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

Die Alternative bietet ein für völlig andere Zwecke bewiesenes Resultat des deutschen Zahlentheoretikers HENSEL, das es erlaubt eine Faktorisierung modulo p fortzusetzen zu einer Faktorisierung modulo jeder beliebiger p-Potenz und, was HENSEL wirklich interessierte, zu den sogenannten p-adischen Zahlen, mit denen wir uns in Rahmen dieser Vorlesung allerdings nicht beschäftigen werden.

§5: Das Henselsche Lemma

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

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

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

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

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

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

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

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

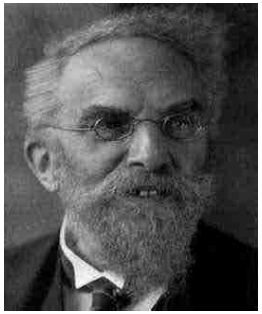
oder

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

Division durch p^n macht daraus

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

denn $g_n \equiv g \bmod p$ und $h_n \equiv h \bmod p$. Die letztere Kongruenz können wir als Gleichung in $\mathbb{F}_p[x]$ auffassen und dort lösen, indem wir den erweiterten EUKLIDischen Algorithmus auf die Polynome $g \bmod p$ und $h \bmod p$ aus $\mathbb{F}_p[x]$ anwenden: Da diese nach Voraussetzung teilerfremd sind, können wir ihren ggT Eins und damit auch jedes andere Polynom über $\mathbb{F}_p$ als Linearkombination der beiden darstellen. Da der Grad von f die Summe der Grade von g und h ist und f^* höchstens denselben Grad wie f hat, können wir dann auch eine Darstellung $f^* = gh^* + hg^*$ in $\mathbb{F}_p[x]$ finden mit $\deg g^* \leq \deg g$ und $\deg h^* \leq \deg h$. Ersetzen wir g^* und h^* durch irgendwelche Repräsentantanten gleichen Grades aus $\mathbb{Z}[x]$, erfüllen $g_{n+1} = g_n + p^n g^*$ und $h_{n+1} = h_n + p^n h^*$ die Kongruenz $f \equiv g_n h_n$ modulo p^{n+1}. ∎



KURT HENSEL wurde 1861 im damaligen Königsberg geboren; als er neun Jahre alt war, zog die Familie nach Berlin. Er studierte dort und in Bonn; 1884 promovierte er in Berlin bei KRONECKER, 1886 folgte die Habilitation. Er blieb bis 1901 als Privatdozent in Berlin; 1901 bekam er einen Lehrstuhl in Marburg, auf dem er bis zu seiner Emeritierung 1930 blieb. Er starb 1941 in Marburg. Seine Arbeiten drehen sich hauptsächlich um die Zahlentheorie und die eng damit verwandte Arithmetik von Funktionenkörpern bekannt wurde er vor allem durch die Einführung der p-adischen Zahlen. Er ist Autor dreier Lehrbücher.

§6: Der Algorithmus von Zassenhaus

Die Werkzeuge aus den vorigen Paragraphen erlauben uns, gemeinsam eingesetzt, nun die Faktorisierung von Polynomen f aus $\mathbb{Z}[x]$ oder $\mathbb{Q}[x]$. Das einzige, was wir noch nicht explizit formuliert haben, ist eine Schranke für die Koeffizienten eines Faktors. Aus Kapitel 2, §8, wissen wir, daß für einen Teiler $g \in \mathbb{C}[z]$ eines Polynoms $f \in \mathbb{C}[z]$ gilt:

$$H(g) \leq \binom{e}{[e/2]} \left| \frac{b_e}{a_d} \right| \|f\|_2 ,$$

wobei e den Grad von g bezeichnet und a_d, b_e die führenden Koeffizienten von f und g. Für $g, f \in \mathbb{Z}[x]$ muß b_e ein Teiler von a_d sein, der Quotient b_e/a_d hat also höchstens den Betrag eins. Der Grad e eines Teilers kann höchstens gleich dem Grad d von f sein, also ist für jeden Teiler $g \in \mathbb{Z}[x]$ von $f \in \mathbb{Z}[x]$

$$H(g) \leq \binom{d}{[d/2]} \|f\|_2 \ .$$

Nach ZASSENHAUS gehen wir zur Faktorisierung eines Polynoms f aus $\mathbb{Z}[x]$ oder $\mathbb{Q}[x]$ nun folgendermaßen vor:

Erster Schritt: Berechne die quadratfreie Zerlegung von f und ersetze die quadratfreien Faktoren durch ihre primitiven Anteile g_i. Dann gibt es eine Konstante c, so daß $f = c\prod_{i=1}^{r} g_i^i$ ist. Falls f in $\mathbb{Z}[x]$ liegt, ist c eine ganze Zahl. Für eine Faktorisierung in $\mathbb{Z}[x]$ muß auch c in seine Primfaktoren zerlegt werden; für eine Faktorisierung in $\mathbb{Q}[x]$ kann c als Einheit aus $\mathbb{Q}^\times$ stehen bleiben. Die folgenden Schritte werden einzeln auf jedes der g_i angewandt, danach werden die Ergebnisse zusammengesetzt zur Faktorisierung von f. Für das Folgende sei g eines der g_i.

Zweiter Schritt: Wir setzen $M = \binom{\deg g}{[\frac{1}{2}\deg g]} \|g\|_2$ und $M = 2L + 1$. Dann wählen wir eine Primzahl p, die weder den führenden Koeffizienten noch die Diskriminante von g teilt.

Dritter Schritt: Wir faktorisiere g mod p nach BERLEKAMP in $\mathbb{F}_p[x]$.

Vierter Schritt: Die Faktorisierung wird nach dem HENSELschen Lemma hochgehoben zu einer Faktorisierung modulo p^n für eine natürliche Zahl n mit $p^n \geq M$.

Fünfter Schritt: Setze $m = 1$ und teste für jeden der gefundenen Faktoren, ob er ein Teiler von g ist. Falls ja, kommt der Faktor in die Liste $\mathcal{L}_1$ der Faktoren von g; andernfalls kommt es in eine Liste $\mathcal{L}_2$.

Sechster Schritt: Falls die Liste $\mathcal{L}_2$ keine Einträge hat, endet der Algorithmus und g ist das Produkt der Faktoren aus $\mathcal{L}_1$. Andernfalls setzen wir $m = m + 1$ und testen für jedes Produkt aus m verschiedenen Polynomen aus $\mathcal{L}_2$, ob ihr Produkt modulo p^n (mit Koeffizienten vom Betrag

höchstens L) ein Teiler von g ist. Falls ja, entfernen wir die m Faktoren aus $\mathcal{L}_2$ und fügen ihr Produkt in die Liste $\mathcal{L}_1$ ein. Wiederhole diesen Schritt.

Auch wenn der sechste Schritt wie eine Endlosschleife aussieht, endet der Algorithmus natürlich nach endlich vielen Schritten, denn $\mathcal{L}_2$ ist eine endliche Liste und spätestens das Produkt aller Elemente aus $\mathcal{L}_2$ muß Teiler von g sein, da sein Produkt mit dem Produkt aller Elemente von $\mathcal{L}_1$ gleich g ist. Tatsächlich kann man schon abbrechen, wenn die betrachteten Faktoren einen größeren Grad haben als $\frac{1}{2}\deg g$, denn falls f reduzibel ist, gibt es einen Faktor, der höchstens diesen Grad hat.

HANS JULIUS ZASSENHAUS wurde 1912 in Koblenz geboren, ging aber in Hamburg zur Schule und zur Universität. Er promovierte 1934 mit einer Arbeit über Permutationsgruppen und habilitierte 1940 über LIE-Ringe in positiver Charakteristik. Da er nicht der NSdAP beitreten wollte, arbeitete er während des Krieges als Meteorologe bei der Marine; nach dem Krieg war er von 1949 bis 1959 Professor in Montreal, dann fünf Jahre lang in Notre Dame und schließlich bis zu seiner Emeritierung an der Ohio State University in Columbus. Dort starb er 1991. Bekannt ist er vor allem für seine Arbeiten zur Gruppentheorie und zur algorithmischen Zahlentheorie.

§ 7: Berechnung von Resultanten und Diskriminanten

Im zweiten Schritt des Algorithmus von ZASSENHAUS wählten wir eine Primzahl, die kein Teiler der Diskriminante des zu faktorisierenden Polynoms sein darf. Um dies zu testen, müssen wir die Diskriminante berechnen oder – was äquivalent ist – die Resultante des Polynoms und seiner Ableitung. Für ein Polynom vom Grad zwanzig ist das eine 39×39-Determinante. Nach dem LAPLACEschen Entwicklungssatz ist dies eine Summe von $39! \approx 2 \cdot 10^{47}$ Summanden, die jeweils Produkte von 39 Zahlen sind. Eine solche Summe zu berechnen liegt weit jenseits der Möglichkeiten heutiger Computer.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Andernfalls ist $\mathrm{Res}_x(f,g) = a_n^{m-s}\,\mathrm{Res}_x(f,h)$, und da diese Formel auch für $h = 0$ gilt, haben wir gezeigt

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

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

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

Beweis: Wir müssen in der SYLVESTER-Matrix m Zeilen zu f mit den n Zeilen zu g vertauschen. Dies kann beispielsweise so realisiert werden, daß wir die unterste f-Zeile nacheinander mit jeder der g-Zeilen vertauschen, bis sie nach n Vertauschungen schließlich unten steht. Dies

müssen wir wiederholen, bis alle f-Zeilen unten stehen, wir haben also insgesamt nm Zeilenvertauschungen. Somit ändert sich das Vorzeichen der Determinante um den Faktor $(-1)^{nm}$. ∎

§8: Swinnerton-Dyer Polynome

Der potentiell problematischste Schritt des obigen Algorithmus ist der sechste: Vor allem, wenn wir auch Produkte von mehr als zwei Faktoren betrachten müssen, kann dieser sehr teuer werden. Ein Beispiel dafür bieten die sogenannten SWINNERTON-DYER-Polynome: Zu n paarweise verschiedenen Primzahlen $p_1, \ldots, p_n$ gibt es genau ein Polynom f vom Grad 2^n mit führendem Koeffizienten eins, dessen Nullstellen genau die 2^n Zahlen

$$\pm\sqrt{p_1} \pm \sqrt{p_2} \pm \cdots \pm \sqrt{p_n}$$

sind. Dieses Polynom hat folgende Eigenschaften:

1. Es hat ganzzahlige Koeffizienten.
2. Es ist irreduzibel über $\mathbb{Z}$.
3. Modulo jeder Primzahl p zerfällt es in Faktoren vom Grad höchstens zwei.

Beweisen läßt sich das am besten mit Methoden der abstrakten Algebra, wie sie in jeder Vorlesung *Algebra I* präsentiert werden. Da hier keine Algebra I vorausgesetzt wird, sei nur kurz die Idee angedeutet: Um den kleinsten Teilkörper von $\mathbb{C}$ zu konstruieren, in dem alle Nullstellen von f liegen, können wir folgendermaßen vorgehen: Wir konstruieren als erstes einen Körper K_1, der $\sqrt{p_1}$ enthält. Das ist einfach: Der Vektorraum $K_1 = \mathbb{Q} \oplus \mathbb{Q}\sqrt{p_1}$ ist offensichtlich so ein Körper. Als nächstes konstruieren wir einen Körper K_2, der sowohl K_1 als auch $\sqrt{p_2}$ enthält. Dazu können wir genauso vorgehen: Wir betrachten einfach den zweidimensionalen K_1-Vektorraum $K_2 = K_1 \oplus K_1\sqrt{p_2}$. Als Vektorraum über $\mathbb{Q}$ ist K_2 natürlich vierdimensional. Weiter geht es mit $K_3 = K_2 \oplus K_2\sqrt{p_3}$ *usw.* bis $K_n = K_{n-1} \oplus K_{n-1}\sqrt{p_n}$. Als $\mathbb{Q}$-Vektorraum hat dieser Körper die Dimension 2^n.

Die GALOIS-Gruppe von K_n über $\mathbb{Q}$ hat somit 2^n Elemente; da sie offensichtlich die Abbildungen $\sqrt{p_i} \mapsto -\sqrt{p_i}$ enthält, ist sie die von diesen

Automorphismen erzeugte elementarabelsche Gruppe. Sie läßt die Nullstellenmenge von f als ganzes betrachtet fest, also nach dem Wurzelsatz von VIÈTE auch die Koeffizienten. Somit hat f rationale Koeffizienten, und da alle Nullstellen ganz sind (im Sinne der algebraischen Zahlentheorie), liegen diese sogar in $\mathbb{Z}$. Außerdem operiert die GALOIS-Gruppe transitiv auf der Nullstellenmenge von f; also ist f irreduzibel in $\mathbb{Q}[x]$ und damit auch $\mathbb{Z}[x]$.

Betrachten wir f modulo einer Primzahl p, so können wir die analoge Konstruktion durchführen ausgehend vom Körper $\mathbb{F}_p$ anstelle von $\mathbb{Q}$. Während wir aber im Falle der rationalen Zahlen sicher sein konnten, daß $\sqrt{p_i}$ nicht bereits im Körper K_{i-1} liegt, ist dies hier nicht mehr der Fall: Für ungerades p gibt es $\frac{1}{2}(p+1)$ Quadrate in $\mathbb{F}_p$; dazu könnte auch p_i gehören. Falls nicht, ist $K = \mathbb{F}_p \oplus \mathbb{F}_p\sqrt{p_i}$ ein Körper mit p^2 Elementen. Wie man in der Algebra lernt, gibt es aber bis auf Isomorphie nur einen solchen Körper; K enthält daher die Quadratwurzeln *aller* Elemente von $\mathbb{F}_p$ und somit *alle* Nullstellen von f mod p. Spätestens über K zerfällt f mod p also in Linearfaktoren, und da alle Koeffizienten in $\mathbb{F}_p$ liegen, lassen sich je zwei Linearfaktoren, die nicht in $\mathbb{F}_p[x]$ liegen, zu einem quadratischen Faktor aus $\mathbb{F}_p[x]$ zusammenfassen. Somit hat f mod p höchstens quadratische Faktoren.

(Für eine ausführlichere und etwas elementarere Darstellung siehe etwa §6.3.2 in MICHAEL KAPLAN: Computeralgebra, *Springer,* 2005.)

Falls wir f nach dem oben angegebenen Algorithmus faktorisieren, erhalten wir daher modulo *jeder* Primzahl p mindestens 2^{n-1} Faktoren. Diese lassen sich über das HENSELsche Lemma liften zu Faktoren über $\mathbb{Z}$, und wir müssen alle Kombinationen aus mindestens 2^{n-2} Faktoren ausprobieren bis wir erkennen, daß f irreduzibel ist, also mindestens $2^{2^{n-2}}$ Möglichkeiten. Für $n = 10$ etwa ist f ein Polynom vom Grad 1024, dessen Manipulation durchaus im Rahmen der Möglichkeiten eines heutigen Computeralgebrasystems liegt. Das Ausprobieren von $2^{256} \approx 10^{77}$ Möglichkeiten überfordert aber selbst heutige Supercomputer oder parallel arbeitende Cluster aus Millionen von Computern ganz gewaltig: Der heutige Sicherheitsstandard der Kryptographie geht davon aus, daß niemand in der Lage ist, 2^{128} (oder sogar nur 2^{100}) Rechenoperationen

in realistischer Zeit (d.h. wenigen Jahren) auszuführen.

SIR HENRY PETER FRANCIS SWINNERTON-DYER, 16th Baronet, wurde 1927 geboren; er studierte und lehrte an der Universität von Cambridge, wo er unter anderem Dean des Trinity College, Master von St. Catherine College und Vizekanzler der Universität war; heute ist er Professor emeritus. Obwohl er hauptsächlich für seine Beiträge zur Zahlentheorie bekannt ist, beschäftigte er sich zunächst mit Differentialgleichungen. Am bekanntesten ist er durch die Vermutung von BIRCH und SWINNERTON-DYER über den Zusammenhang zwischen der Arithmetik einer elliptischen Kurve und analytischen Eigenschaften von deren ζ-Funktion.

§9: Faktoren und Gittervektoren

In diesem Paragraphen soll eine Methode vorgestellt werden, die für Polynome einer Veränderlichen über $\mathbb{Z}$ (der $\mathbb{Q}$), aber leider auch nur für diese, eine Alternative zum stumpfsinnigen Ausprobieren im sechsten Schritt des Algorithmus von ZASSENHAUS bietet.

Sie wurde 1982 vorgestellt in

A.K. LENSTRA, H.W. LENSTRA, L. LOVÁSZ: Factoring Polynomials with Rational Coefficients, *Math. Ann.* **261** (1982), 515–534

und wird nach den Initialen der drei Autoren kurz als LLL bezeichnet.

ARJEN K. LENSTRA wurde 1956 in Groningen geboren und studierte Mathematik an der Universität Amsterdam, wo er 1984 über das Thema Faktorisierung von Polynomen promovierte. Danach arbeitete er zunächst als Gastprofessor an der Informatikfakultät der Universität von Chicago, dann ab 1989 in einem Forschungszentrum von Bellcore. 1996 wurde er Vizepräsident am Corporate Technology Office der City Bank in New York, von 2000 bis 2006 auch Teilzeitprofessor an der Technischen Universität Eindhoven. 2004 wechselte er von der City Bank zu Lucent Technology, die ehemaligen Bell Labs; seit 2006 ist er Professor für Kryptologie an der Eidgenössischen Technischen Hochschule in Lausanne. Seine Arbeiten befassen sich mit der Faktorisierung von Zahlen und Polynomen sowie mit kryptographischen Verfahren und Attacken.

Sein Bruder HENDRIK W. LENSTRA wurde 1949 geboren. Auch er studierte an der Universität Amsterdam, wo er 1977 bei dem Algebraischen Geometer und Zahlentheoretiker FRANS OORT über Zahlkörper mit EUKLIDischem Algorithmus promovierte und 1978 Professor wurde. 1987 wechselte er nach Berkeley; von 1998 bis zu seiner Emeritierung in Berkeley lehrte er sowohl in Berkeley als auch an der Universität Leiden, seither nur noch in Leiden. Seine Arbeiten beschäftigen sich hauptsächlich mit der algorithmischen Seite der Zahlentheorie; außer für den LLL-Algorithmus ist er vor allem bekannt für seinen Algorithmus zur Faktorisierung ganzer Zahlen mit elliptischen Kurven.

LÁSZLÓ LOVÁSZ wurde 1948 in Budapest geboren und promovierte 1971 an der dortigen Universität. Nach kürzeren Aufenthalten an verschiedenen ungarischen und ausländischen Universitäten (darunter 1984/85 in Bonn) ging er 1993 nach Yale, wo er bis 2000 eine Professur hatte; von 1999 bis 2006 war er Senior Researcher bei Microsoft Research. Seit 2006 ist er Direktor des mathematischen Instituts der Eötvös Loránd Universität in Budapest. Für die Wahlperiode 2007–2010 ist er auch Präsident der Internationalen Mathematikervereinigung IMU.

Ausgangspunkt der Methode von LENSTRA, LENSTRA und LOVÁSZ ist das folgende

Lemma: p sei eine Primzahl, k eine beliebige natürliche Zahl. Außerdem sei $f \in \mathbb{Z}[x]$ ein Polynom vom Grad d und $h \in \mathbb{Z}[x]$ eines vom Grad e mit folgenden Eigenschaften:

1.) h hat führenden Koeffizienten eins

2.) $h \bmod p^k$ ist in $(\mathbb{Z}/p^k)[x]$ ein Teiler von $f \bmod p^k$

3.) $h \bmod p$ ist irreduzibel in $\mathbb{F}_p[x]$

4.) $(h \bmod p)^2$ ist kein Teiler von $f \bmod p$ in $\mathbb{F}_p[x]$.

Dann gilt:

a) f hat in $\mathbb{Z}[x]$ einen irreduziblen Faktor h_0, der modulo p ein Vielfaches von $h \bmod p$ ist.

b) h_0 ist bis aufs Vorzeichen eindeutig bestimmt.

c) Für einen beliebigen Teiler g von f in $\mathbb{Z}[x]$ sind folgende Aussagen äquivalent:

(i) $h \bmod p$ teilt $g \bmod p$ in $\mathbb{F}_p[x]$

(ii) $h \bmod p^k$ teilt $g \bmod p^k$ in $(\mathbb{Z}/p^k)[x]$

(iii) h_0 teilt g in $\mathbb{Z}[x]$.

Beweis: Die irreduziblen Faktoren h_i von f in $\mathbb{Z}[x]$ können modulo p in $\mathbb{F}_p[x]$ eventuell weiter zerlegt werden. Da $(h \bmod p)^2$ kein Teiler von $f \bmod p$ ist, teilt $h \bmod p$ genau eines der $h_i \bmod p$; wir setzen $h_0 = h_i$. Da irreduzible Faktoren in $\mathbb{Z}[x]$ bis aufs Vorzeichen eindeutig bestimmt sind, ist auch *b)* klar. In *c)* folgt *(i)* sofort aus *(ii)* wie auch aus *(iii)*; zu zeigen ist die Umkehrung.

Sei also $h \bmod p$ ein Teiler von $g \bmod p$ und $f = gq$. Da $(h \bmod p)^2$ kein Teiler von $f \bmod p$ ist, kann $h \bmod p$ kein Teiler von $q \bmod p$ sein,

also auch h_0 kein Teiler von q. Somit muß h_0 als irreduzibler Teiler von f ein Teiler von g sein und *(iii)* ist bewiesen.

Zum Beweis von *(ii)* beachten wir, daß $h \bmod p$ und $q \bmod p$ teilerfremd sind; der erweiterte EUKLIDische Algorithmus liefert uns also eine Darstellung der Eins als Linearkombination dieser beiden Polynome in $\mathbb{F}_p[x]$. Wir liften die Koeffizienten nach $\mathbb{Z}[x]$ und haben somit Polynome $a, b \in \mathbb{Z}[x]$, für die $ah + bq \equiv 1 \bmod p$ ist, d.h. es gibt ein Polynom $c \in \mathbb{Z}[x]$, so daß $ah + bq = 1 - pc$ ist. Da

$$\left(1 - pc\right)\left(1 + pc + (pc)^2 + \cdots + (pc)^{k-1}\right) = 1 - (pc)^k$$

ist, erhalten wir durch Multiplikation dieser Gleichung mit dem zweiten Faktor eine neue Gleichung der Form

$$\tilde{a}h + \tilde{b}q = 1 - (pc)^k \,.$$

Multiplizieren wir diese noch mit g, so folgt

$$\tilde{a}g \cdot h + \tilde{b}q \cdot g = \tilde{a}gh + \tilde{b}f \equiv g \bmod p^k \,.$$

Hier ist die linke Seite modulo p^k durch h teilbar, also auch die rechte Seite g, womit *(ii)* bewiesen wäre. ∎

Der sechste Schritt des Algorithmus von ZASSENHAUS kann so interpretiert werden, daß er zu jedem irreduziblen Faktor $h \in \mathbb{F}_p[x]$ von $f \bmod p$ den zugehörigen Faktor $h_0 \in \mathbb{Z}[x]$ von f bestimmt, indem er nötigenfalls alle Kombinationen aus h und den anderen Faktoren von $f \bmod p$ durchprobiert. Der Algorithmus von LENSTRA, LENSTRA und LOVÁSZ konstruiert h_0 direkt und ohne Kenntnis der anderen Faktoren, indem er den Vektor der Koeffizienten voN $\vec{h}_0$ als einen „kurzen“ Gittervektor identifiziert.

Wir fixieren dazu eine natürliche Zahl $m \geq e = \deg h$ und betrachten die Menge Λ aller Polynome aus $\mathbb{Z}[x]$ vom Grad höchstens m, die modulo p^k durch $h \bmod p^k$ teilbar sind. Λ ist eine Teilmenge des $(m+1)$-dimensionalen $\mathbb{R}$-Vektorraums aller Polynome vom Grad höchstens m, den wir über die Basis $1, x, \ldots, x^m$ mit $\mathbb{R}^{m+1}$ identifizieren. Dabei ist die L^2-Norm $\|f\|_2$ eines Polynoms aus V gleich der üblichen EULIDischen Länge $|\vec{v}|$ seines Koeffizientenvektors $\vec{v} \in \mathbb{R}^{m+1}$.

Definition: Eine Teilmenge $\Gamma \subset \mathbb{R}^{m+1}$ heißt *Gitter,* wenn es eine Basis $(\vec{b}_0, \ldots, \vec{b}_m)$ von $\mathbb{R}^{m+1}$ gibt, so daß

$$\Gamma = \left\{ \sum_{i=0}^{m} \lambda_i \vec{b}_i \;\middle|\; \lambda_i \in \mathbb{Z} \right\} .$$

Wir bezeichnen diese Basis dann als eine Basis des Gitters Γ und schreiben kurz $\Gamma = \mathbb{Z}\vec{b}_0 \oplus \cdots \oplus \mathbb{Z}\vec{b}_m$.

Da h führenden Koeffizienten eins und Grad e hat, bilden die Polynome $p^k x^i$ mit $0 \le i < e$ und hx^j mit $0 \le j \le m - e$ eine Basis der oben definierten Menge Λ; diese ist also ein Gitter.

Gitterbasen sind genauso wenig eindeutig wie Basen von Vektorräumen. Sind $(\vec{b}_0, \ldots, \vec{b}_m)$ und $(\vec{c}_0, \ldots, \vec{c}_m)$ zwei Basen des Gitters Γ, so sind beide insbesondere Basen von $\mathbb{R}^{m+1}$, es gibt also Matrizen $M, N \in \mathbb{R}^{(m+1)\times(m+1)}$, die diese beiden Basen ineinander überführen. Am einfachsten läßt sich das dadurch ausdrücken, daß wir die Spaltenvektoren $\vec{b}_i$ zu einer Matrix B zusammenfassen und die $\vec{c}_j$ zu einer Matrix C; dann ist $C = MB$ und $B = NC$. Die Einträge von M und N müssen ganzzahlig sein, denn die $\vec{c}_j$ müssen ja ganzzahlige Linearkombinationen der $\vec{b}_i$ sein und umgekehrt. Außerdem ist $MN = NM$ gleich der Einheitsmatrix. Somit sind $\det M$ und $\det N$ ganzzahlig mit Produkt eins, d.h. $\det M = \det N = \pm 1$. Insbesondere unterscheiden sich $\det B$ und $\det C$ höchstens durch das Vorzeichen.

Definition: Der Betrag von $\det B$ heißt Determinante $d(\Gamma)$ des Gitters Γ.

Wie wir gerade gesehen haben, ist $d(\Gamma)$ unabhängig von der gewählten Gitterbasis. Im oben definierten Gitter Λ ist $d(\Lambda) = p^{ke}$, da h den führenden Koeffizienten eins hat und die hinteren Terme der Polynome hx^j durch Zeilenoperationen aus der Determinante entfernt werden können.

Ein Vektorraum hat keinen echten Untervektorraum gleicher Dimension; bei Gittern ist das natürlich anders: Mit Γ ist auch 2Γ ein Gitter und ganz offensichtlich verschieden von Γ. Allgemein sagen wir, ein Gitter $\Gamma \subset \mathbb{R}^{m+1}$ sei ein *Untergitter* des Gitters $\Delta \subset \mathbb{R}^{m+1}$, wenn Γ eine Teilmenge von Δ ist.

Ist in dieser Situation $\vec{b}_0, \ldots \vec{b}_n$ eine Gitterbasis von Γ und $\vec{c}_0, \ldots, \vec{c}_n$ eine von Δ, so lassen sich die $\vec{b}_i$ als Linearkombinationen der $\vec{c}_j$ schreiben. Mit den gleichen Bezeichnungen wie oben ist daher $B = NC$ mit einer ganzzahligen Matrix N. Die inverse Matrix M freilich ist im Falle eines echten Untergitters nicht mehr ganzzahlig, sondern hat nur rationale Einträge. Wir können allerdings die Nenner begrenzen: Die Gleichung NM gleich Einheitsmatrix läßt sich übersetzen in $m + 1$ lineare Gleichungssysteme für die Spalten $\vec{m}_i$ von M, denn $N\vec{m}_i = \vec{e}_i$ ist der i-te Einheitsvektor des $\mathbb{R}^{m+1}$. Lösen wir dieses Gleichungssystem nach der CRAMERschen Regel, so stehen im Zähler der Formeln für die Einträge von $\vec{m}_i$ Determinanten ganzzahliger Matrizen und in Nenner steht jeweils die Determinante D von M. Somit kann höchstens diese als Nenner auftreten und $D \cdot \Delta \subseteq \Gamma \subseteq \Delta$.

In Kürze wird es für uns wichtig sein, daß es zu einer gegebenen Basis von Δ spezielle, daran angepaßte Basen von Γ gibt:

Lemma: Ist Γ ein Untergitter von Δ und $\vec{b}_0, \ldots, \vec{b}_m$ eine Gitterbasis von Δ, so gibt es eine Gitterbasis $\vec{c}_0, \ldots, \vec{c}_m$ von Γ derart, daß

$$\begin{aligned} \vec{c}_0 &= \mu_{00}\vec{b}_0 \\ \vec{c}_1 &= \mu_{10}\vec{b}_0 + \mu_{11}\vec{b}_1 \\ &\cdots \\ \vec{c}_m &= \mu_{m0}\vec{b}_0 + \cdots + \mu_{mm}\vec{b}_m \end{aligned}$$

mit ganzen Zahlen μ_{ij} und $\mu_{ii} \neq 0$.

Beweis: Da $D\vec{b}_i$ in Γ liegt, gibt es in Γ auf jeden Fall für jedes i Vektoren der Form $\mu_{i0}\vec{b}_0 + \cdots + \mu_{ii}\vec{b}_i$ mit $\mu_{ii} \neq 0$. $\vec{c}_i$ sei ein solcher Vektor mit minimalem $|\mu_{ii}|$. Wir wollen zeigen, daß diese Vektoren $\vec{c}_i$ eine Gitterbasis von Γ bilden. Da die lineare Unabhängigkeit trivial ist, muß nur gezeigt werden, daß sich jeder Vektor aus Γ als ganzzahlige Linearkombination der $\vec{c}_i$ schreiben läßt.

Angenommen, es gibt Vektoren $\vec{v} \in \Gamma$, für die das nicht der Fall ist. Da $\vec{v}$ auch in Δ liegt, gibt es auf jeden Fall eine Darstellung

$\vec{v} = \lambda_0\vec{b}_0 + \cdots + \lambda_k\vec{b}_k$ mit ganzen Zahlen λ_i und einem $k \leq m$. Wir wählen einen solchen Vektor $\vec{v}$ mit kleinstmöglichem k.

Da μ_{kk} nach Voraussetzung nicht verschwindet, gibt es eine ganze Zahl q, so daß $|\lambda_k - q\mu_{kk}|$ kleiner ist als der Betrag von μ_{kk}. Dann kann auch der Vektor

$$\vec{v} - q\vec{c}_k = (\lambda_0 - q\mu_{k0})\vec{b}_0 + \cdots + (\lambda_k - z\mu_{kk})\vec{b}_k$$

nicht als ganzzahlige Linearkombination der $\vec{c}_i$ dargestellt werden, denn sonst hätte auch $\vec{v}$ eine solche Darstellung. Wegen der Minimalität von k kann daher $\lambda_k - q\mu_{kk}$ nicht verschwinden. Da aber Betrag von $\lambda_k - q\mu_{kk}$ kleiner ist als der von μ_{kk}, widerspricht dies der Wahl von $\vec{v}$ als Vektor mit minimalem $|\mu_{kk}|$. Somit kann es keinen Gittervektor aus Γ geben, der nicht als ganzzahlige Linearkombination der $\vec{c}_i$ darstellbar ist, und das Lemma ist bewiesen. ∎

Bei der Anwendung von Gittern auf das Faktorisierungsproblem werden die GRAM-SCHMIDT-Orthogonalisierungen von Gitterbasen eine große Rolle spielen; daher sei kurz an diesen Orthogonalisierungsprozeß erinnert. Zunächst die

Definition: *a)* Ein EUKLIDischer Vektorraum ist ein reeller Vektorraum V zusammen mit einer Abbildung

$$\begin{cases} V \times V \to V \\ (\vec{v}, \vec{w}) \mapsto \vec{v} \cdot \vec{w} \end{cases}$$

mit folgenden Eigenschaften:

1.) $(\lambda\vec{u} + \mu\vec{v}) \cdot \vec{w} = \lambda(\vec{u} \cdot \vec{w}) + \mu(\vec{v} \cdot \vec{w})$ für alle $\lambda, \mu \in \mathbb{R}$ und alle $\vec{u}, \vec{v}, \vec{w} \in V$.

2.) $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$ für alle $\vec{v}, \vec{w} \in V$.

3.) $\vec{v} \cdot \vec{v} \geq 0$ für alle $\vec{v} \in V$ und $\vec{v} \cdot \vec{v} = 0$ genau dann, wenn $\vec{v} = 0$.

Die Abbildung $V \times V \to V$ wird als Skalarprodukt bezeichnet.

b) Zwei Vektoren $\vec{v}, \vec{w} \in V$ heißen orthogonal, wenn $\vec{v} \cdot \vec{w} = 0$ ist.

c) Eine Basis $(\vec{c}_0, \ldots, \vec{c}_m)$ eines EUKLIDischen Vektorraums heißt *Orthogonalbasis,* wenn $\vec{c}_i \cdot \vec{c}_j = 0$ für alle $i \neq j$.

Wichtigstes Beispiel ist der Vektorraum $\mathbb{R}^{m+1}$ mit seinem Standardskalarprodukt

$$\begin{pmatrix} v_0 \\ \vdots \\ v_m \end{pmatrix} \cdot \begin{pmatrix} w_0 \\ \vdots \\ w_m \end{pmatrix} = \sum_{i=0}^{m} v_i w_i \,.$$

Das Produkt eines Vektors $\vec{v}$ mit sich selbst ist dann das Quadrat seiner EUKLIDischen Länge, und wenn wir ihn als Koeffizientenvektor eines Polynoms vom Grad m aus $\mathbb{R}[x]$ auffassen, ist das auch das Quadrat der L^2-Norm dieses Polynoms.

Das Orthogonalisierungsverfahren von GRAM und SCHMIDT konstruiert aus einer beliebigen Basis $(\vec{b}_0, \ldots, \vec{b}_m)$ des $\mathbb{R}^{m+1}$ schrittweise eine Orthogonalbasis $(\vec{c}_0, \ldots, \vec{c}_m)$, und zwar so, daß in jedem Schritt der von den Vektoren $\vec{b}_0, \ldots, \vec{b}_r$ erzeugte Untervektorraum gleich dem von $\vec{c}_0, \ldots, \vec{c}_r$ erzeugten ist.

Der erste Schritt ist der einfachste: Da es noch keine Orthogonalitätsbedingung für $\vec{c}_0$ gibt, können wir einfach $\vec{c}_0 = \vec{b}_0$ setzen.

Nachdem wir $r \geq 1$ Schritte durchgeführt haben, haben wir r linear unabhängige Vektoren $\vec{c}_0, \ldots, \vec{c}_{r-1}$ mit $\vec{c}_i \cdot \vec{c}_j = 0$ für $i \neq j$ aus dem von $\vec{b}_0, \ldots, \vec{b}_r$ aufgespannten Untervektorraum. Ist $r = m$, haben wir eine Orthogonalbasis; andernfalls muß ein auf den bisher konstruierten $\vec{c}_i$ senkrecht stehender Vektor $\vec{c}_r$ gefunden werden, der zusammen mit diesen den von $\vec{b}_0$ bis $\vec{b}_r$ erzeugten Untervektorraum erzeugt.

Da $\vec{c}_0, \ldots, \vec{c}_{r-1}$ und $\vec{b}_0, \ldots, \vec{b}_{r-1}$ denselben Untervektorraum erzeugen, gilt dasselbe für $\vec{c}_0, \ldots, \vec{c}_{r-1}, \vec{b}_r$ und $\vec{b}_0, \ldots, \vec{b}_r$; das Problem ist, daß $\vec{b}_r$ im allgemeinen nicht orthogonal zu den $\vec{c}_i$ sein wird. Wir dürfen $\vec{b}_r$ aber abändern um einen beliebigen Vektor aus dem von $\vec{c}_0, \ldots, \vec{c}_{r-1}$ aufgespannten Untervektorraum; also setzen wir

$$\vec{c}_r = \vec{b}_r + \lambda_0\vec{c}_1 - \cdots - \lambda_{r-1}\vec{c}_{r-1}$$

und versuchen, die λ_i so zu bestimmen, daß dieser Vektor orthogonal zu $\vec{c}_0, \ldots, \vec{c}_{r-1}$ wird.

Wegen der Orthogonalität der $\vec{c}_i$ ist

$$\vec{c}_r \cdot \vec{c}_i = \vec{b}_r \cdot \vec{c}_i - \sum_{j=0}^{r-1} \lambda_j (\vec{c}_j \cdot \vec{c}_i) = \vec{b}_r \cdot \vec{c}_i - \lambda_i (\vec{c}_i \cdot \vec{c}_i) ;$$

setzen wir daher

$$\lambda_i = \frac{\vec{b}_{r+1} \cdot \vec{c}_i}{\vec{c}_i \cdot \vec{c}_i} ,$$

so ist $\vec{v} \cdot \vec{c}_i = 0$ für alle $i = 0, \ldots, r-1$.

Nach dem $m+1$-ten Schritt haben wir eine Orthogonalbasis $(\vec{c}_0, \ldots, \vec{c}_m)$ von $\mathbb{R}^{m+1}$ konstruiert.

Der dänische Mathematiker JØRGAN PEDERSEN GRAM (1850–1916) lehrte an der Universität Kopenhagen, war aber gleichzeitig auch noch geschäftsführender Direktor einer Versicherungsgesellschaft und Präsident des Verbands der dänischen Versicherungsunternehmen. Er publizierte anscheinend nur eine einzige mathematische Arbeit *Sur quelque théorèmes fondamentaux de l'algèbre moderne,* die 1874 erschien. Das GRAM-SCHMIDTsche Orthogonalisierungsverfahren, durch das er heute hauptsächlich bekannt ist, stammt wohl von LAPLACE (1749–1827) und wurde auch schon 1836 von CAUCHY verwendet.

ERHARD SCHMIDT (1876–1959) wurde im damals deutschen Ort Dorpat geboren; heute gehört dieser zu Estland und heißt Tartu. Er studierte in Berlin bei SCHWARZ und promovierte 1905 ins Göttingen bei HILBERT mit einer Arbeit über Integralgleichungen. Nach seiner Promotion wechselte er nach Bonn, wo er 1906 habilitierte. Danach lehrte er in Zürich, Erlangen und Breslau, bis er 1917 als Nachfolger von SCHWARZ nach Berlin berufen wurde. Er ist einer der Begründer der modernen Funktionalanalysis; insbesondere geht die Verallgemeinerung EUKLIDischer und HERMITEscher Vektorräume zu sogenannten HILBERT-Räumen auf ihn zurück.

Als Beispiel wollen wir eine Orthogonalbasis des von

$$\vec{b}_0 = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 4 \end{pmatrix}, \quad \vec{b}_1 = \begin{pmatrix} -3 \\ 4 \\ 0 \\ 5 \end{pmatrix} \text{ und } \vec{b}_2 = \begin{pmatrix} -1 \\ -2 \\ 3 \\ 6 \end{pmatrix}$$

aufgespannten Untervektorraums U von $\mathbb{R}^4$ bestimmen.

Als ersten Vektor der Orthogonalbasis wählen wir einfach $\vec{c}_0 = \vec{b}_0$.

Für den zweiten Vektor machen wir den Ansatz $\vec{c}_1 = \vec{b}_1 - \lambda\vec{c}_0$, wobei λ so gewählt werden muß, daß $\vec{c}_1 \cdot \vec{c}_0 = \vec{b}_1 \cdot \vec{c}_0 - \lambda\vec{c}_0 \cdot \vec{c}_0 = 0$ ist. Da

$$\vec{c}_0 \cdot \vec{b}_1 = -3 + 2 \cdot 4 + 4 \cdot 5 = 25 \quad \text{und} \quad \vec{c}_0 \cdot \vec{c}_0 = 1^2 + 2^2 + 2^2 + 4^2 = 25 ,$$

müssen wir $\lambda = 1$ setzen und $\vec{c}_1 = \vec{b}_1 - \vec{c}_0 = \begin{pmatrix} -4 \\ 2 \\ -2 \\ 1 \end{pmatrix}$.

Für den noch fehlenden dritten Vektor der Orthogonalbasis ist der Ansatz entsprechend:

$$\vec{c}_2 = \vec{b}_2 - \lambda\vec{c}_0 - \mu\vec{c}_1 \quad \text{mit} \quad \vec{c}_2 \cdot \vec{c}_0 = \vec{c}_2 \cdot \vec{c}_1 = 0 .$$

$$\vec{c}_2 \cdot \vec{c}_0 = \vec{b}_2 \cdot \vec{c}_0 - \lambda\vec{c}_0 \cdot \vec{c}_0 = (-1 - 4 + 6 + 24) + 25\lambda \Longrightarrow \lambda = 1$$

$$\vec{c}_2 \cdot \vec{c}_1 = \vec{b}_2 \cdot \vec{c}_1 + \mu\vec{c}_1 \cdot \vec{c}_1 = (4 - 4 - 6 + 6) - 25\mu \Longrightarrow \mu = 0$$

$$\vec{c}_2 = \vec{b}_2 - \vec{c}_0 = \begin{pmatrix} -2 \\ -4 \\ 1 \\ 2 \end{pmatrix} .$$

Unsere Orthogonalbasis besteht also aus den drei Vektoren

$$\vec{c}_0 = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 4 \end{pmatrix}, \qquad \vec{c}_1 = \begin{pmatrix} -4 \\ 2 \\ -2 \\ 1 \end{pmatrix} \quad \text{und} \quad \vec{c}_2 = \begin{pmatrix} -2 \\ -4 \\ 1 \\ 2 \end{pmatrix} .$$

Wenn wir den Zusammenhang zwischen der Ausgangsbasis $(\vec{b}_0, \ldots, \vec{b}_m)$ und der Orthogonalbasis $(\vec{c}_0, \ldots, \vec{c}_m)$ explizit festhalten wollen, müssen wir den oben berechneten Koeffizienten Namen geben, die auch vom Schritt abhängen. Wir schreiben

$$\vec{c}_i = \vec{b}_i - \sum_{j=0}^{i-1} \mu_{ij}\vec{c}_j \quad \text{für } i = 0, \ldots, m \quad \text{mit} \quad \mu_{ij} = \frac{\vec{b}_i \cdot \vec{c}_j}{\vec{c}_j \cdot \vec{c}_j} .$$

Im gerade durchgerechneten Beispiel etwa ist

$$\vec{c}_0 = \vec{b}_0, \quad \vec{c}_1 = \vec{b}_1 - \vec{c}_0 \quad \text{und} \quad \vec{c}_2 = \vec{b}_2 - \vec{c}_0\,,$$

also $\mu_{10} = \mu_{20} = 1$ und $\mu_{21} = 0$.

Lösen wir die obigen Formeln auf nach $\vec{b}_i$, kommen wir auf

$$\vec{b}_i = \vec{c}_i + \sum_{j=0}^{i-1} \mu_{ij}\vec{c}_j \quad \text{und} \quad \vec{c}_i \cdot \sum_{j=0}^{i-1} \mu_{ij}\vec{c}_j = \sum_{j=0}^{i-1} \mu_{ij}\vec{c}_i \cdot \vec{c}_j = 0\,.$$

Geometrisch bedeutet dies, daß $\vec{c}_i$ der Lotvektor bei der Projektion von $\vec{b}_i$ auf den von $\vec{c}_1, \ldots, \vec{c}_{i-1}$ aufgespannten Untervektorraum ist oder, anders ausgedrückt, die orthogonale Projektion von $\vec{b}_i$ auf das orthogonale Komplement dieses Raums.

Ist allgemein $\vec{w} = \vec{u} + \vec{v}$ die Summe zweier aufeinander senkrecht stehenden Vektoren, so ist

$$\vec{w} \cdot \vec{w} = (\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v}\,,$$

denn $\vec{v} \cdot \vec{w} = 0$. Insbesondere sind daher die Längen von $\vec{v}$ und $\vec{w}$ höchstens gleich der Länge von $\vec{w}$. In unserer Situation bedeutet dies, daß

$$|\vec{c}_i| \le \left|\vec{b}_i\right| \quad \text{für} \quad i = 0, \ldots, m\,,$$

kein Vektor der Orthogonalbasis kann also länger sein als der entsprechende Vektor der Ausgangsbasis.

Im Falle einer Gitterbasis $(\vec{b}_0, \ldots, \vec{b}_m)$ ist die nach GRAM-SCHMIDT berechnete Orthogonalbasis zwar eine Basis des $\mathbb{R}^{m+1}$, aber im allgemeinen keine Gitterbasis: Es gibt schließlich keinen Grund, warum die μ_{ij} ganze Zahlen sein sollten, so daß die $\vec{c}_j$ oft nicht einmal im Gitter liegen, und tatsächlich muß ein Gitter auch keine Orthogonalbasis haben.

Hätte etwa das Gitter

$$\Lambda = \mathbb{Z}\begin{pmatrix}1\\0\end{pmatrix} \oplus \mathbb{Z}\begin{pmatrix}\sqrt{2}\\1\end{pmatrix}$$

eine Orthogonalbasis $\vec{u}, \vec{v}$, so gäbe es ganze Zahlen a, b, c, d, so daß

$$\vec{u} = a\begin{pmatrix}1\\0\end{pmatrix} + b\begin{pmatrix}\sqrt{2}\\1\end{pmatrix} = \begin{pmatrix}a + b\sqrt{2}\\b\end{pmatrix}$$

und

$$\vec{v} = c\begin{pmatrix}1\\0\end{pmatrix} + d\begin{pmatrix}\sqrt{2}\\1\end{pmatrix} = \begin{pmatrix}c + d\sqrt{2}\\d\end{pmatrix}$$

wäre. Das Skalarprodukt dieser beiden Vektoren wäre null, d.h.

$$(a + b\sqrt{2})(c + d\sqrt{2}) + bd = (ac + 3bd) + (ad + bc)\sqrt{2} = 0\,.$$

Wegen der Irrationalität von $\sqrt{2}$ ist dies nur dann möglich, wenn $ad + bc$ verschwindet. Nun wissen wir aber, daß die Determinante der Matrix für einen Wechsel der Gitterbasis ± 1 sein muß, d.h. $ad - bc = \pm 1$. Addition dieser Gleichung zu $ad + bc = 0$ führt auf $2ad = \pm 1$, was mit ganzen Zahlen a, d offensichtlich nicht gelten kann. Somit hat zumindest dieses Gitter keine Orthogonalbasis.

Trotzdem ist die aus einer Gitterbasis konstruierte Orthogonalbasis auch nützlich zum Verständnis des Gitters. Als erstes Beispiel dafür wollen wir die Determinante des Gitters geometrisch interpretieren:

Die reelle Matrix M, die den Wechsel von der Ausgangsbasis zur Orthogonalbasis beschreibt, ist wie die obigen Formeln zeigen eine Dreiecksmatrix mit lauter Einsen in der Hauptdiagonale, hat also Determinante eins. Obwohl die $\vec{c}_i$ keine Gitterbasis bilden, ist daher die Determinante des Gitters auch gleich dem Betrag der Determinante der Matrix C mit den $\vec{c}_i$ als Spaltenvektoren. Das ist aber einfach das Produkt der Längen der $\vec{c}_i$, denn der ij-Eintrag von CC^T ist $\vec{c}_i \cdot \vec{c}_j$. Da die $\vec{c}_i$ eine Orthogonalbasis bilden, verschwindet dieses Skalarprodukt für $i \neq j$, also ist CC^T eine Diagonalmatrix und ihre Determinante ist das Produkt der Längenquadrate $\vec{c}_i \cdot \vec{c}_i$. Der Betrag der Determinante von C selbst ist daher das Produkt der Längen.

Geometrisch entspricht die Orthogonalisierung nach GRAM-SCHMIDT einer Folge von Scherungen, die aus dem von den Vektoren $\vec{b}_i$ aufgespannten Parallelepiped einen Quader machen. Nach dem Prinzip von CAVALIERI bleibt das Volumen dabei unverändert, und das Volumen

einer Quaders ist natürlich einfach das Produkt seiner Seitenlängen. Somit ist die Determinante eines Gitters gleich dem Volumen des von den Basisvektoren aufgespannten Parallelepipeds, dem sogenannten Fundamentalbereich des Gitters. Je nach Wahl der Basis kann dieser sehr verschiedene Formen haben, aber sein Volumen ist stets dasselbe.

Das wollen nun anwenden, um die Determinante eines Gitters abzuschätzen durch die Längen der Vektoren aus einer beliebigen Gitterbasis $(\vec{b}_0, \ldots, \vec{b}_m)$. Ist $(\vec{c}_0, \ldots, \vec{c}_m)$ die zugehörige Orthogonalbasis, so ist die Determinante des Gitters das Produkt der Längen der Vektoren $\vec{c}_i$. Wie wir oben gesehen haben, kann kein Vektor $\vec{c}_i$ länger sein als der entsprechende Vektor $\vec{b}_i$, und damit folgt die

Ungleichung von Hadamard: Im Gitter $\Gamma = \mathbb{Z}\vec{b}_0 \oplus \cdots \oplus \mathbb{Z}\vec{b}_m$ ist

$$d(\Gamma) \le \prod_{i=0}^{m} |\vec{b}_i| \qquad \blacksquare$$

JACQUES SALOMON HADAMARD wurde 1865 in Versailles geboren, lebte aber ab dem Alter von drei Jahren in Paris. Dort studierte er von 1884-1888 an der Ecole Normale Supérieure; während der anschließenden Arbeit an seiner Dissertation verdiente er seinen Lebensunterhalt als Lehrer. Nach seiner Promotion 1892 ging er zunächst als Dozent, ab 1896 als Professor für Astronomie und Theoretische Mechanik an die Universität von Bordeaux. Während dieser Zeit bewies er unter anderem den berühmten Primzahlsatz, wonach sich die Anzahl der Primzahlen $\le n$ asymtotisch verhält wie $n/\ln n$. Um wieder nach Paris zurückzukommen, akzeptierte er 1897 dort zwei (schlechtere) Stellen an der Sorbonne und am Collège de France; am letzteren erhielt er 1909 einen Lehrstuhl. 1912 wurde er Nachfolger von CAMILLE JORDAN an der Ecole Polytechnique sowie Nachfolger von HENRI PONCARÉ an der Académie des Sciences. 1940 mußte er nach USA emigrieren und lehrte an der Columbia University in New York, kehrte aber sofort nach Kriegsende zurück nach Paris. Unter seinen Arbeiten befinden sich außer dem Primzahlsatz auch fundamentale Beiträge unter anderem zur Theorie der partiellen Differentialgleichung, zu geodätischen Linien und zur Variationsrechnung. Auch politisch war er sehr aktiv, zunächst zugunsten von ALFRED DREYFUS. Nach 1945 engagierte er sich, nachdem drei seiner Söhne in den Weltkriegen gefallen waren, für die Friedensbewegung; zum Internationalen Mathematikerkongress in Cambridge, *Mass.*, dessen Ehrenpräsident er war, erhielt er deshalb nur nach der Intervention zahlreicher amerikanischer Mathematiker ein Einreisevisum für die USA.

Die Ungleichung von HADAMARD spielt eine wichtige Rolle im Beweis des folgenden Lemmas, das bei der Suche nach dem zu Beginn des Paragraphen definierten Polynoms h_0 nützlich sein wird. Alle Bezeichnungen seien wie dort.

Lemma: Erfüllt ein Polynom $v \in \Lambda$ die Ungleichung $\|f\|_2 \cdot \|v\|_2 < p^{ke}$, so ist v durch h_0 teilbar.

Beweis: Für das Nullpolynom ist die Aussage trivial; sei also $v \neq 0$ und $g = \mathrm{ggT}(f, v)$. Nach dem ersten Lemma dieses Paragraphen reicht es zu zeigen, daß h mod p ein Teiler von g mod p ist.

Sollte dies nicht der Fall sein, sind h mod p und g mod p wegen der Irreduzibilität von h mod p teilerfremd; wie oben gibt es also Polynome $a, b, c \in \mathbb{Z}[x]$, so daß gilt

$$ah + bg = 1 - pc\,.$$

Sei $n = \deg g$ und $m' = \deg v$; dann ist $n \le m' \le m$. Wir definieren eine neue Teilmenge

$$\mathsf{M} = \{\lambda f + \mu v \mid \lambda, \mu \in \mathbb{Z}[x],\ \deg\lambda < m' - n,\ \deg\mu < d - n\}$$

des Gitters $\mathbb{Z} \oplus \mathbb{Z}x \oplus \cdots \oplus \mathbb{Z}x^{d+m'-n-1}$; ihre natürliche Projektion auf das Untergitter $\mathbb{Z}x^n \oplus \mathbb{Z}x^{n+1} \oplus \cdots \oplus \mathbb{Z}x^{d+m'-n-1}$ sei M'.

Angenommen, das Element $\lambda f + \mu v \in \mathsf{M}$ wird dabei auf das Nullpolynom projiziert. Dann muß einerseits der Grad von $\lambda f + \mu v$ kleiner als n sein, andererseits ist $\lambda f + \mu v$ durch g teilbar und $n = \deg g$. Also muß $\lambda f + \mu v = 0$ sein und $\lambda f = -\mu v$. Division durch $g = \mathrm{ggT}(f, v)$ führt auf

$$\lambda\frac{f}{g} = -\mu\frac{v}{g} \quad \text{und} \quad \mathrm{ggT}\left(\frac{f}{g}, \frac{v}{g}\right) = 1\,.$$

Somit muß μ ein Vielfaches von f/g sein. Der Grad von μ ist aber nach Definition kleiner als $d - n = \deg f - \deg g$, also ist $\mu = 0$ und damit auch $\lambda = 0$. Daher sind die Projektionen der Polynome

$$x^i f \ \text{ für } \ 0 \le i < m' - n \quad \text{und} \quad x^j v \ \text{ für } \ 0 \le j < d - n$$

nach M' linear unabhängig. Wie die Definition von M zeigt, bilden sie auch ein Erzeugendensystem, also ist M' ein Gitter, und die obigen

Polynome bilden eine Gitterbasis. Darauf können wir die Ungleichung von HADAMARD anwenden:

$$d(\mathsf{M}') \le \|f\|_2^{m'-n} \cdot \|v\|_2^{e-n} \le \|f\|_2^m \, \|v\|_2^d < p^{ke} ,$$

wobei das letzte Kleinerzeichen die Voraussetzung des Lemmas ist.

Im Rest des Beweises wollen wir zeigen, daß $d(\mathsf{M}') \ge p^{ke}$ sein muß, was zusammen mit der gerade gezeigten Ungleichung zu einem Widerspruch führt und damit das Lemma beweist.

Sei dazu $w \in \mathsf{M}$ ein Polynom vom Grad kleiner $n+e$. Als Element von M ist es durch g teilbar. Multiplizieren wir die obige Gleichung $ah + bg = 1 - pc$ mit $1 + pc + \cdots + (pc)^{k-1}$, erhalten wir eine Gleichung der Form $\tilde{a}h + \tilde{b}g = 1 - (pc)^k$ mit $\tilde{a}, \tilde{b} \in \mathbb{Z}[x]$. Multiplikation dieser Gleichung mit dem Polynom w/g führt auf eine neue Gleichung

$$a^* h + b^* w = \frac{w}{g}\left(1 - (pc)^p\right) \equiv \frac{w}{g} \bmod p^k \quad \text{mit} \quad a^*, b^* \in \mathbb{Z}[x] .$$

Als Element von M läßt sich w in der Form $w = \lambda f + \mu v$ schreiben, und nach Voraussetzung sind sowohl f als auch v modulo p^k durch h teilbar. Also ist auch w und damit nach der gerade bewiesenen Gleichung w/g modulo p^k durch h teilbar. Der Grad von w ist kleiner als $n+e$, und g hat Grad n, also ist der Grad von w/g kleiner als $n+e-n = e = \deg h$. Da h führenden Koeffizienten eins hat, wird dieser Grad modulo p^k nicht kleiner; also muß w/g und damit auch w modulo p^k das Nullpolynom sein. Somit ist jedes Polynom aus M mit Grad kleiner $n+e$ durch p^k teilbar.

Das Gitter M' liegt in $\mathbb{Z}x^n \oplus \cdots \oplus \mathbb{Z}x^{d+m'-n-1}$ und hat eine Gitterbasis aus $d+m'-2n$ Elementen, ist also ein Untergitter im Sinne der Definition dieses Paragraphen. Die Polynome x^n bis $x^{d+m'-n-1}$ bilden natürlich eine Basis des größeren Gitters; daher hat M nach dem zweiten Lemma dieses Paragraphen eine Gitterbasis aus Polynomen der Grade $n, n+1, \ldots, d+m'-n-1$. Die ersten e davon müssen, wie wir gerade gesehen haben, durch p^k teilbar sein. Die Determinante von M' ist der Betrag der Determinante der Matrix aus den Basisvektoren; auf Grund der Gradbedingung ist dies eine Dreiecksmatrix, die Determinante ist also einfach das Produkt der führenden Koeffizienten und hat damit mindestens Betrag p^{ke}. Dies liefert den verlangten Widerspruch. ∎

Das gerade bewiesene Lemma legt nahe, daß uns Polynome kleiner L^2-Norm im Gitter Λ zu Faktoren von f verhelfen können, und in der Tat zeigen LENSTRA, LENSTRA und LOVÁSZ, daß wir h_0 konstruieren können als ggT der Polynome aus einer „geeigneten“ Gitterbasis von Λ. Im nächsten Paragraphen soll diese auch für viele andere Aufgaben „geeignete“ Basis allgemein konstruiert werden.

§ 10: Der LLL-Algorithmus zur Basisreduktion

Der hier vorgestellte Algorithmus wurde zwar in der zu Beginn des vorigen Paragraphen zitierten Arbeit von LENSTRA, LENSTRA und LOVÁSZ speziell für die Faktorisierung von Polynomen aus $\mathbb{Z}[x]$ entwickelt, er fand aber inzwischen zahlreiche weitere Anwendungen in der Kryptographie, der diskreten Optimierung und anderswo. Deshalb wird hier nicht von Polynomen, sondern nur allgemein von Vektoren die Rede sein, und wir werden auch, wie dort üblich, die Numerierung nicht wie im vorigen Paragraphen bei der für Polynome sinnvollen Null beginnen, sondern bei eins.

Wir gehen daher aus von einem Gitter $\Gamma = \mathbb{Z}\vec{b}_1 \oplus \cdots \oplus \mathbb{Z}\vec{b}_n \le \mathbb{R}^n$ und wollen dort nach kurzen Vektoren suchen.

Falls die Gitterbasis $\vec{b}_1, \ldots, \vec{b}_n$ eine Orthogonalbasis des $\mathbb{R}^n$ ist, hat ein Vektor $\vec{v} = a_1\vec{b}_1 + \cdots + a_n\vec{b}_n$ aus Γ die Länge

$$|\vec{v}| = \sqrt{a_1^2\, \vec{b}_1 \cdot \vec{b}_1 + \cdots + a_n^2\, \vec{b}_n \cdot \vec{b}_n} ;$$

wenn wir zusätzlich noch annehmen, daß $\left|\vec{b}_1\right| \le \cdots \le \left|\vec{b}_n\right|$ ist, sind daher $\pm\vec{b}_1$ kürzeste Vektoren in Γ. Je nach Länge von $\vec{b}_2$ gilt eventuell dasselbe auch für $\pm\vec{b}_2$; falls $\vec{b}_2$ aber länger ist als $\vec{b}_1$, müssen wir auf der Suche nach zweitkürzesten Vektoren die Längen von $\vec{b}_2$ und $2\vec{b}_1$ miteinander vergleichen und können uns entsprechend weiter hochhangeln, bis wir alle Vektoren unterhalb einer vorgegebenen Länge gefunden haben.