# Diophantische Gleichungen

**Wintersemester 2022/2023**

Universität Bayreuth

Michael Stoll

Inhaltsverzeichnis



Druckversion vom 9. Februar 2023, 12:41 Uhr.

Diese Vorlesung ist eine Vertiefungsvorlesung aus dem Bereich „Algebra/Zahlentheorie/Diskrete Mathematik“. Sie kann im sowohl im Bachelor- als auch im Masterstudiengang Mathematik gehört werden.

Einige Abschnitte in diesem Skript sind kleiner gedruckt. Dabei handelt es sich meistens um ergänzende Bemerkungen zur Vorlesung, die nicht zum eigentlichen Stoff gehören, die Sie aber vielleicht trotzdem interessant finden.

In der Bildschirmversion des Skripts finden sich hin und wieder Links wie dieser hier (er zeigt auf meine Homepage). Die meisten davon verweisen auf Wikipedia-Einträge von Mathematikern.

Für die Zwecke dieser Vorlesung ist Null eine natürliche Zahl:

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\};$$

gelegentlich werden wir die Schreibweise

$$\mathbb{N}_+ = \{1, 2, 3, \ldots\}$$

für die Menge der positiven natürlichen (oder ganzen) Zahlen verwenden. Meistens werden wir zur Vermeidung von Unklarheiten aber $\mathbb{Z}_{\geq 0}$ und $\mathbb{Z}_{>0}$ für diese Mengen schreiben. Wie üblich steht $\mathbb{Z}$ für den Ring der ganzen Zahlen, $\mathbb{Q}$ für den Körper der rationalen Zahlen, $\mathbb{R}$ für den Körper der reellen Zahlen und $\mathbb{C}$ für den Körper der komplexen Zahlen. Außerdem steht $A \subset B$ für die nicht notwendig strikte Inklusion ($A = B$ ist also erlaubt); für die strikte Inklusion schreiben wir $A \subsetneq B$. Wir schreiben $a \perp b$, um auszudrücken, dass die ganzen Zahlen $a$ und $b$ teilerfremd sind (dass also $\mathrm{ggT}(a, b) = 1$ ist).

## 1. Einleitung und Beispiele

Was sind „Diophantische Gleichungen“? Hier ist eine Definition.

1.1. **Definition.** Eine *diophantische Gleichung* ist eine algebraische Gleichung über $\mathbb{Z}$ (in mehreren Variablen), die in *ganzen* oder *rationalen* Zahlen gelöst werden soll. ♢

**DEF** Diophantische Gleichung

Eine *algebraische Gleichung über* $\mathbb{Z}$ ist dabei eine Gleichung der Form

$$F(x_1, x_2, \ldots, x_n) = 0\,,$$

wo $F \in \mathbb{Z}[x_1, x_2, \ldots, x_n]$ ein Polynom mit ganzzahligen Koeffizienten ist.

Das Wesentliche an dieser Definition ist nicht die Form der Gleichung, sondern die Tatsache, dass *ganze* oder *rationale* Lösungen gesucht werden. Was von beiden die interessante Frage ist, hängt vom jeweiligen Problem ab.

Man kann die Definition ausweiten auf *Systeme* von Gleichungen, die gleichzeitig erfüllt werden sollen. (Da für rationale Zahlen $x_1, \ldots, x_m$ die Gleichung $x_1^2 + \ldots + x_m^2 = 0$ nur die Lösung $x_1 = \ldots = x_m = 0$ hat, ist jedes System

$$F_1(x_1, \ldots, x_n) = \ldots = F_m(x_1, \ldots, x_n) = 0$$

äquivalent zu einer einzigen Gleichung

$$F_1(x_1, \ldots, x_n)^2 + \ldots + F_m(x_1, \ldots, x_n)^2 = 0\,;$$

das ist also keine wirkliche Verallgemeinerung.)

Manchmal betrachtet man auch Gleichungen, in denen ein oder mehrere Exponenten als (positive ganzzahlige) Variable auftreten. Solche Gleichungen heißen auch *exponentielle diophantische Gleichungen*.

Die Namensgebung erfolgte zu Ehren von *Diophant(os)* von Alexandria. Man weiß recht wenig über ihn selbst. Einigermaßen sicher kann man sein Schaffen zwischen 150 vor und 350 nach Christus datieren; Experten halten es für wahrscheinlich, dass es sich im 3. Jahrhundert n. Chr. abgespielt hat. Es gibt eine Rätsel-Aufgabe, die sich auf sein Alter bezieht und aus einer Sammlung stammt, die um 500 entstanden ist (Übersetzung von Norbert Schappacher, Lösung als Übungsaufgabe):

> Hier dies Grabmal deckt Diophantos' sterbliche Hülle,
> Und in des Trefflichen Kunst zeigt es sein Alter dir an.
> Knabe zu sein, gewährt' ihm der Gott ein Sechstel des Lebens,
> Und ein Zwölftel der Zeit ward er ein Jüngling genannt.
> Noch ein Siebentel schwand, da fand er des Lebens Gefährtin,
> Und fünf Jahre darauf ward ihm ein liebliches Kind.
> Halb nur hatte der Sohn des Vaters Alter vollendet,
> Als ihn plötzlich der Tod seinem Erzeuger entriss.
> Noch vier Jahre betrauert' er ihn im schmerzlichen Kummer.
> Und nun sage das Ziel, welches er selber erreicht!

Jedoch sind einige von seinen Schriften überliefert. Sein Hauptwerk ist die *Arithmetika*, von deren ursprünglich 13 Büchern sechs oder vielleicht auch zehn erhalten sind. Dort beschäftigt er sich mit der Lösung von Gleichungen in rationalen Zahlen; es ist die erste bekannte systematische Behandlung des Themas. Zu diesem Zweck führt er auch als einer der Ersten symbolische Bezeichnungen für eine Unbestimmte und ihre Potenzen ein.

Die erste brauchbare Übersetzung (ins Lateinische) und Kommentierung des griechischen Textes, die auch allgemein erhältlich war, wurde von Bachet im Jahr 1621 veröffentlicht. Fermat besaß ein Exemplar dieser Ausgabe und wurde dadurch zu eigenen Forschungen angeregt — der Beginn der neuzeitlichen Beschäftigung mit unserem Thema. In diesem Buch befand sich auch die berühmt-berüchtigte Randnotiz mit der Fermatschen Vermutung, die Fermats Sohn (mit den anderen Randbemerkungen) in seine Ausgabe der Arithmetika mit aufnahm.

Hier ist eine (recht willkürliche) Auswahl an Beispielen von diophantischen Gleichungen.

(1) $aX + bY = c$

Hier sind $a, b, c \in \mathbb{Z}$ gegeben und wir suchen nach Lösungen $X, Y \in \mathbb{Z}$ (rationale Lösungen gibt es immer, außer wenn $a = b = 0$ und $c \neq 0$ ist).

Diese einfache lineare Gleichung ist genau dann lösbar, wenn der ggT von $a$ und $b$ ein Teiler von $c$ ist. Ist $(x_0, y_0)$ eine Lösung, dann sind alle Lösungen von der Form $(x_0 + tb', y_0 - ta')$ mit $t \in \mathbb{Z}$, wobei $a' = a/\operatorname{ggT}(a, b)$ und $b' = b/\operatorname{ggT}(a, b)$.

(2) $X^2 + Y^2 = Z^2$

mit $X, Y, Z \in \mathbb{Z}$ (oder auch $\mathbb{Q}$, das macht keinen großen Unterschied). Diese Gleichung ist *homogen* (d.h., jeder Term hat denselben Gesamtgrad, hier 2); deswegen können wir Lösungen *skalieren*, d.h., alle Variablen mit einem Faktor durchmultiplizieren. Abgesehen von der *trivialen Lösung* $X = Y = Z = 0$, die einen hier nicht interessiert, sind dann alle (ganzzahligen oder rationalen) Lösungen Vielfache einer *primitiven* ganzzahligen Lösung, das ist eine Lösung mit $\operatorname{ggT}(X, Y, Z) = 1$.

Wir werden gleich sehen, dass man alle primitiven Lösungen in einer einfachen parametrischen Form beschreiben kann. Wegen des Zusammenhangs mit dem Satz von Pythagoras über rechtwinklige Dreiecke heißen Lösungen dieser Gleichung auch *pythagoreische Tripel*.

(3) $X^n + Y^n = Z^n$

Diese Gleichung ist wieder homogen, sodass man sich auf primitive ganzzahlige Lösungen beschränken kann. Das ist die berühmte *Fermatsche Gleichung:* Fermat behauptete in einer Bemerkung, die er an den Rand seines Exemplars von Bachets Diophant-Übersetzung schrieb, dass diese Gleichung für $n \geq 3$ keine Lösungen in positiven ganzen Zahlen habe (womit Lösungen wie $(1, 0, 1)$ ausgeschlossen sind). Er habe dafür einen „wundervollen Beweis“ entdeckt, der Rand sei aber zu klein dafür. Fermat hat die Aussage für $n = 4$ tatsächlich bewiesen (einen Beweis werden wir bald sehen), möglicherweise auch für $n = 3$. Experten sind sich ziemlich einig, dass Fermats „Beweis“ für den allgemeinen Fall keiner war und Fermat das auch schnell bemerkt hat: Er hat diese Behauptung (für $n \geq 5$) zum Beispiel nie in seinen Briefen an andere Mathematiker erwähnt.

Die Suche nach einem Beweis dieser „Fermatschen Vermutung“ hat in den nachfolgenden Jahrhunderten einen umfangreichen Schatz an mathematischen Entwicklungen hervorgebracht, bis hin zu Wiles' Beweis der Modularitätsvermutung für Elliptische Kurven. Leider gibt es immer noch viele eher unbedarfte Amateure, die glauben, Fermats ursprünglichen (falschen!) Beweis gefunden zu haben...

(4) $X_1^2 + X_2^2 + X_3^2 + X_4^2 = m$

Hier ist $m \in \mathbb{Z}_{\geq 0}$ gegeben, und wir suchen ganzzahlige Lösungen. D.h., wir fragen, welche natürlichen Zahlen man als Summe von vier Quadraten schreiben kann.

Diophant ahnte, Fermat wusste und Lagrange bewies, dass das immer möglich ist. Wir werden später einen Beweis davon sehen.

(5) $X^2 - 409Y^2 = 1$

Wir fragen nach nichttrivialen ($Y \neq 0$) ganzzahligen Lösungen. Gleichungen dieser Form (wo statt 409 eine beliebige positive ganze Zahl stehen kann, die kein Quadrat ist) nennt man *Pellsche Gleichungen*. Die Bezeichnung geht auf Euler zurück und beruht auf einer Verwechslung — Pell hatte mit dieser Art von Gleichungen absolut nichts zu tun.

Wir werden später sehen, dass es stets Lösungen gibt, und dass sie sich aus einer „Fundamentallösung“ erzeugen lassen. Für die Beispielgleichung ist die kleinste Lösung

$$X = 25\,052\,977\,273\,092\,427\,986\,049\,, \qquad Y = 1\,238\,789\,998\,647\,218\,582\,160\,.$$

(6) $X^2 + Y^2 = U^2, \quad X^2 + Z^2 = V^2, \quad Y^2 + Z^2 = W^2, \quad X^2 + Y^2 + Z^2 = T^2$

Wir suchen nach nichttrivialen (oBdA positiven) rationalen Lösungen. Das ist ein Beispiel für ein *System* diophantischer Gleichungen. Es beschreibt einen Quader, dessen Seiten $(X, Y, Z)$, Flächen- $(U, V, W)$ und Raumdiagonalen $(T)$ sämtlich rationale Länge haben. Es ist keine Lösung bekannt, aber auch nicht bewiesen, dass es keine gibt: ein offenes Problem! Wenn man eine der Bedingungen weglässt (also eine Seite, eine Flächendiagonale oder die Raumdiagonale irrationale Länge haben darf), dann gibt es Lösungen.

(7) $Y^2 = X^3 + 7823$

Hier interessieren uns rationale Lösungen (ganzzahlige gibt es nicht). Die Gleichung beschreibt eine *Elliptische Kurve*; solche Kurven haben allgemeiner Gleichungen der Form $Y^2 = X^3 + aX + b$. Dazu gibt es eine extrem reichhaltige Theorie (mit der man mehrere Jahre an Vorlesungen füllen kann). Unter anderem spielen sie eine wesentliche Rolle im Beweis der Fermatschen Vermutung.

In unserem Fall kann man zeigen, dass alle Lösungen von einer Grundlösung erzeugt werden; sie lautet

$$X = \frac{22635821433214215021002092335177 77}{11981673410095561^2}$$

$$Y = \frac{18639815258462330562483755148559677002814477665575 6}{11981673410095561^3}$$

und wurde von mir im Jahr 2002 entdeckt.

(8) $X^2 + Y^3 = Z^7$

Das ist eine *verallgemeinerte Fermatsche Gleichung*. Aus nicht ganz so offensichtlichen, aber sehr guten Gründen fragt man auch hier nach *primitiven*, also teilerfremden, ganzzahligen Lösungen.

Betrachtet man allgemeiner $X^p + Y^q = Z^r$, dann ist bekannt, dass es unendlich viele Lösungen gibt (die jedoch in endlich viele parametrisierte Familien zerfallen), falls $\chi := 1/p + 1/q + 1/r > 1$ ist, und endlich viele, falls $\chi \leq 1$ ist

(der Fall $\chi = 1$ ist dabei klassisch und wurde von Fermat und Euler erledigt, siehe Abschnitt 2 für den Fall $(p, q, r) = (4, 4, 2)$).

Für die Beispielgleichung habe ich zusammen mit zwei Kollegen gezeigt, dass die Liste der bekannten Lösungen

$$(\pm 1, -1, 0), \quad (\pm 1, 0, 1), \quad \pm(0, 1, 1), \quad (\pm 3, -2, 1), \quad (\pm 71, -17, 2),$$
$$(\pm 2213459, 1414, 65), \quad (\pm 15312283, 9262, 113), \quad (\pm 21063928, -76271, 17)$$

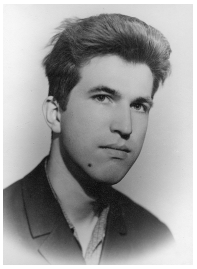
vollständig ist.[1] Wenn eine diophantische Gleichung nur endlich viele Lösungen hat, sind diese oft relativ leicht zu finden. Die Schwierigkeit besteht darin zu beweisen, dass es keine anderen gibt!

D. Hilbert
1862–1943

(9) $\binom{Y}{2} = \binom{X}{5}$ (oder $60Y(Y-1) = X(X-1)(X-2)(X-3)(X-4)$)

Wir suchen ganzzahlige Lösungen. Solche Gleichungen lassen sich im Prinzip lösen, und inzwischen gibt es sogar praktikable Algorithmen.[2] Die Lösungen der Beispielgleichung mit $X > 4$ sind

$$(5, -1),\ (5, 2), \quad (6, -3),\ (6, 4), \quad (7, -6),\ (7, 7),$$
$$(15, -77),\ (15, 78), \quad (19, -152),\ (19, 153)\,.$$



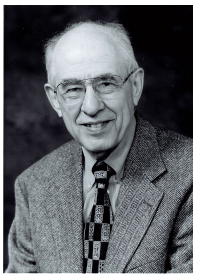
Y. Matiyasevich
1947–
©Y. Matiyasevich

Wieder ist die Schwierigkeit der Beweis, dass dies alle Lösungen sind.

(10) $X^2 + 7 = 2^n$

Wir suchen Lösungen mit $X \in \mathbb{Z}$ und $n \in \mathbb{Z}_{\geq 0}$. Diese Gleichung wurde von Ramanujan vorgeschlagen und von Nagell zuerst vollständig gelöst. Dies ist ein Beispiel einer Gleichung mit einem variablen Exponenten.

Ihre Lösungen sind gegeben durch $n \in \{3, 4, 5, 7, 15\}$.



H. Putnam
1926–2016
©H. Putnam
unverändert, Lizenz

Bevor wir gleich zu ein paar klassischen Beweisen kommen, möchte ich erst noch ein negatives Resultat erwähnen, das uns sagt, dass wir nicht zu viel erwarten können.

Der berühmte Mathematiker David Hilbert hatte in einem berühmten Vortrag auf dem Internationalen Mathematikerkongress in Jahr 1900 eine Liste von 23 Problemen genannt, deren Lösung seiner Meinung nach die Mathematik im 20. Jahrhundert voran bringen sollte. Eines davon, das *Zehnte Hilbert-Problem*, fragte nach einem Verfahren (heute würde man sagen, Algorithmus), das für jedes gegebene Polynom $F \in \mathbb{Z}[X_1, \ldots, X_n]$ Auskunft darüber gibt, ob die Gleichung $F(X_1, \ldots, X_n) = 0$ ganzzahlige Lösungen hat oder nicht. Es dauerte bis 1970, als schließlich Yuri Matiyasevich, aufbauend auf wesentlicher Vorarbeit von Putnam, Davis und Julia Robinson, beweisen konnte, dass ein solcher Algorithmus **nicht existiert**.[3]

M. Davis
1928–
©G. Bergman
Lizenz

So ein Beweis war erst möglich, nachdem der Begriff der Berechenbarkeit theoretisch gefasst und ausreichend verstanden war. Die Beweisidee ist, sehr knapp gefasst, wie folgt. Zunächst einmal sieht man leicht (Übung), dass die gegebene

J. Robinson
1919–1985
©G. Bergman
Lizenz

[1] B. Poonen, E.F. Schaefer, M. Stoll: *Twists of $X(7)$ and primitive solutions to $x^2 + y^3 = z^7$*, Duke Math. J. **137** (2007), 103–158.

[2] Y. Bugeaud, M. Mignotte, S. Siksek, M. Stoll, Sz. Tengely: *Integral points on hyperelliptic curves*, Algebra & Number Theory **2**, No. 8 (2008), 859–885.

[3] Yuri V. Matiyasevich: *Hilbert's tenth problem*, Foundations of Computing Series, MIT Press, Cambridge, MA, 1993.

Formulierung des Problems äquivalent zu einer Formulierung ist, in der ganzzahlige Lösungen durch Lösungen in natürlichen Zahlen (also in $\mathbb{N} = \mathbb{Z}_{\geq 0}$) ersetzt werden. Eine *diophantische Menge* $D$ ist gegeben durch

$$D = \{a \in \mathbb{N} \mid \exists x_1, \ldots, x_n \in \mathbb{N} \colon F(a, x_1, \ldots, x_n) = 0\}$$

für ein geeignetes Polynom $F \in \mathbb{Z}[x_0, x_1, \ldots, x_n]$. Eine *rekursiv aufzählbare Menge* ist die Menge aller $a \in \mathbb{N}$, für die ein geeigneter Algorithmus bei Eingabe von $a$ schließlich anhält. Es ist leicht zu sehen (Übung), dass jede diophantische Menge auch rekursiv aufzählbar ist. Was Matiyasevich eigentlich beweist, ist, dass die Umkehrung ebenfalls gilt: Jede rekursiv aufzählbare Menge ist diophantisch. Da man in der Logik zeigen kann, dass es rekursiv aufzählbare Mengen gibt, die nicht entscheidbar sind (d.h., es gibt keinen Algorithmus, der für ein gegebenes $a \in \mathbb{N}$ entscheidet, ob $a$ in der Menge liegt oder nicht), folgt, dass es ein Polynom $F$ wie oben gibt, sodass man für gegebenes $a \in \mathbb{N}$ nicht entscheiden kann, ob $F(a, x_1, \ldots, x_n) = 0$ eine Lösung in natürlichen (oder, für ein anderes $F$, in ganzen) Zahlen hat.

Man kann sich fragen, wie „kompliziert“ die Polynome $F$ sein müssen, sodass die Lösbarkeit nicht entscheidbar ist. Eine Möglichkeit, die Komplexität zu messen, ist der Grad. Hier ist es so, dass für quadratische Gleichungen (Grad 2) Entscheidungsverfahren existieren. Polynome vom Grad 4 sind hingegen kompliziert genug; für Grad 3 ist die Frage offen. Ein anderes Maß ist die Anzahl der Variablen. Hier ist bekannt, dass neun Unbestimmte für die Unentscheidbarkeit ausreichen. Die Lösbarkeit von Polynomen in einer Variablen lässt sich leicht entscheiden. Für zwei Variable ist die Frage offen; es gibt aber gute Gründe für die Annahme, dass man hier Entscheidbarkeit hat.

Eine Variation des Problems fragt nach einem Entscheidungsverfahren für die Lösbarkeit in $\mathbb{Q}$ (statt in $\mathbb{Z}$). Diese Frage ist offen; allerdings ist die allgemeine Überzeugung, dass es auch hier kein allgemeines Entscheidungsverfahren gibt.

## 2. Appetithappen

Bevor wir uns dem systematischen Studium einiger Arten von diophantischen Gleichungen zuwenden, möchte ich Ihnen vollständige Lösungen von zwei solchen Gleichungen vorführen. Die erste ist die Gleichung der pythagoreischen Tripel,

$$X^2 + Y^2 = Z^2 .$$

Wir wollen ihre primitiven ganzzahligen Lösungen bestimmen.

Zuerst überlegen wir uns, welche der Variablen gerade bzw. ungerade Werte annehmen können. Zunächst einmal können nicht alle gerade sein, denn wir suchen nach teilerfremden Lösungen. Damit die Gleichung „modulo 2“ aufgeht, müssen dann zwei der Variablen ungerade sein, die andere gerade. Es ist jedoch nicht möglich, dass $X$ und $Y$ beide ungerade sind, denn dann ist die linke Seite durch 2, aber nicht durch 4 teilbar ($X^2$ und $Y^2$ lassen beide den Rest 1 bei Division durch 4), kann also kein Quadrat sein. Also ist jedenfalls $Z$ ungerade, und wir können (bis auf eventuelles Vertauschen von $X$ und $Y$) annehmen, dass $X$ gerade und $Y$ ungerade ist.

Für den nächsten Schritt brauchen wir ein Hilfsresultat.

2.1. **Lemma.** *Sind $a, b, c$ ganze Zahlen mit $a$ und $b$ teilerfremd und sodass $ab = c^2$ ist, dann gibt es (teilerfremde) ganze Zahlen $u$ und $v$, sodass entweder*

**LEMMA** $ab = c^2$

$$a = u^2 , \qquad b = v^2 \quad \textit{and} \quad c = uv$$

*oder*

$$a = -u^2 , \qquad b = -v^2 \quad \textit{and} \quad c = uv$$

*gilt.*

*Beweis.* Wir nehmen erst einmal an, dass $c \neq 0$ ist. Wir betrachten die Primfaktorzerlegungen von $a$ und $b$. Sei $p$ eine Primzahl, die (z.B.) $a$ teilt. Da $a$ und $b$ teilerfremd sind, kann $p$ dann nicht auch $b$ teilen. $p$ kommt also genau so oft in $a$ vor, wie in der rechten Seite $c^2$, also ist der Exponent von $p$ in $a$ gerade. Da jede Primzahl mit einem geraden Exponenten in $a$ vorkommt, gibt es $u \in \mathbb{Z}$, sodass $a = \pm u^2$ ist. Ebenso gibt es $v \in \mathbb{Z}$ mit $b = \pm v^2$. Da $ab = c^2 > 0$ ist, müssen die Vorzeichen gleich sein. Außerdem folgt $c = \pm uv$, und wir können falls nötig (z.B.) das Vorzeichen von $u$ ändern, um $c = uv$ zu erhalten.

Wenn $c = 0$ ist, dann ist $a = \pm 1$, $b = 0$, oder umgekehrt, und das Ergebnis stimmt ebenfalls (mit $u = 1$, $v = 0$, oder umgekehrt). ❑

Wir schreiben nun unsere Gleichung in der Form

$$X^2 = Z^2 - Y^2 = (Z - Y)(Z + Y) .$$

Beide Faktoren auf der rechten Seite sind gerade, also gibt es $U, V \in \mathbb{Z}$ mit $2U = Z - Y$ und $2V = Z + Y$. Außerdem können wir $X = 2W$ setzen (denn $X$ ist gerade). Jeder gemeinsame Teiler von $U$ und $V$ muss auch ein gemeinsamer Teiler von $Y = V - U$ und $Z = V + U$ sein, und damit wäre er auch ein Teiler von $X$. Nachdem wir voraussetzen, dass $X, Y, Z$ teilerfremd sind, folgt, dass $U$ und $V$ teilerfremd sind.

Nach dem Lemma gibt es also $S, T \in \mathbb{Z}$ mit

$$U = S^2, \quad V = T^2, \quad W = ST \qquad \text{oder} \qquad U = -S^2, \quad V = -T^2, \quad W = ST .$$

Im ersten Fall ergibt sich

$$X = 2ST, \quad Y = T^2 - S^2, \quad Z = T^2 + S^2\,,$$

im zweiten Fall

$$X = 2TS, \quad Y = S^2 - T^2, \quad Z = -(S^2 + T^2)\,.$$

Beachte noch, dass $S$ und $T$ teilerfremd sind und verschiedene Parität haben (d.h., eines ist gerade, das andere ungerade), denn $S^2+T^2 = \pm Z$ ist ungerade. Wir haben bewiesen:

**SATZ** pythagoreische Tripel

2.2. **Satz.** *Die primitiven pythagoreischen Tripel mit $X$ gerade und $Z > 0$ haben die Form*

$$X = 2ST, \quad Y = S^2 - T^2, \quad Z = S^2 + T^2$$

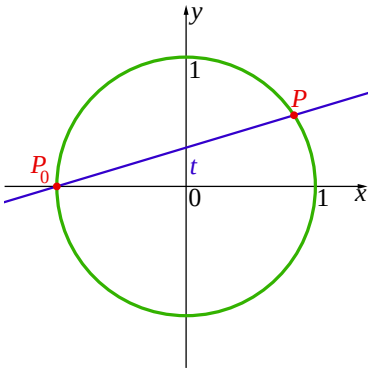
*mit $S, T \in \mathbb{Z}$ teilerfremd und von verschiedener Parität.*

Es ist klar, dass jedes solche Tripel tatsächlich ein primitives pythagoreisches Tripel ist.

Ich werde jetzt für diesen Satz noch einen zweiten „geometrischen“ Beweis (im Gegensatz zum gerade gegebenen „algebraischen“ Beweis) geben. Dazu stellen wir erst einmal fest, dass eine nichttriviale Lösung stets $Z \neq 0$ hat. Wir können die Gleichung also durch $Z^2$ teilen und erhalten

$$x^2 + y^2 = 1 \qquad \text{mit } x = X/Z \text{ und } y = Y/Z.$$

Wir wollen jetzt die *rationalen* Lösungen dieser Gleichung bestimmen; daraus ergeben sich die primitiven Lösungen der ursprünglichen Gleichung (mit $Z > 0$) durch Multiplikation mit dem Hauptnenner $Z$ von $x$ und $y$.




Die Geometrie kommt dadurch ins Spiel, dass wir uns die reellen Lösungen von $x^2 + y^2 = 1$ durch die Punkte des Einheitskreises veranschaulichen können. Die rationalen Lösungen entsprechen dann den Punkten mit rationalen Koordinaten, den so genannten *rationalen Punkten* des Einheitskreises. Vier davon sind offensichtlich, nämlich $(x, y) = (\pm 1, 0)$ und $(x, y) = (0, \pm 1)$. Sei $P_0 = (-1, 0)$ einer davon. Wenn $P = (x, y) \neq P_0$ ein weiterer rationaler Punkt ist, dann hat die Gerade durch $P_0$ und $P$ rationale Steigung $t = \frac{y}{x+1}$. Wir bekommen also alle rationalen Punkte $\neq P_0$, indem wir Geraden mit rationaler Steigung durch $P_0$ legen und den zweiten Schnittpunkt mit dem Einheitskreis betrachten. Dieser zweite Schnittpunkt ist tatsächlich stets rational, was daran liegt, dass er durch eine quadratische Gleichung mit rationalen Koeffizienten bestimmt ist, deren andere Lösung ebenfalls rational ist:

Die Gleichung der Geraden durch $P_0$ mit Steigung $t$ ist

$$y = t(x + 1)\,.$$

Wir setzen die rechte Seite für $y$ in die Kreisgleichung ein:

$$0 = x^2 + y^2 - 1 = x^2 - 1 + t^2(x + 1)^2 = (x + 1)\big(x - 1 + t^2(x + 1)\big)\,.$$

Für den zweiten Schnittpunkt ist $x \neq -1$, also bekommen wir

$$x = \frac{1 - t^2}{1 + t^2}\,, \qquad y = t(x + 1) = \frac{2t}{1 + t^2}\,.$$

Diese „rationale Parametrisierung des Einheitskreises“ liefert alle rationalen Lösungen von $x^2 + y^2 = 1$ mit Ausnahme von $P_0$. Wir können uns $P_0$ als durch den Grenzwert für $t \to \infty$ gegeben vorstellen; tatsächlich bräuchten wir in unserer Konstruktion eine Gerade, die den Kreis in $P_0$ „zweimal“ schneidet, also die Tangente. Diese Tangente in $P_0$ ist aber senkrecht, hat also Steigung $\infty$.

Um zu primitiven Lösungen von $X^2 + Y^2 = Z^2$ zurückzukommen, müssen wir unsere Ausdrücke für $x$ und $y$ als gekürzte Brüche schreiben. Dazu schreiben wir zunächst $t = U/V$ als gekürzten Bruch, das liefert dann

$$x = \frac{V^2 - U^2}{V^2 + U^2}\,, \qquad y = \frac{2UV}{V^2 + U^2}\,.$$

(Hier ist jetzt $P_0$ wieder enthalten, wenn wir $U = 1$, $V = 0$ erlauben.) Dabei ist der Bruch für $x$ gekürzt, falls $U$ und $V$ verschiedene Parität haben. Im anderen Fall ($U$ und $V$ beide ungerade) haben Zähler und Nenner den ggT 2, und das gilt dann auch für den Bruch für $y$. Im ersten Fall erhalten wir also

$$X = V^2 - U^2\,, \quad Y = 2UV\,, \quad Z = V^2 + U^2$$

mit $U$ und $V$ teilerfremd und von verschiedener Parität. Im zweiten Fall schreiben wir $V + U = 2R$, $V - U = 2S$ mit ganzen Zahlen $R$ und $S$; dann sind

$$x = \frac{2RS}{R^2 + S^2}\,, \qquad y = \frac{R^2 - S^2}{R^2 + S^2}$$

gekürzte Brüche, und wir erhalten das primitive pythagoreische Tripel

$$X = 2RS\,, \qquad Y = R^2 - S^2\,, \qquad Z = R^2 + S^2\,.$$

Wir erhalten also wieder Satz 2.2, diesmal beide Versionen ($X$ gerade bzw. $Y$ gerade).

Unser Vorgehen hier ist übrigens recht nahe an dem, was Diophant macht (allerdings rein algebraisch): Wir reduzieren den Grad der Gleichung, sodass sie linear wird.


P. de Fermat
1607–1665

Die rationale Parametrisierung des Einheitskreises hat noch andere Anwendungen. Sie drückt $\sin\alpha$ und $\cos\alpha$ rational durch $t = \tan\frac{\alpha}{2}$ aus und kann beispielsweise benutzt werden, um Integrale über rationale Ausdrücke in $\sin x$ und $\cos x$ in Integrale über rationale Funktionen in $t$ umzuformen, die sich dann berechnen lassen.

Als weiteren Appetithappen möchte ich jetzt den Beweis von Fermat vorführen, dass

$$X^4 + Y^4 = Z^2$$

keine ganzzahligen Lösungen mit $X, Y, Z \neq 0$ hat. Daraus folgt natürlich sofort, dass auch

$$X^4 + Y^4 = Z^4$$

nicht in positiven ganzen Zahlen lösbar ist.

Wir stellen zunächst fest, dass wir nur Lösungen mit paarweise teilerfremden $X$, $Y$, $Z$ berücksichtigen müssen. Denn ist etwa $p$ eine Primzahl, die zwei der Variablen teilt, dann teilt $p$ auch die dritte, und wir erhalten eine kleinere Lösung, indem wir $X$ durch $X/p$, $Y$ durch $Y/p$ und $Z$ durch $Z/p^2$ ersetzen ($Z$ muss durch $p^2$ teilbar sein, denn beide Seiten der Gleichung sind durch $p^4$ teilbar). Wir können so fortfahren, bis $X$, $Y$ und $Z$ paarweise teilerfremd sind.

Die geniale Idee von Fermat war, ausgehend von einer (primitiven) Lösung mit $X, Y, Z > 0$ eine weitere kleinere (d.h. mit kleinerem $Z$) solche Lösung zu konstruieren. Da es keine unendlichen absteigenden Folgen natürlicher Zahlen gibt, führt

das auf einen Widerspruch. Fermat nannte dieses Prinzip den *unendlichen Abstieg* (descente infinie).

Sei also $(X, Y, Z)$ eine primitive Lösung mit $X, Y, Z > 0$. Damit bilden $X^2$, $Y^2$ und $Z$ ein primitives pythagoreisches Tripel. Wir können annehmen, dass $X$ gerade ist; dann gibt es nach Satz 2.2 teilerfremde ganze Zahlen $R$ und $S$ von verschiedener Parität, sodass

$$X^2 = 2RS\,, \qquad Y^2 = R^2 - S^2\,, \qquad Z = R^2 + S^2\,.$$

Wir können annehmen, dass $R$ und $S$ positiv sind. Da $Y$ ungerade ist, folgt aus der zweiten Gleichung, dass $S$ gerade sein muss. Wir schreiben $S = 2T$ und $X = 2W$, dann erhalten wir

$$W^2 = RT$$

mit $R$ und $T$ teilerfremd. Nach Lemma 2.1 gibt es $U, V > 0$ und teilerfremd, so dass

$$R = U^2\,, \qquad T = V^2\,,$$

also $S = 2V^2$ und damit

$$Y^2 = U^4 - 4V^4$$

ist. Außerdem ist $U \leq U^2 = R \leq R^2 < Z$.

Wir sehen, dass $Y$, $2V^2$ und $U^2$ ebenfalls ein primitives pythagoreisches Tripel bilden. Es gibt demnach teilerfremde ganze Zahlen $P, Q > 0$ mit

$$Y = P^2 - Q^2\,, \qquad 2V^2 = 2PQ\,, \qquad U^2 = P^2 + Q^2\,.$$

Auf die mittlere Gleichung können wir wiederum Lemma 2.1 anwenden und finden teilerfremde ganze Zahlen $A, B > 0$, sodass

$$P = A^2 \quad \text{und} \quad Q = B^2\,.$$

In die dritte Gleichung eingesetzt, erhalten wir

$$A^4 + B^4 = U^2\,.$$

Also ist $(A, B, U)$ eine weitere primitive Lösung von $X^4 + Y^4 = Z^2$ mit $A, B, U > 0$ und $U < Z$. Es folgt:

2.3. **Satz.** *Die einzigen primitiven ganzzahligen Lösungen von*

$$X^4 + Y^4 = Z^2$$

*sind gegeben durch* $X = 0$, $Y = \pm 1$, $Z = \pm 1$ *und* $X = \pm 1$, $Y = 0$, $Z = \pm 1$.

**SATZ**
$Y^4 + Y^4 = Z^2$

Während des Beweises haben wir gesehen, dass eine nichttriviale Lösung von $Y^2 = U^4 - 4V^4$ eine nichttriviale Lösung von $X^4 + Y^4 = Z^2$ ergibt. Wir haben also auch die folgende Aussage gezeigt.

2.4. **Satz.** *Die einzigen primitiven ganzzahligen Lösungen von*

$$X^4 - 4Y^4 = Z^2$$

*sind gegeben durch* $X = \pm 1$, $Y = 0$, $Z = \pm 1$.

**SATZ**
$Y^4 - 4Y^4 = Z^2$

## 3. Das Quadratische Reziprozitätsgesetz

Der Chinesische Restsatz und der Euklidische Algorithmus erlauben uns, lineare Kongruenzen oder Systeme von Kongruenzen zu lösen. Der nächste Schritt führt uns zu quadratischen Kongruenzen.

**DEF** quadratischer (Nicht)Rest

3.1. **Definition.** Seien $p$ eine ungerade Primzahl und $a \in \mathbb{Z}$ kein Vielfaches von $p$. Dann heißt $a$ ein *quadratischer Rest mod p*, wenn die Kongruenz $x^2 \equiv a \bmod p$ Lösungen hat. Andernfalls heißt $a$ *quadratischer Nichtrest mod p*. ♢

**BSP** quadratische (Nicht)Reste

3.2. **Beispiel.**

| $p$ | qu. Reste | qu. Nichtreste |
|---|---|---|
| 3 | 1 | 2 |
| 5 | 1, 4 | 2, 3 |
| 7 | 1, 2, 4 | 3, 5, 6 |
| 11 | 1, 3, 4, 5, 9 | 2, 6, 7, 8, 10 |

♣

Sei $g$ eine Primitivwurzel mod $p$ (d.h., die Restklasse $\bar{g} \in \mathbb{F}_p$ von $g$ erzeugt die multiplikative Gruppe $\mathbb{F}_p^\times$); dann ist jedes $a$ mit $p \nmid a$ kongruent zu $g^k \bmod p$ für ein $k \in \mathbb{Z}$, das mod $p-1$ eindeutig bestimmt ist; insbesondere ist die Parität von $k$ eindeutig bestimmt, da $p-1$ gerade ist. Wir schreiben $k = \log_{\bar{g}} \bar{a} \in \mathbb{Z}/(p-1)\mathbb{Z}$.

**SATZ** Euler-Kriterium

3.3. **Satz.** *Seien $p$ eine ungerade Primzahl und $a \in \mathbb{Z}$ mit $p \nmid a$; sei weiter $g$ eine Primitivwurzel mod $p$. Dann sind die folgenden Aussagen äquivalent:*

(1) *$a$ ist quadratischer Rest mod $p$.*

(2) *$\log_{\bar{g}} \bar{a}$ ist gerade.*

(3) *$a^{(p-1)/2} \equiv 1 \bmod p$ (Euler-Kriterium).*

*Beweis.* Sei $a \equiv g^k \bmod p$ mit $k = \log_{\bar{g}} \bar{a}$. Ist $k = 2l$ gerade, dann ist $a \equiv x^2 \bmod p$ für $x = g^l$, also ist $a$ quadratischer Rest mod $p$. Wenn $a$ quadratischer Rest ist, also $a \equiv x^2 \bmod p$ für ein $x \in \mathbb{Z}$, dann ist $a^{(p-1)/2} \equiv x^{p-1} \equiv 1 \bmod p$ nach dem kleinen Satz von Fermat. Ist schließlich $a^{(p-1)/2} \equiv 1 \bmod p$, dann haben wir $g^{k(p-1)/2} \equiv 1 \bmod p$, und da $g$ eine Primitivwurzel ist, bedeutet das, dass $p-1$ den Exponenten $k(p-1)/2$ teilt, woraus wiederum folgt, dass $k$ gerade ist.

Wir haben also (2) ⇒ (1) ⇒ (3) ⇒ (2) und damit die Äquivalenz der drei Aussagen gezeigt. ❑

Da offenbar die Hälfte der möglichen Logarithmen $k$ gerade und die andere Hälfte ungerade ist, sehen wir, dass es genau $(p-1)/2$ quadratische Restklassen und $(p-1)/2$ quadratische Nichtrestklassen mod $p$ gibt.

**FOLG** Kriterium für Nichtrest

3.4. **Folgerung.** *Mit den Notationen von Satz 3.3 gilt*

$$a \text{ quad. Nichtrest mod } p \iff \log_{\bar{g}} \bar{a} \text{ ist ungerade} \iff a^{(p-1)/2} \equiv -1 \bmod p\,.$$

*Beweis.* Es ist nur noch zu zeigen, dass $a^{(p-1)/2} \equiv \pm 1 \bmod p$ ist (falls $p \nmid a$). Sei $b = a^{(p-1)/2}$. Dann ist $b^2 = a^{p-1} \equiv 1 \bmod p$, also $(\bar{b} - \bar{1})(\bar{b} + \bar{1}) = \bar{0}$ im Körper $\mathbb{F}_p$. Es muss also einer der Faktoren verschwinden, und damit ist $b \equiv 1$ oder $b \equiv -1 \bmod p$. ❑

3.5. **Definition.** Wir definieren das *Legendre-Symbol* für eine ungerade Primzahl $p$ und $a \in \mathbb{Z}$ durch

**DEF** Legendre-Symbol

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{wenn } p \nmid a \text{ und } a \text{ quadratischer Rest mod } p \text{ ist,} \\ -1 & \text{wenn } p \nmid a \text{ und } a \text{ quadratischer Nichtrest mod } p \text{ ist,} \\ 0 & \text{wenn } p \mid a. \end{cases}$$

Es gilt dann $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$, falls $a \equiv b \bmod p$. ♢

Aus dem Euler-Kriterium folgt, dass sich $\left(\frac{a}{p}\right)$ effizient berechnen lässt: Die Potenz $\bar{a}^{(p-1)/2} \in \mathbb{F}_p$ lässt sich mit $O\big((\log p)^3\big)$ Bitoperationen berechnen ($O(\log p)$ Multiplikationen in $\mathbb{F}_p$, um die Potenz durch sukzessives Quadrieren zu berechnen; eine Multiplikation lässt sich in $O\big((\log p)^2\big)$ Bitoperationen oder schneller erledigen).

Es ist eine ganz andere Sache, tatsächlich eine Quadratwurzel von $a$ mod $p$ zu *finden* (also $x \in \mathbb{Z}$ mit $x^2 \equiv a \bmod p$), wenn $a$ ein quadratischer Rest mod $p$ ist. Es gibt probabilistische Algorithmen, die polynomiale erwartete Laufzeit haben, aber keinen effizienten deterministischen Algorithmus.[4]

3.6. **Folgerung.** *Die Anzahl der Lösungen von $X^2 = \bar{a}$ in $\mathbb{F}_p$ ist genau $1 + \left(\frac{a}{p}\right)$.*

**FOLG** Anzahl Quadratwurzeln von $\bar{a}$

*Beweis.* Ist $\bar{a} = 0$, dann gibt es genau eine Lösung $X = 0$, und $\left(\frac{a}{p}\right) = 0$.

Ist $\bar{a} \neq 0$ ein Quadrat in $\mathbb{F}_p$, dann gibt es genau zwei Lösungen (die sich nur um das Vorzeichen unterscheiden), und $\left(\frac{a}{p}\right) = 1$.

Ist $\bar{a}$ kein Quadrat in $\mathbb{F}_p$, dann gibt es keine Lösung, und $\left(\frac{a}{p}\right) = -1$. ❑

3.7. **Folgerung.** *Sei $p$ eine ungerade Primzahl, $a \in \mathbb{Z}$. Dann gilt*

**FOLG** Legendre durch Euler

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \bmod p\,,$$

*und der Wert des Legendre-Symbols ist durch diese Kongruenz eindeutig bestimmt.*

*Beweis.* Gilt $p \mid a$, dann sind beide Seiten null mod $p$. In den anderen beiden Fällen folgt die Kongruenz aus Satz 3.3 und Folgerung 3.4. Die Eindeutigkeitsaussage folgt daraus, dass das Legendre-Symbol nur die Werte $-1, 0, 1$ annimmt, die mod $p$ alle verschieden sind (denn $p \geq 3$). ❑

3.8. **Satz.** *Seien $p$ eine ungerade Primzahl und $a, b \in \mathbb{Z}$. Dann gilt*

**SATZ** Legendre-Symbol ist multiplikativ

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

*Beweis.* Nach Folgerung 3.7 gilt

$$\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \equiv a^{(p-1)/2} b^{(p-1)/2} = (ab)^{(p-1)/2} \equiv \left(\frac{ab}{p}\right) \bmod p\,.$$

Wie oben folgt Gleichheit, da die möglichen Werte $-1, 0, 1$ der linken und rechten Seite mod $p$ verschieden sind. ❑

[4] http://en.wikipedia.org/wiki/Quadratic_residue#Complexity_of_finding_square_roots

Insbesondere ist das Produkt von zwei quadratischen Nichtresten mod $p$ ein quadratischer Rest mod $p$.

Die wesentliche Aussage von Satz 3.8 lässt sich auch folgendermaßen ausdrücken: Die Abbildung

$$\mathbb{F}_p^\times \longrightarrow \{\pm 1\}\,, \qquad \bar{a} \longmapsto \left(\frac{a}{p}\right)$$

ist ein Gruppenhomomorphismus. Da es stets quadratische Nichtreste gibt (z.B. ist jede Primitivwurzel mod $p$ ein quadratischer Nichtrest), ist dieser Homomorphismus surjektiv; sein Kern besteht gerade aus den Quadraten in $\mathbb{F}_p^\times$.

**BSP** Legendre-Symbol und Primfaktorzerlegung

3.9. **Beispiel.** Wir können $\left(\frac{a}{p}\right)$ mit Hilfe der Primfaktorzerlegung von $a$ faktorisieren. Sei $a = \pm 2^e q_1^{f_1} q_2^{f_2} \dots q_k^{f_k}$ mit paarweise verschiedenen ungeraden Primzahlen $q_j$. Dann ergibt sich

$$\left(\frac{a}{p}\right) = \left(\frac{\pm 1}{p}\right)\left(\frac{2}{p}\right)^e \left(\frac{q_1}{p}\right)^{f_1} \left(\frac{q_2}{p}\right)^{f_2} \dots \left(\frac{q_k}{p}\right)^{f_k}. \qquad \clubsuit$$

Im Folgenden werden wir uns der Frage zuwenden, wie man die verschiedenen Faktoren in dieser Zerlegung berechnen kann.

Der einfachste Fall ist $a = -1$.

**SATZ** Erstes Ergänzungsgesetz zum QRG

3.10. **Satz.** *Sei $p$ eine ungerade Primzahl. Dann ist*

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} 1 & \text{wenn } p \equiv 1 \bmod 4, \\ -1 & \text{wenn } p \equiv 3 \bmod 4. \end{cases}$$

*Beweis.* Nach Folgerung 3.7 gilt

$$\left(\frac{-1}{p}\right) \equiv (-1)^{(p-1)/2} \bmod p\,.$$

Da beide Seiten $\pm 1$ sind, folgt Gleichheit. ❑

Wenn $p \equiv 1 \bmod 4$ ist, dann gibt es also eine Quadratwurzel aus $-1$ mod $p$. Man kann eine solche sogar hinschreiben. Sei $p = 2m + 1$. Dann gilt

$$\begin{aligned}(m!)^2 &= (-1)^m \cdot 1 \cdot 2 \cdots m \cdot (-m) \cdots (-2) \cdot (-1) \\ &\equiv (-1)^m \cdot 1 \cdot 2 \cdots m \cdot (m+1) \cdots (p-1) \\ &= (-1)^m (p-1)! \equiv (-1)^{m+1} \bmod p\,.\end{aligned}$$

Hier haben wir im letzten Schritt die *Wilsonsche Kongruenz* $(p-1)! \equiv -1 \bmod p$ benutzt. Diese beweist man, indem man in der Fakultät jeden Faktor mit seinem Inversen mod $p$ zusammenfasst. Die einzigen ungepaarten Faktoren sind dann 1 und $-1$.

Wir sehen also, dass $(m!)^2 \equiv -1 \bmod p$ ist, wenn $m$ gerade, also $p \equiv 1 \bmod 4$ ist. Allerdings lässt sich $m! \bmod p$ nicht effizient berechnen, sodass diese Formel für praktische Zwecke nutzlos ist.

Im anderen Fall, für $p \equiv 3 \bmod 4$, ist $(m!)^2 \equiv 1 \bmod p$, also $m! \equiv \pm 1 \bmod p$. Man kann sich fragen, wovon das Vorzeichen abhängt. Es stellt sich heraus, dass das Vorzeichen für $p > 3$ dadurch bestimmt ist, ob die Klassenzahl von $\mathbb{Q}(\sqrt{-p})$ kongruent zu 1 oder zu 3 mod 4 ist. Im ersten Fall ist $m! \equiv -1$, im zweiten Fall $\equiv 1 \bmod p$.

Der nächste Schritt betrifft $a = 2$. Hier ist eine kleine Tabelle:

| $p$ | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 | 23 | 25 | 27 | 29 | 31 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\left(\frac{2}{p}\right)$ | | $-$ | $-$ | $+$ | | $-$ | $-$ | | $+$ | $-$ | | $+$ | | | $-$ | $+$ |

Das Ergebnis lässt vermuten, dass $\left(\frac{2}{p}\right)$ nur von $p$ mod 8 abhängt, und zwar sollte gelten $\left(\frac{2}{p}\right) = 1$ für $p \equiv 1$ oder 7 mod 8 und $\left(\frac{2}{p}\right) = -1$ für $p \equiv 3$ oder 5 mod 8.

Um so eine Aussage zu beweisen, müssen wir das Legendre-Symbol auf andere Weise ausdrücken. Dies wird durch das folgende Resultat von Gauß geleistet.

**LEMMA**
Lemma von Gauß über quadratische Reste

C.F. Gauß (1777–1855)

3.11. **Lemma.** *Sei $p$ eine ungerade Primzahl. Sei weiter $S \subset \mathbb{Z}$ eine Teilmenge mit $\#S = (p-1)/2$, sodass $\{0\} \cup S \cup -S$ ein vollständiges Repräsentantensystem mod $p$ ist. (Zum Beispiel können wir $S = \{1, 2, \ldots, (p-1)/2\}$ nehmen.) Dann gilt für $a \in \mathbb{Z}$ mit $p \nmid a$:*

$$\left(\frac{a}{p}\right) = (-1)^{\#\{s \in S \mid \overline{as} \in -\bar{S}\}}.$$

*Hierbei bezeichnet $\bar{S} = \{\bar{s} \mid s \in S\}$ die Menge der durch Elemente von $S$ repräsentierten Restklassen mod $p$.*

Der quadratische Restcharakter von $a$ hängt also davon ab, wie viele der Reste in $S$ bei Multiplikation mit $a$ „die Seite wechseln“.

*Beweis.* Für alle $s \in S$ gibt es eindeutig bestimmte $t(s) \in S$ und $\varepsilon(s) \in \{\pm 1\}$ mit $as \equiv \varepsilon(s)t(s) \bmod p$. Dann ist $S \ni s \mapsto t(s) \in S$ eine Permutation von $S$: Es genügt, die Injektivität zu zeigen. Seien also $s, s' \in S$ mit $t(s) = t(s')$. Dann ist $as \equiv \pm as' \bmod p$, also (da $a$ mod $p$ invertierbar ist) $s \equiv \pm s' \bmod p$. Das ist auf Grund der Wahl von $S$ nur möglich, wenn $s = s'$ ist.

Modulo $p$ haben wir dann

$$\begin{aligned}\left(\frac{a}{p}\right)\prod_{s\in S} s &\equiv a^{(p-1)/2}\prod_{s\in S} s\\ &= \prod_{s\in S}(as)\\ &\equiv \prod_{s\in S}(\varepsilon(s)t(s))\\ &= \prod_{s\in S}\varepsilon(s)\prod_{s\in S} s\\ &= (-1)^{\#\{s\in S\mid \varepsilon(s)=-1\}}\prod_{s\in S} s\,.\end{aligned}$$

Da $p$ das Produkt $\prod_{s\in S} s$ nicht teilt, folgt

$$\left(\frac{a}{p}\right) \equiv (-1)^{\#\{s\in S\mid \varepsilon(s)=-1\}} = (-1)^{\#\{s\in S\mid \overline{as}\in -\bar{S}\}} \bmod p\,,$$

und daraus die behauptete Gleichheit (beide Seiten sind $\pm 1$). ❑

Wenn wir in diesem Lemma $a = -1$ setzen, bekommen wir wieder Satz 3.10.

Wir können jetzt unsere Vermutung über $\left(\frac{2}{p}\right)$ beweisen.

**SATZ** Zweites Ergänzungsgesetz zum QRG

3.12. **Satz.** *Sei $p$ eine ungerade Primzahl. Dann gilt*

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} 1 & \text{wenn } p \equiv \pm 1 \bmod 8, \\ -1 & \text{wenn } p \equiv \pm 3 \bmod 8. \end{cases}$$

*Beweis.* Wir verwenden Lemma 3.11 mit

$$S = \left\{1, 2, 3, \ldots, \frac{p-1}{2}\right\}.$$

Wir müssen die Elemente von $S$ abzählen, die (mod $p$) außerhalb von $S$ landen, wenn sie verdoppelt werden. Für $s \in S$ gilt das genau dann, wenn $2s > (p-1)/2$, also wenn $(p-1)/4 < s \leq (p-1)/2$ ist. Die Anzahl dieser Elemente ist dann genau

$$n(p) = \frac{p-1}{2} - \left\lfloor \frac{p-1}{4} \right\rfloor.$$

Die folgende Tabelle bestimmt $n(p)$ für die verschiedenen Restklassen mod 8.

| $p$ | $n(p)$ | $\left(\frac{2}{p}\right)$ |
|---|---|---|
| $8k+1$ | $2k$ | $+1$ |
| $8k+3$ | $2k+1$ | $-1$ |
| $8k+5$ | $2k+1$ | $-1$ |
| $8k+7$ | $2k+2$ | $+1$ |

❑

Wie sieht es nun mit $\left(\frac{q}{p}\right)$ aus, wenn $q$ eine feste ungerade Primzahl ist und wir $p$ variieren lassen?

Wenn wir ähnlich wie eben für $a = 2$ die Werte für $a = 3$ und für $a = 5$ tabellieren, dann können wir Folgendes vermuten:

$$\left(\frac{3}{p}\right) = \left\{ \begin{array}{rl} 1 & \text{für } p \equiv \pm 1 \bmod 12, \\ -1 & \text{für } p \equiv \pm 5 \bmod 12; \end{array} \right\} = \begin{cases} \left(\frac{p}{3}\right) & \text{für } p \equiv 1 \bmod 4, \\ -\left(\frac{p}{3}\right) & \text{für } p \equiv -1 \bmod 4; \end{cases}$$

$$\left(\frac{5}{p}\right) = \left\{ \begin{array}{rl} 1 & \text{für } p \equiv \pm 1 \bmod 5, \\ -1 & \text{für } p \equiv \pm 2 \bmod 5. \end{array} \right\} = \left(\frac{p}{5}\right).$$

Für größere $q$ erhalten wir ähnliche Muster: Wenn $q \equiv 1 \bmod 4$ ist, dann hängt das Ergebnis nur von $p \bmod q$ ab, und wenn $q \equiv 3 \bmod 4$ ist, dann hängt das Ergebnis nur von $p \bmod 4q$ ab. Beide Fälle können im folgenden Resultat zusammengefasst werden, das zuerst von Gauß im Jahr 1796 bewiesen wurde.

**SATZ** Quadratisches Reziprozitätsgesetz

3.13. **Satz.** *Wenn $p$ und $q$ verschiedene ungerade Primzahlen sind, dann gilt*

$$\left(\frac{q}{p}\right) = \left(\frac{p^*}{q}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}} \left(\frac{p}{q}\right)$$

$$= \begin{cases} \left(\frac{p}{q}\right) & \textit{falls } p \equiv 1 \bmod 4 \textit{ oder } q \equiv 1 \bmod 4, \\ -\left(\frac{p}{q}\right) & \textit{falls } p \equiv -1 \bmod 4 \textit{ und } q \equiv -1 \bmod 4. \end{cases}$$

*Hier setzen wir $p^* = (-1)^{(p-1)/2} p$, d.h., $p^* = p$ für $p \equiv 1 \bmod 4$ und $p^* = -p$ für $p \equiv -1 \bmod 4$.*

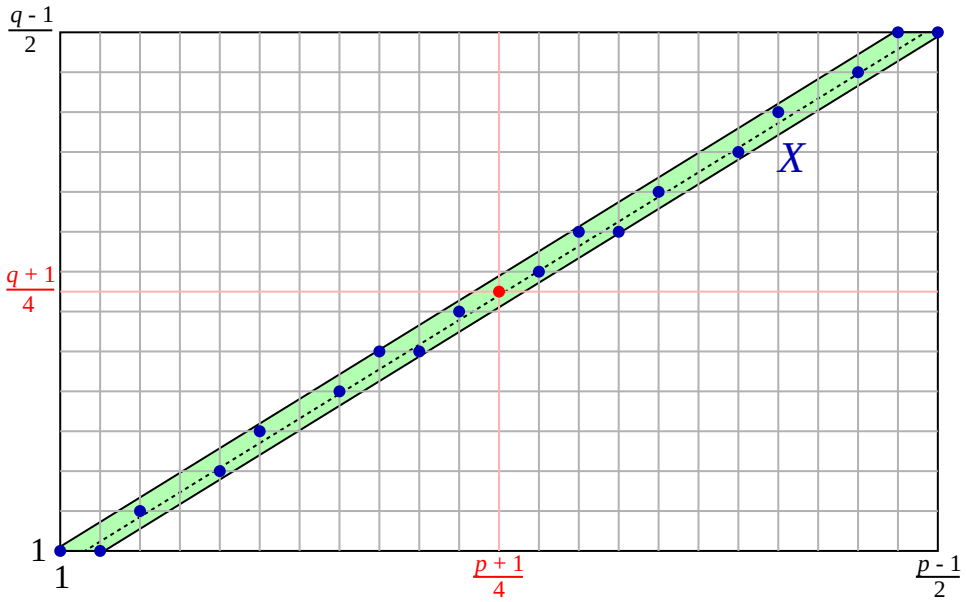



ABBILDUNG 1. Skizze zum Beweis von Satz 3.13. Hier ist $p = 47$, $q = 29$ mit $m = 11$, $n = 7$.

*Beweis.* Wir stützen uns wieder auf das Lemma 3.11 von Gauß. Da wir es mit zwei Legendre-Symbolen zu tun haben, brauchen wir zwei Mengen

$$S = \left\{1, 2, \ldots, \frac{p-1}{2}\right\} \quad \text{und} \quad T = \left\{1, 2, \ldots, \frac{q-1}{2}\right\}.$$

Sei $m = \#\{s \in S \mid \overline{qs} \in -\bar{S}\} \pmod p$ und $n = \#\{t \in T \mid \overline{pt} \in -\bar{T}\} \pmod q$. Dann haben wir

$$\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = (-1)^m(-1)^n = (-1)^{m+n}.$$

Wir müssen also die Parität von $m + n$ bestimmen.

Wenn $qs \equiv -s' \bmod p$ ist für ein $s' \in S$, dann gibt es ein eindeutig bestimmtes $t \in \mathbb{Z}$ mit $pt - qs = s' \in S$, also $0 < pt - qs \le (p-1)/2$. Diese Zahl $t$ muss in $T$ sein, denn

$$pt > qs > 0 \quad \text{und} \quad pt \le \frac{p-1}{2} + qs \le (q+1)\frac{p-1}{2} < p\frac{q+1}{2},$$

also $t < (q+1)/2$, und weil $q$ ungerade ist, heißt das $t \le (q-1)/2$. Damit ist

$$m = \#\left\{(s,t) \in S \times T \;\middle|\; 0 < pt - qs \le \frac{p-1}{2}\right\}.$$

Auf die gleiche Weise sehen wir, dass

$$n = \#\left\{(s,t) \in S \times T \;\middle|\; -\frac{q-1}{2} \le pt - qs < 0\right\}$$

ist. Da $pt - qs$ für $s \in S$, $t \in T$ niemals verschwindet, ergibt sich $m + n = \#X$ mit

$$X = \left\{(s,t) \in S \times T \;\middle|\; -\frac{q-1}{2} \le pt - qs \le \frac{p-1}{2}\right\}.$$

Diese Menge $X$ ist symmetrisch zum Mittelpunkt des Rechtecks $[1, \frac{p-1}{2}] \times [1, \frac{q-1}{2}]$: Punktspiegelung an $(\frac{p+1}{4}, \frac{q+1}{4})$ überführt $(s,t)$ in $(s',t') = (\frac{p+1}{2} - s, \frac{q+1}{2} - t)$, und

$$pt' - qs' = p\left(\frac{q+1}{2} - t\right) - q\left(\frac{p+1}{2} - s\right) = \frac{p-q}{2} - (pt - qs).$$

Also gilt

$$pt - qs \le \frac{p-1}{2} \iff pt' - qs' \ge -\frac{q-1}{2} \qquad \text{und}$$
$$pt - qs \ge -\frac{q-1}{2} \iff pt' - qs' \le \frac{p-1}{2},$$

d.h. $(s',t') \in X \iff (s,t) \in X$. Da der einzige Fixpunkt der Punktspiegelung der Punkt $(\frac{p+1}{4}, \frac{q+1}{4})$ ist und dieser Punkt genau dann in $X$ liegt, wenn er ganzzahlige Koordinaten hat, ergeben sich die folgenden Äquivalenzen:

$$\#X \text{ ist ungerade} \iff \frac{p+1}{4}, \frac{q+1}{4} \in \mathbb{Z} \iff p \equiv -1 \bmod 4 \text{ und } q \equiv -1 \bmod 4.$$

Damit ist der Satz bewiesen. ❑

3.14. **Beispiel.** Wir können das Quadratische Reziprozitätsgesetz dazu benutzen, Legendre-Symbole auf die folgende Weise zu berechnen:

**BSP** Legendre-Symbol durch QRG

$$\left(\frac{67}{109}\right) = \left(\frac{109}{67}\right) = \left(\frac{42}{67}\right) = \left(\frac{2\cdot 3\cdot 7}{67}\right) = \left(\frac{2}{67}\right)\left(\frac{3}{67}\right)\left(\frac{7}{67}\right)$$
$$= (-1)(-\left(\frac{67}{3}\right))(-\left(\frac{67}{7}\right)) = -\left(\frac{1}{3}\right)\left(\frac{4}{7}\right) = -1 \qquad \clubsuit$$

Der Nachteil dabei ist, dass wir die Zahlen, die in den Zwischenschritten auftauchen, faktorisieren müssen, was sehr aufwendig werden kann.

Um dieses Problem zu umgehen, verallgemeinern wir das Legendre-Symbol, indem wir beliebige ungerade Zahlen anstelle nur ungerade Primzahlen im „Nenner“ zulassen.

3.15. **Definition.** Sei $a \in \mathbb{Z}$, und sei $n \in \mathbb{Z}_{>0}$ ungerade mit Primfaktorzerlegung $n = p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}$. Wir definieren das *Jacobi-Symbol* durch

**DEF** Jacobi-Symbol

$$\left(\frac{a}{n}\right) = \prod_{j=1}^{k} \left(\frac{a}{p_j}\right)^{e_j}. \qquad \diamondsuit$$

C.G.J. Jacobi (1804–1851)

Das Jacobi-Symbol hat folgende Eigenschaften, die die entsprechenden Eigenschaften des Legendre-Symbols verallgemeinern:

(1) $\left(\frac{a}{n}\right) = 0$ genau dann, wenn $\gcd(a,n) \neq 1$.

(2) Wenn $a \equiv b \bmod n$, dann $\left(\frac{a}{n}\right) = \left(\frac{b}{n}\right)$.

(3) $\left(\frac{ab}{n}\right) = \left(\frac{a}{n}\right)\left(\frac{b}{n}\right)$.

(3') $\left(\frac{a}{mn}\right) = \left(\frac{a}{m}\right)\left(\frac{a}{n}\right)$. (Das folgt unmittelbar aus der Definition.)

(4) $\left(\frac{a}{n}\right) = 1$ wenn $a \perp n$ und $a$ ein Quadrat mod $n$ ist.

**Warnung.** Wenn $n$ nicht prim ist, gilt im allgemeinen die Umkehrung der letzten Implikation *nicht*. Beispielsweise ist $\left(\frac{2}{15}\right) = 1$, aber 2 ist kein Quadrat mod 15 (denn 2 ist kein Quadrat mod 3 oder mod 5).

Die wichtigste Eigenschaft des Jacobi-Symbols ist jedoch, dass das Quadratische Reziprozitätsgesetz und seine Ergänzungsgesetze gültig bleiben.

**SATZ**
QRG für das Jacobi-Symbol

3.16. **Satz.** *Seien $m, n \in \mathbb{Z}$ positiv und ungerade. Dann gilt*

(1) $\left(\dfrac{-1}{n}\right) = (-1)^{\frac{n-1}{2}}$.

(2) $\left(\dfrac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}}$.

(3) $\left(\dfrac{m}{n}\right) = (-1)^{\frac{m-1}{2}\frac{n-1}{2}} \left(\dfrac{n}{m}\right)$.

*Beweis.* Dazu überlegen wir zuerst, dass $n \mapsto (-1)^{(n-1)/2}$ und $n \mapsto (-1)^{(n^2-1)/8}$ als Abbildungen von $1 + 2\mathbb{Z}$ nach $\{\pm 1\}$ multiplikativ sind. Da der Wert nur von $n$ mod 4 bzw. $n$ mod 8 abhängt, ist das eine endliche Verifikation. Ebenso prüft man nach, dass $(m, n) \mapsto (-1)^{(m-1)(n-1)/4}$ multiplikativ in beiden Argumenten ist. Alternativ können wir so vorgehen:

$$\frac{nn'-1}{2} - \frac{n-1}{2} - \frac{n'-1}{2} = \frac{(n-1)(n'-1)}{2} \in 2\mathbb{Z}\,,$$

da $n$ und $n'$ beide ungerade sind. Ebenso ist

$$\frac{(nn')^2-1}{8} - \frac{n^2-1}{8} - \frac{(n')^2-1}{8} = \frac{(n^2-1)((n')^2-1)}{8} \in 2\mathbb{Z}\,.$$

Damit folgt

$$(-1)^{\frac{nn'-1}{2}} = (-1)^{\frac{n-1}{2}} \cdot (-1)^{\frac{n'-1}{2}} \quad \text{und}$$
$$(-1)^{\frac{(nn')^2-1}{8}} = (-1)^{\frac{n^2-1}{8}} \cdot (-1)^{\frac{(n')^2-1}{8}}\,.$$

Daher sind in den obigen Aussagen jeweils beide Seiten multiplikativ in $m$ und $n$, wir können sie also auf den Fall von Primzahlen und damit auf die bereits bekannten Sätze 3.10, 3.12 und 3.13 reduzieren. ❑

**BSP**
Verwendung des Jacobi-Symbols

3.17. **Beispiel.** Wir wiederholen die Berechnung von $\left(\frac{67}{109}\right)$:

$$\left(\frac{67}{109}\right) = \left(\frac{109}{67}\right) = \left(\frac{42}{67}\right) = \left(\frac{2}{67}\right)\left(\frac{21}{67}\right) = (-1)\left(\frac{67}{21}\right) = -\left(\frac{4}{21}\right) = -1$$

Man sieht, dass man auf diese Weise Legendre-Symbole (oder Jacobi-Symbole) im wesentlichen genauso berechnen kann wie den ggT. Der einzige Unterschied besteht darin, dass man Faktoren 2 herausziehen und extra behandeln muss. ♣

Das Euler-Kriterium 3.7 gilt für das Jacobi-Symbol nicht. Eulers Verallgemeinerung des kleinen Satzes von Fermat sagt, dass $a^{\varphi(n)} \equiv 1 \bmod n$ ist für alle $n, a \in \mathbb{Z}$ mit $n \geq 1$ und $a \perp n$. Dabei bezeichnet $\varphi(n) = \#\{a \in \mathbb{Z} \mid 0 \leq a < n, a \perp n\}$ die Eulersche phi-Funktion. Zum Beispiel gilt

$$a^{\varphi(15)/2} \equiv 1 \bmod 15$$

für alle $a$ mit $a \perp 15$, obwohl das Jacobi-Symbol $\left(\frac{a}{15}\right)$ auch den Wert $-1$ annimmt (z.B. für $a = 7$). Es gilt sogar für jedes positive ungerade $n$, das mindenstens zwei verschiedene Primfaktoren hat, dass $a^{\varphi(n)/2} \equiv 1 \bmod n$ ist für alle $a \in \mathbb{Z}$ mit $a \perp n$ (Übung). Auf der anderen Seite ist $2^{\varphi(9)/2} \equiv -1 \bmod 9$, aber $\left(\frac{a}{9}\right) = 1$ für alle $a \in \mathbb{Z}$ mit $a \perp 9$.

Die Erwartung, dass $a^{(n-1)/2} \equiv \left(\frac{a}{n}\right) \bmod n$ gelten könnte, ist "noch falscher". Dies lässt sich jedoch dazu verwenden, um zu zeigen, dass $n$ keine Primzahl ist. Dazu wählen wir zufällig ein $a \in \{2, 3, \ldots, n-2\}$ und überprüfen, ob die obige Kongruenz erfüllt ist (beide Seiten können effizient mod $n$ berechnet werden). Wenn das nicht der Fall ist, dann kann $n$ keine Primzahl sein. Dies ergibt den sogenannten *Solovay-Strassen-Primzahltest*.

Mit dem Jacobi-Symbol lässt sich folgendes Ergebnis (das man auch für das Legendre-Symbol formulieren kann, wenn man es auf Primzahlen $n$ einschränkt) elegant beweisen.

**SATZ**
$n \mapsto \left(\frac{a}{n}\right)$
ist periodisch

3.18. **Satz.** *Sei $a \in \mathbb{Z} \setminus \{0\}$. Dann hängt der Wert von $\left(\frac{a}{n}\right)$ (für $n > 0$ ungerade) nur von $n \bmod 4a$ ab.*

*Beweis.* Wir schreiben $a = \varepsilon \cdot 2^e \cdot m$ mit $m$ ungerade, $m > 0$, und $\varepsilon = \pm 1$. Dann gilt nach Satz 3.16

$$\begin{aligned}\left(\frac{a}{n}\right) &= \left(\frac{\varepsilon}{n}\right)\left(\frac{2}{n}\right)^e\left(\frac{m}{n}\right)\\ &= \left(\frac{\varepsilon}{n}\right)\left(\frac{2}{n}\right)^e(-1)^{(m-1)(n-1)/4}\left(\frac{n}{m}\right)\\ &= \big(\varepsilon(-1)^{(m-1)/2}\big)^{(n-1)/2}\big((-1)^e\big)^{(n^2-1)/8}\left(\frac{n}{m}\right).\end{aligned}$$

Der erste Faktor hängt höchstens von $n \bmod 4$ ab. Wenn der zweite Faktor nicht trivial ist, dann ist $e > 0$, also $2m \mid a$, und der zweite Faktor hängt nur von $n \bmod 8$ ab. Der dritte Faktor schließlich hängt nur von $n \bmod m$ ab. Insgesamt ist $\left(\frac{a}{n}\right)$ bestimmt durch

- $n \bmod m$, falls $a = 4^k \cdot m'$ mit $m' \equiv 1 \bmod 4$;
- $n \bmod 4m$, falls $a = 4^k \cdot m'$ mit $m' \equiv 3 \bmod 4$;
- $n \bmod 8m$, falls $a = 4^k \cdot m'$ mit $m' \equiv 2 \bmod 4$.

In jedem Fall gilt, dass $m$, $4m$ oder $8m$ ein Teiler von $4a$ ist. ❑

Wir wollen einmal sehen, wie man das Quadratische Reziprozitätsgesetz dazu benutzen kann, die Unlösbarkeit einer diophantischen Gleichung zu beweisen. Die folgende Gleichung wurde von Lind[5] und Reichardt[6] untersucht.

**SATZ**
Satz von Lind und Reichardt

3.19. **Satz.** *Die Gleichung $X^4 - 17\,Y^4 = 2\,Z^2$ hat keine primitiven ganzzahligen Lösungen und somit auch keine nichttrivialen (also mit $(X, Y, Z) \neq (0, 0, 0)$) ganzzahligen Lösungen.*

Man kann zeigen, dass diese Gleichung Lösungen in $\mathbb{R}$ hat (klar), und dass es immer primitive Lösungen mod $n$ gibt für alle $n \geq 1$. ($(x, y, z) \in \mathbb{Z}^3$ ist eine *primitive Lösung mod $n$*, wenn $x^4 - 17y^4 \equiv 2z^2 \bmod n$ und $\mathrm{ggT}(x, y, z, n) = 1$ ist.)

[5] Carl-Erik Lind: *Untersuchungen über die rationalen Punkte der ebenen kubischen Kurven vom Geschlecht Eins,* Uppsala: Diss. 97 S. (1940).

[6] Hans Reichardt: *Einige im Kleinen überall lösbare, im Grossen unlösbare diophantische Gleichungen,* J. reine angew. Math. **184** (1942), 12–18.

Man braucht also bessere Methoden als z.B. Reduktion modulo $n$, um diesen Satz zu beweisen.

*Beweis.* Sei $(X, Y, Z)$ eine primitive Lösung, und sei $p$ ein ungerader Primteiler von $Z$. (Da 17 keine vierte Potenz in $\mathbb{Q}$ ist, kann $Z$ nicht null sein.) Wäre $p = 17$, dann würde 17 auch $X$ und dann $Y$ teilen, was nicht geht. $p$ ist kein Teiler von $X$ oder $Y$ (denn $\gcd(X, Y, Z) = 1$). Modulo $p$ bekommen wir nun die Kongruenz $X^4 \equiv 17\,Y^4$; das zeigt, dass 17 ein quadratischer Rest mod $p$ ist. Nach dem QRG 3.13 folgt

$$\left(\frac{p}{17}\right) = \left(\frac{17}{p}\right) = 1\,.$$

Außerdem gilt nach den beiden Ergänzungsgesetzen 3.10 und 3.12 auch

$$\left(\frac{2}{17}\right) = \left(\frac{-1}{17}\right) = 1\,.$$

Da $Z$ ein Produkt von Potenzen von $-1$, 2 und seinen ungeraden Primteilern ist, folgt, dass $Z$ ein quadratischer Rest mod 17 sein muss. Es gibt also $W \in \mathbb{Z}$ mit $Z \equiv W^2 \bmod 17$. Damit bekommen wir die Kongruenz

$$X^4 \equiv 2\,W^4 \bmod 17$$

mit $W \not\equiv 0 \bmod 17$. Wir können also mit einem Inversen von $W^4$ mod 17 multiplizieren und erhalten

$$U^4 \equiv 2 \bmod 17$$

für geeignetes $U \in \mathbb{Z}$. Diese Kongruenz hat aber keine Lösung (die Quadratwurzeln aus 2 mod 17 sind $\pm 6$, und das sind keine quadratischen Reste mod 17).

Die Reduktion auf den Fall von primitiven Lösungen geht analog wie bei der Gleichung $X^4 + Y^4 = Z^2$; vgl. Abschnitt 2. ❑

## 4. Der Gitterpunktsatz von Minkowski

In diesem Abschnitt stellen wir ein Hilfsmittel bereit, mit dem sich viele interessante Resultate auf recht elegante Weise beweisen lassen. Die Grundidee dieses Prinzips sagt, dass jede „hinreichend große“ konvexe und zentralsymmetrische Teilmenge des $\mathbb{R}^n$ einen von 0 verschiedenen Punkt mit ganzzahligen Koordinaten enthalten muss. Dies ist die Aussage des Gitterpunktsatzes von Minkowski, den wir in diesem Abschnitt beweisen werden.

Zuerst führen wir aber den Begriff eines Gitters ein als Verallgemeinerung von $\mathbb{Z}^n \subset \mathbb{R}^n$.

**DEF**
Gitter
Grundmasche
Kovolumen

4.1. **Definition.** Ein *Gitter* $\Lambda \subset \mathbb{R}^n$ ist die Menge der ganzzahligen Linearkombinationen einer Basis $v_1, \dots, v_n$ von $\mathbb{R}^n$. Insbesondere ist $\Lambda$ dann eine additive Untergruppe von $\mathbb{R}^n$. Die Menge

$$F = \Big\{ \sum_{j=1}^{n} t_j v_j \mid 0 \le t_j < 1 \text{ für alle } j \Big\}$$

heißt eine *Grundmasche* des Gitters $\Lambda$, und $\Delta(\Lambda) = \mathrm{vol}(F) = |\det(v_1, \dots, v_n)|$ heißt das *Kovolumen* von $\Lambda$. ♢

Beachte, dass viele verschiedene Basen dasselbe Gitter erzeugen. Die Grundmasche $F$ hängt von der gewählten Basis ab, während das Kovolumen nur von $\Lambda$ abhängt, da die Basiswechselmatrix $A$ aus $\mathrm{GL}_n(\mathbb{Z})$ kommen muss (d.h., sowohl $A$ als auch $A^{-1}$ haben ganzzahlige Einträge) und somit ihre Determinante $\pm 1$ ist.

Die wichtigste Eigenschaft von $F$ ist, dass jeder Vektor $v \in \mathbb{R}^n$ *eindeutig* geschrieben werden kann als $v = \lambda + w$ mit $\lambda \in \Lambda$ und $w \in F$. Anders gesagt ist $\mathbb{R}^n$ die disjunkte Vereinigung aller Translate $F + \lambda$ von $F$ um Gittervektoren $\lambda \in \Lambda$.

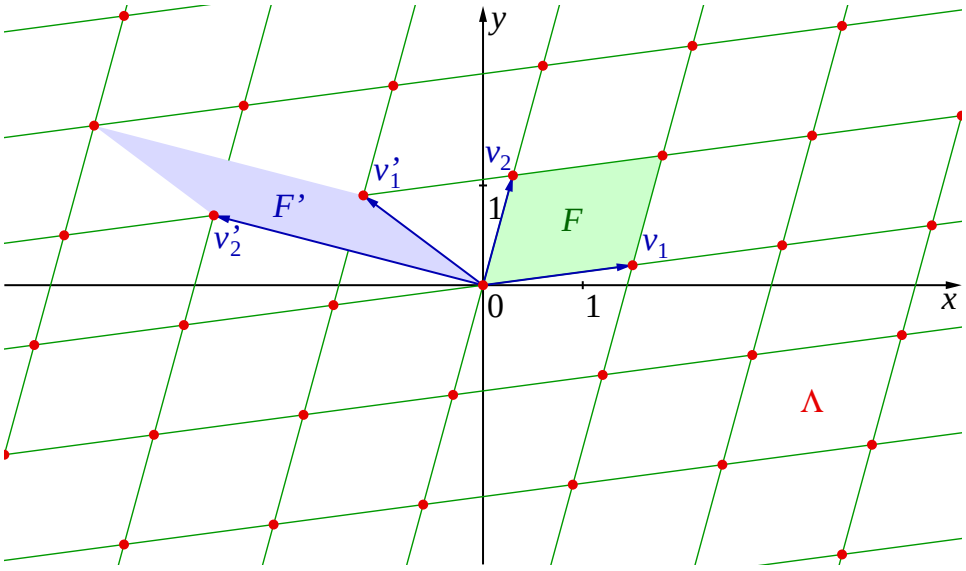



Abbildung 2. Beispiel eines Gitters mit zwei Grundmaschen

4.2. **Beispiel.** Das Standardbeispiel eines Gitters im $\mathbb{R}^n$ ist $\Lambda = \mathbb{Z}^n \subset \mathbb{R}^n$. Es wird von der Standardbasis $e_1, \dots, e_n$ von $\mathbb{R}^n$ erzeugt und hat Kovolumen $\Delta(\mathbb{Z}^n) = 1$. **BSP** Standard-gitter

In gewisser Weise ist dies das einzige Beispiel. Ist nämlich $\Lambda = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_n \subset \mathbb{R}^n$ irgendein Gitter, dann ist $\Lambda$ das Bild von $\mathbb{Z}^n$ unter der invertierbaren linearen Abbildung $T\colon \mathbb{R}^n \to \mathbb{R}^n$, die die Standardbasis $e_1, \dots, e_n$ auf $v_1, \dots, v_n$ abbildet. Das Kovolumen ist dann $\Delta(\Lambda) = |\det(T)|$. ♣

4.3. **Satz.** *Sei $\Lambda \subset \mathbb{R}^n$ ein Gitter und sei $\Lambda' \subset \Lambda$ eine Untergruppe von endlichem Index $m$. Dann ist $\Lambda'$ ebenfalls ein Gitter, und es gilt $\Delta(\Lambda') = m\,\Delta(\Lambda)$.* **SATZ** Untergitter

*Beweis.* Wir zeigen erst einmal, dass $\Lambda'$ ein Gitter ist. Dazu beachten wir, dass gilt $m\Lambda \subset \Lambda' \subset \Lambda$ (denn für $\lambda \in \Lambda$ wird die Restklasse $\bar\lambda \in \Lambda/\Lambda'$ durch die Gruppenordnung $\#(\Lambda/\Lambda') = m$ annulliert; es folgt $m\lambda \in \Lambda'$). Als Untergruppe der endlich erzeugten freien abelschen Gruppe $\Lambda \cong \mathbb{Z}^n$ ist auch $\Lambda'$ endlich erzeugt und frei, mit höchstens $n$ Erzeugern. Da $m\Lambda$ eine Basis von $\mathbb{R}^n$ enthält, muss jedes Erzeugendensystem von $\Lambda'$ ein Erzeugendensystem von $\mathbb{R}^n$ sein. Da $\Lambda'$ sich von höchstens $n$ Elementen erzeugen lässt, folgt, dass diese Erzeuger sogar eine Basis des $\mathbb{R}^n$ bilden; damit ist $\Lambda'$ ein Gitter.

Wenn wir ein Erzeugendensystem von $\Lambda$ fixieren, dann können wir die Erzeuger von $\Lambda'$ als Linearkombinationen der Erzeuger von $\Lambda$ darstellen; das ergibt eine $(n \times n)$-Matrix $A$ mit ganzzahligen Einträgen, in deren $j$-ter Spalte die Koeffizienten des $j$-ten Erzeugers von $\Lambda'$ stehen. Nach dem Elementarteilersatz (auch Smith-Normalform genannt) gibt es invertierbare ganzzahlige Matrizen $U, V \in \mathrm{GL}(n, \mathbb{Z})$, sodass $UAV = \mathrm{diag}(a_1, \dots, a_n)$ eine Diagonalmatrix mit nichtnegativen Einträgen ist. Wenn wir die Erzeugendensysteme von $\Lambda$ und $\Lambda'$ entsprechend den Matrizen $U^{-1}$ und $V$ abändern und die neuen Erzeuger von $\Lambda$ mit $v_1, \dots, v_n$ bezeichnen, dann sind $a_1v_1, \dots, a_nv_n$ Erzeuger von $\Lambda'$. Insbesondere ist $\Lambda/\Lambda' \cong \mathbb{Z}/a_1\mathbb{Z} \times \cdots \times \mathbb{Z}/a_n\mathbb{Z}$. Da $m = \#(\Lambda/\Lambda')$ endlich ist, folgt $a_1 \cdots a_n = m$. Daraus ergibt sich

$$\Delta(\Lambda') = |\det(a_1v_1, \dots, a_nv_n)| = a_1 \cdots a_n\,|\det(v_1, \dots, v_n)| = m\,\Delta(\Lambda)\,. \qquad ❑$$

Dieser Satz gibt uns eine einfache Möglichkeit, Gitter zu konstruieren. In den Anwendungen werden wir uns häufig auf diese Konstruktion stützen.

4.4. **Folgerung.** *Sei $\phi\colon \mathbb{Z}^n \to M$ ein Gruppenhomomorphismus in eine endliche Gruppe $M$. Dann ist der Kern von $\phi$ ein Gitter $\Lambda \subset \mathbb{Z}^n \subset \mathbb{R}^n$ mit Kovolumen $\Delta(\Lambda) = \#\,\mathrm{im}(\phi) \le \#M$.* **FOLG** $\ker(\mathbb{Z}^n \to M)$

*Beweis.* Es gilt $\mathbb{Z}^n/\ker\phi \cong \mathrm{im}(\phi) \le M$, also ist $\Lambda = \ker\phi$ eine Untergruppe des Gitters $\mathbb{Z}^n$ vom endlichen Index $(\mathbb{Z}^n : \Lambda) = \#\,\mathrm{im}(\phi) \le \#M$. Die Behauptung folgt dann aus Satz 4.3 und aus $\Delta(\mathbb{Z}^n) = 1$. ❑

Jetzt können wir den Satz von Minkowski formulieren und beweisen.

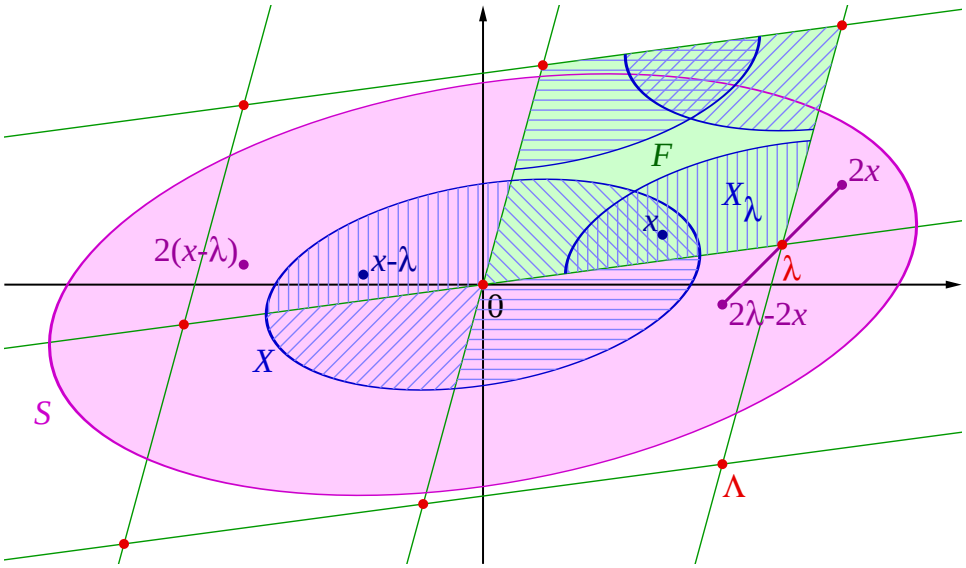



ABBILDUNG 3. Skizze zum Beweis von Satz 4.5

**SATZ** Gitterpunktsatz von Minkowski

4.5. **Satz.** *Sei $\Lambda \subset \mathbb{R}^n$ ein Gitter und sei $S \subset \mathbb{R}^n$ eine (zentral)symmetrische ($S = -S$) und konvexe Teilmenge, sodass $\mathrm{vol}(S) > 2^n\Delta(\Lambda)$ gilt. Dann enthält $S$ einen von Null verschiedenen Gitterpunkt aus $\Lambda$.*

*Beweis.* Wir beweisen zunächst, dass $X = \frac{1}{2}S$ mit einem seiner Translate $X + \lambda$ um ein Element $0 \neq \lambda \in \Lambda$ nichtleeren Durchschnitt hat. Das ist eine Folge davon, dass $\mathrm{vol}(X) > \mathrm{vol}(F)$ ist, sodass die ganzen Translate von $X$ „nicht genug Platz haben“, um disjunkt zu sein.

H. Minkowski (1864–1909)

Sei dazu $F$ eine Grundmasche von $\Lambda$. Wir setzen für $\lambda \in \Lambda$

$$X_\lambda = F \cap (X + \lambda)\,.$$

$\mathbb{R}^n$ ist die disjunkte Vereinigung der $F - \lambda$ mit $\lambda \in \Lambda$: $\mathbb{R}^n = \coprod_{\lambda\in\Lambda}(F - \lambda)$. Es folgt

$$X = \coprod_{\lambda\in\Lambda}\big(X \cap (F - \lambda)\big) = \coprod_{\lambda\in\Lambda}\big(((X + \lambda) \cap F) - \lambda\big) = \coprod_{\lambda\in\Lambda}(X_\lambda - \lambda)\,.$$

Damit ergibt sich (unter Verwendung der Voraussetzung $\mathrm{vol}(S) > 2^n\Delta(\Lambda)$)

$$\sum_{\lambda\in\Lambda}\mathrm{vol}(X_\lambda) = \sum_{\lambda\in\Lambda}\mathrm{vol}(X_\lambda - \lambda) = \mathrm{vol}(X) = 2^{-n}\,\mathrm{vol}(S) > \Delta(\Lambda) = \mathrm{vol}(F)\,.$$

Auf der anderen Seite gilt nach Definition

$$\bigcup_{\lambda\in\Lambda} X_\lambda \subset F\,.$$

Wären die $X_\lambda$ paarweise disjunkt, dann würde daraus

$$\sum_{\lambda\in\Lambda}\mathrm{vol}(X_\lambda) \leq \mathrm{vol}(F)$$

folgen, im Widerspruch zu dem gerade Gezeigten. Also muss es $\lambda, \mu \in \Lambda$ geben mit $\lambda \neq \mu$ und $X_\lambda \cap X_\mu \neq \emptyset$. Wir verschieben um $-\mu$ und erhalten

$$X \cap (X + \lambda - \mu) \supset (X_\mu - \mu) \cap (X_\lambda - \mu) \neq \emptyset\,.$$

(In der Skizze ist $\mu = 0$.)

Sei nun $x \in X \cap (X + \lambda - \mu)$. Dann ist $2x \in 2X = S$ und $2x - 2(\lambda - \mu) \in 2X = S$. Da $S$ symmetrisch ist, haben wir auch $2(\lambda - \mu) - 2x \in S$. Da $S$ außerdem konvex ist, muss der Mittelpunkt der Strecke, die $2x$ und $2(\lambda - \mu) - 2x$ verbindet, ebenfalls in $S$ sein. Dieser Mittelpunkt ist aber gerade $\lambda - \mu \in \Lambda \setminus \{0\}$. Damit ist der Satz bewiesen. ❑

Der erste Teil des Beweises zeigt tatsächlich die folgende etwas stärkere Aussage (Satz von Blichfeldt):

**Satz.** *Seien $\Lambda \subset \mathbb{R}^n$ ein Gitter und $X \subset \mathbb{R}^n$ eine (messbare) Menge mit $\mathrm{vol}(X) > \Delta(\Lambda)$. Dann enthält $X$ zwei Punkte $x$ und $y$ mit $x \neq y$ und $x - y \in \Lambda$.*

**SATZ**
Satz von Blichfeldt

Alle drei Bedingungen an $S$ in Satz 4.5 sind tatsächlich notwendig. Sei zum Beispiel $\Lambda = \mathbb{Z}^n$. Wählen wir $S$ als den offenen $n$-dimensionalen Würfel mit Seitenlänge 2 und Zentrum im Ursprung, dann ist $\mathrm{vol}(S) = 2^n = 2^n\Delta(\Lambda)$, aber dennoch ist $S \cap \Lambda = \{0\}$. Man kann die strikte Ungleichung $\mathrm{vol}(S) > 2^n\Delta(\Lambda)$ abschwächen zu $\mathrm{vol}(S) \geq 2^n\Delta(\Lambda)$, wenn man zusätzlich voraussetzt, dass $S$ kompakt ist (was für den offenen Würfel natürlich nicht gilt), siehe unten.

Es ist auch nicht schwer, Gegenbeispiele zu finden, bei denen $S$ nicht symmetrisch oder nicht konvex ist.

4.6. **Satz.** *Sei $\Lambda \subset \mathbb{R}^n$ ein Gitter und sei $S \subset \mathbb{R}^n$ eine symmetrische, konvexe und kompakte Teilmenge, sodass $\mathrm{vol}(S) \geq 2^n\Delta(\Lambda)$ gilt. Dann enthält $S$ einen von Null verschiedenen Gitterpunkt aus $\Lambda$.*

**SATZ**
Gitterpunktsatz, Variante

*Beweis.* Für $t \in \mathbb{R}_{>0}$ sei $tS = \{tx \mid x \in S\}$. Dann ist $\mathrm{vol}(tS) = t^n \mathrm{vol}(S)$, und für $t > 1$ erfüllt $tS$ die Voraussetzungen in Satz 4.5. Für jedes $m \geq 1$ gibt es also einen Gitterpunkt $0 \neq x_m \in (1 + \frac{1}{m})S \cap \Lambda$. Weil mit $S$ auch $2S$ kompakt und $\Lambda \subset \mathbb{R}^n$ diskret ist, gibt es nur endlich viele Punkte in $2S \cap \Lambda$. Es muss demnach einer dieser Punkte unendlich oft als $x_m$ vorkommen. Sei $x$ ein solcher Punkt. Dann ist $x \neq 0$, $x \in \Lambda$ und $(1 + \frac{1}{m})^{-1}x \in S$ für unendlich viele $m$. Da $S$ abgeschlossen ist, folgt $x = \lim_{m\to\infty}(1 + \frac{1}{m})^{-1}x \in S$. ❑

## 5. Summen von zwei und vier Quadraten

In diesem Abschnitt werden wir die Frage beantworten, welche natürlichen Zahlen als Summe von (höchstens) zwei bzw. vier Quadratzahlen geschrieben werden können.

Wir betrachten zunächst Summen von zwei Quadraten. Sei

$$\begin{aligned}\Sigma_2 &= \{x^2+y^2 \mid x,y\in\mathbb{Z}\}\\ &= \{0,1,2,4,5,8,9,10,13,16,17,18,20,25,26,29,32,34,36,37,40,\dots\}\,.\end{aligned}$$

Offensichtlich sind alle Quadratzahlen in $\Sigma_2$. Fast ebenso klar ist, dass alle Zahlen $n\equiv 3 \bmod 4$ fehlen, denn ein Quadrat ist stets $\equiv 0$ oder $1 \bmod 4$, eine Summe von zwei Quadraten kann also niemals $\equiv 3 \bmod 4$ sein. Außerdem hat $\Sigma_2$ noch die folgende wichtige Eigenschaft:

**LEMMA** $\Sigma_2$ ist mult. abg.

5.1. **Lemma.** *$\Sigma_2$ ist multiplikativ abgeschlossen: $m,n\in\Sigma_2 \Longrightarrow mn\in\Sigma_2$.*

*Beweis.* Wir verifizieren folgende Gleichheit:

$$(x^2+y^2)(u^2+v^2) = (xu\mp yv)^2+(xv\pm yu)^2\,. \qquad \square$$

Eine Möglichkeit, diese Formel zu interpretieren, verwendet komplexe Zahlen:

$$|x+y\boldsymbol{i}|^2 = x^2+y^2 \quad\text{und}\quad |\alpha\beta|^2=|\alpha|^2|\beta|^2\,.$$

Wegen dieser multiplikativen Struktur von $\Sigma_2$ liegt es nahe, sich anzusehen, welche Primzahlen in $\Sigma_2$ liegen. Wir haben schon gesehen, dass $p\notin\Sigma_2$ ist für Primzahlen $p\equiv 3 \bmod 4$. Auf der anderen Seite ist natürlich $2\in\Sigma_2$, und die Aufzählung der Elemente von $\Sigma_2$ oben lässt vermuten, dass alle Primzahlen $p\equiv 1 \bmod 4$ ebenfalls in $\Sigma_2$ sind. Dies ist tatsächlich der Fall, wie bereits von Fermat gezeigt wurde.

**SATZ** 2-□-Satz für Primzahlen

5.2. **Satz.** *Ist $p\equiv 1 \bmod 4$ eine Primzahl, dann ist $p\in\Sigma_2$.*

*Erster Beweis.* Dieser erste Beweis beruht auf der Abstiegsmethode von Fermat.

Wir wissen nach Satz 3.10, dass $-1$ ein quadratischer Rest mod $p$ ist. Also gibt es $a\in\mathbb{Z}$, $k\geq 1$ mit $a^2+1=kp$. Wir können $|a|\leq (p-1)/2$ wählen, dann gilt $k<p/4<p$.

Sei jetzt $k\geq 1$ minimal, sodass es $x,y\in\mathbb{Z}$ gibt mit $x^2+y^2=kp$. Wir müssen zeigen, dass $k=1$ ist. Nehmen wir also $k>1$ an. Es gibt $u\equiv x \bmod k$, $v\equiv -y \bmod k$ mit $|u|,|v|\leq k/2$. Dann ist $u^2+v^2\equiv x^2+y^2\equiv 0 \bmod k$, also

$$u^2+v^2 = kk'$$

mit $0\leq k'\leq k/2<k$. Nun kann $k'$ nicht null sein, sonst wäre $u=v=0$ und damit $k\mid x,y$, also $k^2\mid kp$, was nicht geht, denn $k\nmid p$ (es ist $1<k<p$). Also ist $1\leq k'<k$. Nun gilt

$$xu-yv\equiv x^2+y^2\equiv 0 \bmod k\,,\qquad xv+yu\equiv xy-yx=0 \bmod k$$

und $(xu-yv)^2+(xv+yu)^2=(x^2+y^2)(u^2+v^2)=k^2\,k'p$. Wir setzen

$$x'=\frac{xu-yv}{k}\,,\qquad y'=\frac{xv+yu}{k}\,,$$

dann wird $(x')^2+(y')^2=k'p$ mit $1\leq k'<k$, im Widerspruch zur Minimalität von $k$. Also ist $k>1$ nicht möglich, und wir müssen $k=1$ haben. $\square$

Die Idee in diesem Beweis ist die folgende. Sei $\xi = x + y\boldsymbol{i}$ mit $|\xi|^2 = kp$. Wir konstruieren $\eta = u + v\boldsymbol{i}$ mit $\eta \equiv \bar{\xi} \bmod k$ (im Ring $\mathbb{Z}[\boldsymbol{i}]$) und $|\eta|^2 = kk'$. Dann wird $\xi\eta \equiv |\xi|^2 \equiv 0 \bmod k$, es ist also $\xi\eta = k\xi'$ mit $\xi' \in \mathbb{Z}[\boldsymbol{i}]$, und wir haben $|\xi'|^2 = |\xi\eta|^2/k^2 = k'p$.

Jetzt wollen wir den Satz von Minkowski benutzen, um einen zweiten Beweis von Satz 5.2 zu geben.

*Zweiter Beweis.* Für den Satz von Minkowski 4.5 brauchen wir ein Gitter $\Lambda$ und eine Menge $S \subset \mathbb{R}^n$. Da wir es mit zwei Variablen zu tun haben, ist $n = 2$. Wir werden das Gitter dazu benutzen, um sicherzustellen, dass $x^2 + y^2$ ein Vielfaches von $p$ ist, und wir werden $S$ verwenden, um zu erreichen, dass $x^2 + y^2$ so klein ist, dass $x^2 + y^2 = p$ die einzig verbleibende Möglichkeit ist.

Um das Gitter zu konstruieren, beginnen wir wieder mit der Tatsache, dass $-1$ quadratischer Rest mod $p$ ist. Sei also $a \in \mathbb{Z}$ mit $a^2 + 1 \equiv 0 \bmod p$. Wir definieren

$$\phi\colon \mathbb{Z}^2 \longrightarrow \mathbb{F}_p\,, \quad (x, y) \longmapsto \bar{x} - \bar{a}\bar{y}\,,$$

dann ist $\phi$ ein surjektiver Gruppenhomomorphismus auf die endliche (additive) Gruppe $\mathbb{F}_p$. Nach Folgerung 4.4 ist dann $\Lambda = \ker\phi$ ein Gitter mit Kovolumen $\Delta(\Lambda) = \#\mathbb{F}_p = p$. Sei jetzt $(x, y) \in \Lambda$. Dann gilt $x \equiv ay \bmod p$, also

$$x^2 + y^2 \equiv (ay)^2 + y^2 = (a^2 + 1)y^2 \equiv 0 \bmod p\,,$$

also ist $x^2 + y^2$ durch $p$ teilbar.

Für $S$ nehmen wir die offene Kreisscheibe vom Radius $\sqrt{2p}$ um den Ursprung. Dann gilt für $(x, y) \in S$, dass $x^2 + y^2 < 2p$ ist. Offensichtlich ist $S$ symmetrisch und konvex. Außerdem ist

$$\mathrm{vol}(S) = 2\pi p > 4p = 2^2\Delta(\Lambda)\,,$$

sodass wir Satz 4.5 anwenden können. Der Satz liefert uns $(0, 0) \neq (x, y) \in S \cap \Lambda$. Es folgt, dass $x^2 + y^2$ ein Vielfaches von $p$ ist mit $0 < x^2 + y^2 < 2p$, also muss $p = x^2 + y^2 \in \Sigma_2$ sein. ❑

In der „Einführung in die Zahlentheorie und algebraische Strukturen“ haben wir dafür noch einen anderen Beweis gesehen, der darauf basiert, dass $\mathbb{Z}[\boldsymbol{i}]$ ein euklidischer Ring ist. (Tatsächlich beruht unser Abstiegsbeweis im Grunde auf derselben Eigenschaft von $\mathbb{Z}[\boldsymbol{i}]$.)

Aus dem, was wir bisher bewiesen haben, folgt bereits eine Richtung des folgenden Ergebnisses, das die Elemente von $\Sigma_2$ charakterisiert.

5.3. **Satz.** *Eine positive ganze Zahl $n$ kann genau dann als Summe zweier Quadrate geschrieben werden, wenn jede Primzahl $p \equiv 3 \bmod 4$ in der Primfaktorzerlegung von $n$ mit geradem Exponenten auftritt.*

**SATZ**
2-□-Satz

*Beweis.* „⇐“: Wenn $n$ die angegebene Form hat, dann ist $n = p_1 \cdots p_r m^2$ mit Primzahlen $p_j = 2$ oder $p_j \equiv 1 \bmod 4$. Wir wissen, dass alle Faktoren in $\Sigma_2$ sind (klar für 2 und für $m^2$, Satz 5.2 für $p \equiv 1 \bmod 4$), also ist wegen der multiplikativen Abgeschlossenheit von $\Sigma_2$ auch $n \in \Sigma_2$.

„⇒“: Wir verwenden Induktion: Wir nehmen an, dass $n \in \Sigma_2$ ist und dass wir bereits wissen, dass alle $m \in \Sigma_2$ mit $m < n$ die angegebene Form haben. Hat $n$ keinen Primteiler $p \equiv 3 \bmod 4$, dann ist nichts zu zeigen. Sei also $p \equiv 3 \bmod 4$ ein Primteiler von $n$. Wir können (wegen $n \in \Sigma_2$) $n = x^2 + y^2$ schreiben. Dann muss $p$

sowohl $x$ als auch $y$ teilen: Angenommen, $p$ teilt etwa $x$ nicht. Dann gibt es $a \in \mathbb{Z}$ mit $ax \equiv 1 \bmod p$, und nach Multiplikation mit $a^2$ folgt aus $0 \equiv n = x^2+y^2 \bmod p$

$$-1 \equiv -1 + (ax)^2 + (ay)^2 \equiv (ay)^2 \bmod p\,.$$

Damit wäre $-1$ ein quadratischer Rest mod $p$, was aber wegen $p \equiv 3 \bmod 4$ nach Satz 3.10 nicht sein kann. Also war die Annahme falsch, und $p$ muss $x$ und $y$ teilen. Dann ist aber $p^2$ ein Teiler von $n$. Wir schreiben $n = p^2m$. Es ist $m = (x/p)^2 + (y/p)^2$ ebenfalls Summe von zwei Quadraten, also wissen wir nach unserer Induktionsannahme, dass $m$ die angegebene Form hat. Dann hat aber auch $n$ diese Form. ❑

L. Euler
(1707–1783)

Als nächstes wollen wir uns der Frage zuwenden, welche natürlichen Zahlen Summen von *vier* Quadraten sind. Wenn man etwas herumexperimentiert, wird man feststellen, dass das anscheinend immer möglich ist. Sei also

$$\Sigma_4 = \{x_1^2 + x_2^2 + x_3^2 + x_4^2 \mid x_1, x_2, x_3, x_4 \in \mathbb{Z}\}\,.$$

Euler hat 1748 folgendes Analogon von Lemma 5.1 bewiesen:

5.4. **Lemma.** *$\Sigma_4$ ist multiplikativ abgeschlossen.*

**LEMMA**
$\Sigma_4$ ist
mult. abg.

*Beweis.* Man überzeugt sich davon, dass Folgendes gilt:

$$\begin{aligned}(a^2+b^2+c^2+d^2)&(A^2+B^2+C^2+D^2)\\ &= (aA-bB-cC-dD)^2 + (aB+bA+cD-dC)^2\\ &\quad + (aC+cA-bD+dB)^2 + (aD+dA+bC-cB)^2\,.\end{aligned}$$ ❑

W.R. Hamilton
(1805–1865)

In der gleichen Weise, wie die Multiplikationsformel für Summen zweier Quadrate mit den komplexen Zahlen zusammenhängt, hat diese Formel für vier Quadrate mit den *Quaternionen* zu tun. Sie wurden von Hamilton entdeckt und sind wie folgt definiert: $\mathbb{H}$ ist eine $\mathbb{R}$-Algebra. Als $\mathbb{R}$-Vektorraum ist

$$\mathbb{H} = \{a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \mid a, b, c, d \in \mathbb{R}\}\,,$$

damit ist die Addition in $\mathbb{H}$ definiert. Die Multiplikation ist durch folgende Regeln festgelegt:

$$\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1,$$
$$\boldsymbol{ij} = \boldsymbol{k},\ \boldsymbol{ji} = -\boldsymbol{k}, \quad \boldsymbol{jk} = \boldsymbol{i},\ \boldsymbol{kj} = -\boldsymbol{i}, \quad \boldsymbol{ki} = \boldsymbol{j},\ \boldsymbol{ik} = -\boldsymbol{j}\,.$$

Man sieht, dass $\mathbb{H}$ nicht kommutativ ist.

Zu einer Quaternion $\alpha = a+b\boldsymbol{i}+c\boldsymbol{j}+d\boldsymbol{k}$ definiert man die *konjugierte Quaternion* als $\bar\alpha = a - b\boldsymbol{i} - c\boldsymbol{j} - d\boldsymbol{k}$. Dann prüft man nach, dass

$$N(\alpha) := \alpha\bar\alpha = \bar\alpha\alpha = a^2+b^2+c^2+d^2$$

ist. Da $\overline{\alpha\beta} = \bar\beta\bar\alpha$ gilt, folgt

$$N(\alpha\beta) = \alpha\beta\,\overline{\alpha\beta} = \alpha\beta\bar\beta\bar\alpha = \alpha N(\beta)\bar\alpha = \alpha\bar\alpha N(\beta) = N(\alpha)N(\beta)\,.$$

Dabei haben wir benutzt, dass $N(\beta) \in \mathbb{R}$ ist und daher mit allen Quaternionen kommutiert. Diese Multiplikationsformel für die „Norm“ $N$ ergibt die Formel aus Lemma 5.4, wenn man sie ausschreibt.

Da $N(\alpha) = a^2+b^2+c^2+d^2$ ist, gilt $N(\alpha) \neq 0$ für $\alpha \neq 0$. Demnach ist

$$\alpha^{-1} = N(\alpha)^{-1}\bar\alpha$$

ein Inverses von $\alpha \neq 0$: $\mathbb{H}$ ist ein *Schiefkörper*. Mehr Informationen zu den Quaternionen gibt es in [Z, § 7].

Um zu beweisen, dass alle positiven ganzen Zahlen in $\Sigma_4$ sind, genügt es also zu zeigen, dass alle *Primzahlen* in $\Sigma_4$ sind. Als Startpunkt brauchen wir folgendes Resultat.

**LEMMA**
$a \equiv bu^2 + cv^2$ mod $p$

5.5. **Lemma.** *Sei $p$ eine ungerade Primzahl und seien $a, b, c \in \mathbb{Z}$ keine Vielfachen von $p$. Dann gibt es $u, v \in \mathbb{Z}$ mit*

$$a \equiv bu^2 + cv^2 \bmod p\,.$$

*Beweis.* Wir müssen in $\mathbb{F}_p$ die Gleichung $\bar{a} - \bar{b}\bar{u}^2 = \bar{c}\bar{v}^2$ lösen; hierbei sind $\bar{a}, \bar{b}, \bar{c} \neq 0$. Wir wissen, dass es genau $(p+1)/2$ Quadrate in $\mathbb{F}_p$ gibt (null und die $(p-1)/2$ quadratischen Restklassen). Beide Seiten der Gleichung können also unabhängig voneinander jeweils $(p+1)/2$ verschiedene Werte annehmen. Da $\mathbb{F}_p$ aber nur $p < (p+1)/2 + (p+1)/2$ Elemente hat, können diese beiden Wertemengen nicht disjunkt sein. Es muss also $u, v \in \mathbb{Z}$ geben, sodass beide Seiten gleich werden. ❑

Insbesondere sehen wir, dass die Kongruenz $u^2 + v^2 + 1 \equiv 0 \bmod p$ stets lösbar ist.

**SATZ**
4-□-Satz
für Primzahlen

5.6. **Satz.**
*Sei $p$ eine Primzahl. Dann ist $p$ Summe von vier Quadraten ganzer Zahlen.*

Wir werden wieder zwei Beweise geben.

*Erster Beweis.* Die Aussage ist klar für $p = 2$. Sei also $p$ ungerade. Dann gibt es nach Lemma 5.5 ganze Zahlen $u, v$ mit (oBdA) $|u|, |v| \leq (p-1)/2$, sodass $0^2 + 1^2 + u^2 + v^2 = kp$ ist mit $0 < k < p$. Wir können also jedenfalls ein Vielfaches von $p$ als Summe von vier Quadraten schreiben. Sei nun $k \geq 1$ minimal, sodass es $a, b, c, d \in \mathbb{Z}$ gibt mit $a^2 + b^2 + c^2 + d^2 = kp$. Wir müssen zeigen, dass $k = 1$ ist. Also nehmen wir $k > 1$ an. Analog wie im Beweis des Zwei-Quadrate-Satzes betrachten wir die konjugierte Quaternion mod $k$: Wir wählen $A, B, C, D \in \mathbb{Z}$ mit $|A|, |B|, |C|, |D| \leq k/2$ und

$$A \equiv a, \quad B \equiv -b, \quad C \equiv -c, \quad D \equiv -d \bmod k\,.$$

Es gilt dann jedenfalls $A^2 + B^2 + C^2 + D^2 \leq 4(k/2)^2 = k^2$. Wenn wir hier Gleichheit haben, dann muss $k = 2m$ gerade sein, und es muss $A, B, C, D = \pm m$ gelten. Das heißt dann aber auch, dass $a, b, c, d \equiv m \bmod 2m$ sind; damit wären $a, b, c, d$ alle durch $m$ teilbar: $a = ma'$, $b = mb'$, $c = mc'$, $d = md'$ mit $a', b', c', d'$ ungerade. Dann wäre

$$kp = a^2 + b^2 + c^2 + d^2 = m^2\big((a')^2 + (b')^2 + (c')^2 + (d')^2\big)$$

durch $k^2 = 4m^2$ teilbar (denn $(a')^2, (b')^2, (c')^2, (d')^2 \equiv 1 \bmod 4$), was aber nicht geht, da $k$ wegen $1 < k < p$ kein Teiler von $p$ ist. Ähnlich sieht man, dass $A, B, C, D$ nicht alle null sein können. In jedem Fall ist

$$A^2 + B^2 + C^2 + D^2 \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \bmod k\,.$$

Also ist $A^2 + B^2 + C^2 + D^2 = kk'$ mit $1 \leq k' < k$. Nun ist

$$\begin{aligned} k^2k'p &= (a^2 + b^2 + c^2 + d^2)(A^2 + B^2 + C^2 + D^2) \\ &= (aA - bB - cC - dD)^2 + (aB + bA + cD - dC)^2 \\ &\quad + (aC + cA - bD + dB)^2 + (aD + dA + bC - cB)^2\,. \end{aligned}$$

Man prüft nach, dass alle vier Klammern im letzten Ausdruck durch $k$ teilbar sind. Wir können sie also jeweils durch $k$ teilen und erhalten eine Darstellung von $k'p$ als Summe von vier Quadraten, im Widerspruch zur Minimalität von $k$. ❑

Nun der Beweis mit Hilfe des Gitterpunktsatzes:

*Zweiter Beweis.* Wir müssen wieder ein geeignetes Gitter $\Lambda$ und eine Menge $S$ konstruieren, diesmal im $\mathbb{R}^4$. Die Wahl von $S$ ist ziemlich klar:

$$S = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 < 2p\}\,.$$

Um das Volumen von $S$ zu berechnen, ist es nützlich, die Volumenformel für die $n$-dimensionale Einheitskugel zu kennen:

$$\operatorname{vol}(B^n) = \frac{\pi^{n/2}}{\left(\frac{n}{2}\right)!}$$

(dabei gilt wie üblich $(\frac{n+1}{2})! = \frac{n+1}{2}(\frac{n-1}{2})!$; außerdem ist noch $(-\frac{1}{2})! = \sqrt{\pi}$). Für $n = 4$ ergibt sich $\operatorname{vol}(B^4) = \pi^2/2$, also ist $\operatorname{vol}(S) = \pi^2(2p)^2/2 = 2\pi^2 p^2$.

Daran kann man schon sehen, dass das Gitter $\Lambda$ Kovolumen $p^2$ haben sollte. Wir sollten also einen Homomorphismus $\phi\colon \mathbb{Z}^4 \to \mathbb{F}_p^2$ finden, sodass für alle $(x_1, x_2, x_3, x_4) \in \ker\phi$ gilt, dass $x_1^2 + x_2^2 + x_3^2 + x_4^2$ durch $p$ teilbar ist.

Seien dazu wieder $u, v \in \mathbb{Z}$ mit $u^2 + v^2 + 1 \equiv 0 \bmod p$. Wir definieren

$$\phi\colon \mathbb{Z}^4 \longrightarrow \mathbb{F}_p^2\,, \quad (x_1, x_2, x_3, x_4) \longmapsto (\bar{x}_2 - \bar{u}\bar{x}_1 + \bar{v}\bar{x}_4, \bar{x}_3 - \bar{u}\bar{x}_4 - \bar{v}\bar{x}_1)\,.$$

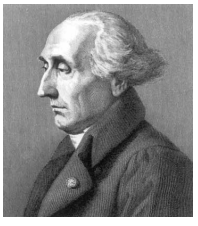
Ist $(x_1, x_2, x_3, x_4) \in \Lambda := \ker\phi$, dann gilt mod $p$:

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 + x_4^2 &\equiv x_1^2 + (ux_1 - vx_4)^2 + (ux_4 + vx_1)^2 + x_4^2 \\ &= (1 + u^2 + v^2)(x_1^2 + x_4^2) \equiv 0\,, \end{aligned}$$

also ist $p$ ein Teiler von $x_1^2 + x_2^2 + x_3^2 + x_4^2$. Es ist klar, dass $\phi$ surjektiv ist, also ist $\Delta(\Lambda) = p^2$. Nun gilt $\operatorname{vol}(S) = 2\pi^2 p^2 > 16p^2 = 2^4\Delta(\Lambda)$, also gibt es

$$(0, 0, 0, 0) \neq (x_1, x_2, x_3, x_4) \in S \cap \Lambda\,,$$

und wie im Beweis des Zwei-Quadrate-Satzes folgt dann $x_1^2 + x_2^2 + x_3^2 + x_4^2 = p$. ❑



J.-L. Lagrange (1736–1813)

Damit folgt der Vier-Quadrate-Satz, dessen erster bekannter Beweis von Lagrange 1770 erbracht wurde.

5.7. **Folgerung.** *Jede nichtnegative ganze Zahl ist Summe von vier Quadratzahlen.*

**FOLG**
4-□-Satz

Jetzt ist es natürlich eine naheliegende Frage, wie es mit Summen von *drei* Quadraten aussieht. Dazu bemerken wir zunächst Folgendes.

**LEMMA**
$m \neq \square+\square+\square$

5.8. **Lemma.** *Ist $m$ von der Form $4^k(8l+7)$ (mit $k, l \geq 0$), dann ist $m$ nicht Summe von drei Quadratzahlen.*

*Beweis.* Zuerst überlegen wir, dass für $m \geq 1$ mit $4m$ auch $m$ Summe von drei Quadraten ist: Gilt $4m = x_1^2 + x_2^2 + x_3^2$, dann ist $x_1^2 + x_2^2 + x_3^2 \equiv 0 \bmod 4$, und das ist nur möglich, wenn $x_1, x_2, x_3$ alle gerade sind. Dann ist aber

$$m = (x_1/2)^2 + (x_2/2)^2 + (x_3/2)^2$$

ebenfalls Summe von drei Quadraten

Es genügt also zu zeigen, dass $m = 8k + 7$ nicht Summe von drei Quadraten sein kann. Das folgt nun aber aus einer Betrachtung modulo 8: Ein Quadrat ist stets $\equiv 0, 1$ oder $4 \bmod 8$; damit kann die Summe dreier Quadrate nicht $\equiv 7 \bmod 8$ sein. ❑

A.-M. Legendre (1752–1833)

Tatsächlich ist das die einzige Einschränkung, wie von Legendre 1797 oder 1798 gezeigt wurde. Gauß gab 1801 einen weiteren Beweis des Satzes.

**SATZ**
3-□-Satz

5.9. **Satz.** *Eine ganze Zahl $m \geq 0$ ist genau dann Summe dreier Quadratzahlen, wenn $m$ nicht in der Form $m = 4^k(8l+7)$ geschrieben werden kann.*

Wir können das jetzt noch nicht beweisen, aber wir können den Satz wenigstens auf eine schwächere Aussage reduzieren.

**LEMMA**
Reduktion auf Lsg. in $\mathbb{Q}$

5.10. **Lemma.** *Ist $m \in \mathbb{Z}$ Summe dreier Quadrate rationaler Zahlen, so ist $m$ auch Summe dreier Quadrate ganzer Zahlen.*

*Beweis.* (Siehe [Sch, S. 198f].) Sei $m = x_1^2 + x_2^2 + x_3^2$ mit $x_1, x_2, x_3 \in \mathbb{Q}$. Wir können annehmen, dass der Hauptnenner $c$ von $x_1, x_2, x_3$ minimal gewählt ist. Wir müssen $c = 1$ zeigen, also nehmen wir $c > 1$ an. Seien $y_1, y_2, y_3$ die zu $x_1, x_2, x_3$ nächstgelegenen ganzen Zahlen (mit willkürlicher Auswahl, wenn es zwei Möglichkeiten gibt). Wir schreiben $\boldsymbol{x} = (x_1, x_2, x_3)$ und $\boldsymbol{y} = (y_1, y_2, y_3)$ und verwenden $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1y_1 + x_2y_2 + x_3y_3$ für das Skalarprodukt. Wir schreiben $|\boldsymbol{x}| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ für die euklidische Länge eines Vektors. Es gilt dann $0 < |\boldsymbol{x} - \boldsymbol{y}|^2 \leq 3/4 < 1$. Außerdem ist

$$c' := c|\boldsymbol{x} - \boldsymbol{y}|^2 = cm - 2\langle c\boldsymbol{x}, \boldsymbol{y} \rangle + c|\boldsymbol{y}|^2 \in \mathbb{Z}\,.$$

Der Punkt

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{2\langle \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{y} \rangle}{|\boldsymbol{x} - \boldsymbol{y}|^2}\,(\boldsymbol{y} - \boldsymbol{x}) = \frac{1}{c'}\big((|\boldsymbol{y}|^2 - m)\,c\boldsymbol{x} + 2(cm - \langle c\boldsymbol{x}, \boldsymbol{y} \rangle)\,\boldsymbol{y}\big)$$

erfüllt ebenfalls $|\boldsymbol{x}'|^2 = m$ ($\boldsymbol{x}'$ ist der zweite Schnittpunkt der Geraden durch $\boldsymbol{x}$ und $\boldsymbol{y}$ mit der Kugeloberfläche $|\boldsymbol{x}|^2 = m$) und hat einen Nenner, der $c' < c$ teilt. Das zeigt, dass $c$ nicht minimal war, und ergibt den gesuchten Widerspruch. ❑

Es bleibt also noch zu zeigen, dass $m$, wenn es nicht die Form $4^k(8l+7)$ hat, als Summe von drei Quadraten rationaler Zahlen geschrieben werden kann. Das folgt aus dem *Hasse-Prinzip* für quadratische Formen, das wir in Abschnitt 8 besprechen werden.

Ein Grund dafür, dass der Drei-Quadrate-Satz schwieriger ist als der Zwei- oder der Vier-Quadrate-Satz, liegt darin, dass die Menge

$$\Sigma_3 = \{x^2 + y^2 + z^2 \mid x, y, z \in \mathbb{Z}\}$$

keine multiplikative Struktur besitzt wie $\Sigma_2$ und $\Sigma_4$. (Zum Beispiel sind 3 und 5 in $\Sigma_3$, $3 \cdot 5 = 15$ jedoch nicht.)

Hier ist noch eine nette Konsequenz des Drei-Quadrate-Satzes. Eine *Dreieckszahl* ist eine ganze Zahl der Form $n(n+1)/2$, also eine Zahl aus der Folge $0, 1, 3, 6, 10, 15, 21, 28, \ldots$.

5.11. **Satz.** *Jede nichtnegative ganze Zahl ist Summe dreier Dreieckszahlen.*

**SATZ**
$n = \triangle + \triangle + \triangle$

*Beweis.* Sei $m \geq 0$ eine ganze Zahl. Dann ist nach dem Drei-Quadrate-Satz 5.9 $8m + 3 = x^2 + y^2 + z^2$ als Summe dreier Quadrate darstellbar. Dabei müssen $x, y, z$ ungerade sein (Betrachtung mod 4). Wir schreiben $x = 2u + 1$, $y = 2v + 1$, $z = 2w + 1$. Es folgt

$$\begin{aligned} m &= \frac{1}{8}\big((2u+1)^2 - 1\big) + \frac{1}{8}\big((2v+1)^2 - 1\big) + \frac{1}{8}\big((2w+1)^2 - 1\big) \\ &= \frac{u(u+1)}{2} + \frac{v(v+1)}{2} + \frac{w(w+1)}{2}\,. \end{aligned}$$

❑

Eine weitere Folgerung ist:

5.12. **Folgerung.** *Sei $f(x, y, z) = x^2 + y^2 + z^2 + z \in \mathbb{Z}[x, y, z]$. Für $x, y, z \in \mathbb{Z}$ gilt dann $f(x, y, z) \geq 0$, und jede nichtnegative ganze Zahl $n$ kann in der Form $n = f(x, y, z)$ mit $x, y, z \in \mathbb{Z}$ geschrieben werden.*

**FOLG**
$\mathbb{Z}_{\geq 0} = f(\mathbb{Z}^3)$

*Beweis.* Übung. ❑

Es gibt eine analoge Identität wie in Lemma 5.1 und in Lemma 5.4 für *acht* Quadrate. (Dahinter steckt die Algebra der *Octonionen* oder *Oktaven*, deren Multiplikation nur noch eine schwächere Bedingung als die Assoziativität erfüllt. Siehe zum Beispiel [Z, § 9].) Hurwitz hat 1898 bewiesen, dass es solche Identitäten nur für Summen von 1, 2, 4 oder 8 Quadraten geben kann. Siehe [Z, § 10].

Man kann sich auch fragen, *wie viele* Möglichkeiten es gibt, eine gegebene natürliche Zahl $m$ als Summe von zwei oder vier Quadraten zu schreiben. Dafür gibt es die folgenden Formeln:

$$R_2(m) := \#\{(x, y) \in \mathbb{Z}^2 \mid x^2 + y^2 = m\} = 4 \sum_{d \mid m} \chi(d)$$

$$R_4(m) := \#\{(x_1, x_2, x_3, x_4) \in \mathbb{Z}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 = m\} = 8 \sum_{d \mid m, 4 \nmid d} d$$

Die Summen laufen jeweils über die positiven Teiler von $m$, und

$$\chi(d) = \begin{cases} 0 & \text{falls } d \text{ gerade,} \\ 1 & \text{falls } d \equiv 1 \bmod 4, \\ -1 & \text{falls } d \equiv 3 \bmod 4. \end{cases}$$

Eine andere natürliche Frage ist die folgende (Waring 1770):

*Gibt es für jedes $k \geq 1$ eine Zahl $g(k)$, sodass jede natürliche Zahl Summe von höchstens $g(k)$ $k$-ten Potenzen natürlicher Zahlen ist?*

Der Vier-Quadrate-Satz sagt, dass $g(2) = 4$ ist. Waring vermutete $g(3) = 9$ und $g(4) = 19$. Hilbert bewies 1909, dass Warings Frage eine positive Antwort hat. Euler vermutete bereits, dass

$$g(k) = 2^k + \left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 2$$

für alle $k$ gilt. (In jedem Fall gilt hier „$\geq$“, da dies die Maximalzahl von $k$-ten Potenzen ist, die man für die Zahlen bis $3^k - 1$ braucht.) Heute ist bekannt, dass das zutrifft, falls

$$2^k\Big(\Big(\frac{3}{2}\Big)^k - \Big\lfloor\Big(\frac{3}{2}\Big)^k\Big\rfloor\Big) + \Big\lfloor\Big(\frac{3}{2}\Big)^k\Big\rfloor \leq 2^k$$

gilt, was vermutungsweise immer der Fall ist. In jedem Fall kann es nur endlich viele Ausnahmen geben (und man hätte dann ebenfalls eine Formel für $g(k)$).

Weit schwieriger ist die Frage, was die kleinste Zahl $G(k)$ ist, sodass jede *hinreichend große* natürliche Zahl Summe von $G(k)$ $k$-ten Potenzen ist. Die einzigen bekannten Werte sind $G(2) = 4$ und $G(4) = 16$; sonst gibt es nur untere und obere Schranken, wie zum Beispiel $4 \leq G(3) \leq 7$ (mit der Vermutung $G(3) = 4$).

## 6. Ternäre quadratische Formen

Wir haben bereits einige Beispiele von *quadratischen Formen* gesehen. Wir erinnern uns an die Definition (vgl. Lineare Algebra II, § 23):

**DEF** quadratische Form

6.1. **Definition.** Eine *quadratische Form* (über $\mathbb{Z}$) in $n$ Variablen ist ein homogenes Polynom vom Grad 2 in $n$ Variablen mit ganzzahligen Koeffizienten. (Man kann quadratische Formen über beliebigen Ringen betrachten.)

Ist $n = 2, 3, 4, \ldots$, so spricht man auch von *binären*, *ternären*, *quaternären*, ... quadratischen Formen. Binäre quadratische Formen haben also die Form

$$Q(x, y) = a\,x^2 + b\,xy + c\,y^2 \qquad \text{mit } a, b, c \in \mathbb{Z}$$

und ternäre quadratische Formen haben die Form

$$Q(x, y, z) = a\,x^2 + b\,y^2 + c\,z^2 + d\,xy + e\,yz + f\,zx$$

mit $a, b, c, d, e, f \in \mathbb{Z}$. ♢

Im vorigen Abschnitt ging es um *Darstellungen* von Zahlen durch die quadratischen Formen $x^2 + y^2$ und $x_1^2 + x_2^2 + x_3^2 + x_4^2$. Eine andere Frage, die man stellen kann, ist, ob eine gegebene quadratische Form eine nichttriviale Nullstelle hat. Im Falle einer ternären quadratischen Form $Q(x, y, z)$ wäre die Frage also, ob es $(0, 0, 0) \neq (x, y, z) \in \mathbb{Z}^3$ gibt mit $Q(x, y, z) = 0$. Damit werden wir uns im Folgenden beschäftigen.

Für *binäre* quadratische Formen ist das keine besonders interessante Frage: Die Existenz einer nichttrivialen Nullstelle ist dazu äquivalent, dass die Form in ein Produkt von zwei Linearformen zerfällt, was genau dann der Fall ist, wenn die *Diskriminante* $b^2 - 4ac$ ein Quadrat ist. Für ternäre Formen ergibt sich aber ein durchaus interessantes Problem.

Wir bemerken, dass wir immer annehmen können, dass eine (nichttriviale) Lösung von $Q(x, y, z) = 0$ *primitiv* ist, d.h. $\mathrm{ggT}(x, y, z) = 1$ erfüllt, denn wir können etwaige gemeinsame Teiler immer abdividieren.

**DEF** Determinante, Diskriminante einer qu. Form (nicht-) ausgeartet singulär

6.2. **Definition.** Eine quadratische Form $Q$ in $n$ Variablen kann auch durch eine symmetrische Matrix $M_Q$ beschrieben werden (mit ganzzahligen Diagonaleinträgen und evtl. halbganzen sonstigen Einträgen), sodass $Q(\boldsymbol{x}) = \boldsymbol{x} M_Q \boldsymbol{x}^\top$ gilt. ($\boldsymbol{x} = (x_1, \ldots, x_n)$ als Zeilenvektor.) Dann nennen wir

$$\det Q = \det(M_Q)$$

die *Determinante* von $Q$, und

$$\operatorname{disc} Q = (-1)^{\binom{n}{2}} 4^{\lfloor n/2 \rfloor} \det Q$$

heißt die *Diskriminante* von $Q$; die Diskriminante ist immer eine ganze Zahl. (Die Potenz von 4 in der Definition von $\operatorname{disc}(Q)$ dient dazu, die Diskriminante ganzzahlig zu machen.)

Zum Beispiel ist

$$\operatorname{disc}(a\,x^2 + b\,xy + c\,y^2) = b^2 - 4\,ac$$

und

$$\operatorname{disc}(a\,x^2 + b\,y^2 + c\,z^2 + d\,xy + e\,yz + f\,zx) = -4\,abc - def + ae^2 + bf^2 + cd^2\,.$$

Eine quadratische Form $Q$ ist *nicht-ausgeartet*, wenn $\operatorname{disc} Q \neq 0$ ist, sonst ist sie *ausgeartet* oder *singulär*. Wenn $Q$ singulär ist, dann gibt es eine lineare Substitution der Variablen, die $Q$ in eine quadratische Form in weniger Variablen als vorher transformiert. (Wähle dazu ein primitives Element des Kerns von $M_Q$ als einen der neuen Basisvektoren.) ♢

**Etwas Geometrie.**

Ternäre quadratische Formen entsprechen *Kegelschnitten* in der Ebene. Wenn wir etwa nach reellen Lösungen von $Q(x, y, z) = 0$ mit (z.B.) $z \neq 0$ suchen, dann können wir die Gleichung durch $z^2$ teilen und $\xi = x/z$, $\eta = y/z$ setzen; wir erhalten dann $Q(\xi, \eta, 1) = 0$; das ist die Gleichung eines Kegelschnitts. (Wenn wir in der *projektiven* Ebene arbeiten, dann brauchen wir die Punkte mit $z = 0$ nicht auszuschließen: Sie kommen auf der unendlich fernen Geraden zu liegen.) Primitive ganzzahlige Lösungen von $Q(x, y, z) = 0$ entsprechen dann *rationalen Punkten* (Punkten mit rationalen Koordinaten) auf dem Kegelschnitt. Dabei entsprechen jeweils die zwei primitiven ganzzahligen Lösungen $(x, y, z)$ und $(-x, -y, -z)$ dem rationalen Punkt $(x/z, y/z)$.

Zum Beispiel (wir haben das bereits ganz am Anfang gesehen) gehört zur quadratischen Form $Q(x, y, z) = x^2+y^2-z^2$ der Einheitskreis, und die primitiven Lösungen (in diesem Fall sind das die primitiven pythagoreischen Tripel) entsprechen den rationalen Punkten auf dem Einheitskreis (es gibt keine Lösungen mit $z = 0$). Wir haben gesehen, wie man ausgehend von dem rationalen Punkt $(-1, 0)$ alle Punkte parametrisieren kann, indem man den zweiten Schnittpunkt von Geraden durch den gewählten Punkt mit dem Kegelschnitt betrachtet. Die gleiche Konstruktion funktioniert mit jedem nicht-ausgearteten Kegelschnitt.

**SATZ**
Parametrisierung von Kegelschnitten

6.3. **Satz.** *Sei $Q(x, y, z)$ eine nicht-ausgeartete ternäre quadratische Form und sei $(x_0, y_0, z_0)$ eine primitive Lösung von $Q(x, y, z) = 0$. Dann gibt es binäre quadratische Formen $R_x$, $R_y$ und $R_z$, sodass bis auf Multiplikation mit einem gemeinsamen (rationalen) Faktor alle ganzzahligen Lösungen von $Q(x, y, z) = 0$ gegeben sind durch*

$$\big(R_x(u, v), R_y(u, v), R_z(u, v)\big)$$

*mit ganzen Zahlen $u$ und $v$.*

*Beweis.* Wir geben hier einen „algebraischen“ Beweis. Man kann auch einen „geometrischen“ Beweis geben analog zu dem für die pythagoreischen Tripel. Unser Beweis hier hat den Vorteil, eine „minimale“ Parametrisierung zu liefern, d.h. eine mit $|\operatorname{disc}(R_x)|$, $|\operatorname{disc}(R_y)|$, $|\operatorname{disc}(R_z)|$ so klein wie möglich.

Wir betrachten erst einmal den Fall $Q = y^2 - xz$. Dann können wir

$$R_x(u, v) = u^2\,, \quad R_y(u, v) = uv\,, \quad R_z(u, v) = v^2$$

wählen. (Siehe Lemma 2.1.)

Als Nächstes nehmen wir an, dass $(x_0, y_0, z_0) = (1, 0, 0)$ ist. Dann ist

$$Q(x, y, z) = b\,y^2 + c\,z^2 + d\,xy + e\,yz + f\,zx$$

(ohne $x^2$-Term). Wenn wir

$$x = b\,X + e\,Y + c\,Z\,, \quad y = -d\,X - f\,Y\,, \quad z = -d\,Y - f\,Z$$

setzen, dann wird $Q(x,y,z) = \mathrm{disc}(Q)\cdot(Y^2 - XZ)$, wie man leicht nachprüft. Da $\mathrm{disc}(Q) \neq 0$ ist, folgt mit dem zuerst betrachteten Spezialfall, dass

$$R_x(u,v) = b\,u^2 + e\,uv + c\,v^2\,, \quad R_y(u,v) = -d\,u^2 - f\,uv\,, \quad R_z(u,v) = -d\,uv - f\,v^2$$

geeignete binäre Formen sind.

Jetzt betrachten wir den allgemeinen Fall. Da $\mathrm{ggT}(x_0,y_0,z_0) = 1$ ist, gibt es eine invertierbare ganzzahlige Matrix $T$ mit $\det(T) = 1$ (also $T \in \mathrm{SL}(3,\mathbb{Z})$), sodass $(x_0\ y_0\ z_0) = (1\ 0\ 0)T$ ist (d.h., $(x_0\ y_0\ z_0)$ ist die erste Zeile von $T$). Wir setzen

$$(x\ y\ z) = (x'\ y'\ z')T$$

und $Q'(x',y',z') = Q(x,y,z)$ (also $M_{Q'} = TM_QT^\top$); dann ist

$$Q'(1,0,0) = Q(x_0,y_0,z_0) = 0\,.$$

Nach dem gerade betrachteten Fall gibt es binäre quadratische Formen $R'_x, R'_y, R'_z$, die die Lösungen von $Q' = 0$ parametrisieren. Dann sind

$$(R_x\ R_y\ R_z) = (R'_x\ R'_y\ R'_z)T$$

die gesuchten binären quadratischen Formen für $Q$. ❑

**FOLG** Normalform lösbarer ternärer qu. Formen

6.4. **Folgerung.** *Ist $Q(x,y,z)$ eine nicht-ausgeartete ternäre quadratische Form, sodass $Q = 0$ eine nichttriviale Lösung hat, dann gibt es eine ganzzahlige lineare Substitution $(x\ y\ z) = (X\ Y\ Z)T$ mit $\det(T) = \mathrm{disc}(Q)$, sodass*

$$Q(x,y,z) = \mathrm{disc}(Q)(Y^2 - XZ)$$

*gilt.*

*Beweis.* Das folgt aus dem vorigen Beweis. Man beachte, dass im Fall $Q(1,0,0) = 0$ die Transformationsmatrix

$$T = \begin{pmatrix} b & -d & 0 \\ e & -f & -d \\ c & 0 & -f \end{pmatrix}$$

die Gleichung $\det(T) = bf^2 + cd^2 - def = \mathrm{disc}(Q)$ erfüllt. Im allgemeinen Fall ändern sich die Diskriminante der quadratischen Form und die Determinante der Transformationsmatrix nicht. ❑

**BSP** pyth. Tripel

6.5. **Beispiel.** Für $Q(x,y,z) = x^2 + y^2 - z^2$ und die Ausgangslösung $(-1,0,1)$ können wir die Matrix $T$ im Beweis von Satz 6.3 wählen als

$$T = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

damit ist $x = -x'$, $y = -y'$, $z = x' + z'$ und

$$Q'(x',y',z') = Q(-x',-y',x'+z') = (y')^2 - (z')^2 - 2x'z'\,.$$

Die quadratischen Formen für $Q'$ sind

$$R'_x(u,v) = u^2 - v^2\,, \quad R'_y(u,v) = 2\,uv\,, \quad R'_z(u,v) = 2\,v^2\,.$$

Für die ursprüngliche Form $Q$ bekommen wir dann

$$\begin{array}{lllll} R_x(u,v) = & -R'_x(u,v) & & & = v^2 - u^2 \\ R_y(u,v) = & & -R'_y(u,v) & & = -2\,uv \\ R_z(u,v) = & R'_x(u,v) & & +R'_z(u,v) & = u^2 + v^2 \end{array}$$

Das ist, bis aufs Vorzeichen, wieder genau die bekannte Parametrisierung der pythagoreischen Tripel. ♣

Wir sehen also, dass wir leicht *alle* Lösungen finden können, wenn wir erst einmal *eine* kennen. Es bleiben noch zwei Fragen zu beantworten: Wie können wir feststellen, *ob* es eine Lösung gibt? Und wie können wir, wenn es eine gibt, eine Lösung *finden*?

Dabei ist es hilfreich, sich darauf beschränken zu können, nur „diagonale“ Formen der speziellen Gestalt

$$Q(x,y,z) = ax^2 + by^2 + cz^2$$

zu betrachten. Dazu brauchen wir einen Äquivalenzbegriff für quadratische Formen.

6.6. **Definition.** Seien $Q$, $Q'$ zwei quadratische Formen in derselben Zahl $n$ von Variablen. Wir nennen $Q$ und $Q'$ *äquivalent*, wenn

**DEF** Äquivalenz von qu. Formen

$$\begin{aligned} Q'(x_1, x_2, \dots, x_n) = \lambda\, Q(&a_{11}x_1 + a_{21}x_2 + \dots + a_{n1}x_n,\\ &a_{12}x_1 + a_{22}x_2 + \dots + a_{n2}x_n,\\ &\dots\\ &a_{1n}x_1 + a_{2n}x_2 + \dots + a_{nn}x_n) \end{aligned}$$

ist mit $\lambda \in \mathbb{Q}^\times$ und einer Matrix

$$T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \in \mathrm{GL}(n, \mathbb{Q})\,.$$

(Es ist dann $Q'(\boldsymbol{x}) = \lambda Q(\boldsymbol{x}T)$.) Für die zugehörigen symmetrischen Matrizen bedeutet das $M_{Q'} = \lambda T M_Q T^\top$; insbesondere folgt $\mathrm{disc}(Q') = \lambda^n \det(T)^2 \mathrm{disc}(Q)$, sodass $Q'$ genau dann nicht-ausgeartet ist, wenn das für $Q$ gilt. Es ist klar, dass wir eine Äquivalenzrelation definiert haben. ♢

Ist $\boldsymbol{x} = (x_1, \dots, x_n)$ eine nichttriviale rationale Lösung von $Q(x_1, x_2, \dots, x_n) = 0$, dann ist $\boldsymbol{x}' = \boldsymbol{x}T^{-1}$ eine nichttriviale Lösung von $Q'(x_1, x_2, \dots, x_n) = 0$, und ist $\boldsymbol{x}'$ eine nichttriviale Lösung von $Q'(x_1, x_2, \dots, x_n) = 0$, dann ist $\boldsymbol{x} = \boldsymbol{x}'T$ eine nichttriviale Lösung von $Q(x_1, x_2, \dots, x_n) = 0$. Da die Existenz einer primitiven ganzzahligen Lösung zur Existenz einer nichttrivialen rationalen Lösung äquivalent ist, haben wir das folgende Resultat gezeigt:

6.7. **Lemma.** *Sind $Q$ und $Q'$ äquivalente quadratische Formen und hat $Q = 0$ eine primitive ganzzahlige Lösung, so hat auch $Q' = 0$ eine primitive ganzzahlige Lösung, und umgekehrt.*

**LEMMA** Lösbarkeit von äquiv. Formen

Folgerung 6.4 lässt sich dann auch so formulieren: Eine nicht-ausgeartete ternäre quadratische Form $Q$ ist genau dann äquivalent zu $y^2 - xz$, wenn es nichttriviale ganzzahlige Lösungen von $Q = 0$ gibt.

**Bemerkung.** Wenn man sich für die *Werte* einer quadratischen Form für ganzzahlige Argumente interessiert statt für ihre Nullstellen, dann muss man einen eingeschränkteren Äquivalenzbegriff verwenden: Skalieren ist nicht erlaubt ($\lambda = 1$), und die Matrix $T$ muss sogar in $\mathrm{GL}(n, \mathbb{Z})$ sein. Unter dieser Voraussetzung sind die Wertemengen von $Q$ und $Q'$ gleich.

Jetzt zeigen wir, dass wir jede quadratische Form „diagonalisieren“ können. (Das wurde für quadratische Formen über beliebigen Körpern der Charakteristik $\neq 2$ auch in der Linearen Algebra II gezeigt.)

**SATZ**
qu. Formen sind diagonalisierbar

6.8. **Satz.** *Sei $Q$ eine quadratische Form in $n$ Variablen. Dann ist $Q$ äquivalent zu einer diagonalen quadratischen Form, d.h., einer Form der Gestalt*

$$Q'(x_1, \ldots, x_n) = a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2 \,.$$

*$Q$ ist genau dann nicht-ausgeartet, wenn $a_1, \ldots, a_n \neq 0$ sind.*

*Beweis.* Wir beweisen zunächst die letzte Aussage. $Q$ ist genau dann nicht-ausgeartet, wenn die Diagonalform nicht-ausgeartet ist. Die Determinante der Diagonalform ist $a_1 a_2 \cdots a_n$, also ist die Diagonalform genau dann nicht-ausgeartet, wenn $a_1, \ldots, a_n \neq 0$ ist.

Die Methode für den Beweis ist sukzessives quadratisches Ergänzen. Der Beweis wird durch Induktion nach $n$ geführt. Der Fall $n = 1$ ist trivial, denn dann ist $Q(x_1) = ax_1^2$ bereits diagonal.

Sei also $n \geq 2$. Wenn $Q$ nicht von $x_1$ abhängt, dann folgt die Behauptung direkt aus der Induktionsannahme (wobei $a_1 = 0$ ist). Im anderen Fall nehmen wir erst einmal an, dass der Koeffizient von $x_1^2$ in $Q$ nicht null ist. In diesem Beweis lassen wir rationale (statt nur ganzzahlige) Koeffizienten zu; am Ende können wir die Nenner wieder wegmultiplizieren. Dann können wir erst einmal $Q$ durch den Koeffizienten von $x_1^2$ teilen. Danach sieht $Q$ so aus:

$$Q(x_1, x_2, \ldots, x_n) = x_1^2 + b_2 x_1 x_2 + b_3 x_1 x_3 + \cdots + b_n x_1 x_n + Q_1(x_2, \ldots, x_n)$$

Hier ist $Q_1$ eine quadratische Form in $n-1$ Variablen. Wir ersetzen jetzt $x_1$ durch $x_1 - \frac{1}{2}(b_2 x_2 + \cdots + b_n x_n)$, dann bekommen wir

$$Q'(x_1, x_2, \ldots, x_n) = x_1^2 + Q_1'(x_2, \ldots, x_n) \,.$$

Nach Induktionsannahme gibt es eine invertierbare lineare Substitution der Variablen $x_2, \ldots, x_n$, die $Q_1'$ diagonalisiert. Anwendung auf $Q'$ liefert

$$Q''(x_1, x_2, \ldots, x_n) = x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 + \cdots + \alpha_n x_n^2 \,.$$

Durch Multiplikation mit dem Hauptnenner der rationalen Zahlen $\alpha_j$ bekommen wir eine diagonale quadratische Form mit ganzzahligen Koeffizienten.

Wir müssen uns noch davon überzeugen, dass wir $Q$ immer so transformieren können, dass der Koeffizient von $x_1^2$ nicht verschwindet. Sei also der Koeffizient von $x_1^2$ in $Q$ gleich null. Weil $Q$ von $x_1$ abhängt, gibt es jedenfalls einen Index $2 \leq j \leq n$, sodass der Koeffizient von $x_1 x_j$ nicht verschwindet. Sei also

$$Q(x_1, x_2, \ldots, x_n) = a x_1 x_j + b x_j^2 + \ldots$$

mit $a \neq 0$, wobei die Punkte für Terme stehen, die ein $x_k$ mit $k \notin \{1, j\}$ enthalten. Sei $c \in \mathbb{Q}^\times$ mit $a + bc \neq 0$. Wenn wir $x_j$ durch $x_j + cx_1$ ersetzen, erhalten wir eine quadratische Form, in der der Koeffizient von $x_1^2$ gegeben ist durch $ac + bc^2 \neq 0$. ❑

6.9. **Beispiel.** Sei $Q(x,y,z) = xy+yz+zx$. Um $Q$ nach dem Verfahren im obigen Beweis zu diagonalisieren, müssen wir erst einmal einen Term mit $x^2$ erzeugen. Dazu ersetzen wir $y$ durch $y+x$ und erhalten

**BSP** Diagonalisierung

$$Q_1(x,y,z) = x(y+x) + (y+x)z + zx = x^2 + xy + 2xz + yz\,.$$

Jetzt ersetzen wir $x$ durch $x - \frac{1}{2}y - z$ (quadratische Ergänzung); das ergibt

$$Q_2(x,y,z) = x^2 - (\tfrac{1}{2}y + z)^2 + yz = x^2 - \tfrac{1}{4}y^2 - z^2\,.$$

Da das Ergebnis ganzzahlige Koeffizienten haben soll, ersetzen wir noch $y$ durch $2y$ und erhalten die diagonale Form

$$Q'(x,y,z) = x^2 - y^2 - z^2\,.$$

Die Matrix $T$ ist hier

$$T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}\,,$$

d.h., $Q'(x,y,z) = Q(x-y-z, x+y-z, z)$. ♣

In praktischen Anwendungen, wenn wir tatsächlich Lösungen berechnen wollen, ist es meistens *keine* gute Idee, die gegebene Form zu diagonalisieren. Es werden dabei nämlich zunächst Nenner eingeführt, die dann am Ende wieder eliminiert werden müssen, wodurch die Diskriminante mit Faktoren multipliziert wird, die wir schlecht kontrollieren können. Da man zur Berechnung einer Lösung die Diskriminante faktorisieren muss, kann das sehr nachteilig sein. Für theoretische Untersuchungen spielt dieser Gesichtspunkt allerdings keine Rolle. Außerdem werden wir auch einen Algorithmus angeben, der ohne Diagonalisierung auskommt, sodass wir in der Praxis das angesprochene Problem umgehen können.

Wir werden jetzt also erst einmal annehmen, die zu untersuchende ternäre quadratische Form sei diagonal:

$$Q(x,y,z) = ax^2 + by^2 + cz^2\,.$$

Wir können natürlich annehmen, dass $\mathrm{ggT}(a,b,c) = 1$ ist (sonst teilen wir die Form durch den ggT von $a,b,c$). Wenn einer der Koeffizienten, sagen wir $a$, durch ein Quadrat $d^2$ (mit $d>1$) teilbar ist, dann können wir $x$ ersetzen durch $x/d$; das hat den Effekt, dass $a$ durch $a/d^2$ ersetzt wird. Wir können also auch annehmen, dass $a$, $b$ und $c$ *quadratfrei* sind. (Um das festzustellen bzw. zu erreichen, müssen wir die Koeffizienten allerdings faktorisieren!)

Wenn jetzt zwei der Koeffizienten, sagen wir $b$ und $c$, einen gemeinsamen Teiler $d>1$ haben (z.B. $d = \mathrm{ggT}(a,b)$), dann können wir $x$ durch $dx$ ersetzen und die Form durch $d$ teilen; dadurch wird aus $(a,b,c)$ das neue Koeffiziententripel $(da, b/d, c/d)$ mit kleinerem Absolutbetrag des Produkts. Wir können einen solchen Schritt also nur endlich oft durchführen, und danach müssen die Koeffizienten *paarweise teilerfremd* sein. (Hierfür ist eine Faktorisierung nicht nötig.)

Wir können also annehmen, dass $a,b,c$ quadratfrei und paarweise teilerfremd sind. Das ist äquivalent dazu, dass das Produkt $abc$ quadratfrei ist. Wir halten das fest:

**LEMMA** Normalform für ternäre qu. Formen

6.10. **Lemma.** *Jede nicht-ausgeartete ternäre quadratische Form ist äquivalent zu einer Diagonalform $ax^2 + by^2 + cz^2$, in der $a, b, c$ paarweise teilerfremd und quadratfrei sind.*

Durch Betrachtung der reellen Lösbarkeit und der Lösbarkeit modulo Potenzen von Primzahlen erhalten wir die folgenden notwendigen Bedingungen für die Lösbarkeit einer ternären quadratischen Form:

**LEMMA** Notwendige Bedingungen für die Lösbarkeit

6.11. **Lemma.** *Sei $abc$ quadratfrei und sei $(x_0, y_0, z_0)$ eine primitive ganzzahlige Lösung von $ax^2+by^2+cz^2 = 0$. Dann sind $ax_0^2$, $by_0^2$ und $cz_0^2$ paarweise teilerfremd, und es müssen folgende Bedingungen an $a, b, c$ erfüllt sein:*

(1) *$a$, $b$ und $c$ haben nicht alle dasselbe Vorzeichen.*

(2) *Wenn $abc$ ungerade ist, dann sind $a$, $b$ und $c$ nicht alle zueinander kongruent mod 4.*

(3) *Wenn $a$ gerade ist, dann ist entweder $b + c \equiv 0$ oder $a + b + c \equiv 0$ mod 8.*

(4) *Wenn $b$ gerade ist, dann ist entweder $a + c \equiv 0$ oder $a + b + c \equiv 0$ mod 8.*

(5) *Wenn $c$ gerade ist, dann ist entweder $a + b \equiv 0$ oder $a + b + c \equiv 0$ mod 8.*

(6) *Wenn $p$ ein ungerader Primteiler von $a$ ist, dann ist $\left(\frac{-bc}{p}\right) = 1$.*

(7) *Wenn $p$ ein ungerader Primteiler von $b$ ist, dann ist $\left(\frac{-ca}{p}\right) = 1$.*

(8) *Wenn $p$ ein ungerader Primteiler von $c$ ist, dann ist $\left(\frac{-ab}{p}\right) = 1$.*

*Beweis.* Wir zeigen zunächst, dass $ax_0^2$, $by_0^2$, $cz_0^2$ paarweise teilerfremd sind. Wir nehmen an, dass eine Primzahl $p$ zwei der Terme teilt, und leiten daraus einen Widerspruch ab. Es muss $p$ dann auch den dritten Term teilen. Da $a$, $b$, $c$ paarweise teilerfremd sind, kann $p$ höchstens einen der Koeffizienten teilen. Dann muss $p$ mindestens zwei der Zahlen $x_0$, $y_0$ und $z_0$ teilen. Da $p^2$ keinen der Koeffizienten teilt, müsste dann auch die dritte der Zahlen durch $p$ teilbar sein, im Widerspruch zu $\mathrm{ggT}(x_0, y_0, z_0) = 1$.

Aussage (1) ist klar: Hätten $a, b, c$ dasselbe Vorzeichen, dann wäre $ax^2 + by^2 + cz^2$ immer positiv oder immer negativ. Zum Beweis von (2) und (3–5) beachten wir, dass von den drei Termen $ax_0^2$, $by_0^2$, $cz_0^2$ genau zwei ungerade sein müssen. Wenn $abc$ ungerade ist, liefert das die Bedingung $a+b \equiv 0$ oder $b+c \equiv 0$ oder $a+c \equiv 0 \bmod 4$; das ist Aussage (2). Wenn zum Beispiel $a$ gerade ist, dann müssen $y_0$ und $z_0$ ungerade sein; $x_0$ kann gerade oder ungerade sein. Betrachtung modulo 8 liefert dann Aussage (3).

Zum Beweis von (6) ((7) und (8) werden ebenso bewiesen) beachten wir, dass $y_0$ und $z_0$ nicht durch $p$ teilbar sein können. Wir haben $by_0^2 + cz_0^2 \equiv 0 \bmod p$, also $(by_0)^2 \equiv -bc \cdot z_0^2 \bmod p$, und weil $z_0$ mod $p$ invertierbar ist, muss $-bc$ ein quadratischer Rest mod $p$ sein. ❑

Für ungerade Primzahlen $p$, die keinen der Koeffizienten teilen, erhalten wir keine Bedingungen, denn nach Lemma 5.5 gibt es immer nichttriviale Lösungen mod $p$.

Um diese notwendigen Bedingungen zu überprüfen (und übrigens auch, um die „Normalform“ mit $abc$ quadratfrei herzustellen), müssen wir die Koeffizienten $a$, $b$, $c$ *faktorisieren*. Man kann zeigen, dass es nicht einfacher gehen kann: wenn man

Nullstellen diagonaler ternärer quadratischer Formen berechnen kann, dann kann man das benutzen, um ganze Zahlen zu faktorisieren.

Es stellt sich heraus, dass diese notwendigen Bedingungen sogar schon hinreichend sind, wie Legendre 1785 gezeigt hat.

**SATZ**
Satz von Legendre über ternäre qu. Formen

6.12. **Satz.** *Sei $Q(x,y,z) = ax^2+by^2+cz^2$ mit $abc$ quadratfrei. Wenn $a$, $b$, $c$ die Bedingungen in 6.11 erfüllen, dann gibt es eine primitive ganzzahlige Lösung von $Q(x,y,z) = 0$.*

A.-M. Legendre (1752–1833)

*Beweis.* Wir geben hier einen Beweis mithilfe des Gitterpunktsatzes 4.5. Wir müssen also wieder ein geeignetes Gitter $\Lambda$ und eine passende symmetrische und konvexe Menge $S$ konstruieren.

Zuerst das Gitter. Sei $D = |abc|$. Wir wollen ein Gitter $\Lambda \subset \mathbb{Z}^3$ konstruieren mit Kovolumen $\Delta(\Lambda) \le 2D$, sodass für $(x,y,z) \in \Lambda$ gilt, dass $2D$ den Wert $Q(x,y,z)$ teilt. Nach Voraussetzung gibt es für jeden Primteiler $p$ von $a$ ein $u_p \in \mathbb{Z}$ mit $bu_p^2 + c \equiv 0 \bmod p$ (auch für $p = 2$, dann ist die Aussage trivial). Nach dem Chinesischen Restsatz gibt es dann $u \in \mathbb{Z}$ mit $u \equiv u_p \bmod p$ für alle $p \mid a$; daraus folgt $bu^2 + c \equiv 0 \bmod a$. Entsprechend gibt es $v, w \in \mathbb{Z}$ mit $cv^2 + a \equiv 0 \bmod b$ und $aw^2 + b \equiv 0 \bmod c$. (Dass die Existenz solcher $u$, $v$, $w$ eine notwendige Bedingung für die Lösbarkeit der Gleichung ist, kann man auch leicht direkt sehen.) Wir brauchen noch ein wenig Information mod 2. Falls $abc$ ungerade ist, setzen wir $\phi_2(x,y,z) = \bar{x} + \bar{y} + \bar{z} \in \mathbb{Z}/2\mathbb{Z}$. Falls $a$ gerade ist, müssen $b$ und $c$ beide ungerade sein. Wir schreiben $b + c = 2m$ und setzen $\phi_2(x,y,z) = \bar{x} + \bar{m}\bar{y} \in \mathbb{Z}/2\mathbb{Z}$. Falls $b$ oder $c$ gerade sind, verfahren wir analog. Wir definieren jetzt

$$\begin{aligned}\Lambda &= \{(x,y,z) \in \mathbb{Z}^3 \mid y \equiv uz \bmod a, z \equiv vx \bmod b, x \equiv wy \bmod c, \phi_2(x,y,z) = \bar{0}\} \\ &= \ker\big((x,y,z) \longmapsto (\bar{y} - \bar{u}\bar{z}, \bar{z} - \bar{v}\bar{x}, \bar{x} - \bar{w}\bar{y}, \phi_2(x,y,z)) \\ &\qquad\qquad \in \mathbb{Z}/a\mathbb{Z} \times \mathbb{Z}/b\mathbb{Z} \times \mathbb{Z}/c\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}\big)\,.\end{aligned}$$

Damit ist klar, dass $\Lambda \subset \mathbb{R}^3$ ein Gitter ist mit $\Delta(\Lambda) \le 2|abc| = 2D$ (es ist nicht schwer zu sehen, dass tatsächlich $\Delta(\Lambda) = 2D$ ist). Wir müssen noch die Teilbarkeitsaussage beweisen. Sei also $(x,y,z) \in \Lambda$. Dann gilt

$$\begin{aligned}ax^2 + by^2 + cz^2 &\equiv b(uz)^2 + cz^2 = (bu^2 + c)z^2 \equiv 0 \bmod a \\ ax^2 + by^2 + cz^2 &\equiv ax^2 + c(vx)^2 = (cv^2 + a)x^2 \equiv 0 \bmod b \\ ax^2 + by^2 + cz^2 &\equiv a(wy)^2 + by^2 = (aw^2 + b)y^2 \equiv 0 \bmod c\,.\end{aligned}$$

Da $a$, $b$, $c$ paarweise teilerfremd sind, folgt jedenfalls schon $D \mid Q(x,y,z)$. Wenn $abc$ ungerade ist, gilt zusätzlich

$$ax^2 + by^2 + cz^2 \equiv x + y + z \equiv 0 \bmod 2\,.$$

Wenn (z.B.) $a$ gerade ist, dann ist $a \equiv 2 \bmod 4$, und wir haben $y \equiv uz \equiv z \bmod 2$ ($u$ muss ungerade sein). Es folgt $y^2 \equiv z^2 \bmod 4$, und mit $b + c = 2m$ dann

$$ax^2 + by^2 + cz^2 \equiv 2x^2 + (b+c)y^2 \equiv 2\big(x + my\big) \equiv 0 \bmod 4\,,$$

weil $\phi_2(x,y,z) = \bar{x} + \bar{m}\bar{y} = \bar{0} \in \mathbb{Z}/2\mathbb{Z}$ ist. In beiden Fällen ergibt sich sogar $2D \mid Q(x,y,z)$, wie behauptet.

Nach Voraussetzung haben die Koeffizienten $a$, $b$, $c$ nicht alle dasselbe Vorzeichen. Sei etwa das Vorzeichen von $c$ anders als das von $a$ und $b$. Dann nehmen wir als Menge $S$ den elliptischen Zylinder

$$S = \{(x,y,z) \in \mathbb{R}^3 : |a|x^2 + |b|y^2 < 2D \text{ und } |c|z^2 < 2D\}\,.$$

Wir berechnen

$$\mathrm{vol}(S) = \pi \frac{2D}{\sqrt{|ab|}} 2 \frac{\sqrt{2D}}{\sqrt{|c|}} = \frac{4\sqrt{2}\pi D\sqrt{D}}{\sqrt{D}} = 4\sqrt{2}\pi D > 16D \geq 8\Delta(\Lambda)\,.$$

Also gibt es nach Satz 4.5 ein $(0,0,0) \neq (x,y,z) \in S \cap \Lambda$. Dann ist

$$|Q(x,y,z)| = \big|(|a|x^2 + |b|y^2) - |c|z^2\big| < 2D\,,$$

denn beide Terme der Differenz liegen im Intervall $[0, 2D[$. Außerdem ist $Q(x,y,z)$ eine ganze Zahl, die durch $2D$ teilbar ist. Beides zusammen erzwingt $Q(x,y,z) = 0$. Wir haben also eine nichttriviale ganzzahlige Lösung gefunden, und damit gibt es auch eine primitive ganzzahlige Lösung. ❑

Wir haben im Beweis die Bedingung mod 4 (für $abc$ ungerade) bzw. mod 8 (für $abc$ gerade) nicht benutzt. Der Beweis zeigt also, dass diese Bedingung aus den anderen folgt.

Es gibt eine Variante des Beweises, die diese Bedingung verwendet, um ein Gitter mit Kovolumen $4D$ zu konstruieren, für dessen Elemente $4D \mid Q(x,y,z)$ gilt. Für die Menge $S$ kann man dann das Ellipsoid $|a|x^2 + |b|y^2 + |c|z^2 < 4D$ nehmen (für unseren Beweis wäre das Ellipsoid mit $2D$ statt $4D$ zu klein). Das zeigt, dass man in diesem Fall die Vorzeichenbedingung nicht braucht: Sie folgt aus den anderen Bedingungen!

Hier zeigt sich ein allgemeineres Phänomen: Man kann die Bedingung an einer „Stelle“ (das heißt entweder die reelle Bedingung an die Vorzeichen oder die Bedingung, dass es Lösungen modulo Potenzen einer bestimmten Primzahl $p$ geben muss) weglassen, und der Satz ist immer noch richtig. Diese Aussage ist äquivalent zum Quadratischen Reziprozitätsgesetz; wir werden sie in Abschnitt 8 beweisen.

Es gibt auch einen Beweis mit der Abstiegsmethode, siehe z.B. [IR, § 17.3].

**FOLG** Schranke für Lösung

6.13. **Folgerung.** *Wenn $ax^2 + by^2 + cz^2 = 0$ (mit $abc$ quadratfrei) eine nichttriviale ganzzahlige Lösung hat, dann gibt es eine Lösung $(x,y,z)$ mit*

$$\max\{|a|\,x^2\,, |b|\,y^2\,, |c|\,z^2\} \leq \frac{16}{\pi^2}|abc| < 1{,}62114\,|abc|\,,$$

*oder äquivalent dazu,*

$$|x| \leq \frac{4}{\pi}\sqrt{|bc|}\,, \quad |y| \leq \frac{4}{\pi}\sqrt{|ca|}\,, \quad |z| \leq \frac{4}{\pi}\sqrt{|ab|}\,.$$

Es ist $4/\pi < 1{,}27324$.

*Beweis.* Mit $2|abc|$ anstelle von $\frac{16}{\pi^2}|abc|$ folgt das aus dem Beweis von Satz 6.12. Der Beweis funktioniert noch, wenn wir für $S$ den abgeschlossenen elliptischen Zylinder

$$S' = \{(x,y,z) \in \mathbb{R}^3 : |a|x^2 + |b|y^2 \leq 2D \text{ und } |c|z^2 \leq \tfrac{16}{\pi^2}D\}\,.$$

nehmen; es gilt nämlich

$$\mathrm{vol}(S') = \pi \frac{2D}{\sqrt{|ab|}} \cdot \frac{4}{\pi} \frac{2\sqrt{D}}{\sqrt{|c|}} = 16D \geq 8\Delta(\Lambda)\,.$$

Dabei verwenden wir die Variante des Satzes von Minkowski für kompakte Mengen $S$ (Satz 4.6). Das ergibt die angegebene Schranke für $z$; die Schranken für $x$ und $y$ folgen dann aus $|a|x^2 + |b|y^2 = |c|z^2$. (Die Beweisvariante mit dem Ellipsoid liefert eine schlechtere Schranke.) ❑

Tatsächlich gilt sogar die folgende stärkere Abschätzung. Sie wurde von L. Holzer gezeigt.[7]

**SATZ** Schranke von Holzer

6.14. **Satz.** *Wenn $ax^2 + by^2 + cz^2 = 0$ (mit abc quadratfrei) eine nichttriviale ganzzahlige Lösung hat, dann gibt es eine Lösung $(x, y, z)$ mit*

$$\max\{|a|\,x^2\,, |b|\,y^2\,, |c|\,z^2\} \le |abc|\,,$$

*oder äquivalent dazu,*

$$|x| \le \sqrt{|bc|}\,, \quad |y| \le \sqrt{|ca|}\,, \quad |z| \le \sqrt{|ab|}\,.$$

Sei zum Beispiel $a, b > 0$ und $c < 0$. Um den Satz zu beweisen, geht man von einer Lösung mit $|z| > \sqrt{ab}$ aus und zeigt dann, dass man eine andere finden kann (als Schnittpunkt einer geeigneten Geraden durch die gegebene Lösung mit dem Kegelschnitt, der der quadratischen Form entspricht), die kleineres $|z|$ hat. Das zeigt, dass die Lösung mit kleinstem $|z|$ die angegebene Schranke erfüllt; die Schranken für $|x|$ und $|y|$ folgen dann.

Der Satz von Holzer (oder auch schon Folgerung 6.13) lässt sich in einen Algorithmus zum Lösen von $ax^2+by^2+cz^2 = 0$ übersetzen: Man suche in dem angegebenen Bereich. Entweder man findet eine Lösung, oder die Gleichung hat keine. Allerdings ist die Größe des Suchraums *exponentiell* in der *Länge* der Eingabe (die ist $O(\log|abc|)$), daher ist dieses Verfahren für die Praxis im allgemeinen nicht brauchbar.

Wir werden jetzt Satz 6.12 auf beliebige ternäre quadratische Formen verallgemeinern.

**SATZ** Lösbarkeit von ternären qu. Formen

6.15. **Satz.** *Sei $Q(x, y, z)$ eine nicht-ausgeartete ternäre quadratische Form. Wir setzen $D = |\operatorname{disc}(Q)|$. Wenn es ein primitives Tripel $(x_0, y_0, z_0) \in \mathbb{Z}^3$ gibt mit $Q(x_0, y_0, z_0) \equiv 0 \bmod D^2$, dann hat $Q(x, y, z) = 0$ eine nichttriviale ganzzahlige Lösung.*

Man beachte, dass wir hier die reelle Lösbarkeit als Bedingung weglassen. Das entspricht der „Ellipsoid-Variante“ im Beweis von Satz 6.12.

Bevor wir mit dem Beweis von Satz 6.15 beginnen, formulieren wir ein Lemma.

**LEMMA** Existenz eines passenden Gitters

6.16. **Lemma.** *Seien $Q(x, y, z)$, $D$ und $(x_0, y_0, z_0)$ wie in Satz 6.15. Dann gibt es ein Gitter $\Lambda \subset \mathbb{Z}^3$ mit Kovolumen $\Delta(\Lambda) = D$ und sodass $Q(x, y, z) \equiv 0 \bmod D$ gilt für alle $(x, y, z) \in \Lambda$.*

*Beweis.* Sei $A \in \mathrm{GL}(3, \mathbb{Z})$. Dann können wir $Q$ durch ${}^AQ$ und $\boldsymbol{x}_0 = (x_0, y_0, z_0)$ durch $\boldsymbol{x}_0 A^{-1}$ ersetzen, wobei ${}^AQ(\boldsymbol{x}) = Q(\boldsymbol{x}A)$ ist (also ist die symmetrische Matrix von ${}^AQ$ gegeben durch $AM_QA^\top$). Wenn wir für ${}^AQ$ beweisen, dass es ein passendes Gitter ${}^A\Lambda$ gibt, dann ist $\Lambda = {}^A\Lambda \cdot A = \{\boldsymbol{x}A \mid \boldsymbol{x} \in {}^A\Lambda\}$ ein geeignetes Gitter für $Q$. Beachte hierfür, dass $\operatorname{disc}({}^AQ) = \operatorname{disc}(Q)\det(A)^2 = \operatorname{disc}(Q)$ ist.

Wir beweisen das Lemma durch Induktion über $D$. Im Fall $D = 1$ tut es $\Lambda = \mathbb{Z}^3$. Wir können also $D > 1$ annehmen. Dann hat $D$ einen Primteiler $p$. Wir behandeln zunächst den Fall, dass $p$ ungerade ist. Die Determinante von $2M_Q$ ist $\pm 2D$, also durch $p$ teilbar. Deswegen hat die Matrix, die aus $2M_Q$ entsteht, indem wir ihre Einträge modulo $p$ reduzieren, einen nichttrivialen Kern. Wir können also ein

[7] L. Holzer: *Minimal solutions of Diophantine equations,* Canadian J. Math. **2** (1950), 238–244

primitives Tripel $(x_1, y_1, z_1) \in \mathbb{Z}^3$ finden mit $(x_1, y_1, z_1)2M_Q \equiv (0,0,0) \bmod p$. Sei $A \in \mathrm{GL}(3,\mathbb{Z})$ mit erster Zeile $(x_1, y_1, z_1)$. Dann hat ${}^AQ$ die folgende Form:

$$ {}^AQ(x,y,z) = pa\,x^2 + b\,y^2 + c\,z^2 + pd\,xy + e\,yz + pf\,zx $$

mit $a, b, c, d, e, f \in \mathbb{Z}$. Sei $(x_0', y_0', z_0') = (x_0, y_0, z_0) \cdot A^{-1}$ die zugehörige primitive Lösung mod $D^2$ von ${}^AQ(x,y,z) = 0$.

1. Fall: $p$ teilt $y_0'$ und $z_0'$. Dann kann $p$ kein Teiler von $x_0'$ sein, und es ist

$$ 0 \equiv {}^AQ(x_0', y_0', z_0') \equiv pa\,(x_0')^2 \bmod p^2\,, $$

also muss $a$ durch $p$ teilbar sein. Die Form

$$ Q'(x,y,z) = \frac{1}{p^2}\,{}^AQ(x,py,pz) = {}^AQ\Big(\frac{x}{p}, y, z\Big) $$

ist ganzzahlig, und $D' = |\operatorname{disc}(Q')| = D/p^2 < D$. Das primitive Tripel

$$ (x_0'', y_0'', z_0'') = \Big(x_0', \frac{y_0'}{p}, \frac{z_0'}{p}\Big) $$

liefert eine Lösung von $Q'(x,y,z) \equiv 0 \bmod (D')^2$. Wir können also die Induktionsannahme auf $Q'$ anwenden und erhalten ein Gitter $\Lambda'$ für $Q'$. Dann ist

$$ {}^A\Lambda = \{(x, py, pz) \mid (x,y,z) \in \Lambda'\} $$

ein Gitter für ${}^AQ$ (beachte ${}^AQ(x,py,pz) = p^2Q'(x,y,z)$ und $\Delta({}^A\Lambda) = p^2\Delta(\Lambda')$), und $\Lambda = {}^A\Lambda \cdot A$ ist das gesuchte Gitter für $Q$.

2. Fall: $p$ teilt $y_0'$ oder $z_0'$ nicht. Dann gilt $p \nmid g = \mathrm{ggT}(y_0', z_0')$. Wir können eine Matrix

$$ B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & y_0'/g & z_0'/g \\ 0 & * & * \end{pmatrix} \in \mathrm{GL}(3,\mathbb{Z}) $$

finden; dann ist $(x_0', y_0', z_0') \cdot B^{-1} = (x_0', g, 0)$. Es ist ${}^B({}^AQ) = {}^{BA}Q$ und

$$ {}^{BA}Q(x,y,z) = pa\,x^2 + pb'\,y^2 + c'\,z^2 + pd'\,xy + e'\,yz + pf'\,zx\,, $$

denn ${}^{BA}Q(x_0', g, 0) \equiv 0 \bmod p$. Die Form

$$ Q'(x,y,z) = \frac{1}{p}\,{}^{BA}Q(x,y,pz) $$

ist ganzzahlig, und $D' = |\operatorname{disc}(Q')| = D/p < D$. Das primitive Tripel

$$ (x_0'', y_0'', z_0'') = (x_0', g, 0) $$

liefert eine Lösung von $Q'(x,y,z) \equiv 0 \bmod (D')^2$. Wir können also die Induktionsannahme auf $Q'$ anwenden und erhalten ein Gitter $\Lambda'$ für $Q'$. Dann ist

$$ {}^{BA}\Lambda = \{(x, y, pz) \mid (x,y,z) \in \Lambda'\} $$

ein Gitter für ${}^{BA}Q$ (beachte ${}^{BA}Q(x,y,pz) = pQ'(x,y,z)$ und $\Delta({}^{BA}\Lambda) = p\Delta(\Lambda')$), und $\Lambda = {}^{BA}\Lambda \cdot BA$ ist das gesuchte Gitter für $Q$.

Wenn $p = 2$ ist, dann gibt es für $Q$ mod 2 nach geeigneter Transformation mit einer Matrix $A \in \mathrm{GL}(3,\mathbb{Z})$ folgende Fälle (Übungsaufgabe):

$$ {}^AQ(x,y,z) \equiv 0, \quad x^2, \quad xy \quad \text{oder} \quad x^2 + xy + y^2 \quad \bmod 2\,. $$

Wir setzen wieder $(x_0', y_0', z_0') = (x_0, y_0, z_0) \cdot A^{-1}$.

Wenn ${}^AQ \equiv 0$ ist, dann können wir einfach ${}^AQ$ durch $Q' = {}^AQ/2$ ersetzen und die Induktionsannahme anwenden. Für das Gitter nehmen wir $\Lambda = 2\Lambda' \cdot A$. Beachte, dass $D' = |\operatorname{disc}(Q')| = D/8$, $\Delta(\Lambda) = 8\Delta(\Lambda')$ und ${}^AQ(2x, 2y, 2z) = 8Q'(x,y,z)$ gilt.

Wenn ${}^AQ \equiv x^2$ ist, dann muss $x_0'$ gerade sein. Die Form $Q'(x,y,z) = {}^AQ(2x,y,z)/2$ ist ganzzahlig, hat kleineres $D' = D/2$ und die primitive Lösung $(x_0'/2, y_0', z_0')$ mod $(D')^2$. Wir wenden die Induktionsannahme auf $Q'$ an und schließen wie im 2. Fall für $p$ ungerade.

Wenn ${}^AQ \equiv xy$ ist, dann muss $x_0'$ oder $y_0'$ gerade sein. Wenn z.B. $x_0'$ gerade ist, können wir wie eben $Q'(x,y,z) = {}^AQ(2x,y,z)/2$ setzen; im anderen Fall verwenden wir $Q'(x,y,z) = {}^AQ(x,2y,z)/2$; das ist ebenfalls analog zum 2. Fall oben.

Wenn schließlich ${}^AQ \equiv x^2 + xy + y^2$ ist, dann müssen $x_0'$ und $y_0'$ beide gerade sein; damit ist $z_0'$ ungerade. Wenn wir

$$ {}^AQ(x,y,z) = (2a+1)\,y^2 + (2b+1)\,y^2 + 2c\,z^2 + (2d+1)\,xy + 2e\,yz + 2f\,zx $$

schreiben, dann wird ${}^AQ(x_0', y_0', z_0') \equiv 2c \bmod 4$. Es folgt, dass $c$ gerade ist. Damit ist $D$ durch 4 teilbar, $Q'(x,y,z) = {}^AQ(x,y,z/2)$ ist ganzzahlig mit primitiver Lösung $(x_0'/2, y_0'/2, z_0')$ mod $(D')^2$, und $D' = D/4 < D$. Wir schließen analog zum 1. Fall für ungerades $p$. ❑

Aus dem Beweis folgt übrigens, dass es genügt, eine primitive Lösung mod $DN$ zu haben, wobei $N$ das Produkt der Primteiler von $D$ ist.

Der Beweis liefert auch die Aussage, dass unter den Voraussetzungen des Lemmas die Form $Q$ zu einer Form mit Diskriminante $\pm 1$ äquivalent ist.

Mithilfe des eben konstruierten Gitters können wir jetzt Satz 6.15 analog zum „Ellipsoid-Beweis“ von Satz 6.12 beweisen.

*Beweis von Satz 6.15.* Sei $\Lambda$ ein Gitter wie in Lemma 6.16. Nach Satz 6.8 gibt es eine Matrix $A \in \mathrm{GL}(3,\mathbb{Q})$, sodass ${}^AQ(x,y,z) = \alpha x^2 + \beta y^2 + \gamma z^2$ (mit $\alpha,\beta,\gamma \in \mathbb{Q}$) diagonal ist. Wir setzen $\Lambda' = \Lambda \cdot A^{-1}$; dann ist $\Lambda'$ ein Gitter im $\mathbb{R}^3$ mit Kovolumen $\Delta(\Lambda') = \Delta(\Lambda)/|\det(A)|$. Weiter definieren wir

$$ S = \{(x,y,z) \in \mathbb{R}^3 : |\alpha|x^2 + |\beta|y^2 + |\gamma|z^2 < D\}\,. $$

Es gilt $|\alpha\beta\gamma| = |\det({}^AQ)| = |\det(A)|^2 D/4$. Damit ergibt sich

$$ \mathrm{vol}(S) = \frac{4\pi}{3}\,\frac{D^{3/2}}{\sqrt{|\alpha\beta\gamma|}} = \frac{8\pi}{3}\frac{D}{|\det(A)|} > 2^3 \frac{\Delta(\Lambda)}{|\det(A)|} = 2^3\Delta(\Lambda')\,, $$

und wir können den Gitterpunktsatz 4.5 anwenden. Er liefert uns einen Punkt $(0,0,0) \neq (x',y',z') \in \Lambda' \cap S$. Wir setzen $(x,y,z) = (x',y',z') \cdot A$, dann folgt

$$ (x,y,z) \neq (0,0,0), \quad (x,y,z) \in \Lambda, $$
$$ |Q(x,y,z)| = |{}^AQ(x',y',z')| \leq |\alpha|(x')^2 + |\beta|(y')^2 + |\gamma|(z')^2 < D\,. $$

Aus $(x,y,z) \in \Lambda$ folgt $Q(x,y,z) \in D\mathbb{Z}$; wegen $|Q(x,y,z)| < D$ muss dann also $Q(x,y,z) = 0$ sein. ❑

Die Konstruktion des Gitters $\Lambda$ im Beweis von Lemma 6.16 führt zu einem Algorithmus zur Berechnung von $\Lambda$. Wir müssen dazu die Primfaktoren von $D$ kennen. Wir können dabei die Lösung mod $D^2$ jeweils in jedem Rekursionsschritt berechnen; dabei genügt es für $p$ ungerade, Quadratwurzeln mod $p$ berechnen zu können. Wenn weder der 1. noch der 2. Fall anwendbar sind, bedeutet das die Unlösbarkeit der Gleichung $Q = 0$.

Auch der Gitterpunktsatz lässt sich algorithmisch nutzen. Wir definieren die positiv definite quadratische Form

$$ Q^+(x,y,z) = |\alpha|(x')^2 + |\beta|(y')^2 + |\gamma|(z')^2 \quad \text{mit } (x',y',z') = (x,y,z)\cdot A^{-1}\,. $$

Es gibt effiziente Algorithmen, mit denen man $(0,0,0) \neq (x,y,z) \in \Lambda$ mit minimalem $Q^+(x,y,z)$ finden kann. Der Gitterpunktsatz sagt, dass für ein solches $(x,y,z)$ dann $|Q(x,y,z)| \leq Q^+(x,y,z) < D$ gilt; es folgt $Q(x,y,z) = 0$.

Eine detailliert ausgearbeitete Variante dieser Methode findet man in einer Arbeit von Denis Simon.[8]

**BSP**
Lösung von $Q(x,y,z) = 0$

6.17. **Beispiel.** Wir betrachten

$$Q(x,y,z) = 983487x^2 + 92527y^2 + 30903z^2 - 603321xy - 106946yz + 348670zx\,.$$

Für diese Form ist $D = 7$. Wir berechnen also die Matrix $2M_Q$ reduziert mod 7; wir erhalten

$$2M_Q \equiv \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \bmod 7\,.$$

Der Kern ist eindimensional und wird erzeugt von $(1,-1,0)$. Wir wählen

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

dann ist

$${}^AQ(x,y,z) \equiv x^2 + (-x+y)^2 - 2z^2 + 2x(-x+y) \equiv y^2 - 2z^2 \bmod 7\,.$$

Eine nichttriviale Lösung mod 7 ist gegeben durch $(y,z) = (1,2)$. Wir sind also im 2. Fall und wählen

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Die Form $Q'(x,y,z) = {}^{BA}Q(x,y,7z)/7$ hat $D' = 1$, also $\Lambda' = \mathbb{Z}^3$. Das liefert uns

$$\begin{aligned} \Lambda &= \mathbb{Z}\cdot(1,0,0)BA + \mathbb{Z}\cdot(0,1,0)BA + \mathbb{Z}\cdot(0,0,7)BA \\ &= \mathbb{Z}\cdot(1,-1,0) + \mathbb{Z}\cdot(0,1,2) + \mathbb{Z}\cdot(0,0,7)\,. \end{aligned}$$

Die Matrix

$$T = \begin{pmatrix} 1 & 0 & 0 \\ 201107 & 655658 & 0 \\ -1502 & 9762 & 25365 \end{pmatrix}$$

diagonalisiert $Q$; wir haben

$${}^TQ(x,y,z) = 983487x^2 + 19402559365y^2 - 25365z^2\,.$$

Wir setzen also wie in der Diskussion vor dem Beispiel

$$Q^+(x,y,z) = 983487(x')^2 + 19402559365(y')^2 + 25365(z')^2 = Q(x,y,z) + \frac{2}{25365}z^2$$

und finden ein nichttriviales Element von $\Lambda$ mit minimalem $Q^+$. Ein geeignetes Computeralgebrasystem liefert uns $(x,y,z) = (8,-45,-123)$; dies ist die gesuchte Lösung.

Wenn wir statt dessen auf die diagonalisierte Form ${}^TQ$ den Beweis des Satzes 6.12 von Legendre anwenden wollten, erhielten wir zunächst die Form

$$8455x^2 + 21y^2 - 327829z^2$$

[8] D. Simon: *Solving quadratic equations using reduced unimodular quadratic forms,* Math. Comp. **74** (2005), 1531–1543

mit paarweise teilerfremden Koeffizienten, die wir dann noch faktorisieren müssen. Hier ist das noch nicht problematisch: $8455 = 5 \cdot 19 \cdot 89$, $21 = 3 \cdot 7$, und 327829 ist prim. Man kann aber schon erkennen, dass man bei noch etwas größeren Koeffizienten der ursprünglichen Form schnell Zahlen bekommt, die man nicht mehr ohne weiteres faktorisieren kann, und dann steckt man fest. Beachte, dass die neu auftretenden Primzahlen 3, 5, 19, 89 und 327829 mit der ursprünglichen Diskriminante 7 nichts zu tun haben! ♣

Beispiele wie das eben behandelte, wo die Form große Koeffizienten, aber kleine Diskriminante hat, treten übrigens in manchen Anwendungen recht häufig auf. In solchen Fällen ist es sehr viel effizienter, direkt mit der gegebenen Form zu arbeiten, als sie zuerst zu diagonalisieren.

## 7. $p$-adische Zahlen

Wir haben gesehen, dass eine diophantische Gleichung nur dann (primitive) ganzzahlige Lösungen haben kann, wenn sie (primitive) Lösungen modulo $m$ hat für jedes $m \geq 1$. Nach dem Chinesischen Restsatz ist dies äquivalent dazu, dass es Lösungen modulo jeder Primzahlpotenz $p^n$ gibt. Wir haben auch die Lösbarkeit in reellen Zahlen als weitere notwendige Bedingung betrachtet. Die reellen Zahlen haben den Vorteil, dass sie einen Körper bilden. Demgegenüber haben die Ringe $\mathbb{Z}/p^n\mathbb{Z}$ zwar die schöne Eigenschaft, endlich zu sein, sie sind jedoch für $n \geq 2$ nicht einmal mehr Integritätsbereiche und deshalb zum Rechnen nicht so praktisch. Es wäre also wünschenswert, eine Struktur zur Verfügung zu haben, die ein Körper oder ein Integritätsbereich ist und außerdem Aussagen modulo $p^n$ für alle $n$ zu formulieren erlaubt. Dies kann erreicht werden, indem man in geeigneter Weise zu einer Art algebraischem Grenzwert für $n \to \infty$ übergeht. Man erhält dann den Ring $\mathbb{Z}_p$ der ganzen $p$-adischen Zahlen, der ein Integritätsbereich ist, und den Körper $\mathbb{Q}_p$ der $p$-adischen Zahlen als seinen Quotientenkörper. Die Existenz (primitiver) ganzzahliger Lösungen mod $p^n$ für alle $n$ wird dann äquivalent zur Existenz einer (primitiven) Lösung in $\mathbb{Z}_p$.

Als Beispiel betrachten wir Lösungen der Gleichung $x^2 + 7 = 0$ modulo Potenzen von 2. In der Tabelle in Abbildung 4 sind die Lösungen für $n \leq 6$ aufgelistet.

$$
\begin{aligned}
&\bmod 2^1: && x \equiv 1\\
&\bmod 2^2: && x \equiv 1, 3\\
&\bmod 2^3: && x \equiv 1, 3, 5, 7\\
&\bmod 2^4: && x \equiv 3, 5, 11, 13\\
&\bmod 2^5: && x \equiv 5, 11, 21, 27\\
&\bmod 2^6: && x \equiv 11, 21, 43, 53
\end{aligned}
$$

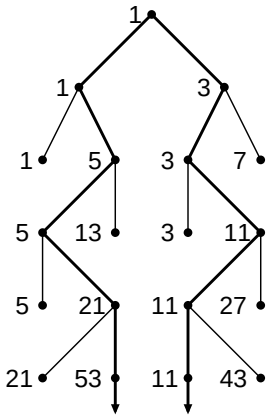



Abbildung 4. Lösungen von $x^2 + 7 \equiv 0 \bmod 2^n$.

Es ist nicht schwer, sich davon zu überzeugen, dass es für $n \geq 3$ stets vier Lösungen mod $2^n$ gibt (Übung). Das wäre nicht möglich, wenn $\mathbb{Z}/2^n\mathbb{Z}$ ein Körper wäre, denn in einem Körper (oder Integritätsbereich) kann eine quadratische Gleichung höchstens zwei Lösungen haben. Auf der anderen Seite sind jeweils zwei der vier Lösungen in einem gewissen Sinn keine „richtigen“ Lösungen, denn sie lassen sich nicht zu Lösungen mod $2^{n+1}$ „hochheben“. Wenn wir jetzt „zum Grenzwert übergehen“ und nur Lösungen betrachten, die sich beliebig weit hochheben lassen, dann bleiben zwei Lösungen übrig, wie wir das erwarten würden (Übung).

**DEF** $p$-adische Zahlen

7.1. **Definition.** Sei $p$ eine Primzahl. Der Ring $\mathbb{Z}_p$ der *ganzen $p$-adischen Zahlen* ist

$$\mathbb{Z}_p = \{(a_n + p^n\mathbb{Z})_{n\geq 1} \mid a_n \equiv a_{n+1} \bmod p^n \text{ für alle } n \geq 1\} \subset \prod_{n=1}^{\infty} \mathbb{Z}/p^n\mathbb{Z}\,.$$

Dabei sind Addition und Multiplikation komponentenweise definiert. Es ist klar, dass $\mathbb{Z}_p$ (als Unterring des Produktrings oben rechts) tatsächlich ein (kommutativer) Ring (mit 1) ist. ♢

Es gibt eine kanonische Einbettung $\mathbb{Z} \hookrightarrow \mathbb{Z}_p$, die gegeben ist durch

$$a \longmapsto (\bar{a}, \bar{a}, \bar{a}, \dots) = (a + p\mathbb{Z}, a + p^2\mathbb{Z}, a + p^3\mathbb{Z}, \dots) .$$

Dies ist ein Ringhomomorphismus.

Die Elemente von $\mathbb{Z}_p$ nach dieser Definition sind also gerade die Folgen

$$(a_1 + p\mathbb{Z}, a_2 + p^2\mathbb{Z}, a_3 + p^3\mathbb{Z}, \dots, a_n + p^n\mathbb{Z}, \dots) ,$$

die „kompatibel“ sind in dem Sinn, dass $a_{n+1} \equiv a_n \bmod p^n$ für jedes $n$ gilt. Im Beispiel oben erhalten wir etwa

$$(1 + 2\mathbb{Z}, 1 + 2^2\mathbb{Z}, 5 + 2^3\mathbb{Z}, 5 + 2^4\mathbb{Z}, 21 + 2^5\mathbb{Z}, 53 + 2^6\mathbb{Z}, \dots) \in \mathbb{Z}_2$$

als eine 2-adische Zahl, deren Quadrat $-7$ ist.

Wir brauchen noch mehr Informationen über die Struktur von $\mathbb{Z}_p$.

**SATZ** Eigenschaften von $\mathbb{Z}_p$

7.2. **Satz.** *$\mathbb{Z}_p$ ist ein Integritätsbereich. Das einzige maximale Ideal ist $p\mathbb{Z}_p$, und alle von null verschiedenen Ideale haben die Form $p^n\mathbb{Z}_p$ für ein $n \geq 0$. ($\mathbb{Z}_p$ ist also insbesondere ein Hauptidealring und damit faktoriell.) Die Einheitengruppe ist $\mathbb{Z}_p^\times = \mathbb{Z}_p \setminus p\mathbb{Z}_p$.*

*Beweis.*

(a) $p\mathbb{Z}_p$ ist ein maximales Ideal: Wir zeigen, dass $\mathbb{Z}_p/p\mathbb{Z}_p \cong \mathbb{Z}/p\mathbb{Z}$ ist; die Behauptung folgt dann, weil $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ ein Körper ist. Die Abbildung

$$\phi\colon \mathbb{Z}_p \ni (a_1 + p\mathbb{Z}, a_2 + p^2\mathbb{Z}, \dots) \longmapsto a_1 + p\mathbb{Z} \in \mathbb{Z}/p\mathbb{Z}$$

ist ein Ringhomomorphismus und surjektiv, denn die zusammengesetzte Abbildung

$$\mathbb{Z} \hookrightarrow \mathbb{Z}_p \to \mathbb{Z}/p\mathbb{Z}$$

ist surjektiv. Es ist klar, dass $p\mathbb{Z}_p \subset \ker\phi$ ist. Sei jetzt $(a_1+p\mathbb{Z}, a_2+p^2\mathbb{Z}, \dots) \in \ker\phi$. Dann ist $a_1$ durch $p$ teilbar. Wegen der Kompatibilität der Folge sind dann alle $a_n$ durch $p$ teilbar: $a_n = pb_n$. Aus der Kompatibilität der $a_n$ folgt $b_{n+1} \equiv b_n \bmod p^{n-1}$. Außerdem ist $pb_{n+1} = a_{n+1} \equiv a_n \bmod p^n$. Daher gilt

$$(a_1 + p\mathbb{Z}, a_2 + p^2\mathbb{Z}, a_3 + p^3\mathbb{Z}, \dots) = p \cdot (b_2 + p\mathbb{Z}, b_3 + p^2\mathbb{Z}, b_4 + p^3\mathbb{Z}, \dots) \in p\mathbb{Z}_p .$$

Also gilt auch $\ker\phi \subset p\mathbb{Z}_p$. Nach dem Homomorphiesatz für Ringe folgt

$$\mathbb{Z}_p/p\mathbb{Z}_p = \mathbb{Z}_p/\ker\phi \cong \operatorname{im}\phi = \mathbb{Z}/p\mathbb{Z} .$$

(b) $\mathbb{Z}_p^\times = \mathbb{Z}_p \setminus p\mathbb{Z}_p$: Die Inklusion „$\subset$“ ist klar (ein Element eines echten Ideals kann keine Einheit sein).
Für „$\supset$“ sei $u \in \mathbb{Z}_p \setminus p\mathbb{Z}_p$. Wir schreiben $u = (u_1 + p\mathbb{Z}, u_2 + p^2\mathbb{Z}, \dots)$; dann ist $u_n \perp p$, also gibt es $v_n$ mit $u_n v_n \equiv 1 \bmod p^n$, und $v_n$ ist mod $p^n$ eindeutig bestimmt.

Da $u_{n+1} \equiv u_n \bmod p^n$ ist, muss dann auch $v_{n+1} \equiv v_n \bmod p^n$ sein. Damit ist dann $v = (v_1 + p\mathbb{Z}, v_2 + p^2\mathbb{Z}, \dots) \in \mathbb{Z}_p$ und $u \cdot v = 1$.

(c) $p\mathbb{Z}_p$ ist das *einzige* maximale Ideal: Wäre nämlich $\mathfrak{m}$ ein weiteres maximales Ideal, dann wäre $\mathfrak{m} \setminus p\mathbb{Z}_p \neq \emptyset$. Nach (b) würde das bedeuten, dass $\mathfrak{m}$ eine Einheit enthält, also wäre $\mathfrak{m} = \mathbb{Z}_p$ und damit kein maximales Ideal.

(d) Es gilt $\bigcap_{n\geq 0} p^n\mathbb{Z}_p = \{0\}$, denn $a = (\bar{a}_1, \bar{a}_2, \dots) \in p^n\mathbb{Z}_p$ bedeutet $\bar{a}_j = 0$ für $j \leq n$.

(e) Für $a \in \mathbb{Z}_p \setminus \{0\}$ gibt es $n \geq 0$ und $u \in \mathbb{Z}_p^\times$, sodass $a = up^n$ ist: Aus (d) folgt, dass es ein $n \geq 0$ gibt, sodass $a \in p^n\mathbb{Z}_p \setminus p^{n+1}\mathbb{Z}_p$ ist. Dann ist $a = up^n$ mit $u \in \mathbb{Z}_p \setminus p\mathbb{Z}_p = \mathbb{Z}_p^\times$.

Beachte: Dies ist die Primfaktorzerlegung in $\mathbb{Z}_p$ (mit der einzigen Primzahl $p$).

(f) Sei jetzt $I \subset \mathbb{Z}_p$ ein von null verschiedenes Ideal. Aus (d) folgt wieder, dass es ein $n \geq 0$ gibt mit $I \subset p^n\mathbb{Z}_p$, aber $I \not\subset p^{n+1}\mathbb{Z}_p$. Es gibt also ein $a \in I$ der Form $a = up^n$ mit $u \in \mathbb{Z}_p^\times$. Weil $u$ invertierbar ist, ist auch $p^n = u^{-1}a \in I$. Es folgt $p^n\mathbb{Z}_p \subset I$, also $I = p^n\mathbb{Z}_p$.

(g) $\mathbb{Z}_p$ ist ein Integritätsbereich: Seien $a, b \in \mathbb{Z}_p$ mit $ab = 0$. Wir schreiben

$$a = (a_1 + p\mathbb{Z}, a_2 + p^2\mathbb{Z}, \dots) \quad \text{und} \quad b = (b_1 + p\mathbb{Z}, b_2 + p^2\mathbb{Z}, \dots)\,.$$

Wir können $a \neq 0$ annehmen. Dann ist $a = up^N$ mit geeigneten $N \geq 0$, $u \in \mathbb{Z}_p^\times$. Es folgt $p^N b = 0$. Für die Komponenten von $b$ bedeutet das $p^N b_{n+N} \equiv 0 \bmod p^{N+n}$ und damit $b_n \equiv b_{N+n} \equiv 0 \bmod p^n$ für alle $n \geq 1$, also $b = 0$. ❑

Ein Hauptidealring mit genau einem maximalen Ideal (das nicht das Nullideal ist) wird ein *diskreter Bewertungsring* genannt. Solche Ringe haben analoge Eigenschaften wie die Ringe $\mathbb{Z}_p$.

Die Aussage in Teil (e) des Beweises von Satz 7.2 motiviert folgende Definition.

7.3. **Definition.** Für $a = (a_1, a_2, \dots) \in \mathbb{Z}_p$ definieren wir die *$p$-adische Bewertung* als

$$v_p(a) = \max(\{0\} \cup \{n \geq 1 \mid a_n = 0\}) = \max\{n \geq 0 \mid a \in p^n\mathbb{Z}_p\}\,,$$

wenn $a \neq 0$, und $v_p(0) = \infty$. Für $0 \neq a \in \mathbb{Z}_p$ gilt dann $a = up^{v_p(a)}$ mit $u \in \mathbb{Z}_p^\times$. Diese Bewertung setzt die $p$-adische Bewertung $v_p$ auf $\mathbb{Z}$ fort; die Schreibweise $v_p$ ist also gerechtfertigt.

**DEF**
$p$-adische Bewertung

$p$-adischer Absolutbetrag

Wir definieren außerdem den *$p$-adischen Absolutbetrag* durch

$$|0|_p = 0\,, \quad |a|_p = p^{-v_p(a)} \quad \text{für } a \neq 0. \qquad \diamondsuit$$

Wie jeden Integritätsbereich können wir auch $\mathbb{Z}_p$ in einen (minimalen) Körper einbetten.

7.4. **Definition.** Der *Körper $\mathbb{Q}_p$ der $p$-adischen Zahlen* ist der Quotientenkörper von $\mathbb{Z}_p$. $\diamondsuit$

**DEF**
Körper $\mathbb{Q}_p$

Wie sehen die Elemente von $\mathbb{Q}_p$ aus? Seien $a, b \in \mathbb{Z}_p \setminus \{0\}$; dann können wir schreiben $a = up^m$, $b = vp^n$ mit Einheiten $u$ und $v$ und $m = v_p(a)$, $n = v_p(b)$. Dann ist $a/b = (uv^{-1})p^{m-n}$. Es genügt also, $p$ zu invertieren, um von $\mathbb{Z}_p$ zu $\mathbb{Q}_p$ zu kommen: $\mathbb{Q}_p = \mathbb{Z}_p[1/p]$.

Wir sehen auch, dass man die $p$-adische Bewertung und den $p$-adischen Absolutbetrag auf $\mathbb{Q}_p$ fortsetzen kann (so wie wir das früher schon mit $\mathbb{Z}$ und $\mathbb{Q}$ gemacht haben):

$$v_p(a/b) = v_p(a) - v_p(b) \quad \text{und} \quad |a/b|_p = |a|_p/|b|_p\,.$$

$v_p(a/b)$ kann jetzt natürlich eine beliebige ganze Zahl sein (oder $\infty$). Es gilt wieder für alle $a \in \mathbb{Q}_p^\times$, dass

$$a = up^{v_p(a)}$$

ist mit $u \in \mathbb{Z}_p^\times$.

Der $p$-adische Absolutbetrag hat Eigenschaften, die wir vom gewöhnlichen Absolutbetrag (auf $\mathbb{R}$ oder $\mathbb{C}$) kennen.

**LEMMA**
Eigenschaften des $p$-adischen Absolutbetrags

7.5. **Lemma.** *Für alle $a, b \in \mathbb{Q}_p$ gilt*

(1) *(Multiplikativität)* $|ab|_p = |a|_p\,|b|_p$ ;
*für die Bewertung gilt* $v_p(ab) = v_p(a) + v_p(b)$.

(2) *(Dreiecksungleichung)* $|a+b|_p \le \max\{|a|_p, |b|_p\} \le |a|_p + |b|_p$ ;
*für die Bewertung gilt* $v_p(a+b) \ge \min\{v_p(a), v_p(b)\}$.

*Beweis.* Die Aussagen sind klar, wenn $a = 0$ oder $b = 0$ ist. Seien also $a$ und $b$ von null verschieden. Dann ist $a = up^{v_p(a)}$ und $b = vp^{v_p(b)}$ mit $u, v \in \mathbb{Z}_p^\times$ und damit $ab = uvp^{v_p(a)+v_p(b)}$. Wegen $uv \in \mathbb{Z}_p^\times$ folgt $v_p(ab) = v_p(a) + v_p(b)$, also auch $|ab|_p = |a|_p\,|b|_p$.

Sei jetzt ohne Einschränkung $v_p(a) \le v_p(b)$. Dann haben wir

$$a + b = (u + vp^{v_p(b)-v_p(a)})p^{v_p(a)}\,.$$

Da der erste Faktor wegen $v_p(b) - v_p(a) \ge 0$ in $\mathbb{Z}_p$ ist, folgt $v_p(a+b) \ge v_p(a)$ und damit wiederum die Behauptung. ❑

Die Dreiecksungleichung gilt hier in einer verschärften Form, die auch *ultrametrische Dreiecksungleichung* heißt.

Aus dem Beweis ergibt sich auch, dass im Fall $|a|_p \ne |b|_p$ sogar $|a+b|_p = \max\{|a|_p, |b|_p\}$ gilt (denn dann ist $u + vp^{v_p(b)-v_p(a)} \in \mathbb{Z}_p^\times$). Das folgt auch direkt aus der ultrametrischen Dreiecksungleichung.

Wir können also (wie mit dem gewöhnlichen Absolutbetrag) eine *Metrik* auf $\mathbb{Z}_p$ und auf $\mathbb{Q}_p$ definieren durch

$$d(a, b) = |a - b|_p\,.$$

Bezüglich dieser Metrik ist dann zum Beispiel $\mathbb{Z}_p$ die abgeschlossene Einheitskugel in $\mathbb{Q}_p$ (denn es gilt $\quad a \in \mathbb{Z}_p \iff v_p(a) \ge 0 \iff |a|_p \le 1$).

**SATZ**
$\mathbb{Z}_p$ ist kompakt

$\mathbb{Q}_p$ ist vollständig

7.6. **Satz.** *Der metrische Raum $(\mathbb{Z}_p, d)$ ist kompakt. $\mathbb{Q}_p$ ist ein lokal-kompakter Körper und damit vollständig.*

*Beweis.* Wir müssen zeigen, dass jede Folge $(a_n)$ in $\mathbb{Z}_p$ einen Häufungspunkt hat. Dazu schreiben wir wie in Def. 7.1 $a_n = (a_n^{(1)}, a_n^{(2)}, \dots) \in \prod_{m\ge1} \mathbb{Z}/p^m\mathbb{Z}$. Da es für die erste Komponente $a_n^{(1)}$ nur endlich viele Möglichkeiten gibt, muss einer der möglichen Werte unendlich oft vorkommen. Sei $a^{(1)}$ ein solcher Wert. Es gibt dann unendlich viele $n \ge 1$ mit $a_n^{(1)} = a^{(1)}$. Für $a_n^{(2)}$ gibt es ebenfalls nur endlich viele Möglichkeiten, also gibt es einen Wert $a^{(2)}$, sodass für unendlich viele $n$ gilt $a_n^{(1)} = a^{(1)}$ und $a_n^{(2)} = a^{(2)}$. Offenbar können wir dieses Verfahren fortsetzen und bekommen eine Folge $(a^{(1)}, a^{(2)}, \dots) \in \prod_{m\ge1} \mathbb{Z}/p^m\mathbb{Z}$, sodass es für jedes $k \ge 1$ jeweils unendlich viele $n$ gibt mit $a_n^{(m)} = a^{(m)}$ für alle $1 \le m \le k$. Es folgt, dass die Folge der $a^{(m)}$ kompatibel ist, d.h., es ist $a = (a^{(1)}, a^{(2)}, \dots) \in \mathbb{Z}_p$. Wir definieren nun rekursiv $n_0 = 1$, und

$$n_k = \min\{n > n_{k-1} \mid a_n^{(m)} = a^{(m)} \text{ für alle } 1 \le m \le k\}$$

für $k \geq 1$. Nach Konstruktion ist die Folge $(n_k)$ wohldefiniert, und es gilt

$$|a_{n_k} - a|_p \leq p^{-k}\,.$$

Also konvergiert die Teilfolge $(a_{n_k})_{k\geq 1}$ gegen $a \in \mathbb{Z}_p$.

Für die zweite Aussage bemerken wir, dass $\mathbb{Z}_p = \{a \in \mathbb{Q}_p \mid |a|_p < p\}$ auch offen in $\mathbb{Q}_p$ ist. Für jedes $a \in \mathbb{Q}_p$ ist dann $a+\mathbb{Z}_p$ eine kompakte Umgebung von $a$ in $\mathbb{Q}_p$. Ist nun $(a_n)$ eine Cauchy-Folge in $\mathbb{Q}_p$, dann liegen (für $n$ groß genug, aber tatsächlich für alle $n$) die Glieder in einer kompakten Teilmenge, also gibt es eine konvergente Teilfolge. Eine Cauchy-Folge mit einer konvergenten Teilfolge muss aber schon selbst konvergieren. Ein lokal-kompakter Körper ist also vollständig. ❑

Ein anderer (und weniger konstruktiver) Beweis der Kompaktheit von $\mathbb{Z}_p$ kann wie folgt geführt werden:

(1) Die Topologie auf $\mathbb{Z}_p$ stimmt mit der von $\prod_{n\geq 1} \mathbb{Z}/p^n\mathbb{Z}$ induzierten Teilraumtopologie überein, wobei das Produkt die Produkttopologie bezüglich der diskreten Topologie auf jedem Faktor trägt.

(2) Endliche diskrete Räume sind kompakt, also ist $\prod_{n\geq 1} \mathbb{Z}/p^n\mathbb{Z}$ nach dem Satz von Tychonoff kompakt.

(3) $\mathbb{Z}_p$ ist in $\prod_{n\geq 1} \mathbb{Z}/p^n\mathbb{Z}$ abgeschlossen, also als abgeschlossene Teilmenge eines kompakten Raums ebenfalls kompakt.

7.7. **Satz.** *$\mathbb{Z}$ liegt dicht in $\mathbb{Z}_p$, und $\mathbb{Q}$ liegt dicht in $\mathbb{Q}_p$. Insbesondere ist $\mathbb{Q}_p$ die Vervollständigung von $\mathbb{Q}$ bezüglich der durch $|\cdot|_p$ gegebenen Metrik.*

**SATZ**
$\mathbb{Z}$ dicht in $\mathbb{Z}_p$

*Beweis.* Sei $a = (a^{(1)}, a^{(2)}, \dots) \in \mathbb{Z}_p$. Dann können wir für jedes $n \geq 1$ die Restklasse $a^{(n)} \in \mathbb{Z}/p^n\mathbb{Z}$ durch eine ganze Zahl $a_n$ repräsentieren. Es gilt dann $|a - a_n|_p \leq p^{-n}$. Es gibt also eine Folge ganzer Zahlen, die gegen $a$ konvergiert.

Sei nun $a \in \mathbb{Q}_p$. Dann gibt es $m \geq 0$, sodass $p^m a \in \mathbb{Z}_p$ ist. Sei $(b_n)$ eine Folge ganzer Zahlen, die in $\mathbb{Z}_p$ gegen $p^m a$ konvergiert. Wegen

$$|a - p^{-m}b_n|_p = |p^{-m}(p^m a - b_n)|_p = p^m |p^m a - b_n|_p$$

folgt, dass $(b_n/p^m)$ gegen $a$ konvergiert.

Damit erfüllt $\mathbb{Q}_p$ die Bedingungen, um die Vervollständigung von $\mathbb{Q}$ bezüglich $|\cdot|_p$ zu sein: $\mathbb{Q}_p$ ist vollständig bezüglich $|\cdot|_p$, und $\mathbb{Q} \subset \mathbb{Q}_p$ ist dicht. ❑

Wir sehen also, dass die $p$-adischen Körper $\mathbb{Q}_p$ eine sehr ähnliche Rolle spielen wie die reellen Zahlen $\mathbb{R}$, die ja die Vervollständigung von $\mathbb{Q}$ bezüglich des gewöhnlichen Absolutbetrags $|\cdot|$ sind.

7.8. **Definition.** Sei $K$ ein Körper. Ein *Absolutbetrag* $|\cdot|$ auf $K$ ist eine Funktion $K \ni x \mapsto |x| \in \mathbb{R}_{\geq 0}$ mit den folgenden Eigenschaften:

**DEF**
Absolutbetrag

(1) $\forall x \in K\colon |x| = 0 \iff x = 0$,

(2) (Multiplikativität) $\quad \forall x, y \in K\colon |xy| = |x|\,|y|$,

(3) (Dreiecksungleichung) $\quad \forall x, y \in K\colon |x + y| \leq |x| + |y|$. ♢

Man kann zeigen, dass bis auf eine natürliche Äquivalenz die $p$-adischen Absolutbeträge $|\cdot|_p$ und der gewöhnliche Absolutbetrag $|\cdot|_\infty := |\cdot|$ die einzigen nichttrivialen Absolutbeträge auf $\mathbb{Q}$ sind. (Der triviale Absolutbetrag ist $|x| = 1$ für alle $x \neq 0$.) Das kommt auch in der folgenden Relation zum Ausdruck.

**SATZ**
Produktformel

7.9. **Satz.** *Für alle $a \in \mathbb{Q}^\times$ gilt*

$$\prod_{v=p,\infty} |a|_v = 1\,.$$

*Das Produkt läuft dabei über alle Primzahlen $p$ und $\infty$.*

*Beweis.* Zuerst müssen wir uns davon überzeugen, dass das unendliche Produkt auf der linken Seite wohldefiniert ist. Das liegt daran, dass nur endlich viele Primzahlen im Zähler und im Nenner von $a$ vorkommen; für alle anderen $p$ ist $|a|_p = 1$. Also sind nur endlich viele Faktoren von 1 verschieden.

Alle Absolutbeträge sind multiplikativ, und jedes $a \in \mathbb{Q}^\times$ ist ein Produkt von Potenzen (mit möglicherweise negativen Exponenten) von $-1$ und Primzahlen. Es genügt also, die Fälle $a = -1$ und $a = p$ zu betrachten. Für $a = -1$ ist $|a|_v = 1$ für alle $v$ (das gilt für jeden Absolutbetrag auf jedem Körper). Für $a = p$ ist $|a|_\infty = p$ und $|a|_p = p^{-1}$; alle anderen Faktoren sind 1. ❑

Es ist ein allgemeines Prinzip in der Zahlentheorie, das oft sehr nützlich ist, alle Vervollständigungen von $\mathbb{Q}$ gleichberechtigt zu betrachten.

**Bemerkung.** Die Definition der $p$-adischen Zahlen in Def. 7.1 entspricht der Konstruktion der reellen Zahlen durch Intervallschachtelungen: Die $n$-te Komponente in der Folge fixiert die Restklasse mod $p^n$ und schränkt die $p$-adische Zahl damit auf ein kompaktes „Intervall“ vom Durchmesser $p^{-n}$ ein.

Die starke ultrametrische Dreiecksungleichung

$$|a + b|_p \leq \max\{|a|_p, |b|_p\}$$

hat ungewohnte Konsequenzen. Wir haben bereits gesehen, dass $\mathbb{Z}$ in der $p$-adischen Metrik *beschränkt* ist: $|a|_p \leq 1$ für $a \in \mathbb{Z}$ (oder sogar $a \in \mathbb{Z}_p$). Insbesondere gilt das vom Arbeiten mit $\mathbb{R}$ her gewohnte *Archimedische Prinzip* nicht, demzufolge es zu jedem Element $\alpha$ des betrachteten Körpers eine ganze Zahl $a$ geben sollte mit $|a| > |\alpha|$. Man spricht deshalb auch von *nicht-archimedischen* Körpern.

Eine Interpretation der starken Dreiecksungleichung ist, dass sich „Rundungsfehler“ nicht verstärken können, wenn man $p$-adische Zahlen addiert. Das bedeutet, dass sich numerische Rechnungen sehr viel besser kontrollieren lassen, als wenn man reelle Näherungen verwendet. Das ist in vielen Fällen nützlich.

Das folgende Lemma illustriert die Wirkung der starken Dreiecksungleichung.

**LEMMA**
Konvergenz in $\mathbb{Q}_p$

7.10. **Lemma.**

(1) *Eine Reihe $\sum_{n=0}^\infty a_n$ mit $a_n \in \mathbb{Q}_p$ konvergiert genau dann in $\mathbb{Q}_p$, wenn $(a_n)$ eine Nullfolge in $\mathbb{Q}_p$ ist.*

(2) *Jede Reihe $\sum_{n=0}^\infty c_n p^n$ mit $c_n \in \mathbb{Z}_p$ konvergiert in $\mathbb{Z}_p$.*

(3) *Jedes $a \in \mathbb{Z}_p$ hat eine eindeutige Darstellung der Form*

$$a = \sum_{n=0}^\infty c_n p^n$$

*mit $c_n \in \{0, 1, \ldots, p-1\}$ für alle $n \geq 0$.*

*Beweis.*

(1) Eine unendliche Reihe kann nur dann konvergieren, wenn ihre Glieder eine Nullfolge bilden (Beweis wie für $\mathbb{R}$). Sei also umgekehrt $(a_n)$ eine Nullfolge in $\mathbb{Q}_p$, und sei $m \in \mathbb{Z}$. Dann ist $|a_n|_p \leq p^{-m}$ für $n \geq N_m$. Aus der starken Dreiecksungleichung folgt, dass

$$\Big|\sum_{n=\nu}^{\nu+k} a_n\Big|_p \leq \max\{|a_n|_p \mid \nu \leq n \leq \nu+k\} \leq p^{-m}$$

gilt für $\nu \geq N_m$ und $k \geq 0$. Das zeigt, dass die Folge der Partialsummen eine Cauchy-Folge ist; nach Satz 7.6 konvergiert die Reihe also.

(2) Es ist $|c_n p^n|_p = |c_n|_p\, p^{-n} \leq p^{-n}$, also bilden die Glieder der Reihe eine Nullfolge. Nach Teil (1) konvergiert die Reihe dann.

(3) Übungsaufgabe. ❑

Die Aussage von Teil (3) kann als $p$-adisches Analogon der Dezimalbruchentwicklung in $\mathbb{R}$ interpretiert werden.

**BSP** geometrische Reihe

7.11. **Beispiel.** In jedem Körper $K$ mit Absolutbetrag $|\cdot|$ gilt, dass für $|x| < 1$ die geometrische Reihe $\sum_{n=0}^{\infty} x^n$ konvergiert, und zwar gegen $1/(1-x)$:

$$\Big|\frac{1}{1-x} - \sum_{n=0}^{N-1} x^n\Big| = \Big|\frac{1}{1-x} - \frac{1-x^N}{1-x}\Big| = \Big|\frac{x^N}{1-x}\Big| = \frac{1}{|1-x|}\,|x|^N \longrightarrow 0 \quad \text{für } N \to \infty$$

In $\mathbb{Q}_p$ gilt zum Beispiel mit $x = p$ also

$$1 + p + p^2 + p^3 + \cdots = \frac{1}{1-p}\,.$$

Daraus folgt die folgende Darstellung von $-1$ wie in Teil (3) von Lemma 7.10:

$$-1 = (p-1) + (p-1)\cdot p + (p-1)\cdot p^2 + \ldots\,. \qquad ♣$$

Für uns war die ursprüngliche Motivation, die $p$-adischen Zahlen einzuführen, dass wir zu einem besseren Verständnis von Kongruenzen mod $p^n$ für alle $n$ (bei fester Primzahl $p$) gelangen wollten. Dieser Zusammenhang wird im folgenden Satz formuliert.

**SATZ** Lösbarkeit in $\mathbb{Z}_p$ vs. Lösbarkeit mod $p^n$

7.12. **Satz.** *Sei $F \in \mathbb{Z}[X_1, \ldots, X_k]$ ein Polynom mit ganzzahligen Koeffizienten, und sei $p$ eine Primzahl.*

(1)
$$\begin{aligned} &\forall n \geq 1\, \exists (x_1, \ldots, x_k) \in \mathbb{Z}^k \colon p^n \mid F(x_1, \ldots, x_k) \\ &\qquad \Longleftrightarrow\ \exists (x_1, \ldots, x_k) \in \mathbb{Z}_p^k \colon F(x_1, \ldots, x_k) = 0\,. \end{aligned}$$

(2) *Wenn $F$ homogen ist, gilt*

$$\begin{aligned} &\forall n \geq 1\, \exists (x_1, \ldots, x_k) \in \mathbb{Z}^k \setminus (p\mathbb{Z})^k \colon p^n \mid F(x_1, \ldots, x_k) \\ &\qquad \Longleftrightarrow\ \exists (x_1, \ldots, x_k) \in \mathbb{Z}_p^k \setminus (p\mathbb{Z}_p)^k \colon F(x_1, \ldots, x_k) = 0 \\ &\qquad \Longleftrightarrow\ \exists (x_1, \ldots, x_k) \in \mathbb{Q}_p^k \setminus \{0\} \colon F(x_1, \ldots, x_k) = 0\,. \end{aligned}$$

*Beweis.*

(1) Wenn $(x_1, \dots, x_k) \in \mathbb{Z}_p^k$ eine Lösung ist, dann ist das Tupel $(x_1^{(n)}, \dots, x_k^{(n)})$ der $n$-ten Komponenten der ganzen $p$-adischen Zahlen $x_j$ eine Lösung in $\mathbb{Z}/p^n\mathbb{Z}$. Wir können $x_j^{(n)}$ durch eine ganze Zahl $x_j'$ repräsentieren; dann folgt, dass $F(x_1', \dots, x_k')$ durch $p^n$ teilbar ist.

Die Gegenrichtung ist die eigentlich interessante Aussage. Sei für $n \geq 1$ das Tupel $(x_1^{(n)}, \dots, x_k^{(n)}) \in \mathbb{Z}^k$ so gewählt, dass $p^n \mid F(x_1^{(n)}, \dots, x_k^{(n)})$ gilt. Da mit $\mathbb{Z}_p$ auch $\mathbb{Z}_p^k$ kompakt ist, gibt es eine Teilfolge, die in $\mathbb{Z}_p^k$ konvergiert; $(x_1, \dots, x_k)$ sei der Grenzwert. Da $F$ als Polynom stetig ist (das folgt aus den Eigenschaften eines Absolutbetrages), gilt

$$F(x_1, \dots, x_k) = \lim_{m\to\infty} F(x_1^{(n_m)}, \dots x_k^{(n_m)}) = 0\,,$$

denn $|F(x_1^{(n)}, \dots, x_k^{(n)})|_p \leq p^{-n}$. Dabei ist $(n_m)_m$ eine Teilfolge der Indizes, sodass $(x_1^{(n_m)}, \dots x_k^{(n_m)})$ in $\mathbb{Z}_p^k$ konvergiert.

(2) Die erste Äquivalenz wird wie in (1) gezeigt (und gilt auch ganz allgemein). Dazu beachte man, dass $\mathbb{Z}_p^k \setminus (p\mathbb{Z}_p)^k$ ebenfalls kompakt ist, denn $(p\mathbb{Z}_p)^k$ ist offen.

Die Richtung „$\Rightarrow$“ der zweiten Äquivalenz ist trivial. Zum Beweis der anderen Richtung sei $(x_1, \dots, x_k) \in \mathbb{Q}_p^k \setminus \{0\}$ mit $F(x_1, \dots, x_k) = 0$, und $F$ sei homogen vom Grad $d$. Da nicht alle $x_j$ verschwinden, ist $v = \min\{v_p(x_j) \mid 1 \leq j \leq k\}$ eine wohldefinierte ganze Zahl. Wir setzen $y_j = p^{-v}x_j$; dann ist

$$v_p(y_j) = v_p(x_j) - v \geq 0 \quad \text{mit Gleichheit für wenigstens ein } j.$$

Das bedeutet gerade, dass $(y_1, \dots, y_k) \in \mathbb{Z}_p^k \setminus (p\mathbb{Z}_p)^k$ ist. Außerdem gilt

$$F(y_1, \dots, y_k) = F(p^{-v}x_1, \dots, p^{-v}x_k) = p^{-dv}F(x_1, \dots, x_k) = 0\,. \qquad \square$$

7.13. **Folgerung.** *Sei $F \in \mathbb{Z}[x_1, \dots, x_k]$ ein Polynom mit ganzzahligen Koeffizienten. Dann sind äquivalent:*

**FOLG** Lösbarkeit in allen $\mathbb{Z}_p$ vs. Lösbarkeit mod jedem $n$

(i) *Für jedes $n \geq 1$ gibt es $(x_1, \dots, x_k) \in \mathbb{Z}^k$ mit $F(x_1, \dots, x_k) \equiv 0 \bmod n$.*
(ii) *Für jede Primzahl $p$ gibt es $(x_1, \dots, x_k) \in \mathbb{Z}_p^k$ mit $F(x_1, \dots, x_k) = 0$.*

*Wenn $F$ homogen ist, dann sind die folgenden Aussagen äquivalent:*

(i) *Für jedes $n \geq 1$ gibt es $(x_1, \dots, x_k) \in \mathbb{Z}^k$ mit $F(x_1, \dots, x_k) \equiv 0 \bmod n$ und $\mathrm{ggT}(x_1, \dots, x_k, n) = 1$.*
(ii) *Für jede Primzahl $p$ gibt es $(x_1, \dots, x_k) \in \mathbb{Q}_p^k \setminus \{0\}$ mit $F(x_1, \dots, x_k) = 0$.*

*Beweis.* Nach dem Chinesischen Restsatz sind die ersten Aussagen jeweils äquivalent zu der gleichen Aussage mit $n$ eingeschränkt auf Primzahlpotenzen. Nach Satz 7.12 ist für jede feste Primzahl $p$ die erste Aussage für alle $n = p^m$ äquivalent zur zweiten Aussage für $p$. $\square$

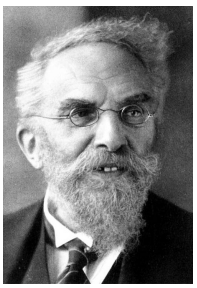
Zur Vereinheitlichung der Schreibweise setzen wir $\mathbb{Q}_\infty = \mathbb{R}$ (analog zu $|\cdot|_\infty = |\cdot|$).

Das nächste Resultat ist eines der wichtigsten in der Theorie der $p$-adischen Zahlen. Es erlaubt uns nämlich, die Existenz einer Lösung in $\mathbb{Z}_p$ auf ein endliches Problem zu reduzieren.

**SATZ** Henselsches Lemma

7.14. **Satz.** *Sei $f \in \mathbb{Z}_p[x]$. Wir schreiben $\bar f \in \mathbb{F}_p[x]$ für das Polynom, dessen Koeffizienten durch Reduktion der Koeffizienten von $f$ modulo $p$ enstehen. Sei weiter $a \in \mathbb{F}_p$ mit $\bar f(a) = 0$ und $\bar f'(a) \neq 0$ (d.h., $a$ ist eine einfache Nullstelle von $\bar f$). Dann gibt es ein eindeutig bestimmtes $\alpha \in \mathbb{Z}_p$ mit $f(\alpha) = 0$ und $\bar\alpha = a$ (in $\mathbb{Z}_p/p\mathbb{Z}_p = \mathbb{F}_p$).*

Die Voraussetzung kann auch so formuliert werden: Es gibt ein $a \in \mathbb{Z}_p$, sodass $f(a) \equiv 0 \bmod p$ und $f'(a) \not\equiv 0 \bmod p$ gelten. Es ist dann $\alpha \equiv a \bmod p$.



K. Hensel (1861–1941)

Die Ableitung $f'$ wird formal entsprechend den Ableitungsregeln gebildet:

$$f = c_n x^n + \cdots + c_1 x + c_0 \quad \Longrightarrow \quad f' = n c_n x^{n-1} + \cdots + c_1 \,.$$

Es gilt dann

$$f(y) = f(x) + (y-x)f'(x) + (y-x)^2 g(x,y)$$

mit einem Polynom $g \in \mathbb{Z}_p[x,y]$.

*Beweis.* Die Idee für den Beweis kommt (vielleicht etwas überraschend) vom Newton-Verfahren.

Zuerst bemerken wir, dass für alle $\alpha \in \mathbb{Z}_p$ mit $\bar\alpha = a$ gilt $f(\alpha) \equiv 0 \bmod p$ und $f'(\alpha) \not\equiv 0 \bmod p$ (genauer gilt $\overline{f'(\alpha)} = \bar f'(a)$). Insbesondere ist $f'(\alpha) \in \mathbb{Z}_p^\times$, denn $|f'(\alpha)|_p = 1$.

Sei jetzt $\alpha_0 \in \mathbb{Z}_p$ beliebig mit $\bar\alpha_0 = a$. Wir definieren rekursiv

$$\alpha_{n+1} = \alpha_n - f'(\alpha_n)^{-1} f(\alpha_n) \,.$$

Ein einfaches Induktionsargument zeigt für alle $n \geq 0$, dass $f(\alpha_n) \equiv 0 \bmod p$ gilt; damit ist $\alpha_{n+1} \equiv \alpha_n \equiv \alpha_0 \bmod p$ und auch $f'(\alpha_n) \in \mathbb{Z}_p^\times$. Insbesondere sind die $\alpha_n$ wohldefiniert. Die auf diese Weise definierte Folge $(\alpha_n)$ konvergiert in $\mathbb{Z}_p$. Sei nämlich $g \in \mathbb{Z}_p[x,y]$ wie oben; dann gilt

$$\begin{aligned} f(\alpha_{n+1}) &= f(\alpha_n) + (\alpha_{n+1} - \alpha_n) f'(\alpha_n) + (\alpha_{n+1} - \alpha_n)^2 g(\alpha_n, \alpha_{n+1}) \\ &= f'(\alpha_n)^{-2} f(\alpha_n)^2 g(\alpha_n, \alpha_{n+1}) \,. \end{aligned}$$

Das zeigt

$$|f(\alpha_{n+1})|_p = |f(\alpha_n)|_p^2 \, |g(\alpha_n, \alpha_{n+1})|_p \leq |f(\alpha_n)|_p^2 \,.$$

Da $|f(\alpha_0)|_p < 1$ ist, gilt $f(\alpha_n) \to 0$ für $n \to \infty$. Aus der Definition von $(\alpha_n)$ folgt dann, dass auch $\alpha_{n+1} - \alpha_n$ gegen null geht. Nach Lemma 7.10 konvergiert demnach die Reihe $\sum_{n=0}^\infty (\alpha_{n+1} - \alpha_n)$; das heißt aber gerade, dass die Folge $(\alpha_n)$ konvergiert. Sei $\alpha = \lim_{n\to\infty} \alpha_n$ der Grenzwert. Da $f$ als Polynom stetig ist, folgt

$$f(\alpha) = \lim_{n\to\infty} f(\alpha_n) = 0 \,.$$

Es bleibt die Eindeutigkeit von $\alpha$ zu zeigen. Sei also $\alpha'$ eine weitere Nullstelle von $f$ mit $\bar\alpha' = a$. Dann gilt

$$0 = f(\alpha') - f(\alpha) = (\alpha' - \alpha)\big(f'(\alpha) + (\alpha' - \alpha) g(\alpha, \alpha')\big) \,.$$

Da $|f'(\alpha)|_p = 1$, $|\alpha' - \alpha|_p < 1$ und $|g(\alpha,\alpha')|_p \leq 1$ gilt, kann der zweite Faktor nicht verschwinden. Es folgt also $\alpha' = \alpha$. ❑

Der Beweis ist konstruktiv: Er sagt uns, wie wir $\alpha$ mit beliebig vorgegebener $p$-adischer Genauigkeit bestimmen können.

Das folgende Resultat zeigt die Nützlichkeit des Henselschen Lemmas.

7.15. **Lemma.** *Sei $p$ eine ungerade Primzahl und sei $a \in \mathbb{Z}_p^\times$. $a$ ist genau dann ein Quadrat in $\mathbb{Z}_p$, wenn $a$ ein quadratischer Rest mod $p$ ist (d.h., $\bar{a} \in \mathbb{F}_p^\times$ ist ein Quadrat).*

*Wenn $a \in \mathbb{Z}_2^\times$ ist, dann ist $a$ genau dann ein Quadrat in $\mathbb{Z}_2$, wenn $a \equiv 1 \bmod 8$ ist.*

**LEMMA** Quadrate in $\mathbb{Z}_p$

*Beweis.* Es ist klar, dass ein Quadrat in $\mathbb{Z}_p$ insbesondere ein Quadrat mod $p$ (bzw. mod 8 für $p = 2$) sein muss. Für die Gegenrichtung sei zunächst $p$ ungerade. Wir betrachten $f(x) = x^2 - a$. Wenn $a$ ein quadratischer Rest mod $p$ ist, dann gibt es $s \in \mathbb{F}_p$ mit $\bar{f}(s) = 0$. Außerdem ist $\bar{f}'(s) = 2s \neq 0$. Nach Satz 7.14 gibt es also (genau) ein $\sigma \in \mathbb{Z}_p$ mit ($\bar{\sigma} = s$ und) $\sigma^2 = a$.

Im Fall $p = 2$ müssen wir etwas anders vorgehen, da die Ableitung von $x^2 - a$ immer gerade ist. Wir schreiben $a = 8A + 1$ mit $A \in \mathbb{Z}_2$ und betrachten jetzt $f(x) = 2x^2 + x - A$. Dann ist $f(A) = 2A^2$ gerade und $f'(A) = 4A + 1$ ungerade. Wir können also wieder Satz 7.14 anwenden: $f$ hat eine Nullstelle $\alpha \in \mathbb{Z}_2$. Es folgt $(4\alpha + 1)^2 - a = 8f(\alpha) = 0$. ❑

Das zeigt zum Beispiel, dass $-7$ ein Quadrat in $\mathbb{Z}_2$ ist.

Allgemeiner haben wir folgendes Resultat über Quadrate in $\mathbb{Q}_p$.

7.16. **Lemma.** *Sei $a \in \mathbb{Q}_p^\times$; dann ist $a = p^n u$ mit $u \in \mathbb{Z}_p^\times$ und $n = v_p(a) \in \mathbb{Z}$. $a$ ist genau dann ein Quadrat in $\mathbb{Q}_p$, wenn $n$ gerade ist und $u$ ein Quadrat in $\mathbb{Z}_p$ ist.*

**LEMMA** Quadrate in $\mathbb{Q}_p$

*Beweis.* Dass die Bedingung hinreichend ist, ist klar. Ist umgekehrt $a = b^2$, dann muss $n = v_p(a) = 2v_p(b)$ gerade sein, und $u = (b/p^{n/2})^2 \in \mathbb{Z}_p^\times$ ist ein Quadrat in $\mathbb{Q}_p$. Es ist aber $v_p(b/p^{n/2}) = 0$, also ist $u$ das Quadrat eines Elements von $\mathbb{Z}_p$. ❑

Als eine weitere Folgerung sehen wir, dass die $(p-1)$-ten Einheitswurzeln in $\mathbb{Z}_p$ enthalten sind.

7.17. **Folgerung.** *Sei $p$ eine Primzahl. Dann hat $x^{p-1} - 1$ in $\mathbb{Z}_p$ genau $(p-1)$ verschiedene Nullstellen.*

**FOLG** Einheitswurzeln in $\mathbb{Z}_p$

*Beweis.* Nach dem kleinen Satz von Fermat gilt $a^{p-1} = 1$ für jedes $a \in \mathbb{F}_p^\times$. Sei $f(x) = x^{p-1} - 1$, dann gilt also für jedes $0 \neq a \in \mathbb{F}_p$, dass

$$\bar{f}(a) = 0 \qquad \text{und} \qquad \bar{f}'(a) = (p-1)a^{p-2} = -a^{p-2} \neq 0$$

ist. Nach Satz 7.14 folgt, dass es genau eine Nullstelle $\alpha \in \mathbb{Z}_p$ von $f$ gibt mit $\bar{\alpha} = a$. Das zeigt, dass es mindestens $(p-1)$ verschiedene Nullstellen in $\mathbb{Z}_p$ gibt. Auf der anderen Seite kann ein Polynom vom Grad $(p-1)$ im Integritätsbereich $\mathbb{Z}_p$ aber auch nicht mehr als $(p-1)$ verschiedene Nullstellen besitzen. ❑

## 8. Der Satz von Hasse und das Normrestsymbol

Wir wollen jetzt das, was wir über $p$-adische Zahlen gelernt haben, auf quadratische Formen anwenden.

Zum Beispiel sehen wir, dass die notwendigen Bedingungen 6.11 für die nichttriviale Lösbarkeit von $ax^2 + by^2 + cz^2 = 0$ (mit $abc$ quadratfrei) die nichttriviale Lösbarkeit in $\mathbb{R}$ (das ist einfach Bedingung (1)) und in allen $\mathbb{Q}_p$ impliziert. Ist etwa $p$ ungerade mit $p \mid a$, dann sagt Bedingung (6), dass es $\bar{y}, \bar{z} \in \mathbb{F}_p^\times$ gibt mit $b\bar{y}^2 + c\bar{z}^2 = 0$. Sei $z \in \mathbb{Z}_p$ ein Repräsentant von $\bar{z}$. Dann hat das Polynom $f(X) = bX^2 + cz^2 \in \mathbb{Z}_p[X]$ eine einfache Nullstelle $\bar{y}$ in $\mathbb{F}_p$ (denn $f'(\bar{y}) = 2\bar{b}\bar{y} \neq 0$), also gibt es nach Satz 7.14 eine Nullstelle in $\mathbb{Z}_p$. Alternativ können wir mit Lemma 7.15 argumentieren, dass $-bc$ ein Quadrat in $\mathbb{Z}_p$ sein muss und daraus eine Lösung in $\mathbb{Z}_p$ konstruieren. Falls $p$ ungerade ist und $abc$ nicht teilt, dann gibt es eine Lösung mod $p$ nach Lemma 5.5, und wir können wie eben verfahren (wobei wir die Variable im Polynom so wählen, dass die Nullstelle mod $p$ nicht bei null liegt). Im Fall $p = 2$ können wir erreichen, dass wir Quadratwurzeln aus Zahlen ziehen müssen, die $\equiv 1 \bmod 8$ sind, was nach Lemma 7.15 in $\mathbb{Z}_2$ immer möglich ist.

Man beachte, dass wir den Satz von Legendre in dieser Überlegung nicht benutzt haben. Natürlich kann man auch argumentieren, dass es nach dem Satz von Legendre sogar eine nichttriviale ganzzahlige Lösung geben muss, die dann auch eine reelle bzw. eine $p$-adische Lösung ist. Der Punkt ist, dass wir unabhängig von Resultaten über Lösungen in $\mathbb{Z}$ oder in $\mathbb{Q}$ mit endlichem Aufwand entscheiden können, ob es reelle und $p$-adische Lösungen (für alle Primzahlen $p$) gibt. Dies gilt auch noch in sehr viel allgemeineren Situationen, wo es kein dem Satz von Legendre vergleichbares Resultat gibt.

In jedem Fall sehen wir, dass der Satz von Legendre 6.12 und auch seine allgemeinere Form in Satz 6.15 zu folgender Aussage äquivalent sind.

**SATZ**
Hasse-Prinzip für ternäre qu. Formen

8.1. **Satz.** *Sei $Q(x, y, z)$ eine nicht-ausgeartete ternäre quadratische Form. Dann sind äquivalent:*

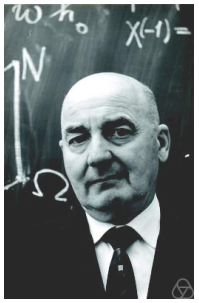
(1) *$Q(x, y, z) = 0$ hat eine primitive ganzzahlige Lösung.*

(2) *$Q(x, y, z) = 0$ hat eine nichttriviale Lösung in $\mathbb{R}$ und in $\mathbb{Q}_p$ für jede Primzahl $p$.*

*Hierbei kann in der zweiten Aussage entweder auf die Lösbarkeit in $\mathbb{R}$ oder auf die Lösbarkeit in $\mathbb{Q}_2$ verzichtet werden.*

*Beweis.* Zunächst einmal ist klar, dass beide Aussagen wahr bzw. falsch bleiben, wenn wir $Q$ durch eine äquivalente Form $Q'$ ersetzen. Es genügt also, sich auf diagonale Formen $Q = ax^2 + by^2 + cz^2$ mit $abc$ quadratfrei zu beschränken.

Die Implikation (1) $\Rightarrow$ (2) ist trivial. Die umgekehrte Implikation folgt aus dem Satz von Legendre 6.12; der Zusatz folgt aus den beiden Beweisvarianten, die entweder die 2-adische oder die reelle Lösbarkeit nicht verwenden. ❑



H. Hasse (1898–1979) © MFO

Dieses Ergebnis wird das *Hasse-Prinzip* oder das *Lokal-Global-Prinzip* für ternäre quadratische Formen genannt. Es besagt, dass die Existenz „lokaler“ Lösungen (in $\mathbb{R}$ und in $\mathbb{Q}_p$ für alle $p$) die Existenz „globaler“ Lösungen (in $\mathbb{Q}$) zur Folge hat.

Die Aussage gilt allgemeiner für quadratische Formen in beliebig vielen Variablen; wir werden das später sehen.

Das Lokal-Global-Prinzip gilt keineswegs immer. Ein berühmtes Gegenbeispiel[9] ist die Gleichung

$$3\,x^3 + 4\,y^3 + 5\,z^3 = 0\,.$$

Man kann (relativ leicht) zeigen, dass sie nichttriviale reelle und $p$-adische Lösungen hat (für alle Primzahlen $p$), aber (das ist schwieriger) keine nichttriviale rationale Lösung.

Da wir die folgende Überlegung mehrfach verwenden werden, formulieren wir sie als ein Lemma.

**LEMMA**
hinr. Bed. für $p$-adische Lösbarkeit

8.2. **Lemma.** *Sei $Q(x_1,\dots,x_n) = a_1x_1^2 + \cdots + a_nx_n^2$ eine nicht-ausgeartete diagonale quadratische Form. Wenn $p$ eine ungerade Primzahl ist und mindestens drei der Koeffizienten $a_j$ $p$-adische Einheiten sind, dann hat $Q = 0$ nichttriviale Lösungen in $\mathbb{Q}_p$.*

*Beweis.* Wir können (eventuell nach Vertauschen der Variablen) annehmen, dass $a_1$, $a_2$ und $a_3$ $p$-adische Einheiten, also nicht durch $p$ teilbare ganze Zahlen sind. Nach Lemma 5.5 gibt es dann $u, v \in \mathbb{Z}$ mit $a_1 \equiv -a_2u^2 - a_3v^2 \bmod p$. Dann ist (in $\mathbb{Z}_p$) $(-a_2u^2 - a_3v^2)/a_1 \equiv 1 \bmod p$; nach Lemma 7.15 ist also

$$(-a_2u^2 - a_3v^2)/a_1 = t^2 \quad \text{für ein } t \in \mathbb{Z}_p.$$

Damit ist $(x_1,\dots,x_n) = (t,u,v,0,\dots,0)$ eine nichttriviale Lösung von $Q = 0$. ❑

Wir wollen jetzt das Phänomen, dass man eine der lokalen Bedingungen in Satz 8.1 weglassen kann, genauer untersuchen. Dazu führen wir das *Normrestsymbol* ein. Es beschreibt die reelle bzw. $p$-adische Lösbarkeit ternärer quadratischer Formen.

Wir hatten gesehen, dass jede nicht-ausgeartete ternäre quadratische Form zu einer Diagonalform $ax^2+by^2+cz^2$ (mit $abc \neq 0$) äquivalent ist. Diese ist wiederum äquivalent zu $(-ac)x^2 + (-bc)y^2 - z^2$ (multipliziere mit $-c$ und ersetze $z$ durch $z/c$). Wir können uns also auf Formen der Gestalt $ax^2 + by^2 - z^2$ beschränken.

**DEF**
Normrest-symbol

8.3. **Definition.** Seien $a, b \in \mathbb{Q}^\times$. Für eine Primzahl $p$ setzen wir

$$\Big(\frac{a,b}{p}\Big) = 1\,,$$

falls $a\,x^2 + b\,y^2 = z^2$ eine nichttriviale Lösung in $\mathbb{Q}_p$ hat. Andernfalls setzen wir

$$\Big(\frac{a,b}{p}\Big) = -1\,.$$

In ähnlicher Weise definieren wir

$$\Big(\frac{a,b}{\infty}\Big) = 1\,,$$

falls es eine nichttriviale reelle Lösung gibt, andernfalls

$$\Big(\frac{a,b}{\infty}\Big) = -1\,.$$

Das Symbol $\big(\frac{a,b}{v}\big)$ heißt *(quadratisches) Hilbertsches Normrestsymbol* oder kürzer *Normrestsymbol* oder *Hilbertsymbol*. ♢

D. Hilbert (1862–1943)

[9] E.S. Selmer: *The Diophantine equation $ax^3 + by^3 + cz^3 = 0$*, Acta Math. **85** (1951), 203–362.

Satz 8.1 sagt dann, dass $a\,x^2+b\,y^2=z^2$ genau dann eine nichttriviale Lösung in $\mathbb{Q}$ hat, wenn

$$\left(\frac{a,b}{v}\right)=1 \qquad \text{ist für alle „Stellen“ } v=p,\ v=\infty.$$

Wir wollen jetzt den Wert des Normrestsymbols bestimmen. Wir beginnen mit dem einfachsten Fall.

8.4. **Lemma.** *Für $a,b\in\mathbb{Q}^\times$ gilt*

$$\left(\frac{a,b}{\infty}\right)=-1 \iff a<0 \text{ und } b<0\,.$$

**LEMMA** $\left(\frac{a,b}{\infty}\right)$

*Insbesondere ist das Symbol $\left(\frac{a,b}{\infty}\right)$ in beiden Argumenten multiplikativ.*

*Beweis.* Das ist klar. ❑

Für die anderen Fälle ist folgende Eigenschaft nützlich.

8.5. **Lemma.** *Für $a,b,c\in\mathbb{Q}^\times$ und jede „Stelle“ $v$ gilt*

$$\left(\frac{ab,ac}{v}\right)=\left(\frac{ab,-bc}{v}\right).$$

**LEMMA** Relation für Normrestsymbol

*Beweis.* Die Formen $ab\,x^2+ac\,y^2-z^2$ und $ab\,x^2-bc\,y^2-z^2$ sind äquivalent (multipliziere mit $-ab$, skaliere $x$ und $y$, sodass die Quadrate in den Koeffizienten verschwinden, und vertausche $x$ und $z$). ❑

Eine weitere einfache Beobachtung ist, dass

$$\left(\frac{as,bt}{v}\right)=\left(\frac{a,b}{v}\right) \quad \text{ist, wenn } s \text{ und } t \text{ Quadrate in } \mathbb{Q}_v \text{ sind,}$$

und natürlich ist das Symbol symmetrisch:

$$\left(\frac{a,b}{v}\right)=\left(\frac{b,a}{v}\right).$$

8.6. **Lemma.** *Seien $p$ eine ungerade Primzahl und $u_1,u_2\in\mathbb{Q}^\times\cap\mathbb{Z}_p^\times$. Dann gilt*

$$\left(\frac{u_1,u_2}{p}\right)=1\,,\quad \left(\frac{u_1p,u_2}{p}\right)=\left(\frac{u_2}{p}\right) \quad \textit{und} \quad \left(\frac{u_1p,u_2p}{p}\right)=\left(\frac{-u_1u_2}{p}\right).$$

**LEMMA** Normrestsymbol für $p$ ungerade

*Das Symbol $\left(\frac{a,b}{p}\right)$ ist in beiden Argumenten multiplikativ.*

*Beweis.* Die Gleichung $u_1x^2+u_2y^2=z^2$ hat nach Lemma 8.2 stets eine nichttriviale in $\mathbb{Q}_p$.

Wenn $u_2$ ein quadratischer Rest mod $p$ ist, dann gibt es nach Lemma 7.15 eine Quadratwurzel $s$ von $u_2$ in $\mathbb{Z}_p$; damit ist $(x,y,z)=(0,1,s)$ eine nichttriviale Lösung von $u_1px^2+u_2y^2=z^2$. Gibt es umgekehrt eine nichttriviale Lösung in $\mathbb{Q}_p$, dann gibt es auch eine primitive Lösung in $\mathbb{Z}_p$ (hier heißt „primitiv“, dass nicht alle Variablen durch $p$ teilbar sind). In einer primitiven Lösung kann $p$ nicht $y$ und $z$ teilen; Reduktion mod $p$ zeigt dann, dass $u_2$ ein quadratischer Rest mod $p$ sein muss.

Die dritte Formel folgt aus Lemma 8.5 und der zweiten Formel.

Die Multiplikativität prüft man für die verschiedenen Fälle nach; beachte, dass Potenzen von $p^2$ in den Argumenten ignoriert werden können. ❑

Der komplizierteste Fall ist $v = 2$.

**LEMMA** Normrestsymbol für $p = 2$

8.7. **Lemma.** *Seien $u_1, u_2 \in \mathbb{Q}^\times \cap \mathbb{Z}_2^\times$. Dann gilt*

$$\Big(\frac{u_1, u_2}{2}\Big) = (-1)^{\frac{u_1-1}{2}\,\frac{u_2-1}{2}}\,, \quad \Big(\frac{2u_1, u_2}{2}\Big) = (-1)^{\frac{u_2^2-1}{8}}(-1)^{\frac{u_1-1}{2}\,\frac{u_2-1}{2}}\,,$$
$$\Big(\frac{2u_1, 2u_2}{2}\Big) = (-1)^{\frac{u_1^2u_2^2-1}{8}}(-1)^{\frac{u_1-1}{2}\,\frac{u_2-1}{2}}\,.$$

*Das Symbol $\left(\frac{a,b}{2}\right)$ ist in beiden Argumenten multiplikativ.*

*Beweis.* Da wir $u_1$ und $u_2$ mit beliebigen rationalen Zahlen multiplizieren können, die Quadrate in $\mathbb{Q}_2$ sind, genügt es, die Fälle $u_j = \pm 1, \pm 3$ zu betrachten. In jedem Fall sieht man, dass man entweder mithilfe von Lemma 7.15 eine Lösung konstruieren kann, oder dass man einen Widerspruch mod 8 erhält, entsprechend dem behaupteten Wert des Symbols.

Die Multiplikativität kann man wiederum fallweise nachprüfen. ❑

Um zu verstehen, woher der Name „*Norm*restsymbol“ kommt, schreiben wir die Gleichung $ax^2 + by^2 = z^2$ um als $z^2 - ax^2 = by^2$. Wenn es nichttriviale Lösungen gibt, dann gibt es auch eine Lösung mit $y = 1$ (siehe Lemma 8.11 unten). Es gibt dann also $X, Y \in \mathbb{Q}_v$ mit $X^2 - aY^2 = b$ (mit $(x, y, z) = (Y, 1, X)$).

Wenn $a$ kein Quadrat in $\mathbb{Q}_v$ ist, dann ist die linke Seite $X^2 - aY^2$ gerade die *Norm* des Elements $X + Y\sqrt{a}$ in der Körpererweiterung $\mathbb{Q}_v \subset \mathbb{Q}_v(\sqrt{a})$. Ganz allgemein definiert man die Norm für eine endliche Körpererweiterung $K \subset L$ wie folgt. Zu jedem $\alpha \in L$ hat man die Abbildung $m_\alpha \colon L \to L$, $x \mapsto \alpha x$; diese Abbildung ist $K$-linear (sogar $L$-linear). Man setzt dann

$$N_{L/K} \colon L \longrightarrow K, \quad \alpha \longmapsto \det(m_\alpha)$$

(die Determinante ist über $K$ zu verstehen). Dann gilt $N_{L/K}(\alpha) = 0 \iff \alpha = 0$ (denn anderenfalls ist $m_\alpha$ invertierbar) und

$$N_{L/K}(\alpha\beta) = \det(m_{\alpha\beta}) = \det(m_\alpha \circ m_\beta) = \det(m_\alpha)\det(m_\beta) = N_{L/K}(\alpha)N_{L/K}(\beta)\,;$$

die Norm ist also multiplikativ und ergibt somit einen Gruppenhomomorphismus $N_{L/K} \colon L^\times \to K^\times$. Insbesondere ist die Menge $N_{L/K}(L^\times)$ aller Normen in $K^\times$ eine Untergruppe.

Für die quadratische Erweiterung $\mathbb{Q}_v \subset \mathbb{Q}_v(\sqrt{a})$ ist eine $\mathbb{Q}_v$-Basis gegeben durch $(1, \sqrt{a})$. Wenn wir für $\alpha = x + y\sqrt{a}$ die Abbildung $m_\alpha$ in dieser Basis ausdrücken, sehen wir, dass

$$N_{L/K}(x + y\sqrt{a}) = \begin{vmatrix} x & ay \\ y & x \end{vmatrix} = x^2 - ay^2$$

ist. Das Symbol $\left(\frac{a,b}{v}\right)$ hat also genau dann den Wert 1, wenn ($a$ ein Quadrat in $\mathbb{Q}_v$ oder) $b$ eine Norm dieser Körpererweiterung ist. Da die Normen eine Untergruppe $N_a$ von $\mathbb{Q}_v^\times$ bilden, folgt

$$\Big(\frac{a, b_1}{v}\Big) = 1\,, \quad \Big(\frac{a, b_2}{v}\Big) = 1 \quad \Longrightarrow \quad \Big(\frac{a, b_1b_2}{v}\Big) = 1 \qquad \text{und}$$
$$\Big(\frac{a, b_1}{v}\Big) = 1\,, \quad \Big(\frac{a, b_2}{v}\Big) = -1 \quad \Longrightarrow \quad \Big(\frac{a, b_1b_2}{v}\Big) = -1\,.$$

(Diese Eigenschaften des Normrestsymbols kann man übrigens verwenden, um die Bestimmung der Werte bei $v = 2$ etwas zu vereinfachen.)

Die für die Multiplikativität des Normrestsymbols noch fehlende Aussage

$$\left(\frac{a,b_1}{v}\right) = -1\,, \quad \left(\frac{a,b_2}{v}\right) = -1 \quad \Longrightarrow \quad \left(\frac{a,b_1b_2}{v}\right) = 1$$

ist dann äquivalent dazu, dass die Normgruppe $N_a$ Index 2 in $\mathbb{Q}_v^\times$ hat. In jedem Fall sehen wir, dass es zum Nachweis der Multiplikativität genügt, nur diesen letzten Fall zu betrachten.

Die Aussage, dass der Index der Normuntergruppe in diesem Fall immer 2 ist, ist ein Spezialfall eines allgemeineren Resultats der sogenannten „lokalen Klassenkörpertheorie."

Wir kommen jetzt zum Hauptergebnis über das Normrestsymbol.

**SATZ** Produktformel für das Normrestsymbol

8.8. **Satz.** *Seien $a, b \in \mathbb{Q}^\times$. Dann ist für alle bis auf endlich viele Primzahlen $p$*

$$\left(\frac{a,b}{p}\right) = 1\,,$$

*und wir haben die Produktformel*

$$\prod_{v=p,\infty} \left(\frac{a,b}{v}\right) = 1\,.$$

*(Dabei läuft $v$ über alle Primzahlen und $\infty$.)*

Die Produktformel besagt, dass es immer eine (endliche und) *gerade* Anzahl von Stellen $v$ gibt, sodass $ax^2 + by^2 = z^2$ keine nichttrivialen Lösungen in $\mathbb{Q}_v$ hat.

*Beweis.* Wir können annehmen, dass $a$ und $b$ quadratfreie ganze Zahlen sind (denn wir können $a$ und $b$ mit Quadraten rationaler Zahlen multiplizieren, ohne den Wert der Symbole zu verändern). Nach Lemma 8.6 ist dann $\left(\frac{a,b}{p}\right) = 1$ für alle ungeraden Primzahlen $p$, die weder $a$ noch $b$ teilen. Das zeigt die erste Behauptung. Insbesondere ist das Produkt definiert.

Wir haben gesehen, dass alle Symbole im Produkt symmetrisch und in beiden Argumenten multiplikativ sind. Daher genügt es, folgende Fälle zu betrachten:

$$(a,b) = (-1,-1),\ (-1,2),\ (-1,p),\ (2,2),\ (2,p),\ (p,p),\ (p,q)\,.$$

Dabei sind $p$ und $q$ verschiedene ungerade Primzahlen. Wegen

$$\left(\frac{a,a}{v}\right) = \left(\frac{-1,a}{v}\right)$$

reduzieren sich die Fälle $(a,b) = (2,2)$ und $(p,p)$ auf $(a,b) = (-1,2)$ und $(-1,p)$. Für die verbleibenden Fälle ergibt sich die folgende Tabelle (alle anderen Symbole sind 1):

| $(a,b)$ | $\left(\frac{a,b}{\infty}\right)$ | $\left(\frac{a,b}{2}\right)$ | $\left(\frac{a,b}{p}\right)$ | $\left(\frac{a,b}{q}\right)$ |
|---|---|---|---|---|
| $(-1,-1)$ | $-1$ | $-1$ | $+1$ | $+1$ |
| $(-1,2)$ | $+1$ | $+1$ | $+1$ | $+1$ |
| $(-1,p)$ | $+1$ | $(-1)^{\frac{p-1}{2}}$ | $\left(\frac{-1}{p}\right)$ | $+1$ |
| $(2,p)$ | $+1$ | $(-1)^{\frac{p^2-1}{8}}$ | $\left(\frac{2}{p}\right)$ | $+1$ |
| $(p,q)$ | $+1$ | $(-1)^{\frac{p-1}{2}\,\frac{q-1}{2}}$ | $\left(\frac{q}{p}\right)$ | $\left(\frac{p}{q}\right)$ |

Wir sehen, dass die Produktformel genau zum Quadratischen Reziprozitätsgesetz 3.13 und seinen beiden Ergänzungsgesetzen 3.10 und 3.12 äquivalent ist. ❑

Die Produktformel für das Normrestsymbol ist also nur eine andere Möglichkeit, das Quadratische Reziprozitätsgesetz samt Ergänzungsgesetzen zu formulieren. In gewisser Weise ist die Produktformel schöner, da sie eine einzige einfache Aussage darstellt anstelle von drei verschiedenen.

Die Produktformel für das Normrestsymbol führt zu folgender Verbesserung von Satz 8.1.

**SATZ**
Hasse-Prinzip für ternäre qu. Formen: Verbesserung

8.9. **Satz.** *Sei $Q(x,y,z)$ eine nicht-ausgeartete ternäre quadratische Form. Dann sind äquivalent:*

(1) *$Q(x,y,z) = 0$ hat eine primitive ganzzahlige Lösung.*

(2) *$Q(x,y,z) = 0$ hat eine nichttriviale Lösung in $\mathbb{Q}_v$ für alle bis auf eventuell eine Stelle $v$ ($v = p$ Primzahl oder $v = \infty$).*

*Beweis.* Wir können $Q$ durch eine äquivalente Form der Gestalt $ax^2 + by^2 - z^2$ ersetzen. Die Implikation „(1) ⇒ (2)“ ist wieder trivial. Für die Gegenrichtung treffe (2) zu. Aus der Produktformel 8.8 folgt dann, dass es tatsächlich nichttriviale Lösungen in $\mathbb{Q}_v$ für *alle* $v$ geben muss. Aussage (1) folgt dann aus Satz 8.1. ❑

P.G.L. Dirichlet (1805–1859)

Wir wollen jetzt das Hasse-Prinzip auf quadratische Formen in beliebig vielen Variablen verallgemeinern. Der entscheidende Schritt ist der von drei zu vier Variablen; hierfür brauchen wir den berühmten Satz von Dirichlet über Primzahlen in Restklassen („arithmetischen Progressionen“).

**SATZ**
Satz von Dirichlet

8.10. **Satz.** *Sei $n \geq 1$ und $a \in \mathbb{Z}$ mit $a \perp n$. Dann gibt es unendlich viele Primzahlen $p$ mit $p \equiv a \bmod n$.*

Es ist klar, dass die Voraussetzung $a \perp n$ notwendig ist, denn anderenfalls kann es höchstens eine Primzahl $\equiv a \bmod n$ geben.

*Beweis.* Der Beweis wird mit analytischen Hilfsmitteln (Dirichletschen $L$-Funktionen) geführt. Nachlesen kann man einen Beweis etwa in [IR, Ch. 16] oder [Sch, Kap. 8]. ❑

Man kann die Aussage von Satz 8.10 so formulieren:

*Hat $f = nx + a \in \mathbb{Z}[x]$ die Eigenschaft, dass nicht sämtliche Werte $f(m) \in \mathbb{Z}$ für $m \in \mathbb{Z}$ durch eine feste Primzahl $p$ teilbar sind, dann ist $f(m)$ eine Primzahl für unendlich viele $m \in \mathbb{Z}$.*

(Die Voraussetzung ist äquivalent zu $a \perp n$.) Die „H-Vermutung“ von Schinzel verallgemeinert das zu folgender Aussage:

*Sind $f_1, \ldots, f_k \in \mathbb{Z}[x]$ irreduzible Polynome mit positiven Leitkoeffizienten und mit der Eigenschaft, dass nicht alle Produkte $f_1(m) \cdots f_k(m)$ für $m \in \mathbb{Z}$ durch eine feste Primzahl $p$ teilbar sind, dann gibt es unendlich viele $m \in \mathbb{Z}$, sodass alle Werte $f_1(m), \ldots, f_k(m)$ Primzahlen sind.*

Allerdings ist Satz 8.10 nach wie vor der einzige Fall, in dem diese Vermutung bewiesen werden konnte. Zum Beispiel ist immer noch unbekannt, ob $m^2 + 1$ unendlich oft eine Primzahl ist!

Wir brauchen noch zwei Hilfsaussagen.

**LEMMA**
Nullstelle
⇒ universell
Lösungen
mit $x_n = 1$

8.11. **Lemma.** *Sei $Q(x_1, x_2, \ldots, x_n)$ eine nicht-ausgeartete quadratische Form in $n$ Variablen und sei $\mathbb{Q} \subset K$ eine Körpererweiterung.*

(1) *Wenn $Q$ eine nichttriviale Nullstelle $\boldsymbol{x} \in K^n$ hat, dann gibt es für jedes $a \in K$ einen Vektor $\boldsymbol{y} \in K^n$ mit $Q(\boldsymbol{y}) = a$.*

(2) *Wenn $Q$ die Form $Q'(x_1, \ldots, x_{n-1}) + cx_n^2$ hat und es nichttriviale Lösungen von $Q = 0$ in $K$ gibt, dann gibt es auch Lösungen in $K$ mit $x_n = 1$.*

*Beweis.*

(1) Sei $M$ die symmetrische Matrix mit $Q(\boldsymbol{z}) = \boldsymbol{z}M\boldsymbol{z}^\top$. Da $Q$ nicht-ausgeartet ist, ist $M$ invertierbar; da $\boldsymbol{x} \neq 0$ ist, ist dann $M\boldsymbol{x}^\top \neq 0$. Es gibt daher einen Vektor $\boldsymbol{z} \in K^n$ mit $\boldsymbol{z}M\boldsymbol{x}^\top = \frac{1}{2}$. Wir setzen

$$\boldsymbol{y} = \boldsymbol{z} + (a - Q(\boldsymbol{z}))\boldsymbol{x}\,;$$

dann ist

$$Q(\boldsymbol{y}) = Q(\boldsymbol{z}) + 2(a - Q(\boldsymbol{z}))\cdot \boldsymbol{z}M\boldsymbol{x}^\top + (a - Q(\boldsymbol{z}))^2 \cdot Q(\boldsymbol{x}) = Q(\boldsymbol{z}) + 2(a - Q(\boldsymbol{z}))\cdot \tfrac{1}{2} = a\,.$$

(2) Da $Q$ nicht-ausgeartet ist, ist $c \neq 0$. Es muss dann $n \geq 2$ sein, da es keine nichttriviale Lösung von $cx^2 = 0$ gibt.

Ist $x_n \neq 0$, dann ist $x_n^{-1} \cdot \boldsymbol{x}$ eine Lösung mit letzter Koordinate 1. Ist $x_n = 0$, dann ist $(x_1, \ldots, x_{n-1})$ eine nichttriviale Nullstelle von $Q'$. Nach Teil (1) gibt es $(y_1, \ldots, y_{n-1}) \in K^{n-1}$ mit $Q'(y_1, \ldots, y_{n-1}) = -c$; dann ist $(y_1, \ldots, y_{n-1}, 1)$ eine Lösung von $Q = 0$. ❑

Wir können jetzt das Hasse-Prinzip für quadratische Formen in vier Variablen beweisen.

**SATZ**
Hasse-Prinzip
für qu. Formen
in 4 Variablen

8.12. **Satz.** *Sei $Q(x_1, x_2, x_3, x_4)$ eine nicht-ausgeartete quadratische Form in vier Variablen. Dann sind äquivalent:*

(1) *$Q = 0$ hat eine primitive ganzzahlige Lösung.*

(2) *$Q = 0$ hat eine nichttriviale Lösung in $\mathbb{Q}_v$ für alle Stellen $v$ ($v = p$ Primzahl oder $v = \infty$).*

*Beweis.* Es ist nur „(2) ⇒ (1)“ zu zeigen. Beide Aussagen bleiben wahr oder falsch, wenn wir zu einer äquivalenten quadratischen Form übergehen. Wir können also ohne Einschränkung annehmen, dass $Q = a_1x_1^2 + a_2x_2^2 - a_3x_3^2 - a_4x_4^2$ diagonal ist mit ganzen Zahlen $a_1, \ldots, a_4$. Da die Koeffizienten nicht alle dasselbe Vorzeichen haben können (reelle Lösbarkeit), können wir (nach eventueller Permutation der Variablen) annehmen, dass $a_1, a_4 > 0$ sind.

Für jede Primzahl $p$, die $D := 2a_1a_2a_3a_4$ teilt, gibt es nach Annahme einen gemeinsamen Wert $u_p \in \mathbb{Q}_p$ von $a_1x_1^2 + a_2x_2^2$ und $a_3x_3^2 + a_4x_4^2$. Dabei ist $(x_1, x_2, y_3, x_4) \neq 0$. Wir können sogar $(x_1, x_2), (x_3, x_4) \neq (0, 0)$ annehmen: Denn ist z.B. $x_1 = x_2 = 0$, dann hat $a_3x_3^2 + a_4x_4^2$ eine nichttriviale Nullstelle, stellt also nach Lemma 8.11 (1) jedes Element von $\mathbb{Q}_p$ dar, sodass wir $(x_1, x_2)$ beliebig mit $a_1x_1^2 + a_2x_2^2 \neq 0$ vorgeben können. Wenn $u_p = 0$ ist, dann gibt es nach Lemma 8.11 (2), angewendet auf $a_1x^2 + a_2y^2 - z^2$ und $a_3x^2 + a_4y^2 - z^2$ auch eine Lösung, die $u_p = 1$ liefert. Wir können also $u_p \in \mathbb{Q}_p^\times$ annehmen. Offenbar können wir $u_p$ mit einem beliebigen Quadrat in $\mathbb{Q}_p^\times$ multiplizieren; damit können wir $u_p \in \mathbb{Z}_p^\times \cup p\mathbb{Z}_p^\times$ erreichen. Sei $d$ das Produkt der Primzahlen $p$ mit $u_p \notin \mathbb{Z}_p^\times$.

Nach dem Chinesischen Restsatz gibt es $u \in \mathbb{Z}$ mit $u \equiv u_p \bmod p^2$ für alle ungeraden $p \mid D$ und $u \equiv u_2 \bmod 16$; es ist $u = du'$ mit $\mathrm{ggT}(u', D) = 1$. Nach Satz 8.10 gibt es eine Primzahl $q$ mit $q \equiv u' \bmod 4D^2$. Wir betrachten jetzt die nicht-ausgearteten ternären quadratischen Formen

$$Q_1(x, y, z) = a_1x^2 + a_2y^2 - dqz^2 \quad \text{und} \quad Q_2(x, y, z) = a_3x^2 + a_4y^2 - dqz^2 \,.$$

Wir wollen zeigen, dass beide nichttriviale Nullstellen in $\mathbb{Q}$ haben. Wegen $a_1, a_4 > 0$ gibt es jedenfalls reelle Lösungen von $Q_j = 0$. Für Primzahlen $p \nmid qD$ gibt es nach Lemma 8.2 immer $p$-adische Lösungen. Gilt $p \mid D$, dann ist $dq \equiv u_p \bmod p^2$ (bzw. mod 16 für $p = 2$). Da $v_p(u_p) \in \{0, 1\}$ ist, folgt daraus, dass $dq/u_p \equiv 1 \bmod p$ (bzw. mod 8) und damit ein Quadrat in $\mathbb{Q}_p$ ist (vergleiche Lemma 7.16); also gibt es eine $p$-adische Lösung von $Q_1 = 0$ und von $Q_2 = 0$.

Aus Satz 8.9 folgt jetzt, dass $Q_1 = 0$ und $Q_2 = 0$ nichttriviale Lösungen in $\mathbb{Q}$ haben (wir haben die Existenz von Lösungen in $\mathbb{Q}_v$ für alle $v$ mit der einen Ausnahme $v = q$ nachgewiesen). Nach Lemma 8.11 (2) gibt es dann auch Lösungen mit $z = 1$. Das bedeutet, dass $dq$ sowohl in der Form $a_1x_1^2 + a_2x_2^2$ als auch in der Form $a_3x_3^2 + a_4x_4^2$ (mit $x_1, x_2, x_3, x_4 \in \mathbb{Q}$) geschrieben werden kann. Damit ist $(x_1, x_2, x_3, x_4)$ eine Lösung von $Q = 0$, und nichttrivial, weil $dq \neq 0$ ist. ❑

Damit können wir nun endlich den Drei-Quadrate-Satz 5.9 beweisen.

*Beweis des Drei-Quadrate-Satzes.* Nach Lemma 5.10 genügt es zu zeigen, dass jede natürliche Zahl $n$, die nicht die Form $4^k(8l + 7)$ hat, Summe von drei rationalen Quadraten ist. Dazu betrachten wir die nicht-ausgeartete quadratische Form

$$Q(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 - nx_4^2 \,.$$

Die Gleichung $Q = 0$ hat sicher reelle Lösungen und nach Lemma 8.2 auch Lösungen in allen $\mathbb{Q}_p$ für $p \neq 2$. Es bleibt zu zeigen, dass es auch eine 2-adische Lösung gibt. Dazu können wir annehmen, dass $n$ nicht durch 4 teilbar ist (denn die Formen mit $n$ und mit $4n$ sind äquivalent). Dann gibt es (für $n \not\equiv 7 \bmod 8$) stets eine Darstellung

$$n \equiv x_1^2 + x_2^2 + x_3^2 \bmod 8 \,,$$

in der wenigstens ein $x_j$ ungerade ist. Wiederum nach Lemma 7.15 führt das zu einer nichttrivialen Lösung in $\mathbb{Z}_2$.

Aus Satz 8.12 folgt jetzt, dass es eine nichttriviale rationale Lösung von $Q = 0$ gibt. Darin muss $x_4 \neq 0$ sein; wir können also durch $x_4^2$ teilen und erhalten eine Darstellung von $n$ als Summe dreier rationaler Quadrate. ❑

Beachte, dass in Satz 8.12 im Unterschied zu Satz 8.9 tatsächlich die Existenz von Lösungen in $\mathbb{Q}_v$ für *alle* $v$ verlangt werden muss! Zum Beispiel hat die Form $x_1^2 + x_2^2 + x_3^2 - 7x_4^2$ nichttriviale Nullstellen in $\mathbb{R}$ und in allen $\mathbb{Q}_p$ für $p$ ungerade, aber nicht in $\mathbb{Q}_2$.

Die Aussage von Satz 8.12 gilt jetzt ganz allgemein für nicht-ausgeartete quadratische Formen in beliebig vielen Variablen.

**SATZ**
Hasse-Prinzip für quadrat. Formen

8.13. **Satz.** *Sei $Q(x_1, \ldots, x_n)$ eine nicht-ausgeartete quadratische Form in $n$ Variablen. Dann sind äquivalent:*

(1) *$Q = 0$ hat eine primitive ganzzahlige Lösung.*

(2) *$Q = 0$ hat eine nichttriviale Lösung in $\mathbb{Q}_v$ für alle Stellen $v$ ($v = p$ Primzahl oder $v = \infty$).*

*Beweis.* Wir können wieder annehmen, dass $Q(x_1, \dots, x_n) = a_1x_1^2 + \dots + a_nx_n^2$ diagonal ist mit ganzzahligen Koeffizienten $a_j$.

Der Fall $n = 1$ ist trivial ($a_1x_1^2 = 0$ hat niemals eine nichttriviale Lösung, weder in $\mathbb{Q}$ noch in $\mathbb{Q}_v$).

Im Fall $n = 2$ ist $Q(x_1, x_2) = a_1x_1^2 + a_2x_2^2$, und die Existenz einer nichttrivialen Lösung ist äquivalent dazu, dass $-a_2/a_1$ ein Quadrat ist (in $\mathbb{Q}$ bzw. in $\mathbb{Q}_v$). Es ist aber leicht zu sehen, dass $a \in \mathbb{Q}^\times$ genau dann ein Quadrat in $\mathbb{Q}$ ist, wenn $a$ ein Quadrat in $\mathbb{Q}_p$ ist für alle bis auf endlich viele Primzahlen $p$ (Übungsaufgabe).

Der Fall $n = 3$ wurde in Satz 8.9 erledigt, und der Fall $n = 4$ in Satz 8.12.

Den Fall $n \geq 5$ beweisen wir durch Induktion mit einem ähnlichen Argument wie Satz 8.12. Wir nehmen an, dass $Q = 0$ nichttriviale Lösungen in allen $\mathbb{Q}_v$ hat. Da $Q = 0$ insbesondere nichttriviale reelle Lösungen hat, können wir (möglicherweise nach Umordnen der Variablen) annehmen, dass $a_1 > 0$ und $a_n < 0$ ist. Sei wieder $D = 2a_1 \dots a_n$. Für jede Primzahl $p \mid D$ gibt es dann $u_p \in \mathbb{Q}_p$ und $(x_1, \dots, x_n) \in \mathbb{Q}_p^n \setminus \{0\}$ mit

$$u_p = a_1x_1^2 + \dots + a_{n-2}x_{n-2}^2 = -a_{n-1}x_{n-1}^2 - a_nx_n^2 \,.$$

Wie im Beweis von Satz 8.12 können wir annehmen, dass $u_p \in \mathbb{Z}_p^\times \cup p\mathbb{Z}_p^\times$ ist. Sei $d$ das Produkt der Primzahlen $p$ mit $u_p \notin \mathbb{Z}_p^\times$.

Nach dem Chinesischen Restsatz gibt es wieder $u \in \mathbb{Z}$ mit $u \equiv u_p \bmod p^2$ für alle ungeraden $p \mid D$ und $u \equiv u_2 \bmod 16$; es ist $u = du'$ mit $\mathrm{ggT}(u', D) = 1$. Nach Satz 8.10 gibt es eine Primzahl $q$ mit $q \equiv u' \bmod 4D^2$. Wir betrachten die nicht-ausgearteten quadratischen Formen

$$Q_1(x_1, \dots, x_{n-2}, y) = a_1x^2 + \dots + a_{n-2}x_{n-2}^2 - dqy^2 \quad \text{und}$$
$$Q_2(x_{n-1}, x_n, z) = a_{n-1}x_{n-1}^2 + a_nx_n^2 + dqz^2 \,.$$

Wie vorher sieht man, dass $Q_1$ und $Q_2$ jeweils nichttriviale Nullstellen in allen $\mathbb{Q}_v$ mit $v \neq q$ haben. Da $Q_2$ eine ternäre Form ist, folgt dann nach Satz 8.9 bereits, dass $Q_2 = 0$ eine nichttriviale rationale Lösung hat. Auf der anderen Seite sind die (wegen $n \geq 5$) mindestens drei Koeffizienten $a_1, \dots, a_{n-2}$ nicht durch die ungerade Primzahl $q$ teilbar; daher hat $Q_1 = 0$ nach Lemma 8.2 auch nichttriviale Lösungen in $\mathbb{Q}_q$. Damit gibt es nichttriviale Lösungen von $Q_1 = 0$ in *allen* $\mathbb{Q}_v$. Nach Induktionsvoraussetzung ($Q_1$ ist eine Form in $n - 1$ Variablen) hat $Q_1 = 0$ dann auch nichttriviale rationale Lösungen.

Nach Lemma 8.11 gibt es auch Lösungen mit $y = z = 1$. Das bedeutet, dass $dq$ sowohl in der Form $a_1x_1^2 + \dots + a_{n-2}x_{n-2}^2$ als auch in der Form $-a_{n-1}x_{n-1}^2 - a_nx_n^2$ (mit $x_1, \dots, x_n \in \mathbb{Q}$) geschrieben werden kann. Damit ist $(x_1, \dots, x_n)$ eine Lösung von $Q = 0$, und nichttrivial, weil $dq \neq 0$ ist. ❑

8.14. **Definition.** Eine nicht-ausgeartete quadratische Form $Q$ heißt *indefinit*, wenn $Q = 0$ eine nichttriviale reelle Nullstelle hat. ♢

**DEF** indefinite qu. Form

8.15. **Folgerung.** *Sei $Q$ eine indefinite nicht-ausgeartete quadratische Form in $n \geq 5$ Variablen. Dann hat $Q = 0$ eine primitive ganzzahlige Lösung.*

**FOLG** Nullstellen von indefiniten qu. Formen

*Beweis.* Ist $p$ eine Primzahl, dann hat $Q = 0$ stets nichttriviale Lösungen in $\mathbb{Q}_p$ (Übungsaufgabe). Nach Voraussetzung hat $Q = 0$ auch eine nichttriviale Lösung in $\mathbb{R}$. Insgesamt sind also die Voraussetzungen von Satz 8.13 erfüllt, sodass $Q = 0$ eine primitive ganzzahlige Lösung haben muss. ❑

Damit können wir den Vier-Quadrate-Satz 5.7 noch einmal beweisen. Sei $n \geq 1$. Wir betrachten

$$Q(x_1, x_2, x_3, x_4, x_5) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - nx_5^2 .$$

Dann ist $Q$ offensichtlich indefinit, also gibt es, wie wir eben gesehen haben, nichttriviale rationale Lösungen von $Q = 0$. Nach Lemma 8.11 gibt es dann auch eine Lösung mit $x_5 = 1$, d.h., $n$ ist jedenfalls Summe von vier *rationalen* Quadraten.

Wir müssen noch zeigen, dass daraus folgt, dass $n$ auch Summe von vier *ganzzahligen* Quadraten ist. Dazu gehen wir wie im Beweis von Lemma 5.10 vor: Sei $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ eine rationale Lösung von $|\boldsymbol{x}|^2 = n$ mit Nenner $c$. Wir nehmen für den Augenblick einmal an, dass wir nicht den Fall $c = 2$ mit $2x_1, \ldots, 2x_4$ ungerade haben. Dann gibt es $\boldsymbol{y} \in \mathbb{Z}^4$ mit $|\boldsymbol{y} - \boldsymbol{x}| < 1$, und man kann wie in Lemma 5.10 schließen, dass

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{2\langle \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{y}\rangle}{|\boldsymbol{x} - \boldsymbol{y}|^2} (\boldsymbol{y} - \boldsymbol{x})$$

ebenfalls eine Lösung ist und kleineren Nenner hat.

Im verbleibenden Fall $\boldsymbol{x} = (m_1/2, m_2/2, m_3/2, m_4/2)$ mit $m_1, m_2, m_3, m_4$ ungerade schreiben wir $\mu = m_1 + m_2 i + m_3 j + m_4 k$ als Quaternion mit ganzen Koeffizienten. Wir wählen $s_1, s_2, s_3, s_4 = \pm 1$ mit $s_1 \equiv m_1 \bmod 4$, $s_2 \equiv -m_2 \bmod 4$, $s_3 \equiv -m_3 \bmod 4$, $s_4 \equiv -m_4 \bmod 4$ und setzen $\sigma = s_1 + s_2 i + s_3 j + s_4 k$. Dann ist

$$\mu\sigma \equiv \mu\bar{\mu} = N(\mu) = m_1^2 + m_2^2 + m_3^2 + m_4^2 \equiv 1 + 1 + 1 + 1 \equiv 0 \bmod 4 ,$$

also ist $\mu\sigma/4$ eine Quaternion mit ganzen Koeffizienten. Außerdem gilt

$$N(\mu\sigma/4) = N(\mu/2)N(\sigma)/4 = n \cdot 4/4 = n .$$

Damit haben wir auch in diesem Fall eine ganzzahlige Lösung gefunden.

Man kann natürlich auch den Zwei-Quadrate-Satz 5.3 mit Hilfe des Satzes 8.9 beweisen, wobei man am einfachsten die Bedingung der 2-adischen Lösbarkeit weglässt (Übungsaufgabe).

Wir zeigen noch, dass Folgerung 8.15 „scharf“ ist.

**LEMMA**
unlösbare qu. Formen in 4 Variablen

8.16. **Lemma.** *Für jede Primzahl $p$ gibt es nicht-ausgeartete quadratische Formen in vier Variablen, die keine nichttriviale $p$-adische Nullstelle haben.*

*Beweis.* Für $p$ ungerade sei $a$ ein quadratischer Nichtrest mod $p$; dann hat

$$Q(x_1, x_2, x_3, x_4) = x_1^2 - ax_2^2 + px_3^2 - apx_4^2$$

keine nichttriviale $p$-adische Nullstelle. Für $p = 2$ kann man

$$Q(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

nehmen. ❑

Die Produktformel 8.8 ist tatsächlich die einzige Relation zwischen den verschiedenen Normrestsymbolen.

**SATZ**
Keine weitere Relation zw. Normrest-symbolen

8.17. **Satz.** *Seien $\varepsilon_v = \pm 1$ für alle Stellen $v$ ($v = p$ oder $v = \infty$) gegeben mit $\varepsilon_v = 1$ für alle bis auf endlich viele $v$ und $\prod_v \varepsilon_v = 1$. Dann gibt es $a, b \in \mathbb{Q}^\times$ mit $\left(\frac{a,b}{v}\right) = \varepsilon_v$ für alle $v$.*

*Beweis.* Seien $p_1, \ldots, p_k$ die (endlich vielen) ungeraden Primzahlen mit $\varepsilon_{p_j} = -1$. Für jedes $1 \leq j \leq k$ sei $d_j$ ein quadratischer Nichtrest mod $p_j$. Nach dem Chinesischen Restsatz und dem Satz von Dirichlet 8.10 gibt es eine Primzahl $q$ mit $q \equiv \varepsilon_\infty d_j \bmod p_j$ für alle $1 \leq j \leq k$ und

$$q \equiv \left\{ \begin{array}{ll} 1 & \text{falls } \varepsilon_\infty = \phantom{-}1 \text{ und } \varepsilon_2 = \phantom{-}1 \\ 3 & \text{falls } \varepsilon_\infty = -1 \text{ und } \varepsilon_2 = -1 \\ 5 & \text{falls } \varepsilon_\infty = \phantom{-}1 \text{ und } \varepsilon_2 = -1 \\ 7 & \text{falls } \varepsilon_\infty = -1 \text{ und } \varepsilon_2 = \phantom{-}1 \end{array} \right\} \bmod 8 .$$

Sei $a = \varepsilon_\infty q$ und $b = -2p_1 \cdots p_k$. Dann ist jedenfalls $\left(\frac{a,b}{\infty}\right) = \varepsilon_\infty$; siehe Lemma 8.4. Aus $\varepsilon_\infty q \equiv 1 \bmod 4$ folgt nach Lemma 8.7, dass $\left(\frac{a,b}{2}\right) = (-1)^{(q^2-1)/8} = \varepsilon_2$ ist. Für alle $1 \le j \le k$ ist $\varepsilon_\infty q$ ein quadratischer Nichtrest mod $p_j$, also ist $\left(\frac{a,b}{p_j}\right) = -1 = \varepsilon_{p_j}$ nach Lemma 8.6. Für alle Primzahlen $p \notin \{2, p_1, \ldots, p_k, q\}$ sind $a$ und $b$ nicht durch $p$ teilbar, also ist $\left(\frac{a,b}{p}\right) = 1 = \varepsilon_p$; siehe ebenfalls Lemma 8.6. Aus der Produktformel 8.8 folgt schließlich, dass auch $\left(\frac{a,b}{q}\right) = \prod_{v \ne q} \varepsilon_v = \varepsilon_q = 1$ ist. ❑

Man kann auch zeigen, dass die Werte des Normrestsymbols ternäre quadratische Formen bis auf Äquivalenz klassifizieren. Genauer gilt:

*Zwei nicht-ausgeartete ternäre quadratische Formen $Q$ und $Q'$ sind genau dann äquivalent, wenn für jedes $v$ die Gleichungen $Q = 0$ und $Q' = 0$ entweder beide nichttriviale Lösungen oder beide keine nichttrivialen Lösungen in $\mathbb{Q}_v$ haben.*

Zum Abschluss dieses Abschnitts folgt hier noch eine Anwendung der Produktformel für das Normrestsymbol auf eine konkrete diophantische Gleichung.

**SATZ**
Anwendung der Produktformel

8.18. **Satz.** *Ist $n \in \mathbb{Z}_{>0}$ ungerade, dann hat die Gleichung*

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} = n$$

*keine Lösung in positiven ganzen (oder rationalen) Zahlen $a, b, c$.*

Für *gerade* ganze Zahlen $n > 0$ kann es positive ganzzahlige Lösungen geben. Für $n = 4$ ist zum Beispiel die kleinste positive Lösung (bis auf Permutation) gegeben durch

$a = 437361267792869725786125260237139015281653755816161361862143799337842346777 2036$

$b = 368751317941299998271978115652254748254929799689719709962831374716372246340 55579$

$c = 154476802108746166441951315019919837485664325669565431700026634898253202035277999$

Für ungerade ganze Zahlen $n > 0$ kann es (dann aber nicht positive) ganzzahlige Lösungen geben. Das geschieht zum ersten Mal für $n = 19$, mit z.B.

$$a = 53753\,, \quad b = -55862\,, \quad c = 57017\,.$$

*Beweis.* Es gelte die Gleichung im Satz. Mit

$$x = 4(n+3)\frac{a+b+2c}{c-(n+2)(a+b)} \quad \text{und} \quad y = 4(n+3)(2n+5)\frac{a-b}{c-(n+2)(a+b)}$$

gilt dann

$$y^2 = x(x^2 + Ax + B) \qquad \text{mit } A = 4n(n+3) - 3 \text{ und } B = 32(n+3). \tag{8.1}$$

Sind $a, b, c > 0$, dann ist $c < (n+2)(a+b)$ (sonst wäre der Term $c/(a+b)$ schon größer als $n$). Es folgt, dass positive Lösungen $(a, b, c)$ auf rationale Lösungen $(x, y)$ von (8.1) mit $x < 0$ abgebildet werden. Wir werden zeigen, dass es solche Lösungen nicht geben kann, wenn $n > 0$ ungerade ist.

Sei also jetzt $n \ge 1$ und ungerade. Sei $D = 2n + 5$, dann ist $D$ positiv, ungerade, teilerfremd zu $B = 32(n+3)$ und teilt die Diskriminante $A^2 - 4B = (2n-3)(2n+5)^3$ des quadratischen Faktors in (8.1). Sei $(\xi, \eta)$ eine Lösung von (8.1) mit $\xi < 0$. Wir wollen einen Widerspruch herleiten. Dazu zeigen wir

$$\left(\frac{\xi, -D}{p}\right) = 1 \qquad \text{für alle Primzahlen } p.$$

Aus der Produktformel 8.8 für das Normrestsymbol folgt dann, dass auch

$$\left(\frac{\xi,-D}{\infty}\right)=1 \quad \text{sein muss, aber es ist} \quad \left(\frac{\xi,-D}{\infty}\right)=-1\,,$$

da sowohl $\xi$ als auch $-D$ negativ sind (Lemma 8.4). Das ergibt dann den gewünschten Widerspruch.

Sei also $p$ eine Primzahl. Da das Produkt $\xi(\xi^2+A\xi+B)$ ein Quadrat (nämlich $\eta^2$) ist, folgt

$$\left(\frac{\xi,-D}{p}\right)=\left(\frac{\xi^2+A\xi+B,-D}{p}\right) \qquad \text{für alle } p.$$

(Die Diskriminante $A^2-4B=((2n+1)^2-16)(2n+5)^2$ ist nur für $n=2$ ein Quadrat, aber niemals für ungerades $n>0$. Deswegen ist stets $\xi^2+A\xi+B\neq 0$.) Wenn $p$ ungerade und $v_p(\xi)<0$ ist, dann ist $1+A\xi^{-1}+B\xi^{-2}\equiv 1 \bmod p$ und damit ein Quadrat $s^2$ in $\mathbb{Q}_p$, womit auch $\xi=(\eta/\xi s)^2$ ein Quadrat in $\mathbb{Q}_p$ sein muss. In diesem Fall ist $\left(\frac{\xi,-D}{p}\right)=1$. Wir können also $\xi\in\mathbb{Z}_p$ annehmen. Wir unterscheiden die folgenden Fälle.

(1) $p$ ungerade mit $p\nmid BD$.
Falls $\xi\in\mathbb{Z}_p^\times$ ist, dann ist $\left(\frac{\xi,-D}{p}\right)=1$ nach Lemma 8.6, da sowohl $\xi$ als auch $-D$ $p$-adische Einheiten sind. Im verbleibenden Fall $\xi\in p\mathbb{Z}_p$ ist $\xi^2+A\xi+B\equiv B \bmod p$, also eine $p$-adische Einheit, und es folgt ebenfalls $\left(\frac{\xi,-D}{p}\right)=\left(\frac{\xi^2+A\xi+B,-D}{p}\right)=1$.

(2) $2\neq p\mid B$.
Dann ist $n\equiv -3 \bmod p$ und damit $-D\equiv 1 \bmod p$. Es folgt, dass $-D$ ein Quadrat in $\mathbb{Q}_p$ ist; damit ist $\left(\frac{\xi,-D}{p}\right)=1$.

(3) $p\mid D$ (dann ist $p\neq 2$, da $D$ ungerade ist).
Dann ist $A\equiv -8 \bmod p$ und $B\equiv 16 \bmod p$, also ist

$$\xi^2+A\xi+B\equiv(\xi-4)^2 \bmod p\,.$$

Im Fall $\xi\equiv 4 \bmod p$ ist $\xi$ ein Quadrat in $\mathbb{Q}_p$, und es folgt $\left(\frac{\xi,-D}{p}\right)=1$. Anderenfalls ist der quadratische Faktor ein Quadrat in $\mathbb{Q}_p$, und es folgt ebenfalls $\left(\frac{\xi,-D}{p}\right)=\left(\frac{\xi^2+A\xi+B,-D}{p}\right)=1$.

(4) $p=2$ und $n\equiv 1 \bmod 4$.
Dann ist $-D\equiv 1 \bmod 8$, also ist $\left(\frac{\xi,-D}{2}\right)=1$ für alle $\xi$ nach Lemma 8.7.

(5) $p=2$ und $n\equiv 3 \bmod 4$.
Dann ist $-D\equiv 5 \bmod 8$, also nach Lemma 8.7 $\left(\frac{\xi,-D}{2}\right)=(-1)^{v_2(\xi)}$. Wir müssen also zeigen, dass $v_2(\xi)$ gerade ist. Wegen $n+3\equiv 2 \bmod 4$ ist hier $v_2(B)=6$, und es ist $A\equiv -3 \bmod 8$. Wir nehmen an, dass $v_2(\xi)$ ungerade ist und leiten einen Widerspruch her. Die 2-adischen Bewertungen von $\xi^3$, $A\xi^2$ und $B\xi$ sind $3v_2(\xi)$, $2v_2(\xi)$ und $6+v_2(\xi)$. Da $v_2(\xi)$ ungerade ist, sind die ersten beiden und die letzten beiden verschieden. Ist $3v_2(\xi)=6+v_2(\xi)$, dann ist $v_2(\xi)=3$ und damit $2v_2(\xi)<3v_2(\xi)=6+v_2(\xi)$. Die kleinste Bewertung der drei Terme wird also genau einmal angenommen; diese Bewertung muss gerade sein, denn die Summe der Terme ist ein Quadrat. Damit muss der mittlere Term $A\xi^2$ die kleinste Bewertung haben. Es kommen dann nur noch $\nu:=v_2(\xi)=1,3,5$ infrage. Sei $\xi=2^\nu\xi_0$ mit $\xi_0\in\mathbb{Z}_2^\times$. Dann ist

$$\xi(\xi^2+A\xi+B)=4^\nu(2^\nu\xi_0^3+A\xi_0^2+2^{5-\nu}(n+3)\xi_0)=4^\nu u\,.$$

Im Fall $\nu = 1$ ist $u \equiv 2\xi_0^3 - 3\xi_0^2 \equiv 2 + 1 = 3 \bmod 4$.
Im Fall $\nu = 3$ ist $u \equiv -3\xi_0^2 + 4(n+3)\xi_0 \equiv 5 \bmod 8$.
Im Fall $\nu = 5$ ist $u \equiv -3\xi_0^2 + (n+3)\xi_0 \equiv 1 + 2 = 3 \bmod 4$.
In allen Fällen ist also $u \in \mathbb{Z}_2^\times$ mit $u \not\equiv 1 \bmod 8$; damit ist $4^\nu u$ jeweils kein Quadrat in $\mathbb{Q}_2$, was den gewünschten Widerspruch liefert. ❑

Da die Produktformel für das Normrestsymbol, die im Beweis wesentlich benutzt wird, äquivalent zum Quadratischen Reziprozitätsgesetz und seinen Ergänzungsgesetzen ist, kann man einen Beweis alternativ auch darauf stützen.[10]

[10] A. Bremner, A. Macleod: *An unusual cubic representation problem*, Ann. Math. Inform. **43** (2014), 29–41.

## 9. Die Pellsche Gleichung

Sei $d > 0$ eine ganze Zahl, die kein Quadrat ist. Die Gleichung

$$x^2 - dy^2 = 1 \quad \text{oder auch} \quad x^2 - dy^2 = \pm 1\,,$$

die in ganzen Zahlen $x$ und $y$ zu lösen ist, wird *Pellsche Gleichung* genannt. Diese Bezeichnung geht auf Euler zurück, der offenbar irrtümlich annahm, dass Pell (ein englischer Mathematiker) an der Entwicklung eines Lösungsverfahrens beteiligt war. Tatsächlich hatte Brouncker auf eine Herausforderung von Fermat hin ein solches Verfahren entwickelt, das später unter anderem von Wallis in einem Buch, das mit Pells Hilfe entstand, wiedergegeben wurde. Es waren aber schon indische Mathematiker im 11. oder 12. Jahrhundert im Besitz eines äquivalenten Verfahrens. Lagrange gab dann eine präzise Formulierung der Methode und bewies, dass sie immer zum Ziel führt.

**DEF** Pellsche Gleichung

Die Voraussetzung „$d > 0$ und kein Quadrat“ schließt uninteressante oder triviale Fälle aus. Die *rationalen* Lösungen lassen sich für jedes $d$, ausgehend von der offensichtlichen Lösung $(x, y) = (1, 0)$, leicht parametrisieren, vergleiche Satz 6.3.

9.1. **Beispiele.** Wir betrachten ein paar Beispiele. Für $d = 2$ erhalten wir folgende Lösungen von $x^2 - 2y^2 = 1$ mit $x, y \geq 0$.

**BSP** Lösungen der Pellschen Gleichung

| $x$ | 1 | 3 | 17 | 99 | 577 | 3363 | 19601 |
|---|---|---|---|---|---|---|---|
| $y$ | 0 | 2 | 12 | 70 | 408 | 2378 | 13860 |

Für $d = 3$ finden wir:

| $x$ | 1 | 2 | 7 | 26 | 97 | 362 | 1351 |
|---|---|---|---|---|---|---|---|
| $y$ | 0 | 1 | 4 | 15 | 56 | 209 | 780 |

Und für $d = 409$ sind die beiden kleinsten Lösungen (die wir schon in der Einleitung gesehen haben):

| $x$ | 1 | 2505297727309242798604 9 |
|---|---|---|
| $y$ | 0 | 123878999864721858216 0 |

In den betrachteten Fällen gibt es also immer nichttriviale Lösungen (als *trivial* wollen wir hier die Lösungen mit $y = 0$ betrachten). Mindestens für $d = 2$ und $d = 3$ scheint es eine Folge regelmäßig wachsender Lösungen zu geben. Wir sehen aber auch, dass die kleinste nichttriviale Lösung im Vergleich zu $d$ recht groß sein kann. ♣

Wir wollen zunächst die Struktur der Lösungsmenge studieren. Analog zu der Beobachtung

$$x^2 + y^2 = (x + y\boldsymbol{i})(x - y\boldsymbol{i})\,,$$

die beim Studium der Summen zweier Quadrate wichtig war, gilt hier

$$x^2 - dy^2 = (x + y\sqrt{d})(x - y\sqrt{d})\,.$$

Wenn wir die zwei Ausdrücke $x + y\sqrt{d}$ und $u + v\sqrt{d}$ multiplizieren, erhalten wir die Relation

$$(x^2 - dy^2)(u^2 - dv^2) = (xu + dyv)^2 - d(xv + yu)^2\,.$$

Lemma 5.1 ist der Spezialfall $d = -1$ dieser Beziehung.

Analog zum Ring $\mathbb{Z}[\boldsymbol{i}]$ der ganzen Gaußschen Zahlen können wir hier den Ring $R = \mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}$ betrachten. Die Abbildung $N\colon R \to \mathbb{Z}$,

$a + b\sqrt{d} \mapsto a^2 - db^2$ ist dann gerade die *Norm*; wir haben schon gesehen, dass die Norm multiplikativ ist (siehe Seite 61). Wir sehen auch, dass $a + b\sqrt{d}$ in $R$ invertierbar ist, falls $N(a + b\sqrt{d}) = \pm 1$ ist. (Das Inverse ist dann $\pm(a - b\sqrt{d})$). Umgekehrt muss die Norm einer Einheit ein Teiler von 1 sein, also ist

$$R^\times = \{a + b\sqrt{d} \mid a^2 - db^2 = \pm 1\}\,.$$

Die Einheiten mit Norm 1 bilden eine Untergruppe

$$R_+^\times = \{a + b\sqrt{d} \mid a^2 - db^2 = 1\}\,.$$

Es ist entweder $R_+^\times = R^\times$ (zum Beispiel für $d \equiv 3 \bmod 4$ ist $a^2 - db^2$ niemals $\equiv -1 \bmod 4$), oder $R_+^\times$ hat Index 2 in $R^\times$ (die nichttriviale Nebenklasse wird dann von den Einheiten mit Norm $-1$ gebildet).

Wir geben den Lösungsmengen der beiden Versionen der Pellschen Gleichung eine Bezeichnung.

**DEF**
$S_d$, $T_d$

9.2. **Definition.** Für $d \in \mathbb{Z}_{>0}$, $d$ kein Quadrat, setzen wir

$$S_d = \{(x, y) \in \mathbb{Z}^2 \mid x^2 - dy^2 = 1\} \quad \text{und} \quad T_d = \{(x, y) \in \mathbb{Z}^2 \mid x^2 - dy^2 = \pm 1\}\,. \ \diamondsuit$$

Obige Überlegung zeigt dann die folgende Aussage:

**LEMMA**
Lösungsmenge hat Gruppenstruktur

9.3. **Lemma.** *Die Mengen $S_d$ und $T_d$ haben eine natürliche Struktur als abelsche Gruppen. Dabei ist die Verknüpfung gegeben durch*

$$(x, y) * (x', y') = (xx' + dyy', xy' + yx')\,.$$

*Die Abbildung $\mathbb{Z}^2 \ni (x, y) \mapsto x + y\sqrt{d} \in R = \mathbb{Z}[\sqrt{d}]$ liefert Isomorphismen $S_d \cong R_+^\times$ und $T_d \cong R^\times$.*

Wir wollen jetzt die Struktur dieser Gruppen ermitteln. Dazu betrachten wir

$$\phi\colon S_d \longrightarrow \mathbb{R}^\times\,, \quad (x, y) \longmapsto x + y\sqrt{d}\,.$$

Hier ist $\sqrt{d} \in \mathbb{R}$ die positive reelle Quadratwurzel (d.h., wir verwenden die offensichtliche Einbettung von $R$ in $\mathbb{R}$). Es ist $1/\phi(x, y) = x - y\sqrt{d}$, also haben wir

$$x = \frac{\phi(x, y) + 1/\phi(x, y)}{2} \quad \text{und} \quad y = \frac{\phi(x, y) - 1/\phi(x, y)}{2\sqrt{d}}\,.$$

Das zeigt, dass $\phi$ injektiv ist. Außerdem gilt

$$\phi(x, y) > 0 \iff x > 0 \quad \text{und} \quad \phi(x, y) > 1 \iff x, y > 0\,.$$

Nach dem oben Gesagten ist $\phi$ ein Gruppenhomomorphismus. Weil $(-1, 0) \in S_d$ ist, ist der Homomorphismus $S_d \to \{\pm 1\}$, $(x, y) \mapsto \operatorname{sign}(\phi(x, y))$ surjektiv; sein Kern

$$S_d^+ = \{(x, y) \in S_d \mid x > 0\}$$

ist also eine Untergruppe von $S_d$ vom Index 2. Es gilt $(-1, 0) * (x, y) = (-x, -y)$.

**LEMMA** $S_d^+ \neq \{(1,0)\}$ $\Rightarrow S_d^+ \cong \mathbb{Z}$

9.4. **Lemma.** *Wir nehmen an, dass $S_d^+$ nicht trivial ist (d.h., es gibt Lösungen $(x,y)$ von $x^2-dy^2=1$ mit $x>1$). Sei $(x_1,y_1)$ die Lösung mit $x_1>1$ minimal und $y_1>0$. Dann wird die Gruppe $S_d^+$ von $(x_1,y_1)$ erzeugt und ist unendlich.*

*Beweis.* Sei $\alpha = \phi(x_1,y_1)$; es ist $\alpha>1$, und es gibt dann kein $(x,y)\in S_d^+$ mit $1<\phi(x,y)<\alpha$ (denn sonst wäre $y>0$ und $1<x<x_1$).

Sei jetzt $(x,y)\in S_d^+$ beliebig; setze $\beta=\phi(x,y)>0$. Da $\alpha>1$ ist, gibt es ein $n\in\mathbb{Z}$ mit $\alpha^n\le\beta<\alpha^{n+1}$. Wenn wir das $m$-fache Produkt von $(x,y)$ in der Gruppe $S_d^+$ mit $(x,y)^{*m}$ bezeichnen, dann haben wir

$$1\le \beta\alpha^{-n} = \phi\big((x,y)*(x_1,y_1)^{*(-n)}\big) < \alpha\,.$$

Dann muss $\phi\big((x,y)*(x_1,y_1)^{*(-n)}\big)=1$ sein, also

$$(x,y)*(x_1,y_1)^{*(-n)}=(1,0) \quad\Longrightarrow\quad (x,y)=(x_1,y_1)^{*n}\,.$$

Damit ist gezeigt, dass jedes Element $(x,y)\in S_d^+$ eine Potenz von $(x_1,y_1)$ ist. Da $\phi(S_d^+)=\{\alpha^n\mid n\in\mathbb{Z}\}$ unendlich ist, ist auch $S_d^+$ unendlich. ❑

Hinter dem Beweis steht folgende Beobachtung: $\log\circ\phi$ liefert einen Homomorphismus von $S_d^+$ in die additive Gruppe $\mathbb{R}$ mit diskretem Bild. Eine nichttriviale diskrete Untergruppe $\Gamma$ von $\mathbb{R}$ hat aber stets die Form $\mathbb{Z}\cdot r$ für ein $r>0$; genauer: $r=\min(\Gamma\cap\mathbb{R}_{>0})$ (hier ist $r=\log\alpha$).

Wir sehen also, dass es, sobald es *eine* nichttriviale Lösung gibt, schon unendlich viele Lösungen geben muss, die (bis aufs Vorzeichen) alle die Form $(x_n,y_n)$ haben mit $n\in\mathbb{Z}$, wobei gilt

$$x_n+y_n\sqrt{d}=(x_1+y_1\sqrt{d})^n\,.$$

Da wir für großes $n$ die folgenden Näherungen haben:

$$x_n=\frac{\alpha^n+\alpha^{-n}}{2}\approx\frac{\alpha^n}{2}\quad\text{und}\quad y_n=\frac{\alpha^n-\alpha^{-n}}{2\sqrt{d}}\approx\frac{\alpha^n}{2\sqrt{d}}\,,$$

sehen wir, dass die Lösungen exponentiell wachsen, wie wir das in den ersten beiden Beispielen auch beobachtet haben.

Es bleibt noch zu zeigen, dass es tatsächlich immer nichttriviale Lösungen geben muss.

**LEMMA** Approximation von $\sqrt{d}$

9.5. **Lemma.** *Seien $x,y\in\mathbb{Z}_{>0}$. Dann gilt*

$$x^2-dy^2=1 \iff 0<\frac{x}{y}-\sqrt{d}<\frac{1}{2\sqrt{d}\,y^2}\,.$$

Eine Lösung von $x^2-dy^2=1$ mit $x,y>0$ liefert also eine sehr gute rationale Näherung von $\sqrt{d}$ und umgekehrt.

*Beweis.*
„⇒“: Aus $x^2-dy^2=1$ und $x,y>0$ folgt $\frac{x}{y}>\sqrt{d}$ und damit

$$0<\frac{x}{y}-\sqrt{d}=\frac{x-y\sqrt{d}}{y}=\frac{x^2-dy^2}{y(x+y\sqrt{d})}=\frac{1}{y^2(x/y+\sqrt{d})}<\frac{1}{2\sqrt{d}\,y^2}\,.$$

„⇐“: Aus $0<\frac{x}{y}-\sqrt{d}<1/(2\sqrt{d}\,y^2)$ folgt

$$0<x^2-dy^2=y^2\Big(\frac{x}{y}-\sqrt{d}\Big)\Big(\frac{x}{y}+\sqrt{d}\Big)<\frac{1}{2\sqrt{d}}\Big(2\sqrt{d}+\frac{1}{2\sqrt{d}y^2}\Big)=1+\frac{1}{4dy^2}<2\,.$$

Da $x^2-dy^2\in\mathbb{Z}$ ist, muss $x^2-dy^2=1$ sein. ❑

Wir müssen also zeigen, dass es stets so gute Näherungen von $\sqrt{d}$ wie in Lemma 9.5 gibt. Als ersten Schritt beweisen wir folgendes Ergebnis, das immerhin bis auf einen konstanten Faktor an unser Ziel herankommt. Dieser Satz geht auf Dirichlet zurück.

**LEMMA**
Diophantische Approximation

9.6. **Lemma.** *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ eine irrationale reelle Zahl. Dann gibt es unendlich viele rationale Zahlen $p/q$ mit*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}\,.$$

*Beweis.* Wir bezeichnen mit $\langle x\rangle = x - \lfloor x\rfloor$ den gebrochenen Anteil von $x \in \mathbb{R}$. Der Beweis benutzt das „Schubfachprinzip“. Sei $n \geq 1$. Von den $n+1$ Zahlen

$$0, \langle\alpha\rangle, \langle 2\alpha\rangle, \ldots, \langle n\alpha\rangle$$

im halb-offenen Intervall $[0,1[$ muss es (wenigstens) zwei geben, die im selben Teilintervall $[k/n, (k+1)/n[$ zu liegen kommen, für ein $0 \leq k < n$. Es gibt dann also $0 \leq l < m \leq n$ mit

$$\frac{1}{n} > |\langle m\alpha\rangle - \langle l\alpha\rangle| = \big|(m-l)\alpha - (\lfloor m\alpha\rfloor - \lfloor l\alpha\rfloor)\big|\,.$$

Mit $p = \lfloor m\alpha\rfloor - \lfloor l\alpha\rfloor$ und $q = m - l$ heißt das

$$0 < \left|\alpha - \frac{p}{q}\right| < \frac{1}{nq} \leq \frac{1}{q^2}\,.$$

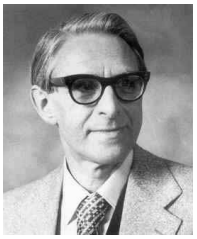
(Da $\alpha$ irrational ist, ist $\alpha \neq p/q$.) Wenn $p/q$ eine gegebene Näherung ist, können wir $n$ so groß wählen, dass $|\alpha - p/q| > 1/n$ ist; die neue Näherung $p'/q'$, die wir dann finden, erfüllt $|\alpha - p'/q'| < 1/nq' \leq 1/n$ und ist daher von $p/q$ verschieden. Indem wir also $n$ immer größer wählen, finden wir unendlich viele Näherungen mit der verlangten Eigenschaft. ❑

Zwei Bemerkungen:

(1) Die Eigenschaft, dass die Menge

$$\left\{\frac{p}{q} \in \mathbb{Q} \;\middle|\; 0 < \left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}\right\}$$

unendlich ist, charakterisiert die irrationalen Zahlen unter den reellen Zahlen $\alpha$. Denn ist $\alpha = r/s \in \mathbb{Q}$, dann ist $|\alpha - p/q|$ entweder gleich null oder mindestens $1/qs$, sodass $0 < |\alpha - p/q| < 1/q^2$ nur für $q < s$ möglich ist. Für jedes gegebene $q$ gibt es höchstens ein $p$ (oder zwei für $q = 1$), das die Ungleichung erfüllt. Wenn $q$ beschränkt ist, muss die Menge also endlich sein.



K.F. Roth
(1925–2015)
© unbekannt

(2) Wenn $\alpha \in \mathbb{R}$ *algebraisch* ist (also Nullstelle eines normierten Polynoms mit rationalen Koeffizienten), dann lässt sich $\alpha$ nicht wesentlich besser durch rationale Zahlen approximieren als in Lemma 9.6. Genauer gilt für jedes $\varepsilon > 0$, dass es nur endlich viele $p/q \in \mathbb{Q}$ gibt mit

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{2+\varepsilon}}\,.$$

Dieses Ergebnis ist nicht einfach zu beweisen. Es ist als *Satz von Roth* bekannt (oder auch als *Satz von Thue-Siegel-Roth*, denn Thue und Siegel bewiesen schwächere Versionen). Der Beweis ist allerdings nicht *effektiv*, d.h., er liefert keinen Algorithmus (egal wie ineffizient), der diese endliche Menge von Brüchen bestimmt.

Man kann zum Beispiel aus diesem Satz folgern, dass eine Gleichung

$$F(x,y) = m$$

mit $F \in \mathbb{Z}[x,y]$ homogen und irreduzibel vom Grad $\geq 3$ und $0 \neq m \in \mathbb{Z}$ nur endlich viele ganzzahlige Lösungen haben kann. Solche Gleichungen heißen *Thue-Gleichungen*, denn Thue benutzte sein oben erwähntes Resultat, um die Endlichkeit der Lösungsmenge zu beweisen.

A. Thue (1863–1922)

Der Beweis geht etwa so. Sei $F$ wie oben irreduzibel vom Grad $d \geq 3$. Wir setzen $f(x) = F(x,1) \in \mathbb{Z}[x]$; dann ist $f$ irreduzibel mit $\deg(f) = d$. Wir nehmen an, dass wir ein Resultat wie den Satz von Roth zur Verfügung haben mit Exponent von $q$ echt kleiner als $d$, wobei $d$ der Grad des Minimalpolynoms von $\alpha$ über $\mathbb{Q}$ ist. Daraus folgt, dass es für jedes vorgegebene $C > 0$ nur endlich viele $(p,q)$ gibt mit $|\alpha - \frac{p}{q}| \leq Cq^{-d}$. Seien $\alpha_1, \dots, \alpha_d \in \mathbb{C}$ die Nullstellen von $f$, d.h., $f = c(x-\alpha_1)\cdots(x-\alpha_d)$. Gilt $F(x,y) = m$ mit $y \neq 0$, dann folgt

$$|c| \prod_{j=1}^{d} \Big|\frac{x}{y} - \alpha_j\Big| = \frac{|m|}{|y|^d}\,.$$

Ist $|y|$ groß, dann ist die rechte Seite sehr klein, also muss ein Faktor links klein sein. Da die $\alpha_j$ paarweise verschieden sind (wegen der Irreduzibilität von $f$), kann nur ein Faktor beliebig klein werden. Wir setzen $\delta = \min\{|\alpha_j - \alpha_k| \mid 1 \leq j < k \leq d\}$. Sei $|y|$ groß genug, sodass wir ohne Einschränkung annehmen können, dass $|\frac{x}{y} - \alpha| \leq \delta/2$ ist für ein $\alpha_k =: \alpha$. Dann ist $|\frac{x}{y} - \alpha_j| \geq \delta/2$ für $j \neq k$, also

$$\Big|\frac{x}{y} - \alpha\Big| = \frac{1}{|c| \prod_{j\neq k} \left|\frac{x}{y} - \alpha_j\right|} \frac{|m|}{|y|^d} \leq \frac{2^{d-1}|m|}{|c|\delta^{d-1}|y|^d}\,.$$

Nach der Überlegung oben folgt dann, dass $|y|$ beschränkt ist. Für festes $y$ gibt es maximal $d$ Lösungen $x$ von $F(x,y) = m$. Aus einer Schranke für $y$ folgt also die Endlichkeit der Lösungsmenge.

Wir werden jetzt das Approximationslemma 9.6 dazu verwenden, die Existenz von nichttrivialen Lösungen der Pellschen Gleichung nachzuweisen.

**SATZ** Existenz nichttrivialer Lösungen

9.7. **Satz.** *Die Pellsche Gleichung $x^2 - dy^2 = 1$ hat (für $d > 0$ kein Quadrat) stets nichttriviale Lösungen. Die Lösungsmenge $S_d$ trägt eine natürliche Struktur als abelsche Gruppe; es gilt $S_d \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$.*

*Beweis.* Wir haben bereits gesehen (Lemma 9.4), dass $S_d^+ \cong \mathbb{Z}$ ist, sobald es nichttriviale Lösungen gibt. Dann ist $S_d$ das direkte Produkt der zweielementigen Gruppe $\{(1,0),(-1,0)\}$ und $S_d^+$. Es genügt also, die Existenz einer nichttrivialen Lösung zu zeigen.

**Behauptung:** *Es gibt unendlich viele Paare $(x,y) \in \mathbb{Z}^2$ mit $|x^2 - dy^2| < 2\sqrt{d}+1$.*

Zum Beweis beachten wir, dass es nach Lemma 9.6 unendlich viele Paare $(x,y)$ ganzer Zahlen gibt mit $|x/y - \sqrt{d}| < 1/y^2$ (hier benutzen wir die Voraussetzung $d > 0$ und $d$ kein Quadrat, also $\sqrt{d} \in \mathbb{R} \setminus \mathbb{Q}$). Es folgt

$$\begin{aligned} |x^2 - d\,y^2| = y^2\Big|\frac{x}{y} - \sqrt{d}\Big|\Big(\frac{x}{y} + \sqrt{d}\Big) &< \frac{x}{y} + \sqrt{d} = 2\sqrt{d} + \Big(\frac{x}{y} - \sqrt{d}\Big) \\ &\leq 2\sqrt{d} + \Big|\frac{x}{y} - \sqrt{d}\Big| < 2\sqrt{d} + \frac{1}{y^2} \leq 2\sqrt{d} + 1\,. \end{aligned}$$

Es gibt nur endliche viele ganze Zahlen $m$ mit $|m| < 2\sqrt{d}+1$, also muss es ein $m$ geben mit $x^2 - dy^2 = m$ für unendlich viele $(x,y)$ (hier verwenden wir die unendliche Version des Schubfachprinzips). Um daraus eine Lösung von $x^2 - dy^2 = 1$

zu konstruieren, wollen wir zwei Paare mit $x^2 - dy^2 = m$ durcheinander „dividieren.“ Damit dabei ganze Zahlen herauskommen, müssen die beiden Paare $(x, y)$ zueinander mod $m$ kongruent sein. Da es nur endlich viele Paare von Restklassen mod $m$ gibt, gibt es jedenfalls $0 < x < u$ und $0 < y, v$ mit $x^2 - dy^2 = u^2 - dv^2 = m$ und $x \equiv u \bmod m$, $y \equiv v \bmod m$. Dann wird

$$(xu - d\,yv)^2 - d(uy - xv)^2 = m^2\,,$$

und

$$xu - d\,yv \equiv x^2 - d\,y^2 = m \equiv 0 \bmod m\,, \qquad uy - xv \equiv xy - xy = 0 \bmod m\,,$$

also haben wir mit

$$\Big(\frac{xu - d\,yv}{m}, \frac{uy - xv}{m}\Big) \in S_d^+$$

eine nichttriviale Lösung gefunden. (Wäre die Lösung trivial, dann wäre $uy = xv$ und $xu - dyv = \pm m$, woraus man $mx = x(u^2 - dv^2) = (xu - dyv)u = \pm mu$, also $x = \pm u$ folgern könnte, im Widerspruch zu $0 < x < u$.) ❑

9.8. **Definition.** Der Erzeuger $(x_1, y_1)$ von $S_d^+$ mit $x_1, y_1 > 0$ (und damit $x_1 > 0$ minimal in einer nichttrivialen Lösung) heißt *Grundlösung* der Pellschen Gleichung $x^2 - dy^2 = 1$. ♢

**DEF**
Grundlösung

9.9. **Beispiele.** Hier ist eine Tabelle mit Beispielen: ♣

**BSP**
Grund-
lösungen

| $d$ | $x_1$ | $y_1$ | | $d$ | $x_1$ | $y_1$ | | $d$ | $x_1$ | $y_1$ | | $d$ | $x_1$ | $y_1$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 2 | | 7 | 8 | 3 | | 12 | 7 | 2 | | 17 | 33 | 8 |
| 3 | 2 | 1 | | 8 | 3 | 1 | | 13 | 649 | 180 | | 18 | 17 | 4 |
| 5 | 9 | 4 | | 10 | 19 | 6 | | 14 | 15 | 4 | | 19 | 170 | 39 |
| 6 | 5 | 2 | | 11 | 10 | 3 | | 15 | 4 | 1 | | 20 | 9 | 2 |

Es bleibt die Frage zu klären, wie man diese Grundlösungen findet, beziehungsweise, wie man die guten rationalen Näherungen effizient finden kann, deren Existenz wir in Lemma 9.6 bewiesen haben. Das kann mithilfe von *Kettenbrüchen* geschehen, die wir im nächsten Abschnitt besprechen.

## 10. Kettenbrüche

Sei $\alpha \in \mathbb{R}\setminus\mathbb{Q}$ eine irrationale reelle Zahl. Wir setzen $\alpha_0 = \alpha$ und definieren rekursiv für $n \geq 0$

$$a_n = \lfloor \alpha_n \rfloor\,, \qquad \alpha_{n+1} = \frac{1}{\alpha_n - a_n}\,.$$

Es ist klar, dass alle $\alpha_n$ ebenfalls irrational sind; daher ist stets $\alpha_n \neq a_n$, und $a_n$ ist für alle $n \geq 0$ definiert. Für $n \geq 1$ ist $\alpha_n > 1$ und damit auch $a_n \geq 1$.

Es gilt dann

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_{n-1} + \cfrac{1}{\alpha_n}}}}}\,.$$

Zur Vereinfachung der Notation kürzen wir den geschachtelten Bruch auf der rechten Seite ab durch

$$[a_0; a_1, a_2, \ldots, a_{n-1}, \alpha_n]\,.$$

(Diese Schreibweise ist für beliebige $a_0, \ldots, a_{n-1}, \alpha_n$ sinnvoll.) Wir haben dann die Rekursion

$$[x] = x\,, \qquad [a_0; a_1, \ldots, a_{n-2}, a_{n-1}, x] = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} + \tfrac{1}{x}] \quad (n \geq 1)\,.$$

**DEF** Kettenbruchentwicklung

10.1. **Definition.** Der formale Ausdruck

$$[a_0; a_1, a_2, a_3, \ldots]$$

heißt die *Kettenbruchentwicklung* von $\alpha$. ♢

Wenn $a_0, a_1, a_2, \ldots$ ganze Zahlen sind mit $a_1, a_2, \ldots \geq 1$, dann ist $[a_0; a_1, \ldots, a_n]$ für jedes $n$ eine rationale Zahl. Wir können diese Zahl berechnen, indem wir den geschachtelten Bruch von innen her auflösen (d.h., wir verwenden obige Rekursion). Das hat den Nachteil, dass wir jedes Mal wieder neu anfangen müssen, wenn wir den Kettenbruch verlängern. Das folgende Lemma zeigt eine bessere Alternative auf.

**LEMMA** Berechnung von Kettenbrüchen

10.2. **Lemma.** *Sei $a_0, a_1, a_2, \ldots$ eine Folge ganzer Zahlen mit $a_1, a_2, \ldots \geq 1$. Wir setzen $p_{-2} = 0$, $q_{-2} = 1$, $p_{-1} = 1$, $q_{-1} = 0$ und definieren rekursiv*

$$p_{n+1} = a_{n+1}p_n + p_{n-1}\,, \qquad q_{n+1} = a_{n+1}q_n + q_{n-1}\,.$$

*Die Folgen $(p_n)$ und $(q_n)$ haben folgende Eigenschaften.*

(1) $p_{n+1}q_n - p_nq_{n+1} = (-1)^n$ *für alle* $n \geq -2$. *Insbesondere gilt* $p_n \perp q_n$.

(2) $[a_0; a_1, \ldots, a_n] = p_n/q_n$ *für alle* $n \geq 0$.

(3) *Wenn* $[a_0; a_1, a_2, \ldots]$ *die Kettenbruchentwicklung von* $\alpha$ *ist, dann gilt für* $n \geq 0$

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

*Außerdem ist* $\operatorname{sign}(\alpha - p_n/q_n) = (-1)^n$.

(4) *Unter den Annahmen von (3) gilt*

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

**DEF** Näherungsbruch

Der Bruch $p_n/q_n = [a_0; a_1, \ldots, a_n]$ heißt der *n-te Näherungsbruch von* $\alpha$ (engl.: *nth convergent*), wenn $[a_0; a_1, a_2, \ldots]$ die Kettenbruchentwicklung von $\alpha$ ist.

*Beweis.*

(1) Die Behauptung gilt nach Definition für $n = -2$. Wir beweisen den allgemeinen Fall durch Induktion. Sei $n \geq -1$ und die Behauptung für $n-1$ schon gezeigt. Dann gilt

$$\begin{aligned} p_{n+1}q_n - p_nq_{n+1} &= (a_{n+1}p_n + p_{n-1})q_n - p_n(a_{n+1}q_n + q_{n-1}) \\ &= -(p_nq_{n-1} - p_{n-1}q_n) = -(-1)^{n-1} = (-1)^n\,. \end{aligned}$$

(2) Es gilt allgemeiner für $n \geq -1$ und beliebiges $x$:

$$[a_0; a_1, \ldots, a_n, x] = \frac{p_nx + p_{n-1}}{q_nx + q_{n-1}}\,.$$

Das ist klar für $n = -1$: $x = (p_{-1}x + p_{-2})/(q_{-1}x + q_{-2})$. Unter der Annahme, dass die Beziehung für $n-1$ gilt, folgt

$$\begin{aligned} [a_0; a_1, \ldots, a_{n-1}, a_n, x] &= [a_0; a_1, \ldots, a_{n-1}, a_n + \tfrac{1}{x}] \\ &= \frac{p_{n-1}\big(a_n + \frac{1}{x}\big) + p_{n-2}}{q_{n-1}\big(a_n + \frac{1}{x}\big) + q_{n-2}} = \frac{(a_np_{n-1} + p_{n-2})x + p_{n-1}}{(a_nq_{n-1} + q_{n-2})x + q_{n-1}} \\ &= \frac{p_nx + p_{n-1}}{q_nx + q_{n-1}} \end{aligned}$$

Wenn wir hierin $x = a_{n+1}$ setzen, erhalten wir die Aussage des Lemmas.

(3) Es ist $\alpha = [a_0; a_1, \ldots, a_n, \alpha_{n+1}]$. Aus der eben bewiesenen Aussage und Teil (1) folgt

$$\alpha - \frac{p_n}{q_n} = \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}} - \frac{p_n}{q_n} = \frac{-(p_nq_{n-1} - p_{n-1}q_n)}{q_n(\alpha_{n+1}q_n + q_{n-1})} = \frac{(-1)^n}{q_n(\alpha_{n+1}q_n + q_{n-1})}\,.$$

Das zeigt die Behauptung über das Vorzeichen der Differenz. Weiter ist $\alpha_{n+1} > a_{n+1}$, also $\alpha_{n+1}q_n + q_{n-1} > a_{n+1}q_n + q_{n-1} = q_{n+1}$ (beachte, dass $1 = q_0 \leq q_1 < q_2 < \ldots$ gilt) und daher

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_nq_{n+1}}\,.$$

Schließlich ist $q_{n+1} = a_{n+1}q_n + q_{n-1} \geq a_{n+1}q_n \geq q_n$.

(4) Die Differenz

$$\frac{p_{n+2}}{q_{n+2}} - \frac{p_n}{q_n} = \frac{a_{n+2}(p_{n+1}q_n - p_nq_{n+1})}{q_nq_{n+2}} = \frac{a_{n+2}(-1)^n}{q_nq_{n+2}}$$

ist positiv für gerades $n$ und negativ für ungerades $n$. ❑

**FOLG**
Konvergenz von Kettenbrüchen

10.3. **Folgerung.**

(1) *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Ist $[a_0; a_1, a_2, \ldots]$ die Kettenbruchentwicklung von $\alpha$, so gilt*

$$\lim_{n\to\infty} \frac{p_n}{q_n} = \lim_{n\to\infty} [a_0; a_1, \ldots, a_n] = \alpha\,.$$

(2) *Seien $a_0, a_1, a_2, \ldots \in \mathbb{Z}$ mit $a_1, a_2, \ldots \geq 1$. Sei*

$$\frac{p_n}{q_n} = [a_0; a_1, a_2, \ldots, a_n]$$

*wie in Lemma 10.2. Dann konvergiert die Folge $\frac{p_n}{q_n}$ gegen einen Grenzwert $\alpha$, und $[a_0; a_1, a_2, \ldots]$ ist die Kettenbruchentwicklung von $\alpha$.*

*Beweis.*

(1) Das folgt sofort aus Lemma 10.2, (3).

(2) Nach Lemma 10.2, (1) gilt $|p_{n+1}/q_{n+1} - p_n/q_n| = 1/(q_n q_{n+1})$. Für $n \geq 2$ ist $q_n \geq n$ (siehe oben im Beweis von Lemma 10.2 (3)), also konvergiert die Reihe $\sum_{n\geq 1} 1/(q_n q_{n+1})$. Das zeigt, dass die Folge $(p_n/q_n)$ eine Cauchy-Folge ist; sie konvergiert daher gegen einen Grenzwert $\alpha$. (Man könnte auch verwenden, dass die Reihe

$$\sum_{n=0}^{\infty} \Big( \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \Big) = \sum_{n=0}^{\infty} \frac{(-1)^n}{q_n q_{n+1}}$$

alternierend ist mit Termen, deren Beträge monoton gegen 0 gehen.)
Es ist $a_0 = p_0/q_0 < p_2/q_2 \leq \alpha \leq p_3/q_3 < p_1/q_1 \leq a_0 + 1$, also gilt $a_0 = \lfloor \alpha \rfloor$. Dann ist

$$\alpha_1 = \frac{1}{\alpha - a_0} = \lim_{n\to\infty} [a_1; a_2, \ldots, a_n]\,.$$

Es folgt, dass $a_1 = \lfloor \alpha_1 \rfloor$ ist, und auf dieselbe Weise ergibt sich induktiv, dass $[a_0; a_1, a_2, \ldots]$ die Kettenbruchentwicklung von $\alpha$ sein muss. ❑

Man kann dieses Ergebnis auch so formulieren: Die Abbildungen

$$\alpha \longmapsto \text{Kettenbruchentwicklung von } \alpha$$

und

$$[a_0; a_1, a_2, \ldots] \longmapsto \lim_{n\to\infty} [a_0; a_1, \ldots, a_n]$$

sind zueinander inverse Bijektionen zwischen $\{(a_n)_{n\geq 0} \mid a_0 \in \mathbb{Z}, a_1, a_2, \ldots \in \mathbb{Z}_{\geq 1}\}$ und $\mathbb{R} \setminus \mathbb{Q}$.

Für rationale Zahlen $\alpha$ bricht die Kettenbruchentwicklung ab, und man erhält jeweils zwei Kettenbrüche mit Wert $\alpha$, z.B.

$$\frac{17}{13} = [1; 3, 4] = [1; 3, 3, 1]\,.$$

Das kommt daher, dass $[a_0; a_1, \ldots, a_{n-1}, a_n, 1] = [a_0; a_1, \ldots, a_{n-1}, a_n + 1]$ ist.

**BSP** Kettenbruch-entwicklungen

10.4. **Beispiele.** Es ist zum Beispiel

$$\begin{aligned}
\frac{1+\sqrt{5}}{2} &= [1;1,1,1,\ldots] \\
\sqrt{2} &= [1;2,2,2,\ldots] \\
\sqrt{409} &= [20;4,2,7,1,1,1,4,2,2,13,13,2,2,4,1,1,1,7,2,4,40, \\
&\qquad 4,2,7,1,1,1,4,2,2,13,13,2,2,4,1,1,1,7,2,4,40,\ldots] \\
\sqrt[3]{2} &= [1;3,1,5,1,1,4,1,1,8,1,14,1,10,2,1,4,12,2,3,2,1,3,4, \\
&\qquad 1,1,2,14,3,12,1,15,3,1,4,534,1,1,\ldots] \\
e &= [2;1,2,1,1,4,1,1,6,1,1,8,1,1,10,1,1,12,1,1,14,1,\ldots] \\
\pi &= [3;7,15,1,292,1,1,1,2,1,3,1,14,2,1,1,2,2,2,2,1,84,\ldots]
\end{aligned}$$

Nach Lemma 10.2 (3) bekommt man besonders gut approximierende Näherungsbrüche $p_n/q_n$, wenn $a_{n+1}$ groß ist. Das ergibt zum Beispiel die Näherungen

$$\pi \approx \quad [3;7] \quad = \frac{22}{7} = 3{,}14285714\ldots \quad \text{mit} \quad \left|\pi - \frac{22}{7}\right| < \frac{1}{16\cdot 7^2} \quad \text{und}$$

$$\pi \approx [3;7,15,1] = \frac{355}{113} = 3{,}141592920\ldots \quad \text{mit} \quad \left|\pi - \frac{355}{113}\right| < \frac{1}{293\cdot 113^2}\,. \qquad \clubsuit$$

Wir müssen jetzt noch zeigen, dass jede hinreichend gute rationale Approximation an $\alpha$ auch als Näherungsbruch in der Kettenbruchentwicklung von $\alpha$ auftaucht.

**LEMMA** gute Näherungen sind Näherungsbrüche

10.5. **Lemma.** *Sei $\alpha \in \mathbb{R}\setminus\mathbb{Q}$, und sei $p/q \in \mathbb{Q}$ mit*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}\,.$$

*Dann ist $p/q = p_n/q_n$ der $n$-te Näherungsbruch von $\alpha$ für ein $n \geq 0$.*

*Beweis.* Wir behandeln zuerst den Fall $q = 1$. Dann ist $p/q = p$ die zu $\alpha$ nächst gelegene ganze Zahl. Wenn $p < \alpha$ ist, dann haben wir $p = a_0 = p_0/q_0$. Wenn $p > \alpha$ ist, dann ist $p = a_0 + 1$, und wir haben $a_0 + 1/2 < \alpha$, woraus sich $\alpha_1 < 2$, also $a_1 = 1$ ergibt, und wir erhalten $p_1/q_1 = a_0 + 1 = p$.

Für $q \geq 2$ behaupten wir, dass sogar die folgende etwas stärkere Aussage gilt:

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q(2q-1)} \quad \Longrightarrow \quad \exists n\colon \frac{p}{q} = \frac{p_n}{q_n}\,.$$

Wir können dabei annehmen, dass $p/q$ vollständig gekürzt ist. Der Vorteil dieser Variante ist, dass damit der Induktionsbeweis funktioniert. Wir beweisen diese Behauptung also durch Induktion über $q$. Wir können $\alpha$ und $p/q$ durch $\alpha - a_0$ und $p/q - a_0$ ersetzen und daher annehmen, dass $0 < \alpha < 1$ ist. Dann muss auch $0 < p/q < 1$ gelten: Wegen $q > 1$ kann Gleichheit nicht eintreten, und aus (z.B.) $p/q < 0$ würde folgen, dass

$$\frac{1}{q(2q-1)} > \left|\alpha - \frac{p}{q}\right| > \left|\frac{p}{q}\right| \geq \frac{1}{q}$$

ist im Widerspruch zu $q \geq 2$. Ähnliches ergibt sich für $p/q > 1$. Es muss also $0 < p < q$ gelten.

Wir haben nun

$$\left|\frac{1}{\alpha} - \frac{q}{p}\right| = \left|\alpha - \frac{p}{q}\right| \frac{q}{p\alpha} < \frac{1}{p\alpha(2q-1)}\,.$$

Außerdem ist

$$\alpha(2q-1) \geq \Big(\frac{p}{q} - \frac{1}{q(2q-1)}\Big)(2q-1) = 2p - \frac{p}{q} - \frac{1}{q} \geq 2p-1\,;$$

damit ist

$$\Big|\frac{1}{\alpha} - \frac{q}{p}\Big| < \frac{1}{p(2p-1)}\,.$$

Wenn $p \geq 2$ ist, dann folgt nach Induktionsannahme, dass $q/p$ ein Näherungsbruch der Kettenbruchentwicklung von $1/\alpha = \alpha_1$ ist; damit ist $p/q$ ein Näherungsbruch der Kettenbruchentwicklung von $\alpha$. Im Fall $p = 1$ haben wir

$$\frac{2q-2}{q(2q-1)} = \frac{1}{q} - \frac{1}{q(2q-1)} < \alpha < \frac{1}{q} + \frac{1}{q(2q-1)} = \frac{2}{2q-1}$$

und damit

$$q - \tfrac{1}{2} < \frac{1}{\alpha} < q + \frac{q}{2(q-1)} \leq q+1\,.$$

Es folgt $a_1 = q$ oder $a_1 = q-1$. Im ersten Fall ist $p_1/q_1 = 1/q = p/q$; im zweiten Fall ist $a_2 = 1$ und dann $p_2/q_2 = 1/q = p/q$. ❑

Damit können wir die Pellsche Gleichung lösen.

**SATZ** Lösung der Pellschen Gleichung

10.6. **Satz.** *Sei $d > 0$ kein Quadrat. Seien $p_n/q_n$ die Näherungsbrüche der Kettenbruchentwicklung von $\sqrt{d}$. Dann ist die Grundlösung der Pellschen Gleichung*

$$x^2 - dy^2 = 1$$

*gegeben durch $(x_1, y_1) = (p_n, q_n)$, wobei $n \geq 0$ minimal ist mit $p_n^2 - dq_n^2 = 1$.*

*Beweis.* Wir haben in Lemma 9.5 gesehen, dass jede nichttriviale Lösung $(x, y)$ mit $x, y > 0$ die Ungleichung

$$\Big|\sqrt{d} - \frac{x}{y}\Big| < \frac{1}{2\sqrt{d}y^2} < \frac{1}{2y^2}\,.$$

erfüllt. Nach Lemma 10.5 folgt, dass $(x, y) = (p_n, q_n)$ sein muss für ein geeignetes $n$. Da alle $a_n$ (einschließlich $a_0$) positiv sind, ist die Folge $(p_n)_{n\geq 0}$ streng monoton wachsend; die Grundlösung ist also durch das kleinste $n$ gegeben, das eine Lösung liefert. ❑

**BSP** Grundlösung aus Kettenbruch

10.7. **Beispiel.** Zur Illustration berechnen wir die Grundlösung für $d = 31$. Es gilt $5 < \sqrt{31} < 6$. Wir erhalten folgende Tabelle.

| $n$ | $\alpha_n$ | $a_n$ | $p_n$ | $q_n$ | $p_n^2 - 31\,q_n^2$ |
|---|---|---|---|---|---|
| 0 | $\sqrt{31}$ | 5 | 5 | 1 | $-6$ |
| 1 | $\frac{1}{\sqrt{31}-5} = \frac{\sqrt{31}+5}{6}$ | 1 | 6 | 1 | 5 |
| 2 | $\frac{6}{\sqrt{31}-1} = \frac{\sqrt{31}+1}{5}$ | 1 | 11 | 2 | $-3$ |
| 3 | $\frac{5}{\sqrt{31}-4} = \frac{\sqrt{31}+4}{3}$ | 3 | 39 | 7 | 2 |
| 4 | $\frac{3}{\sqrt{31}-5} = \frac{\sqrt{31}+5}{2}$ | 5 | 206 | 37 | $-3$ |
| 5 | $\frac{2}{\sqrt{31}-5} = \frac{\sqrt{31}+5}{3}$ | 3 | 657 | 118 | 5 |
| 6 | $\frac{3}{\sqrt{31}-4} = \frac{\sqrt{31}+4}{5}$ | 1 | 863 | 155 | $-6$ |
| 7 | $\frac{5}{\sqrt{31}-1} = \frac{\sqrt{31}+1}{6}$ | 1 | 1520 | 273 | 1 |
| 8 | $\frac{6}{\sqrt{31}-5} = \sqrt{31}+5$ | | | | |

Die Grundlösung ist also $(1520, 273)$. ♣

Wenn wir zuerst eine Lösung $(x_0, y_0)$ von $x^2 - dy^2 = -1$ finden, dann gilt (mit einem analogen Beweis wie für Lemma 9.4, wenn man $\phi(x, y)^2$ verwendet), dass $(x_0, y_0)$ die Gruppe $T_d^+ = \{(x, y) \in T_d \mid x > 0\}$ erzeugt. Es folgt, dass $S_d^+$ gerade aus allen $(x, y)^{*2}$ mit $(x, y) \in T_d^+$ besteht. Insbesondere ist die Grundlösung dann

$$(x_1, y_1) = (x_0, y_0)^{*2} = (x_0^2 + dy_0^2, 2x_0y_0) = (2x_0^2 + 1, 2x_0y_0)\,.$$

10.8. **Beispiel.** Für $d = 13$ erhalten wir beispielsweise: **BSP** $x^2-dy^2=-1$

| $n$ | $\alpha_n$ | $a_n$ | $p_n$ | $q_n$ | $p_n^2 - 13\,q_n^2$ |
|---|---|---|---|---|---|
| 0 | $\sqrt{13}$ | 3 | 3 | 1 | $-4$ |
| 1 | $\frac{1}{\sqrt{13}-3} = \frac{\sqrt{13}+3}{4}$ | 1 | 4 | 1 | 3 |
| 2 | $\frac{4}{\sqrt{13}-1} = \frac{\sqrt{13}+1}{3}$ | 1 | 7 | 2 | $-3$ |
| 3 | $\frac{3}{\sqrt{13}-2} = \frac{\sqrt{13}+2}{3}$ | 1 | 11 | 3 | 4 |
| 4 | $\frac{3}{\sqrt{13}-1} = \frac{\sqrt{13}+1}{4}$ | 1 | 18 | 5 | $-1$ |

Also ist $(x_0, y_0) = (18, 5)$ ein Erzeuger von $T_{13}^+$, und

$$(x_1, y_1) = (2 \cdot 18^2 + 1, 2 \cdot 18 \cdot 5) = (649, 180)$$

ist die Grundlösung in $S_{13}^+$. ♣

Es gibt einen weiteren ähnlichen Fall, in dem es eine „Abkürzung“ gibt. Wenn man zuerst ein $n \geq 0$ findet mit $p_n^2 - dq_n^2 = \pm 2$, dann liefert „Quadrieren“ von $(p_n, q_n)$ die Relation

$$\big(2(p_n^2 \mp 1)\big)^2 - d(2\,p_nq_n)^2 = (p_n^2 + dq_n^2)^2 - d(2p_nq_n)^2 = 4\,,$$

also eine Lösung $(x_1, y_1) = (p_n^2 \mp 1, p_nq_n)$ von $x^2 - dy^2 = 1$, die wiederum die Grundlösung ist. Im ersten Beispiel oben mit $d = 31$ war etwa $(p_3, q_3) = (39, 7)$ mit $p_3^2 - 31q_3^2 = 2$, also ist die Grundlösung $(x_1, y_1) = (39^2 - 1, 39 \cdot 7) = (1520, 273)$.

Anhand der Tabellen in Beispiel 10.7 und Beispiel 10.8 macht man folgende Beobachtungen:

- $\alpha_n = (\sqrt{d} + u_n)/v_n$ mit $u_n \in \mathbb{Z}$, $v_n \in \mathbb{Z}_{>0}$ und $v_n \mid d - u_n^2$.
- $p_n^2 - dq_n^2 = (-1)^{n+1}v_{n+1}$.

Das werden wir jetzt beweisen.

10.9. **Lemma.** *Sei $\alpha = \sqrt{d}$ (mit $d > 0$ kein Quadrat). Dann gilt für die Größen $\alpha_n$, $a_n$, $p_n$, $q_n$, die zur Kettenbruchentwicklung von $\alpha$ gehören:* **LEMMA** Kettenbruch von $\sqrt{d}$

(1) *Für $n \geq 0$ gibt es $u_n \in \mathbb{Z}$, $v_n \in \mathbb{Z}_{>0}$ mit $v_n \mid d - u_n^2$ und $\alpha_n = \frac{\sqrt{d}+u_n}{v_n}$.*

(2) *Für $n \geq -1$ gilt $p_np_{n-1} - dq_nq_{n-1} = (-1)^n u_{n+1}$ und $p_n^2 - dq_n^2 = (-1)^{n+1}v_{n+1}$.*

Aus Lemma 10.11 unten wird sich noch ergeben, dass für $n \geq 1$ gilt:

$$0 < u_n < \sqrt{d} \qquad \text{und} \qquad 0 < \sqrt{d} - u_n < v_n < u_n + \sqrt{d} < 2\sqrt{d}\,.$$

*Beweis.*

(1) Induktion nach $n$. Für $n = 0$ ist $\alpha_0 = \alpha = \sqrt{d}$; die Behauptung gilt also mit $u_0 = 0$ und $v_0 = 1$.

Sei jetzt $n > 0$. Wir haben

$$\alpha_n = \frac{1}{\alpha_{n-1} - a_{n-1}} = \frac{v_{n-1}}{\sqrt{d} + u_{n-1} - a_{n-1}v_{n-1}}$$
$$= \frac{\sqrt{d} + a_{n-1}v_{n-1} - u_{n-1}}{v_n} = \frac{\sqrt{d} + u_n}{v_n}$$

mit $u_n = a_{n-1}v_{n-1} - u_{n-1}$ und $v_n = (d - u_n^2)/v_{n-1}$. Wir müssen noch zeigen, dass $v_n$ hier ganz ist. Nach Induktionsannahme ist $d \equiv u_{n-1}^2 \bmod v_{n-1}$. Da $u_n \equiv -u_{n-1} \bmod v_{n-1}$ ist, folgt $d \equiv u_n^2 \bmod v_{n-1}$, also ist $d - u_n^2 = v_{n-1}v_n$ mit $v_n \in \mathbb{Z}$. Aus Teil (2) (dessen Beweis nicht benutzt, dass $v_n > 0$ ist) folgt $\operatorname{sign} v_n = (-1)^n \operatorname{sign}(p_{n-1}^2 - dq_{n-1}^2) = 1$.

(2) Induktion nach $n$. Für $n = -1$ sind die Behauptungen klar. Sei jetzt $n \geq 0$. Dann ist (unter Verwendung der Induktionsannahme)

$$p_n p_{n-1} - dq_n q_{n-1} = a_n(p_{n-1}^2 - dq_{n-1}^2) + p_{n-1}p_{n-2} - dq_{n-1}q_{n-2}$$
$$= a_n(-1)^n v_n + (-1)^{n-1}u_n = (-1)^n u_{n+1}$$

und

$$(p_n^2 - dq_n^2)(p_{n-1}^2 - dq_{n-1}^2) = (p_n p_{n-1} - dq_n q_{n-1})^2 - d(p_n q_{n-1} - p_{n-1}q_n)^2$$
$$= -(d - u_{n+1}^2) = -v_{n+1}v_n$$
$$= (-1)^{n+1}v_{n+1}(p_{n-1}^2 - dq_{n-1}^2)\,,$$

also ist $p_n^2 - dq_n^2 = (-1)^{n+1}v_{n+1}$. ❑

Wir können die Berechnung also stoppen, sobald $v_{n+1} = 1$ wird. Dann ist entweder $p_n^2 - dq_n^2 = 1$ (falls $n$ ungerade), und $(p_n, q_n)$ ist die Grundlösung, oder (falls $n$ gerade) $p_n^2 - dq_n^2 = -1$, dann ist $(2p_n^2 + 1, 2p_n q_n)$ die Grundlösung.

Wenn $v_{n+1} = 2$ zuerst auftritt, dann haben wir bereits gesehen, dass $(x, y) = (p_n^2 + (-1)^n, p_n q_n)$ eine Lösung liefert. Wir müssen noch zeigen, dass dies die Grundlösung ist. Andernfalls wäre $(x, y) = (x_k, y_k) = (x_1, y_1)^{*k}$ mit $k \geq 2$; insbesondere müsste gelten $2x_1^2 - 1 = x_2 \leq x_k = p_n^2 + (-1)^n$. Es folgt

$$x_1^2 \leq \frac{1}{2}(p_n^2 + 1 + (-1)^n) \leq \frac{p_n^2}{2} + 1\,.$$

Wir wissen, dass $x_1 = p_m$ ist für ein geeignetes $m$. Aus der obigen Ungleichung ergibt sich

$$p_m = x_1 \leq \sqrt{\frac{p_n^2}{2} + 1} \leq p_n\,,$$

außer möglicherweise, wenn $p_n = 1$ ist. Wegen des streng monotonen Wachstums von $(p_k)_{k\geq 0}$ (beachte, dass $a_0 = \lfloor\sqrt{d}\rfloor \geq 1$ ist) muss dann $n = 0$ sein. Es ist $p_0^2 - dq_0^2 = 1 - d = -2$, also $d = 3$, und die Grundlösung ist $(x_1, y_1) = (2, 1) = (p_0^2 + 1, p_0 q_0)$ wie behauptet. In allen anderen Fällen ist $p_m \leq p_n$, also $m \leq n$, und weil $m = n$ nicht möglich ist, folgt $m < n$. Dann war aber bereits $v_{m+1} = 1$, und wir hätten $v_{n+1} = 2$ gar nicht erst erreicht. Dieser Widerspruch beweist die Behauptung.

---

**ALGO**
Berechnung einer Grundlösung

Wir erhalten folgenden Algorithmus für die Berechnung der Grundlösung der Pellschen Gleichung $x^2 - dy^2 = 1$.

(1) **Initialisierung.**
Setze $p_{-2} = 0$, $q_{-2} = 1$, $p_{-1} = 1$, $q_{-1} = 0$, $k = \lfloor\sqrt{d}\rfloor$, $u_0 = 0$, $v_0 = 1$.

(2) **Iteration.**
Für $n = 0, 1, 2, \ldots$ berechne $a_n = \lfloor(k + u_n)/v_n\rfloor$, $p_n = a_n p_{n-1} + p_{n-2}$, $q_n = a_n q_{n-1} + q_{n-2}$, $u_{n+1} = a_n v_n - u_n$ und $v_{n+1} = (d - u_{n+1}^2)/v_n$.

(3) **Abbruch.**
Wenn $v_{n+1} = 1$ ist, dann gib $(p_n, q_n)$ aus, wenn $n$ ungerade ist, anderenfalls gib $(2p_n^2 + 1, 2p_n q_n)$ aus.
Wenn $v_{n+1} = 2$ ist, dann gib $(p_n^2 + (-1)^n, p_n q_n)$ aus.

---

Man beachte, dass (sobald $k = \lfloor\sqrt{d}\rfloor$ bestimmt ist) in diesem Algorithmus nur Operationen mit ganzen Zahlen vorkommen.

Für die Berechnung von $u_n$ und $v_n$ ist es nicht einmal nötig, $a_n$ zu berechnen:
Es gilt $u_{n+1} \equiv -u_n \bmod v_n$, und aus der Formel für $a_n$ folgt

$$a_n \le \frac{k + u_n}{v_n} < a_n + 1 \quad \Longrightarrow \quad k - v_n < a_n v_n - u_n = u_{n+1} \le k\,,$$

sodass $u_{n+1}$ durch die Kongruenz und die Ungleichungen eindeutig bestimmt ist. (Für die Bestimmung von $p_n$ und $q_n$ wird $a_n$ aber in jedem Fall benötigt.)

10.10. **Definition.** Wir nennen eine Zahl $\alpha = (\sqrt{d} + u)/v$ (mit $u \in \mathbb{Z}$, $v \in \mathbb{Z}_{>0}$ und $u^2 \equiv d \bmod v$) *reduziert*, wenn $\alpha > 1$ und $-1 < \bar\alpha = (-\sqrt{d} + u)/v < 0$ gilt. ♢

**DEF** $(\sqrt{d} + u)/v$ reduziert

10.11. **Lemma.**

**LEMMA** reduzierte Zahlen und Kettenbrüche

(1) *Ist $\alpha = (\sqrt{d} + u)/v$ reduziert, so ist auch $\alpha' = 1/(\alpha - \lfloor\alpha\rfloor)$ reduziert. Es gilt $\alpha' = (\sqrt{d} + u')/v'$ mit $u' = \lfloor\alpha\rfloor v - u$ und $v' = (d - (u')^2)/v$.*

(2) *In der Kettenbruchentwicklung von $\alpha = \sqrt{d}$ ist $\alpha_n$ reduziert, sobald $n \ge 1$ ist.*

(3) *Ist $\alpha = (\sqrt{d} + u)/v$ reduziert, so gibt es ein eindeutig bestimmtes reduziertes $\alpha' = (\sqrt{d} + u')/v'$, sodass $\alpha = 1/(\alpha' - \lfloor\alpha'\rfloor)$ ist.*

*Beweis.* Die Abbildung $\alpha = x + y\sqrt{d} \mapsto \bar\alpha = x - y\sqrt{d}$ ist ein Automorphismus des Körpers $\mathbb{Q}(\sqrt{d}) = \{x + y\sqrt{d} \mid x, y \in \mathbb{Q}\}$: Es gilt $\overline{\alpha + \beta} = \bar\alpha + \bar\beta$, $\overline{\alpha\beta} = \bar\alpha \cdot \bar\beta$, $\overline{\alpha^{-1}} = \bar\alpha^{-1}$.

(1) Aus $\alpha > 1$ und $-1 < \bar\alpha < 0$ folgt mit $a = \lfloor\alpha\rfloor$, dass $0 < \alpha - a < 1$ und $\bar\alpha - a < -1$ gilt, also

$$\alpha' = \frac{1}{\alpha - a} > 1 \quad \text{und} \quad -1 < \overline{\alpha'} = \frac{1}{\bar\alpha - a} < 0\,.$$

Die Formeln für $u'$ und $v'$ ergeben sich wie in Lemma 10.9.

(2) Es ist $\alpha_1 = 1/(\sqrt{d} - \lfloor\sqrt{d}\rfloor) > 1$, und $\bar\alpha_1 = -1/(\sqrt{d} + \lfloor\sqrt{d}\rfloor)$ ist negativ und hat Betrag $< 1$. Also ist $\alpha_1$ reduziert. Nach Teil (1) folgt mit Induktion, dass auch alle folgenden $\alpha_n$ reduziert sind.

(3) Wir zeigen zunächst die Eindeutigkeit. Sei $\alpha'$ ein reduzierter Vorgänger von $\alpha$. Die Ungleichungen in Definition 10.10 für $\alpha'$ bedeuten $|\sqrt{d} - v'| < u' < \sqrt{d}$. Außerdem muss $vv' = d - u^2$ und $u' \equiv -u \bmod v'$ gelten. Dadurch sind erst $v'$ und dann $u'$ eindeutig durch $d$, $u$, $v$ bestimmt.

Es bleibt die Existenz von $\alpha'$ zu zeigen. Sei $\alpha_0 = \alpha$, und für $n \geq 0$ sei $\alpha_{n+1} = 1/(\alpha_n - \lfloor \alpha_n \rfloor)$. Da für $(\sqrt{d} + u)/v$ reduziert stets $0 < u < \sqrt{d}$ und $0 < v < u + \sqrt{d} < 2\sqrt{d}$ gilt, gibt es nur endlich viele reduzierte Zahlen dieser Form. Nach Teil (1) sind alle $\alpha_n$ reduziert, also gibt es $m \geq 1$ und $k \geq 0$ mit $\alpha_{k+m} = \alpha_k$. Sei $k$ minimal gewählt. Dann muss $k = 0$ sein, denn sonst wären $\alpha_{k-1}$ und $\alpha_{k+m-1}$ zwei verschiedene reduzierte Zahlen, die beide auf $\alpha_k = \alpha_{k+m}$ abgebildet werden, was der schon bewiesenen Eindeutigkeit widerspräche. Es gilt also $\alpha_m = \alpha$; damit leistet $\alpha' = \alpha_{m-1}$ das Gewünschte. ❑

10.12. **Folgerung.** *Sei $\alpha = \sqrt{d}$ mit $d > 0$ kein Quadrat. Sei weiter $m \geq 1$ die kleinste Zahl mit $v_m = 1$. Dann gilt:*

**FOLG** Kettenbruch für $\sqrt{d}$ ist periodisch

(1) *$a_{n+m} = a_n$ für alle $n \geq 1$. Insbesondere ist die Kettenbruchentwicklung von $\sqrt{d}$ ab $a_1$ periodisch mit Periode $m$.*

(2) *$a_m = 2a_0 = 2\lfloor \sqrt{d} \rfloor$.*

*Beweis.*

(1) Nach Lemma 10.9 ist $\alpha_m = \sqrt{d} + u_m$, also $a_m = \lfloor \sqrt{d} \rfloor + u_m$ und

$$\alpha_{m+1} = \frac{1}{\sqrt{d} - \lfloor \sqrt{d} \rfloor} = \alpha_1 \,.$$

Da $\alpha_{n+1}$ nur von $\alpha_n$ abhängt, folgt mit Induktion $\alpha_{n+m} = \alpha_n$ für alle $n \geq 1$; damit ist auch $a_{n+m} = a_n$.

(2) Es bleibt zu zeigen, dass $u_m = \lfloor \sqrt{d} \rfloor$ ist. Das folgt aber aus der Beobachtung, dass $\sqrt{d} + \lfloor \sqrt{d} \rfloor$ der eindeutig bestimmte reduzierte Vorgänger von $\alpha_{m+1} = \alpha_1 = 1/(\sqrt{d} - \lfloor \sqrt{d} \rfloor)$ ist: $\sqrt{d} + \lfloor \sqrt{d} \rfloor > 1$ und $-1 < -\sqrt{d} + \lfloor \sqrt{d} \rfloor < 0$. ❑

Wir sehen, dass die Kettenbruchentwicklung einer Quadratwurzel schließlich periodisch wird. Wir wollen nun die Zahlen charakterisieren, deren Kettenbruchentwicklung dieselbe Eigenschaft hat.

10.13. **Lemma.** *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, und sei $[a_0; a_1, a_2, \ldots]$ die Kettenbruchentwicklung von $\alpha$. Wenn es $k \geq 0$ und $m \geq 1$ gibt, sodass $a_{n+m} = a_n$ ist für alle $n \geq k$, dann ist $\alpha$ Nullstelle eines irreduziblen Polynoms $x^2 + ax + b \in \mathbb{Q}[x]$.*

**LEMMA** periodische Kettenbrüche

Man sagt dann auch, $\alpha$ sei eine *quadratische Irrationalität*.

**DEF** quadratische Irrationalität

*Beweis.* Zunächst folgt, dass auch $\alpha_{n+m} = \alpha_n$ ist für alle $n \geq k$, denn

$$\alpha_{n+m} = [a_{n+m}; a_{n+m+1}, a_{n+m+2}, \ldots] = [a_n; a_{n+1}, a_{n+2}, \ldots] = \alpha_n \,.$$

Wenn $p_n/q_n$ die Näherungsbrüche der Kettenbruchentwicklung von $\alpha$ und $p'_n/q'_n$ diejenigen der Kettenbruchentwicklung von $\alpha_k = \alpha_{k+m}$ sind, dann gilt nach dem Beweis von Lemma 10.2, Teil (2):

$$\alpha = \frac{p_{k-1}\alpha_k + p_{k-2}}{q_{k-1}\alpha_k + q_{k-2}} \quad \Longrightarrow \quad \alpha_k = \frac{q_{k-2}\alpha - p_{k-2}}{-q_{k-1}\alpha + p_{k-1}}$$

und

$$\alpha_k = \frac{p'_{m-1}\alpha_{k+m} + p'_{m-2}}{q'_{m-1}\alpha_{k+m} + q'_{m-2}} = \frac{p'_{m-1}\alpha_k + p'_{m-2}}{q'_{m-1}\alpha_k + q'_{m-2}}\,.$$

Aus der zweiten Gleichung folgt

$$\alpha_k^2 + \frac{q'_{m-2} - p'_{m-1}}{q'_{m-1}}\alpha_k - \frac{p'_{m-2}}{q'_{m-1}} = 0\,,$$

und aus der ersten folgt eine ähnliche Gleichung für $\alpha$. Das Polynom muss irreduzibel sein, weil es keine rationale Nullstelle hat. ❑

10.14. **Beispiel.** Der einfachste Kettenbruch ist $[1;1,1,1,\dots]$. Für die zugehörige quadratische Irrationalität $\phi$ ergibt sich die Gleichung

**BSP** $[1;1,1,\dots]$

$$\phi = 1 + \frac{1}{\phi} \quad\Longrightarrow\quad \phi^2 - \phi - 1 = 0 \quad\Longrightarrow\quad \phi = \frac{\sqrt{5}+1}{2}\,.$$

(Das Vorzeichen der Wurzel ist positiv, da $\phi > 0$ ist.) Das ist der *Goldene Schnitt.* ♣

10.15. **Beispiel.** Wir finden die Zahl mit dem Kettenbruch $[1;2,3,3,3,\dots]$. Zunächst ist $\alpha_2 = [3,3,3,\dots] = \alpha_3$. Wir haben $p'_0 = 3$, $q'_0 = 1$, $p'_{-1} = 1$, $q'_{-1} = 0$ und damit $\alpha_2^2 - 3\alpha_2 - 1 = 0$, also $\alpha_2 = (3+\sqrt{13})/2$ (das Vorzeichen ist positiv, da $\alpha_2 > 3$ ist). Außerdem ist $k = 2$, $p_0 = 1$, $q_0 = 1$, $p_1 = 3$, $q_1 = 2$ und damit

**BSP** periodischer Kettenbruch

$$\alpha = \frac{3\alpha_2 + 1}{2\alpha_2 + 1} = \frac{\frac{11}{2} + \frac{3}{2}\sqrt{13}}{4 + \sqrt{13}} = \frac{5+\sqrt{13}}{6}\,.$$

Tatsächlich ergibt sich für die Folge $(\alpha_n)$:

$$\frac{\sqrt{13}+5}{6} \mapsto \frac{\sqrt{13}+1}{2} \mapsto \frac{\sqrt{13}+3}{2} \mapsto \frac{\sqrt{13}+3}{2} \mapsto \dots\,,$$

was die korrekte Folge $(a_n) = (1,2,3,3,3,\dots)$ liefert. ♣

Wir wollen nun die Umkehrung beweisen: Die Kettenbruchentwicklung einer quadratischen Irrationalität wird schließlich periodisch.

Der wesentliche Schritt wird im folgenden Lemma erledigt.

10.16. **Lemma.** *Sei $\alpha \in \mathbb{R}\setminus\mathbb{Q}$ eine quadratische Irrationalität. Dann gilt für die Größen $\alpha_n$ in der Kettenbruchentwicklung von $\alpha$*

**LEMMA** $\alpha_n$ schließlich reduziert

$$\alpha_n > 1 \qquad \textit{und} \qquad -1 < \bar\alpha_n < 0 \qquad \textit{für } n \textit{ hinreichend groß.}$$

*Beweis.* Nach dem Beweis von Lemma 10.2 gilt

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

also

$$\alpha_{n+1} = \frac{q_{n-1}\alpha - p_{n-1}}{-q_n\alpha + p_n} \qquad \text{und damit} \qquad \bar\alpha_{n+1} = \frac{q_{n-1}\bar\alpha - p_{n-1}}{-q_n\bar\alpha + p_n}\,.$$

Es folgt

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

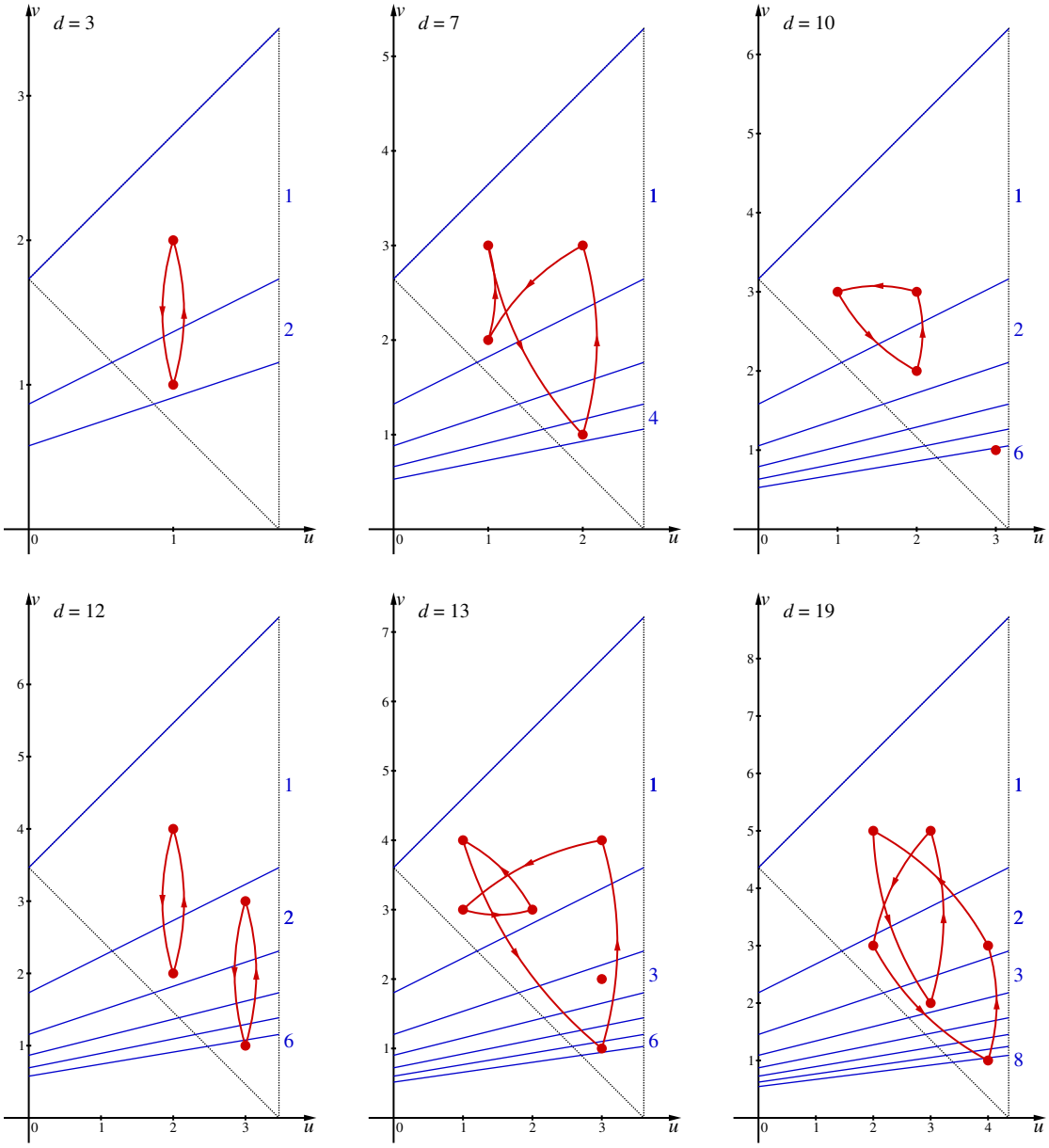



ABBILDUNG 5. Mengen $R_d$ (als Paare $(u, v)$) mit der durch den Kettenbruch-Algorithmus gegebenen Permutation

Für $n \to \infty$ konvergiert $\frac{p_n}{q_n}$ gegen $\alpha \neq \bar{\alpha}$, also konvergiert der zweite Faktor gegen 1. Genauer gilt

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

mit $\varepsilon_n \to 0$ für $n \to \infty$. Es folgt

$$-\frac{1}{\bar{\alpha}_{n+1}} = \frac{q_n}{q_{n-1}} + \frac{(-1)^n}{q_{n-1}^2(\bar{\alpha} - \alpha + \varepsilon_{n-1})} > 1$$

für hinreichend großes $n$, denn der erste Summand ist mindestens $1 + 1/q_{n-1}$, und der zweite Summand ist vom Betrag kleiner als $1/q_{n-1}$, wenn $n$ genügend groß ist. Es folgt (für diese $n$) $-1 < \bar{\alpha}_n < 0$. Nach Konstruktion gilt in jedem Fall $\alpha_n > 1$ für alle $n \geq 1$. ❑

Jetzt müssen wir uns noch davon überzeugen, dass $\alpha_n$ wie oben auch in unserem Sinn reduziert ist. Wir definieren zunächst:

10.17. **Definition.** Sei $d > 0$ kein Quadrat. Wir setzen **DEF** $R_d$

$$R_d = \left\{ \alpha = \frac{\sqrt{d}+u}{v} \;\middle|\; \alpha \text{ reduziert} \right\}. \qquad \diamondsuit$$

Wir erinnern uns daran, dass $\frac{\sqrt{d}+u}{v} \in R_d$ genau dann gilt, wenn $|\sqrt{d}-v| < u < \sqrt{d}$ und $v \mid d-u^2$ gelten. Teile (1) und (3) von Lemma 10.11 lassen sich dann auch so formulieren, dass der Kettenbruch-Algorithmus eine Bijektion $\phi\colon R_d \to R_d$ liefert.

10.18. **Beispiele.** Die Grafiken in Abbildung 5 zeigen verschiedene Mengen $R_d$ zusammen mit der durch den Kettenbruch-Algorithmus gegebenen Permutation. Die (blauen) Zahlen rechts geben $a_n$ an für die $\alpha_n$, die zu Paaren $(u, v)$ gehören, die zwischen den benachbarten blauen Geraden liegen. ♣ **BSP** Iteration auf $R_d$

10.19. **Lemma.** *Ist $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ eine quadratische Irrationalität mit $\alpha > 1$ und $-1 < \bar{\alpha} < 0$, dann ist $\alpha \in R_D$ für ein geeignetes $D$.* **LEMMA** $\alpha$ reduziert $\Rightarrow \alpha \in R_D$

*Beweis.* Wir können jedenfalls schreiben $\alpha = \frac{r\sqrt{d}+s}{t}$ für eine quadratfreie positive ganze Zahl $d$ und ganze Zahlen $r, s, t$ mit $r \neq 0$ und $t > 0$. Dann ist $\alpha = \frac{\sqrt{r^2t^2d}+st}{t^2}$ und erfüllt die verlangten Ungleichungen. Außerdem gilt $r^2t^2d-(st)^2 = t^2(r^2d-s^2)$. Damit ist $\alpha \in R_D$ mit $D = r^2t^2d$ (und $u = st$, $v = t^2$). ❑

Das eben konstruierte $D$ muss nicht optimal sein; es kann echte Teiler von $D$ geben, die ebenfalls funktionieren.

10.20. **Satz.** *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Die Kettenbruchentwicklung von $\alpha$ ist genau dann schließlich periodisch, wenn $\alpha$ eine quadratische Irrationalität ist.* **SATZ** periodische Kettenbruch-entwicklung

*Beweis.* Die eine Richtung wurde in Lemma 10.13 bewiesen. Sei also für die Gegenrichtung $\alpha$ eine quadratische Irrationalität. Nach Lemma 10.16 ist $\alpha_n > 1$, $-1 < \bar{\alpha}_n < 0$ für ein $n$. Nach Lemma 10.19 ist $\alpha_n \in R_D$ für ein $D$, und nach Lemma 10.11 ist dann $\alpha_m \in R_D$ für alle $m \geq n$. Da $R_D$ endlich ist, muss die Folge der $\alpha_m$ und damit auch die Folge der $a_m$ periodisch werden. Genauer gilt, dass die Folgen der Reste $\alpha_m$ und die der Zahlen $a_m$ genau ab $m = n$ periodisch werden, wenn $n$ der kleinste Index ist, für den $\alpha_n > 1$ und $-1 < \bar{\alpha}_n < 0$ gilt. ❑

**BSP** Kettenbruchentwicklung einer qu. Irrationalität

10.21. **Beispiel.** Wir bestimmen die Kettenbruchentwicklung von $\alpha = -\frac{\sqrt{21}}{5}$. Wir finden (unter Beachtung von $\lfloor 5\sqrt{21} \rfloor = \lfloor \sqrt{525} \rfloor = 22$):

| $n$ | $\alpha_n$ | $a_n$ |
|---|---|---|
| 0 | $-\frac{\sqrt{21}}{5}$ | $-1$ |
| 1 | $\frac{5\sqrt{21}+25}{4}$ | 11 |
| 2 | $\frac{5\sqrt{21}+19}{41}$ | 1 |
| 3 | $5\sqrt{21}+22$ | 44 |
| 4 | $\frac{5\sqrt{21}+22}{41}$ | 1 |
| 5 | $\frac{5\sqrt{21}+19}{4}$ | 10 |
| 6 | $\frac{5\sqrt{21}+21}{21}$ | 2 |
| 7 | $\frac{5\sqrt{21}+21}{4}$ | 10 |
| 8 | $\frac{5\sqrt{21}+19}{41}$ | 1 |
| $\vdots$ | $\vdots$ | $\vdots$ |

Hier ist $\alpha_2 \in R_{5^2 \cdot 21}$, und die Kettenbruchentwicklung ist

$$[-1; 11, \overline{1, 44, 1, 10, 2, 10}]$$

(wobei der periodische Teil durch den Überstrich gekennzeichnet ist). ♣

Zum Abschluss der Diskussion der Kettenbruchentwicklung von $\sqrt{d}$ wollen wir noch zeigen, dass die Periode symmetrisch ist.

**LEMMA** Umkehrung der Periode

10.22. **Lemma.** *Sei $\alpha \in R_d$ für ein $d > 0$, $d$ kein Quadrat, und sei*

$$\alpha = [\overline{a_0, a_1, \ldots, a_{n-1}}]$$

*die (rein periodische) Kettenbruchentwicklung von $\alpha$. Dann ist*

$$[\overline{a_{n-1}, a_{n-2}, \ldots, a_1, a_0}] = -\frac{1}{\bar{\alpha}}\,.$$

*Beweis.* Seien $p_k/q_k$ die Näherungsbrüche der Kettenbruchentwicklung von $\alpha$, und seien $p'_k/q'_k$ die Näherungsbrüche von $[\overline{a_{n-1}, \ldots, a_1, a_0}]$. Die Rekursionen für die $p_k$, $q_k$, $p'_k$, $q'_k$ lassen sich recht elegant durch Matrizen beschreiben:

$$\begin{pmatrix} p_k & q_k \\ p_{k-1} & q_{k-1} \end{pmatrix} = \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_{k-1} & q_{k-1} \\ p_{k-2} & q_{k-2} \end{pmatrix}$$

und

$$\begin{pmatrix} p'_k & q'_k \\ p'_{k-1} & q'_{k-1} \end{pmatrix} = \begin{pmatrix} a_{n-1-k} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p'_{k-1} & q'_{k-1} \\ p'_{k-2} & q'_{k-2} \end{pmatrix}$$

(für $0 \le k \le n-1$). Es folgt

$$\begin{pmatrix} p_{n-1} & q_{n-1} \\ p_{n-2} & q_{n-2} \end{pmatrix} = \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix},$$

also

$$\begin{pmatrix} p'_{n-1} & q'_{n-1} \\ p'_{n-2} & q'_{n-2} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_{n-1} & q_{n-1} \\ p_{n-2} & q_{n-2} \end{pmatrix}^\top .$$

Sei $\alpha'$ der Wert von $[\overline{a_{n-1}, \dots, a_1, a_0}]$. Dann gilt

$$\alpha = \frac{p_{n-1}\alpha + p_{n-2}}{q_{n-1}\alpha + q_{n-2}} \qquad \text{und} \qquad \alpha' = \frac{p'_{n-1}\alpha' + p'_{n-2}}{q'_{n-1}\alpha' + q'_{n-2}} = \frac{p_{n-1}\alpha' + q_{n-1}}{p_{n-2}\alpha' + q_{n-2}}\,.$$

$\alpha$ ist also eine Nullstelle des Polynoms $f(x) = q_{n-1}x^2 + (q_{n-2} - p_{n-1})x - p_{n-2}$, und $\alpha'$ ist eine Nullstelle von $p_{n-2}x^2 + (q_{n-2} - p_{n-1})x - q_{n-1} = -x^2 f(-1/x)$. Es muss also $\alpha' = -1/\alpha$ oder $-1/\bar{\alpha}$ sein (denn $\alpha$ und $\bar{\alpha}$ sind die Nullstellen von $f$). Da $-1/\alpha < 0$ ist, kommt nur die zweite Möglichkeit infrage. ❑

10.23. **Satz.** *Sei $d > 0$ und kein Quadrat. Sei*

$$\sqrt{d} = [a_0; \overline{a_1, a_2, \dots, a_{n-1}, 2a_0}]$$

*die Kettenbruchentwicklung von $\sqrt{d}$. Dann gilt $a_k = a_{n-k}$ für alle $1 \le k < n$.*

**SATZ**
Symmetrie der Kettenbruch-entwicklung von $\sqrt{d}$

*Beweis.* $\alpha = \sqrt{d} + a_0 = \sqrt{d} + \lfloor\sqrt{d}\rfloor$ ist reduziert, und $\alpha = [\overline{2a_0, a_1, \dots, a_{n-1}}]$. Nach Lemma 10.22 ist

$$[\overline{a_{n-1}, \dots, a_1, 2a_0}] = -\frac{1}{\bar{\alpha}} = \frac{1}{\sqrt{d} - a_0}\,.$$

Also ist

$$[\overline{2a_0, a_{n-1}, a_{n-2}, \dots, a_1}] = 2a_0 + (\sqrt{d} - a_0) = \sqrt{d} + a_0 = \alpha = [\overline{2a_0, a_1, a_2, \dots, a_{n-1}}]\,,$$

woraus sich die behauptete Symmetrie ergibt. ❑

10.24. **Beispiel.** Eine Quadratwurzel, die eine etwas längere Periode liefert, ist etwa

$$\sqrt{163} = [12; \overline{1, 3, 3, 2, 1, 1, 7, 1, 11, 1, 7, 1, 1, 2, 3, 3, 1, 24}]\,. \qquad \clubsuit$$

**BSP**
lange Periode

## 11. Verallgemeinerte Pellsche Gleichung

Wir wollen nun die Pellsche Gleichung etwas verallgemeinern und Gleichungen der Form

$$x^2 - dy^2 = n$$

betrachten. Hier ist $d$ wie immer positiv und kein Quadrat, und $n$ ist irgendeine von null verschiedene ganze Zahl.

11.1. **Definition.** Wir setzen **DEF** $S_d(n)$

$$S_d(n) = \{(x, y) \in \mathbb{Z}^2 \mid x^2 - dy^2 = n\}\,.$$

(In unserer bisherigen Schreibweise ist also $S_d = S_d(1)$ und $T_d = S_d(1) \cup S_d(-1)$.) ♢

11.2. **Satz.** *Die Verknüpfung $(x, y) * (x', y') = (xx' + dyy', xy' + yx')$ liefert eine Operation der Gruppe $S_d = S_d(1)$ auf $S_d(n)$. Die Menge $S_d(n)$ zerfällt in endlich viele Bahnen unter dieser Operation.* **SATZ** Operation von $S_d$ auf $S_d(n)$

*Beweis.* Die Verknüpfung $*$ ist assoziativ und definiert allgemeiner eine Abbildung $S_d(n_1) \times S_d(n_2) \to S_d(n_1 n_2)$, die für $n_1 = n_2 = 1$ gerade die Gruppenstruktur von $S_d = S_d(1)$ liefert. Daraus folgt die erste Behauptung (unter Beachtung von $(1, 0) * (x, y) = (x, y)$).

Zum Beweis der zweiten Aussage betrachten wir wieder die Abbildung $\phi\colon \mathbb{Z}^2 \to \mathbb{R}$, $(x, y) \mapsto x + y\sqrt{d}$. Sei $(x_1, y_1)$ die Grundlösung der Gleichung $x^2 - dy^2 = 1$, und sei $\varepsilon = \phi(x_1, y_1) > 1$. Sei $(x, y) \in S_d(n)$. Aus $\phi\big((x, y) * (x', y')\big) = \phi(x, y)\phi(x', y')$ folgt, dass es ein eindeutig bestimmtes $m \in \mathbb{Z}$ gibt mit

$$\sqrt{\frac{|n|}{\varepsilon}} \le \varepsilon^m |\phi(x, y)| < \sqrt{|n|\varepsilon}\,.$$

Dann ist $(x', y') = (x_1, y_1)^{*m} * (x, y) \in S_d(n)$ in derselben Bahn wie $(x, y)$. Durch „Multiplikation“ mit $(-1, 0) \in S_d$ können wir noch erreichen, dass $\phi(x', y') > 0$ ist. Weiterhin gilt

$$\frac{1}{\phi(x', y')} = \frac{1}{x' + y'\sqrt{d}} = \frac{x' - y'\sqrt{d}}{n}\,.$$

Es folgt

$$y' = \frac{\phi(x', y') - n\phi(x', y')^{-1}}{2\sqrt{d}} \qquad \Longrightarrow \qquad |y'| < \frac{\sqrt{|n|\varepsilon} + \sqrt{|n|\varepsilon}}{2\sqrt{d}} = \sqrt{\frac{|n|\varepsilon}{d}}\,.$$

Analog gilt $|x'| < \sqrt{|n|\varepsilon}$. Damit gibt es nur endlich viele Möglichkeiten für $x'$ und $y'$. Also gibt es auch nur endlich viele Bahnen. ❑

Der Beweis liefert ein Lösungsverfahren, das allerdings nicht sehr effizient ist, wenn $\sqrt{|n|\varepsilon/d}$ groß wird.

**BSP** $x^2-5y^2=-4$

11.3. **Beispiel.** Wir bestimmen $S_5(-4)$, also die Lösungen von

$$x^2 - 5y^2 = -4\,.$$

Die Grundlösung $(x_1, y_1)$ von $x^2 - 5y^2 = 1$ ist $(9, 4)$, also ist

$$\sqrt{\frac{|n|\varepsilon}{d}} = \sqrt{4\frac{9+4\sqrt{5}}{5}} \approx 3{,}79\,;$$

damit ist $|y| \le 3$ für einen Repräsentanten jeder Bahn. Wir probieren die verschiedenen Möglichkeiten aus:

$$\begin{aligned} y = 0 &\implies \text{keine Lösung} \\ y = \pm 1 &\implies x = \pm 1 \\ y = \pm 2 &\implies x = \pm 4 \\ y = \pm 3 &\implies \text{keine Lösung} \end{aligned}$$

und berechnen $\phi(x, y)$:

| $x$ | $-1$ | $-1$ | $1$ | $1$ | $-4$ | $-4$ | $4$ | $4$ |
|---|---|---|---|---|---|---|---|---|
| $y$ | $-1$ | $1$ | $-1$ | $1$ | $-2$ | $2$ | $-2$ | $2$ |
| $\phi(x,y)$ | $-3{,}24$ | $1{,}24$ | $-1{,}24$ | $3{,}24$ | $-8{,}47$ | $0{,}47$ | $-0{,}47$ | $8{,}47$ |

Es ist $\sqrt{|n|/\varepsilon} \approx 0{,}47$ und $\sqrt{|n|\varepsilon} \approx 8{,}47$. Alle positiven Werte sind im richtigen Bereich, mit Ausnahme von $(x, y) = (4, 2)$; hier ist $|\phi(x, y)| = \sqrt{|n|\varepsilon}$. Es gilt nämlich $\varepsilon = (2+\sqrt{5})^2$, also $4+2\sqrt{5} = \sqrt{4\varepsilon}$. Es folgt, dass es genau drei Bahnen von Lösungen gibt, die von $(x, y) = (-1, 1)$, $(1, 1)$ und $(-4, 2)$ repräsentiert werden. (Und $(4, 2) = (9, 4) * (-4, 2)$ ist in derselben Bahn wie $(-4, 2)$.) ♣

Wenn es ganzzahlige Lösungen von $x^2 - dy^2 = n$ gibt, dann sicher auch rationale. Bevor man also versucht, eine Lösung zu finden, sollte man prüfen, ob die ternäre quadratische Form $x^2 - dy^2 - nz^2$ nichttriviale Nullstellen hat. Zum Beispiel kann es keine ganzzahligen Lösungen von $x^2 - 5y^2 = \pm 2$ oder $\pm 3$ geben, da die rechten Seiten keine quadratischen Reste mod 5 sind.

Für kleine $n$ findet man die relevanten Lösungen gewissermaßen „auf dem Weg“ zur Berechnung der Grundlösung der zugehörigen Pellschen Gleichung. Wir zeigen zunächst ein paar Hilfsaussagen. Dafür erweitern wir die Definition der Verknüpfung „$*$“ auf Paare von rationalen Zahlen. Dann ist

$$\mathbb{Q}^2 \longrightarrow \mathbb{Q}(\sqrt{d})\,, (x, y) \longmapsto x + y\sqrt{d}$$

ein Isomorphismus der Monoide $(\mathbb{Q}^2, *)$ und $(\mathbb{Q}(\sqrt{d}), \cdot)$.

Insbesondere gilt für $(x, y), (x', y'), (u, v) \in \mathbb{Q}^2$ mit $(u, v) \ne (0, 0)$ die Äquivalenz

$$(x, y) = (x', y') \iff (x, y) * (u, v) = (x', y') * (u, v)\,.$$

**LEMMA** Rekursion für $p_n$, $q_n$ Symmetrie von $u_n$, $v_n$

11.4. **Lemma.** *Sei $d > 0$ kein Quadrat, sei $[a_0; \overline{a_1, \ldots, a_{m-1}, 2a_0}]$ die Kettenbruchentwicklung von $\sqrt{d}$ mit (minimaler) Periode $m$ und seien $p_n$, $q_n$, $u_n$, $v_n$ die zugehörigen Größen. Dann gilt:*

(1) *Für alle $n \ge 0$ ist* $\quad (p_n, q_n) = \Big(\dfrac{u_{n+1}}{v_n}, \dfrac{1}{v_n}\Big) * (p_{n-1}, q_{n-1})\,.$

(2) *Für alle $1 \le k \le m$ ist $u_{m+1-k} = u_k$ und für alle $0 \le k \le m$ ist $v_{m-k} = v_k$.*

*Beweis.*

(1) Nach Lemma 10.9, (2) und Lemma 10.2, (1) gilt

$$(p_n, q_n) * (p_{n-1}, -q_{n-1}) = (-1)^n (u_{n+1}, 1)$$

und

$$(p_{n-1}, q_{n-1}) * (p_{n-1}, -q_{n-1}) = (-1)^n (v_n, 0)\,.$$

Die behauptete Gleichheit ist äquivalent zu der, die wir durch Verknüpfen mit $(p_{n-1}, -q_{n-1})$ bekommen; diese folgt aus den obigen Relationen.

(2) Es ist $\alpha_m = \sqrt{d} + \lfloor \sqrt{d} \rfloor$ und somit $v_m = 1 = v_0$. Sei jetzt $1 \le k \le m$. Wir schreiben $\sigma(\alpha)$ für das konjugierte Element $\bar{\alpha}$. Mit Lemma 10.22 und Satz 10.23 haben wir dann

$$\begin{aligned}
\frac{\sqrt{d} + u_k}{v_k} = \alpha_k &= \overline{[a_k, \ldots, a_{m-1}, 2a_0, a_1, \ldots, a_{k-1}]} \\
&= -\frac{1}{\sigma\big(\overline{[a_{k-1}, \ldots, a_1, 2a_0, a_{m-1}, \ldots, a_k]}\big)} \\
&= -\frac{1}{\sigma\big(\overline{[a_{m+1-k}, \ldots, a_{m-1}, 2a_0, a_1, \ldots, a_{m-k}]}\big)} \\
&= -\frac{1}{\sigma(\alpha_{m+1-k})} = \frac{v_{m+1-k}}{\sqrt{d} - u_{m+1-k}} = \frac{\sqrt{d} + u_{m+1-k}}{v_{m-k}}\,,
\end{aligned}$$

denn $v_{m-k} v_{m+1-k} = d - u_{m+1-k}^2$. Die Aussage folgt durch Koeffizientenvergleich. ❑

**LEMMA** Gruppenoperation und Näherungsbrüche

11.5. **Lemma.** *Sei $d > 0$ und kein Quadrat, und seien $p_k/q_k$ die Näherungsbrüche der Kettenbruchentwicklung von $\sqrt{d}$; diese habe die minimale Periode $m$. Dann gilt für alle $k \ge -1$:*

$$(p_{k+m}, q_{k+m}) = (p_{m-1}, q_{m-1}) * (p_k, q_k)\,.$$

*Beweis.* Wir setzen $U_k = (u_{k+1}/v_k, 1/v_k) \in \mathbb{Q}^2$. Nach Lemma 11.4 (1) gilt dann

$$(p_k, q_k) = U_k * (p_{k-1}, q_{k-1}) \qquad \text{für } k \ge 0$$

und somit für $k \ge -1$ ($(p_{-1}, q_{-1}) = (1, 0)$ ist das neutrale Element von „$*$“)

$$(p_k, q_k) = \mathop{\ast}_{n=0}^{k} U_n\,.$$

Außerdem gilt $U_{k+m} = U_k$ für alle $k \ge 0$. Es folgt für $k \ge -1$

$$\begin{aligned}
(p_{k+m}, q_{k+m}) &= \mathop{\ast}_{n=0}^{k+m} U_n = \mathop{\ast}_{n=k+1}^{k+m} U_k * \mathop{\ast}_{n=0}^{k} U_k \\
&= \mathop{\ast}_{n=0}^{m-1} U_n * (p_k, q_k) = (p_{m-1}, q_{m-1}) * (p_k, q_k)\,.
\end{aligned}$$ ❑

Wir können die eben bewiesene Aussage dazu verwenden, die Folge der $(p_k, q_k)$ „nach links“ fortzusetzen. Da sich für $k = -2$ etwas anderes ergibt als die bisherige Festlegung $(p_{-2}, q_{-2}) = (0, 1)$, ändern wir die Bezeichnung ein wenig.

11.6. **Definition.** In der Situation des vorangehenden Lemmas definieren wir **DEF** $p_k^*, q_k^*$

$$(p_k^*, q_k^*) = (p_k, q_k) \quad \text{für } k \geq -1$$

und rekursiv

$$(p_k^*, q_k^*) = \big((-1)^m p_{m-1}, (-1)^{m-1} q_{m-1}\big) * (p_{k+m}^*, q_{k+m}^*) \quad \text{für } k < -1.$$

Wegen

$$\big((-1)^m p_{m-1}, (-1)^{m-1} q_{m-1}\big) * (p_{m-1}, q_{m-1}) = \big((-1)^m (p_{m-1}^2 - d q_{m-1}^2), 0\big) = (1, 0)$$

gilt dann $(p_{k+m}^*, q_{k+m}^*) = (p_{m-1}, q_{m-1}) * (p_k^*, q_k^*)$ für alle $k \in \mathbb{Z}$. ♢

11.7. **Lemma.** *In der Situation von Definition 11.6 gilt* **LEMMA** Symmetrie von $p_k^*, q_k^*$

$$(p_{-1-k}^*, q_{-1-k}^*) = \big((-1)^k p_{-1+k}^*, (-1)^{k-1} q_{-1+k}^*\big)$$

*für alle $k \in \mathbb{Z}$.*

*Beweis.* Wir definieren $U_k$ für $k \geq 0$ wie im Beweis von Lemma 11.5 und für $k < 0$ rekursiv als $U_{k+m}$. Dann ist $(U_k)_{k\in\mathbb{Z}}$ periodisch mit Periode $m$, und es gilt

$$(p_k^*, q_k^*) = U_k * (p_{k-1}^*, q_{k-1}^*) \qquad \text{für alle } k \in \mathbb{Z}.$$

Für $k \geq 0$ hatten wir das bereits gesehen. Für $k < 0$ ergibt sich die Relation induktiv via

$$\begin{aligned}(p_k^*, q_k^*) &= ((-1)^m p_{m-1}, (-1)^{m-1} q_{m-1}) * (p_{k+m}^*, q_{k+m}^*) \\ &\overset{\text{IV}}{=} ((-1)^m p_{m-1}, (-1)^{m-1} q_{m-1}) * U_{k+m} * (p_{k+m-1}^*, q_{k+m-1}^*) \\ &= U_k * ((-1)^m p_{m-1}, (-1)^{m-1} q_{m-1}) * (p_{k+m-1}^*, q_{k+m-1}^*) = U_k * (p_{k-1}^*, q_{k-1}^*)\,.\end{aligned}$$

Es genügt, die Behauptung für $k \geq 0$ zu beweisen (die Aussage für $-k$ ist zu der für $k$ äquivalent). Für $k = 0$ ist sie klar (denn $q_{-1} = 0$). Sei also $k > 0$. Die Behauptung ist dann äquivalent zu

$$\begin{aligned}(p_{-1-k}^*, q_{-1-k}^*) * (p_{-1+k}, q_{-1+k}) &= ((-1)^k p_{-1+k}, (-1)^{k-1} q_{-1+k}) * (p_{-1+k}, q_{-1+k}) \\ &= ((-1)^k (p_{-1+k}^2 - d q_{-1+k}^2), 0) = (v_k, 0)\,.\end{aligned}$$

Wir haben induktiv

$$\begin{aligned}(p_{-1-k}^*, q_{-1-k}^*) * (p_{-1+k}, q_{-1+k}) &= (U_{-k})^{*-1} * (p_{-k}^*, q_{-k}^*) * U_{k-1} * (p_{k-2}, q_{k-2}) \\ &\overset{\text{IV}}{=} (U_{-k})^{*-1} * U_{k-1} * (v_{k-1}, 0)\,,\end{aligned}$$

sodass wir noch Folgendes zeigen müssen:

$$v_k U_{-k} = v_{k-1} U_{k-1} = (u_k, 1)\,.$$

Wegen der Periodizität können wir $1 \leq k \leq m$ annehmen; dann ist die linke Seite

$$v_k U_{m-k} = \frac{v_k}{v_{m-k}} (u_{m-k+1}, 1)\,,$$

und das Resultat folgt aus Lemma 11.4 (2). ❑

Da $p_k$ und $q_k$ immer teilerfremd sind, können wir in dieser Form nur Lösungen mit teilerfremden $x$ und $y$ erwarten.

11.8. **Definition.** Eine Lösung $(x, y) \in S_d(n)$ heißt *primitiv*, wenn $x \perp y$ gilt. Wir schreiben $S_d^\perp(n)$ für die Menge der primitiven Lösungen von $x^2 - dy^2 = n$.

**DEF** primitive Lösung

Man beachte, dass der ggT von $x$ und $y$ auf Bahnen konstant ist. Daher können wir eine $S_d$-Bahn in $S_d(n)$ *primitiv* nennen, wenn ihre Elemente primitiv sind, und $S_d^\perp(n)$ ist Vereinigung von Bahnen in $S_d(n)$.

Offensichtlich gilt

$$S_d(n) = \bigcup_{a>0, a^2|n} \{(ax, ay) \mid (x, y) \in S_d^\perp(n/a^2)\},$$

und die Vereinigung ist disjunkt. ♢

Jetzt können wir die oben angedeutete Aussage formulieren und beweisen, dass Lösungen für "kleines" $n$ aus der Kettenbruchentwicklung gewonnen werden können.

11.9. **Satz.** *Sei $d > 0$ und kein Quadrat und sei $|n| < \sqrt{d}$. Seien weiter $p_k/q_k$ die Näherungsbrüche der Kettenbruchentwicklung von $\sqrt{d}$; diese habe die minimale Periode $m$. Sei $m' = \mathrm{kgV}(2, m)$. Dann ist die Menge der $(p_k, q_k)$ mit*

**SATZ** Lösungen für $|n| < \sqrt{d}$

$$p_k^2 - dq_k^2 = n \qquad \text{und} \qquad -1 \le k < m' - 1$$

*ein vollständiges Repräsentantensystem der Bahnen von $S_d^\perp(n)$ unter $S_d$.*

*Beweis.* Zunächst stellen wir fest, dass $(x_1, y_1) = (p_{m'-1}, q_{m'-1})$ die Grundlösung von $x^2 - dy^2 = 1$ ist (vergleiche Satz 10.6 und die nachfolgenden Ergebnisse.) Nach Definition 11.6 gilt dann (für $m' = 2m$ beachte $(x_1, y_1) = (p_{m-1}, q_{m-1})^{*2}$)

$$(p^*_{k+m'}, q^*_{k+m'}) = (x_1, y_1) * (p^*_k, q^*_k) \quad \text{für alle } k \in \mathbb{Z}.$$

Es folgt, dass jede Bahn von $S_d^\perp(n)$, die ein Paar der Form $(p^*_k, q^*_k)$ enthält, einen eindeutigen Repräsentanten mit $-1 \le k < m' - 1$ besitzt. Es bleibt zu zeigen, dass tatsächlich alle Bahnen auftreten. Dazu verwenden wir wieder Lemma 10.5. Sei zunächst $(x, y) \in S_d^\perp(n)$ mit $x, y > 0$. Aus $n > -\sqrt{d} > -2\sqrt{d} + 1/y^2$ folgt

$$x^2 = dy^2 + n > dy^2 - 2\sqrt{d} + \frac{1}{y^2} = \left(\sqrt{d}y - \frac{1}{y}\right)^2 \quad \Longrightarrow \quad x > \sqrt{d}y - \frac{1}{y}$$

und damit

$$\left|\sqrt{d} - \frac{x}{y}\right| = \frac{|n|}{y(x + y\sqrt{d})} < \frac{|n|}{2\sqrt{d}y^2 - 1} < \frac{1}{2y^2}$$

für $y$ hinreichend groß. Wir können $(x, y)$ durch $(x', y') = (x_1, y_1)^{*N} * (x, y)$ für $N \gg 0$ ersetzen, dann sind $x', y' > 0$ und $y'$ wird beliebig groß. Nach Lemma 10.5 ist dann $(x', y') = (p_k, q_k)$ für ein geeignetes $k$, also hat die Bahn von $(x, y)$ einen Vertreter der gewünschten Form.

Sei nun $x > 0$ und $y < 0$. Nach dem schon Bewiesenen gibt es ein $k$, sodass $(p_k, q_k)$ in der Bahn von $(x, -y)$ liegt; dann liegt $(p_k, -q_k)$ in der Bahn von $(x, y)$. Nun gilt aber nach Lemma 11.7, dass $(p^*_{-k-2}, q^*_{-k-2}) = \pm(p_k, -q_k)$, also in der relevanten Bahn ist.

Ist $x > 0$ und $y = 0$, dann muss $x = 1$ sein, also ist $(x, y) = (p_{-1}, q_{-1})$.

Im Fall $x < 0$ können wir $(x, y)$ durch $(-x, -y) = (-1, 0) * (x, y)$ ersetzen; damit ist die Aussage auf den Fall $x > 0$ zurückgeführt. ❑

**BSP** Lösungen für kleines $n$

11.10. **Beispiel.** Sei $d = 163$. Wir haben $m' = m = 18$ und erhalten die folgende Tabelle:

| $k$ | $-1$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_k^2 - dq_k^2$ | 1 | $-19$ | 6 | $-7$ | 9 | $-11$ | 14 | $-3$ | 21 | $-2$ | 21 | $-3$ | 14 | $-11$ | 9 | $-7$ | 6 | $-19$ |

Für alle $n$ mit $0 < |n| \leq 12$ finden wir Repräsentanten aller Bahnen von $S_{163}^{\perp}(n)$ unter $S_{163}$ in der Form $(p_k, q_k)$ mit $k$ im Bereich dieser Tabelle. Insbesondere gibt es keine primitiven Lösungen für Werte $|n| \leq 12$, die nicht in der Tabelle vorkommen.

Für größere $n$ muss das nicht mehr gelten, unabhängig davon, ob sie in der Tabelle auftreten oder nicht. Es ist etwa $64^2 - 163 \cdot 5^2 = 21$, aber $(64, 5)$ ist nicht von der Form $(p_k, q_k)$. Ähnliches gilt für $(x, y) = (217, 17)$ mit $x^2 - 163y^2 = -18$.

In jedem Fall gilt für jedes $n$, das in der Tabelle vorkommt, $|n| < 2\sqrt{d}$ (denn $|n|$ ist der Nenner $v$ eines Elements $\alpha \in R_d$, vergleiche den Beweis von Lemma 10.11). ♣

Zum Abschluss dieses Kapitels möchte ich noch eine Anwendung der Theorie vorführen.

**SATZ** Anwendung

11.11. **Satz.** *Seien $F_0 = 0$, $F_1 = 1$, $F_{n+2} = F_{n+1} + F_n$ die Fibonacci-Zahlen. Dann sind die Lösungen von*

$$\binom{n+1}{k} = \binom{n}{k+1} \qquad \text{mit } n > k \geq 0$$

*gegeben durch $(n, k) = (F_{2j+2}F_{2j+3} - 1, F_{2j}F_{2j+3})$ für $j = 0, 1, 2, \ldots$.*

*Beweis.* Wir formen die Gleichung erst einmal um:

$$\begin{aligned} \binom{n+1}{k} = \binom{n}{k+1} &\iff \frac{(n+1)!}{k!(n-k+1)!} = \frac{n!}{(k+1)!(n-k-1)!} \\ &\iff (n+1)(k+1) = (n-k+1)(n-k) \\ &\iff n^2 - 3nk + k^2 - 2k - 1 = 0 \end{aligned}$$

Wir multiplizieren mit 4, dann ergibt quadratische Ergänzung

$$(2n - 3k)^2 - 5k^2 - 8k - 4 = 0\,.$$

Jetzt multiplizieren wir mit 5 und ergänzen wieder; das ergibt

$$5(2n - 3k)^2 - (5k + 4)^2 - 4 = 0$$

oder

$$(5k + 4)^2 - 5(2n - 3k)^2 = -4\,.$$

Wir erkennen die Gleichung $x^2 - 5y^2 = -4$ aus Beispiel 11.3. Wir suchen also alle Lösungen $(x, y) \in S_5(-4)$, sodass

$$k = \frac{x - 4}{5} \qquad \text{und} \qquad n = \frac{y + 3k}{2}$$

ganzzahlig sind. Die Bedingungen dafür sind $x \equiv 4 \bmod 5$ und $y \equiv k \equiv x \bmod 2$. Die letzte Bedingung $y \equiv x \bmod 2$ ist immer erfüllt.

Wir hatten gesehen, dass $S_5(-4)$ unter der Operation von $S_5$ in drei Bahnen zerfällt, die von $(1, 1)$, $(-1, 1)$ und $(-4, 2)$ repräsentiert werden. Für das Folgende ist es günstiger, die Zerlegung in Bahnen unter der kleineren Gruppe $S_5^+ = \langle (9, 4) \rangle$ zu betrachten, die von der Grundlösung $(9, 4)$ erzeugt wird. Jede $S_5$-Bahn zerfällt in zwei $S_5^+$-Bahnen; unsere drei $S_5$-Bahnen zerfallen in sechs $S_5^+$-Bahnen mit Repräsentanten $(1, 1)$, $(-1, -1)$, $(-1, 1)$, $(1, -1)$, $(-4, 2)$ und $(4, -2)$.

Um herauszufinden, welche Lösungen die Bedingung $x \equiv 4 \bmod 5$ erfüllen, betrachten wir die Lösungen modulo 5. Unter Verknüpfung mit der Grundlösung $(9,4) \equiv (-1,-1) \bmod 5$ erhalten wir

$$(1,*) \mapsto (-1,*) \mapsto (1,*)\,,$$

also erfüllt jede zweite Lösung in jeder Bahn die Bedingung. Die gesuchte Menge ist also Vereinigung von sechs Bahnen unter $\langle (9,4)^{*2} \rangle$:

$$\{(x,y) \in S_5(-4) \mid x \equiv 4 \bmod 5\} = \{(9,4)^{*2m} * (x',y') \mid m \in \mathbb{Z}, (x',y') \in U\}$$

mit

$$U = \{(-11,5), (-1,-1), (-1,1), (-11,-5), (4,2), (4,-2)\}\,.$$

(Die Repräsentanten sind entweder die oben gegebenen ursprünglichen Repräsentanten $(x,y)$ oder $(9,\pm 4) * (x,y)$.)

Man führt nun am besten den Goldenen Schnitt $\varphi = (1+\sqrt{5})/2$ ein. Wir schreiben $\bar\varphi = (1-\sqrt{5})/2 = -1/\varphi$. Man berechnet folgende Tabelle von Potenzen von $\varphi$:

| $\ell$ | $-5$ | $-3$ | $-1$ | $1$ | $3$ | $5$ |
|---|---|---|---|---|---|---|
| $2\varphi^\ell$ | $-11+5\sqrt{5}$ | $-4+2\sqrt{5}$ | $-1+\sqrt{5}$ | $1+\sqrt{5}$ | $4+2\sqrt{5}$ | $11+5\sqrt{5}$ |

Außerdem ist $\varphi^6 = 9+4\sqrt{5}$. Es folgt (mit $\bar\varphi = -\varphi^{-1}$)

$$\begin{aligned}\{(x,y) \in S_5(-4) \mid x \equiv 4 \bmod 5\}& \\ = \{(x,y) \mid x \pm y\sqrt{5} = 2\varphi^{4m-1} \text{ für ein } m \in \mathbb{Z}\}\,.&\end{aligned}$$

Dann ist

$$x = \varphi^{4m-1} + \bar\varphi^{4m-1} = L_{4m-1}$$

(dabei sind $L_0 = 2$, $L_1 = 1$, $L_{m+2} = L_{m+1} + L_m$ die *Lucas-Zahlen*; es gilt allgemein $L_m = \varphi^m + \bar\varphi^m$) und

$$y = \pm\frac{1}{\sqrt{5}}(\varphi^{4m-1} - \bar\varphi^{4m-1}) = \pm F_{4m-1}\,.$$

Es ist $L_{-m} = (-1)^m L_m$, also muss oben $m \geq 1$ sein, damit $x = 5k+4 > 0$ ist. Nun beachten wir, dass für $L_m$ und $F_m$ allgemein folgende Relationen gelten:

$$\begin{aligned}5F_mF_{m+l} &= (\varphi^m - \bar\varphi^m)(\varphi^{m+l} - \bar\varphi^{m+l}) = \varphi^{2m+l} + \bar\varphi^{2m+l} - (-1)^m(\varphi^l + \bar\varphi^l) \\ &= L_{2m+l} - (-1)^m L_l \\ 2L_{m\pm 2} &= 2\varphi^{m\pm 2} + 2\bar\varphi^{m\pm 2} = (3 \pm \sqrt{5})\varphi^m + (3 \mp \sqrt{5})\bar\varphi^m \\ &= 3(\varphi^m + \bar\varphi^m) \pm \sqrt{5}(\varphi^m - \bar\varphi^m) = 3L_m \pm 5F_m\end{aligned}$$

Für die ursprünglichen Variablen $k$ und $n$ folgt also mit $m = j+1$, $j \geq 0$:

$$\begin{aligned}k &= \frac{x-4}{5} = \frac{L_{4j+3}-4}{5} = \frac{L_{4j+3}-L_3}{5} = F_{2j}F_{2j+3} \\ n &= \frac{y+3k}{2} = \frac{5F_{4j+3}+3L_{4j+3}-12}{10} = \frac{L_{4j+5}-L_1}{5} - 1 = F_{2j+2}F_{2j+3} - 1\end{aligned}$$

oder

$$= \frac{-5F_{4j+3}+3L_{4j+3}-12}{10} = \frac{L_{4j+1}-L_1}{5} - 1 = F_{2j}F_{2j+1} - 1$$

Bei der zweiten Möglichkeit für $n$ gilt allerdings $n < k$, also kommt sie nicht infrage. ❑

Satz 11.11 gibt eine unendliche Folge von Lösungen der Gleichung

$$\binom{n}{k} = \binom{n'}{k'}$$

an, die fast alle den zusätzlichen Bedingungen $1 < k \le n/2$, $1 < k' \le n'/2$, $n' < n$ genügen (die man sinnvollerweise stellt, um triviale Lösungen auszuschließen). Die ersten dieser Lösungen sind

$$\binom{2}{0} = \binom{1}{1}, \quad \binom{15}{5} = \binom{14}{6}, \quad \binom{104}{39} = \binom{103}{40}, \quad \binom{714}{272} = \binom{713}{273}, \quad \dots.$$

Es sind nur noch einzelne weitere Lösungen bekannt, nämlich

$$\binom{16}{2} = \binom{10}{3}, \quad \binom{56}{2} = \binom{22}{3}, \quad \binom{120}{2} = \binom{36}{3}, \quad \binom{21}{2} = \binom{10}{4},$$
$$\binom{78}{2} = \binom{15}{5} \left( = \binom{14}{6} \right), \qquad \binom{153}{2} = \binom{19}{5} \quad \text{und} \quad \binom{221}{2} = \binom{17}{8}.$$

Für die Paare

$$(k, k') = (2,3),\ (2,4),\ (2,5),\ (2,6),\ (2,8),\ (3,4),\ (3,6),\ (4,6),\ (4,8)$$

ist die obige Gleichung vollständig gelöst.[11] Ob es noch weitere Lösungen gibt, ist eine offene Frage.

[11] R.J. Stroeker und B.M.M. de Weger: *Elliptic binomial Diophantine equations*, Math. Comp. **68** (1999), 1257–1281; Y. Bugeaud, M. Mignotte, S. Siksek, M. Stoll, S. Tengely: *Integral points on hyperelliptic curves*, Algebra Number Theory **2** (2008), 859–885.

## 12. Verwendung von $p$-adischen Potenzreihen

Manche diophantische Gleichungen lassen sich mithilfe von $p$-adischen Potenzreihen lösen. Folgendes Resultat spielt dabei eine wichtige Rolle:

**SATZ**
Satz von Straßmann

12.1. **Satz.** *Es sei*

$$f = \sum_{n=0}^{\infty} a_n x^n \in \mathbb{Q}_p[\![x]\!]$$

*eine Potenzreihe mit Koeffizienten in $\mathbb{Q}_p$, die nicht die Nullreihe ist. Es gelte*

$$v_p(a_n) \to \infty \qquad \text{für } n \to \infty.$$

*Dann konvergiert $f$ auf $\mathbb{Z}_p$. Außerdem existiert $m = \min\{v_p(a_n) \mid n \in \mathbb{Z}_{\geq 0}\}$; sei $N = N(f) = \max\{n \in \mathbb{Z}_{\geq 0} \mid v_p(a_n) = m\}$. Dann hat $f$ höchstens $N$ Nullstellen in $\mathbb{Z}_p$ (mit Vielfachheit gezählt).*

*Beweis.* Dass $f$ auf $\mathbb{Z}_p$ konvergiert, folgt aus Lemma 7.10 (1).

Durch Skalieren von $f$ mit $a_N^{-1}$ können wir erreichen, dass alle $a_n \in \mathbb{Z}_p$ sind mit $a_N = 1$ und $a_n \in p\mathbb{Z}_p$ für $n > N$. Es ist dann also $m = 0$. Wir zeigen die Behauptung durch Induktion über $N$.

Ist $N = 0$, dann gilt $f(\alpha) \equiv 1 \bmod p$ für alle $\alpha \in \mathbb{Z}_p$, also kann $f$ keine Nullstelle in $\mathbb{Z}_p$ haben.

Für den allgemeinen Fall zeigen wir zunächst, dass für $\alpha \in \mathbb{Z}_p$ die Potenzreihe

$$f(x+\alpha) = \sum_{n=0}^{\infty} a_n (x+\alpha)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(\alpha)}{n!} x^n = \sum_{n=0}^{\infty} b_n x^n$$

die Voraussetzungen des Satzes mit $m = 0$ und demselben $N$ wie $f$ erfüllt. Es ist

$$b_n = \sum_{k=n}^{\infty} a_k \binom{k}{n} \alpha^{k-n} \in \mathbb{Z}_p\,.$$

Für $n = N$ ist

$$b_N = 1 + \sum_{k=N+1}^{\infty} a_k \binom{k}{N} \alpha^{k-N} \in 1 + p\mathbb{Z}_p\,.$$

Außerdem gilt $v_p(b_n) \geq \min\{v_p(a_k) \mid k \geq n\}$, was für $n > N$ echt positiv ist und für $n \to \infty$ beliebig groß wird. Damit sind die Voraussetzungen nachgewiesen.

Sei also jetzt $N \geq 1$ und sei $\alpha \in \mathbb{Z}_p$ eine Nullstelle von $f$ (wenn es keine gibt, sind wir schon fertig). Sei $\tilde{f} = f(x+\alpha)$; dann ist $\tilde{f}(0) = f(\alpha) = 0$, also $\tilde{f} = xg$ mit einer Potenzreihe $g$. Es ist klar, dass $g$ die Voraussetzungen des Satzes mit $N(g) = N(\tilde{f}) - 1 = N(f) - 1 = N - 1$ erfüllt. Nach Induktionsvoraussetzung hat $g$ höchstens $N-1$ Nullstellen (mit Vielfachheit) in $\mathbb{Z}_p$. Es folgt, dass $f$ höchstens $N$ Nullstellen in $\mathbb{Z}_p$ hat. ❑

Sei $f$ so skaliert wie im Beweis und sei $\bar{f} = \sum_{n=0}^{N} \bar{a}_n x^n \in \mathbb{F}_p[x]$ das normierte Polynom vom Grad $N$, das man durch Reduktion der Koeffizienten von $f$ modulo $p$ erhält. Jede Nullstelle $\alpha \in \mathbb{Z}_p$ von $f$ reduziert sich modulo $p$ auf eine Nullstelle $a = \bar{\alpha} \in \mathbb{F}_p$ von $\bar{f}$. Ist $a$ eine *einfache* Nullstelle von $\bar{f}$, dann zeigt eine leichte Verallgemeinerung des Henselschen Lemmas 7.14 auf Potenzreihen, dass $f$ *genau eine* Nullstelle $\alpha \in \mathbb{Z}_p$ mit $\bar{\alpha} = a$ hat.

Man kann auch zeigen, dass die Anzahl der Nullstellen $\alpha$ von $f$ mit $\bar{\alpha} = a$ höchstens die Vielfachheit von $a$ als Nullstelle von $\bar{f}$ ist. Dazu sei $\alpha \in \mathbb{Z}_p$ beliebig mit $\bar{\alpha} = a$. Dann erfüllt die Reihe $f(\alpha + px)$ die Voraussetzungen im Satz, wobei $N$ höchstens die Vielfachheit der Nullstelle $a$ von $\bar{f}$ ist. (Durch Verschieben können wir ohne Einschränkung $\alpha = 0$ annehmen. Sei $\nu$ die Vielfachheit der Nullstelle 0 von $\bar{f}$; dann gilt $v_p(a_n) \geq 0$ für alle $n$ und $v_p(a_\nu) = 0$. Die Koeffizienten von $f(px)$ sind $p^n a_n$; es folgt $v_p(p^n a_n) = n + v_p(a_n) > \nu = v_p(p^\nu a_\nu)$ für $n > \nu$. Da alle Koeffizienten mit $n > \nu$ echt größere Bewertung haben als der Koeffizient für $n = \nu$, muss $N \leq \nu$ sein.) Die Behauptung folgt dann aus dem Satz, denn die Nullstellen von $f$ in der Restklasse von $\alpha$ entsprechen bijektiv den Nullstellen in $\mathbb{Z}_p$ von $f(\alpha + px)$.

Noch etwas genauer gilt:

**LEMMA** Faktorisierung von Potenzreihen

**Lemma.** *Seien $f$, $m$ und $N$ wie in Satz 12.1. Dann gibt es eine Konstante $c \in \mathbb{Q}_p^\times$, ein normiertes Polynom $P \in \mathbb{Z}_p[x]$ vom Grad $N$ und eine Potenzreihe $h \in 1 + px\mathbb{Z}_p[\![x]\!]$, die auf $\mathbb{Z}_p$ konvergiert, sodass $f(x) = cP(x)h(x)$ ist.*

Da für $\alpha \in \mathbb{Z}_p$ stets $h(\alpha) \equiv 1 \bmod p$ und damit insbesondere $h(\alpha) \neq 0$ gilt, sind die Nullstellen von $f$ in $\mathbb{Z}_p$ dieselben wie die Nullstellen von $P$.

*Beweis.* Der Beweis wird hier nicht ausgeführt. Man zeigt die Existenz der Faktorisierung induktiv modulo $p^n$ und bildet dann geeignete Grenzwerte. ❑

Eine wichtige Potenzreihe ist die Logarithmusreihe. Sie hat im $p$-adischen ähnlich schöne Eigenschaften wie über $\mathbb{R}$ oder $\mathbb{C}$.

**LEMMA** $p$-adischer Logarithmus

12.2. **Lemma.** *Die Potenzreihe*

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}$$

*konvergiert für $x = \alpha \in p\mathbb{Z}_p$; es gilt $\log(1+\alpha) \in p\mathbb{Z}_p$. Für $\alpha, \beta \in p\mathbb{Z}_p$ gilt*

$$\log(1 + \alpha + \beta + \alpha\beta) = \log\big((1+\alpha)(1+\beta)\big) = \log(1+\alpha) + \log(1+\beta)\,.$$

*Damit ist $\log\colon (1 + p\mathbb{Z}_p, \cdot) \to (p\mathbb{Z}_p, +)$ ein Gruppenhomomorphismus; er lässt sich eindeutig zu einem Gruppenhomomorphismus $\log\colon \mathbb{Z}_p^\times \to p\mathbb{Z}_p$ fortsetzen.*

*Beweis.* (Vergleiche Übungsblatt 7, Aufgaben (6) und (7).) Die Gleichung

$$\log(1 + x + y + xy) = \log(1+x) + \log(1+y) \tag{12.1}$$

ist eine Identität von formalen Potenzreihen in $\mathbb{Q}[\![x, y]\!]$. Man kann sie zum Beispiel zeigen, indem man beide Seiten als Potenzreihen in $x$ mit Koeffizienten in $\mathbb{Q}[\![y]\!]$ auffasst. Eine Potenzreihe ist durch ihren konstanten Term und ihre Ableitung eindeutig bestimmt. Für $x = 0$ sind beide Seiten $\log(1+y)$, und die Ableitung nach $x$ beider Seiten ist $\frac{1}{1+x}$.

Für $n \geq 1$ sei $e = v_p(n)$. Dann ist $p^e \leq n$, also $|n|_p = p^{-e} \geq 1/n$. Für $\alpha \in p\mathbb{Z}_p$ ist dann $|\pm\alpha^n/n|_p \leq np^{-n} \to 0$ für $n \to \infty$; nach Lemma 7.10 (1) konvergiert also die Reihe $\log(1+\alpha)$. Aus $np^{-n} \leq p^{-1}$ für $n \geq 1$ folgt, dass $\log(1+\alpha) \in p\mathbb{Z}_p$ ist. Dass die so definierte Funktion $\log\colon (1 + p\mathbb{Z}_p, \cdot) \to (p\mathbb{Z}_p, +)$ ein Gruppenhomomorphismus ist, folgt dann aus (12.1).

Für $\alpha \in \mathbb{Z}_p^\times$ ist nach dem kleinen Satz von Fermat $\alpha^{p-1} \in 1 + p\mathbb{Z}_p$. Wir definieren $\log\colon \mathbb{Z}_p^\times \to \mathbb{Z}_p$ durch

$$\log(\alpha) = \frac{\log(\alpha^{p-1})}{p-1} \in p\mathbb{Z}_p\,.$$

Da log auf $1+p\mathbb{Z}_p$ ein Gruppenhomomorphismus ist, stimmt diese Definition auf $1+p\mathbb{Z}_p$ mit der ursprünglichen überein. Der fortgesetzte Logarithmus ist ebenfalls ein Gruppenhomomorphismus, denn es ist

$$\begin{aligned}\log(\alpha\beta) &= \frac{\log((\alpha\beta)^{p-1})}{p-1} = \frac{\log(\alpha^{p-1}\beta^{p-1})}{p-1} \\ &= \frac{\log(\alpha^{p-1})+\log(\beta^{p-1})}{p-1} = \log(\alpha)+\log(\beta)\,.\end{aligned}$$

Es ist auch klar, dass log als Gruppenhomomorphismus nicht anders auf $\mathbb{Z}_p^\times$ fortgesetzt werden kann. ❑

S. Ramanujan (1887–1920)
© MFO

Wir bemerken, dass $\log(\zeta)=0$ ist, wenn $\zeta\in\mathbb{Z}_p^\times$ eine Einheitswurzel ist.

Die Gleichung $x^2+7=2^n$ heißt *Ramanujan-Nagell-Gleichung* nach Srinivasa Ramanujan, der sie vorgeschlagen, und Trygve Nagell, der sie als Erster vollständig gelöst hat. (Siehe auch das einführende Beispiel auf Seite 6.)

**BSP**
$x^2+7=2^n$

12.3. **Beispiel.** Als Anwendungsbeispiel zeigen wir, wie man die Ramanujan-Nagell-Gleichung

$$x^2+7=2^n$$

in ganzen Zahlen $x$ und $n$ lösen kann. Die erste Beobachtung ist, dass $n\geq 3$ sein muss; damit ist dann $x$ ungerade. Sei $\sqrt{-7}=\sqrt{7}\boldsymbol{i}\in\mathbb{C}$ und $\omega=\frac{1+\sqrt{-7}}{2}$. Dann ist $\omega^2=\omega-2$, also ist $\mathbb{Z}[\omega]=\{a+b\omega\mid a,b\in\mathbb{Z}\}$. Ähnlich wie für $\mathbb{Z}[\boldsymbol{i}]$ zeigt man, dass $\mathbb{Z}[\omega]$ ein euklidischer Ring und damit ein Hauptidealring ist. Die Matrix der Multiplikation mit $a+b\omega$ bezüglich der Basis $(1,\omega)$ ist

T. Nagell (1895–1988)
© unbekannt

$$\begin{pmatrix} a & -2b \\ b & a+b \end{pmatrix}$$

mit Determinante

$$N(a+b\omega)=a(a+b)+2b^2=a^2+ab+2b^2=|a+b\omega|^2\,.$$

Die quadratische Form $a^2+ab+2b^2$ ist positiv definit; man stellt leicht fest, dass die einzigen Elemente mit Norm 2 gegeben sind durch die Primelemente $\pm\omega$ und $\pm\bar\omega=\pm(1-\omega)$; weitere Primelemente mit Norm eine Potenz von 2 gibt es nicht. Das sieht man ähnlich wie bei der Klassifikation der Primelemente von $\mathbb{Z}[\boldsymbol{i}]$ (vgl. die „Einführung in die Zahlentheorie und algebraische Strukturen“): Für eine Primzahl $p$ ist entweder $p$ ein Primelement von $\mathbb{Z}[\omega]$, oder $p=N(\alpha)$ ist eine Norm, und dann sind $\alpha$ und $\bar\alpha$ bis auf Multiplikation mit Einheiten alle Primelemente, deren Norm eine Potenz von $p$ ist.

Für $x\in\mathbb{Z}$ ungerade ist $\frac{x\pm\sqrt{-7}}{2}\in\mathbb{Z}[\omega]$; aus $x^2+7=2^n$ folgt

$$N\Big(\frac{x+\sqrt{-7}}{2}\Big)=2^{n-2}\,.$$

Weil $\mathbb{Z}[\omega]$ ein Hauptidealring ist und damit faktoriell, muss dann gelten (beachte auch $\mathbb{Z}[\omega]^\times=\{\pm1\}$)

$$\frac{x+\sqrt{-7}}{2}=\pm\omega^k\bar\omega^{n-2-k}$$

mit $0\leq k\leq n-2$. Da aber $\omega\bar\omega=2$ die linke Seite nicht teilt, muss $k=0$ oder $k=n-2$ sein. Im Fall $k=0$ haben wir

$$\frac{-x+\sqrt{-7}}{2}=-\overline{\frac{x+\sqrt{-7}}{2}}=\mp\omega^{n-2}\,.$$

In jedem Fall gilt nach einem eventuellen Vorzeichenwechsel von $x$:

$$\frac{x+\sqrt{-7}}{2} = \pm\omega^{n-2} \qquad \text{und damit auch} \qquad \frac{x-\sqrt{-7}}{2} = \pm(1-\omega)^{n-2}\,. \tag{12.2}$$

Sei jetzt $p$ eine ungerade Primzahl, sodass $-7$ ein quadratischer Rest mod $p$ ist. Dann gibt es $s \in \mathbb{Z}_p^\times$ mit $s^2 = -7$; wir setzen $w = (1+s)/2 \in \mathbb{Z}_p^\times$. Dann ist $\mathbb{Z}[\omega] \to \mathbb{Z}_p$, $a + b\omega \mapsto a + bw$, ein Ringhomomorphismus. Es folgt

$$\log\Big(\frac{x+s}{2}\Big) = (n-2)\log(w) \qquad \text{und} \qquad \log\Big(\frac{x-s}{2}\Big) = (n-2)\log(1-w)\,,$$

also auch

$$\log(1-w)\log\Big(\frac{x+s}{2}\Big) - \log(w)\log\Big(\frac{x-s}{2}\Big) = 0\,.$$

Wir schreiben $x = x_0 + px_1$ für $x_0 = 0, 1, \ldots, p-1$. Wir können testen, ob $x_0^2 + 7$ kongruent mod $p$ zu einer Potenz von 2 ist; falls nicht, können wir die betreffende Restklasse mod $p$ ausschließen. Anderenfalls ist $(x_0 \pm s)/2 \in \mathbb{Z}_p^\times$, und die obige Gleichung kann geschrieben werden als

$$\begin{aligned}\log(1-w)\Big(\log\Big(\frac{x_0+s}{2}\Big) + \log\Big(1 + \frac{p}{x_0+s}x_1\Big)\Big)\\ - \log(w)\Big(\log\Big(\frac{x_0-s}{2}\Big) + \log\Big(1 + \frac{p}{x_0-s}x_1\Big)\Big) = 0\,;\end{aligned}$$

die linke Seite ist eine Potenzreihe in $x_1$. Auf diese Reihe können wir Satz 12.1 anwenden. Das liefert uns in jedem Fall eine obere Schranke für die Anzahl der Lösungen (und insbesondere, dass es nur endlich viele gibt). Wenn wir Glück haben, dann stimmt die Schranke mit der Zahl der bekannten Lösungen in der betreffenden Restklasse überein, und wir haben gezeigt, dass es keine weiteren Lösungen gibt. Die bekannten Lösungen haben $x \in \{\pm1, \pm3, \pm5, \pm11, \pm181\}$. Dabei wird von jedem Paar jeweils nur entweder $x$ oder $-x$ auftreten, da wir das Vorzeichen von $x$ oben passend gewählt haben.

Konkret verwenden wir $p = 11$; dann ist $-7 \equiv 2^2 \bmod 11$. Die Rechnung ergibt folgende obere Schranken:

| $x_0$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Schranke | 1 | 1 | – | 1 | 1 | 2 | 1 | 1 | 1 | – | 1 |
| Lsg. mod $11^2$ | $-11$ | 1 | – | $-30$ | $-18$ | $5, 60$ | $-49$ | 40 | $-3$ | – | $-34$ |
| bekannte Lösungen | $-11$ | 1 | – | – | – | $5, 181$ | – | – | $-3$ | – | – |

Wir müssen also noch ausschließen, dass es Lösungen mit

$$x \equiv -30,\ -18,\ -49,\ 40 \text{ oder } -34 \bmod 11^2$$

gibt. Dazu stellen wir fest, dass dann

$$n \equiv 54,\ 60,\ 35,\ 50 \text{ oder } 43 \bmod 110 = \varphi(11^2)$$

sein müsste. Wir können für jede andere Primzahl $q$ feststellen, welche Potenzen $2^n$ die Eigenschaft haben, dass $2^n - 7$ ein Quadrat mod $q$ ist. Das ergibt für $q = 23$, dass $n \equiv 1, 3, 4, 5, 7 \bmod 11$ sein muss, was $n \equiv 54, 35, 50, 43 \bmod 110$ ausschließt. Der verbleibende Fall $n \equiv 60 \bmod 110$ ist etwas schwieriger zu eliminieren. Wir können die Nullstelle der betreffenden Potenzreihe etwas genauer approximieren als $x \equiv 466 \bmod 11^3$; daraus ergibt sich $n \equiv 170 \bmod 10 \cdot 11^2$. Wir betrachten jetzt $q = 727 = 6 \cdot 11^2 + 1$. Es ist $170 \equiv 49 \bmod 11^2$; die Ordnung von 2 in $\mathbb{F}_q$ ist $11^2$, und $2^{49} - 7$ ist kein Quadrat modulo $q$. Das schließt die verbleibende Restklasse aus.

Insgesamt haben wir gezeigt, dass

$$(x,n) = (\pm 1,3),\ (\pm 3,4),\ (\pm 5,5),\ (\pm 11,7),\ (\pm 181,15)$$

tatsächlich alle Lösungen der Gleichung sind. ♣

Hier ist eine genauere Beschreibung der für das obige Beispiel durchzuführenden Rechnungen. Wir nehmen für $s$ die Quadratwurzel aus $-7$ mit $s \equiv 2 \bmod 11$. Mittels des Algorithmus aus dem Beweis des Henselschen Lemmas 7.14 erhalten wir

$$s = 9 + 2\cdot 11 + 2\cdot 11^2 + 3\cdot 11^3 + \dots$$

und damit

$$w = 5 + 11 + 11^2 + 7\cdot 11^3 + \dots \quad \text{und} \quad 1 - w = 7 + 9\cdot 11 + 9\cdot 11^2 + 3\cdot 11^3 + \dots .$$

Wir berechnen (am besten mithilfe eines Computeralgebrasystems)

$$\log(1-w) = 11\cdot(4 + 10\cdot 11 + 4\cdot 11^2 + \dots) \quad \text{und} \quad \log(w) = 11\cdot(2 + 5\cdot 11 + 7\cdot 11^2 + \dots).$$

Die einzigen Werte von $x_0$, die man a priori ausschließen kann (weil $x_0^2 + 7 \equiv 2^n \bmod 11$ nicht lösbar ist), sind $x_0 = 2$ und $x_0 = 9$. Für die anderen Werte bekommt man:

| $x_0$ | $\log((x_0+s)/2)/11$ | $\log((x_0-s)/2)/11$ |
|---|---|---|
| 0 | $4 + 7\cdot 11 + 6\cdot 11^2 + \dots$ | $4 + 7\cdot 11 + 6\cdot 11^2 + \dots$ |
| 1 | $2 + 5\cdot 11 + 7\cdot 11^2 + \dots$ | $4 + 10\cdot 11 + 4\cdot 11^2 + \dots$ |
| 3 | $8 + 9\cdot 11 + 9\cdot 11^2 + \dots$ | $4 + 10\cdot 11 + 3\cdot 11^2 + \dots$ |
| 4 | $7 + 9\cdot 11 + 3\cdot 11^2 + \dots$ | $5 + 11 + 3\cdot 11^2 + \dots$ |
| 5 | $6 + 4\cdot 11 + \dots$ | $1 + 9\cdot 11 + 3\cdot 11^2 + \dots$ |
| 6 | $4 + 10\cdot 11^2 + \dots$ | $2 + 4\cdot 11 + 5\cdot 11^2 + \dots$ |
| 7 | $3 + 2\cdot 11^2 + \dots$ | $1 + 6\cdot 11 + 4\cdot 11^2 + \dots$ |
| 8 | $6 + 9\cdot 11 + 8\cdot 11^2 + \dots$ | $7 + 6\cdot 11 + 7\cdot 11^2 + \dots$ |
| 10 | $2\cdot 11 + 9\cdot 11^2 + \dots$ | $3 + 2\cdot 11 + 3\cdot 11^2 + \dots$ |

Die Potenzreihen

$$f_{x_0}(x_1) = 11^{-2}\Big(\log(1-w)\log\Big(\frac{x_0+s}{2}\Big) - \log(w)\log\Big(\frac{x_0-s}{2}\Big) + \log(1-w)\log\Big(1+11\frac{x_1}{x_0+s}\Big) - \log(w)\log\Big(1+11\frac{x_1}{x_0-s}\Big)\Big)$$

berechnen sich dann modulo $11^2$ zu

| $x_0$ | $f_{x_0}$ | $N$ |
|---|---|---|
| 0 | $(8+11) + (8+9\cdot 11)x_1 + 8\cdot 11x_1^2$ | 1 |
| 1 | $10x_1 + 3\cdot 11x_1^2$ | 1 |
| 3 | $(2+11) + (8+10\cdot 11)x_1 + 2\cdot 11x_1^2$ | 1 |
| 4 | $(7+3\cdot 11) + (9+4\cdot 11)x_1 + 9\cdot 11x_1^2$ | 1 |
| 5 | $5\cdot 11x_1 + 10\cdot 11x_1^2$ | 2 |
| 6 | $(1+11) + (9+2\cdot 11)x_1 + 9\cdot 11x_1^2$ | 1 |
| 7 | $(10+2\cdot 11) + (4+7\cdot 11)x_1 + 6\cdot 11x_1^2$ | 1 |
| 8 | $(10+5\cdot 11) + (10+9\cdot 11)x_1 + 4\cdot 11x_1^2$ | 1 |
| 10 | $(5+10\cdot 11) + (4+6\cdot 11)x_1 + 2\cdot 11x_1^2$ | 1 |

Es ist klar, dass die Bewertung der Koeffizienten von $x_1^n$ für $n \geq 3$ stets mindestens 2 ist. Somit ergeben sich die angegebenen Werte für die Schranke $N$ für die Anzahl der Nullstellen. (Die Reihen $f_1$ und $f_5$ sind exakt durch $x_1$ teilbar, denn $x = 1$ und $x = 5$ sind Lösungen.)

Man kann sich einen beträchtlichen Teil der Rechnung sparen, indem man beachtet, dass der Koeffizient von $x_1$ in $f_{x_0}$ gegeben ist durch

$$\frac{\log(1-w)}{11(x_0+s)} - \frac{\log(w)}{11(x_0-s)} = \frac{(\log(1-w)-\log(w))x_0 - (\log(1-w)+\log(w))s}{11(x_0^2+7)}\,.$$

Für

$$x_0 \not\equiv \frac{\log(1-w)+\log(w)}{\log(1-w)-\log(w)}s \equiv 5 \bmod 11$$

ist das eine Einheit. Da der Koeffizient von $x_1^2$ stets Bewertung $\geq 1$ hat, folgt $N = 1$ für alle $x_0 \neq 5$. Für $x_0 = 5$ ist der Koeffizient von $x_1$ durch 11 teilbar, aber der Koeffizient von $x_1^2$ nicht durch $11^2$ teilbar, woraus sich $N = 2$ ergibt.

Im Beweis im Beispiel oben haben wir die Variable $n$ eliminiert und dann Potenzreihen (im Wesentlichen) in $x$ erhalten. Man kann auch umgekehrt vorgehen und $x$ eliminieren, um Potenzreihen in $n$ zu bekommen. Dazu brauchen wir folgende Beobachtung. Wir erinnern uns an die Exponentialreihe

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots .$$

**LEMMA** Konvergenz von exp

12.4. **Lemma.** *Ist $p > 2$, dann konvergiert $\exp(\alpha)$ für $\alpha \in p\mathbb{Z}_p$ und es ist $\exp(\alpha) \in 1+p\mathbb{Z}_p$. Für $\alpha, \beta \in p\mathbb{Z}_p$ ist $\exp(\alpha+\beta) = \exp(\alpha)\cdot\exp(\beta)$, $\log(\exp(\alpha)) = \alpha$ und $\exp(\log(1+\alpha)) = 1+\alpha$.*

*Die analogen Aussagen gelten für $p = 2$ und $\alpha, \beta \in 4\mathbb{Z}_2$ (mit $\exp(\alpha) \in 1 + 2\mathbb{Z}_2$).*

*Beweis.* (Vergleiche Übungsblatt 7, Aufgaben (5) und (6).) Es ist

$$v_p(n!) = \sum_{k=1}^{\infty} \Big\lfloor \frac{n}{p^k} \Big\rfloor < \frac{n}{p-1} .$$

Für $p > 2$ und $\alpha \in p\mathbb{Z}_p$ folgt

$$v_p\Big(\frac{\alpha^n}{n!}\Big) \geq n - \frac{n}{p-1} = \frac{p-2}{p-1} \cdot n \to \infty \qquad \text{für } n \to \infty,$$

also konvergiert $\exp(\alpha)$ und alle Terme außer dem für $n = 0$ haben positive $p$-adische Bewertung. Für $p = 2$ und $\alpha \in 4\mathbb{Z}_2$ gilt analog

$$v_2\Big(\frac{\alpha^n}{n!}\Big) \geq 2n - n = n \to \infty .$$

$\exp(x+y) = \exp(x)\exp(y)$, $\log(\exp(x)) = x$ und $\exp(\log(1+x)) = 1+x$ sind formale Identitäten von Potenzreihen; die restlichen Aussagen folgen durch Auswerten in $\alpha$ (und $\beta$) unter Beachtung der Tatsache, dass der Wert der inneren Reihe jeweils im Konvergenzbereich der äußeren Reihe liegt. ❑

Wie in der reellen Analysis auch kann man die Exponentialreihe und die Logarithmusreihe kombinieren, um Exponentialfunktionen mit anderen Basen zu definieren.

**LEMMA** Potenzen als Potenzreihe

12.5. **Lemma.** *Sei $p$ ungerade und $\alpha \in 1+p\mathbb{Z}_p$ oder $p = 2$ und $\alpha \in 1+4\mathbb{Z}_2$, und sei*

$$f(n) = \exp(n \log(\alpha)) \in \mathbb{Q}_p[\![n]\!] .$$

*Dann ist*

$$f(n) = 1 + a_1 n + a_2 n^2 + \dots \in \mathbb{Z}_p[\![n]\!]$$

*eine Potenzreihe mit $v_p(a_k) \to \infty$ für $k \to \infty$. Insbesondere konvergiert $f$ auf $\mathbb{Z}_p$. Für $n \in \mathbb{Z}$ gilt $f(n) = \alpha^n$.*

*Beweis.* Es ist $\log(\alpha) \in p\mathbb{Z}_p$ bzw. $4\mathbb{Z}_2$. Aus dem Beweis von Lemma 12.4 ergibt sich, dass die Koeffizienten $a_k = \log(\alpha)^k/k!$ von $f$ in $\mathbb{Z}_p$ sind und für $n \to \infty$ gegen null konvergieren. Die letzte Aussage folgt aus Lemma 12.4 und der Funktionalgleichung $\exp(nx) = \exp(x)^n$. ❑

**DEF** Potenzen mit $p$-adischen Exponenten

12.6. **Definition.** Seien $p$ eine ungerade Primzahl und $\alpha \in 1 + p\mathbb{Z}_p$ oder $p = 2$ und $\alpha \in 1 + 4\mathbb{Z}_2$. Für $n \in \mathbb{Z}_p$ definieren wir

$$\alpha^n = \exp(n \log(\alpha)) \in 1 + p\mathbb{Z}_p \,.$$

Für $n \in \mathbb{Z} \subset \mathbb{Z}_p$ ist das die übliche Potenz.

Es gilt $\alpha^{m+n} = \alpha^m \cdot \alpha^m$ für $m, n \in \mathbb{Z}_p$. ♢

Nach dem kleinen Satz von Fermat ist $\alpha^{p-1} \in 1 + p\mathbb{Z}_p$, wenn $\alpha \in \mathbb{Z}_p^\times$ ist. Das wird im folgenden Beispiel benutzt, um Potenzen mit solchen $\alpha$ als Potenzreihen zu schreiben.

**BSP** $x^2 + 7 = 2^n$

12.7. **Beispiel.** Wir betrachten wieder die Gleichung $x^2 + 7 = 2^n$. Diesmal verwenden wir die beiden Gleichungen in (12.2), um $x$ zu eliminieren. Das ergibt

$$\omega^{n-2} - (1 - \omega)^{n-2} = \pm\sqrt{-7}$$

und damit in $\mathbb{Z}_{11}$

$$w^{n-2} - (1 - w)^{n-2} = \pm s \,.$$

Nach dem kleinen Satz von Fermat ist $w^{10} \equiv (1-w)^{10} \equiv 1 \bmod 11$. Wir schreiben $n - 2 = n_0 + 10n_1$ mit $n_0 \in \{0, 1, \ldots, 9\}$ und erhalten

$$w^{n_0} \exp\big(n_1 \log(w^{10})\big) - (1 - w)^{n_0} \exp\big(n_1 \log((1 - w)^{10})\big) \mp s = 0 \,. \tag{12.3}$$

Das ergibt insgesamt 20 Potenzreihen in $n_1$, deren Nullstellen in $\mathbb{Z}_{\geq 0} \subset \mathbb{Z}_{11}$ wir suchen. Satz 12.1 liefert uns Schranken für ihre Anzahl.

Konkret können wir so vorgehen: Eine Wahl von $s$ ist $s \equiv 2 - 3 \cdot 11 \bmod 11^2$ und damit

$$w \equiv -4 - 11 \bmod 11^2 \qquad \text{und} \qquad 1 - w \equiv 5 + 11 \bmod 11^2 \,.$$

Das liefert

$$w^{10} \equiv 1 - 4 \cdot 11 \bmod 11^2 \qquad \text{und} \qquad (1 - w)^{10} \equiv 1 - 2 \cdot 11 \bmod 11^2 \,.$$

Dann ist

$$\exp\big(n_1 \log(w^{10})\big) = 1 + 11a_1 n_1 + 11^2 a_2 n_1^2 + 11^3 a_3 n_1^3 + \ldots$$

mit $a_1 \equiv -4 \bmod 11$, $a_2 \equiv -3 \bmod 11$ und $v_{11}(11^k a_k) > 2$ für $k \geq 3$, sowie

$$\exp\big(n_1 \log((1 - w)^{10})\big) = 1 + 11b_1 n_1 + 11^2 b_2 n_1^2 + 11^3 b_3 n_1^3 + \ldots$$

mit $b_1 \equiv -2 \bmod 11$, $b_2 \equiv 2 \bmod 11$ und $v_{11}(11^k b_k) > 2$ für $k \geq 3$. Damit ist die uns interessierende Reihe aus (12.3) gegeben durch

$$\big(w^{n_0} - (1-w)^{n_0} \mp s\big) + 11(w^{n_0} a_1 - (1-w)^{n_0} b_1) n_1 + 11^2 (w^{n_0} a_2 - (1-w)^{n_0} b_2) n_1^2 + \ldots \,.$$

Diese Reihe erfüllt die Voraussetzungen von Satz 12.1. Falls

$$w^{n_0} - (1 - w)^{n_0} \mp s \not\equiv 0 \bmod 11$$

ist, ist $N = 0$ im Satz, und wir haben keine Lösung. Das trifft für alle $n_0$ zu außer $n_0 = 1, 2, 3, 5$. In den anderen Fällen ist $w^{n_0} - (1 - w)^{n_0} \mp s = 0$ für eines der Vorzeichen (denn $n - 2 = n_0$ ergibt eine Lösung). Ist dann

$$w^{n_0} a_1 - (1 - w)^{n_0} b_1 \equiv -4w^{n_0} + 2(1 - w)^{n_0} \not\equiv 0 \bmod 11 \,,$$

so ist ($m = 1$ und) $N = 1$ im Satz, und die bekannte Lösung ist die einzige. Das trifft für $n_0 = 1, 2, 5$ zu. Im verbleibenden Fall $n_0 = 3$ ist

$$w^{n_0} a_2 - (1 - w)^{n_0} b_2 \equiv -3w^{n_0} - 2(1 - w)^{n_0} \not\equiv 0 \bmod 11 \,,$$

und es ist $N = 2$ im Satz. Insgesamt ergibt das folgende Tabelle (die Schranken für die beiden Vorzeichen von $s$ sind jeweils addiert):

| $n_0$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Schranke | 0 | 1 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| bekannte Lösungen $n$ | – | 3 | 4 | 5, 15 | – | 7 | – | – | – | – |

Wir sehen, dass die resultierende Schranke für die Anzahl der Lösungen in diesem Fall sogar scharf ist. ♣

Der wesentliche Punkt in den Beispielen 12.3 und 12.7 war, dass wir mithilfe von etwas Information über die Struktur des Rings $\mathbb{Z}[\omega]$ die *eine* Gleichung in *zwei* Unbekannten

$$x^2 + 7 = 2^n$$

übersetzen konnten in die *zwei* Gleichungen in *zwei* Unbekannten

$$\frac{x + \sqrt{-7}}{2} = \pm\omega^{n-2} \qquad \text{und} \qquad \frac{x - \sqrt{-7}}{2} = \pm\bar{\omega}^{n-2}\,.$$

Man erwartet, dass $m$ Gleichungen in $m$ Unbekannten normalerweise nur endlich viele Lösungen haben. Wir haben diese Lösungen bestimmt, indem wir eine der Unbekannten eliminiert haben und die entstehende (transzendente) Gleichung für die verbliebene Unbekannte mit $p$-adischen Methoden gelöst haben.

Wir wollen diese Vorgehensweise verwenden, um Gleichungen der Form

$$F(x, y) = c \tag{12.4}$$

zu studieren, wobei $F \in \mathbb{Z}[x, y]$ homogen und $c \in \mathbb{Z} \setminus \{0\}$ ist und man Lösungen $(x, y) \in \mathbb{Z}^2$ sucht.

Ist $\deg(F) \le 1$, dann ist diese Gleichung sehr einfach zu behandeln. Wir können auch annehmen, dass $F$ nicht die Form $F = aG^m$ hat mit $m \ge 2$ und $a \in \mathbb{Z}$, denn in diesem Fall muss $c = ac'^m$ sein, und die Gleichung reduziert sich auf $G(x, y) = c'$ (falls $m$ ungerade) oder $G(x, y) = \pm c'$ (falls $m$ gerade).

Ist $F$ reduzibel (und nicht von der Form $aG^m$), dann ist $F = F_1F_2$ mit Faktoren $F_1$, $F_2$ ohne gemeinsamen Faktor. Es muss dann gelten

$$F_1(x, y) = c_1 \quad \text{und} \quad F_2(x, y) = c_2 \quad \text{mit} \quad c_1c_2 = c\,.$$

Da es nur endlich viele Faktorisierungen von $c$ gibt, führt das auf endlich viele Paare von algebraischen Gleichungen in $x$ und $y$, die jeweils nur endlich viele (und berechenbare) Lösungen haben. Wir können also annehmen, dass $F$ irreduzibel ist.

Im Fall $\deg(F) = 2$ (und irreduzibel) ist $F$ eine nicht-ausgeartete quadratische Form. Ist $F$ positiv oder negativ definit, dann gibt es nur endlich viele Lösungen, die man explizit bestimmen kann. Ist $F$ indefinit, dann kann man die Gleichung durch Umformungen wie im Beweis von Satz 11.11 auf eine verallgemeinerte Pellsche Gleichung

$$X^2 - DY^2 = n$$

(eventuell mit Kongruenzbedingungen an $X$ und/oder $Y$) reduzieren. Unsere Ergebnisse sagen dann, dass die Lösungsmenge die disjunkte Vereinigung von endlich vielen Bahnen unter der von einem Paar $(X_1, Y_1) \in S_D^+$ erzeugten Gruppe ist.

Der interessante Fall ist somit, dass $\deg(F) \ge 3$ und $F$ irreduzibel ist. In diesem Fall heißt Gleichung (12.4) eine *Thue-Gleichung* nach Axel Thue, der gezeigt hat,

**DEF** Thue-Gleichung


A. Thue
(1863–1922)

dass es stets nur endlich viele Lösungen gibt; siehe die Bemerkungen nach Lemma 9.6. Wir werden diese Aussage beweisen im Fall, dass $F(x,1) \in \mathbb{Z}[x]$ Grad 3 und ein Paar echt komplexer Nullstellen hat.

Sei ab jetzt $d = \deg(F) \geq 3$ und $f = F(x,1) \in \mathbb{Z}[x]$; dann ist auch $\deg(f) = d$ (denn $F$ ist irreduzibel, also nicht durch $y$ teilbar). Als weitere Vereinfachung können wir annehmen, dass $f$ normiert ist. Denn sei

$$F(x,y) = a_0 x^d + a_1 x^{d-1} y + \ldots + a_d y^d \,;$$

dann ist

$$\begin{aligned} a_0^{d-1} F(x,y) &= (a_0 x)^d + a_1 (a_0 x)^{d-1} y + \ldots + a_{d-1} a_0^{d-1} (a_0 x) y^{d-1} + a_d a_0^{d-1} y^d \\ &= \tilde{F}(a_0 x, y) \end{aligned}$$

mit einem geeigneten homogenen Polynom $\tilde{F} \in \mathbb{Z}[x,y]$ vom Grad $d$, sodass $\tilde{F}(x,1)$ normiert ist. Die Gleichung $F(x,y) = c$ ist äquivalent zu $\tilde{F}(a_0 x, y) = a_0^{d-1} c$. Wenn wir also die Lösungen von $\tilde{F}(x,y) = a_0^{d-1} c$ bestimmen können, dann erhalten wir auch die Lösungen von $F(x,y) = c$.

Da $f$ irreduzibel ist, ist $K = \mathbb{Q}[x]/\langle f \rangle$ ein Körper vom Grad $d$ über $\mathbb{Q}$, in dem $f$ eine Nullstelle $\theta$ hat ($\theta$ ist die Restklasse von $x$). So ein Körper heißt ein *algebraischer Zahlkörper*. Der Körper $K = \mathbb{Q}(\theta)$ enthält den Unterring $\mathbb{Z}[\theta]$. Da mittels der Gleichung $f(\theta) = 0$ die Potenz $\theta^d$ als $\mathbb{Z}$-Linearkombination von niedrigeren Potenzen geschrieben werden kann, ist

**DEF**
algebraischer Zahlkörper

$$\mathbb{Z}[\theta] = \mathbb{Z} + \mathbb{Z}\theta + \mathbb{Z}\theta^2 + \ldots + \mathbb{Z}\theta^{d-1} \,.$$

Wir erinnern uns daran, dass die *Norm* $N(\alpha)$ eines Elements $\alpha \in K$ definiert ist als die Determinante des $\mathbb{Q}$-linearen Endomorphismus von $K$, der durch Multiplikation mit $\alpha$ gegeben ist; siehe Seite 61.

12.8. **Lemma.** *In unserer Situation gilt* $F(x,y) = N(x - \theta y)$.

**LEMMA**
$F(x,y)$ ist Norm

*Beweis.* Wir schreiben $f = x^d + a_1 x^{d-1} + \ldots + a_d$. Die Matrix der Multiplikation mit $\theta$ bezüglich der $\mathbb{Q}$-Basis $(1, \theta, \theta^2, \ldots, \theta^{d-1})$ von $K$ ist

$$M_d = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_d \\ 1 & 0 & 0 & \cdots & 0 & -a_{d-1} \\ 0 & 1 & 0 & \cdots & 0 & -a_{d-2} \\ 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -a_1 \end{pmatrix},$$

und $N(x - \theta y) = \det(x I_d - y M_d)$. Für $d \geq 2$ liefert Entwicklung der Determinante nach der ersten Spalte

$$\det(x I_d - y M_d) = x \det(x I_{d-1} - y M_{d-1}) + a_d y^d \,;$$

Induktion ergibt die Behauptung. ❑

Unser Problem ist also äquivalent dazu, die Elemente $x - y\theta \in \mathbb{Z}[\theta]$ zu finden, deren Norm $c$ ist. Die Norm ist multiplikativ, und die Normen von Elementen von $\mathbb{Z}[\theta]$ sind in $\mathbb{Z}$. Die Elemente $\alpha \in \mathbb{Z}[\theta]$ mit $N(\alpha) = c$ bilden also eine Vereinigung von Bahnen unter der Operation der Gruppe $\mathbb{Z}[\theta]^\times_+ := \{\gamma \in \mathbb{Z}[\theta] \mid N(\gamma) = 1\}$ durch Multiplikation. Wir brauchen an dieser Stelle etwas Input aus der algebraischen Zahlentheorie (das ist das Teilgebiet der Mathematik, in dem algebraische Zahlkörper studiert werden).

**SATZ** Struktur von $N^{-1}(c)$

12.9. **Satz.** *Sei $f \in \mathbb{Z}[x]$ normiert und irreduzibel und sei $R = \mathbb{Z}[\theta]$, wobei $\theta$ eine Nullstelle von $f$ ist.*

(1) *Sei $r$ die Anzahl der reellen Nullstellen von $f$ und $s$ die Anzahl der Paare von echt komplexen Nullstellen von $f$ (dann ist $r + 2s = d = \deg(f)$). Dann gibt es Elemente $\varepsilon_1, \dots, \varepsilon_{r+s-1} \in R$ mit $N(\varepsilon_j) = 1$ für alle $j$, sodass jedes Element von $R_+^\times = \{\gamma \in R \mid N(\gamma) = 1\}$ eindeutig geschrieben werden kann als*

$$\gamma = \zeta \varepsilon_1^{n_1} \dots \varepsilon_{r+s-1}^{n_{r+s-1}}$$

*mit einer Einheitswurzel $\zeta \in R_+^\times$ und $n_j \in \mathbb{Z}$.*

(2) *Sei $c \in \mathbb{Z} \setminus \{0\}$. Dann zerfällt die Menge $\{\alpha \in R \mid N(\alpha) = c\}$ in endlich viele Bahnen unter der Operation von $R_+^\times$ auf $R$ durch Multiplikation.*

Die Aussage in (1) ist eine Version des Dirichletschen Einheitensatzes.

Die Erzeuger $\varepsilon_j$ in (1) und ein Repräsentantensystem der Bahnen in (2) können für gegebenes $f$ explizit bestimmt werden. Geeignete Computeralgebrasysteme stellen dafür Funktionen zur Verfügung.

Ist $d \in \mathbb{Z}_{>0}$ kein Quadrat, dann erfüllt $f = x^2 - d$ die Voraussetzungen von Satz 12.9 mit $r = 2$ und $s = 0$. Via $(x, y) \mapsto x + y\sqrt{d}$ haben wir einen Isomorphismus von $S_d$ mit $R_+^\times$ und von $S_d(c)$ mit der Menge der Elemente von $R$, die Norm $c$ haben. In diesem Fall entsprechen die Aussagen von Satz 12.9 unseren früheren Ergebnissen Satz 9.7 (es ist $\zeta = \pm 1$) und Satz 11.2 über die (verallgemeinerte) Pellsche Gleichung. Man kann also Satz 12.9 als eine weit reichende Verallgemeinerung dieser Ergebnisse betrachten.

Wenn wir die Einheitswurzeln in $R_+^\times$ (das sind alle in $R$, wenn der Grad gerade ist, und nur 1, wenn der Grad ungerade ist) mit den Repräsentanten der Bahnen multiplizieren, erhalten wir eine endliche Menge $\{\gamma_1, \dots, \gamma_m\} \subset R$ mit der Eigenschaft, dass jedes $\alpha \in R$ mit $N(\alpha) = c$ geschrieben werden kann als

$$\alpha = \gamma_j \varepsilon_1^{n_1} \cdots \varepsilon_{r+s-1}^{n_{r+s-1}}$$

mit $j \in \{1, \dots, m\}$ und $n_1, \dots, n_{r+s-1} \in \mathbb{Z}$. Für uns ist die Frage also, wann ein solches $\alpha$ die Form $x - y\theta$ haben kann. Um einer Antwort näher zu kommen, fixieren wir eine Primzahl $p$, die die Eigenschaft hat, dass $f$ mod $p$ in ein Produkt von $d$ verschiedenen Linearfaktoren zerfällt. (Der Tschebotarjowsche Dichtigkeitssatz garantiert, dass es immer unendlich viele solche Primzahlen gibt.) Nach dem Henselschen Lemma 7.14 ist dann

$$f(x) = (x - \theta_1)(x - \theta_2) \cdots (x - \theta_d)$$

in $\mathbb{Z}_p[x]$. Wir erhalten $d$ verschiedene Einbettungen

$$\sigma_i \colon \mathbb{Z}[\theta] \longrightarrow \mathbb{Z}_p\,, \quad \theta \longmapsto \theta_i$$

für $i = 1, 2, \dots, d$. Das ergibt (für festes $j \in \{1, 2, \dots, m\}$) $d$ Gleichungen

$$x - y\theta_i = \sigma_i(\gamma_j)\sigma_i(\varepsilon_1)^{n_1} \cdots \sigma_i(\varepsilon_{r+s-1})^{n_{r+s-1}}\,, \qquad i \in \{1, 2, \dots, d\}$$

in den $r + s + 1$ Unbekannten $x, y, n_1, \dots, n_{r+s-1}$. Ist $r + s + 1 \leq d = r + 2s$, also $s \geq 1$, dann kann man erwarten, dass es für jedes $j$ jeweils nur endlich viele Lösungen dieses Gleichungssystems gibt.

Wir betrachten jetzt den einfachsten Fall $d = 3$. Die Bedingung ist dann, dass $f$ eine reelle Nullstelle und ein paar konjugiert komplexer Nullstellen hat; das ist damit gleichbedeutend, dass die Diskriminante $\operatorname{disc}(f)$ negativ ist. Dann ist

$r+s-1=1+1-1=1$, und das Gleichungssystem vereinfacht sich (für festes $\gamma=\gamma_j$) zu

$$x-y\theta_1=\gamma^{(1)}(\varepsilon^{(1)})^n\,,\quad x-y\theta_2=\gamma^{(2)}(\varepsilon^{(2)})^n\,,\quad x-y\theta_3=\gamma^{(3)}(\varepsilon^{(3)})^n\,,$$

wobei wir $\varepsilon=\varepsilon_1$ und $\alpha^{(i)}=\sigma_i(\alpha)$ schreiben. Dieses Gleichungssystem ist linear in $x$ und $y$. Elimination dieser beiden Unbekannten ergibt

$$\alpha_1\gamma^{(1)}(\varepsilon^{(1)})^n+\alpha_2\gamma^{(2)}(\varepsilon^{(2)})^n+\alpha_3\gamma^{(3)}(\varepsilon^{(3)})^n=0\,, \tag{12.5}$$

wobei $(\alpha_1,\alpha_2,\alpha_3)$ eine nichttriviale Lösung von

$$\alpha_1+\alpha_2+\alpha_3=\theta_1\alpha_1+\theta_2\alpha_2+\theta_3\alpha_3=0$$

ist (da $\theta_1,\theta_2,\theta_3$ paarweise verschieden sind, ist diese Lösung bis auf einen skalaren Faktor eindeutig; konkret können wir $(\alpha_1,\alpha_2,\alpha_3)=(\theta_2-\theta_3,\theta_3-\theta_1,\theta_1-\theta_2)$ nehmen).

Nach dem kleinen Satz von Fermat ist $(\varepsilon^{(i)})^{p-1}\in 1+p\mathbb{Z}_p$; sei $p\lambda_i=\log\bigl((\varepsilon^{(i)})^{p-1}\bigr)$ mit $\lambda_i\in\mathbb{Z}_p$. Wenn wir $n=n_0+(p-1)t$ schreiben, dann ist (12.5) dasselbe wie

$$\alpha_1\gamma^{(1)}(\varepsilon^{(1)})^{n_0}\exp(p\lambda_1 t)+\alpha_2\gamma^{(2)}(\varepsilon^{(2)})^{n_0}\exp(p\lambda_2 t)+\alpha_3\gamma^{(3)}(\varepsilon^{(3)})^{n_0}\exp(p\lambda_3 t)=0\,.$$

Wir kürzen ab und setzen $\beta_i=\alpha_i\gamma^{(i)}(\varepsilon^{(i)})^{n_0}$. Obige Gleichung lautet dann

$$(\beta_1+\beta_2+\beta_3)+(\beta_1\lambda_1+\beta_2\lambda_2+\beta_3\lambda_3)pt+\frac{\beta_1\lambda_1^2+\beta_2\lambda_2^2+\beta_3\lambda_3^2}{2}p^2t^2+\dots$$
$$+\frac{\beta_1\lambda_1^k+\beta_2\lambda_2^k+\beta_3\lambda_3^k}{k!}p^kt^k+\dots=0\,.$$

**LEMMA** Gleichung degeneriert nicht

12.10. **Lemma.** *Die Potenzreihe auf der linken Seite in der Gleichung oben ist nicht die Nullreihe.*

*Beweis.* Es ist $\beta_i\neq 0$ für $i=1,2,3$. Wäre die Reihe die Nullreihe, dann gälte insbesondere

$$\beta_1+\beta_2+\beta_3=\beta_1\lambda_1+\beta_2\lambda_2+\beta_3\lambda_3=\beta_1\lambda_1^2+\beta_2\lambda_2^2+\beta_3\lambda_3^2=0\,.$$

Die Matrix $M=(\lambda_j^i)_{0\le i\le 2,1\le j\le 3}$ hätte damit nichttrivialen Kern, also wäre (Vandermonde-Determinante)

$$\det(M)=(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)=0\,,$$

also zum Beispiel $\lambda_1=\lambda_2=:\lambda$. Dann hätten wir

$$(\beta_1+\beta_2)+\beta_3=(\beta_1+\beta_2)\lambda+\beta_3\lambda_3=0\,,$$

woraus (wegen $\beta_3\neq 0$) auch $\lambda_3=\lambda$ folgen würde. Daraus würde sich aber

$$(\varepsilon^{(1)})^{p-1}=(\varepsilon^{(2)})^{p-1}=(\varepsilon^{(3)})^{p-1}=\exp(p\lambda)$$

ergeben, d.h., $\varepsilon^{p-1}$ hätte unter allen drei Einbettungen in $\mathbb{Z}_p$ dasselbe Bild $\varepsilon'$. Diagonalisiert man die Matrix der $\mathbb{Q}$-linearen Abbildung $\alpha\mapsto\varepsilon^{p-1}\alpha$ über $\mathbb{Q}_p$, dann sieht man, dass $\varepsilon^{p-1}$ eine Nullstelle des Polynoms

$$\mathbb{Q}[x]\ni\chi_{\varepsilon^{p-1}}(x)=N(x-\varepsilon^{p-1})=(x-\varepsilon')^3$$

ist. Da dieses Polynom rationale Koeffizienten hat, muss (beachte $N(q)=q^3$ für $q\in\mathbb{Q}$)

$$\varepsilon^{p-1}=\varepsilon'\in\mathbb{Q}\cap\mathbb{Z}[\theta]_+^\times=\{1\}\,,$$

also $\varepsilon^{p-1}=1$ sein, und das wäre ein Widerspruch zu Teil (1) von Satz 12.9. ❑

Es folgt:

**SATZ** kubische Thue-Gleichungen

12.11. **Satz.** *Sei $f = x^3 + a_1x^2 + a_2x + a_3 \in \mathbb{Z}[x]$ irreduzibel mit $\operatorname{disc}(f) < 0$ und sei $0 \neq c \in \mathbb{Z}$. Dann hat die Gleichung*

$$x^3 + a_1x^2y + a_2xy^2 + a_3y^3 = c$$

*nur endlich viele Lösungen $(x, y) \in \mathbb{Z}^2$.*

*Beweis.* Aus der Diskussion oben ergab sich, dass es einen Erzeuger $\varepsilon$ von $\mathbb{Z}[\theta]_+^\times$ und endlich viele Elemente $\gamma_1, \dots, \gamma_m \in \mathbb{Z}[\theta]$ gibt, sodass es für jede Lösung $(x, y) \in \mathbb{Z}^2$ von $F(x, y) = c$ ein $j \in \{1, \dots, m\}$, ein $n_0 \in \{0, \dots, p-2\}$ und ein $n \in \mathbb{Z}$ gibt mit

$$\begin{aligned} x - y\theta_1 &= \gamma_j^{(1)}(\varepsilon^{(1)})^{n_0} \exp(p\lambda_1 n) \\ x - y\theta_2 &= \gamma_j^{(2)}(\varepsilon^{(2)})^{n_0} \exp(p\lambda_2 n) \\ x - y\theta_3 &= \gamma_j^{(3)}(\varepsilon^{(3)})^{n_0} \exp(p\lambda_3 n)\,, \end{aligned}$$

wobei $p\lambda_i = \log((\varepsilon^{(i)})^{p-1}) \in p\mathbb{Z}_p$ ist. Dieses Gleichungssystem ist äquivalent zu

$$\begin{aligned} x &= \xi_1 \exp(p\lambda_1 n) + \xi_2 \exp(p\lambda_2 n) \\ y &= \eta_1 \exp(p\lambda_1 n) + \eta_2 \exp(p\lambda_2 n) \\ 0 &= \beta_1 \exp(p\lambda_1 n) + \beta_2 \exp(p\lambda_2 n) + \beta_3 \exp(p\lambda_3 n) \end{aligned}$$

mit gewissen Koeffizienten $\xi_i, \eta_i, \beta_i \in \mathbb{Z}_p$, die von $j$ und $n_0$ abhängen. Lemma 12.10 zeigt, dass die Potenzreihe in $n$ auf der rechten Seite der letzten Gleichung nicht die Nullreihe ist. Für die Koeffizienten

$$b_k = \frac{\beta_1\lambda_1^k + \beta_2\lambda_2^k + \beta_3\lambda_3^k}{k!}p^k$$

der Reihe gilt $v_p(b_k) \to \infty$ für $k \to \infty$ (vergleiche Lemma 12.4; wir haben $p > 2$, da $f$ mod 2 keine drei verschiedenen Nullstellen haben kann). Satz 12.1 ist damit auf die letzte Gleichung anwendbar und zeigt, dass sie nur endlich viele Lösungen $n$ hat. Jede dieser Lösungen liefert eindeutige $x, y \in \mathbb{Z}_p$. Es folgt, dass es für jedes Paar $(j, n_0)$ höchstens endlich viele Lösungen $(x, y) \in \mathbb{Z}$ geben kann. Da es nur endlich viele (nämlich $m(p-1)$) solcher Paare gibt, ist die Zahl der Lösungen insgesamt endlich. ❑

Der Beweis liefert für jedes $p$, das die Voraussetzungen erfüllt, eine explizite Schranke für die Anzahl der Lösungen. Ist die Schranke nicht scharf, dann kann man ähnlich wie in Beispiel 12.3 versuchen, durch Reduktion modulo anderer Primzahlen die Fälle auszuschließen, die keinen bekannten Lösungen entsprechen.

**BSP** kubische Thue-Gleichung

12.12. **Beispiel.** Wir betrachten die Gleichung

$$x^3 + 2y^3 = 62\,.$$

Das Polynom $f(x) = x^3 + 2$ ist irreduzibel (nach Eisenstein) und hat genau eine reelle Nullstelle $\theta = -\sqrt[3]{2}$; die Methode ist also anwendbar. Der relevante Ring ist hier $R = \mathbb{Z}[\theta] = \mathbb{Z}[\sqrt[3]{2}]$. Man kann zeigen, dass $R$ ein Hauptidealring ist und dass $\varepsilon = \sqrt[3]{2} - 1$ die Gruppe $R_+^\times$ erzeugt. Bis auf Multiplikation mit Potenzen von $\varepsilon$ gibt es genau drei Elemente von $R$ mit Norm 62, nämlich

$$\gamma_1 = 4 - \sqrt[3]{2}\,, \quad \gamma_2 = 2 - \sqrt[3]{2} + 2\sqrt[3]{2}^2 \quad \text{und} \quad \gamma_3 = 2 + 3\sqrt[3]{2}\,.$$

Die kleinste Primzahl $p$, sodass $x^3 + 2$ modulo $p$ in Linearfaktoren zerfällt, ist $p = 31$. Wir bestimmen zunächst für jedes $j \in \{1, 2, 3\}$ die $n_0 \in \{0, 1, \ldots, 29\}$, für die

$$\alpha_1 \gamma_j^{(1)} (\varepsilon^{(1)})^{n_0} + \alpha_2 \gamma_j^{(2)} (\varepsilon^{(2)})^{n_0} + \alpha_3 \gamma_j^{(3)} (\varepsilon^{(3)})^{n_0} \equiv 0 \bmod 31$$

ist (für die anderen $n_0$ wird $N = 0$ in Satz 12.1, also gibt es keine Lösungen mit $n \equiv n_0 \bmod 30$). Das ergibt:

$$j = 1\colon n_0 = 0, 10, 20\,; \quad j = 2\colon n_0 = 4, 14, 24\,; \quad j = 3\colon n_0 = 0, 5, 10, 15, 20, 25\,.$$

Der Koeffizient $b_1$ von $n$ in der jeweiligen Potenzreihe hat in jedem Fall $v_{31}(b_1) = 1$; für $k \geq 2$ gilt $v_{31}(b_k) \geq 2$. Damit ist $N = 1$ in Satz 12.1, sodass die Reihe höchstens (und sogar genau) eine Nullstelle in $\mathbb{Z}_{31}$ hat. Für jedes $j$ entspricht dem kleinsten Wert von $n_0$ tatsächlich auch eine Lösung in $\mathbb{Z}$, nämlich

$$(x, y) = (4, -1)\,, \qquad (-34, 27) \qquad \text{und} \qquad (2, 3)\,.$$

Die verbleibenden Werte von $n_0$ scheinen zu keiner ganzzahligen Lösung zu gehören und müssen daher noch ausgeschlossen werden. Wenn man die Gleichungen modulo $q = 1801$ betrachtet, dann stellt man analog zum ersten Schritt in der Rechnung oben fest, dass man folgende Kongruenzen haben muss:

$$j = 1\colon n \equiv 0 \bmod 600\,; \quad j = 2\colon n \equiv 4 \bmod 600\,; \quad j = 3\colon n \equiv 0, 514 \bmod 600\,.$$

Damit können die „überflüssigen“ Möglichkeiten für $n_0$ in allen Fällen eliminiert werden. Wir haben also gezeigt, dass die Gleichung

$$x^3 + 2y^3 = 62$$

genau die drei ganzzahligen Lösungen

$$(x, y) = (4, -1)\,, \qquad (-34, 27) \qquad \text{und} \qquad (2, 3)$$

hat. ♣

**Exkurs über die ABC-Vermutung.**

Die (etwas überraschende) Lösung $(x, y) = (-34, 27)$ der oben betrachteten Gleichung führt auf $3^9 = 2^2 \cdot 17^3 + 31$, was ein sogenanntes *gutes ABC-Tripel* liefert. Das ist ein Tripel $(A, B, C)$ von teilerfremden positiven ganzen Zahlen $A, B, C$ mit $C = A + B$ und der Eigenschaft, dass

$$C > \prod_{p \mid ABC} p$$

ist (rechts steht das Produkt der verschiedenen Primteiler von $A$, $B$ und $C$). Im konkreten Fall ist

$$3^9 = 19\,683 > 3\,162 = 2 \cdot 3 \cdot 17 \cdot 31\,.$$

Die ABC-Vermutung sagt, dass so etwas nur selten vorkommt: Für jedes $\varepsilon > 0$ gibt es eine Konstante $C_\varepsilon > 0$, sodass für alle $(A, B, C)$ wie oben gilt

$$C \leq C_\varepsilon \Big( \prod_{p \mid ABC} p \Big)^{1+\varepsilon}\,.$$

Äquivalent dazu ist die Aussage, dass es für jedes $\varepsilon > 0$ nur endlich viele Tripel $(A, B, C)$ wie oben gibt mit

$$C \geq \Big( \prod_{p \mid ABC} p \Big)^{1+\varepsilon}\,.$$

Aus dieser Vermutung würden allerhand interessante Endlichkeitsaussagen folgen; sie ist aber bisher (gemäß der Meinung der überwiegenden Zahl der Experten) nicht bewiesen. (Mochizuki behauptet, einen Beweis zu haben, konnte ihn aber bisher nicht so formulieren, dass die Experten ihn verifizieren können. Der oben verlinkte Wikipedia-Eintrag hat etwas mehr Informationen dazu.)

Hätte man die ABC-Vermutung zur Verfügung für ein $\varepsilon < 1/2$, dann könnte man die Endlichkeit der Lösungsmenge von Gleichungen der Form

$$ax^3 + by^3 = c$$

sehr leicht beweisen: Wir können annehmen, dass die drei Terme in der Gleichung teilerfremd sind (nicht-primitive Lösungen lassen sich auf primitive Lösungen von endlich vielen Gleichungen derselben Art reduzieren). Aus ABC folgt dann mit $X = \max\{|ax^3|, |by^3|\}$

$$X \le C_\varepsilon\Big(\prod_{p|abcxy} p\Big)^{1+\varepsilon} \le C_\varepsilon(abcxy)^{1+\varepsilon} \le C_\varepsilon(ab)^{2(1+\varepsilon)/3}c^{1+\varepsilon}X^{2(1+\varepsilon)/3}$$

und damit

$$X \le \big(C_\varepsilon(ab)^{2(1+\varepsilon)/3}c^{1+\varepsilon}\big)^{3/(1-2\varepsilon)}.$$

Mit einer expliziten Konstante $C_\varepsilon$ bekäme man sogar eine explizite Schranke an die Lösungen.

Tatsächlich folgt aus ABC auch der Satz von (Thue-Siegel-)Roth und damit die Endlichkeit der Lösungsmenge *jeder* Thue-Gleichung.

Wenn wir diese $p$-adische Methode auf Thue-Gleichungen höheren Grades anwenden wollen, dann müssen wir Gleichungssysteme aus mehreren Potenzreihen in mehreren Variablen lösen. Eine Potenzreihe in $m$ Variablen $x_1, \dots, x_m$ hat die Form

$$f(\boldsymbol{x}) = f(x_1, \dots, x_m) = \sum_{i_1,\dots,i_m=0}^{\infty} a_{i_0,\dots,i_m} x_1^{i_1} \cdots x_m^{i_m} = \sum_{\boldsymbol{i}\ge\boldsymbol{0}} a_{\boldsymbol{i}} \boldsymbol{x}^{\boldsymbol{i}}\,;$$

dabei steht $\boldsymbol{i}$ für das Tupel $(i_1, \dots, i_m)$, und $\boldsymbol{x}^{\boldsymbol{i}}$ ist eine Abkürzung für das Monom $x_1^{i_1} \cdots x_m^{i_m}$. Der Ring der (formalen) Potenzreihen in $x_1, \dots, x_m$ mit Koeffizienten in einem Ring $R$ wird $R[\![\boldsymbol{x}]\!] = R[\![x_1, \dots, x_m]\!]$ notiert. Wir schreiben $|\boldsymbol{i}|$ für die Summe $i_1 + \ldots + i_m$. Die Reihe $f$ oben konvergiert auf $\mathbb{Z}_p^m$, falls $v_p(a_{\boldsymbol{i}}) \to \infty$ für $|\boldsymbol{i}| \to \infty$.

Eine Verallgemeinerung von Satz 12.1 sieht dann so aus:

**SATZ**
Nullstellen von Systemen von Potenzreihen

12.13. **Satz.** *Es seien $f_1, \dots, f_n \in \mathbb{Z}_p[\![x_1, \dots, x_m]\!]$ mit $f_j = \sum_{\boldsymbol{i}} a_{\boldsymbol{i}}^{(j)} \boldsymbol{x}^{\boldsymbol{i}}$, wobei $v_p(a_{\boldsymbol{i}}^{(j)}) \to \infty$ für $|\boldsymbol{i}| \to \infty$ gelte für jedes $j \in \{1, \dots, n\}$. Wir schreiben $\bar f_j \in \mathbb{F}_p[\boldsymbol{x}]$ für das Polynom, das man erhält, wenn man die Koeffizienten von $f_j$ modulo $p$ reduziert.*

(1) *Hat das Gleichungssystem $\bar f_1(\boldsymbol{x}) = \ldots = \bar f_n(\boldsymbol{x}) = 0$ nur endlich viele Lösungen in $\bar{\mathbb{F}}_p^m$ (dabei ist $\bar{\mathbb{F}}_p$ der algebraische Abschluss von $\mathbb{F}_p$), dann hat das Gleichungssystem $f_1(\boldsymbol{x}) = \ldots = f_n(\boldsymbol{x}) = 0$ nur endlich viele Lösungen in $\mathbb{Z}_p^m$.*

(2) *Es sei jetzt $m = n$. Gilt $\bar a_{\boldsymbol{0}}^{(j)} = 0$ für alle $j$ und ist die Matrix*

$$M = (\bar a_{\mathbf{e}_i}^{(j)})_{1\le i,j\le n} \in \mathrm{Mat}(n, \mathbb{F}_p)$$

*invertierbar, dann hat das Gleichungssystem $f_1(\boldsymbol{x}) = \ldots = f_n(\boldsymbol{x}) = 0$ genau eine Lösung in $(p\mathbb{Z}_p)^n$.*

(3) *Ist das reduzierte Gleichungssystem $\bar f_1(\boldsymbol{x}) = \ldots = \bar f_n(\boldsymbol{x}) = 0$ ein lineares Gleichungssystem mit eindeutiger Lösung, dann hat das Gleichungssystem $f_1(\boldsymbol{x}) = \ldots = f_n(\boldsymbol{x}) = 0$ höchstens eine Lösung in $\mathbb{Z}_p^m$. Ist zusätzlich $m = n$, dann hat es genau eine Lösung in $\mathbb{Z}_p^m$.*

*Beweis.* Teil (1) werden wir hier nicht beweisen. Die Bedingung ist dazu äquivalent, dass das modulo $p$ reduzierte System eine nulldimensionale (oder leere) algebraische Varietät über $\mathbb{F}_p$ beschreibt. Das kann man mithilfe geeigneter Algorithmen (Stichwort Gröbnerbasen) entscheiden. Es gilt sogar genauer, dass dann

die Anzahl der Lösungen des Gleichungssystems über $\mathbb{Z}_p$ höchstens so groß ist wie die Anzahl der Lösungen in $\mathbb{F}_p^m$ (mit Vielfachheit gezählt) des modulo $p$ reduzierten Systems.

Teil (2) ist eine mehrdimensionale Version des Henselschen Lemmas 7.14. Der Beweis verwendet das mehrdimensionale Newton-Verfahren. Wir nehmen an, dass wir eine Lösung $\boldsymbol{y}^{(k)} \in (p\mathbb{Z}_p)^n$ modulo $p^k$ kennen. Zu Beginn ist $k = 1$, und wir können $\boldsymbol{y}^{(1)} = \mathbf{0}$ nehmen (denn $f_j(\mathbf{0}) = a_{\mathbf{0}}^{(j)} \equiv 0 \bmod p$). Es ist klar, dass $\boldsymbol{y}^{(1)}$ modulo $p$ eindeutig bestimmt ist.

Nach Voraussetzung ist also $f_j(\boldsymbol{y}^{(k)}) = p^k b_j$ mit $b_j \in \mathbb{Z}_p$. Wir schreiben $\partial_i f_j$ für die partielle Ableitung von $f_j$ nach $x_i$. Es gilt dann

$$\begin{aligned} f_j(\boldsymbol{y}^{(k)} + p^k \boldsymbol{w}) &= f_j(\boldsymbol{y}^{(k)}) + p^k \sum_{i=1}^{n} \partial_i f_j(\boldsymbol{y}^{(k)}) w_i + p^{2k}(\ldots) \\ &= p^k \Big( b_j + \sum_{i=1}^{n} a_{\mathbf{e}_i}^{(j)} w_i \Big) + p^{k+1}(\ldots)\,. \end{aligned}$$

(Beachte, dass $\partial_i f_j(\boldsymbol{y}^{(k)}) \equiv \partial_i f_j(\mathbf{0}) = a_{\mathbf{e}_i}^{(j)} \bmod p$ ist.) Die rechte Seite ist also genau dann durch $p^{k+1}$ teilbar, wenn $\bar{\boldsymbol{w}} M = -\bar{\boldsymbol{b}}$ ist. Da $M$ invertierbar ist, hat diese Gleichung eine eindeutige Lösung. Ist $\boldsymbol{w} \in \mathbb{Z}_p^n$ beliebig, sodass sich $\boldsymbol{w}$ modulo $p$ auf diese Lösung reduziert, dann ist $\boldsymbol{y}^{(k+1)} = \boldsymbol{y}^{(k)} + p^k \boldsymbol{w}$ eine Lösung modulo $p^{k+1}$, und $\boldsymbol{y}^{(k+1)}$ ist modulo $p^{k+1}$ eindeutig bestimmt.

Die Folge $(\boldsymbol{y}^{(k)})_{k\geq 1}$ konvergiert in $\mathbb{Z}_p^n$ gegen einen Grenzwert $\boldsymbol{y} \in (p\mathbb{Z}_p)^n$. Die durch die Potenzreihen $f_j$ definierten Funktionen sind stetig auf $\mathbb{Z}_p^n$; es folgt, dass $\boldsymbol{y}$ eine Lösung des Gleichungssystems ist. Die Eindeutigkeit ergibt sich daraus, dass $\boldsymbol{y}$ modulo $p^k$ eindeutig bestimmt ist für jedes $k \geq 1$.

Die erste Aussage von Teil (3) ist ein Spezialfall von (1). Wir zeigen, dass sie aus (2) folgt. Zunächst können wir eine geeignete Verschiebung vornehmen, sodass die eindeutige Lösung des reduzierten Systems der Nullvektor ist. Dann muss jede Lösung in $\mathbb{Z}_p^m$ des ursprünglichen Systems schon in $(p\mathbb{Z}_p)^m$ liegen. Aus der Voraussetzung folgt $n \geq m$. Lassen wir passende $n - m$ der Gleichungen weg, dann sind wir in der Situation von (2); die verbleibenden $m$ Gleichungen haben also eine eindeutige Lösung in $(p\mathbb{Z}_p)^m$ und damit auch in $\mathbb{Z}_p^m$. Im Fall $m = n$ sind dies alle Gleichungen, also hat das ursprüngliche System genau eine Lösung in $\mathbb{Z}_p^m$. Anderenfalls kann das ursprüngliche System jedenfalls nicht mehr Lösungen haben. ❑

Teil (3) von Satz 12.13 ermöglicht es uns, für konkrete Gleichungssysteme zu zeigen, dass sie keine unbekannten Lösungen haben. Dazu betrachtet man zunächst das modulo $p$ reduzierte System. Ist dies linear mit eindeutiger Lösung, dann ist Teil (3) anwendbar. Anderenfalls kann man für jede Lösung $\bar{\boldsymbol{b}} \in \mathbb{F}_p^m$ des reduzierten Systems das System $\tilde{f}_j(\boldsymbol{x}) := f_j(p\boldsymbol{x} + \boldsymbol{b}) = 0$ betrachten (wobei $\boldsymbol{b} \in \mathbb{Z}_p^m$ ein Vektor ist, der sich auf $\bar{\boldsymbol{b}}$ reduziert). Hier sind die linken Seiten jeweils durch $p$ teilbar. Wir ersetzen die linken Seiten daher durch eine $\mathbb{Z}_p$-Basis von $(\mathbb{Q}_p \tilde{f}_1 + \ldots + \mathbb{Q}_p \tilde{f}_n) \cap \mathbb{Z}_p[\![\boldsymbol{x}]\!]$. Wir testen dann, ob Teil (3) des Satzes auf das neue System anwendbar ist. Gegebenenfalls müssen wir diesen Verfeinerungsschritt mehrfach wiederholen. Allerdings gibt es keine Garantie, dass das Verfahren nach endlich vielen Schritten zum Ende kommt. In jedem Fall können wir auch alternativ mit einer anderen Primzahl $p$ neu beginnen.

Sei jetzt

$$F(x,y) = c$$

eine Thue-Gleichung vom Grad $d \geq 4$ (nach wie vor mit $F$ irreduzibel und $f(x) = F(x,1)$ normiert); wir nehmen an, dass $f$ mindestens ein Paar echt komplexer Nullstellen hat. Wie schon diskutiert, gibt es $\gamma_1, \dots, \gamma_m \in R = \mathbb{Z}[\theta]$ und $\varepsilon_1, \dots, \varepsilon_{r+s-1} \in R_+^\times$, sodass jede Lösung $(x,y)$ eine Gleichung

$$x - y\theta = \gamma_j \varepsilon_1^{n_1} \cdots \varepsilon_{r+s-1}^{n_{r+s-1}}$$

erfüllt mit $j \in \{1, \dots, m\}$ und $n_1, \dots, n_{r+s-1} \in \mathbb{Z}$. Wählen wir die Primzahl $p$ so, dass $f$ modulo $p$ über $\mathbb{F}_p$ in verschiedene Linearfaktoren zerfällt, dann liefert uns das $d$ Gleichungen der Form

$$x - y\theta_i = \gamma_j^{(i)} (\varepsilon_1^{(i)})^{n_1} \cdots (\varepsilon_{r+s-1}^{(i)})^{n_{r+s-1}}$$

in $\mathbb{Z}_p$, mit $i = 1, \dots, d$. Um dieses Gleichungssystem zu lösen, gibt es im Wesentlichen zwei Ansätze: Entweder eliminieren wir $n_1, \dots, n_{r+s-1}$, oder wir eliminieren $x$ und $y$.

Für den ersten Ansatz logarithmieren wir die Gleichungen; das ergibt

$$\log(x - y\theta_i) = \log(\gamma_j^{(i)}) + n_1 \log(\varepsilon_1^{(i)}) + \ldots + n_{r+s-1} \log(\varepsilon_{r+s-1}^{(i)})$$

für $i = 1, \dots, d$. Diese Gleichungen sind linear in $n_1, \dots, n_{r+s-1}$, sodass wir diese Variablen ohne Probleme eliminieren können. Man erwartet, dass die Matrix $\big(\log(\varepsilon_j^{(i)})\big)_{1 \leq i \leq d, 1 \leq j \leq r+s-1}$ (vollen) Rang $r+s-1$ hat (das ist die sogenannte Leopoldt-Vermutung); es ergeben sich dann $d - (r+s-1) = s+1$ Gleichungen der Art

$$\alpha_0 + \alpha_1 \log(x - y\theta_1) + \ldots + \alpha_d \log(x - y\theta_d) = 0\,.$$

Da wir $s \geq 1$ angenommen haben, sind es mindestens zwei Gleichungen. Um sie mittels Potenzreihen zu schreiben, fixieren wir $x_0, y_0 \in \{0, 1, \dots, p-1\}$ und schreiben $x = x_0 + px_1$, $y = y_0 + py_1$. Das ergibt

$$\alpha_0 + \sum_{i=1}^{d} \alpha_i \log(x_0 - y_0\theta_i) + \sum_{i=1}^{d} \alpha_i \log\Big(1 + \frac{p}{x_0 - y_0\theta_i}(x_1 + y_1\theta_i)\Big) = 0\,.$$

Die Logarithmen sind durch $p$ teilbar, und dasselbe gilt für $\alpha_0$ (die Konstante ist eine Linearkombination der $\log(\gamma_j^{(i)})$); wir können die linken Seiten also noch mit $1/p$ skalieren, bevor wir versuchen, Satz 12.13 anzuwenden. Das Verfahren ist analog zur Vorgehensweise in Beispiel 12.3, nur dass wir jetzt zwei Unbekannte $x$ und $y$ haben statt nur einer.

**BSP**
Thue-Gleichung vom Grad 4

12.14. **Beispiel.** Wir betrachten die Gleichung

$$x^4 - 5y^4 = 1\,.$$

Das Polynom $f(x) = x^4 - 5$ hat zwei reelle und ein paar konjugiert komplexer Nullstellen; die Voraussetzung ist also erfüllt. Der relevante Ring ist hier $R = \mathbb{Z}[\theta]$ mit $\theta = \sqrt[4]{5}$. Erzeuger für $R_+^\times$ sind die Einheitswurzel $-1$ sowie

$$\varepsilon = \theta^2 - 2 \qquad \text{und} \qquad \eta = 2\theta + 3\,.$$

Da hier $c = 1$ ist, erhalten wir die Gleichung

$$x - y\theta = \pm \varepsilon^m \eta^n$$

mit $m, n \in \mathbb{Z}$.

Die kleinste Primzahl $p$ mit der Eigenschaft, dass $x^4 - 5$ über $\mathbb{F}_p$ in verschiedene Linearfaktoren zerfällt, ist $p = 101$: Es ist

$$x^4 - 5 \equiv (x - 34)(x - 37)(x + 37)(x + 34) \bmod 101\,.$$

Da $\gamma_j = \pm 1$ ist, ist $\log(\gamma_j) = 0$. Weiter ist

$$\begin{pmatrix} \log(\varepsilon^{(1)}) & \cdots & \log(\varepsilon^{(4)}) \\ \log(\eta^{(1)}) & \cdots & \log(\eta^{(4)}) \end{pmatrix} \equiv 101 \cdot \begin{pmatrix} 33 & -33 & -33 & 33 \\ 2 & -43 & -23 & -37 \end{pmatrix} \bmod 101^2\,.$$

Es folgt

$$\begin{aligned} \alpha_1 \log(x - y\theta_1) + \alpha_2 \log(x - y\theta_2) + \alpha_3 \log(x - y\theta_3) + \alpha_4 \log(x - y\theta_4) = 0 \\ \beta_1 \log(x - y\theta_1) + \beta_2 \log(x - y\theta_2) + \beta_3 \log(x - y\theta_3) + \beta_4 \log(x - y\theta_4) = 0 \end{aligned}$$

mit

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix} \equiv \begin{pmatrix} 35 & 36 & 0 & 1 \\ -34 & -35 & 1 & 0 \end{pmatrix} \bmod 101\,.$$

Wir wollen jetzt erst einmal die Menge von a priori $101^2$ Paaren $(x_0, y_0)$ etwas eindampfen. Als ersten Schritt können wir, prüfen, ob $x_0^4 - 5y_0^4 \equiv 1 \bmod 101$ ist. Das reduziert die Anzahl schon einmal auf 104 Paare. Als Nächstes testen wir, ob es $m, n \in \mathbb{Z}$ gibt, sodass $x_0 - y_0\theta_i \equiv \pm(\varepsilon^{(i)})^m(\eta^{(i)})^n \bmod 101$ ist für alle $i \in \{1, 2, 3, 4\}$. Das reduziert die relevante Menge auf die 10 Paare

$$(\pm 1, 0), \quad (\pm 3, \pm 2), \quad (\pm 3, \pm 20)\,.$$

Für diese Paare berechnen wir die (ersten paar Koeffizienten der) beiden Potenzreihen. Es stellt sich heraus, dass Satz 12.13 (3) in jedem Fall anwendbar ist. Damit wissen wir bereits, dass es höchstens 10 Lösungen gibt. Tatsächliche Lösungen erhalten wir für $(\pm 1, 0)$ und $(\pm 3, \pm 2)$, und zwar gilt

$$\pm 1 = \pm\varepsilon^0\eta^0\,, \quad \pm(3 - 2\theta) = \pm\varepsilon^2\eta^{-1} \quad \text{und} \quad \pm(3 + 2\theta) = \pm\varepsilon^0\eta^1\,.$$

Für die verbleibenden vier Paare stellen wir fest, dass

$$(m, n) \equiv (6, -7) \text{ oder } (-8, 7) \bmod 25$$

sein müsste. Wir können verifizieren, dass das modulo $q = 401$ unmöglich ist.

Wir haben also gezeigt, dass die Gleichung

$$x^4 - 5y^4 = 1$$

genau die Lösungen $(x, y) = (\pm 1, 0)$ und $(\pm 3, \pm 2)$ in $\mathbb{Z}$ hat. ♣

Die Alternative ist, die beiden Unbekannten $x$ und $y$ zu eliminieren, analog zur Vorgehensweise in Beispiel 12.7. Das liefert $d - 2$ Gleichungen der Form

$$\alpha_1\gamma_j^{(1)}(\varepsilon_1^{(1)})^{n_1} \cdots (\varepsilon_{r+s-1}^{(1)})^{n_{r+s-1}} + \ldots + \alpha_d\gamma_j^{(d)}(\varepsilon_1^{(d)})^{n_1} \cdots (\varepsilon_{r+s-1}^{(d)})^{n_{r+s-1}} = 0\,.$$

Für Tupel $(n_1, \ldots, n_{r+s-1})$ in einer fixierten Restklasse modulo $p - 1$ können die Produkte als Potenzreihen in $r + s - 1$ Unbestimmten geschrieben werden. Auf die entstehenden Gleichungen kann man dann wieder versuchen, Satz 12.13 anzuwenden.

**BSP** Thue-Gleichung vom Grad 4

12.15. **Beispiel.** Wir betrachten wieder die Gleichung

$$x^4 - 5y^4 = 1$$

und $p = 101$. Es stellt sich heraus, dass $\varepsilon^{50}$ und $\eta^{50}$ sich unter allen vier Einbettungen $R \to \mathbb{Z}_{101}$ sich auf ein Element $\equiv 1 \bmod 101$ abbilden; es genügt also, $(m_0, n_0) \in \{0, 1, \ldots, 49\}^2$ zu betrachten. Die Rechnung, die wir schon für Beispiel 12.14 durchgeführt haben, zeigt, dass zehn solche Paare zu einer Lösung modulo 101 führen, nämlich

$$(0,0)\,,\ (0,1)\,,\ (2,49)\,,\ (6,43)\,,\ (17,32)\,,\ (25,25)\,,$$
$$(25,26)\,,\ (27,14)\,,\ (31,18) \quad \text{und} \quad (42,7)\,.$$

Die Rechnung modulo $q = 401$ lässt sich erweitern und zeigt, dass davon nur die drei Paare

$$(0,0)\,, \quad (0,1)\,, \quad \text{und} \quad (2,49)$$

möglich sind, die den bekannten Lösungen $\pm(1,0)$, $\pm(3,-2)$ und $\pm(3,2)$ entsprechen.

Die relevanten $p$-adischen Logarithmen sind

$$\begin{pmatrix} \log((\varepsilon^{(1)})^{50}) & \cdots & \log((\varepsilon^{(4)})^{50}) \\ \log((\eta^{(1)})^{50}) & \cdots & \log((\eta^{(4)})^{50}) \end{pmatrix} =: 101 \begin{pmatrix} \lambda_1 & \cdots & \lambda_4 \\ \mu_1 & \cdots & \mu_4 \end{pmatrix}$$
$$\equiv 101 \cdot \begin{pmatrix} 34 & -34 & -34 & 34 \\ -1 & -29 & -39 & -32 \end{pmatrix} \bmod 101^2\,.$$

Die Gleichungen, die wir durch Elimination von $x$ und $y$ und die Substitution $m = m_0 + 50u$, $n = n_0 + 50v$ erhalten, haben die Form

$$\sum_{i=1}^{4} \alpha_i (\varepsilon^{(i)})^{m_0} (\eta^{(i)})^{n_0} \exp(101(\lambda_i u + \mu_i v)) = 0$$
$$\sum_{i=1}^{4} \beta_i (\varepsilon^{(i)})^{m_0} (\eta^{(i)})^{n_0} \exp(101(\lambda_i u + \mu_i v)) = 0$$

mit

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix} \equiv \begin{pmatrix} 10 & -11 & 0 & 1 \\ 9 & -10 & 1 & 0 \end{pmatrix} \bmod 101\,.$$

Man prüft dann leicht nach, dass in allen drei verbleibenden Fällen das Kriterium von Satz 12.13 (3) anwendbar ist (nachdem man die Gleichungen durch $p = 101$ geteilt hat). Damit folgt wieder, dass die sechs bekannten Lösungen alle Lösungen der Gleichung sind. ♣

Für Thue-Gleichungen vom Grad 4 (mit $s = 1$; im Fall $s = 2$ ist $F(x, y)$ definit, und man bekommt leicht eine Schranke für die Lösungen) führen beide Ansätze auf ein System aus zwei Potenzreihen in zwei Unbekannten. Ist der Grad größer, dann ist es eher günstiger, den ersten Ansatz zu verwenden, da man hier stets bei einem System von Potenzreihen-Gleichungen in zwei Unbekannten landet, während die Anzahl der Unbekannten (und Gleichungen) beim zweiten Ansatz mit dem Grad der Thue-Gleichung wächst.

12.16. **Einige abschließende Bemerkungen.**

(1) Es ist nicht wirklich nötig, eine Primzahl $p$ zu verwenden mit der Eigenschaft, dass $f(x)$ modulo $p$ in verschiedene Linearfaktoren zerfällt. Man kann eine beliebige Primzahl $p \nmid \operatorname{disc}(f)$ verwenden und statt in $\mathbb{Z}_p^d$ im Ring $\mathbb{Z}_p[\theta] = \mathbb{Z}_p[x]/\langle f\rangle$ arbeiten. (Falls $f$ modulo $p$ in verschiedene Linearfaktoren zerfällt, dann ist nach dem Chinesischen Restsatz der Ring $\mathbb{Z}_p[\theta]$ zum Ring $\mathbb{Z}_p^d$ isomorph.) Der Faktorring $\mathbb{Z}_p[\theta]/\langle p\rangle \cong \mathbb{Z}[\theta]/\langle p\rangle$ ist im Allgemeinen kein Körper, aber ein endlicher (kommutativer) Ring. Seine Einheitengruppe ist endlich, woraus folgt, dass es $N \in \mathbb{Z}_{>0}$ gibt mit $\varepsilon_j^N \equiv 1 \bmod p$ in $\mathbb{Z}[\theta]$ für alle $j$. Mit diesem $N$ anstelle von $p-1$ kann man dann im zweiten Ansatz arbeiten. In jedem Fall kann man die Potenzreihen mit Koeffizienten in $\mathbb{Z}_p[\theta]$, die man erhält, auch als ein Element von $\mathbb{Z}_p[\![\boldsymbol{x}]\!][\theta]$ auffassen. Schreibt man das Produkt $\gamma_j\varepsilon_1^{n_1}\cdots\varepsilon_{r+s-1}^{n_{r+s-1}}$ (für eine bestimmte Restklasse der $n_j$ modulo $N$) in dieser Form, dann erhält man die relevanten Gleichungen, indem man die Koeffizienten von $\theta^2,\dots,\theta^{d-1}$ null setzt. Siehe Aufgabe (1) auf dem letzten Übungsblatt.

A. Baker
1939–2018



(2) Thue hat 1909 als Erster bewiesen, dass Thue-Gleichungen stets nur endlich viele Lösungen haben. Allerdings ist dieser Beweis nicht „effektiv“, d.h., er führt nicht zu einem Algorithmus, der (wenigstens im Prinzip) die Lösungsmenge bestimmt. Alan Baker gab 1967 einen neuen Beweis, der auf seinen Resultaten über „Linearformen in Logarithmen“ beruht (für die er 1970 die Fields-Medaille bekam). Dieser Beweis ist effektiv; er ergibt eine berechenbare Schranke für $|x|$ und $|y|$ in einer Lösung bzw. eine Schranke für die Exponenten $n_j$ in der Gleichung

$$x - y\theta = \gamma\varepsilon_1^{n_1}\cdots\varepsilon_{r+s-1}^{n_{r+s-1}}\,.$$

Diese Schranken sind viel zu groß (typischerweise von der Art $10^{\text{einige }100}$ für die Exponenten $n_k$, nach etlichen Verbesserungen seit Bakers Arbeiten), um direkt für eine explizite Lösung anwendbar zu sein. Es gibt jedoch ein Reduktionsverfahren, mit dessen Hilfe man die Schranken in der Praxis so weit verkleinern kann, dass man den verbleibenden Suchraum in vernünftiger Zeit absuchen kann. Dieses Lösungsverfahren ist in Computeralgebrasystemen wie Magma implementiert. Es lässt sich verallgemeinern auf sogenannte Thue-Mahler-Gleichungen

$$F(x,y) = cp_1^{m_1}\cdots p_k^{m_k}\,.$$

Hier sind $p_1,\dots,p_k$ verschiedene Primzahlen, und man sucht Lösungen mit $x,y\in\mathbb{Z}$ und $m_1,\dots,m_k\in\mathbb{Z}_{\ge 0}$.

(3) Für Gleichungen der Art wie wir sie in den Beispielen 12.14 und 12.15 betrachtet haben, gibt es ein sehr starkes allgemeines Resultat: Mike Bennett[12] hat gezeigt, dass die Gleichung

$$|ax^n - by^n| = 1$$

mit $a,b\in\mathbb{Z}\setminus\{0\}$ und $n\ge 3$ höchstens eine Lösung $(x,y)\in\mathbb{Z}_{>0}^2$ hat. Daraus folgt dann zum Beispiel sofort, dass $(\pm1,0)$ und $(\pm3,\pm2)$ alle Lösungen von $x^4-5y^4=1$ sind.

(4) Die $p$-adische Lösungsmethode, die wir in diesem Abschnitt diskutiert haben, wurde von dem norwegischen Mathematiker Thoralf Skolem in den 1930er Jahren entwickelt. Er benutzte sie dafür, die Endlichkeit der Lösungsmenge einer

[12] M. Bennett: *Rational approximation to algebraic numbers of small height: the Diophantine equation* $|ax^n - by^n| = 1$, J. reine angew. Math. **535** (2001), 1–49.

Thue-Gleichung mit $s \geq 1$ zu zeigen. Was man dazu über die hier entwickelte Theorie hinaus noch braucht, ist ein Argument, das zeigt, dass das $p$-adische Gleichungssystem eine Lösungsmenge hat, die aus (dann endlich vielen) isolierten Punkten besteht, also keine positiv-dimensionalen Komponenten hat. Dieser Beweis ist allerdings auch nicht effektiv, denn man kann ohne weitere Informationen nicht entscheiden, ob eine $p$-adische Lösung von einer ganzzahligen Lösung herkommt. (Genauer: Man kann nicht zeigen, dass das nicht der Fall ist.) In der Praxis kann man dieses Problem umgehen, indem man weitere Bedingungen betrachtet, die von einer Betrachtung modulo anderer Primzahlen kommen (so haben wir es auch in den Beispielen gemacht). Man kann auch die Schranken verwenden, die sich aus dem Beweis von Baker ergeben. Diese sagen einem, wie genau man eine Lösung $p$-adisch approximieren muss, um zeigen zu können, dass die Lösung nicht ganzzahlig ist (weil es keine ganze Zahl unterhalb der Schranke gibt, die $p$-adisch hinreichend nahe an der Lösung des Gleichungssystems liegt).

## Literatur

[IR] Kenneth Ireland und Michael I. Rosen: *A classical introduction to modern number theory,* Springer Graduate texts in mathematics **84**, 2nd edition, 1990. Buch online (aus dem UBT-Netz)

[Z] Heinz-Dieter Ebbinghaus et al.: *Zahlen,* Springer Grundwissen Mathematik **1**, 3. verb. Auflage, 1992.

[Sch] Alexander Schmidt: *Einführung in die algebraische Zahlentheorie,* Springer-Verlag 2007. Buch online (aus dem UBT-Netz)