Kapitel 0
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 Intel-

ligenz. 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 Konkurrenz 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 Konkurrenz 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, exaktes und symbolisches Rechnen

Mit vielen Fragestellungen der Computeralgebra wie etwa der Lösung von Polynomgleichungen oder Systemen solcher Gleichungen beschäftigt sich auch die numerische Mathematik; um die unterschiedlichen Ansätze beider Gebiete zu verstehen, müssen wir uns die Unterschiede zwischen numerischem Rechnen, exaktem Rechnen und symbolischem Rechnen klar machen.

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 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** einer Programmiersprache 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 Exponenten 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)

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

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

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

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*-Bedingung ignoriert und mit dem Ergebnis $+\infty$ weiter rechnet, 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: Beispielsweise 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 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.

Um zu sehen, wie sich Probleme mit Rundungsfehlern bei algebraischen Fragestellungen auswirken können, wollen wir zum Abschluß dieses Paragraphen ein Beispiel aus WILKINSONs Buch betrachten. Er geht aus vom Polynom zwanzigsten Grades

$$f(x) = (x-1)(x-2)(x-3)\cdots(x-18)(x-19)(x-20)$$

mit den Nullstellen $1, 2, \ldots, 20$. In ausmultiplizierter Form würde es mehrere Zeilen benötigen: Der größte Koeffizient, der von x^2, hat zwanzig Dezimalstellen, und die meisten anderen haben nicht viel weniger.

Der Koeffizient von x^{19} ist allerdings noch überschaubar: Wie man sich leicht überlegt, ist er gleich der negativen Summe der Zahlen von eins bis zwanzig, also -210.

WILKINSON stört nun diesen Koeffizienten um einen kleinen Betrag und berechnet die Nullstellen des so modifizierten Polynoms. Betrachten wir etwa die Nullstellen von $g(x) = f(x) - 10^{-9}x^{19}$; wir ersetzen in f also den Koeffizienten -210 durch $-210{,}000000001$. Die neuen Nullstellen sind, auf fünf Nachkommastellen gerundet,

$$1{,}0000, \quad 2{,}0000, \quad 3{,}0000, \quad 4{,}0000, \quad 5{,}0000,$$

$$6{,}0000,\quad 7{,}0000,\quad 8{,}0001,\quad 8{,}9992,\quad 10{,}008,$$
$$10{,}957,\quad 12{,}383 \pm 0{,}10867i,\quad 14{,}374 \pm 0{,}77316i,$$
$$16{,}572 \pm 0{,}88332i,\quad 18{,}670 \pm 0{,}35064i,\quad 20{,}039\,.$$

Durch kleinste Veränderungen an einem einzigen Koeffizienten, wie sie beispielsweise jederzeit durch Rundungen entstehen können, kann sich also selbst das qualitative Bild ändern: Hier etwa reduziert sich die Anzahl der (für viele Anwendungen einzig relevanten) reellen Nullstellen von zwanzig auf zwölf. Schon wenn wir verläßliche Aussagen über die Anzahl reeller Nullstellen brauchen, können wir uns also nicht allein auf numerische Berechnungen verlassen, sondern brauchen alternative Methoden wie zum Beispiel explizite Lösungsformeln, mit denen wir auch theoretisch arbeiten können.

§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.

Dieses negative Ergebnis kam hat seinen Ausgangspunkt in einem po-

sitiv formulierten Problem von DAVID HILBERT. Dieser stellte auf dem Internationalen Mathematikerkongress 1900 in Paris 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:

Man gebe ein Verfahren an, das für eine beliebige diophantische Gleichung entscheidet, ob sie lösbar ist.

Wie sich zeigte, war HILBERT hier zu optimistisch: 1970 bewies YURI V. MATIYASEVICH, daß es kein solches Verfahren geben kann, da sich jedes sogenannte rekursiv aufzählbare Problem auf die Frage nach der Lösbarkeit eines diophantischen Gleichung zurückführen läßt. Da zu den rekursiv aufzählbaren Problem auch unlößbare wie das Halteproblem für TURING-Maschinen gehören, folgte daraus die Unmöglichkeit des von HILBERT geforderten Verfahrens.

Da reelle Zahlen $x_1, \ldots, x_n$ genau dann ganz sind, wenn $\sum_{i=1}^{n} \sin^2 \pi x_i$ verschwindet, übersetzte DANIEL RICHARDSON dies in den folgenden Unmöglichkeitssatz für reeller Zahlen:

Satz von Richardson: Es gibt kein Verfahren, das in endlich vielen Schritten entscheidet, ob ein beliebig vorgegebener Ausdruck bestehend aus rationalen Zahlen, π, einer Variablen x sowie den Funktionen +, ·, Sinus und Betrag gleich Null ist.

Tatsächlich bewies RICHARDSON ein etwas schwächeres Resultat, denn seine Arbeit erschien bereits 1969, also ein Jahr vor der von MATIYASEVICH, so daß er nur ein schwächeres Resultat verwenden konnte. Zusammen mit dem Resultat von MATIYASEVICH zeigt seine Methode aber sofort den angegebenen Satz.

Mehr zum zehnten HILBERTschen Problem und seinen Konsequenzen findet man bei

YURI V. MATIYASEVICH: Hilbert's Tenth Problem, *MIT Press,* 1993

Kapitel 1
Algebraische Vorbereitungen

Auch wenn es in der Computeralgebra vor allem um algorithmische Verfahren geht, können wir doch nicht auf alle strukturellen Aussagen der klassischen Algebra verzichten; die für uns wichtigsten sind in diesem Kapitel zusammengestellt.

§1: Teilbarkeit in Integritätsbereichen

Ringe sollte dem Leser aus den Anfängervorlesungen bekannt sein, seien aber der Vollständigkeit halber kurz definiert:

Definition: *a)* Ein Ring ist eine Menge R zusammen mit zwei Rechenoperationen „+“ und „$\cdot$“ von $R \times R$ nach R, so daß gilt:

1.) R bildet bezüglich „+“ eine abelsche Gruppe, d.h. für die Addition gilt das Kommutativgesetz $f + g = g + f$ sowie das Assoziativgesetz $(f + g) + h = f + (g + h)$ für alle $f, g, h \in \mathbb{R}$, es gibt ein Element $0 \in R$, so daß $0 + f = f + 0 = f$ für alle $f \in R$, und zu jedem $f \in R$ gibt es ein Element $-f \in R$, so daß $f + (-f) = 0$ ist.

2.) Die Verknüpfung „$\cdot$“: $R \times R \to R$ erfüllt das Assoziativgesetz $f(gh) = (fg)h$, und es gibt ein Element $1 \in R$, so daß $1f = f1 = f$.

3.) „+“ und „$\cdot$“ erfüllen die Distributivgesetze $f(g + h) = fg + fh$ und $(f + g)h = fh + gh$.

b) Ein Ring heißt *kommutativ,* falls zusätzlich noch das Kommutativgesetz $fg = gf$ der Multiplikation gilt.

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

Natürlich ist jeder Körper ein Ring; für einen Körper werden schließlich genau dieselben Eigenschaften gefordert und zusätzlich auch noch die Kommutativität der Multiplikation sowie die Existenz multiplikativer Inverser. Ein Körper ist somit insbesondere auch ein Integritätsbereich.

Das bekannteste Beispiel eines Rings, der kein Körper ist, sind die ganzen Zahlen; auch sie bilden einen Integritätsbereich.

Auch die Menge

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

aller Polynome mit Koeffizienten aus einem Körper k ist ein Integritätsbereich; ersetzt man den Körper k durch einen beliebigen kommutativen Ring R, ist $R[X]$ immerhin noch ein Ring; wir bezeichnen ihn als den *Polynomring* in X über R. Da der führende Koeffizient eines Produkts zweier Polynome das Produkt der führenden Koeffizienten der Faktoren ist, folgt leicht, daß $R[X]$ genau dann ein Integritätsbereich ist, wenn auch R einer ist.

Indem wir Polynomringe über Polynomringen betrachten, erhalten wir Polynomringe in zwei und mehr Variablen über einem Ring R; diese werden bezeichnet mit $R[X_1, \ldots, X_n]$ und sind offensichtlich auch genau dann Integritätsbereiche, wenn R einer ist.

Auch $\mathbb{Z}/m = \{0, 1, \ldots, m-1\}$ mit der Addition und Multiplikation modulo m ist ein Ring; falls m eine zusammengesetzte Zahl ist, hat er offensichtlich Nullteiler: Beispielsweise ist in $\mathbb{Z}/6$ das Produkt $2 \cdot 3 = 6 \bmod 6 = 0$.

Als Beispiel eines nichtkommutativen Rings können wir die Menge aller $n \times n$-Matrizen über einem Körper betrachten; dieser Ring hat auch Nullteiler, denn beispielsweise ist

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

obwohl keiner der beiden Faktoren die Nullmatrix ist.

Quotienten müssen in beliebigen Ringen nicht existieren, und wenn, müssen sie im Falle von Nullteiler auch nicht eindeutig sein. Da sie dann auch nicht sonderlich nützlich sind, beschränken wir für Teilbarkeitsfragen auf Integritätsbereiche.

Definition: R sei ein Integritätsbereich.
a) Ein Element $h \in R$ heißt Teiler von $f \in R$, in Zeichen $h|f$, wenn es ein $q \in R$ gibt, so daß $f = qh$ ist.
b) $h \in R$ heißt *größter gemeinsamer Teiler* (kurz ggT) der beiden Elemente f und g aus R, wenn h Teiler von f und von g ist und wenn für jeden anderen gemeinsamen Teiler r von f und g gilt: $r|h$.
c) Zwei Elemente $f, g \in R$ heißen *assoziiert*, wenn f Teiler von g und g Teiler von f ist.
d) Ein Element $u \in R$ heißt *Einheit*, falls es ein $v \in R$ gibt mit $uv = 1$. Die Menge aller Einheiten von R bezeichnen wir mit $R^\times$.

In einem Körper ist natürlich jedes von null verschiedene Element Teiler eines jeden anderen Elements und damit auch eine Einheit; in $\mathbb{Z}$ dagegen sind ± 1 die beiden einzigen Einheiten, und zwei ganze Zahlen sind genau dann assoziiert, wenn sie sich höchstens im Vorzeichen unterscheiden.

Zu zwei Elementen f, g eines Intergritätsbereichs muß es keinen größten gemeinsamen Teiler geben, und wenn einer existiert muß er nicht eindeutig sein: Da wir seine „Größe“ über Teilbarkeit definieren; von schon in $\mathbb{Z}$ außer 2 auch -2 ein größter gemeinsamer Teiler von 8 und 10.

In einem Polynomring über einem Integritätsbereich ist der Grad des Produkts zweier Polynome gleich der Summe der Grade der Faktoren; da das konstante Polynom eins Grad null hat, muß daher jede Einheit Grad null haben; die Einheiten von $\mathbb{R}[x]$ sind also genau die Einheiten von R. Speziell für Polynomringe über Körpern sind dies genau die von null verschiedenen Konstanten.

Damit wissen wir auch, wann zwei Polynome assoziiert sind:

Lemma: Zwei von null verschiedene Elemente f, g eines Integritätsbereichs sind genau dann assoziiert, wenn es eine Einheit u gibt, so daß $f = ug$ ist.

Beweis: Eine Einheit $u \in R$ hat nach Definition ein Inverses $u^{-1} \in R$, und aus $f = ug$ folgt $g = u^{-1}f$. Somit ist f Teiler von g und g Teiler von f; die beiden Elemente sind also assoziiert.

Sind umgekehrt $f, g \in R \setminus \{0\}$ assoziiert, so gibt es Elemente $u, v \in R$ derart, daß $g = uf$ und $f = vg$ ist. Damit ist $g = uf = uvg$ und $f = vg = vuf$, also $(1 - uv)g = 0$ und $(1 - vu)f = 0$. Da wir in einem Integritätsbereich sind und f, g nicht verschwinden, muß somit $uv = vu = 1$ sein, d.h. u und v sind Einheiten.

■

Damit sind also zwei Polynome über einem Körper genau dann assoziiert, wenn sie sich nur um eine von null verschiedene multiplikative Konstante unterscheiden. Nur bis auf eine solche Konstante können wir auch den größten gemeinsamen Teiler zweier Polynome bestimmen, denn allgemein gilt:

Lemma: Der größte gemeinsame Teiler zweier Polynome ist bis bis auf Assoziiertheit eindeutig. Sind also h und $\tilde{h}$ zwei größte gemeinsame Teiler der beiden Elemente f und g, so sind h und $\tilde{h}$ assoziiert; ist umgekehrt h ein größter gemeinsamer Teiler von f und g und ist $\tilde{h}$ assoziiert zu h, so ist auch $\tilde{h}$ ein größter gemeinsamer Teiler von f und g.

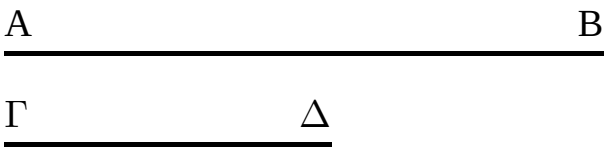
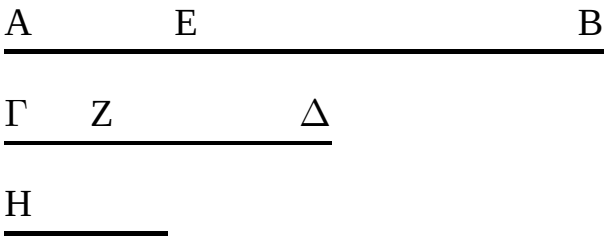
Beweis: Sind h und $\tilde{h}$ größte gemeinsame Teiler, so sind sie insbesondere gemeinsame Teiler und damit Teiler eines jeden größte gemeinsamen Teilers. Somit müssen h und $\tilde{h}$ einander teilen, sind also assoziiert. Ist h ein größter gemeinsamer Teiler und $\tilde{h}$ assoziiert zu h, so teilt $\tilde{h}$ jedes Vielfache von h, ist also auch ein gemeinsamer Teiler, und da h jeden gemeinsamen Teiler teilt, gilt dasselbe auch für $\tilde{h}$. Somit ist auch $\tilde{h}$ ein größter gemeinsamer Teiler.

■

§2: Euklidische Ringe

Das wichtigste Verfahren zur Berechnung des größten gemeinsamen Teilers zweier natürlicher Zahlen ist der EUKLIDische Algorithmus. Bei EUKLID, in Proposition 2 des siebten Buchs seiner *Elemente,* wird er so

beschrieben (nach der Übersetzung von CLEMENS THAER in Oswalds Klassiker der exakten Wissenschaften, Band 235):

> *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 $AB, \Gamma\Delta$. Man soll das größte gemeinsame Maß von $AB, \Gamma\Delta$ finden.
>
> 
>
>
> Wenn $\Gamma\Delta$ hier AB mißt – sich selbst mißt es auch – dann ist $\Gamma\Delta$ gemeinsames Maß von $\Gamma\Delta, AB$. Und es ist klar, daß es auch das größte ist, denn keine Zahl größer $\Gamma\Delta$ kann $\Gamma\Delta$ messen.
>
> Wenn $\Gamma\Delta$ aber AB nicht mißt, und man nimmt bei $AB, \Gamma\Delta$ 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 $AB, \Gamma\Delta$ gegeneinander prim sein, gegen die Voraussetzung. Also muß eine Zahl übrig bleiben, die die vorangehende mißt. $\Gamma\Delta$ lasse, indem es BE mißt, EA, kleiner als sich selbst übrig; und EA lasse, indem es ΔZ mißt, $Z\Gamma$, 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 $\Gamma\Delta$ messen. $\Gamma\Delta$ 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.

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.

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 der 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 schließlich nur einen SOKRATES. Erst bei seinen Nachfolgern, den Peripatetikern,

setzte sich langsam auch die Eins als Zahl durch; ihr Zeitgenosse EUKLID macht noch brav eine Fallunterscheidung: In Proposition 1, unmittelbar vor der hier abgedruckten Proposition 2, führt er praktisch dieselbe Konstruktion durch für teilerfremde Zahlen.

Als zweites fällt auf, daß EUKLID seine Konstruktion rein geometrisch durchführt; wenn er von einer Strecke eine andere Strecke abträgt solange es geht, ist das arithmetisch ausgedrückt gerade die Konstruktion des Rests bei der Division der beiden Streckenlängen durcheinander.

Die wesentliche Operation beim EUKLIDischen Algorithmus ist somit die Division mit Rest; wir wollen daher Ringe betrachten, in der diese möglich ist. Rein formal ist sie natürlich immer möglich; wir könnten sagen, daß $f : g = 0$ Rest f ist, aber so eine Division nützt offensichtlich nichts: Wir brauchen einen Rest, der in irgendeinem Sinne kleiner ist als der Divisor. Dies führt zur folgenden

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 $f = gh$, so ist $\nu(f) \geq \max\big(\nu(g), \nu(h)\big)$, und zu je zwei Elementen $f, g \in R$ gibt es Elemente $q, r \in R$ mit

$$f = qq + r \quad \text{und} \quad r = 0 \text{ oder } \nu(r) < \nu(g)\,.$$

Wir schreiben auch $f : g = q$ Rest r und bezeichnen r als Divisionsrest bei der Division von f durch g.

Standardbeispiel sind die ganzen Zahlen, wo wir als ν einfach die Betragsfunktion nehmen können. Quotient und Divisionsrest sind durch die Forderung $\nu(r) < \nu(y)$ allerdings nicht eindeutig festgelegt, beispielsweise ist im Sinne dieser Definition

$$11 : 3 = 3 \text{ Rest } 2 \quad \text{und} \quad 11 : 3 = 4 \text{ Rest } -1\,.$$

Die Definition des EUKLIDischen Rings verlangt nur, daß es *mindestens* eine Darstellung gibt; Eindeutigkeit ist nicht gefordert.

Das für uns in der Computeralgebra wichtigste Beispiel ist der Polynomring $k[X]$ über einem Körper k; hier zeigt die bekannte Polynomdivision

mit Rest, daß die Bedingungen erfüllt sind bezüglich der Gradabbildung

$$\nu\colon \begin{cases} k[X] \setminus \{0\} \to \mathbb{N}_0 \\ f \mapsto \deg f \end{cases} .$$

Hier ist es allerdings wichtig, daß k ein Körper ist: Bei der Polynomdivision mit Rest müssen wir schließlich die führenden Koeffizienten durcheinander dividieren, und das wäre etwa im Polynomring $\mathbb{Z}[X]$ nicht möglich.

Dies beweist freilich nicht, daß $\mathbb{Z}[X]$ *kein* EUKLIDischer Ring wäre, denn in der Definition war ja nur gefordert, daß es für *irgendeine* Funktion ν *irgendein* Divisionsverfahren gibt; dessen Nichtexistenz ist sehr schwer zu zeigen – es sei denn, eine der im folgenden hergeleiteten Eigenschaften eines EUKLIDischen Rings ist nicht erfüllt. Bei $\mathbb{Z}[X]$ ist dies, wie wir bald sehen werden, bei der linearen Kombinierbarkeit des ggT in der Tat der Fall, so daß $\mathbb{Z}[X]$ kein EUKLIDischer Ring sein kann.

Doch zunächst müssen wir uns überlegen, daß in einem EUKLIDischen Ring in der Tat größte gemeinsame Teiler existieren und mit dem EUKLIDischen Algorithmus auch berechnet werden können.

In heutiger Sprache ausgedrückt beruht der EUKLIDische Algorithmus auf folgenden beiden Tatsachen:

1. Wenn wir zwei Elemente f, g eines EUKLIDischen Rings mit Rest durcheinander dividieren, so ist $f : g = q$ Rest r äquivalent zu jeder der beiden Gleichungen

   $$f = qg + r \quad \text{und} \quad r = f - qg .$$

   Diese zeigen, daß jeder gemeinsame Teiler von f und g auch ein gemeinsamer Teiler von g und r ist und umgekehrt. Die beiden Paare (f, g) und (g, r) haben also dieselben gemeinsamen Teiler und damit auch denselben größten gemeinsamen Teiler:

   $$\text{ggT}(f, g) = \text{ggT}(g, r) .$$

2. $\text{ggT}(f, 0) = f$, denn jedes Element eines Integritätsbereichs teilt die Null.

Aus diesen beiden Beobachtungen folgt nun leicht

Satz: In einem EUKLIDischen Ring gibt es zu je zwei Elementen $f, g \in R$ stets einen größten gemeinsamen Teiler. Dieser kann nach folgendem Algorithmus berechnet werden:
Schritt 0: Setze $r_0 = f$ und $r_1 = g$
Schritt i, $i \geq 1$: Falls $r_i = 0$ ist, endet der Algorithmus mit dem Ergebnis ggT$(f, g) = r_{i-1}$; andernfalls wird r_{i-1} mit Rest durch r_i dividiert, wobei r_{i+1} der Divisionsrest sei.
Der Algorithmus endet nach endlich vielen Schritten und liefert den größten gemeinsamen Teiler.

Beweis: Wir überlegen uns als erstes, daß im i-ten Schritt für $i \geq 1$ stets ggT(f, g) = ggT(r_{i-1}, r_i) ist. Für $i = 1$ gilt dies nach der Konstruktion im nullten Schritt. Falls es im i-ten Schritt für ein $i \geq 1$ gilt und der Algorithmus nicht mit dem i-ten Schritt abbricht, wird dort r_{i+1} als Rest bei der Division von r_{i-1} durch r_i berechnet; wie wir oben gesehen haben, ist somit ggT(r_i, r_{i+1}) = ggT(r_{i-1}, r_i), und das ist nach Induktionsvoraussetzung gleich dem ggT von f und g.

Falls der Algorithmus im i-ten Schritt abbricht, ist dort $r_i = 0$. Außerdem ist dort wie in jedem anderen Schritt auch ggT(f, g) = ggT(r_{i-1}, r_i). Somit ist r_{i-1} der ggT von f und g.

Schließlich muß noch gezeigt werden, daß der Algorithmus nach endlich vielen Schritten abbricht. Dazu dient die Funktion ν: Nach Definition eines EUKLIDischen Rings ist im i-ten Schritt entweder $\nu(r_i) < \nu(r_{i-1})$ oder $r_i = 0$. Da ν nur natürliche Zahlen und die Null als Werte annimmt und es keine unendliche absteigende Folge solcher Zahlen gibt, muß nach endlich vielen Schritten $r_i = 0$ sein, womit der Algorithmus abbricht.

■

Als erstes Beispiel wollen wir den EUKLIDischen Algorithmus anwenden auf zwei ganze Zahlen: Um den ggT von 200 und 148 zu berechnen, müssen wir als erstes 200 durch 148 dividieren: 200 : 148 = 1 Rest 52
Als nächstes wird 148 durch 52 dividiert: 148 : 52 = 2 Rest 44
Weiter geht es mit der Division von 52 durch 44: 52 : 44 = 1 Rest 8
Im nächsten Schritt dividieren wir 44 : 8 = 5 Rest 4 und kommen schließlich mit 8 : 4 = 2 Rest 0 zu einer Division,die aufgeht. Somit haben 200 und 148 den größten gemeinsamen Teiler vier.

Als zweites Beispiel wollen wir den größten gemeinsamen Teiler der 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 $\mathbb{Q}[x]$ berechnen. Da Polynomdivision aufwendiger ist als die obigen Rechnungen, wollen wir die Rechenarbeit von Maple erledigen lassen. Wir brauchen dazu im wesentlichen nur den Befehl `rem(f, g, X)`, der den Rest bei der Division von f durch g berechnet, wobei f und g als Polynome in X aufgefaßt werden. Falls uns auch der Quotient interessiert, können wir den durch `quo(f, g, X)` berechnen lassen. Alternativ können wir aber auch dem Befehl `rem` noch ein viertes Argument geben: Die Eingabe `rem(f, g, X, 'q')` führt auf dasselbe Ergebnis wie `rem(f, g, X)`, weist aber zusätzlich noch der Variablen q den Wert des Quotienten zu. Das q muß dabei in Hochkommata stehen, weil auf der linken Seite einer Zuweisung eine Variable stehen muß. Falls der Quotient etwa das Polynom X^2+X+1 wäre und die Variable q aus einer vorigen Rechnung den Wert $X-3$ hätte, würde `rem(f, g, X, q)` versuchen, die Zuweisung $X-3 := X^2+X+1$ auszuführen, was natürlich Unsinn ist und auf eine Fehlermeldung führt. Die Hochkommata in `'q'` sorgen dafür, daß unabhängig von einem etwaigen vorigen Wert von q in jedem Fall nur der Variablenname q verwendet wird, so daß die sinnvolle Anweisung $q := X^2 + X + 1$ ausgeführt wird.

```
> f := X^8 + X^6 - 3*X^4 - 3*X^3 + 8*X^2 + 2*X - 5;
```

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

```
> g := 3*X^6 + 5*X^4 - 4*X^2 - 9*X + 21;
```

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

```
> r2 := rem(f, g, X, 'q'); q;
```

$$r2 := -\frac{5}{9}X^4 + \frac{1}{9}X^2 - \frac{1}{3}$$

$$\frac{X^2}{3} - \frac{2}{9}$$

```
> r3 := rem(g, r2, X);
```

$$r3 := -\frac{117}{25}X^2 - 9X + \frac{441}{25}$$

```
> r4 := rem(r2, r3, X);
```

$$r4 = \frac{233150}{6591}X - \frac{102500}{2197}$$

```
> r5 := rem(r3, r4, X);
```

$$r5 := \frac{1288744821}{543589225}$$

```
> r6 := rem(r4, r5, X);
```

$$r6 := 0$$

Der ggT von f und g ist somit $r_5 = \frac{1288744821}{543589225}$. Da der ggT nur bis auf eine multiplikative Konstante bestimmt ist, können wir freilich genauso gut sagen, der ggT von f und g sei eins. In der Tat liefert uns Maple auch diese Antwort, wenn wir direkt nach dem ggT von f und g fragen:

```
> gcd(f, g);
```

$$1$$

Die Frage ist nun: Müssen wir wirklich mit so riesigen Brüchen wie r_5 rechnen, um auf diese einfache Antwort zu kommen?

Da der größte gemeinsame Teiler ohnehin nur bis auf eine multiplikative Konstante bestimmt ist, bestünde ein einfacher Ausweg darin, vor jeder Polynomdivision den Dividenden mit einer geeigneten Konstanten zu multiplizieren um so sicherzustellen, daß beim Dividieren keine Nenner auftreten. Bei der Division eines Polynoms vom Grad n durch ein Polynom vom Grad $m \leq n$ wird bis zu $n - m + 1$ mal durch den führenden Koeffizienten a des Divisors dividiert; wir müssen als den Dividenden vorher mit a^{n-m+1} multiplizieren. Im obigen Beispiel führt das auf folgende Rechnung:

```
> r2 := rem(3^3*f, g, X);
```

$$r2 := -15X^4 + 3X^2 - 9$$

```
> r3 := rem((-15)^3*g, r2, X);
```

$$r3 := 15795X^2 + 30375X - 59535$$

```
> r4 := rem(15795^3*r2, r3, X);
```

$$r4 := 1254542875143750X - 1654608338437500$$

```
> r5 := rem(1254542875143750^2*r3, r4, X);
```

$$r5 := 1259333879550074310093114199218750 0$$

Verglichen mit der Größe der Ausgangsdaten und des Ergebnisses entstehen auch hier wieder riesige Zahlen. Das ist leider kein Einzelfall: Auch wenn es sich hier um ein (von DONALD E. KNUTH für sein Buch *The Art of Computer Programming,* Abschnitt 4.6.1) konstruiertes besonders extremes Beispiel handelt, zeigt die Erfahrung, daß wir es beim EUKLIDischen Algorithmus für Polynome über den rationalen Zahlen oft mit einer Explosion der Koeffizienten zu tun haben, die in keiner Weise der Komplexität des Ergebnisses entspricht. Wenn wir den Algorithmus ernsthaft auf größere Polynome anwenden wollen, sollten wir nach Wegen suchen, um dieses Problem zu umschiffen.

Solche Wege gibt es in der Tat; zwei davon werden wir gegen Ende der Vorlesung noch kennenlernen.

§3: Der erweiterte Euklidische Algorithmus

Zur Bestimmung des ggT zweier Elemente eines EUKLIDischen Rings R berechnen wir eine Reihe von Elementen r_i, wobei r_0 und r_1 die Ausgangsdaten sind und alle weiteren r_i durch Division mit Rest ermittelt werden:

$$r_{i-1} : r_i = q_i \text{ Rest } r_{i+1}$$

Damit ist $r_{i+1} = r_{i-1} - q_i r_i$ als Linearkombination seiner beiden Vorgänger r_i und r_{i-1} mit Koeffizienten aus R darstellbar, die wiederum R-Linearkombinationen *ihrer* Vorgänger sind, *usw.* Wenn wir alle diese Darstellungen ineinander einsetzen, erhalten wir r_i schließlich als Linearkombination der Ausgangselemente. Dies gilt insbesondere für das letzte nichtverschwindende r_i, den ggT. Der ggT zweier Elemente f, g eines EUKLIDischen Rings ist somit darstellbar als R-Linearkombination von f und g.

Algorithmisch sieht dies folgendermaßen aus:

Schritt 0: Setze $r_0 = f$, $r_1 = g$, $\alpha_0 = \beta_1 = 1$ 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

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

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

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

man setze also

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

Da die Schritte hier einfach Erweiterungen der entsprechenden Schritte des klassischen EUKLIDischen Algorithmus sind, ist klar, daß auch dieser Algorithmus nach endlich vielen Schritten abbricht und als Ergebnis den ggT liefert. Da die beiden Relationen aus Schritt 0 in allen weiteren Schritten erhalten bleiben, ist auch klar, daß dieser ggT am Ende als Linearkombination dargestellt ist.

Obwohl es keinerlei Anhaltspunkt dafür gibt, daß diese Erweiterung EUKLID bekannt gewesen sein könnte, bezeichnet man sie als den *erweiterten* EUKLIDischen Algorithmus, Vor allem in der französischen Literatur wird die Darstellung des ggT als Linearkombination auch als Identität von BÉZOUT bezeichnet, da dieser sie 1766 in einem Lehrbuch beschrieb und als erster auch auf Polynome anwandte. Für Zahlen ist die Erweiterung jedoch bereits 1624 zu finden in der zweiten Auflage des Buchs *Problèmes plaisants et délectables qui se fonts par les nombres* von BACHET DE MÉZIRIAC. (Eine vereinfachte Ausgabe dieses Buchs von 1874 wurde 1993 bei Blanchard neu aufgelegt; sie ist auch online verfügbar unter `cnum.cnam.fr/DET/8PY45.html`.)

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

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

Als Beispiel wollen wir den ggT von 200 und 148 als Linearkombination darstellen. Der Rechengang ist natürlich genau derselbe wie in §3, nur daß wir jetzt noch in jedem Schritt den Divisionsrest als ganzzahlige Linearkombination von 200 und 148 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 $52 \neq 0$ ist, 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 $44 \not| 0$, wir dividieren also weiter: $52 = 1 \cdot 44 + 8$ und

$$8 = 52 - 44 = (1 \cdot 200 - 1 \cdot 148) - (3 \cdot 148 - 2 \cdot 200)$$
$$= 3 \cdot 200 - 4 \cdot 148 .$$

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

$$4 = 44 - 5 \cdot 8 = (3 \cdot 148 - 2 \cdot 200) - 5 \cdot (3 \cdot 200 - 4 \cdot 148)$$
$$= 23 \cdot 148 - 17 \cdot 200 .$$

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. Diese Darstellung ist freilich nicht eindeutig: Beispielsweise können wir beliebige Vielfache der trivialen Darstellung $200 \cdot 148 - 148 \cdot 200 = 0$ der Null addieren. Tatsächlich können wir diese auch noch durch den ggT kürzen zu $50 \cdot 148 - 36 \cdot 200 = 0$; wir haben also für jede ganze Zahl k eine Darstellung $1 = (23 + 50k) \cdot 148 - (17 + 36k) \cdot 200$.

Wir können den erweiterten EUKLIDischen Algorithmus natürlich auch auf die beiden Polynome f und g aus dem vorigen Paragraphen anwenden, allerdings ist das Ergebnis alles andere als schön: $1 = \alpha f + \beta g$, wobei α ein Polynom vom Grad fünf und β eines vom Grad sieben ist. Der Hauptnenner der Koeffizienten ist in beiden Fällen 130 354. Wir können den Algorithmus allerdings verwenden, um ein negatives Resultat zu beweisen:

Lemma: Der Ring $\mathbb{Z}[X]$ aller Polynome mit ganzzahligen Koeffizienten ist nicht EUKLIDisch.

Beweis: Wir wissen zwar noch nicht, daß zwei beliebige Elemente von $\mathbb{Z}[X]$ auch in $\mathbb{Z}[X]$ einen größten gemeinsamen Teiler haben, es ist aber klar, daß der größte gemeinsame Teiler der beiden Polynome X und 2 existiert und eins ist: Die einzigen Teiler von 2 sind ± 1 und ± 2, und ± 2 sind keine Teiler von X. Wäre $\mathbb{Z}[X]$ ein EUKLIDischer Ring, gäbe es also Polynome $\alpha, \beta \in \mathbb{Z}[X]$, so daß $\alpha X + 2\beta = 1$ wäre. Der konstante Koeffizient von $\alpha X + 2\beta$ ist aber das Doppelte des konstanten Koeffizienten von β, also eine gerade Zahl. Somit kann es keine solche Darstellung geben.

■

(In $\mathbb{Q}[x]$ gibt es selbstverständlich so eine Darstellung: $1 = 0 \cdot X + \frac{1}{2} \cdot 2$. Allerdings ist dort 2 auch ein Teiler von X.)

§4: Faktorielle Ringe

In einfachen Fällen berechnet man in der Schule den größten gemeinsamen Teiler zweier Zahlen über deren Primzerlegung. Auch diesen Zugang wollen wir auf eine größere Klasse von Ringen verallgemeinern: Er wird sich zwar für die effiziente Berechnung des ggT als nicht sonderlich nützlich erweisen, gibt uns aber für die Computeralgebra wichtiges Hintergrundwissen.

Definition: *a)* Ein Element f eines Integritätsbereichs R heißt *irreduzibel*, falls gilt: f ist keine Einheit, und ist $f = gh$ das Produkt zweier Elemente aus R, so muß g oder h eine Einheit sein.
b) Ein Integritätsbereich R heißt *faktoriell* oder *ZPE*-Ring, wenn gilt: Jedes Element $f \in R$ läßt sich bis auf Reihenfolge und Assoziiertheit eindeutig schreiben als Produkt $f = u \prod_{i=1}^{n} p_i^{e_i}$ mit einer Einheit $u \in R^\times$, irreduziblen Elementen $p_i \in R$ und natürlichen Zahlen e_i.
(*ZPE* steht für **Z**erlegung in **P**rimfaktoren **E**indeutig.)

Lemma: In einem faktoriellen Ring R gibt es zu je zwei Elementen $f, g \in R$ einen größten gemeinsamen Teiler.

Beweis: Sind $f = u \prod_{i=1}^{n} p_i^{e_i}$ und $g = v \prod_{j=1}^{m} q_j^{d_j}$ mit $u, v \in R^\times$ und p_i, q_j irreduzibel die Zerlegungen von f und g in Primfaktoren, so können wir, indem wir nötigenfalls Exponenten null einführen, o.B.d.A. annehmen, daß $n = m$ ist und $p_i = q_i$ für alle i. Dann ist offenbar $\prod_{i=1}^{n} p_i^{\min(e_i, d_i)}$ ein ggT von f und g, denn $h = \prod_{i=1}^{r} p_i^{a_i}$ ist genau dann Teiler von f, wenn $a_i \leq e_i$ für alle i, und Teiler von g, wenn $a_i \leq d_i$. ∎

Wie wir bald sehen werden, bedeutet dies keineswegs, daß jeder faktorielle Ring EUKLIDisch wäre. Umgekehrt gilt allerdings

Satz: Jeder EUKLIDische Ring ist faktoriell.

Beweis: Wir müssen zeigen, daß jedes Element $f \neq 0$ 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 f überhaupt so darstellen läßt.

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

Ist $\nu(f) = 0$, so ist f eine Einheit: Bei der Division $1 : f = g$ Rest r ist nämlich entweder $r = 0$ oder aber $\nu(r) < \nu(f) = 0$. Letzteres ist nicht möglich, also ist $gf = 1$ und f 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 f irreduzibel, so ist $f = f$ eine Darstellung der gewünschten Form, und wir sind fertig.

Andernfalls läßt sich $f = gh$ als Produkt zweier Elemente schreiben, die beide keine Einheiten sind. Nach Definition eines EUKLIDischen Rings sind dann $\nu(g), \nu(h) \leq \nu(f)$. Wir wollen uns überlegen, daß sie tatsächlich sogar echt kleiner sind.

Dazu dividieren wir g mit Rest durch f; das Ergebnis sei q Rest r, d.h. $g = qf + r$ mit $r = 0$ oder $\nu(r) < \nu(f)$. Wäre $r = 0$, wäre g ein Vielfaches von f, es gäbe also ein $u \in R$ mit $g = uf = u(gh) = (uh)g$. Damit wäre $uh = 1$, also h eine Einheit, im Widerspruch zur Annahme. Somit ist $\nu(r) < \nu(f)$. Außerdem ist g ein Teiler von $r = g - qf = g(1 - qh)$, also muß gelten $\nu(g) \leq \nu(r) < \nu(f)$.

Genauso folgt die strikte Ungleichung $\nu(h) < \nu(f)$.

Nach Induktionsvoraussetzung lassen sich daher g und h als Produkte von Einheiten und Potenzen irreduzibler Elemente schreiben, und damit läßt sich auch $f = gh$ 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 fg teilt, teilt es mindestens einen der beiden Faktoren.

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

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

als Linearkombination von p und f schreiben. Multiplikation mit g macht daraus $g = \alpha pf + \beta fg$, und hier sind beide Summanden auf der rechten Seite durch p teilbar: Bei αpf ist das klar, und bei βfg folgt es daraus, daß nach Voraussetzung p ein Teiler von fg ist. Also ist p Teiler von g, und die Zwischenbehauptung ist bewiesen.

Induktiv folgt sofort:

Falls ein irreduzibles Element $p \in R$ ein Produkt $\prod_{i=1}^{n} f_i$ 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 $\nu(f) \leq N$ eine bis auf Reihenfolge und Einheiten *eindeutige* Primfaktorzerlegung haben.

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

Für $N \geq 1$ betrachten wir ein Element

$$f = u \prod_{i=1}^{n} p_i^{e_i} = v \prod_{j=1}^{m} q_j^{d_j}$$

mit zwei Zerlegungen, wobei wir annehmen können, daß alle $e_i, f_j \geq 1$ sind. Dann ist p_1 trivialerweise Teiler des ersten Produkts, also auch des zweiten. Wegen der Zwischenbehauptung teilt p_1 daher mindestens eines der Elemente q_j, d.h. $p_1 = wq_j$ ist bis auf eine Einheit w gleich q_j. Da p_i keine Einheit ist, ist $\nu(f/p_i) < \nu(f)$; nach Induktionsannahme hat also $f/p_i = x/(wq_j)$ eine bis auf Reihenfolge und Einheiten eindeutige Zerlegung in irreduzible Elemente. Damit hat auch x diese Eigenschaft. ∎

Als nächstes wollen wir Produktzerlegungen in $\mathbb{Q}[X]$ vergleichen mit solchen in $\mathbb{Z}[X]$. Das entsprechende Argument von GAUSS wird uns

auch nützlich sein für den Beweis der Faktorialität von Polynomringen in mehreren Veränderlichen; wir fassen es daher gleich etwas allgemeiner.

Dazu brauchen wir als erstes für einen beliebigen Integritätsbereich eine Entsprechung für die rationalen Zahlen, den sogenannten Quotientenkörper, der genauso konstruiert wird wie die rationalen Zahlen aus den ganzen:

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

$$(f, g) \sim (r, s) \iff fs = gr\,;$$

die Äquivalenzklasse von (f, g) bezeichnen wir als den *Bruch* $\dfrac{f}{g}$.

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

$$\frac{f}{g} + \frac{r}{s} = \frac{fs + rg}{gs} \quad \text{und} \quad \frac{f}{g} \cdot \frac{r}{s} = \frac{fr}{gs}\,.$$

Dies ist wohldefiniert, denn sind $(f, g) \sim (\tilde{f}, \tilde{g})$ und $(r, s) \sim (\tilde{r}, \tilde{s})$, so ist

$$\frac{\tilde{f}}{\tilde{g}} + \frac{\tilde{r}}{\tilde{s}} = \frac{\tilde{f}\tilde{s} + \tilde{r}\tilde{g}}{\tilde{g}\tilde{s}} \quad \text{und} \quad \frac{\tilde{f}}{\tilde{g}} \cdot \frac{\tilde{r}}{\tilde{s}} = \frac{\tilde{f}\tilde{r}}{\tilde{g}\tilde{s}}\,.$$

Wegen $f\tilde{g} = \tilde{f}g$ und $r\tilde{s} = \tilde{r}s$ ist

$$\begin{aligned}(\tilde{f}\tilde{s} + \tilde{r}\tilde{g}) \cdot gs &= \tilde{f}\tilde{s}gs + \tilde{r}\tilde{g}gs = \tilde{f}gs\tilde{s} + \tilde{r}sg\tilde{g} \\ &= g\tilde{g}s\tilde{s} + r\tilde{s}g\tilde{g} = (gs + ry)\tilde{g}\tilde{s}\end{aligned}$$

und $(\tilde{f}\tilde{r})(gs) = \tilde{f}g\tilde{r}s = g\tilde{g}r\tilde{s} = (gr)(\tilde{g}\tilde{s})$, d.h. auch die Ergebnisse sind äquivalent.

Man rechnet leicht nach (wie bei $\mathbb{Q}$), daß diese Äquivalenzklassen einen Ring bilden mit $\frac{0}{1}$ als Null und $\frac{1}{1}$ als Eins; er ist sogar ein Körper, denn für $f, g \neq 0$ ist $\frac{g}{s}$ ein multiplikatives Inverses zu $\frac{f}{g}$, da $(fg, fg) \sim (1, 1)$. Identifizieren wir schließlich ein Element $f \in R$ mit dem Bruch $\frac{f}{1}$, so können wir R in den Körper K einbetten.

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

Das Standardbeispiel ist natürlich $\mathbb{Q} = \operatorname{Quot}\mathbb{Z}$, aber auch der Quotientenkörper $k(X) \underset{\text{def}}{=} \operatorname{Quot} k[X]$ eines Polynomrings über einem Körper k ist wichtig: $k(X)$ heißt rationaler Funktionenkörper in einer Veränderlichen über k. Seine Elemente sind rationale Funktionen in X, d.h. Quotienten von Polynomen in X, wobei der Nenner natürlich nicht das Nullpolynom sein darf.

Für Polynome, die statt über einem Körper nur über einem faktoriellen Ring definiert sind, sind die beiden folgenden Begriffe sehr wesentlich:

Definition: *a)* Der *Inhalt* eines Polynoms $f = a_d X^d + \cdots + a_0 \in R[X]$ ist der größte gemeinsame Teiler $I(f)$ seiner 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 faktorieller Ring. Für zwei Polynome

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

und

$$g = b_e X^e + b_{e-1} X^{e-1} + \cdots + b_1 X + b_0$$

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

$$fg = c_{d+e} X^{d+e} + c_{d+e-1} X^{d+e-1} + \cdots + c_1 X + c_0 \,;$$

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, der keine Einheit ist. Wegen der Faktorialität von R gibt es dann auch ein irreduzibles Element p, das alle Koeffizienten c_r teilt.

Insbesondere ist p ein Teiler von $c_0 = a_0 b_0$; da p irreduzibel ist, muß mindestens einer der beiden Faktoren a_0, b_0 durch p teilbar sein. Da es im Lemma nicht auf die Reihenfolge von f und g ankommt, können wir o.B.d.A. annehmen, daß a_0 Vielfaches von p ist.

Da f ein primitives Polynom ist, kann nicht jeder Koeffizient a_i durch p teilbar sein; ν sei der kleinste Index, so daß a_ν kein Vielfaches von p ist. Genauso gibt es auch einen kleinsten Index $\mu \geq 0$, für den b_μ nicht durch p teilbar ist. In

$$c_{\mu+\nu} = \sum_{i,j \text{ mit } i+j=\mu+\nu} a_i b_j$$

ist dann der Summand $a_\nu b_\mu$ nicht durch p teilbar, denn für jeden anderen Summanden $a_i b_j$ ist entweder $i < \nu$ oder $j < \mu$, so daß mindestens einer der Faktoren und damit auch das Produkt durch p teilbar ist. Insgesamt ist daher $c_{\mu+\nu}$ nicht durch p teilbar, im Widerspruch zur Annahme.

Somit muß fg ein primitives Polynom sein.

■

Satz von Gauß: R sei ein faktorieller Ring und $K = \operatorname{Quot} R$. Falls sich ein Polynom $f \in R[X]$ in $K[X]$ als Produkt zweier Polynome $g, h \in K[X]$ schreiben läßt, gibt es ein $\lambda \in K$, so daß $\widetilde{g} = \lambda g$ und $\widetilde{h} = \lambda^{-1} h$ in $R[X]$ liegen und $f = \widetilde{g} \cdot \widetilde{h}$.

Beweis: Durch Multiplikation mit einem gemeinsamen Vielfache 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.

■

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

Für nichtprimitive Polynome gilt diese Aussage natürlich nicht: Das Polynom $2X + 2$ ist zwar irreduzibel in $\mathbb{Q}[X]$, hat aber in $\mathbb{Z}[X]$ die beiden irreduziblen Faktoren 2 und $X + 1$.

CARL FRIEDRICH GAUSS (1777–1855) leistete wesentliche Beiträge zur Zahlentheorie, zur nichteuklidischen Geometrie, zur Differentialgeometrie und Kartographie, zur Fehlerrechnung und Statistik, zur Astronomie und Geophysik *usw.* Als Direktor der Göttinger Sternwarte baute er zusammen mit dem Physiker Weber den ersten Telegraphen. Er leitete die erste Vermessung und Kartierung des Königreichs Hannover, was sowohl seine Methode der kleinsten Quadrate als auch sein *Theorema egregium* motivierte, und zeitweise auch den Witwenfond der Universität Göttingen. Seine hierbei gewonnene Erfahrung nutzte er für erfolgreiche Spekulationen mit Aktien.

Aus dem Satz von GAUSS folgt induktiv sofort, daß seine Aussage auch für Produkte von mehr als zwei Polynomen gilt, und daraus folgt

Satz: Der Polynomring über einem faktoriellen Ring R ist faktoriell.

Beweis: Wir müssen zeigen, daß sich jedes $f \in R[X]$ bis auf Reihenfolge und Einheiten eindeutig als Produkt von Potenzen irreduzibler Elemente aus $R[X]$ und einer Einheit schreiben läßt. Dazu schreiben wir $f = I(f) \cdot f^*$ mit einem primitiven Polynom $f^* \in R[X]$ und zerlegen zunächst den Inhalt $I(f)$ in R. Da R nach Voraussetzung faktoriell ist, ist diese Zerlegung eindeutig bis auf Reihenfolge und Einheiten in R, und wie wir aus §1 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 Polynome in mehreren Veränderlichen über $\mathbb{Z}$ oder über einem Körper in Produkte irreduzibler Polynome zerlegt werden können; insbesondere existieren daher auch in dieser Ringen größte gemeinsame Teiler.

Der Beweis des obigen Satzes ist allerdings nicht konstruktiv; wir werden im nächsten Kapitel noch viel Arbeit investieren müssen, bevor wir Polynome über $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{F}_p$ oder anderen Körpern, in denen wir rechnen können, wirklich in ihre irreduziblen Faktoren zerlegen können.

§5: Resultanten

In diesem Paragraphen wollen wir ein Kriterium dafür herleiten, daß zwei Polynome einen gemeinsamen Teiler vom Grad mindestens r haben ohne daß wir so einen Teiler wirklich berechnen müssen. Wir betrachten also zwei Polynome

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

und

$$g = b_e X^e + b_{e-1} X^{e-1} + \cdots + b_1 X + b_0$$

über einem faktoriellen Ring R, von denen wir annehmen wollen, daß sie wirklich Grad d *bzw.* e haben, d.h. $a_d \neq 0$ und $b_e \neq 0$.

Wir nehmen an, es gebe ein Polynom $h \in R[X]$ vom Grad mindestens $r \geq 1$, das sowohl f als auch g teilt. Dann ist

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

ein gemeinsames Vielfaches von f und q, dessen Grad

$$\deg f + \deg g - \deg h = d + e - \deg h$$

höchstens gleich $d + e - r$ ist.

Haben umgekehrt f und g ein gemeinsames Vielfaches vom Grad höchstens $d + e - r$, so hat auch ihr kleinstes gemeinsames Vielfaches S höchstens den Grad $d + e - r$. (Ein kleinstes gemeinsames Vielfaches existiert, da mit R auch $R[x]$ faktoriell ist.)

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

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

es teilt also sowohl f als auch g und sein Grad $d + e - \deg S$ ist mindestens r. Damit ist gezeigt:

Lemma: Zwei Polynome $f, g \in R[X]$ haben genau dann einen gemeinsamen Teiler vom Grad mindestens r, wenn es nichtverschwindende Polynome $u, v \in R[X]$ mit $uf = vg$ und Graden $\deg u \leq \deg g - r$ und $\deg v \leq \deg f - r$.

■

Diese Bedingung schreiben wir um in ein lineares Gleichungssystem für die Koeffizienten von u und v: Da $\deg u \leq \deg g - r = e - r$ ist und

$\deg v \leq \deg f - r = d - r$, lassen sich die beiden Polynome schreiben als

$$u = u_{e-r}X^{e-r} + u_{e-r-1}X^{e-r-1} + \ldots + u_1X + u_0$$

und

$$v = v_{d-r}X^{d-r} + v_{d-r-1}X^{d-r-1} + \ldots + v_1X + v_0\,.$$

Die Koeffizienten von X^r in uf und vg sind

$$\sum_{i,j \text{ mit } i+j=r} a_iu_j \quad \text{und} \quad \sum_{i,j \text{ mit } i+j=r} b_iv_j\,,$$

f und g haben daher genau dann einen gemeinsamen Teiler vom Grad mindestens r, wenn es nicht allesamt verschwindende Körperelemente $u_0, \ldots, u_{e-r}$ und $v_0, \ldots, v_{d-r}$ gibt, so daß

$$\sum_{i,j \text{ mit } i+j=r} a_iu_j - \sum_{i,j \text{ mit } i+j=r} b_iv_j = 0 \quad \text{für } r = 0, \ldots, n+m-d$$

ist. Dies ist ein homogenes lineares Gleichungssystem aus $d + e + 1 - r$ Gleichungen für die $d + e + 2 - 2e$ Unbekannten $u_0, \ldots, u_{e-r}$ und $v_0, \ldots, v_{d-r}$; es hat genau dann eine nichttriviale Lösung, wenn seine Matrix kleineren Rang als $d + e + 2 - 2r$ hat. Im Falle $r = 1$ ist die Matrix quadratisch; hier bedeutet dies einfach, daß ihre Determinante verschwindet. Für $r > 1$ müssen wir r Determinanten von quadratischen Untermatrizen betrachten

Ausgeschrieben wird das Gleichungssystem, wenn wir mit dem Koeffizienten von x^{d+e-r} anfangen, zu

$$\begin{aligned}
a_du_{e-r} - b_ev_{d-r} &= 0\\
a_{d-1}u_{e-r} + a_du_{e-r-1} - b_{e-1}v_{d-r} - b_ev_{d-r-1} &= 0\\
a_{d-2}u_{e-r} + a_{d-1}u_{e-r-1} + a_du_{e-r-2} & \\
-b_{e-2}v_{d-r} - b_{e-1}v_{d-r-1}b_ev_{d-r-2} &= 0\\
\ldots & \\
a_0u_3 + a_1u_2 + a_2u_1 + a_3u_0 - b_0v_3 - b_1v_2 - b_2v_1 - b_3v_0 &= 0\\
a_0u_2 + a_1u_1 + a_2u_0 - b_0v_2 - b_1v_1 - b_2v_0 &= 0\\
a_0u_1 + a_1u_0 - b_0v_1 - b_1v_0 &= 0\\
a_0u_0 - b_0v_0 &= 0
\end{aligned}$$

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 $(d + e + 2 - 2r) \times (d + e + 1 - r)$-Matrix

$$\begin{pmatrix}
a_d & a_{d-1} & a_{d-2} & \dots & a_1 & a_0 & 0 & 0 & \dots & 0 \\
0 & a_d & a_{d-1} & \dots & a_2 & a_1 & a_0 & 0 & \dots & 0 \\
0 & 0 & a_d & \dots & a_3 & a_2 & a_1 & a_0 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \dots & a_d & a_{d-1} & a_{d-2} & a_{d-3} & \dots & a_0 \\
b_e & b_{e-1} & b_{e-2} & \dots & b_2 & b_1 & b_0 & 0 & \dots & 0 \\
0 & b_e & b_{e-1} & \dots & b_3 & b_2 & b_1 & b_0 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \dots & 0 & b_e & b_{e-1} & b_{e-2} & \dots & b_0
\end{pmatrix},$$

in der $e + 1 - r$ Zeilen aus Koeffizienten von f stehen und $d + 1 - r$ Zeilen aus Koeffizienten von g.

Für $r = 1$ ist diese Matrix quadratisch; das Gleichungssystem hat also genau dann nichttriviale Lösungen, wenn ihre Determinante verschwindet.

Definition: Im Fall $r = 1$ wird die obige Matrix als SYLVESTER-Matrix bezeichnet und ihre Determinante als die *Resultante*

$$\mathrm{Res}(f, g) = \mathrm{Res}_X(f, g)$$

der beiden Polynome f und g bezüglich der Variablen X.

Der Index X ist dann notwendig, wenn f und g Polynome in mehreren Veränderlichen sind und nicht klar ist, bezüglich welcher von ihnen die Resultante gebildet wird.

Für $r > 1$ betrachten wir für $j = d + e + 1 - r, \dots, d + e + 2 - 2r$ jene quadratische Matrix M_j, die aus den ersten $d + e + 2 - r$ Spalten sowie der j-ten Spalte der obigen Matrix besteht. Offensichtlich hat letztere

genau dann einen kleineren Rang als $d + e + 1 - r$, wenn alle r Matrizen M_j singulär sind, wenn also deren Determinanten verschwinden. Diese Determinanten bezeichnet man als *Subresultanten.*

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 Rücksicht auf 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 noch heute mit die wichtigste mathematische Fachzeitschrift Amerikas ist.

Sowohl die Resultante als auch die Subresultanten sind Polynome in den Koeffizienten von f und g; wir haben daher den

Satz: Zwei Polynome $f, g \in R[X]$ über dem faktoriellen Ring R haben genau dann einen gemeinsamen Faktor vom Grad mindestens d, wenn gewisse Polynome in ihren Koeffizienten verschwinden. Insbesondere gibt es genau dann einen gemeinsamen Faktor positiven Grades, wenn die Resultante der beiden Polynome verschwindet.

■

§6: Die Berechnung der Resultante

Die Resultante zweier Polynome der Grade 30 und 40 ist eine 70×70-Determinante – nichts, was man mit den aus der Linearen Algebra bekannten Algorithmen leicht und schnell ausrechnen könnte. Tatsächlich verwendet aber natürlich ohnehin niemand den Entwicklungssatz von LAGRANGE um eine große Determinante zu berechnen;

dessen Nützlichkeit beschränt sich definitiv auf kleineren Spielzeugdeterminanten, wie sie vor allem in Mathematikklausuren vorkommen. In realistischen Anwendungen wird man die Matrix durch Zeilen- und/oder Spaltenoperationen auf Dreiecksform bringen und dann die Determinante einfach als Produkt der Diagonaleinträge berechnen oder man tat dies über eine LR- oder QR-Zerlegung. 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 einmal der EUKLIDische Algorithmus.

Angenommen, wir haben zwei Polynome f, g in einer Variablen X über einem faktoriellen Ring R:

$$f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0 \quad \text{und}$$
$$g = b_e X^e + b_{e-1} X^{e-1} + \cdots + b_1 X + b_0 \quad \text{mit} \quad d \leq e\,.$$

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

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

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

Das ist freilich nur dann möglich, wenn $R[X]$ ein EUKLIDischer Ring ist, also im wesentlichen nur dann, wenn R ein Körper ist. Das ist aber keine so große Einschränkung wie es scheint, denn wir können statt in $R[X]$ im Polynomring über dem Quotientenkörper von R rechnen; da die Resultante ein eindeutig bestimmtes Element von R ist, muß spätestens das Endergebnis unserer Rechnung in R liegen. Die Zwischenergebnisse können freilich recht große Nenner bekommen – ein wohlbekanntes Problem der Computeralgebra, das uns bereits beim EUKLIDischen Algorithmus für Polynome begegnet ist.

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 $\mathrm{Res}_X(f,g)$.

Bei der Polynomdivision berechnen wir 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

$$\deg g - \deg f = e - d\,.$$

Wir wollen uns überlegen, wie sich die SYLVESTER-Matrix ändert, wenn wir dort die Koeffizienten von $g_0 = g$ nacheinander durch die der nachfolgenden g_i ersetzen. Um die Gestalt der Matrix nicht zu verändern, betrachten wir dazu auch die g_i als Polynome vom Grad m, indem wir die Koeffizienten aller x-Potenzen mit einem Exponent oberhalb $\deg g_i$ auf Null setzen.

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

Im ersten Divisionschritt subtrahieren wir von g ein Vielfaches $\lambda X^j f$ mit $j = e - d$; damit subtrahieren wir auch von jeder Potenz $X^i g$ das Polynom $\lambda X^{i+j} f$. Für $0 \le i < d$ und $0 \le j \le e + d$ ist $0 \le i + j < e$, 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 wir in der SYLVESTER-Matrix jede Zeile mit Koeffizienten von g ersetzen durch die entsprechende

Zeile mit Koeffizienten von h, wobei h als ein Polynom vom Grad e behandelt wird, dessen führende Koeffizienten verschwinden.

Ist $h = c_s X^s + \cdots + c_1 X + c_0$, so ist also $\mathrm{Res}_X(f,g)$ gleich

$$\begin{vmatrix}
a_d & a_{d-1} & a_{d-2} & \dots & a_1 & a_0 & 0 & 0 & \dots & 0 \\
0 & a_d & a_{d-1} & \dots & a_2 & a_1 & a_0 & 0 & \dots & 0 \\
0 & 0 & a_d & \dots & a_3 & a_2 & a_1 & a_0 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \dots & a_d & a_{d-1} & a_{d-2} & a_{d-3} & \dots & a_0 \\
c_e & c_{e-1} & c_{e-2} & \dots & c_2 & c_1 & c_0 & 0 & \dots & 0 \\
0 & c_e & c_{e-1} & \dots & c_3 & c_2 & c_1 & c_0 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \dots & 0 & c_e & c_{e-1} & c_{e-2} & \dots & c_0
\end{vmatrix},$$

wobei die Koeffizienten $c_e, \dots, c_{s+1}$ alle verschwinden.

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

In den ersten $e - s$ Spalten der Matrix stehen daher nur noch Koeffizienten von f: In der ersten ist dies ausschließlich der führende Koeffizient a_d von f in der ersten Zeile. Entwickeln wir nach der ersten Zeile, können wir also einfach die erste Zeile und die erste Spalte streichen; die Determinante ist dann a_d mal der Determinante der übrigbleibenden Matrix. Diese hat (falls $e > s+1$) wieder dieselbe Gestalt, wir können also wieder einen Faktor a_d ausklammern und bekommen eine Determinante mit einer Zeile und einer Spalte weniger *usw.* Das Ganze funktioniert $e - s$ mal; dann ist der führende Koeffizient von h in die erste Spalte gerutscht und die übriggebliebene Matrix ist die SYLVESTER-Matrix 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 $\mathrm{Res}_X(f,g) = a_d^{e-s} \mathrm{Res}_X(f,h)$, 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 $\mathrm{Res}_X(f,g) = a_d^{e-s} \mathrm{Res}_X(f,h)$. ∎

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 d und ein Polynom g vom Grad e ist $\mathrm{Res}_X(f,g) = (-1)^{de}\,\mathrm{Res}_X(g,f)$.

Beweis: Wir müssen in der SYLVESTER-Matrix e Zeilen zu f mit den d 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 d Vertauschungen schließlich unten steht. Dies müssen wir wiederholen, bis alle f-Zeilen unten stehen, wir haben also insgesamt de Zeilenvertauschungen. Somit ändert sich das Vorzeichen der Determinante um den Faktor $(-1)^{de}$.

■

Zum Abschluß dieses Paragraphen wollen wir uns noch überlegen, daß die Resultante zweier Polynome noch aus einem anderen Grund für jede gemeinsame Nullstelle verschwinden muß: Sie läßt sich nämlich als Linearkombination der beiden Polynome darstellen:

Lemma: R sei ein Ring und $f, g \in R[X]$ seien Polynome über R. Dann gibt es Polynome $p, q \in R[X]$, so daß $\mathrm{Res}_X(f,g) = pf + qg$ ist.

Man beachte, daß p, q, f und g zwar Polynome sind, die Resultante aber nur ein Element von R.

Beweis: Wir schreiben

$$f = a_d X^d + \cdots + a_1 X + a_0 \quad \text{und} \quad g = b_e X^e + \cdots + b_1 X + b_0\,,$$

wobei wir annehmen können, daß a_d und b_e beide nicht verschwinden. Die Gleichungen

$$X^i f = a_d X^{d+i} + \cdots + a_1 X^{1+i} + a_0 X^i \quad \text{für } i = 0, \ldots, e-1$$

und

$$X^j g = b_e X^{e+j} + \cdots + b_1 X^{1+j} + b_0 X^j \quad \text{für } j = 0, \ldots, d-1$$

können wir in Vektorschreibweise so zusammenfassen, daß wir den $(d+e)$-dimensionalen Vektor

$$F = \left(X^{e-1} f, \ldots, Xf, f, X^{d-1} g, \ldots, Xg, g\right)^T \in R[X]^{d+e}$$

darstellen in der Form

$$F = X^{d+e-1} r_1 + \cdots + X^1 r_{d+e-1} + X^0 r_{d+e}$$

mit Vektoren $r_k \in R^{d+e}$, deren Einträge Koeffizienten von f und g sind. Die Resultante ist nach Definition gleich der Determinanten der $(d+e) \times (d+e)$-Matrix mit den r_k als Spaltenvektoren.

Nun gehen wir vor, wie bei der Herleitung der CRAMERschen Regel: Wir betrachten obige Vektorgleichung als ein lineares Gleichungssystem mit rechter Seite F in den „Unbekannten“ X^k und tun so, als wollten wir den Wert von $X^0 = 1$ aus diesem Gleichungssystem bestimmen. Dazu ersetzen wir nach CRAMER in der Determinante des Gleichungssystems die letzte Spalte durch die rechte Seite, berechnen also die Determinante

$$\begin{aligned}
\det(r_1, \ldots, r_{d+e-1}, F) &= \det\left(r_1, \ldots, r_{d+e-1}, \sum_{k=1}^{d+e} x^{d+e-k} r_k\right) \\
&= \sum_{k=1}^{d+e} x^{d+e-k} \det(r_1, \ldots, r_{d+e-1}, r_k) \\
&= \det(r_1, \ldots, r_{d+e-1}, r_{d+e}),
\end{aligned}$$

denn für $k \neq d+e$ steht die Spalte r_k zweimal in der Matrix, so daß die Determinante verschwindet.

Wenn wir bei der Berechnung von $\det(f_1, \ldots, r_{d+e-1})$ nach dem LAGRANGEschen Entwicklungssatz die Polynome f und g in F stehen lassen, erhalten wir die Determinante als Ausdruck der Form $pf + qg$ mit Polynomen p und q aus $R[X]$: Da f und g beide nur in der letzten Spalte vorkommen, dort aber in jedem Eintrag genau eines der beiden, enthält jedes der $(d+e)!$ Produkte, die nach LAGRANGE aufsummiert

werden, genau eines der beiden Polynome. Nach der obigen Rechnung ist $pf + qg$ gleich der Determinante der r_k, also die Resultante.

■