Diophantische Gleichungen

Wintersemester 2008/2009

Universität Bayreuth

Michael Stoll

Inhaltsverzeichnis

1. Einleitung und Beispiele

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

1.1. **Definition.** Eine *diophantische Gleichung* ist eine algebraische Gleichung (in mehreren Variablen), die in *ganzen* oder *rationalen* Zahlen gelöst werden soll.

Eine *algebraische Gleichung* 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. Manchmal betrachtet man auch Gleichungen, in denen ein oder mehrere Exponenten als (positive ganzzahlige) Variable auftreten.

Die Namensgebung erfolgte zu Ehren von *Diophant(os)* von Alexandria. Man weiß recht wenig über ihn selbst.[1] 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 Rätselaufgabe, die sich auf sein Alter bezieht und aus einer Sammlung stammt, die um 500 entstanden ist:

> 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.

[1] http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Diophantus.html

Hier ist eine (recht willkürliche) Auswahl an Beispielen.

(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 $a = b = 0$ und $c \neq 0$).

Diese einfache lineare Gleichung ist lösbar genau dann, wenn der ggT von a und b ein Teiler von c ist. Ist (x_0, y_0) eine Lösung, dann sind alle weiteren von der Form $(x_0 + tb', y_0 - ta')$, 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 den selben 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, so dass 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 (siehe unten), 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. Siehe § 6 in diesem Skript.

(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 in § 10 sehen, dass es stets Lösungen gibt, und dass sie sich aus einer „Fundamentallösung“ erzeugen lassen. Für die Beispielgleichung ist die kleinste Lösung

$$X = 25052977273092427986049\,, \qquad Y = 1238789998647218582160\,.$$

(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

$$X = \frac{2263582143321421502100209233517777}{11981673410095561^2}$$

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

und wurde von mir vor einigen Jahren entdeckt.[2]

(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 § 2 für den Fall $(p, q, r) = (4, 4, 2)$).

[2] http://listserv.nodak.edu/cgi-bin/wa.exe?A2=ind0201&L=nmbrthry&T=0&F=&S=&P=1432

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.[3] 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!

(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.[4] Die Lösungen der Beispielgleichung mit $X > 4$ sind

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

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 $n \in \{3, 4, 5, 7, 15\}$.

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 in die 1970er Jahre, bis schließlich Yuri Matiyasevich, aufbauend auf wesentlicher Vorarbeit von Putnam, Davis und Julia Robinson, beweisen konnte, dass ein solcher Algorithmus **nicht** existiert.

So ein Beweis war erst möglich, nachdem der Begriff der Berechenbarkeit theoretisch gefasst und ausreichend verstanden war. Er beruht darauf, dass man logische Probleme, von denen man weiß, dass es unlösbare gibt, in diophantische Gleichungen übersetzen kann.

---

[3] 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**, 103–158 (2007).

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

2. Appetithappen

Bevor wir uns dem systematischen Studium einiger Arten von diophantischen Gleichungen zuwenden, möchte ich Ihnen vollständige Lösungen von zwei diophantischen 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 und b teilerfremde ganze Zahlen, so dass $ab = c^2$ ist (mit $c \in \mathbb{Z}$), dann gibt es (teilerfremde) ganze Zahlen u und v, so dass entweder $a = u^2$, $b = v^2$, $c = uv$ oder $a = -u^2$, $b = -v^2$, $c = uv$ ist.*

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 linken Seite c^2, also ist der Exponent von p in a gerade. Da jede Primzahl mit einem geraden Exponenten in a vorkommt, gibt es $u \in \mathbb{Z}$, so dass $a = \pm u^2$ ist. Ebenso gibt es $v \in \mathbb{Z}$ mit $b = \pm v^2$. Da $ab = c^2 > 0$, müssen die Vorzeichen gleich sein. Außerdem folgt $c = \pm uv$, und wir können (z.B.) falls nötig 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:

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

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

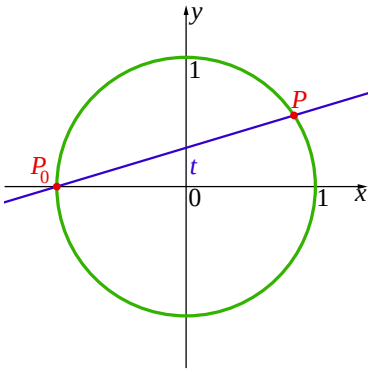
mit $S, T \in \mathbb{Z}$ teilerfremd und von verschiedener Parität.

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 ∞.

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 wieder

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

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, so dass sie linear wird.

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, so dass

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

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

$$Y^2 = U^4 - 4V^4\,.$$

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 $A, B > 0$, so dass

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

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

3. Grundlagen

In diesem Abschnitt stellen wir die grundlegenden Begriffe und Sätze über den Ring der ganzen Zahlen, das Rechnen mit Kongruenzen und die Restklassenringe $\mathbb{Z}/m\mathbb{Z}$ und speziell die endlichen Körper $\mathbb{F}_p$ zusammen.

Die Basis für alles weitere ist durch die folgenden grundlegenden Eigenschaften der ganzen Zahlen $\mathbb{Z}$ gegeben:

(1) $\mathbb{Z}$ ist ein *Integritätsring* (also ein kommutativer Ring mit 1, in dem aus $ab = 0$ stets $a = 0$ oder $b = 0$ folgt).
(2) $\mathbb{Z}_{\geq 0}$ ist *wohlgeordnet:* Jede nichtleere Menge nichtnegativer ganzer Zahlen hat ein kleinstes Element.
(3) In $\mathbb{Z}$ gilt das *Archimedische Prinzip:* Ist $n > 0$, dann gibt es für jedes $m \in \mathbb{Z}$ ein $k \in \mathbb{Z}$ mit $kn > m$.

Die *Teilbarkeitsrelation* spielt eine wichtige Rolle.

3.1. **Definition.** Seien a und b ganze Zahlen. Wir sagen „a teilt b“, geschrieben

$$a \mid b\,,$$

wenn es eine ganze Zahl c gibt mit $b = ac$. In diesem Fall sagen wir auch, „a ist ein Teiler von b“ oder „b ist ein Vielfaches von a“.

Es gelten die folgenden leicht zu beweisenden Aussagen (für alle $a, b, c \in \mathbb{Z}$).

(1) $a \mid a$, $1 \mid a$, $a \mid 0$.
(2) Aus $0 \mid a$ folgt $a = 0$.
(3) Aus $a \mid 1$ folgt $a = \pm 1$.
(4) Aus $a \mid b$ und $b \mid c$ folgt $a \mid c$.
(5) Aus $a \mid b$ folgt $a \mid bc$.
(6) Aus $a \mid b$ und $a \mid c$ folgt $a \mid b \pm c$.
(7) Aus $a \mid b$ und $|b| < |a|$ folgt $b = 0$.
(8) Aus $a \mid b$ und $b \mid a$ folgt $a = \pm b$.

3.2. **Definition.** Wir sagen „d ist der größte gemeinsame Teiler (ggT) von a und b“ und schreiben

$$d = \mathrm{ggT}(a, b)\,,$$

wenn $d \mid a$ und $d \mid b$, $d \geq 0$, und wenn für alle $k \in \mathbb{Z}$ mit $k \mid a$ und $k \mid b$ gilt, dass $k \mid d$.

Analog sagen wir „m ist das kleinste gemeinsame Vielfache (kgV) von a und b“ und schreiben

$$m = \mathrm{kgV}(a, b)\,,$$

wenn $a \mid m$ und $b \mid m$, $m \geq 0$, und wenn für alle $n \in \mathbb{Z}$ mit $a \mid n$ und $b \mid n$ gilt, dass $m \mid n$.

Auf die gleiche Weise können wir den größten gemeinsamen Teiler und das kleinste gemeinsame Vielfache einer beliebigen Menge S von ganzen Zahlen definieren. ggT und kgV haben folgende einfach zu beweisende Eigenschaften:

(1) $\mathrm{ggT}(\emptyset) = 0$, $\mathrm{kgV}(\emptyset) = 1$.
(2) $\mathrm{ggT}(S_1 \cup S_2) = \mathrm{ggT}(\mathrm{ggT}(S_1), \mathrm{ggT}(S_2))$,
$\mathrm{kgV}(S_1 \cup S_2) = \mathrm{kgV}(\mathrm{kgV}(S_1), \mathrm{kgV}(S_2))$.
(3) $\mathrm{ggT}(\{a\}) = \mathrm{kgV}(\{a\}) = |a|$.
(4) $\mathrm{ggT}(ac, bc) = |c|\ \mathrm{ggT}(a, b)$.
(5) $\mathrm{ggT}(a, b) = \mathrm{ggT}(a, ka + b)$.

Wie kann man den ggT zweier gegebener ganzer Zahlen berechnen? Der Schlüssel dazu liegt in der letzten Eigenschaft in der Liste oben. Zunächst brauchen wir aber die „Division mit Rest“.

3.3. **Lemma.** *Sind a und b ganze Zahlen mit $b \neq 0$, dann gibt es eindeutig bestimmte ganze Zahlen q („Quotient“) und r („Rest“) mit $0 \leq r < |b|$ und $a = bq + r$.*

Beweis. Existenz: Sei $S = \{a - kb \mid k \in \mathbb{Z}, a - kb \geq 0\}$. Dann ist $S \subset \mathbb{Z}_{\geq 0}$ nicht leer und hat demnach ein kleinstes Element $r = a - qb$ (für ein $q \in \mathbb{Z}$). Nach Definition von S ist $r \geq 0$, und wenn $r \geq |b|$ wäre, dann wäre $r - |b|$ ebenfalls in S, und r könnte nicht das kleinste Element von S sein.

Eindeutigkeit: Wenn $a = bq + r = bq' + r'$ mit $0 \leq r, r' < |b|$, dann folgt $b \mid r - r'$ und $0 \leq |r - r'| < |b|$, also muss $r = r'$ sein. Es folgt $bq = bq'$, also auch $q = q'$ (denn $b \neq 0$). □

3.4. **Algorithmus.** (Euklidischer Algorithmus)
a und b seien ganze Zahlen.

(1) (Initialisierung) Setze $n = 0$, $a_0 = |a|$, $b_0 = |b|$.
(2) (Ende) Wenn $b_n = 0$, gib a_n als Ergebnis aus.
(3) (Division) Finde q_n und r_n mit $a_n = b_n\, q_n + r_n$ und $0 \leq r_n < b_n$.
(4) (Neue Werte) Setze $a_{n+1} = b_n$, $b_{n+1} = r_n$.
(5) (Wiederholen) Ersetze n durch $n + 1$ und gehe zu Schritt 2.

Zunächst sehen wir, dass $0 \leq b_{n+1} < b_n$ gilt, so dass nach endlich vielen Durchläufen $b_n = 0$ sein muss und der Algorithmus ein Ergebnis liefert. Wir behaupten, dass dieses Ergebnis gerade $\mathrm{ggT}(a, b)$ ist.

Beweis. Wir zeigen, dass für alle n, die im Algorithmus vorkommen, $\mathrm{ggT}(a_n, b_n) = \mathrm{ggT}(a, b)$ gilt. Daraus folgt die Behauptung wegen $a_n = \mathrm{ggT}(a_n, 0) = \mathrm{ggT}(a_n, b_n)$ für das letzte n.

Zu Beginn ($n = 0$) haben wir $\mathrm{ggT}(a_0, b_0) = \mathrm{ggT}(|a|, |b|) = \mathrm{ggT}(a, b)$. Jetzt nehmen wir an, dass wir bereits $\mathrm{ggT}(a_n, b_n) = \mathrm{ggT}(a, b)$ gezeigt haben und dass $b_n \neq 0$ ist (d.h., wir sind noch nicht fertig). Dann ist

$$\mathrm{ggT}(a_{n+1}, b_{n+1}) = \mathrm{ggT}(b_n, a_n - b_n q_n) = \mathrm{ggT}(b_n, a_n) = \mathrm{ggT}(a_n, b_n)$$

wegen Eigenschaft (5) des ggT. □

3.5. **Bemerkung.** Man kann zeigen (Übungsaufgabe!), dass die Anzahl der Schleifendurchläufe $\leq C \max\{\log(1+|a|), \log(1+|b|)\}$ ist für eine Konstante C. Für Zahlen a, b mit $|a|, |b| < N$ ist damit der Gesamtaufwand höchstens von der Größenordnung $O\big((\log N)^3\big)$.

3.6. **Satz.** *Seien $a, b \in \mathbb{Z}$. Die ganzen Zahlen der Form $xa + yb$ mit $x, y \in \mathbb{Z}$ sind genau die Vielfachen von $d = \mathrm{ggT}(a, b)$. Insbesondere gibt es $x, y \in \mathbb{Z}$ mit $d = xa + yb$.*

Beweis. Da d sowohl a als auch b teilt, muss d auch $xa + yb$ teilen. Für den Beweis der Gegenrichtung genügt es zu zeigen, dass d in der Form $xa + yb$ geschrieben werden kann. Dies folgt aus dem Euklidischen Algorithmus: Sei N der letzte Wert von n. Dann ist $d = 1 \cdot a_N + 0 \cdot b_N$. Außerdem folgt aus $d = x_{n+1}a_{n+1} + y_{n+1}b_{n+1}$, dass $d = y_{n+1}a_n + (x_{n+1} - q_n y_{n+1})b_n$, also bekommen wir mit $x_n = y_{n+1}$ und $y_n = x_{n+1} - q_n y_{n+1}$, dass $d = x_n a_n + y_n b_n$. Am Ende dieser rückwärts laufenden Induktion bekommen wir $d = x_0 a_0 + y_0 b_0$. □

Es gibt eine einfache Erweiterung des Euklidischen Algorithmus, die zusätzlich zum ggT auch ganze Zahlen x und y bestimmt, so dass $\mathrm{ggT}(a, b) = xa + yb$.

3.7. **Algorithmus.** (Erweiterter Euklidischer Algorithmus)
a und b seien ganze Zahlen.

(1) (Initialisierung) Setze $n = 0$, $a_0 = |a|$, $b_0 = |b|$,
$x_0 = \mathrm{sign}(a)$, $y_0 = 0$, $u_0 = 0$, $v_0 = \mathrm{sign}(b)$.
(2) (Ende) Wenn $b_n = 0$, gib (a_n, x_n, y_n) als Ergebnis zurück.
(3) (Division) Finde q_n und r_n mit $a_n = b_n\, q_n + r_n$ und $0 \leq r_n < b_n$.
(4) (Neue Werte) Setze $a_{n+1} = b_n$, $b_{n+1} = r_n$,
$x_{n+1} = u_n$, $y_{n+1} = v_n$, $u_{n+1} = x_n - u_n\, q_n$, $v_{n+1} = y_n - v_n\, q_n$.
(5) (Wiederholen) Ersetze n durch $n + 1$ und gehe zu Schritt 2.

Wie vorhin zeigt man durch Induktion, dass $a_n = x_n\, a + y_n\, b$ und $b_n = u_n\, a + v_n\, b$. Das Ergebnis (d, x, y) erfüllt also $xa + yb = d = \mathrm{ggT}(a, b)$.

3.8. **Lemma.** *Wenn $n \mid ab$ und $\mathrm{ggT}(n, a) = 1$, dann $n \mid b$.*

Beweis. Nach Satz 3.6 gibt es x und y mit $xa + yn = 1$. Wir multiplizieren mit b und erhalten $b = xab + ynb$. Da n ein Teiler von ab ist, teilt n die rechte Seite und damit auch b. □

Als nächstes behandeln wir Primzahlen und die Primfaktorzerlegung.

3.9. **Definition.** Eine positive ganze Zahl p heißt *Primzahl* (oder ist *prim*), wenn $p > 1$ ist, und 1 und p die einzigen positiven Teiler von p sind.

(Das ist eigentlich die Definition eines „irreduziblen Elements“ in einem Ring.)

3.10. **Lemma.** *Sei p eine Primzahl, und seien a, b ganze Zahlen mit $p \mid ab$. Dann gilt $p \mid a$ oder $p \mid b$.*

(Das ist die eigentliche Definition eines „Primelements“ in einem Ring!)

Beweis. Wir können annehmen, dass $p \nmid a$ (sonst sind wir schon fertig). Dann ist $\mathrm{ggT}(a, p) = 1$ (weil der andere positive Teiler p von p nach Annahme a nicht teilt). Aus Lemma 3.8 folgt dann, dass p b teilt. □

Sei jetzt $n > 0$ eine positive ganze Zahl. Dann ist entweder $n = 1$, oder n ist eine Primzahl, oder n hat einen „echten“ Teiler d, d.h. mit $1 < d < n$. Damit ist $n = de$, wo ebenfalls $1 < e < n$. Wenn wir in der gleichen Weise mit d und e fortfahren, bekommen wir schließlich eine Darstellung von n als Produkt von Primzahlen. (Für $n = 1$ ist das das leere Produkt von Primzahlen; ein leeres Produkt hat definitionsgemäß den Wert 1.)

3.11. **Satz.** *Diese Primfaktorzerlegung von n ist (bis auf die Reihenfolge der Faktoren) eindeutig.*

Beweis. Induktion nach n. Für $n = 1$ ist die Aussage klar (es gibt nur eine leere Menge von Primzahlen). Wir können also annehmen, dass $n > 1$ ist und dass die Aussage für alle kleineren n stimmt. Aus

$$n = p_1 p_2 \dots p_k = q_1 q_2 \dots q_l$$

(mit Primzahlen p_i, q_j und $k, l \geq 1$) folgt, dass p_1 das Produkt der q_j teilt, also muss p_1 nach Lemma 3.10 einen der Faktoren q_j teilen. Nachdem wir die q_j evtl. umsortiert haben, können wir annehmen, dass $p_1 \mid q_1$. Da $p_1 \neq 1$ und q_1 prim ist, folgt $p_1 = q_1$. Sei $n' = n/p_1 = n/q_1$; dann ist

$$n' = p_2 p_3 \dots p_k = q_2 q_3 \dots q_l\,.$$

Da $n' < n$, folgt aus der Induktionsannahme, dass $k = l$ und dass wir die q_j so umordnen können, dass $p_j = q_j$ für $j = 2, \dots, k$. Damit ist die Aussage auch für n gezeigt. □

3.12. **Definition.** Aus dem Satz folgt, dass jede ganze Zahl $n \neq 0$ eindeutig geschrieben werden kann als

$$n = \pm \prod_p p^{v_p(n)},$$

wobei das Produkt über alle Primzahlen p läuft und die Exponenten $v_p(n)$ nichtnegative ganze Zahlen sind, die nur für endliche viele p von null verschieden sind. Dadurch ist $v_p(n)$ für Primzahlen p und ganze Zahlen $n \neq 0$ definiert. $v_p(n)$ heißt die *Bewertung* von n bei p.

Es gelten die folgenden leicht zu beweisenden Eigenschaften.

(1) $v_p(mn) = v_p(m) + v_p(n)$.
(2) $m \mid n \iff \forall p : v_p(m) \leq v_p(n)$.
(3) $\mathrm{ggT}(m,n) = \prod_p p^{\min(v_p(m),v_p(n))}$, $\quad \mathrm{kgV}(m,n) = \prod_p p^{\max(v_p(m),v_p(n))}$.
(4) $v_p(m+n) \geq \min(v_p(m), v_p(n))$, mit Gleichheit im Fall $v_p(m) \neq v_p(n)$.

Aus Eigenschaft (3) folgt $\mathrm{ggT}(m,n)\ \mathrm{kgV}(m,n) = mn$ für positive m, n. Für $m, n \in \mathbb{Z}$ beliebig haben wir $\mathrm{ggT}(m,n)\ \mathrm{kgV}(m,n) = |mn|$.

Wenn wir noch $v_p(0) = +\infty$ setzen, dann gelten alle Eigenschaften für alle ganzen Zahlen m, n (einschließlich 0). Dabei verwenden wir die üblichen Regeln wie $\min\{e, \infty\} = e$, $e + \infty = \infty$ etc.

Wir können die Bewertung v_p von $\mathbb{Z}$ auf $\mathbb{Q}$ fortsetzen, indem wir

$$v_p\left(\frac{r}{s}\right) = v_p(r) - v_p(s)$$

definieren. Damit ist v_p auf $\mathbb{Q}$ wohldefiniert, und Eigenschaften (1) und (4) oben gelten auch für rationale Zahlen (Übungsaufgabe). Eine rationale Zahl x ist genau dann ganz, wenn $v_p(x) \geq 0$ für alle Primzahlen p gilt.

Als nächstes behandeln wir Kongruenzen.

3.13. **Definition.** Seien a, b und n ganze Zahlen mit $n \neq 0$. Wir sagen, „a ist *kongruent* zu b *modulo* n“ und schreiben

$$a \equiv b \bmod n,$$

wenn n die Differenz $a - b$ teilt.

3.14. **Kongruenz ist eine Äquivalenzrelation.**
Für $a, b, c, n \in \mathbb{Z}$ mit $n \neq 0$ gilt:

(1) $a \equiv a \bmod n$.
(2) Aus $a \equiv b \bmod n$ folgt $b \equiv a \bmod n$.
(3) Aus $a \equiv b \bmod n$ und $b \equiv c \bmod n$ folgt $a \equiv c \bmod n$.

Beweis.

(1) n teilt $a - a = 0$.
(2) Wenn $n \mid a - b$, dann auch $n \mid b - a$.
(3) Aus $n \mid a - b$ und $n \mid b - c$ folgt $n \mid (a - b) + (b - c) = a - c$.

□

Wir können $\mathbb{Z}$ also in *Kongruenzklassen* (oder auch *Restklassen*) mod n zerlegen: Wir schreiben

$$\bar{a} = a + n\mathbb{Z} = \{a + nx \mid x \in \mathbb{Z}\} = \{b \in \mathbb{Z} \mid a \equiv b \bmod n\}$$

(wenn wir $\bar{a}$ schreiben, muss n aus dem Zusammenhang klar sein) und

$$\mathbb{Z}/n\mathbb{Z} = \{a + n\mathbb{Z} \mid a \in \mathbb{Z}\}\,.$$

Es gilt dann

$$a \equiv b \bmod n \iff b \in \bar{a} \iff \bar{a} = \bar{b}\,.$$

3.15. **Satz.** *Die Abbildung*

$$\{0, 1, \ldots, n-1\} \longrightarrow \mathbb{Z}/n\mathbb{Z}\,, \quad r \longmapsto \bar{r} = r + n\mathbb{Z}$$

ist bijektiv. Insbesondere hat $\mathbb{Z}/n\mathbb{Z}$ genau n Elemente.

Beweis. Injektivität: Seien $0 \le r, s < n$ mit $\bar{r} = \bar{s}$. Dann gilt $r \equiv s \bmod n$, also $n \mid r - s$. Wegen $|r - s| < n$ folgt $r = s$.
Surjektivität: Sei $\bar{a}$ eine Kongruenzklasse. Wir schreiben $a = nq + r$ mit $0 \le r < n$. Dann ist $\bar{a} = \bar{r}$. □

Die Tatsache, dass der Repräsentant r einer Klasse $\bar{a}$ durch den Rest von a bei Division durch n gegeben ist, erklärt den Namen „Restklasse“.

3.16. **Die Restklassen mod n bilden einen kommutativen Ring.**
Wir definieren Addition und Multiplikation auf $\mathbb{Z}/n\mathbb{Z}$ durch

$$\bar{a} + \bar{b} = \overline{a + b}\,, \qquad \bar{a} \cdot \bar{b} = \overline{ab}$$

Wir müssen nachprüfen, dass diese Operationen wohldefiniert sind. Das bedeutet hier, dass aus $a \equiv a' \bmod n$ und $b \equiv b' \bmod n$ folgt, dass auch $a+b \equiv a'+b' \bmod n$ und $ab \equiv a'b' \bmod n$. Nun ist

$$(a + b) - (a' + b') = (a - a') + (b - b') \quad \text{teilbar durch } n,$$

und auch

$$ab - a'b' = (a - a')b + a'(b - b') \quad \text{ist teilbar durch } n.$$

Sobald diese Addition und Multiplikation wohldefiniert sind, folgt die Gültigkeit der Axiome für einen kommutativen Ring mit 1 aus ihrer Gültigkeit für $\mathbb{Z}$.

Wir nennen daher $\mathbb{Z}/n\mathbb{Z}$ auch den *Restklassenring* mod n.

3.17. **Kongruenzen sind nützlich.** Warum? Wenn wir modulo n rechnen, erhalten wir ein vergröbertes Abbild $\mathbb{Z}/n\mathbb{Z}$ der ganzen Zahlen, dessen Hauptvorteil darin besteht, dass es nur endlich viele Elemente gibt. Die Lösbarkeit von Gleichungen in $\mathbb{Z}/n\mathbb{Z}$ lässt sich also mit endlichem Aufwand überprüfen, im Gegensatz zur Lösbarkeit in $\mathbb{Z}$, wo wir im Prinzip unendlich viele Möglichkeiten durchprobieren müssten. Wir können also etwa nachweisen, dass eine Gleichung *nicht* in $\mathbb{Z}/n\mathbb{Z}$ lösbar ist; in vielen Fällen folgt daraus, dass es auch keine Lösung in $\mathbb{Z}$ gibt.

Wir können zum Beispiel die Gleichung $x^2 + y^2 - 15\,z^2 = 7$ betrachten. Hat sie ganzzahlige Lösungen? Das ist auf den ersten Blick schwer zu beantworten. Auf der anderen Seite können wir aber leicht feststellen, welche Werte die linke Seite mod 8 annehmen kann: Ein Quadrat ist stets kongruent zu 0, 1 oder 4 mod 8, und $-15 \equiv 1 \bmod 8$, also ist die linke Seite kongruent zu einer Summe dreier Quadrate, und man stellt schnell fest, dass dies niemals $\equiv 7 \bmod 8$ sein kann (alle anderen Restklassen können jedoch auftreten). Es gibt also keine Lösung in $\mathbb{Z}/8\mathbb{Z}$. Wenn

es eine Lösung in $\mathbb{Z}$ gäbe, dann wäre aber ihr Bild in $\mathbb{Z}/8\mathbb{Z}$ ebenfalls eine Lösung, also kann es auch keine ganzzahlige Lösung geben.

Beim Studium der Restklassenringe $\mathbb{Z}/n\mathbb{Z}$ stellt sich die Frage, welche ihrer Elemente invertierbar sind. Sei also $\bar{a} \in \mathbb{Z}/n\mathbb{Z}$. Wir müssen also untersuchen, wann sich die Kongruenz $ax \equiv 1 \bmod n$ lösen lässt. Das ist äquivalent dazu, dass es ganze Zahlen x und y gibt, so dass $ax + ny = 1$. Satz 3.6 sagt uns, dass das genau dann der Fall ist, wenn $\mathrm{ggT}(a, n) = 1$.

In diesem Fall sagen wir, dass a und n *teilerfremd* (oder auch *relativ prim*) sind, und wir schreiben gelegentlich dafür $a \perp n$. Der Erweiterte Euklidische Algorithmus 3.7 berechnet die passenden x und y; die Restklasse $\bar{x} \in \mathbb{Z}/n\mathbb{Z}$ ist dann das gesuchte Inverse (und die Berechnung von y können wir uns in diesem Zusammenhang sparen).

Ein Ring ist genau dann ein Körper, wenn alle von null verschiedenen Elemente invertierbar sind. Was bedeutet das für $\mathbb{Z}/n\mathbb{Z}$?

3.18. **Satz.** *Der Ring $\mathbb{Z}/n\mathbb{Z}$ ist genau dann ein Körper, wenn n eine Primzahl ist.*

Beweis. Das ist klar für $n = 1$ (ein Körper hat mindestens die zwei verschiedenen Elemente 0 und 1, und 1 ist keine Primzahl). Sei also $n > 1$. Dann ist $\mathbb{Z}/n\mathbb{Z}$ genau dann kein Körper, wenn es ein $a \in \mathbb{Z}$ gibt mit $n \nmid a$ und $d = \mathrm{ggT}(n, a) > 1$. Dann muss $1 < d < n$ ein echter Teiler von n sein, also ist n keine Primzahl. Wenn umgekehrt $d > 1$ ein echter Teiler von n ist, dann ist $\mathrm{ggT}(d, n) = d > 1$, und $\bar{d}$ ist nicht invertierbar, also ist $\mathbb{Z}/n\mathbb{Z}$ kein Körper. □

Wenn $\mathrm{ggT}(n, a) = 1$, dann heißt $\bar{a}$ auch eine *prime Restklasse* mod n; ihr eindeutig bestimmtes Inverses in $\mathbb{Z}/n\mathbb{Z}$ wird mit $\bar{a}^{-1}$ bezeichnet. Die primen Restklassen bilden eine Gruppe, nämlich die multiplikative Gruppe (oder Einheitengruppe) $(\mathbb{Z}/n\mathbb{Z})^\times$ des Rings $\mathbb{Z}/n\mathbb{Z}$.

Wenn $n = p$ eine Primzahl ist, schreiben wir für den Körper $\mathbb{Z}/p\mathbb{Z}$ auch $\mathbb{F}_p$; es ist dann $\mathbb{F}_p^\times = \mathbb{F}_p \setminus \{\bar{0}\}$; insbesondere $\#\mathbb{F}_p^\times = p - 1$.

Beachte, dass viele Autoren auch die Schreibweise $\mathbb{Z}_n$ anstelle von $\mathbb{Z}/n\mathbb{Z}$ (oder $\mathbb{Z}_p$ statt $\mathbb{F}_p$) verwenden. Wir werden das nicht tun, denn wir werden $\mathbb{Z}_p$ als Bezeichnung für den Ring der *p-adischen Zahlen* verwenden, siehe Abschnitt 8.

Die Ordnung der primen Restklassengruppe ist eine wichtige Größe.

3.19. **Definition.** Die *Eulersche ϕ-Funktion* ist für $n > 0$ definiert durch

$$\phi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times = \#\{a \in \mathbb{Z} \mid 0 \le a < n, a \perp n\}\,.$$

Hier ist eine kleine Tabelle:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\phi(n)$	1	1	2	2	4	2	6	4	6	4	10	4	12	6	8	8	16	6	18	8

Nach Satz 3.18 ist n prim genau dann, wenn $\phi(n) = n - 1$.

Wie können wir $\phi(n)$ für allgemeines n berechnen?

3.20. **Lemma.** *Sei p eine Primzahl und $e \geq 1$. Dann ist $\phi(p^e) = p^{e-1}(p-1)$.*

Beweis. Klar für $e = 1$. Für $e > 1$ beachte, dass $a \perp p^e \iff a \perp p \iff p \nmid a$, also ist

$$\begin{aligned}\phi(p^e) &= \#\{a \in \mathbb{Z} \mid 0 \leq a < p^e, p \nmid a\} \\ &= p^e - \#\{a \in \mathbb{Z} \mid 0 \leq a < p^e, p \mid a\} \\ &= p^e - p^{e-1} = (p-1)p^{e-1}\,.\end{aligned}$$

□

3.21. **Eine Rekursionsformel.** Wenn wir die Zahlen zwischen 0 (einschließlich) und n (ausschließlich) entsprechend ihrem ggT mit n zusammenfassen und dann abzählen, erhalten wir die Beziehung

$$\sum_{d|n} \phi\Big(\frac{n}{d}\Big) = \sum_{d|n} \phi(d) = n\,.$$

(Denn die Zahlen $0 \leq a < n$ mit $\mathrm{ggT}(a,n) = d$ sind gerade von der Form $a = bd$ mit $0 \leq b < n/d$ und $\mathrm{ggT}(b, n/d) = 1$.)

Dies kann als eine Rekursionsformel für $\phi(n)$ gelesen werden:

$$\phi(n) = n - \sum_{d|n, d<n} \phi(d)\,.$$

Wir berechnen zum Beispiel

$$\phi(1) = 1 - 0 = 1\,, \quad \phi(2) = 2 - \phi(1) = 1\,, \quad \phi(3) = 3 - \phi(1) = 2\,,$$
$$\phi(4) = 4 - \phi(2) - \phi(1) = 2\,, \quad \phi(6) = 6 - \phi(3) - \phi(2) - \phi(1) = 2\,, \quad \text{usw.}$$

Offenbar ist ϕ durch diese Rekursion eindeutig bestimmt: Wenn ganze (oder auch komplexe) Zahlen a_n für $n \geq 1$ die Gleichung

$$\sum_{d|n} a_d = n$$

erfüllen, dann gilt $a_n = \phi(n)$. Das gilt auch noch, wenn n eingeschränkt wird auf die Teiler einer festen Zahl N.

Was wir bisher gelernt haben, erlaubt uns, *lineare Kongruenzen* zu lösen: Wenn a, b und n gegeben sind, was sind dann die Lösungen x von

$$ax \equiv b \bmod n\,?$$

Anders gesagt, für welche $x \in \mathbb{Z}$ gibt es $y \in \mathbb{Z}$ mit $ax + ny = b$?

3.22. **Satz.** *Die Kongruenz $ax \equiv b \bmod n$ ist lösbar genau dann, wenn b ein Vielfaches von $\mathrm{ggT}(a,n)$ ist. In diesem Fall bilden die Lösungen eine Restklasse modulo $n/\mathrm{ggT}(a,n)$.*

Beweis. Nach Satz 3.6 ist die Bedingung $\mathrm{ggT}(a,n) \mid b$ notwendig und hinreichend für die Existenz von Lösungen. Sei also $g = \mathrm{ggT}(a,n)$ ein Teiler von b. Dann können wir die Gleichung $ax + ny = b$ durch g teilen und erhalten $a'x + n'y = b'$, wo wir $a = a'g$, $n = n'g$ und $b = b'g$ gesetzt haben. Dann ist $\mathrm{ggT}(a', n') = 1$, und wir können die Gleichung $\bar{a}'\,\bar{x} = \bar{b}'$ in $\mathbb{Z}/n'\mathbb{Z}$ nach $\bar{x}$ auflösen: $\bar{x} = \bar{b}'\,(\bar{a}')^{-1}$. Die Lösungsmenge ist also durch die Restklasse $\bar{b}'\,(\bar{a}')^{-1}$ modulo n' gegeben. □

Ein wichtiges Resultat ist der Chinesische Restsatz, der es uns erlaubt, Kongruenzen modulo verschiedener Zahlen n miteinander in Beziehung zu setzen.

Wir betrachten ein System von zwei simultanen Kongruenzen:

$$x \equiv a \bmod m\,, \qquad x \equiv b \bmod n$$

Gibt es eine Lösung? Wie sieht die Lösungsmenge aus? Ist $d = \mathrm{ggT}(m,n)$, dann ist $a \equiv b \bmod d$ offenbar eine notwendige Bedingung für die Lösbarkeit. Wenn $m \perp n$, dann ist dies keine Einschränkung. Tatsächlich gibt es dann auch immer Lösungen:

3.23. **Chinesischer Restsatz.** *Wenn m und n teilerfremd sind, dann hat das System*

$$x \equiv a \bmod m\,, \qquad x \equiv b \bmod n$$

stets Lösungen $x \in \mathbb{Z}$, und die Lösungen bilden eine Restklasse modulo mn.

Beweis. Da $m \perp n$, können wir nach Satz 3.6 u und v finden mit $mu + nv = 1$. Sei $x = anv + bmu$, dann ist

$$x = anv + bmu \equiv anv = a - amu \equiv a \bmod m$$

und analog $x \equiv b \bmod n$. Wir haben also eine Lösung gefunden. Wir zeigen noch, dass $y \in \mathbb{Z}$ genau dann eine Lösung ist, wenn $y \equiv x \bmod mn$: Es gilt

$$\begin{aligned} y \equiv a \bmod m\,, \quad y \equiv b \bmod m \;&\iff\; y \equiv x \bmod m\,, \quad y \equiv x \bmod n \\ &\iff\; m \mid y - x\,, \quad n \mid y - x \\ &\iff\; \mathrm{kgV}(m,n) = mn \mid y - x\,. \end{aligned}$$

□

Zum Beweis der Verallgemeinerung auf mehr als zwei simultane Kongruenzen brauchen wir noch ein kleines Lemma.

3.24. **Lemma.** *Seien $a, b, c \in \mathbb{Z}$ mit $a \perp b$ und $a \perp c$. Dann gilt auch $a \perp bc$.*

Beweis. Andernfalls gäbe es eine Primzahl p, die sowohl a als auch bc teilt. Dann teilt p aber auch b oder c, im Widerspruch zur Voraussetzung. □

3.25. **Chinesischer Restsatz.** *Wenn die Zahlen $m_1, m_2, \ldots, m_k$ paarweise teilerfremd sind, dann ist jedes System von Kongruenzen*

$$x \equiv a_1 \bmod m_1\,, \quad x \equiv a_2 \bmod m_2\,, \ldots\,, \quad x \equiv a_k \bmod m_k$$

lösbar, und die Lösungen bilden eine Restklasse modulo $m_1 m_2 \ldots m_k$.

Beweis. Induktion nach k. Der Fall $k = 1$ ist trivial. Sei also $k \geq 2$. Nach Satz 3.23 gibt es eine Lösung b des Systems

$$x \equiv a_{k-1} \bmod m_{k-1}\,, \quad x \equiv a_k \bmod m_k\,,$$

und jede Lösung ist $\equiv b \bmod m_{k-1}m_k$. Das ursprüngliche System ist also äquivalent zu

$$x \equiv a_1 \bmod m_1\,, \quad \ldots\,, \quad x \equiv a_{k-2} \bmod m_{k-2}\,, \quad x \equiv b \bmod m_{k-1}m_k\,.$$

Nach Lemma 3.24 sind $m_1, \ldots, m_{k-2}, m_{k-1}m_k$ paarweise teilerfremd. Wir können also die Induktionsannahme auf dieses System anwenden, was dann direkt die Behauptung liefert. □

Jetzt können wir auch die Frage beantworten, wann ein beliebiges System von zwei simultanen Kongruenzen eine Lösung hat.

3.26. **Satz.** *Das System*

$$x \equiv a \bmod m\,, \qquad x \equiv b \bmod n$$

ist genau dann lösbar, wenn $a \equiv b \bmod \mathrm{ggT}(m,n)$*. In diesem Fall bilden die Lösungen eine Restklasse modulo* $\mathrm{kgV}(m,n)$*.*

Beweis. Die Notwendigkeit der Bedingung hatten wir schon gesehen. Nach Satz 3.6 gibt es, wenn die Bedingung erfüllt ist, ganze Zahlen u und v mit $mu+nv = b-a$. Dann ist $x = a+mu = b-nv$ eine Lösung. Wie im Beweis von Satz 3.23 sehen wir, dass y genau dann eine weitere Lösung ist, wenn sowohl m als auch n die Differenz $y-x$ teilen. Das ist gleichbedeutend mit $y \equiv x \bmod \mathrm{kgV}(m,n)$. □

Wir geben noch eine mehr algebraische Formulierung des Chinesischen Restsatzes:

3.27. **Chinesischer Restsatz.** *Die Zahlen* $m_1, m_2, \ldots, m_k$ *seien paarweise teilerfremd. Dann ist der natürliche Ringhomomorphismus*

$$\mathbb{Z}/m_1m_2\ldots m_k\mathbb{Z} \longrightarrow \mathbb{Z}/m_1\mathbb{Z} \times \mathbb{Z}/m_2\mathbb{Z} \times \cdots \times \mathbb{Z}/m_k\mathbb{Z}$$

ein Isomorphismus. Insbesondere haben wir einen Isomorphismus der Einheitengruppen

$$(\mathbb{Z}/m_1m_2\ldots m_k\mathbb{Z})^\times \cong (\mathbb{Z}/m_1\mathbb{Z})^\times \times (\mathbb{Z}/m_2\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/m_k\mathbb{Z})^\times\,.$$

Beweis. Satz 3.25 sagt, dass der Homomorphismus bijektiv ist. □

3.28. **Eine Formel für $\phi(n)$.** Daraus können wir nun eine Formel für die Eulersche ϕ-Funktion ableiten: Der Chinesische Restsatz 3.27 hat zur Folge, dass

$$\phi(mn) = \phi(m)\,\phi(n) \qquad \text{falls } m \perp n.$$

Eine ähnliche Beziehung gilt für Produkte mit mehr Faktoren, wenn diese paarweise teilerfremd sind. Wir wenden das auf die Primfaktorzerlegung von n an und erhalten

$$\phi(n) = \prod_{p|n} p^{v_p(n)-1}(p-1) = n\prod_{p|n}\Big(1-\frac{1}{p}\Big)\,.$$

Die endlichen Körper $\mathbb{F}_p$ (und alle ihre Erweiterungen, d.h. die Körper der Charakteristik p) haben die schöne Eigenschaft, dass Potenzieren mit p, $x \mapsto x^p$, nicht nur eine multiplikative, sondern auch eine additive Abbildung ist.

3.29. **Satz.** („Freshman's Dream“) *Sei p eine Primzahl. Dann gilt für alle $x, y \in \mathbb{F}_p$, dass $(x+y)^p = x^p + y^p$. In anderen Worten: Sind $a, b \in \mathbb{Z}$, dann gilt*

$$(a+b)^p \equiv a^p + b^p \bmod p\,.$$

Beweis. Nach dem Binomialsatz ist

$$(a+b)^p - a^p - b^p = \binom{p}{1}a^{p-1}b + \binom{p}{2}a^{p-2}b^2 + \cdots + \binom{p}{k}a^{p-k}b^k + \cdots + \binom{p}{p-1}ab^{p-1}\,.$$

Alle Binomialkoeffizienten $\binom{p}{k}$ für $1 \le k \le p-1$ sind ganze Zahlen, die durch p teilbar sind (warum?), also ist die rechte Seite durch p teilbar, woraus die Behauptung folgt. □

3.30. **Satz.** („Kleiner Satz von Fermat“) *Sei p eine Primzahl und $a \in \mathbb{Z}$. Dann gilt $a^p \equiv a \bmod p$.*

Gilt zusätzlich $p \nmid a$, dann ist die Aussage äquivalent zu $a^{p-1} \equiv 1 \bmod p$.

Beweis. Für $p \mid a$ sind beide Seiten $\equiv 0 \bmod p$. Andernfalls ist $\bar{a} \in \mathbb{F}_p^\times$, und $\mathbb{F}_p^\times$ ist eine Gruppe der Ordnung $p-1$. Daher ist $\bar{a}^{p-1} = \bar{1}$, was zu $a^{p-1} \equiv 1 \bmod p$ äquivalent ist. □

Diese Aussage und ihr Beweis lassen sich verallgemeinern.

3.31. **Satz.** (Euler) *Sei $m \ge 1$ und $a \in \mathbb{Z}$ mit $a \perp m$. Dann gilt $a^{\phi(m)} \equiv 1 \bmod m$.*

3.32. **Beispiel.** Was ist $7^{11^{13}} \bmod 15$? Es ist $\phi(15) = 8$, also genügt es, $11^{13} \bmod 8$ zu kennen. $\phi(8) = 4$, also $11^{13} \equiv 3^{13} \equiv 3 \bmod 8$, und daher $7^{11^{13}} \equiv 7^3 = 343 \equiv 13 \bmod 15$.

3.33. **Beispiel.** Wir finden die letzten sechs Stellen der zur Zeit (Oktober 2008) größten bekannten Primzahl[5] $2^{43\,112\,609} - 1$. Da 2 nicht teilerfremd zu 10^6 ist, rechnen wir zunächst modulo 5^6. Wir haben $\phi(5^6) = 4 \cdot 5^5 = 12\,500$, und $43\,112\,609 \equiv 109 \bmod 12\,500$. Wir erhalten sukzessive

$$2^1 = 2, \quad 2^2 = 4, \quad 2^3 = 8, \quad 2^6 = 64, \quad 2^{12} = 4096, \quad 2^{13} = 8192,$$
$$2^{26} = 67108864 \equiv 15114 \bmod 5^6, \quad 2^{27} \equiv 30228 \equiv 14603 \bmod 5^6,$$
$$2^{54} \equiv 213247609 \equiv 13234 \bmod 5^6, \quad 2^{108} \equiv 175138756 \equiv 13756 \bmod 5^6,$$
$$2^{109} \equiv 27512 \equiv 11887 \bmod 5^6$$

Auf der anderen Seite ist $2^{43\,112\,609}$ natürlich durch 2^6 teilbar. Wir können mit Hilfe des Chinesischen Restsatzes die Restklasse mod 10^6 finden, die $\equiv 0 \bmod 2^6$ und $\equiv 11887 \bmod 5^6$ ist; wir erhalten 152512. Damit sind die gesuchten letzten sechs Ziffern 152 511.

Ein wichtiges Ergebnis gibt uns die Struktur der multiplikativen Gruppe $\mathbb{F}_p^\times$ des endlichen Körpers $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

[5] http://prime.haugk.co.uk/

3.34. **Satz.** *Sei K ein Körper und $G \subset K^\times$ eine endliche Untergruppe der multiplikativen Gruppe von K. Dann ist G zyklisch.*

Beweis. Sei $n = \#G$ die Ordnung von G. Angenommen, G ist nicht zyklisch. Dann gibt es einen echten Teiler d von n, so dass $g^d = 1$ für alle $g \in G$. [G ist isomorph zu einem Produkt

$$G \cong \mathbb{Z}/e_1\mathbb{Z} \times \mathbb{Z}/e_2\mathbb{Z} \times \cdots \times \mathbb{Z}/e_r\mathbb{Z}$$

von zyklischen Gruppen (die hier additiv geschrieben sind) mit $e_1 \geq 2$, $e_1 \mid e_2$, $\ldots$, $e_{r-1} \mid e_r$, und $r \geq 2$ (sonst wäre G zyklisch). Dann ist $d = e_r$ ein echter Teiler von $n = e_1 \cdots e_r$ und hat die verlangte Eigenschaft.] Das bedeutet, dass alle Elemente von G Nullstellen in K des Polynoms $X^d - 1$ sind. Ein Polynom vom Grad d kann in einem Körper aber höchstens d Nullstellen haben, ein Widerspruch zu $d < n$. Also muss G zyklisch sein. □

3.35. **Folgerung.** *Die Gruppe $\mathbb{F}_p^\times$ ist zyklisch.*

Beweis. $\mathbb{F}_p^\times$ ist eine endliche Untergruppe von $\mathbb{F}_p^\times$. □

Ist $G = \langle g \rangle$ zyklisch der Ordnung $n = \#G$, dann sind die Erzeuger von G gerade die Elemente g^a mit $a \perp n$. (Denn ein Element $h \in G$ erzeugt G genau dann, wenn $g = h^b$ ist für ein b. Ist $h = g^a$, dann heißt das $ab \equiv 1 \bmod n$, d.h. a muss modulo n invertierbar sein.) Die Erzeuger entsprechen also gerade den primen Restklassen modulo n, und es gibt $\phi(n)$ Erzeuger.

Insbesondere sehen wir, dass es $\phi(p-1)$ Erzeuger von $\mathbb{F}_p^\times$ gibt.

3.36. **Definition.** Sei p eine Primzahl. Dann heißt $g \in \mathbb{Z}$ *Primitivwurzel mod p*, wenn $\bar{g} \in \mathbb{F}_p$ ein Erzeuger der multiplikativen Gruppe $\mathbb{F}_p^\times$ ist.

Hier ist eine kleine Tabelle von Primitivwurzeln:

Primzahl p	$\phi(p-1)$	Primitivwurzeln
2	1	1
3	1	2
5	2	2, 3
7	2	3, 5
11	4	2, 3, 8, 9
13	4	2, 6, 7, 11

Eine berühmte Vermutung[6] von Emil Artin behauptet, dass jede ganze Zahl $g \neq -1$, die keine Quadratzahl ist, eine Primitivwurzel mod p ist für unendlich viele Primzahlen p. (Warum ist das für Quadratzahlen falsch?) Es gibt einen Beweis dafür unter der Annahme der „Extended Riemann Hypothesis“ (erweiterte Riemannsche Vermutung). Es genügt, die Behauptung für Primzahlen g zu beweisen. Das beste Resultat, das sich nicht auf unbewiesene Vermutungen stützt, sagt, dass die Aussage für alle Primzahlen g stimmt, *mit höchstens zwei Ausnahmen.*[7] Allerdings konnte die Aussage noch für kein einziges konkretes g bewiesen werden!

Die nächste offensichtliche Frage ist jetzt, wann $(\mathbb{Z}/m\mathbb{Z})^\times$ zyklisch ist. Da wir die Antwort im Folgenden nicht benötigen, werde ich nur das Resultat angeben.

[6] http://en.wikipedia.org/wiki/Artin_conjecture_on_primitive_roots

[7] D.R. Heath-Brown, *Artin's conjecture for primitive roots,* Quart. J. Math. Oxford Ser. (2) **37**, 27–38 (1986)

3.37. **Satz.** *Sei p eine ungerade Primzahl und g eine Primitivwurzel mod p. Dann können wir $h = g$ oder $h = g+p$ wählen, so dass $h+p^n\mathbb{Z}$ für alle $n \geq 1$ die Gruppe $(\mathbb{Z}/p^n\mathbb{Z})^\times$ erzeugt. Insbesondere ist $(\mathbb{Z}/p^n\mathbb{Z})^\times$ zyklisch.*

3.38. **Satz.** *Die Gruppe $(\mathbb{Z}/2^n\mathbb{Z})^\times$ ist zyklisch für $n \leq 2$. Für $n \geq 3$ ist $(\mathbb{Z}/2^n\mathbb{Z})^\times$ ein Produkt von zwei zyklischen Gruppen der Ordnung 2 und 2^{n-2}. Als Erzeuger können wir -1 und 5 wählen.*

3.39. **Satz.** *Die Gruppe $(\mathbb{Z}/m\mathbb{Z})^\times$ ist zyklisch genau dann, wenn $m = 1$, 2, 4, p^n oder $2p^n$ ist mit einer ungeraden Primzahl p und $n \geq 1$.*

4. Quadratische Reste und 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.

4.1. **Definition.** Sei 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*.

4.2. **Beispiele.**

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; dann ist jedes a mit $p \nmid a$ kongruent zu $g^k \bmod p$ (für ein k, das mod $p-1$ eindeutig bestimmt ist; insbesondere ist die Parität von k eindeutig bestimmt, da $p-1$ gerade ist). Wir schreiben $k = \log_{\bar{g}} \bar{a} \in \mathbb{Z}/(p-1)\mathbb{Z}$.

4.3. **Satz.** *Sei p eine ungerade Primzahl und $a \in \mathbb{Z}$, $p \nmid a$; 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) $\Rightarrow$ (1) $\Rightarrow$ (3) $\Rightarrow$ (2) und damit die Äquivalenz der Aussagen gezeigt. □

Daraus sehen wir bereits, dass es genau $(p-1)/2$ quadratische Restklassen und $(p-1)/2$ quadratische Nichtrestklassen mod p gibt.

4.4. **Bemerkung.** *Es gilt*

$$a \text{ qu. Nichtrest mod } p \iff k \text{ ist ungerade} \iff a^{(p-1)/2} \equiv -1 \bmod p\,.$$

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 $b \equiv -1 \bmod p$. □

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

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

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

4.6. **Bemerkung.** 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.[8]

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

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

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

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

Beweis. Wenn $p \mid a$, dann sind beide Seiten null mod p. In den anderen beiden Fällen folgt die Kongruenz aus Satz 4.3 und der nachfolgenden Bemerkung. Die Eindeutigkeitsaussage folgt daraus, dass das Legendre-Symbol nur die Werte $-1, 0, 1$ annimmt, die mod p alle verschieden sind (denn $p \geq 3$). □

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

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

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

Beweis. Nach Satz 4.8 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. □

4.10. **Bemerkung.** Insbesondere ist das Produkt von zwei quadratischen Nichtresten mod p ein quadratischer Rest mod p.

4.11. **Bemerkung.** Die wesentliche Aussage von Satz 4.9 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 $\mathbb{F}_p^\times$.

4.12. **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

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

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

Der einfachste Fall ist $a = -1$.

4.13. **Satz.** (Erstes Ergänzungsgesetz zum Quadratischen Reziprozitätsgesetz) *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 Satz 4.8 gilt

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

Da beide Seiten ± 1 sind, folgt Gleichheit. □

4.14. **Bemerkung.** 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 $(m!)^2 \equiv -1 \bmod p$, wenn m gerade, also $p \equiv 1 \bmod 4$ ist. Allerdings lässt sich $m!$ mod p nicht effizient berechnen, so dass 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.

4.15. **Lemma.** *Sei p eine ungerade Primzahl. Sei weiter $S \subset \mathbb{Z}$ eine Teilmenge mit $\#S = (p-1)/2$, so dass $\{0\} \cup S \cup -S$ 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 $p \nmid a$:*

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

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 $s = s'$.

Modulo p haben wir dann

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

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

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

und daraus die behauptete Gleichheit (beide Seiten sind ± 1). □

Wenn wir in diesem Lemma $a = -1$ setzen, bekommen wir wieder Satz 4.13.

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

4.16. **Satz.** (Zweites Ergänzungsgesetz zum Quadratischen Reziprozitätsgesetz) *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 4.15 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
$8k+7$	$2k+2$	$+1$

□

4.17. 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 vermuten, dass

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

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

Für größere q erhalten wir ähnliche Muster: Wenn $q \equiv 1 \bmod 4$, 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.

4.18. **Satz.** (Quadratisches Reziprozitätsgesetz; Gauß 1796)
Wenn p und q verschiedene ungerade Primzahlen sind, dann gilt

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

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

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

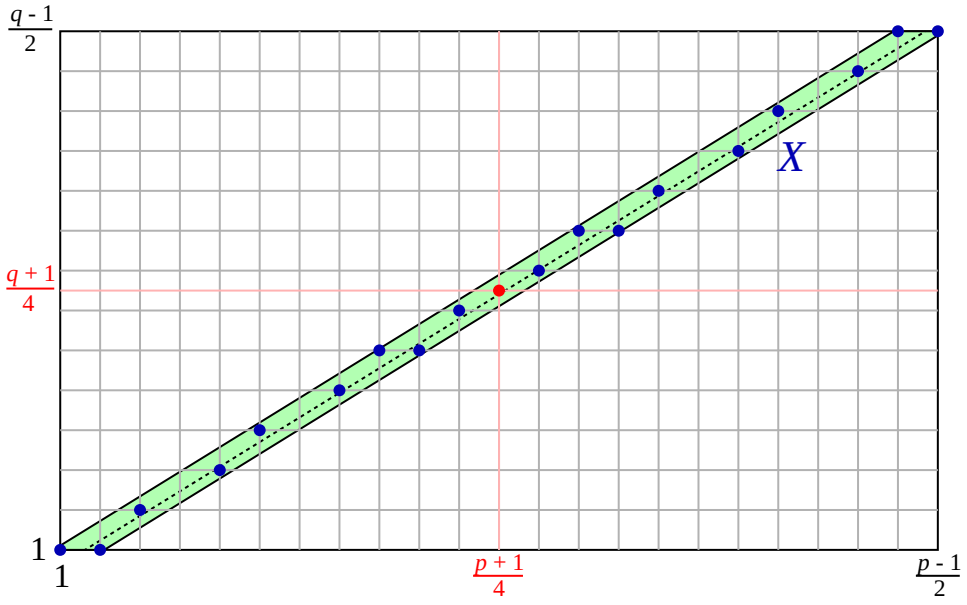


Abbildung 1. Skizze zum Beweis von Satz 4.18. Hier ist $p = 47$, $q = 29$ mit $m = 11$, $n = 7$.

Beweis. Wir stützen uns wieder auf das Lemma 4.15 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 $qs \equiv -s' \bmod p$ 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

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

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, sehen wir, dass

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

Damit ist der Satz bewiesen. □

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

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

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.

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

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

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

(1) $\left(\frac{a}{n}\right) = 0$ genau dann, wenn $\mathrm{ggT}(a, n) \neq 1$.
(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') $\left(\frac{a}{mn}\right) = \left(\frac{a}{m}\right)\left(\frac{a}{n}\right)$.
(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.

4.21. **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 reduzieren, wo wir die bereits bekannten Sätze 4.13, 4.16 und 4.18 erhalten. □

4.22. **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 verarzten muss.

4.23. **Bemerkung.** Das Euler-Kriterium 4.8 gilt für das Jacobi-Symbol nicht. Zum Beispiel gilt

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

für alle a mit $a \perp 15$, obwohl das Jacobi-Symbol $\left(\frac{a}{15}\right)$ auch den Wert -1 annimmt (z.B. für $a = 7$).

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.

4.24. **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 4.21

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

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. □

4.25. **Beispiel.** 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[9] und Reichardt[10] untersucht.

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

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

Satz. (Lind, Reichardt)
Die Gleichung $X^4 - 17\,Y^4 = 2\,Z^2$ *hat keine primitiven 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 $n \geq 1$. Man braucht also bessere Methoden, 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 $\mathrm{ggT}(X, Y, Z) = 1$). Modulo p bekommen wir nun die Kongruenz $X^4 \equiv 17\,Y^4$; das zeigt, dass 17 quadratischer Rest mod p ist. Es folgt nach dem QRG 4.18, dass

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

Außerdem gilt nach den beiden Ergänzungsgesetzen 4.13 und 4.16 noch

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

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 $Z \equiv W^2$ mod 17. Damit bekommen wir die Kongruenz

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

mit $W \not\equiv 0$ mod 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 ± 6, und das sind keine quadratischen Reste mod 17). □

5. Der Gitterpunktsatz von Minkowski

Bevor wir uns nun dem eigentlichen Thema der Vorlesung zuwenden, müssen wir noch ein Hilfsmittel bereitstellen, mit dem sich viele der folgenden Aussagen 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 $\mathbb{Z}^n \subset \mathbb{R}^n$.

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

$$F = \left\{ \sum_{j=1}^{n} t_j v_j \mid 0 \leq t_j < 1 \text{ für alle } j \right\}$$

heißt eine *Grundmasche* des Gitters Λ, und $\Delta(\Lambda) = \mathrm{vol}(F) = |\det(v_1, \ldots, v_n)|$ heißt das *Kovolumen* von Λ. Beachte, dass viele verschiedene Basen das selbe Gitter erzeugen. Die Grundmasche F hängt von der gewählten Basis ab, während das Kovolumen nur von Λ 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 ± 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 $\lambda \in \Lambda$.

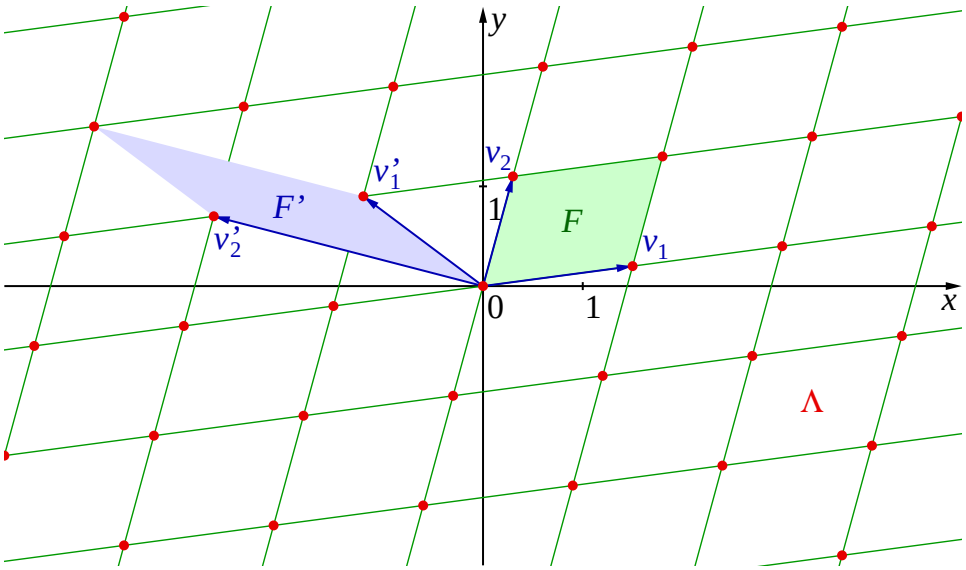


Abbildung 2. Beispiel eines Gitters mit zwei Grundmaschen

5.2. **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 $\Delta(\mathbb{Z}^n) = 1$.

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 Λ das Bild von $\mathbb{Z}^n$ unter der invertierbaren linearen Abbildung $T : \mathbb{R}^n \to \mathbb{R}^n$, die die Standardbasis $e_1, \dots, e_n$ auf $v_1, \dots, v_n$ abbildet. Das Kovolumen ist dann $\Delta(\Lambda) = |\det(T)|$.

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

Beweis. Wir haben $\Lambda \cong \mathbb{Z}^n$ als abelsche Gruppen. Nach dem Struktursatz für endlich erzeugte abelsche Gruppen gibt es einen Isomorphismus $\phi : \Lambda \to \mathbb{Z}^n$, der Λ' auf $a_1\mathbb{Z} \times \cdots \times a_n\mathbb{Z}$ abbildet mit geeigneten nichtnegativen ganzen Zahlen $a_1, \dots, a_n$. Da der Index von Λ' in Λ endlich ist, sind die a_j sogar positiv, und $a_1 \cdots a_n = m$. Sei $v_1, \dots, v_n$ die Basis von Λ, die unter ϕ auf die Standardbasis abgebildet wird. Dann ist $\Lambda' = \mathbb{Z}a_1v_1 + \cdots + \mathbb{Z}a_nv_n \subset \mathbb{R}^n$, woran wir sehen, dass Λ' ein Gitter ist. Außerdem gilt

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

□

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

5.4. **Folgerung.** *Sei $\phi : \mathbb{Z}^n \to M$ ein Gruppenhomomorphismus auf eine endliche Gruppe M (d.h., $\phi(\mathbb{Z}^n) = M$). Dann ist der Kern von ϕ ein Gitter $\Lambda \subset \mathbb{Z}^n \subset \mathbb{R}^n$ mit Kovolumen $\Delta(\Lambda) = \#M$.*

Beweis. Es gilt $\mathbb{Z}^n / \ker \phi \cong M$, also ist $\Lambda = \ker \phi$ eine Untergruppe des Gitters $\mathbb{Z}^n$ von endlichem Index $(\mathbb{Z}^n : \Lambda) = \#M$. Die Behauptung folgt dann aus Satz 5.3 und aus $\Delta(\mathbb{Z}^n) = 1$. □

Jetzt können wir den Satz von Minkowski formulieren und beweisen.

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

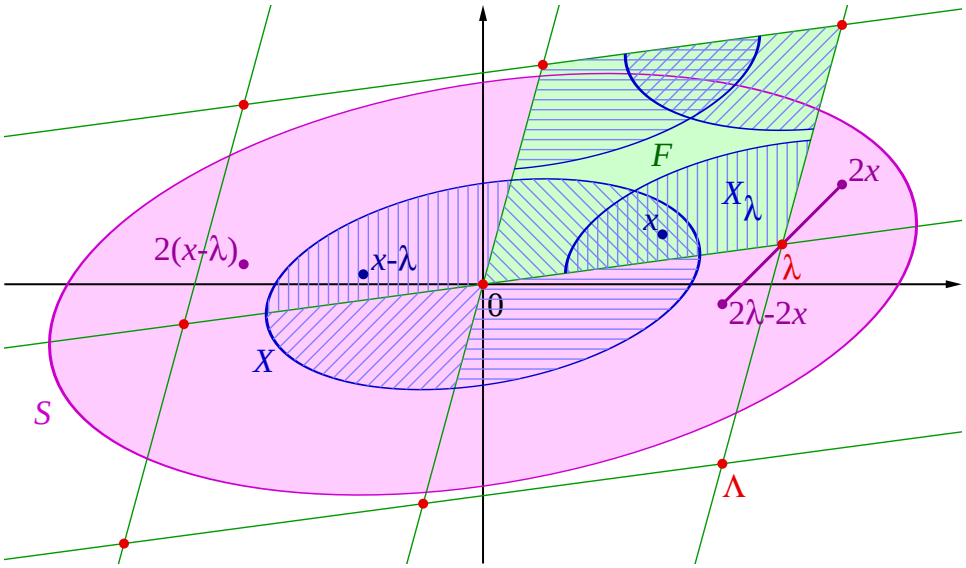


Abbildung 3. Skizze zum Beweis von Satz 5.5

Beweis. Wir beweisen zunächst, dass $X = \frac{1}{2}S$ mit einem seiner Translate $X + \lambda$ um Elemente $\lambda \in \Lambda$ nichtleeren Durchschnitt hat. Das liegt daran, dass $\mathrm{vol}(X) > \mathrm{vol}(F)$ ist, so dass die ganzen Translate von X „nicht genug Platz haben“ um disjunkt zu sein.

Sei dazu F eine Grundmasche von Λ. 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_λ paarweise disjunkt, dann würde daraus

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

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. □

5.6. **Bemerkung.** Alle drei Bedingungen an S in Satz 5.5 sind 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 $S \cap \Lambda = \{0\}$. Man kann die strikte Ungleichung $\mathrm{vol}(S) > 2^n\Delta(\Lambda)$ abschwächen zu $\mathrm{vol}(S) \ge 2^n\Delta(\Lambda)$, 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.

5.7. **Satz.** (Variante zum Gitterpunktsatz) *Sei $\Lambda \subset \mathbb{R}^n$ ein Gitter, und sei $S \subset \mathbb{R}^n$ eine symmetrische, konvexe und kompakte Teilmenge, so dass* $\mathrm{vol}(S) \ge 2^n\Delta(\Lambda)$. *Dann enthält S einen von Null verschiedenen Gitterpunkt aus Λ.*

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 5.5. Für jedes $m \ge 1$ 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 $x = \lim_{m\to\infty}(1 + \frac{1}{m})^{-1}x \in S$. □

6. 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 Σ_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 Σ_2 noch die folgende wichtige Eigenschaft:

6.1. **Lemma.** *Σ_2 ist multiplikativ abgeschlossen: $m, n \in \Sigma_2 \Longrightarrow mn \in \Sigma_2$.*

Beweis. Wir verifizieren, dass

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

□

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

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

Wegen dieser multiplikativen Struktur von Σ_2 liegt es nahe, sich anzusehen, welche Primzahlen in Σ_2 liegen. Wir haben schon gesehen, dass $p \notin \Sigma_2$, wenn $p \equiv 3 \bmod 4$ ist. Auf der anderen Seite ist natürlich $2 \in \Sigma_2$, und die Aufzählung der Elemente von Σ_2 oben lässt vermuten, dass alle Primzahlen $p \equiv 1 \bmod 4$ ebenfalls in Σ_2 sind. Dies ist tatsächlich der Fall.

6.2. **Satz.** (Zwei-Quadrate-Satz für Primzahlen)
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 4.13, 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, so dass 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, weil $k \nmid p$ (denn $1 < k < p$). 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. □

Die Idee in diesem Beweis ist die folgende. Sei $\xi = x + iy$ mit $|\xi|^2 = kp$. Wir konstruieren $\eta = u + iv$ mit $\eta \equiv \bar{\xi} \bmod k$ (im Ring $\mathbb{Z}[i]$) und $|\eta|^2 = kk'$. Dann wird $\xi\eta \equiv |\xi|^2 \equiv 0 \bmod k$, es ist also $\xi\eta = k\xi'$ mit $\xi' \in \mathbb{Z}[i]$, 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 6.2 zu geben.

Zweiter Beweis. Für den Satz von Minkowski 5.5 brauchen wir ein Gitter Λ 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

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

dann ist ϕ ein surjektiver Gruppenhomomorphismus auf die endliche (additive) Gruppe $\mathbb{F}_p$. Nach Folgerung 5.4 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

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

so dass wir Satz 5.5 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. □

Aus dem, was wir bisher bewiesen haben, folgt bereits eine Richtung des folgenden Ergebnisses, das die Elemente von Σ_2 charakterisiert.

6.3. **Satz.** (Zwei-Quadrate-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.*

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 Σ_2 sind (klar für 2 und für m^2, eben bewiesen für $p \equiv 1 \bmod 4$), also ist wegen der multiplikativen Abgeschlossenheit von Σ_2 auch $n \in \Sigma_2$.

Für die Gegenrichtung nehmen wir 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. Sei $p \equiv 3 \bmod 4$ 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 4.13 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 $n = p^2 m$. 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. □

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

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

6.4. **Lemma.** (Euler 1748) Σ_4 *ist multiplikativ abgeschlossen.*

Beweis. Man überzeugt sich davon, dass

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

□

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

$$\mathbb{H} = \{a+bi+cj+dk \mid a,b,c,d \in \mathbb{R}\}\,,$$

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

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

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

Zu einer Quaternion $\alpha = a+bi+cj+dk$ definiert man die *konjugierte Quaternion* als $\bar\alpha = a-bi-cj-dk$. Dann prüft man nach, dass

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

Da $\overline{\alpha\beta} = \bar\beta\bar\alpha$, 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 6.4, 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 Σ_4 sind, genügt es also zu zeigen, dass alle *Primzahlen* in Σ_4 sind. Als Startpunkt brauchen wir folgendes Resultat.

6.5. **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, so dass beide Seiten gleich werden. □

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

6.6. **Satz.** (Vier-Quadrate-Satz für Primzahlen)
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 6.5 ganze Zahlen u, v mit (oBdA) $|u|, |v| \leq (p-1)/2$, so dass $0^2+1^2+u^2+v^2 = kp$ 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, so dass 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 Λ 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:

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

(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 $\mathrm{vol}(B^4) = \pi^2/2$, also ist $\mathrm{vol}(S) = \pi^2(2p)^2/2 = 2\pi^2p^2$.

Daran kann man schon sehen, dass das Gitter Λ Kovolumen p^2 haben sollte. Wir sollten also einen Homomorphismus $\phi : \mathbb{Z}^4 \to \mathbb{F}_p^2$ finden, so dass 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

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

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 ϕ surjektiv ist, also ist $\Delta(\Lambda) = p^2$. Nun gilt $\mathrm{vol}(S) = 2\pi^2 p > 16p^2 = 2^4\Delta(\Lambda)$, 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 $x_1^2 + x_2^2 + x_3^2 + x_4^2 = p$. □

6.7. **Folgerung.** (Vier-Quadrate-Satz; Lagrange 1770)
Jede nichtnegative ganze Zahl ist Summe von vier Quadratzahlen.

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

6.8. **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. □

Tatsächlich ist das die einzige Einschränkung.

6.9. **Satz.** (Drei-Quadrate-Satz; Gauß)
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.

6.10. **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 $x = (x_1, x_2, x_3)$ und $y = (y_1, y_2, y_3)$ und verwenden $\langle x, y\rangle = x_1y_1 + x_2y_2 + x_3y_3$ für das Skalarprodukt. Wir schreiben $|x| = \sqrt{\langle x, x\rangle}$ für die euklidische Länge eines Vektors. Es gilt dann $0 < |x - y|^2 \leq 3/4 < 1$. Außerdem ist

$$c' := c|x - y|^2 = cm - 2(cx) \cdot y + cy^2 \in \mathbb{Z}\,.$$

Der Punkt

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

erfüllt ebenfalls $|x'|^2 = m$ (x' ist der zweite Schnittpunkt der Geraden durch x und y mit der Kugeloberfläche $|x|^2 = m$) 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 9 besprechen werden.

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 $0, 1, 3, 6, 10, 15, 21, 28, \ldots$.

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

Beweis. Sei $m \geq 0$ eine ganze Zahl. Dann ist nach dem Drei-Quadrate-Satz 6.9 $8m + 3 = x^2 + y^2 + z^2$ 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

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

□

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 Σ_2 und Σ_4. (Zum Beispiel sind 3 und 5 in Σ_3, $3 \cdot 5 = 15$ jedoch nicht.)

Es gibt eine analoge Identität wie in Lemma 6.1 und in Lemma 6.4 für acht Quadrate.[11] (Dahinter steckt die Algebra der *Octonionen* oder *Oktaven*, deren Multiplikation nur noch eine schwächere Bedingung als die Assoziativität erfüllt.

[11] http://en.wikipedia.org/wiki/Degen%27s_eight-square_identity

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:[12]

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

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

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):[13]

Gibt es für jedes $k \geq 1$ eine Zahl $g(k)$, so dass 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

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

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, so dass 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 $G(3) = 4$).

7. Ternäre quadratische Formen

Wir haben bereits einige Beispiele von *quadratischen Formen* gesehen.

[12] http://en.wikipedia.org/wiki/Jacobi%27s_four-square_theorem
[13] http://en.wikipedia.org/wiki/Waring_problem

7.1. **Definition.** Eine *quadratische Form* in n Variablen ist ein homogenes Polynom von Grad 2 in n Variablen mit ganzzahligen Koeffizienten. (Man kann aber quadratische Formen auch allgemeiner 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

$$Q(x, y) = a\,x^2 + b\,xy + c\,y^2\,.$$

Ternäre quadratische Formen sehen so aus:

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

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 Formen ist das keine besonders interessante Frage: Die Existenz einer nichttrivialen Nullstelle ist damit 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.

7.2. **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), so dass $Q(\mathbf{x}) = \mathbf{x} M_Q \mathbf{x}^\top$. Dann nennen wir $\det Q = \det(M_Q)$ die *Determinante* von Q, und

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

heißt die *Diskriminante* von Q; die Diskriminante ist immer eine ganze Zahl. (Die Potenz von 2 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 $\operatorname{disc} Q \neq 0$, 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 Element des Kerns von M_Q als einen der neuen Basisvektoren.)

7.3. **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 dann 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.

7.4. **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, so dass 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 $(x_0,y_0,z_0) = (1,0,0)$. 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 $Q(x,y,z) = -\operatorname{disc}(Q)(Y^2 - XZ)$, wie man leicht nachprüft. Da $\operatorname{disc}(Q) \neq 0$, folgt nach 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 $\operatorname{ggT}(x_0,y_0,z_0) = 1$, gibt es eine invertierbare ganzzahlige Matrix T (so dass T^{-1} ebenfalls ganzzahlige Einträge hat: $T \in \mathrm{GL}_3(\mathbb{Z})$), so dass $(x_0\ y_0\ z_0) = (1\ 0\ 0)T$ (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. □

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

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

Beweis. Das folgt aus dem vorigen Beweis. □

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

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

damit ist $x = -x'$, $y = y'$, $z = x' + z'$, so dass

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

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

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

Das ist 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 Gestalt $Q(x, y, z) = ax^2 + by^2 + cz^2$ zu betrachten. Dazu brauchen wir einen Äquivalenzbegriff für quadratische Formen.

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

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

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

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

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

Ist $\mathbf{x} = (x_1, \dots, x_n)$ eine nichttriviale rationale Lösung von $Q(x_1, x_2, \dots, x_n) = 0$, dann ist $\mathbf{x}' = \mathbf{x}(T^\top)^{-1}$ eine nichttriviale Lösung von $Q'(x_1, x_2, \dots, x_n) = 0$, und ist $\mathbf{x}'$ eine nichttriviale Lösung von $Q'(x_1, x_2, \dots, x_n) = 0$, dann ist $\mathbf{x} = \mathbf{x}'T^\top$ 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:

7.8. **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.*

Folgerung 7.5 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.

7.9. **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.

7.10. **Satz.** *Sei Q eine quadratische Form in n Variablen. Dann ist Q äquivalent zu einer diagonalen quadratischen Form, d.h., einer der Gestalt $a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2$. Q ist genau dann nicht-ausgeartet, wenn $a_1, \dots, a_n \neq 0$.*

Beweis. Wir beweisen zunächst die letzte Aussage. Q ist genau dann nicht-ausgeartet, wenn die Diagonalform nicht-ausgeartet ist. Die Determinante der Diagonalform ist $a_1a_2\cdots a_n$, also ist die Diagonalform genau dann nicht-ausgeartet, wenn $a_1, \dots, a_n \neq 0$.

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 $a_1 = 0$). 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 α_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 $2 \leq j \leq n$, so dass 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 $x_j + c x_1$ ersetzen, erhalten wir eine quadratische Form, in der der Koeffizient von x_1^2 gegeben ist durch $ac + bc^2 \neq 0$. □

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

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

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

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

Um das Ergebnis etwas schöner zu machen, können wir noch y durch $2y$ ersetzen und erhalten die diagonale Form

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

Die Matrix T ist hier

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

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

7.12. **Bemerkung.** 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 wegmultipliziert 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, so dass 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 $\mathrm{ggT}(a, b, c) = 1$ (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.

Wenn jetzt zwei der Koeffizienten, sagen wir b und c, einen gemeinsamen Primteiler p haben, dann muss in jeder ganzzahligen Lösung x durch p teilbar sein. Wir ersetzen x durch px und teilen die Form durch p; dadurch wird aus (a, b, c) das neue Koeffiziententripel $(pa, b/p, c/p)$ mit kleinerem Absolutbetrag des Produkts. Wir können einen solchen Schritt also nur endlich oft durchführen, und danach müssen die Koeffizienten *paarweise teilerfremd* sein.

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

7.13. **Notwendige Bedingungen.** *Sei abc quadratfrei und (x_0, y_0, z_0) eine primitive ganzzahligen 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 $-bc$ quadratischer Rest mod p.*
(7) *Wenn p ein ungerader Primteiler von b ist, dann ist $-ca$ quadratischer Rest mod p.*
(8) *Wenn p ein ungerader Primteiler von c ist, dann ist $-ab$ quadratischer Rest mod p.*

Beweis. Wir zeigen zunächst, dass ax_0^2, by_0^2, cz_0^2 paarweise teilerfremd sind. Wenn eine Primzahl p zwei der Terme teilt, dann auch den dritten. 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 $\mathrm{ggT}(x, y, z) = 1$.

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^2, by^2, cz^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 und z ungerade sein; x 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 und z nicht durch p teilbar sein können. Wir haben $by^2 + cz^2 \equiv 0 \bmod p$, also $(by)^2 \equiv -bc \cdot z^2 \bmod p$, und weil $z \bmod p$ invertierbar ist, muss $-bc$ quadratischer Rest mod p sein. □

Für ungerade Primzahlen p, die keinen der Koeffizienten teilen, erhalten wir keine Bedingungen, denn nach Lemma 6.5 gibt es immer nichttriviale Lösungen mod p.

7.14. **Bemerkung.** 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 dazu benutzen, ganze Zahlen zu faktorisieren.

Es stellt sich heraus, dass diese notwendigen Bedingungen sogar schon hinreichend sind.

7.15. **Satz.** (Legendre 1785) *Sei $Q(x, y, z) = a\,x^2 + b\,y^2 + c\,z^2$ mit abc quadratfrei. Wenn a, b, c die Bedingungen in 7.13 erfüllen, dann gibt es eine primitive ganzzahlige Lösung von $Q(x, y, z) = 0$.*

Beweis. Wir geben hier einen Beweis mit Hilfe des Gitterpunktsatzes 5.5. Wir müssen also wieder ein geeignetes Gitter Λ 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 $\Delta(\Lambda) \le 2D$, so dass 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 wenn $p = 2$, 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 $aw^2 + b \equiv 0 \bmod c$. 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

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

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, dass 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 5.5 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. □

7.16. **Bemerkung.** 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 9 beweisen.

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

7.17. **Folgerung.** *Wenn $a\,x^2+b\,y^2+c\,z^2=0$ eine nichttriviale ganzzahlige Lösung hat, dann gibt es eine mit*

$$\max\{|a|\,x^2\,,|b|\,y^2\,,|c|\,z^2\}\le 4\pi^{-2/3}|abc|<1,865|abc|\,,$$

oder äquivalent dazu,

$$|x|\le 2\pi^{-1/3}\sqrt{|bc|}\,,\quad |y|\le 2\pi^{-1/3}\sqrt{|ca|}\,,\quad |z|\le 2\pi^{-1/3}\sqrt{|ab|}\,.$$

Es ist $2\pi^{-1/3}<1,3656$.

Beweis. Mit $2|abc|$ anstelle von $4\pi^{-2/3}|abc|$ folgt das aus dem Beweis von Satz 7.15. Der Beweis funktioniert noch, wenn wir für S den abgeschlossenen elliptischen Zylinder nehmen, bei dem wir $2D$ durch αD mit $\alpha=4\pi^{-2/3}$ ersetzen. Das ergibt die angegebene Schranke. (Die Beweisvariante mit dem Ellipsoid liefert eine schlechtere Schranke.) □

Tatsächlich gilt eine stärkere Abschätzung.

7.18. **Satz.** (Holzer[14]) *Wenn $a\,x^2+b\,y^2+c\,z^2=0$ (mit abc quadratfrei) eine nichttriviale ganzzahlige Lösung hat, dann gibt es eine 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 7.17) 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 7.15 auf beliebige ternäre quadratische Formen verallgemeinern.

7.19. **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 7.15.

Bevor wir mit dem Beweis von Satz 7.19 beginnen, formulieren wir ein Lemma.

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

7.20. **Lemma.** *Seien $Q(x,y,z)$ und D wie in Satz 7.19. Dann gibt es ein Gitter $\Lambda \subset \mathbb{Z}^3$ mit Kovolumen $\Delta(\Lambda) = D$ und so dass $Q(x,y,z) \equiv 0 \bmod D$ gilt für alle $(x,y,z) \in \Lambda$.*

Beweis. Sei $A \in \mathrm{GL}_3(\mathbb{Z})$. Dann können wir Q durch ${}^AQ$ und $\mathbf{x}_0 = (x_0, y_0, z_0)$ durch $\mathbf{x}_0 \cdot A^{-1}$ ersetzen, wobei ${}^AQ(\mathbf{x}) = Q(\mathbf{x} \cdot A)$ (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 $\Lambda = {}^A\Lambda \cdot A$ ein geeignetes Gitter für Q. Beachte auch, dass $\mathrm{disc}({}^AQ) = \mathrm{disc}(Q)\det(A)^2 = \mathrm{disc}(Q)$.

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 primitives Tripel $(x_1, y_1, z_1) \in \mathbb{Z}^3$ finden mit $(x_1, y_1, z_1)M_Q \equiv (0,0,0) \bmod p$. Sei $A \in \mathrm{GL}_3(\mathbb{Z})$ mit erster Zeile (x_1, y_1, z_1). Dann hat ${}^AQ$ die folgende Form:

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

Sei $(x_0', y_0', z_0') = (x_0, y_0, z_0) \cdot A^{-1}$ die zugehörige primitive Lösung mod D^2.

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

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

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 $D' = |\,\mathrm{disc}(Q')| = D/p^2 < D$. 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 Λ' 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 $D' = |\,\mathrm{disc}(Q')| = D/p < D$. 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 Λ' 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, evtl. nach geeigneter Transformation mit einer Matrix $A \in \mathrm{GL}_3(\mathbb{Z})$, folgende Fälle (Übungsaufgabe):

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

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

Wenn $Q \equiv x^2$, dann muss x_0 gerade sein, Die Form $Q'(x,y,z) = Q(2x,y,z)/2$ 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 oben.

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

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

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

schreiben, dann wird $Q(x_0,y_0,z_0) \equiv 2c \bmod 4$. Es folgt, dass c gerade ist. Damit ist D durch 4 teilbar, $Q'(x,y,z) = Q(x,y,z/2)$ 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, wo N das Produkt der Primteiler von D ist.

Jetzt können wir Satz 7.19 beweisen.

Beweis. Sei Λ ein Gitter wie in Lemma 7.20. Nach Satz 7.10 gibt es eine Matrix $A \in \mathrm{GL}_3(\mathbb{Q})$, so dass $^AQ(x,y,z) = \alpha x^2 + \beta y^2 + \gamma z^2$ diagonal ist. Wir setzen $\Lambda' = \Lambda \cdot A^{-1}$; das ist ein Gitter im $\mathbb{R}^3$ mit Kovolumen $\Delta(\Lambda') = \Delta(\Lambda)/|\det(A)|$. Weiter definieren wir

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

Beachte, dass $|\alpha\beta\gamma| = |\det(^AQ)| = |\det(A)|^2 D/4$. Es ergibt sich

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

damit können wir den Gitterpunktsatz 5.5 anwenden. Wir erhalten 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. □

7.21. **Bemerkung.** Die Konstruktion des Gitters Λ im Beweis von Lemma 7.20 führt zu einem Algorithmus zur Berechnung von Λ. 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) $|Q(x,y,z)| \leq Q^+(x,y,z) < D$ ist, damit ist dann $Q(x,y,z) = 0$.

Eine detailliert ausgearbeitete Variante dieser Methode findet man in einer Arbeit von Denis Simon.[15]

7.22. **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

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

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

Wir setzen also

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

und finden ein nichttriviales Element von Λ 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 von Legendre anwenden wollten, erhielten wir zunächst die Form

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

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 Koeffizienten bekommt, die man nicht mehr ohne weiteres faktorisieren kann, und dann steckt man fest. Beachte, dass die neu aufgetretenen 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.

8. p-adische Zahlen

8.1. **Motivation.** Wir haben gesehen, dass eine diophantische Gleichung nur dann (primitive) ganzzahlige Lösungen haben kann, wenn sie (primitive) Lösungen modulo m hat für jedes $m \geq 2$. Nach dem Chinesischen Restsatz 3.23 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 Integritätsringe und deshalb zum Rechnen nicht so praktisch. Es wäre also wünschenswert, eine Struktur zur Verfügung zu haben, die ein Körper oder ein Integritätsring 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 Integritätsring 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.

Es ist nicht schwer, sich davon zu überzeugen, dass es für $n \geq 3$ stets vier Lösungen mod 2^n gibt. Das wäre nicht möglich, wenn $\mathbb{Z}/2^n\mathbb{Z}$ ein Körper wäre, denn in einem Körper (oder Integritätsring) 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.

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

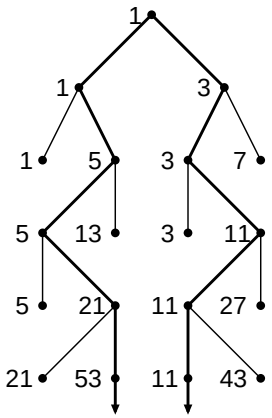


ABBILDUNG 4. Lösungen von $x^2 + 7 \equiv 0 \bmod 2^n$.

8.2. **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 sollte klar sein, dass $\mathbb{Z}_p$ 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; d.h., $a_{n+1} \equiv a_n \bmod p^n$. 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 müssen noch mehr über die Struktur von $\mathbb{Z}_p$ wissen.

8.3. **Satz.** *$\mathbb{Z}_p$ ist ein Integritätsring. 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 $\mathbb{Z}_p/p\mathbb{Z}_p \cong \mathbb{Z}/p\mathbb{Z}$; die Behauptung folgt dann, weil $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ ein Körper ist. Die Abbildung

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

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 $p\mathbb{Z}_p \subset \ker\phi$. 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 Isomorphiesatz für Ringhomomorphismen folgt

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

(b) Es ist klar, dass $\mathbb{Z}_p^\times \subset \mathbb{Z}_p \setminus p\mathbb{Z}_p$ (ein Element eines maximalen Ideals kann keine Einheit sein). Wir müssen die umgekehrte Inklusion zeigen. Sei also $u \in \mathbb{Z}_p \setminus p\mathbb{Z}_p$. 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 $u_{n+1} \equiv u_n \bmod p^n$, 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) Es folgt, dass $p\mathbb{Z}_p$ das *einzige* maximale Ideal ist. 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 $\bigcap_{n\geq 1} p^n\mathbb{Z}_p = \{0\}$: 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 $u \in \mathbb{Z}_p^\times$, so dass $a = up^n$: Aus (d) folgt, dass es ein $n \geq 0$ gibt, so dass $a \in p^n\mathbb{Z}_p \setminus p^{n+1}\mathbb{Z}_p$. Dann ist $a = up^n$, wobei $u \in \mathbb{Z}_p \setminus p\mathbb{Z}_p = \mathbb{Z}_p^\times$.

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 Integritätsring: Seien $a, b \in \mathbb{Z}_p$ mit $ab = 0$, wobei

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

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 $b = 0$. □

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

Die Aussage in Teil (e) des Beweises von Satz 8.3 motiviert folgende Definition.

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

$$v_p(a) = \max(\{0\} \cup \{n \geq 1 \mid a_n = 0\})\,,$$

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 $\mathbb{Z}$ (siehe Def. 3.12) fort; die Schreibweise ist also gerechtfertigt.

Wir definieren außerdem den *p-adischen Absolutbetrag* durch

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

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

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 ∞). Es gilt wieder für alle $a \in \mathbb{Q}_p^\times$, dass

$$a = p^{v_p(a)}u$$

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.

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

(1) $|ab|_p = |a|_p\,|b|_p$;
(2) $|a+b|_p \le \max\{|a|_p, |b|_p\} \le |a|_p + |b|_p$.

Beweis. Die Aussagen sind klar, wenn $a = 0$ oder $b = 0$. 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 $|ab|_p = |a|_p\,|b|_p$.

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 in $\mathbb{Z}_p$ ist, folgt $v_p(a+b) \ge v_p(a)$ und damit wiederum die Behauptung. □

Wir sehen auch, dass $|a+b|_p = \max\{|a|_p, |b|_p\}$ gilt, wenn $|a|_p \ne |b|_p$ (denn dann ist $u + vp^{v_p(b)-v_p(a)} \in \mathbb{Z}_p^\times$).

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 $a \in \mathbb{Z}_p \iff v_p(a) \ge 0 \iff |a|_p \le 1$).

8.8. **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. 8.2 $a_n = (a_n^{(1)}, a_n^{(2)}, \dots) \in \prod_{m\ge1} \mathbb{Z}/p^m\mathbb{Z}$. 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 $a^{(2)}$, so dass 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 $(a^{(1)}, a^{(2)}, \dots) \in \prod_{m\geq 1} \mathbb{Z}/p^m\mathbb{Z}$, so dass es für jedes $k \geq 1$ jeweils unendlich viele n gibt mit $a_n^{(m)} = a^{(m)}$ für alle $1 \leq m \leq k$. Es folgt, dass die Folge der $a^{(m)}$ kompatibel ist, d.h., $a = (a^{(1)}, a^{(2)}, \dots) \in \mathbb{Z}_p$. Wir definieren nun rekursiv $n_0 = 1$, und

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

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 $(a_{n_k})_k$ 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. □

8.9. **Bemerkung.** 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.

8.10. **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.*

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 $m \geq 0$, so dass $p^m a \in \mathbb{Z}_p$. 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 Absolutbetrages $|\cdot|$ sind.

Man definiert allgemein einen Absolutbetrag $|\cdot|$ auf einem Körper K als eine Funktion $K \ni x \mapsto |x| \in \mathbb{R}_{\geq 0}$, die auf $K^\times$ positiv ist, die multiplikativ ist: $|xy| = |x|\,|y|$, und die Dreiecksungleichung erfüllt: $|x + y| \leq |x| + |y|$. Man kann dann 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.

8.11. **Satz.** (Produktformel) *Für alle $a \in \mathbb{Q}^\times$ gilt*

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

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.

8.12. **Bemerkung.** Die Definition der p-adischen Zahlen in Def. 8.2 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, manchmal auch „ultrametrisch“ genannte, Dreiecksungleichung

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

hat 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 α 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 sehr nützlich.

8.13. **Bemerkung.** Für $\alpha \in \mathbb{Q}_p$ gilt:

(1) $\alpha \in \mathbb{Z}_p \iff |\alpha|_p \leq 1$.
(2) $\alpha \in \mathbb{Z}_p^\times \iff |\alpha|_p = 1$.
(3) $\alpha \in p^n\mathbb{Z}_p \iff |\alpha|_p \leq p^{-n}$.

Das folgt sofort aus entsprechenden Aussagen über $v_p(\alpha)$.

Das folgende Lemma, dessen erste Aussage auch eine Art „Freshman's Dream“ darstellt, beleuchtet die Wirkung der starken Dreiecksungleichung.

8.14. **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

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

für $\nu \geq N_m$ und $k \geq 0$. Das zeigt, dass die Folge der Partialsummen eine Cauchy-Folge ist; nach Satz 8.8 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.

8.15. **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

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

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.

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

(1) $$\forall n \geq 1\, \exists (x_1, \ldots, x_k) \in \mathbb{Z}^k : p^n \mid F(x_1, \ldots, x_k)$$
$$\iff \exists (x_1, \ldots, x_k) \in \mathbb{Z}_p^k : F(x_1, \ldots, x_k) = 0\,.$$

(2) *Wenn F homogen ist, gilt*

$$\begin{aligned}\forall n \geq 1\, \exists (x_1,\dots,x_k) \in \mathbb{Z}^k \setminus (p\mathbb{Z})^k : p^n \mid F(x_1,\dots,x_k)\\ \iff \exists (x_1,\dots,x_k) \in \mathbb{Z}_p^k \setminus (p\mathbb{Z}_p)^k : F(x_1,\dots,x_k) = 0\\ \iff \exists (x_1,\dots,x_k) \in \mathbb{Q}_p^k \setminus \{0\} : F(x_1,\dots,x_k) = 0\,.\end{aligned}$$

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 $p^n \mid F(x_1',\dots,x_k')$.
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 $p^n \mid F(x_1^{(n)},\dots,x_k^{(n)})$. 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, so dass $(x_1^{(n_m)},\dots x_k^{(n_m)})$ 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 „⇒“ 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 $(y_1,\dots,y_k) \in \mathbb{Z}_p^k \setminus (p\mathbb{Z}_p)^k$. Außerdem ist

$$F(y_1,\dots,y_k) = F(p^{-v}x_1,\dots,p^{-v}x_k) = p^{-dv}F(x_1,\dots,x_k) = 0\,.$$

□

8.17. **Folgerung.** *Sei $F \in \mathbb{Z}[x_1,\dots,x_k]$ ein Polynom mit ganzzahligen Koeffizienten. Dann sind äquivalent:*

(1) *Für jedes $n \geq 1$ gibt es $(x_1,\dots,x_k) \in \mathbb{Z}^k$ mit $F(x_1,\dots,x_k) \equiv 0 \bmod n$.*
(2) *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 folgende Aussagen äquivalent:

(1) *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 $\operatorname{ggT}(x_1,\dots,x_k,n) = 1$.*
(2) *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 3.25 sind die ersten Aussagen jeweils äquivalent zu der gleichen Aussage mit n eingeschränkt auf Primzahlpotenzen. Nach Satz 8.16 ist für jede feste Primzahl p die erste Aussage für alle $n = p^m$ äquivalent zur zweiten Aussage für p. □

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.

8.18. **Satz.** (Henselsches Lemma) *Sei $f \in \mathbb{Z}[x]$ (oder auch in $\mathbb{Z}_p[x]$), und sei $a \in \mathbb{F}_p$ mit $f(a) = 0$ und $f'(a) \neq 0$. 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 $a \in \mathbb{Z}$, so dass $f(a) \equiv 0 \bmod p$ und $f'(a) \not\equiv 0 \bmod p$. Es gilt dann $\alpha \equiv a \bmod p$.

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 $g \in \mathbb{Z}[x,y]$ (oder $\mathbb{Z}_p[x,y]$).

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 $\overline{f'(\alpha)} = f'(a)$). 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

$$\alpha_{n+1} = \alpha_n - f'(\alpha_n)^{-1} f(\alpha_n) \,.$$

Es gilt dann (wie man mit Induktion sieht) $f(\alpha_n) \equiv 0 \bmod p$, also $\alpha_{n+1} \equiv \alpha_n \equiv \alpha_0 \bmod p$ und damit $f'(\alpha_n) \in \mathbb{Z}_p^\times$. Ich behaupte jetzt, dass die Folge (α_n) in $\mathbb{Z}_p$ konvergiert. Sei nämlich $g \in \mathbb{Z}_p[x,y]$ das Polynom mit

$$f(y) = f(x) + (y-x) f'(x) + (y-x)^2 g(x,y) \,,$$

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 $|f(\alpha_0)|_p < 1$, gilt $f(\alpha_n) \to 0$ für $n \to \infty$. Aus der Definition von (α_n) folgt dann, dass auch $\alpha_{n+1} - \alpha_n$ gegen null geht. Nach Lemma 8.14 konvergiert demnach die Reihe $\sum_{n=0}^{\infty} (\alpha_{n+1} - \alpha_n)$, das heißt aber gerade, dass die Folge (α_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 α zu zeigen. Sei also α' eine 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 $|g(\alpha, \alpha')|_p \leq 1$, kann der zweite Faktor nicht verschwinden. Es folgt also $\alpha' = \alpha$. □

Der Beweis ist konstruktiv: Er sagt uns, wie wir α mit beliebig vorgegebener p-adischer Genauigkeit bestimmen können.

Das folgende Resultat zeigt die Nützlichkeit des Henselschen Lemmas.

8.19. **Lemma.** *Sei p eine ungerade Primzahl und $a \in \mathbb{Z}_p$ mit $p \nmid a$. Dann ist a genau dann ein Quadrat in $\mathbb{Z}_p$, wenn a ein quadratischer Rest mod p ist.*

Wenn $a \in \mathbb{Z}_2$ ungerade ist, dann ist a genau dann ein Quadrat in $\mathbb{Z}_2$, wenn $a \equiv 1 \bmod 8$ ist.

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 $f(s) = 0$. Außerdem ist $f'(s) = 2s \neq 0$. Nach Satz 8.18 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 8.18 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 $\mathbb{Q}_p$.

8.20. **Lemma.** *Sei $a \in \mathbb{Q}_p^\times$; dann ist $a = p^n u$ mit $u \in \mathbb{Z}_p^\times$ (und $n = v_p(a)$). a ist ein Quadrat in $\mathbb{Q}_p$ genau dann, wenn n gerade ist und u ein Quadrat in $\mathbb{Z}_p$ ist.*

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.

8.21. **Folgerung.** *Sei p eine Primzahl. Dann hat $x^{p-1} - 1$ in $\mathbb{Z}_p$ genau $(p-1)$ verschiedene Nullstellen.*

Beweis. Nach dem kleinen Satz von Fermat 3.30 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 $f(a) = 0$ und $f'(a) = (p-1)a^{p-2} = -a^{p-2} \neq 0$. Nach Satz 8.18 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 Integritätsring $\mathbb{Z}_p$ aber auch nicht mehr als $(p-1)$ verschiedene Nullstellen besitzen. □

9. 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 7.13 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 8.18 eine Nullstelle in $\mathbb{Z}_p$. Alternativ können wir mit Lemma 8.19 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 6.5, 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 8.19 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 7.15 und auch seine allgemeinere Form in Satz 7.19 zu folgender Aussage äquivalent sind.

9.1. **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 7.15, und der Zusatz aus den beiden Beweisvarianten, die entweder die 2-adische oder die reelle Lösbarkeit nicht verwenden. □

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.

9.2. **Bemerkung.** Das Lokal-Global-Prinzip gilt keineswegs immer. Ein berühmtes Gegenbeispiel[16] 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.

[16] E.S. Selmer: *The Diophantine equation $ax^3 + by^3 + cz^3 = 0$*, Acta Math. **85**, 203–362 (1951).

9.3. **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 hier nicht durch p teilbare ganze Zahlen sind. Nach Lemma 6.5 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 8.19 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 9.1 weglassen kann, genauer untersuchen. Dazu führen wir das *Normrestsymbol* ein. Es kodiert 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. Dies 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.

9.4. **Definition.** Seien $a, b \in \mathbb{Q}^\times$. Für eine Primzahl p setzen wir

$$\left(\frac{a,b}{p}\right) = 1\,,$$

falls $a\,x^2 + b\,y^2 = z^2$ eine nichttriviale Lösung in $\mathbb{Q}_p$ hat. Andernfalls setzen wir

$$\left(\frac{a,b}{p}\right) = -1\,.$$

In ähnlicher Weise definieren wir

$$\left(\frac{a,b}{\infty}\right) = 1\,,$$

falls es eine nichttriviale reelle Lösung gibt, andernfalls

$$\left(\frac{a,b}{\infty}\right) = -1\,.$$

Das Symbol $\left(\frac{a,b}{v}\right)$ heißt *Hilbertsches Normrestsymbol* oder kürzer *Normrestsymbol* oder *Hilbertsymbol*.

Satz 9.1 sagt dann, dass $a\,x^2+b\,y^2 = z^2$ genau dann eine nichttriviale Lösung in $\mathbb{Q}$ hat, wenn

$$\left(\frac{a,b}{v}\right) = 1 \qquad \text{für alle „Stellen“ } v = p,\ v = \infty.$$

Wir wollen jetzt den Wert des Normrestsymbols bestimmen. Wir beginnen mit dem einfachsten Fall.

9.5. **Lemma.**

$$\left(\frac{a,b}{\infty}\right) = -1 \iff a < 0 \text{ und } b < 0\,.$$

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.

9.6. **Lemma.** *Für $a, b, c \in \mathbb{Q}^\times$ und jede „Stelle“ v gilt*

$$\left(\frac{ab, ac}{v}\right) = \left(\frac{ab, -bc}{v}\right).$$

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 so dass die Quadrate in den Koeffizienten verschwinden, und vertausche x und z). □

9.7. **Definition.** Zur Vereinheitlichung der Schreibweise setzen wir $\mathbb{Q}_\infty = \mathbb{R}$ (analog zu $|\cdot|_\infty = |\cdot|$).

Eine weitere einfache Beobachtung ist, dass

$$\left(\frac{as, bt}{v}\right) = \left(\frac{a, b}{v}\right), \quad \text{wenn } s, t \text{ Quadrate in } \mathbb{Q}_v \text{ sind,}$$

und natürlich

$$\left(\frac{a, b}{v}\right) = \left(\frac{b, a}{v}\right).$$

9.8. **Lemma.** *Sei p eine ungerade Primzahl und $u_1, u_2 \in \mathbb{Q}^\times \cap \mathbb{Z}_p^\times$. Dann gilt*

$$\left(\frac{u_1, u_2}{p}\right) = 1\,, \quad \left(\frac{u_1 p, u_2}{p}\right) = \left(\frac{u_2}{p}\right) \quad \textit{und} \quad \left(\frac{u_1 p, u_2 p}{p}\right) = \left(\frac{-u_1 u_2}{p}\right).$$

Das Symbol $\left(\frac{a,b}{p}\right)$ ist in beiden Argumenten multiplikativ.

Beweis. Die Gleichung $u_1 x^2 + u_2 y^2 = z^2$ hat nach Lemma 9.3 stets eine nichttriviale in $\mathbb{Q}_p$.

Wenn u_2 ein quadratischer Rest mod p ist, dann gibt es nach Lemma 8.19 eine Quadratwurzel s von u_2 in $\mathbb{Z}_p$, und $(x, y, z) = (0, 1, s)$ ist eine nichttriviale Lösung von $u_1 p x^2 + u_2 y^2 = z^2$. Gibt es umgekehrt eine nichttriviale Lösung in $\mathbb{Q}_p$, dann 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 9.6 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 $v = 2$.

9.9. **Lemma.** *Seien $u_1, u_2 \in \mathbb{Q}^\times \cap \mathbb{Z}_2^\times$. Dann gilt*

$$\left(\frac{u_1, u_2}{2}\right) = (-1)^{\frac{u_1-1}{2}\frac{u_2-1}{2}}\,, \quad \left(\frac{2u_1, u_2}{2}\right) = (-1)^{\frac{u_2^2-1}{8}}(-1)^{\frac{u_1-1}{2}\frac{u_2-1}{2}}\,,$$

$$\left(\frac{2u_1, 2u_2}{2}\right) = (-1)^{\frac{u_1^2u_2^2-1}{8}}(-1)^{\frac{u_1-1}{2}\frac{u_2-1}{2}}\,.$$

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 mit Hilfe von Lemma 8.19 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 „Normrestsymbol“ 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 $y \neq 0$ (wenn man eine Lösung mit $y = 0$ hat, dann ist $a = s^2$ ein Quadrat, und man kann jedes Element als $(z - sx)(z + sx)$ schreiben). Es gilt also

$$\left(\frac{a, b}{v}\right) = 1 \iff \exists X, Y \in \mathbb{Q}_v : X^2 - aY^2 = b\,.$$

(Wir teilen durch y^2 und schreiben $X = z/y$, $Y = x/y$.)

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 $\mathbb{Q}_v(\sqrt{a})$. Das Symbol $\left(\frac{a,b}{v}\right)$ ist also 1 genau dann, wenn b eine Norm von $\mathbb{Q}_v(\sqrt{a})$ ist. Die Norm ist multiplikativ (leichte Rechnung; vergleiche auch Lemma 6.1), also bildet die Menge N_a der Normen von Elementen in $\mathbb{Q}_v(\sqrt{a})^\times$ eine Untergruppe von $\mathbb{Q}_v^\times$. Daraus folgt sofort:

$$\left(\frac{a, b_1}{v}\right) = 1\,, \quad \left(\frac{a, b_2}{v}\right) = 1 \quad \Longrightarrow \quad \left(\frac{a, b_1 b_2}{v}\right) = 1$$
$$\left(\frac{a, b_1}{v}\right) = 1\,, \quad \left(\frac{a, b_2}{v}\right) = -1 \quad \Longrightarrow \quad \left(\frac{a, b_1 b_2}{v}\right) = -1$$

(Diese Eigenschaften des Normrestsymbols kann man übrigens verwenden, um die Bestimmung der Werte bei $v = 2$ etwas zu vereinfachen.)

Die noch fehlende Aussage

$$\left(\frac{a, b_1}{v}\right) = -1\,, \quad \left(\frac{a, b_2}{v}\right) = -1 \quad \Longrightarrow \quad \left(\frac{a, b_1 b_2}{v}\right) = 1$$

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.

9.10. **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 ∞.)

Die Produktformel besagt, dass es immer eine (endliche und) *gerade* Anzahl von Stellen v gibt, so dass $ax^2 + by^2 = z^2$ 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 9.8 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 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 4.18 und seinen beiden Ergänzungsgesetzen 4.13 und 4.16 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 9.1.

9.11. **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 9.10 folgt dann, dass es tatsächlich nichttriviale Lösungen in $\mathbb{Q}_v$ für *alle* v geben muss. Aussage (1) folgt dann aus Satz 9.1. □

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 arithmetischen Progressionen.

9.12. **Satz.** (Dirichlet 1837) *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. Wir werden ihn am Ende der Vorlesung in Abschnitt 11 nachholen. Nachlesen kann man einen Beweis etwa in [IR, Ch. 16] oder [Sch, Kap. 8]. □

9.13. **Lemma.** *Sei $Q(x_1, x_2, \ldots, x_n)$ eine nicht-ausgeartete quadratische Form in n Variablen. Wenn es nichttriviale Lösungen in $\mathbb{Q}$ (oder $\mathbb{Q}_v$) gibt von $Q = 0$, dann gibt es auch Lösungen in $\mathbb{Q}$ (oder $\mathbb{Q}_v$) mit $x_n = 1$.*

Beweis. Eine nicht-ausgeartete quadratische Form in einer Variablen hat die Form ax_1^2 mit $a \neq 0$, hat also keine nichttrivialen Nullstellen. Wir können also $n \geq 2$ voraussetzen. Sei $K = \mathbb{Q}$ oder $K = \mathbb{Q}_v$ der betrachtete Körper.

Es genügt zu zeigen, dass es eine Lösung mit $x_n \neq 0$ gibt (dann können wir die Lösung durch x_n teilen). Sei $(\xi_1, \ldots, \xi_{n-1}, 0)$ eine Lösung. Seien $\eta_1, \ldots, \eta_{n-1} \in K$; dann ist

$$Q(\xi_1 + t\eta_1, \ldots, \xi_{n-1} + t\eta_{n-1}, t) = t\big(L(\eta_1, \ldots, \eta_{n-1}, 1) + Q(\eta_1, \ldots, \eta_{n-1}, 1)t\big)\,,$$

wobei L eine Linearform ist. (L ist $\mathbf{y} \mapsto \mathbf{x} \cdot 2M_Q \cdot \mathbf{y}^\top$; hier ist $\mathbf{x} = (\xi_1, \ldots, \xi_{n-1}, 0)$.) Wäre $L(\eta_1, \ldots, \eta_{n-1}, 1) = 0$ für alle Wahlen der η_j, dann wäre $\mathbf{x}$ im Kern von M_Q; das hieße aber $\det(M_Q) = 0$, und Q wäre ausgeartet, im Widerspruch zur Voraussetzung. Es gibt also $\eta_1, \ldots, \eta_{n-1} \in \mathbb{Q}$ mit $L(\eta_1, \ldots, \eta_{n-1}, 1) \neq 0$. Gilt dann $Q(\eta_1, \ldots, \eta_{n-1}, 1) = 0$, so haben wir eine Lösung mit $x_n = 1$ gefunden. Anderenfalls setzen wir $t = -L(\eta_1, \ldots, \eta_{n-1}, 1)/Q(\eta_1, \ldots, \eta_{n-1}, 1) \neq 0$; dann ist $(\xi_1 + t\eta_1, \ldots, \xi_{n-1} + t\eta_{n-1}, t)$ eine Lösung mit $x_n \neq 0$. □

Dieser Beweis hat einen geometrischen Hintergrund: Wir betrachten Geraden durch die gegebene Lösung und zeigen, dass es eine Gerade gibt, deren zweiter Schnittpunkt mit der Quadrik $Q = 0$ die Bedingung $x_n \neq 0$ erfüllt.

Wir können jetzt das Hasse-Prinzip für quadratische Formen in vier Variablen beweisen.

9.14. **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 quadratfreien und teilerfremden 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 $a_1, a_3 > 0$.

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 $a_3x_3^2 + a_4x_4^2$ (für $(x_1, x_2), (x_3, x_4) \in \mathbb{Q}_p^2 \setminus \{0\}$). Wenn $u_p = 0$ ist, dann gibt es nach Lemma 9.13, 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, so dass $u_p \notin \mathbb{Z}_p^\times$.

Nach dem Chinesischen Restsatz 3.25 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 9.12 gibt es eine Primzahl q mit $q \equiv u' \bmod 4D^2/d$. 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 $a_1, a_3 > 0$ gibt es jedenfalls reelle Lösungen von $Q_j = 0$. Für Primzahlen $p \nmid qD$ gibt es nach Lemma 9.3 immer p-adische Lösungen. Wenn $p \mid D$, dann ist $dq \equiv u_p \bmod p^2$ (oder $\bmod 16$, wenn $p = 2$). Da $v_p(u_p) \in \{0, 1\}$, folgt daraus, dass dq/u_p ein Quadrat in $\mathbb{Q}_p$ ist (vergleiche Lemma 8.20), also gibt es eine p-adische Lösung von $Q_1 = 0$ und von $Q_2 = 0$.

Aus Satz 9.11 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 9.13 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 6.9 beweisen.

9.15. **Beweis des Drei-Quadrate-Satzes.** Nach Lemma 6.10 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 9.3 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 mit n ist auch $4n$ Summe von drei Quadraten). 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 8.19 führt das zu einer nichttrivialen Lösung in $\mathbb{Z}_2$.

Aus Satz 9.14 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 9.14 im Unterschied zu Satz 9.11 tatsächlich die Existenz von Lösungen in $\mathbb{Q}_v$ für *alle* v verlangt werden muss!

Die Aussage von Satz 9.14 gilt jetzt ganz allgemein für nicht-ausgeartete quadratische Formen in beliebig vielen Variablen.

9.16. **Satz.** *Sei $Q(x_1, \dots, x_n)$ 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 $Q(x_1, \dots, x_n) = a_1 x_1^2 + \dots + a_n x_n^2$ diagonal ist mit quadratfreien Koeffizienten a_j.

Der Fall $n = 1$ ist trivial ($a_1 x_1^2 = 0$ hat niemals eine nichttriviale Lösung, weder in $\mathbb{Q}$ noch in $\mathbb{Q}_v$).

Im Fall $n = 2$ ist $Q(x_1, x_2) = a_1 x_1^2 + a_2 x_2^2$, 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 9.11 erledigt, und der Fall $n = 4$ in Satz 9.14.

Den Fall $n \geq 5$ beweisen wir durch Induktion mit einem ähnlichen Argument wie Satz 9.14. 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 $a_n < 0$. 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

$$u_p = a_1 x_1^2 + \dots + a_{n-2} x_{n-2}^2 = -a_{n-1} x_{n-1}^2 - a_n x_n^2 \,.$$

Wie im Beweis von Satz 9.14 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, so dass $u_p \notin \mathbb{Z}_p^\times$.

Nach dem Chinesischen Restsatz 3.25 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 9.12 gibt es eine Primzahl q mit $q \equiv u' \bmod 4D^2/d$. Wir betrachten die nicht-ausgearteten quadratischen Formen

$$Q_1(x_1, \dots, x_{n-2}, y) = a_1 x^2 + \dots + a_{n-2} x_{n-2}^2 - dqy^2 \quad \text{und}$$
$$Q_2(x_{n-1}, x_n, z) = a_{n-1} x_{n-1}^2 + a_n x_n^2 + dqz^2 \,.$$

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 9.11 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 9.3 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) also auch nichttriviale rationale Lösungen.

Nach Lemma 9.13 gibt es dann auch Lösungen mit $y = z = 1$. Das bedeutet, dass dq sowohl in der Form $a_1 x_1^2 + \dots + a_{n-2} x_{n-2}^2$ als auch in der Form $-a_{n-1} x_{n-1}^2 - a_n x_n^2$ (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. □

9.17. **Definition.** Eine nicht-ausgeartete quadratische Form Q heißt *indefinit*, wenn $Q = 0$ eine nichttriviale reelle Nullstelle hat.

9.18. **Folgerung.** *Sei Q eine indefinite nicht-ausgeartete quadratische Form in $n \geq 5$ Variablen. Dann hat $Q = 0$ eine primitive ganzzahlige Lösung.*

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 9.16 erfüllt, so dass $Q = 0$ eine primitive ganzzahlige Lösung haben muss. □

Damit können wir den Vier-Quadrate-Satz 6.7 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 9.13 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 6.10 vor: Sei $x = (x_1, x_2, x_3, x_4)$ von $|x|^2 = n$ eine rationale Lösung 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 $y \in \mathbb{Z}^4$ mit $|y - x| < 1$, und man kann wie in Lemma 6.10 schließen, dass

$$x' = x + \frac{2\langle x, x - y\rangle}{|x - y|^2}\,(y - x)$$

ebenfalls eine Lösung ist und kleineren Nenner hat.

Im verbleibenden Fall $x = (m_1/2, m_2/2, m_3/2, m_4/2)$ 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 6.3 mit Hilfe des Satzes 9.11 beweisen, wobei man am einfachsten die Bedingung der 2-adischen Lösbarkeit weglässt (Übungsaufgabe).

9.19. **Bemerkung.** Für jede Primzahl p gibt es nicht-ausgeartete quadratische Formen in vier Variablen, die keine nichttriviale p-adische Nullstelle haben. 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.

Zum Abschluss werden wir noch zeigen, dass die Produktformel 9.10 die einzige Relation zwischen den verschiedenen Normrestsymbolen ist.

9.20. **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 $1 \le j \le k$ sei d_j ein quadratischer Nichtrest mod p_j. Nach dem Chinesischen Restsatz 3.25 und dem Satz von Dirichlet 9.12 gibt es eine Primzahl q mit $q \equiv \varepsilon_\infty d_j \bmod p_j$ für alle $1 \le j \le k$ und

$$q \equiv \left\{ \begin{array}{ll} 1 & \text{falls } \varepsilon_\infty = 1 \text{ und } \varepsilon_2 = 1 \\ 3 & \text{falls } \varepsilon_\infty = -1 \text{ und } \varepsilon_2 = -1 \\ 5 & \text{falls } \varepsilon_\infty = 1 \text{ und } \varepsilon_2 = -1 \\ 7 & \text{falls } \varepsilon_\infty = -1 \text{ und } \varepsilon_2 = 1 \end{array} \right\} \bmod 8\,.$$

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 9.5. Da $\varepsilon_\infty q \equiv 1 \bmod 4$, ist auch $\left(\frac{a,b}{2}\right) = (-1)^{(q^2-1)/8} = \varepsilon_2$, siehe Lemma 9.9. Für alle $1 \le j \le k$ ist $\varepsilon_\infty q$ ein quadratischer Nichtrest mod p_j, also ist $\left(\frac{a,b}{p_j}\right) = -1 = \varepsilon_{p_j}$, siehe Lemma 9.8. 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 9.8. Aus der Produktformel 9.10 folgt schließlich, dass $\left(\frac{a,b}{q}\right) = \prod_{v \ne q} \varepsilon_v = \varepsilon_q = 1$ ist. □

9.21. **Bemerkung.** 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.

10. Die Pellsche Gleichung

Sei $d > 0$ eine ganze Zahl, die kein Quadrat ist. Die Gleichung

$$x^2 - d\,y^2 = 1 \quad \text{oder auch} \quad x^2 - d\,y^2 = \pm 1\,,$$

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. Tatsächlich 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.

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 7.4.

Wir betrachten ein paar Beispiele. Für $d = 2$ erhalten wir folgende Lösungen von $x^2 - 2y^2 = 1$ mit $x, y \ge 0$.

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.

10.1. **Struktur der Lösungsmenge.** Analog zu der Beobachtung

$$x^2 + y^2 = (x + iy)(x - iy)\,,$$

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, sehen wir, dass

$$(x^2 - dy^2)(u^2 - dv^2) = (xu + dyv)^2 - d(xv + yu)^2\,.$$

Lemma 6.1 ist der Spezialfall $d = -1$ dieser Beziehung.

Analog zum Ring $\mathbb{Z}[i]$ 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 $N : R \to \mathbb{Z}$, $a+b\sqrt{d} \mapsto a^2 - db^2$ ist dann gerade die *Norm*; wir haben eben schon gesehen, dass die Norm multiplikativ ist. 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).

Übersetzt auf die Lösungsmengen

$$S_d = \{(x, y) \in \mathbb{Z}^2 \mid x^2 - dy^2 = 1\} \quad \text{und} \quad T_d = \{(x, y) \in \mathbb{Z}^2 \mid x^2 - dy^2 = \pm 1\}$$

der Pellschen Gleichung heißt das, dass sie eine natürliche Struktur als abelsche Gruppen haben, wobei die Verknüpfung gegeben ist durch

$$(x, y) * (x', y') = (xx' + dyy', xy' + yx')\,.$$

Die Abbildung $\mathbb{Z}^2 \ni (x, y) \mapsto x + y\sqrt{d} \in R$ liefert Isomorphismen $S_d \cong R^\times_+$ und $T_d \cong R^\times$.

Wir wollen jetzt die Struktur dieser Gruppen ermitteln. Dazu betrachten wir

$$\phi : S_d \longrightarrow \mathbb{R}^\times\,, \quad (x, y) \longmapsto x + y\sqrt{d}\,.$$

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 ϕ 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 ϕ ein Gruppenhomomorphismus. Weil $(-1,0) \in S_d$, 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 $(-1,0) * (x,y) = (-x,-y)$.

10.2. **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 minimal, so dass $x_1, y_1 > 0$. 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 Γ 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

$$x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n\,.$$

Da

$$x_n = \frac{\alpha^n + \alpha^{-n}}{2} \approx \frac{\alpha^n}{2} \quad \text{und} \quad y_n = \frac{\alpha^n - \alpha^{-n}}{2\sqrt{d}} \approx \frac{\alpha^n}{2\sqrt{d}}\,,$$

(die Näherungen gelten für $n \gg 0$) 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.

10.3. **Diophantische Approximation.** Eine Lösung von $x^2 - dy^2 = 1$ mit $x, y > 0$ liefert eine sehr gute rationale Näherung von $\sqrt{d}$:

$$0 < \frac{x}{y} - \sqrt{d} = \frac{x^2 - dy^2}{y^2(\sqrt{d} + x/y)} < \frac{1}{2\sqrt{d}\, y^2}$$

(beachte, dass $\frac{x}{y} > \sqrt{d}$).

Umgekehrt liefert eine Näherung $\frac{x}{y}$ mit $0 < \frac{x}{y} - \sqrt{d} < 1/(2\sqrt{d}\, y^2)$ eine Lösung, denn

$$0 < x^2 - dy^2 = y^2\Big(\frac{x}{y} - \sqrt{d}\Big)\Big(\frac{x}{y} + \sqrt{d}\Big) < \frac{1}{2\sqrt{d}}\Big(2\sqrt{d} + \frac{1}{2\sqrt{d}}\Big) = 1 + \frac{1}{4d} < 2\,.$$

Wir müssen also zeigen, dass es so gute Näherungen stets gibt. Als ersten Schritt beweisen wir folgendes Ergebnis, das immerhin bis auf einen konstanten Faktor an unser Ziel herankommt.

10.4. **Lemma.** *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ eine irrationale reelle Zahl. Dann gibt es unendlich viele rationale Zahlen p/q mit*

$$\Big|\alpha - \frac{p}{q}\Big| < \frac{1}{q^2}\,.$$

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 $0 \leq l < m \leq n$, so dass

$$\frac{1}{n} > |\langle m\alpha\rangle - \langle l\alpha\rangle| = \big|(m-l)\alpha - (\lfloor m\alpha\rfloor - \lfloor l\alpha\rfloor)\big|\,.$$

Mit $p = \lfloor m\alpha\rfloor - \lfloor l\alpha\rfloor$ und $q = m - l$ heißt das

$$0 < \Big|\alpha - \frac{p}{q}\Big| < \frac{1}{nq} \leq \frac{1}{q^2}\,.$$

(Da α 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. □

10.5. **Bemerkungen.**

(1) Die Eigenschaft, dass die Menge

$$\Big\{\frac{p}{q} \in \mathbb{Q} \;\Big|\; 0 < \Big|\alpha - \frac{p}{q}\Big| < \frac{1}{q^2}\Big\}$$

unendlich ist, charakterisiert die irrationalen Zahlen unter den reellen Zahlen α. Denn ist $\alpha = r/s \in \mathbb{Q}$, dann ist $|\alpha - p/q|$ entweder gleich null oder mindestens $1/qs$, so dass $0 < |\alpha - p/q| < 1/q^2$ 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.

(2) Wenn $\alpha \in \mathbb{R}$ *algebraisch* ist (also Nullstelle eines normierten Polynoms mit rationalen Koeffizienten), dann lässt sich α nicht wesentlich besser durch rationale Zahlen approximieren als in Lemma 10.4. 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). Man kann zum Beispiel daraus folgern, dass eine Gleichung

$$F(x,y) = m$$

mit $F \in \mathbb{Z}[x,y]$ homogen und irreduzibel vom Grad ≥ 3 und $0 \neq m \in \mathbb{Z}$ nur endliche viele ganzzahlige Lösungen haben kann. (Solche Gleichungen heißen *Thue-Gleichungen*, denn Thue benutzte sein oben erwähntes Resultat, um die Endlichkeit der Lösungsmenge zu beweisen.)

Wir werden jetzt das Approximationslemma 10.4 dazu verwenden, die Existenz von nichttrivialen Lösungen der Pellschen Gleichung nachzuweisen.

10.6. **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 10.2), dass $S_d^+ \cong \mathbb{Z}$, 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 10.4 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

$$\begin{aligned} |x^2 - d\,y^2| = y^2\left|\frac{x}{y} - \sqrt{d}\right|\left(\frac{x}{y} + \sqrt{d}\right) &< \frac{x}{y} + \sqrt{d} = 2\sqrt{d} + \left(\frac{x}{y} - \sqrt{d}\right) \\ &\leq 2\sqrt{d} + \left|\frac{x}{y} - \sqrt{d}\right| < 2\sqrt{d} + \frac{1}{y^2} \leq 2\sqrt{d} + 1\,. \end{aligned}$$

Es gibt nur endliche viele ganze Zahlen m mit $|m| < 2\sqrt{d}+1$, also muss es ein m geben, so dass $x^2 - dy^2 = m$ 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 $0 < x < u$.) □

10.7. **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 $x^2 - dy^2 = 1$.

Hier ist eine Tabelle mit Beispielen:

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 10.4 bewiesen haben.

10.8. **Kettenbrüche.** Sei $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ 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 α_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

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{\vdots}{\cfrac{1}{a_{n-1} + \cfrac{1}{\alpha_n}}}}}}\,.$$

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

$$[a_0] = a_0\,, \qquad [a_0; a_1, \ldots, a_{n-2}, a_{n-1}, x] = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} + \tfrac{1}{x}] \quad (n \geq 1)\,.$$

10.9. **Definition.** Der formale Ausdruck

$$[a_0; a_1, a_2, a_3, \dots]$$

heißt die *Kettenbruchentwicklung* von α.

Wenn $a_0, a_1, a_2, \dots$ ganze Zahlen sind mit $a_1, a_2, \dots \geq 1$, dann ist $[a_0; a_1, \dots, a_n]$ 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 jedesmal wieder neu anfangen müssen, wenn wir den Kettenbruch verlängern. Das folgende Lemma zeigt eine bessere Alternative auf.

10.10. **Lemma.** *Sei $a_0, a_1, a_2, \dots$ eine Folge ganzer Zahlen mit $a_1, a_2, \dots \geq 1$. 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) *$[a_0; a_1, \dots, a_n] = p_n/q_n$ für alle $n \geq 0$.*

(3) *Wenn $[a_0; a_1, a_2, \dots]$ die Kettenbruchentwicklung von α ist, dann gilt für $n \geq 0$*

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_nq_{n+1}} \leq \frac{1}{q_n^2}\,.$$

Außerdem ist $\operatorname{sign}(\alpha - p_n/q_n) = (-1)^n$.

(4) *Unter den Annahmen von (3) gilt*

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

Der Bruch $p_n/q_n = [a_0; a_1, \dots, a_n]$ heißt der *n-te Näherungsbruch von α* (engl.: *nth convergent*), wenn $[a_0; a_1, a_2, \dots]$ die Kettenbruchentwicklung von α 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:

$$[a_0; a_1, \dots, a_n, x] = \frac{p_nx + p_{n-1}}{q_nx + q_{n-1}}\,.$$

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

$$\begin{aligned} [a_0; a_1, \dots, a_{n-1}, a_n, x] &= [a_0; a_1, \dots, a_{n-1}, a_n + \tfrac{1}{x}] \\ &= \frac{p_{n-1}\big(a_n + \frac{1}{x}\big) + p_{n-2}}{q_{n-1}\big(a_n + \frac{1}{x}\big) + q_{n-2}} = \frac{(a_np_{n-1} + p_{n-2})x + p_{n-1}}{(a_nq_{n-1} + q_{n-2})x + q_{n-1}} \\ &= \frac{p_nx + p_{n-1}}{q_nx + q_{n-1}} \end{aligned}$$

Wenn wir hierin $x = a_{n+1}$ setzen, erhalten wir die Aussage des Lemmas.

(3) Es ist $\alpha = [a_0; a_1, \dots, a_n, \alpha_{n+1}]$. 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 $\alpha_{n+1}q_n + q_{n-1} > a_{n+1}q_n + q_{n-1} = q_{n+1}$ und daher

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_nq_{n+1}}\,.$$

Schließlich ist $q_{n+1} = a_{n+1}q_n + q_{n-1} \geq q_n$.

(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.

□

10.11. **Folgerung.**

(1) *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Ist $[a_0; a_1, a_2, \dots]$ die Kettenbruchentwicklung von α, so gilt*

$$\lim_{n\to\infty} \frac{p_n}{q_n} = \lim_{n\to\infty} [a_0; a_1, \dots, a_n] = \alpha\,.$$

(2) *Seien $a_0, a_1, a_2, \dots \in \mathbb{Z}$ mit $a_1, a_2, \dots \geq 1$. Sei*

$$\frac{p_n}{q_n} = [a_0; a_1, a_2, \dots, a_n]$$

wie in Lemma 10.10. Dann konvergiert die Folge $\frac{p_n}{q_n}$ gegen einen Grenzwert α, und $[a_0; a_1, a_2, \dots]$ ist die Kettenbruchentwicklung von α.

Beweis.

(1) Das folgt sofort aus Lemma 10.10, (3).

(2) Nach Lemma 10.10, (1) gilt $|p_{n+1}/q_{n+1} - p_n/q_n| = 1/(q_nq_{n+1})$. Für $n \geq 2$ ist $q_n \geq n$ (Induktion: $q_0 = 1$, $q_1 = a_1 \geq 1$, $q_n = a_nq_{n-1} + q_{n-2} \geq q_{n-1} + 1$ für $n \geq 2$), also konvergiert die Reihe $\sum_{n\geq 1} 1/(q_nq_{n+1})$. Das zeigt, dass die Folge (p_n/q_n) eine Cauchy-Folge ist; sie konvergiert daher gegen einen Grenzwert α.
Für $n \geq 2$ ist $a_0 = p_0/q_0 \leq p_n/q_n < p_1/q_1 \leq a_0 + 1$, also gilt $a_0 = \lfloor\alpha\rfloor$. Dann ist

$$\alpha_1 = \frac{1}{\alpha - a_0} = \lim_{n\to\infty} [a_1; a_2, \dots, a_n]\,.$$

Es folgt, dass $a_1 = \lfloor\alpha_1\rfloor$ ist, und auf gleiche Weise ergibt sich induktiv, dass $[a_0; a_1, a_2, \dots]$ die Kettenbruchentwicklung von α sein muss.

□

Man kann dieses Ergebnis auch so formulieren: Die Abbildungen

$$\alpha \longmapsto \text{Kettenbruchentwicklung von } \alpha$$

und

$$[a_0; a_1, a_2, \dots] \longmapsto \lim_{n\to\infty} [a_0; a_1, \dots, a_n]$$

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 $\mathbb{R} \setminus \mathbb{Q}$.

Wir müssen jetzt noch zeigen, dass jeder hinreichend gute Näherungsbruch für α auch als Näherungsbruch in der Kettenbruchentwicklung von α auftaucht.

10.12. **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 α für ein $n \geq 0$.

Beweis. Wir behandeln zuerst den Fall $q = 1$. Dann ist $p/q = p$ die zu α nächst gelegene ganze Zahl. Wenn $p < \alpha$, dann haben wir $p = a_0 = p_0/q_0$. Wenn $p > \alpha$, 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 eine etwas stärkere Aussage gilt: $|\alpha - p/q| < 1/(q(2q-1))$ hat bereits zur Folge, dass $p/q = p_n/q_n$ für ein n. Wir können annehmen, dass p/q vollständig gekürzt ist.

Wir beweisen die Behauptung durch Induktion über q. Wir können α und p/q durch $\alpha - a_0$ und $p/q - a_0$ ersetzen und daher annehmen, dass $0 < \alpha < 1$. 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

$$\frac{1}{q(2q-1)} > \left|\alpha - \frac{p}{q}\right| > \left|\frac{p}{q}\right| \geq \frac{1}{q}\,,$$

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

$$\alpha(2q-1) \geq \Big(\frac{p}{q} - \frac{1}{q(2q-1)}\Big)(2q-1) = 2p - \frac{p}{q} - \frac{1}{q} \geq 2p - 1\,,$$

so dass

$$\left|\frac{1}{\alpha} - \frac{q}{p}\right| < \frac{1}{p(2p-1)}\,.$$

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 α. 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

$$q - \tfrac{1}{2} < \frac{1}{\alpha} < q + \frac{q}{2(q-1)} < q + 1\,.$$

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 jetzt die Pellsche Gleichung lösen.

10.13. **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, so dass $p_n^2 - dq_n^2 = 1$.

Beweis. Wir haben in 10.3 gesehen, dass jede nichttriviale Lösung (x, y) mit $x, y > 0$ die Ungleichung

$$\left|\sqrt{d} - \frac{x}{y}\right| < \frac{1}{2\sqrt{d}y^2} \leq \frac{1}{2y^2}\,.$$

erfüllt. Nach Lemma 10.12 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. □

10.14. **Beispiel.** Zur Illustration berechnen wir die Grundlösung für $d = 31$. Es gilt $5 < \sqrt{31} < 6$. Wir erhalten folgende Tabelle.

n	α_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 10.2, 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

$$(x_1, y_1) = (x_0, y_0)^{*2} = (x_0^2 + dy_0^2, 2x_0y_0) = (2x_0^2 + 1, 2x_0y_0)\,.$$

Für $d = 13$ erhalten wir beispielsweise:

n	α_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}^+.

Eine weitere Bemerkung, die die Arbeit erleichtern kann, ist die folgende. 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

$$\big(2(p_n^2 \mp 1)\big)^2 - d(2\,p_n q_n)^2 = (p_n^2 + dq_n^2)^2 - d(2p_n q_n)^2 = 4\,,$$

also eine Lösung $(x_1, y_1) = (p_n^2 \mp 1, p_n q_n)$ 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 $(x_1, y_1) = (39^2 - 1, 39 \cdot 7) = (1520, 273)$.

An Hand der obigen Tabellen 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.

10.15. **Lemma.** *Sei $\alpha = \sqrt{d}$ (mit $d > 0$ kein Quadrat). Dann gilt für die Größen α_n, a_n, p_n, q_n, die zur Kettenbruchentwicklung von α gehören:*

(1) *Für $n \geq 0$ gibt es $u_n \in \mathbb{Z}$, $v_n \in \mathbb{Z}_{>0}$ mit $v_n \mid d - u_n^2$, so dass $\alpha_n = \frac{\sqrt{d}+u_n}{v_n}$.*
(2) *Für $n \geq -1$ gilt $p_n p_{n-1} - dq_n q_{n-1} = (-1)^n u_{n+1}$ und $p_n^2 - dq_n^2 = (-1)^{n+1} v_{n+1}$.*

Aus Lemma 10.18 unten wird sich noch ergeben, dass für $n \geq 1$

$$0 < u_n < \sqrt{d} \qquad \text{und} \qquad 0 < \sqrt{d} - u_n < v_n < u_n + \sqrt{d} < 2\sqrt{d}$$

gilt.

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

$$\begin{aligned} \alpha_n &= \frac{1}{\alpha_{n-1} - a_{n-1}} = \frac{v_{n-1}}{\sqrt{d} + u_{n-1} - a_{n-1} v_{n-1}} \\ &= \frac{\sqrt{d} + a_{n-1} v_{n-1} - u_{n-1}}{v_n} = \frac{\sqrt{d} + u_n}{v_n} \end{aligned}$$

mit $u_n = a_{n-1} v_{n-1} - u_{n-1}$ und $v_n = (d - u_n^2)/v_{n-1}$. Wir müssen zeigen, dass v_n hier ganz ist. Nach Induktionsannahme ist $d \equiv u_{n-1}^2 \bmod v_{n-1}$. Da $u_n \equiv -u_{n-1} \bmod v_{n-1}$, 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)

$$\begin{aligned} p_n p_{n-1} - dq_n q_{n-1} &= a_n (p_{n-1}^2 - dq_{n-1}^2) + p_{n-1} p_{n-2} - dq_{n-1} q_{n-2} \\ &= a_n (-1)^n v_n + (-1)^{n-1} u_n = (-1)^n u_{n+1} \end{aligned}$$

und

$$\begin{aligned} (p_n^2 - dq_n^2)(p_{n-1}^2 - dq_{n-1}^2) &= \big((p_n p_{n-1} - dq_n q_{n-1})^2 - d(p_n q_{n-1} - p_{n-1} q_n)^2\big) \\ &= -(d - u_{n+1}^2) = -v_{n+1} v_n \\ &= (-1)^{n+1} v_{n+1} (p_{n-1}^2 - dq_{n-1}^2)\,, \end{aligned}$$

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 $(x, y) = (x_n, y_n) = (x_1, y_1)^{*n}$ mit $n \geq 2$; insbesondere müsste gelten $x_2 = 2x_1^2 - 1 \leq x = p_n^2 + (-1)^n$. 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 folgt, dass

$$p_m = x_1 \leq \sqrt{\frac{p_n^2}{2} + 1} \leq p_n\,,$$

außer möglicherweise, wenn $p_n = 1$ ist. Wegen des streng monotonen Wachstums von $(p_k)_{k\geq 0}$ 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.

10.16. **Algorithmus.** 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 $v_{n+1} = 1$, dann gib (p_n, q_n) aus, wenn n ungerade ist, anderenfalls gib $(2p_n^2 + 1, 2p_n q_n)$ aus.
Wenn $v_{n+1} = 2$, 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

$$a_n \leq \frac{k + u_n}{v_n} < a_n + 1 \quad \Longrightarrow \quad k - v_n < a_n v_n - u_n = u_{n+1} \leq k\,,$$

so dass u_{n+1} durch die Kongruenz und die Ungleichungen eindeutig bestimmt ist.

10.17. **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 $-1 < \bar{\alpha} = (-\sqrt{d} + u)/v < 0$.

10.18. **Lemma.**

(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* α_n *reduziert, sobald* $n \geq 1$ *ist.*
(3) *Ist* $\alpha = (\sqrt{d}+u)/v$ *reduziert, so gibt es ein eindeutig bestimmtes reduziertes* $\alpha' = (\sqrt{d}+u')/v'$, *so dass* $\alpha = 1/(\alpha' - \lfloor\alpha'\rfloor)$.

Beweis. Die Abbildung $\alpha = x + y\sqrt{d} \mapsto \bar{\alpha} = x - y\sqrt{d}$ ist ein Automorphismus des Körpers $\mathbb{Q}(\sqrt{d}) = \{x + y\sqrt{d} \mid x, y \in \mathbb{Q}\}$. D.h. 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 $\bar{\alpha} - a < -1$, 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 10.15.
(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 α_1 reduziert. Nach Teil (1) folgt mit Induktion, dass auch alle folgenden α_n reduziert sind.
(3) Die Ungleichungen in Definition 10.17 für α' bedeuten $|\sqrt{d} - v'| < u' < \sqrt{d}$. Außerdem muss $vv' = d - u^2$ und $u' \equiv -u \bmod v'$ gelten. Dadurch sind v' und dann u' eindeutig bestimmt, falls sie existieren.
Es bleibt die Existenz von α' 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 α_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 α_{k-1} und α_{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.

□

10.19. **Folgerung.** *Sei* $\alpha = \sqrt{d}$ *mit* $d > 0$ *kein Quadrat. Sei weiter* $m \geq 1$ *die kleinste Zahl mit* $v_m = 1$. *Dann gilt:*

(1) $a_{n+m} = a_n$ *für alle* $n \geq 1$. *Insbesondere ist die Kettenbruchentwicklung von* $\sqrt{d}$ *ab* a_1 periodisch.
(2) $a_m = 2a_0 = 2\lfloor\sqrt{d}\rfloor$.

Beweis. Nach Lemma 10.15 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 α_{n+1} nur von α_n abhängt, folgt mit Induktion, dass $\alpha_{n+m} = \alpha_n$ für alle $n \geq 1$; damit ist auch $a_{n+m} = a_n$. Das beweist Teil (1). Für Teil (2) 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.

10.20. **Lemma.** *Sei $\alpha \in \mathbb{R}\setminus\mathbb{Q}$, und sei $[a_0; a_1, a_2, \dots]$ die Kettenbruchentwicklung von α. Wenn es $k \geq 0$ und $m \geq 1$ gibt, so dass $a_{n+m} = a_n$ für alle $n \geq k$, dann ist α Nullstelle eines irreduziblen Polynoms $x^2 + ax + b \in \mathbb{Q}[x]$.*

Man sagt dann auch, α sei eine *quadratische Irrationalität*.

Beweis. Zunächst folgt, dass auch $\alpha_{n+m} = \alpha_n$ ist für alle $n \geq k$, denn

$$\alpha_{n+m} = [a_{n+m}; a_{n+m+1}, a_{n+m+2}, \dots] = [a_n; a_{n+1}, a_{n+2}, \dots] = \alpha_n\,.$$

Wenn p_n/q_n die Näherungsbrüche der Kettenbruchentwicklung von α und p'_n/q'_n diejenigen der Kettenbruchentwicklung von $\alpha_k = \alpha_{k+m}$ sind, dann gilt nach dem Beweis von Lemma 10.10, 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 α. Das Polynom muss irreduzibel sein, weil es keine rationale Nullstelle hat. □

10.21. **Beispiel.** Der einfachste Kettenbruch ist $[1; 1, 1, 1, \dots]$. Für die zugehörige quadratische Irrationalität ϕ ergibt sich die Gleichung

$$\phi = 1 + \frac{1}{\phi} \quad\Longrightarrow\quad \phi^2 - \phi - 1 = 0 \quad\Longrightarrow\quad \phi = \frac{\sqrt{5}+1}{2}\,.$$

(Das Vorzeichen der Wurzel ist positiv, da $\phi > 0$.) Das ist der *Goldene Schnitt*.

10.22. **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 $\alpha_2 > 3$). Außerdem ist $k = 2$, $p_0 = 1$, $q_0 = 1$, $p_1 = 3$, $q_1 = 2$ und damit

$$\alpha = \frac{3\alpha_2 + 1}{2\alpha_2 + 1} = \frac{\frac{11}{2} + \frac{3}{2}\sqrt{13}}{4 + \sqrt{13}} = \frac{5 + \sqrt{13}}{6}\,.$$

Tatsächlich ergibt sich für die Folge (α_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.

10.23. **Lemma.** *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ eine quadratische Irrationalität. Dann gilt für die „Reste“ α_n in der Kettenbruchentwicklung von α, dass*

$$\alpha_n > 1 \qquad \textit{und} \qquad -1 < \bar{\alpha}_n < 0 \qquad \textit{für } n \gg 0.$$

Beweis. Nach dem Beweis von Lemma 10.10 gilt

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

also

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

und damit

$$\bar{\alpha}_{n+1} = \frac{q_{n-1}\bar{\alpha} - p_{n-1}}{-q_n \bar{\alpha} + p_n}.$$

Es folgt

$$-\frac{1}{\bar{\alpha}_{n+1}} = \frac{q_n \bar{\alpha} - p_n}{q_{n-1}\bar{\alpha} - p_{n-1}} = \frac{q_n}{q_{n-1}} \, \frac{\bar{\alpha} - \frac{p_n}{q_n}}{\bar{\alpha} - \frac{p_{n-1}}{q_{n-1}}}.$$

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 $\varepsilon_n \to 0$. 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 $n \gg 0$, 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 α_n wie oben auch in unserem Sinn reduziert ist. Wir definieren zunächst:

10.24. **Definition.** Sei $d > 0$ kein Quadrat. Wir setzen

$$R_d = \left\{ \alpha = \frac{\sqrt{d} + u}{v} \;\middle|\; \alpha \text{ reduziert} \right\}$$

und erinnern uns daran, dass $\frac{\sqrt{d}+u}{v} \in R_d$ genau dann gilt, wenn $|\sqrt{d} - v| < u < \sqrt{d}$ und $v \mid d - u^2$. Teile (1) und (3) von Lemma 10.18 lassen sich dann auch so formulieren, dass der Kettenbruchalgorithmus eine Bijektion $\phi : R_d \to R_d$ liefert.

10.25. **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.*

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.

10.26. **Satz.** *Sei $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Die Kettenbruchentwicklung von α ist genau dann schließlich periodisch, wenn α eine quadratische Irrationalität ist.*

Beweis. Die eine Richtung wurde in Lemma 10.20 bewiesen. Sei also für die Gegenrichtung α eine quadratische Irrationalität. Nach Lemma 10.23 ist $\alpha_n > 1$, $-1 < \bar{\alpha}_n < 0$ für ein n. Nach Lemma 10.25 ist $\alpha_n \in R_D$ für ein D, und nach Lemma 10.18 ist dann $\alpha_m \in R_D$ für alle $m \geq n$. Da R_D endlich ist, muss die Folge der α_m und damit auch die Folge der a_m periodisch werden. Genauer gilt, dass die Folgen der Reste α_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. □

10.27. **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	α_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	$\frac{5\sqrt{21}+21}{2}$	2
7	$\frac{5\sqrt{21}+21}{4}$	10
8	$\frac{5\sqrt{21}+19}{41}$	1
⋮	⋮	⋮

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.

10.28. **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 α. 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 α, 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 α' der Wert von $[\overline{a_{n-1}, \ldots, a_1, a_0}]$. Dann gilt

$$\alpha = \frac{p_{n-1}\alpha + p_{n-2}}{q_{n-1}\alpha + q_{n-2}} \qquad \text{und} \qquad \alpha' = \frac{p'_{n-1}\alpha' + p'_{n-2}}{q'_{n-1}\alpha' + q'_{n-2}} = \frac{p_{n-1}\alpha' + q_{n-1}}{p_{n-2}\alpha' + q_{n-2}}\,.$$

α ist also eine Nullstelle des Polynoms $f(x) = q_{n-1}x^2 + (q_{n-2} - p_{n-1})x - p_{n-2}$, und α' 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 α und $\bar{\alpha}$ sind die Nullstellen von f). Da $-1/\alpha < 0$ ist, kommt nur die zweite Möglichkeit in Frage. □

10.29. **Satz.** *Sei $d > 0$ und kein Quadrat. Sei*

$$\sqrt{d} = [a_0; \overline{a_1, a_2, \ldots, a_{n-1}, 2a_0}]$$

die Kettenbruchentwicklung von $\sqrt{d}$. Dann gilt $a_k = a_{n-k}$ für alle $1 \le k < n$.

Beweis. $\alpha = \sqrt{d} + a_0 = \sqrt{d} + \lfloor \sqrt{d} \rfloor$ ist reduziert, und $\alpha = [\overline{2a_0, a_1, \ldots, a_{n-1}}]$. Nach Lemma 10.28 ist

$$[\overline{a_{n-1}, \ldots, a_1, 2a_0}] = -\frac{1}{\bar{\alpha}} = \frac{1}{\sqrt{d} - a_0}\,.$$

Also ist

$$[\overline{2a_0, a_{n-1}, a_{n-2}, \ldots, a_1}] = 2a_0 + (\sqrt{d} - a_0) = \sqrt{d} + a_0 = \alpha = [\overline{2a_0, a_1, a_2, \ldots, a_{n-1}}]\,,$$

woraus sich die behauptete Symmetrie ergibt. □

10.30. **Beispiel.** Eine Quadratwurzel, die eine etwas längere Periode liefert, ist etwa

$$\sqrt{163} = [12; \overline{1, 3, 3, 2, 1, 1, 7, 1, 11, 1, 7, 1, 1, 2, 3, 3, 1, 24}]\,.$$

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 irgend eine von null verschiedene ganze Zahl.

10.31. **Definition.** Wir setzen

$$S_d(n) = \{(x, y) \in \mathbb{Z}^2 \mid x^2 - dy^2 = n\}\,.$$

(In unserer bisherigen Schreibweise ist also $S_d = S_d(1)$ und $T_d = S_d(1) \cup S_d(-1)$.)

10.32. **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.*

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 $\phi : \mathbb{Z}^2 \to \mathbb{R}$, $(x, y) \mapsto x + y\sqrt{d}$. Sei (x_1, y_1) die Grundlösung der Gleichung $x^2 - dy^2 = 1$, und sei $\varepsilon = \phi(x_1, y_1) > 1$. Sei $(x, y) \in S_d(n)$. Da $\phi\big((x, y) * (x', y')\big) = \phi(x, y)\phi(x', y')$, gibt es ein eindeutig bestimmtes $m \in \mathbb{Z}$ 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.

10.33. **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 $(-4, 2)$.)

10.34. **Bemerkung.** 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 ± 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. Zunächst ein Lemma.

10.35. **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 $k \ge -1$, dass*

$$(p_{k+m}, q_{k+m}) = (p_{m-1}, q_{m-1}) * (p_k, q_k) = (p_{m-1}p_k + dq_{m-1}q_k, p_{m-1}q_k + q_{m-1}p_k)\,.$$

Beweis. Sei dazu $\sqrt{d} = [a_0; \overline{a_1, \ldots, a_{m-1}, 2a_0}]$ und

$$A_k = \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix}$$

sowie $A = A_0 A_{m-1} \cdots A_2 A_1 A_0$. Es ist dann

$$A = A_0 \begin{pmatrix} p_{m-1} & q_{m-1} \\ p_{m-2} & q_{m-2} \end{pmatrix} = \begin{pmatrix} a_0 p_{m-1} + p_{m-2} & a_0 q_{m-1} + q_{m-2} \\ p_{m-1} & q_{m-1} \end{pmatrix}\,.$$

Da die A_k symmetrisch sind und $A_{m-k} = A_k$ gilt für $1 \le k < m$, folgt, dass auch A symmetrisch ist: $a_0 q_{m-1} + q_{m-2} = p_{m-1}$. Wegen $\det A_k = -1$ haben wir $\det A = (-1)^{m+1}$, also folgt

$$p_{m-1}^2 - dq_{m-1}^2 = (-1)^m = -\det A = p_{m-1}^2 - (a_0 p_{m-1} + p_{m-2}) q_{m-1}$$

und damit $a_0 p_{m-1} + p_{m-2} = dq_{m-1}$. Es ist also

$$A = \begin{pmatrix} dq_{m-1} & p_{m-1} \\ p_{m-1} & q_{m-1} \end{pmatrix}.$$

Wir setzen $a_m = 2a_0$ und $a_{k+m} = a_k$ für alle $k > 0$; A_k sei entsprechend definiert. Dann gilt

$$A_m = \begin{pmatrix} 2a_0 & 1 \\ 1 & 0 \end{pmatrix} = A_0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} A_0 \,.$$

Für jedes $k \ge 0$ ist

$$\begin{pmatrix} p_k & q_k \\ p_{k-1} & q_{k-1} \end{pmatrix} = A_k A_{k-1} \cdots A_1 A_0 \,,$$

also haben wir für $k \ge 0$ (unter Benutzung von $A_{k+m} = A_k$ für $k \ge 1$)

$$\begin{aligned} \begin{pmatrix} p_{k+m} & q_{k+m} \\ p_{k+m-1} & q_{k+m-1} \end{pmatrix} &= A_{k+m} \cdots A_{m+1} A_m \cdots A_1 A_0 \\ &= A_k \cdots A_1 A_0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} A_0 A_{m-1} \cdots A_1 A_0 \\ &= \begin{pmatrix} p_k & q_k \\ p_{k-1} & q_{k-1} \end{pmatrix} \begin{pmatrix} p_{m-1} & q_{m-1} \\ dq_{m-1} & p_{m-1} \end{pmatrix} \\ &= \begin{pmatrix} p_k p_{m-1} + dq_k q_{m-1} & p_k q_{m-1} + q_k p_{m-1} \\ * & * \end{pmatrix}. \end{aligned}$$

Das zeigt die Behauptung (für $k = -1$ ist sie trivial). □

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.

10.36. **Definition.** In der Situation des vorangehenden Lemmas definieren wir

$$(p_k^*, q_k^*) = (p_k, q_k) \quad \text{für } k \ge -1$$

und

$$(p_k^*, q_k^*) = \big((-1)^m p_{m-1}, -(-1)^m q_{m-1}\big) * (p_{k+m}^*, q_{k+m}^*) \quad \text{für } k < -1.$$

Wegen

$$\big((-1)^m p_{m-1}, -(-1)^m q_{m-1}\big) * (p_{m-1}, q_{m-1}) = \big((-1)^m (p_{m-1}^2 - dq_{m-1}^2), 0\big) = (1, 0)$$

gilt dann $(p_{k+m}^*, q_{k+m}^*) = (p_{m-1}, q_{m-1}) * (p_k^*, q_k^*)$ für alle $k \in \mathbb{Z}$.

10.37. **Lemma.** *In der Situation von Definition 10.36 gilt*

$$p_{-1-k}^* = (-1)^k p_{-1+k}^* \qquad \textit{und} \qquad q_{-1-k}^* = (-1)^{k-1} q_{-1+k}^*$$

für alle $k \in \mathbb{Z}$.

Beweis. Mit der Schreibweise aus dem Beweis von Lemma 10.35 und den Abkürzungen

$$M_k = \begin{pmatrix} p_k^* & q_k^* \\ p_{k-1}^* & q_{k-1}^* \end{pmatrix} \qquad \text{und} \qquad A' = \begin{pmatrix} p_{m-1} & q_{m-1} \\ dq_{m-1} & p_{m-1} \end{pmatrix}$$

gilt für alle $k \in \mathbb{Z}$ (mit einem $N \geq 0$, so dass $Nm + k > 0$):

$$\begin{aligned} M_k (A')^N = M_{Nm+k} &= A_{Nm+k} A_{Nm+k-1} \cdots A_1 A_0 \\ &= A_{Nm+k} M_{Nm+k-1} = A_{Nm+k} M_{k-1} (A')^N \,, \end{aligned}$$

also (wegen der Periodizität der A_k für $k \geq 1$) $M_k = A_{k'} M_{k-1}$ mit $k' \equiv k \bmod m$ und $0 < k' \leq m$. Wegen der Symmetrie der Periode (Satz 10.29) gilt dann auch $M_{-k} = A_{k'} M_{-k-1}$. Man findet zum Beispiel

$$M_{-1} = A_m^{-1} M_0 = \begin{pmatrix} 0 & 1 \\ 1 & -2a_0 \end{pmatrix} \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -a_0 & 1 \end{pmatrix} ,$$

was die Behauptung für $k = \pm 1$ zeigt. Allgemein gilt

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} M_{-k} = (-1)^{k-1} M_{-1+k} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Für $k = 1$ haben wir das gerade gesehen, und für $k \geq 2$ folgt es durch Induktion unter Beachtung von

$$A_{k'}^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} A_{k'} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} ,$$

denn

$$\begin{aligned} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} M_{-(k+1)} &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} A_{k'}^{-1} M_{-k} \\ &= -A_{k'} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} M_{-k} \\ &= -A_{k'} (-1)^{k-1} M_{-1+k} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= (-1)^k M_{-1+(k+1)} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \end{aligned}$$

Obige Gleichung lautet ausgeschrieben

$$\begin{pmatrix} -p_{-k-1}^* & -q_{-k-1}^* \\ p_{-k}^* & q_{-k}^* \end{pmatrix} = (-1)^{k-1} \begin{pmatrix} p_{k-1}^* & -q_{k-1}^* \\ p_{k-2}^* & -q_{k-2}^* \end{pmatrix} ;$$

die beiden oberen Einträge ergeben die Behauptung für $|k| \geq 1$; für $k = 0$ folgt sie aus $q_{-1}^* = q_{-1} = 0$. □

Da p_k und q_k immer teilerfremd sind, können wir in dieser Form nur Lösungen mit teilerfremden x und y erwarten.

10.38. **Definition.** Eine Lösung $(x, y) \in S_d(n)$ heißt *primitiv*, wenn $x \perp y$. Wir schreiben $S_d^\perp(n)$ für die Menge der primitiven Lösungen von $x^2 - dy^2 = n$.

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

$$S_d(n) = \bigcup_{a>0, a^2 \mid n} \{(ax, ay) \mid (x, y) \in S_d^\perp(n/a^2)\} ,$$

und die Vereinigung ist disjunkt.

10.39. **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*

$$p_k^2 - dq_k^2 = n \qquad \text{und} \qquad -1 \le k < m' - 1$$

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 10.13 und die nachfolgenden Ergebnisse.) Nach Lemma 10.35 gilt dann (für $m' = 2m$ beachte $(x_1, y_1) = (p_{m-1}, q_{m-1})^{*2}$)

$$(p^*_{k+m'}, q^*_{k+m'}) = (x_1, y_1) * (p^*_k, q^*_k)\,.$$

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 10.12. Sei zunächst $(x, y) \in S_d^\perp(n)$ mit $x, y > 0$. Aus $x^2 - dy^2 = n$ folgt

$$\left|\sqrt{d} - \frac{x}{y}\right| = \frac{|n|}{y(x + y\sqrt{d})} < \frac{|n|}{2\sqrt{d}y^2 - 1} < \frac{1}{2y^2}$$

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 10.12 ist dann $(x', y') = (p_k, q_k)$ für ein geeignetes k, also hat die Bahn von (x, y) einen Vertreter der gewünschten Form.

Wenn $x, y < 0$, dann können wir (x, y) durch $(-x, -y) = (-1, 0) * (x, y)$ ersetzen und wie eben schließen.

Sei nun zum Beispiel $x > 0$ und $y < 0$. Nach dem schon Bewiesenen gibt es ein k, so dass (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 10.37, dass $(p^*_{-k-2}, q^*_{-k-2}) = \pm(p_k, -q_k)$, also in der relevanten Bahn ist. Der Fall $x < 0$, $y > 0$ kann genauso behandelt werden. Ist schließlich $y = 0$, dann muss $x = \pm 1$ sein, und $(x, y) = \pm(p_{-1}, q_{-1})$. □

10.40. **Beispiel.** Sei wieder $d = 163$. Wir haben $m' = m = 18$ und erhalten 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 $0 < |n| \le 12$ 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 $|n| \le 12$, die nicht vorkommen.

Für größere n, die in der Tabelle auftreten, muss das nicht mehr gelten. 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 10.18).

Zum Abschluss dieses Kapitels möchte ich noch eine Anwendung der Theorie vorführen. Unser Ziel ist es, folgenden Satz zu beweisen.

10.41. **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 $k = F_{2j}F_{2j+3}$, $n = F_{2j+2}F_{2j+3} - 1$ für $j = 0, 1, 2, \ldots$.

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 10.33. Wir suchen also alle Lösungen $(x, y) \in S_5(-4)$, so dass

$$k = \frac{x-4}{5} \qquad \text{und} \qquad n = \frac{y+3k}{2}$$

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. Um herauszufinden, welche Lösungen die Bedingung $x \equiv 4 \bmod 5$ erfüllen, betrachten wir die Lösungen modulo 5. Unter „Multiplikation“ 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:

$$\{(x, y) \in S_5(-4) \mid x \equiv 4 \bmod 5\} = \{(9, 4)^{*2m} * (x', y') \mid m \in \mathbb{Z}, (x', y') \in U\}$$

mit

$$U = \{(29, 13), (-1, -1), (-1, 1), (-11, -5), (4, 2), (4, -2)\}\,.$$

Man führt nun am besten den Goldenen Schnitt $\phi = (1+\sqrt{5})/2$ ein. Man berechnet folgende Tabelle von Potenzen von ϕ:

ℓ	-3	-1	1	3	5	7
$2\phi^\ell$	$-4 + 2\sqrt{5}$	$-1 + \sqrt{5}$	$1 + \sqrt{5}$	$4 + 2\sqrt{5}$	$11 + 5\sqrt{5}$	$29 + 13\sqrt{5}$

Außerdem ist $\phi^6 = 9 + 4\sqrt{5}$. Es folgt (mit $\bar{\phi} = -\phi^{-1}$)

$$\begin{aligned}&\{(x, y) \in S_5(-4) \mid x \equiv 4 \bmod 5\} \\ &\qquad = \{(x, y) \mid x \pm y\sqrt{5} = 2\phi^{4m-1} \text{ für ein } m \in \mathbb{Z}\}\,.\end{aligned}$$

Dann ist

$$x = \phi^{4m-1} + \bar{\phi}^{4m-1} = L_{4m-1}$$

(dabei sind $L_0 = 2$, $L_1 = 1$, $L_{n+2} = L_{n+1} + L_n$ die *Lucas-Zahlen*) und

$$y = \pm\frac{1}{\sqrt{5}}(\phi^{4m-1} - \bar{\phi}^{4m-1}) = \pm F_{4m-1}\,.$$

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:

$$\begin{aligned}
5F_mF_{m+j} &= (\phi^m - \bar{\phi}^m)(\phi^{m+j} - \bar{\phi}^{m+j}) = \phi^{2m+j} + \bar{\phi}^{2m+j} - (-1)^m(\phi^j + \bar{\phi}^j)\\
&= L_{2m+j} - (-1)^m L_j\\
2L_{m+2} &= 2\phi^{m+2} + 2\bar{\phi}^{m+2} = (3+\sqrt{5})\phi^m + (3-\sqrt{5})\bar{\phi}^m\\
&= 3(\phi^m + \bar{\phi}^m) + \sqrt{5}(\phi^m - \bar{\phi}^m) = 3L_m + 5F_m\\
2L_{m-2} &= 2\phi^{m-2} + 2\bar{\phi}^{m-2} = (3-\sqrt{5})\phi^m + (3+\sqrt{5})\bar{\phi}^m\\
&= 3(\phi^m + \bar{\phi}^m) - \sqrt{5}(\phi^m - \bar{\phi}^m) = 3L_m - 5F_m
\end{aligned}$$

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 in Frage.

Satz 10.41 gibt eine unendliche Folge von Lösungen der Gleichung

$$\binom{n}{k} = \binom{n'}{k'}$$

an, die fast alle den zusätzlichen Bedingungen $1 < k \leq n/2$, $1 < k' \leq n'/2$, $(n,k) \neq (n',k')$ 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

$$(k,k') = (2,3),\ (2,4),\ (2,5),\ (2,6),\ (2,8),\ (3,4),\ (3,6),\ (4,6)$$

ist die obige Gleichung vollständig gelöst. Ob es noch weitere Lösungen gibt, ist eine offene Frage.

11. Primzahlen in Restklassen

Zum Abschluss dieser Vorlesung wollen wir noch eine Lücke stopfen: Wir haben mehrfach den Satz 9.12 von Dirichlet über Primzahlen in „arithmetischen Progressionen“ (d.h. Restklassen) verwendet, etwa für Satz 9.14, Satz 9.16 oder Satz 9.20. Insbesondere hing auch der Drei-Quadrate-Satz 6.9 davon ab. Diesen Satz von Dirichlet wollen wir jetzt beweisen.

11.1. **Satz.** (Dirichlet 1837) *Sei $N \geq 1$ und $a \in \mathbb{Z}$ mit $a \perp N$. Dann gibt es unendlich viele Primzahlen p mit $p \equiv a \bmod N$.*

Sicher ist jedem der klassische Beweis von Euklid bekannt, der zeigt, dass es unendlich viele Primzahlen gibt. Man kann leichte Variationen dieser Idee benutzen, um Satz 11.1 für gewisse Restklassen zu beweisen.

11.2. **Satz.** *Es gibt unendlich viele Primzahlen $p \equiv 3 \bmod 4$.*

Beweis. Wir zeigen, dass es zu jeder endlichen Menge $\{p_1, p_2, \dots, p_k\}$ von Primzahlen $p_j \equiv 3 \bmod 4$ eine weitere solche Primzahl gibt. Dazu bilden wir die Zahl $N = 4p_1p_2 \cdots p_k - 1 \geq 3$. Dann ist $N \equiv 3 \bmod 4$, also muss N einen Primteiler $p \equiv 3 \bmod 4$ haben (warum?), und p muss von allen p_j verschieden sein. (Das funktioniert auch mit der leeren Menge, dann ist $N = p = 3$.) □

Auf die gleiche Weise zeigt man, dass es unendlich viele Primzahlen $p \equiv 2 \bmod 3$ (oder $\equiv 5 \bmod 6$) gibt.

Für die jeweils andere prime Restklasse ist der Beweis etwas schwieriger.

11.3. **Satz.** *Es gibt unendlich viele Primzahlen $p \equiv 1 \bmod 4$.*

Beweis. Wir zeigen, dass es zu jeder endlichen Menge $\{p_1, p_2, \dots, p_k\}$ von Primzahlen $p_j \equiv 1 \bmod 4$ eine weitere solche Primzahl gibt. Dazu bilden wir die Zahl $N = 4(p_1p_2 \cdots p_k)^2 + 1 \geq 5$. Sei p ein Primteiler von N. Dann ist -1 ein quadratischer Rest mod p, also muss $p \equiv 1 \bmod 4$ sein, und natürlich ist $p \notin \{p_1, \dots, p_k\}$. □

Man kann diese Ansätze verallgemeinern und zum Beispiel zeigen, dass es für jede quadratfreie ganze Zahl d unendlich viele Primzahlen p gibt, so dass d quadratischer Rest mod p ist. Allerdings kommt man auf diesem Weg nicht sehr weit, wenn man etwas über einzelne Restklassen beweisen möchte.

Man braucht also eine neue Idee. Dazu diskutieren wir erst einmal einen anderen Beweis für die Unendlichkeit der Menge der Primzahlen, der auf Euler zurückgeht. Euler betrachtet die Funktion

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

die für reelle $s > 1$ (oder komplexe s mit $\operatorname{Re} s > 1$) definiert ist, denn dann konvergiert die Reihe auf der rechten Seite. Diese Funktion ist unter dem Namen „Riemannsche Zetafunktion“ bekannt, seit Riemann in einer berühmten Arbeit von 1860 diese Funktion und ihre funktionentheoretischen Eigenschaften mit dem Ziel studierte, eine Strategie für den Beweis des Primzahlsatzes (der angibt, wie viele Primzahlen $< x$ es asymptotisch für großes x gibt; er wurde von Gauß und Legendre um 1800 vermutet und kurz vor Ende des 19. Jahrhunderts von Hadamard und de la Vallée Poussin unabhängig voneinander bewiesen) zu formulieren.

11.4. **Lemma.** *Für $s > 1$ (oder $\operatorname{Re} s > 1$) gilt*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1-p^{-s}}\,,$$

wobei das Produkt über alle Primzahlen p läuft und absolut konvergiert.

Beweis. Formal ist das einfach eine Formulierung der eindeutigen Primfaktorzerlegung: Jeder Faktor im Produkt ergibt eine geometrische Reihe $\sum_{m=0}^{\infty} p^{-ms}$, und wenn man diese Reihen zusammen multipliziert, dann bekommt man jeden Term n^{-s} genau einmal.

Da die Reihe absolut konvergiert, kann man die Terme beliebig umordnen, und man erhält jedenfalls

$$\zeta(s) = \prod_{p<X} \frac{1}{1-p^{-s}} + \sum_{n\in N_X} n^{-s}\,,$$

wobei N_X die Menge der ganzen Zahlen ≥ 1 ist, die einen Primteiler $p \geq X$ haben. Nun ist sicher $N_X \subset \mathbb{Z}_{\geq X}$, wegen der absoluten Konvergenz der Reihe $\sum_{n\geq 1} n^{-s}$ geht der Fehlerterm für $X \to \infty$ also gegen 0. Damit ist die behauptete Gleichheit bewiesen.

Nach Definition konvergiert ein unendliches Produkt $\prod_n (1+a_n)$ absolut, wenn die Reihe $\sum_n a_n$ absolut konvergiert. Hier ist

$$\frac{1}{1-p^{-s}} = 1 + (p^{-s} + p^{-2s} + p^{-3s} + \dots)\,,$$

und der Term in der Klammer lässt sich abschätzen gegen $2|p^{-s}|$. Die absolute Konvergenz folgt dann aus der der Reihe $\sum_n n^{-s}$. □

Nun gilt

$$\lim_{s\to 1+} \zeta(s) = +\infty\,.$$

Die Funktion $\zeta(s)$ ist nämlich für $s > 1$ streng monoton fallend (denn das gilt für jeden Summanden in der definierenden Reihe), und die harmonische Reihe $\sum_{n\geq 1} 1/n$ divergiert. Es folgt

$$\begin{aligned}\lim_{s\to 1+} \log\zeta(s) &= \lim_{s\to 1+} \sum_p \log\frac{1}{1-p^{-s}} = \lim_{s\to 1+} \sum_p \Big(p^{-s} + \frac{p^{-2s}}{2} + \frac{p^{-3s}}{3} + \dots\Big)\\ &= \lim_{s\to 1+} \sum_p p^{-s} + f(s) = +\infty\end{aligned}$$

(denn $f(s)$ ist beschränkt), also muss es unendlich viele Primzahlen geben. Es gilt sogar, dass $\sum_p 1/p$ divergiert.

Dass $f(s)$ beschränkt ist, sieht man so:

$$\begin{aligned}f(s) &= \sum_p \sum_{k\geq 2} \frac{1}{k} p^{-ks} < \sum_p p^{-2} \frac{1}{2} \sum_{k\geq 0} p^{-k}\\ &= \sum_p \frac{1}{p^2} \frac{1}{2(1-p^{-1})} < \sum_p \frac{1}{p^2} < \sum_{n\geq 1} \frac{1}{n^2} = \frac{\pi^2}{6} < \infty\,.\end{aligned}$$

Die Idee für den Beweis von Satz 11.1 ist jetzt, entsprechend zu zeigen, dass

$$\lim_{s\to 1+} \sum_{p\equiv a \bmod N} p^{-s} = +\infty\,.$$

Dazu muss man einen Umweg gehen, denn ganz direkt lässt sich der Beweis nicht übertragen. Wir zeigen am Beispiel $N = 4$, wie es funktioniert. Dazu definieren wir die Funktion

$$L(s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = 1 - \frac{1}{3^s} + \frac{1}{5^s} - \frac{1}{7^s} + - \cdots = \prod_p \frac{1}{1 - \chi(p)p^{-s}}\,,$$

wobei $\chi(n) = 0$ für n gerade, und $\chi(n) = (-1)^{(n-1)/2}$ für n ungerade. Die Reihe konvergiert für $s > 0$ (alternierende Reihe mit gegen Null abfallenden Beträgen der Glieder), und

$$L(1) = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots = \arctan(1) = \frac{\pi}{4} > 0\,.$$

Ähnlich wie für die Zetafunktion ergibt sich für $s > 1$

$$\log L(s) = \sum_p \log \frac{1}{1 - \chi(p)p^{-s}} = \sum_p \chi(p)p^{-s} + f_\chi(s)$$

mit $f_\chi(s)$ beschränkt für $s \to 1+$. Es folgt, dass $\log \zeta(s) \pm \log L(s)$ für $s \to 1+$ gegen Unendlich geht, also gilt dasselbe auch für

$$\sum_{p \equiv 1 \bmod 4} p^{-s} \qquad \text{und} \qquad \sum_{p \equiv 3 \bmod 4} p^{-s}\,.$$

Damit ist wieder gezeigt, dass es unendlich viele Primzahlen $p \equiv 1 \bmod 4$ und unendlich viele Primzahlen $p \equiv 3 \bmod 4$ gibt. Genauer folgt sogar, dass

$$\lim_{s \to 1+} \frac{\sum_{p \equiv 1 \bmod 4} p^{-s}}{\sum_p p^{-s}} = \lim_{s \to 1+} \frac{\sum_{p \equiv 3 \bmod 4} p^{-s}}{\sum_p p^{-s}} = \frac{1}{2}\,,$$

was die Tatsache ausdrückt, dass beide Restklassen ihren fairen Anteil an Primzahlen erhalten.

Für allgemeines N müssen wir mehrere Funktionen wie L oben verwenden, mit einer passenden Menge von Funktionen χ, die es uns erlauben, die relevante Restklasse mod N herauszufiltern.

11.5. **Definition.** Ein *Dirichlet-Charakter* mod N ist eine Funktion $\chi : \mathbb{Z} \to \mathbb{C}$, die die folgenden Eigenschaften hat:

(1) $\chi(n)$ hängt nur von der Restklasse von n mod N ab.
(2) $\chi(n) \neq 0$ genau dann, wenn $n \perp N$.
(3) $\chi(mn) = \chi(m)\chi(n)$ für alle $m, n \in \mathbb{Z}$.

Die *L-Reihe* zu χ ist für $s > 1$ durch

$$L(\chi, s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_p \frac{1}{1 - \chi(p)p^{-s}}\,.$$

definiert. Dabei wird die zweite Gleichheit wie für $\zeta(s)$ bewiesen. Ein Produkt dieser Art heißt *Euler-Produkt*.

Der *triviale* Dirichlet-Charakter mod N ist χ_0 mit $\chi_0(n) = 1$ falls $n \perp N$ (und $= 0$ sonst). Es gilt

$$L(\chi_0, s) = \prod_{p \nmid N} \frac{1}{1 - p^{-s}} = \prod_{p \mid N} (1 - p^{-s}) \cdot \zeta(s)\,.$$

Insbesondere gilt wieder

$$\lim_{s\to 1+} \log L(\chi_0, s) = +\infty\,.$$

11.6. **Beispiel.** Wir betrachten noch einmal den Fall $N = 4$. Ein Dirichlet-Charakter χ mod 4 ist durch die Werte $\chi(1)$ und $\chi(3)$ bestimmt. Für jeden Dirichlet-Charakter gilt $\chi(1) = 1$; außerdem haben wir $\chi(3)^2 = \chi(3^2) = \chi(9) = \chi(1) = 1$, also muss $\chi(3) = \pm 1$ sein. Beide Wahlen sind möglich, und wir erhalten den trivialen Charakter χ_0 und den nichttrivialen Charakter χ_1, der gleich der Funktion χ ist, die wir oben benutzt haben.

	0	1	2	3
χ_0	0	1	0	1
χ_1	0	1	0	-1

11.7. **Beispiel.** Für einen Dirichlet-Charakter χ mod 5 gilt $\chi(4) = \chi(2)^2$, $\chi(3) = \chi(2)^3$, $1 = \chi(1) = \chi(2)^4$, also ist χ durch den Wert $\chi(2)$ eindeutig bestimmt, und $\chi(2) \in \{1, i, -1, -i\}$. Wir erhalten die vier Dirichlet-Charaktere der folgenden Tabelle:

	0	1	2	3	4
χ_0	0	1	1	1	1
χ_1	0	1	i	$-i$	-1
χ_2	0	1	-1	-1	1
χ_3	0	1	$-i$	i	-1

Wir müssen noch ein wenig allgemeine Theorie betreiben.

11.8. **Definition.** Sei G eine Gruppe. Ein *Charakter* von G ist ein Gruppenhomomorphismus $\chi : G \to \mathbb{C}^\times$. Wir schreiben $\hat{G}$ für die Menge der Charaktere von G; $\hat{G}$ ist eine (abelsche) Gruppe unter punktweiser Multiplikation: $(\chi\psi)(g) = \chi(g)\psi(g)$; $\hat{G}$ heißt daher die *Charaktergruppe* von G.

11.9. **Bemerkung.** Die Menge der Dirichlet-Charaktere mod N steht in Bijektion mit der Charaktergruppe von $(\mathbb{Z}/N\mathbb{Z})^\times$. Der Zusammenhang ist wie folgt. Wenn χ ein Dirichlet-Charakter mod N ist, dann definiert $\psi(\bar{a}) = \chi(a)$ einen Charakter auf $(\mathbb{Z}/N\mathbb{Z})^\times$ (da $\chi(a)$ nur von $\bar{a}$ abhängt, ist ψ wohldefiniert). Ist umgekehrt ψ ein Charakter von $(\mathbb{Z}/N\mathbb{Z})^\times$, dann definiert

$$\chi(a) = \begin{cases} \psi(\bar{a}) & \text{falls } a \perp N, \\ 0 & \text{sonst} \end{cases}$$

einen Dirichlet-Charakter mod N. Wir werden im Folgenden häufig beide Sichtweisen nebeneinander verwenden.

Für endliche abelsche Gruppen wie $(\mathbb{Z}/N\mathbb{Z})^\times$ hat die Charaktergruppe besonders schöne Eigenschaften.

11.10. **Satz.** *Ist G eine endliche abelsche Gruppe, dann ist $\hat{G}$ isomorph zu G.*

Beweis. Sei zunächst G zyklisch der Ordnung n, und sei g ein Erzeuger von G. Für jeden Charakter χ von G haben wir dann $\chi(g^k) = \chi(g)^k$, also ist χ eindeutig durch $\chi(g)$ bestimmt, und $\chi(g)$ muss eine $n$te Einheitswurzel sein. Man prüft nach, dass jede der n möglichen Wahlen dieser Einheitswurzel einen Charakter ergibt. Bezeichnet $\mu_n = \{z \in \mathbb{C} \mid z^n = 1\}$ die Gruppe der n-ten Einheitswurzeln, dann haben wir also einen Isomorphismus

$$\mu_n \longrightarrow \hat{G}\,, \qquad \zeta \longmapsto (g^k \mapsto \zeta^k)\,;$$

wegen $\mu_n \cong \mathbb{Z}/n\mathbb{Z} \cong G$ gilt die Behauptung also in diesem Fall.

Jetzt zeigen wir, dass $\widehat{G \times H} \cong \hat{G} \times \hat{H}$. Im Gegensatz zum gerade behandelten Fall (wo der Isomorphismus von der Wahl von Erzeugern von μ_n und von G abhängt) ist dieser Isomorphismus kanonisch:

$$\begin{aligned} \widehat{G \times H} &\longrightarrow \hat{G} \times \hat{H} \\ \varphi &\longmapsto (g \mapsto \varphi(g,1), h \mapsto \varphi(1,h)) \\ \big((g,h) \mapsto \chi(g)\psi(h)\big) &\longmapsto (\chi,\psi) \end{aligned}$$

Man überzeugt sich davon, dass die angegebenen Abbildungen wohldefiniert und invers zueinander sind.

Für allgemeines G benutzen wir die Tatsache, dass G ein Produkt von zyklischen Gruppen $G_1, \ldots, G_k$ ist:

$$G = G_1 \times \cdots \times G_k \cong \hat{G}_1 \times \cdots \times \hat{G}_k \cong \widehat{(G_1 \times \cdots \times G_k)} = \hat{G}$$

□

Es folgt zum Beispiel, dass es genau $\#(\mathbb{Z}/N\mathbb{Z})^\times = \phi(N)$ verschiedene Dirichlet-Charaktere mod N gibt.

11.11. **Bemerkung.** Aus dem Beweis von Satz 11.10 sehen wir, dass die Werte eines Charakters χ einer endlichen Gruppe G stets Einheitswurzeln sind. Insbesondere gilt $\overline{\chi(g)} = \chi(g)^{-1} = \chi(g^{-1})$, oder auch kurz $\bar{\chi} = \chi^{-1}$.

Als nächstes beweisen wir die „Orthogonalitätsrelationen“ für Charaktere. Sie werden es uns gestatten, aus den Logarithmen der verschiedenen L-Reihen die gewünschte Restklasse herauszufischen.

11.12. **Lemma.** *Sei G eine endliche abelsche Gruppe und $1 \neq g \in G$. Dann gibt es einen Charakter $\chi \in \hat{G}$ mit $\chi(g) \neq 1$.*

Beweis. Angenommen, $\chi(g) = 1$ für alle $\chi \in \hat{G}$. Sei $H = \langle g \rangle$ die von g erzeugte Untergruppe von G. Dann hängt $\chi(g')$ nur von der Restklasse $g'H$ ab, induziert also einen Charakter auf der Quotientengruppe G/H. Wir bekommen also eine injektive Abbildung $\hat{G} \to \widehat{G/H}$. Auf der anderen Seite ist aber

$$\#\widehat{G/H} = \#G/H < \#G = \#\hat{G}\,,$$

also kann es eine solche Injektion nicht geben. □

11.13. **Satz.** *Sei G eine endliche abelsche Gruppe.*

(1) *Für alle $\chi \in \hat{G}$ gilt*

$$\sum_{g\in G} \chi(g) = \begin{cases} \#G & \text{falls } \chi = 1_{\hat{G}} \\ 0 & \text{sonst} \end{cases}$$

und deshalb für alle $\chi_1, \chi_2 \in \hat{G}$:

$$\frac{1}{\#G} \sum_{g\in G} \overline{\chi_1(g)}\chi_2(g) = \begin{cases} 1 & \text{falls } \chi_1 = \chi_2 \\ 0 & \text{sonst} \end{cases}$$

(2) *Für alle $g \in G$ gilt*

$$\sum_{\chi\in\hat{G}} \chi(g) = \begin{cases} \#G & \text{falls } g = 1_G \\ 0 & \text{sonst} \end{cases}$$

und deshalb für alle $g_1, g_2 \in G$:

$$\frac{1}{\#G} \sum_{\chi\in\hat{G}} \overline{\chi(g_1)}\chi(g_2) = \begin{cases} 1 & \text{falls } g_1 = g_2 \\ 0 & \text{sonst} \end{cases}$$

Beweis.

(1) Falls $\chi = 1_{\hat{G}}$ der triviale Charakter ist, dann ist $\chi(g) = 1$ für alle $g \in G$, und die erste Aussage ist trivial, Andernfalls gibt es ein $h \in G$ mit $\chi(h) \neq 1$. Dann folgt

$$(1-\chi(h)) \sum_{g\in G} \chi(g) = \sum_{g\in G} \chi(g) - \sum_{g\in G} \chi(hg) = \sum_{g\in G} \chi(g) - \sum_{g\in G} \chi(g) = 0\,,$$

also ist auch $\sum_g \chi(g) = 0$ (denn $1-\chi(h) \neq 0$). Die Behauptung über χ_1 und χ_2 folgt, indem man die erste Aussage auf $\chi = \overline{\chi}_1\chi_2 = \chi_1^{-1}\chi_2$ anwendet.

(2) Wenn $g = 1_G$ ist, dann ist $\chi(g) = 1$ für alle $\chi \in \hat{G}$, und die erste Aussage ist wiederum trivial (unter Beachtung von $\#\hat{G} = \#G$). Andernfalls gibt es nach Lemma 11.12 ein $\psi \in \hat{G}$ mit $\psi(g) \neq 1$. Dann gilt wie eben

$$(1-\psi(g)) \sum_{\chi\in\hat{G}} \chi(g) = \sum_{\chi\in\hat{G}} \chi(g) - \sum_{\chi\in\hat{G}} (\psi\chi)(g) = \sum_{\chi\in\hat{G}} \chi(g) - \sum_{\chi\in\hat{G}} \chi(g) = 0\,,$$

also $\sum_\chi \chi(g) = 0$. Die Behauptung über g_1 und g_2 folgt wieder, indem man die erste Aussage auf $g = g_1^{-1}g_2$ anwendet und dabei beachtet, dass $\chi(g_1^{-1}) = \chi(g_1)^{-1} = \overline{\chi(g_1)}$.

□

11.14. **Folgerung.** *Wenn wir Satz 11.13 auf die Dirichlet-Charaktere mod N anwenden, erhalten wir für $a \perp N$ die Relation*

$$\frac{1}{\phi(N)} \sum_{\chi} \overline{\chi(a)}\chi(n) = \begin{cases} 1 & \text{falls } n \equiv a \bmod N, \\ 0 & \text{sonst,} \end{cases}$$

wobei die Summe über alle Dirichlet-Charaktere mod N läuft.

11.15. **Lemma.** *Sei χ ein Dirichlet-Charakter mod N. Dann gilt für $s > 1$*

$$\log L(\chi, s) = \sum_p \chi(p) p^{-s} + f_\chi(s)$$

mit einer Funktion f_χ, die für $s \to 1+$ beschränkt bleibt.

Beweis. Wie oben haben wir

$$\begin{aligned}\log L(\chi, s) &= \sum_p \log \frac{1}{1 - \chi(p)p^{-s}} = \sum_p \Big(\chi(p)p^{-s} + \sum_{k \geq 2} \frac{1}{k} \chi(p^k) p^{-ks} \Big) \\ &= \sum_p \chi(p) p^{-s} + f_\chi(s)\end{aligned}$$

mit

$$|f_\chi(s)| = \Big| \sum_p \sum_{k \geq 2} \frac{1}{k} \chi(p^k) p^{-ks} \Big| \leq \sum_p \sum_{k \geq 2} \frac{1}{k} p^{-ks} = f(s) \leq \frac{\pi^2}{6}\,.$$

Dabei haben wir verwendet, dass $|\chi(p)| \leq 1$ ist. □

11.16. **Folgerung.** *Für $s \to 1+$ und $a \perp N$ gilt*

$$\frac{1}{\phi(N)} \sum_\chi \overline{\chi(a)} \log L(\chi, s) = \sum_{p \equiv a \bmod N} p^{-s} + f_{a,N}(s)$$

mit einer für $s \to 1+$ beschränkten Funktion $f_{a,N}$. (Die erste Summe läuft über die Dirichlet-Charaktere mod N, die zweite über alle Primzahlen $p \equiv a \bmod N$.)

Beweis. Es gilt nach Folgerung 11.14

$$\begin{aligned}\frac{1}{\phi(N)} \sum_\chi \overline{\chi(a)} \log L(\chi, s) &= \frac{1}{\phi(N)} \sum_\chi \overline{\chi(a)} \Big(\sum_p \chi(p) p^{-s} + f_\chi(s) \Big) \\ &= \sum_p \frac{1}{\phi(N)} \Big(\sum_\chi \overline{\chi(a)} \chi(p) \Big) p^{-s} + \frac{1}{\phi(N)} \sum_\chi \overline{\chi(a)} f_\chi(s) \\ &= \sum_{p \equiv a \bmod N} p^{-s} + f_{a,N}\end{aligned}$$

mit

$$f_{a,N}(s) = \frac{1}{\phi(N)} \sum_\chi \overline{\chi(a)} f_\chi(s)$$

beschränkt für $s \to 1+$ (da die f_χ nach Lemma 11.15 alle beschränkt bleiben). □

Um den Beweis von Satz 11.1 abzuschließen, müssen wir also noch zeigen, dass für $\chi \neq \chi_0$ die Funktionen $\log L(\chi, s)$ für $s \to 1+$ alle beschränkt bleiben.

Für den trivialen Dirichlet-Charakter χ_0 mod N gilt, wie oben schon gesehen, die folgende Aussage.

11.17. **Lemma.** *Sei χ_0 der triviale Dirichlet-Charakter mod N. Dann ist für $s > 1$*

$$L(\chi_0, s) = \prod_{p|N} \Big(1 - \frac{1}{p^s}\Big) \zeta(s)\,,$$

also ist

$$\log L(\chi_0, s) - \log \zeta(s) = \sum_{p|N} \log(1 - p^{-s})$$

für $s \to 1+$ beschränkt.

Beweis. Man vergleiche die Produktentwicklungen

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}$$

und

$$L(\chi_0, s) = \prod_p (1 - \chi_0(p) p^{-s})^{-1} = \prod_{p \nmid N} (1 - p^{-s})^{-1}\,.$$

□

Anders gesagt,

$$\lim_{s \to 1+} \frac{\log L(\chi_0, s)}{\sum_p p^{-s}} = 1\,.$$

Wir müssen nun erst einmal etwas über das Konvergenzverhalten von sogenannten Dirichlet-Reihen $\sum_{n=1}^\infty a_n n^{-s}$ beweisen. Für uns wird später wichtig sein, von Null- und Polstellenordnungen reden zu können, was für holomorphe oder meromorphe Funktionen möglich ist (aber nicht für beliebige glatte reelle Funktionen); deshalb betrachten wir hier komplexe Argumente s.

11.18. **Satz.**

(1) *(„Abelsche partielle Summation“) Seien (a_n) und (b_n) Folgen komplexer Zahlen, und sei weiter $A_n = \sum_{k=1}^n a_k$. Dann gilt*

$$\sum_{n=1}^N a_n b_n = \sum_{n=1}^{N-1} A_n (b_n - b_{n+1}) + A_N b_N\,.$$

(2) *(Verallgemeinerung des Konvergenzkriteriums für alternierende Reihen) Falls zusätzlich (A_n) beschränkt und (b_n) eine monoton fallende Folge reeller Zahlen mit $b_n \to 0$ ist, dann konvergiert $\sum_{n=1}^\infty a_n b_n$, und*

$$\sum_{n=1}^\infty a_n b_n = \sum_{n=1}^\infty A_n (b_n - b_{n+1})\,.$$

(3) *Konvergiert (A_n) und ist (b_n) beschränkt mit $\sum_{n=1}^\infty |b_n - b_{n+1}| < \infty$, dann konvergiert $\sum_{n=1}^\infty a_n b_n$ ebenfalls.*

(4) *Sei $f(s) = \sum_{n=1}^\infty c_n n^{-s}$. Wenn die Reihe für $s = s_0 \in \mathbb{C}$ konvergiert, dann konvergiert sie für alle $s \in \mathbb{C}$ mit $\operatorname{Re}(s - s_0) > 0$. Die Konvergenz ist gleichmäßig im Bereich $|s - s_0| \le \lambda \operatorname{Re}(s - s_0)$ für jedes $\lambda > 1$. Insbesondere ist f holomorph auf der rechten Halbebene $\operatorname{Re} s > \operatorname{Re} s_0$, und die Ableitungen von f können gliedweise berechnet werden:*

$$f^{(k)}(s) = (-1)^k \sum_{n=1}^\infty \frac{c_n (\log n)^k}{n^s}$$

Beweis. (1) Wir formen um (beachte, dass $A_0 = 0$):

$$\begin{aligned}\sum_{n=1}^{N} a_n b_n &= \sum_{n=1}^{N} (A_n - A_{n-1}) b_n \\ &= \sum_{n=1}^{N} A_n b_n - \sum_{n=0}^{N-1} A_n b_{n+1} \\ &= \sum_{n=1}^{N-1} A_n (b_n - b_{n+1}) + A_N b_N \,.\end{aligned}$$

(2) Nach (1) gilt für $M < N$:

$$\sum_{n=M+1}^{N} a_n b_n = \sum_{n=M}^{N-1} A_n (b_n - b_{n+1}) + A_N b_N - A_M b_M \,.$$

Sei nun $|A_n| \le A$ für alle n. Es folgt

$$\Big|\sum_{n=M+1}^{N} a_n b_n\Big| \le A\Big(\sum_{n=M}^{N-1} (b_n - b_{n+1}) + b_M + b_N\Big) = A(b_M - b_N + b_M + b_N) = 2Ab_M \to 0$$

für $M \to \infty$, also bilden die Partialsummen von $\sum_n a_n b_n$ eine Cauchy-Folge. Aus $A_N b_N \to 0$ und (1) folgt dann auch die behauptete Formel für den Wert der Reihe.

(3) Sei $A = \lim_{n\to\infty} A_n$, dann ist $A_n = A + \varepsilon_n$ mit $\varepsilon_n \to 0$. Dann gilt auch $\delta_N = \sup\{|\varepsilon_n| \mid n \ge N\} \to 0$ für $N \to \infty$. Sei weiter $|b_n| \le B$. Wie eben haben wir

$$\begin{aligned}\sum_{n=M+1}^{N} a_n b_n &= \sum_{n=M}^{N-1} A_n (b_n - b_{n+1}) + A_N b_N - A_M b_M \\ &= A(b_M - b_N) + \sum_{n=M}^{N-1} \varepsilon_n (b_n - b_{n+1}) + A(b_N - b_M) + \varepsilon_N b_N - \varepsilon_M b_M \\ &= \sum_{n=M}^{N-1} \varepsilon_n (b_n - b_{n+1}) + \varepsilon_N b_N - \varepsilon_M b_M \,.\end{aligned}$$

Es folgt

$$\Big|\sum_{n=M+1}^{N} a_n b_n\Big| \le \delta_M\Big(\sum_{n=M}^{N-1} |b_n - b_{n+1}| + |b_M| + |b_N|\Big) \le \delta_M\Big(\sum_{n=M}^{\infty} |b_n - b_{n+1}| + 2B\Big) \to 0$$

für $M \to \infty$, und Konvergenz folgt wie in (2).

(4) Wir setzen $a_n = c_n n^{-s_0}$ und $b_n = n^{-(s-s_0)}$. Dann ist für $\operatorname{Re} s \ge s_0$ jedenfalls $|b_n| = n^{-\operatorname{Re}(s-s_0)} \le 1$. Außerdem gilt

$$\begin{aligned}|b_n - b_{n+1}| = |n^{-(s-s_0)} - (n+1)^{-(s-s_0)}| &= \Big|-(s-s_0)\int_n^{n+1} t^{-(s-s_0)-1}\,dt\Big| \\ &\le |s-s_0| \int_n^{n+1} t^{-\operatorname{Re}(s-s_0)-1}\,dt\end{aligned}$$

und damit

$$\sum_{n=1}^{\infty} |b_n - b_{n+1}| \le |s - s_0| \int_1^{\infty} t^{-\operatorname{Re}(s-s_0)-1}\, dt = \frac{|s-s_0|}{\operatorname{Re}(s-s_0)}\,.$$

Weiterhin konvergiert (A_n) gegen $f(s_0)$. Die Voraussetzungen von (3) sind damit erfüllt. Aus der eben hergeleiteten Abschätzung sieht man auch, dass die Konvergenz gleichmäßig ist, wenn $|s - s_0| \le \lambda \operatorname{Re}(s - s_0)$ ist. Eine gleichmäßig konvergente Folge holomorpher Funktionen hat einen holomorphen Grenzwert, und weil $\lambda > 1$ beliebig groß gewählt werden kann, ist die Grenzfunktion holomorph auf $\operatorname{Re} s > s_0$. Außerdem konvergiert die Folge der Ableitungen gegen f', damit kann man f' gliedweise berechnen, und für die höheren Ableitungen gilt dasselbe. □

11.19. **Bemerkungen.**

(1) Die Formel von Teil (2) für den Wert der Reihe braucht in der Situation von (3) nicht zu gelten (denn $A_N b_N$ muss keine Nullfolge sein). Stattdessen hat man mit den Bezeichnungen aus dem Beweis von (3)

$$\sum_{n=1}^{\infty} a_n b_n = A b_1 + \sum_{n=1}^{\infty} \varepsilon_n (b_n - b_{n+1})\,.$$

(2) Der Bereich $|s - s_0| \le \lambda \operatorname{Re}(s - s_0)$ ist ein sich nach rechts öffnender Winkelbereich mit Scheitel s_0, symmetrisch zur Geraden $\operatorname{Im}(s - s_0) = 0$ und mit Öffnungswinkel $\alpha = 2 \arccos \lambda^{-1}$.

11.20. **Folgerung.** *Sei $f(s) = \sum_{n=1}^{\infty} c_n n^{-s}$ eine Dirichlet-Reihe. Dann tritt einer der folgenden drei Fälle ein.*

(1) *Die Reihe konvergiert für kein $s \in \mathbb{C}$.*
(2) *Die Reihe konvergiert für jedes $s \in \mathbb{C}$. Dann ist f eine ganze Funktion (also holomorph auf $\mathbb{C}$).*
(3) *Es gibt $s_0 \in \mathbb{R}$, die sogenannte* Konvergenzabszisse*, so dass die Reihe für $\operatorname{Re} s > s_0$ konvergiert und für $\operatorname{Re} s < s_0$ divergiert. In diesem Fall ist f eine holomorphe Funktion auf der rechten Halbebene $\operatorname{Re} s > s_0$.*

Beweis. Wir nehmen an, dass Fall (1) nicht eintritt. Es gibt also Zahlen $s \in \mathbb{C}$, so dass die Reihe konvergiert. Sei $s_0 = \inf\{\operatorname{Re} s \mid f(s) \text{ konvergiert}\} \in \{-\infty\} \cup \mathbb{R}$. Ist $s_0 = -\infty$, dann gibt es für jedes $s_1 \in \mathbb{C}$ ein $s' \in \mathbb{C}$ mit $\operatorname{Re} s' < \operatorname{Re} s_1$, so dass die Reihe für $s = s'$ konvergiert. Nach Satz 11.18, (4) konvergiert die Reihe auch für $s = s_1$, und f ist holomorph in der rechten Halbebene $\operatorname{Re} s > \operatorname{Re} s'$, also in einer Umgebung von s_1. Da s_1 beliebig war, ist f holomorph auf ganz $\mathbb{C}$.

Im verbleibenden Fall ist $s_0 \in \mathbb{R}$. Für $s \in \mathbb{C}$ mit $\operatorname{Re} s > s_0$ zeigt dasselbe Argument wie eben, dass die Reihe konvergiert und dass f auf der Halbebene $\operatorname{Re} s > s_0$ holomorph ist. Ist $\operatorname{Re} s < s_0$, dann kann die Reihe nach Definition von s_0 nicht konvergieren. □

11.21. **Bemerkungen.**

(1) Alle drei Fälle können auftreten. Zum Beispiel konvergiert $\sum_{n=1}^{\infty} 2^n n^{-s}$ nirgends, während $\sum_{n=1}^{\infty} 2^{-n} n^{-s}$ überall in $\mathbb{C}$ konvergiert. Die Zetafunktion $\sum_{n=1}^{\infty} n^{-s}$ hat Konvergenzabszisse $s_0 = 1$.

(2) Sehr ähnlich (sogar einfacher) lässt sich die analoge Aussage für das Gebiet der absoluten Konvergenz zeigen. Man erhält eine Abszisse s_1 der absoluten Konvergenz ($+\infty$ im ersten Fall, $-\infty$ im zweiten Fall). Wenn $s_0 \in \mathbb{R}$ ist, dann gilt $s_0 \le s_1 \le s_0 + 1$, und beide Grenzfälle sind möglich (Übungsaufgabe!). In diesem Punkt unterscheiden sich Dirichlet-Reihen in ihrem Konvergenzverhalten von Potenzreihen, bei denen der Konvergenzradius nicht davon abhängt, ob man absolute Konvergenz oder nur Konvergenz betrachtet.

Wir werden diese Ergebnisse jetzt auf unsere L-Reihen anwenden. Zunächst zeigen wir, dass sich die Zetafunktion bis auf einen einfachen Pol bei $s = 1$ ein Stück weit nach links fortsetzen lässt.

11.22. **Lemma.** *Es gibt eine auf* $\operatorname{Re} s > 0$ *holomorphe Funktion* f*, so dass für* $\operatorname{Re} s > 1$ *gilt*

$$\zeta(s) = \frac{1}{s-1} + f(s)\,.$$

Wir definieren $\zeta(s)$ *durch diese Formel für alle* $s \neq 1$ *mit* $\operatorname{Re} s > 0$ *Dann ist* ζ *meromorph auf* $\operatorname{Re} s > 0$*, holomorph bis auf einen einfachen Pol bei* $s = 1$ *mit Residuum* 1.

Beweis. Für $\operatorname{Re} s > 1$ können wir die Reihe für die Zetafunktion wie folgt umformen:

$$\begin{aligned}\sum_{n=1}^{\infty} n^{-s} &= \sum_{n=1}^{\infty}\Big(\int_n^{n+1} x^{-s}\,dx + \int_n^{n+1} (n^{-s} - x^{-s})\,dx\Big)\\ &= \int_1^{\infty} x^{-s}\,dx + \int_1^{\infty} \big(\lfloor x\rfloor^{-s} - x^{-s}\big)\,dx = \frac{1}{s-1} + f(s)\,.\end{aligned}$$

Es gilt

$$\lfloor x\rfloor^{-s} - x^{-s} = \int_{\lfloor x\rfloor}^{x} s t^{-s-1}\,dt\,.$$

Damit können wir $f(s)$ noch umschreiben:

$$f(s) = \int_1^{\infty} \big(\lfloor x\rfloor^{-s} - x^{-s}\big)\,dx = s\int_1^{\infty}\int_{\lfloor x\rfloor}^{x} t^{-s-1}\,dt\,dx = s\int_1^{\infty} \big(\lceil t\rceil - t\big)t^{-s-1}\,dt$$

Für $\operatorname{Re} s \ge \delta > 0$ hat der Integrand in diesem Ausdruck die gleichmäßige integrierbare Majorante $t^{-\delta-1}$; es folgt, dass $f(s)$ für $\operatorname{Re} s > 0$ eine holomorphe Funktion definiert.

Damit ist $\zeta(s) = 1/(s-1) + f(s)$ meromorph auf $\operatorname{Re} s > 0$ mit einem einzigen Pol, bei $s = 1$, der einfach ist und das Residuum $\lim_{s\to 1}(s-1)\zeta(s) = 1$ hat. □

11.23. **Bemerkung.** Tatsächlich lässt sich $\zeta(s)$ sogar zu einer auf ganz $\mathbb{C}$ meromorphen Funktion fortsetzen, die immer noch $s = 1$ als einzigen Pol hat. Eine

Möglichkeit, zu dieser Fortsetzung zu gelangen, ist wiederholte partielle Integration. Um auf $\operatorname{Re} s > -1$ fortzusetzen, schreiben wir (für $\operatorname{Re} s > 0$)

$$f(s) = s\int_1^\infty \big(\lceil t\rceil - t\big)t^{-s-1}\,dt = s\int_1^\infty \Big(\lceil t\rceil - t - \frac{1}{2}\Big)t^{-s-1}\,dt + \frac{1}{2}$$
$$= s(s+1)\int_1^\infty g(t)t^{-s-2}\,dt + \frac{1}{2}\,,$$

wobei

$$g(x) = \int_0^x \Big(\lceil t\rceil - t - \frac{1}{2}\Big)\,dt = \frac{1}{2}\big(\lceil t\rceil - t\big)\big(\lceil t\rceil - t - 1\big)$$

beschränkt ist ($0 \le g(x) \le 1/8$). Es folgt, dass das Integral für $\operatorname{Re} s > -1$ lokal gleichmäßig konvergiert, also dort eine holomorphe Funktion definiert. Wir haben dann also für $\operatorname{Re} s > -1$:

$$\zeta(s) = \frac{1}{s-1} + \frac{1}{2} + s(s+1)\int_1^\infty g(t)t^{-s-2}\,dt\,.$$

Insbesondere sieht man, dass $\zeta(0) = -1/2$ ist.

Alternativ kann man auch zunächst für $0 < \operatorname{Re} s < 1$ die *Funktionalgleichung*

$$\pi^{-s/2}\Gamma\Big(\frac{s}{2}\Big)\zeta(s) = \pi^{-(1-s)/2}\Gamma\Big(\frac{1-s}{2}\Big)\zeta(1-s)$$

beweisen (was aber nicht so ganz trivial ist) und damit dann $\zeta(s)$ auf einen Schlag auf ganz $\mathbb{C}$ fortsetzen.

Jetzt wenden wir uns der L-Reihe $L(\chi, s)$ zu.

11.24. **Satz.** *Sei χ ein nichttrivialer Dirichlet-Charakter mod N. Dann konvergiert die Reihe*

$$L(\chi, s) = \sum_{n=1}^\infty \frac{\chi(n)}{n^s}$$

für $\operatorname{Re} s > 0$ und definiert dort eine holomorphe Funktion.

Beweis. Da χ nichttrivial ist, gilt nach Satz 11.13 $\sum_{n=1}^N \chi(n) = 0$. Daraus folgt, dass die Folge der Summen $\sum_{n=1}^M \chi(n)$ (für $M \ge 0$) beschränkt ist. Ist nun $\delta > 0$, so folgt aus Satz 11.18, Teil (2) (mit $a_n = \chi(n)$ und $b_n = n^{-\delta}$), dass $L(\chi, \delta)$ konvergiert. Nach Teil (4) desselben Satzes folgt dann die Behauptung. □

11.25. **Bemerkung.** Tatsächlich ist $s_0 = 0$ die Konvergenzabszisse von $L(\chi, s)$. Die Abszisse der absoluten Konvergenz ist dagegen $s_1 = 1$.

11.26. **Satz.** (Landau) *Sei (a_n) eine Folge nichtnegativer reeller Zahlen, und sei $f(s)$ eine auf $\operatorname{Re} s > \sigma_0$ holomorphe Funktion, so dass für $\operatorname{Re} s > \sigma_1 \ge \sigma_0$ gilt*

$$f(s) = \sum_{n=1}^\infty a_n n^{-s}\,.$$

Dann konvergiert die rechts stehende Dirichlet-Reihe für alle s mit $\operatorname{Re} s > \sigma_0$ gegen $f(s)$.

Beweis. Indem wir a_n durch $a_n n^{-\sigma_0}$ ersetzen, können wir annehmen, dass $\sigma_0 = 0$ ist. Wenn die Behauptung falsch ist, dann gibt es $\delta > 0$, so dass die Reihe für $s = \delta$ nicht konvergiert, aber für $\operatorname{Re} s \geq \frac{3}{2}\delta$ konvergiert. Wir betrachten jetzt die Taylorreihe von f im Entwicklungspunkt 2δ. Weil f auf $\operatorname{Re} s > 0$ holomorph ist, ist der Konvergenzradius dieser Reihe mindestens 2δ. Die Koeffizienten der Reihe sind

$$c_k = \frac{f^{(k)}(2\delta)}{k!} = \frac{(-1)^k}{k!} \sum_{n=1}^{\infty} \frac{a_n (\log n)^k}{n^{2\delta}} .$$

Wir werten die Taylorreihe jetzt bei $s = \delta$ aus:

$$\begin{aligned} f(\delta) = \sum_{k=0}^{\infty} c_k (\delta - 2\delta)^k &= \sum_{k=0}^{\infty} \frac{\delta^k}{k!} \sum_{n=1}^{\infty} \frac{a_n (\log n)^k}{n^{2\delta}} \\ &\overset{(*)}{=} \sum_{n=1}^{\infty} \frac{a_n}{n^{2\delta}} \sum_{k=0}^{\infty} \frac{(\delta \log n)^k}{k!} = \sum_{n=1}^{\infty} \frac{a_n}{n^{2\delta}} e^{\delta \log n} = \sum_{n=1}^{\infty} \frac{a_n}{n^{\delta}} \end{aligned}$$

(An der Stelle $(*)$ verwenden wir, dass die Doppelreihe absolut konvergiert, da alle Terme ≥ 0 sind.) Am Ende steht aber gerade die Dirichlet-Reihe für $s = \delta$, die also doch konvergiert. Dieser Widerspruch beweist, dass die Reihe für alle s mit $\operatorname{Re} s > 0$ konvergiert. Nach dem Identitätssatz der Funktionentheorie muss sie dann auch gegen $f(s)$ konvergieren. □

11.27. **Bemerkungen.**

(1) Die Positivitätsvoraussetzung an die Koeffizienten a_n ist wesentlich. Man kann zum Beispiel (ähnlich wie für die Zetafunktion) zeigen, dass sich jede L-Reihe $L(\chi, s)$ zu einem nichttrivialen Dirichlet-Charakter χ zu einer auf ganz $\mathbb{C}$ holomorphen Funktion fortsetzen lässt. Trotzdem konvergiert die definierende Dirichlet-Reihe nur für $\operatorname{Re} s > 0$.

(2) Aus dem Beweis ergibt sich auch folgende Aussage:

Ist σ die Konvergenzabszisse der Dirichlet-Reihe im Satz, dann lässt sich f nicht holomorph in eine Umgebung von σ fortsetzen.

Wir können dazu $\sigma = 0$ annehmen. Wenn sich f holomorph in eine Umgebung von 0 fortsetzen lässt, dann gibt es $\varepsilon > 0$, so dass der Konvergenzradius der Taylorreihe von f bei ε größer als ε ist. Mit der gleichen Rechnung wie im obigen Beweis stellt man fest, dass die Dirichlet-Reihe auch noch links von 0 konvergieren müsste.

Wir werden den Satz von Landau dazu benutzen, das Nicht-Verschwinden der Werte $L(\chi, 1)$ für $\chi \neq \chi_0$ zu beweisen. Dazu brauchen wir noch zwei Lemmas.

11.28. **Lemma.** *Sei G eine endliche abelsche Gruppe und $g \in G$ ein Element der Ordnung m. Dann gilt*

$$\prod_{\chi \in \hat{G}} \big(1 - \chi(g) X\big) = (1 - X^m)^{\#G/m}$$

im Polynomring $\mathbb{C}[X]$.

Beweis. Die Abbildung $\hat{G} \ni \chi \mapsto \chi(g) \in \mu_m$ ist ein surjektiver Gruppenhomomorphismus. (Surjektiv: sonst wäre $\chi(g^k) = 1$ für alle $\chi \in \hat{G}$ für $k = m/\#H > 1$, wo

H das Bild des Homomorphismus ist, im Widerspruch zu Lemma 11.12.) Es folgt

$$\prod_{\chi\in\hat{G}}\big(1-\chi(g)X\big)=\prod_{\zeta\in\mu_m}(1-\zeta X)^{\#G/m}=(1-X^m)^{\#G/m}\,,$$

denn die m-ten Einheitswurzeln (die Elemente von μ_m) sind gerade die Nullstellen von X^m-1. □

11.29. **Lemma.** *Sei $F(s)=\prod_\chi L(\chi,s)$ das Produkt der L-Reihen zu allen Dirichlet-Charakteren mod N. Dann gilt für $\operatorname{Re} s>1$, dass*

$$F(s)=\sum_{n=1}^{\infty}\frac{a_n}{n^s}$$

mit $a_n\geq 0$ für alle n und $a_n\geq 1$ für alle n der Form $n=m^{\phi(N)}$ mit $m\geq 1$, $m\perp N$.

Beweis. Wir betrachten das Euler-Produkt von $F(s)$, gültig für $\operatorname{Re} s>1$. Dazu sei f_p die Ordnung der Restklasse von p in der Einheitengruppe $(\mathbb{Z}/N\mathbb{Z})^\times$ und $g_p=\phi(N)/f_p$. Dann gilt

$$F(s)=\prod_\chi L(\chi,s)=\prod_\chi\prod_p\big(1-\chi(p)p^{-s}\big)^{-1}=\prod_p\Big(\prod_\chi\big(1-\chi(p)p^{-s}\big)\Big)^{-1}$$

und nach Lemma 11.28:

$$=\prod_{p\nmid N}(1-p^{-f_ps})^{-g_p}=\prod_{p\nmid N}\sum_{k=0}^{\infty}\binom{g_p-1+k}{k}p^{-kf_ps}=\sum_{n=1}^{\infty}a_nn^{-s}\,.$$

In jedem Faktor des letzten Produkts haben die Glieder der Reihe nichtnegative Koeffizienten, und die Terme $p^{-l\phi(N)s}$ treten für alle $l\geq 0$ mit positivem Koeffizienten auf. Die Behauptung folgt durch Ausmultiplizieren des Produkts. □

Jetzt können wir endlich beweisen, dass $L(\chi,1)\neq 0$ ist.

11.30. **Satz.** (Dirichlet) *Ist χ ein nichttrivialer Dirichlet-Charakter mod N, dann ist $L(\chi,1)\neq 0$.*

Beweis. Angenommen, $L(\chi,1)=0$ für einen nichttrivialen Charakter χ. Dann hebt die Nullstelle von $L(\chi,s)$ bei $s=1$ den einfachen Pol von $L(\chi_0,s)$ im Produkt $F(s)$ auf. Da alle L-Reihen bis auf diesen einfachen Pol von $L(\chi_0,s)$ für $\operatorname{Re} s>0$ holomorph sind, muss $F(s)$ dann ebenfalls auf $\operatorname{Re} s>0$ holomorph sein.

Nach Lemma 11.29 ist $F(s)=\sum_{n\geq 1}a_nn^{-s}$ mit $a_n\geq 0$ und $a_{m^{\phi(N)}}\geq 1$ für alle $m\perp N$. Die $F(s)$ darstellende Reihe konvergiert daher für $s=1/\phi(N)$ nicht:

$$\sum_{n=1}^{\infty}a_nn^{-1/\phi(N)}\geq\sum_{m\geq 1,m\perp N}(m^{\phi(N)})^{-1/\phi(N)}=\sum_{m\geq 1,m\perp N}\frac{1}{m}=\infty$$

Nach dem Satz 11.26 von Landau müsste die Reihe aber für $\operatorname{Re} s>0$ konvergieren. Dieser Widerspruch beweist, dass unsere Annahme „$L(\chi,1)=0$ für ein χ“ falsch sein muss. □

Jetzt haben wir alles zusammen.

11.31. **Beweis von Satz 11.1.** Sei $N > 1$ und $a \perp N$. Nach Folgerung 11.16 gilt für $s > 1$

$$\sum_{p \equiv a \bmod N} p^{-s} = \frac{1}{\phi(N)} \sum_{\chi} \overline{\chi(a)} \log L(\chi, s) - f_{a,N}(s)$$
$$= \frac{\log L(\chi_0, s)}{\phi(N)} + \frac{1}{\phi(N)} \sum_{\chi \neq \chi_0} \overline{\chi(a)} \log L(\chi, s) - f_{a,N}(s)$$

mit $f_{a,N}(s)$ beschränkt für $s \to 1+$. Nach Satz 11.30 sind alle $L(\chi, 1) \neq 0$ für $\chi \neq \chi_0$, also ist $\log L(\chi, s)$ für $s \to 1+$ beschränkt. Nach Lemma 11.17 gilt $\log L(\chi_0, s) \to +\infty$ für $s \to 1+$, also folgt auch

$$\sum_{p \equiv a \bmod N} p^{-s} \to +\infty \qquad \text{für } s \to 1+.$$

Das ist nur möglich, wenn die links stehende Summe unendlich viele Terme hat. Also muss es unendlich viele Primzahlen $p \equiv a \bmod N$ geben. □

Wir wollen jetzt noch eine Verschärfung beweisen, nämlich dass in einem gewissen Sinn die Primzahlen sich gleichmäßig auf die primen Restklassen mod N verteilen. Dazu müssen wir ein Maß dafür definieren, welcher Anteil aller Primzahlen in einer gegebenen Menge von Primzahlen steckt.

11.32. **Definition.** Sei P eine Menge von Primzahlen. Falls der Grenzwert

$$\delta(P) = \lim_{s \to 1+} \frac{\sum_{p \in P} p^{-s}}{\sum_{p} p^{-s}}$$

existiert, heißt $\delta(P)$ die *Dirichlet-Dichte* der Menge P.

11.33. **Satz.** (Dirichlet) *Sei $N > 1$. Dann hat für jedes $a \perp N$ die Menge*

$$P_{a,N} = \{p \mid p \text{ Primzahl}, p \equiv a \bmod N\}$$

die Dirichlet-Dichte $\delta(P_{a,N}) = 1/\phi(N)$.

Beweis. Es ist

$$\log L(\chi_0, s) = \log \zeta(s) + \text{beschränkt} = \sum_{p} p^{-s} + \text{beschränkt}\,,$$

also haben wir

$$\sum_{p \equiv a \bmod N} p^{-s} = \frac{1}{\phi(N)} \sum_{p} p^{-s} + \text{beschränkt}\,,$$

woraus die Behauptung folgt. □

11.34. **Bemerkung.** Es erscheint vielleicht naheliegender, die *natürliche Dichte*

$$\delta'(P) = \lim_{X \to \infty} \frac{\#\{p \in P \mid p \leq X\}}{\pi(X)}$$

zu betrachten (hier ist $\pi(X)$ die Anzahl der Primzahlen $\leq X$). Tatsächlich gilt auch $\delta'(P_{a,N}) = 1/\phi(N)$; der Beweis ist aber schwieriger und benutzt Methoden, die auch beim Beweis des Primzahlsatzes $\pi(X) \sim X/\log X$ gebraucht werden.

Allgemein gilt, dass eine Menge P von Primzahlen, die eine natürliche Dichte besitzt, auch eine Dirichlet-Dichte hat, und beide stimmen überein. Das sieht man wie folgt: Sei $\pi_P(x) = \#\{p \in P \mid p \leq x\}$, dann gilt nach Voraussetzung

$$\pi_P(x) = \delta'(P)\pi(x) + \varepsilon(x)\pi(x)$$

mit $|\varepsilon(x)| \leq 1$ und $\varepsilon(x) \to 0$ für $x \to \infty$. Für $s > 1$ und $n \geq 1$ ist

$$n^{-s} = s \int_n^\infty x^{-s-1}\, dx\,.$$

Für eine beliebige Menge M natürlicher Zahlen und entsprechender Definition von $\pi_M(x)$ folgt dann

$$\sum_{n\in M} n^{-s} = \sum_{n\in M} s \int_n^\infty x^{-s-1}\, dx = s \int_1^\infty \pi_M(x) x^{-s-1}\, dx\,.$$

Wir erhalten also mit $a > 1$

$$\begin{aligned}
\sum_{p\in P} p^{-s} &= s \int_1^\infty \pi_P(x) x^{-s-1}\, dx \\
&= \delta'(P) s \int_1^\infty \pi(x) x^{-s-1}\, dx + s \int_1^\infty \varepsilon(x)\pi(x) x^{-s-1}\, dx \\
&= \delta'(P) \sum_p p^{-s} + s \int_1^a \varepsilon(x)\pi(x) x^{-s-1}\, dx + s \int_a^\infty \varepsilon(x)\pi(x) x^{-s-1}\, dx\,.
\end{aligned}$$

Der Betrag des zweiten Summanden in der letzten Zeile ist beschränkt durch as, denn $|\varepsilon(x)| \leq 1$ und $\pi(x) \leq x$. Sei $\varepsilon_a = \sup\{|\varepsilon(x)| \mid x \geq a\}$, dann ist der Betrag des dritten Summanden beschränkt durch

$$\varepsilon_a s \int_a^\infty \pi(x) x^{-s-1}\, dx \leq \varepsilon_a \sum_p p^{-s}\,.$$

Es folgt

$$\left| \frac{\sum_{p\in P} p^{-s}}{\sum_p p^{-s}} - \delta'(P) \right| \leq \frac{as}{\sum_p p^{-s}} + \varepsilon_a\,.$$

Wenn wir s gegen 1 gehen lassen, sehen wir, dass der Betrag der Differenz im Grenzwert durch ε_a beschränkt ist. Da wir a beliebig groß wählen können und $\varepsilon_a \to 0$ für $a \to \infty$, ergibt sich wie gewünscht, dass

$$\lim_{s\to 1+} \frac{\sum_{p\in P} p^{-s}}{\sum_p p^{-s}} = \delta'(P)\,.$$

Die umgekehrte Implikation, d.h., dass die Existenz der Dirichlet-Dichte die Existenz der natürlichen Dichte zur Folge hat, ist aber falsch.

11.35. **Bemerkung.** Obwohl wir also eine Art Gleichverteilung der Primzahlen auf die verschiedenen primen Restklassen mod N haben, kann es interessante Phänomene geben, wenn man die Anzahl der Primzahlen $\leq x$ in den Restklassen miteinander vergleicht. Zum Beispiel scheint es fast immer mehr Primzahlen $\equiv 3 \bmod 4$ zu geben als Primzahlen $\equiv 1 \bmod 4$. Wer darüber mehr wissen möchte, dem sei der sehr schöne Artikel „Prime Number Races“[17] von Andrew Granville und Greg Martin ans Herz gelegt.

Wir wollen nun zum Abschluss noch in einigen Fällen den Wert $L(\chi, 1)$ berechnen.

11.36. **Definition.** Ein Dirichlet-Charakter χ mod N heißt *primitiv*, falls χ nicht von einem Dirichlet-Charakter χ' modulo einem echten Teiler N' von N im folgenden Sinn induziert wird:

$$\chi(n) = \begin{cases} \chi'(n) & \text{falls } n \perp N, \\ 0 & \text{sonst.} \end{cases}$$

Wenn wir χ und χ' mit Charakteren der Einheitengruppen von $\mathbb{Z}/N\mathbb{Z}$ und $\mathbb{Z}/N'\mathbb{Z}$ identifizieren, dann bedeutet obige Relation, dass das folgende Diagramm kommutiert:

$$\begin{array}{ccc} (\mathbb{Z}/N\mathbb{Z})^\times & \longrightarrow & (\mathbb{Z}/N'\mathbb{Z})^\times \\ & {\scriptstyle\chi}\searrow \quad \swarrow{\scriptstyle\chi'} & \\ & \mathbb{C}^\times & \end{array}$$

11.37. **Lemma.** *Ist χ ein primitiver Dirichlet-Charakter mod N, und ist $N = dN'$ mit $d > 1$, dann gilt*

$$\sum_{j=0}^{d-1} \chi(jN' + k) = 0$$

für alle $k \in \mathbb{Z}$.

Beweis. Ist $\mathrm{ggT}(jN' + k) > 1$ für alle j, dann sind alle Terme in der Summe null. Anderenfalls gibt es $j_0 \in \mathbb{Z}$, so dass $(j_0N' + k) \perp N$. Die Summe ändert sich nicht, wenn wir k durch $j_0N' + k$ ersetzen, also können wir annehmen, dass $k \perp N$. Sei $l \in \mathbb{Z}$ mit $kl \equiv 1 \bmod N$. Dann ist $\chi(l) \neq 0$ und

$$\chi(l) \sum_{j=0}^{d-1} \chi(jN' + k) = \sum_{j=0}^{d-1} \chi(l)\chi(jN' + k) = \sum_{j=0}^{d-1} \chi(jlN' + 1)\,;$$

hier summieren wir über die Untergruppe $(1+N'\mathbb{Z}/N\mathbb{Z}) \cap (\mathbb{Z}/N\mathbb{Z})^\times$ von $(\mathbb{Z}/N\mathbb{Z})^\times$, also den Kern des kanonischen Homomorphismus $(\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N'\mathbb{Z})^\times$. Da χ primitiv ist, ist χ ein nichttrivialer Charakter auf dieser Untergruppe, also muss die Summe nach Satz 11.13 verschwinden. □

11.38. **Definition.** Sei χ ein Dirichlet-Charakter mod N, und sei $a \in \mathbb{Z}$. Wir definieren die *Gauß-Summe* von χ als

$$g_a(\chi) = \sum_{n=0}^{N-1} \chi(n) e^{2\pi i a n/N}\,.$$

Im Fall $a = 1$ schreiben wir auch $g(\chi)$ für $g_1(\chi)$.

[17] Amer. Math. Monthly **113**, no. 1, 1–33 (2006)

11.39. **Satz.** *Ist χ ein primitiver Dirichlet-Charakter mod N, dann gilt*

$$g_a(\chi) = \overline{\chi(a)} g(\chi)$$

und

$$|g(\chi)|^2 = N\,.$$

Beweis. Ist $\chi(a) = 0$, also $d = \operatorname{ggT}(a, N) > 1$, dann ist $a = da'$, $N = dN'$, und

$$\begin{aligned} g_a(\chi) &= \sum_{n=0}^{N-1} \chi(n) e^{2\pi i a n/N} = \sum_{n=0}^{N-1} \chi(n) e^{2\pi i a' n/N'} \\ &= \sum_{k=0}^{N'-1} \Big(\sum_{j=0}^{d-1} \chi(jN' + k) \Big) e^{2\pi i a' k/N'} = 0 = \overline{\chi(a)} g(\chi) \end{aligned}$$

nach Lemma 11.37.

Im Fall $a \perp N$, also $\chi(a) \neq 0$, gilt

$$\begin{aligned} \chi(a) g_a(\chi) &= \chi(a) \sum_{n=0}^{N-1} \chi(n) e^{2\pi i a n/N} = \sum_{n=0}^{N-1} \chi(an) e^{2\pi i a n/N} \\ &= \sum_{n=0}^{N-1} \chi(n) e^{2\pi i n/N} = g(\chi)\,, \end{aligned}$$

denn $n \mapsto an$ induziert eine Permutation von $\mathbb{Z}/N\mathbb{Z}$. Multiplikation mit $\overline{\chi(a)}$ (unter Beachtung von $|\chi(a)|^2 = 1$) liefert die Behauptung.

Für die zweite Aussage rechnen wir

$$\begin{aligned} \phi(N) |g(\chi)|^2 &= \sum_{a=0}^{N-1} g_a(\chi) \overline{g_a(\chi)} \\ &= \sum_{a=0}^{N-1} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \chi(n) \overline{\chi(n')} e^{2\pi i a (n-n')/N} \\ &= \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \chi(n) \overline{\chi(n')} \sum_{a=0}^{N-1} e^{2\pi i a (n-n')/N} \\ &= N \sum_{n=0}^{N-1} |\chi(n)|^2 = N\phi(N)\,. \end{aligned}$$

□

Damit haben wir die diskrete Fourier-Transformation von χ berechnet:

11.40. **Folgerung.** *Sei χ ein primitiver Dirichlet-Charakter mod $N > 1$. Dann gilt*

$$\chi(n) = \frac{1}{g(\bar{\chi})} \sum_{m=1}^{N-1} \overline{\chi(m)} e^{2\pi i m n/N}\,.$$

Beweis. Wende Satz 11.39 auf $\bar{\chi}$ an und beachte, dass $\chi(0) = 0$ ist. □

Wir brauchen noch ein Lemma über die Logarithmusreihe.

11.41. **Lemma.** *Sei $z = e^{i\varphi}$ mit $0 < \varphi < 2\pi$. Dann konvergiert die Reihe $\sum_{n=1}^{\infty} z^n/n$, und ihr Wert ist*

$$\sum_{n=1}^{\infty} \frac{z^n}{n} = -\log(1-z) = -\log\Big(2\sin\frac{\varphi}{2}\Big) + \frac{\pi-\varphi}{2}\,i\,.$$

Beweis. Die Reihe konvergiert nach Satz 11.18, Teil (2), denn

$$\Big|\sum_{n=1}^{N} z^n\Big| = \Big|z\frac{1-z^N}{1-z}\Big| \le \frac{2}{|z-1|}$$

ist beschränkt. Dass der Wert dann $-\log(1-z)$ ist, folgt aus dem Abelschen Grenzwertsatz, und die Zerlegung in Real- und Imaginärteil ergibt sich aus

$$1 - z = -e^{i\varphi/2}(e^{i\varphi/2} - e^{-i\varphi/2}) = -ie^{i\varphi/2}2\sin\frac{\varphi}{2} = 2\sin\frac{\varphi}{2}\cdot e^{i(\varphi-\pi)/2}\,.$$

□

Damit können wir jetzt die Reihe für $L(\chi,1)$ umformen. Wir setzen dabei χ als primitiv und $N > 1$ voraus.

$$\begin{aligned}
L(\chi,1) &= \sum_{n=1}^{\infty}\frac{\chi(n)}{n}\\
&= \frac{1}{g(\bar\chi)}\sum_{n=1}^{\infty}\sum_{m=1}^{N-1}\overline{\chi(m)}\,\frac{e^{2\pi imn/N}}{n}\\
&= \frac{1}{g(\bar\chi)}\sum_{m=1}^{N-1}\overline{\chi(m)}\sum_{n=1}^{\infty}\frac{\big(e^{2\pi im/N}\big)^n}{n}\\
&= \frac{1}{g(\bar\chi)}\sum_{m=1}^{N-1}\overline{\chi(m)}\Big(-\log\Big(2\sin\frac{\pi m}{N}\Big) + \Big(\frac{\pi}{2} - \frac{\pi m}{N}\Big)i\Big)\\
&= -\frac{1}{g(\bar\chi)}\sum_{m=1}^{N-1}\overline{\chi(m)}\Big(\log\sin\frac{\pi m}{N} + \frac{\pi m}{N}i\Big)
\end{aligned}$$

Im vorletzten Schritt haben wir Lemma 11.41 benutzt, und im letzten, dass die Summe $\sum_{m \bmod N}\chi(m)$ verschwindet (um die konstanten Terme $\log 2$ und $\pi i/2$ loszuwerden).

Wir beschränken uns jetzt auf spezielle Charaktere.

11.42. **Definition.** Ein Dirichlet-Charakter χ mod N heißt *quadratisch* (oder *reell*), wenn $\chi^2 = \chi_0$ (oder äquivalent $\bar\chi = \chi$) ist. Das bedeutet, dass χ nur die Werte $0, 1, -1$ annimmt.

11.43. **Definition.** Sei χ ein Dirichlet-Charakter mod N. Dann gilt $\chi(-1)^2 = \chi(1) = 1$, also ist $\chi(-1) = \pm 1$. χ heißt *gerade*, wenn $\chi(-1) = 1$ und *ungerade*, wenn $\chi(-1) = -1$.

χ ist dann tatsächlich eine gerade bzw. ungerade Funktion $\mathbb{Z} \to \mathbb{C}$, denn

$$\chi(-n) = \chi(-1)\chi(n) = \pm\chi(n)\,.$$

11.44. **Satz.** *Sei χ ein primitiver quadratischer Dirichlet-Charakter mod $N > 1$. Dann ist für χ gerade*

$$L(\chi,1) = -\frac{1}{g(\chi)} \sum_{m=1}^{N-1} \chi(m) \log\sin\frac{\pi m}{N}$$

und für χ ungerade

$$L(\chi,1) = -\frac{\pi i}{N g(\chi)} \sum_{m=1}^{N-1} \chi(m) m\,.$$

Beweis. In jedem Fall ist $\bar\chi = \chi$, also gilt

$$L(\chi,1) = -\frac{1}{g(\chi)} \sum_{m=1}^{N-1} \chi(m) \Big(\log\sin\frac{\pi m}{N} + \frac{\pi m}{N} i \Big)\,.$$

Wir ersetzen in der Summe m durch $N-m$, dann erhalten wir

$$\begin{aligned} L(\chi,1) &= -\frac{1}{g(\chi)} \sum_{m=1}^{N-1} \chi(-m) \Big(\log\sin\Big(\pi - \frac{\pi m}{N}\Big) + \Big(\pi - \frac{\pi m}{N}\Big) i \Big) \\ &= -\chi(-1) \frac{1}{g(\chi)} \sum_{m=1}^{N-1} \chi(m) \Big(\log\sin\frac{\pi m}{N} - \frac{\pi m}{N} i \Big) \end{aligned}$$

Dabei haben wir benutzt, dass $\sin(\pi - x) = \sin x$ ist, und dass wir den konstanten Term πi in der Summe weglassen können.

Ist jetzt χ gerade, also $\chi(-1) = 1$, dann ergibt Addition beider Ausdrücke und Halbieren

$$L(\chi,1) = -\frac{1}{g(\chi)} \sum_{m=1}^{N-1} \chi(m) \log\sin\frac{\pi m}{N}\,,$$

denn die Imaginärteile in der Summe heben sich weg. Wenn χ ungerade ist, bekommen wir entsprechend

$$L(\chi,1) = -\frac{\pi i}{N g(\chi)} \sum_{m=1}^{N-1} \chi(m) m\,.$$

□

11.45. **Beispiel.** Für den nichttrivialen Charakter χ mod 4 ergibt sich

$$g(\chi) = i - i^3 = 2i$$

und dann

$$L(\chi,1) = -\frac{\pi i}{4 \cdot 2i}(1-3) = \frac{\pi}{4}\,,$$

was den Wert bestätigt, den man aus der Arkustangensreihe $x - x^3/3 + x^5/5 - \ldots$ für $x = 1$ mit dem Abelschen Grenzwertsatz bekommt.

Für den (geraden) nichttrivialen quadratischen Charakter $\chi(n) = \left(\frac{n}{5}\right)$ mod 5 haben wir mit $\zeta = e^{2\pi i/5}$

$$g(\chi) = \zeta - \zeta^2 - \zeta^3 + \zeta^4 = \sqrt{5}$$

und dann

$$L(\chi,1) = -\frac{1}{\sqrt5} \log \frac{\sin^2\frac{\pi}{5}}{\sin^2\frac{2\pi}{5}} = \frac{2}{\sqrt5} \log(2\cos\tfrac{\pi}{5}) = \frac{2}{\sqrt5} \log\frac{1+\sqrt5}{2}\,.$$

Wenn χ ein quadratischer Charakter ist, kann man den genauen Wert der zugehörigen Gauß-Summe bestimmen.

11.46. **Satz.** *Sei χ ein primitiver quadratischer Dirichlet-Charakter mod N. Dann gilt*

$$g(\chi) = \begin{cases} \sqrt{N} & \text{für } \chi \text{ gerade,} \\ i\sqrt{N} & \text{für } \chi \text{ ungerade.} \end{cases}$$

Wenn $N = p$ prim ist, dann ist χ genau dann gerade, wenn $p \equiv 1 \bmod 4$ ist.

Beachte, dass es keine primitiven Charaktere mod 2 gibt, so dass p automatisch ungerade ist.

Beweis. Wir können $g(\chi)$ sehr einfach bis aufs Vorzeichen bestimmen. Sei dazu χ ein primitiver quadratischer Charakter mod N. Dann gilt

$$N = |g(\chi)|^2 = g(\chi)\overline{g(\chi)} = g(\chi)g_{-1}(\chi) = \chi(-1)g(\chi)^2\,,$$

also ist $g(\chi) = \pm\sqrt{N}$, wenn χ gerade ist, und $g(\chi) = \pm i\sqrt{N}$, wenn χ ungerade ist.

Für $N = p$ ist der einzige primitive (oder auch nur nichttriviale) quadratische Charakter das Legendre-Symbol $n \mapsto \chi_p(n) = \left(\frac{n}{p}\right)$. (Die Charaktergruppe von $(\mathbb{Z}/p\mathbb{Z})^\times$ ist zyklisch, hat also nur ein Element der Ordnung 2.) Wir wissen, dass

$$\chi_p(-1) = \left(\frac{-1}{p}\right) = \begin{cases} 1 & \text{für } p \equiv 1 \bmod 4, \\ -1 & \text{für } p \equiv 3 \bmod 4, \end{cases}$$

also ist χ_p genau dann gerade, wenn $p \equiv 1 \bmod 4$ ist.

Es bleibt zu zeigen, dass wir das richtige Vorzeichen haben. Der Beweis ist nicht ganz einfach (Gauß selbst hat vier Jahre gebraucht, bis er einen fand); ich verweise deswegen hier einfach auf [IR, § 6.4]. □

11.47. **Folgerung.** *Sei χ ein primitiver quadratischer Charakter mod $N > 1$. Falls χ gerade ist, gilt*

$$L(\chi, 1) = -\frac{1}{\sqrt{N}} \sum_{m=1}^{N-1} \chi(m) \log \sin \frac{\pi m}{N}\,,$$

und falls χ ungerade ist,

$$L(\chi, 1) = -\frac{\pi}{N\sqrt{N}} \sum_{m=1}^{N-1} \chi(m) m\,.$$

11.48. **Lemma.** *Sei χ ein nichttrivialer quadratischer Dirichlet-Charakter. Dann ist $L(\chi, 1) > 0$.*

Beweis. Für $s > 1$ ist

$$L(\chi, s) = \prod_p \frac{1}{1 - \chi(p)p^{-s}} > 0\,,$$

denn alle Faktoren sind positiv ($1 - \chi(p)p^{-s} \geq 1 - p^{-s} > 1/2$) und das Produkt konvergiert. Da $L(\chi, 1) \neq 0$ nach Satz 11.30, muss $L(\chi, 1)$ auch positiv sein. □

11.49. **Folgerung.** *Sei $p \equiv 3 \bmod 4$ eine Primzahl. Dann ist die Summe der quadratischen Nichtreste mod p zwischen 1 und $p-1$ größer als die Summe der quadratischen Reste im selben Intervall.*

Im Durchschnitt sind die Nichtreste also größer als die Reste.

Beweis. Die Differenz der Summen (Reste minus Nichtreste) ist gerade

$$\sum_{m=1}^{p-1} \chi_p(m)m = -L(\chi_p, 1)\frac{p\sqrt{p}}{\pi} < 0$$

nach Folgerung 11.47 und Lemma 11.48. □

Wir können den Ausdruck $\sum_m \chi(m)m$ noch umschreiben.

11.50. **Lemma.** *Für $p \equiv 3 \bmod 4$ prim gilt*

$$\sum_{m=1}^{p-1} \chi_p(m)m = -\frac{p}{2-\chi_2(p)} \sum_{m=1}^{(p-1)/2} \chi_p(m)\,.$$

Beweis. Wir haben einerseits

$$\begin{aligned}\sum_{m=1}^{p-1} \chi_p(m)m &= \sum_{m=1}^{(p-1)/2} \chi_p(m)m + \sum_{m=1}^{(p-1)/2} \chi_p(p-m)\cdot(p-m)\\ &= \sum_{m=1}^{(p-1)/2} \chi_p(m)\cdot(2m-p)\\ &= 2\sum_{m=1}^{(p-1)/2} \chi_p(m)m - p\sum_{m=1}^{(p-1)/2} \chi_p(m)\end{aligned}$$

und andererseits

$$\begin{aligned}\sum_{m=1}^{p-1} \chi_p(m)m &= \sum_{m=1}^{(p-1)/2} \chi_p(2m)2m + \sum_{m=1}^{(p-1)/2} \chi_p(p-2m)\cdot(p-2m)\\ &= \sum_{m=1}^{(p-1)/2} \chi_p(2m)\cdot(4m-p)\\ &= 4\chi_p(2)\sum_{m=1}^{(p-1)/2} \chi_p(m)m - \chi_p(2)p\sum_{m=1}^{(p-1)/2} \chi_p(m)\,.\end{aligned}$$

Wir multiplizieren die erste Gleichung mit $2\chi_p(2)$ und subtrahieren die zweite, dann ergibt sich die behauptete Formel. □

11.51. **Folgerung.** *Für $p \equiv 3 \bmod 4$ prim gilt*

$$L(\chi_p, 1) = \frac{\pi}{(2-\chi_2(p))\sqrt{p}} \sum_{m=1}^{(p-1)/2} \chi_p(m)\,.$$

Es gibt mehr quadratische Reste mod p zwischen 1 und $(p-1)/2$ als quadratische Nichtreste.

Beweis. Wir setzen die Formel aus Lemma 11.50 in den Ausdruck für $L(\chi_p, 1)$ aus Folgerung 11.47 ein. Die zweite Aussage folgt dann wieder aus $L(\chi_p, 1) > 0$. □

11.52. **Beispiele.** Für die ersten paar $p \equiv 3 \bmod 4$ ergibt sich folgende Tabelle. Dabei steht unter R die Zahl der quadratischen Reste mod p von 1 bis $(p-1)/2$ und unter N die Zahl der Nichtreste.

p	$\chi_p(2)$	R	N	$R-N$	$L(\chi_p,1)$
3	-1	1	0	1	$\frac{\pi}{3\sqrt{3}}$
7	1	2	1	1	$\frac{\pi}{\sqrt{7}}$
11	-1	4	1	3	$\frac{\pi}{\sqrt{11}}$
19	-1	6	3	3	$\frac{\pi}{\sqrt{19}}$
23	1	7	4	3	$\frac{3\pi}{\sqrt{23}}$
31	1	9	6	3	$\frac{3\pi}{\sqrt{31}}$
43	-1	12	9	3	$\frac{\pi}{\sqrt{43}}$
47	1	14	9	5	$\frac{5\pi}{\sqrt{47}}$

Man sieht, dass für $p > 3$ ($p = 3$ ist ein spezieller Fall) immer

$$L(\chi_p, 1) = h\frac{\pi}{\sqrt{p}}$$

gilt mit einer positiven ungeraden ganzen Zahl

$$h = \begin{cases} R-N & \text{falls } p \equiv 7 \bmod 8, \\ \frac{1}{3}(R-N) & \text{falls } p \equiv 3 \bmod 8. \end{cases}$$

Hier bezeichnen R und N wie oben die Anzahl der quadratischen Reste bzw. Nichtreste mod p in $\{1, 2, \ldots, \frac{p-1}{2}\}$.

Diese Zahl h hat eine Bedeutung: Sie ist die *Klassenzahl* des Körpers $\mathbb{Q}(\sqrt{-p})$, also die Ordnung der *Idealklassengruppe* des zugehörigen Rings $\mathbb{Z}\left[\frac{1+\sqrt{-p}}{2}\right]$. Mehr dazu findet man in [IR, § 12,13].

Literatur

[IR] Kenneth Ireland und Michael I. Rosen: *A classical introduction to modern number theory,* Springer Graduate texts in mathematics **84**, 2nd edition, 1990.

[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)