Kapitel 5. Verschiedene Entwicklungen reeller Zahlen

Während bisher in diesem Buch die Untersuchung ganzer Zahlen weitgehend im Vordergrund stand, verlagert sich nun der Schwerpunkt der Thematik hin zu den reellen Zahlen. Insbesondere wird dabei die g–adische Entwicklung reeller Zahlen als Verallgemeinerung der geläufigen Dezimalbruchentwicklung behandelt ebenso wie die regelmäßige Kettenbruchentwicklung. Beide Darstellungen haben sich historisch bei dem Bemühen herausgebildet, reelle Irrationalzahlen möglichst gut durch rationale Zahlen anzunähern. Zusätzlich erfüllen dabei die Kettenbrüche die Forderung guter Approximation selbst bei Verwendung relativ kleiner, geeignet gewählter Nenner; dagegen sind die g–adischen Brüche vor allem für das praktische Rechnen von Vorteil, während bei ihnen das Verhältnis von erzielter Approximationsgüte zur Größe der verwendeten Nenner viel ungünstiger ausfällt.

In § 1 wird insbesondere die Rationalität einer reellen Zahl durch die Periodizität ihrer g–adischen Entwicklung charakterisiert. Auch die feineren Periodizitätseigenschaften der g–adischen Entwicklung rationaler Zahlen werden dort vollständig aufgedeckt.

§ 2 bringt die auf Cantor zurückgehende Verallgemeinerung der g–adischen Entwicklung reeller Zahlen und fügt auf diesem Wege den in 1.1.9, 4.3.2 und in § 1 gefundenen Irrationalitätskriterien weitere hinzu.

Während der euklidische Algorithmus bereits in 1.2.10 den regelmäßigen Kettenbruch einer *rationalen* Zahl lieferte, wird der allgemeine Fall *reeller* Zahlen in § 3 behandelt. Insbesondere läßt sich die Tatsache, daß eine reelle Zahl algebraisch vom Grade 1 (also rational) bzw. 2 ist, jeweils durch Eigenschaften ihrer Kettenbruchentwicklung charakterisieren. Damit zeigt sich am Ende, daß die Eulersche Zahl e eine nichtquadratische Irrationalzahl ist; dies leitet dann zu den arithmetischen Untersuchungen in Kap. 6 über.

§ 1. Die g–adische Entwicklung

In diesem Paragraphen sei $g \geq 2$ stets eine feste natürliche Zahl und S_g das in 2.1.4 eingeführte kleinste nichtnegative Restsystem $\{0, 1, \dots, g-1\}$ modulo g.

1. Entwicklung natürlicher Zahlen. Möchte man eine bestimmte natürliche Zahl identifizieren, z.B. um sie schriftlich zu übermitteln, so bedient man sich üblicherweise ihrer *dezimalen* (auch *dekadischen*)*) Darstellung. So wurde dem Leser in 1.1.8 die viertkleinste vollkommene Zahl in der Form 8128 mitgeteilt, was eine Kurzschreibweise ist für $8 \cdot 10^3 + 1 \cdot 10^2 + 2 \cdot 10^1 + 8 \cdot 10^0$.

Diese Art der Darstellung natürlicher Zahlen wird nun verallgemeinert zu folgendem

Satz. *Jede natürliche Zahl n hat eine eindeutige Darstellung der Form*

$$n = \sum_{i=0}^{k} a_i g^i \tag{1}$$

mit $a_0, \dots, a_k \in S_g$ und $a_k \neq 0$.

Beweis. Um die *Existenz* einer Summendarstellung des Typs (1) für n zu zeigen, setzt man erst $\nu_0 := n$ und wendet dann folgendes sukzessive Konstruktionsprinzip an: Sei $j \in \mathbb{N}_0$ und seien ganze $\nu_0, \dots, \nu_j, a_0, \dots, a_{j-1}$ schon so erhalten, daß sie den Bedingungen

$$\nu_i = \nu_{i+1} g + a_i, \qquad 0 \leq a_i < g \leq \nu_i \tag{2}$$

für $i = 0, \dots, j-1$ genügen. Sicher ist dann $\nu_i / g \geq \nu_{i+1} > 0$ für $i = 0, \dots, j-1$, also $\nu_0 g^{-j} \geq \nu_1 g^{-(j-1)} \geq \dots \geq \nu_j > 0$. Ist nun $\nu_j \geq g$, so wendet man auf das Paar (ν_j, g) den Divisionsalgorithmus 1.2.2 an und erhält (2) für $i = j$ mit ganzen ν_{j+1}, a_j. Ist jedoch $\nu_j < g$, so setzt man $a_j := \nu_j$ und hört auf; in diesem Fall ist $0 < a_j < g$.

Wegen $\nu_0 g^{-j} \geq \nu_j$ muß hier die zweite Alternative eintreten, sobald j größer als $\frac{\log n}{\log g}$ ist. Sie trete nach genau k (≥ 0) Schritten ein. Dann hat man also (2) für $i = 0, \dots, k-1$ und überdies $0 < a_k := \nu_k < g$. Daraus sieht man induktiv für $j = 0, \dots, k$

$$\nu_0 = \nu_j g^j + \sum_{i=0}^{j-1} a_i g^i,$$

woraus man (1) speziell für $j = k$ erhält; überdies haben die a_i die behaupteten Eigenschaften.

*) Abgeleitet vom lateinischen *decem* (bzw. griechischen $\delta\acute{\varepsilon}\kappa\alpha$) für *zehn.*

Um noch die *Eindeutigkeit* der Darstellung (1) einzusehen, beachte man, daß (1) und die Eigenschaften der a_i die Ungleichungen $g^k \leq n < g^{k+1}$ implizieren; deshalb ist k die größte ganze, $\frac{\log n}{\log g}$ nicht übersteigende Zahl und somit eindeutig festgelegt. Hat man nun eine weitere Darstellung für n, etwa

$$n = \sum_{i=0}^{k} a_i' g^i$$

mit $a_0', \dots, a_k' \in S_g$, $a_k' \neq 0$, so folgt aus

$$\sum_{i=0}^{k} (a_i' - a_i) g^i = 0 \tag{3}$$

die Teilbarkeit von $a_0' - a_0$ durch g, welche wegen $|a_0' - a_0| < g$ zu $a_0' = a_0$ führt. Berücksichtigt man dies in (3), so erweist sich $(a_1' - a_1)g$ als durch g^2 teilbar, d.h. $a_1' - a_1$ als durch g teilbar usw. Induktiv findet man $a_i' = a_i$ für $i = 0, \dots, k$. □

Die Darstellung (1) heißt die *g–adische Darstellung* von n, die a_i heißen die *Ziffern* dieser Darstellung, $1 + k = 1 + [\frac{\log n}{\log g}]$ ihre *Stellenzahl*. Die Zahl g, nach der entwickelt wird, heißt *Basis* der Darstellung. Weiter heißen $\sum_{i=0}^{k} a_i$ bzw. $\sum_{i=0}^{k} (-1)^i a_i$ in Verallgemeinerung zweier schon aus dem Schulunterricht bekannter Begriffe *g–adische Quersumme* bzw. *alternierende g–adische Quersumme* von n.

Bemerkung. Im Fall $g = 2$ spricht man von *dualer* (auch *binärer* oder *dyadischer*) Darstellung.

2. Teilbarkeitsregeln. Der Leser wird sich an die bekannten, den Fall $g = 10$ betreffenden Teilbarkeitsregeln natürlicher Zahlen durch 3, 9 bzw. 11 erinnern. Diese Regeln werden verallgemeinert in folgendem

Satz. *Ist d ein beliebiger Teiler von $g - 1$, so gilt: Dann und nur dann ist n durch d teilbar, wenn die g–adische Quersumme von n durch d teilbar ist. Für beliebige Teiler d von $g + 1$ hat man: n ist genau dann durch d teilbar, wenn dies für seine alternierende g–adische Quersumme zutrifft.*

Beweis. Die Behauptungen ergeben sich direkt aus

$$n = \sum_{i=0}^{k} a_i \big((g-1) + 1\big)^i \equiv \sum_{i=0}^{k} a_i \pmod{g-1}$$

bzw.

$$n = \sum_{i=0}^{k} a_i\big((g+1)-1\big)^i \equiv \sum_{i=0}^{k} (-1)^i a_i \pmod{g+1}. \qquad \square$$

Insbesondere ist also 3 bzw. 9 ein Teiler von n genau dann, wenn 3 bzw. 9 die (dezimale) Quersumme von n teilt; weiter ist 11 ein Teiler von n genau dann, wenn 11 in der alternierenden (dezimalen) Quersumme von n aufgeht.

Der *Vorteil derartiger Kriterien* besteht offenbar darin, daß sie die Frage nach der Teilbarkeit einer natürlichen Zahl n durch ein d reduzieren auf die Frage, ob die absolut genommen viel kleinere g–adische Quersumme bzw. alternierende Quersumme von n durch d teilbar ist für ein ganzes $g \geq 2$, welches $d|(g-1)$ bzw. $d|(g+1)$ genügt.

Beispielsweise ist die dezimal als 91813843 notierte natürliche Zahl n_0 nicht durch 3 (also erst recht nicht durch 9) teilbar, weil ihre (dezimale) Quersumme gleich 37 ist; ihre alternierende (dezimale) Quersumme ist -11 und so ist n_0 durch 11 teilbar. Bei der Entscheidung der Teilbarkeit der Zahl n_0 durch 7 hilft der hier gezeigte Satz offenbar nicht, da 7 weder in $10-1$ noch in $10+1$ aufgeht. Dagegen hat man für die Teilbarkeit durch 7 folgendes Kriterium

$$7|\big(\sum_{i=0}^{k} a_i 10^i\big) \quad \Leftrightarrow \quad 7|\big(\sum_{j\geq 0}(-1)^j(a_{3j}+3a_{3j+1}+2a_{3j+2})\big),$$

dessen Beweis dem Leser als Übung überlassen sei. Die Summe rechts ist gleich 39 für obiges n_0, welches somit nicht durch 7 teilbar ist. Selbstverständlich kann man eine Vielzahl derartiger Kriterien angeben, worauf hier jedoch nicht weiter eingegangen werden soll.

3. Der gebrochene Teil reeller Zahlen. Sind $c_1, c_2, \ldots \in S_g$, jedoch nicht alle gleich $g-1$, so definiert die Reihe $\sum_{i\geq 1} c_i g^{-i}$ wegen $0 \leq \sum_{i\geq 1} c_i g^{-i} < (g-1)\sum_{i\geq 1} g^{-i} = 1$ eine reelle Zahl des halboffenen Intervalls $[0,1[$. Daß auch umgekehrt jedes reelle $\alpha \in [0,1[$ eine Reihendarstellung obiger Art besitzt, die überdies unter einer geringfügigen Zusatzforderung an die c_i eindeutig ist, beinhaltet folgender

Satz. *Jedes reelle $\alpha \in [0,1[$ hat genau eine Entwicklung der Gestalt*

$$\alpha = \sum_{i=1}^{\infty} c_i g^{-i} \tag{1}$$

mit allen $c_1, c_2, \ldots \in S_g$*, von denen unendlich viele ungleich* $g-1$ *sind; dabei ergeben sich die* c_i *rekursiv aus*

$$(2) \qquad \alpha_1 := \alpha \ \ \textit{und} \ \ c_i := [\alpha_i g], \quad \alpha_{i+1} := \{\alpha_i g\} \qquad \textit{für } i \in \mathbb{N}.$$

Es sei hier an die in 4.3.2 eingeführte Bezeichnung $\{z\} := z - [z]$ für den gebrochenen Teil einer reellen Zahl z erinnert.

Beweis für die Existenz. Nach (2) ist α_i als gebrochener Teil einer reellen Zahl nichtnegativ, aber kleiner als 1, also $0 \leq \alpha_i g < g$ und somit $c_i \in S_g$ für jedes $i \geq 1$. Daß unendlich viele dieser c_i ungleich $g-1$ sind, sieht man so: Aus (2) folgt für $j \geq 1$ die Gleichung $\alpha_j = c_j g^{-1} + \alpha_{j+1} g^{-1}$ und daraus induktiv

$$(3) \qquad \alpha_j = \sum_{i=j}^{j+k-1} c_i g^{j-1-i} + \alpha_{j+k} g^{-k}$$

für alle ganzen $j \geq 1$, $k \geq 0$. Gäbe es nun ein ganzes $j \geq 1$ mit $c_i = g-1$ für alle $i \geq j$, so würde (3) durch Grenzübergang $k \to \infty$ zu $\alpha_j = (g-1)\sum_{t \geq 1} g^{-t} = 1$ führen, was $0 \leq \alpha_j < 1$ widerspricht. Wendet man nun (3) an mit $j = 1$, so bekommt man

$$\alpha = \sum_{i=1}^{k} c_i g^{-i} + \alpha_{k+1} g^{-k},$$

woraus man (1) bei $k \to \infty$ erhält.

Zur Eindeutigkeit : Hat man neben (1) eine weitere Darstellung

$$(4) \qquad \alpha = \sum_{i=1}^{\infty} c_i' g^{-i}$$

von α mit $c_1', c_2', \ldots \in S_g$, aber unendlich oft $\neq g-1$, so wird $c_i' = c_i$ für alle $i \geq 1$ behauptet. Nimmt man an, dies treffe nicht zu, so sei $j \geq 1$ der kleinste Index mit $c_j' \neq c_j$. Aus (1) und (4) folgt dann

$$\begin{aligned} 1 \leq |c_j - c_j'| &= \Big|\sum_{i>j} (c_i' - c_i) g^{j-i}\Big| \\ &\leq \sum_{i>j} |c_i' - c_i| g^{j-i} \leq (g-1) \sum_{t \geq 1} g^{-t} = 1, \end{aligned}$$

weshalb insbesondere $|c_i' - c_i| = g-1$ für alle $i > j$ gelten muß und überdies müssen entweder alle $c_i' - c_i$ mit $i > j$ positiv sein oder negativ. Dann ist $c_i' = g-1$, $c_i = 0$ für alle $i > j$ oder $c_i' = 0$, $c_i = g-1$ für dieselben i, was der Voraussetzung widerspricht, daß von den c_i bzw. c_i' jeweils unendlich viele von $g-1$ verschieden sein sollten. $\square$

Bemerkung. Würde man die soeben wiederholte Zusatzforderung an die c_i, c_i' nicht stellen, so wäre obiger Eindeutigkeitsbeweis genau dann undurchführbar, wenn bei geeignetem $\ell \in \mathbb{N}$ die Gleichungen $c_1' = c_1, \ldots, c_{\ell-1}' = c_{\ell-1}$ und (o.B.d.A.) $c_\ell' = c_\ell - 1$ (also $c_\ell \geq 1$) sowie $c_i' = g-1$, $c_i = 0$ für $i > \ell$ bestehen würden. Dann ist die durch (1) dargestellte Zahl

$$\sum_{i=1}^{\ell} c_i g^{-i} \tag{5}$$

rational und diese hätte tatsächlich (man beachte $c_\ell \geq 1$) die zweite Entwicklung

$$\sum_{i=1}^{\ell-1} c_i g^{-i} + (c_\ell - 1)g^{-\ell} + \sum_{i=\ell+1}^{\infty} (g-1)g^{-i}.$$

4. Entwicklung reeller Zahlen. Während in Satz 3 nur reelle Zahlen des Intervalls $[0, 1[$ behandelt wurden, erhält man aus dem dortigen Ergebnis direkt den

Satz. *Jede reelle Zahl α hat genau eine Darstellung der Form*

$$\alpha = [\alpha] + \sum_{i=1}^{\infty} c_i g^{-i} \tag{1}$$

mit allen $c_1, c_2, \ldots \in S_g$, von denen unendlich viele ungleich $g-1$ sind. Dabei ergeben sich die c_i rekursiv aus

$$\alpha_1 := \{\alpha\} \quad \textit{und} \quad c_i := [\alpha_i g], \quad \alpha_{i+1} := \{\alpha_i g\} \quad \textit{für } i \in \mathbb{N}. \tag{2}$$

Wenigstens dann, wenn $\alpha \geq 0$ ist, kann man auch noch die Ausnahmestellung von $[\alpha]$ rechts in (1) beseitigen: Man schreibt nämlich gemäß 1 eindeutig

$$[\alpha] = \sum_{i=0}^{k} a_i g^i \tag{3}$$

mit geeigneten $k \in \mathbb{N}_0$ und $a_0, \ldots, a_k \in S_g$, $a_k > 0$, falls $\alpha \geq 1$ gilt, bzw. mit $k = 0$, $a_0 = 0$ im Fall $0 \leq \alpha < 1$. Dann hat man aus (1) und (3) für die reelle Zahl $\alpha \geq 0$ die eindeutige Darstellung

$$\alpha = \sum_{i=-k}^{\infty} c_i g^{-i} \tag{4}$$

mit $k \in \mathbb{N}_0$ und $c_i := a_{-i}$ für $-k \le i \le 0$, während die c_i mit $i \ge 1$ aus (1) übernommen sind. Auch in (4) sind alle $c_{-k}, \ldots, c_0, c_1, \ldots \in S_g$, aber unendlich oft von $g-1$ verschieden.

Die unter den erwähnten Bedingungen an die c_i eindeutige Darstellung (1) bzw. (4) der reellen Zahl α nennt man deren *g–adische Entwicklung* (*nach fallenden Potenzen*), die Koeffizienten c_i heißen die g–Ziffern, kurz *Ziffern* dieser Entwicklung. Die Entwicklung nennt man *abbrechend*, wenn wie etwa in 3(5) höchstens endlich viele Ziffern von Null verschieden sind.

In Anlehnung an den historisch und für die Praxis besonders bedeutsamen Spezialfall $g = 10$ (wo man regelmäßig von der *Dezimalbruchentwicklung* einer Zahl spricht und nicht von “10–adischer Entwicklung nach fallenden Potenzen”) schreibt man die rechte Seite von (4) oft auch in der Form

$$c_{-k} \ldots c_0, c_1 c_2 \ldots c_i \ldots \tag{5}$$

bzw. $c_{-k} \ldots c_0, c_1 \ldots c_\ell$. Dabei ist letzteres üblich, wenn die g–adische Entwicklung abbricht und c_ℓ die letzte nichtverschwindende Ziffer mit positivem Index ist, vgl. 3(5). Die Zahl g, nach der man entwickelt hat, bleibt bei der Schreibweise (5) zwar unerwähnt, ist aber der jeweils vorab getroffenen Konvention zweifelsfrei zu entnehmen.

Bemerkung. Zur Abtrennung der Ziffern von ganzem und gebrochenem Teil einer nichtnegativen reellen Zahl scheint das Komma (oder im angloamerikanischen Sprachraum der Punkt) etwa um 1600 erstmals von J. NAPIER (auch NEPER geschrieben) bei Dezimalbruchentwicklungen verwendet worden zu sein.

5. Entwicklung rationaler Zahlen. Sind a, b teilerfremde ganze Zahlen mit $b > 0$, so sind die Ziffern der g–adischen Entwicklung der rationalen Zahl $\frac{a}{b}$ nach Satz 4, insbesondere nach 4(2), rekursiv zu ermitteln aus

$$\alpha_1 := \{\frac{a}{b}\} \quad \text{und} \quad c_i := [\alpha_i g], \quad \alpha_{i+1} := \{\alpha_i g\} \qquad \text{für } i \ge 1. \tag{1}$$

Setzt man noch

$$b_i := \alpha_i b, \tag{2}$$

so ist $b_i \in S_b$ für alle $i \ge 1$ aus (1) leicht einzusehen. Daher muß es ganze s, t mit $1 \le s < t$ geben, für die $b_s = b_t$, also auch $\alpha_s = \alpha_t$ zutrifft. Wegen (1) gilt dann $\alpha_{s+i} = \alpha_{t+i}$ für alle $i \ge 0$ und somit $c_{s+i} = c_{t+i}$ für dieselben i. Damit ist bewiesen der folgende

Satz. *Die Ziffernfolge der g–adischen Entwicklung jeder rationalen Zahl ist periodisch.*

Bemerkung. 1) Wegen $\alpha_i g = c_i + \alpha_{i+1}$ genügen die in (2) erklärten $b_i \in S_b$ der Rekursion

$$b_1 = \{\frac{a}{b}\}b, \quad b_i g = c_i b + b_{i+1} \qquad \text{für } i \in \mathbb{N}. \tag{3}$$

Dies läßt folgende Interpretation zu: Hat man a, b wie oben und *definiert* nun b_1 durch die erste Gleichung in (3), so gilt $b_1 \in S_b$ und man erhält c_1, b_2 durch Anwendung des Divisionsalgorithmus 1.2.2 auf $b_1 g$, b, vgl. (3). Hierbei ergibt sich $b_2 \in S_b$ und daraus $c_1 \in S_g$ nach (3) und man gewinnt induktiv die unendliche Folge $b_1, b_2, \ldots \in S_b$ ebenso wie die Ziffernfolge $c_1, c_2, \ldots \in S_g$ der g–adischen Entwicklung der rationalen Zahl $\frac{a}{b}$.

Die obige Folge (c_i) heißt (analog zu 2.3.2) *periodisch*, da es ein $p \in \mathbb{N}$ und ein $\ell \in \mathbb{N}_0$ gibt derart, daß

$$c_{i+p} = c_i \tag{4}$$

für alle ganzen $i > \ell$ gilt. Das minimale $p \in \mathbb{N}$, für das (4) für alle großen i gilt, heißt die *Periodenlänge* der Folge. Hat die Folge die Periodenlänge p, so heißt das minimale $\ell \in \mathbb{N}_0$, so daß (4) für alle $i > \ell$ zutrifft, die *Vorperiodenlänge* der Folge. Bei $\ell = 0$ heißt (c_i) *reinperiodisch*, bei $\ell \geq 1$ *gemischtperiodisch*.

Bevor in 6 für jede rationale Zahl Vorperioden– und Periodenlänge der Ziffernfolge ihrer g–adischen Entwicklung völlig explizit bestimmt werden, sei die Aussage des obigen Satzes noch ergänzt zu folgendem

Irrationalitätskriterium. *Eine reelle Zahl ist genau dann rational, wenn die Ziffernfolge ihrer g–adischen Entwicklung periodisch ist.*

Beweis. Ist die g–adische Ziffernfolge $(c_i)_{i\geq 1}$ eines reellen α periodisch, d.h. gilt für die c_i in 4(1) die Gleichung (4) für alle $i > \ell$, so ist

$$\begin{aligned}\{\alpha\} &= \sum_{i=1}^{\ell} c_i g^{-i} + g^{-\ell}\Big(\sum_{j=0}^{\infty} g^{-jp}\Big)\Big(\sum_{k=1}^{p} c_{k+\ell} g^{-k}\Big) \\ &= \frac{1}{g^\ell(g^p-1)}\Big((g^p-1)\sum_{i=1}^{\ell} c_i g^{\ell-i} + \sum_{k=1}^{p} c_{k+\ell} g^{p-k}\Big),\end{aligned} \tag{5}$$

was die Rationalität von α zeigt. Dies zusammen mit dem Satz liefert die Aussage des Kriteriums. □

Dieses Kriterium wird in 9 zu einem Irrationalitätsbeweis angewandt. Generell ist die entscheidende Hürde für seine Anwendungsfähigkeit offenbar die Tatsache, daß für mindestens ein ganzes $g \geq 2$ die g–adische Entwicklung der zu untersuchenden Zahl bekannt sein muß, d.h. man muß die entsprechende g–Ziffernfolge (c_i) explizit kennen.

Bemerkung. 2) In Anlehnung an den Spezialfall $g = 10$ bringt man die Tatsache, daß die Ziffernfolge einer rationalen Zahl aus $[0, 1[$ Vorperiodenlänge ℓ und Periodenlänge p hat, durch die Schreibweise $0, c_1 \dots c_\ell \overline{c_{\ell+1} \cdots c_{\ell+p}}$ zum Ausdruck, vgl. 4(5). Den (eventuell fehlenden) Ziffernblock $c_1 \dots c_\ell$ nennt man die *Vorperiode*, den Block $c_{\ell+1} \dots c_{\ell+p}$ die *Periode*.

6. Periodizitätseigenschaften der Ziffernfolge. Im folgenden Satz zeigt sich, daß Vorperioden– und Periodenlänge der g–Ziffernfolge einer gekürzten rationalen Zahl $\frac{a}{b}$ nur von b und g, nicht jedoch von a abhängen.

Satz. *Seien a, b teilerfremde ganze Zahlen mit $b > 0$; sei b^* der größte positive, zu g teilerfremde Teiler von b und sei $b^{**} := \frac{b}{b^*}$. Dann ist die Ziffernfolge der g–adischen Entwicklung der rationalen Zahl $\frac{a}{b}$ periodisch mit der Periodenlänge $\mathrm{ord}_{b^*} g$ und der Vorperiodenlänge $\mathrm{Min}\{\mu \in \mathbb{N}_0 : b^{**} | g^\mu\}$. Diese Entwicklung ist abbrechend genau dann, wenn $b^* = 1$ gilt, d.h. wenn jeder Primfaktor von b in g aufgeht.*

Es sei daran erinnert, daß hier $\mathrm{ord}_{b^*} g$ die in 2.3.2 eingeführte Ordnung von g modulo b^* bedeutet, also das kleinste $q \in \mathbb{N}$ mit $g^q \equiv 1 \pmod{b^*}$.

Beweis. Nach Satz 5 ist die g–adische Ziffernfolge $(c_i)_{i \geq 1}$ von $\frac{a}{b}$ periodisch; ℓ bzw. p seien die Vorperioden– bzw. Periodenlänge. Bei geeigneten ganzen A, B ist nach 5(5)

$$\frac{a}{b} = \left[\frac{a}{b}\right] + \frac{A}{g^\ell (g^p - 1)} = \frac{B}{g^\ell (g^p - 1)},$$

woraus wegen $(a, b) = 1$ die Bedingung $b | g^\ell (g^p - 1)$ folgt, also $b^* | (g^p - 1)$ und $b^{**} | g^\ell$. Mit den Festsetzungen

(1) $$q := \mathrm{ord}_{b^*} g, \quad m := \mathrm{Min}\{\mu \in \mathbb{N}_0 : b^{**} | g^\mu\}$$

folgt daraus $m \leq \ell$ und $q \leq p$.

Andererseits besagt (1) erstens $b^* | (g^q - 1)$, zweitens $b^{**} | g^m$, insgesamt also $b | g^m (g^q - 1)$ und so ist $\{\frac{a}{b}\} g^m (g^q - 1) \in \mathbb{N}_0$. Nach dem Divisionsalgorithmus 1.2.2 existieren ganze u, v mit

(2) $$\left\{\frac{a}{b}\right\} g^m (g^q - 1) = u(g^q - 1) + v$$

und $0 \le v < g^q - 1$, was dann $0 \le u < g^m$ nach sich zieht. Wegen Satz 1 gelten mit $u_1, \ldots, u_m, v_1, \ldots, v_q \in S_g$ die Gleichungen

$$(3) \qquad u = u_m + u_{m-1}g + \ldots + u_1 g^{m-1}, \quad v = v_q + v_{q-1}g + \ldots + v_1 g^{q-1},$$

wobei nicht alle $v_1, \ldots, v_q$ gleich $g-1$ sind, man beachte $v < g^q - 1$. Eintragen der Darstellungen (3) für u, v in (2) führt zu

$$\begin{aligned}\{\frac{a}{b}\} &= ug^{-m} + v(1-g^{-q})^{-1}g^{-m-q} \\ &= \sum_{i=1}^{m} u_i g^{-i} + g^{-m}\Big(\sum_{j=1}^{q} v_j g^{-j}\Big)\Big(\sum_{k=0}^{\infty} g^{-kq}\Big) = \sum_{i=1}^{\infty} d_i g^{-i},\end{aligned}$$

wobei

$$(4) \qquad d_i := u_i \ \ (1 \le i \le m), \qquad d_{m+jq+k} := v_k \ \ (j \in \mathbb{N}_0,\ 1 \le k \le q)$$

gesetzt ist. Danach sind alle d_i $(i \ge 1)$ aus S_g, jedoch gilt $d_i \ne g-1$ unendlich oft, weil $v_1, \ldots, v_q$ nicht alle gleich $g-1$ sind. So ist $[\frac{a}{b}] + \sum_{i\ge 1} d_i g^{-i}$ die g–adische Entwicklung von $\frac{a}{b}$, deren Ziffernfolge (d_i) nach (4) eine Periodenlänge höchstens q und eine Vorperiodenlänge höchstens m hat. Dies besagt $p \le q$ und $\ell \le m$, insgesamt also $p = q$, $\ell = m$.

Die g–adische Entwicklung von $\frac{a}{b}$ ist abbrechend genau dann, wenn $\{\frac{a}{b}\} = \sum_{i=1}^{\ell} c_i g^{-i}$ mit geeignetem $\ell \in \mathbb{N}_0$ gilt; dies letztere ist mit $\frac{a}{b} = \frac{A}{g^\ell}$ bei geeignetem ganzem A äquivalent, was wegen $(a,b) = 1$ wiederum mit $b|g^\ell$ gleichwertig ist. Aus $b|g^\ell$ folgt offenbar $b^* = 1$; ist umgekehrt $b^* = 1$, so hat man $b = b^{**}|g^m = g^\ell$, also $b|g^\ell$. Dabei wurde die Definition von m in (1) benützt ebenso wie die bereits eingesehene Gleichung $m = \ell$. □

Bemerkungen. 1) Bei $b^* > 1$ sind zwei Fälle möglich: Erstens $b^* = b > 1$, $b^{**} = 1$, was $\ell = 0$, also die Reinperiodizität der Ziffernfolge bedeutet. Zweitens $1 < b^*, b^{**} < b$; hier ist $\ell > 0$ und somit die Ziffernfolge gemischtperiodisch. In beiden Fällen gilt aber bei $p = 1$: Alle Ziffern nach Ablauf der Vorperiode sind gleich einer festen Zahl aus $\{1, \ldots, g-2\}$.

2) Ist speziell $(b,g) = 1$, also $b^* = b$, $b^{**} = 1$, und ist außerdem g eine Primitivwurzel modulo b, so ist $p = \mathrm{ord}_b g = \varphi(b)$. Z.B. ist 10 eine Primitivwurzel modulo 7 und so fällt bei $g = 10$ und rationalen Zahlen der Form $\frac{a}{7}$ mit $a \in \mathbb{Z}$, $7 \nmid a$ die Ziffernfolge stets reinperiodisch mit der Periodenlänge $\varphi(7) = 6$ aus.

7. Dezimalbruchentwicklungen. Die stets periodische Dezimalbruchentwicklung einer rationalen Zahl $\frac{a}{b}$ in gekürzter Darstellung (d.h. mit teilerfemden

a, b) bricht nach Satz 6 genau dann ab, wenn b höchstens die Primfaktoren 2 und 5 besitzt. Weiter ist sie reinperiodisch genau dann, wenn $b^{**} = 1$, $b^* = b$ ist, d.h. wenn $2\not| b$ und $5\not| b$ gilt, und hier ist die Periodenlänge gleich der Ordnung von 10 modulo b, insbesondere also ein Teiler von $\varphi(b)$. Die Zahl $\frac{a}{b}$ hat eine reinperiodische Dezimalbruchentwicklung der Periodenlänge $\varphi(b)$ genau dann, wenn 10 eine Primitivwurzel modulo b ist. Der nachfolgenden kleinen Tabelle entnimmt man, daß 1, 7, 17, 19, 23, 29 sämtliche natürliche b unterhalb 30 sind, modulo derer 10 Primitivwurzel ist.

b	1	3	7	9	11	13	17	19	21	23	27	29
$\varphi(b)$	1	2	6	6	10	12	16	18	12	22	18	28
$\mathrm{ord}_b 10$	1	1	6	1	2	6	16	18	6	22	3	28

Bei der Herstellung solcher Tabellen kann man sich für die dritte Zeile der Tatsache bedienen, daß bei paarweise teilerfremden b_1, b_2, 10 nach Lemma 2.3.5 gilt: $\mathrm{ord}_{b_1 b_2} 10 = \mathrm{kgV}(\mathrm{ord}_{b_1} 10, \mathrm{ord}_{b_2} 10)$.

In der vorstehenden Tabelle liest man in der letzten Zeile die Periodenlänge der (reinperiodischen) Dezimalbruchentwicklung von gekürzten $\frac{a}{b}$, b wie in der ersten Zeile, ab. Z.B. haben die rationalen Zahlen $\frac{1}{7}$, $\frac{1}{13}$, $\frac{1}{21}$ sämtliche eine reinperiodische Dezimalbruchentwicklung der Periodenlänge 6, nämlich $0,\overline{142857}$; $0,\overline{076923}$; $0,\overline{047619}$. Bei der Entwicklung von $\frac{3}{17}$ bzw. $\frac{3}{29}$ treten die Periodenlängen 16 bzw. 28 auf und man hat die Dezimalbrüche $0,\overline{1764705882352941}$ bzw. $0,\overline{1034482758620689655172413793}$.

Nach Satz 6 hat ein gekürztes $\frac{a}{b}$ genau dann eine nicht abbrechende gemischtperiodische Dezimalbruchentwicklung, wenn unter den Primfaktoren von b mindestens einer aus $\{2, 5\}$ und mindestens einer aus $\mathbb{P} \setminus \{2, 5\}$ vorkommt. Die nachfolgende kleine Tabelle enthält alle die Zahl 30 nicht übersteigenden natürlichen b mit dieser Eigenschaft; außerdem sind die jeweils zur vollständigen Klärung der Periodizitätsverhältnisse der Dezimalbruchentwicklungen gekürzter $\frac{a}{b}$ benötigten Angaben gemacht:

b	6	12	14	15	18	22	24	26	28	30
b^*	3	3	7	3	9	11	3	13	7	3
$\mathrm{ord}_{b^*} 10$	1	1	6	1	1	2	1	6	6	1
b^{**}	2	4	2	5	2	2	8	2	4	10
$\mathrm{Min}\{\mu \in \mathbb{N}_0 : b^{**} \mid 10^\mu\}$	1	2	1	1	1	1	3	1	2	1

Der dritten bzw. fünften Zeile entnimmt man die Perioden– bzw. Vorperiodenlängen. Z.B. ist bei $\frac{1}{12}$ und bei $\frac{1}{28}$ die Vorperiodenlänge 2, während die Periodenlängen 1 und 6 sind; die Dezimalbruchentwicklungen lauten hier $0,08\overline{3}$ und $0,03\overline{571428}$.

8. Rationale Zahlen mit gleichen Nennern. Der Leser wird bereits im frühen Schulunterricht folgende Bemerkung gemacht haben, wenn er dort Dezimalbruchentwicklungen zu üben hatte: $\frac{1}{7} = 0,\overline{142857}$; $\frac{2}{7} = 0,\overline{285714}$; $\frac{3}{7} = 0,\overline{428571}$; $\frac{4}{7} = 0,\overline{571428}$; $\frac{5}{7} = 0,\overline{714285}$; $\frac{6}{7} = 0,\overline{857142}$. Dies Phänomen wird vollständig geklärt im nachfolgenden

Satz. *Sei p eine nicht in g aufgehende Primzahl, modulo der g Primitivwurzel ist. Dann haben die rationalen Zahlen $\frac{a}{p}$ $(a = 1, \ldots, p-1)$ reinperiodische g–adische Entwicklungen der Periodenlänge $p-1$ und ihre Perioden gehen auseinander durch zyklische Vertauschung hervor.*

Beweis. Man startet den Algorithmus 5(3) mit $a = 1$, $b = p$ und erhält zunächst $b_1 = 1$. Da jedes b_i $(i \geq 1)$ sowohl c_i als auch b_{i+1} eindeutig bestimmt und da nach Satz 6 die Periodenlänge der reinperiodischen Entwicklung von $\frac{1}{p}$ bereits als $p-1$ feststeht, müssen $b_1, \ldots, b_{p-1}$ paarweise verschieden und $\neq 0$ sein, während b_p wieder gleich $b_1 = 1$ ist. Würde nämlich ein $b_i = 0$ sein, so wären auch $c_i = b_{i+1} = 0$ nach 5(3) und $\frac{1}{p}$ hätte eine abbrechende Entwicklung; wäre $b_p \neq b_1$, so müßte $b_p \in \{b_2, \ldots, b_{p-1}\}$ sein und dann wäre die Periodenlänge kleiner als $p-1$. Die ersten $p-1$ Gleichungen rechts in 5(3) lauten daher

$$(1) \qquad b_1 g = c_1 p + b_2, \ldots, b_i g = c_i p + b_{i+1}, \ldots, b_{p-1} g = c_{p-1} p + b_1.$$

Startet man in 5(3) nun mit $b = p$ und beliebigem $a \in \{2, \ldots, p-1\}$, so wird dort $b_1 = a$ und a ist gleich genau einem der in (1) anfallenden $b_2, \ldots, b_{p-1}$, etwa gleich b_i. Dann ist klar, daß $0, \overline{c_i \ldots c_{p-1} c_1 \ldots c_{i-1}}$ die g–adische Entwicklung von $\frac{a}{p}$ sein muß. □

9. Eine Anwendung des Irrationalitätskriteriums. Man definiert α_0 durch die Dezimalbruchentwicklung $0,1234567891011\ldots$, die dadurch entsteht, daß man hinter dem Komma nacheinander die gemäß Satz 1 dezimal dargestellten natürlichen Zahlen hinschreibt. Präziser und allgemeiner gefaßt sieht dies so aus: Es ist $s(n) := 1 + [\frac{\log n}{\log g}]$ die Stellenzahl der g–adischen Darstellung von n gemäß 1. Diese Darstellung selbst möge

$$\sum_{i=0}^{s(n)-1} a_i(n) g^i$$

lauten, wobei also die $a_0(n), a_1(n), \ldots, a_{s(n)-1}(n)$ die Ziffern (die letzte ist hier $\neq 0$) in der Darstellung von n nach Satz 1 sind. Setzt man dann

$$\alpha(g) := \sum_{n=1}^{\infty} \Big(\sum_{i=0}^{s(n)-1} a_i(n) g^i \Big) g^{-\sum_{m=1}^{n} s(m)}, \tag{3}$$

so ist gerade $\alpha(10) = \alpha_0$ und allgemein wird behauptet die

Proposition. *Die in (3) definierte reelle Zahl $\alpha(g)$ ist irrational.*

Beweis. Man überlegt sich, daß in der Ziffernfolge der g–adischen Entwicklung von $\alpha(g)$ beliebig lange Folgen sukzessiver Nullen vorkommen, daß die Entwicklung aber nicht abbricht: Ist nämlich n von der Form g^N, so ist $s(g^N) = N + 1$ und $a_0(g^N) = \ldots = a_{N-1}(g^N) = 0$, $a_N(g^N) = 1$. In $\sum_{i \geq 1} c_i g^{-i}$ ist also $c_i = 1$ für $i = 1 + \sum_{m=1}^{g^N - 1} s(m) =: T(N)$, aber $c_i = 0$ für $i = 1 + T(N), \ldots, N + T(N)$ und so hat man N sukzessive Nullen und N kann dabei beliebig groß gewählt werden. □

Bemerkung. MAHLER hat 1937 sogar die Transzendenz aller $\alpha(g)$ gezeigt, vgl. Bemerkung zu 6.2.2.

10. Existenz transzendenter Zahlen. Hier soll Satz 3 noch benützt werden, um zu zeigen, daß es transzendente reelle Zahlen gibt. Das dafür verwendete Argument geht auf eine Idee von G. CANTOR (Gesammelte Abhandlungen, 115–118) aus dem Jahre 1874 zurück: Man zeigt, daß es überabzählbar viele reelle Zahlen gibt, wovon nur abzählbar viele algebraisch sein können.

Vorab sei daran erinnert, daß man eine Menge genau dann *abzählbar* nennt, wenn sie sich bijektiv auf die Menge $\mathbb{N}$ (oder $\mathbb{N}_0$) abbilden läßt. Ist eine Menge weder endlich noch abzählbar, so heißt sie *überabzählbar*. Damit formuliert man

Satz A. *Die Menge $\mathbb{R}$ aller reeller Zahlen (und damit jede $\mathbb{R}$ umfassende Menge, z.B. $\mathbb{C}$) ist überabzählbar.*

Beweis. Es reicht, die Überabzählbarkeit einer geeigneten Teilmenge von $\mathbb{R}$ zu beweisen, etwa diejenige des halboffenen Intervalls $[0, 1[$. Man nimmt an, diese letztere Menge sei abzählbar und $r_1, r_2, \ldots$ sei eine Abzählung von ihr. Sodann fixiert man ein beliebiges ganzes $g \geq 3$ und schreibt jedes r_n gemäß Satz 3 in seiner g–adischen Entwicklung

$$r_n = \sum_{i=1}^{\infty} c_i(n) g^{-i} \qquad \text{für } n = 1, 2, \ldots,$$

wobei man $0 \leq r_n < 1$ für alle $n \geq 1$ beachtet hat. Nun definiert man eine neue Ziffernfolge $c_1, c_2, \ldots$ durch die Festsetzung

$$(1) \qquad c_i := \begin{cases} 1, & \text{falls } c_i(i) \neq 1, \\ 0, & \text{falls } c_i(i) = 1, \end{cases} \qquad i = 1, 2, \ldots .$$

Hier sind sogar alle $c_1, c_2, \ldots \in \{0, \ldots, g-2\}$ und so gehört die durch die Reihe $\sum_{i \geq 1} c_i g^{-i}$ definierte reelle Zahl r sicher zum Intervall $[0, 1[$, d.h. mit geeignetem $m \in \mathbb{N}$ ist $r = r_m$. Die Eindeutigkeit der g–adischen Entwicklung zeigt dann $c_i = c_i(m)$ für alle $i \in \mathbb{N}$, insbesondere $c_i = c_i(i)$ entgegen (1). □

Für das nächste Ergebnis werden folgende Vorbemerkungen über Polynome und algebraische Zahlen benötigt. Für $P \in \mathbb{C}[X]$, etwa $P = \sum a_i X^i$, wird $L(P) := \sum |a_i|$ gesetzt; ist außerdem $P \neq 0$ und $\partial(P)$ der Grad von P, so sei $s(P) = L(P) + \partial(P)$ geschrieben. Für $P \in \mathbb{Z}[X]$, $P \neq 0$ ist dann $s(P) \in \mathbb{N}$ klar.

Weiter bezeichne $W(P)$ für $P \in \mathbb{C}[X]$, $P \neq 0$ die Menge der verschiedenen komplexen Wurzeln von P; als Konsequenz des Abspaltungslemmas 1.5.8 ist die Abschätzung $\#W(P) \leq \partial(P)$ klar.

Nun gilt folgender

Satz B. *Die Menge aller komplexen algebraischen Zahlen ist abzählbar.*

Beweis. Klar ist, daß es zu jedem vorgegebenen $s \in \mathbb{N}$ nur eine endliche (von s abhängige) Anzahl von $P \in \mathbb{Z}[X]$, $P \neq 0$ mit $s(P) = s$ geben kann. Daher ist die Menge aller $P \in \mathbb{Z}[X]$ abzählbar. Die im Satz genannte Menge ist aber gerade die Vereinigung der $W(P)$ über die abzählbar vielen $P \in \mathbb{Z}[X]$ mit $P \neq 0$ und ist somit bekanntlich selbst abzählbar. □

Durch Kombination der Sätze A und B erhält man unmittelbar

Korollar A. *Es gibt überabzählbar viele transzendente reelle Zahlen.*

Wie man aus den vorstehenden Argumentationen ersehen konnte, ist der CANTORsche Beweis für die Existenz transzendenter Zahlen *nicht konstruktiv*. In diesem Zusammenhang ist zu bemerken, daß bereits 30 Jahre vor CANTOR im Jahre 1844 LIOUVILLE einen *konstruktiven* Existenzbeweis geführt hat, der in 6.1.2ff. gebracht wird.

Da abzählbare Teilmengen von $\mathbb{R}$ bzw. $\mathbb{C}$ jeweils LEBESGUE–Maß Null haben, folgt aus Satz B noch als Verschärfung von Korollar A

Korollar B. *Im Sinne des* LEBESGUE*–Maßes ist fast jede reelle bzw. komplexe Zahl transzendent.*

So einfach diese maßtheoretische Aussage zu erhalten ist, so schwer ist es doch in aller Regel, die Transzendenz "in der Natur vorkommender" Zahlen zu beweisen. Einen gewissen Eindruck wird man davon in Kapitel 6 erhalten.

11. Dezimalbruchentwicklung und Dichtung. Wie schon am Ende von 5 erwähnt, benötigt man zur Klärung gewisser Fragen (z.B. für Irrationalitätsbeweise mittels Kriterium 5) möglichst umfassende Kenntnisse über die g–adische Entwicklung reeller Zahlen für mindestens ein ganzes $g \geq 2$.

Hier aber beginnen die heute noch weitestgehend offenen Probleme. Vielleicht die naheliegendste Frage ist die nach der effektiven Angabe der g–adischen Entwicklung (für wenigstens ein $g \geq 2$) gewisser "in der Natur vorkommender" reeller Zahlen. Darunter seien Zahlen verstanden, die bei manchen Problemen in der Mathematik in natürlicher Weise auftreten wie z.B. $\sqrt{2}$, e, π oder die EULERsche Konstante $\gamma := \lim_{n\to\infty}(\sum_{\nu=1}^{n} \frac{1}{\nu} - \log n)$. Bei den erstgenannten Zahlen ist zwar die Irrationalität bekannt (vgl. 1.1.9, 2.2, 6.3.2), die bei γ auch noch offen ist, jedoch kennt man in keinem Fall eine g–adische Entwicklung.

Selbstverständlich hat man im Zeitalter der immer leistungsfähigeren Computer versucht, für Zahlen wie $\sqrt{2}$, e, π mehr und mehr Stellen der entsprechenden Dezimalbruchentwicklungen zu berechnen, vielleicht in der (bisher unerfüllt gebliebenen) Hoffnung, doch einmal in einem Einzelfall eine "einfache" Gesetzmäßigkeit experimentell entdecken und dann beweisen zu können. Bei solchen numerischen Rechnungen sucht man stets zunächst eine geeignete Darstellung der zu entwickelnden Zahl durch einen genügend rasch konvergenten Grenzprozeß. Im Fall $\sqrt{2}$ bzw. e kann man sich z.B. der Entwicklung in einen sogenannten regelmäßigen Kettenbruch bedienen (vgl. 3.4 und 3.12). Hier seien noch die derzeit gültigen Rekorde für die gesicherten Dezimalstellen–Anzahlen bei $\sqrt{2}$, e, π mitgeteilt:

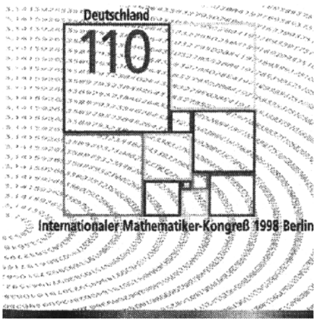
$\sqrt{2}$	$2 \cdot 10^{11}$ Stellen	S. KONDO (2006)
e	10^{11} Stellen	S. KONDO, S. PAGLIARULO (2007)
π	$12,4 \cdot 10^{11}$ Stellen	Y. KANADA (2002).

Für π beginnt der Computer–Ausdruck wie oben auf Seite 215 angegeben.

Ergänzend sei auch auf die Versuche hingewiesen, die in verschiedenen Sprachen unternommen wurden, Merkverse zu ersinnen, um den Beginn der Dezimalbruchentwicklung von π mnemotechnisch zu verschlüsseln, vgl. etwa H. TIETZE (*Gelöste und ungelöste mathematische Probleme aus alter und neuer Zeit*, Bie-

3,14159	26535	89793	23846	26433	83279	50288	41971	69399	37510
58209	74944	59230	78164	06286	20899	86280	34825	34211	70679
82148	08651	32823	06647	09384	46095	50582	23172	53594	08128
48111	74502	84102	70193	85211	05559	64462	29489	54930	38196
44288	10975	66593	34461	28475	64823	37867	83165	27120	19091
45648	56692	34603	48610	45432	66482	13393	60726	02491	41273
72458	70066	06315	58817	48815	20920	96282	92540	91715	36436
78925	90360	01133	05305	48820	46652	13841	46951	94151	16094
33057	27036	57595	91953	09218	61173	81932	61173	31051	18548
07446	23799	62749	56735	18857	52724	89122	79381	83011	94912
98336	73362	44065	66430	86021	39494	63952	24737	19070	21798
60943	70277	05392	17176	29317	67523	84674	81846	76694	05132
00056	81271	45263	56082	77857	71342	75778	96091	73637	17872
14684	40901	22495	34301	46549	58537	10507	92279	68925	89235
42019	95611	21290	21960	86403	44181	59813	62977	47713	09960
51870	72113	49999	99837	29780	49951	05973	17328	16096	31859
50244	59455	34690	83026	42522	30825	33446	85035	26193	11881
71010	00313	78387	52886	58753	32083	81420	61717	76691	47303
59825	34904	28755	46873	11595	62863	88235	37875	93751	95778
18577	80532	17122	68066	13001	92787	66111	95909	21642	01989

derstein, München, 1949). Bei solchen Merkversen hat man jedes Wort durch die Anzahl seiner Buchstaben zu ersetzen. Allerdings bildet die 32. Ziffer nach dem Komma jeder derartigen dichterischen Übung, unabhängig von der Sprache, eine ganz natürliche Grenze.



Der nebenstehend abgebildete Entwurf der Sonderbriefmarke zum Internationalen Mathematiker–Kongreß 1998 in Berlin zeigt (unter anderem) immer längere Ziffernblöcke vom Anfang der Entwicklung von π, kreisrund angeordnet. Vielleicht Symbolik für ein vollbesetztes aufsteigendes Auditorium?

12. Historische Anmerkungen. Das Prinzip der g–adischen Entwicklung von (rationalen) Zahlen nach fallenden Potenzen war bereits im Altertum geläufig. Insbesondere haben die Sumerer ab dem dritten vorchristlichen Jahrtausend und dann die nachfolgenden semitischen Babylonier beim praktischen Rechnen das Sexagesimalsystem (also $g = 60$) verwendet. So findet sich z.B. ein Keilschrifttext (etwa 2000 v. Chr.; Yale Babylonian Collection 7289), in dem das Verhältnis von Diagonalen– zu Seitenlänge eines Quadrats (hier aus der Keilschrift umgeschrieben) zu 1 24 51 10, also $1 + 24 \cdot 60^{-1} + 51 \cdot 60^{-2} + 10 \cdot 60^{-3} = \frac{30547}{21600}$ angegeben wurde. Dies entspricht einer dezimal geschriebenen Approximation $1,4142129\ldots$ für $\sqrt{2}$, die immerhin auf fünf Nachkommaziffern korrekt ist.

Während sich das Sexagesimalsystem in der Astronomie aller Kulturvölker halten konnte — man denke etwa an die Minuten–Sekunden–Einteilung von Grad oder Stunde —, wurde es aus den übrigen Naturwissenschaften und der Mathematik zwischen 1000 und 1500 durch das Dezimalsystem langsam verdrängt. Einer der Gründe hierfür war die Verbreitung der praktischen zehn Symbole 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 über die ganze (arabische) Welt von Indien, wo die letzten neun ursprünglich herkamen (die Null stammt aus Griechenland) über Kleinasien, Nordafrika und Südeuropa. Eine der ersten systematischen Darstellungen der Dezimalbrüche gab S. STEVIN (*De Thiende*, 1585). Übrigens findet der Leser eine faszinierende Schilderung dieses hier angesprochenen Teils der Mathematikgeschichte bei B.L. VAN DER WAERDEN (*Erwachende Wissenschaft*, 2. Aufl., Birkhäuser, Basel–Stuttgart, 1966).

Die erste einigermaßen erschöpfende neuzeitliche Darstellung der Theorie der allgemeinen g–adischen Entwicklung scheint von O. STOLZ (*Vorlesungen über allgemeine Arithmetik*, Teubner, Leipzig, 1886) zu stammen.

Schließlich sei noch auf das Dualsystem (also $g = 2$) hingewiesen, das sich in Ansätzen bereits bei australischen Eingeborenen vor vielen Jahrhunderten findet. Vor rund 300 Jahren hat dann insbesondere LEIBNIZ dieses System ausführlich untersucht, welches für die heutigen Computer wegen seiner einfachen Eigenschaften (z.B. kommt es mit zwei Ziffern aus) außerordentliche Bedeutung erlangt hat.

§ **2. Die Cantorsche Entwicklung. Weitere Irrationalitätskriterien**

In diesem Paragraphen sei stets $(g_i)_{i\geq 1}$ eine unendliche Folge ganzer Zahlen, sämtliche nicht kleiner als 2; mit diesen werde $P_i := g_1 \cdot \ldots \cdot g_i$ für alle $i \in \mathbb{N}_0$ gesetzt.

1. Beschreibung der Entwicklung. CANTOR (Gesammelte Abhandlungen, 35–42) hat 1869 den in 1.4 beschriebenen Algorithmus zur Gewinnung der g–adischen Entwicklung einer reellen Zahl verallgemeinert. Diese Verallgemeinerung soll hier kurz dargestellt werden; in 2 wird dann das zugehörige notwendige und hinreichende Irrationalitätskriterium nebst Anwendungen präsentiert.

Ist $\alpha \in \mathbb{R}$ beliebig, so läuft der CANTORsche Algorithmus wie folgt ab: Man schreibe analog zu 1.4(2) zunächst

$$\alpha_1 := \{\alpha\}. \tag{1}$$

Sei weiter $k \geq 1$ und seien $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$; $c_1, \ldots, c_{k-1} \in \mathbb{Z}$ bereits so bestimmt, daß sie den Bedingungen

$$\alpha_i g_i = c_i + \alpha_{i+1} \tag{2}$$

für $i = 1, \ldots, k-1$ sowie

$$0 \le \alpha_i < 1 \tag{3}$$

für $i = 1, \ldots, k$ genügen. (3) führt insbesondere zu $c_i \in S_{g_i}$ für $i = 1, \ldots, k-1$. Nun definiert man

$$c_k := [\alpha_k g_k], \quad \alpha_{k+1} := \{\alpha_k g_k\} \tag{4}$$

und hat somit (2) bzw. (3) (einschließlich $c_k \in S_{g_k}$) auch für $i = k$ bzw. $i = k+1$. Nun kann man behaupten den

Satz. *Jede reelle Zahl α hat die eindeutig bestimmte* CANTOR*sche Entwicklung*

$$\alpha = [\alpha] + \sum_{i=1}^{\infty} c_i P_i^{-1},$$

wobei die c_i aus (1) und (4) rekursiv zu ermitteln sind und den Bedingungen $c_i \in \{0, \ldots, g_i - 1\}$ für alle $i \in \mathbb{N}$, jedoch unendlich oft von $g_i - 1$ verschieden, genügen.

Wählt man hier alle $g_1, g_2, \ldots$ gleich ein und demselben ganzen $g \ge 2$, so reduziert sich der behauptete Satz auf Satz 1.4.

Beweis. Zunächst sieht man induktiv aus (1) und (2)

$$\alpha = [\alpha] + \sum_{i=1}^{n} c_i P_i^{-1} + \alpha_{n+1} P_n^{-1} \qquad \text{für } n = 0, 1, \ldots . \tag{5}$$

Wegen (3) und der Voraussetzung $g_j \ge 2$ ist hier der nichtnegative letzte Summand rechts kleiner als 2^{-n}, so daß man aus (5) durch Grenzübergang sofort eine Entwicklung für α der im Satz gewünschten Art erhält, wenn man noch ausschalten kann, daß bei geeignetem n für alle $i > n$ die Gleichheit $c_i = g_i - 1$ eintritt. In diesem Falle wäre nämlich

$$\sum_{i>n} c_i P_i^{-1} = \sum_{i>n} P_{i-1}^{-1} - \sum_{i>n} P_i^{-1} = P_n^{-1}, \tag{6}$$

was wegen (5) mit $\alpha_{n+1} = 1$ äquivalent wäre; dies aber widerspricht (3). Hat man eine weitere CANTORsche Entwicklung $c' + \sum_{i\geq 1} c'_i P_i^{-1}$ für α, so sei bereits $c' = [\alpha]$, $c'_1 = c_1, \ldots, c'_{n-1} = c_{n-1}$ als richtig erkannt. Benützt man diese Gleichungen, so erhält man durch Gleichsetzen der beiden Reihen für α nach Multiplikation mit P_n und bei Beachtung von (6)

$$|c'_n - c_n| = |\sum_{i>n}(c_i - c'_i) \prod_{j=n+1}^{i} g_j^{-1}| < \sum_{i>n}(g_i - 1) \prod_{j=n+1}^{i} g_j^{-1} = 1.$$

Hier muß die *strenge* Ungleichheit gelten, weil sonst entweder $c_i - c'_i = g_i - 1$ für alle $i > n$ oder $c_i - c'_i = 1 - g_i$ für dieselben i gelten müßte; somit hat man $c'_n = c_n$, also die Eindeutigkeit der CANTORschen Entwicklung. □

2. Cantorsche Reihen und Irrationalität. Basierend auf der CANTORschen Entwicklung wird nun ein weiteres *notwendiges und hinreichendes* Kriterium für die Irrationalität reeller Zahlen angegeben.

Irrationalitätskriterium. *Sei die Folge $(g_i)_{i\geq 1}$ so beschaffen, daß es zu jeder Primzahl p unendlich viele j gibt mit $p|g_j$. Eine reelle Zahl α ist genau dann irrational, wenn für ihre* CANTOR*sche Entwicklung gemäß Satz 1 unendlich viele c_i von Null verschieden sind.*

Beweis. Sei zuerst α rational, etwa gleich a/b mit $a \in \mathbb{Z}$, $b \in \mathbb{N}$. Nach der gegenüber Satz 1 neu hinzugekommenen Voraussetzung über die g_j gibt es ein $n \in \mathbb{N}_0$, so daß $b|P_n$ gilt. Die CANTORsche Entwicklung für $\alpha = a/b$ gemäß 1 führt nach Multiplikation mit P_n zu

$$\{\frac{a}{b}\}P_n - \sum_{i=1}^{n} c_i \prod_{j=i+1}^{n} g_j = \sum_{i>n} c_i \prod_{j=n+1}^{i} g_j^{-1}.$$

Die Summe rechts ist nicht negativ; wegen $c_i \leq g_i - 1$ und 1(6) ist diese Summe aber kleiner als 1, da unendlich oft $c_i < g_i - 1$ gelten muß. Auf der linken Seite steht offensichtlich eine ganze Zahl, die somit gleich 0 sein muß. Daraus wiederum folgt $c_i = 0$ für alle $i > n$ wie behauptet. Die umgekehrte Richtung ist trivial. □

Dies Irrationalitätskriterium eignet sich besonders gut für Anwendungen, bei denen man z.B. $g_j = j + 1$ für alle $j \in \mathbb{N}$ zu nehmen wünscht, wo die gestellte Zusatzbedingung ersichtlich erfüllt ist. Sind $c_0 \in \mathbb{Z}$, $c_i \in \{0, \ldots, i\}$ für alle $i \in \mathbb{N}$, jedoch unendlich oft $c_i \neq i$, so ist $\sum_{i\geq 1} c_{i-1}/i!$ irrational, falls unendlich viele c_i von Null verschieden sind. Daraus ergibt sich mit der Wahl $c_i := 1$ für alle $i \geq 0$ die Irrationalität von $e - 1$, also die

Proposition A. *Die Zahl e ist irrational.*

Die Irrationalität von e wurde erstmals 1767 von LAMBERT mit Hilfe einer Kettenbruchentwicklung bewiesen; vgl. hierzu auch 3.11.

Proposition B. *Bezeichnet τ bzw. φ die in 1.1.3 bzw. 1.4.11 eingeführte Teileranzahlfunktion bzw. EULERsche Phi-Funktion, so sind $\sum_{i\geq 1}\tau(i)i!^{-1}$ und $\sum_{i\geq 1}\varphi(i)i!^{-1}$ irrational.*

Beweis. Für $i \geq 3$ ist offenbar $1 \leq \tau(i) \leq i-1$; denn dann ist $i-1$ eine natürliche Zahl größer als 1, die i nicht teilen kann. Nun setzt man für die erste Reihe $c_{i-1} := \tau(i)$ für $i \geq 3$ und etwa $c_0 := 0$, $c_1 := 0$; für jede Primzahl $i > 3$ ist $c_{i-1} = 2 < i-1$.

Zur Behandlung der zweiten Reihe beachtet man $1 \leq \varphi(i) \leq i-1$ für $i \geq 2$ und setzt $c_{i-1} := \varphi(i)$ für diese i und $c_0 := \varphi(1) = 1$. Um $c_{i-1} < i-1$ für unendlich viele i zu erhalten, bemerkt man, daß nach Korollar 1.4.11(iii) in der Ungleichung $\varphi(i) \leq i-1$ $(i \geq 2)$ Gleichheit genau dann eintritt, wenn i Primzahl ist. □

Wegen $\sigma(i) \geq i+1$ für alle $i \geq 2$ kann auf dem hier gezeigten Wege die Irrationalität der Reihe $\sum_{i\geq 1}\sigma(i)i!^{-1}$ nicht erhalten werden, wobei σ die in 1.1.7 eingeführte Teilersummenfunktion bedeutet. Allerdings ist die Irrationalität dieser Reihe auf etwas anderem Wege bewiesen worden. Auch die Irrationalität von $e^2 = \sum_{i\geq 0} 2^i i!^{-1}$ ist hier nicht zu haben.

3. Verwandte Irrationalitätskriterien. Aus der Fülle der zum Hauptergebnis aus 2 verwandten Reihenkriterien für Irrationalität sei noch eines herausgegriffen:

Satz. *Sei $g_1 \leq g_2 \leq \ldots$ und $(c_i)_{i=1,2,\ldots}$ eine beschränkte Folge ganzer Zahlen, von denen unendlich viele nicht Null sein sollen; es werde gesetzt*

$$\alpha := \sum_{i=1}^{\infty} c_i P_i^{-1}.$$

Wenn $(g_i)_{i=1,2,\ldots}$ unbeschränkt ist, ist α irrational.

Bemerkung. Es kann α auch dann irrational sein, wenn (g_i) beschränkt ist: Man braucht ja nur alle g_i gleich einem festen g zu nehmen und dann eine nichtperiodische Folge (c_i) mit allen $c_i \in S_g$.

Beweis. Sei $|c_i| \leq c$ für alle $i \geq 1$ und es werde angenommen, α sei rational, etwa gleich a/b. Dann gilt für alle $n \in \mathbb{N}_0$

$$
\begin{aligned}
bP_n\left|\frac{a}{b} - \sum_{i=1}^{n} c_i P_i^{-1}\right| &= b\left|\sum_{i>n} c_i \prod_{j=n+1}^{i} g_j^{-1}\right| \\
&\leq \frac{bc}{g_{n+1}} \sum_{i>n} 2^{1+n-i} = \frac{2bc}{g_{n+1}} .
\end{aligned} \tag{1}
$$

Wird nun $N \in \mathbb{N}$ so fixiert, daß $g_{N+1} > 2bc$ ist, so gilt $g_{n+1} > 2bc$ für alle $n \geq N$, da (g_i) nicht fällt. Da die linke Seite in (1) ganzzahlig ist, muß also für alle $n \geq N$

$$\frac{a}{b} = \sum_{i=1}^{n} c_i P_i^{-1}$$

zutreffen, also $c_i = 0$ für alle $i > N$, was einer Voraussetzung über (c_i) widerspricht. □

Sind unter den zunächst genannten Voraussetzungen über die g_i alle c_i gleich, so erhält man erneut ein notwendiges und hinreichendes Irrationalitätskriterium aus obigem Satz.

Korollar. *Gilt $g_1 \leq g_2 \leq \ldots$, so ist die Beschränktheit der Folge (g_i) notwendig und hinreichend für die Rationalität von*

$$\alpha := \sum_{i=1}^{\infty} P_i^{-1}.$$

Beweis. Die eine Richtung folgt durch Spezialisierung aus obigem Satz. Ist umgekehrt (g_i) beschränkt, so gibt es ein $n \in \mathbb{N}$, so daß $2 \leq g_1 \leq \ldots \leq g_n = g_{n+1} = \ldots =: g$ gilt. Dann ist

$$\left(\alpha - \sum_{i=1}^{n-1} P_i^{-1}\right) P_{n-1} = \sum_{i=n}^{\infty} g^{n-i-1} = (g-1)^{-1},$$

also α rational. □

4. Anwendungen. Wählt man wieder $g_i := i + 1$ für alle $i \in \mathbb{N}$, so erhält man aus dem vorstehenden Korollar *erneut die Irrationalität von e*. Im wesentlichen auf diesem Wege hat J.B. FOURIER 1815 diese arithmetische Aussage bewiesen.

Die nächste Anwendung des Korollars bezieht sich auf die in 2.1.2 eingeführten FERMAT–Zahlen $F_n := 2^{2^n} + 1$, $n = 0, 1, \ldots$, sowie auf die dort ebenfalls betrachteten Zahlen $G_n := F_n - 2$, die der unten benötigten Beziehung

2.1.2(1) $$G_k \prod_{j=k}^{i-1} F_j = G_i \qquad \text{für } 0 \leq k \leq i$$

genügen.

Proposition A. *Beide Reihen* $\sum_{i \geq 0} F_i^{-1}$, $\sum_{i \geq 0} G_i^{-1}$ *sind irrational.*

Beweis. Wegen $G_0 = 1$ und 2.1.2(1) ist

$$G := \sum_{i=0}^{\infty} G_i^{-1} = 1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} F_{j-1}^{-1};$$

auf die Reihe rechts wendet man nun das Korollar mit $g_i := F_{i-1}$ für $i \in \mathbb{N}$ an und erhält direkt $G \notin \mathbb{Q}$. Aus 2.1.2(1) folgt auch $G_i = F_{i-1}G_{i-1}$ und somit $2G_i^{-1} = G_{i-1}^{-1} - F_{i-1}^{-1}$ für alle $i \in \mathbb{N}$, also

$$F := \sum_{i=0}^{\infty} F_i^{-1} = G - 2(G - 1) = 2 - G,$$

woraus die Irrationalität von F folgt. □

Übrigens hat MAHLER 1929/30 eine analytische Methode entwickelt, deren Hauptresultate die *Transzendenz von F und G* bei weitem umfassen.

Für die Gewinnung des letzten Ergebnisses dieses Paragraphen ist es bequem, sich direkt auf Satz 3 zu stützen.

Während die bloße Aussage der Irrationalität von e äquivalent ist damit, daß es kein $P \in \mathbb{Z}[X]$, $P \neq 0$, $\partial(P) \leq 1$ gibt mit $P(e) = 0$, geht Proposition B noch einen Schritt weiter:

Proposition B. *Es gibt kein* $P \in \mathbb{Z}[X]$, $P \neq 0$, $\partial(P) \leq 2$ *mit* $P(e) = 0$, *d.h.* e *ist weder rational noch eine quadratische Irrationalität.*

Im Rahmen der Theorie der regelmäßigen Kettenbrüche wird hierfür in 3.11 ein weiterer Beweis gegeben; in Kapitel 6 wird dann sogar die Transzendenz von e bewiesen.

Beweis. Angenommen, es gäbe nicht sämtlich verschwindende $a_0, a_1, a_2 \in \mathbb{Z}$, so daß $a_0 + a_1 e + a_2 e^2 = 0$ oder gleichbedeutend

$$a_0 e^{-1} + a_2 e = -a_1 \tag{1}$$

gilt; sicher ist dann $(a_0, a_2) \neq (0,0)$. Gleichung (1) besagt gerade

$$\sum_{i=2}^{\infty} \frac{a_0(-1)^i + a_2}{i!} = -a_1 - 2a_2. \tag{2}$$

Nimmt man $g_i := i+1$, $c_i := a_0(-1)^{i+1} + a_2$ für $i = 1, 2, \ldots$, beachtet $(c_i, c_{i+1}) \neq (0,0)$ für dieselben i wegen $(a_0, a_2) \neq (0,0)$, so ist die linke Seite von (2) nach Satz 3 irrational und so erweist sich die anfangs gemachte Annahme als falsch. □

§ 3. Die regelmäßige Kettenbruchentwicklung

1. Der Kettenbruchalgorithmus. In 1.2.10 wurde im Anschluß an den euklidischen Algorithmus die Entwicklung einer rationalen Zahl r_0/r_1 mit $r_0, r_1 \in \mathbb{Z}$, $r_1 > 0$ in einen *endlichen regelmäßigen Kettenbruch*

$$[a_0; a_1, \ldots, a_j] \tag{1}$$

besprochen. Dabei genügten die *Elemente* (oft auch *Teilnenner*) a_i dieses Kettenbruchs stets den Bedingungen $a_0 \in \mathbb{Z}$; $a_1, \ldots, a_j \in \mathbb{N}$; $a_j \geq 2$. Wegen $a_j = (a_j - 1) + \frac{1}{1}$ ist aber klar, daß man (1) genausogut als

$$[a_0; a_1, \ldots, a_{j-1}, a_j - 1, 1] \tag{2}$$

schreiben kann, wobei die Eigenschaft erhalten bleibt, daß alle nach dem Strichpunkt aufgeführten Elemente natürliche Zahlen sind. Auf diese Möglichkeit, eine rationale Zahl auf (mindestens) zwei Weisen in einen endlichen Kettenbruch zu entwickeln, wird in 3 zurückzukommen sein.

Zunächst stellt man fest, daß der Algorithmus 1.2.9(1) mit der Bezeichnung $\alpha_i := r_i/r_{i+1}$ $(i = 0, \ldots, j)$ äquivalent ist zu

$$\alpha_0 = a_0 + \alpha_1^{-1}, \quad \alpha_1 = a_1 + \alpha_2^{-1}, \ldots, \alpha_{j-1} = a_{j-1} + \alpha_j^{-1}, \quad \alpha_j = a_j$$

mit $\alpha_1, \ldots, \alpha_j > 1$, vgl. auch 1.2.9(3). Dabei gilt, wie bereits seinerzeit bemerkt, in der jetzigen Terminologie

$$a_i = [\alpha_i] \text{ für } i = 0, \ldots, j \quad \text{bzw.} \quad \alpha_{i+1} = \{\alpha_i\}^{-1} \text{ für } i = 0, \ldots, j-1.$$

In dieser Form ist der Algorithmus auch zur Anwendung auf irrationale $\alpha \in \mathbb{R}$ geeignet: Man setzt $\alpha_0 := \alpha$, $a_0 := [\alpha_0]$, hat $0 < \{\alpha_0\} < 1$ wegen $\alpha_0 \notin \mathbb{Q}$ und setzt damit weiter $\alpha_1 := \{\alpha_0\}^{-1}$; es ist $a_0 \in \mathbb{Z}$, $\alpha_0 = a_0 + \alpha_1^{-1}$, $\alpha_1 > 1$, $\alpha_1 \notin \mathbb{Q}$. Nun setzt man voraus, es sei $i \geq 1$ und man habe bereits

$$
\begin{aligned}
&\alpha_j = a_j + \alpha_{j+1}^{-1} \text{ für } j = 0, \ldots, i-1 \\
&a_0 \in \mathbb{Z}, \quad a_1, \ldots, a_{i-1} \in \mathbb{N} \qquad \text{und} \\
&\alpha_j > 1, \quad \alpha_j \notin \mathbb{Q} \qquad \text{für } j = 1, \ldots, i
\end{aligned} \tag{3}
$$

gewonnen. Definiert man dann $a_i := [\alpha_i]$, so ist $\{\alpha_i\}$ aus $]0,1[$ und irrational, also ist $\alpha_{i+1} := \{\alpha_i\}^{-1}$ größer als 1 und irrational und genügt der Gleichung $\alpha_i = a_i + \alpha_{i+1}^{-1}$. Induktiv hat man somit (3) für alle $i \in \mathbb{N}$ bestätigt und hat überdies

$$\alpha = [a_0; a_1, \ldots, a_{i-1}, \alpha_i] \qquad \text{für } i = 0, 1, \ldots . \tag{4}$$

Wegen der obigen Festsetzung $\alpha_0 := \alpha$ und der Konvention 1.2.10(1) ist (4) für $i = 0$ richtig. Ist (4) für ein $i \geq 0$ in Ordnung, so trägt man in (4) für α_i aus (3) $a_i + \alpha_{i+1}^{-1}$ ein; mittels 1.2.10(2) erhält man gerade (4) für $i+1$ anstelle von i.

Der beschriebene Algorithmus ordnet jedem vorgegebenen $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ eine wohlbestimmte unendliche Folge $a_0, a_1, a_2, \ldots$ ganzer Zahlen zu, wobei die $a_1, a_2, \ldots$ sämtliche positiv sind, oder genauer gesagt den *unendlichen regelmäßigen Kettenbruch*

$$[a_0; a_1, \ldots, a_i, \ldots].$$

Es sei hier sogleich angemerkt, daß der Zusatz "regelmäßig" im weiteren stets unterdrückt wird, weil andersartige Kettenbrüche in diesem Buch nicht diskutiert werden.

2. Konvergenz unendlicher Kettenbrüche. Zunächst muß die Konvergenz solcher Kettenbrüche untersucht werden. Sei also ganz allgemein eine endliche oder unendliche Folge $a_0, a_1, \ldots, a_k(,\ldots)$ ganzer Zahlen mit $a_0 \in \mathbb{Z}$, $a_1, \ldots \in \mathbb{N}$ vorgegeben und es werde verabredet, daß a_k ihr letztes Glied sein soll, wenn sie endlich ist. Zu ihr definiert man zwei neue Folgen (p_i), (q_i), im endlichen Fall nur für $i \leq k$, gemäß den Rekursionsformeln

$$
\begin{aligned}
p_{-2} &:= 0, \quad p_{-1} := 1, \quad p_i := a_i p_{i-1} + p_{i-2}, \\
q_{-2} &:= 1, \quad q_{-1} := 0, \quad q_i := a_i q_{i-1} + q_{i-2}
\end{aligned} \tag{1}
$$

für $i = 0, 1, \ldots, k(,\ldots)$. Daraus sieht man $p_i \in \mathbb{Z}$, $q_i \in \mathbb{N}$ für alle in Frage kommenden $i \geq 0$. Genauer ist $q_0 = 1$, $q_i \geq q_{i-1} + q_{i-2}$ für $i \geq 1$, was induktiv $q_i \geq i$ für $i \geq 1$ liefert. Insbesondere ist (q_i) unbeschränkt, wenn die vorgegebene Folge (a_i) unendlich ist.

Mit den so eingeführten p_i, q_i hat man das

Lemma A. *Für alle in Frage kommenden $i \geq 0$ ist mit einer Unbestimmten X*

$$[a_0; a_1, \ldots, a_{i-1}, X] = \frac{p_{i-1}X + p_{i-2}}{q_{i-1}X + q_{i-2}}\,.$$

Beweis. Für $i = 0$ hat man (1) und die Konvention 1.2.10(1) zu beachten. Ist die zu beweisende Gleichung für $i \geq 0$ richtig, so führt 1.2.10(2) und (1) zu

$$\begin{aligned}[a_0; a_1, \ldots, a_i, X] &= [a_0; a_1, \ldots, a_{i-1}, a_i + \frac{1}{X}] \\ &= \frac{p_{i-1}(a_i + \frac{1}{X}) + p_{i-2}}{q_{i-1}(a_i + \frac{1}{X}) + q_{i-2}} = \frac{p_iX + p_{i-1}}{q_iX + q_{i-1}}\,.\end{aligned}$$ □

Indem man nun a_i für X einsetzt und erneut (1) beachtet, folgt aus Lemma A

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

für $i \geq 0$ (und gegebenenfalls $i \leq k$). A_i heißt der *i–te Näherungsbruch* des (eventuell endlichen) Kettenbruchs $[a_0; a_1, \ldots, a_k, \ldots]$; p_i bzw. q_i heißt der *i–te Näherungszähler* bzw. *–nenner* dieses Kettenbruchs.

Als weiteres technisches Hilfsmittel für die Konvergenzuntersuchung unendlicher Kettenbrüche ist nützlich das folgende

Lemma B. *Für $i \geq -1$ ist*

(2) $$p_{i-1}q_i - p_iq_{i-1} = (-1)^i;$$

insbesondere gilt $(p_i, q_i) = 1$ für $i \geq -2$ sowie

(3) $$A_{i-1} - A_i = \frac{(-1)^i}{q_{i-1}q_i} \qquad \text{für } i \geq 1.$$

Für $i \geq 0$ ist

(4) $$p_{i-2}q_i - p_iq_{i-2} = (-1)^{i-1}\,a_i$$

und daher

(5) $$A_i - A_{i-2} = \frac{(-1)^i a_i}{q_{i-2}q_i} \qquad \text{für } i \geq 2,$$

jeweils gegebenenfalls nur für $i \leq k$.

Beweis. (2) ist nach (1) für $i = -1$ richtig und aus der Richtigkeit für ein $i \geq -1$ folgt nach (1) und der Induktionsannahme $p_iq_{i+1} - p_{i+1}q_i = p_i(a_{i+1}q_i + q_{i-1}) - (a_{i+1}p_i + p_{i-1})q_i = p_iq_{i-1} - p_{i-1}q_i = (-1)^{i+1}$. (2) und (3) sind äquivalent. Völlig analog beweist man (4) und damit (5). □

Proposition. *Die (gegebenenfalls endlichen) Folgen $A_0, A_2, \ldots$ bzw. $A_1, A_3, \ldots$ sind streng monoton wachsend bzw. fallend und jedes Glied der ersten Folge ist kleiner als jedes Glied der zweiten Folge.*

Beweis. Bei geradem $i \geq 2$ folgt nämlich $A_i > A_{i-2}$ aus (5), bei ungeradem $i \geq 3$ jedoch $A_i < A_{i-2}$. Werden mit $s, t \in \mathbb{N}_0$ nun A_{2s}, A_{2t+1} miteinander verglichen, so gilt bei $s \leq t$ die Ungleichung $A_{2s} \leq A_{2t} < A_{2t+1}$ nach (3) für $i = 2t+1$; bei $s > t$ ist $A_{2s} < A_{2s+1} < A_{2t+1}$ nach dem bereits vorher Festgestellten. □

Nun stehen alle Hilfsmittel bereit für den

Konvergenzsatz. *Ist ein unendlicher Kettenbruch $[a_0; a_1, a_2, \ldots]$ mit $a_0 \in \mathbb{Z}$ und allen $a_1, a_2, \ldots \in \mathbb{N}$ vorgelegt, so konvergiert die unendliche Folge $(A_i)_{i=0,1,\ldots}$ der Näherungsbrüche dieses Kettenbruchs; ist $\alpha \in \mathbb{R}$ der Grenzwert, so hat man überdies $A_0 < A_2 < \ldots < \alpha < \ldots < A_3 < A_1$ und $\alpha \notin \mathbb{Q}$.*

Bemerkung. Unter dem Wert von $[a_0; a_1, a_2, \ldots]$ versteht man den Grenzwert α und schreibt direkt $[a_0; a_1, \ldots] = \alpha$.

Beweis. Nach vorstehender Proposition ist die Folge $A_0, A_2, \ldots$ monoton wachsend und (etwa durch A_1) nach oben beschränkt, also konvergent, etwa gegen α'. Analog ist $A_1, A_3, \ldots$ gegen ein reelles $\alpha'' \geq \alpha'$ konvergent. Weiter ist

$$0 \leq \alpha'' - \alpha' < A_{2s-1} - A_{2s} = q_{2s-1}^{-1} q_{2s}^{-1} \leq \frac{1}{(2s-1)2s}$$

für alle $s \geq 1$ wegen (3) und $q_i \geq i$ für $i \geq 1$. Der letzten Ungleichung entnimmt man $\alpha'' = \alpha' =: \alpha$.

Wäre $\alpha \in \mathbb{Q}$, etwa $\alpha = \frac{a}{b}$, so würde $\alpha \neq A_i$ für alle $i \geq 0$ gelten, also nach (3)

$$\frac{1}{bq_i} \leq |\alpha - A_i| < |A_{i+1} - A_i| = \frac{1}{q_i q_{i+1}}.$$

Dies würde aber $q_{i+1} < b$ für alle $i \geq 0$ implizieren, während doch die Folge (q_i) unbeschränkt ist. □

3. Eindeutigkeit. Irrationalität. Der folgende Satz enthält offenbar wieder ein notwendiges und hinreichendes Kriterium für die Irrationalität einer reellen Zahl.

Satz. *Jedes $\alpha \in \mathbb{R}$ läßt sich in eindeutiger Weise in einen Kettenbruch entwickeln und dieser ist endlich genau dann, wenn α rational ist. Dabei hat man im endlichen Fall zu verbieten, daß das letzte Element gleich Eins ist, falls es einen positiven Index hat.*

Beweis. Daran, daß sich eine rationale Zahl auf mindestens eine Weise in einen (endlich ausfallenden) Kettenbruch entwickeln läßt, wurde in 1 erinnert. Für $\alpha \notin \mathbb{Q}$ gilt nach 1(3) und 1(4) die Gleichung $\alpha = [a_0; a_1, \ldots, a_i, \alpha_{i+1}]$ mit irrationalen $\alpha_{i+1} > 1$ für $i \geq 0$. Nach Lemma 2A ist $\alpha = (p_i\alpha_{i+1} + p_{i-1})/(q_i\alpha_{i+1} + q_{i-1})$ für $i \geq 0$, also für dieselben i

$$(1) \qquad \alpha - \frac{p_i}{q_i} = \frac{q_i(p_i\alpha_{i+1} + p_{i-1}) - p_i(q_i\alpha_{i+1} + q_{i-1})}{q_i(q_i\alpha_{i+1} + q_{i-1})} = \frac{(-1)^i}{q_i(q_i\alpha_{i+1} + q_{i-1})},$$

woraus insbesondere $|\alpha - A_i| < q_i^{-2}$ folgt, also $\alpha = \lim_{i\to\infty} A_i$ und dieser letzte Grenzwert ist nach dem Konvergenzsatz 2 gleich α. Der in der zweiten Hälfte von 1 beschriebene Algorithmus liefert somit tatsächlich *eine* Entwicklung des irrationalen α in einen unendlichen Kettenbruch.

Zur Eindeutigkeit der Kettenbruchentwicklung nimmt man nun an, $\alpha \in \mathbb{R}$ habe zwei Entwicklungen,

$$(2) \qquad [a_0; a_1, \ldots] = \alpha = [a_0'; a_1', \ldots].$$

Diese sind nach dem Bisherigen entweder beide endlich oder beide unendlich. Ist nun $i \geq 0$ und bereits $a_j' = a_j$ für $j = 0, \ldots, i-1$ eingesehen (vorausgesetzt wird dabei im endlichen Fall selbstverständlich, daß beide Kettenbrüche in (2) mindestens $i+1$ Elemente haben), so gilt nach Lemma A und (2) mit den Bezeichnungen $\alpha_i := [a_i; a_{i+1}, \ldots]$, $\alpha_i' := [a_i'; a_{i+1}', \ldots]$:

$$\frac{p_{i-1}\alpha_i + p_{i-2}}{q_{i-1}\alpha_i + q_{i-2}} = [a_0; \ldots, a_{i-1}, \alpha_i] = \alpha = [a_0; \ldots, a_{i-1}, \alpha_i'] = \frac{p_{i-1}\alpha_i' + p_{i-2}}{q_{i-1}\alpha_i' + q_{i-2}},$$

woraus $\alpha_i' = \alpha_i$ wegen 2(2) folgt. Sind in (2) beide Kettenbrüche unendlich, so bedeutet dies $a_i + 1/\alpha_{i+1} = a_i' + 1/\alpha_{i+1}'$, also $|a_i' - a_i| = |1/\alpha_{i+1} - 1/\alpha_{i+1}'| < 1$ wegen $\alpha_{i+1}, \alpha_{i+1}' > 1$ und somit $a_i' = a_i$. Dieselbe Schlußfolgerung kann gezogen werden, wenn in (2) zwei endliche Kettenbrüche mit jeweils mindestens $i+2$ Gliedern stehen; dabei ist ganz wesentlich zu beachten, daß keiner der beiden Kettenbrüche nach der getroffenen Konvention mit einem Element 1 enden darf. Haben beide endliche Kettenbrüche die Länge $i+1$, so ist $a_i' = \alpha_i' = \alpha_i = a_i$; hat einer, etwa der rechts in (2), die Länge $i+1$, der andere aber mindestens die Länge $i+2$, so ist $a_i' - a_i = \alpha_{i+1}^{-1} \in]0,1[$ wegen $\alpha_{i+1} > 1$. Dies widerspricht der Ganzzahligkeit von a_i, a_i' und auch hier hat man wieder ausgenutzt, daß ein endlicher Kettenbruch nicht auf 1 enden darf. $\square$

4. Periodische Kettenbrüche. Satz von Euler. Da z.B. die Zahlen $\sqrt{2}$ und $\frac{1}{2}(\sqrt{5}+1)$ nach 1.1.9 irrational sind, haben sie nach dem letzten Satz unendliche Kettenbruchentwicklungen, welche zunächst bestimmt werden sollen.

Ausgehend von der Gleichung $(\sqrt{2}-1)(\sqrt{2}+1)=1$ findet man $\sqrt{2}=1+\frac{1}{1+\sqrt{2}}$; trägt man hier rechts im Nenner für $\sqrt{2}$ erneut die gesamte rechte Seite ein, so folgt $\sqrt{2}=[1;2,1+\sqrt{2}]$ und induktiv kommt man zu $\sqrt{2}=[1;2,2,2,\ldots]$.

Analog beachtet man $\frac{1}{2}(\sqrt{5}-1)=(\frac{1}{2}(\sqrt{5}+1))^{-1}$, um aus $\frac{1}{2}(\sqrt{5}+1)=1+\frac{1}{2}(\sqrt{5}-1)$ zu schließen $\frac{1}{2}(\sqrt{5}+1)=[1;\frac{1}{2}(\sqrt{5}+1)]$, also $\frac{1}{2}(\sqrt{5}+1)=[1;1,1,1,\ldots]$. Offenbar ist $\frac{1}{2}(\sqrt{5}+1)$ (bis auf einen ganzzahligen Summanden) diejenige reelle Irrationalzahl mit den kleinstmöglichen Elementen.

Für die Zwecke von 6 wird nun in Verallgemeinerung des obigen Beispiels $\sqrt{2}$ behauptet:

Proposition. *Für $D\in\mathbb{N}$ gilt $(1+D^2)^{1/2}=[D;2D,2D,\ldots]$.*

Beweis. Setzt man $\alpha:=[2D;2D,2D,\ldots]$, so ist $\alpha=2D+\alpha^{-1}$ und die positive Irrationalzahl α genügt der quadratischen Gleichung $\alpha^2-2D\alpha-1=0$. Deren einzige positive Wurzel ist $D+(1+D^2)^{1/2}$, woraus bereits die Behauptung folgt. □

Selbstverständlich wird man einen unendlichen Kettenbruch

$$[a_0;a_1,a_2,\ldots] \tag{1}$$

periodisch nennen, wenn die Folge $a_1,a_2,\ldots$ natürlicher Zahlen periodisch ist. Die Begriffe *Vorperiodenlänge* (ℓ) und *Periodenlänge* (hier h statt p), *Vorperiode*, *Periode* werden aus 1.5 übernommen. Ist insbesondere $a_{\ell+1},\ldots,a_{\ell+h}$ die Periode und $a_1,\ldots,a_\ell$ die (bei $\ell=0$ fehlende) Vorperiode von (1), so schreibt man (1) als $[a_0;a_1,\ldots,a_\ell,\overline{a_{\ell+1},\ldots,a_{\ell+h}}]$ in Anlehnung an die entsprechende Bezeichnungsweise in 1.5 bei g–adischen Entwicklungen. Die bisher betrachteten Zahlen $\sqrt{2}$, $\frac{1}{2}(\sqrt{5}+1)$, $(1+D^2)^{1/2}$ haben sämtliche periodische Kettenbruchentwicklungen, wie gesehen nämlich $[1;\overline{2}]$, $[1;\overline{1}]$, $[D;\overline{2D}]$. Durch den Satz von Lagrange in 5 wird diese Beobachtung in einen allgemeineren Rahmen gestellt; jener Satz ist die Umkehrung des folgenden (einfacheren) Satzes von Euler (Opera Omnia Ser. 1, XIV, 187–215) aus dem Jahre 1737:

Satz von Euler. *Ein unendlicher periodischer Kettenbruch definiert eine reell–quadratische Irrationalzahl.*

Beweis. Sei (1) periodisch und α der Wert dieses Kettenbruchs. Mit $\alpha_i := [a_i; a_{i+1}, \ldots]$ für $i \geq 0$ (also $\alpha_0 := \alpha$) hat man mittels Lemma 2A

$$\alpha = [a_0; a_1, \ldots, a_{i-1}, \alpha_i] = \frac{p_{i-1}\alpha_i + p_{i-2}}{q_{i-1}\alpha_i + q_{i-2}} \qquad \text{für } i \geq 0. \tag{2}$$

Für dieselben i erhält man hieraus

$$\alpha_i = \frac{p_{i-2} - q_{i-2}\alpha}{q_{i-1}\alpha - p_{i-1}}; \tag{3}$$

der Nenner verschwindet dabei nicht. $a_{i+h} = a_i$ für alle $i > \ell$ impliziert natürlich $\alpha_{i+h} = \alpha_i$ für dieselben i, woraus sich mit (3) ergibt

$$\frac{p_{i+h-2} - \alpha q_{i+h-2}}{\alpha q_{i+h-1} - p_{i+h-1}} = \frac{p_{i-2} - \alpha q_{i-2}}{\alpha q_{i-1} - p_{i-1}} \qquad \text{für } i > \ell. \tag{4}$$

Setzt man

$$\begin{aligned} R_i &:= q_{i+h-1}q_{i-2} - q_{i+h-2}q_{i-1}, \\ S_i &:= p_{i+h-2}q_{i-1} + p_{i-1}q_{i+h-2} - p_{i-2}q_{i+h-1} - p_{i+h-1}q_{i-2}, \\ T_i &:= p_{i+h-1}p_{i-2} - p_{i-1}p_{i+h-2}, \end{aligned}$$

so genügt α wegen (4) der quadratischen Gleichung $R_i\alpha^2 + S_i\alpha + T_i = 0$ für $i > \ell$. Dem sogleich nachgeschobenen Lemma entnimmt man $R_i \neq 0$ für $i > \ell$ und somit ist α algebraisch von einem Grad höchstens 2; wegen der Unendlichkeit seines Kettenbruchs ist $\alpha \notin \mathbb{Q}$, also sein Grad tatsächlich genau 2. □

Lemma. *Für verschiedene $j, k \in \mathbb{N}_0$ gilt $q_j q_{k-1} \neq q_{j-1} q_k$.*

Wendet man das Lemma an mit $j := i - 1$, $k := i + h - 1$, so sind diese beiden ≥ 0 für $i > 0$ und außerdem wegen $h \geq 1$ voneinander verschieden; also ist $R_i \neq 0$ für $i > 0$.

Beweis des Lemmas. Sei o.B.d.A. $k > j \geq 0$. Für $j = 0$ ist $q_{j-1}q_k = 0$, aber $q_j q_{k-1} \neq 0$. Sei also bereits $k > j \geq 1$; dann sind alle im Lemma vorkommenden q's nicht Null. Wegen $(q_{i-1}, q_i) = 1$, vgl. 2(2), folgt aus der Annahme $q_j q_{k-1} = q_{j-1} q_k$, daß $q_j | q_k$ gelten muß, also $q_k = g q_j$ und damit auch $q_{k-1} = g q_{j-1}$ mit einem $g \in \mathbb{N}$. Wegen $(q_{k-1}, q_k) = g(q_{j-1}, q_j)$ muß $g = 1$ sein, also $q_{k-1} = q_{j-1}$, $q_k = q_j$. Andererseits ist aber wegen $k \geq 2$: $q_k \geq q_{k-1} + q_{k-2} > q_{k-1} \geq q_j$. □

5. Der Satz von Lagrange. Die Umkehrung des Eulerschen Satzes 4 geht auf Lagrange (Oeuvres II, 581–652) zurück; sie lautet

Satz von Lagrange. *Jede reell–quadratische Irrationalzahl hat einen unendlichen periodischen Kettenbruch.*

Beweis. Sei $\alpha \in \mathbb{R}$ algebraisch vom Grade 2 und sei 4(1) sein (nach Satz 3 unendlicher) Kettenbruch. Nach Voraussetzung gibt es ein $P := RX^2+SX+T \in \mathbb{Z}[X]$, $P \neq 0$, mit α als Wurzel. Wegen $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ muß $R \neq 0$ gelten und die Diskriminante $S^2 - 4RT$ von P muß positiv sein, jedoch keine Quadratzahl. Nach 4(2) ist

$$R(p_{i-1}\alpha_i + p_{i-2})^2 + S(p_{i-1}\alpha_i + p_{i-2})(q_{i-1}\alpha_i + q_{i-2}) + T(q_{i-1}\alpha_i + q_{i-2})^2 = 0$$

für alle $i \geq 0$ und mit den Bezeichnungen

$$\begin{aligned} R_i^* &:= Rp_{i-1}^2 + Sp_{i-1}q_{i-1} + Tq_{i-1}^2, \\ S_i^* &:= 2Rp_{i-1}p_{i-2} + S(p_{i-1}q_{i-2} + p_{i-2}q_{i-1}) + 2Tq_{i-1}q_{i-2}, \\ T_i^* &:= R_{i-1}^* \end{aligned} \tag{1}$$

bedeutet dies: α_i ist Wurzel von $P_i := R_i^*X^2+S_i^*X+T_i^* \in \mathbb{Z}[X]$ für $i = 0, 1, \ldots$. Wegen der leicht nachzuprüfenden Matrizengleichung

$$\begin{pmatrix} 2R_i^* & S_i^* \\ S_i^* & 2T_i^* \end{pmatrix} = \begin{pmatrix} p_{i-1} & q_{i-1} \\ p_{i-2} & q_{i-2} \end{pmatrix} \begin{pmatrix} 2R & S \\ S & 2T \end{pmatrix} \begin{pmatrix} p_{i-1} & p_{i-2} \\ q_{i-1} & q_{i-2} \end{pmatrix}$$

ist nach Übergang zu den entsprechenden Determinanten und bei Beachtung von 2(2)

$$S_i^{*2} - 4R_i^*T_i^* = S^2 - 4RT \qquad \text{für alle } i \geq 0. \tag{2}$$

Schätzt man in 3(1) für $i-1$ statt i ab, so erhält man

$$|q_{i-1}\alpha - p_{i-1}| = (q_{i-1}\alpha_i + q_{i-2})^{-1} < (q_{i-1}a_i + q_{i-2})^{-1} = q_i^{-1}$$

für $i \geq 1$, also mit einem $\vartheta_i \in \mathbb{R}$, $0 < |\vartheta_i| < 1$:

$$p_{i-1} = \alpha q_{i-1} + \vartheta_i q_i^{-1} \qquad \text{für } i \geq 0.$$

Trägt man dies in der rechten Seite der ersten Gleichung von (1) ein, so folgt

$$R_i^* = P(\alpha)q_{i-1}^2 + (2R\alpha + S)\vartheta_i q_{i-1}q_i^{-1} + R\vartheta_i^2 q_i^{-2}.$$

$P(\alpha) = 0$ führt unmittelbar zu $|R_i^*| \leq |2R\alpha + S| + |R|$ für alle $i \geq 0$, d.h. (R_i^*) und damit nach (1) auch (T_i^*) erweist sich als beschränkte Folge ganzer Zahlen. Wegen (2) trifft dasselbe nun für (S_i^*) zu, und so muß es ein $(R^*, S^*, T^*) \in \mathbb{Z}^3$ mit $S^{*2} - 4R^*T^* = S^2 - 4RT$ geben, so daß für unendlich viele Indizes i gilt: $R_i^* = R^*$, $S_i^* = S^*$, $T_i^* = T^*$. Da $S^2 - 4RT$ kein Quadrat ist, gilt $R^*T^* \neq 0$ und also hat das Polynom $P^* := R^*X^2 + S^*X + T^*$ zwei verschiedene reelle irrationale Wurzeln. Andererseits ist $P^*(\alpha_i) = 0$ unendlich oft und so existieren $j, k \in \mathbb{N}_0$ mit $j < k$ und $\alpha_k = \alpha_j$, d.h. $[a_k; a_{k+1}, \ldots] = [a_j; a_{j+1}, \ldots]$. Nach der Eindeutigkeitsaussage von Satz 3 ist dann mit $h := k - j \in \mathbb{N}$

$$a_{i+h} = a_i \qquad \text{für alle } i \geq j$$

und α hat somit eine periodische Kettenbruchentwicklung. □

Mit Satz 3 und den Sätzen von EULER und LAGRANGE scheint sich für *algebraische* $\alpha \in \mathbb{R}$ eine Serie von Charakterisierungen durch Eigenschaften ihrer Kettenbruchentwicklung anzubahnen: α ist vom Grade 1 genau dann, wenn sein Kettenbruch endlich ist; α ist vom Grade 2 genau dann, wenn sein Kettenbruch unendlich und periodisch ist. Hier jedoch endet bereits die Serie nach heutiger Kenntnis der Dinge: Man kennt z.B. keine Charakterisierung des Grades 3 durch Kettenbrucheigenschaften, ja man kennt für kein algebraisches $\alpha \in \mathbb{R}$ mit Grad ≥ 3 den Kettenbruch. Insbesondere weiß man nicht, ob es darunter welche gibt, für die die Folge der Kettenbruchelemente z.B. unbeschränkt ist.

6. Zur Minimallösung der Pellschen Gleichung. In 4.3.3 wurde gezeigt, daß die PELLsche Gleichung

$$X^2 - dY^2 = 1 \tag{1}$$

bei nicht quadratischem $d \in \mathbb{N}$ unendlich viele Lösungen hat. Der Beweis dort begann (vgl. 4.3.3(2)) mit einer Anwendung des zweiten Teils des DIRICHLETschen Approximationssatzes 4.3.2. Es sei hier zunächst bemerkt, daß jene *zweite Teilaussage des* DIRICHLET*schen Approximationssatzes genausogut aus der* in diesem Paragraphen entwickelten *Kettenbruchtheorie ableitbar ist.*

Sind nämlich p_i, q_i der i–te Näherungszähler bzw. –nenner des Kettenbruchs von α, so folgt aus 3(1) unmittelbar durch Abschätzung $|q_i\alpha - p_i| < 1/q_i$ für alle $i \geq 0$ und nach Lemma 2 sind p_i, q_i teilerfremd für jedes solche i.

Ferner wurde in 4.3.4 das Problem der vollständigen Lösung von (1) auf das Problem der Auffindung der Minimallösung von (1) zurückgespielt. Am Ende von 4.3.4 wurde dann ein Probierverfahren zur Auffindung dieser Minimallösung angegeben und ein systematischer Weg via Kettenbruchtheorie zur Erledigung dieser Aufgabe in Aussicht gestellt.

Dieser Weg soll hier zunächst für die speziellen $d \in \mathbb{N}$ der Form $d = 1 + D^2$, $D \in \mathbb{N}$ dargelegt werden, die dann automatisch keine Quadrate sind. Behauptet wird

Satz A. *Ist $d := 1 + D^2$, $D \in \mathbb{N}$, und ist p_i bzw. q_i der i–te Näherungszähler bzw. –nenner der Kettenbruchentwicklung von $\sqrt{d}$, so ist (p_1, q_1) die Minimallösung der Pellschen Gleichung (1). Überdies sind $(p_{2k-1}, q_{2k-1})_{k=1,2,\ldots}$ alle Lösungen von (1) in $\mathbb{N}^2$.*

Dieser Satz klärt auch das am Ende von 4.3.4 angegebene Beispiel der Primzahl $d = 98597$ auf: Es ist nämlich $d = 1 + 314^2$, also $\frac{p_1}{q_1} = [314; 628] = \frac{197193}{628}$ nach Proposition 4 und somit ist $(197193, 628)$ nach Satz A die Minimallösung der entsprechenden Gleichung (1) wie in 4.3.4 in Aussicht gestellt.

Beweis von Satz A. Wenn für diesen Beweis stets d für $1 + D^2$ steht, wird zunächst behauptet

$$p_i^2 - dq_i^2 = (-1)^{i+1} \qquad \text{für } i \geq -1, \tag{2}$$

$$p_{i-1}p_i - dq_{i-1}q_i = (-1)^i D \qquad \text{für } i \geq 0. \tag{3}$$

Für $i = -1, 0$ ist (2) und für $i = 0$ ist (3) richtig nach 2(1) und $p_0 = a_0$, $q_0 = 1$; man beachte, daß $a_0 = D$, $a_i = 2D$ für $i \geq 1$ wegen Proposition 4 gilt. Sei nun $i \geq 0$ und (2), (3) für dieses i schon als richtig erkannt. Dann ist unter Beachtung von (2), (3) für $i + 1$

$$\begin{aligned} p_{i+1}^2 - dq_{i+1}^2 &= (2Dp_i + p_{i-1})^2 - d(2Dq_i + q_{i-1})^2 \\ &= 4D^2(p_i^2 - dq_i^2) + 4D(p_{i-1}p_i - dq_{i-1}q_i) + (p_{i-1}^2 - dq_{i-1}^2) \\ &= (-1)^{i+1}(4D^2 - 4D^2 - 1) \\ &= (-1)^{i+2}, \end{aligned}$$

was (2) für $i+1$ beweist. Dabei hat man 2(1) benützt, wie auch in der nächsten Formelzeile, die zu (3) für $i + 1$ führt:

$$\begin{aligned} p_i p_{i+1} - dq_i q_{i+1} &= p_i(2Dp_i + p_{i-1}) - dq_i(2Dq_i + q_{i-1}) \\ &= 2D(p_i^2 - dq_i^2) + (p_i p_{i-1} - dq_i q_{i-1}) \\ &= (-1)^{i+1}(2D - D) \\ &= (-1)^{i+1}D. \end{aligned}$$

Aus (2) sieht man bereits, daß alle $(p_i, q_i) \in \mathbb{N}^2$ mit ungeradem $i \geq 1$ die Pellsche Gleichung (1) bei $d = 1 + D^2$ lösen. Ist (x_1, y_1) deren Minimallösung, so ist bei geeignetem $n \in \mathbb{N}$ nach Satz 4.3.4

$$p_1 = \sum_{\nu=0}^{[n/2]} \binom{n}{2\nu} x_1^{n-2\nu} y_1^{2\nu} (1 + D^2)^\nu; \tag{4}$$

wegen $x_1^2 \equiv 0, 1$, aber $2 + D^2 \equiv 2, 3 \pmod 4$ ist $y_1 \neq 1$, also $y_1 \geq 2$ und damit würde aus (4) und der Annahme $n \geq 2$ folgen

$$p_1 \geq x_1^n + \binom{n}{2} x_1^{n-2} y_1^2 (1 + D^2) > 4(1 + D^2),$$

was $p_1 = 2D^2 + 1$ widerspricht, vgl. 2(1). Damit ist $n = 1$, also $p_1 = x_1$ nach (4) und schließlich (p_1, q_1) die Minimallösung.

Ist weiter $k \geq 0$, so folgt aus 2(1), aus (2) für $i = 2k-1$, aus (3) und 2(2) für $i = 2k$ sowie schließlich aus $q_1 = 2D$

$$\begin{aligned}\frac{p_{2k+1}+\sqrt{d}q_{2k+1}}{p_{2k-1}+\sqrt{d}q_{2k-1}} &= 2D(p_{2k}+\sqrt{d}q_{2k})(p_{2k-1}-\sqrt{d}q_{2k-1})+1\\ &= 2D((p_{2k-1}p_{2k}-dq_{2k-1}q_{2k})\\ &\quad +\sqrt{d}(p_{2k-1}q_{2k}-p_{2k}q_{2k-1}))+1\\ &= 2D(D+\sqrt{d})+1\\ &= p_1+\sqrt{d}q_1.\end{aligned}$$

Induktiv folgt hieraus $p_{2k-1}+\sqrt{d}q_{2k-1} = (p_1+\sqrt{d}q_1)^k$ für $k \geq 1$ und nun erhält man aus 4.3.4 direkt $x_k = p_{2k-1}$, $y_k = q_{2k-1}$ für $k \geq 1$, wo (x_k, y_k) die in Satz 4.3.4 aufgeführten sämtlichen Lösungen von (1) in $\mathbb{N}^2$ sind, wobei jetzt k statt n geschrieben ist. □

Wie man bei beliebigem, nicht quadratischem $d \in \mathbb{N}$ vorgeht, entnimmt man den ausführlichen Darlegungen bei PERRON [19], §§ 24–26. Die Beweismethode in diesem allgemeinen Fall ist grundsätzlich ähnlich zu der für Satz A, wenngleich technisch deutlich aufwendiger; das Ergebnis sei hier ohne Beweis mitgeteilt als

Satz B. *Ist $d \in \mathbb{N}$ kein Quadrat und bezeichnen p_i, q_i den i-ten Näherungszähler, –nenner der Kettenbruchentwicklung von $\sqrt{d}$, deren Periodenlänge h sei, so gilt mit $m := \frac{1}{2}(3-(-1)^h)$: Das Paar (p_{mh-1}, q_{mh-1}) ist die Minimallösung von (1) und $(p_{mhk-1}, q_{mhk-1})_{k=1,2,\ldots}$ sind sämtliche Lösungen von (1) in $\mathbb{N}^2$.*

Offenbar ist m gleich 1 bzw. 2, wenn h gerade bzw. ungerade ist. In Satz A ist insbesondere $h = 1$.

Teilweise wurde Satz B um 1766 herum von LAGRANGE gefunden; einen vollständigen Beweis hat aber erst LEGENDRE einige Jahrzehnte später liefern können. Satz B macht besonders schön deutlich, was die Kettenbruchtheorie bei der Untersuchung gewisser diophantischer Gleichungen zu leisten imstande ist.

7. Annäherung reeller Zahlen durch rationale. Aus 3(1) folgt für die Näherungsbrüche $\frac{p_0}{q_0}, \ldots, \frac{p_k}{q_k}, \ldots$ von α (wobei diese Folge für $\alpha \in \mathbb{Q}$ mit $\frac{p_k}{q_k}$ enden möge)

(1) $$q_i^{-1}(q_{i+1}+q_i)^{-1} < |\alpha - \frac{p_i}{q_i}| = q_i^{-1}(\alpha_{i+1}q_i+q_{i-1})^{-1} \leq q_i^{-1}q_{i+1}^{-1},$$

jedenfalls solange $i < k$ im Falle rationaler α gilt und rechts tritt Gleichheit ein genau für $\alpha \in \mathbb{Q}$, $i = k - 1$. Ungleichung (1) bietet *eine* Lösung der in der Praxis oft wichtigen Aufgabe an, eine vorgegebene reelle Zahl α durch rationale Zahlen anzunähern und den dabei begangenen Approximationsfehler sehr genau zu kontrollieren: Dabei approximiert man *hier* das vorgegebene α speziell durch die Näherungsbrüche der Kettenbruchentwicklung von α.

Die Frage drängt sich nun natürlich auf, ob man die oben angesprochene Aufgabe nicht besser durch andere rationale Zahlen als die Näherungsbrüche der zu approximierenden Zahl lösen sollte. In diesem und dem folgenden Abschnitt soll sie in präzisere Form gefaßt und dann gelöst werden. Dabei werden die Resultate in 8 glatter, wenn man die Annäherung von α durch rationale $\frac{a}{b}$ nicht wie in (1) durch $|\alpha - \frac{a}{b}|$, sondern durch $|b\alpha - a|$ mißt, was sogleich geschehen wird.

Der nun zu formulierende Hilfssatz enthält die wichtigsten Mittel, die für den Vergleich der Approximation reeller Zahlen durch beliebige rationale Zahlen bzw. durch Näherungsbrüche benötigt werden.

Lemma.

(i) *Bei $\alpha \in \mathbb{Q}$ hat man für alle $c \in \mathbb{Z}$, $d \in \mathbb{N}$ die Ungleichung $|q_k\alpha - p_k| \leq |d\alpha - c|$ mit Gleichheit genau für $\frac{c}{d} = \frac{p_k}{q_k}$ $(= \alpha)$.*

(ii) *Gilt $q_k > 1$ für $\alpha \in \mathbb{Q}$, so hat man für alle $c \in \mathbb{Z}$, $d \in \mathbb{N}$ mit $d < q_k$ die Ungleichung $|q_{k-1}\alpha - p_{k-1}| \leq |d\alpha - c|$ mit Gleichheit genau dann, wenn (c, d) gleich (p_{k-1}, q_{k-1}) bzw. $(p_k - p_{k-1}, q_k - q_{k-1})$ ist.*

(iii) *Ist $\alpha \in \mathbb{Q}$, $0 \leq i \leq k - 2$, bzw. $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, $i \geq 0$, jedoch nicht gleichzeitig $i = 0$, $a_1 = 1$, so hat man für alle $c \in \mathbb{Z}$, $d \in \mathbb{N}$ mit $d < q_{i+1}$ die Ungleichung $|q_i\alpha - p_i| \leq |d\alpha - c|$ mit Gleichheit genau für $c = p_i$, $d = q_i$.*

Bemerkung. Im Fall (ii) ist $k \geq 1$ und $q_k = a_k q_{k-1} + q_{k-2} \geq 2q_{k-1}$ nach 2(1) und also genügen die voneinander verschiedenen Paare $(c, d) = (p_{k-1}, q_{k-1})$, $(p_k - p_{k-1}, q_k - q_{k-1})$ beide der Bedingung $0 < d < q_k$. Daß übrigens für diese beiden Paare in der Ungleichung von (ii) die Gleichheit eintritt, ist wegen $q_k\alpha = p_k$ einfach zu sehen.

Der erste Unterfall in (iii) kann offenbar nur bei $k \geq 2$ eintreten; weiter gilt stets $q_i \leq q_{i+1}$ mit Gleichheit genau für $(a_{i+1} - 1)q_i + q_{i-1} = 0$, d.h. für $i = 0$, $a_1 = 1$, was in (iii) gerade ausgeschlossen wurde. Unter den Bedingungen von (iii) ist also $q_i < q_{i+1}$ und so kommt hier (p_i, q_i) unter den zugelassenen Paaren (c, d) vor.

Beweis des Lemmas. Während (i) trivial ist, gewinnt man (ii) und (iii) folgendermaßen. Man betrachtet das System linearer Gleichungen

$$(2) \qquad p_i X + p_{i+1} Y = c, \qquad q_i X + q_{i+1} Y = d,$$

welches nach 2(2) die Determinante $(-1)^{i+1}$ hat und also seine Lösung (x, y) in $\mathbb{Z}^2$. Dabei ist $x \neq 0$, weil sonst $q_{i+1} | d$ aus der zweiten Gleichung (2) folgen würde, entgegen der einheitlichen Voraussetzung $d < q_{i+1}$, wenn man für (ii) das Bisherige mit $i = k-1$ anwendet.

In diesem letztgenannten Fall folgt aus (2) die Gleichung $x(q_{k-1}\alpha - p_{k-1}) = d\alpha - c$. Bei $|x| \geq 2$ hat man wegen $\frac{p_{k-1}}{q_{k-1}} \neq \alpha$ die gewünschte Ungleichung $|d\alpha - c| > |q_{k-1}\alpha - p_{k-1}|$. Wegen (2) und 2(2) ist $|x| = 1$ äquivalent zu $cq_k - dp_k = \varepsilon$ mit einem $\varepsilon \in \{1, -1\}$. Diese diophantische Gleichung vom Typ 1.3.3(2) hat nach Satz 1.3.3 genau die Lösungen

$$(3) \qquad (\varepsilon c_0 + tp_k, \varepsilon d_0 + tq_k), \qquad t \in \mathbb{Z},$$

wobei $(c_0, d_0) \in \mathbb{Z}^2$ nach 1.3.3(1) gemäß $c_0 q_k - d_0 p_k = 1$ zu wählen ist, etwa $c_0 = (-1)^k p_{k-1}$, $d_0 = (-1)^k q_{k-1}$ nach 2(2). Daß übrigens die Bedingung $ab \neq 0$ in Satz 1.3.3 hier erfüllt ist, folgt aus $p_k \neq 0$; denn aus $p_k = 0$ würde $q_k = 1$ folgen, also $k = 1$ (vgl. Bemerkung) und somit $a_1 = 1$, was nicht geht, da das letzte Element des Kettenbruchs einer rationalen, nicht ganzen Zahl stets größer als 1 ist. $|d\alpha - c| = |q_{k-1}\alpha - p_{k-1}|$ tritt also nach (3) genau für die Paare (c, d) der Form

$$(4) \qquad c = \varepsilon' p_{k-1} + tp_k, \qquad d = \varepsilon' q_{k-1} + tq_k$$

ein, wobei $\varepsilon' \in \{1, -1\}$, $t \in \mathbb{Z}$ so zu wählen sind, daß $0 < d < q_k$ gilt. Diese letzte Bedingung führt zu $t = 0$, $\varepsilon' = 1$ bzw. $t = 1$, $\varepsilon' = -1$ und so führt (4) zu den beiden in (ii) genannten Paaren (c, d).

Im Falle (iii) ist für eventuelle (c, d) der Form $(\lambda p_i, \lambda q_i)$, $\lambda = 2, 3, \ldots$, wegen $q_i\alpha \neq p_i$ die strenge Ungleichung klar. Sei jetzt also (c, d) von allen $(\lambda p_i, \lambda q_i)$, $\lambda = 1, 2, 3, \ldots$, verschieden, d.h. $\frac{c}{d} \neq \frac{p_i}{q_i}$. Dann genügt die Lösung (x, y) von (2) der Bedingung $xy < 0$. Denn $y = 0$ würde zu $\frac{p_i}{q_i} = \frac{c}{d}$ führen und $xy > 0$ wäre mit $x, y > 0$ oder $x, y < 0$ äquivalent, was wegen $0 < d < q_{i+1}$ beides der zweiten Gleichung in (2) widerspricht. Da $q_i\alpha - p_i$ und $q_{i+1}\alpha - p_{i+1}$ nach 3(1) verschiedenes Vorzeichen haben – man beachte, daß sie auch bei $\alpha \in \mathbb{Q}$ wegen $i \leq k-2$ beide nicht verschwinden –, folgt aus $xy < 0$, daß $x(q_i\alpha - p_i)$ und $y(q_{i+1}\alpha - p_{i+1})$ gleiches Vorzeichen haben und so wird nach (2)

$$\begin{aligned} |d\alpha - c| &= |x(q_i\alpha - p_i) + y(q_{i+1}\alpha - p_{i+1})| \\ &= |x|\,|q_i\alpha - p_i| + |y|\,|q_{i+1}\alpha - p_{i+1}| > |q_i\alpha - p_i|. \end{aligned} \qquad \square$$

8. Beste Näherungen. Man nennt eine rationale Zahl $\frac{a}{b}$ mit $a \in \mathbb{Z}$, $b \in \mathbb{N}$ eine *beste Näherung für* $\alpha \in \mathbb{R}$, wenn für alle $c \in \mathbb{Z}$, $d \in \mathbb{N}$ mit $\frac{c}{d} \neq \frac{a}{b}$ und $d \leq b$ gilt: $|d\alpha - c| > |b\alpha - a|$.

Mit diesem Begriff kann man behaupten

Satz A. *Jede beste Näherung für $\alpha \in \mathbb{R}$ ist ein Näherungsbruch von α.*

Beweis. Es sei $\frac{a}{b} \neq \frac{p_0}{q_0}, \ldots, \frac{p_k}{q_k}, \ldots$, wo bei $\alpha \in \mathbb{Q}$ dieselbe Verabredung wie zu Anfang von 7 getroffen sei. Ist $\alpha \in \mathbb{Q}$ und $q_k \leq b$, so nehme $c = p_k$, $d = q_k$; man erhält damit $\frac{c}{d} \neq \frac{a}{b}$, $d \leq b$ und $0 = |q_k\alpha - p_k| = |d\alpha - c| < |b\alpha - a|$ und so ist $\frac{a}{b}$ keine beste Näherung für α.

Sei nun entweder $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ oder $\alpha \in \mathbb{Q}$ habe $q_k > b$. Man fixiert i gemäß $q_i \leq b < q_{i+1}$ und erhält $i < k$ und $1 < q_{i+1} = a_{i+1}q_i + q_{i-1}$, also entweder $i \geq 1$ oder $i = 0$, $a_1 > 1$. Nach (ii) und (iii) von Lemma 7 ist $|q_i\alpha - p_i| \leq |b\alpha - a|$ wegen $b < q_{i+1}$; wählt man hier $c = p_i$, $d = q_i$, so ist $\frac{c}{d} \neq \frac{a}{b}$, $d \leq b$, $|d\alpha - c| \leq |b\alpha - a|$ erfüllt und $\frac{a}{b}$ ist wieder keine beste Näherung für α. □

Das nächste Ergebnis zeigt, daß sich Satz A in weitest möglichem Maße umkehren läßt.

Satz B. *Jeder Näherungsbruch von $\alpha \in \mathbb{R}$ ist eine beste Näherung für α, wenn man für die α der Form $[a_0; 2]$, $[a_0; 1, a_2, \ldots, a_k]$ mit einem $k \geq 2$ bzw. $[a_0; 1, a_2, \ldots]$ von den Näherungsbrüchen nullter Ordnung absieht. In diesen Ausnahmefällen ist der nullte Näherungsbruch tatsächlich keine beste Näherung für α.*

Beweis. Ist $\frac{p_i}{q_i}$ ein Näherungsbruch von α (mit $i \geq 1$ in den Ausnahmefällen), so ist für alle $c \in \mathbb{Z}$, $d \in \mathbb{N}$ mit $\frac{c}{d} \neq \frac{p_i}{q_i}$, $d \leq q_i$ die Ungleichung $|d\alpha - c| > |q_i\alpha - p_i|$ zu zeigen.

Ist $\alpha \in \mathbb{Q}$, so ist dies bei $i = k$ nach Lemma 7(i) in Ordnung. Da im Falle $\alpha \in \mathbb{Q}$ nicht gleichzeitig $k = 1$ und $a_k = 2$ gilt, ist $q_k - q_{k-1} > q_{k-1}$ und so gilt hier das Gewünschte für $i = k - 1$ nach (ii) des Lemmas. Da im Falle $\alpha \in \mathbb{Q}$, $0 \leq i \leq k - 2$ bzw. $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, $i \geq 0$ nicht gleichzeitig $i = 0$, $a_1 = 1$ gilt, folgt jetzt das Gewünschte aus Lemma 7(iii).

Im ersten Ausnahmefall $\alpha = a_0 + \frac{1}{2}$ ist $p_0 = a_0$, $q_0 = 1$, $p_1 = 2a_0 + 1$, $q_1 = 2$, also $|q_0\alpha - p_0| = |(q_1 - q_0)\alpha - (p_1 - p_0)|$ $(= \frac{1}{2})$ nach (ii) und so ist hier $\frac{p_0}{q_0}$ keine beste Näherung für α. Im zweiten und dritten Ausnahmefall ist $\alpha - a_0 > \frac{p_2}{q_2} - a_0 = (1 + a_2^{-1})^{-1} \geq \frac{1}{2}$ für $\alpha \notin \mathbb{Q}$ bzw. $\alpha \in \mathbb{Q}$, $k \geq 3$; bei $\alpha \in \mathbb{Q}$, $k = 2$ ist hier $\alpha - a_0 = \frac{p_2}{q_2} - a_0 = (1 + a_2^{-1})^{-1} \geq \frac{2}{3}$ wegen $a_2 \geq 2$. Insgesamt ist in den letztgenannten Ausnahmefällen $\alpha - a_0 > \frac{1}{2}$ oder also $\alpha - a_0 > (a_0+1) - \alpha$ (> 0), was zeigt, daß hier erneut $\frac{p_0}{q_0}$ keine beste Näherung für α ist. $\square$

Unter Verwendung der in der Fußnote zu 4.3.2 eingeführten Bezeichnung $\|\cdot\|$ für den Abstand zur nächsten ganzen Zahl soll aus Satz B noch gefolgert werden die

Proposition. *Für jedes reelle α gilt $\|q_0\alpha\| \geq \|q_1\alpha\| > \ldots > \|q_k\alpha\| > \ldots$, wobei diese Ungleichungskette im Falle $\alpha \in \mathbb{Q}$, $\alpha = \frac{p_k}{q_k}$ mit dem Gliede $\|q_k\alpha\| = 0$ abbricht. Dabei tritt an der einzig möglichen Stelle Gleichheit ein genau dann, wenn das erste Kettenbruchelement a_1 von α gleich 1 ist, was im Falle $\alpha \in \mathbb{Q}$ nur bei $k \geq 2$ vorkommen kann.*

Beweis. Sei $\alpha \notin \mathbb{Z}$, etwa $\alpha = [a_0; a_1, \ldots, a_k, \ldots]$ mit einem $k \geq 1$ für rationales α. Nach Satz B ist $\frac{p_i}{q_i}$ für $i \geq 1$ eine beste Näherung für α und so hat man wegen $q_{i-1} \leq q_i$, $\frac{p_{i-1}}{q_{i-1}} \neq \frac{p_i}{q_i}$ die Ungleichung $|q_{i-1}\alpha - p_{i-1}| > |q_i\alpha - p_i|$; ebensoleicht sieht man $|q_0\alpha - (p_0+1)| \geq |q_1\alpha - p_1|$ mit Gleichheit genau dann, wenn $a_1 = 1$ gilt. Dies liefert die Ungleichungskette

$$\text{(1)} \qquad \mathrm{Min}(|q_0\alpha - p_0|, |q_0\alpha - p_0 - 1|) \geq |q_1\alpha - p_1| > \ldots > |q_k\alpha - p_k| > \ldots,$$

die gegebenenfalls nach dem $(k+1)$–tem Glied abbricht. Nach 7(1) ist mit α_{i+1} wie in 4

$$|q_i\alpha - p_i| = (q_i\alpha_{i+1} + q_{i-1})^{-1} \leq (\alpha_{i+1}+1)^{-1} < \frac{1}{2}$$

für $i \geq 1$, wenn $i < k$ im Falle eines rationalen α erfüllt ist; daher gilt $|q_i\alpha - p_i| = \|q_i\alpha\|$ für $i \geq 1$ stets. Da das Minimum links in (1) gleich $\|q_0\alpha\|$ ist, folgt daraus die Behauptung der Proposition. $\square$

Die zuletzt bewiesenen Sätze A und B zeigen, daß — abgesehen von gewissen seltenen, vollständig zu charakterisierenden Ausnahmen — die besten Näherungen für eine reelle Zahl α *genau* deren Näherungsbrüche sind.

9. Anmerkungen dazu. Ist $\frac{a}{b}$ eine beste Näherung für α, so ist nach der anfangs von 8 gegebenen Definition $|\alpha - \frac{a}{b}| < \frac{d}{b}|\alpha - \frac{c}{d}|$, a fortiori also $|\alpha - \frac{a}{b}| < |\alpha - \frac{c}{d}|$ für alle rationalen $\frac{c}{d} \neq \frac{a}{b}$ mit $d \leq b$. Dies kann man unter Beachtung von Satz 8B in einer für das numerische Rechnen sehr wichtigen Weise interpretieren: Ersetzt man α durch seinen i–ten Näherungsbruch $\frac{p_i}{q_i}$, $i \geq 1$, so ist der dabei begangene Fehler absolut kleiner, als wenn man α durch irgendeine andere rationale Zahl mit Nenner höchstens q_i ersetzt.

Es sei an dieser Stelle angemerkt, daß die Kettenbruchentwicklung der Zahl π bisher unbekannt ist. Sie beginnt mit $[3; 7, 15, 1, 292, \ldots]$ und daher sind $3, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103993}{33102}, \ldots$ die ersten fünf Näherungsbrüche von π. Während in China bereits vor über 1500 Jahren $\frac{355}{113}$ als Näherungswert für π auftrat, scheint sich etwa gleichzeitig im Abendland, ausgehend von den römischen Agrimensoren, die Zahl $\frac{22}{7}$ als sehr gute Annäherung an π verbreitet zu haben. Nach den Sätzen 8A und 8B sind die besten Näherungen für π genau die Näherungsbrüche von π. So wird klar, warum man keine rationale Zahl mit Nenner unterhalb 8 finden konnte, die die Zahl π so gut approximiert wie eben $\frac{22}{7}$.

Bereits 1685 hat Wallis in seinem *Tractatus de Algebra* unter Benutzung der ersten Ziffern der Dezimalbruchentwicklung von π die ersten 34 Kettenbruchelemente von π berechnet. Seit 2003 sind vom Kettenbruch von π mehr als 10^8 Anfangselemente bekannt (E.W. Weisstein, http://mathworld.wolfram.com/PiContinuedFraction.html).

Das in 8 besprochene Problem der möglichst guten Annäherung beliebiger reeller Zahlen durch rationale Zahlen möglichst kleiner Nenner war historisch sogar die entscheidende Triebfeder zur Entdeckung und Untersuchung des Kalküls der regelmäßigen Kettenbrüche. Wie man nämlich seinen *Opuscula Posthuma, Descriptio Automati Planetarii* (1703) entnimmt, wurde C. Huygens auf dieses Problem bei der Herstellung eines Zahnradmodells des Sonnensystems geführt. Dabei mußte er die Anzahlen der Zähne in seinem Modell so wählen, daß ihr Quotient für zwei gekoppelte Zahnräder (d.h. der Quotient der Umlaufzeiten dieser beiden Räder) den Quotienten α der Umlaufzeiten der beiden darzustellenden Planeten möglichst gut annäherte. Aus technischen Gründen sollten gleichzeitig die Anzahlen der Zähne nicht zu groß werden.

10. Approximation algebraischer Zahlen zweiten Grades durch rationale. Die Annäherung einer beliebigen Zahl α durch solche rationale, die speziell Näherungsbrüche sind, wird durch 3(1) kontrolliert. Für Zahlen α mit geeigneten Voraussetzungen über ihre Kettenbruchentwicklung (oder besser noch: Kenntnis derselben) kann durch Kombination von 3(1) mit Lemma 7 die Annäherung durch *beliebige* rationale Zahlen angegriffen werden. Ein Beispiel dieser Art liefert der

Satz. *Ist für reelle irrationale $\alpha = [a_0; a_1, a_2, \ldots]$ die Folge $(a_i)_{i=1,2,\ldots}$ durch A (≥ 1) beschränkt, so gilt für alle $p \in \mathbb{Z}$, $q \in \mathbb{N}$ die Ungleichung*

$$|\alpha - \frac{p}{q}| > (A+2)^{-1} q^{-2}.$$

Beweis. Sei ein Paar (p, q) wie im Satz beliebig vorgegeben. Man definiere i in eindeutiger Weise durch die Forderung $q_i \le q < q_{i+1}$: es ist daher $i \ge 0$ und sogar $i \ge 1$, falls $a_1 = 1$ ist. Daher ist nach Lemma 7(iii) und wegen $q_{i-1} \le q_i$

$$|q\alpha - p| \ge |q_i\alpha - p_i| > ((a_{i+1}+1)q_i + q_{i-1})^{-1} \ge (A+2)^{-1}q_i^{-1},$$

woraus mit $q_i \le q$ die Behauptung folgt. □

Bemerkung. Für die im Satz behandelten α gibt es also eine nur von α abhängige Konstante $c(\alpha) > 0$, so daß $|\alpha - \frac{p}{q}| > c(\alpha)q^{-2}$ für alle rationalen $\frac{p}{q}$ gilt. Andererseits hat $|\alpha - \frac{p}{q}| < q^{-2}$ für jedes irrationale α unendlich viele rationale Lösungen $\frac{p}{q}$ und so ist die obige Aussage bis auf den Wert von $c(\alpha)$ bestmöglich.

Ein Ergebnis aus der sogenannten metrischen Kettenbruchtheorie besagt, daß für fast alle reellen α (im Sinne des LEBESGUE–Maßes in $\mathbb{R}$) die Folge $(a_i(\alpha))_{i=1,2,\dots}$ unbeschränkt ist, wobei $\alpha = [a_0(\alpha); a_1(\alpha), \dots]$. Obschon also die Voraussetzung über α im obigen Satz fast nie zutrifft, gehört doch eine wichtige Klasse reeller Zahlen sicher zur Ausnahmemenge, nach dem LAGRANGEschen Satz 5 nämlich alle reell–quadratischen Zahlen, da ja die Periodizität der Folge (a_i) deren Beschränktheit impliziert. Somit hat man folgendes

Korollar. *Zu jeder algebraischen Zahl α zweiten Grades gibt es eine Konstante $c(\alpha) > 0$, so daß für alle rationalen $\frac{p}{q}$ gilt*

$$|\alpha - \frac{p}{q}| > c(\alpha)q^{-2}.$$

Für reelle α der genannten Art wurde dies bereits oben gefolgert. Ist $\alpha \in \mathbb{C} \setminus \mathbb{R}$, was im Korollar ja vorkommen kann, so ist trivialerweise und ohne jede arithmetische Voraussetzung an α

$$|\alpha - \frac{p}{q}| \ge |\operatorname{Im}\alpha| \ge c(\alpha)q^{-2}$$

für alle p, q und mit $c(\alpha) := |\operatorname{Im}\alpha| > 0$.

Bemerkung. Dies Korollar wird in 6.1.2 durch den Satz von LIOUVILLE verallgemeinert.

11. Eine arithmetische Eigenschaft von $e^{2/k}$. Bisher wurden Kettenbruchentwicklungen nur für gewisse reelle algebraische Zahlen höchstens zweiten

Grades hier explizit angegeben. In 12 soll nun noch diejenige von e abgeleitet werden, eine etwas schwierigere Aufgabe.

Dazu betrachtet man vorbereitend für $n = 0, 1, \ldots$ die folgenden positiven reellen Zahlen

$$(1) \qquad \xi_n := \frac{1}{n!}\int_0^1 x^n(1-x)^n e^{2x/k}dx, \quad \eta_n := \frac{1}{n!}\int_0^1 x^{n+1}(1-x)^n e^{2x/k}dx$$

mit fixiertem $k \in \mathbb{N}$.

Das erste Ziel ist die Herleitung einer zweigliedrigen Rekursionsformel für die ξ_n; die η_n sind lediglich bequeme Hilfsgrößen. Indem man auf das erste Integral in (1) zunächst partielle Integration anwendet, findet man für $n \geq 1$

$$\frac{2}{k}n!\xi_n = n\int_0^1 (2x-1)x^{n-1}(1-x)^{n-1}e^{2x/k}dx = n!(2\eta_{n-1} - \xi_{n-1}),$$

also

$$(2) \qquad \frac{2}{k}\xi_n + \xi_{n-1} = 2\eta_{n-1} \qquad \text{für } n \geq 1.$$

Völlig analog erhält man aus dem zweiten Integral in (1) durch partielle Integration

$$(2n+1)\xi_n = \eta_{n-1} - \frac{2}{k}\eta_n \qquad \text{für } n \geq 1.$$

Eliminiert man hieraus die η's mittels (2), so bekommt man

$$(3) \qquad \frac{2}{k}\xi_{n+1} + (2n+1)k\xi_n = \frac{k}{2}\xi_{n-1} \qquad (n \geq 1)$$

oder äquivalent für dieselben n

$$(3') \qquad \frac{k\xi_{n-1}}{2\xi_n} = (2n+1)k + \frac{2\xi_{n+1}}{k\xi_n}\,.$$

Weiter erhält man aus (1)

$$(4) \qquad \xi_0 = \frac{k}{2}(e^{2/k}-1), \quad \eta_0 = \frac{k}{2}e^{2/k} - \left(\frac{k}{2}\right)^2(e^{2/k}-1),$$

was mittels (2), angewandt für $n = 1$, direkt zu

$$(5) \qquad \xi_1 = \left(\frac{k}{2}\right)^2(e^{2/k} + 1 - k(e^{2/k}-1))$$

führt. Aus (4) und (5) folgt $(e^{2/k}+1)/(e^{2/k}-1) = k + (2\xi_1)/(k\xi_0)$ oder in Kettenbruchschreibweise $(e^{2/k}+1)/(e^{2/k}-1) = [k; (k\xi_0)/(2\xi_1)]$. Wendet man hier rechts (3′) mit $n = 1$ an, so wird $(e^{2/k}+1)/(e^{2/k}-1) = [k; 3k, (k\xi_1)/(2\xi_2)]$ und schließlich folgt mittels (3′) induktiv

$$(6) \qquad \frac{e^{2/k}+1}{e^{2/k}-1} = [k; 3k, 5k, 7k, \ldots] \qquad \text{für } k \in \mathbb{N}.$$

Hieraus kann geschlossen werden auf die

Proposition *Die Zahl $e^{2/k}$ ist für kein ganzes $k \neq 0$ algebraisch von einem Grade höchstens zwei.*

Bemerkung. Für $k = 2$ wurde dies bereits in Proposition 2.4B eingesehen.

Beweis. Bei $k \in \mathbb{N}$ werde $z := e^{2/k}$ und $y := \frac{z+1}{z-1}$ gesetzt. Gäbe es ein nichttriviales Tripel $(a, b, c) \in \mathbb{Z}^3$ mit $az^2 + bz + c = 0$, so wäre dies wegen $z = \frac{y+1}{y-1}$ mit $a^*y^2 + b^*y + c^* = 0$, $a^* := a + b + c$, $b^* := 2(a - c)$, $c^* := a - b + c$ äquivalent. Nach (6), Satz 3 und dem LAGRANGEschen Satz 5 ist y nicht algebraisch von einem Grade höchstens zwei und also folgt $a^* = b^* = c^* = 0$, was seinerseits zu $a = b = c = 0$ führt, entgegen der gemachten Annahme. Das Additionstheorem der Exponentialfunktion erlaubt es schließlich, den Fall $-k \in \mathbb{N}$ auf den soeben abgeschlossenen zurückzuspielen. □

12. Kettenbruchentwicklung von e. Für $k = 2$ ergibt sich aus 11(6)

$$\frac{e+1}{e-1} = [2; 6, 10, 14, \ldots]. \tag{1}$$

Die beiden Kettenbrüche 11(6), (1) sind von EULER 1737 gefunden worden.

Es liegt nun nahe zu versuchen, aus 11(6) den Kettenbruch für $e^{2/k}$ selbst zu ermitteln. EULER gelang dies 1737 für gerade k; der Fall ungerader k konnte erst um die Wende zum 20. Jahrhundert von K.F. SUNDMAN (für $k = 1$) und schließlich STIELTJES (für $k = 1, 3, 5, \ldots$) erledigt werden. Die beiden hier unterschiedenen Fälle für k sind bei PERRON [19], Band I, §§ 31, 32 ausführlich dargestellt.

Hier soll lediglich noch der Fall $k = 2$ behandelt werden; d.h. aus (1) soll die Kettenbruchentwicklung für e abgeleitet werden. Dazu wird die EULERsche Verfahrensweise angewandt: Durch explizite Berechnung genügend vieler Anfangsglieder der Elementenfolge des Kettenbruchs von e hat er diesen erraten und seine Vermutung anschließend mit Hilfe von (1) bewiesen. Behauptet wird

$$e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, \ldots]. \tag{2}$$

Zum *Beweis* setzt man $a_i := 4i + 2$ $(i = 0, 1, \ldots)$ und bezeichnet mit p_i bzw. q_i die Näherungszähler bzw. –nenner des Kettenbruchs $[a_0; a_1, a_2, \ldots] = \frac{e+1}{e-1}$, vgl. (1). Weiter definiert man

$$a'_0 := 2 \quad \text{sowie} \quad a'_{3j-2} := 1, \quad a'_{3j-1} := 2j, \quad a'_{3j} := 1 \quad (j \geq 1) \tag{3}$$

und $e' := [a'_0; a'_1, \ldots]$. Dieser letzte Kettenbruch habe die Näherungszähler bzw. –nenner p'_i, q'_i. Wegen (3) und der Definition von e' ist für den Beweis von (2) $e' = e$ zu zeigen. Kernpunkt dieses Beweises ist die Bestätigung der Formeln

$$(4) \qquad p'_{3j+1} = p_j + q_j, \quad q'_{3j+1} = p_j - q_j \qquad \text{für } j \geq 0.$$

Wegen der leicht zu verifizierenden Gleichungen $p_0 = 2$, $q_0 = 1$, $p_1 = 13$, $q_1 = 6$, $p'_1 = 3$, $q'_1 = 1$, $p'_4 = 19$, $q'_4 = 7$ gelten (4) für $j = 0, 1$. Sei $j \geq 2$ und (4) bereits für $j-2$ und $j-1$ als richtig erkannt. Aus den Rekursionsformeln 2(1) für die gestrichenen p's folgt unter Beachtung von (3)

$$\begin{aligned}
p'_{3j-3} &= p'_{3j-4} + p'_{3j-5},\\
p'_{3j-2} &= p'_{3j-3} + p'_{3j-4},\\
p'_{3j-1} &= 2jp'_{3j-2} + p'_{3j-3},\\
p'_{3j} &= p'_{3j-1} + p'_{3j-2},\\
p'_{3j+1} &= p'_{3j} + p'_{3j-1}.
\end{aligned}$$

Multipliziert man hier die erste,...,fünfte Gleichung mit $1, -1, 2, 1, 1$ und addiert anschließend alles, so erhält man unter Berücksichtigung der Induktionsvoraussetzung

$$\begin{aligned}
p'_{3j+1} &= (4j+2)p'_{3j-2} + p'_{3j-5}\\
&= (4j+2)(p_{j-1} + q_{j-1}) + (p_{j-2} + q_{j-2})\\
&= \{(4j+2)p_{j-1} + p_{j-2}\} + \{(4j+2)q_{j-1} + q_{j-2}\}\\
&= p_j + q_j.
\end{aligned}$$

Dabei wurden zuletzt auch die Formeln 2(1) für die ungestrichenen p, q ausgenützt. Somit hat man die erste Behauptung in (4) und die zweite folgt ganz analog.

Nach (4) ist

$$e' = \lim_{j\to\infty} \frac{p'_{3j+1}}{q'_{3j+1}} = \lim_{j\to\infty} \frac{\frac{p_j}{q_j} + 1}{\frac{p_j}{q_j} - 1} = \frac{\frac{e+1}{e-1} + 1}{\frac{e+1}{e-1} - 1} = e. \qquad \square$$

Bemerkung. Die Herleitung der arithmetischen Aussagen in diesem und dem letzten Abschnitt benutzte sowohl das Additionstheorem wie die Differentialgleichung der Exponentialfunktion, letztere bei den partiellen Integrationen nach 11(1). Beide Eigenschaften werden wieder eine entscheidende Rolle spielen, wenn in Kap. 6, §§ 3–5 mit analytischen Methoden gewisse mit der Exponentialfunktion zusammenhängende komplexe Zahlen auf Transzendenz untersucht werden. Dort werden sich insbesondere alle Zahlen $e^{2/k}$, $k \in \mathbb{Z} \setminus \{0\}$, und damit die in 11(6) als transzendent erweisen. (Vgl. hierzu auch die Anmerkung am Ende von 5 über Kettenbrüche algebraischer Zahlen mindestens dritten Grades.)