BAYERISCHE AKADEMIE DER WISSENSCHAFTEN
MATHEMATISCH-NATURWISSENSCHAFTLICHE KLASSE

SITZUNGSBERICHTE

JAHRGANG
1966

MÜNCHEN 1967

VERLAG DER BAYERISCHEN AKADEMIE DER WISSENSCHAFTEN
In Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

Über das Kettenbruchverfahren von *Patz* und *Arwin* zur Darstellung von Zahlen durch positiv definite binäre quadratische Formen

Von **Hermann Schmidt** in Würzburg

Vorgetragen am 14. Januar 1966

Die folgenden Bemerkungen haben das Ziel, den Anregungen von Herrn Perron ([2], Vorbemerkung) folgend und über die Beiträge von Herrn Steuerwald [5] hinaus, soweit möglich, einwandfreie allgemeine Grundlagen für das Rechenverfahren zu geben, das Herr Patz [2] Beispiel 1. 2. zur Gewinnung von Lösungen in numerisch unbequemen Fällen so erfolgreich angewendet hat. Es zeigt sich, daß dieses im Grundgedanken mit einem von Arwin [1] § 6, S. 66–68 vor Jahren kurz angegebenen übereinstimmt; nur sind dort die benützten Kettenbruchteilnenner nach einem anderen Gesetz ausgewählt (entsprechend der Gaußschen Reduktionstheorie), so daß das Patzsche Verfahren für sich untersucht werden muß. Das Ergebnis ist (**Satz 1**), daß für die Gleichung (1) der mit $L = Q_0$ begonnene Algorithmus stets in einer von vornherein abschätzbaren Anzahl von Schritten zum Ziele führen muß, wenn eine Lösung vorhanden ist, so daß er also gleichzeitig den Entscheid darüber, wie auch bejahendenfalls alle Lösungen selbst ergibt. Sucht man dagegen den nämlichen Algorithmus (d.h. wieder den für Q_0, nicht für cQ_0) für die Gleichung (17) auszunützen, so ist das Ergebnis weniger befriedigend (**Satz 2**). Bei einem dem Einzelfall anzupassenden Verfahren würde er zwar ebenfalls jede Lösung liefern, allein die Auswahl der Teilnenner ist nicht von vornherein bekannt, so daß man nicht sicher ist, ob man praktisch alle gewünschten Lösungen erhält. So kann man hier wohl auch heute noch nur von einem allerdings wohlmotivierten Probierverfahren sprechen (vgl. auch die Bemerkungen über [5] am Ende sowie das Summar S. 8* dieses Bandes).

Es seien D, L aus dem Bereich N der natürlichen Zahlen fest vorgegeben; Γ bezeichne den Ring der ganzen rationalen Zahlen. Gesucht sind die Lösungen $\mathfrak{x} = (x, y)$ von

$$(1) \qquad F(\mathfrak{x}) = x^2 + Dy^2 = L \text{ mit } |x|, |y| \in N, \ (x, y) = 1,$$

so daß auch

$$(2) \qquad (y, L) = 1;$$

es zähle jeweils nur ein Vertreter für 4 nur durch die Vorzeichen verschiedene Paare $(\pm x, \pm y)$. Als notwendige Bedingung ergibt sich sofort die Lösbarkeit der Kongruenz

$$(3) \qquad X^2 + D \equiv 0 \,(L); \qquad \text{ist } X = P_0 \in \Gamma$$

eine Lösung, und schreibt man noch $L = Q_0$, so gilt also mit $Q_{-1} \in N$ eine Beziehung

$$(4)_0 \qquad P_0^2 + D = Q_0 Q_{-1}.$$

Sei nun $(4)_0$ erfüllt; dann bilden wir mit einer Folge ganzer Zahlen b_ν die Folge imaginär-quadratischer Zahlen $\zeta_\nu = \frac{P_\nu + \sqrt{-D}}{Q_\nu}$ $(\nu = 0, 1, \ldots)$ nach dem Rekursionsgesetz

(5.1)

$$\zeta_{\nu+1} = \frac{1}{\bar{\zeta}_\nu - b_\nu} \text{ oder also } \zeta_\nu = b_\nu + \frac{1}{\bar{\zeta}_{\nu+1}} \ (\zeta, \bar{\zeta} \text{ konjugiert imaginär}).$$

Wie man leicht induktiv erkennt, sind hier P_ν, Q_ν (eindeutig bestimmte) Zahlen $\in \Gamma$, und es gilt die Matrixgleichung

$$(5.2) \qquad \mathfrak{Q}_{\nu+1} = \mathfrak{B}_\nu \mathfrak{Q}_\nu \mathfrak{B}_\nu; \text{ hierbei ist gesetzt}$$

$$\mathfrak{Q}_\nu = \begin{pmatrix} Q_\nu, & -P_\nu \\ -P_\nu, & Q_{\nu-1} \end{pmatrix}, \qquad \mathfrak{B}_\nu = \begin{pmatrix} b_\nu & 1 \\ 1 & 0 \end{pmatrix}.$$

Schreibt man noch $F_\nu(\mathfrak{x}) = \mathfrak{x} \mathfrak{Q}_\nu \tilde{\mathfrak{x}}$, so gehen nach (5.2) die quadratischen Formen F_ν vermöge $\mathfrak{x}_\nu = \mathfrak{x}_{\nu+1} \mathfrak{B}_\nu$ paarweise auseinander hervor (der untere Index diene zur Unterscheidung der in

den einzelnen Formen zugrunde gelegten Variablen), sind also sämtlich (eigentlich oder uneigentlich) miteinander äquivalent, und zwar alle positiv definit mit der gemeinsamen Determinante

$$(4)_\nu \qquad \det \mathfrak{Q}_\nu = Q_\nu Q_{\nu-1} - P_\nu^2 = D.$$

Ferner ist im einzelnen nach (5.2)

$$(6.1) \quad P_{\nu+1} = P_\nu - b_\nu Q_\nu,$$

$$(6.2) \quad Q_{\nu+1} = F_\nu(b_\nu, 1) =$$
$$= b_\nu^2 Q_\nu - 2 b_\nu P_\nu + Q_{\nu-1} = Q_{\nu-1} - b_\nu (P_\nu + P_{\nu+1})$$

(6.1), (6.2) (letzte Form) liefern ein recht bequemes Rechenschema (es gibt bei der Annahme (10) genau die Patzschen Werte; der von der üblichen Kettenbruchtheorie – vgl. etwa [2], S. 22 oder [4], S. 174 – abweichende Ansatz (5.1) ist dem positivdefiniten Fall angemessen und befreit von lästigen Vorzeichenschwankungen, s. sogleich (8)). Setzt man noch

$$(7) \qquad \mathfrak{C}_0 = \mathfrak{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathfrak{C}_{\nu+1} = \mathfrak{B}_\nu \mathfrak{C}_\nu \ (\nu = 0,1,2, \ldots),$$

so folgt

$$(8) \qquad \mathfrak{Q}_\nu = \mathfrak{C}_\nu \mathfrak{Q}_0 \tilde{\mathfrak{C}}_\nu \quad \text{mit} \quad \mathfrak{C}_\nu = \begin{pmatrix} A_{\nu-1} & B_{\nu-1} \\ A_{\nu-2} & B_{\nu-2} \end{pmatrix},$$

worin also jetzt $A_\varkappa$, $B_\varkappa$ die Näherungszähler bzw. -nenner des folgenden Kettenbruches sind (der freilich kein regelmäßiger zu sein braucht): $b_0 + 1/b_1 + 1/b_2 + \ldots$

Aus (8) folgt dann noch $Q_\nu = F_0(A_{\nu-1}, B_{\nu-1})$, und wegen $Q_0 F_0(x,y) = F(Q_0 x - P_0 y, y)$ ist

$$(9) \qquad (Q_0 A_{n-1} - P_0 B_{n-1}, B_{n-1})$$

eine Lösung von (1), *falls* $Q_n = 1$ ausfällt.

Übrigens folgt aus (8) in der Gestalt $\mathfrak{Q}_\nu \tilde{\mathfrak{C}}_\nu^{-1} = \mathfrak{C}_\nu \mathfrak{Q}_0$ sofort

$$(9.1) \qquad Q_0 A_{\nu-1} - P_0 B_{\nu-1} = (-1)^\nu (Q_\nu B_{\nu-2} + P_\nu B_{\nu-1}),$$

so daß für $Q_n = 1$ in (9) auch $x = B_{n-2} + P_n B_{n-1}$ genommen werden kann. Gilt nun das Gesetz

$$b_\nu = \left[\frac{P_\nu}{Q_\nu}\right] \tag{10}$$

speziell für $\nu = n$ (was ohne Änderung der in (9) gebrauchten Werte auch durch nachträgliche Änderung eingerichtet werden kann), dann wird schließlich aus (9) die Lösung

$$\mathfrak{x} = (B_n, B_{n-1}). \tag{11}$$

(Falls $b_n = 0$, folgt aus $(4)_n$ $Q_{n-1} = D$, und damit $F_0 \sim F_n = F$ vermöge $\mathfrak{x}_0 = \mathfrak{x}\,\mathfrak{E}_n$, wodurch (11) aus $\mathfrak{x}_0 = (1, 0)$ entspringt; für $b_n > 0$ wird $Q_{n-1} = D + b_n^2$, ferner nach (6) $\mathfrak{Q}_{n+1} = \begin{pmatrix} D & 0 \\ 0 & 1 \end{pmatrix}$, $\mathfrak{Q}_{n+2} = \begin{pmatrix} 1 & 0 \\ 0 & D \end{pmatrix}$, somit $F_0 \sim F_{n+2} = F$, falls man (10) auch noch für $\nu = n + 1$ festsetzt).

Bisher haben wir nicht untersucht, ob bzw. für welche Auswahl der b_ν das Verfahren tatsächlich zu einer Lösung führt, bzw. ob, falls dies nicht eintritt, auf die *Unlösbarkeit* von (1) geschlossen werden kann. In dieser Hinsicht beweisen wir nun

Satz 1. *In dem durch (5.1), (5.2) beschriebenen Algorithmus gelte stets (10). Dann ist (1) genau dann lösbar, wenn für mindestens eine Lösung P_0 von (3) mit $0 < P_0 < Q_0$ einmal $Q_n = 1$ auftritt, und man erhält alle Lösungen in der Form (11), wenn man P_0 alle diese Werte durchlaufen läßt (wobei nur einer der Werte P_0 und $Q_0 - P_0$ berücksichtigt zu werden braucht).*

Wenn überhaupt, tritt $Q_n = 1$ nach höchstens $Q_0 - 1$ Schritten auf, so daß das Verfahren die Entscheidung über die Lösbarkeit gleichzeitig mit allen etwaigen Lösungen liefert.

Beweis. Es sei die Lösung $x = a$, $y = b$ ($\in N$) von (1) vorgelegt. Dann entwickeln wir den Quotienten $a : b$ auf eine der beiden möglichen Arten in einen endlichen, regelmäßigen Kettenbruch, den wir schreiben

$$a : b = [b_n, b_{n-1}, \ldots b_1] \qquad (n \geq 1). \tag{12}$$

Mit diesen Werten b_ν sowie $b_0 = 0$ berechnen wir A_ν, B_ν nach (7). $(-2 \leq \nu \leq n)$; A_ν, B_ν sind dann Näherungszähler und -nenner des Kettenbruchs $^1/_{b_1} + ^1/_{b_2} + \ldots ^1/_{b_n}$, der allerdings für $b_n = 0$ kein regelmäßiger ist. Alsdann ist nach einer bekannten Formel $a : b = B_n : B_{n-1}$, und wegen $(a, b) = 1$ (s. (1)) ist

$$a = B_n, b = B_{n-1}. \tag{13}$$

Nunmehr setzen wir

(14)

$$\mathfrak{Q}_n^* = \begin{pmatrix} 1, & -b_n \\ -b_n, & D + b_n^2 \end{pmatrix} \text{ und } \mathfrak{Q}_0^* = \mathfrak{E}_n^{-1} \mathfrak{Q}_n^* \tilde{\mathfrak{E}}_n^{-1} = \begin{pmatrix} Q_0^*, & -P_0^* \\ -P_0^*, & Q_{-1}^* \end{pmatrix}.$$

Wählt man jetzt als Anfang der Rekursion (5.2) die Matrix $\mathfrak{Q}_0^*$, so erscheint wegen (7) nach n Schritten die Matrix $\mathfrak{Q}_n^*$. Ferner ist $\det \mathfrak{Q}_0^* = D$, somit ist $(4)_0$ für P_0^*, Q_0^*, Q_{-1}^* befriedigt; daraus folgt aber wegen $b_n = P_n^*$, daß nach (11) $B_n^2 + B_{n-1}^2 D = Q_0^*$, also $Q_0^* = Q_0$ wegen (13). Die Sterne können und sollen jetzt weggelassen werden; wir haben gezeigt, daß das mit den Werten aus (12) und einer geeigneten Lösung von $(4)_0$ gebildete Kettenbruchverfahren jedenfalls die beliebig vorgegebene Lösung (a, b) liefert. Als nächstes zeigen wir, daß hier stets (10) erfüllt ist, so daß es also mit dem durch P_0 und (10) festgelegten Verfahren zusammenfällt. In der Tat sind zunächst alle $Q_\nu > 0$, da ja die F_ν positiv definit sind. Somit gilt nach (6.1)

$$P_\nu = P_{\nu+1} + b_\nu Q_\nu \geq 1 \ (\nu = n-1, n-2, \ldots 1, 0), \text{ also} \tag{6.1)'}$$

$$P_\nu - b_\nu Q_\nu = P_{\nu+1} \geq 1 \text{ für } 0 \leq \nu \leq n-2, \text{ also}$$

$$b_\nu \leq \frac{P_\nu}{Q_\nu} \text{ für } 0 \leq \nu \leq n \text{ (vgl. (14) für } \nu = n-1, n). \tag{15}$$

Endlich ist $Q_{n-1} = b_n^2 + D > P_n = b_n$, und für $1 \leq \nu \leq n-1$ wegen der aus (6.1) (6.2) (bzw. aus der Umkehrung von (5.2)) fließenden Beziehung

$$Q_{\nu-1} = Q_{\nu+1} + 2 b_\nu P_{\nu+1} + b_\nu^2 Q_\nu \ (= F_{\nu+1}(1, -b_\nu)) \tag{6.2)'}$$

schließlich

(6.3)
$$Q_{\nu-1} - P_\nu = Q_{\nu+1} + (2b_\nu - 1)\, P_{\nu+1} + b_\nu (b_\nu - 1)\, Q_\nu \geq Q_{\nu+1} > 0,$$

und dies besagt, daß

$$(16) \qquad b_\nu = \frac{P_\nu}{Q_\nu} - \frac{P_{\nu+1}}{Q_\nu} > \frac{P_\nu}{Q_\nu} - 1 \text{ für } 0 \leq \nu \leq n-1,$$

womit (10) gezeigt ist. Insbesondere wird nach (6.1)', (6.3) $0 < P_0 < Q_0$. Endlich wird, solange $b_\nu \geq 1$, d.h. für $\nu = 1, 2 \ldots n-1$ nach (6.2)', und für $\nu = n$ nach (14)

$Q_\nu < Q_{\nu-1}$, die Matrix (14) entspricht also der *ersten* auftretenden $1 = Q_n \leq Q_0 - n$, $n \leq Q_0 - 1$ (für $D = 1$ bei der dann zulässigen Vor. $a > b$).

Es bleibt noch zu zeigen, daß $Q_0 - P_0$ zur nämlichen Lösung führt wie gegebenenfalls P_0. Hierzu werde wieder von dem Kettenbruch (12) ausgegangen, und es sei darin etwa $b_1 \geq 2$ gewählt worden. Für die andere Möglichkeit, die im folgenden durch Überstreichen angedeutet werde, ist dann $\bar{n} = n + 1$, und für $\nu \geq 2$ $\bar{b}_{\nu+1} = b_\nu$, daher

$$\bar{P}_{\nu+1} = P_\nu,\ \bar{Q}_\nu = Q_{\nu-1},\ \text{ferner } \bar{b}_2 = b_1 - 1,\ \bar{b}_1 = 1,\ \bar{b}_0 = b_0 = 0.$$

Das gibt (mit $P_3 = 0$ für $n = 1$)

$$\bar{P}_2 = \bar{P}_3 + \bar{b}_2 \bar{Q}_2 = P_2 + (b_1 - 1)\, Q_1 = P_1 - Q_1$$
$$\bar{P}_1 = \bar{P}_2 + 1 \cdot \bar{Q}_1 = P_1 - Q_1 + \bar{Q}_1, \quad \text{wobei}$$
$$\bar{Q}_1 \bar{Q}_2 = \bar{Q}_1 Q_1 = D + (P_1 - Q_1)^2 = Q_1 (Q_0 - 2P_1 + Q_1),$$
somit
$$\bar{P}_0 = \bar{P}_1 = P_1 - Q_1 + Q_0 - 2P_1 + Q_1$$
$$= Q_0 - P_0.$$

Umgekehrt entspricht daher die Wahl $\bar{P}_0 = Q_0 - P_0$ nur der anderen Möglichkeit für die Wahl des Kettenbruches (12) w.z.b.w.

Mit Satz 1 sind nun zwar grundsätzlich alle Gleichungen (1) erledigt bzw. auf ein fertiges Rechenschema zurückgeführt; doch wäre es für die Praxis erwünscht, bei zusammengesetztem

$L = cQ_0$, um allzu große Zahlen zu vermeiden, sich auf den durch Q_0 erzeugten Algorithmus zu beschränken, zum mindesten vielleicht für die kleineren zulässigen c. Wir fragen daher jetzt danach, was dieser (einerlei, ob (1) mit $L = Q_0$ lösbar ist oder nicht) für die Gleichung

$$x^2 + Dy^2 = Q_0 z \tag{17}$$

leistet.

Entsprechend wie früher gilt hier: wird für passendes n

$$Q_n = c, \text{ so befriedigt entsprechend (9), (9.1)}$$

$$x = cB_{n-2} + P_n B_{n-1}, \quad y = B_{n-1}, \quad z = c \tag{18}$$

die Gleichung (17). Dabei ist allerdings nicht notwendig $(x, y) = 1$. Umgekehrt beweisen wir

Satz 2. *Ist (a,b,c) mit $(a, b) = 1$, $b \geq 2$ eine Lösung von (17) (in natürlichen Zahlen), so gibt es eine Wahl von P_0 als Lösung von $(4)_0$ und von $b_\nu \in \Gamma$ ($b_0 = 0$, $b_\nu \geq 1$ für $1 \leq \nu \leq n-1$, $b_n \geq 0$) derart, daß der Algorithmus mit P_0, Q_0 vermöge (18) diese Lösung ergibt. Jetzt brauchen aber nicht alle Lösungen, auch nicht diejenigen mit kleinstem c, durch ein Verfahren mit der Vorschrift (10) hervorzugehen.*

Beweis. Wegen $(b, c) = 1$ ist die Kongruenz für P

$$a - Pb \equiv 0 \ (c) \text{ bei der Nebenbedingung}$$

$$0 \leq a - Pb < cb \tag{19}$$

eindeutig lösbar. Man setze jetzt $\mathfrak{Q}_n^* = \begin{pmatrix} c, & -P \\ -P, & \frac{D+P^2}{c} \end{pmatrix}$; dann ist auch $Q_{n-1}^* = \frac{D+P^2}{c} \in N$, was sofort aus

$$a^2 + Db^2 \equiv a^2 - P^2 b^2 \equiv 0 \ (c) \text{ und } (b, c) = 1$$

folgt. Ferner definiere man wieder $\mathfrak{Q}_0^*$ wie in (14), und schließlich die b_ν durch die Entwicklung

$$\frac{a-Pb}{bc} = [0, b_{n-1}, b_{n-2}, \ldots b_1], \ b_0 = 0. \text{ (Es wird } n \geq 2).$$

Dann ergibt sich entsprechend wie bei Satz 1, daß

$$b = B_{n-1}, \quad a = cB_{n-2} + P_n B_{n-1}, \text{ wie verlangt.}$$

Nach (19) muß jedoch diesmal auch ($P =$) $P_n < 0$ zugelassen werden, und dann ist nach (6.1) (15) nicht mehr erfüllt für $\nu = n - 1$.

Wollte man sich auf die Annahme (10) beschränken, was nach Satz 1 für $c = 1$ genügt, so können tatsächlich Lösungen entgehen, wie das Beispiel $Q_0 = 38$, $c = 2$, $D = 3$, $P_0 = 15$ (bzw. 23) zeigt, wo zwar $Q_n = 2 = c_{min}$ (entsprechend einer Abschätzung von Steuerwald [5]) auftritt, aber nur die Lösungen (7,3) und (8,2) hervorgehen (letztere für uns belanglos), aber nicht die Lösung (1,5), die im Sinne von Satz 2 aus $b_1 = 2$, $b_2 = b_3 = 1$, $P_4 = -1$ hervorgeht, und hier ist $b_3 = -\left[\frac{-P_3}{Q_3}\right]$.

In der erwähnten Arbeit gibt Herr Steuerwald obere Abschätzungen für $M = \min Q_n$ ($\geq m = \min z$) bei einem bestimmten Entwicklungsprinzip; im Falle (10) gilt $M \leq \frac{1+D}{2}$. Zeigte das vorhergehende Beispiel, daß auch bei $M = m$ das Verfahren nicht *alle* zugehörigen Lösungen zu liefern braucht, so erläutert das Beispiel $D = 5$, $Q_0 = 23$, $M = 3$, $m = 2$ den Fall $M > m$. Die Lösung (1, 3, 2) wird durch (10) nicht erhalten. Die Anwendung des Verfahrens mit Q_0 auf (17) kann daher im Gegensatz zu $c = 1$ theoretisch nicht voll befriedigen, solange man nicht weitere Einblicke gewonnen hat.

Schriftenverzeichnis

[1] Arwin A., Periodically closed chains of reduced fractions Ann. of Math. 24 (1923), 39–68.

[2] Patz W., Über die Gleichung $X^2 - DY^2 = \pm c\,(2^{31} - 1)$, wo c möglichst klein. Sitz.-Ber. Bay. Akad. Wiss., Math. Nat. Klasse 1948, 21–30.

[3] Perron O., Die Lehre von den Kettenbrüchen I, 3. Aufl. Stuttgart 1954.

[4] Schmidt Hermann, Zur Approximation und Kettenbruchentwicklung quadratischer Zahlen. Math. Z. 52 (1950), 168–192.

[5] Steuerwald R., Über die Gleichung $x^2 - Dy^2 = Cz$. Math. Z. 73 (1960), 382–385.