J. Wolfart WiSe 2012/13

Diophantische Approximationen

Teilweise basiert dieses Skriptum auf meiner gleichnamigen vierstündigen Vorlesung vom Winter 1995/96, damals freundlicherweise ausgearbeitet von zwei sehr fleißigen Studenten, Roger und Marc Fischlin, inzwischen arrivierte Mathematiker/Informatiker.

Literatur

G.H. Hardy, E.M. Wright: An Introduction to the Theory of Numbers, Oxford, Clarendon Press 1979

W.M. Schmidt: Diophantine Approximation, Springer Lecture Notes in Mathematics 785 (1980)

J. Steuding: Diophantine Analysis, Chapman and Hall 2005

Y. Bugeaud: Approximation by Algebraic Numbers, Cambridge 2004

1 Zwei Probleme aus der Astronomie

1.1 Huygens und das Planetarium

Huygens hatte 1682 ein automatisches Planetarium zu konstruieren. In einem Jahr mit 365 Tagen bewegt sich die Erde um einen Winkel $359^0 45' 40'' 30'''$ um die Sonne, Saturn um einen Winkel $12^0 13' 34'' 18'''$. Ins Dezimalsystem umgerechnet, stehen die Winkelgeschwindigkeiten in einem Verhältnis

$$\frac{77\,708\,431}{2\,640\,858} \;=\; 29,42544 \quad ,$$

das durch eine Zahnradübersetzung in einem mechanischen Getriebe dargestellt werden muss. Man könnte hier ein Zahnradverhältnis $\frac{294}{10} = \frac{147}{5}$ nehmen, handelt sich damit aber einem Fehler von $0,025$, also $2,5\%$ ein. Gibt es ein besseres Zahlenverhältnis, ohne dass man die Zahnräder unverhältnismäßig kompliziert macht? In der Tat gibt

$$\frac{206}{7} - \frac{77\,708\,431}{2\,640\,858} \;=\; 0,00312\ldots$$

ein viel genaueres Resultat – mit kleinerem Zähler und Nenner. Wie hat Huygens dieses gefunden?

1.2 Kalender

Die Umlaufzeit der Erde um die Sonne und die Drehung der Erde um ihre Achse sind schlecht aufeinander abgestimmt: Das Sonnenjahr besteht aus 365 Tagen, 5 Stunden, 48 Minuten und 45,8 Sekunden, also etwa $365+\frac{419}{1730}$ Tagen. Eine erste Approximation besteht also darin, dass man das Kalenderjahr immer mit 365 Tagen ausstattet; so wurde im alten Ägypten gerechnet. Damit nimmt man aber in Kauf, dass sich innerhalb von, sagen wir, 40 Jahren das Hochwasser des Nils – lebenswichtig für die Landwirtschaft – um bereits 10 Tage im Kalender verschiebt. In der Tat wurde der Zeitpunkt der Nilüberschwemmung von den Astronomen immer neu bestimmt, und diese Angaben sind heute für Historiker ein sehr wichtiges Datierungsinstrument.

Der nächste Schritt war eine Verbesserung des Kalenders, die durch Julius Caesar Gesetz wurde: die Approximation des Sonnenjahrs durch $365+\frac{1}{4}$. Der *Julianische Kalender* trägt dem Rechnung dadurch, dass alle vier Jahre ein weiterer Tag eingefügt wird. Nun ist

$$\frac{1}{4} - \frac{419}{1730} \approx 0,0078$$

ein Unterschied, der in 100 Jahren auch schon wieder fast einen Tag ausmacht, der also bis zum Jahr 1582 (Papst Gregor XIII) durch eine Verschiebung von Jahreszeiten deutlich spürbar wurde. Wir werden sehen, dass $\frac{194}{801}$ eine wesentlich bessere Approximation an $\frac{419}{1730}$ ist als $1/4$. Mit

$$\frac{419}{1730} \approx \frac{194}{801} \approx \frac{194}{800} = \frac{97}{400}$$

lautet nun das Rezept des immer noch gebräuchlichen *Gregorianischen Kalenders*, innerhalb von 400 Jahren dreimal den Schalttag wegfallen zu lassen, nämlich in den Jahren mit einer Jahreszahl $\equiv 100,\ 200$ oder $300 \bmod 400$. In 10.000 Jahren macht man damit einen Fehler von 3 Tagen.

1.3 Der mathematische Hintergrund

besteht darin, reelle Zahlen möglichst gut durch rationale Zahlen mit beschränkten Nennern oder gar Nennern spezieller Bauart zu approximieren.

- Wie präzisiert man ,,gut approximieren"?
- Wie stellt man gut approximierende rationale Zahlen effektiv her?
- Was sagt die Qualität der Approximierbarkeit über die Natur der Zahl aus?

2 Der Approximationssatz von Kronecker

Zu $a \in \mathbf{R}$ bezeichne $[a]$ die größte ganze Zahl $\leq a$ (Gaußklammer) und $\langle a \rangle := a - [a]$ den *gebrochenen Anteil* von a.

Satz 2.1 *(*KRONECKER*) Sei $a \in \mathbf{R}$. Dann gilt die folgende Alternative:*

- $\{\langle na \rangle \mid n \in \mathbf{N}\}$ *ist genau dann endlich, wenn $a \in \mathbf{Q}$.*
- $\{\langle na \rangle \mid n \in \mathbf{N}\}$ *ist genau dann dicht im Intervall $[0,1]$, wenn $a \notin \mathbf{Q}$. M.a.W.: Dann gibt es für jedes $r \in \mathbf{R}$ und jedes $\varepsilon > 0$ ganze Zahlen $n \in \mathbf{N}$, $m \in \mathbf{Z}$ mit $|na - m - r| < \varepsilon$.*

Beweis. Wenn $a \in \mathbf{Q}$, dann ist $a = \frac{q}{m}$ mit teilerfremden $q, m \in \mathbf{Z}$, $m \neq 0$, und dann gilt

$$n \equiv n' \bmod m \;\Leftrightarrow\; n - n' = km\,,\ k \in \mathbf{Z} \;\Leftrightarrow\; na - n'a \in \mathbf{Z} \;\Leftrightarrow\; \langle na \rangle = \langle n'a \rangle\,,$$

die gebrochenen Anteile $\langle na \rangle$ durchlaufen also genau $|m|$ verschiedene Werte. Wenn sie andererseits nur endlich viele verschiedene Werte durchlaufen, gibt es $n \neq n' \in \mathbf{N}$ mit $na - n'a = (n - n')a \in \mathbf{Z}$, also ist a rational.

Insbesondere heißt das umgekehrt: Wenn a irrational ist, sind die gebrochenen Anteile $\langle na \rangle$ alle paarweise verschieden, besitzen daher in dem kompakten Intervall $[0,1]$ einen Häufungspunkt. Nach dem Cauchy–Kriterium gibt es also zu jedem $\varepsilon > 0$ Vielfache $n''a \neq n'a$ mit $|\langle n''a \rangle - \langle n'a \rangle| < \varepsilon$. Sei o.B.d.A. $n'' > n'$, dann heißt das: Es gibt ein $m \in \mathbf{Z}$ und ein $k := n'' - n'$ mit $|ka - m| < \varepsilon$. Die aufeinander folgenden Vielfachen von ka haben somit die Eigenschaft $|\langle (n+1)ka \rangle - \langle nka \rangle| < \varepsilon$, ihre gebrochenen Anteile also einen Abstand $< \varepsilon$. Da man ε beliebig klein wählen darf, liegt $\{\langle na \rangle \mid n \in \mathbf{N}\}$ also dicht im Einheitsintervall. □

Anwendung. Wann beschreibt eine Billardkugel auf einem rechteckigen Billardtisch eine periodische Bahn? Dazu macht man natürlich einige vereinfachende Annahmen. Die Kugel bewege sich völlig reibungsfrei, also ungebremst auf unbestimmte Zeit, dazu auf genau geradlinigen Bahnen, wird ferner bei jedem Anstoßen an die Bande exakt reflektiert (Einfallswinkel = Ausfallswinkel) und treffe niemals auf einen Eckpunkt des Tischs. Zur Vereinfachung der Rechnung nehmen wir an, dass der Tisch quadratisch ist, dass zwei Tischseiten von den Koordinatenachsen des $\mathbf{R}^2$ gebildet werden und dass die Eckpunkte in

$$(0,0)\,,\quad (\frac{1}{2},0)\,,\quad (0,\frac{1}{2})\,,\quad (\frac{1}{2},\frac{1}{2})$$

liegen. Der Anfangsweg der Kugel sei durch die Geradengleichung $y = y_0 + mx$ beschrieben. Dann hat das Problem eine einfache Lösung:

Folgerung 2.1 *Die Billardkugel beschreibt genau dann eine periodische Bahn, wenn m eine rationale Zahl ist. Andernfalls verläuft die Bahn der Kugel dicht auf dem Tisch.*

Zum *Beweis* denke man sich den $\mathbf{R}^2$ mit einem quadratischen Gitter der Maschengröße $1/2$ überzogen, gebildet durch die Geraden $x = k\,,\ y = j\,,\ k,j \in \frac{1}{2}\mathbf{Z}$. Der Billardtisch ist dabei eine der Gittermaschen. Anstelle einer Reflexion der Kugelbahn an der Bande spiegele man nun lieber die Masche selbst an der berandenden Geraden; dabei verläuft die gespiegelte Kugelbahn genau auf der alten Geraden. Periodizität liegt offenbar genau dann vor, wenn die Gerade $y = y_0 + mx$ durch einen nichttriviale ganzzahlige Translation $(x,y) \mapsto (x+\nu, y+\mu)\,,\ \nu,\mu \in \mathbf{Z}$, in sich übergeführt wird. Das ist gerade für $m = \mu/\nu$ erfüllt.

Wenn andererseits $m \notin \mathbf{Q}$ ist, wissen wir nach dem Kroneckerschen Satz, dass die $\langle km\rangle$ für $k \in \mathbf{N}$ dicht im Einheitsintervall liegen. Zu jedem Punkt (x,y) des Billardtischs gibt es daher Punkte

$$(x+k, y_0 + m(x+k)) \;=\; (x+k, y_0 + mx + [mk] + \langle mk\rangle)$$

auf der Geraden, die bis auf Translation um eine Vektor aus $\mathbf{Z}^2$ beliebig nahe an (x,y) liegen, wenn man k so wählt, dass $\langle mk\rangle$ hinreichend dicht an $y - y_0 - mx - [mk]$ liegt. □

Das gleiche Argument hat noch eine andere geometrische Interpretation, wenn man keine Reflexionen, sondern nur die Translationen aus $\mathbf{Z}^2$ betrachtet. Ebenso wie man $\mathbf{R}/\mathbf{Z}$, dessen Elemente vom Intervall $[0,1[$ repräsentiert werden, vermöge $x \mapsto e^{2\pi i x}$ mit dem Einheitskreis identifizieren kann, so das kann man Einheitsquadrat ansehen als Repräsentantenmenge von $\mathbf{R}^2/\mathbf{Z}^2$, wenn man gegenüberliegende Seiten identifiziert. Topologisch entsteht dabei ein *Torus* (Fahrradschlauch, Schwimmring), und das Resultat der Folgerung kann man für den Torus so interpretieren, dass eine Gerade im $\mathbf{R}^2$ je nach der arithmetischen Natur ihrer Steigung entweder eine endliche (periodische, kreuzungsfreie) Kurve auf dem Torus definiert oder eine immer noch kreuzungsfreie, aber auf dem Torus dicht liegende Kurve unendlicher Länge wird.

Es gibt auch eine höherdimensionale Version des Kroneckerschen Satzes. Man darf hier allerdings nicht erwarten, dass für beliebige irrationale a, b die gebrochenen Anteile der Vielfachen, also die $(\langle na\rangle, \langle nb\rangle)$ dicht im Einheitsquadrat $[0,1]^2$ liegen: Für $a = b$ liegen sie nur auf der Diagonalen $x = y$ dicht, man braucht also offenbar schärfere Voraussetzungen. Hier ist der Satz in zwei Versionen:

Satz 2.2 *Seien* $n \in \mathbf{N}$ *und* $1, a_1, \ldots, a_n \in \mathbf{R}$ *linear unabhängig über* $\mathbf{Q}$. *Dann liegt die Menge* $\{\,(\langle ma_1\rangle, \ldots \langle ma_n\rangle) \mid m \in \mathbf{N}\,\}$ *dicht in* $[0,1]^n$.

Satz 2.3 *Seien* $k \in \mathbf{N}$, $b_1, \ldots, b_k \in \mathbf{R}$ *linear unabhängig über* $\mathbf{Q}$. *Dann gibt es für alle* $(x_1, \ldots, x_k) \in \mathbf{R}^k$, $T > 0$, $\varepsilon > 0$ *ganze Zahlen* $p_1, \ldots, p_k$ *und reelle* $t > T$ *mit*

$$|tb_j - p_j - x_j| < \varepsilon \qquad \textit{für alle} \qquad j = 1, \ldots, k\,.$$

Beweise. a) Satz 2.3 ⇒ Satz 2.2 : Wähle $n+1 = k$, $b_k = 1$, o.B.d.A. alle $a_j = b_j = \langle b_j\rangle \in]0,1[$ für $j = 1, \ldots, k-1 = n$, $x_{n+1} = 0$. Dann folgt aus Satz 2.3 die Existenz von

genügend großen $t \in \mathbf{R}$ und $p_j \in \mathbf{Z}$ mit

$$|t - p_k| < \varepsilon \,,\ m := p_k \,, \quad \text{alle} \quad |ta_j - p_j - x_j| < \varepsilon \quad \text{für alle} \quad j \le n$$

für beliebige $(x_1, \ldots, x_n) \in \mathbf{R}^n$. Dann ist aber auch durch geeignete $m = p_k \in \mathbf{N}$ die Ungleichung

$$|ma_j - p_j - x_j| = |(m-t)a_j + ta_j - p_j - x_j| \le |m-t|\,|a_j| + |ta_j - p_j - x_j| < (|a_j| + 1)\varepsilon$$

erfüllbar, also liegt $\{\,(\langle ma_1\rangle, \ldots \langle ma_n\rangle) \mid m \in \mathbf{N}\,\}$ dicht in $[0,1]^n$.

b) Satz 2.2 $\Rightarrow$ Satz 2.3 : Sei wieder $k := n+1$, $b_1, \ldots, b_k$ linear unabhängig über $\mathbf{Q}$. Wir setzen o.B.d.A. $b_k > 0$ voraus (warum geht das?). Dann sind

$$a_1 := \frac{b_1}{b_k} \,, \quad \ldots \quad , \ a_n := \frac{b_n}{b_k} \,, \ 1$$

linear unabhängig über $\mathbf{Q}$, nach Satz 2.2 also $\{\,(\langle ma_1\rangle, \ldots \langle ma_n\rangle) \mid m \in \mathbf{N}\,\}$ dicht in $[0,1]^n$. Dies bleibt richtig, wenn man nur $m > M =: Tb_k$ zulässt, da man dann nur endlich viele m außer Acht lässt. Für alle $\varepsilon > 0$ und alle $(x_1, \ldots, x_n) \in \mathbf{R}^n$ gibt es also $p'_1, \ldots, p'_n \in \mathbf{Z}$ und $m \in \mathbf{N}$, $m > M$, die für alle $j = 1, \ldots, n$

$$|ma_j - p'_j - x_j| \ = \ |m\frac{b_j}{b_k} - p'_j - x_j| \ < \ \varepsilon$$

erfüllen. Mit $t' := m/b_k$ ist also $|t'b_j - p'_j - x_j| < \varepsilon$ für alle $j \le n$, dazu $t' = m/b_k > M/b_k = T$ und $|t'b_k - m - 0| = 0 < \varepsilon$ erfüllt für $p'_k := m$ und $x_k = x_{n+1} = 0$. Satz 2.3 ist also gültig für alle $(x_1, \ldots, x_n, 0) \in \mathbf{R}^{n+1}$. Mit der gleichen Schlussweise kann man $t'' > T$ und $p''_1, \ldots, p''_{n+1} \in \mathbf{Z}$ finden, welche

$$|t''b_j - p''_j - 0| < \varepsilon \quad \text{für alle} \quad j = 0, \ldots, n \quad \text{und} \quad |t''b_k - p''_k - x_k| < \varepsilon$$

für beliebige $x_k = x_{n+1}$ erfüllen. Addition der beiden Ungleichungstypen zeigt, dass mit $t := t' + t''$ und $p_j := p'_j + p''_j$ für alle $(x_1, \ldots, x_k) \in \mathbf{R}^k$

$$|tb_j - p_j - x_j| < 2\varepsilon \qquad \text{für alle} \qquad j = 1, \ldots, k$$

erfüllt ist.

c) Beweis von Satz 2.3 nach Harald Bohr : 1. Sei $e(x) := e^{2\pi i x}$. Es gilt

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T e^{ict} dt \ = \ \lim_{T\to\infty} \frac{e^{ciT} - 1}{ciT} \ = 0 \text{ für } c \in \mathbf{R} \,,\ c \neq 0 \quad \text{und} \ = 1 \text{ für } c = 0 \,.$$

Wenn also alle c_ν reell und paarweise verschieden sind, kann man die Koeffizienten d_ν einer Funktion

$$\chi(t) = \sum_{\nu=1}^{r} d_\nu e^{c_\nu it} \qquad \text{durch} \qquad d_\nu = \lim_{T\to\infty} \frac{1}{T} \int_0^T \chi(t) e^{-c_\nu it} dt$$

bestimmen. Dies wird nun auf die Funktion $F(t) := 1 + \sum_{j=1}^{k} e(b_j t - x_j)$, $F : \mathbf{R} \to \mathbf{C}$ bzw. ihren Absolutbetrag $\Phi(t) := |F(t)|$ und deren Potenzen angewandt werden. Da die Werte der Funktion e alle auf dem Einheitskreis in $\mathbf{C}$ liegen, gilt offenbar $\Phi(t) \leq k+1$, und das Gleichheitszeichen tritt genau dann auf, wenn alle Argumente $b_j t - x_j$ ganzzahlig sind. Genauso überlegt man sich: $\Phi(t)$ liegt um so näher an $k+1$, je näher alle $b_j t - x_j$ an ganzzahligen Werten liegen. Satz 2.3 ist also äquivalent zu der Aussage über den Limes superior für $t \to \infty$

$$\overline{\lim}\, \Phi(t) \;=\; k+1\;,$$

und diese werden wir jetzt beweisen.

2. Dazu werden wir nicht nur F und Φ, sondern auch deren Potenzen betrachten. F ist von der Bauart $\psi(y_1, \ldots, y_k) = 1 + y_1 + \ldots + y_k$, also wird

$$\psi^p(y_1, \ldots, y_k) \;=\; \sum a_{n_1,\ldots,n_k} y_1^{n_1} \cdot \ldots \cdot y_k^{n_k} \quad \text{mit} \quad \sum a_{n_1,\ldots,n_k} = \psi^p(1, \ldots, 1) = (k+1)^p\;,$$

und die Anzahl der Koeffizienten $a_{n_1,\ldots,n_k} > 0$ ist $\leq (p+1)^k$, wie man durch Induktion über k einsieht: Klar für $k = 1$ nach dem binomischen Satz, und für den Induktionsschritt benutze man ihn nochmals in der Form

$$(1 + y_1 + \ldots y_k)^p \;=\; \sum_{l=0}^{p} \binom{p}{l} (1 + y_1 + \ldots y_{k-1})^l y_k^{p-l}\;.$$

3. Nun ist $F^p(t) = (1 + e(b_1 t - x_1) + \ldots + e(b_k t - x_k))^p = \psi^p(\ldots, e(b_j t - x_j), \ldots) =$

$$= \sum a_{n_1,\ldots,n_k} e(n_1 b_1 t - n_1 x_1 + n_2 b_2 t - n_2 x_2 + \ldots + n_k b_k t - n_k x_k) \;=$$

$$= \sum a_{n_1,\ldots,n_k} e(-n_1 x_1 - \ldots - n_k x_k)\, e((n_1 b_1 + \ldots + n_k b_k) t)\;.$$

Die Faktoren von t in den Exponenten, also die $2\pi(n_1 b_1 + \ldots + n_k b_k)$, sind alle paarweise verschieden, weil die b_j linear unabhängig über $\mathbf{Q}$ sind. Sie können also die Rolle der c_ν aus Abschnitt 1. des Beweises übernehmen; die zugehörigen Koeffizienten d_ν dieser Funktion χ sind dann von der Form

$$a_{n_1,\ldots,n_k} e(-n_1 x_1 - \ldots - n_k x_k) \quad \text{mit} \quad \sum |d_\nu| = \sum a_{n_1,\ldots,n_k} = (k+1)^p\;,$$

denn für reelle x ist stets $|e(x)| = 1$.

4. Angenommen, die Aussage des Satzes wäre falsch, also $\overline{\lim}\, \Phi(t) < k+1$, dann gäbe es ein $\lambda < k+1$ und ein t_0 so, dass $|F(t)| \leq \lambda$ für alle $t \geq t_0$ wäre, somit auch $|F^p(t)| \leq \lambda^p$. Daraus folgt aber

$$\overline{\lim}\, \frac{1}{T} \int_0^T |F(t)|^p dt \leq \lim \frac{1}{T} \int_0^T \lambda^p dt \;=\; \lambda^p \qquad \Rightarrow$$

$$|d_\nu| = \left| \lim \frac{1}{T} \int_0^T F(t)^p e^{-c_\nu i t} dt \right| \leq \overline{\lim}\, \frac{1}{T} \int_0^T |F(t)|^p dt\;,$$

zusammen also $a_{n_1,\ldots,n_k} = |d_\nu| \le \lambda^p$ für alle $(n_1,\ldots,n_k)$. Für die Summe der Koeffizienten hat man somit

$$\sum a_{n_1,\ldots,n_k} = (k+1)^p \le (p+1)^k \lambda^p \quad \Rightarrow \quad (\frac{k+1}{\lambda})^p \le (p+1)^k \,,$$

und das steht für große p im Widerspruch zu $\lambda < k+1$. $\square$

Anwendung. k Planeten bewegen sich auf kreisförmigen Bahnen einer festen Ebene im Raum um die Sonne mit Winkelgeschwindigkeiten $2\pi b_1, \ldots 2\pi b_k$. Zum Zeitpunkt $t = 0$ sollen sie sich in den Winkelpositionen $2\pi y_1, \ldots, 2\pi y_k$ befinden, zum Zeitpunkt t also in den Winkelpositionen $2\pi(y_1+b_1 t), \ldots, 2\pi(y_k+b_k t)$. Nun sagt Satz 2.3: Wenn die $b_1, \ldots, b_k$ linear unabhängig über $\mathbf{Q}$ sind, wird zu einem geeigneten Zeitpunkt $t > T$ jedes k–tupel $(x_1, \ldots, x_k)$ von Winkelpositionen beliebig genau erreicht.

Die Zahlentheorie liefert auch noch schöne **Beispiele** für reelle Zahlen b_j, welche über $\mathbf{Q}$ linear unabhängig sind: Man nehme $b_j := \log p_j$ für paarweise verschiedene Primzahlen. Gäbe es dann eine nichttriviale lineare Relation $r_1 b_1 + \ldots + r_k b_k = 0$ mit rationalen r_j, dann auch (Multiplikation mit dem Hauptnenner) eine mit ganzzahligen Koeffizienten n_j. Das hieße aber $p_1^{n_1} \cdot \ldots \cdot p_k^{n_k} = 1$ im Widerspruch zur Eindeutigkeit der Primfaktorzerlegung.

3 Der Approximationssatz von Dirichlet

Satz 3.1 *Seien a, $Q \in \mathbf{R}$, $Q \ge 1$. Dann gibt es $p, q \in \mathbf{Z}$ mit*

$$1 \le q \le Q \quad \textit{und} \quad |aq - p| \le \frac{1}{Q} \quad \textit{, d.h.} \quad |\, a - \frac{p}{q}| \;\le\; \frac{1}{qQ} \;\le\; \frac{1}{q^2} \quad .$$

Beweis. Sei zunächst $Q \in \mathbf{N}$. Mindestens zwei der $Q+1$ Zahlen

$$0\,,\ 1\,,\ \langle a\rangle\,,\ \langle 2a\rangle\,,\ \ldots\ ,\ \langle (Q-1)a\rangle \in [0,1]$$

liegen dann in einem der Q Intervalle $[\frac{u}{Q}, \frac{u+1}{Q}[$, $u = 0, \ldots, Q-2$ und $[\frac{Q-1}{Q}, 1]$ (das ist der DIRICHLETsche Schubfachschluss), d.h. es gibt $r_1 \ne r_2$, $s_1, s_2 \in \mathbf{Z}$ mit $0 \le r_i < Q$ und

$$|\langle r_1 a\rangle - \langle r_2 a\rangle| \le \frac{1}{Q}\,, \quad \text{d.h.} \quad |\, r_1 a - s_1 - (r_2 a - s_2)| \le \frac{1}{Q}\,.$$

$|r_1 - r_2| =: q$ tut's also mit $p = \pm(s_1 - s_2)$. Wenn $Q \notin \mathbf{N}$, verwende man den gleichen Beweis mit $Q' := [Q] + 1$. Es ergibt sich dabei ein q mit $1 \le q < Q'$ also $q \le [Q] < Q$, und wegen $Q' > Q$ ist dabei sogar $\frac{1}{Q'} < \frac{1}{Q}$, die Behauptung ist also nicht nur für Q', sondern erst recht für Q richtig.

Folgerung 3.1 *Sei $a \in \mathbf{R}$ irrational. Dann existieren unendlich viele teilerfremde Paare $(p, q) \in \mathbf{Z}^2$, $q \ge 1$, mit*

$$|\, a - \frac{p}{q}| \;<\; \frac{1}{q^2}\,.$$

Beweis. Satz 3.1 ist erst recht richtig für teilerfremde Paare (p,q). Gäbe es nur endlich viele Lösungen, so hätte auch $|aq-p| \le \frac{1}{Q}$ nur endlich viele Lösungen, allerdings alle $\neq 0$, weil $a \notin \mathbf{Q}$. Die Lösungen hätten also ein Minimum > 0, darum gäbe es ein Q_0, für das $|aq-p| \le \frac{1}{Q_0}$ unlösbar wäre im Widerspruch zum Satz. □

Für $a = \frac{u}{v} \in \mathbf{Q}$ hat $|a - \frac{p}{q}| < \frac{1}{q^2}$ nur endlich viele Lösungen, denn für $\frac{p}{q} \neq \frac{u}{v}$ (beide gekürzt) ist

$$|\frac{u}{v} - \frac{p}{q}| = |\frac{uq-pv}{vq}| \ge \frac{1}{|vq|} \ge \frac{1}{q^2} \quad \text{für} \quad q \ge |v| .$$

Da $|v|$ durch a bereits festliegt, ergibt sich genauer dadurch sogar ein Irrationalitätskriterium:

Folgerung 3.2 *Für jedes $0 < \delta \le 1$ gilt: $a \in \mathbf{R}$ ist genau dann irrational, wenn*

$$|a - \frac{p}{q}| < |q|^{-1-\delta}$$

unendlich viele teilerfremde Paare $p, q \in \mathbf{Z}$ als Lösung besitzt.

Um die Qualität der Approximation durch rationale Zahlen zu präzisieren, können wir folgenden Begriff einführen.

Definition 3.1 *$a \in \mathbf{R}$ heißt „approximierbar von Ordnung k" durch rationale Zahlen, wenn eine Konstante $c = c(a)$ existiert, so dass die Ungleichung*

$$\left|a - \frac{p}{q}\right| < \frac{c}{q^k}$$

unendlich viele (gekürzte) rationale Lösungen p/q besitzt.

In diesem Sinne kann man also sagen, dass rationale a nicht von Ordnung $k > 1$ approximierbar sind, Irrationalzahlen jedoch mindestens von Ordnung 2. Aus dem Dirichletschen Satz folgt außerdem eine quantitative Version des Kroneckerschen Satzes.

Satz 3.2 *Sei $a \in \mathbf{R}$ irrational, $x \in \mathbf{R}$, $N \in \mathbf{N}$. Dann gibt es $n \in \mathbf{N}$, $n > N$, $m \in \mathbf{Z}$ mit*

$$|na - m - x| < \frac{3}{n} .$$

Beweis. Wähle nach Folgerung 3.1 eine teilerfremde Lösung $(p,q) \in \mathbf{Z}^2$ mit $q > 2N$ von $|qa - p| < \frac{1}{q}$, dazu $Q \in \mathbf{Z}$ mit $|qx - Q| \le \frac{1}{2}$. Dazu gibt es $u, v \in \mathbf{Z}$ mit $Q = vp - uq$, o.B.d.A. mit $|v| \le \frac{1}{2}q$, folglich

$$|q(va - u - x)| = |qav - pv + Q - qx| < \frac{1}{q}|v| + \frac{1}{2} < \frac{1}{q} \cdot \frac{1}{2} \cdot q + \frac{1}{2} = 1 .$$

Setzt man nun $n := q + v$, $m := p + u$, dann wird $N < \frac{1}{2}q \le n \le \frac{3}{2}q$ und

$$|na - m - x| \le |va - u - x| + |qa - p| < \frac{1}{q} + \frac{1}{q} = \frac{2}{q} \le \frac{3}{n} . \quad \square$$

4 Liouville–Zahlen

Vorbemerkung über *algebraische Zahlen:* Dabei handelt es sich um komplexe Zahlen a, welche einer *algebraischen Gleichung* $P(a) = 0$ genügen, wo $P(x) \in \mathbf{Q}[x]$ ein Polynom mit rationalen Koeffizienten ist. Natürlich gibt es für jedes a viele solcher Polynome: Mit je zwei solchen Polynomen P, Q hat auch die Summe $P + Q \in \mathbf{Q}[x]$ die Eigenschaft $(P + Q)(a) = 0$, und für ein beliebiges $q \in \mathbf{Q}[x]$ auch das Produkt Pq. Wenn eine Untermenge eines kommutativen Rings diese Eigenschaften hat, nennt man die Untermenge ein *Ideal.* Hier kommt hinzu, dass im Polynomring über einem Körper eine Division mit Rest existiert. Man nehme nun im Ideal aller Polynome mit der Eigenschaft $Q(a) = 0$ ein Polynom kleinsten Grades P; für jedes andere Polynom Q im Ideal schreibe man

$$Q = bP + r \quad \text{mit} \quad \operatorname{grad} r < \operatorname{grad} P\,, \quad \text{aber offenbar} \quad r(a) = Q(a) - (bP)(a) = 0\,,$$

also gehört auch r zum Ideal dazu. Wegen der Annahme über den Grad geht das nur, wenn r das Nullpolynom ist. Daraus folgt:

Alle Polynome in $\mathbf{Q}[x]$ *mit Nullstelle* a *sind Vielfache des Polynoms* P*.*

Man kann es natürlich so normieren, dass der führende Koeffizient 1 wird, dann heißt es *das irreduzible Polynom zu* a, und sein Grad heißt der *Grad der algebraischen Zahl* a. (Natürlich kann man P in $\mathbf{Q}[x]$ nicht in ein Produkt pq nicht–konstanter Polynome zerlegen, warum?) Insbesondere: Wenn a den Grad 1 hat, ist $a \in \mathbf{Q}$, im Fall von Grad 2 heißt a eine *quadratische Irrationalzahl* etc. Ein erstes Ergebnis darüber, dass sich algebraische Zahlen durch rationale nicht beliebig gut approximieren lassen, stammt von Liouville aus dem Jahr 1844.

Satz 4.1 *Sei* $a \in \mathbf{R}$ *algebraisch vom Grad* d*. Dann gibt es eine Konstante* $c = c(a)$*, so dass für alle* $\frac{p}{q} \neq a$*,* $p, q \in \mathbf{Z}$*,* $q > 0$

$$|\, a - \frac{p}{q}| \;>\; \frac{c}{q^d}\,.$$

Die Approximationsordnung von a *ist also nicht größer als* d*.*

Beweis. Sei P das irreduzible Polynom zu a, nun aber durch Multiplikation mit dem Hauptnenner der Koeffizienten so normiert, dass $P \in \mathbf{Z}[x]$ mit teilerfremden Koeffizienten und führendem Koeffizienten > 0 ist. Taylorentwicklung um den Punkt a ergibt

$$P(\frac{p}{q}) \;=\; \sum_{i=1}^{d} \frac{1}{i!}\, P^{(i)}(a)\, (\frac{p}{q} - a)^i \quad \Rightarrow \quad |\, P(\frac{p}{q})| \;\leq\; \frac{1}{c} \cdot |\, \frac{p}{q} - a|\,,$$

wenn man $|\, \frac{p}{q} - a| \leq 1$ annimmt (darf man natürlich), etwa mit dem Faktor $\frac{1}{c(a)} := \sum_{i=1}^{d} \frac{1}{i!} |\, P^{(i)}(a)|$. O.B.d.A. darf man außerdem $P(\frac{p}{q}) \neq 0$ annehmen, andernfalls wäre

$d = 1$ und $a = \frac{p}{q}$. Da P ganzzahlige Koeffizienten hat, gilt aber nun

$$|\,P(\frac{p}{q})|\ \geq\ \frac{1}{q^d}\quad\Rightarrow\quad|\,a-\frac{p}{q}|\ >\ \frac{c}{q^d}\,.\quad\square$$

Inzwischen gibt es zu diesem Thema weit bessere Resultate (s.u.), damals war es aber der Schlüssel dazu, dass man die Existenz *transzendenter Zahlen* einsehen konnte, das sind reelle oder komplexe nicht–algebraische Zahlen, die also keiner algebraischen Gleichung genügen. Heutzutage haben wir es leichter: Man kann mit CANTORs Diagonalverfahren zeigen, dass alle algebraischen Zahlen eine abzählbare Menge bilden, $\mathbf{R}$ aber überabzählbar ist, in einem mengentheoretischen Sinn sind ,,die meisten" Zahlen darum transzendent — sogar die reellen unter ihnen. Abzählbarkeitsargumente gab es aber erst viel später.

Folgerung 4.1 *Zu $a \in \mathbf{R} \setminus \mathbf{Q}$ und jedem $d \in \mathbf{N}$ existiere eine rationale Zahl*

$$\frac{p}{q}\quad\textit{mit}\quad p,q\in\mathbf{Z}\,,\ q>0\,,\quad\textit{so dass}\quad|\,a-\frac{p}{q}|\ <\ \frac{1}{q^d}\,.$$

Dann ist a transzendent. Mit anderen Worten: Wenn die Approximationsordnung von a beliebig groß ist, ist a transzendent.

Solche Zahlen, die sich extrem gut durch rationale approximieren lassen, heißen *Liouville–Zahlen*. Zum *Beweis* überlege man sich zunächst, dass aus der Bedingung ,,für alle d" folgt, dass sogar jeweils unendlich viele Lösungen $p, q \in \mathbf{Z}$ existieren müssen: Man vergrößere einfach d, bis es der alte Bruch p/q nicht mehr tut. Genauso kann man q^{-d} ersetzen durch cq^{-d} für beliebige positive Konstanten c. Die Voraussetzung $a \notin \mathbf{Q}$ könnte man übrigens auch ersetzen durch die Bedingung, dass $a \neq \frac{p}{q}$ zu sein hat. Auch dann kann a nicht selbst rational sein (Folgerung 3.2). Dann folgt die Aussage direkt aus Satz 4.1. $\square$

Bleibt nur die Frage: Gibt es überhaupt solche Liouville–Zahlen?

Beispiel: Sei $a := \sum_{s=1}^{\infty} e_s 2^{-s!}$, dabei alle $e_s = 0$ oder 1, unendlich viele davon $= 1$. Man wähle p/q z.B. als abbrechenden Binärbruch, also mit

$$q = 2^{k!}\,,\quad p = 2^{k!}\cdot\sum_{s=1}^{k} e_s 2^{-s!}\,.$$

Dann wird

$$|\,a-\frac{p}{q}|\ =\ |\sum_{s>k}^{\infty} e_s 2^{-s!}|\ <\ 2\cdot 2^{-(k+1)!}\ =\ \frac{2}{q^{k+1}}\,.$$

Man sieht an der Konstruktion gleichzeitig, dass es sicher überabzählbar viele Liouville–Zahlen gibt, da für die e_s nahezu beliebige $0,1$–Folgen gewählt werden können. Andererseits lässt sich zeigen, dass in einem maßtheoretischen Sinn sehr wenig Liouville–Zahlen existieren:

Satz 4.2 *Die Menge L aller Liouville–Zahlen besitzt das (Lebesgue–)Maß 0.*

Beweis. 1. Es genügt zu zeigen, dass die Liouville–Zahlen im Intervall $[0,1]$ eine Nullmenge bilden, denn a ist Lioville–Zahl genau dann, wenn $a+n$, $n \in \mathbf{Z}$, eine Liouville–Zahl ist, also entsteht L als abzählbare Vereinigung von $\mathbf{Z}$–Translaten von $L \cap [0,1]$.
2. Zu jedem $a \in L$ und jedem $d \in \mathbf{N}$ gibt es sogar beliebig große $q \in \mathbf{N}$ und dazu $p \in \mathbf{Z}$ mit der Eigenschaft $|a - \frac{p}{q}| < q^{-d}$, siehe den Beweis von Folgerung 4.1. Wir dürfen also o.B.d.A. voraussetzen, dass $d > D$ und $q > Q$ ist.
3. $L \cap [0,1]$ liegt also für alle $d > D$ in der Vereinigung der Intervalle

$$\bigcup_{q>Q} \bigcup_{p=0}^{q} \left[\frac{p}{q} - q^{-d} \, , \, \frac{p}{q} + q^{-d} \right]$$

mit einer Gesamtlänge $< \sum_{q>Q} 2(q+1)q^{-d} < 3 \sum_{q>Q} q^{1-d} < \frac{3}{d-2} Q^{2-d}$, wenn wir die übliche Abschätzung der Reihe durch ein uneigentliches Integral verwenden.
4. Für $d > 2$ wird mit wachsendem Q die Gesamtlänge dieser Intervalle demnach beliebig klein, L ist somit eine Nullmenge. □

Der Liouvillesche Satz ist nur der Beginn einer längeren Entwicklung gewesen, die gezeigt hat, dass algebraische Zahlen nicht ,,allzu gut" durch rationale Zahlen approximierbar sind. Nach Vorarbeiten von Axel Thue (1908) und Carl Ludwig Siegel (1921) hat Klaus Friedrich Roth (1955) ein kaum zu verbesserndes Resultat gezeigt:

Satz 4.3 *Zu jeder algebraischen Zahl a und jedem $\delta > 0$ gibt es eine Konstante $c = c(a,\delta)$, so dass für alle $\frac{p}{q} \neq a$, $p, q \in \mathbf{Z}$, $q > 0$*

$$|\, a - \frac{p}{q} \,| \; > \; \frac{c}{q^{2+\delta}} \, .$$

Mit anderen Worten: algebraische Irrationalzahlen sind nicht von Ordnung $k > 2$ approximierbar.

,,Kaum zu verbessern" muss allerdings mit einer wichtigen Einschränkung versehen werden: Der Beweis von Roth erlaubt es nicht, die Konstante c explizit zu bestimmen, sie ist — so der Fachausdruck — *ineffektiv*. Das liegt u.a. an der Verwendung des Dirichletschen Schubfachschlusses, aus dem man zwar die Existenz, nicht aber die genaue Lage eines Elements ablesen kann. Insofern wird aktuell immer noch an Verbesserungen gearbeitet, z.T. für spezielle Klassen von Zahlen; man nimmt dabei schlechtere (d.h. größere) Exponenten des Nenners q in Kauf, wenn man dafür die Konstante effektiv berechnen kann.

Auch eine andere Schattenseite dieser Sätze sei nicht verschwiegen: Als Transzendenzbeweise kann man sie nur anwenden auf Zahlen, die man extra zu diesem Zweck konstruiert wie z.B. die Liouville–Zahlen. Für Zahlen, welche in der ,,Natur" vorkommen wie z.B. e und π, lassen sich Transzendenzbeweise nur mit anderen Mitteln führen.

5 Der Kettenbruchformalismus

Seien zunächst $a_0, \ldots, a_n$ Variablen, $p_0, q_0, \ldots, p_n, q_n$ Polynome in $a_0, \ldots, a_n$ für $n \in \mathbf{N}_0$ über einem beliebigem Grundkörper. Wir definieren die Polynome p_i, q_i rekursiv durch $p_0 := a_0$, $q_0 := 1$ und für $p'_k = p_k(a_1, \ldots, a_{k+1})$ und $q'_k = q_k(a_1, \ldots, a_{k+1})$ sei

$$p_{k+1} := a_0 p'_k + q'_k, \qquad q_{k+1} := p'_k.$$

Insbesondere gilt:

$$\begin{array}{llll} p_0 = a_0, & p'_0 = a_1, & p_1 = a_0a_1 + 1, & p'_1 = a_1a_2 + 1, \quad p_2 = a_0a_1a_2 + a_0 + a_2 \\ q_0 = 1, & q'_0 = 1, & q_1 = a_1, & q'_1 = a_2, \qquad\quad q_2 = a_1a_2 + 1 \end{array}$$

Dann heißt $[a_0, \ldots, a_n] := \dfrac{p_n}{q_n}$ *(endlicher) Kettenbruch* und ist eine rationale Funktion von $a_0, a_1, \ldots, a_n$. **Achtung:** Die eckigen Klammern hier haben nichts zu tun mit Grenzen abgeschlossener Intervalle, kgV oder Gaußklammern! Insbesondere gilt:

$$\begin{aligned} [a_0] &= \frac{p_0}{q_0} = a_0 \\ [a_0, a_1] &= \frac{p_1}{q_1} = a_0 + \frac{1}{a_1} \\ [a_0, a_1, a_2] &= \frac{p_2}{q_2} = \frac{a_0a_1a_2 + a_0 + a_2}{a_1a_2 + 1} = a_0 + \frac{a_2}{a_1a_2 + 1} \\ &= a_0 + \frac{1}{\frac{a_1a_2+1}{a_2}} = a_0 + \frac{1}{a_1 + \frac{1}{a_2}} \\ &\vdots \\ [a_0, \ldots, a_n] &= \frac{p_n}{q_n} = \frac{a_0p'_{n-1} + q'_{n-1}}{p'_{n-1}} = a_0 + \frac{1}{\frac{p_{n-1}(a_1,\ldots,a_n)}{q_{n-1}(a_1,\ldots,a_n)}} \\ &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}} \end{aligned}$$

Die Werte $\dfrac{p_k}{q_k}$ heißen *Nährungsbrüche* von $\dfrac{p_n}{q_n}$, die a_k die *Teilnenner*, die p_k heißen *Näherungszähler* und die q_k *Näherungsnenner*.

Lemma 5.1 *Für alle $n \geq 2$ gilt:*

$$p_n = a_n p_{n-1} + p_{n-2}, \qquad q_n = a_n q_{n-1} + q_{n-2}$$

oder äquivalent:

$$\begin{bmatrix} p_n \\ q_n \end{bmatrix} = \begin{bmatrix} p_{n-1} & p_{n-2} \\ q_{n-1} & q_{n-2} \end{bmatrix} \cdot \begin{bmatrix} a_n \\ 1 \end{bmatrix}$$

Der *Beweis* erfolgt per Induktion über n. Für $n = 2$ gilt:

$$p_2 = a_0a_1a_2 + a_0 + a_2 = a_2(a_0a_1 + 1) + a_0 = a_2p_1 + p_0$$
$$q_2 = a_1a_2 + 1 = a_2q_1 + q_0$$

Sei $n > 2$ und die Behauptung für $n - 1$ gezeigt:

$$p'_{n-1} = a_np'_{n-2} + p'_{n-3}, \qquad q'_{n-1} = a_nq'_{n-2} + q'_{n-3}.$$

Nach Definition gilt

$$\begin{aligned} p_n &= a_0p'_{n-1} + q'_{n-1} \\ &= a_0(a_np'_{n-2} + p'_{n-3}) + a_nq'_{n-2} + q'_{n-3} \\ &= a_n(a_0p'_{n-2} + q'_{n-2}) + (a_0p'_{n-3} + q'_{n-3}) \\ &= a_np_{n-1} + p_{n-2} \end{aligned}$$

und

$$\begin{aligned} q_n &= p'_{n-1} \\ &= a_np'_{n-2} + p'_{n-3} \\ &= a_nq_{n-1} + q_{n-2} \end{aligned}$$

Die Äquivalenz zur Matrizenschreibweise ist offensichtlich. □

Setzen wir $p_{-2} := q_{-1} := 0$ und $p_{-1} := q_{-2} := 1$, gilt Lemma 5.1 auch für $n = 0, 1$:

$$\begin{bmatrix} p_0 \\ q_0 \end{bmatrix} = \begin{bmatrix} a_0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} = \begin{bmatrix} a_0a_1 + 1 \\ a_1 \end{bmatrix} = \begin{bmatrix} a_0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ 1 \end{bmatrix}$$

Lemma 5.2 *Sei* $1 \le k \le n$ *und* $r_k = [a_k, a_{k+1}, \ldots, a_n]$. *Dann gilt:*

$$[a_0, \ldots, a_n] = [a_0, \ldots, a_{k-1}, r_k] = \frac{p_{k-1}r_k + p_{k-2}}{q_{k-1}r_k + q_{k-2}}$$

Beweis. Die zweite Gleichung erhält man aus Lemma 5.1, indem man n als k und a_n als r_k wählt. Die erste Gleichung beweisen wir per Induktion über k. Sei $k = 1$. Dann ist:

$$\begin{aligned} [a_0, \ldots, a_n] = \frac{p_n}{q_n} = \frac{a_0p'_{n-1} + q'_{n-1}}{p'_{n-1}} &= a_0 + \frac{1}{\frac{p'_{n-1}}{q'_{n-1}}} \\ = a_0 + \frac{1}{[a_1, \ldots, a_n]} &= a_0 + \frac{1}{r_1} = [a_0, r_1] \end{aligned}$$

Sei $k > 1$ und die Behauptung gelte für $k - 1$. Dann folgt nach Induktionsvoraussetzung:

$$[a_0, \ldots, a_n] = a_0 + \frac{1}{[a_1, \ldots, a_n]} = a_0 + \frac{1}{[a_1, \ldots, a_{k-1}, r_k]} = [a_0, \ldots, a_{k-1}, r_k] \qquad . \quad \square$$

Lemma 5.3 *Für alle* $n \geq -1$ *ist* $q_n p_{n-1} - p_n q_{n-1} = (-1)^n$.

Beweis durch Induktion über $n \geq -1$. Der Induktionsanfang für $n = -1$ ist trivial. Sei $n \geq 0$. Da Addition eines Vielfachen einer Spalte zur einer anderen die Determinante nicht ändert, gilt nach Lemma 5.1

$$\det \begin{bmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{bmatrix} = \det \begin{bmatrix} p_{n-1} & p_{n-2} \\ q_{n-1} & q_{n-2} \end{bmatrix} = -\det \underbrace{\begin{bmatrix} p_{n-2} & p_{n-1} \\ q_{n-2} & q_{n-1} \end{bmatrix}}_{= (-1)^{n-1} \text{ nach Ind.Annahme}} = (-1)(-1)^{n-1} = (-1)^n,$$

was zu zeigen war. □

Lemma 5.4 *Für alle* $n \geq 0$ *gilt:* $p_{n-2} q_n - p_n q_{n-2} = (-1)^{n-1} a_n$.

Beweis. Nach Lemma 5.1 und Lemma 5.3 gilt:

$$\det \begin{bmatrix} p_{n-2} & p_n \\ q_{n-2} & q_n \end{bmatrix} = a_n \cdot \det \begin{bmatrix} p_{n-2} & p_{n-1} \\ q_{n-2} & q_{n-1} \end{bmatrix} = (-1)^{n-1} a_n. \quad \square$$

Aus der Kettenbruchdarstellung von $\dfrac{p_n}{q_n} = [a_0, \ldots, a_n]$ erhalten wir eine für den Quotienten $\dfrac{q_n}{q_{n-1}}$:

Lemma 5.5 *Für alle* $n \geq 1$ *gilt:* $\dfrac{q_n}{q_{n-1}} = [a_n, \ldots, a_2, a_1]$.

Der *Beweis* erfolgt per Induktion über n . Für $n = 1$ ist

$$\frac{q_1}{q_0} = a_1 = [a_1] \quad .$$

Sei $n > 1$. Dann gilt nach Induktionsvoraussetzung:

$$\frac{q_n}{q_{n-1}} = \frac{a_n q_{n-1} + q_{n-2}}{q_{n-1}} = a_n + \frac{1}{\dfrac{q_{n-1}}{q_{n-2}}} = a_n + \frac{1}{[a_{n-1}, \ldots, a_1]} = [a_n, \ldots, a_2, a_1] \qquad . \qquad \square$$

6 Kettenbruchentwicklung reeller Zahlen

Das folgende Lemma werden wir später in diesem Abschnitt zum Beweis brauchen, daß die Folge der Näherungsbrüche gegen einen Grenzwert konvergiert.

Lemma 6.1 *Seien* $a_0, a_1, \ldots, a_n \in \mathbf{R}$, *alle* $a_n > 0$. *Für die daraus gebildeten Kettenbrüche ist:*

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \cdots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}$$

Beweis. Nach Lemma 5.4 gilt:

$$\frac{p_{n-2}}{q_{n-2}} - \frac{p_n}{q_n} = \frac{(-1)^{n-1}a_n}{q_{n-2}q_n}$$

Dabei ist $q_0 = 1$ und $\dfrac{q_n}{q_{n-1}} = [a_n, a_{n-1}, \ldots, a_1] > 0$ nach Lemma 5.5. Wegen $a_n, q_n > 0$ für alle n folgt für die rechte Seite der Gleichung:

$$\frac{(-1)^{n-1}a_n}{q_{n-2}q_n} \qquad \begin{cases} < 0 & \text{für gerade} \quad n \geq 2 \\ > 0 & \text{für ungerade} \quad n \end{cases} .$$

Mit Lemma 5.3 und

$$\frac{p_{n-1}}{q_{n-1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n q_{n-1}}$$

folgt daraus die Behauptung. □

Lemma 6.2 *Sei* $a_0 \in \mathbf{Z}$ *und für alle* $n \in \mathbf{N}$ *seien die* $a_n \in \mathbf{N}$. *Dann gilt* $[a_0, a_1, \ldots, a_n] \in \mathbf{Q}$. *Sei umgekehrt* $\frac{u}{v} \in \mathbf{Q}$, *dann existieren* $n \in \mathbf{N}_0$, $a_0 \in \mathbf{Z}$ *und* $a_1, \ldots, a_n \in \mathbf{N}$ *mit*

$$\frac{u}{v} = [a_0, a_1, \ldots, a_n] \quad ,$$

wobei $a_0 \geq 1$ *für* $\frac{u}{v} \geq 1$.

Die erste Aussage ist offenbar richtig. Für die zweite Behauptung können wir o.B.d.A. annehmen, daß $v \in \mathbf{N}$ und teilerfremd zu u ist. *Beweis* durch Induktion über v:

- Induktionsanfang $v = 1$: Es gilt $\frac{u}{v} = u \in \mathbf{Z}$. Setze $n := 0$ und $a_0 := u$.
- Induktionsschluss: Es gibt $q, r \in \mathbf{Z}$ mit $1 \leq r < v$ und $u = vq + r$ (Division mit Rest). Somit:

  $$\frac{u}{v} = q + \frac{r}{v} = q + \frac{1}{\frac{v}{r}}$$

  Wegen $1 \leq r < v$ existiert nach Induktionsannahme eine Kettenbruchentwicklung

  $$\frac{v}{r} = [a_1, \ldots, a_n] \quad ,$$

  alle $a_k \in \mathbf{N}$. Nach §5 erhalten wir mit $a_0 := q \in \mathbf{Z}$:

  $$\frac{u}{v} = [a_0, a_1, \ldots, a_n] \qquad \square$$

Definition 6.1 (Einfacher Kettenbruch, Näherungsbruch, vollständiger Quotient) *Ein (endlicher) einfacher Kettenbruch für* $r \in \mathbf{Q}$ *ist*

$$r = [a_0, a_1, \ldots, a_n]$$

mit $a_0 \in \mathbf{Z}$*, alle anderen* $a_k \in \mathbf{N}$*. Wir nennen*

$$\frac{p_i}{q_i} := [a_0, a_1, \ldots, a_i]$$

den i*-ten Näherungsbruch,* a_i *den* i*-ten Teilnenner und*

$$\alpha_i := [a_i, a_{i+1}, \ldots, a_n]$$

seinen i*-ten vollständigen Quotienten.*

Nach Lemma 5.3 ist der i-te Näherungsbruch $\dfrac{p_i}{q_i}$ gekürzt.

Lemma 6.3 *1. Jedes* $r \in \mathbf{Z}$ *hat genau zwei einfache Kettenbruchentwicklungen, nämlich* $r = [r]$ *und* $[r-1, 1] = r - 1 + \frac{1}{1}$*.*

2. Jedes $r \in \mathbf{Q} \setminus \mathbf{Z}$ *hat genau zwei einfache Kettenbruchentwicklungen, nämlich (mit* $a_n \geq 2$*)*

$$r = [a_0, a_1, \ldots, a_n] = [a_0, a_1, \ldots, a_n - 1, 1]$$

Der *Beweis* der Gleichungen ist klar. Eindeutigkeitsaussage:

1. Sei $r \in \mathbf{Z}$ mit einfacher Kettenbruchentwicklung $r = [a_0, a_1, \ldots, a_n]$, $a_0 \in \mathbf{Z}$, alle anderen $a_n \in \mathbf{N}$. somit

$$[a_1, \ldots, a_n] = a_1 + \frac{1}{[a_2, a_3, \ldots, a_n]} \geq 1$$

Die Gleichheit

$$r = [a_0, a_1, \ldots, a_n] = a_0 + [a_1, \ldots, a_n]^{-1}$$

gilt darum genau dann, wenn $a_1 = 1$ und $n = 1$.

2. Sei jetzt $r = \frac{u}{v}$ mit $(u, v) = 1$, $v \in \mathbf{N}$ und $v \geq 2$: Dann ist

$$r = a_0 + \frac{1}{[a_1, \ldots, a_n]}$$

mit $a_0 = [r]$ (diesmal ist's die Gaußklammer!), denn $[a_1, \ldots, a_n]^{-1} \in (0, 1]$, und der Wert 1 wird nur für $r \in \mathbf{Z}$ angenommen, darum ist a_0 eindeutig bestimmt. Es gilt $[a_1, \ldots, a_n] > 1$. Setze

$$r_1 := \frac{v_1}{u_1} = [a_1, \ldots, a_n] > 1$$

mit $0 < u_1 < v_1$. Wir führen einen Induktionsbeweis über den Nenner v:

- Induktionsanfang $v = 1 \Leftrightarrow r_1 \in \mathbf{Z}$, vgl. den schon bewiesenen ersten Teil des Lemmas.
- Induktionsschluss: Sei die Aussage für alle Nenner kleiner v gezeigt.

  $$r = \frac{u}{v} = a_0 + \frac{1}{r_1} = a_0 + \frac{u_1}{v_1}$$

  Da u_1 und v_1 teilerfremd sind, gilt $v_1 = v$, d.h. r_1 hat den Nenner $u_1 < v$ und nach Induktionsannahme hat r_1 nur die beiden einfachen Kettenbruchentwicklungen

  $$r_1 = [a_1, \ldots, a_n] = [a_1, a_2, \ldots, a_n - 1, 1] \quad .$$

  Einsetzen nach Lemma 5.2 liefert die Behauptung. □

Definition 6.2 (Kettenbruchentwicklung reeller Zahlen) *Die Kettenbruchentwicklung von $r \in \mathbf{R}$ entsteht durch ein Verfahren analog zum euklidischen Algorithmus: Sei $r_0 := r$. Wir berechnen $a_0 \in \mathbf{Z}$ als Gaußklammer von $r_0 = r$ und ersetzen, wenn $a_0 \neq r$, dieses r_0 durch $r_1 := \langle r_0 \rangle^{-1} > 1$, beide eindeutig bestimmt mit der Eigenschaft*

$$r_0 = a_0 + \frac{1}{r_1},$$

Wir iterieren diesen Prozess

$$r_i = a_i + \frac{1}{r_{i+1}} \quad \textit{bzw.} \quad r_{i+1} := \frac{1}{\langle r_i \rangle} \quad ,$$

so dass

$$r = [a_0, a_1, \ldots, a_{i-1} + \frac{1}{r_i}]$$

Wenn $r_i = a_i \in \mathbf{N}$ ist, stoppen wir (was natürlich nur vorkommt, wenn $r \in \mathbf{Q}$ ist).

Die Kettenbruchentwicklung hat darüber hinaus eine geometrische Interpretation. Sei $r = \frac{u}{v} \in \mathbf{Q}$ mit teilerfremden $u, v \in \mathbf{Z}$ und $v > 1$. Der Gitterpunkt $(v, u) \in \mathbf{Z}^2$ ist (bezüglich $\mathbf{Z}^2$) ein *primitiver Gittervektor*, d.h. einziger Gitterpunkt auf der Strecke $\overrightarrow{(0,0)(v,u)}$. Dem i-ten Näherungsbruch $\frac{p_i}{q_i}$ entspricht ebenso ein primitiver Gitterpunkt $(q_i, p_i) \in \mathbf{Z}^2$. Nach Lemma 5.3 ist

$$\det \begin{bmatrix} q_i & q_{i-1} \\ p_i & p_{i-1} \end{bmatrix} = (-1)^i$$

Äquivalent dazu ist die Aussage: Das Dreieck mit den Punkten $0, \begin{bmatrix} q_i \\ p_i \end{bmatrix}, \begin{bmatrix} q_{i-1} \\ p_{i-1} \end{bmatrix}$ enthält keinen weiteren Gitterpunkt in $\mathbf{Z}^2$, denn die Vektoren $\begin{bmatrix} q_i \\ p_i \end{bmatrix}$ und $\begin{bmatrix} q_{i-1} \\ p_{i-1} \end{bmatrix}$ bilden eine

Gitterbasis von $\mathbf{Z}^2$. Nach Lemma 6.1 liegt $r = \dfrac{p_n}{q_n}$ zwischen den aufeinander folgenden Näherungsbrüchen $\dfrac{p_i}{q_i}$ und $\dfrac{p_{i-1}}{q_{i-1}}$. Wir beginnen die Kettenbruchentwicklung mit der größten ganzen Zahl kleiner als r, also mit $\dfrac{p_0}{q_0} = \dfrac{p_0}{1}$ und beachten, dass

$$q_0 = 1 < q_1 < q_2 < \cdots < q_n \qquad ,$$

und dass es keine anderen Gitterpunkte des $\mathbf{Z}^2$ im Dreieck $0, \begin{bmatrix} q_i \\ p_i \end{bmatrix}, \begin{bmatrix} q_{i-1} \\ p_{i-1} \end{bmatrix}$ gibt, wie man an der Determinanten–Bedingung sieht.

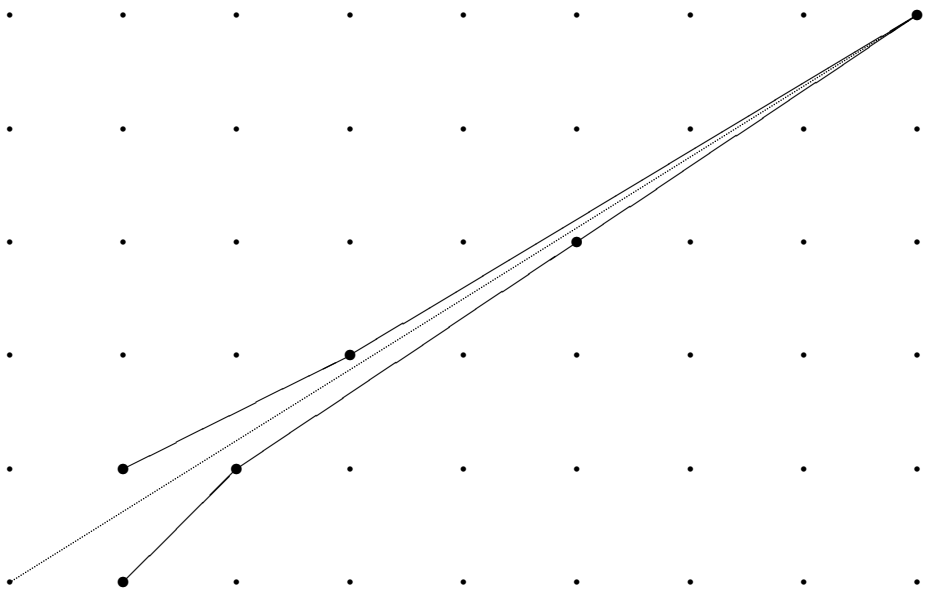

Abbildung 1: Der Kettenbruch $[0, 1, 1, 1, 1, 1]$

Wir wollen diese geometrische Interpretation anhand des **Beispiels** $r = \frac{5}{8}$ darstellen. Die Kettenbruchentwicklung lautet:

$$r = 0 + \frac{1}{\frac{8}{5}} = 0 + \frac{1}{1 + \frac{1}{\frac{5}{3}}} = 0 + \frac{1}{1 + \dfrac{1}{1 + \dfrac{1}{1 + \frac{1}{2}}}} = [0, 1, 1, 1, 2] = [0, 1, 1, 1, 1, 1]$$

Die Näherungsbrüche sind

$$\frac{p_0}{q_0} = 0 \qquad \frac{p_1}{q_1} = [0,1] = 1$$
$$\frac{p_2}{q_2} = [0,1,1] = \frac{1}{2} \qquad \frac{p_3}{q_3} = [0,1,1,1] = \frac{2}{3}$$

und $\frac{p_4}{q_4} = [0,1,1,1,1] = \frac{3}{5}$, schließlich $\frac{p_5}{q_5} = [0,1,1,1,1,1] = \frac{5}{8}$. Die Abbildung zeigt die Näherungsbrüche in Form von Gitterpunkten des $\mathbf{Z}^2$.

Ein weiteres Beispiel ist $r = \frac{5}{17}$ mit der Kettenbruchentwicklung

$$r = 0 + \frac{1}{\frac{17}{5}} = 0 + \frac{1}{3 + \frac{1}{\frac{5}{2}}} = 0 + \frac{1}{3 + \frac{1}{2 + \frac{1}{1 + \frac{1}{1}}}} = [0,3,2,2] = [0,3,2,1,1]$$

und den Näherungsbrüchen

$$\frac{p_0}{q_0} = 0 \qquad \frac{p_1}{q_1} = [0,3] = \frac{1}{3}$$
$$\frac{p_2}{q_2} = [0,3,2] = \frac{2}{7} \qquad \frac{p_3}{q_3} = [0,3,2,1] = \frac{3}{10}$$

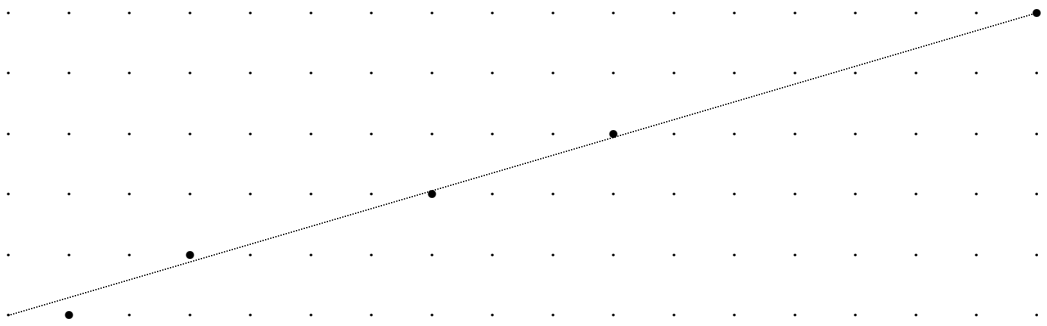

Abbildung 2: Der Kettenbruch $[0,3,2,1,1]$

und schließlich $\frac{p_4}{q_4} = [0,3,2,1,1] = \frac{5}{17}$

Die Verbindungsgerade zwischen den Punkten 0 und $\begin{bmatrix} q_n \\ p_n \end{bmatrix}$ wird durch zwei Streckenzüge eingefasst, nämlich

$$\begin{bmatrix} q_0 \\ p_0 \end{bmatrix}, \begin{bmatrix} q_2 \\ p_2 \end{bmatrix}, \ldots \qquad \text{und} \qquad \begin{bmatrix} q_1 \\ p_1 \end{bmatrix}, \begin{bmatrix} q_3 \\ p_3 \end{bmatrix}, \ldots \quad .$$

Innerhalb dieser Streckenzüge liegen keine weiteren Gitterpunkte (Beweis?). Die Streckenzüge sind nach oben bzw. nach unten konvex (Beweis??).

Satz 6.1 *1. Sei* $a_0 \in \mathbf{Z}$ *und für alle* $i \geq 1$ *sei* $a_i \in \mathbf{N}$. *Für die Näherungsbrüche* $\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n]$ *existiert der Grenzwert:*

$$\lim_{n\to\infty} \frac{p_n}{q_n} =: \alpha \in \mathbf{R} \setminus \mathbf{Q}$$

2. Für alle $\alpha \in \mathbf{R} \setminus \mathbf{Q}$ *existieren eindeutig bestimmte* $a_0 \in \mathbf{Z}$ *und* $a_1, a_2, \ldots \in \mathbf{N}$, *so dass für die Näherungsbrüche* $\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n]$

$$\lim_{n\to\infty} \frac{p_n}{q_n} = \alpha.$$

Wir nennen den Grenzwert

$$\lim_{n\to\infty} [a_0, a_1, \ldots, a_n] =: [a_0, a_1, \ldots]$$

den *einfachen Kettenbruch* für $\alpha \in \mathbf{R} \setminus \mathbf{Q}$. Entsprechend heißt $\frac{p_i}{q_i} := [a_0, a_1, \ldots, a_i]$ der *i-te Näherungsbruch,* a_i *der i-te Teilnenner und* $\alpha_i := [a_i, a_{i+1}, \ldots]$ *der i-te vollständige Quotient.* Der i-te Näherungsbruch liefert eine explizite Form des Dirichlet'schen Approximationssatzes 3.1 :

$$\left|\alpha - \frac{p_i}{q_i}\right| < \frac{1}{q_i^2}$$

Beweis. 1. Nach Lemma 6.1 gilt

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \cdots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1},$$

so dass die Folge der Näherungsbrüche mit geraden Indizies monoton wachsend und nach oben beschränkt ist und somit der Grenzwert $\lim_{n\to\infty} \frac{p_{2n}}{q_{2n}}$ existiert. Analog folgt, dass der der Grenzwert $\lim_{n\to\infty} \frac{p_{2n+1}}{q_{2n+1}}$ existiert. Nach Lemma 5.4 gilt

$$\frac{p_{n-1}}{q_{n-1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_{n-1}q_n}$$

und aus $q_n = a_n q_{n-1} + q_{n-2}$ folgt $q_n \geq q_{n-1} + 1 > q_{n-1}$, so dass wir

$$\lim_{n\to\infty} \left(\frac{p_{n-1}}{q_{n-1}} - \frac{p_n}{q_n}\right) = 0$$

erhalten. Folglich gilt:

$$\lim_{n\to\infty} \frac{p_{2n}}{q_{2n}} \;=\; \lim_{n\to\infty} \frac{p_{2n+1}}{q_{2n+1}}$$

Sei α dieser Grenzwert, der zwischen $\dfrac{p_n}{q_n}$ und $\dfrac{p_{n+1}}{q_{n+1}}$ liegt. Wir erhalten eine effektives Verfahren für den Dirichlet'schen Approximationssatzes 3.1, denn es gilt:

$$\left|\alpha - \frac{p_n}{q_n}\right| \;<\; \frac{1}{q_{n+1}q_n} \;<\; \frac{1}{q_n^2}$$

Es existieren unendlich viele gekürzte $\frac{p}{q}$ mit $|\alpha - \frac{p}{q}| < \frac{1}{q^2}$ und somit gilt $\alpha \notin \mathbf{Q}$ (denn nach §3 gibt es für $\alpha \in \mathbf{Q}$ nur endlich viele).

2. Sei $\alpha \in \mathbf{R} \setminus \mathbf{Q}$, $a_0 := [\alpha]$ (Gaußklammer) mit $\alpha = a_0 + \frac{1}{\alpha_1}$ mit $\alpha_1 > 1$ (sonst wäre $a_0 \neq [\alpha]$) und $\alpha_1 \notin \mathbf{Q}$ (sonst wäre $\alpha \in \mathbf{Q}$). Induktiv folgt, dass für alle $k \in \mathbf{N}$

$$\alpha = [a_0, a_1, \ldots, a_{k-1}, \alpha_k] \qquad \text{mit } \alpha_k \notin \mathbf{Q} \text{ und } \alpha_k > 1 \;.$$

Betrachte den endlichen (weil nicht einfachen) Kettenbruch

$$\alpha \;=\; [a_0, a_1, \ldots, a_n, \alpha_{n+1}]$$

mit Näherungsbrüchen $\dfrac{\overline{p}_{n-1}}{\overline{q}_{n-1}} = \dfrac{p_{n-1}}{q_{n-1}}$, $\dfrac{\overline{p}_n}{\overline{q}_n} = \dfrac{p_n}{q_n}$ und $\dfrac{\overline{p}_{n+1}}{\overline{q}_{n+1}} = \alpha$. Es gilt

$$\overline{q}_n\alpha - \overline{p}_n = \overline{q}_n \cdot \frac{\overline{p}_{n+1}}{\overline{q}_{n+1}} - \overline{p}_n = \frac{-\left(\overline{p}_n\overline{q}_{n+1} - \overline{p}_{n+1}\overline{q}_n\right)}{\overline{q}_{n+1}} \;.$$

Nach Lemma 5.3 und Lemma 5.1 folgt

$$\overline{q}_n\alpha - \overline{p}_n = \frac{(-1)^n}{\overline{q}_{n+1}} = \frac{(-1)^n}{\overline{q}_n\alpha_{n+1} + \overline{q}_{n-1}} \;.$$

Wir erhalten wegen $\alpha_{n+1} > 1$

$$\left|\alpha - \frac{p_n}{q_n}\right| \;=\; \frac{1}{\left|q_n\left(\,\overline{q}_n\alpha_{n+1} + \overline{q}_{n-1}\right)\right|} \;<\; \frac{1}{q_n^2} \;. \tag{1}$$

Weil $q_n \in \mathbf{N}$ monoton wächst ($q_{n+1} = a_{n+1}q_n + q_{n-1}$), gilt somit $\lim_{n\to\infty} \dfrac{p_n}{q_n} = \alpha$.

Die Eindeutigkeit beweist man durch Induktion über n. Für den Induktionsanfang beachte, dass $a_0 = [\alpha]$ gilt. Im Induktionsschritt verwende, dass nach Annahme $a_0, a_1, \ldots, a_{n-1}$ und $\alpha_n > 1$ eindeutig bestimmt sind, so dass auch $a_n = [\alpha_n]$ eindeutig bestimmt ist. □

7 Optimale Approximation durch rationale Zahlen

Satz 7.1 (Legendre) *Sei* $\alpha \in \mathbf{R}$, *dazu* $p \in \mathbf{Z}$ *und* $q \in \mathbf{N}$ *teilerfremd mit* $|\alpha - \frac{p}{q}| < \frac{1}{2q^2}$. *Dann ist* $\frac{p}{q}$ *Näherungsbruch der Kettenbruchentwicklung von* α.

Beweis. O.B.d.A. sei $\alpha \neq \frac{p}{q}$ und $\alpha - \frac{p}{q} = \frac{\epsilon\vartheta}{q^2}$ mit $\epsilon = \pm 1$ und $0 < \vartheta < \frac{1}{2}$. Es existiert eine einfache Kettenbruchentwicklung von $\frac{p}{q} =: [b_0, b_1, \ldots, b_{n-1}]$ mit $(-1)^{n-1} = \epsilon$ (wegen der Zweideutigkeit der Kettenbruchentwicklung ist $[b_0, \ldots, b_{n-1}] = [b_0, \ldots, b_{n-1} - 1, +1]$ möglich) und Näherungsbrüchen $\frac{p_i}{q_i}$. Sei $\omega \in \mathbf{R}$ so definiert, dass

$$\alpha \;=\; \frac{\omega p_{n-1} + p_{n-2}}{\omega q_{n-1} + q_{n-2}} \tag{2}$$

erfüllt ist. Dies ist möglich, da $x \mapsto \frac{xp_{n-1} + p_{n-2}}{xq_{n-1} + q_{n-2}}$ eine Bijektion $\mathbf{R} \cup \{\infty\} \to \mathbf{R} \cup \{\infty\}$ ist, wobei genau dann $\infty \mapsto \alpha$, wenn $\alpha = \frac{p_{n-1}}{q_{n-1}} = \frac{p}{q}$ ist, was wir ausgeschlossen haben. Mit Hilfe von Lemma 5.3 – und nachdem man sich einige Male verrechnet hat – erhält man

$$\frac{\epsilon\vartheta}{q^2} \;=\; \alpha - \frac{p}{q} \;=\; \alpha - \frac{p_{n-1}}{q_{n-1}} \;=\; \frac{1}{q_{n-1}}\left(\alpha q_{n-1} - p_{n-1}\right) \;=\; \frac{1}{q_{n-1}} \cdot \frac{(-1)^{n-1}}{\omega q_{n-1} + q_{n-2}} \quad ,$$

$$\text{und} \qquad \vartheta \;=\; \frac{q_{n-1}}{\omega q_{n-1} + q_{n-2}} \quad \Rightarrow \quad \omega \;=\; \frac{1}{\vartheta} - \frac{q_{n-2}}{q_{n-1}} \;>\; 2 - 1 \;=\; 1\,,$$

also hat ω eine Kettenbruchentwicklung $[b_n, b_{n+1}, \ldots]$, alle $b_j \in \mathbf{N}$ für $j \geq n$. Folglich ist

$$\alpha \;=\; [b_0, b_1, \ldots, b_{n-1}, [b_n, b_{n+1} \ldots]] \;=\; [b_0, b_1, \ldots, b_{n-1}, b_n, \ldots]$$

eine einfache Kettenbruchentwicklung mit Näherungsbruch $\frac{p}{q} \;=\; \frac{p_{n-1}}{q_{n-1}}$, eindeutig bestimmt nach Lemma 6.3 und Satz 6.1. □

Satz 7.2 (Lagrange 1770) *Sei* $\alpha \in \mathbf{R} \setminus \mathbf{Q}$ *mit Näherungsbrüchen* $\frac{p_0}{q_0}, \frac{p_1}{q_1}, \ldots$ *der Kettenbruchentwicklung. Dann gilt:*

1. $|\alpha q_0 - p_0| > |\alpha q_1 - p_1| > |\alpha q_2 - p_2| > \ldots$
2. *Für* $n \in \mathbf{N}$ *und* $1 \leq q \leq q_n$ *gilt: Wenn* $(q, p) \neq (q_n, p_n)$ *und* $(q, p) \neq (q_{n-1}, p_{n-1})$, *ist*

$$|\alpha q - p| \;>\; |\alpha q_{n-1} - p_{n-1}|$$

Beweis. Wegen $\alpha \notin \mathbf{Q}$ gilt für den n–ten vollständigen Quotienten $\alpha_n > 1$. Wir wissen aus Lemma 5.2

$$\alpha \;=\; \frac{p_n\alpha_{n+1} + p_{n-1}}{q_n\alpha_{n+1} + q_{n-1}}$$

und erhalten durch Auflösen dieser Gleichung nach α_{n+1} mit Lemma 5.3

$$|\alpha q_n - p_n| \;=\; \frac{1}{\alpha_{n+1}q_n + q_{n-1}} \;<\; \frac{1}{q_n + q_{n-1}} \quad .$$

Ganz entsprechend ist wegen $q_{n-1}a_n + q_{n-2} = q_n$

$$|\alpha q_{n-1} - p_{n-1}| \;=\; \frac{1}{\alpha_n q_{n-1} + q_{n-2}} \;>\; \frac{1}{(a_n+1)q_{n-1} + q_{n-2}} \;=\; \frac{1}{q_n + q_{n-1}} \quad ,$$

und aus beiden Ungleichungen folgt die Behauptung 1.

Für Aussage 2 beachte, dass die Vektoren $\begin{bmatrix} q_n \\ p_n \end{bmatrix}$ und $\begin{bmatrix} q_{n-1} \\ p_{n-1} \end{bmatrix}$ eine Gitterbasis von $\mathbf{Z}^2$ bilden. Daher existieren zu $\begin{bmatrix} q \\ p \end{bmatrix}$ Koeffizienten $\mu, \nu \in \mathbf{Z}$ mit

$$\begin{aligned} \mu q_n + \nu q_{n-1} &= q \\ \mu p_n + \nu p_{n-1} &= p \quad . \end{aligned}$$

Die Fälle $\mu = 0$ oder $\nu = 0$ können wir ausschließen wegen $(q,p) \neq (q_n, p_n)$ und $(p,q) \neq (q_{n-1}, p_{n-1})$. Wegen $q \leq q_n$ haben μ und ν verschiedene Vorzeichen, d.h. es gilt $\nu\mu < 0$ (sonst wäre $\mu q_n + \nu q_{n-1} > q$ wegen $q_n, q_{n-1} \in \mathbf{N}$).

Die Zahlen $\alpha q_n - p_n$ und $\alpha q_{n-1} - p_{n-1}$ haben verschiedene Vorzeichen, so dass $\mu(\alpha q_n - p_n)$ und $\nu(\alpha q_{n-1} - p_{n-1})$ die gleichen Vorzeichen haben. Es folgt:

$$|\alpha q - p| \;=\; |\mu(\alpha q_n - p_n)| + |\nu(\alpha q_{n-1} - p_{n-1})|$$

Wegen $|\mu|, |\nu| \geq 1$ erhalten wir die Behauptung von Aussage 2 :

$$|\alpha q - p| \;>\; |\alpha q_{n-1} - p_{n-1}| \qquad \square$$

Definition 7.1 *Eine Zahl $\alpha \in \mathbf{R} \setminus \mathbf{Q}$ heißt „schlecht approximierbar" (durch rationale Zahlen), wenn eine Konstante $c = c(\alpha) > 0$ existiert, so dass*

$$\left|\alpha - \frac{p}{q}\right| \;>\; \frac{c}{q^2}$$

für alle $\frac{p}{q} \in \mathbf{Q}$ mit $p \in \mathbf{Z}$ und $q \in \mathbf{N}$.

Wir werden im nächsten Paragraphen sehen, dass alle quadratischen Irrationalzahlen schlecht approximierbar sind.

Satz 7.3 *Eine Zahl $\alpha \in \mathbf{R} \setminus \mathbf{Q}$ ist genau dann schlecht approximierbar, wenn eine Konstante $K = K(\alpha)$ mit $a_n \leq K$ existiert, d.h. wenn die Teilnenner der Kettenbruchentwicklung $\alpha = [a_0, a_1, \ldots]$ beschränkt sind.*

Beweis. Nach Satz 7.1 genügt es, $c < \frac{1}{2}$ und die Näherungsbrüche $\frac{p_n}{q_n}$ der Kettenbruchentwicklung von α zu betrachten. Die vollständigen Quotienten sind hier $\alpha_i = [a_i, a_{i+1}, \ldots] > 1$ und außerdem sind alle $\beta_i := \frac{q_{i-2}}{q_{i-1}} < 1$. Wie im Beweis von Satz 7.2 ist

$$|\alpha q_n - p_n| \;=\; \frac{1}{\alpha_{n+1} q_n + q_{n-1}}$$

und somit

$$\left|\alpha - \frac{p_n}{q_n}\right| = \frac{1}{q_n^2(\alpha_{n+1} + \beta_{n+1})} \;=\; \frac{1}{q_n^2([a_{n+1}, a_{n+2}, \ldots] + \beta_{n+1})} \quad .$$

Wegen $[a_{n+1}, a_{n+2}, \ldots] < a_{n+1} + 1$ und $\beta_{n+1} < 1$ erhalten wir

$$\left|\alpha - \frac{p_n}{q_n}\right| \;>\; \frac{1}{q_n^2(a_{n+1} + 2)} \quad .$$

Andererseits ist $\alpha_{n+1} > a_{n+1}$ und $\beta_{n+1} > 0$, d.h.

$$\left|\alpha - \frac{p_n}{q_n}\right| \;<\; \frac{1}{q_n^2 a_{n+1}} \quad .$$

Wenn $K(\alpha)$ existiert, wähle $c := \frac{1}{K+2}$, so dass $\left|\alpha - \frac{p_n}{q_n}\right| \;>\; \frac{c}{q_n^2}$.

Umgekehrt, wenn eine Schranke $K(\alpha)$ nicht existiert, gibt es eine Teilfolge von $(a_n)_{n \in \mathbf{N}}$, die gegen unendlich geht. Dann gibt es auch kein geeignetes c wegen $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2 a_{n+1}}$. □

Für quadratische Irrationalzahlen existiert eine solche Konstante $K = K(\alpha)$. Das Problem für algebraisches $\alpha \in \mathbf{R}$ mit $[\mathbf{Q}(\alpha) : \mathbf{Q}] > 2$ ist offen, ob alle a_k eine gemeinsame obere Schranke K besitzen (dann ist die Kettenbruchentwicklung nicht periodisch, vergleiche den nächsten §). Man vermutet, dass die Teilnenner dieser Kettenbruchentwicklungen unbeschränkt sind.

8 Periodische Kettenbrüche und reell–quadratische Zahlkörper

Wir erinnern zunächst an die *reell–quadratischen Zahlkörper* $\mathbf{Q}(\sqrt{d})$. Sei $d \geq 2$, $d \in \mathbf{N}$ quadratfrei. Dann ist

$$\mathbf{Q}(\sqrt{d}) \;:=\; \{r + s\sqrt{d} \mid r, s \in \mathbf{Q}\}$$

ein Körper, insbesondere ist das Inverse durch

$$\frac{1}{r+s\sqrt{d}} \;=\; \frac{r-s\sqrt{d}}{r^2-s^2d}$$

gegeben (beachte, dass für $r,s \neq 0$ gilt $r^2-s^2d \neq 0$, da d nach Voraussetzung quadratfrei ist). Der *Grad* der Körpererweiterung $[\mathbf{Q}(\sqrt{d}) : \mathbf{Q}]$ (= Dimension des $\mathbf{Q}$–Vektorraums $\mathbf{Q}(\sqrt{d})$) ist 2.

$\mathbf{Q}(\sqrt{d})$ besitzt einen *Ring ganzer Zahlen*

$$\mathcal{O}_d \;:=\; \mathbf{Z}[\alpha] \;=\; \{n+m\alpha \mid n,m \in \mathbf{N}\}\,,$$

wobei

$$\alpha \;:=\; \begin{cases} \sqrt{d} & \text{wenn} \quad d \equiv 2,3 \bmod 4 \\ \frac{1}{2}\,(1+\sqrt{d}) & \text{wenn} \quad d \equiv 1 \bmod 4 \end{cases}\quad . \tag{3}$$

$\mathcal{O}_d$ ist der größte Unterring von $\mathbf{Q}(\sqrt{d})$ mit $\mathcal{O}_d \cap \mathbf{Q} = \mathbf{Z}$ und besteht aus allen Lösungen normierter quadratischer Gleichungen $x^2+bx+c=0$ mit $b,c \in \mathbf{Z}$ und $x \in \mathbf{Q}(\sqrt{d})$. Die Teilbarkeit in $\mathcal{O}_d$ ist definiert wie in allen Ringen:

$$\beta \mid \gamma \quad :\Longleftrightarrow \quad \exists\, \delta \in \mathcal{O}_d \quad \text{mit} \quad \gamma = \delta\beta$$

Man beachte aber, dass nicht immer die eindeutige Primfaktorzerlegung gegeben ist. Ein Automorphismus von $\mathcal{O}_d$ (bzw. $\mathbf{Q}(\sqrt{d})$) ist die Abbildung auf das *algebraisch konjugierte* Element:

$$\beta = r+s\sqrt{d} \quad \mapsto \quad \beta' = r-s\sqrt{d}$$

Die Teilbarkeitseigenschaft bleibt durch durch Übergang auf die algebraisch konjugierten Elemente erhalten, d.h. genau dann ist $\beta \mid \gamma$, wenn $\beta' \mid \gamma'$. Als *Norm* in $\mathbf{Q}(\sqrt{d})$ verwenden wir:

$$N(\beta) \;:=\; \beta\beta'$$

Für $\mathcal{O}_d$ bedeutet dies (vergleiche (3)):

$$N(n+m\alpha) \;=\; \begin{cases} n^2-dm^2 & \text{für} \quad d \equiv 2,3 \pmod 4 \\ n^2+mn-m^2\cdot\frac{d-1}{4} & \text{für} \quad d \equiv 1 \pmod 4 \end{cases}\quad .$$

Insbesondere gilt $N(n+m\alpha) \in \mathbf{Z}$. Die *Einheitengruppe* von $\mathcal{O}_d$ ist

$$\mathcal{O}_d^* \;:=\; \{\beta \in \mathcal{O}_d \mid \quad \beta \mid 1\} \quad .$$

Aus $\beta \mid 1$ folgt $N(\beta) = \beta\beta' \mid N(1) = 1$. Für $\beta \in \mathcal{O}_d^*$ ist dies äquivalent zu $\beta\beta' = \pm 1$ bzw. gleichbedeutend zu $\beta' = \pm\beta^{-1}$. Zum Beispiel ist $\beta = \frac{1}{2}\left(1+\sqrt{5}\,\right) \in \mathcal{O}_5^*$ mit $N(\beta) = -1$.

Lemma 8.1 $\beta \in \mathcal{O}_d^* \iff N(\beta) = \pm 1$

Es besteht eine Bijektion zwischen $\mathcal{O}_d^*$ und den Lösungen der *Pell'schen Gleichung* (der Name geht zurück auf Eulers irrtümliche Annahme, J. Pell habe sie zuerst behandelt).

Lemma 8.2 *Es besteht eine Bijektion zwischen $\mathcal{O}_d^*$ und den Lösungen $(\pm x, \pm y) \in \mathbf{Z}^2$ der sogenannten Pell'schen Gleichung*

$$\begin{aligned} x^2 - dy^2 &= \pm 1 \qquad \text{falls } d \equiv 2, 3 \pmod 4 \\ x^2 - dy^2 &= \pm 4 \qquad \text{falls } d \equiv 1 \pmod 4. \end{aligned}$$

Beweis. Für $d \neq 1 \pmod 4$ folgt die Aussage unmittelbar aus Lemma 8.1, da $N(n+m\alpha) = n^2 - dm^2$. Für $d \equiv 1 \pmod 4$ beachte:

$$\tfrac{1}{2}\left(n + m\sqrt{d}\right) \in \mathcal{O}_d \iff n, m \in \mathbf{Z} \text{ und } n \equiv m \pmod 2$$

Einerseits ist

$$N\left(\tfrac{1}{2}\left(n + m\sqrt{d}\right)\right) = \tfrac{1}{4}\left(n^2 - m^2 d\right) = \pm 1 \iff n^2 - m^2 d = \pm 4 \quad .$$

Andererseits erfüllen alle Lösungen von $n^2 - m^2 d = \pm 4$ auch $n \equiv m \pmod 2$. □

Definition 8.1 (periodischer, einfacher Kettenbruch) *Ein einfacher Kettenbruch $[a_0, a_1, \ldots, a_n, \ldots]$ heißt „periodisch", wenn es $m \in \mathbf{N}$ und $M \in \mathbf{N}_0$ gibt, so dass für alle $n \geq M$*

$$a_{n+km} = a_n \text{ für } k = 1, 2, \ldots .$$

Wir verwenden dafür die folgende Notation:

$$[a_0, a_1, \ldots, a_{M-1}, \dot{a}_M, \dot{a}_{M+1}, \ldots, \dot{a}_{M+m-1}]$$

Der einfacher Kettenbruch $[a_0, \ldots, a_n, \ldots]$ heißt reinperiodisch, wenn er als $[\dot{a}_0, \dot{a}_1, \ldots, \dot{a}_{m-1}]$ geschrieben werden kann.

Wann hat eine Zahl α eine einfache, periodische Kettenbruchentwicklung? L. Euler hat eine hinreichende Bedingung gegeben, nämlich $\alpha \in \mathbf{Q}(\sqrt{d}) \setminus \mathbf{Q}$ für $d \geq 2$ quadratfrei, und J.L. Lagrange hat später bewiesen, dass dieses Kriterium auch notwendig ist:

Lemma 8.3 (Euler 1737) *Reinperiodische Kettenbrüche $r = [\dot{a}_0, \dot{a}_1, \ldots, \dot{a}_{m-1}]$ sind reell–quadratische Irrationalzahlen.*

Beweis. Nach Lemma 5.2 ist mit $p_m, q_m \in \mathbf{Z}$

$$r \;=\; [\dot{a}_0, \dot{a}_1, \ldots, \dot{a}_{m-1}] \;=\; \frac{rp_{m-1} + p_{m-2}}{rq_{m-1} + q_{m-2}} \quad .$$

Man erhält quadratische Gleichungen für r. Es ist $r \notin \mathbf{Q}$, weil r sonst eine endlichen Kettenbruchentwicklung hätte. Somit gilt $r \in \mathbf{Q}(\sqrt{d})$ mit (Diskriminante der quadratischen Gleichung)

$$d \;=\; (q_{m-2} - p_{m-1})^2 + 4q_{m-1}p_{m-2} \quad ,$$

gegebenenfalls nach Division durch quadratische Teiler. □

Satz 8.1 (Lagrange 1748) *Genau dann hat $\alpha \in \mathbf{R}$ eine periodische Kettenbruchentwicklung*

$$[a_0, a_1, \ldots, a_{M-1}, \dot{a}_M, \dot{a}_{M+1}, \ldots, \dot{a}_{M+m-1}] \quad ,$$

wenn α eine reell–quadratische Irrationalzahl ist.

Beweis. 1. Sei $\alpha = [a_0, a_1, \ldots, a_{M-1}, r]$ mit reinperiodischer Kettenbruchentwicklung für r. Nach Lemma 8.3 ist $r \in \mathbf{Q}(\sqrt{d}) \setminus \mathbf{Q}$, dabei $d \geq 2$ quadratfrei,

$$\alpha \;=\; \frac{rp_{M-1} + p_{M-2}}{rq_{M-1} + q_{M-2}} \in \mathbf{Q}(\sqrt{d})$$

und $\alpha \notin \mathbf{Q}$ (da α eine unendliche Kettenbruchentwicklung hat).
2. Sei $\alpha \in \mathbf{Q}(\sqrt{d}) \setminus \mathbf{Q}$ mit irreduzibler Gleichung

$$A_0\alpha^2 + B_0\alpha + C_0 \;=\; 0 \quad ,$$

wobei $A_0, B_0, C_0 \in \mathbf{Z}$ teilerfremd sind. Die Zahl α hat die Kettenbruchentwicklung $[a_0, a_1, \ldots, a_{n-1}, \alpha_n]$ mit

$$\alpha \;=\; \frac{p_{n-1}\alpha_n + p_{n-2}}{q_{n-1}\alpha_n + q_{n-2}} \quad .$$

Dies liefert eine quadratische Gleichung für α_n, nämlich

$$A_0 \left(p_{n-1}\alpha_n + p_{n-2}\right)^2 + B_0 \left(p_{n-1}\alpha_n + p_{n-2}\right)\left(q_{n-1}\alpha_n + q_{n-2}\right) + C_0 \left(q_{n-1}\alpha_n + q_{n-2}\right)^2 \;=\; 0 \quad ,$$

gleichbedeutend mit $A_n\alpha_n^2 + B_n\alpha_n + C_n = 0$, wobei

$$\begin{aligned} A_n &:= A_0p_{n-1}^2 + B_0p_{n-1}q_{n-1} + C_0q_{n-1}^2 \\ B_n &:= 2A_0p_{n-1}p_{n-2} + B_0\left(p_{n-1}q_{n-2} + p_{n-2}q_{n-1}\right) + 2C_0q_{n-1}q_{n-2} \\ C_n &:= A_0p_{n-2}^2 + B_0p_{n-2}q_{n-2} + C_0q_{n-2}^2 \quad . \end{aligned}$$

Nach §6, Beweis von Satz 6.1, wissen wir

$$\alpha q_{n-1} - p_{n-1} \;=\; \frac{(-1)^{n-1}}{\alpha_n q_{n-1} + q_{n-2}} \quad ,$$

und daraus folgt

$$p_{n-1} \;=\; \alpha q_{n-1} + \frac{\delta}{q_{n-1}}$$

mit $|\delta| < 1$, da $\alpha_n \geq 1$ und $q_{n-2} > 0$, also

$$\begin{aligned} A_n = \; & A_0 \left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right)^2 + B_0 \left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right) q_{n-1} + C_0 q_{n-1}^2 \\ = \; & q_{n-1}^2 \underbrace{\left(A_0\alpha^2 + B_0\alpha + C_0\right)}_{=\,0 \text{ nach Voraussetzung}} + 2A_0\alpha\delta + B_0\delta + A_0\frac{\delta^2}{q_{n-1}^2} \end{aligned}$$

Auch B_n und C_n sind unabhängig von n beschränkt. Somit ist die Menge $\{\alpha_n \mid n \in \mathbf{N}\}$ endlich und die Kettenbruchentwicklung von α ist periodisch. □

9 Diophantische Approximationen und diophantische Gleichungen

Eine Erinnerung an den Satz 4.3:

Satz 9.1 (Thue–Siegel–Roth) *Sei $\alpha \in \mathbf{R} \cap \overline{\mathbf{Q}}$, $d := [\mathbf{Q}(\alpha) : \mathbf{Q}] \geq 2$ und $\mu > 2$. Dann existiert ein $b := b(\alpha, \mu) > 0$, so dass die Ungleichung*

$$\left|\alpha - \frac{p}{q}\right| < \frac{b}{q^\mu}$$

nur endlich viele Lösungen $\frac{p}{q} \in \mathbf{Q}$ mit teilerfremden $p, q \in \mathbf{Z}$ besitzt. Anders gesagt: Für irrationale algebraische α existiert eine Konstante $c := c(\alpha, \mu) > 0$ mit

$$\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^\mu}$$

für alle $\frac{p}{q} \in \mathbf{Q}$.

J. Liouville(1844) hat die Aussage für $\mu \geq d$ gezeigt. A. Thue (1909) hat diese Schranke zu $\mu \geq \frac{1}{2}d + 1$, C.L. Siegel (1921) zu $\mu \geq 2\sqrt{d}$ und schließlich K.F. Roth (1955) zu $\mu > 2$ verbessert.

Satz 9.2 (Thue 1908) *Sei* $F(X,Y) = \sum_{i=0}^{d} a_i X^{d-i} Y^i$ *binäre Form mit rationalen Koeffizienten, d.h. homogenes Polynom*

$$F(\lambda X, \lambda Y) \;=\; \lambda^d \cdot F(X,Y),$$

mit mindestens drei paarweise linear unabhängigen Linearfaktoren aus $\overline{\mathbf{Q}}[X,Y]$ *(insbesondere ist* $d \geq 3$*). Wenn* $m \in \mathbf{Z} \setminus \{0\}$*, dann hat die Thue–Gleichung* $F(x,y) = m$ *(wenn überhaupt) nur endlich viele Lösungen* $(x,y) \in \mathbf{Z}^2$*.*

Bevor wir den Satz beweisen, zwei Vorüberlegungen:

- Warum $d \geq 3$? Sei $F(X,Y) = X^2 - DY^2$. Die Gleichung $F(X,Y) = \pm m$ (*Pell'sche Gleichung*) hat unendlich viele Lösungen (vergleiche Mini–Projekt in den Übungen).
- Warum Zerfällung in Linearformen? Für $d \geq 3$ ist mit $a_0 \neq 0$

  $$\frac{1}{Y^d} F(X,Y) = \sum_{i=0}^{d} a_i \underbrace{\left(\frac{X}{Y}\right)}_{=:z}^{d-i} = a_0 z^d + a_1 z^{d-1} + \cdots + a_d = a_0 \prod_{i=1}^{d} (z - \alpha_i)$$

  mit $\alpha_i \in \overline{\mathbf{Q}}$. Multiplikation mit Y^d liefert

  $$F(X,Y) \;=\; a_0 \prod_{i=1}^{d} (X - \alpha_i Y) \quad .$$

  Für $a_0 = 0$ gilt $F(X,Y) = Y \cdot F_1(X,Y)$, dabei $F_1(X,Y)$ binäre Form vom Grad $d-1$. Wiederhole diese Zerlegung, bis der führende Koeffizient $\neq 0$ ist.

  Stets gilt: $F(X,Y)$ zerfällt in lineare homogene Polynome aus $\overline{\mathbf{Q}}[X,Y]$ (gewisse dieser Polynome können X oder Y sein).
- Warum mindestens drei verschiedene Linearformen? Hätte F nur zwei verschiedene Linearformen, könnte man z.B. unendlich viele Lösungen von $(X^2 - DY^2)^2 = 1$ finden.

Beweis des Thue'schen Satzes. Wie oben findet man stets eine Zerlegung

$$F(X,Y) \;=\; a\,(\gamma_1 X + \delta_1 Y)^{e_1}\,(\gamma_2 X + \delta_2 Y)^{e_2} \cdot \ldots \cdot (\gamma_s X + \delta_s Y)^{e_s}$$

mit $s \geq 3$. Dabei sind $(\gamma_i X + \delta_i Y)$ paarweise linear unabhängig über $\overline{\mathbf{Q}}$ für $i = 1, \ldots, s$ und so normalisiert, dass $\gamma_i = 1$ oder $\gamma_i = 0$ und $\delta_i = 1$, d.h. δ_i sind Nullstellen von $F(X,1)$ und $\delta_i \in \overline{\mathbf{Q}}$. Notwendig ist $a \in \mathbf{Q}$ (sonst wäre $F(X,Y) \notin \mathbf{Q}[X,Y]$).

Sei $\mathbf{z} := (x,y) \in \mathbf{Z}^2$ eine Lösung der Thue–Gleichung. Dann kann man die Faktoren von F o.B.d.A. so anordnen, dass

$$|\gamma_s x + \delta_s y| \geq |\gamma_{s-1} x + \delta_{s-1} y| \geq \ldots \geq |\gamma_1 x + \delta_1 y| > 0$$

gilt. Mittelbildung in der vorletzten Ungleichung liefert

$$|\gamma_s x + \delta_s y| \geq \ldots \geq |\gamma_2 x + \delta_2 y| \geq \frac{|\gamma_1 x + \delta_1 y| + |\gamma_2 x + \delta_2 y|}{2} > 0 \quad .$$

Letzteres ist sogar $\geq c_1 \max\{|x|, |y|\} := c_1 ||\mathbf{z}||$ (Maximumsnorm) für eine Konstante c_1, die nicht von $\mathbf{z}$ und m, sondern nur von $(\gamma_1, \delta_1), (\gamma_2, \delta_2)$ abhängt. Diese beiden Zeilen sind nämlich linear unabhängig nach Voraussetzung über die Linearfaktoren von F, es gibt also eine Matrix

$$A := \begin{bmatrix} \gamma_1 & \delta_1 \\ \gamma_2 & \delta_2 \end{bmatrix}^{-1} \quad ,$$

und für diese ist

$$\begin{bmatrix} x \\ y \end{bmatrix} = A \begin{bmatrix} \gamma_1 & \delta_1 \\ \gamma_2 & \delta_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = A \begin{bmatrix} \gamma_1 x + \delta_1 y \\ \gamma_2 x + \delta_2 y \end{bmatrix}$$

Mit Hilfe der Maximumxnorm für die Matrix ergibt sich daraus

$$|x|, |y| \leq ||A|| \cdot (|\gamma_1 x + \delta_1 y| + |\gamma_2 x + \delta_2 y|)$$

wie behauptet. Auf die anderen Linearfaktoren von F lässt sich die gleiche Schlussweise anwenden. Produktbildung über alle diese Ungleichungen zeigt daher

$$|F(x, y)| \geq c_2 \cdot |\gamma_1 x + \delta_1 y|^{e_1} \cdot ||\mathbf{z}||^{d - e_1} \quad .$$

Dabei hängt c_2 von den γ_j, δ_j und a, also von F ab, aber nicht von (x, y). Fallunterscheidung:

1. Sei $\gamma_1, \delta_1 \in \mathbf{Q}$. Wegen $x, y \in \mathbf{Z}$ ist dann

$$|\gamma_1 x + \delta_1 y| \geq c_3 > 0$$

(zum Beispiel mit $\frac{1}{c_3}$ als Hauptnenner von γ_1, δ_1). Weiter ist $d > e_1$, denn
$d = e_1 + e_2 + \cdots + e_s$ mit $s \geq 3$. Folglich gilt

$$|F(x, y)| \geq c_2 \cdot c_3^{e_1} \cdot ||(x, y)||^{d - e_1} \quad .$$

Wegen $c_3 \neq 0$ geht $|F(x, y)|$ gegen unendlich für $||(x, y)|| \to \infty$, d.h. es kann nur endlich viele Lösungen der Thue–Gleichung geben.

2. Es gilt $\gamma_1 = 1$ und $\delta_1 \in \overline{\mathbf{Q}}$ mit Grad $l \geq 2$. Für alle $\delta > 0$ gilt nach dem Satz von Roth (den wir noch beweisen müssen)

$$\left| \frac{x}{y} + \delta_1 \right| \geq c_4(\delta_1, \delta) \cdot |y|^{-2-\delta} \quad .$$

Damit folgt:

$$|\gamma_1 x + \delta_1 y| \geq c_4 \cdot |y|^{-1-\delta}$$

Dabei setzen wir $y \neq 0$ voraus. (Für $y = 0$ haben wir nur endlich viele Lösungen.) Erst recht ist

$$|\gamma_1 x + \delta_1 y| \ \geq \ c_5 \cdot ||(x,y)||^{-1-\delta} \quad .$$

Damit ist

$$|F(x,y)| \ \geq \ c_6 \cdot ||(x,y)||^{d-e_1-e_1(1+\delta)} \ \geq \ c_6 \cdot ||(x,y)||^{d-e_1(2+\delta)}$$

Um zu zeigen, dass $F(x,y) = m$ nur endlich viele Lösungen hat, genügt also der Nachweis, dass dieser Exponent $d - e_1(2+\delta) > 0$ ist.

Mit $-\delta_1$ (Nullstelle von $F(X,1) \in \mathbf{Q}[X]$) sind auch alle $\ell \geq 2$ Konjugierten von $-\delta_1$ Nullstellen von $F(X,1)$. Alle Konjugierten treten mit gleicher Multiplizität e_1 auf (man zerlege $F(X,1)$ in irreduzible Polynome, dann muss $F(X,1) = f(X)^{e_1} \cdots$ sein, wobei $-\delta_1$ Nullstelle von $f(X)$ ist ebenso wie die algebraisch konjugierten zu $-\delta_1$). Folglich gilt

$$d \geq le_1 \geq 2e_1.$$

Für $l = 2$ ist $d > 2e_1$ wegen $s \geq 3$ (sonst gäbe es nur zwei Linearfaktoren). In jedem Fall gilt $d > 2e_1$. Mit hinreichend kleinem $\delta > 0$ folgt also

$$|F(X,Y)| \ \geq \ c_6 \cdot ||(x,y)||^{\varepsilon}$$

für $\varepsilon > 0$. $\square$

Man hätte diesen Beweis bereits mit der Thue–Ungleichung führen können. Aus dem Satz von Roth ergibt sich allerdings zum Beispiel der Beweis folgender stärkerer Aussage:

Satz 9.3 *Sei $F(X,Y)$ eine binäre Form vom Grad $d \geq 3$ mit rationalen Koeffizienten ohne mehrfache Linearfaktoren (d.h. $e_1 = e_2 = \cdots = e_d = 1$) und $\nu < d-2$. Dann gibt es nur endlich viele $\mathbf{z} = (x,y) \in \mathbf{Z}^2$ mit*

$$0 \ < \ |F(\mathbf{z})| \ < \ ||\mathbf{z}||^{\nu}$$

Insbesondere: Sei $G(X,Y) \in \mathbf{C}[X,Y]$ mit Gesamtgrad kleiner $d-2$, so hat die Diophantische Gleichung $F(\mathbf{z}) = G(\mathbf{z})$ nur endlich viele Lösungen $\mathbf{z} \in \mathbf{Z}^2$ mit $F(\mathbf{z}) \neq 0$.

10 Höhen. Index. Das Siegelsche Lemma

Wir definieren die *Höhe* eines Polynoms aus $\mathbf{Z}[X_1, \ldots, X_m]$ als die Maximumsnorm des Koeffizientenvektors:

Definition 10.1 *Die ,,Höhe" von* $P(X_1,\ldots,X_m)=\sum C(j_1,\ldots,j_m)\,X_1^{j_1}X_2^{j_2}\cdots X_m^{j_m}\in\mathbf{Z}[X_1,\ldots,X_m]$ *ist*

$$\overline{|P|}\ :=\ \max|C(j_1,\ldots,j_m)|$$

und für alle $i_1,\ldots,i_m\in\mathbf{N}_0$ *sei:*

$$P_{i_1,\ldots,i_m}\ :=\ \frac{1}{i_1!\,i_2!\ldots i_m!}\cdot\frac{\partial^{i_1+i_2+\cdots+i_m}}{\partial X_1^{i_1}\partial X_2^{i_2}\ldots\partial X_m^{i_m}}P(X_1,\ldots,X_m)\quad .$$

Für $i=(i_1,\ldots,i_m)$ schreiben im folgenden kurz P_i statt $P_{i_1,\ldots,i_m}$.

Lemma 10.1 *Sei* $P\in\mathbf{Z}[X_1,\ldots,X_m]$. *Dann gilt* $P_i\in\mathbf{Z}[X_1,\ldots,X_m]$ *für alle* $i=(i_1,\ldots,i_m)\in\mathbf{N}_0^m$. *Wenn* $\deg_{X_h}P\le r_h$ *für* $h=1,\ldots,m$ *ist, gilt*

$$\overline{|P_i|}\ \le\ 2^{r_1+r_2+\cdots+r_m}\cdot\overline{|P|}$$

Beweis. Sei $P(X_1,\ldots,X_m)=\sum C(j_1,\ldots,j_m)\,X_1^{j_1}X_2^{j_2}\cdots X_m^{j_m}$. Dann gilt für alle $i=(i_1,\ldots,i_m)\in\mathbf{N}_0^m$

$$P_i=\sum_{j_1=0}^{r_1}\sum_{j_2=0}^{r_2}\cdots\sum_{j_m=0}^{r_m}\underbrace{\binom{j_1}{i_1}\binom{j_2}{i_2}\ldots\binom{j_m}{i_m}C(j_1,\ldots,j_m)}_{\in\mathbf{Z}}\cdot X_1^{j_1-i_1}X_2^{j_2-i_2}\cdots X_m^{j_m-i_m}$$

Vereinbarungsgemäß sei $\binom{j}{i}=0$ für $i>j$. Es folgt $P_i\in\mathbf{Z}[X_1,\ldots,X_m]$. Die Höhenabschätzung folgt aus $\binom{j_h}{i_h}\le 2^{j_h}\le 2^{r_h}$ für $h=1,\ldots,m$.

Definition 10.2 *Sei* $P(X_1,\ldots,X_m)\in\mathbf{Z}[X_1,\ldots,X_m]$, $r_1,\ldots,r_m\in\mathbf{N}$ *und* $(\alpha_1,\ldots,\alpha_m)\in\mathbf{R}^m$. *Für* $P\not\equiv 0$ *heißt*

$$\operatorname{Ind}P\ :=\ \min\{\frac{i_1}{r_1}+\frac{i_2}{r_2}+\ldots+\frac{i_m}{r_m}\mid P_{i_1,\ldots,i_m}(\alpha_1,\ldots,\alpha_m)\neq 0\}$$

der ,,Index" von P *bezüglich* $(\alpha_1,\ldots,\alpha_n,r_1,\ldots,r_m)$. *Für* $P\equiv 0$ *sei* $\operatorname{Ind}P:=\infty$.

Insbesondere ist genau dann $\operatorname{Ind}P=0$, wenn $P(\alpha_1,\ldots,\alpha_m)\neq 0$.

Lemma 10.2 *Sei* $r_1,\ldots,r_m\in\mathbf{N}$ *und* $(\alpha_1,\ldots,\alpha_m)\in\mathbf{R}^m$. *Dann gilt:*

1. $\operatorname{Ind}P_i\ge\operatorname{Ind}P-\sum_{h=1}^{m}\frac{i_h}{r_h}$
2. $\operatorname{Ind}\left(P^{(1)}+P^{(2)}\right)\ge\min\left\{\operatorname{Ind}P^{(1)},\operatorname{Ind}P^{(2)}\right\}$

3. $\operatorname{Ind}\left(P^{(1)} \cdot P^{(2)}\right) = \operatorname{Ind} P^{(1)} + \operatorname{Ind} P^{(2)}$

Beweis. 1.) Sei $T := P_i$ und $j = (j_1, \ldots, j_m)$ mit $T_j(\alpha_1, \ldots, \alpha_m) \neq 0$. Dann ist $P_{i+j}(\alpha_1, \ldots, \alpha_m) \neq 0$, da beide Polynome sich nur um einen konstanten Faktor unterscheiden. Nach Definition des Index gilt

$$\frac{i_1 + j_1}{r_1} + \frac{i_2 + j_2}{r_2} + \ldots + \frac{i_m + j_m}{r_m} \;\geq\; \operatorname{Ind} P \quad .$$

Man erhält

$$\frac{j_1}{r_1} + \frac{j_2}{r_2} + \ldots + \frac{j_m}{r_m} \;\geq\; \operatorname{Ind} P - \frac{i_1}{r_1} + \frac{i_2}{r_2} + \cdots + \frac{i_m}{r_m} \quad .$$

Wähle j so, dass $\frac{j_1}{r_1} + \frac{j_2}{r_2} + \ldots + \frac{j_m}{r_m} = \operatorname{Ind} T = \operatorname{Ind} P_i$.

2.) Sei o.B.d.A. $P^{(1)} + P^{(2)} \neq 0$ und j so gewählt, dass

$$\operatorname{Ind}\left(P^{(1)} + P^{(2)}\right) \;=\; \sum_{h=1}^{m} \frac{j_h}{r_h}$$

und $\left(P^{(1)} + P^{(2)}\right)_j (\alpha_1, \ldots, \alpha_m) \neq 0$. Dann gilt

$$P_j^{(1)}(\alpha_1, \ldots, \alpha_m) \neq 0 \qquad \text{oder} \qquad P_j^{(2)}(\alpha_1, \ldots, \alpha_m) \neq 0$$

und somit:

$$\operatorname{Ind} P^{(1)} \leq \sum_{h=1}^{m} \frac{j_h}{r_h} \qquad \text{oder} \qquad \operatorname{Ind} P^{(2)} \leq \sum_{h=1}^{m} \frac{j_h}{r_h}$$

Es folgt die Behauptung.

3.) Es ist für alle $j \in \mathbf{N}_0^m$

$$\left(P^{(1)} \cdot P^{(2)}\right)_j \;=\; \sum_{j=i+i'} P_i^{(1)} \cdot P_{i'}^{(2)} \cdot C(i, i') \quad . \tag{4}$$

mit Konstanten $C(i, i') \neq 0$. Wähle j so, dass

$$\sum_{h=1}^{m} \frac{j_h}{r_h} \;=\; \operatorname{Ind}\left(P^{(1)} \cdot P^{(2)}\right)$$

und $\left(P^{(1)} \cdot P^{(2)}\right)_j (\alpha_1, \ldots, \alpha_m) \neq 0$. Folglich existieren $i, i' \in \mathbf{N}_0^m$ mit $i + i' = j$ und

$$\operatorname{Ind} P_i^{(1)}(\alpha_1, \ldots, \alpha_m) \;\neq\; 0 \qquad \text{und} \qquad \operatorname{Ind} P_{i'}^{(2)}(\alpha_1, \ldots, \alpha_m) \;\neq\; 0 \quad .$$

Damit gilt

$$\operatorname{Ind} P^{(1)} \le \sum_{h=1}^{m} \frac{i_h}{r_h} \qquad \text{und} \qquad \operatorname{Ind} P^{(2)} \le \sum_{h=1}^{m} \frac{i'_h}{r_h}$$

und man erhält

$$\operatorname{Ind} P^{(1)} + \operatorname{Ind} P^{(2)} \le \sum_{h=1}^{m} \left(\frac{i_h}{r_h} + \frac{i'_h}{r_h} \right) \;=\; \sum_{h=1}^{m} \frac{j_h}{r_h} \;=\; \operatorname{Ind} \left(P^{(1)} \cdot P^{(2)} \right) \quad .$$

Umgekehrt seien $i, i' \in \mathbf{N}_0^m$ gegeben mit

$$\operatorname{Ind} P^{(1)} = \sum_{h=1}^{m} \frac{i_h}{r_h} \qquad \text{und} \qquad \operatorname{Ind} P^{(2)} = \sum_{h=1}^{m} \frac{i'_h}{r_h} \quad ,$$

dabei $P_i^{(1)}(\alpha_1, \ldots, \alpha_m) \neq 0$ und $P_{i'}^{(2)}(\alpha_1, \ldots, \alpha_m) \neq 0$. Unter den i, i' mit dieser Eigenschaft wähle jeweils das lexikographisch erste aus. Für diese i, i' setze $j := i + i'$. Damit gilt

$$\begin{aligned} &\left(P^{(1)} \cdot P^{(2)} \right)_j (\alpha_1, \ldots, \alpha_m) \\ &= C(i, i') \cdot P_i^{(1)}(\alpha_1, \ldots, \alpha_m) \cdot P_{i'}^{(2)}(\alpha_1, \ldots, \alpha_m) \quad . \end{aligned} \tag{5}$$

Sei nämlich $C(\bar{i}, \bar{i}') \cdot P_{\bar{i}}^{(1)} \cdot P_{\bar{i}'}^{(2)}$ ein anderer Summand in (4), etwa mit $\bar{i} > i$, so muss $\bar{i}' <_{\text{Lex}} i'$ sein, da die Summen j sind. Daraus folgt

$$P_{\bar{i}'}^{(2)}(\alpha_1, \ldots, \alpha_m) \;=\; 0 \quad ,$$

so dass der Summand 0 ist. Aus (5) folgt

$$\sum_{h=1}^{m} \frac{j_h}{r_h} \;\ge\; \operatorname{Ind} \left(P^{(1)} \cdot P^{(2)} \right) \quad . \qquad \square$$

Satz 10.1 (Das Siegelsche Lemma) *Seien $L_j(\mathbf{z}) = \sum_{k=1}^{N} a_{jk} z_k$ für $j = 1, \ldots, M$, $\mathbf{z} = (z_1, \ldots, z_N)$, Linearformen mit Koeffizienten aus $\mathbf{Z}$. Sei $N > M$ und $|a_{jk}| \le A$ für alle $1 \le j \le M$ und $1 \le k \le N$. Dann existiert eine Lösung $\mathbf{z} \in \mathbf{Z}^N \setminus \{\mathbf{0}\}$ des Gleichungssystems $L_1(\mathbf{z}) = L_2(\mathbf{z}) = \cdots = L_M(\mathbf{z}) = 0$ mit*

$$||\mathbf{z}|| \;\le\; \left[(NA)^{\frac{M}{N-M}} \right] \;:=\; Z \in \mathbf{Z} \quad .$$

Beweis. Wegen $N > M$ ist der Kern nichttrivial, d.h. es existieren Lösungen $\mathbf{z} \in \mathbf{Q}^N \setminus \{\mathbf{0}\}$; wenn wir mit dem Hauptnenner multiplizieren, sogar in $\mathbf{Z}^N \setminus \{\mathbf{0}\}$. Aus

$$Z + 1 \;>\; (N \cdot A)^{\frac{M}{N-M}}$$

folgt $N \cdot A < (Z+1)^{\frac{N-M}{M}}$, so dass

$$N \cdot A \cdot Z + 1 \ < \ N \cdot A(Z+1) \ < \ (Z+1)^{\frac{N}{M}} \quad .$$

Für die $(Z+1)^N$ Punkte $\mathbf{z} \in \mathbf{Z}^N$ mit $0 \leq z_\nu \leq Z$ ist

$$-B_j \cdot Z \ \leq \ L_j(\mathbf{z}) \leq C_j \cdot Z \qquad \text{für} \quad j = 1, \ldots, M \quad ,$$

wobei C_j die Summe der positiven $a_{j\nu}$ und $-B_j$ die Summe der negativen $a_{j\nu}$ bezeichne. Damit gilt

$$B_j + C_j \ \leq \ N \cdot A \quad ,$$

da jeder Summand in L_j absolut durch A beschränkt ist. Alle $L_j(\mathbf{z})$ liegen in einem Intervall der Länge maximal $N \cdot A \cdot Z$. Folglich haben alle Werte in $\mathbf{Z}^N \cap [0, Z]^N$ höchstens

$$(N \cdot A \cdot Z + 1)^M \ < \ (Z+1)^N$$

Bildpunkte, also kann die Abbildung nach dem Schubfachprinzip nicht injektiv sein, d.h. es existieren $\mathbf{z}^{(1)}, \mathbf{z}^{(2)} \in \mathbf{Z}^N \cap [0, Z]^N$ mit $\mathbf{z}^{(1)} \neq \mathbf{z}^{(2)}$ und

$$L_j\left(\mathbf{z}^{(1)}\right) \ = \ L_j\left(\mathbf{z}^{(2)}\right) \qquad \text{für} \quad j = 1, \ldots, M \quad ,$$

Somit gilt $L_j\left(\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\right) = 0$ für $j = 1, \ldots, M$. Damit erfüllt $\mathbf{z} := \mathbf{z}^{(1)} - \mathbf{z}^{(2)}$ die Ungleichung $||\mathbf{z}|| \leq Z$. $\square$

11 Der Indexsatz

Eine Zahl $\alpha \in \overline{\mathbf{Q}}$ heißt *ganz–algebraisch,* wenn das zugehörige irreduzible, normierte Polynom $f(x) \in \mathbf{Q}[x]$ sogar aus $\mathbf{Z}[x]$ ist. Äquivalent: α erfüllt eine algebraische Gleichung

$$\alpha^d + a_1\alpha^{d-1} + \cdots + a_{d-1}\alpha + a_d \ = \ 0$$

mit $a_1, \ldots, a_d \in \mathbf{Z}$. Für alle $\beta \in \overline{\mathbf{Q}}$ mit irreduzibler, *primitiver* Gleichung

$$b_0\beta^d + b_1\beta^{d-1} + \cdots + b_{d-1}\beta + b_d \ = \ 0 \quad ,$$

d.h. mit $b_0 \in \mathbf{N}$ und teilerfremden $b_0, \ldots, b_d \in \mathbf{Z}$, ist $\alpha = b_0\beta$ ganz–algebraisch (multipliziere mit b_0^{d-1}, so dass man ein irreduzibles, normiertes Polynom für α erhält). Wenn für ein $\delta > 0$ und $\frac{p}{q} \in \mathbf{Q}$ mit teilerfremden $p, q \in \mathbf{Z}$ gilt

$$\left|\beta - \frac{p}{q}\right| \ \leq \ c \cdot q^{-2-\delta},$$

so ergibt es sich durch Multiplikation mit $b_0 \in \mathbf{N}$

$$\left|b_0\beta - \frac{b_0 p}{q}\right| \ \leq \ cb_0 \cdot q^{-2-\delta}$$

und umgekehrt. Somit genügt es, den Satz von Roth für ganz-algebraische α zu zeigen. Die Konstante c hängt ohnehin nur von α und δ ab.

Lemma 11.1 *Sei α ganz–algebraisch mit irreduziblem, normiertem Polynom*

$$Q(X) \;=\; X^d + a_1 X^{d-1} + \cdots + a_{d-1}X + a_d \in \mathbf{Z}[X] \quad .$$

Für alle $\ell \in \mathbf{N}$ existieren $a_1^{(\ell)}, a_2^{(\ell)}, \ldots, a_d^{(\ell)} \in \mathbf{Z}$ mit

$$\alpha^\ell \;=\; a_1^{(\ell)}\alpha^{d-1} + \cdots + a_{d-1}^{(\ell)}\alpha + a_d^{(\ell)}$$

und $|a_i^{(\ell)}| \le \left(\overline{|Q|}+1\right)^\ell$ für $i = 1, \ldots, d$.

Beweis durch Induktion über ℓ:

- Induktionsanfang: Für $\ell = 0, \ldots, d-1$ ist die Aussage trivial.
- Induktionsannahme: Die Aussage sei für $\ell - 1$ anstelle von ℓ bereits bewiesen.
- Induktionsschluss: für $\ell \ge d$ gilt

  $$\alpha^\ell \;=\; \alpha \cdot \alpha^{\ell-1} \;=\; \alpha\left(a_1^{(\ell-1)}\alpha^{d-1} + \cdots + a_{d-1}^{(\ell-1)}\alpha + a_d^{(\ell-1)}\right) .$$

  Da α ganz–algebraisch ist, folgt

  $$\alpha^d \;=\; -a_1\alpha^{d-1} - a_2\alpha^{d-2} - \cdots - a_d$$

  und wir erhalten

  $$\begin{aligned}\alpha^\ell &= a_1^{(\ell-1)}\left(-a_1\alpha^{d-1} - \cdots - a_d\right) + a_2^{(\ell-1)}\alpha^{d-1} + \cdots + a_2^{(\ell-1)}\alpha^2 + a_d^{(\ell-1)}\alpha \\ &= \left(a_2^{(\ell-1)} - a_1 a_1^{(\ell-1)}\right)\alpha^{d-1} + \cdots + \left(a_d^{(\ell-1)} - a_{d-1}a_1^{(\ell-1)}\right)\alpha - a_d a_1^{(\ell-1)}\end{aligned}$$

  Alle Koeffizienten sind ganzzahlig und wir können mit der Induktionsannahme den Absolutbetrag der Koeffizienten nach oben beschränken durch

  $$\underbrace{\left(\overline{|Q|}+1\right)^{\ell-1}}_{a_i^{(\ell-1)}} + \underbrace{\overline{|Q|}}_{a_i} \cdot \underbrace{\left(\overline{|Q|}+1\right)^{\ell-1}}_{a_1^{(\ell-1)}} = \left(\overline{|Q|}+1\right)^\ell \quad . \qquad \square$$

Satz 11.1 *Sei α ganz-algebraisch vom Grad $d := [\mathbf{Q}(\alpha) : \mathbf{Q}] \ge 2$, $\epsilon > 0$ und $n \in \mathbf{N}$ sowie $m > 16\epsilon^{-2}\log(4d)$ und $r_1, \ldots, r_m \in \mathbf{N}$. Dann existiert ein Polynom $P(X_1, \ldots, X_m) \in \mathbf{Z}[X_1, \ldots, X_m]$ mit $P \not\equiv 0$ und folgenden Eigenschaften:*

1. *$\deg_{X_h} P \le r_h$, P hat also höchstens $N := (r_1+1)\ldots(r_m+1)$ Koeffizienten.*
2. *Bezüglich $(\alpha, \ldots, \alpha, r_1, \ldots, r_m)$ hat P den Index $\operatorname{Ind} P \ge \frac{1}{2}m(1-\epsilon)$.*

3. Es existiert eine Konstante $B := B(\alpha)$ *mit* $\overline{|P|} \leq B^{r_1+r_2+\cdots+r_m}$.

Beweis. Ansatz:

$$P(X_1,\ldots,X_m) \;=\; \sum_{j_1=0}^{r_1}\sum_{j_2=0}^{r_2}\cdots\sum_{j_m=0}^{r_m} C(j_1,\ldots,j_m)\cdot X_1^{j_1}X_2^{j_2}\ldots X_m^{j_m}$$

mit $N := (r_1+1)(r_2+1)\cdots(r_m+1)$ ganzrationalen Koeffizienen $C(j_1,\ldots,j_m)$.
P erfülle folgende Gleichungen

$$P_i(\alpha,\alpha,\ldots,\alpha) \;=\; 0 \qquad \text{für alle } \; i=(i_1,\ldots,i_m)\in\mathbf{N}_0^m \tag{6}$$

mit $\left(\sum_{h=1}^m \frac{i_h}{r_h}\right) - \frac{m}{2} < -\frac{\epsilon m}{2}$. Mit einigem Aufwand an Kombinatorik oder Wahrscheinlichkeitsrechnung kann man nun zeigen, dass wegen der Voraussetzung $m > 16\epsilon^{-2}\log(4d)$ die Anzahl der i–Vektoren mit dieser Eigenschaft höchstens

$$2(r_1+1)(r_2+1)\cdots(r_m+1)\cdot e^{-\left(\frac{\epsilon}{2}\right)^2\cdot\frac{m}{4}} \;<\; 2N\cdot e^{-\log(4d)} \;=\; \frac{N}{2d}$$

werden kann. Die Idee dabei ist, dass man die Summanden i_h/r_h als unabhängige Zufallsvariable mit Erwartungswert $\frac{1}{2}$ auffasst; ihre Summe hat dann den Erwartungswert $\frac{m}{2}$, und die Behauptung über die Häufigkeit des Abstands der Summe zum Erwartungswert folgt dann aus einer Art Tschebyscheff–Ungleichung. Das Gleichungssystem (6) definiert lineare Gleichungen für die $C(j_1,\ldots,j_m)$ mit Koeffizienten in $\mathbf{Z}\cdot\alpha^\ell$ mit $\ell\in\mathbf{N}_0$. Zerlege alle α^ℓ gemäß Lemma 11.1 in $\mathbf{Z}$–Linearkombinationen von $1,\alpha,\alpha^2,\ldots,\alpha^{d-1}$. Dann zerfällt jede der einzelnen Gleichungen (6) in d lineare Gleichungen für die $C(j_1,\ldots,j_m)$ mit Koeffizienten in $\mathbf{Z}$, also insgesamt $M < d\cdot\frac{N}{2d} = \frac{1}{2}N$ lineare Gleichungen für N Unbekannte $C(j_1,\ldots,j_m)$.

Um eine nichttriviale und nicht zu große Lösung des Gleichungssystems zu finden, wenden wir das Siegelsche Lemma (Satz 10.1) an. Dazu benötigen wir eine Schranke A für die Absolutbeträge der Koeffizienten. Nach Lemma 10.1 ist können wir die Höhe der P_i durch die Höhe von P abschätzen, die Koeffizienten des Gleichungssystems (6) sind also beschränkt durch $2^{r_1+\cdots+r_m}$; schreiben wir das Gleichungssystem in ein d–faches Gleichungssystem mit Koeffizienten in $\mathbf{Z}$ um, multiplizieren sich diese Koeffizienten höchstens mit dem Faktor $\left(\overline{|Q|}+1\right)^\ell$ (Lemma 11.1), wobei ℓ die Exponenten von α in den $P_i(\alpha,\ldots,\alpha)$ durchläuft. Letztlich erhält man also eine Abschätzung

$$A \;\leq\; \left[2\cdot\left(\overline{|Q|}+1\right)\right]^{r_1+r_2+\cdots+r_m} \qquad ,$$

das Siegelsche Lemma liefert also ganzzahlige Koeffizienten $C(j_1,\ldots,j_m)$ von P der Größe

$$|C(j_1,\ldots,j_m)| \leq (NA)^{\frac{M}{N-M}} = NA \leq (r_1+1)(r_2+1)\ldots(r_m+1)\cdot\left[2\cdot\left(\overline{|Q|}+1\right)\right]^{r_1+r_2+\cdots+r_m} .$$

Mit $B := B(\alpha) := \left[4\cdot\left(\overline{|Q|}+1\right)\right]$ folgt daraus die Behauptung des Indexsatzes. □

Satz 11.2 *Seien die Voraussetzungen des Indexsatzes gegeben, P ebenso konstruiert, $0 < \delta < 1$ mit $0 < \epsilon < \frac{1}{36}\delta$, $D = D(\alpha) \in \mathbf{R}$, $\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}$ rationale Approximationen an α mit*

$$q_h^{\delta} > D \qquad \text{und} \qquad \left|\alpha - \frac{p_h}{q_h}\right| < q_h^{-2-\delta}$$

für $h = 1, \ldots, m$. Ferner seien die r_h so gewählt, dass

$$r_1 \cdot \log q_1 \leq r_h \cdot \log q_h \leq (1+\epsilon) \cdot r_1 \cdot \log q_1$$

für $h = 1, \ldots, m$. Dann ist $\operatorname{Ind} P \geq \epsilon m$ bezüglich $\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m\right)$; insbesondere gilt $P\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right) = 0$.

Beweis. Seien $j_1, \ldots, j_m \in \mathbf{N}_0$ mit $\sum_h \frac{j_h}{r_h} < \epsilon m$, $j = (j_1, \ldots, j_h)$ und

$$T(X_1, \ldots, X_m) := P_j(X_1, \ldots, X_m) \quad .$$

Zu zeigen ist $T\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right) = 0$. Nach Lemma 10.1 ist

$$\overline{|T|} \leq (2B)^{r_1+\cdots+r_m}$$

und ebenso

$$\overline{|T_j|} \leq (4B)^{r_1+\cdots+r_m} \qquad \text{für alle } j\,.$$

Jedes Monom in $T_j(\alpha, \ldots, \alpha)$ hat also einen Absolutbetrag von höchstens

$$(4B)^{r_1+\cdots+r_m} \cdot \max\{1, |\alpha|\}^{r_1+\cdots+r_m} \quad .$$

Weiterhin ist die Anzahl der Monome höchstens

$$(r_1+1)\cdots(r_m+1) \leq 2^{r_1+\cdots+r_m} \quad .$$

Folglich gilt für eine nur von α abhängige Kontante c:

$$|T_j(\alpha, \ldots, \alpha)| \leq (8B \cdot \max\{1, |\alpha|\})^{r_1+\cdots+r_m} =: c^{r_1+\cdots+r_m}$$

Gemäß Indexsatz ist $\operatorname{Ind} P \geq \frac{m}{2}(1-\epsilon)$ bezüglich $(\alpha, \ldots, \alpha, r_1, \ldots, r_m)$. Somit gilt nach Lemma 10.1

$$\operatorname{Ind} T \geq \frac{1}{2}m \cdot (1-\epsilon) - \epsilon m = \frac{1}{2}m \cdot (1-3\epsilon) \qquad .$$

Die Taylorentwicklung von T um $(\alpha, \ldots, \alpha)$ ergibt

$$T\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right) = \sum_{i_1=0}^{r_1} \cdots \sum_{i_m=0}^{r_m} T_{i_1,\ldots,i_m}(\alpha, \ldots, \alpha) \cdot \left(\frac{p_1}{q_1} - \alpha\right)^{i_1} \cdots \left(\frac{p_m}{q_m} - \alpha\right)^{i_m} \quad ,$$

wobei $T_i(\alpha,\dots,\alpha)=0$ für alle i mit $\dfrac{i_1}{r_1}+\dots+\dfrac{i_m}{r_m}\le\dfrac{m}{2}(1-3\epsilon)$. Damit haben wir

$$\left|T\left(\frac{p_1}{q_1},\dots,\frac{p_m}{q_m}\right)\right| \;\le\; \sum_i c^{r_1+\dots+r_m}\left(q_1^{i_1}\cdots q_m^{i_m}\right)^{-2-\delta} \quad .$$

Es wird hier nur noch über alle i mit $\dfrac{i_1}{r_1}+\dots+\dfrac{i_m}{r_m}>\dfrac{m}{2}(1-3\epsilon)$ summiert. In diesen verbleibenden Summanden ist

$$\begin{aligned} q_1^{i_1}\cdots q_m^{i_m} &= q_1^{r_1\frac{i_1}{r_1}}\cdots q_m^{r_m\frac{i_m}{r_m}} \\ &\ge q_1^{r_1\left(\frac{i_1}{r_1}+\dots+\frac{i_m}{r_m}\right)} && (\text{da } r_1\log q_1\le r_h\log q_h) \\ &> q_1^{r_1 m\left(\frac12-2\epsilon\right)} && (\text{da } \tfrac12(1-3\epsilon)>\tfrac12-2\epsilon) \quad . \end{aligned}$$

Wegen $r_1\log q_1\ge(1+\epsilon)r_h\log q_h$ gilt ferner

$$q_1^{i_1}\cdots q_m^{i_m} \;>\; \left(\prod_{h=1}^m q_1^{r_1}\right)^{\frac12-2\epsilon} \;\ge\; \left(q_1^{r_1}\cdots q_m^{r_m}\right)^{(\frac12-2\epsilon)(1+\epsilon)} \;\ge\; \left(q_1^{r_1}\cdots q_m^{r_m}\right)^{\frac12(1-6\epsilon)} \quad .$$

Die Anzahl der Summanden beträgt höchstens $2^{r_1+\dots+r_m}$, so dass

$$\left|T\left(\frac{p_1}{q_1},\dots,\frac{p_m}{q_m}\right)\right| \;\le\; \prod_{h=1}^m\left(2cq_h^{-\frac12(1-6\epsilon)(2+\delta)}\right)^{r_h} \quad .$$

Wegen $\delta<1$ und $\epsilon<\frac{1}{36}\delta$ ist

$$\tfrac12(1-6\epsilon)(2+\delta) \;>\; 1+\tfrac12\delta-9\epsilon \;>\; 1+\tfrac14\delta \quad .$$

Für $q_h^\delta>(2c)^4=:D$ gilt folglich

$$2cq_h^{-\frac12(1-6\epsilon)(2+\delta)} < 2cq_h^{-1-\frac{\delta}{4}} \;<\; q_h^{-1} \quad .$$

Daraus folgt

$$\left|T\left(\frac{p_1}{q_1},\dots,\frac{p_m}{q_m}\right)\right| \;<\; q_1^{-r_1}\cdots q_m^{-r_m} \quad .$$

Andererseits ist $T\in\mathbf{Z}[X_1,\dots,X_m]$ und $\deg_{X_h}T\le r_h$, deshalb

$$T\left(\frac{p_1}{q_1},\dots,\frac{p_m}{q_m}\right) \;=\; \frac{N}{q_1^{r_1}\cdots q_m^{r_m}}$$

mit einem Zähler $N\in\mathbf{Z}$. Aus dieser Gleichung und der letzten Ungleichung folgt $N=0$, d.h. $T\left(\dfrac{p_1}{q_1},\dots,\dfrac{p_m}{q_m}\right)=0$. $\square$

12 Das ,,Lemma von Roth" und der Beweis des Satzes

Definition 12.1 *Seien* $\varphi, \varphi_1, \ldots, \varphi_k \in \mathbf{R}(X_1, \ldots, X_m)$ *rationale Funktionen in* m *Variablen und für* $\mathbf{i} = (i_1, \ldots, i_m)$ *sei ein Differentialoperator*

$$D_{\mathbf{i}} \ : \ \varphi \mapsto \varphi_{\mathbf{i}}$$

der ,,Ordnung" $i := i_1 + \cdots + i_m$ *, definiert wie in Definition 10.1.* Δ_i *bezeichne einen solchen Differentialoperator der Ordnung* $\leq i-1$ *. Eine ,,verallgemeinerte Wronski-Determinante" von* $\varphi_1, \ldots, \varphi_k$ *ist*

$$W \ := \ \det(\Delta_i \varphi_j), \qquad i, j = 1, \ldots, k \quad .$$

Für $m = 1$ ist notwendigerweise $\Delta_1 = D_0 = \mathrm{id}$, $\Delta_2 = D_1 = \frac{\partial}{\partial X}$, $\ldots$, $\Delta_k = D_{k-1} = \frac{1}{(k-1)!}\frac{\partial^{k-1}}{\partial X^{k-1}}$ das einzige Beispiel, für das $W \not\equiv 0$ (sonst tritt ein Operator mehrfach auf, so dass $W \equiv 0$).

Lemma 12.1 *Seien* $\varphi_1, \ldots, \varphi_k \in \mathbf{R}(X_1, \ldots, X_m)$ *linear unabhängig über* $\mathbf{R}$ *. Dann existiert eine verallgemeinerte Wronski–Determinante* $W = W(\varphi_1, \ldots, \varphi_k) \not\equiv 0$ *.*

Beweis durch Induktion über k. Der Induktionsanfang $k = 1$ ist trivial, weil nur der Differentialoperator $D_{\mathbf{0}} = \mathrm{id}$ in Frage kommt und $W(\varphi_1) = \varphi_1 \not\equiv 0$ für nichttriviales φ_1 erfüllt.

Für den Induktionsschritt beachte man, dass aus der linearen Unabhängigkeit der Funktionen $\varphi_1, , \ldots, \varphi_k$ die lineare Unabhängigkeit der Funktionen

$$1\,, \ \frac{\varphi_2}{\varphi_1}\,, \ \ldots\,, \ \frac{\varphi_k}{\varphi_1}$$

folgt; mehr noch: es gibt eine Variable X_j, so dass die partiellen Ableitungen

$$\frac{\partial}{\partial X_j}\frac{\varphi_2}{\varphi_1}\,, \ \ldots\,, \ \frac{\partial}{\partial X_j}\frac{\varphi_k}{\varphi_1}$$

ebenfalls linear unabhängig sind. Nach Induktionsannahme gibt es Differentialoperatoren $\Delta_1, \ldots, \Delta_{k-1}$, die – angewandt auf die Funktionen φ_j/φ_1 – eine Wronskideterminante $\not\equiv 0$ ergeben. Im Induktionsschritt wende man auf die Funktionen $1, \varphi_2/\varphi_1, \ldots$ die Differentialoperatoren $D_{\mathbf{0}}$ und $\frac{\partial}{\partial X_j}\Delta_1, \ldots, \frac{\partial}{\partial X_j}\Delta_{k-1}$ an und überzeuge sich, dass eine nichtsinguläre Matrix entsteht. □

Satz 12.1 (,,Roths Lemma") *Seien* $0 < \epsilon < \frac{1}{12}$ *und* $m \in \mathbf{N}$ *fest gewählt sowie*

$$\omega \ = \ \omega(\epsilon, m) \ := \ 24 \cdot 2^{-m} \left(\frac{\epsilon}{12}\right)^{2^{m-1}} \quad .$$

Seien $r_1, \dots, r_m \in \mathbf{N}$ *mit* $\omega r_h \geq r_{h+1}$ *für alle* $h = 1, \dots, m-1$ *und* $(p_1, q_1), \dots, (p_m, q_m)$ *teilerfremde Paare in* $\mathbf{Z}^2$ *mit* $q_h \in \mathbf{N}$ *und* $q_h^{r_h} \geq q_1^{r_1}$, $q_h^{\omega} \geq 2^{3m}$. *Sei* $P(X_1, \dots, X_m) \in \mathbf{Z}[X_1, \dots, X_m]$, $P \neq 0$, $\deg_{X_h} P \leq r_h$ *mit* $\overline{|P|} \leq q_1^{\omega r_1}$. *Dann ist* $\operatorname{Ind} P \leq \epsilon$ *bezüglich* $\left(\frac{p_1}{q_1}, \dots, \frac{p_m}{q_m}, r_1, \dots, r_m\right)$.

Der *Beweis des Satzes von Thue–Siegel–Roth* aus dem Indexsatz und Roths Lemma wird als Widerspruchsbeweis geführt. Wir nehmen also an, es gäbe eine algebraische Irrationalzahl α vom Grad $d \geq 2$, so dass für jedes $\delta > 0$ die Ungleichung

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{2+\delta}} \tag{7}$$

unendlich viele Lösungen $\frac{p}{q} \in \mathbf{Q}$ mit teilerfremden $p, q \in \mathbf{Z}$ besitzt. O.B.d.A. dürfen wir annehmen, dass alle $q > 0$ sind, dass α ganzalgebraisch ist (vgl. den Anfang von Sec. 11) und zwischen 0 und 1 liegt (durch Übergang zu $\langle\alpha\rangle = \alpha - [\alpha]$), und dass $\delta < 1$ gilt.
Wähle $\epsilon \in \mathbf{R}$ mit $0 < \epsilon < \frac{\delta}{36}$ und $m \in \mathbf{N}$ mit $m > 16\epsilon^{-2} \log 4d$. Der Indexsatz besagt:

> Für alle $r_1, \dots, r_m \in \mathbf{N}$ gibt es ein Polynom $P(X_1, \dots, X_m) \in \mathbf{Z}[X_1, \dots, X_m]$, $P \not\equiv 0$, so dass
>
> - $\deg_{X_h} P \leq r_h$
> - Bezüglich $(\alpha, \dots, \alpha, r_1, \dots, r_m)$ ist $\operatorname{Ind} P \geq \frac{m}{2}(1-\epsilon)$
> - Es existiert eine Konstante $B = B(\alpha)$ mit $\overline{|P|} \leq B^{r_1+\dots+r_m}$

Nach Satz 11.2 wissen wir:

> Es existiert ein $D = D(\alpha) \in \mathbf{R}$, so dass für die rationalen Approximationen $\frac{p_1}{q_1}, \dots, \frac{p_m}{q_m}$ an α mit (7) und $q_h^{\delta} > D$ sowie $r_1 \log q_1 \leq r_h \log q_h \leq (1+\epsilon) r_1 \log q_1$ der Index $\operatorname{Ind} P \geq \epsilon m$ ist bezüglich $\left(\frac{p_1}{q_1}, \dots, \frac{p_m}{q_m}, r_1, \dots, r_m\right)$.

Es ist $0 < \epsilon < \frac{1}{12}$, $\omega := \omega(\epsilon, m) := 24 \cdot 2^{-m} \left(\frac{\epsilon}{12}\right)^{2^{m-1}}$. Nach Roths Lemma gilt:

> Wenn $q_h^{r_h} \geq q_1^{r_1}$ und $q_h^{\omega} \geq 2^{3m}$ für alle h sowie $\overline{|P|} \leq q_1^{\omega r_1}$, dann ist bezüglich $\left(\frac{p_1}{q_1}, \dots, \frac{p_m}{q_m}, r_1, \dots, r_m\right)$ der Index $\operatorname{Ind} P \leq \epsilon$.

Wähle $\frac{p_1}{q_1} \in \mathbf{Q}$, so dass (7) erfüllt ist und $q_1^{\omega} > B^m$ (ein solches q_1 existiert, da es unendlich viele Lösungen von (7) gibt). Weiter sei $q_1^{\delta} > D := D(\alpha) := (16 \cdot \max\{1, \alpha\})^4 = 2^{16}$ (vgl.

Beweis von Satz 11.2). Wähle $B \geq 2^3$, so dass $q_1^\omega \geq 2^{3m}$. Wähle weiterhin sukzessive Lösungen $\frac{p_2}{q_2}, \ldots, \frac{p_m}{q_m} \in \mathbf{Q}$ von (7), für welche

$$\omega \log q_{h+1} \geq 2 \log q_h \quad .$$

Insbesondere gilt dann $q_1 < q_2 < \ldots < q_m$. Nach deren Wahl gilt:

$$q_h^\omega \; > \; B^m \; \geq \; 2^{3m} \, , \qquad q_h^\delta \; > \; D$$

für alle h. Wähle die r_h wie folgt: $r_1 \in \mathbf{N}$ sei so groß, dass

$$\epsilon r_1 \log q_1 \; \geq \; \log q_m \qquad .$$

Für $h = 2, \ldots, m$ sei

$$r_h \; := \; \left[\frac{r_1 \log q_1}{\log q_h} \right] + 1 \qquad ,$$

so dass $r_h \log q_h > r_1 \log q_1$. Da $\log q_h \leq \log q_m \leq \epsilon r_1 \log q_1$, gilt andererseits

$$r_h \log q_h \; \leq \; r_1 \log q_1 + \log q_h \; \leq \; (1+\epsilon) r_1 \log q_1 \qquad .$$

Wähle P nach dem Indexsatz passend zu α mit diesen $r_1, \ldots, r_m$. Damit sind die Voraussetzungen des Satzes 11.2 erfüllt:

$$\overline{|P|} \; \leq \; B^{r_1 + \cdots + r_m} \; \leq \; B^{r_1 m} \qquad ,$$

denn nach Konstruktion ist $\frac{\log q_1}{\log q_h} < 1$ und somit $r_h \leq r_1$. Folglich ist

$$\overline{|P|} \; \leq \; B^{r_1 m} \; \leq \; q_1^{\omega r_1} \qquad .$$

Nach dem Roth'schen Lemma ist damit

$$\operatorname{Ind} P \; \leq \; \epsilon \; < \; \epsilon m$$

im Widerspruch zur Annahme $\operatorname{Ind} P \geq \epsilon m$. $\square$

Der Satz von Roth ist nicht effektiv, d.h. es ist nicht möglich, eine obere Schranke für q aus (7) abzuleiten, die nur von α und δ abhängt. Grund: Schon die Wahl von q_1 lässt sich nicht nach oben einschränken.

Die Anzahl der sehr guten Approximationen $\frac{p}{q}$ mit (7) von α lässt sich (schlecht, aber) effektiv nach oben abschätzen, $|\{q \in \mathbf{N} \mid q < B^{\frac{m}{\omega}}\}|$ ist allerdings sehr groß. Durch

$$q_{h+1}^\omega \; \geq \; q_h^2 \quad \Longleftrightarrow \quad q_{h+1} \; \geq \; q_h^{\frac{2}{\omega}}$$

werden $[q_h^{2/\omega}]$ Nenner ausgeschlossen. Aufgabe: Diese Anzahl möglicher Nenner durch eine Größe zu beschränken, welche nur von α, δ, h abhängt (lösbar!). Dies folgt aus Varianten

des Beweises nach Dyson, Esnault, Viehweg.

Beweis von Roths Lemma durch Induktion über m. Sei $m = 1$, also $\omega = \epsilon$. Dann ist

$$P(X) \;=\; \left(X - \frac{p_1}{q_1}\right)^{\ell} \cdot M(X) \qquad \text{mit } \; M(X) \in \mathbf{Q}[X]\,.$$

Wähle ℓ so groß, dass $M\left(\frac{p_1}{q_1}\right) \neq 0$, d.h. ℓ maximal. Damit genügt es, zu zeigen, dass

$$\operatorname{Ind} P \;=\; \frac{\ell}{r_1} \;\overset{!}{\leq}\; \epsilon \quad .$$

Es ist ebenfalls folgende Zerlegung möglich:

$$P(X) \;=\; (q_1 X - p_1)^{\ell} \cdot R(X) \qquad \text{mit } \; R = \frac{1}{q_1^{\ell}} \cdot M\,.$$

Das Gauß'sche Lemma besagt, dass $R(X) \in \mathbf{Z}[X]$. Der führende Koeffizient von $(q_1 X - p_1)^{\ell}$ wird von q_1^{ℓ} geteilt, somit auch $P(X)$ und damit

$$q_1^{\ell} \;\leq\; \overline{|P|} \;\leq\; q_1^{\omega r_1} \quad .$$

Aus der Voraussetzung $q_1^{\omega} \geq 2^{3m} = 8$ und $\omega = \epsilon$ folgt $q_1^{\epsilon} \geq 8$, d.h. $q_1 > 1$. Wir erhalten

$$\ell \;\leq\; \omega r_1 = \epsilon r_1 \qquad .$$

Induktionsschritt: Sei das Roth'sche Lemma für $1, \ldots, m-1$ und alle zulässigen ϵ bewiesen. Dann setzen wir

$$P(X_1, \ldots, X_m) \;=\; \sum_{j=1}^{k} \varphi_j(X_1, \ldots, X_{m-1}) \cdot \psi_j(X_m) \in \mathbf{Z}[X_1, \ldots, X_m] \qquad ,$$

wobei $\varphi_1, \ldots, \varphi_k \in \mathbf{Q}[X_1, \ldots, X_{m-1}]$, z.B. mit $k = r_m + 1$, und $\psi_j = X_m^{j-1}$ für $j = 1, \ldots, k$. Wähle die Zerlegung so, dass k minimal wird. Insbesondere ist $k \leq r_m + 1$.

Behauptung: $\varphi_1, \ldots, \varphi_k$ sind linear unabhängig über $\mathbf{R}$.

Begründung: Andernfalls sei o.B.d.A. $\varphi_k = \sum_{j=1}^{k-1} c_j \varphi_j$. Ersetze φ_k durch diese Linearkombination, dann steht

$$P(X_1, \ldots, X_m) \;=\; \sum_{j=1}^{k-1} \varphi_j(\psi_j + c_j \psi_k)$$

im Widerspruch zur Minimalität von k.

Analog zeigt man, dass $\psi_1, \ldots, \psi_k$ linear unabhängig über $\mathbf{R}$ sind. Sei

$$U(X_m) \;:=\; \det\left(\frac{1}{(i-1)!} \cdot \frac{\partial^{i-1}}{\partial X_m^{i-1}} \psi_j(X_m)\right)_{i,j=1,\ldots,k} \qquad ,$$

dann ist $U \not\equiv 0$ die (einzige) nichttriviale verallgemeinerte Wronski-Determinante. Nach Lemma 12.1 gibt es Operatoren

$$\Delta_i' \;:=\; \frac{1}{i_1! \cdots i_{m-1}!} \cdot \frac{\partial^{i_1+\cdots+i_{m-1}}}{\partial X_1^{i_1} \cdots \partial X_{m-1}^{i_{m-1}}} \quad , \qquad i = 1, \ldots, k$$

mit Ordnungen $i_1 + \cdots + i_m \leq i - 1 \leq k - 1 \leq r_m$, so dass

$$V(X_1, \ldots, X_{m-1}) \;=\; \det\left(\Delta_i' \varphi_j\right)_{i,j} \;\not\equiv 0$$

ist. Setze

$$\begin{aligned} W(X_1, \ldots, X_m) &:= \det\left(\sum_{r=1}^{k} (\Delta_i' \varphi_r) \cdot \left(\frac{1}{(j-1)!} \cdot \frac{\partial^{j-1}}{\partial X_m^{j-1}} \psi_r\right)\right)_{i,j} \\ &= \det\left(\underbrace{\frac{1}{(j-1)!} \cdot \frac{\partial^{j-1}}{\partial X_m^{j-1}} \Delta_i' P}_{\in \mathbf{Z}[X_1, \ldots, X_m]}\right)_{i,j} \quad ; \end{aligned}$$

mit anderen Worten erfolgt die Zeilenindizierung nach Ableitung der ersten $m-1$ Variablen und die Spaltenindizierung nach Ableitung der m-ten Variablen. Damit ergibt sich

$$0 \;\not\equiv\; W(X_1, \ldots, X_m) \;=\; V(X_1, \ldots, X_{m-1}) \cdot U(X_m) \;=\; V^*(X_1, \ldots, X_{m-1}) \cdot U^*(X_m)$$

mit $U^*, V^* \in \mathbf{Z}[X_1, \ldots, X_m]$. (Man multipliziere V mit Nenner v und dividiere U durch v. Nach dem Gauß'schen Lemma folgt $U^*, V^* \in \mathbf{Z}[X_1, \ldots, X_m]$.)

Behauptung: Bezüglich $\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m\right)$ gilt für den Index von W:

$$\Theta \;:=\; \operatorname{Ind} W \;\leq\; \frac{1}{6} k \epsilon^2$$

Begründung: Sei $\vartheta := \operatorname{Ind} P$ bezüglich $\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m\right)$. Nach Konstruktion ist

$$W \;=\; \det P_{i_1, \ldots, i_{m-1}, j_m} \quad , \qquad j_m = j - 1 \quad ,$$

für den Zeilenindex $i_1, \ldots, i_{m-1}$ und Spaltenindex j_m. Nach Lemma 10.2.1 folgt aus $\omega r_h \geq r_{h+1}$

$$\begin{aligned} \operatorname{Ind} P_{i_1, \ldots, i_{m-1}, j_m} &\geq\; \vartheta - \frac{i_1}{r_1} - \cdots - \frac{i_{m-1}}{r_{m-1}} - \frac{j-1}{r_m} \\ &\geq\; \vartheta - \frac{i_1 + \cdots + i_{m-1}}{r_{m-1}} - \frac{j-1}{r_m} \quad . \end{aligned}$$

Aus $\sum_{\nu=1}^{m-1} i_\nu \leq i-1 \leq k-1 \leq r_m$ erhalten wir

$$\begin{aligned} \operatorname{Ind} P_{i_1,\ldots,i_{m-1},j_m} &\geq \vartheta - \frac{r_m}{r_{m-1}} - \frac{j-1}{r_m} \\ &\geq \vartheta - \omega - \frac{j-1}{r_m} \end{aligned}$$

Wegen $m \geq 2$ gilt

$$\operatorname{Ind} P_{i_1,\ldots,i_{m-1},j_m} \geq \vartheta - \frac{\epsilon^2}{24} - \frac{j-1}{r_m}$$

Da W die Form

$$W = \sum_{k!\text{ Glieder}} \prod_{\substack{k \text{ Glieder} \\ j=1,\ldots,k}} P_{i_1,\ldots,i_{m-1},j-1}$$

hat und nach Lemma 10.2 gilt

$$\begin{aligned} \operatorname{Ind}(P^{(1)} + P^{(2)}) &\geq \min\{\operatorname{Ind} P^{(1)}, \operatorname{Ind} P^{(2)}\} \\ \operatorname{Ind} P^{(1)} \cdot P^{(2)}) &= \operatorname{Ind} P^{(1)} + \operatorname{Ind} P^{(2)} \end{aligned} ,$$

erhalten wir

$$\begin{aligned} \Theta = \operatorname{Ind} W &\geq \min_{k!\text{ Glieder}} \left\{ \sum_{j=1}^{k} \operatorname{Ind} P_{i_1,\ldots,i_{m-1},j-1} \right\} \\ &\geq \sum_{j=1}^{k} \max\left\{ (\vartheta - \tfrac{\epsilon^2}{24} - \tfrac{j-1}{i_m}), 0 \right\} \\ &\geq -\tfrac{4}{24}\epsilon^2 + \sum_{j=1}^{k} \max\{\vartheta - \tfrac{i}{r_m}, 0\} \end{aligned}$$

In der zweiten Zeile haben wir ausgenutzt, dass jedes $P_{i_1,\ldots,i_{m-1},j-1}$ der Abschätzung genügt, insbesondere das Minimum. Ferner ist der Index eines Polynoms nie negativ, so dass wir das Maximum mit 0 bilden können. Es folgt:

$$\sum_{i=0}^{k-1} \max\{\vartheta - \tfrac{i}{r_m}, 0\} \leq \Theta + \tfrac{1}{24}k\epsilon^2 < \tfrac{1}{4}k\epsilon^2 \tag{8}$$

Wir führen eine Fallunterscheidung durch:

1. Sei $\vartheta > \frac{k-1}{r_m}$. Aus Ungleichung (8) folgt $\frac{1}{2}k\left(2\vartheta - \frac{k-1}{r_m}\right) < k \cdot \frac{\epsilon^2}{4}$ oder, äquivalent dazu:

$$\vartheta + \Big(\underbrace{\vartheta - \frac{k-1}{r_m}}_{>0}\Big) < \frac{\epsilon^2}{2}$$

   Wir erhalten $\vartheta < \frac{1}{2}\epsilon^2 < \epsilon$.

2. Sei $\vartheta \leq \frac{k-1}{r_m}$. Dann hat Ungleichung (8) die Form:

$$\sum_{i=0}^{[\vartheta r_m]} \left(\vartheta - \frac{i}{r_m} \right) \; < \; \tfrac{1}{4} k \epsilon^2$$

Daraus erhält man

$$\tfrac{1}{2} \vartheta ([\vartheta r_m] + 1) \; < \; \tfrac{1}{4} k \epsilon^2$$

Es folgt $\frac{1}{2} \vartheta^2 r_m < \frac{1}{4} k \epsilon^2$, wegen $k \leq r_m + 1$, d.h. $k \leq 2 r_m$, ist $\vartheta^2 < \epsilon^2$.

Aus beiden Fällen folgt die Behauptung. □