Kapitel 0
Semialgebraische Mengen

Das klassische Thema der algebraischen Geometrie sind die Nullstellenmengen von Polynomgleichungen, also beispielsweise der Einheitskreis um den Nullpunkt als Nullstellenmenge des Polynoms $x^2 + y^2 - 1$ oder der durch die beiden Gleichungen $x^2 + y^2 + z^2 - 1 = 0$ und $2z - 1 = 0$ bestimmte Breitenkreis auf der Kugel mit Radius eins um den Nullpunkt auf Höhe $z = \frac{1}{2}$. Allgemein definiert man:

Definition: k sei ein Körper. Eine algebraische Varietät X in k^n ist die Nullstellenmenge einer Menge $f_1, \ldots, f_m$ von Polynomen aus $k[X_1, \ldots, X_n]$:

$$X = V(f_1, \ldots, f_m)$$
$$= \left\{(x_1, \ldots, x_n) \in k^n \;\middle|\; f_j(x_1, \ldots, x_n) = 0 \text{ für } j = 1, \ldots, m\right\}.$$

Traditionellerweise beschäftigte sich die algebraische Geometrie vor allem mit algebraisch abgeschlossenen Körpern k, da hier jedes nichtkonstante Polynom Nullstellen hat und sich dadurch eine einfachere Theorie entwickeln läßt. Selbst da reichen allerdings algebraische Varietäten nicht aus: Projiziert man etwa die Hyperbel $xy = 1$ auf die x-Achse, erhält man die x-Achse ohne den Nullpunkt, was keine algebraische Varietät ist, sondern durch die *Ungleichung* $x \neq 0$ definiert wird. Für algebraisch abgeschlossene Körper k sagt ein Satz von CLAUDE CHEVALLEY, daß das Bild jeder algebraischen Varietät unter einer Projektion oder allgemeiner unter irgendeiner durch Polynome gegebenen Abbildung stets durch Polynomgleichungen und -ungleichungen definiert werden kann.

Über dem Körper $k = \mathbb{R}$ ist dies nicht der Fall: Projizieren wir den Kreis $x^2 + y^2 = 1$ auf die x-Achse, erhalten wir das Intervall $[-1, 1]$

das durch die beiden Bedingungen $x \geq -1$ und $x \leq 1$ definiert ist; wir brauchen also noch zusätzlich die Größer- und die Kleiner-Relation. Das sind keine rein algebraischen Relationen mehr; für die meisten Körper lassen sie sich nicht einmal sinnvoll definieren. Wir führen daher einen neuen Begriff ein:

Definition: Eine semialgebraische Menge X in $\mathbb{R}^n$ ist gegeben durch Polynome $f_{ij} \in \mathbb{R}[X_1, \ldots, X_n]$ und Relationen

$$R_{ij} \in \{=, \neq, <, >, \leq, \geq\}$$

als $X = \bigcup_{i=1}^{r} \bigcap_{j=1}^{s_i} \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid f_{ij}(x_1, \ldots, x_n)\ R_j\ 0\}$.

Nach dieser Definition ist klar, daß endliche Durchschnitte und Vereinigungen semialgebraischer Mengen wieder semialgebraisch sind; einer der zentralen Sätze dieser Vorlesung mit vielen interessanten und teils unerwarteten Anwendungen, der Satz von TARSKI und SEIDENBERG, besagt, daß darüber hinaus auch das Bild einer semialgebraischen Menge unter einer Projektion oder sonst einer durch Polynome gegebenen Abbildung wieder semialgebraisch ist. Das ist alles andere als selbstverständlich: Ersetzt man die Polynome f_j durch analytische Funktionen, erhält man semianalytische Mengen, für die kein entsprechender Satz gilt; hier muß man den Begriff nochmals erweitern zur sogenannten subanalytischen Menge.

Für algebraische Varietäten über einem algebraisch abgeschlossenen Körper k kann man zeigen, daß zur Definition einer d-dimensionalen Varietät in k^n mindestens $n - d$ Polynomgleichungen notwendig sind. Über $k = \mathbb{R}$ dagegen gilt

Lemma: Jede algebraische Varietät $X \subseteq \mathbb{R}^n$ kann durch eine einzige Gleichung definiert werden.

Beweis: Nach Definition gibt es Polynome $f_1, \ldots, f_m$, durch deren simultanes Verschwinden X definiert ist. Das Verschwinden dieser m Polynome ist äquivalent zum Verschwinden des einen Polynoms $f = f_1^2 + \ldots + f_m^2$.

■

Kapitel 1
Die reellen Nullstellen eines Polynoms

Kehren wir zurück zum Grundthema dieser Vorlesung, der Lösung algebraischer Gleichungen. Der letzte Paragraph zeigte uns, wie man reduzible Polynome in einer Veränderlichen in irreduzible zerlegen kann; die damit verbundene Reduktion der Grade kann dazu führen, daß die Nullstellen danach durch explizite Formeln berechnet werden können. Es gibt aber über $\mathbb{Z}$, $\mathbb{Q}$ sowie über jedem endlich erzeugten Körper irreduzible Polynome beliebig hohen Grades, und da kann keine Faktorisierung weiterhelfen. Wir brauchen daher zusätzliche Methoden, um auch etwas über die Nullstellen solcher Polynome aussagen zu können.

In diesem Kapitel beschäftigen wir uns nur mit reellen Nullstellen. Erstens sind das für viele Anwendungen ohnehin die einzig interessanten, und zweitens läßt sich die Lokalisierung komplexer Nullstellen zurückführen auf die von reellen.

§1: Die Regel von Descartes

Die älteste Aussage über reelle Nullstellen eines beliebigen Polynoms geht im wesentlichen zurück auf RENÉ DESCARTES; da sie gelegentlich auch als Regel von CARDANO-DESCARTES bezeichnet wird, kannte der rund hundert Jahre vor DESCARTES lebende GIROLAMO CARDANO wahrscheinlich auch schon zumindest einige Spezialfälle.

Wie bei allen im folgenden besprochenen Sätzen wird die Nullstellenanzahl in Verbindung gebracht mit der Anzahl der Vorzeichenwechsel in einer Folge $(a_0, \ldots, a_n)$ reeller Zahlen. Um diese zu definieren, streichen wir alle Nullen aus der Folge und zählen dann, wie oft eine der

verbleibenden Zahlen ein anderes Vorzeichen hat als ihr Nachfolger. Die Folge $(1, 0, 0, 1, 0, -1, 0, -2, 0, 0, 3)$ hat also *zwei* Vorzeichenwechsel: von 1 auf -1 und von -2 auf 3.

Bei der Regel von DESCARTES betrachten wir für ein reelles Polynom $f = a_n X^n + \cdots + a_1 X + a_0$ die Anzahl $v(f)$ der Vorzeichenwechsel in der Folge $(a_0, \ldots, a_n)$ seiner Koeffizienten.

Regel von Descartes: *a)* Die Anzahl m der mit Vielfachheiten gezählten positiven Nullstellen eines nicht identisch verschwindenden reellen Polynoms $f = a_n X^n + \cdots + a_1 X + a_0$ ist höchstens gleich $v(f)$.
b) $m \equiv v(f) \bmod 2$.

Beweis: Wir können zunächst o.B.d.A. annehmen, daß der konstante Term a_0 nicht verschwindet, denn andernfalls können wir das Polynom durch eine X-Potenz dividieren, ohne daß sich an den positiven Nullstellen und an $v(f)$ etwas ändert. Auch Multiplikation mit -1 läßt beides unverändert; wir können uns daher auf den Fall $a_0 > 0$ beschränken.

Wir führen den Beweis durch vollständige Induktion nach dem Grad n von f. Für $n = 0$ gibt es weder Nullstellen noch Vorzeichenwechsel, so daß die Behauptung trivialerweise erfüllt ist.

Für $n > 0$ vergleichen wir f mit seiner Ableitung

$$f' = n a_n X^{n-1} + \cdots + 2 a_2 X + a_1 \,.$$

Um sicherzustellen, daß auch hier der konstante Term nicht verschwindet, nehmen wir den kleinsten Index $q \geq 1$ mit $a_q \neq 0$ und setzen $f_1 = f'/X^{q-1}$. Dann hat f_1 dieselben positiven Nullstellen wie f' und auch dieselbe Anzahl von Vorzeichenwechsel. Nach Induktionsannahme wissen wir, daß die Anzahl m' der positiver Nullstellen von f_1 (und damit f') höchstens gleich $v(f_1) = v(f')$ ist und $m' \equiv v(f') \bmod d$.

Die positiven Nullstellen von f seien $0 < x_1 < \cdots < x_r$, wobei x_i die Vielfachheit e_i habe. Die Anzahl m der positiver Nullstellen von f ist also, mit Vielfachheiten gerechnet, die Summe der e_i.

Eine e_i-fache Nullstelle von f ist eine $(e_i - 1)$-fache Nullstelle von f', außerdem liegt zwischen je zwei Nullstellen von f mindestens eine Null-

stelle der Ableitung f' und damit auch von f_1. Somit hat f_1 mindestens

$$r - 1 + \sum_{i=1}^{r}(e_i - 1) = \sum_{i=1}^{r} e_i - 1 = m - 1$$

positive Nullstellen im abgeschlossenen Intervall $[x_1, x_r]$.

Wir können noch mehr sagen: Unmittelbar rechts von einer Nullstelle x_i geht der Graph von f entweder nach oben und kommt dann auch von oben auf die nächste Nullstelle x_{i+1} zu, oder aber er geht nach unten und von dort aus zu x_{i+1}. In jedem Fall hat aber f' unmittelbar rechts von x_i ein anderes Vorzeichen als unmittelbar links von x_{i+1}. Daher muß die Anzahl der Nullstellen von f' im offenen Intervall (x_1, x_{i+1}) mit Vielfachheiten gezählt ungerade sein. Die Anzahl der Nullstellen von f' im abgeschlossenen Intervall $[x_1, x_r]$ ist daher modulo zwei kongruent zur gerade berechneten Mindestanzahl $m - 1$.

Bleiben noch die Nullstellen von f' rechts von x_r und links von x_1. Falls f' unmittelbar rechts von x_r positiv ist, geht f für $x \to \infty$ gegen $+\infty$; daher ist auch f' für $x \to \infty$ positiv. Somit hat f', wenn überhaupt, eine gerade Anzahl von Nullstellen rechts von x_r.

Für die Nullstellen zwischen 0 und x_1 können wir ähnlich argumentieren: $f(0) = a_0$ ist positiv; der Graph von f kommt also von oben nach x_1, d.h. f' ist unmittelbar links von x_1 negativ. An der Stelle $x = 0$ ist $f'(0) = a_q$, wir haben also für positives a_q eine ungerade Anzahl von Nullstellen im offenen Intervall $(0, x_1)$ und für negatives a_q eine gerade. Somit ist

$$v(f_1) \geq m' \geq \left\{ \begin{array}{ll} m-1 & \text{falls } a_q < 0 \\ m & \text{falls } a_q > 0 \end{array} \right\}$$

und

$$m' \equiv \begin{cases} m - 1 \bmod 2 & \text{falls } a_q < 0 \\ m \bmod 2 & \text{falls } a_q > 0 \end{cases}.$$

Die Koeffizientenfolge $(na_n, (n-1)a_{n-1}, \ldots, qa_q)$ hat dieselben Vorzeichen wie $(a_n, \ldots, a_q)$. Falls a_q positiv ist, ist die Anzahl der Vorzeichenwechsel dort gleich $v(f)$, denn zwischen a_q und a_0 gibt es dann keinen Vorzeichenwechsel mehr. Bei negativem a_q gibt es einen, dann

ist also $v(f_1) = v(f) - 1$. Vergleichen wir dies mit den obigen Formeln für m', folgt die Behauptung.

■

Der Mathematiker und Philosoph RENÉ DESCARTES wurde 1596 im französischen La Haye en Touraine geboren. 1802 wurde der Ort umbenannt in La Haye Descartes, seit 1967 heißt er einfach Descartes. Von 1604 bis 1612 war RENÉ DESCARTES Schüler am Jesuitenkolleg in Anjou, später studierte er Jura an der Universität von Poitiers. Nach seinem Abschluß im Jahr 1616 ging er an die Militärschule von Breda, wo er unter anderem Mathematik und Naturwissenschaften studierte. Nach zweijähriger Reise durch Europa schloß er sich 1619 der Bayrischen Armee an. Weitere Reisen quer durch Europa folgten, bis er sich 1628 in Holland niederließ. Er schrieb dort ein physikalisches Werk unter dem Titel *Le Monde, ou Traité de la Lumière,* das er aber, nachdem er von GALILEIS Verurteilung hörte, nicht veröffentlichte. Erst 1637 erschien es als philosophisches Werk unter dem Titel *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences* mit drei Anhängen über Optik, Meteore und Geometrie. Im letzteren führte er algebraischen Methoden ein, unter anderem die kartesischen Koordinaten.

Nach der Regel von DESCARTES hat beispielsweise das Polynom

$$f(X) = \frac{X^n - 1}{X - 1} = X^{n-1} + X^{n-2} + \cdots + X + 1$$

keine positive reelle Nullstelle, denn alle seine Koeffizienten sind positiv, so daß es keine Vorzeichenwechsel gibt. (Natürlich ist hier auch so klar, daß es keine positive Nullstelle geben kann.) Die negativen Nullstellen dieses Polynoms entsprechen den positiven Nullstellen von

$$f(-X) = (-1)^{n-1}X^{n-1} + (-1)^{n-2}X^{n-2} + \cdots - X + 1\,.$$

Hier wechselt die Koeffizientenfolge ständig zwischen 1 und -1; es gibt daher $n-1$ Vorzeichenwechsel und somit höchstens $n-1$ negative Nullstellen von f. Deren Anzahl ist gerade für ungerade n und ungerade für gerade n; insbesondere muß es also für gerade n mindestens eine negative Nullstelle von f geben. In der Tat ist dann -1 eine Nullstelle, und da die Nullstellen von f allesamt n-te Einheitswurzeln sind, ist es die einzige.

Durch einen Trick von JACOBI kann man mit der Regel von DESCARTES auch etwas über die Anzahl der Nullstellen in einem vorgegebenen Intervall $(a,\, b)$ aussagen kann: Wie man sich leicht überlegt (siehe aktuelles Übungsblatt), bildet

$$\varphi\colon \mathbb{R} \smallsetminus \{-1\} \to \mathbb{R}\,;\ x \mapsto \frac{a + bx}{1 + x}$$

die positiven reellen Zahlen bijektiv ab nach $(a,\, b)$, das Intervall $(-1,\, 0)$ auf die Zahlen kleiner a und die Zahlen kleiner -1 auf die Zahlen größer b. Betrachten wir daher zu einem vorgegebenen Polynom f aus $\mathbb{R}[X]$ vom Grad n das neue Polynom $g(X) = (1 + X)^n f\big(\varphi(X)\big)$, so entsprechen dessen positive Nullstellen genau den Nullstellen von f aus $(a,\, b)$. Die Berechnung von g ist freilich etwas mühsam und muß für jedes neue Intervall von neuem durchgeführt werden.

§2: Der Satz von Budan-Fourier

Dieses Problem vermeidet ein in diesem Paragraphen vorgestellte Resultat, das BUDAN 1807 und FOURIER 1820 unabhängig voneinander veröffentlichten und das FOURIER anscheinend schon ab 1796 in seinen Vorlesungen an der *Ecole Polytechnique* lehrte.

JEAN BAPTISTE JOSEPH FOURIER (1768–1830) begann zunächst eine Ausbildung zum Priester, beendete diese jedoch nicht, sondern wurde stattdessen Mathematiklehrer. 1793 trat er dem lokalen Revolutionskomittee bei, 1798 begleitete er Napoleon auf dessen Ägyptenfeldzug. Nach dem Rückzug aus Ägypten ernannte ihn dieser zum Präfekten von Isère; dort in Grenoble begann er mit seinen Arbeiten über Wärmeleitung, aus denen die FOURIER-Reihen hervorgingen. Nach Napoleons endgültiger Vertreibung wurde FOURIER 1817 in die Akademie der Wissenschaften gewählt; 1822 wurde er Sekretär der mathematischen Sektion.

FERDINAND FRANÇOIS DÉSIRÉ BUDAN DE BOISLAURENT wurde 1761 auf Haiti geboren. Im Alter von acht Jahren wurde in eine Klosterschule in der Nähe von Paris geschickt, wo er acht Jahre lang vor allem klassische Sprachen lernte. Mathematik und Naturwissenschaften waren kein Teil des Lehrplans; wegen seines großen Interesses erhielt er aber zweimal wöchentlich Zusatzstunden bei dem Mathematiker J.-C. FARCOT. Danach studierte er

vor allem Rhetorik und Philosophie; ab 1803 arbeitete er als Schulrat. Im gleichen Jahr reichte er auch seine Arbeit mit dem hier vorgestellten Resultat ein, sie wurde aber erst 1807 veröffentlicht und, da BUDAN eher als Amateurmathematiker galt, wenig beachtet. 1811 präsentierte er der Akademie der Wissenschaften in Paris einen Beweis, der 1822 nach Begutachtung (und Korrektur kleinerer Lücken) durch LAGRANGE veröffentlicht wurde. Er starb 1840 in Paris.

BUDAN und FOURIER suchten nach Aussagen über die Anzahl der Nullstellen eines reellen Polynoms f in einem Intervall $[a,\, b]$. Sie betrachten dazu für ein $x \in \mathbb{R}$ die Anzahl $S_f(x)$ der Vorzeichenwechsel in der Folge $\big(f(x), f'(x), f''(x), \ldots, f^{(n)}(x)\big)$, wobei n den Grad von f bezeichnet, und zeigen den

Satz von Budan-Fourier: Falls weder f noch eine seiner Ableitungen in den Punkten $a < b$ verschwindet, ist die Anzahl reeller Nullstellen von f in $[a,\, b]$ mit Vielfachheiten gezählt höchstens gleich $S_f(a) - S_f(b)$, und sie ist modulo zwei kongruent zu dieser Differenz.

Beweis: Wir betrachten die Menge aller Zahlen $x_i \in (a,\, b)$, für die f oder eine seiner Ableitungen verschwindet, und ordnen sie der Größe nach an:

$$a < x_1 < x_2 < \cdots < x_r < b\,.$$

Der Einfachheit halber setzen wir $x_0 = a$ und $x_{r+1} = b$, obwohl das natürlich nach Voraussetzung keine Nullstellen sind.

In jedem der offenen Intervalle $(x_i,\, x_{i+1})$ ist $S_f(x)$ konstant, denn solange weder f noch eine seiner Ableitungen verschwindet, ändert sich nichts an den Vorzeichen und damit der Anzahl der Vorzeichenwechsel.

Wir wählen für $i = 1, \ldots, r+1$ je eine reelle Zahl c_i aus dem Intervall (x_{i-1}, x_i). Dann ist x_i der einzige Wert im Intervall $[c_i,\, c_{i+1}]$, an dem f oder eine seiner Ableitungen verschwinden kann. Insbesondere ist die Anzahl der Nullstellen von f in $[a, b]$ gleich der Summe der Nullstellenanzahlen in den Intervallen $[c_i,\, c_{i+1}]$, und

$$S_f(a) - S_f(b) = S_f(c_1) - S_f(c_{r+1}) = \sum_{i=0}^{r} \big(S_f(c_i) - S_f(c_{i+1})\big)\,.$$

Daher genügt es, den Satz für die Intervalle $[c_i,\, c_{i+1}]$ zu beweisen; wir können also o.B.d.A. annehmen, daß es im Intervall $[a,\, b]$ genau einen

Punkt x gibt, so daß sowohl f als auch seine Ableitungen höchstens dort verschwinden. m sei die Nullstellenordnung von f in x, d.h. $m = 0$, falls x keine Nullstelle von f, sondern nur von einer der Ableitungen ist, und $m \geq 1$ sonst. Mit dieser Bezeichnung ist der Satz äquivalent zu den beiden Aussagen

$$m \leq S_f(a) - S_f(b) \quad \text{und} \quad m \equiv S_f(a) - S_f(b) \bmod 2\,.$$

Wir beweisen ihn durch Induktion nach dem Grad von f.

Im Falle einer linearen Funktion $f(x) = px + q$ können wir o.B.d.A. annehmen, daß $p > 0$ ist; andernfalls ersetzen wir einfach f durch $-f$. Die einzige Nullstelle $x = -q/p$ liegt genau dann im Intervall (a, b), wenn wenn $f(a)$ negativ und $f(b)$ positiv ist. Die Ableitung $f'(x) = p$ ist überall positiv, also ist $S_f(u) = 1$ genau dann, wenn $f(u)$ negativ ist, und $S_f(u) = 0$ sonst. Damit ist der Induktionsanfang erledigt.

Nun sei f ein Polynom vom Grad mindestens zwei. Wir unterscheiden drei Fälle:

1. Fall: $m \geq 1$. Dann ist $f(x) = 0$ und f' hat in x eine $m-1$-fache Nullstelle. Nach Induktionsannahme ist

$$S_{f'}(a) - S_{f'}(b) \geq m - 1 \quad \text{und} \quad m - 1 \equiv S_{f'}(a) - S_{f'}(b) \bmod 2\,.$$

Wenn $f(a)$ positiv ist, muß $f'(a)$ negativ sein, da der Graph nach *unten* zur Nullstelle geht; entsprechend ist $f'(a)$ positiv für negatives $f(a)$. Somit ist $S_f(a) = S_{f'}(a) + 1$. Am anderen Intervallende ist dagegen $S_f(b) = S_{f'}(b)$, denn ist $f(b)$ positiv, so muß der Graph von f steigen, und ist $f(b)$ negativ, so muß er fallen. Damit ist

$$S_f(a) - S_f(b) \geq m \quad \text{und} \quad m \equiv S_f(a) - S_f(b) \bmod 2\,,$$

was die Behauptung auch für f beweist.

2. Fall: Weder f noch f' haben in $[a, b]$ eine Nullstelle. Dann hat sowohl f als auch f' überall im Intervall dasselbe Vorzeichen, also ist entweder $S_f(a) = S_{f'}(a)$ und $S_f(b) = S_{f'}(b)$ oder $S_f(a) = S_{f'}(a) + 1$ und $S_f(b) = S_{f'}(b) + 1$. In beiden Fällen ist

$$S_f(a) - S_f(b) = S_{f'}(a) - S_{f'}(b)$$

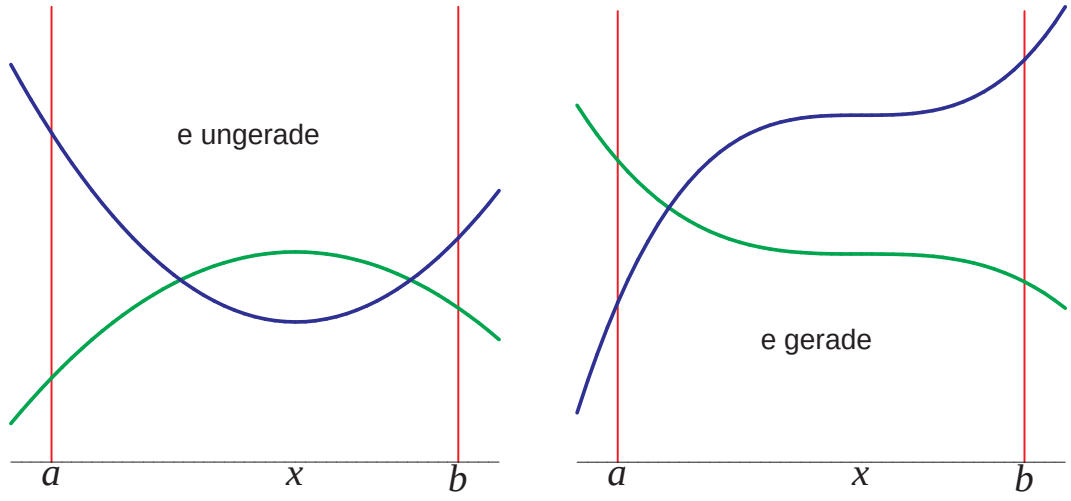
nach Induktionsvoraussetzung größer oder gleich null und gerade, wie es in diesem Fall auch sein muß.

3. Fall: $f'(x) = 0$; die Vielfachheit dieser Nullstelle sei e. Nach der TAYLORschen Formel ist in der Umgebung von x

$$f'(x+h) \approx \frac{f^{(e+1)}(x)}{e!}h^e \quad \text{und} \quad f(x+h) \approx f(x) + \frac{f^{(e+1)}(x)}{(e+1)!}h^{e+1}\,.$$

Da f' außer in x nirgends sein Vorzeichen wechseln kann, können wir daraus die Vorzeichen von $f'(a)$ und $f'(b)$ bestimmen: Für gerades e sind beide gleich dem Vorzeichen von $f^{(e+1)}(x)$, für ungerades e haben $f^{(e+1)}(x)$ und $f'(b)$ dasselbe Vorzeichen und $f'(a)$ das andere.

Die beiden folgenden Diagramme zeigen das Verhalten von f zwischen a und b, wobei die dicke blaue Kurve jeweils einem positiven Wert von $f^{(e+1)}(x)$ entspricht und die dünnere grüne einem negativen.



Über die Vorzeichen von $f(a)$ und $f(b)$ wissen wir nur, daß sie gleich sind, da f in $[a,\ b]$ keine Nullstelle hat. Um $S_f(a)$ und $S_f(b)$ auf $S_{f'}(a)$ und $S_{f'}(b)$ zurückzuführen, müssen wir daher die verschiedenen Kombinationen aus Vorzeichen von $f(a)$ und $f^{(e+1)}(x)$ betrachten.

Um die Diskussion kurz und übersichtlich zu machen, benutzen wir die Vorzeichen- oder Signum-Funktion

$$\operatorname{sgn} x = \begin{cases} +1 & \text{falls } x > 0 \\ 0 & \text{falls } x = 0 \\ -1 & \text{falls } x < 0 \end{cases}\,.$$

Wie wir uns gerade überlegt haben, ist

$$\operatorname{sgn} f'(a) = (-1)^e \operatorname{sgn} f^{(e+1)}(x) \quad \text{und} \quad \operatorname{sgn} f'(b) = \operatorname{sgn} f^{(e+1)}(x)\,;$$

außerdem ist $\operatorname{sgn} f(a) = \operatorname{sgn} f(b)$.

Im Falle $\operatorname{sgn} f(a) = \operatorname{sgn} f^{(e+1)}(x)$ ist daher

$$\operatorname{sgn} f(a) = (-1)^e \operatorname{sgn} f'(a) \quad \text{und} \quad \operatorname{sgn} f(b) = \operatorname{sgn} f'(b)\,,$$

also $S_f(b) = S_{f'}(b)$ und

$$S_f(a) = \begin{cases} S_{f'}(a) & \text{falls } e \text{ gerade} \\ S_{f'}(a) + 1 & \text{falls } e \text{ ungerade} \end{cases}.$$

Die Differenz $S_f(a) - S_f(b)$ ist somit für gerade e gleich der zwischen $S_{f'}(a)$ und $S_{f'}(b)$, für ungerade e ist sie um eins größer.

Falls $f(a)$ und $f^{(e+1)}(x)$ verschiedene Vorzeichen haben, ist

$$\operatorname{sgn} f(a) = (-1)^{e+1} \operatorname{sgn} f'(a) \quad \text{und} \quad \operatorname{sgn} f(b) = -\operatorname{sgn} f'(b)\,.$$

Hier ist also $S_f(b) = S_{f'}(b) + 1$ und

$$S_f(a) = \begin{cases} S_{f'}(a) & \text{falls } e \text{ ungerade} \\ S_{f'}(a) + 1 & \text{falls } e \text{ gerade} \end{cases}.$$

Die Differenz $S_f(a) - S_f(b)$ bleibt also wieder für gerade e unverändert, für ungerade aber wird sie nun um eins kleiner.

Nach Induktionsvoraussetzung ist

$$S_{f'}(a) - S_{f'}(b) \geq e \quad \text{und} \quad S_{f'}(a) - S_{f'}(b) \equiv e \bmod 2\,.$$

Für gerade e ist daher

$$S_f(a) - S_f(b) \geq e \geq 0 \quad \text{und} \quad S_f(a) - S_f(b) \equiv e \equiv 0 \bmod 2$$

und für ungerade e erhalten wir

$$S_f(a) - S_f(b) \geq e - 1 \geq 0 \quad \text{und} \quad S_f(a) - S_f(b) \equiv e - 1 \equiv 0 \bmod 2\,,$$

wie behauptet.

Damit ist der Satz von BUDAN-FOURIER bewiesen. ∎

§3: Der Satz von Sturm

Die Regel von DESCARTES und auch der Satz von BUDAN-FOURIER geben nur Obergrenzen und Kongruenzbedingungen modulo zwei für Nullstellenanzahlen; der erste, der eine Formel für die genaue Anzahl angeben konnte, war 1835 CHARLES-FRANÇOIS STURM. Im Unterschied zu DESCARTES, BUDAN und FOURIER geht es beim Satz von STURM allerdings um die Anzahl *verschiedener* Nullstellen; Vielfachheiten spielen hier keine Rolle.

JACQUES CHARLES-FRANÇOIS STURM wurde 1803 in Genf als Sohn eines Mathematiklehrers geboren. Ab 1821 studierte er an der dortigen Akademie Mathematik. 1823, nach Ende seines Studiums, wurde er Tutor des Sohns von Mme DE STAËL, was ihm genügend Zeit für mathematische Arbeiten ließ. Als die Familie für sechs Monate nach Paris zog, traf er dort im Haus von ARAGO unter anderem LAPLACE, POISSON, FOURIER, GAY-LUSSAC und AMPÈRE. 1825 kehrte er nach Paris zurück, wo er zwar als Tutor für ARAGOs Sohn arbeitete, vor allem aber Vorlesungen besuchte. Zeitweise arbeitete er auch als Assistent von FOURIER. Nach der Revolution von 1830 wurde es auch für einen Protestanten möglich, eine Professor in Frankreich zu bekommen; so kam er 1830 ans *Collège Rollin* und 1838 an die *Ecole normale supérieure;* 1833 wurde er französischer Staatsbürger. In seinen späten Arbeiten beschäftige er sich vor allem, zusammen mit LIOUVILLE, mit Differentialgleichungen. Er starb 1855 in Paris.

STURMs Ansatz beruht auf einer leichten Modifikation des EUKLIDischen Algorithmus: Die Folge der Divisionsreste bei der Berechnung des ggT zweier Polynome f und g nach EUKLID können wir beschreiben durch die Rekursionsvorschrift

$$r_0 = f, \quad r_1 = g, \quad r_{i+2} = \text{Rest bei der Division von } r_i \text{ durch } r_{i+1},$$

wobei abgebrochen wird, sobald ein r_j verschwindet. Für $i \geq 2$ ist also $r_i = q_i r_{i+1} + r_{i+2}$. STURM nimmt stattdessen als jeweils nächstes Â¤r_i den *negativen* Divisionsrest:

Definition: Die STURMsche Kette zum Polynom $f \in \mathbb{R}[X]$ wird berechnet nach der Rekursionsvorschrift

$$f_0 = f, \quad f_1 = f', \quad f_{i+2} = -\text{Rest bei der Division von } f_i \text{ durch } f_{i+1},$$

wobei abgebrochen wird, sobald ein f_j verschwindet. Für $i \geq 2$ ist also $f_i = q_i f_{i+1} - f_{i+2}$, wobei q_i der Quotient bei der Polynomdivision mit Rest von f_i durch f_{i+1} ist.

Da der Grad von f_{i+1} stets echt kleiner ist als der von f_i, bricht jede Sturmsche Kette ab. Ihr letztes Polynom f_s ist ein ggT von f und f', denn für die Teilbarkeitsargumente beim EUKLIDischen Algorithmus spielen Vorzeichen keine Rolle. Insbesondere ist also f_s konstant, falls f keine mehrfache Nullstellen hat.

Beschränken wir uns zunächst auf diesen Fall eines Polynoms mit höchstens einfachen Nullstellen. Seine STURMsche Kette $(f_0, \ldots, f_s)$ hat folgende Eigenschaften:

a) $f_0 = f$

b) f_s hat keine Nullstellen

c) Ist für ein i mit $0 < i < s$ der Punkt x eine Nullstelle von f_i, so ist $f_{i-1}(x)f_{i+1}(x) < 0$, d.h. $f_{i-1}(x)$ und $f_{i+1}(x)$ haben verschiedene Vorzeichen, denn $f_{i-1}(x) = q_{i-1}f_i(x) - f_{i-1}(x) = -f_{i-1}(x)$, falls $f_i(x)$ verschwindet.

d) Ist x_0 eine Nullstelle von $f = f_0$, so sind in einer Umgebung von x_0 die Funktionswerte von $f_0(x)f_1(x)$ links von x_0 negativ und rechts davon positiv, denn da f keine mehrfache Nullstellen hat, kann $f'(x_0)$ nicht verschwinden, und $\frac{d}{dx}f_0(x)f_1(x) = f'(x)^2 - f(x)f''(x)$ hat bei x_0 den positiven Wert $f'(x_0)^2$.

Wir wollen in Zukunft jede Folge $(f_0, \ldots, f_s)$ zu einem Polynom f mit den Eigenschaften *a)* bis *d)* als eine STURMsche *Folge* zu f bezeichnen. Für Polynome ohne mehrfache Nullstellen ist also die STURMsche Kette eine STURMsche Folge.

Als nächstes definieren wir für jede Folge $(f_0, \ldots, f_s)$ von Polynomen ihre *Variation* $v(a)$ in einem Punkt $a \in \mathbb{R}$ als Anzahl der Vorzeichenwechsel in der Folge reeller Zahlen $f_0(a), f_1(a), \ldots, f_s(a)$.

Für jede STURMsche Folge zu einem Polynom f gilt:

Satz: Die Anzahl der Nullstellen des Polynoms f mit $a < x \leq b$ ist $v(a) - v(b)$.

Beweis: Wir überlegen uns zunächst, in der Umgebung welcher Punkte sich in der Folge $\big(f_0(x), f_1(x), \ldots, f_s(x)\big)$ etwas an den Vorzeichen ändern kann.

Sind alle $f_i(x) \neq 0$, so bleiben auch in einer Umgebung von x alle Vorzeichen gleich, also ist $v(x)$ konstant in der Umgebung von x.

Ist $f(x) \neq 0$, aber (mindestens) ein $f_i(x) = 0$, so ist $i > 0$ und nach *b)* ist $i < s$. Damit gibt es Funktionen f_{i-1} und f_{i+1}; nach *c)* ist $f_{i-1}(x)f_{i+1}(x) < 0$. Somit haben $f_{i-1}(x)$ und $f_{i+1}(x)$ verschiedene Vorzeichen, sind also insbesondere ungleich null. Die Vorzeichen von $f_{i-1}(x)$ und $f_{i+1}(x)$ sind daher in einer Umgebung von x konstant und verschieden; egal welche Werte f in dieser Umgebung annimmt, gibt es also von f_{i-1} nach f_{i+1} genau einen Vorzeichenwechsel, so daß $v(x)$ auch in der Umgebung dieses Punkts konstant ist.

Bleibt noch der Fall, daß f selbst im Punkt x verschwindet. Falls $f = f_0$ bei x einen Vorzeichenwechsel von – nach + hat, muß f_1 wegen *d)* in einer Umgebung von x positiv sein; also haben f_0 und f_1 vor x verschiedene Vorzeichen, danach gleiche. Entsprechendes gilt, wenn f bei x von + nach – wechselt, denn dann muß f_1 in einer Umgebung von x negativ sein, d.h. beim Durchgang durch x wird v um eins kleiner.

Ist f sowohl links als auch rechts von x positiv, so muß f_1 wegen *d)* links negativ und rechts positiv sein; wieder geht also beim Durchgang durch x ein Vorzeichenwechsel verloren, genauso im Fall, daß f links und recht von x negativ ist.

Damit ist gezeigt, daß die Funktion v genau in den Punkten um eins kleiner wird, in denen f eine Nullstelle hat; und damit ist der Satz bewiesen. ∎

Satz von Sturm: Ist $[a, b]$ ein Intervall, an dessen Endpunkten a, b das Polynom f nicht verschwindet, und bezeichnet $v(x)$ die Variation der STURMschen Kette zu f, so hat f in $[a, b]$ genau $v(a) - v(b)$ Nullstellen.

Beweis: Falls f keine mehrfachen Nullstellen hat, ist die STURMsche Kette eine STURMsche Folge, also folgt die Behauptung aus dem gerade bewiesenen Satz.

Andernfalls sei $(f_0, \ldots, f_s)$ die STURMsche Kette von f; wegen deren Konstruktion über den EUKLIDischen Algorithmus ist dann $g = f_s$ ein

größter gemeinsamer Teiler von f und f', und alle f_i sind durch g teilbar. Somit besteht auch die Folge $(f_0/g, \ldots, f_s/g)$ nur aus Polynomen, und in jedem Punkt, in dem g keine Nullstelle hat, ist ihre Variation gleich der der STURMschen Kette. Da die Intervallenden a und b nach Voraussetzung keine Nullstellen sind, hat sie also insbesondere an den Stellen a und b dieselbe Variation wie die STURMsche Kette von f.

Die Funktion f/g hat dieselben Nullstellen wie f, aber jeweils nur einfach. Falls wir also zeigen können, daß $(f_0/g, \ldots, f_s/g)$ eine STURMsche Folge zu f/g ist, folgt der Satz auch in diesem Fall aus dem gerade bewiesenen.

Die Eigenschaften *a)* und *b)* einer STURMsche Folge sind trivial.

Für *c)* müssen wir eine Nullstelle x von f_i/g für ein i zwischen null und $s-1$ betrachten. Ist $f_{i-1} = g_i f_i - f_{i+1}$ die Gleichung aus der Definition der STURMschen Kette zu f, so ist natürlich auch $f_{i-1}/g = g_i \cdot f_i/g - f_{i+1}/g$, also haben f_{i-1}/g und f_{i+1}/g in jeder Nullstelle von f_i/g verschiedene Vorzeichen. (Sie können nicht null sein, denn wenn zwei aufeinanderfolgende f_i/g in einem Punkt verschwinden, müßte $f_s/g = 1$ dort wegen der gerade gezeigten Rekursionsbeziehung ebenfalls verschwinden.)

Bleibt noch *d):* Wir betrachten eine k-fache Nullstelle x_0 von f und schreiben $f(x) = (x-x_0)^k h(x)$. Dann ist

$$f'(x) = (x-x_0)^k h'(x) + k(x-x_0)^{k-1} h(x) \quad \text{und}$$
$$g(x) = (x-x_0)^{k-1} q(x) \quad \text{mit} \quad q(x_0) \neq 0\,,$$

also
$$\frac{f_1(x)}{g(x)} = (x-x_0)\frac{h'(x)}{q(x)} + k\frac{h(x)}{q(x)}\,.$$

Somit ist

$$\frac{f_0(x)}{g(x)} \cdot \frac{f_1(x)}{g(x)} = (x-x_0)^2 \frac{h(x)h'(x)}{q(x)^2} + k(x-x_0)\frac{h(x)^2}{q(x)^2}\,.$$

Der Koeffizient von $(x-x_0)$ ist somit ein Quadrat, also positiv; damit folgt *d)*, und der STURMsche Satz ist bewiesen.

■

Als erstes Beispiel betrachten wir das Polynom

$$f = X^4 + 3X^3 + 2X^2 + X + \tfrac{1}{2}\,.$$

Seine STURMsche Kette ist

$$\left(f,\ 4X^3 + 9X^2 + 4X + 1,\ \tfrac{11}{16}X^2 - \tfrac{5}{16},\ -\tfrac{64}{11}X - \tfrac{56}{11},\ -\tfrac{219}{1024}\right).$$

Für eine Zahl x mit hinreichend großem Betrag wird das Vorzeichen des Werts einer Polynomfunktion durch den höchsten Term bestimmt; wir haben also für stark negative Werte von x die Vorzeichenverteilung $(+,-,+,+,-)$ mit drei Vorzeichenwechseln; für große positive x erhalten wir $(+,+,+,-,-)$ mit nur einem Vorzeichenwechsel. Somit gibt es insgesamt zwei reelle Nullstellen.

Um deren Vorzeichen zu bestimmen, werten wir die STURMsche Kette an der Stelle null aus: $(\frac{1}{2}, 1, -\frac{5}{16}, -\frac{56}{11}, -\frac{219}{1024})$ hat einen Vorzeichenwechsel, also sind alle Nullstellen negativ. Wenn wir $x = -1$ in die STURMsche Kette einsetzen, erhalten wir die Folge $(-\frac{1}{2}, 2, \frac{3}{8}, \frac{8}{11}, -\frac{219}{1024})$ mit zwei Vorzeichenwechsel; somit gibt es eine Nullstelle z_1 mit $-1 < z_1 \le 0$ und eine Nullstelle $z_2 \le -1$.

Für $x = -2$ erhalten wir die Folge $(-\frac{3}{2}, -3, \frac{39}{16}, \frac{72}{11}, -\frac{219}{1024})$ mit ebenfalls zwei Vorzeichenwechseln, so daß $z_2 \le -2$ sein muß. und für $x = -3$ haben wir in $(\frac{31}{2}, -38, \frac{47}{8}, \frac{136}{11}, -\frac{219}{1024})$ drei Vorzeichenwechsel, also ist $-3 < z_2 \le -2$. Damit kennen wir immerhin schon die ganzzahligen Anteile der beiden Nullstellen.

Als nächstes „Beispiel“ wollen wir untersuchen, wie viele reelle Nullstellen das quadratische Polynom $f = aX^2 + bX + c$ mit $a \neq 0$ hat. Seine Ableitung ist $f_1 = 2aX + b$, und

$$(aX^2 + bX + c) : (2aX + b) = \frac{X}{2} + \frac{b}{4a} \text{ Rest } \frac{b^2 - 4ac}{4a}\,.$$

Also ist f_2 die Konstante $\Delta/4a$ mit $\Delta = b^2 - 4ac$, und die STURMsche Kette von f ist

$$\left(aX^2 + bX + c,\ \ 2aX + b, \frac{\Delta}{4a}\right).$$

Ist $a > 0$, so haben wir für große x die Vorzeichenfolge $\big(+,+,\mathrm{sgn}(\Delta)\big)$, für sehr negative x erhalten wir $\big(+,-,\mathrm{sgn}(\Delta)\big)$.

Für $\Delta > 0$ haben wir daher für $x \to \infty$ die Variation $v(x) = 0$, für $x \to -\infty$ dagegen $v(x) = 2$. Somit gibt es zwei Nullstellen. Für $\Delta = 0$ folgt entsprechend, daß es nur eine gibt. Für $\Delta < 0$ haben wir die beiden Vorzeichenverteilungen $(+,+,-)$ und $(+,-,-)$, die beide Variation eins haben, also gibt es für $\Delta < 0$ keine reelle Nullstelle. Für $a < 0$ drehen sich alle Vorzeichen um, an den Variationen und somit am Ergebnis ändert sich nichts. Beruhigenderweise stimmen alle diese Ergebnisse überein mit dem, was wir auch direkt aus der Lösungsformel für quadratische Gleichungen ablesen können.

§4: Kubische Gleichungen

Bevor wir den Satz von STURM auch kubische Gleichungen anwenden, wollen wir uns zunächst überlegen, wie man solche Gleichungen explizit lösen kann.

Auch für die Gleichung $X^3 + aX^2 + bX + c = 0$ können wir einen ähnlichen Ansatz machen wie den der quadratischen Ergänzung: Schreiben wir $X = Y + \frac{1}{3}a$, erhalten wir die neue Gleichung

$$Y^3 + \left(b - \frac{a^2}{3}\right) Y + \frac{2a^3}{27} - \frac{ab}{3} + c = 0\,;$$

es reicht also, wenn wir Gleichungen der Form

$$Y^3 + pY + q = 0$$

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

Der Trick, der schließlich zum Erfolg führte, ist folgender: Wir schreiben die Variable Y als Summe zweier neuer Variablen U und V und machen dadurch das Problem auf den ersten Blick nur schwieriger. Andererseits ist diese Summendarstellung natürlich alles andere als eindeutig; wir können daher hoffen, daß es auch dann noch Lösungen gibt, wenn wir

an U und V zusätzliche Forderungen stellen und dadurch das Problem vielleicht vereinfachen.

Einsetzen von $Y = U + V$ führt auf die Bedingung

$$(U+V)^3 + p(U+V) + q = U^3 + 3U^2V + 3UV^2 + V^3 + p(U+V) + q = 0\,.$$

Dies können wir auch anders zusammenfassen als

$$(U^3 + V^3 + q) + (3UV + p)(U+V) = 0\,,$$

und natürlich verschwindet diese Summe insbesondere dann, wenn beide Summanden einzeln verschwinden. Falls es uns also gelingt, eine Lösung des Gleichungssystems

$$U^3 + V^3 = -q \quad \text{und} \quad 3UV = -p\,,$$

haben wir eine Lösung der kubischen Gleichung gefunden.

Jede Lösung (u, v) des obigen Gleichungssystems erfüllt erst recht die beiden Gleichungen

$$U^3 + V^3 = -q \quad \text{und} \quad U^3 \cdot V^3 = -\frac{p^3}{27}\,,$$

wir kennen also die Summe und das Produkt von u^3 und v^3 Damit kennen wir aber auch u^3 und v^3:

Haben zwei Zahlen h, k das Produkt r und die Summe s, so sind h und k die beiden Nullstellen der Gleichung

$$(Z-h)(Z-k) = Z^2 - (h+k)Z + hk = Z^2 - sZ + r = 0\,;$$

falls wir r und s kennen, erhalten wir h und k also einfach als Lösungen einer quadratischen Gleichung:

$$h = \frac{s}{2} + \sqrt{\frac{s^2}{4} - r} \quad \text{und} \quad k = \frac{s}{2} - \sqrt{\frac{s^2}{4} - r}$$

oder umgekehrt.

In unserem Fall ist daher

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

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

wobei es auf die Reihenfolge natürlich nicht ankommt.

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

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

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

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

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

Betrachten wir dazu als einfaches Beispiel die Gleichung

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

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

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

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

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

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

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

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

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

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

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

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

Die beiden anderen Kubikwurzeln erhalten wir, indem wir die reelle

Kubikwurzel mit einer der beiden komplexen dritten Einheitswurzeln multiplizieren, d.h. also mit

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

Im ersten Fall ist

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

und

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

wir erhalten somit die Lösung $y = u + v = -1$.

Die dritte Kubikwurzel

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

schließlich führt auf

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

und liefert so die Lösung $y = u + v = 1$.

Etwas komplizierter wird es bei der Gleichung

$$X^3 - 7X + 6 = 0\,.$$

Da sie keinen x^2-Term hat, können wir gleich $p = -7$ und $q = 6$ in die Formel einsetzen und erhalten

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

Was nun? Wenn wir einen Ansatz der Form $u = r + is$ machen, kommen wir auf ein System von zwei kubischen Gleichungen in zwei Unbekannten, also ein schwierigeres Problem als unsere Ausgangsgleichung. Wir können auch Maple nach dem Wert dieser Kubikwurzel fragen: Die imaginäre Einheit i wird dort als großes I eingegeben und Wurzeln

(abgesehen von der auch als `sqrt` darstellbaren Quadratwurzel) durch Potenzen mit gebrochenem Exponenten; wir tippen also

```
> (-3 + 10/9*sqrt(3)*I)^(1/3);
```

$$\left(-3+\frac{10}{9}I\sqrt{3}\right)^{\left(\frac{1}{3}\right)}$$

und erhalten unsere Eingabe unausgerechnet zurück. Mit dem Befehl `evalc` können wir Maple veranlassen, das Ergebnis – falls möglich – in der Form $a + bi$ darzustellen:

```
> evalc(%);
```

$$\begin{aligned}&\frac{1}{9}7^{\left(\frac{1}{3}\right)}9^{\left(\frac{2}{3}\right)}21^{\left(\frac{1}{6}\right)}\left(\cos\left(-\frac{1}{3}\arctan\left(\frac{10\sqrt{3}}{27}\right)+\frac{\pi}{3}\right)\right.\\&\left.+\frac{1}{9}i7^{\left(\frac{1}{3}\right)}9^{\left(\frac{2}{3}\right)}21^{\left(\frac{1}{6}\right)}\left(\sin\left(-\frac{1}{3}\arctan\left(\frac{10\sqrt{3}}{27}\right)+\frac{\pi}{3}\right)\right.\right.\end{aligned}$$

Dieses Ergebnis ist offensichtlich noch nicht in der bestmöglichen Weise dargestellt; mit dem Kommando `simplify` können wir Maple dazu überreden, sich etwas mehr Mühe zu geben:

```
> u := simplify(%);
```

$$\begin{aligned}u:=\frac{1}{3}\sqrt{7}\sqrt{3}\left(\cos\left(-\frac{1}{3}\arctan\left(\frac{10\sqrt{3}}{27}\right)+\frac{\pi}{3}\right)\right.\\\left.+\sin\left(-\frac{1}{3}\arctan\left(\frac{10\sqrt{3}}{27}\right)+\frac{\pi}{3}\right)i\right)\end{aligned}$$

Maple arbeitet hier also mit der Polarkoordinatendarstellung komplexer Zahlen: Jede komplexe Zahl z läßt sich darstellen in der Form

$$z = |z|\left(\cos\varphi + i\sin\varphi\right),$$

und

$$\sqrt[3]{z} = \sqrt[3]{|z|}\left(\cos\tfrac{\varphi}{3} + i\sin\tfrac{\varphi}{3}\right).$$

Leider gibt es keine einfache Formel, die Sinus und Kosinus von $\frac{\varphi}{3}$ durch $\cos\varphi$ und $\sin\varphi$ ausdrückt. Aus den Additionstheoremen können wir uns

natürlich leicht Formeln für $\cos 3\alpha$ und $\sin 3\alpha$ verschaffen; wenn wir das nicht selbst ausrechnen wollen, tut es Maple für uns:

```
> expand(cos(3*alpha));
```

$$4\cos(\alpha)^3 - 3\cos(\alpha)$$

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

Trotzdem ist die obige Darstellung der Lösung nicht völlig nutzlos: Sie gibt uns immerhin Formeln für den Real- und den Imaginärteil der Lösung, und diese Formeln können wir numerisch auswerten. Der Maple-Befehl dafür heißt `evalf`, wobei das `f` für *floating point* steht, d.h. also für Gleitkommaarithmetik.

```
> evalf(u);
```

$$1.000000001 + 1.154700538\,I$$

Wie jedes numerische Ergebnis stimmt diese Zahl natürlich nur näherungsweise, und zumindest in diesem Fall ist die Hypothese, daß es sich beim Realteil um eine durch Rundungsfehler verfälschte Eins handeln kann, eine Überlegung wert. Falls dem so sein sollte, ist

$$\cos\left(-\frac{1}{3}\arctan\left(\frac{10}{27}\sqrt{3}\right)+\frac{\pi}{3}\right) = \frac{3}{\sqrt{3}\sqrt{7}} = \sqrt{\frac{3}{7}}\,,$$

und daraus folgt dann über die Beziehung $\sin^2\alpha + \cos^2\alpha = 1$, daß

$$\sin\left(-\frac{1}{3}\arctan\left(\frac{10}{27}\sqrt{3}\right)+\frac{\pi}{3}\right) = \pm\sqrt{1-\frac{3}{7}} = \pm\sqrt{\frac{4}{7}} = \pm\frac{2\sqrt{7}}{7}$$

und

$$u = 1 \pm \frac{1}{3}\sqrt{3}\sqrt{7}\cdot\frac{2\sqrt{7}}{7}i = 1 \pm \frac{2\sqrt{3}}{3}i$$

ist. Bislang war alles noch Spekulation; nun kommt die Probe:

$$\left(1 \pm \frac{2\sqrt{3}}{3}i\right)^3 = 1 \pm 2\sqrt{3}\,i - 4 \mp \frac{8}{9}\sqrt{3}\,i = -3 \pm \frac{10}{9}\sqrt{3}\,i\,,$$

also ist $u = 1 + \frac{2\sqrt{3}}{3}i$. Das zugehörige v ist

$$v = \frac{7}{3u} = \frac{7}{3 + 2\sqrt{3}\,i} = \frac{7(3 - 2\sqrt{3}\,i)}{3^2 + 2^2 \cdot 3} = 1 - \frac{2\sqrt{3}}{3}i\,.$$

Damit ist die erste Lösung $x = u + v = 2$ gefunden. Die beiden anderen sind nun (etwas langwierige) Routine, können also beruhigt Maple überlassen werden. Wir verwenden dabei den Befehl `evalc`, der eine komplexe Zahl – sofern möglich – auf die Form $a + ib$ bringt und den Befehl `conjugate`, der die konjugiert komplexe Zahl berechnet:

```
> rho := -1/2 + sqrt(3)/2*I; u := 1 + 2/3*sqrt(3)*I;
```

$$\rho := -\frac{1}{2} + \frac{1}{2}I\sqrt{3}$$

$$u := 1 + \frac{2}{3}I\sqrt{3}$$

```
> evalc(u*rho + 7/(3*u*rho));
```

$$-3$$

```
> evalc(u*conjugate(rho) + 7/(3*u*conjugate(rho)));
```

$$1$$

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

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

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

Hier setzen wir $X = Y + 1$ und erhalten die neue Gleichung

$$(Y+1)^3 - 3(Y+1)^2 + 9(Y+1) + 13$$
$$= Y^3 + 3Y^2 + 3Y + 1 - 3(Y^2 + 2Y + 1) + 9Y + 9 + 13$$
$$= Y^3 + 6Y + 20 = 0$$

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

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

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

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

Damit haben wir im Prinzip eine Lösung gefunden, die auch Maple nicht weiter vereinfachen kann:

```
> u := (-10 + 6*sqrt(3))^(1/3); simplify(u - 2/u);
```

$$u := (-10+6\sqrt{3})^{(\frac{1}{3})}$$

$$\frac{(-10+6\sqrt{3})^{(\frac{2}{3})} - 2}{(-10+6\sqrt{3})^{(\frac{1}{3})}}$$

Wenn wir das allerdings numerisch auswerten, drängt sich wieder die Hypothese auf, daß hier tatsächlich etwas sehr viel einfacheres steht:

```
> evalf(%);
```

$$-1.999999986$$

Einsetzen von $y = -2$ in unsere kubische Gleichung zeigt in der Tat, daß

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

ist. Aber warum ist

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

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

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

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

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

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

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

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

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

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

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

und

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

Damit ist

$$\frac{2}{u\rho} = \frac{4\big((1-\sqrt{3}) - (3-\sqrt{3})i\big)}{(1-\sqrt{3})^2 + (3-\sqrt{3})^2} = \frac{4\big((1-\sqrt{3}) - (3-\sqrt{3})i\big)}{16 - 8\sqrt{3}}$$
$$= \frac{\big((1-\sqrt{3}) - (3-\sqrt{3})i\big)(2+\sqrt{3})}{2(2-\sqrt{3})(2+\sqrt{3})} = \frac{-(1+\sqrt{3}) - (3+\sqrt{3})i}{2}\,,$$

also

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

Entsprechend folgt $y_3 = u\bar{\rho} - \dfrac{2}{u\bar{\rho}} = 1 - 3i$.

Die Mathematiker des fünfzehnten und sechzehnten Jahrhunderts, auf die die Lösungsformel für kubische Gleichungen zurückgeht, hatten natürlich weder Computer noch Taschenrechner; auch kannten sie weder Dezimalbrüche noch komplexe Zahlen. Trotzdem konnten sie erstaunlich gut mit der Lösungsformel umgehen. In §3.2 des Buchs

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

sind zwei Beispiele für ihre Vorgehensweise zu finden:

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

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

Der numerische Näherungswert 2,414213562 für diese (reelle) Wurzel hilft uns nicht weiter. Wenn wir aber auf gut Glück versuchen, eine Wurzel zu finden, die sich auch in der Form $a + b\sqrt{2}$ mit ganzen Zahlen a und b schreiben läßt, Dann ist

$$(a + b\sqrt{2})^3 = a^3 + 3a^2b\sqrt{2} + 6ab^2 + 2b^3\sqrt{2} = 7 + 5\sqrt{2}\,,$$

also

$$a^3 + 6ab^2 = 7 \quad \text{und} \quad 3a^2b + 2b^3 = 5\,.$$

Damit haben wir, wie schon oben erwähnt, ein System von *zwei* kubischen Gleichungen anstelle von einer, jetzt allerdings suchen wir nur nach ganzzahligen Lösungen. Aus der ersten Gleichung können wir a ausklammern und erhalten $a(a^2 + 6b^2) = 7$. Somit muß a ein Teiler von sieben sein, d.h. $a = \pm 1$ oder $a = \pm 7$. Die negativen Zahlen scheiden aus, da die Klammer nicht negativ werden kann, und auch $a = 7$ ist nicht möglich, denn dann wäre die linke Seite mindestens gleich 7^3. Wenn es eine ganzzahlige Lösung gibt, muß daher $a = 1$ sein; durch Einsetzen folgt, daß dann mit $b = \pm 1$ die erste Gleichung in der Tat erfüllt ist. Die zweite Gleichung $b(3a^2 + 2b^2) = 5$ zeigt, daß auch b positiv sein muß und $a = b = 1$ beide Gleichungen erfüllt. Somit ist

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

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

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

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

Wir interessieren und (zumindest gelegentlich) auch für komplexe Zahlen; die beiden noch fehlenden Lösungen können wir nun entweder berechnen als $u\rho + v\rho^2$ und $u\rho^2 + v\rho$, oder aber wir dividieren die Gleichung durch $x-2$ und erhalten das quadratische Polynom x^2+2x+7 mit den Nullstellen $-1 \pm \sqrt{6}\,i$.

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

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

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

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

$$a^3 - 9ab^2 = 10 \quad \text{und} \quad 3a^2b - 3b^3 = 9\,.$$

Ausklammern von a *bzw.* b und Kürzen der zweiten Gleichung durch drei führt auf

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

Wenn es ganzzahlige Lösungen gibt, muß wegen der zweiten Gleichung $b = \pm 1$ oder $b = \pm 3$ sein. $b = \pm 1$ führt auf $a^2 - 1 = \pm 3$, also $b = 1$ und $a = \pm 2$; für $b = \pm 3$ läßt sich kein ganzzahliges a finden. Einsetzen in die erste Gleichung zeigt, daß $a = -2, b = 1$ das System löst, also ist

$-2+\sqrt{-3}$ eine der drei Wurzeln. Die anderen könnten wir finden, indem wir das Problem durch Abdividieren auf eine quadratische Gleichung reduzieren; alternativ – und das war wohl die Methode des 17. Jahrhunderts – können wir die Einschränkung aufheben, daß a und b ganze Zahlen sein müssen und auch Brüche mit kleinen Nennern zulassen.

Der kleinstmögliche Nenner ist zwei; der Ansatz

$$\left(\frac{a}{2}+\frac{b}{2}\sqrt{-3}\right)^3 = 10+9\sqrt{-3}$$

führt auf die Gleichungen $a(a^2-9b^2) = 80$ und $b(a^2-b^2) = 24$, wobei mindestens eine der Zahlen a und b ungerade sein muß, da wir ansonsten nichts neues bekommen. Da rechts jeweils gerade Zahlen stehen, sieht man leicht, daß dann beide Zahlen ungerade sein müssen; damit bleiben also für a nur die Möglichkeiten $a = \pm 1$ oder ± 5 und für b entsprechend $b = \pm 1$ oder ± 3. Einsetzen zeigt, daß wir mit $a = 5, b = 1$ und $a = -1, b = -3$ Lösungen bekommen. Die drei Kubikwurzeln von $-10+9\sqrt{-3}$ sind somit

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

Zu jedem dieser drei möglichen Werte von u müssen wir jene Zahl v finden, für die $uv = \frac{21}{3} = 7$ ist; in allen drei Fällen erhält man den Summanden $v = 7/u$ dadurch, daß man einfach das Vorzeichen des Koeffizienten von $\sqrt{-3}$ ändert. Die drei mit so großem Aufwand ermittelten Lösungen der kubischen Gleichung sind also einfach die drei ganzen Zahlen

$$(-2+\sqrt{-3})+(-2-\sqrt{-3}) = -4\,,$$
$$\left(\frac{5}{2}+\frac{1}{2}\sqrt{-3}\right)+\left(\frac{5}{2}-\frac{1}{2}\sqrt{-3}\right) = 5 \quad \text{und}$$
$$\left(-\frac{1}{2}-\frac{3}{2}\sqrt{-3}\right)+\left(-\frac{1}{2}+\frac{3}{2}\sqrt{-3}\right) = -1\,.$$

Wie die Beispiele in diesem Paragraphen zeigen, haben wir beim exakten Lösen kubischer Gleichungen nach der hier betrachteten Formel oft

mit komplizierten Ausdrücken zu tun, von denen sich nachher (nach teilweise recht trickreichen Ansätzen) herausstellt, daß sie sich tatsächlich sehr viel einfacher darstellen lassen. Dies ist ein allgemeines Problem der der Computeralgebra, zu dem es leider keine allgemeine Lösung gibt: Wie wir im nächsten Kapitel sehen werden, hat D. RICHARDSON 1968 gezeigt, daß es keinen Algorithmus geben kann, der von zwei beliebigen reellen Ausdrücken entscheidet, ob sie gleich sind oder nicht. Dabei reicht es schon, wenn wir nur Ausdrücke betrachten, die aus ganzen Zahlen, den Grundrechenarten, der Sinus- und der Betragsfunktion sowie der Zahl π aufgebaut werden können. Wir werden allerdings auch sehen, daß wir für wichtige Teilmengen von $\mathbb{R}$ solche Algorithmen haben.

Dies gilt insbesondere im Falle aller in diesem Paragraphen aufgetretenen Ausdrücke; wir werden im Laufe der Vorlesung noch mehrere Strategien kennen lernen, wie wir zumindest bei den hier betrachteten Beispielen die Lösungen erheblich einfacher und schneller gefunden hätten als über die allgemeine Lösungsformel.

Nachdem wir einigermaßen mit der Lösung kubischer Gleichungen vertraut sind, wollen wir das kubische Polynom $f = X^3 + pX + q$ mit Hilfe der STURMschen Theorie untersuchen. Seine Ableitung ist $f_1 = 3X^2 + p$ und

$$(X^3 + pX + q) : (3X^2 + p) = \frac{X}{3} \text{ Rest } \frac{3p}{2}X + q\,,$$

so daß $f_2 = -\frac{3p}{2}X - q$ ist. Weiter ist

$$(3X^2 + p) : \left(-\frac{3p}{2}X - q\right) = -\frac{9X}{2p} + \frac{27q}{4p^2} \text{ Rest } \left(p^3 + \frac{27}{4}q^2\right) \Big/ p^2\,,$$

die STURMsche Kette endet also mit $f_3 = -\left(p^3 + \frac{27}{4}q^2\right) \Big/ p^2$.

Für die Anzahl reeller Lösungen ist das asymptotische Verhalten relevant: Da $f = X^3 + pX + q$ durch den führenden Term X^3 dominiert wird, ist hier das Vorzeichen unabhängig von p und q für große negative x stets negativ und für große positive x stets positiv. Entsprechend haben

wir für $f_1 = 3X^2 + p$ in beiden Fällen positive Vorzeichen. Auch bei der linearen Funktion $f_2 = -\frac{3}{2p}X - q$ ist unabhängig von q das Vorzeichen für stark negative x stets gleich dem von p, für positive dagegen gleich dem von $-p$. (Der ziemlich triviale Fall $p = 0$ sei dem Leser überlassen.) Das Vorzeichen von f_3 schließlich ist das von $-\Delta$ mit $\Delta = p^3 + \frac{27}{4}q^2$, denn $p^2 \geq 0$.

Die Vorzeichenfolge wird dann für große negative Werte von x zu $(-, +, \operatorname{sgn} p, -\operatorname{sgn}\Delta)$; für große positive zu $(+, +, -\operatorname{sgn} p, -\operatorname{sgn}\Delta)$. Für $\Delta > 0$ und haben wir also die Folgen $(-, +, \operatorname{sgn} p, -)$ und $(+, +, -\operatorname{sgn} p, -)$; da $\pm\operatorname{sgn} p$ zwischen einem + und einem $-$ steht, haben wir im ersten Quadrupel immer zwei Vorzeichenwechsel und im zweiten immer nur einen; es gibt daher für $\Delta < 0$ nur eine reelle Nullstelle (und zwei komplexe).

Im Fall $\Delta = 0$ haben wir die Folgen $(-, +, \operatorname{sgn} p, 0)$ und $(+, +, -\operatorname{sgn} p, 0)$. Da $q^2 \geq 0$ aber $\Delta = 0$ ist, muß hier entweder $p = q = 0$ sein oder $p < 0$. Im letzteren Fall hat $(-, +, \operatorname{sgn} p, 0)$ zwei Vorzeichenwechsel und $(+, +, -\operatorname{sgn} p, 0)$ keinen; es gibt also zwei reelle Nullstellen (von denen eine die Vielfachheit zwei hat). Im ersten Fall hat $(+, +, -\operatorname{sgn} p, 0)$ nur einen Vorzeichenwechsel und $(+, +, -\operatorname{sgn} p, 0)$ wieder keinen; es gibt also nur eine reelle Nullstelle. Da wir für $p = q = 0$ die Gleichung $y^3 = 0$ haben, ist das die dreifache Nullstelle $y = 0$.

Für $\Delta < 0$ schließlich ist notwendigerweise $p < 0$, denn q^2 kann nicht negativ werden. Wir bekommen daher die Folgen $(-, +, -, +)$ mit drei Vorzeichenwechseln und $(+, +, +, +)$ ohne Vorzeichenwechsel; hier gibt es also drei reelle Nullstellen.

Wenn wir mit der Lösungsformel

$$y = u - \frac{p}{3u} \quad \text{mit} \quad u = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{\Delta}{27}}}$$

vergleichen, sehen wir, warum wir oben bei den Beispielen mit drei verschiedenen reellen Nullstellen Schwierigkeiten hatten: Wie wir gerade gesehen haben, muß Δ dann negativ sein; die Quadratwurzel in der Lösungsformel liefert also einen imaginären Wert. Obwohl es drei reelle

Nullstellen gibt, müssen wir also zu deren Berechnung die Kubikwurzel einer nichtreellen komplexen Zahl finden.

§5: Isolation der reellen Nullstellen

Der Satz von STURM sagt uns für jedes Intervall $[a, b]$, wie viele Nullstellen des Polynoms f dort liegen. Wenn wir uns für die Nullstellen eines Polynoms interessieren, geht es aber eher darum, eine Liste möglichst kleiner Intervalle zu finden die jeweils genau eine Nullstelle von f enthalten. STURM hat auch gezeigt, wie das möglich ist: Sobald wir ein Intervall kennen, in dem alle reellen Nullstellen liegen, können wir durch fortgesetzte Intervallhalbierungen zu einer Liste von Intervallen kommen, die jeweils genau eine Nullstelle enthalten. Durch weitere Halbierungen können wir gegebenenfalls auch die Länge dieser Intervalle beliebig kurz machen.

Für das Ausgangsintervall brauchen wir eine Abschätzung für die Größe der reellen Nullstellen eines Polynoms

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

An den Nullstellen dieses Polynoms ändert sich nichts, wenn wir es mit einer von Null verschiedenen Konstanten multiplizieren; eine gute Schranke sollte daher nur von den Quotienten a_i/a_n abhängen. Um nicht ständig mit diesen Quotienten hantieren zu müssen, beschränken wir uns in der folgenden Diskussion auf normierte Polynome.

Es gibt eine ganze Reihe von Schranken für die Nullstellen eines Polynoms; am einfachsten und für uns völlig ausreichend ist die folgende, die CAUCHY bereits 1829 veröffentlichte:

Lemma: z sei eine reelle Nullstelle des Polynoms

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

und J sei die Menge aller Indizes i mit $a_i < 0$; die Elementanzahl von J sei m. Für $m = 0$ ist $z \leq 0$, ansonsten ist z kleiner als das Maximum aus eins und den Zahlen $\sqrt[n-k]{|m a_k|}$ für $k \in J$.

Beweis: Ist $J = \emptyset$, so kann es nach der Regel von DESCARTES keine positiven Nullstellen geben, also ist $z \leq 0$. Andernfalls sei S das Maximum aus eins und den Zahlen $\sqrt[n-k]{|ma_k|}$ mit $k \in J$. Für jedes solche k ist dann $|ma_k| \leq S^{n-k}$. Damit ist auch für jedes $x > S$

$$x^n > |ma_k|\, x^k = -ma_k x^k \,.$$

Addieren wir diese Gleichungen für alle $k \in J$, folgt, daß

$$mx^n > -m \sum_{k \in J} a_k x^k \quad \text{und} \quad x^n + \sum_{k \in J} a_k x^k > 0 \,.$$

Da $a_k x^k$ für alle $k \neq J$ größer oder gleich null ist, ist damit auch $f(x) > 0$; ein $x > S$ kann also keine Nullstelle sein. ∎

Als Beispiel betrachten wir das Polynoms $f = x^5 - 2x^4 - 3x^3 + 2x^2 - 1$. Hier ist $J = \{0, 3, 4\}$, also $m = 3$. Unter den Zahlen 6, $\sqrt{9} = 3$ und $\sqrt[5]{3}$ ist 6 die größte, also ist jede reelle Nullstelle kleiner oder gleich sechs.

Um auch eine untere Schranke für die reellen Nullstellen zu erhalten, betrachten wir das Polynom $f(-X)$ oder besser, da $f(-X)$ den höchsten Term $-X^5$ hat, das Polynom $-f(-X) = X^5 + 2X^4 - 3X^3 - 2X^2 + 1$. Hier ist $J = \{2, 3\}$, also $m = 2$, und unter den Zahlen $\sqrt{6}$ und $\sqrt[3]{4}$ ist $\sqrt{6}$ die größere, denn $\sqrt[3]{4} < \sqrt[3]{8} = 2$, aber $\sqrt{6} > \sqrt{4} = 2$. Damit wissen wir, daß alle reellen Nullstellen z von f die Ungleichung $-\sqrt{6} \leq z \leq 6$ erfüllen.

Baron AUGUSTIN LOUIS CAUCHY (1789–1857) stellte als erster durch die exakte Definition von Begriffen wie *Konvergenz* und *Stetigkeit* die Analysis auf ein sicheres Fundament. In insgesamt 789 Arbeiten beschäftigte er sich u.a. auch mit komplexer Analysis, Variationsrechnung, Differentialgleichungen, FOURIER-Analysis, Permutationsgruppen, der Diagonalisierung von Matrizen und der theoretischen Mechanik. Als überzeugter Royalist hatte er häufig Schwierigkeiten mit den damaligen Regierungen; er lebte daher mehrere Jahre im Exil in Turin und später in Prag, wo er (mit sehr mäßigem Erfolg) den französischen Thronfolger unterrichtete.

Sobald wir ein Intervall $[a, b]$ kennen, in dem alle reellen Nullstellen eines Polynoms f liegen, ist eigentlich klar, wie das STURMsche Intervallhalbierungsverfahren funktioniert: Wir suchen eine Liste $\mathcal{N}$ von Intervallen $[a_i, b_i]$ mit der Eigenschaft, daß jedes dieser Intervalle genau eine Nullstelle von f enthält. eventuell können wir auch noch fordern, daß die Länge jedes dieser Intervalle unter einer gewissen Schranke s liegt.

Wir beginnen mit einer Liste $\mathcal{L}$ bestehend aus noch zu untersuchenden Intervallen $[c, d]$ zusammen mit den Anzahlen der Vorzeichenwechsel in den STURMschen Ketten $S_f(c)$ und $S_f(d)$. Zu Beginn enthält $\mathcal{L}$ nur das Intervall $[a, b]$, in dem alle reellen Nullstellen liegen, zusammen mit den Vorzeichenwechseln von $S_f(a)$ und $S_f(b)$.

So lange die Liste $\mathcal{L}$ nicht leer ist, wählen wir eines der dort befindlichen Intervalle $[c, d]$ aus und berechnen nach STURM die Anzahl der dort befindlichen Nullstellen. Wenn es keine gibt, eliminieren wir das Intervall; falls es nur eine ist (und gegebenenfalls die Intervallänge unter der Schranke s liegt), kommt das Intervall in die Ergebnisliste $\mathcal{N}$. Andernfalls wählen wir einen Punkt $t \in (c, d)$, z.B. den Mittelpunkt $t = \frac{1}{2}(c + d)$, und berechnen die STURMsche Kette $S_f(t)$; danach wird $[c, d]$ in der Liste $\mathcal{L}$ ersetzt durch die beiden Intervalle $[c, t]$ und $[t, d]$.

Um diesen Algorithmus auf das obige Beispiel anwenden zu können, müssen wir zunächst die STURMsche Kette von f berechnen, am besten nachdem wir uns vergewissert haben, daß f irreduzibel ist:

```
> f := X^5 - 2*X^4 - 3*X^3 + 2*X^2 - 1:
```

```
> factor(f);
```

$$X^5 - 2X^4 - 3X^3 + 2X^2 - 1$$

```
> f1 := diff(f, X);
```

$$f1 := 5X^4 - 8 * X^3 - 9X^2 + 4X$$

```
> f2 := -rem(f, f1, X);
```

$$f2 := \frac{46}{25}X^3 - \frac{12}{25}X^2 + 1 - \frac{8}{25}X$$

Da es uns nur um Vorzeichen geht, multiplizieren wir mit dem (Haupt-) nenner der Koeffizienten:

```
> f2 := 25*f;
```
$$f2 := 46X^3 - 12X^2 + 25 - 8X$$

```
> f3 := -rem(f1, f2, X);
```
$$f3 := \frac{5225}{529}X^2 - \frac{125}{1058}X - \frac{1925}{529}$$

```
> f3 := 1058/25*f3;
```
$$f3 := 418X^2 - 5X - 154$$

```
> f4 := -rem(f2, f3, X);
```
$$f4 := -\frac{82524}{3971} - \frac{769695}{87362}X$$

```
> f5 := -rem(f3, f4, X);
```
$$f5 := -\frac{513601198}{235225}$$

Wir wissen, daß alle reellen Nullstellen zwischen $-\sqrt{6}$ und 6 liegen; um ganze Zahlen zu haben. starten wir mit dem Intervall $[-3, 6]$ und werten die STURMsche an dessen Endpunkten aus. Dazu müssen wir in der Liste `[f, f1, f2, f3, f4, f5]` den gerade betrachteten Punkt für x einsetzen und die Vorzeichenwechsel zählen.

Zum Einsetzen können wir natürlich den `subs`-Befehl von Maple benutzen, allerdings müssen wir ihn für die sechs Elemente der Liste sechsmal eintippen. Zum Glück kennt Maple ein Kommando, das uns dies erspart: Der Befehl `map(`G`, [`$f_1, \ldots, f_n$`])` wendet G auf jedes Element der Liste an, liefert also die Liste $[G(f_1), \ldots, G(f_n)]$. Das `subs`-Kommando können wir hier allerdings nicht für G einsetzen, denn dieses Kommando hat ja *zwei* Argumente. Dafür gibt es den Befehl `map2([`G`, `A`, [`$f_1, \ldots, f_n$`])`, der die Liste aus den Elementen $G(A, f_i)$ konstruiert. Mit

```
> map2(subs, x=-3, [f, f1, f2, f3, f4, f5]);
```

erhalten wir also die Funktionswerte an der Stelle $x = -3$:

$$[-307, 528, -1301, 3623, 493557, -1]$$

Bequemer wird es, wenn wir noch die Signum-Funktion `sign` darauf anwenden und das ganze als eine Funktion schreiben:

```
> Sturm := t -> map(sign,
>          map2(subs, x=t, [f, f1, f2, f3, f4, f5])):
```

Damit können wir nun direkt die Vorzeichenfolge an einer Stelle x berechnen:

```
> Sturm(-3);
```
$[-1, 1, -1, 1, 1, -1]$

```
> Sturm(6);
```
$[1, 1, 1, 1, -1, -1]$

Für $x = -3$ haben wir also vier Vorzeichenwechsel, für $x = 6$ nur einen. Damit hat f drei reelle Nullstellen.

Wir unterteilen das Intervall an der Stelle $x = 2$ und berechnen die STURMsche Kette:

```
> Sturm(2);
```
$[-1, -1, 1, 1, -1, -1]$

Hier gibt es zwei Vorzeichenwechsel, also haben wir zwei Nullstellen in $[-3, 2]$ und eine in $[2, 6]$. Letzteres Intervall enthält also bereits nur eine einzige Nullstelle und kommt somit, falls wir keine Ansprüche an die Intervallängen stellen, in die Ergebnisliste.

Das Intervall $[-3, 2]$ muß weiter zerlegt werden, z.B. an der Stelle $x = 0$:

```
> Sturm(0);
```
$[-1, 1, 1, -1, -1, -1]$

Wieder zwei Vorzeichenwechsel, also gibt es keine Nullstelle in $[0, 2]$, aber zwei in $[-3, 0]$. Wir zerlegen weiter an der Stelle $x = -1$:

```
> Sturm(-1);
```
$[1, 1, -1, 1, -1, -1]$

Das sind drei Vorzeichenwechsel, also haben wir eine Nullstelle in $[-3, -1]$ und eine in $[-1, 0]$. Wenn uns die Intervallängen nicht interessieren, sind wir damit fertig; wenn wir allerdings Intervalle der Länge eins wollen, müssen wir $[-3, -1]$ und $[2, 6]$ noch weiter unterteilen:

```
> Sturm(-2);
```
$[-1, 1, -1, 1, -1, -1]$

Vier Vorzeichenwechsel, die Nullstelle liegt also in $[-2, -1]$.

```
> Sturm(4);
```

$$[1, 1, 1, 1, -1, -1]$$

Ein Vorzeichenwechsel; die Nullstelle liegt in $[2, 4]$.

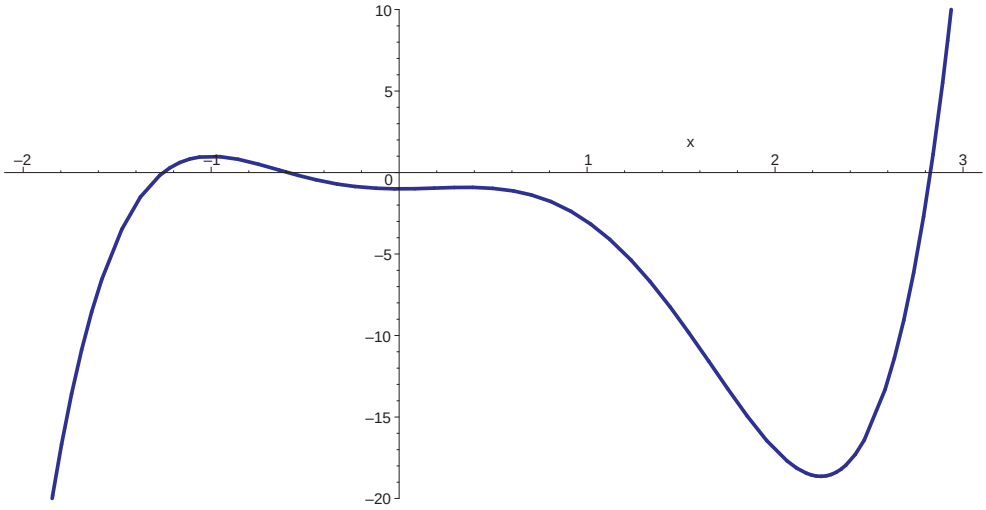
```
> Sturm(3);
```

$$[1, 1, 1, 1, -1, -1]$$

Dieselbe Folge, also liegt die Nullstelle in $[2, 3]$.

Wir haben also drei reelle Nullstellen, und sie liegen in den Intervallen $[-2, -1]$, $[-1, 0]$ und $[2, 3]$. Durch eine Zeichnung können wir uns vergewissern, daß die Nullstellen wirklich so liegen:

```
> plot(f, x=-3..2, color=blue, thickness=5);
```



Wie gut (und schnell) man die Nullstellen im konkreten Fall isolieren kann, hängt natürlich ab von deren Abstand. Erfahrungsgemäß funktioniert der Algorithmus recht gut; er ist auch in vielen Computeralgebrasystemen standardmäßig vorhanden. In Maple bestimmt `realroot(f)` für ein Polynom mit *ganzzahligen* Koeffizienten diese Intervalle; gibt man noch ein zweites Argument ℓ an, so werden Intervalle einer Länge von höchstens ℓ berechnet.

Da der Algorithmus in der Praxis gut funktioniert, könnte man es dabei bewenden lassen und an eine Bemerkung von ZASSENHAUS denken, der in einem Kolloquiumsvortrag an der Universität Karlsruhe einmal

sagte: „In der experimentellen Mathematik haben wir es nicht nötig, die Sätze zu beweisen, die wir für wahr halten.“ Tatsächlich aber ist von ZASSENHAUS so gut wie keine unbewiesene Behauptung überliefert, und auch für unser Problem gibt es bewiesene Resultate: Wie KURT MAHLER 1964 zeigte, ist der Betrag des Abstands zwischen zwei verschiedenen Nullstellen eines Polynoms $f \in \mathbb{C}[X]$ vom Grad n mit Diskriminante Δ (siehe Kapitel 2, §7) mindestens gleich

$$\frac{\sqrt{3\Delta}}{n^{(n+2)/2}\,|f|_1^{n-1}}\,.$$

Der Beweis verwendet abgesehen von den üblichen Techniken, die wir schon mehrfach beim Beweis von Schranken kennen gelernt haben, vor allem die Ungleichung von HADAMARD; da er relativ lang ist, sei hier darauf verzichtet. Interessenten finden ihn gut lesbar in der (auch frei im Netz zugänglichen) Originalarbeit

K. MAHLER: An inequality for the discriminant of a polynomial, *Michigan Math. J.* **11** (1964), 257–262

oder in §7.2.4 des Buchs

ALKIVAIDIS G. AKRITAS: Elements of Computer Algebra with Applications, *Wiley,* 1989

KURT MAHLER wurde 1903 in Krefeld als Sohn eines Buchdruckers geboren. Da er seit früher Kindheit an Knochen-Tuberkulose litt, ging er nur 1 1/2 Jahre zur Vorschule und 2 1/2 Jahre zur Volksschule. Um Feinmechaniker zu werden besuchte er ab 1917 zwei Jahre lang elementare technische Schulen in Krefeld. Dabei entdeckte er sein Interesse an Mathematik und kaufte sich entsprechende Bücher, die er parallel zur Schule studierte. Der Direktor seiner Schule schickte einige seiner Arbeiten an FELIX KLEIN, der sie seinem damaligen Assistenten CARL LUDWIG SIEGEL zur Begutachtung gab. Dieser befand, daß man MAHLER ein Mathematikstudium ermöglichen sollte. Mit Hilfe mehrerer Lehrer seiner Schule konnte er das Abitur bestehen und studierte dann ab 1923 bei SIEGEL in Frankfurt, ab 1925 bei HILBERT, COURANT, EMMY NOETHER, BORN, HEISENBERG und anderen in Göttingen, wo er auch eine Zeitlang als unbezahlter Assistent von NORBERT WIENER arbeitete. 1927 wurde er in Frankfurt

promoviert mit einer Arbeit über die Nullstellen der Γ-Funktion. 1933 erhielt er eine Stelle an der Universität Königsberg, konnte diese jedoch als Jude wegen der Machtergreifung der Nationalsozialisten nicht antreten. Auf Einladung von MORDELL ging er stattdessen nach Manchester, wo er abgesehen von zwei Jahren in Groningen trotz einer dreimonatigen Internierung als feindlicher Ausländer bis 1962 blieb. Seine letzten sechs Berufsjahre verbrachte er in Canberra, Australien, wo er auch nach seiner Emeritierung noch regelmäßig publizierte. Er starb dort im Februar 1988; seine letzte mathematische Arbeit erschien 1989. Praktisch alle seiner vielen Arbeiten befassen sich mit der Zahlentheorie; besonders berühmt sind seine Beiträge zur Theorie der transzendenten Zahlen.

Kapitel 2
Resultanten und Diskriminanten

§1: Der Begriff der Resultante

Auch wenn uns in dieser Vorlesung vor allem Polynomringe über $\mathbb{R}$ oder $\mathbb{C}$ interessieren, betrachten wir das folgende Problem zunächst über einem beliebigen faktoriellen Ring R. Zur Bequemlichkeit des Lesers seien die wesentlichen Definitionen kurz zusammengestellt:

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.

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) 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 $R^\times$.
d) 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.
e) Ein Integritätsbereich R heißt *faktoriell* oder *ZPE*-Ring, wenn gilt: Jedes Element $f \in R$ läßt sich bis auf Reihenfolge und Einheiten eindeutig schreiben als Produkt $f = u \prod_{i=1}^{n} p_i^{e_i}$ mit $u \in R^\times$, irreduziblen

Elementen $p_i \in R$ und natürlichen Zahlen e_i.
(*ZPE* steht für **Z**erlegung in **P**rimfaktoren **E**indeutig.)

Der Ring $\mathbb{Z}$ der ganzen Zahlen ist wegen der Existenz der Primzerlegung faktoriell; die Einheiten hier sind nur ± 1, die irreduziblen Elemente sind die Primzahlen und ihre Negativen. Auch Körper sind natürlich faktoriell; hier sind alle Elemente außer der Null Einheiten, und es gibt keine irreduziblen Elemente.

Nach einem Satz von GAUSS ist auch der Polynomring über einem faktoriellen Ring faktoriell; induktiv folgt also, daß alle Ringe der Art $\mathbb{Z}[X_1, \ldots, X_n]$ und $k[X_1, \ldots, X_n]$ mit einem Körper k faktoriell sind. Für den Beweis dieses Satzes sei auf eine der beiden Vorlesungen *Algebra* oder *Computeralgebra* verwiesen.

Resultanten sollen zeigen, ob zwei vorgegebene Polynome

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

und

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

mit Koeffizienten aus einem faktoriellen Ring R einen gemeinsamen Faktor positiven Grades haben.

Angenommen, es gibt ein Polynom $h \in R[X]$ vom Grad $d \geq 1$, 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

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

höchstens gleich $n + m - d$ ist.

Haben umgekehrt f und g ein gemeinsames Vielfaches vom Grad höchstens $n + m - d$, so hat auch ihr kleinstes gemeinsames Vielfaches S höchstens den Grad $n + m - d$. (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 p 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 $n + m - \deg S$ ist mindestens d. Damit ist gezeigt:

Lemma: Zwei Polynome $f, g \in R[X]$ haben genau dann einen gemeinsamen Teiler vom Grad mindestens d, wenn es Polynome $u \neq 0$ und $v \neq 0$ aus $R[X]$ gibt mit $\deg u \leq \deg g - d$ und $\deg v \leq \deg f - d$, für die $uf = vg$ ist.

■

Indem wir v durch $-v$ ersetzen, können wir auch sagen, dies sei genau dann der Fall, wenn es Polynome $u \neq 0$ und $v \neq 0$ aus $R[X]$ gibt mit $\deg u \leq \deg g - d$ und $\deg v \leq \deg f - d$, für die $uf + vg = 0$ ist.

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

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

und

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

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 d, wenn es nicht allesamt verschwindende Körperelemente $u_0, \ldots, u_{m-d}$ und $v_0, \ldots, v_{n-d}$ 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 $n+m+1-d$ Gleichungen für die $n + m + 2 - 2d$ Unbekannten $u_0, \ldots, u_{m-d}$ und

$v_0, \ldots, v_{n-d}$; es hat genau dann eine nichttriviale Lösung, wenn seine Matrix kleineren Rang als $n + m + 2 - 2d$ hat. Für $d = 1$, wenn die Matrix quadratisch ist, bedeutet dies einfach, daß ihre Determinante verschwindet; für $d > 1$ müssen wir d Determinanten von quadratischen Untermatrizen betrachten

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

$$\begin{aligned}
a_n u_{m-d} + b_m v_{n-d} &= 0 \\
a_{n-1} u_{m-d} + a_n u_{m-d-1} + b_{m-1} v_{n-d} + b_m v_{n-d-1} &= 0 \\
a_{n-2} u_{m-d} + a_{n-1} u_{m-d-1} + a_n u_{m-d-2} & \\
+ b_{m-2} v_{n-d} + b_{m-1} v_{n-d-1} + b_m v_{n-d-2} &= 0 \\
\ldots & \\
a_0 u_3 + a_1 u_2 + a_2 u_1 + a_3 u_0 + b_0 v_3 + b_1 v_2 + b_2 v_1 + b_3 v_0 &= 0 \\
a_0 u_2 + a_1 u_1 + a_2 u_0 + b_0 v_2 + b_1 v_1 + b_2 v_0 &= 0 \\
a_0 u_1 + a_1 u_0 + b_0 v_1 + b_1 v_0 &= 0 \\
a_0 u_0 + b_0 v_0 &= 0
\end{aligned}$$

Es hat genau dann eine nichttriviale Lösung, wenn der Rang seiner Matrix nicht maximal ist. Das ist äquivalent dazu, daß der Rang der transponierten Matrix nicht maximal ist, und da diese etwas übersichtlicher ist, betrachtet man im allgemeinen diese $(n+m+2-2d) \times (n+m+1-d)$-Matrix

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

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

Für $d = 1$ ist diese Matrix quadratisch; man bezeichnet sie als SYLVESTER-Matrix und ihre Determinante als die *Resultante* $\mathrm{Res}(f, g)$ 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 $\mathrm{Res}_X(f, g)$. Wir haben damit gezeigt

Satz: Die beiden Polynome $f, g \in R[X]$ haben genau dann einen einen gemeinsamen Faktor positiven Grades, wenn $\mathrm{Res}(f, g)$ verschwindet. ∎

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.

Für $d > 1$ hat die Matrix die obige Matrix $j = d - 1$ Spalten mehr als sie Zeilen hat. Wir betrachten die quadratische Matrix M_j, die durch Streichen der letzten j Spalten entsteht und bezeichnen ihre Determinante $r_j(f, g) = \det M_j$ als die j-te Subresultante von f und g. Für $j = d - 1 = 0$ ist M_j einfach die SYLVESTER-Matrix, und $r_0 = \mathrm{Res}_X(f, g)$.

Satz: Zwei Polynome $f, g \in R[X]$ über dem faktoriellen Ring R haben genau dann einen gemeinsamen Faktor vom Grad mindestens d, wenn

$r_0(f,g) = \cdots = r_{d-1}(f,g) = 0$ ist. Ihr ggT hat genau dann den Grad d, wenn zusätzlich noch $r_d(f,g) \neq 0$ ist.

Beweis durch Induktion nach d: Der Fall $d = 1$ ist bereits im obigen Satz gezeigt; sei also $d > 1$ und $j = d - 1$. Wir nehmen an, der Satz sei für Faktoren vom Grad mindestens j bewiesen.

Als erstes betrachten wir den Fall, daß f und g einen Faktor vom Grad mindestens d gemeinsam haben. Da $d > j$ ist, ist nach Induktionsannahme $r_0(f,g) = \cdots = r_{j-1}(f,g) = 0$. Wir müssen zeigen, daß auch $r_j(f,g)$ verschwindet.

Da f und g einen Faktor vom Grad d gemeinsam haben, gibt es Polynome $u, v \in R[X]$ vom Grad höchstens $m - d$ *bzw.* $n - d$, für die $uf + bg = 0$ ist, so daß die obige $(n + m + 2 - 2d) \times (n + m + 1 - d)$-Matrix nicht den maximalen Rang $n + m + 2 - 2d = n + m - 2j$ hat. Daher müssen alle quadratischen $(n + m - 2j) \times (n + m - 2j)$-Untermatrizen singulär sein, insbesondere auch die, die durch streichen der letzten j Spalten entsteht. Das ist aber gerade die Matrix M_j, so daß ihre Determinante $r_j(f,g)$ verschwinden muß.

Für die Umkehrung nehmen wir an, daß $r_0(f,g) = \cdots = r_{d-1}(f,g)$ alle verschwinden. Nach Induktionsannahme wissen wir dann, daß f und g einen Faktor vom Grad mindestens j gemeinsam haben. Nach Annahme verschwindet außerdem die Subresultante $r_{d-1}(f,g) = r_j(f,g)$, also die Determinante der Matrix M_j. Damit verschwindet auch die Determinante der Matrix

$$A = \begin{pmatrix} M_j & 0 \\ 0 & E_j \end{pmatrix},$$

wobei E_j die $j \times j$-Einheitsmatrix bezeichnet. Bezeichnet z den zu $(u_{m-d}, \ldots u_0, v_{n-d}, \ldots, v_0, w_{j-1}, \ldots, w_j)$ transponierten Spaltenvektor, so hat also das homogene lineare Gleichungssystem $A^T z = 0$ eine nichttriviale Lösung. Wenn wir mit der Argumentation zur Herleitung der SYLVESTER-Matrix vergleichen, sehen wir, daß die äquivalent ist zur Existenz von Polynomen u, v, w der Grade höchstens $m - d$, $n - d$ *bzw.* $j - 1$, die nicht allesamt verschwinden und für die $uf + vg + w = 0$ ist. Da f und g einen gemeinsamen Faktor vom Grad mindestens j haben, muß dieser dann auch w teilen; da w höchstens Grad $j - 1$ hat,

geht das nur, wenn w das Nullpolynom ist. Also gibt es Polynome u, v der Grade höchstens $m - d$ *bzw.* $n - d$, die nicht beide verschwinden und für die $uf + vg = 0$ ist. Dies zeigt, daß f und g einen gemeinsmen Faktor vom Grad mindestens d haben.

Damit ist der erste Teil des Satzes bewiesen, und der zweite folgt natürlich sofort daraus. ∎

§2: 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 beschränkt 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 dieser 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:

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

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

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

$$\deg g - \deg f = m - n \,.$$

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^{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 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 m behandelt wird, dessen führende Koeffizienten verschwinden.

Ist $h = c_s X^s + \cdots + c_0$, so ist also $\operatorname{Res}_X(f, g)$ 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 *usw.;* das Ganze funktioniert $m - s$ 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 $\operatorname{Res}_X(f, g) = a_n^{m-s} \operatorname{Res}_X(f, h)$, 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 $(-1)^{nm}$. ∎

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 nicht verschwinden. Die Gleichungen

$$X^i f = a_d X^{d+i} + \cdots + a_1 X^{1+i} + a_0 X^i \quad \text{und} \quad X^j g = b_e X^{e+j} + \cdots + b_1 X^{1+j} + b_0 X^j$$

für $i = 0, \ldots, e-1$ und $j = 0, \ldots, d-1$ können wir in Vektorschreibweise so zusammenfassen, daß wir den $(d+e)$-dimensionalen Vektor F mit Komponenten $X^{e-1}f, \ldots, Xf, f, X^{d-1}g, \ldots, Xg, g$ darstellen in der Form

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

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

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

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. ∎

§3: Variablenelimination mit Resultanten

In der reell-algebraischen Geometrie interessieren wir uns für die Nullstellenmengen reeller Polynome in mehreren Veränderlichen. Auch wenn es in diesem Kapitel noch in erster Linie um Polynome einer Veränderlichen geht, wollen wir uns doch schon kurz überlegen, wie uns Resultanten dabei helfen können. Da die speziellen Eigenschaften reeller Zahlen hier noch keine Rolle spielen, arbeiten wir über einem beliebigen Körper k.

Im Gleichungssystem

$$f_1(X_1, \ldots, X_n) = \cdots = f_m(X_1, \ldots, X_n) = 0$$

betrachten wir die $f_i \in k[X_1, \ldots, X_n]$ als Polynome in X_n mit Koeffizienten aus $k[X_1, \ldots, X_{n-1}]$. Falls die Resultante $\mathrm{Res}_{X_n}(f_i, f_j)$ für zwei Polynome f_i, f_j das Nullpolynom ist, haben f_i und f_j einen gemeinsamen Faktor; dies wird wohl nur selten der Fall sein. Falls wir die Polynome vorher faktorisieren und dann das eine Gleichungssystem ersetzen durch mehrere Systeme aus Polynomen kleineren Grades, können wir das sogar ausschließen.

Häufiger und interessanter ist der Fall, daß die Resultante nur für gewisse $(n-1)$-tupel $(x_1, \ldots, x_{n-1}) \in k^{n-1}$ verschwindet. Dann wissen wir, daß die Polynome

$$f_i(x_1, \ldots, x_{n-1}, X) \quad \text{und} \quad f_j(x_1, \ldots, x_{n-1}, X)$$

aus $k[X]$ zumindest in einem Erweiterungskörper von k eine gemeinsame Nullstelle haben. Falls wir $x_1, \ldots, x_{n-1}$ kennen, können wir diese Nullstelle(n) bestimmen, indem wir die Nullstellen zweier Polynome in einer Veränderlichen berechnen und miteinander vergleichen.

Um das obige Gleichungssystem zu lösen, führen wir es also zurück auf das Gleichungssystem

$$\mathrm{Res}_{X_n}(f_i, f_{i+1})(X_1, \ldots, X_{n-1}) = 0 \quad \text{für } i = 1, \ldots, m-1\,,$$

lösen dieses und betrachten für jedes Lösungstupel jenes Gleichungssystem in X_n, das entsteht, wenn wir im Ausgangssystem für die ersten $n-1$ Variablen die Werte aus dem Tupel einsetzen. Die Lösungen dieses Gleichungssystems sind gerade die Nullstellen des größten gemeinsamen Teilers aller Gleichungen.

Man beachte, daß dieser ggT durchaus gleich eins sein kann, daß es also nicht notwendigerweise eine Erweiterung des Tupels $(x_1, \ldots, x_{n-1})$ zu einer Lösung des gegebenen Gleichungssystems gibt: Wenn alle Resultanten verschwinden, haben nach Einsetzen zwar f_1 und f_2 eine gemeinsame Nullstelle und genauso auch f_2 und f_3, aber diese beiden Nullstellen können verschieden sein. Es muß also keine gemeinsame Nullstelle von f_1, f_2 und f_3 geben.

Als Beispiel für die Lösung eines nichtlinearen Gleichungssystems mit Resultanten betrachten wir die beiden Gleichungen

$$f(X,Y) = X^2 + 2Y^2 + 8X + 8Y - 40 \qquad \text{und}$$
$$g(X,Y) = 3X^2 + Y^2 + 18X + 4Y - 50\,.$$

Ihre Resultante bezüglich X ist

$$\mathrm{Res}_X(f,g) = 25Y^4 + 200Y^3 - 468Y^2 - 3472Y + 6820\,;$$

Maple gibt deren Nullstellen an als $y = -2 \pm \frac{1}{5}\sqrt{534 \pm 24\sqrt{31}}$. Diese können wir beispielsweise in g einsetzen, die entstehende quadratische Gleichung für x lösen, um dann zu testen, ob das Lösungspaar (x, y) auch eine Nullstelle von g ist. Zumindest mit Maple ist das durchaus machbar.

Einfacher wird es aber, wenn wir Y statt X eliminieren:

$$\mathrm{Res}_Y(f,g) = (5X^2 + 28X - 60)^2$$

ist das Quadrat eines quadratischen Polynoms, dessen Nullstellen

$$x = -\frac{14}{5} \pm \frac{4}{5}\sqrt{31}$$

uns die wohlbekannte Lösungsformel liefert. Diese Werte können wir nun in f oder g einsetzen, die entstehende Gleichung lösen und das Ergebnis ins andere Polynom einsetzen.

Alternativ können wir auch mit *beiden* Resultanten arbeiten: Ist (x, y) eine gemeinsame Nullstelle von f und g, so muß x eine Nullstelle von $\mathrm{Res}_Y(f, g)$ sein und y eine von $\mathrm{Res}_X(f, g)$. Da es nur $4 \times 2 = 8$ Kombinationen gibt, können wir diese hier einfach durch Einsetzen testen. Wie sich zeigt, hat das System die vier Lösungen

$$\left(-\frac{14}{5} + \frac{4}{5}\sqrt{31},\ -2 - \frac{1}{5}\sqrt{534 - 24\sqrt{31}}\right)$$
$$\left(-\frac{14}{5} + \frac{4}{5}\sqrt{31},\ -2 + \frac{1}{5}\sqrt{534 - 24\sqrt{31}}\right)$$
$$\left(-\frac{14}{5} - \frac{4}{5}\sqrt{31},\ -2 - \frac{1}{5}\sqrt{534 + 24\sqrt{31}}\right)$$
$$\left(-\frac{14}{5} - \frac{4}{5}\sqrt{31},\ -2 + \frac{1}{5}\sqrt{534 + 24\sqrt{31}}\right).$$

§4: Der Wurzelsatz von Viète und symmetrische Funktionen

Angenommen, wir haben ein Polynom

$$f = X^N + a_{n-1}X^{n-1} + a_{n-2}X^{n-2} + \cdots + a_2X^2 + a_1X + a_0\,,$$

mit (nicht notwendigerweise verschiedenen) Nullstellen $z_1, \ldots, z_n$. Dann ist auch $f = (X - z_1)(X - z_2)\cdots(X - z_n)$. Ausmultiplizieren und Koeffizientenvergleich liefert uns die Gleichungen

$$\begin{aligned}
a_{n-1} &= -(z_1 + \cdots + z_n)\\
a_{n-2} &= \sum_{i<j} z_i z_j\\
a_{n-3} &= -\sum_{i<j<k} z_i z_j z_k\\
&\vdots \qquad\qquad \vdots\\
a_0 &= (-1)^n z_1 \cdots z_n\,.
\end{aligned}$$

Allgemein ist a_{n-r} bis aufs Vorzeichen gleich der Summe alle Produkte aus r Werten z_i mit verschiedenem Index. Diese Summen bezeichnet

man als die *elementarsymmetrischen Funktionen* in $z_1, \ldots, z_n$ und die obigen Gleichungen als den Wurzelsatz von VIÈTE.

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

Für eine quadratische Gleichung $X^2 + pX + q = 0$ besagt er einfach, daß die Summe der Lösungen gleich $-p$ und das Produkt gleich q ist. Das hatten wir bereits in bei der Lösung kubischer Gleichungen ausgenutzt, um zwei Zahlen mit vorgegebenen Werten für Summe und Produkt zu berechnen.

Diese Summen, die sogenannten elementarsymmetrischen Funktionen, sind für r-Werte im mittleren Bereich recht umfangreich, die beiden Fälle $r = 0$ und $r = n - 1$ können aber gelegentlich ganz nützlich sein, um Lösungen zu erraten:

Falls wir aus irgendeinem Grund erwarten, daß alle Nullstellen ganzzahlig sind, folgt aus der Tatsache, daß ihr Produkt gleich $(-1)^n a_0$ ist, daß sie allesamt Teiler von a_0 sein müssen. Außerdem ist ihre Summe gleich $-a_{n-1}$.

Bei der Gleichung $f = X^3 - 7X + 6 = 0$ etwa, die uns in §3 so viele Schwierigkeiten machte, ist das Produkt aller Nullstellen gleich -6; *falls* sie alle ganzzahlig sind, kommen also nur $\pm 1, \pm 2, \pm 3$ und ± 6 in Frage. Aus diesen acht Zahlen müssen wir drei (nicht notwendigerweise verschiedene) auswählen mit Produkt -6 und Summe null. Das geht offensichtlich nur mit $1, 2$ und -3; Einsetzen zeigt, daß dies auch tatsächlich Nullstellen sind.

Man beachte, daß dieses Einsetzen unbedingt notwendig ist: Bei der

Gleichung $g(x) = X^3 - 6X + 6 = 0$ hätten wir genauso vorgehen können und wären auf dieselben drei Kandidaten gekommen, aber $g(1) = 1, g(2) = 2$ und $g(-3) = -3$. Hier führt aber die Lösungsformel aus §3 relativ schnell ans Ziel: Einsetzen der Parameter $p = -6$ und $q = 6$ in die Lösungsformel führt zunächst auf

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

für die reelle Wurzel; die erste Lösung ist also

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

Für die zweite und dritte Lösung müssen wir mit $u\rho$ *bzw.* $u\bar{\rho}$ anstelle von u arbeiten und erhalten

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

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

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

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

führt. Natürlich erfüllen auch diese Zahlen den Satz von VIÈTE, jedoch nützt uns dieser nichts, um sie zu erraten.

§5: Symmetrische Polynome

Die elementarsymmetrischen Funktionen σ_i sind Polynome, die sich nicht verändern, wenn ihre Variablen in irgendeiner Weise permutiert werden. Solche Polynome bezeichnen wir als symmetrisch:

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

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

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

Satz: $f \in k[X_1, \ldots, X_n]$ sei ein symmetrisches Polynom. Dann gibt es ein Polynom $g \in k[Y_1, \ldots, Y_n]$, so daß

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

Beweis: Zur Konstruktion von g ordnen wir die Monome $X_1^{e_1} \cdots X_n^{e_n}$ lexikographisch an, d.h. das Monom $X_1^{e_1} \cdots X_n^{e_n}$ kommt vor $X_1^{f_1} \cdots X_n^{f_n}$, wenn die erste von Null verschiedene Differenz $e_i - f_i$ positiv ist. Da f ein symmetrisches Polynom ist, muß im ersten, dem sogenannten *führenden* Monom, $e_1 \geq e_2 \geq \cdots \geq e_n$ sein, denn wegen der Symmetrie muß mit $X_1^{e_1} \cdots X_n^{e_n}$ auch jedes Monom $X_1^{e_{\pi(1)}} \cdots X_n^{e_{\pi(n)}}$ vorkommen. Der Koeffizient von $X_1^{e_1} \cdots X_n^{e_n}$ sei a.

Das führende Monom von σ_i ist $X_1 \cdots X_i$; setzen wir $\delta_i = e_i - e_{i-1}$ für $i = 1, \ldots, n-1$ und $\delta_n = e_n$, hat daher $\sigma_1^{\delta_1} \cdots \sigma_n^{\delta_n}$ ebenfalls das führende Monom $X_1^{e_1} \cdots X_n^{e_n}$.

Im der Differenz $f_1 = f - a\sigma_1^{\delta_1} \cdots \sigma_n^{\delta_n}$ heben sich somit die führenden Monome weg, und f_1 hat ein führendes Monom, das kleiner ist als $X_1^{e_1} \cdots X_n^{e_n}$. Wir schreiben $f = a\sigma_1^{\delta_1} \cdots \sigma_n^{\delta_n} + f_1$, wenden die gleiche Konstruktion an auf f_1, und so weiter. Da die führenden Monome dabei immer kleiner werden und es nur endlich viele Monome gibt, die lexikographisch kleiner sind als ein gegebenes Monom, endet diese Konstruktion nach endlich vielen Schritten und drückt f aus als Linearkombination von Monomen in den σ_i. Das gesuchte Polynom g ist nun einfach die entsprechende Linearkombination mit Monomen in den Y_i an Stelle der σ_i.

∎

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

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

und erhalten

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

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

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

Somit ist

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

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

Satz: P sei ein symmetrisches Polynom aus $k[Z_1,\dots,Z_n]$, und für jedes n-tupel $z = (z_1,\dots,z_n)$ sei

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

Dann gibt es ein Polynom $Q \in k[a_0,\dots,a_{n-1}]$ mit der Eigenschaft, daß $P(z) = Q\left(a_0^{(z)},\dots,a_{n-1}^{(z)}\right)$ für alle $z \in k^n$.

∎

§6: Die Resultante als Funktion der Wurzeln

Die Resultante gibt an, ob zwei Polynome einen gemeinsamen Faktor haben. Wenn wir sie über einem Körper betrachten, der alle Nullstellen der beiden Polynome enthält, läßt sich das leicht entscheiden: Sind $\alpha_1,\dots,\alpha_n$ und $\beta_1,\dots,\beta_m$ die Nullstellen von

$$f = a_nX^n + \dots + a_1X + a_0 = a_n(X-\alpha_1)\cdots(X-\alpha_n)$$

und

$$g = b_m X^m + \cdots + b_1 X + b_0 = b_m (X - \beta_1) \cdots (X - \beta_m)$$

mit $a_n \neq 0$ und $b_m \neq 0$, so gibt es genau dann einen gemeinsamen Faktor und damit eine gemeinsame Nullstelle, wenn mindestens eine der Zahlen $f(\beta_j)$ verschwindet, wenn also das Produkt

$$\prod_{j=1}^{m} f(\beta_j) = a_n^m \prod_{j=1}^{m} \prod_{i=1}^{n} (\beta_j - \alpha_i)$$

verschwindet. Entsprechend können wir natürlich auch sagen, daß dies genau dann der Fall ist, wenn

$$\prod_{i=1}^{n} g(\alpha_i) = b_m^n \prod_{i=1}^{n} \prod_{j=1}^{m} (\alpha_i - \beta_j)$$

verschwindet. Bis auf den Vorfaktor und eventuell das Vorzeichen sind die beiden Ausdrücke gleich; um etwas möglichst symmetrisches zu haben, betrachten wir

$$S = a_n^m b_m^n \prod_{i=1}^{n} \prod_{j=1}^{m} (\alpha_i - \beta_j) = a_n^m \prod_{i=1}^{n} g(\alpha_i) = (-1)^{nm} b_m^n \prod_{j=1}^{m} f(\beta_j)\,,$$

und natürlich verschwindet auch dieser Ausdruck genau dann, wenn f und g einen gemeinsamen Faktor haben. Alle drei betrachteten Ausdrücke sollten also etwas mit der Resultante von f und g zu tun haben.

Dazu betrachten wir a_n, b_m sowie die α_i und die β_j als Unbestimmte und S als ein Polynom in diesen Variablen. Auch die Resultante können wir als ein solches Polynom betrachten, denn die Koeffizienten von f lassen sich über den Wurzelsatz von VIÈTE ausdrücken durch a_n und die α_i, die von g durch b_m und die β_j.

Falls ein α_i gleich einem β_j ist, haben f und g den Faktor $X - \alpha_i$ gemeinsam; also muß ihre Resultante verschwinden. Somit ist $\mathrm{Res}_X(f, g) = 0$ wann immer $\alpha_i - \beta_j$ verschwindet. Wir wollen uns überlegen, daß dann $\alpha_i - \beta_j$ ein Teiler der Resultante sein muß:

Wie wir wissen, ist für ein Polynom $h \in k[X]$ über einem Körper k für jede Nullstelle c das lineare Polynom $X - c$ ein Teiler von h. Zum

Beweis dividieren wir einfach h durch $(X-c)$; ist $q \in k[X]$ der Quotient und $r \in k$ der Rest, gilt dann

$$f = q \cdot (X - c) + r \,,$$

Setzen wir in alle vorkommenden Polynome den Wert c ein, verschwindet alles außer eventuell r; also muß $r = 0$ sein.

Ersetzen wir den Körper k durch einen beliebigen Integritätsbereich R, haben wir keinen allgemeinen Algorithmus zur Polynomdivision mehr; für ein Element $c \in R$ können wir aber immer noch jedes $h \in R[X]$ durch $X - c$ dividieren, denn da der führende Koeffizient von $X - c$ gleich eins ist, müssen wir bei Anwendung des gewohnten Algorithmus zur Polynomdivision nie durch ein anderes Element als die Eins dividieren, und das geht natürlich in jedem Integritätsbereich. Also können wir genauso argumentieren wie im Körperfall.

In unserem Fall nehmen wir als R den Polynomring in den Variablen a_n, b_m, allen α_k außer α_i und allen β_j und betrachten darüber den Polynomring $R[\alpha_i]$. Division der Resultante durch $\alpha_i - \beta_j$ gibt uns dann eine Darstellung

$$\mathrm{Res}_X(f, g) = q \cdot (\alpha_i - \beta_j) + r \,,$$

wobei q und r in R liegen. Setzen wir $\alpha_i = \beta_j$, verschwindet alles außer eventuell r; also muß auch $r = 0$ sein und die Resultante ist ein Vielfaches von $\alpha_i - \beta_j$.

Da dies für alle Paare (i, j) gilt und a_n, b_m als nicht verschwindende Konstanten jedes Polynom teilen, folgt, daß S ein Teiler von $\mathrm{Res}_X(f, g)$ sein muß. Aus den Darstellungen

$$S = b_m^n \prod_{j=1}^{m} f(\beta_j) = (-1)^{nm} a_n^m \prod_{i=1}^{n} g(\alpha_i)$$

folgt, daß S als Polynom in den a_i homogen vom Grad m ist, und als Polynom in den b_j homogen vom Grad n. Genau die gleichen Homogenitätseigenschaften hat auch die Resultante; da sie ein Vielfaches von S ist, kann der Faktor daher nur eine Konstante sein. Diesen können wir bestimmen, indem wir in beiden Ausdrücken das Monom betrachten, das die höchste Potenz des konstanten Koeffizienten b_0 von g enthält. In

der Resultante ist dies das Produkt der Diagonalelemente, also $a_n^m b_0^n$, und in S wegen der Darstellung $S = a_n^m \prod_{i=1}^n g(\alpha_i)$ ebenfalls. Also ist der Faktor gleich eins und wir haben gezeigt

Satz: Für zwei Polynome

$$f = a_n X^n + \cdots + a_1 X + a_0 = a_n(X - \alpha_1) \cdots (X - \alpha_n)$$

und

$$g = b_m X^m + \cdots + b_1 X + b_0 = b_m(X - \beta_1) \cdots (X - \beta_m)$$

mit $a_n, b_m \neq 0$ ist

$$\begin{aligned} \mathrm{Res}_X(f,g) &= a_n^m b_m^n \prod_{i=1}^n \prod_{j=1}^m (\alpha_i - \beta_j) \\ &= a_n^m \prod_{i=1}^n g(\alpha_i) = (-1)^{nm} b_m^n \prod_{j=1}^m f(\beta_j)\,, \end{aligned}$$

Korollar: Die Resultante als Polynom in $a_0, \ldots, a_n, b_0, \ldots, b_m$ ist irreduzibel.

Beweis: Wäre $\mathrm{Res}_X(f,g) = p \cdot q$ Produkt zweier nichtkonstanter Polynome, so ließen sich auch diese als symmetrische Funktionen in den α_i und β_j schreiben. Insbesondere müßte einer der beiden Faktoren durch $\alpha_1 - \beta_1$ teilbar sein. Wegen der Symmetrie müßte er dann durch *jedes* $\alpha_i - \beta_j$ teilbar sein, also auch durch deren Produkt. Für den anderen Faktor bliebe somit nur ein Teiler des Polynoms $a_n^m b_m^n$. Da die Resultante als Polynom in den a_i und b_j aber weder durch a_n noch durch b_m teilbar ist, gibt es keinen solchen Teiler außer Konstanten. Also müßte der andere Faktor konstant sein, im Widerspruch zur Annahme. ∎

§7: Die Diskriminante eines Polynoms

Als weitere Anwendung des gerade bewiesenen Satzes wollen wir die *Diskriminante* eines Polynoms f definieren. Sie soll angeben, ob f mehrfache Nullstellen hat. Mehrfache Nullstellen sind gleichzeitig

Nullstellen der Ableitung und umgekehrt; daher gibt es genau dann mehrfache Nullstellen, wenn die Resultante des Polynoms und seiner Ableitung verschwindet. Nach obigem Satz ist

$$\mathrm{Res}_X(f, f') = a_n^{n-1} \prod_{i=1}^{n} f'(\alpha_i)\,,$$

wobei $\alpha_1, \ldots, \alpha_n$ die Nullstellen von f sind, und nach der Produktregel, angewandt auf

$$f = a_n(X - \alpha_1)\cdots(X - \alpha_n)$$

ist

$$f' = a_n \sum_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} (X - \alpha_j)\,.$$

Der i-te Summand verschwindet also für alle α_j mit $j \neq i$; daher ist

$$f'(\alpha_i) = \prod_{\substack{j=1 \\ j \neq i}}^{n} (\alpha_i - \alpha_j) \quad \text{und} \quad \mathrm{Res}_X(f, f') = a_n^{2n-1} \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} (\alpha_i - \alpha_j)\,.$$

Schreiben wir die Resultante als Determinante, so stehen in der ersten Spalte a_n und na_n; die erste Spalte und damit die Resultante ist also durch a_n teilbar.

Definition: Die Diskriminante von f ist $\Delta = a_n^{2n-2} \prod_{i<j} (\alpha_i - \alpha_j)^2$.

Wir wir gerade gesehen haben, läßt sie sich als ein Polynom in den Koeffizienten von f schreiben, und $\Delta = \pm a_n \,\mathrm{Res}_X(f, f')$.

Kapitel 4
Rechnen mit reellen Zahlen

§ 1: Unentscheidbarkeitsprobleme

Trotz ihres Namens sind die reellen Zahlen alles andere als real: Es wäre beispielsweise völlig sinnlos, von einem realen Gegenstand zu sagen, er haben eine Länge von $\sqrt{2}\,m$ oder eine Masse von $\pi\,kg$. Die reellen Zahlen bilden schließlich eine überabzählbare Menge, während wir in einer endlichen Welt leben. Auch unsere Gehirne und unsere Computer sind endlich. Trotzdem zeigt die Erfahrung von über zwei Jahrtausenden, daß die reellen Zahlen gerade wegen der idealisierten Annahmen, die ihnen zugrunde liegen, extrem nützlich sind für die Beschreibung naturwissenschaftlicher und teilweise auch wirtschaftlicher und sozialer Phänomene.

Wirklich rechnen können wir aber nicht mit reellen Zahlen: wie DANIEL RICHARDSON 1969 zeigte, können wir nicht einmal immer entscheiden, ob eine durch einen relativ einfachen Ausdruck gegebene reelle Zahl verschwindet oder nicht. Mit dem, was wir heute wissen, können wir seinen Satz so formulieren:

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 +, $\cdot$, Sinus und Betrag verschwindet.

DANIEL RICHARDSON wurde 1941 in Chicago geboren. Er studierte Mathematik in New York und in San Francisco, wo er 1962 seinen Bachelor bekam. Danach ging er an die Universität von Bristol und promovierte dort 1965 in mathematischer Logik. Nach verschiedenen ein- bis zweijährigen Tätigkeiten an Universitäten und in der Industrie

wurde er 1988 zunächst Lecturer, später Senior Lecturer, an der Universität von Bath. http://people.bath.ac.uk/masdr/

Ein vollständiger Beweis des Satzes von RICHARDSON wäre fast eine eigene zweistündige Vorlesung; deshalb sollen hier nur die wesentlichen Ideen kurz skizziert werden.

Das erste Unentscheidbarkeitsresultat in der Geschichte der Mathematik war ein berühmter Satz von KURT GÖDEL (1906–1978), wonach jedes Axiomensystem zur Charakterisierung der natürlichen Zahlen entweder Widersprüche enthält oder aber nicht jede wahre Aussage als Folgerung ableiten läßt.

Die damit bewiesene Unentscheidbarkeit der Zahlentheorie wurde wenig später mit anderen Ansätzen auch von ALAN TURING (1912–1954), EMIL POST (1897–1954) und von ALONZO CHURCH (1903–1995) gezeigt. Am einfachsten verständlich ist wohl der Ansatz von TURING in seiner (auch vielfach im Netz zu findenden) Arbeit

ALAN K. TURING: On computable numbers with an application to the Entscheidungsproblem, *Proc. London Math. Soc.* **42** (1936), 230–265 *und* **43** (1937), 544–546

ALAN MATHISON TURING wurde 1912 in London geboren. Ab 1931 studierte er Mathematik an der Universität Cambridge; 1935 wurde er auf Grund einer Arbeit über den zentralen Grenzwertsatz *fellow* am King's College; dort publizierte er 1936 die oben zitierte Arbeit. Da CHURCH im gleichen Jahr mit anderen Methoden dasselbe Unentscheidbarkeitsresultat veröffentlicht hatte, kam es zu Diskussionen zwischen den beiden und TURING ging 1937 zu CHURCH an die Universität Princeton. 1938 kehrte er nach England zurück und begann mit Plänen zum Bau eines Computers zur Berechnung der Nullstellen der RIEMANNschen ζ-Funktion. Diese Arbeit wurde unterbrochen durch den zweiten Weltkrieg; ab 1939 arbeitete er in der Code and Cypher School in Bletchley Park, wo er unter anderem Rechenmaschinen (die sogenannten Bomben) zum Knacken der deutschen Enigma-Verschlüsselung entwickelte. 1942/43 war er in den USA an einem britisch-amerikanischen Projekt zur Sprachverschlüsselung beteiligt.

Nach Kriegsende entwarf und baute er am National Physical Laboratory in London einen der ersten Computer; 1948 wurde er nach einem kurzen Zwischenaufenthalt in Cambridge

Mathematikprofessor in Manchester. Dort arbeitete er unter anderem am Wortproblem für Gruppen und über Reaktions-Diffusionsgleichungen zur Modellierung der Morphogenese, d.h. der Entstehung biologischer Formen und Strukturen. Sein Tod 1954 wurde offiziell als Selbstmord bezeichnet, könnte aber auch ein Unfall bei einem chemischen Experiment gewesen sein.

In der zitierten Arbeit zeigt er, daß es kein Verfahren geben kann, das für einen beliebigen Algorithmus zeigt, ob er abbricht: Gäbe es nämlich so ein Verfahren $V(A, E)$, das genau dann die Antwort „bricht ab“ liefert, wenn der Algorithmus A mit Eingabe E nach endlich vielen Schritten abbricht, und „bricht nicht ab“ sonst, so könnten wir V auch auf den folgenden Algorithmus $W(A)$ anwenden: Breche ab, falls $V(A, A)$ zum Ergebnis „bricht nicht ab“ kommt, und gehe sonst in eine Endlosschleife. Offensichtlich bricht $W(W)$ genau dann ab, wenn $V(W, W)$ uns sagt, daß $W(W)$ *nicht* abbricht und umgekehrt. Somit kann V nicht immer das richtige Ergebnis liefern. Er stellt auch die Verbindung mit natürlichen Zahlen her, indem er Programmen Zahlen zuordnet.

Wie man heute weiß, gibt es unentscheidbare Mengen auch im Zusammenhang mit Polynomgleichungen. Ausgangspunkt dafür ist eines der 23 Probleme, die DAVID HILBERT 1900 auf dem Internationalen Mathematikerkongreß in Paris vorstellte und 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 gelöst, andere sind auch heute nach mehr als ein Jahrhundert noch offen.

DAVID HILBERT (1862–1943) wurde in Königsberg geboren, wo er auch zur Schule und zur Universität ging. Er promovierte dort 1885 mit einem Thema aus der Invariantentheorie, habilitierte sich 1886 und bekam 1893 einen Lehrstuhl. 1895 wechselte er an das damalige Zentrum der deutschen wie auch internationalen Mathematik, die Universität Göttingen, wo er bis zu seiner Emeritierung im Jahre 1930 lehrte. Seine Arbeiten umfassen ein riesiges Spektrum aus unter anderem Invariantentheorie, Zahlentheorie, Geometrie, Funktionalanalysis, Logik und Grundlagen der Mathematik sowie auch zur Relativitätstheorie. Er gilt als einer der Väter der modernen Algebra.

HILBERTs zehnte Problem war:

> 10. Entscheidung der Lösbarkeit einer Diophantischen Gleichung
>
> Eine Diophantische Gleichung mit irgend welchen Unbekannten und mit ganzen rationalen Zahlencoeffizienten sei vorgelegt: *man soll ein Verfahren angeben, nach welchem sich mittelst einer endlichen Anzahl von Operationen entscheiden läßt, ob die Gleichung in ganzen rationalen Zahlen lösbar ist.*

HILBERT war, wie einige seiner späteren Arbeiten zeigen, davon überzeugt, daß es ein solches Verfahren geben müsse; er glaubte sogar an einen Entscheidungsalgorithmus für die gesamte Analysis.

1970 bewies dann aber YURI MATIYASEVICH im Alter von 22 Jahren in seiner Doktorarbeit daß es *kein solches Verfahren geben kann.* Er zeigte dies über ein positives Resultat: Jede *rekursiv aufzählbare* Teilmenge M von $\mathbb{N}^n, \mathbb{N}_0^n$ und $\mathbb{Z}^n$ ist *diophantisch,* d.h. es gibt eine nichtnegative ganze Zahl m und ein Polynom

$$f \in \mathbb{Z}[A_1, \ldots, A_n, X_1, \ldots, X_m]$$

derart, daß

$$M = \left\{(a_1, \ldots, a_n) \;\middle|\; \exists (x_1, \ldots, x_m) : f(a_1, \ldots, a_n, x_1, \ldots, x_m) = 0\right\}.$$

Die a_i und die x_j müssen dabei jeweils in $\mathbb{N}, \mathbb{N}_0$ *bzw.* $\mathbb{Z}$ liegen.

YURI VLADIMIROVICH MATIYASEVICH (Юрий Владимирович Матиясевич) wurde 1947 in Sankt Petersburg (damals Leningrad) geboren, ging dort zur Schule und studierte an der dortigen Universität. 1969-1979 arbeitete er am Petersburger Steklov Institut für Mathematik an seiner Dissertation, dem gerade erwähnten Resultat. Seit 1980 leitet er die dortige Arbeitsgruppe für Logik; außerdem hat er einen Lehrstuhl für Algebra und Zahlentheorie an der Sankt Petersburger Universität.
http://logic.pdmi.ras.ru/~yumat/

Eine rekursiv aufzählbare Menge ist eine (in allen interessanten Fällen unendliche) Menge, für die es einen Algorithmus gibt, der nach hinreichend langer Anwendung jedes ihrer Elemente produziert. Sie heißt *rekursiv entscheidbar,* wenn es auch einen entsprechenden Algorithmus für ihre Komplementärmenge gibt; dann läßt sich für jedes Element der Grundgesamtheit in endlich vielen Schritten entscheiden, ob es in der Menge liegt oder aber in der Komplementärmenge.

Beispiel einer rekursiv aufzählbaren Menge ist etwa die Menge aller Primzahlen; sie ist rekursiv aufzählbar, etwa durch den einfachsten (und dümmsten) Algorithmus, der nacheinander alle natürlichen Zahlen n darauf untersucht, ob es unter den Zahlen $1 < m < n$ eine gibt, die m teilt; falls nicht, ist n eine Primzahl. In der Tat zeigten

JAMES J. JONES, DAIHACHIRO SATO, HIDEO WADA, DOUGLAS WIENS: Diophantine representations of the set of prime numbers, *Am. Math. Monthly* **83** (1976), 449-464,

daß das Polynom

$$\begin{aligned}
&[wz+h+j+q]^2+[(gk+2g+k+1)(h+j)+h+z]^2+[2n+p+q+z+e]^2\\
&+[16(k+1)^3(k+2)(n+1)^2+1+f^2]^2+[e^3(e+2)(a+1)^2+1+o^2]^2\\
&+[(a^2+1)y^2+1+x^2]^2+[16r^2y^4(a^2+1)+1+u^2]^2\\
&+[((a+u^2(u^2+a))^2+1)(n+4dy)^2+1+(x+cu)^2]^2\\
&+[n+\ell+v+y]^2+[(a^2+1)\ell^2+1+m^2]^2+[ai+k+1+\ell+i]^2\\
&+[p+\ell(a+n+1)+b(2an+2a+n^2+2n+2)+m]^2\\
&+[q+y(a+p+1)+s(2ap+2a+p^2+2p+2)+x]^2\\
&+[z+p\ell(a+p)+t(2ap+p^2+1)+pm]^2
\end{aligned}$$

genau für die $k \in \mathbb{N}_0$, für die $k+2$ prim ist, eine Nullstelle

$$(a,b,c,d,e,f,g,h,i,j,k,\ell,m,n,o,p,q,r,s,t,u,v,w,x,y,z) \in \mathbb{N}_0^{26}$$

hat. Da eine Summe von Quadraten reeller Zahlen x_i genau dann verschwindet, wenn jede einzelne x_i verschwindet, ist das gleichbedeutend damit, daß das System aus den 14 Polynomen in den eckigen Klammern eine entsprechende Lösung hat.

Die Komplementärmenge der Primzahlen, die Menge der zusammengesetzten Zahlen plus der Eins, ist natürlich auch rekursiv aufzählbar: Dazu müssen wir einfach die Menge $(\mathbb{N} \setminus 1) \times (\mathbb{N} \setminus 1)$ irgendwie anordnen und jedes Produkt ab eines Paars (a,b) als zusammengesetzt betrachten. Die Menge der Primzahlen sowie die Menge der zusammengesetzten Zahlen sind damit auch rekursiv entscheidbar.

Die Menge M aller Paare (A, E) bestehend aus einem Algorithmus A und dessen Eingabedaten E derart, daß A mit Eingabe E nach endlich vielen Schritten stoppt, ist ebenfalls rekursiv aufzählbar. Zum Beweis kann man beispielsweise die Paare (A, E) irgendwie anordnen und dann nach Art der Zeitscheibentechnik eines Betriebssystems quasiparallel ausführen in einer solchen Weise, daß jedes Paar (A, E), für das $A(E)$ nach endlich vielen Schritten ein Ergebnis liefert, dieses Ergebnis auch hier nach endlich vielen Schritten liefert. Setzt man nach Ende eines Programms $A(E)$ das Paar (A, E) auf eine Liste endender Programme, so muß offensichtlich jedes solche Paar nach und nach auf dieser Liste erscheinen, die Menge aller dieser Paare ist also rekursiv aufzählbar. Nach TURING können wir die Menge *aller* Paare (A, E) in Bijektion mit $\mathbb{N}$ setzen und so M mit einer Teilmenge von $\mathbb{N}$ identifizieren.

Diese Menge kann nach TURING nicht entscheidbar sein, ist nach MATIYASEVICH aber diophantisch. Somit muß es Familien von diophantischen Gleichungen geben, bei denen nicht entscheidbar ist, für welche Parameterwerte sie lösbar sind.

Um den Satz von RICHARDSON zu beweisen, müssen wir daraus ein Problem mit reellen Zahlen machen. Das ist relativ einfach: Eine reelle Zahl x ist genau dann ganz, wenn $\sin \pi x$ verschwindet. Wenn wir daher von einem Polynom $f \in \mathbb{Z}[X_1, \ldots, X_m]$ nicht entscheiden können, ob es eine Nullstelle $(x_1, \ldots, x_m) \in \mathbb{N}_0^m$ hat, können wir von der Funktion

$$f\left(|x_1|, \ldots, |x_m|\right)^2 + \sum_{i=1}^{m} \sin^2 \pi x_i$$

nicht entscheiden, ob sie eine reelle Nullstelle hat.

Für Einzelheiten und Folgerungen sei etwa auf das Buch

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

verwiesen. Einen einfachen und kurzen Beweis des Satzes von MATIYASEVICH findet man im (via `www.jstor.org` auch online erhältlichen) Artikel

J. P. JONES, Y. V. MATIJASEVIČ: Proof of Recursive Unsolvability of Hilbert's Tenth Problem, *American Mathematical Monthly* **98** (1991), 689–709

§2: Rechnen in Teilkörpern von $\mathbb{R}$

Das für uns wesentliche Ergebnis des vorigen Paragraphen ist, daß wir im Körper der reellen Zahlen nicht wirklich rechnen können. In diesem Paragraphen wollen wir uns überlegen, daß wir zumindest in einem Teilkörper der reellen Zahlen exakt rechnen und auch die gängigen Relationen entscheiden können:

Definition: Eine reelle Zahl $z \in \mathbb{R}$ heißt algebraisch, wenn es ein Polynom $f \in \mathbb{Q}[X]$ gibt, so daß $f(z)$ verschwindet; andernfalls heißt z transzendent.

So ist beispielsweise $\sqrt{2}$ als Nullstelle des Polynoms $x^2 - 2$ algebraisch, π dagegen ist nach einem 1882 bewiesenen Resultat von FERDINAND VON LINDEMANN (1852–1939) transzendent.

Eine reelle algebraische Zahl kann eindeutig beschrieben werden durch Angabe eines irreduziblen Polynoms f aus $\mathbb{Q}[X]$ und eines Intervalls $[a, b]$, in dem f nur eine Nullstelle hat. Die Intervallenden wählen wir – um exakt rechnen zu können – als rationale Zahlen. Wir wollen uns überlegen, daß wir mit solchen Paaren $\big(f, [a, b]\big)$ alle Grundrechenarten durchführen können, was gleichzeitig zeigen wird, daß die reellen algebraischen Zahlen einen Körper bilden und daß wir auch Gleichheit sowie Größenbeziehungen entscheiden können.

Dazu seien u und v zwei reelle algebraische Zahlen, gegeben durch die Paare $\big(f, [a, b]\big)$ und $\big(g, [c, d]\big)$.

Beginnen wir mit der Addition: Für $w = u+v$ ist $g(w-u) = g(v) = 0$. Wir führen eine neue Variable Z ein und schreiben $g(Z - X)$ als Polynom in X mit Koeffizienten aus $\mathbb{Q}[Z]$. Natürlich können wir auch $f \in \mathbb{Q}[X]$ als ein solches Polynom auffassen: Die Koeffizienten von f sind rationale Zahlen und damit auch (konstante) Polynome aus $\mathbb{Q}[Z]$. Wir haben damit zwei Polynome in X mit Koeffizienten aus $\mathbb{Q}[Z]$, die für $z = u+v$ die gemeinsame Nullstelle $X = u$ haben.

Aus dem vorigen Kapitel wissen wir, daß zwei Polynome in X über einem faktoriellen Ring R genau dann eine gemeinsame Nullstelle haben, wenn ihre Resultante verschwindet. Diese ist hier ein Polynom

$h \in \mathbb{Q}[Z]$, das somit für $z = u + v$ verschwinden muß. Damit haben wir ein Polynom h mit rationalen Koeffizienten gefunden, das $w = u + v$ als Nullstelle hat.

Da $a \le u \le b$ und $c \le v \le d$, liegt w natürlich im Intervall $[a+c,\ b+d]$; es ist allerdings nicht klar, daß h in diesem Intervall nur w als Nullstelle hat. Dies läßt sich aber nach STURM (oder DESCARTES-JACOBI oder BUDAN-FOURIER) überprüfen; gegebenenfalls müssen die Intervalle $[a,\ b]$ und $[c,\ d]$ so lange verkleinert werden, bis auch das Intervall $[a + c,\ b + d]$ hinreichend klein ist.

Damit wissen wir, wie man Summen berechnet; um auch Differenzen berechnen zu können, müssen wir uns nur überlegen, wie man für eine durch $\big(f, [a,\ b]\big)$ dargestellte reelle algebraische Zahl u ihr Negatives darstellt. Mit

$$f = \sum_{i=0}^{n} a_i x^i \quad \text{ist} \quad h = \sum_{i=0}^{n} (-1)^i a_i x^i$$

offensichtlich ein Polynom, das $-u$ als Nullstelle hat, und $-u$ liegt im Intervall $[-b,\ -a]$. Es ist dort auch die einzige Nullstelle, denn jede Nullstelle von h ist das Negative einer Nullstelle von f.

Für Produkte können wir fast genauso vorgehen wir für Summen: Ist $w = uv$ und $u \neq 0$, so ist $g(w/u) = 0$. Um aus $g(Z/X)$ ein Polynom zu machen, müssen wir mit X^m multiplizieren, wobei m den Grad von g bezeichnet: Für $g = b_m X^m + \cdots + b_0$ betrachten wir also

$$X^m g(Z/X) = b_m Z^m + b_{m-1} Z^{m-1} X + \cdots + b_1 Z X^{m-1} + b_0 X^m\ ,$$

ein Polynom vom Grad m in X mit Koeffizienten aus $\mathbb{Q}[Z]$. Auch f können wir als so ein Polynom auffassen und erhalten wie oben, daß die Resultante dieser beiden Polynome ein rationales Polynom ist, das in w verschwindet. Das Intervall, in dem w liegt, läßt sich leicht aus a, b, c und d berechnen, allerdings sind Fallunterscheidungen bezüglich der Vorzeichen nötig. Auch hier kann es wieder notwendig werden, die Ausgangsintervalle zu verkleinern um sicherzustellen, daß w die einzige Nullstelle im Intervall ist.

Fehlen schließlich noch die Quotienten, und dazu reicht es, wenn wir zu einer gegebenen reellen algebraischen Zahl u ein Polynom und ein

Intervall für ihren Kehrwert finden. Ist u eine Nullstelle von f, so ist $1/u$ offensichtlich Nullstelle von

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

Beim Intervall $[a,\ b]$ für u müssen wir zunächst sicherstellen, daß es die Null nicht enthält, daß also a und b dasselbe Vorzeichen haben; dann liegt $1/u$ im Intervall $[1/b,\ 1/a]$.

Damit lassen sich alle vier Grundrechenarten algorithmisch ausführen, und wir haben auch gezeigt, daß die reellen algebraischen Zahlen einen Körper bilden.

Als Beispiel wollen wir Summe und Produkt von $\sqrt{2}$ und $\sqrt{3}$ auf diese Weise behandeln. $\sqrt{2}$ ist Nullstelle des irreduziblen Polynoms $f = x^2 - 2$ und liegt im Intervall $[0,\ 2]$, entsprechend ist $\sqrt{3}$ Nullstelle des Polynoms $g = X^2 - 3$ und liegt im Intervall $[0,\ 3]$.

$$g(Z - X) = (Z - X)^2 - 3 = X^2 - 2ZX + Z^2 - 3\,;$$

wir müssen also die Resultante

$$\begin{vmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & -2 \\ 1 & -2Z & Z^2 - 3 & 0 \\ 0 & 1 & -2Z & Z^2 - 3 \end{vmatrix} = Z^4 - 10Z^2 + 1$$

von f und diesem Polynom berechnen.

Da f auch die Nullstelle $-\sqrt{2}$ hat und g entsprechend die Nullstelle $-\sqrt{3}$, sind alle vier Zahlen $\pm\sqrt{2} \pm \sqrt{3}$ Nullstellen dieser Resultante; da sie Grad vier hat kennen wir somit ihre sämtlichen Nullstellen.

Die Summe einer Zahl aus $[0,\ 2]$ und einer aus $[0,\ 3]$ liegt in $[0,\ 5]$; somit ist $\sqrt{2} + \sqrt{3}$ somit eine Nullstelle des Polynoms $X^4 - 10X^2 + 1$ aus diesem Intervall; wir müssen überprüfen, ob es die einzige ist.

Da wir alle vier Nullstellen kennen, brauchen wir dazu keinen Satz von STURM, sondern sehen auch so, daß die Nullstelle $\sqrt{3} - \sqrt{2}$ ebenfalls in diesem Intervall liegt. Wir müssen wir die Ausgangsintervalle also verkleinern.

Da $1^2 < 2 < 3 < 2^2$, liegen sowohl $\sqrt{2}$ als auch $\sqrt{3}$ im Intervall $[1,\,2]$; ihre Summe liegt daher in $[2,\,4]$. In diesem Intervall liegt keine weitere Nullstelle, denn $\sqrt{3}-\sqrt{2} \in [0,\,1]$. Somit können wir $\sqrt{2}+\sqrt{3}$ charakterisieren als *die* Nullstelle von X^4-10X^2+1 im Intervall $[2,\,4]$.

Für das Produkt $\sqrt{2}\cdot\sqrt{3} = \sqrt{6}$ müssen wir, wenn wir strikt nach Schema vorgehen, zunächst das Polynom

$$X^2 g(Z/X) = X^2\left(\frac{Z^2}{X^2}-3\right) = Z^2-3X^2$$

berechnen und seine Resultante mit f berechnen:

$$\begin{vmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & -2 \\ -3 & 0 & Z^2 & 0 \\ 0 & -3 & 0 & Z^2 \end{vmatrix} = (z^2-6)^2\,.$$

Dieses Polynom hat nur die Nullstellen $\pm\sqrt{6}$, wir können das Produkt von $\sqrt{2}$ und $\sqrt{3}$ also charakterisieren als die Nullstelle von $(Z^2-6)^2$ in $[0,\,6]$. Natürlich reicht auch Z^2-6; die Resultante liefert offensichtlich nicht immer das kleinstmögliche Ergebnis. Glücklicherweise erlaubt uns die Computeralgebra, jedes Polynom über $\mathbb{Q}$ zu faktorisieren, und wir können dann nachprüfen, welcher irreduzible Faktor für die betrachtete Zahl verschwindet.

Wir müssen uns noch überlegen, daß wir auch entscheiden können, wann zwei Darstellungen $\big(f,[a,\,b]\big)$ und $\big(g,[c,\,d]\big)$ die gleiche Zahl darstellen. Wenn der Durchschnitt der beiden Intervalle leer ist, kann das unmöglich der Fall sein, ebenso wenig, wenn f und g teilerfremd sind, denn dann gibt es keine gemeinsame Nullstelle.

Falls ggT(f,g) positiven Grad hat und $[a,\,b]\cap[c,\,d]$ nicht leer ist. können wir zunächst (z.B. nach STURM) überprüfen, ob der ggT im Durchschnitt der beiden Intervalle eine Nullstelle hat. Falls nein, können die Ausgangszahlen nicht gleich sein.

Andernfalls ist diese Nullstelle auch eine Nullstelle von f, und sie liegt insbesondere im Intervall $[a,\,b]$. Da f dort nur die eine Nullstelle x hat,

muß sie also gleich x sein. Genauso folgt, daß sie gleich y sein muß, und das zeigt die Gleichheit von x und y.

Die Relationen $<$ und $>$ lassen sich ebenfalls leicht entscheiden, z.B. durch Verkleinerung der Intervalle oder durch Berechnung der Differenz und Untersuchung auf positive oder negative Nullstellen von deren definierendem Polynom.

Damit haben wir einen Teilkörper von $\mathbb{R}$ gefunden, in dem wir (wenn auch mit relativ großem Aufwand) exakt rechnen und Relationen entscheiden können.

Kapitel 5
Die Abhängigkeit der Nullstellen von den Koeffizienten

§1: Symmetrische Potenzen

Ein reelles oder komplexes Polynom vom Grad n in einer Veränderlichen hat nach dem Fundamentalsatz der Algebra mit Vielfachheiten gezählt n Nullstellen. Diese können wir nicht in natürlicher Weise als n-tupel betrachten, denn es gibt keine natürliche Reihenfolge der Nullstellen. Wir können auch nicht einfach die *Menge* der Nullstellen betrachten, denn dann gehen die Vielfachheiten verloren. Was wir brauchen, sind n-tupel modulo Permutationen oder, äquivalent, Mengen, deren Elemente noch eine Vielfachheit haben. Das bieten uns die symmetrischen Potenzen:

Definition: X sei eine Menge. Die n-te symmetrische Potenz $X^{(n)}$ von X ist das kartesische Produkt $X^n = X \times \cdots \times X$ modulo folgender Äquivalenzrelation: $(x_1, \ldots, x_n) \sim (y_1, \ldots, y_n)$ genau dann, wenn es eine Permutation $\pi \in \mathfrak{S}_n$ gibt, so daß $y_i = x_{\pi(i)}$ für alle i. Dabei bezeichnet $\mathfrak{S}_n$, wie im vorigen Kapitel, die symmetrische Gruppe aller Permutationen der Menge $\{1, \ldots, n\}$.

Wir bezeichnen die Äquivalenzklasse von $(x_1, \ldots, x_n)$ mit $[x_1, \ldots, x_n]$; falls einige der x_i mehrfach vorkommen, schreiben wir auch kurz $[m_1 x_1 \ldots, m_r x_r]$ für die Klasse eines n-tupels, in dem x_i mit Vielfachheit m_i auftritt. Dabei muß die Summe der m_i natürlich gleich n hat.

Das System der Nullstellen eines Polynoms vom Grad n über $\mathbb{C}$ ist somit in natürlicher Weise ein Element von $\mathbb{C}^{(n)}$, wir haben also eine

Abbildung $h\colon \mathbb{C}^n \to \mathbb{C}^{(n)}$, die jedem Vektor $a = (a_0, \ldots, a_{n-1})$ aus $\mathbb{C}^n$ das System der Nullstellen des Polynoms

$$P_a(X) = X^n + a_{n-1}X^{n-1} + a_{n-2}X^{n-2} + \cdots + a_1X + a_0$$

zuordnet. Diese Abbildung ist bijektiv, denn ein Polynom vom Grad n mit höchstem Koeffizienten eins ist sowohl durch seine übrigen Koeffizienten als auch durch seine Nullstellen eindeutig bestimmt.

Wenn wir uns fragen, ob die Nullstellen in stetiger Weise von den Koeffizienten abhängen, haben wir das Problem, daß wir Stetigkeit nur kennen für Funktionen zwischen Teilmengen von Räumen $\mathbb{R}^n$ oder $\mathbb{C}^n$; die symmetrische Potenz $\mathbb{C}^{(n)}$ ist aber keine Teilmenge eines $\mathbb{C}^m$, so daß wir Stetigkeit hier erst definieren müssen. Die notwendigen Begriffe stellt der nächste Paragraph zusammen.

§2: Topologische Grundbegriffe

In der Analysis wird Stetigkeit meist so definiert, daß es zu jedem $\varepsilon > 0$ ein $\delta > 0$ geben muß, so daß $\|f(x) - f(y)\| < \varepsilon$, wann immer $\|x - y\| < \delta$ ist. Für diese Definition brauchen wir Differenzen und eine Norm, was wir in vielen Fällen nicht haben. Äquivalent läßt sich die Stetigkeit einer Abbildung aber auch so charakterisieren, daß das Urbild jeder offenen Menge offen ist. Dazu brauchen wir nur den Begriff der offenen Menge.

Ein topologischer Raum ist eine Menge, von deren Teilmengen einige als offene Mengen ausgezeichnet sind derart, daß die vom $\mathbb{R}^n$ gewohnten Eigenschaften offener Mengen gelten. Formal heißt das

Definition: *a)* Ein *topologischer Raum* X ist eine Menge X zusammen mit einer Menge $\mathcal{T}$ von Teilmengen von X, genannt die *Topologie* von X, für die gilt:

1.) $X \in \mathcal{T}$ und $\emptyset \in \mathcal{T}$
2.) Ist I eine beliebige Indexmenge und ist $U_i \in \mathcal{T}$ für alle $i \in I$, so liegt auch die Vereinigung $\bigcup_{i \in I} U_i$ der U_i in $\mathcal{T}$.

3.) Sind für eine natürliche Zahl r die Mengen $U_1, \ldots, U_r$ Elemente von $\mathcal{T}$, so auch ihr Durchschnitt $\bigcap_{i=1}^{r} U_i$.

b) Die Elemente von $\mathcal{T}$ heißen *offene Mengen.*
c) Eine Teilmenge $A \subseteq X$ heißt *abgeschlossen*, wenn $X \setminus A$ offen ist.
d) Eine Abbildung $f\colon X \to Y$ zwischen zwei topologischen Räumen X und Y heißt *stetig*, wenn für jede offene Teilmenge $U \subset Y$ auch die Urbildmenge $f^{-1}(U)$ offen in X ist.
e) Eine stetige Abbildung $f\colon X \to Y$ heißt *Homöomorphismus*, wenn es eine stetige Abbildung $g\colon Y \to X$ gibt, so daß $f \circ g$ die Identität auf Y ist und $g \circ f$ die Identität auf X. (f und g sind dann natürlich bijektiv.)
f) Zwei topologische Räume heißen *homöomorph*, wenn es einen Homöomorphismus $f\colon X \to Y$ gibt.

Es ist klar, daß der $\mathbb{R}^n$ ein topologischer Raum ist, wenn wir für $\mathcal{T}$ die Menge aller (im Sinne der Analysis) offenen Teilmengen nehmen; genauso ist auch $\mathbb{C}^n$ ein topologischer Raum, wenn wir ihn einfach mit $\mathbb{R}^{2n}$ identifizieren.

Um auch die symmetrische Potenz $\mathbb{C}^{(n)}$ und Teilmengen von $\mathbb{C}^n$ zu topologischen Räumen zu machen, definieren wir:

Definition: *a)* Z sei eine Teilmenge eines topologischen Raums X. Eine Teilmenge $U \subset Z$ ist genau dann offen (in der sogenannten *Spurtopologie*), wenn es eine offene Teilmenge $V \subset X$ gibt, so daß $U = V \cap Z$ ist.
b) X sei ein topologischer Raum und $\pi\colon X \to Y$ sei eine surjektive Abbildung. Eine Teilmenge $U \subseteq Y$ ist genau dann offen (in der sogenannten *Quotiententopologie*), wenn das Urbild $\pi^{-1}(U)$ offen in X ist.

Da Durchschnitte und Vereinigungen sowohl beim Schneiden mit Z als auch bei Anwendung von π^{-1} erhalten bleiben, $Z \cap \emptyset$ sowie $\pi^{-1}(\emptyset)$ beide leer sind und $Z \cap X = Z$ und $\pi^{-1}(Y) = X$ ist, definiert dies in der Tat Topologien auf Z beziehungsweise Y.

Wir versehen somit $\mathbb{C}^{(n)}$ mit der Quotiententopologie bezüglich der

natürlichen Projektion

$$\pi\colon \begin{cases} \mathbb{C}^n \to \mathbb{C}^{(n)} \\ (x_1, \ldots, x_n) \mapsto [x_1, \ldots, x_n] \end{cases},$$

die jedem n-tupel seine Äquivalenzklasse unter Permutation der Einträge zuordnet. Wir wollen uns im nächsten Paragraphen überlegen, daß bezüglich dieser Topologie die Abbildung $h\colon \mathbb{C}^n \to \mathbb{C}^{(n)}$ ein Homöomorphismus ist.

Wir werden dort ziemlich einfach zeigen können, daß die Umkehrabbildung zu h stetig ist; jedoch muß die Umkehrabbildung einer bijektiven stetigen Abbildung nicht stetig sein muß. Als Beispiel betrachten wir die Einheitskreisline

$$S^1 = \left\{ z \in \mathbb{C} \;\middle|\; |z| = 1 \right\},$$

die wir mit der Quotiententopologie bezüglich der Abbildung

$$\varphi\colon \begin{cases} \mathbb{R} \to \mathbb{C} \\ t \mapsto e^{it} \end{cases}$$

versehen; wir legen also wieder fest, daß eine Teilmenge $U \subseteq S^1$ genau dann offen sein soll, wenn $\varphi^{-1}(U)$ offen in $\mathbb{R}$ ist.

Wenn wir das halboffene Intervall $[0,\, 2\pi)$ mit seiner Spurtopologie als Teilmenge von $\mathbb{R}$ versehen, ist die Einschränkung von φ auf $[0,\, 2\pi)$ bijektiv und wegen der Stetigkeit der Exponentialfunktion auch stetig. Trotzdem ist die Umkehrabbildung $S^1 \to [0,\, 2\pi)$ nicht stetig, denn $[0,\, \pi/2) = (-\pi/2,\, \pi/2) \cap [0,\, 2\pi)$ ist eine offene Menge, deren Urbild der Viertelkreis zwischen 1 und i ist, wobei 1 im Urbild liegt. Damit ist dieses Urbild keine offene Menge, denn jede offene Teilmenge von S^1, die 1 enthält, muß auch Punkte mit negativem Imaginärteil enthalten.

Da wir die Abbildung h im Gegensatz zu ihrer Umkehrabbildung nicht explizit kennen, können wir ihre Stetigkeit nur über die Umkehrabbildung beweisen. Zum Glück ist zumindest für kompakte Räume die Umkehrabbildung einer bijektiven stetigen Abbildung wieder stetig.

Definition: *a)* Ein topologischer Raum X heißt HAUSDORFFsch, wenn es zu je zwei Punkten $x \neq y$ aus X offene Mengen $U, V \subseteq X$ gibt mit

$x \in U$, $y \in V$ und $U \cap V = \emptyset$.
b) Ein System $\mathfrak{U} = \{U_i \mid i \in I\}$ von offenen Teilmengen $U_i \subset X$, wobei I eine beliebige Indexmenge bezeichnet, heißt *offene Überdeckung* der Teilmenge $Z \subseteq X$, wenn

$$Z \subseteq \bigcup_{i \in I} U_i$$

in der Vereinigungsmenge aller U_i liegt.
c) Ist $J \subseteq I$ eine Teilmenge von I und liegt Z bereits in der Vereinigung aller U_i mit $i \in J$, bezeichnen wir $\mathfrak{V} = \{U_i \mid i \in J\}$ als eine *Teilüberdeckung* von $\mathfrak{U}$. Ist speziell J eine endliche Menge, so sprechen wir von einer *endlichen Teilüberdeckung.*
d) Ein topologischer Raum X heißt *kompakt,* wenn er HAUSDORFFsch ist und wenn jede offene Überdeckung $\mathfrak{U}$ von X eine endliche Teilüberdekkung hat.

Lemma: Jede kompakte Teilmenge eines HAUSDORFFschen topologischen Raums X ist abgeschlossen. Ist X kompakt, so ist umgekehrt auch jede abgeschlossene Teilmenge von X kompakt.

Beweis: $Z \subseteq X$ sei kompakt; wir wollen zeigen, daß dann $X \setminus Z$ offen ist. Dazu sei y ein beliebiger Punkt von $X \setminus Z$ und z ein beliebiger Punkt von Z. Da X HAUSDORFFsch ist, gibt es dazu zwei offene Mengen $U_z, V_z \subseteq X$ derart, daß $z \in U_z$, $y \in V_z$ und $U_z \cap V_z = \emptyset$ ist. Die sämtlichen Mengen U_z bilden natürlich eine offene Überdeckung von Z; wegen der Kompaktheit von Z hat diese eine endliche Teilüberdeckung $U_{z_1}, \ldots, U_{z_r}$. Wir betrachten den Durchschnitt V der zugehörigen offenen Mengen V_{z_i}. Da $U_{z_i} \cap V_{z_i} = \emptyset$ ist, ist insbesondere $U_{z_i} \cap V = \emptyset$ für alle U_{z_i}, d.h. $Z \cap V = \emptyset$. Wir haben damit gesehen, daß es zu jedem Punkt $y \in X \setminus Z$ eine offene Menge V gibt, die ganz in $X \setminus Z$ liegt. Damit ist $X \setminus Z$ als Vereinigung aller dieser offener Mengen offen, Z also abgeschlossen.

Umgekehrt sei X kompakt und $Z \subseteq X$ abgeschlossen. Wir wollen zeigen, daß X kompakt ist.

Dazu sei $\mathfrak{U} = \{U_i \mid i \in I\}$ eine offene Überdeckung von X und V sei das Komplement von Z in X. Wegen der Abgeschlossenheit von Z ist

V offen, und da die U_i bereits Z überdecken, überdecken sie zusammen mit V ganz Z. Wegen der Kompaktheit von X reichen dazu bereits endlich viele Mengen $U_{i_1}, \ldots, U_{i_r}$, eventuell zusammen mit V. Da kein Punkt von Z in $V = X \setminus Z$ liegt, bilden $U_{i_1}, \ldots, U_{i_r}$ eine endliche Teilüberdeckung von $\mathfrak{U}$.

∎

Das Bild einer abgeschlossenen Menge unter einer stetigen Abbildung muß bekanntlich im allgemeinen nicht abgeschlossen sein: Das Bild der Hyperbel $xy = 1$ unter der Projektion auf die x-Achse ist beispielsweise die offene Menge $\mathbb{R} \setminus \{0\}$. Für kompakte Mengen gilt aber

Lemma: Ist $f\colon X \to Y$ eine stetige Abbildung und ist $Z \subseteq X$ kompakt, so ist auch $f(Z) \subseteq Y$ kompakt.

Beweis: Ist $\mathfrak{U} = \{U_i \mid i \in I\}$ eine offene Überdeckung von $f(Z)$, so bilden die Urbildmengen $f^{-1}(U_i)$ eine offene Überdeckung von Z. Wegen der Kompaktheit von Z hat diese eine endliche Teilüberdeckung; diese bestehe etwa aus den Mengen $f^{-1}(U_{i_1})$ bis $f^{-1}(U_{i_r})$. Da Z in der Vereinigung dieser Urbilder enthalten ist, liegt $f(Z)$ in der Vereinigung der Mengen U_{i_1} bis U_{i_r}, die somit eine endliche Teilüberdeckung von $\mathfrak{U}$ bilden.

∎

Lemma: Ist $f\colon X \to Y$ bijektiv und stetig, und ist X kompakt, so ist f ein Homöomorphismus.

Beweis: $g\colon Y \to X$ sei die Umkehrabbildung von f; wir müssen zeigen, daß g stetig ist. Sei also $U \subseteq X$ eine offene Menge. Ihr Urbild unter g ist $g^{-1}(U) = f(U)$; wir müssen zeigen, daß diese Menge offen ist.

Das Komplement $Z = X \setminus U$ von U ist abgeschlossen im kompakten topologischen Raum X, also selbst kompakt. Damit ist auch $f(Z) \subseteq Y$ kompakt, also abgeschlossen. Wegen der Bijektivität von f ist das Bild $f(U) = Y \setminus f(Z)$; also ist $f(U)$ offen.

∎

Die uns interessierenden Räume $\mathbb{C}^n$ und $\mathbb{C}^{(n)}$ sind beide nicht kompakt; trotzdem ist das gerade bewiesene Lemma auch in solchen Situationen gelegentlich nützlich. Es gilt nämlich

Lemma: $f: X \to Y$ sei eine bijektive stetige Abbildung zwischen zwei topologischen Räumen und in X gebe es eine Folge kompakter Teilmengen $X_1 \subseteq X_2 \subseteq X_3 \subseteq \cdots$, deren Vereinigung ganz X sei. Falls jeder Punkt $y \in Y$ in einer offenen Menge V liegt, die ganz in einer der Mengen $f(X_k)$ liegt, ist f ein Homöomorphismus.

Beweis: Wie beim vorigen Lemma müssen wir zeigen, daß das Bild $f(U)$ einer jeden offenen Teilmenge $U \subset X$ offen in Y ist. Da die Teilmengen X_k kompakt sind, sind auch ihre Bilder $Y_k = f(X_k)$ kompakt und die Einschränkung f_k von f zu einer Abbildung $f_k: X_k \to Y_k$ ist nach dem vorigen Lemma ein Homöomorphismus, d.h. $f(U \cap X_k)$ ist für jedes k offen in Y_k. Es gibt daher eine offene Teilmenge $W_k \subset Y$, so daß $f(U \cap X_k) = W_k \cap Y_k$ ist.

Sei nun $x \in U$ ein beliebiger Punkt aus U. Nach Voraussetzung gibt es eine offene Menge $V \subseteq Y$, die $f(x)$ enthält und ganz in einer der Mengen Y_k liegt. Damit liegt $x \in X_k$ und $V \cap W_k$ ist eine offene Teilmenge von Y, die ganz in $W_k \cap Y_k = f(U \cap X_k) \subseteq f(U)$ liegt.

Damit haben wir gezeigt, daß es zu jedem Punkt $x \in U$ eine offene Teilmenge von Y gibt, die ganz in $f(U)$ liegt. $f(U)$ ist natürlich die Vereinigung aller dieser Teilmengen und somit offen.

∎

Man beachte, daß die zweite Voraussetzung des Lemmas hier wichtig ist: Für die Abbildung $f: [0, 2\pi) \to S^1$, die den Punkt t auf e^{it} abbildet, sind die Mengen $X_k = [0, 2\pi - \frac{1}{k})$ kompakt, und ihre Vereinigung ist ganz $[0, 2\pi)$; nur weil $1 \in S^1$ keine offene Umgebung hat, die ganz in einem der $f(X_k)$ liegt, können wir das obige Lemma nicht anwenden und fälschlicherweise schließen, daß f ein Homöomorphismus sei.

§3: Stetigkeit der Nullstellen

Wir identifizieren $\mathbb{C}^n$ mit der Menge aller normierter Polynome vom Grad n, indem wir den Punkt $a = (a_0, \ldots, a_{n-1}) \in \mathbb{C}^n$ identifizieren mit dem Polynom

$$f_a = X^n + a_{n-1}X^{n-1} + a_{n-2}X^{n-2} + \cdots + a_1X + a_0 \in \mathbb{C}[X]$$

und bekommen dann eine Abbildung $\widetilde{g}\colon \mathbb{C}^n \to \mathbb{C}^n$, die jedem n-tupel $z = (z_1, \dots z_n) \in \mathbb{C}^n$ das Polynom

$$\begin{aligned}\widetilde{g}(z) &= (X - z_1)(X - z_2)\cdots(X - z_n)\\ &= X^n - \sigma_1(z)X^{n-1} + \cdots + (-1)^{n-1}\sigma_{n-1}(z)X + (-1)^n\sigma_n(z)\end{aligned}$$

zuordnet. Da die elementarsymmetrischen Funktionen $\sigma_j(z)$ Polynome in den z_i sind, ist dies eine stetige Abbildung.

Das Bild $g(z)$ hängt nur ab von der Äquivalenzklasse $[z_1, \dots, z_r]$ des Tupels $z = (z_1, \dots, z_n)$; daher induziert $\widetilde{g}$ eine Abbildung $g\colon \mathbb{C}^n \to \mathbb{C}^{(n)}$, die jeder Äquivalenzklasse aus $\mathbb{C}^{(n)}$ das normierte Polynom mit genau diesen Nullstellen zuordnet.

Auch g ist eine stetige Abbildung: Für die offene Teilmenge $U \subset \mathbb{C}^n$ sei $V = g^{-1}(U)$ das Urbild in $\mathbb{C}^{(n)}$. Nach unserer Definition der Topologie auf $\mathbb{C}^{(n)}$ ist V genau dann offen, wenn $\pi^{-1}(V) \subset \mathbb{C}^n$ offen ist. Da $\widetilde{g} = g \circ \pi$ ist, ist

$$\pi^{-1}(V) = \pi^{-1}\big(g^{-1}(U)\big)\big)(g \circ \pi)^{-1}(U) = \widetilde{g}^{-1}(U)\,,$$

und das ist wegen der Stetigkeit von $\widetilde{g}$ offen.

Wir interessieren uns für die Umkehrabbildung $h\colon \mathbb{C}^n \to \mathbb{C}^{(n)}$, die jedem Punkt $a \in \mathbb{C}^n$ die Nullstellen des normierten Polynoms mit Koeffizienten a zuordnet und wollen zeigen

Satz: $h\colon \mathbb{C}^n \to \mathbb{C}^{(n)}$ ist ein Homöomorphismus.

Beweis: Der Fall $n = 1$ ist trivial: Hier ist $\mathbb{C}^{(1)} = \mathbb{C}$, und h ist einfach die Abbildung $z \mapsto -z$, denn das normierte lineare Polynom mit Nullstelle z ist $X - z$. Sei daher im folgenden $n \geq 2$.

Wir wissen bereits, daß die Umkehrabbildung g stetig und bijektiv ist; daher wollen wir das letzte Lemma des vorigen Paragraphen anwenden.

Wir definieren dazu in $\mathbb{C}^n$ für jedes $k \in \mathbb{N}$ die kompakte Teilmenge

$$X_k = \left\{ z = [z_1, \dots, z_n] \in \mathbb{C}^{(n)} \;\middle|\; |z_i| \leq nk \;\forall k \right\};$$

offensichtlich ist die Vereinigung aller X_k ganz $\mathbb{C}^{(n)}$. Um den Satz aus dem Lemma folgern zu können, müssen wir noch zeigen, daß jeder

Punkt aus $\mathbb{C}^n$ eine offene Umgebung hat, die ganz in einer der Mengen $g(X_k)$ liegt.

Sei also $a = (a_0, \ldots, a_{n-1}) \in \mathbb{C}^n$ ein beliebiger Punkt und k eine natürliche Zahl, die größer ist als der Betrag eines jeden a_j. Dann ist

$$V_k = \left\{ b = (b_0, \ldots, b_{n-1}) \in \mathbb{C}^n \;\middle|\; |b_j| < k \;\forall j \right\}$$

eine offene Umgebung von a, und wir wollen uns überlegen, daß sogar deren Abschluß ganz in $g(X_k)$ liegt.

Wir betrachten also ein Tupel $b = (b_0, \ldots, b_{n-1}) \in \overline{V}_k$, d.h. $|b_j| \leq k$ für alle j. Weiter sei $h(b) = [z_1, \ldots, z_n] \in \mathbb{C}^{(n)}$. Wir müssen zeigen, daß alle z_i einen Betrag von höchstens k haben. Da k eine natürliche Zahl ist, gilt dies natürlich für alle z_i vom Betrag kleiner eins. Wir betrachten daher im folgenden nur Nullstellen z_i mit $|z_i| \geq 1$. Wie für jede Nullstelle ist $f_b(z_i) = z_i^n + b_{n-1}z_i^{n-1} + \cdots + b_1 z_i + b_0 = 0$, also

$$\left|z_i^n\right| = \left|b_{n-1}z_i^{n-1} + \cdots + b_1 z_i + b_0\right| \leq k\left|z_i^{n-1} + \cdots z_i + 1\right| \leq nk|z_i|^{n-1}$$

denn wegen $|z_i| \geq 1$ ist $\left|z_i^j\right| \leq \left|z_i^{n-1}\right|$ für alle $j \leq n-1$. Somit ist $|z_i^n| \leq nk\left|z_i^{n-1}\right|$; indem wir durch $\left|z_i^{n-1}\right|$ kürzen, erhalten wir die gewünschte Abschätzung $|z_i| \leq nk$.

■