$; wir müssen uns also überlegen, wann das der Fall ist. Die elementare Zahlentheorie sagt uns:

Lemma: Genau dann gibt es im endlichen Körper $\mathbb{F}_p$ ein Element a mit $a^2=-1$, wenn $p=2$ oder $p\equiv 1 \bmod 4$ ist.

Beweis: Für $p=2$ ist natürlich $1^2=1=-1$ die Lösung. Für $p\equiv 1 \bmod 4$ schreiben wir $p=4k+1$. Nach dem kleinen Satz von FERMAT ist für alle $x\in\mathbb{F}_p^\times$

$$(x^{p-1}-1)=(x^{2k}+1)(x^{2k}-1)=0\,,$$

das linksstehende Polynom hat also $p-1=4k$ Nullstellen und zerfällt damit über $\mathbb{F}_p$ in Linearfaktoren. Damit gilt dasselbe für die beiden rechtsstehenden Faktoren; insbesondere gibt es also ein $x\in\mathbb{F}_p$ mit $x^{2k}+1=0$. Für $a=x^k$ ist dann $a^2=x^{2k}=-1$.

Ist $p\equiv 3 \bmod 4$ und $a^2=-1$ für ein $a\in\mathbb{F}_p$, so ist $a^4=1$. Außerdem ist nach dem kleinen Satz von FERMAT $a^{p-1}=1$. Da $p\equiv 3 \bmod 4$ ist $\mathrm{ggT}(4,p-1)=2$ als Linearkombination von 2 und $p-1$ darstellar, also ist auch $a^2=1$, im Widerspruch zu Annahme $a^2=-1$. Somit gibt es in $\mathbb{F}_p$ keine Elemente mit Quadrat -1. ∎

Damit ist X^2+1 genau dann irreduzibel über $\mathbb{F}_p$, wenn $p\equiv 3 \bmod 4$; in allen anderen Fällen zerfällt das Polynom in zwei Linearfaktoren. Nach DIRICHLETs Satz über Primzahlen in arithmetischen Progressionen bleibt X^2+1 damit nur modulo der Hälfte aller Primzahlen irreduzibel.

Noch schlimmer ist es bei X^4+1: Auch dieses Polynom ist irreduzibel über $\mathbb{Z}$: Da seine Nullstellen $\frac{1}{2}\sqrt{2}(\pm 1\pm i)$ nicht in $\mathbb{Z}$ liegen, gibt es keinen linearen Faktor, und wäre

$$\begin{aligned}X^4+1&=(X^2+aX+b)(X^2+cX+d)\\&=X^4+(a+c)X^3+(b+d+ac)X^2+(ad+bc)X+bd\end{aligned}$$

eine Zerlegung in quadratische Faktoren, so zeigen die Koeffizieten von X^3 und der konstante Term, daß $c=-a$ und $b=d=\pm 1$ sein

müßte. Die Produkte

$$(X^2+aX+1)(X^2-aX+1)=X^4+(2-a^2)X^2+1$$

und

$$(X^2+aX-1)(X^2-aX-1)=X^4-(2+a^2)X^2+1$$

zeigen aber, daß beides nur für $a^2=\pm 2$ zu einer Faktorisierung führen könnte, was in $\mathbb{Z}$ nicht erfüllbar ist.

In den Körpern $\mathbb{F}_p$ dagegen kann es sehr wohl Elemente geben, deren Quadrat ± 2 ist, und dann zeigen die obigen Formeln, daß X^4+1 dort in ein Produkt zweier quadratischer Polynome zerlegt werden kann. Auch wenn es ein Element $a\in\mathbb{F}_p$ gibt mit $a^2=-1$, können wir X^4+1 als Produkt schreiben, nämlich genau wie oben im Falle X^2+1 als

$$X^4+1=(X^2+a)(X^2-a)\,.$$

Somit ist X^4+1 über dem Körper $\mathbb{F}_p$ zumindest dann reduzibel, wenn dort wenigstens eines der drei Elemente -1 und ± 2 ein Quadrat ist. Um zu sehen, daß X^4+1 über jedem dieser Körper zerfällt, müssen wir uns also überlegen, daß in keinem der Körper $\mathbb{F}_p$ alle drei Elemente *keine* Quadrate sind. Da $-2=-1\cdot 2$ ist, folgt dies aus

Lemma: Sind im Körper $\mathbb{F}_p$ die beiden Elemente a,b nicht als Quadrate darstellbar, so ist ab ein Quadrat.

Beweis: Für $p=2$ ist jedes Element ein Quadrat und nicht zu beweisen. Ansonsten betrachten wir die Abbildung $\mathbb{F}_p^\times\to\mathbb{F}_p^\times$, die jedes von Null verschiedene Element von $\mathbb{F}_p$ auf sein Quadrat abbildet. Sie ist offensichtlich ein Homomorphismus von multiplikativen Gruppen, und da $1\neq p-1=-1$ ist, besteht ihr Kern aus genau zwei Elementen. Nach dem Homomorphiesatz hat das Bild somit $\frac{1}{2}(p-1)$ Elemente, d.h. die Hälfte alle Elemente von $\mathbb{F}_p^\times$ sind Quadrate. Ist a keines, so ist auch ax^2 für kein $x\in\mathbb{F}_p^\times$ ein Quadrat y, denn sonst wäre $a=(y^2x^{-2}$ selbst ein Quadrat.

Da es $\frac{1}{2}(p-1)$ Quadrate und genauso viele Nichtquadrate gibt, läßt sich somit jedes Nichtquadrat b als $b = ax^2$ schreiben mit einem geeigneten Element $x \in \mathbb{F}_p$. Damit ist $ab = (ax)^2$ ein Quadrat.

∎

Die Situation kann hier also deutlich schlechter werden als im Fall des EUKLIDischen Algorithmus, wo wir sicher sein konnten, daß es höchstens endlich viele schlechte Primzahlen gibt: Hier können *alle* Primzahlen schlecht sein in dem Sinne, daß ein irreduzibles Polynom aus $\mathbb{Z}[X]$ modulo p reduzibel wird, und oft wird zumindest die Hälfte aller Primzahlen schlecht sein. Der Ansatz über den chinesischen Restesatz empfiehlt sich hier also definitiv nicht: Wenn wir die Faktorisierung modulo verschiedener Primzahlen durchführen, können wir praktisch sicher sein, daß es darunter auch schlechte gibt, und meist werden auch die Ergebnisse modulo verschiedener Primzahlen entweder nicht zusammenpassen, oder aber wir haben mehrere Faktoren gleichen Grades, von denen wir nicht wissen, welche wir via chinesischen Restesatz miteinander kombinieren sollen. Es hat daher keinen Zweck, zufällig Primzahlen zu wählen und dann eine Rückfallstrategie für schlechte Primzahlen

Der Weg über endliche Körper verfolgt daher im Falle der Faktorisierung eine andere Strategie als beim EUKLIDischen Algorithmus: Wir beschränken uns auf eine einzige Primzahl – unabhängig davon, ob diese nun gut oder schlecht dafür geeignet ist.

Wir kennen bereits aus dem Kapitel über den EUKLIDischen Algorithmus Schranken für die Koeffizienten der Faktoren eines Polynoms; wir könnten also eine Primzahl wählen, die größer ist als das Doppelte dieser Schranke und modulo dieser rechnen.

Der Nachteil dabei ist, daß das Rechnen modulo einer Primzahl p umso teurer wird, je größer die Primzahl ist: Die Kosten für Multiplikationen wachsen quadratisch mit der Stellenzahl von p, die Kosten für Divisionen modulo p nach dem erweiterten EUKLIDischen Algorithmus können sogar bis zu kubisch ansteigen.

Die Alternative bietet ein für völlig andere Zwecke bewiesenes Resultat des deutschen Zahlentheoretikers HENSEL, das es erlaubt eine Faktorisierung modulo p fortzusetzen zu einer Faktorisierung modulo jeder

beliebiger p-Potenz und, was HENSEL wirklich interessierte, zu den p-adischen Zahlen, mit denen wir uns in Rahmen dieser Vorlesung nicht beschäftigen werden.

§4: Das Henselsche Lemma

Lemma: f, g, h seien Polynome aus $\mathbb{Z}[X]$ derart, daß $f \equiv gh \bmod p$; dabei seien $g \bmod p$ und $h \bmod p$ teilerfremd über $\mathbb{F}_p[X]$. Dann gibt es für jede natürliche Zahl n Polynome g_n, h_n derart, daß

$$g_n \equiv g \bmod p, \quad h_n \equiv hh \bmod p \quad \text{und} \quad f \equiv g_n h_n \bmod p^n .$$

Beweis durch vollständige Induktion: Der Fall $n = 1$ ist die Voraussetzung des Lemmas. Ist das Lemma für ein n bewiesen, machen wir den Ansatz

$$g_{n+1} = g_n + p^n g^* \quad \text{und} \quad h_{n+1} = h_n + p^n h^* .$$

Nach Induktionsvoraussetzung ist $f \equiv g_n h_n \bmod p^n$, die Differenz $f - g_n h_n$ ist also durch p^n teilbar und es gibt ein Polynom $f^* \in \mathbb{Z}[X]$, so daß $f = g_n h_n + p^n f^*$ ist. Wir möchten, daß

$$f \equiv (g_n + p^n g^*)(h_n + p^n h^*) = g_n h_n + p^n (g_n h^* + h_n g^*) + p^{2n} \bmod p^{n+1}$$

ist. Da $2n \geq n + 1$ ist, können wir den letzten Summanden vergessen; zu lösen ist also die Kongruenz

$$f \equiv g_n h_n + p^n f^* = g_n h_n + p^n (g_n h^* + h_n g^*) \bmod p^{n+1}$$

oder

$$p^n f^* \equiv p^n (g_n h^* + h_n g^*) \bmod p^{n-1} .$$

Division durch p^n macht daraus

$$f^* \equiv g_n h^* + h_n g^* \bmod p \quad \text{oder} \quad f^* \equiv gh^* + hg^* \bmod p ,$$

denn $g_n \equiv g \bmod p$ und $h_n \equiv h \bmod p$. Die letztere Kongruenz können wir als Gleichung in $\mathbb{F}_p[X]$ auffassen und dort lösen, indem wir den erweiterten EUKLIDischen Algorithmus auf die Polynome $g \bmod p$ und $h \bmod p$ aus $\mathbb{F}_p[X]$ anwenden: Da diese nach Voraussetzung teilerfremd sind, können wir ihren ggT Eins und damit auch jedes andere Polynom

über $\mathbb{F}_p$ als Linearkombination der beiden darstellen. Da der Grad von f die Summe der Grade von g und h ist und f^* höchstens denselben Grad wie f hat, können wir dann auch eine Darstellung $f^* = gh^* + hg^*$ in $\mathbb{F}_p[X]$ finden mit $\deg g^* \leq \deg g$ und $\deg h^* \leq \deg h$. Ersetzen wir g^* und h^* durch irgendwelche Repräsentantanten gleichen Grades aus $\mathbb{Z}[X]$ erfüllen dann $g_{n+1} = g_n + p^n g^*$ und $h_{n+1} = h_n + p^n h^*$ die Kongruenz $f \equiv g_n h_n$ modulo p^{n+1}.

■

§5: Der Algorithmus von Zassenhaus

Die Werkzeuge aus den vorigen Paragraphen erlauben, gemeinsam eingesetzt, nun die Faktorisierung von Polynomen über $\mathbb{Z}$. Nach ZASSENHAUS geht man dabei folgendermaßen vor:

Erster Schritt: Auch wenn es nicht unbedingt nötig ist, sollte man beginnen mit einer quadratfreien Zerlegung des Polynoms $f \in \mathbb{Z}[X]$ oder besser seines primitiven Anteils über $\mathbb{Q}[X]$; die primitiven Anteile davon liefern dann eine quadratfreie Zerlegung über $\mathbb{Z}[X]$, und auf diese quadratfreie Faktoren werden die folgenden Schritte angewendet.

Zweiter Schritt: Berechne eine obere Schranke L für die Koeffizienten der Faktoren des Polynoms und setze $M = 2L + 1$. Wähle eine Primzahl p, die den führenden Koeffizienten nicht teilt und führe eine quadratfreie Zerlegung über $\mathbb{F}_p$ durch. (Dies ist trotz des ersten Schritts notwendig, denn ein in $\mathbb{Z}[X]$ quadratfreies Polynom muß keine quadratfreie Reduktion in $\mathbb{F}_p[X]$ haben.) Wende die folgenden Schritte an für jeden der quadratfreien Faktoren.

Dritter Schritt: Faktorisiere das Polynom nach BERLEKAMP in $\mathbb{F}_p[X]$.

Vierter Schritt: Hebe die Faktorisierung nach dem HENSELschen Lemma hoch zu einer Faktorisierung modulo p^n für eine natürliche Zahl n derart, daß $p^n \geq M$ ist.

Fünfter Schritt: Setze $m = 1$ und teste für jeden der gefundenen Faktoren, ob er das zu faktorisierende Polynom teilt. Falls ja, kommt der Faktor in die Liste L_1 der Faktoren von f; andernfalls kommt es in eine Liste L_2.

Sechster Schritt: Falls die Liste L_2 keine Einträge hat, endet der Algorithmus und f ist das Produkt der Faktoren aus L_1. Andernfalls setze $m = m + 1$ und teste für jedes Produkt aus m verschiedenen Polynomen aus L_2, ob ihr Produkt modulo p^n (mit Koeffizienten vom Betrag höchstens L) ein Teiler von f ist. Falls ja, entferne man die m Faktoren aus L_1 und füge ihr Produkt in die Liste L_1 ein. Wiederhole diesen Schritt.

Auch wenn der sechste Schritt wie eine Endlosschleife aussieht, endet der Algorithmus natürlich nach endlich vielen Schritten, denn L_2 ist eine endliche Liste und spätestens das Produkt aller Elemente aus L_2 muß Teiler von f sein, da sein Produkt mit dem Produkt aller Elemente von L_1 gleich f ist.

§6: Ausblicke

Der sechste Schritt des obigen Algorithmus kann sehr teuer werden, insbesondere wenn man auch Produkte von mehr als zwei Faktoren testen muß. Es gibt einen Algorithmus von LENSTRA, LOVACZ und LENSTRA, den sogenannten LLL-Algorithmus, mit dem man auf systematischere Weise geeignete Kandidaten für zu testende Produkte finden kann. Für Einzelheiten fehlt hier leider die Zeit; man findet sie aber in praktisch jedem neueren Lehrbuch der Computeralgebra oder der algorithmischen Zahlentheorie. LLL findet ganz allgemein kürzeste Vektoren in Gittern der Form $\bigoplus_{i=1}^{m} \mathbb{Z}\vec{v}_i \subset \mathbb{R}^n$ und hat daher noch zahlreiche weitere Anwendungen abgesehen von der Faktorisierung.

Mit nur geringfügiger Modifikation kann der obige Algorithmus auch zur Faktorisierung von Polynomen in mehr als einer Veränderlichen benutzt werden. Auch hier wird das Problem zurückgeführt auf den BERLEKAMP-Algorithmus in $\mathbb{F}_p[X]$, und zwar indem man alle Variablen mit einer Ausnahme auf spezielle Werte setzt. Auch für Faktorisierungen modulo einem Polynom $(X_i - a_i)$ gibt es ein HENSELsches Lemma, mit dem man diese hochheben kann zu Faktorisierungen modulo $(X_i - a_i)^r$, wobei man für r mindestens den Grad von f in X_i wählen muß. Wie oben ist der sechste Schritt oft der umfangreichste; im Gegensatz zur

Situation über $\mathbb{Z}[X]$ gibt es jedoch keinen LLL-Algorithmus, um diesen Schritt zu beschleunigen. Für Einzelheiten sei auf die Originalarbeiten verwiesen, z.B.

P.S. WANG: An improved multivariable polynomial factorising algorithm, *Mathematics of Computation* **32** (1978), 1215-1231

Kapitel 4
Symbolische Integration

Symbolische Integration wurde noch vor dreißig Jahren als typische Anwendung der *Künstlichen Intelligenz* betrachtet und arbeitete viel mit heuristischen Strategien. Inzwischen ist das Problem als weitgehend algebraisch erkannt und es gibt deterministische Algorithmen, die nicht nur Stammfunktionen finden können, sondern gegebenenfalls auch beweisen, daß es keine Stammfunktion in einer gewissen Klasse von Funktionen gibt. Das Gebiet zerfällt im wesentlichen in drei große Teilgebiete: Die Integration von rationalen Funktionen, die von (elementaren) transzendenten Funktionen und schließlich die von algebraischen Funktionen.

Algorithmen zur Integration rationaler Funktionen (jenseits der meist unrealistischen Zerlegung des Nenners in höchstens quadratische Faktoren) kannten bereits BERNOULLI vor dreihundert und HERMITE vor über hundert Jahren; wir werden diese gleich kennenlernen.

Im Falle der elementaren transzendenten Funktionen, d.h. der trigonometrischen und Exponentialfunktionen und ihrer Umkehrfunktionen geht von der mathematischen Theorie her auf etwa genauso alte Sätze von LIOUVILLE zurück, aus denen RISCH 1969 einen Algorithmus machte.

Am schwierigsten sind algebraische Integranden, d.h. solche, die auch Wurzeln aus Polynomen enthalten. Im allgemeinsten Fall, wenn es sich um Polynome in transzendenten Funktionen handelt, sind bislang noch keine allgemein anwendbare Algorithmen bekannts. Bei rein algebraischen Integranden betrachtet die einzig bekannte Strategie die RIEMANNsche Fläche des Integranden, was sehr aufwendig sein kann;

siehe dazu

JAMES HAROLD DAVENPORT: Integration of algebraic functions, *Springer Lecture notes in Computer Science* **102**, 1981.

§ 1: Integration rationaler Funktionen

Bekanntlich muß die Stammfunktion einer rationalen Funktion im allgemeinen nicht mehr rational sein; beispielsweise ist die Stammfunktion von $1/x$ der natürliche Logarithmus. Die Strategie zur Integration rationaler Funktionen besteht darin, die Anteile mit rationalen und logarithmischen Stammfunktionen zu identifizieren und dann getrennt zu bearbeiten.

a) Partialbruchzerlegung

Die klassische, aus der Analysis-Grundvorlesung bekannte Methode zur Integration rationaler Funktionen beruht auf der Partialbruchzerlegung des Integranden. Auch effizientere Methode beruhen auf Partialbruchzerlegungen, allerdings kann man dabei verschieden weit gehen.

Grundlage jeder Partialbruchzerlegung ist das folgende

Lemma: k sei ein Körper und $f, g \in k[X]$ seien Polynome für die gelte $\deg f < \deg g$. Ist $g = g_1 g_2$ ein Produkt zweier teilerfremder Faktoren g_1 und g_2, so gibt es Funktionen f_1, f_2 mit $\deg f_1 < \deg g_1$ und $\deg f_2 < \deg g_2$, so daß gilt

$$\frac{f}{g} = \frac{f_1}{g_1} + \frac{f_2}{g_2} .$$

Beweis: Da g_1 und g_2 teilerfremd sind, gibt es Polynome $h_1, h_2 \in k[X]$, so daß $h_1 g_1 + h_2 g_2 = 1$ ist. Damit ist

$$\frac{f}{g} = \frac{f(h_1 g_1 + h_2 g_2)}{g_1 g_2} = \frac{f h_2}{g_1} + \frac{f h_1}{g_2} .$$

Wir dividieren die Zähler mit Rest durch die Nenner:

$$f h_2 : g_1 = q_1 \quad \text{Rest } f_1 \quad \text{und} \quad f h_1 : g_2 = q_2 \quad \text{Rest } f_2 ;$$

dann ist

$$\frac{f}{g} = q_1 + q_2 + \frac{f_1}{g_1} + \frac{f_2}{g_2}$$

mit $\deg f < \deg g$, $\deg f_1 < \deg g_1$ und $\deg f_2 < \deg g_2$. Wäre $q_1 + q_2$ ein von Null verschiedenes Polynom, müßte auf der rechten Seite bei der Addition zu einem einzigen Bruch eine Summe entstehen, in der der Zähler mindestens den gleichen Grad hätte wie der Nenner. Dies ist unmöglich, da die Summe gleich f/g sein muß. Somit ist wie gewünscht

$$\frac{f}{g} = \frac{f_1}{g_1} + \frac{f_2}{g_2}\,.$$

■

Induktiv folgt sofort

Lemma: k sei ein Körper und $f, g \in k[X]$ seien Polynome für die gelte $\deg f < \deg g$. Ist $g = g_1 \cdots g_r$ ein Produkt paarweise teilerfremder Faktoren $g_1, \ldots, g_r$, so gibt es Funktionen $f_1, \ldots, f_r$ mit $\deg f_i < \deg g_i$, so daß gilt

$$\frac{f}{g} = \frac{f_1}{g_1} + \ldots + \frac{f_r}{g_r}\,.$$

■

Der obige Beweis ist zwar im Prinzip konstruktiv; nachdem wir wissen, daß es eine solche Zerlegung gibt, können wir sie uns allerdings auch einfacher verschaffen: Wir machen für die Zähler f_i einen Ansatz

$$f_i = a_{im}X^m + a_{i,m-1}X^{m-1} + \ldots + a_{i1}X + a_{i0}$$

mit unbestimmten Koeffizienten a_{ij}, wobei $m = \deg g_i - 1$ ist. Dann bringen wir die Summe auf der rechten Seite auf den Nenner $g = g_1 \cdots g_r$ und machen Koeffizientenvergleich im Zähler. Dies führt auf ein lineares Gleichungssystem für die Koeffizienten a_{ij} und damit schnell und einfach zu den Polynomen f_i.

Die extremste Form der Partialbruchzerlegung besteht darin, daß wir einen Körper k wählen, über dem das Nennerpolynom g ganz in Linearfaktoren zerfällt: Ist $g = (X-c_1)^{e_1} \cdots (X-c_r)^{e_r}$, so gibt es Polynome f_i

vom Grad kleiner e_i, für die

$$\frac{f}{g} = \frac{f_1}{(X-c_1)^{e_1}} + \ldots + \frac{f_r}{(X-c_r)^{e_r}}$$

ist. Das Polynom f_i kann auch als Polynom in $X - c_i$ geschrieben werden als

$$f_i = \sum_{j=0}^{e_i-1} a_{ij}(X-c_i)^j\,;$$

dann ist

$$\frac{f}{g} = \sum_{i=1}^{r}\sum_{j=0}^{e_i-1} \frac{a_{ij}}{(e_i-j+1)(X-c_i)^{e_i-j}}\,,$$

und in dieser Form läßt sich die Funktion leicht integrieren:

$$\int \frac{f}{g}\,dX = -\sum_{i=1}^{r}\sum_{j=0}^{e_i-2} \frac{a_{ij}}{(e_i-j-1)(X-c_i)^{e_i-j-1}} + \sum_{i=1}^{r} a_{i,e_i-1}\log(X-c_i)\,.$$

Insbesondere folgt

Lemma: Eine rationale Funktion $\frac{f}{g}$ hat genau dann eine logarithmische Stammfunktion, wenn g quadratfrei ist und $\deg f < \deg g$.

■

b) Die Methode von Bernoulli

In $\mathbb{R}[X]$ zerfällt nicht jedes Polynom in Linearfaktoren, jedoch gilt

Lemma: Jedes Polynom aus $\mathbb{R}$ zerfällt in lineare und quadratische Faktoren.

Beweis: Über den komplexen Zahlen zerfällt das Polynom (bis auf einen konstanten Faktor) in ein Produkt von Linearfaktoren $X - c$, wobei c die Nullstellenmenge durchläuft.

Für ein reelles Polynom $f = a_nX^n + \cdots + a_0$ mit Nullstelle c ist

$$f(\overline{c}) = a_n\overline{c}^n + \cdots + a_0 = \overline{a_nc^n + \cdots + a_0} = \overline{f(c)} = 0\,,$$

so daß auch $\overline{c}$ eine Nullstelle ist. Ihre Vielfachheit gleich derer von c, denn für die Ableitungen von f können wir genauso argumentieren wie gerade eben für f. Da das Produkt

$$(X-c)(X-\overline{c}) = X^2 - (c+\overline{c})X + c\overline{c} = X^2 - 2\,\mathfrak{Re}\,c\cdot X + |c|^2$$

reelle Koeffizienten hat, folgt die Behauptung. ∎

Zur Integration einer rationalen Funktion $f/g \in \mathbb{R}(X)$ können wir somit folgendermaßen vorgehen: Zunächst läßt sich durch Polynomdivision mit Rest ein polynomialer Anteil abspalten und problemlos integrieren; damit ist das Problem zurückgeführt auf den Fall, daß $\deg f < \deg g$ ist.

Der Nenner g läßt sich zerlegen in ein Produkt

$$g = \prod_{i=1}^{r} \ell_i^{e_i} \prod_{j=1}^{s} q_j^{d_j}$$

von linearen Faktoren ℓ_i und quadratischen Faktoren q_j; nach der allgemeinen Theorie der Partialbruchzerlegung läßt sich f/g also schreiben als

$$\frac{f}{g} = \sum_{i=1}^{r} \frac{f_i}{\ell_i^{e_i}} + \sum_{j=1}^{s} \frac{h_j}{q_j^{d_j}}$$

mit $\deg f_i < e_i$ und $\deg h_j < 2d_j$.

Wir wissen bereits aus dem letzten Abschnitt, daß sich die erste Summe weiter zerlegen läßt in eine Summe von Brüchen mit konstanten Zählern und ℓ_i-Potenzen als Nennern. Ähnlich können wir wir auch bei den Brüchen mit quadratischen Nennern vorgehen: Wir dividieren den Zähler h_j zunächst durch $q_j^{d_j-1}$; der Quotient sei b_{jd_j-1} und r_{jd_j-1} sei der Rest. Diesen dividieren wir durch $q_j^{d_j-2}$ mit Quotient b_{jd_j-2} und Rest r_{jd_j-2}, und so weiter. Die Quotienten sind jeweils höchstens linear, so daß wir insgesamt eine Darstellung von h_j als Linearkombination von q_j-Potenzen bekommen mit höchstens linearen Funktionen als Koeffizienten. Division durch $q_j^{d_j}$ macht daraus eine Summe von Quotienten linearer Polynome durch Potenzen von q_j, so daß wir insgesamt eine

Darstellung

$$\frac{f}{g} = \sum_{i=1}^{r}\sum_{\nu=1}^{e_i} \frac{a_{i\nu}}{\ell_i^{\nu}} + \sum_{j=1}^{s}\sum_{\mu=1}^{d_j} \frac{b_{j\mu}X + c_{j\mu}}{q_j^{\mu}}$$

bekommen. Dieser Ansatz ist im wesentlichen bereits 1703 bei LEIBNIZ zu finden, allerdings konnte dieser nicht alle auftretenden Summanden integrieren: Dies schaffte erst JOHANN BERNOULLI.

BARON GOTTFRIED WILHELM VON LEIBNIZ (1646–1716) gilt als der letzte Universalgelehrte, der das gesamte Wissen seiner Zeit überblickte. In der Mathematik ist er vor allem berühmt durch die Entwicklung der Infinitesimalrechnung (bezüglich derer es einen langen Prioritätsstreit mit NEWTON gab); Bezeichnungen wie $\frac{dy}{dx}$ und $\int f(x)\,dx$ gehen auf ihn zurück. Durch seine Begründung der symbolischen Logik legte er auch einen wesentlichen Grundstein der späteren Informatik. Weitere Arbeiten befassen sich mit den Naturwissenschaften und der Technik, der Philosophie, Theologie und der Geschichte.

JOHANN BERNOULLI (1667–1748) sollte den Gewürzhandel seiner Familie übernehmen, zeigte aber Zeit seines Lebens kein Interesse an Handel und Wirtschaft. Schließlich erlaubte ihm sein Vater, sich an der Universität Basel für Medizin zu immatrikulieren; tatsächlich aber studierte er wie schon sein zwölf Jahre älterer Bruder JACOB Mathematik. Insbesondere studierten die Brüder die LEIBNIZschen Arbeiten zur Infinitesimalrechnung. 1691 ging er via Genf nach Paris, 1695 folgten Rufe nach Halle und nach Groningen; letzteren nahm er an. Er arbeitete auch dort weiter am Ausbau der Analysis sowie an Problemen der Variationsrechnung.

Bei den linearen Anteilen der Partialbruchzerlegung gibt es natürlich, wie wir bereits im vorigen Abschnitt gesehen haben, keine Schwierigkeiten. Die logarithmischen Anteile bei den quadratischen Nenner sind ebenfalls aus der Analysis-Grundvorlesung bekannt: Für einen quadratischen Faktor

$$(X-c)(X-\overline{c}) = X^2 - 2\,\mathfrak{Re}\,c\cdot X + |c|^2 = X^2 - 2rX + b^2$$

mit zwei konjugiert komplexen Nullstellen ist

$$\int \frac{dx}{x^2 - 2rx + b^2} = \frac{1}{\sqrt{b^2 - r^2}} \arctan \frac{x - r}{\sqrt{b^2 - r^2}} ,$$

wobei die Quadratwurzel eine reelle Zahl ist, da der Realteil r der komplexen Zahl c kleineren Betrag hat als ihr Betrag b.

Den Fall eines linearen Polynoms im Zähler kann man leicht darauf zurückführen, indem man einen Zähler subtrahiert, der ein Vielfaches der Ableitung des Nenners ist; für einen solchen Zähler ist die Stammfunktion einfach das entsprechende Vielfache des Logarithmus des Nenners.

Bleibt noch das Problem der Potenzen quadratischer Nenner; hierfür fand BERNOULLI eine Rekursionsformel, auf die hier nicht weiter eingegangen sei, da wir im nächsten Abschnitt eine allgemeinere und nützlichere entsprechende Formel von HERMITE betrachten werden.

Anwendungen des Verfahrens von BERNOULLI sind im wesentlichen Übungsaufgaben zu den mathematischen Anfängervorlesungen; ansonsten ist es wenig gebräuchlich. Sein hauptsächlicher Nachteil ist die Schwierigkeit, die Faktorisierung des Nenners zu finden: Ist etwa der Nenner g ein über q irreduzibles Polynom vom Grad mindestens fünf, gibt es im allgemeinen keine Möglichkeit, die Faktoren auf einfache Weise hinzuschreiben. Die alternativen Verfahren, mit denen Computeralgebrasysteme wirklich arbeiten, versuchen daher insbesondere, eine Stammfunktion zu bestimmen *ohne* den Nenner allzu weit zu faktorisieren.

c) Die Methode von Hermite

CHARLES HERMITE schlug 1872 die folgende Methode vor, um den rationalen Anteil der Stammfunktion einer rationalen Funktion f/g zu bestimmen: Er startet mit einer quadratfreien Zerlegung $g = \prod g_i^i$ des Nenners; da die Faktoren g_i paarweise teilerfremd sind, führt diese zu einer Partialbruchzerlegung

$$\frac{f}{g} = \frac{f}{\prod_{i=1}^{r} g_i^i} = \sum_{i=1}^{r} \frac{f_i}{g_i^i} .$$

Der Summand f_1/g_1 hat einen quadratfreien Nenner und damit eine Summe von Logarithmen als Stammfunktion. Bei den übrigen Summanden reduziert HERMITE sukzessive die Potenz im Nenner, bis auch hier nur noch das quadratfreie Polynom g_i übrigbleibt.

Um Indizes zu vermeiden, betrachten wir allgemein eine rationale Funktion f/g^i mit quadratfreiem g und $i \geq 2$ und versuchen, ihre Stammfunktion durch die von f/g^{i-1} auszudrücken.

Da g quadratfrei ist, hat es ggT eins mit seiner Ableitung; mit dem erweiterten EUKLIDischen Algorithmus können wir daher Polynome α, β finden, so daß $\alpha g + \beta g' = 1$ ist. Somit ist

$$\int \frac{f}{g^i}\,dx = \int \frac{(f\alpha g + \beta g')}{g^i}\,dx = \int \frac{\alpha f}{g^{i-1}}\,dx + \int \frac{\beta f g'}{g^i}\,dx .$$

Das erste Integral hat bereits Nenner g^{i-1} und kann somit stehenbleiben; am zweiten müssen wir noch arbeiten.

Dazu beachten wir, daß

$$\left(\frac{\beta f}{g^{i-1}}\right)' = \frac{g^{i-1}(\beta f)' - \beta f \cdot (i-1) g^{i-2} g'}{g^{2i-2}} = \frac{(\beta f)'}{g^{i-1}} - (i-1)\frac{\beta f g'}{g^i}$$

ist, also

$$\frac{\beta f g'}{g^i} = \frac{1}{i-1}\frac{(\beta f)'}{g^{i-1}} - \frac{1}{i-1}\left(\frac{\beta f}{g^{i-1}}\right)' .$$

Somit ist

$$\int \frac{\beta f g'}{g^i}\,dx = \frac{1}{i-1}\int \frac{(\beta f)'}{g^{i-1}}\,dx - \frac{1}{i-1}\frac{\beta f}{g^{i-1}}$$

und

$$\begin{aligned}\int \frac{f}{g^i}\,dx &= \int \frac{\alpha f}{g^{i-1}}\,dx + \frac{1}{i-1}\int \frac{(\beta f)'}{g^{i-1}}\,dx - \frac{1}{i-1}\frac{\beta f}{g^{i-1}} \\ &= -\frac{1}{i-1}\frac{\beta f}{g^{i-1}} + \int \frac{\alpha f + (\beta f)'/(i-1)}{g^{i-1}}\,dx ,\end{aligned}$$

womit die gewünschte Reduktion des Nennergrads erreicht ist.

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, der Nachweis, daß e eine transzendente Zahl ist, also keiner algebraischen Gleichung über $\mathbb{Q}$ genügt, 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.

d) Die Methode von Horowitz

Die Methode von HERMITE funktioniert zwar grundsätzlich sehr gut, hat aber aus Sicht eines Programmierers den Nachteil, daß man mehrere verschiedener Algorithmen benötigt und die meisten davon auch noch mehrfach anwenden muß: Zunächst muß eine quadratfreie Zerlegung durchgeführt werden, danach für jeden von deren Faktoren außer dem ersten eine Reduktion des Nennerexponenten sowie anschließend für jeden Term die Bestimmung des logarithmischen Anteils.

Die 1969 von HOROWITZ vorgeschlagene Alternativmethode ist stromlinienförmiger und fast identisch mit einer bereits 1845 von MIKHAIL VASILEVIČ OSTROGRADSKI (1801–1862) angegebenen Algorithmus, der jedoch außerhalb Rußlands in Vergessenheit geraten war.

Wir gehen wieder aus von einem gekürzten Bruch f/g mit der üblichen Bedingung $\deg f < \deg g$. Der größte gemeinsame Teiler von g und g' ist durch alle Faktoren von g teilbar, die dort mit Vielfachheit mindestens zwei vorkommen, und ihre Vielfachheit im ggT ist um eins kleiner als die in g. Mit derselben Vielfachheit kommen sie auch im rationalen Anteil der Stammfunktion vor; wir können diesen Anteil also schreiben als

$$\frac{f_1}{g_1} \quad \text{mit} \quad g_1 = \operatorname{ggT}(g, g') .$$

Damit ist

$$\int \frac{f_1}{g_1}\,dx = \frac{f_1}{g_1} + \int \frac{f_2}{g_2}\,dx , \qquad (*)$$

wobei das zweite Integral auf rein logarithmische Terme führt. Somit ist g_2 das Produkt aller Faktoren von g (über dem Zerfällungskörper), d.h.

$$g_2 = \frac{g}{\operatorname{ggT}(g, g')} = \frac{g}{g_1} .$$

Differentiation von $(*)$ führt auf die Gleichung

$$\frac{f}{g} = \left(\frac{f_1}{g_1}\right)' + \frac{f_2}{g_2} = \frac{g_1 f_1' - f_1 g_1'}{g_1^2} + \frac{f_2}{g_2} = \frac{f_1' - f_1 \frac{g_1'}{g_1}}{g_1} + \frac{f_2}{g_2}$$
$$= \frac{f_1' g_2 - f_1 \frac{g_1' g_2}{g_1} + f_2 g_1}{g} .$$

Der Faktor $g_1' g_2 / g_1$ ist ein Polynom, denn jeder Linearfaktor von g_1 tritt in g_1' mit um eins kleinerer Vielfachheit auf und in g_2 mit Vielfachheit eins. Betrachten wir nur die Zähler, haben wir also eine Identität

$$f = f_1' g_2 - f_1 \frac{g_1' g_2}{g_1} + f_2 g_1$$

von Polynomen.

Die Polynome f_1 und f_2 sind uns nicht bekannt; wir wissen aber, daß f_1 kleineren Grad als g_1 hat und f_2 kleineren Grad als g_2. Somit können wir einen Ansatz mit unbestimmten Koeffizienten machen und erhalten ein lineares Gleichungssystem für diese Koeffizienten. Dessen Lösung liefert uns die Polynome f_1 und f_2 und damit den rationalen Anteil f_1/g_1 der Stammfunktion sowie den Anteil f_2/g_2 des Integranden, dessen Stammfunktion der logarithmische Anteil ist.

e) Die Methode von Rothstein und Trager

Egal ob wir nach HERMITE oder nach HOROWITZ vorgehen, bleibt noch das Problem, den logarithmischen Anteil der Stammfunktion zu finden. Dies liefert eine Methode, die 1976 gleichzeitig und unabhängig voneinander von ROTHSTEIN und TRAGER gefunden wurde.

Lemma: Sind $f, g \in \mathbb{R}[X]$ zwei Polynome mit $\deg f < \deg g$ und g quadratfrei, so ist die Stammfunktion von f/g über $\mathbb{C}$ als Summe von

Logarithmen darstellbar, über $\mathbb{R}$ also als Summe von Logarithmen und Arcustangenstermen.

Beweis: Da der Nenner von g in $\mathbb{C}[X]$ in ein Produkt verschiedener Linearfaktoren zerfällt, ist die Partialbruchzerlegung von f/g eine Summe von Brüchen mit konstantem Zähler und linearem Nenner. Deren Stammfunktionen sind Logarithmen, und falls eine davon komplex sind, lassen sie sich paarweise zu Arkustangenstermen kombinieren. ∎

Der Einfachheit sollten und wollen wir im folgenden voraussetzen, daß f/g ein gekürzter Bruch ist, daß Zähler und Nenner also teilerfremd sind. Außerdem können wir, indem wir den Bruch durch den führenden Koeffizienten von g kürzen, annehmen, daß der führende Koeffizient von g gleich eins ist.

Nach obigem Lemma ist

$$\int \frac{f}{g}\,dx = \sum_{i=1}^{r} a_i \log g_i \qquad (*)$$

mit rationalen Funktionen $g_i \in \mathbb{C}[X]$ und Konstanten $a_i \in \mathbb{C}$. Diese Darstellung ist alles andere als eindeutig, da wir ja Logarithmen gemäß ihrer Funktionalgleichung $\log(uv) = \log u + \log v$ zerlegen und kombinieren können. Die maximal zerlegte Darstellung wäre zwar im wesentlichen eindeutig, allerdings ist sie auch die Darstellung, an der wir das wenigste Interesse haben, denn sie setzt ja voraus, daß wir das Nennerpolynom in Linearfaktoren zerlegt haben, was für irreduzible Polynome hohen Grades große Körpererweiterungen erfordert.

Um die mögliche Vielfalt der Darstellungen einzuschränken und nicht zu weit zerlegen zu müssen, stellen wir folgende Forderungen:

1.) *Alle g_i sind Polynome mit führendem Koeffizienten eins.* Das läßt sich problemlos erreichen, da der Logarithmus einer rationalen Funktion die Differenz der Logarithmen von Zähler und Nenner ist.

2.) *Die Polynome g_i sind paarweise teilerfremd.* Auch damit gibt es keine Probleme, denn ist h ein gemeinsamer Teiler von g_i und g_j, etwa $g_i = hG_i$ und $g_j = hG_j$, so ist

$$a_i \log g_i + a_j \log g_j = a_i \log G_i + a_j \log G_j + (a_i + a_j) \log h\,.$$

3.) *Alle Koeffizienten a_i sind verschieden.* Ist $a_i = a_j$, können wir nämlich die Terme $a_i \log g_i$ und $a_j \log g_j$ zu $a_i \log(g_i g_j)$ zusammenfassen.

Wir gehen nun aus von einer Darstellung

$$\int \frac{f}{g}\,dx = \sum_{i=1}^{r} a_i \log g_i\,,$$

die die Bedingungen *1.) – 3.)* erfüllt, und versuchen, die Polynome g_i sowie die Koeffizienten a_i so darzustellen, daß wir sie mit den verfügbaren Informationen über f und g berechnen können.

Differentiation von $(*)$ führt auf die Gleichung

$$\frac{f}{g} = \sum_{i=1}^{r} a_i \frac{g_i'}{g_i}\,.$$

Da f und g teilerfremd sind, ist dann g das Produkt der g_i. Da g quadratfrei vorausgesetzt war, sind auch die g_i quadratfrei, und somit läßt sich auch keiner der Brüche g_i'/g_i kürzen. Damit gibt es keinerlei Kürzungen, wenn wir die Summe auf der rechten Seite in der üblichen Weise ausrechnen, indem wir alle Summanden auf den Hauptnenner bringen.

Dabei müssen wir den i-ten Summanden mit dem Produkt aller g_j außer g_i selbst multiplizieren; wir setzen zur Abkürzung

$$u_i = \prod_{j \neq i} g_i = \frac{g}{g_i}\,.$$

Nach der LEIBNIZ-Regel ist dann

$$g' = \left(\prod_{i=1}^{r} g_i\right)' = \sum_{i=1}^{r} g_i' u_i\,,$$

und die obige Summe wird zu

$$\frac{f}{g} = \sum_{i=1}^{r} a_i \frac{g_i' u_i}{g} = \frac{\sum_{i=1}^{r} a_i g_i' u_i}{g}\,, \quad \text{d.h.} \quad f = \sum_{i=1}^{r} a_i g_i' u_i\,.$$

Dann ist

$$g_i = \mathrm{ggT}(g_i, 0) = \mathrm{ggT}\Big(g_i, f - \sum_{j=1}^{r} a_j g_i' u_j\Big) = \mathrm{ggT}\big(g_i, f - a_i g_i' u_i\big)\,,$$

denn für alle $j \neq i$ ist g_i ein Teiler von u_j, so daß wir den j-ten Summanden weglassen können, ohne den ggT zu verändern. Ebenso können wir natürlich auch Vielfache von g_i addieren, ohne den ggT zu ändern; damit ist auch

$$g_i = \mathrm{ggT}\Big(g_i, f - a_i \sum_{j=1}^{n} g_j' u_j\Big) = \mathrm{ggT}\big(g_i, f - a_i g'\big)\,.$$

Für ein $k \neq i$ ist

$$\begin{aligned}\mathrm{ggT}\Big(g_k, f - a_i \sum_{j=1}^{n} g_j' u_j\Big) &= \mathrm{ggT}\Big(g_k, \sum_{j=1}^{n} (a_j - a_i) g_j' u_j\Big) \\ &= \mathrm{ggT}\big(g_k, (a_k - a_i) g_k' u_k\big) = 1\,,\end{aligned}$$

denn u_k ist teilerfremd zu g_k, da die verschiedenen g_j paarweise teilerfremd sind, g_k' ist teilerfremd zu g_k, da g_k quadratfrei vorausgesetzt war, und $a_k - a_i \neq 0$, da alle a_i verschieden sind.

Multiplikation des ersten Arguments mit einem zum zweiten Argument teilerfremden Faktor ändert nichts am ggT; daher ist auch

$$g_i = \mathrm{ggT}(g_i, f - a_i g') = \mathrm{ggT}(g, f - a_i g')\,.$$

Rechts kennen wir alle vorkommenden Größen bis auf a_i. Nun ist aber für jedes a der ggT von g und $f - ag'$ ein Teiler von g und die verschiedenen solchen ggT sind teilerfremd zueinander, da g und g' keinen gemeinsamen Teiler haben. Da g bereits das Produkt von g_1 bis g_r ist, müssen g und $f - ag'$ daher teilerfremd sein, sofern a keines der a_i ist. Somit sind die a_i genau diejenigen Zahlen a, für die g und $f - ag'$ einen gemeinsamen Teiler positiven Grades haben, also die Nullstellen der Resultante von g und $f - ag'$.

Damit ist klar, wie wir die Zahlen a_i und die Polynome g_i finden können: Wir bestimmen alle Nullstellen a_i von $\mathrm{Res}(g, f - ag')$ und berechnen für

jede dieser Nullstellen $g_i = \mathrm{ggT}(g, f - ag')$. Falls wir den ggT jeweils so normieren, daß sein höchster Koeffizient gleich eins ist, ist dann

$$\int \frac{f}{g}\,dx = \sum_{i=1}^{r} a_i g_i\,.$$

§ 2: Elementare Erweiterungen von Differentialkörpern

Wenn wir die Nullstellen eines Polynoms $f \in k[X]$ bestimmen möchten, brauchen wir im allgemeinen eine Erweiterung des Körpers k, den Zerfällungskörper von k, und die Struktur dieser Körpererweiterung gibt uns Informationen über Nullstellenmenge von f. Entsprechend können wir auch für Differentialgleichung Körpererweiterungen definieren, allerdings müssen wir dazu den Körpern eine zusätzliche Struktur geben, damit wir von Ableitungen reden können.

In diesem Paragraphen betrachten wir grundsätzlich nur Körper der Charakteristik null.

a) Differentialkörper

Definition: Ein Differentialkörper K über dem Konstantenkörper k ist ein Erweiterungskörper K von k zusammen mit einer k-linearen Abbildung $D\colon K \to K$, für die gilt
a) $D(f) = 0 \iff f \in k$
b) $D(fg) = D(f)g + fD(g)$ für alle $f, g \in K$.

Wir schreiben kurz $D(f) = f'$.

Typisches Beispiel eines Differentialkörpers ist etwa der rationale Funktionenkörper $\mathbb{R}(X)$ bestehend aus allen Quotienten von reellen Polynomen in X mit der üblichen Differentiation. Wir können aber auch noch weitere Funktionen dazu nehmen, beispielsweise die Funktion $y = \sqrt{x}$ mit $D(y) = 1/2y$ oder $z = \ln x$ mit $D(z) = 1/x$.

In einem Differentialkörper gelten die üblichen Ableitungsregeln, beispielweise auch die Quotientenregel: Da $\frac{f}{g} \cdot g = f$ ist, führt die Produkt-

regel auf

$$D\left(\frac{f}{g}\right)\cdot g + \frac{f}{g}\cdot D(g) = D(f) \Longrightarrow D\left(\frac{f}{g}\right) = \frac{gD(f) - fD(g)}{g^2}\,.$$

Die Kettenregel hat insofern keine Entsprechung, als wir in einem Differentialkörper keine Verkettungsoperation definiert haben. Ein Spezialfall ist aber sinnvoll:

Lemma: Ist K ein Differentialkörper mit Konstantenkörper k und P ein Polynom aus $k[X]$, so gilt für jedes Element $f \in K$ die Kettenregel $D\big(P(f)\big) = P'(f)\cdot D(f)$.

Beweis: Wegen der Linearität von D genügt es, dies für X-Potenzen P zu beweisen, und da folgt die Behauptung induktiv aus der Produktregel. ∎

b) Erweiterungen

Schon wenn wir eine rationale Funktion, d.h. ein Element $f \in \mathbb{R}(X)$, integrieren wollen, brauchen wir zusätzlich zu den Elementen des Differentialkörpers $\mathbb{R}(X)$ im allgemeinen auch noch Logarithmen; die Stammfunktion von f liegt also in einem größeren Differentialkörper – genau wie wir auch bei der Bestimmung der Nullstellen eines Polynoms $f \in \mathbb{Q}[X]$ im allgemeinen einen größeren Körper als $\mathbb{Q}$ brauchen.

Unter einer *Körpererweiterung* L/K eines Körpers K versteht man bekanntlich einen Körper L, der K als Teilkörper enthält. Diese Erweiterung macht L zu einem K-Vektorraum; falls dieser endlichdimensional ist, reden wir von einer *endlichen Körpererweiterung.*

Für jedes Element $t \in L$ eines Erweiterungskörpers L/K bezeichnen wir mit $K(t)$ den kleinsten Teilkörper von L, der sowohl K als auch das Element t enthält. Falls $K(t)/K$ eine endliche Körpererweiterung ist, sagen wir, t sei *algebraisch* über K; andernfalls heißt t *transzendent.*

Falls t algebraisch über K ist, können die t-Potenzen t^n für $n \in \mathbb{N}_0$ nicht allesamt linear unabhängig sein; damit ist klar, daß t einer Polynomgleichung $t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ mit Koeffizienten aus K genügen muß, und umgekehrt ist auch jedes Element, das einer solchen

Gleichung genügt, algebraisch über K, denn die Potenzen $1, t, \ldots, t^{n-1}$ erzeugen dann den K-Vektorraum $K(t)$:

$$\begin{aligned}
t^n &= -a_{n-1}t^{n-1} - \cdots - a_0\,,\\
t^{n+1} &= -a_{n-1}t^n - \cdots - a_0 t \quad \textit{usw.},\\
t^{-1} &= -a_0^{-1}\big(a_{n-1}t^{n-2} + \cdots + a_2 t + a_1\big)\,,\\
t^{-2} &= -a_0^{-1}\big(a_{n-1}t^{n-3} + \cdots + a_2 + a_1 t^{-1}\big) \quad \textit{usw.}
\end{aligned}$$

Für transzendentes t müssen entsprechend alle Potenzen von t über K linear unabhängig sein; $K(t)$ enthält also einen zum Polynomring in einer Veränderlichen isomorphen Teilring, den wir mit $K[t]$ bezeichen, und $K(t)$ ist der Quotientenkörper von $K[t]$.

Definition: *a)* Eine Erweiterung L/K eines Differentialkörpers K heißt Erweiterung von Differentialkörpern, wenn auch L ein Differentialkörper ist und die Differentiation auf K die Einschränkung derer auf L ist.
b) Eine Erweiterung L/K von Differentialkörpern heißt *logarithmisch,* wenn L über K durch ein Element g erzeugt wird derart, daß $D(g) = D(f)/f$ ist mit einem $f \in K$.
c) Eine Erweiterung L/K von Differentialkörpern heißt *exponentiell,* wenn L über K durch ein Element g erzeugt wird derart, daß $D(g) = fg$ ist mit einem $f \in K$.
d) Eine Erweiterung L/K von Differentialkörpern heißt *Konstantenerweiterung,* wenn L über K durch ein Element g erzeugt wird derart, daß $D(g) = 0$ ist. *e)* Eine Erweiterung L/K von Differentialkörpern heißt *elementar,* wenn es eine Folge von Zwischenkörpern $K = L_0 < L_1 < \cdots < L_r = L$ gibt derart, daß jede der Erweiterungen L_{i+1}/L_i endlich, logarithmisch, exponentiell oder eine Konstantenerweiterung ist.

Über $K = \mathbb{R}(X)$ erzeugt also beispielsweise für $f \in \mathbb{R}(x)$ die Funktion $g = \log f$ eine logarithmische Erweiterung und $h = e^f$ eine exponentielle.

Definition: Eine Funktion $f \in K$ aus einem Differentialkörper K heißt

elementar integrierbar, wenn es eine elementare Erweiterung L/K gibt und ein Element $g \in L$ mit $D(g) = f$.

Der Name „elementar“ kommt daher, daß Exponentialfunktion, Logarithmus, trigonometrische Funktionen und deren Umkehrfunktionen klassisch unter dem Namen „elementare transzdendete Funktionen“ bekannt waren. Über $\mathbb{C}$ lassen sich alle diese Funktionen durch Exponentialfunktionen und Logarithmen ausdrücken. Beispielsweise liegen für $f \in \mathbb{C}(X)$ auch $\sin f, \cos f, \tan f$ in einer elementaren Erweiterung von $\mathbb{C}(X)$, denn $u = e^{if}$ definiert eine exponentielle Erweiterung, und

$$\sin f = \frac{1}{2i}\left(u - \frac{1}{u}\right), \quad \cos f = \frac{1}{2}\left(u + \frac{1}{u}\right), \quad \tan f = \frac{u - u^{-1}}{i(u + u^{-1})}.$$

c) Algebraische Erweiterungen von Differentialkörpern

L/K sei eine endliche Körpererweiterung. Dann ist jedes Element $t \in L$ algebraisch über K, es gibt also ein Polynom $P \in K[T]$, so daß $P(t) = 0$ ist. Sind P, Q zwei solche Polynome, verschwindet auch deren ggT in t; daher gibt es ein bis auf multiplikative Konstanten eindeutig bestimmtes kleinstes von Null verschiedenes Polynom, das in t verschwindet. Dieses Polynom bezeichnen wir als das *Minimalpolynom* von t. Es hat t als einfache Nullstelle, denn sonst wäre t auch Nullstelle der Ableitung, die nicht das Nullpolynom sein kann, da wir in diesem Paragraphen nur Körper der Charakteristik null betrachten.

Z sei ein Körper, über dem P ganz in Linearfaktoren zerfällt, und $t = t_1, t_2, \ldots, t_n$ seien die sämtlichen Nullstellen von P. Wir bezeichnen die Summe der t_i als die *Spur* von t und ihr Produkt als Norm:

$$\mathrm{Sp}_{K(t)/K}(t) = \sum_{i=1}^{n} t_i \quad \text{und} \quad \mathrm{N}_{K(t)/K}(t) = \prod_{i=1}^{n} t_i .$$

Nach dem Wurzelsatz von VIÈTE ist die Sput von t gleich dem Negativen des Koeffizienten von T^{n-1} in P, und die Norm ist gleich $(-1)^n$ mal dem konstanten Koeffizienten. Insbesondere liegen also beide im Körper K.

Im allgemeinen wird $K(t)$ ein kleinerer Körper als L sein; ist r die Dimension von L als $K(t)$-Vektorraum, so definieren wir Norm und

Spur von t in der Erweiterung L/K als

$$\mathrm{Sp}_{L/K}(t) = r\,\mathrm{Sp}_{K(t)/K}(t) \quad \text{und} \quad \mathrm{N}_{L/K}(t) = \mathrm{N}_{K(t)/K}(t)^r .$$

In dieser ist für jedes $t \in L$ aus einer endlichen Erweiterung L/K eine Norm und eine Spur definiert; sie sind allerdings so nicht gerade einfach auszurechnen. Dazu empfiehlt sich eine andere Konstruktion:

Wir betrachten zunächst wieder den Fall $L = K(t)$, wobei t das Minimalpolynom

$$P(T) = T^n + a_{n-1}T^{n-1} + \cdots + a_1 T + a_0 \in K[T]$$

habe, und wir betrachten die Abbildung $\mu_t : L \to L$, die jedes Element von L mit t multipliziert. Die Elemente $1, t, \ldots, t^{n-1}$ bilden eine K-Vektorraumbasis von L, und

$$\mu_t(1) = 1, \quad \mu_t(t) = t^2, \quad \ldots, \quad \mu_t(t^{n-2}) = t^{n-1},$$
$$\mu_t(t^{n-1}) = t^n = -a_{n-1}T^{n-1} - \cdots - a_1 T - a_0 .$$

Die Abbildungsmatrix M_t von μ_t bezüglich der betrachteten Basis ist daher

$$M_t = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & -a_{n-2} \\ 0 & 0 & 0 & \ldots & 1 & -a_{n-1} \end{pmatrix}$$

Ihre Spur, d.h. die Summe der Diagonaleinträge, ist $-a_{n-1}$ und damit gleich der Spur von t.

Ihre Determinante können wir leicht nach dem LAPLACEschen Entwicklungssatz berechnen, indem wir nach der ersten Zeile entwicklen:

$$\begin{vmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & -a_{n-2} \\ 0 & 0 & 0 & \ldots & 1 & -a_{n-1} \end{vmatrix} = (-1)^n a_0 \begin{vmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 1 \end{vmatrix},$$

und da rechts die Determinante der Einheitsmatrix steht, ist das einfach gleich $(-1)^n a_0$, also die Norm von t.

Ist L ein Erweiterungskörper von $K(t)$, so wählen wir zunächst eine $K(t)$-Basis $f_1, \ldots, f_r$ von $L/K(t)$ und wie oben die K-Basis $1, t, \ldots, t^{n-1}$ von $K(t)/K$. Dann ist

$$f_1, tf_1, \ldots, tf_1, f_2, tf_2, \ldots, tf_2, \ldots, f_r, tf_r, \ldots, tf_r$$

eine K-Basis von L, bezüglich derer die Abbildungsmatrix der Multiplikation mit t aus r Blöcken besteht, die allesamt gleich der oben betrachteten Matrix M_t sind. Somit ist die Spur der Abbildungsmatrix gleich dem r-fachen der Spur von M_t, also gleich $\mathrm{Sp}_{L/K}(t)$, und die Determinante ist die r-te Potenz der Determinante von M_t, also $\mathrm{N}_{L/K}(t)$.

Wir können Spur und Norm daher auch alternativ so definieren:

Definition: L/K sei eine endliche Körpererweiterung und für jedes $t \in L$ sei M_t die Abbildungsmatrix der K-linearen Abbildung $L \to L$, die jedes Element von L mit t multipliziert. Die Abbildungen

$$\mathrm{N}_{L/K}\colon \begin{cases} L \to K \\ x \mapsto \det M_t \end{cases} \quad \text{und} \quad \mathrm{Sp}_{L/K}\colon \begin{cases} L \to K \\ x \mapsto \mathrm{Sp}\, M_t \end{cases}$$

heißen *Norm* und *Spur* von L/K.

(Aus der Linearen Algebra ist bekannt, daß beide Abbildungen unabhängig sind von der Basis bezüglich derer die Abbildungsmatrizen betrachtet werden: Berechnet man das charakteristische Polynom $\det(M_x - \lambda E)$ der $n \times n$-Matrix M_t nach der definierenden Formel der Determinante, sieht man, daß $\mathrm{Sp}_{L/K}$ bis auf einen Vorzeichenfaktor $(-1)^{n-1}$ gleich dem Koeffizienten von λ^{n-1} ist, und setzt man $\lambda = 0$, so ist klar, daß der konstante Term des charakteristischen Polynoms gleich der Determinante ist. Das charakteristische Polynom ist unabhängig von der Basis, und damit auch Norm und Spur.)

Aufgrund dieser neuen Definition ist klar, daß die Spur eine K-lineare Abbildung ist und die Norm eine multiplikative; nicht so klar ist, was die Spur mit Produkten macht. Für Potenzen von t ist klar, daß $M_{t^r} = M_t^r$ ist; da es zu jeder linearen Abbildung eine Basis gibt, bezüglich

derer die Abbildungsmatrix Dreiecksgestalt hat, ist außerdem klar, daß die Spur einer Matrix gleich der Summe ihrer Eigenwerte ist, wobei jeder Eigenwert so oft gezählt wird, wie seine algebraische Vielfachheit angibt. Um damit etwas anfangen zu können, sollten wir die Eigenwerte von M_t bestimmen.

Wir betrachten zunächst nur den Fall $L = K(t)$, wo wir mit der Basis aus den t-Potenzen arbeiten können. Nach dem LAPLACEschen Entwicklungssatz, angewandt auf die erste Zeile, ist

$$\begin{aligned}
\det(M_t - \lambda E) &= \begin{vmatrix} -\lambda & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & -\lambda & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & -\lambda & \ldots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\lambda & -a_{n-2} \\ 0 & 0 & 0 & \ldots & 1 & -a_{n-1} - \lambda \end{vmatrix} \\
&= -\lambda \begin{vmatrix} -\lambda & 0 & \ldots & 0 & -a_1 \\ 1 & -\lambda & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & -\lambda & -a_{n-2} \\ 0 & 0 & \ldots & 1 & -a_{n-1} - \lambda \end{vmatrix} \\
&\quad + (-1)^n a_0 \begin{vmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 1 \end{vmatrix}.
\end{aligned}$$

Der zweite Summand ist einfach $(-1)^n a_0$; auf den ersten können wir Induktion nach n anwenden und erhalten

$$\det(M_t - \lambda E) = (-1)^n \left(\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0\right) = (-1)^n P(\lambda)\,.$$

Die Eigenwerte von M_t sind also genau die Nullstellen t_i des Minimalpolynoms, und da die Eigenwerte von M_t^r die r-ten Potenzen der Eigenwerte von M_t sind, folgt:

$$\mathrm{Sp}_{L/K}(t^r) = t_1^r + \cdots + t_n^r\,.$$

Dies läßt sich auch anders ausdrücken: Ist Z/K ein Körper, über dem P in Linearfaktoren zerfällt, so haben wir für jedes i einen Isomorphismus σ_i von $K(t)$ auf einen Unterkörper von M, indem wir K festlassen und t auf t_i abbilden. Natürlich ist dabei $\sigma_i(t^r) = t_i^r$, und da die $K(t)$ über K eine Basis aus Potenzen von t hat, folgt damit allgemein, daß für jedes $f \in K(t)$ gilt

$$\mathrm{Sp}_{K(t)/K}(f) = \sigma_1(f) + \cdots + \sigma_n(f)\,.$$

Norm und Spur spielen eine zentrale Rolle für endliche Körpererweiterungen im Rahmen der klassischen Algebra und Zahlentheorie. Ihre Nützlichkeit auch für Differentialkörpern ergibt sich unter anderem aus dem folgenden

Lemma: *a)* Ist K ein Differentialkörper und L/K eine endliche Erweiterung von K, so läßt sich die Differentiation auf K in genau einer Weise fortsetzen zu einer Differentiation von L.
b) Für alle Elemente $f \in L$ ist

$$\mathrm{Sp}_{L/K}\big(D(f)\big) = D\big(\mathrm{Sp}_{L/K}(f)\big)$$

und

$$\mathrm{Sp}_{L/K}\left(\frac{D(f)}{f}\right) = \frac{D\big(\mathrm{N}_{L/K}(f)\big)}{\mathrm{N}_{L/K}(f)}\,.$$

Beweis: a) Wir können ohne Beschränkung der Allgemeinheit annehmen, daß $L = K(t)$ von einem einzigen Element erzeugt ist, denn jede endliche Erweiterung läßt sich schrittweise aus solche Erweiterungen aufbauen. Das Minimalpolynom von t sei

$$P(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \quad \text{mit} \quad a_i \in K\,.$$

Dann ist auch

$$D\big(P(t)\big) = \sum_{i=0}^{n} D(a_i)t^i + \sum_{i=0}^{n}(i-1)a_i t^{i-1}D(t) = 0\,,$$

also ist

$$D(t) = -\frac{\sum_{i=0}^{n} D(a_i)t^i}{\sum_{i=0}^{n}(i-1)a_i t^{i-1}} = -\frac{\sum_{i=0}^{n} D(a_i)t^i}{P'(t)}\,.$$

(Bei der Division durch $P'(t)$ gibt es keine Probleme, da das Minimalpolynom in Charakteristik null keine mehrfachen Nullstellen haben kann.)

Somit ist $D(t)$ durch die Differentiation von K und das Minimalpolynom P eindeutig bestimmt, und da jedes Element von $K(t)$ eindeutig als Polynom vom Grad kleiner n in t geschrieben werden kann, können wir in eindeutiger Weise fortsetzen auf $L = K(t)$. Die k-Linearität der Fortsetzung ist klar, und auch die Produktregel läßt sich leicht nachrechnen.

b) Auch hier beschränken wir uns zunächst auf den Fall $L = K(f)$. Dazu betrachten wir einen Körper Z/K, über dem das Minimalpolynom P von f in Linearfaktoren zerfällt, und wir betrachten wieder die Isomorphismen σ_j von $K(f)$ auf Teilkörper von Z, die K festlassen und f auf die j-te Nullstelle f_j von P abbilden. Dann ist

$$\mathrm{Sp}_{L/K}(f) = \sigma_1(f) + \cdots + \sigma_n(f) = f_1 + \cdots + f_n$$

und

$$\mathrm{Sp}_{L/K}\big(D(f)\big) = \sigma_1\big(D(f)\big) + \cdots + \sigma_n\big(D(f)\big)\,.$$

Nach *a)* ist

$$D(f) = -\frac{\sum_{i=0}^{n} D(a_i)t^i}{\sum_{i=0}^{n}(i-1)a_i f^{i-1}} = -\frac{\sum_{i=0}^{n} D(a_i)f^i}{P'(f)}\,,$$

also –da alle a_i, $D(a_i)$ sowie alle Koeffizienten von P' in K liegen–

$$\begin{aligned} D\big(\sigma_j(f)\big) &= -\frac{\sum_{i=0}^{n} D(a_i)\sigma_j(f)^i}{P'\big(\sigma_j(f)\big)} = \sigma_j\left(-\frac{\sum_{i=0}^{n} D(a_i)f^i}{P'(f)}\right) \\ &= \sigma_j\big(D(f)\big)\,. \end{aligned}$$

Somit ist

$$\begin{aligned} \mathrm{Sp}_{L/K}\big(D(f)\big) &= \sum_{j=1}^{n}\sigma_j\big(D(f)\big) = \sum_{j=1}^{n} D\big(\sigma_j(f)\big) \\ &= D\left(\sum_{j=1}^{n}\sigma_j(f)\right) = D\big(\mathrm{Sp}_{L/K}(f)\big)\,. \end{aligned}$$

Nach der Produktregel ist

$$D\big(\mathrm{N}_{L/K}(f)\big) = D\left(\prod_{j=1}^{n}\sigma_j(f)\right) = \sum_{j=1}^{n}\prod_{\substack{\ell=1\\ \ell\neq j}}^{n}\sigma_\ell(f)\cdot D\big(\sigma_j(f)\big)$$

$$= \sum_{j=1}^{n}\prod_{\substack{\ell=1\\ \ell\neq j}}^{n}\sigma_\ell(f)\cdot\sigma_j\big(D(f)\big)$$

und damit

$$\frac{D\big(\mathrm{N}_{L/K}(f)\big)}{\mathrm{N}_{L/K}(f)} = \sum_{j=1}^{n}\frac{\sigma_j\big(D(f)\big)}{\sigma_i(f)} = \sum_{j=1}^{n}\sigma_j\left(\frac{D(f)}{f}\right) = \mathrm{Sp}_{L/K}\left(\frac{D(f)}{f}\right).$$

Dies beweist die Behauptung im Fall $L = K(f)$. Ist L ein größerer Körper, etwa ein r-dimensionaler $K(t)$-Vektorraum, wird die Spur einfach mit r multipliziert und die Norm mit r potenziert. Die erste Formel ist damit völlig unproblematisch und die zweite eine einfache Anwendung der Produktregel. ∎

d) Bewertungen

Zur genaueren Untersuchung transzendenter Erweiterungen brauchen wir Aussagen über die multiplikative Struktur von Ableitungen. Wie wir wissen, ist hier $K(t)$ der Quotientenkörper des Rings

$$K[t] = \Big\{\sum_{i=0}^{n} a_i t^i \;\Big|\; n\in\mathbb{N}_0, a_i\in K\Big\}.$$

Da T transzendent ist, ist dieser Ring isomorph zum Polynomring $K[T]$ über K in einer unbestimmten T; wir wollen daher die Elemente von $K[t]$ auch kurz als Polynome bezeichnen.

Da $K[T]$ faktoriell ist, ist es auch $K[t]$; wir können daher für jedes von Null verschiedenen Element von $K(t) = \mathrm{Quot}\,K[t]$ Zähler und Nenner als Produkte von Potenzen irreduzibler Polynome schreiben. Kombinieren wir die beiden Darstellungen, erhalten wir ein Produkt von Potenzen mit positiven und negativen Exponenten. Auch diese Zerlegung ist bis

auf Einheiten, d.h. bis auf Assoziiertheit, eindeutig. Wenn wir uns auf normierte Polynome beschränken und noch eine Einheit als Vorfaktor zulassen, ist die Zerlegung eindeutig.

Definition: Hat $f\in K(t)^\times$ die Zerlegung $f = c\prod_{i=1}^{r} p_i^{e_i}$ mit $c\in K^\times$ und normierten irreduziblen Polynomen p_i, so setzen wir für ein normiertes irreduzibles Polynom $p\in K[t]$

$$v_p(f) = \begin{cases} e_i & \text{falls } p = p_i \\ 0 & \text{sonst}\end{cases}.$$

v_p heißt die diskrete Bewertung von $K(t)$ zur Basis p.

Aufgrund der Faktorialität von $K[t]$ und der üblichen Regeln der Bruchrechnung folgt sofort

Lemma: Für zwei Elemente $f, g\in K(t)$ ist $v_p(fg) = v_p(f) + v_p(g)$ und

$$v_p(f+g)\begin{cases} = \max\big(v_p(f), v_p(g)\big) & \text{falls } v_p(f)\neq v_p(g)\\ \geq v_p(f) & \text{falls } v_p(f) = v_p(g)\end{cases}.$$

∎

Um die Bewertungen von Ableitungen abzuschätzen, betrachten wir zunächst die Ableitung eines Polynoms in t, etwa

$$p(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \quad\text{mit}\quad a_i\in K.$$

Nach der Produktregel ist

$$D\big(p(t)\big) = D(a_n)t^n + D(a_{n-1})t^{n-1} + \cdots + D(a_1)t + D(a_0)$$
$$+ \big(na_n t^{n-1} + (n-1)a_{n-1}t^{n-2} + \cdots + 2a_2 t + a_1\big)D(t).$$

Für ein irreduzibles Polynom p ist dies entweder teilerfremd zu $p(t)$ oder ein Vielfaches von $p(t)$, denn der ggT muß ein Teiler von $p(t)$ sein.

Definition: Ein Polynom $p\in K[t]$ heißt *normal* bezüglich D, wenn p und $D(p)$ teilerfremd sind; es heißt *speziell*, wenn p ein Teiler von $D(p)$ ist.

Man beachte, daß ein *reduzibles* Polynom p weder normal noch speziell sein muß; hier könnte der ggT von p und $D(p)$ auch ein nichttrivialer

Teiler von p sein. „Konstante“ Polynome, d.h. Funktionen $p \in K$, sind in diesem Sinne sowohl speziell als auch normal, da mit p auch $D(p)$ in K liegt und dort natürlich ein Vielfaches von p ist, und natürlich ist auch der ggT zweier Elemente von K in $K[t]$ gleich eins.

Lemma: Ist $K(t)/K$ eine logarithmische Erweiterung von K, so sind alle irreduziblen Polynome $p \in K[t]$ normal.

Beweis: In diesem Fall gibt es ein Element $f \in K$, so daß $D(t) = D(f)/f$ ist; insbesondere liegt also $D(t)$ in K. Ist daher p irreduzibel und normiert (d.h. $a_n = 0$), so zeigt die obige Formel, daß $D(p)$ kleineren Grad als p hat, so daß der ggT gleich eins sein muß. Ein beliebiges irreduzibles Polynom p läßt sich schreiben als Produkt $p = wq$ mit $w \in K$ und einem normierten irreduziblen Polynom q. Falls p und $D(p) = D(w)q + wD(q)$ einen gemeinsamen Teiler in $K[t] \setminus K$ hätten, müßte auch q einer sein; dann wäre aber q auch Teiler von $D(q)$, was wir bereits ausgeschlossen haben. ∎

Damit gibt es in logarithmischen Erweiterungen außer den Elementen von K keine speziellen Polynome. Für exponentiellen Erweiterungen gilt dies nicht mehr: Ist beispielsweise $t = e^g$ mit einem Element $g \in K$, so ist $D(t) = D(g)t$ ein Vielfaches von t, bereits das Polynom t selbst ist also speziell.

Allgemein gilt

Lemma: Ein Produkt $p = w \prod_{i=1}^{r} p_i^{e_i}$ mit irreduziblen Polynomen p_i und einem $w \in K$ ist genau dann speziell, wenn alle p_i speziell sind.

Beweis: Falls alle p_i speziell sind, ist in

$$D(p) = \sum_{j=1}^{r} e_j p_j^{e_j-1} D(p_j) \prod_{\substack{i=1 \\ i\neq j}}^{r} p_i^{e_i}$$

jeder Faktor $D(p_j)$ durch p_j teilbar, und damit auch $D(p)$ durch p.

Umgekehrt ist in obiger Formel für $D(p)$ ist jeder Summand außer möglicherweise dem j-ten durch $p_j^{e_j}$ teilbar; falls $D(p)$ durch p und

damit auch $p_j^{e_j}$ teilbar ist, muß also auch der j-te Summand teilbar sein. Da die p_i voneinander verschieden sind, muß daher $p_j^{e_j}$ ein Teiler von $p_j^{e_j-1} D(p_j)$ sein, also p_j ein Teiler von $D(p_j)$. Somit ist jedes p_j speziell. ∎

Für normale Polynome gilt selbstverständlich keine entsprechende Aussage; beispielsweise kann eine echte Potenz eines Polynoms nie normal sein.

Nach diesen Vorbereitungen können wir damit beginnen, die Bewertungen einer Ableitung abzuschätzen.

Ein Element $f \in L$ mit $v_p(f) = r$ läßt sich schreiben als $f = p^r \cdot \frac{g}{h}$ mit zwei zu p teilerfremden Polynomen g und h. Nach der Produktregel ist

$$D(f) = rp^{r-1} D(p) \frac{g}{h} + p^r D\left(\frac{g}{h}\right) ,$$

wobei

$$D\left(\frac{g}{h}\right) = \frac{hD(g) - gD(h)}{g^2}$$

einen zu p teilerfremden Nenner hat und im Zähler ein Polynom, das möglicherweise durch p teilbar sein könnte, d.h.

$$v_p\left(D\left(\frac{g}{h}\right)\right) \geq 0 .$$

Somit ist stets

$$v_p\big(D(f)\big) \geq v_p(f) - 1 \quad \text{und} \quad D\left(\frac{D(f)}{f}\right) \geq -1$$

mit Gleichheit im Falle $v_p(f) \neq 0$ genau dann, wenn p normal ist. Für alle p gilt

$$v_p(f) = 0 \Longrightarrow v_p\big(D(f)\big) \geq 0 .$$

...

e) Der Satz von Liouville

Satz: K sei ein Differentialkörper mit Konstantenkörper k. Falls es zu einem Element $f \in K$ ein elementare Erweiterung L/K gibt mit Konstantenkörper k, so daß f eine Stammfunktion in L hat, gibt es Elemente $v \in K$ und $u_1, \ldots, u_n \in K^\times$ sowie Konstanten $c_1, \ldots, c_n \in k$ derart, daß

$$f = D(v) + \sum_{i=1}^{n} c_i \frac{D(u_i)}{u_i} .$$

Beweis: $g \in L$ sei eine Stammfunktion von f und L/K elementar. Dann gibt es eine Folge $K = L_0 < L_1 < \cdots < L_r = L$ von Zwischenkörpern derart, daß $L_i = L_{i-1}(t_i)$ ist, wobei t_i entweder eine logarithmische oder eine exponentielle Erweiterung von L_{i-1} definiert.

Wir beweisen den Satz durch Induktion über r.

Für $r = 0$ ist $L = K$ und wir können einfach $n = 0$ setzen und $v = g$.

Für $r \geq 1$ betrachten wir den ersten Zwischenkörper $M = K(t_1)$. Da L eine elementare Erweiterung von M ist, die über nur $r - 1$ Körpererweiterungen definiert ist und $f \in K$ erst recht auch in M liegt, gibt es nach Induktionsannahme Elemente $v \in M$ und $u_1, \ldots, u_n \in M^\times$ sowie Konstanten $c_i \in k$ derart, daß

$$f = D(v) + \sum_{i=1}^{n} c_i \frac{D(u_i)}{u_i}$$

ist.

Wir müssen zeigen, daß es auch eine solche Darstellung mit Elementen aus K statt M gibt.

Betrachten wir zunächst den Fall, daß t_1 algebraisch über K ist. Da f in K liegt, ist $\mathrm{Sp}_{M/K}(f) = mf$, wobei $m = (M : K)$ den Grad der Körpererweiterung bezeichnet. Außerdem ist nach den oben bewiesenen

Identitäten über die Ableitungen von Norm und Spur

$$\begin{aligned} \mathrm{Sp}_{M/K}(f) &= \mathrm{Sp}_{M/K}\left(D(v) + \sum_{i=1}^{n} c_i \frac{D(u_i)}{u_i} \right) \\ &= D\big(\mathrm{Sp}_{M/K}(v)\big) + \sum_{i=1}^{n} c_i D\big(\mathrm{N}_{M/K}(u_i)\big) . \end{aligned}$$

Da Normen und Spuren in K liegen hat somit $f = \frac{1}{m}\,\mathrm{Sp}_{M/K}(f)$ eine Darstellung der verlangten Art.

Die Hauptarbeit des Beweises erfordert der Fall, daß t_1 transzendent über K ist. Dann ist $M = K(t_1)$ ein rationaler Funktionenkörper über K, also der Quotientenkörper des Polynomrings $K[t_1]$. Nach Definition einer elementaren Erweiterung ist t_1 entweder logarithmisch oder exponentiell über K, d.h. $D(t_1)$ liegt entweder in K oder ist ein gleich t_1 mal einem Element von K. Insbesondere ist also die Ableitung eines jeden Polynoms in t_1 wieder als Polynom in t_1 darstellbar.

...

§3: Der Algorithmus von Risch

...

Kapitel 5
Gröbner-Basen

§ 1: Gauß und Euklid

Wir haben bereits ziemlich am Anfang der Vorlesung ein Verfahren zur Lösung nichtlinearer Gleichungssysteme kennengelernt, die Elimination von Variablen durch Resultanten. Hier im letzten Kapitel soll es um ein alternatives Verfahren gehen, dessen Bedeutung in der Computeralgebra – genau wie im Falle der Resultanten – weit über die Lösung nichtlinearer Gleichungssysteme hinausgeht.

Ausgangspunkt sind der GAUSS-Algorithmus zur Lösung linearer Gleichungssysteme und der Algorithmus zur Polynomdivision, wie er im EUKLIDischen Algorithmus zur Berechnung des ggT zweier Polynome verwendet wird:

Wenn wir ein lineares Gleichungssystem durch GAUSS-Elimination lösen, bringen wir es zunächst auf eine Treppengestalt, indem wir die erste vorkommende Variable aus allen Gleichungen außer der ersten eliminieren, die zweite aus allen Gleichungen außer den ersten beiden, uns so weiter, bis wir schließlich Gleichungen haben, deren letzte entweder nur eine Variable enthält oder aber eine Relation zwischen Variablen, für die es sonst keine weiteren Bedingungen mehr gibt. Konkret sieht ein Eliminationsschritt folgendermaßen aus: Wenn wir im Falle der beiden Gleichungen

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = r \qquad (1)$$
$$b_1x_1 + b_2x_2 + \cdots + b_nx_n = s \qquad (2)$$

die Variable x_1 mit Hilfe von (1) aus (2) eliminieren wollen, ersetzen wir die zweite Gleichung durch ihre Summe mit $-b_1/a_1$ mal der ersten. Die theoretische Rechtfertigung für diese Umformung besteht darin, daß das Gleichungssystem bestehend aus (1) und (2) sowie das neue Gleichungssystem dieselbe Lösungsmenge haben, und daran ändert sich auch dann nichts, wenn noch weitere Gleichungen dazukommen.

Ähnlich können wir vorgehen, wenn wir ein nichtlineares Gleichungssystem in nur einer Variablen betrachten: Am schwersten sind natürlich die Gleichungen vom höchsten Grad, also versuchen wir, die zu reduzieren auf Polynome niedrigeren Grades. Das kanonische Verfahren dazu ist die Polynomdivision: Haben wir zwei Polynome

$$f = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \quad \text{und}$$
$$g = b_mX^n + b_{m-1}X^{m-1} + \ldots + b_1X + b_0$$

mit $m \le n$, so dividieren wir f durch g, d.h. wir berechnen einen Quotienten q und einen Rest r derart, daß $f = qg + r$ ist und r kleineren Grad als g hat. Konkret: Bei jedem Divisionsschritt haben wir ein Polynom

$$f = a_nX^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0\,,$$

das wir mit Hilfe des Divisors

$$g = b_mX^m + b_{m-1}X^{m-1} + \cdots + b_1X + b_0$$

reduzieren, indem wir es ersetzen durch

$$f - \frac{b_m}{a_n}X^{n-m}g\,,$$

und das führen wir so lange fort, bis f auf ein Polynom von kleinerem Grad als g reduziert ist: Das ist dann der Divisionsrest r. Auch hier ist klar, daß sich nichts an der Lösungsmenge ändert, wenn man die beiden Gleichungen f, g ersetzt durch g, r, denn

$$f = qg + r \quad \text{und} \quad r = f - qg\,,$$

d.h. f und g verschwinden genau dann für einen Wert x, wenn g und r an der Stelle x verschwinden.

In beiden Fällen ist die Vorgehensweise sehr ähnlich: Wir vereinfachen das Gleichungssystem schrittweise, indem wir eine Gleichung ersetzen durch ihre Summe mit einem geeigneter Vielfachen einer anderen Gleichung.

Dieselbe Strategie wollen wir auch anwenden Systeme von Polynomgleichungen in mehreren Veränderlichen. Erstes Problem dabei ist, daß wir nicht wissen, wie wir die Monome eines Polynoms anordnen sollen und damit, was der führende Term ist. Dazu gibt es eine ganze Reihe verschiedener Strategien, von denen je nach Anwendung mal die eine, mal die andere vorteilhaft ist. Wir wollen uns daher zunächst damit beschäftigen.

§ 2: Monomordnungen

Wir betrachten Polynome in n Variablen $X_1, \ldots, X_n$ und setzen zur Abkürzung

$$X^\alpha = X_1^{\alpha_1} \cdots X_n^{\alpha_n} \quad \text{mit} \quad \alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0 \ .$$

Eine Anordnung der Monome ist offensichtlich äquivalent zu einer Anordnung auf $\mathbb{N}_0^n$, und es gibt sehr viele Möglichkeiten, diese Menge anzuordnen. Für uns sind allerdings nur Anordnungen interessant, die einigermaßen kompatibel sind mit der algebraischen Struktur des Polynomrings $k[X_1, \ldots, X_n]$; beispielsweise wollen wir sicherstellen, daß der führende Term des Produkts zweier Polynome das Produkt der führenden Terme der Faktoren ist – wie wir es auch vom Eindimensionalen her gewohnt sind. Daher definieren wir

Definition: *a)* Eine Monomordnung ist eine Ordnungsrelation $<$ auf $\mathbb{N}_0^n$, für die gilt

1. $<$ ist eine Linear- oder Totalordnung, d.h. für zwei Elemente $\alpha, \beta \in \mathbb{N}_0^n$ ist entweder $\alpha < \beta$ oder $\beta < \alpha$ oder $\alpha = \beta$.
2. Für $\alpha, \beta, \gamma \in \mathbb{N}_0^n$ gilt $\alpha < \beta \Longrightarrow \alpha + \gamma < \beta + \gamma$.
3. $<$ ist eine Wohlordnung, d.h. jede Teilmenge $I \subseteq \mathbb{N}_0^n$ hat ein kleinstes Element.

b) Für $f = \sum_{\alpha \in I} c_\alpha X^\alpha \in k[X_1, \ldots, X_n]$ mit $c_\alpha \neq 0$ für alle $\alpha \in I \subset \mathbb{N}_0$ sei γ das größte Element von I bezüglich einer fest gewählten Monomordnung. Dann bezeichnen wir bezüglich dieser Monomordnung

- γ = multideg f als Multigrad von f
- X^γ = FM f als führendes Monom von f
- c_γ = FK f als führenden Koeffizienten von f
- $c_\gamma X^\gamma$ = FT f als führenden Term von f

Der Grad deg f von f ist, wie in der Algebra üblich, der höchste Grad eines Monoms von f; je nach gewählter Monomordnung muß das nicht unbedingt der Grad des führenden Monoms sein. Aus der zweiten Forderung an eine Monomordnung folgt aber, daß für ein Produkt stets gilt FM(fg) = FM(f) · FM(g)

Beispiele von Monomordnungen sind

1) Die lexikographische Ordnung: Hier ist $\alpha < \beta$ genau dann, wenn für den ersten Index i, in dem sich α und β unterscheiden, $\alpha_i < \beta_i$ ist. Betrachtet man Monome X^α als Worte über dem (geordneten) Alphabet $\{X_1, \ldots, X_n\}$, kommt hier ein Monom X^α genau dann vor X^β, wenn die entsprechenden Worte im Lexikon in dieser Reihenfolge gelistet werden. Die ersten beiden Forderungen an eine Monomordnung sind klar, und auch die Wohlordnung macht keine großen Probleme: Man betrachtet zunächst die Teilmenge aller Exponenten $\alpha \in I$ mit kleinstmöglichem α_1, unter diesen die Teilmenge mit kleinstmöglichem α_2, *usw.*, bis man bei α_n angelangt ist. Spätestens hier ist die verbleibende Teilmenge einelementig, und ihr einziges Element ist das gesuchte kleinste Element von I.

2) Die graduierte lexikographische Ordnung: Hier ist der Grad eines Monoms erstes Ordnungskriterium: Ist deg $X^\alpha <$ deg X^β, so definieren wir $\alpha < \beta$. Falls beide Monome gleichen Grad haben, soll $\alpha < \beta$ genau dann gelten, wenn α im lexikographischen Sinne kleiner als β ist. Auch hier sind offensichtlich alle drei Forderungen erfüllt.

3) Die inverse lexikographische Ordnung: Hier ist $\alpha < \beta$ genau dann, wenn für den *letzten* Index i, in dem sich α und β unterschei-

den. Das entspricht offensichtlich gerade der lexikographischen Anordnung bezüglich des rückwärts gelesenen Alphabets $X_n, \ldots, X_1$. Entsprechend läßt sich natürlich auch bezüglich jeder anderen Permutation des Alphabets eine Monomordnung definieren.

§3: Der Hilbertsche Basissatz

Definition: Ein Ideal $I \lhd R = k[X_1, \ldots, X_n]$ heißt *monomial,* wenn es von (nicht notwendigerweise endlich vielen) Monomen erzeugt wird.

Nehmen wir an, I werde erzeugt von den Monomen X^α mit α aus einer Indexmenge A. Ist dann X^β irgendein Monom aus I, kann es als endliche Linearkombination

$$X^\beta = \sum_{i=1}^{r} f_i X^{\alpha_i} \quad \text{mit} \quad \alpha_i \in A$$

geschrieben werden, wobei die f_i irgendwelche Polynome aus R sind. Da sich jedes Polynom als Summe von Monomen schreiben läßt, können wir f_i als k-Linearkombination von Monomen X^γ schreiben und bekommen damit eine neue Darstellung von X^β als Summe von Termen der Form $cX^\gamma X^\alpha$ mit $\alpha \in A, \beta \in \mathbb{N}_0^n$ und $c \in k$. Sortieren wir diese Summanden nach den resultierenden Monomen $X^{\gamma+\alpha}$, entsteht eine k-Linearkombination verschiedener Monome, die insgesamt gleich X^β ist. Das ist aber nur möglich, wenn diese Summe aus dem einen Summanden X^β besteht, d.h. β läßt sich schreiben in der Form $\beta = \alpha + \gamma$ mit einem $\alpha \in A$ und einem $\gamma \in \mathbb{N}_0^n$.

Dies zeigt, daß ein Monom X^β genau dann in I liegt, wenn $\beta = \alpha + \gamma$ ist mit einem $\alpha \in A$ und einem $\gamma \in \mathbb{N}_0^n$; das Ideal I selbst besteht also genau aus den Polynomen f, die sich als k-Linearkombinationen solcher Monome schreiben lassen.

Damit folgt insbesondere, daß ein Polynom f genau dann in einem monomialen Ideal I liegt, wenn jedes seiner Monome dort liegt.

Lemma von Dickson: Jedes monomiale Ideal in $R = k[X_1, \ldots, X_n]$ kann von endlich vielen Monomen erzeugt werden.

Der *Beweis* wird durch vollständige Induktion nach n geführt. Im Fall $n = 1$ ist alles klar, denn da sind die Monome gerade die Potenzen der einzigen Variable, und natürlich erzeugt jede Menge von Potenzen genau dasselbe Ideal wie die Potenz mit dem kleinsten Exponenten aus dieser Menge. Hier kommt man also sogar mit einem einzigen Monom aus.

Für $n > 1$ betrachten wir jenes Ideal $J \lhd k[X_1, \ldots, X_{n-1}]$, das erzeugt wird von allen jenen Monomen $X'^{\alpha'} = X_1^{\alpha_1} \cdots X_{n-1}^{\alpha_{n-1}}$, für die es ein $\alpha_n \in \mathbb{N}_0$ gibt derart, daß $X^\alpha = X_1^{\alpha_1} \cdots X_{n-1}^{\alpha_{n-1}} X_n^{\alpha_n}$ in I liegt. Dieses Ideal wird nach Induktionsvoraussetzung erzeugt von endlich vielen Monomen $X'^{\alpha'}$, wobei die Striche bedeuten sollen, daß wir jeweils nur Indizes bis $n-1$ betrachten.

Zu jedem der $X'^{\alpha'}$ aus dem endlichen Erzeugendensystem von J gibt es nach Definition von J ein $\alpha_n \in \mathbb{N}_0$ derart, daß X^α für das damit komplettierte α in I liegt. Sei r die größte der Zahlen α_n; dann liegt $X'^{\alpha'} X_n^r$ für jedes Monom aus dem Erzeugendensystem von J in I und damit für jedes Monom aus J. Die endlich vielen Monome $X^{\alpha'} X_n^r$ erzeugen also zumindest ein Teilideal von I.

Es gibt aber natürlich auch noch Monome in I, in denen X_n mit einem kleineren Exponenten als r auftritt. Um auch diese Elemente zu erfassen, betrachten wir für jedes $s < r$ das Ideal $J_s \lhd k[X_1, \ldots, X_{n-1}]$, das von allen jeden Monomen $X'^{\alpha'}$ erzeugt wird, für die $X'^{\alpha'} X_n^s$ in I liegt. Auch jedes der J_s wird nach Induktionsannahme erzeugt von endlich vielen Monomen $X'^{\alpha'}$, und wenn wir die sämtlichen Monome $X'^{\alpha'} X_n^s$ zu unserem Erzeugendensystem hinzunehmen (für alle $s = 0, 1, \cdots, r-1$), haben wir offensichtlich ein endliches Erzeugendensystem aus Monomen für I gefunden. ∎

Beliebige Ideale sind im allgemeinen nicht monomial; schon das von $X + 1$ erzeugte Ideal in $k[X]$ ist ein Gegenbeispiel, denn es enthält weder das Monom X noch das Monom 1, im Widerspruch zu der oben gezeigten Eigenschaft eines monomialen Ideals, zu jedem seiner Elemente auch dessen Monome sämtliche zu enthalten. Um monomiale

Ideale auch für die Untersuchung solcher Ideale nützlich zu machen, wählen wir eine Monomordnung auf R und definieren für ein beliebiges Ideal $I \lhd R \underset{\text{def}}{=} k[X_1, \ldots, X_n]$ das monomiale Ideal

$$\text{FM}(I) = \left\{ \text{FM}(f) \;\middle|\; f \in I \setminus \{0\} \right\},$$

das von den führenden Monomen *aller* Elemente von I erzeugt wird – außer natürlich dem nicht existierenden führenden Term der Null.

Nach dem Lemma von DICKSON ist FM(I) erzeugt von endlich vielen Monomen. Jedes dieser Monome ist, wie wir eingangs gesehen haben, ein Vielfaches eines der erzeugenden Monome, also eines führenden Monoms eines Elements von I. Ein Vielfaches des führenden Monoms ist aber das führende Monom des entsprechenden Vielfachen des Elements von I, denn FM($X^\gamma f$) = X^γ FM(f), da für jede Monomordnung gilt $\alpha < \beta \Longrightarrow \alpha + \beta < \alpha + \gamma$. Somit wird FM($I$) erzeugt von endlich vielen Monomen der Form FM(f_i), wobei die f_i Elemente von I sind. Wir wollen sehen, daß die Elemente f_i das Ideal I erzeugen; damit folgt insbesondere

Hilbertscher Basissatz: Jedes Ideal $I \lhd R = k[X_1, \ldots, X_n]$ hat ein endliches Erzeugendensystem.

Beweis: Wie wir bereits wissen, gibt es Elemente $f_1, \ldots f_m \in I$, so daß FM(I) von den Monomen FM(f_i) erzeugt wird. Um zu zeigen, daß die Elemente f_i das Ideal I erzeugen, betrachten wir ein beliebiges Element $f \in I$ und versuchen, es als R-Linearkombination der f_i zu schreiben. Division von f durch $f_1, \ldots, f_r$ zeigt, daß es Polynome $a_1, \ldots, a_m$ und r in R gibt derart, daß

$$f = a_1 f_1 + \cdots + a_m f_m + r\,.$$

Wir sind fertig, wenn wir zeigen können, daß der Divisionsrest r verschwindet.

Falls r *nicht* verschwindet, zeigt der Divisionsalgorithmus, daß das führende Monom FM(r) von r durch kein führendes Monom FM(f_i) eines der Divisoren f_i teilbar ist. Andererseits ist aber

$$r = f - (a_1 f_1 + \cdots + a_m f_m)$$

ein Element von I, und damit liegt FM(r) im von den FM(f_i) erzeugten Ideal FM(I). Somit muß FM(r) Vielfaches eines FM(f_i) sein, ein Widerspruch. Also ist $r = 0$.

■

§ 4: Gröbner-Basen und der Buchberger-Algorithmus

Definition: Eine endliche Teilmenge $G = \{g_1, \ldots, g_m\} \subset I$ eines Ideals $I \lhd R = k[X_1, \ldots, X_n]$ heißt Standardbasis oder GRÖBNER-Basis von I falls die Monome FM(g_i) das Ideal FM(I) erzeugen.

Wie der obige Beweis des HILBERTschen Basissatzes zeigt, hat jedes Ideal außer dem Nullideal eine GRÖBNER-Basis, und diese erzeugt das Ideal. Bevor wir uns damit beschäftigen, wie man diese berechnen kann, wollen wir zunächst eine wichtige Eigenschaften betrachten.

Sei $g_1, \ldots, g_m$ GRÖBNER-Basis eines Ideals $I \lhd R$. Wir wollen ein beliebiges Element $f \in R$ durch $g_1, \ldots, g_m$ dividieren. Dies liefert als Ergebnis

$$f = a_1 g_1 + \cdots + a_m g_m + r\,,$$

wobei kein Monom von r durch eines der Monome FM(g_i) teilbar ist. Wie wir wissen, sind allerdings bei der Polynomdivision weder der Divisionsrest r noch die Koeffizienten a_i auch nur im entferntesten eindeutig. Sei etwa

$$f = a_1 g_1 + \cdots + a_m g_m + r = b_1 g_1 + \cdots + b_m g_m + s\,.$$

Dann ist

$$(a_1 - b_1) g_1 + \cdots + (a_m - b_m) g_m = s - r\,.$$

Links steht ein Element von I, also auch rechts. Andererseits enthält aber weder r noch s ein Monom, das durch eines der Monome FM(g_i) teilbar ist, d.h. $r - s = 0$. Somit ist bei der Division durch die Elemente einer GRÖBNER-Basis der Divisionsrest eindeutig bestimmt. Insbesondere ist f genau dann ein Element von I, wenn der Divisionsrest verschwindet. Wenn wir eine GRÖBNER-Basis haben, können wir als leicht entscheiden, ob ein gegebenes Element $f \in R$ im Ideal I liegt.

Nachdem im Fall einer GRÖBNER-Basis der Divisionsrest nicht von der Reihenfolge der Basiselemente abhängt, können wir ihn durch ein Symbol bezeichnen, das nur von der *Menge* $G = \{g_1, \ldots, g_m\}$ abhängt; wir schreiben $\overline{f}^G$.

Als nächstes wollen wir uns überlegen, wie sich eine GRÖBNER-Basis eines vorgegebenen Ideals I finden läßt.

Dazu müssen wir uns als erstes überlegen, *wie* das Ideal vorgegeben sein soll. Wenn wir damit rechnen wollen, müssen wir irgendeine Art von endlicher Information haben; was sich anbietet ist natürlich ein endliches Erzeugendensystem.

Wir gehen also aus von einem Ideal $I = (f_1, \ldots, f_m)$ und suchen eine GRÖBNER-Basis. Das Problem ist, daß die Monome $\mathrm{FM}(f_i)$ im allgemeinen nicht ausreichen, um das monomiale Ideal $\mathrm{FM}(I)$ zu erzeugen, denn dieses enthält ja *jedes* Monom eines jeden Elements von I und nicht nur das führende. Wir müssen daher neue Elemente produzieren, deren führende Monome in den gegebenen Elementen f_i oder auch anderen Elementen von I erst weiter hinten vorkommen.

BUCHBERGERS Idee dazu war die Konstruktion sogenannter S-Polynome: Seien $f, g \in R$ zwei Polynome; $\mathrm{FM}(f) = X^\alpha$s und $\mathrm{FM}(g) = X^\beta$ seien ihre führenden Monome, und X^γ sei das kgV von X^α und X^β, d.h. $\gamma_i = \max(\alpha_i, \beta_i)$ für alle $i = 1, \ldots, n$. Das S-Polynom von f und g ist

$$S(f,g) = \frac{X^\gamma}{\mathrm{FT}(f)} \cdot f - \frac{X^\gamma}{\mathrm{FT}(g)} \cdot g\,.$$

Da $\frac{X^\gamma}{\mathrm{FT}(f)} \cdot f$ und $\frac{X^\gamma}{\mathrm{FT}(g)} \cdot g$ beide nicht nur dasselbe führende Monom X^γ haben, sondern es wegen der Division durch den führenden *Term* statt nur das führende Monom auch beide mit Koeffizient eins enthalten, fällt es bei der Bildung von $S(f,g)$ weg, d.h. $S(f,g)$ hat ein kleineres führendes Monom. Das folgende Lemma ist der Kern des Beweises, daß S-Polynome alles sind, was wir brauchen, um GRÖBNER-Basen zu berechnen.

Lemma: Für die Polynome $f_1, \ldots, f_m \in R$ sei für ein $\delta \in \mathbb{N}_0^n$

$$S = \sum_{i=1}^m \lambda_i X^{\alpha_i} f_i \quad \text{mit} \quad \lambda_i \in k \quad \text{und} \quad \alpha_i \in \mathbb{N}_0^n$$

derart, daß $\alpha_i + \text{multideg}\, f_i = \delta$ für $i = 1, \ldots, n$. Falls $\text{multideg}\, S < \delta$ ist, gibt es Elemente $\lambda_{ij} \in k$, so daß

$$S = \sum_{i=1}^m \sum_{j=1}^m \lambda_{ij} X^{\gamma_{ij}} S(f_i, f_j)$$

ist mit $X^{\gamma_{ij}} = \mathrm{kgV}\big(\mathrm{FM}(f_i), \mathrm{FM}(f_j)\big)$.

Beweis: Der führende Koeffizient von f_i sei μ_i; dann ist $\lambda_i\mu_i$ der führende Koeffizient von $\lambda_i X^{\alpha_i} f_i$. Somit ist $\text{multideg}\, S$ genau dann kleiner als δ wenn $\sum_{i=1}^m \lambda_i\mu_i$ verschwindet. Wir normieren alle $X^{\alpha_i} f_i$ auf führenden Koeffizienten eins, indem wir $p_i = X^{\alpha_i} f_i/\mu_i$ betrachten; dann ist

$$\begin{aligned} S = \sum_{i=1}^m \lambda_i\mu_i p_i &= \lambda_1\mu_1(p1 - p_2) + (\lambda_1\mu_1 + \lambda_2\mu_2)(p_2 - p_3) + \ldots \\ &\quad + (\lambda_1\mu_1 + \cdots + \lambda_{m-1}\mu_{m-1})(p_{m-1} - p_m) \\ &\quad + (\lambda_1\mu_1 + \cdots + \lambda_m\mu_m)p_m\,. \end{aligned}$$

Da alle p_i denselben Multigrad δ und denselben führenden Köeffizienten eins haben, kürzen sich in den Differenzen $p_i - p_j$ die führenden Terme weg, genau wie in den S-Polynomen. In der Tat: Bezeichnen wir den Multigrad von $\mathrm{kgV}\big(\mathrm{FM}(f_i), \mathrm{FM}(f_j)\big)$ mit γ_{ij}, so ist

$$p_i - p_j = X^{\delta - \gamma_{ij}} S(f_i, f_j)\,.$$

Damit hat die obige Summendarstellung von S die gewünschte Form. ∎

. . .