Math. Semesterber. (2003)

Digital Object Identifier (DOI) 10.1007/s00591-003-0060-3

Christian Elsholtz

Kombinatorische Beweise des Zweiquadratesatzes und Verallgemeinerungen

Eingegangen am 5. Juli 2002 / Angenommen am 15. August 2002

Zusammenfassung. Ein klassischer Satz der Zahlentheorie besagt, dass man alle Primzahlen p der Form $4k+1$ als Summe von zwei Quadraten natürlicher Zahlen schreiben kann. Die meisten Beweise verwenden die Eigenschaft, dass $-1 \bmod p$ ein Quadrat ist. Ein hiervon wesentlich verschiedener Beweis geht auf ein Abzählargument von Liouville zurück, das von Heath-Brown und Zagier vereinfacht wurde. Dieser Beweis gilt derzeit als der kürzeste. Unklar blieb aber, wie man diesen Beweis motivieren kann, da er eine außerordentlich mysteriöse Abbildung verwendet. In dieser Arbeit zeigen wir, wie man den Beweis in einem allgemeineren Rahmen motivieren kann. Zudem gelingt es, den Beweis auf andere Fälle aus der Theorie der binären quadratischen Formen zu übertragen. Außerdem zeigen wir, dass man auch einen anderen der älteren Beweise in der Kurzform schreiben kann.

1. Einleitung

Der folgende auf Fermat und Euler zurückgehende Satz ist einer der bekanntesten Sätze der Zahlentheorie. Er wird auch als der Zweiquadratesatz von Fermat bezeichnet.

Satz 1. *Jede Primzahl* $p \equiv 1 \bmod 4$ *kann als Summe von zwei Quadraten natürlicher Zahlen geschrieben werden.*

Der derzeit kürzeste Beweis hierzu stammt von Zagier [24] und ist im englischen Original nur einen Satz lang: *Die durch*

$$(x,y,z) \mapsto \begin{cases} (x+2z, z, y-x-z) & \textit{falls } x < y-z \\ (2y-x, y, x-y+z) & \textit{falls } y-z < x < 2y \\ (x-2y, x-y+z, y) & \textit{falls } x > 2y \end{cases}$$

auf der endlichen Menge $\mathcal{S} = \{(x,y,z) \in \mathbb{N}^3 : x^2+4yz = p\}$ *definierte Involution hat genau einen Fixpunkt, so dass* $|\mathcal{S}|$ *ungerade ist; daher hat aber auch die Involution* $(x,y,z) \to (x,z,y)$ *einen Fixpunkt.* □

C. Elsholtz: Institut für Mathematik, TU Clausthal, Erzstraße 1,
38678 Clausthal-Zellerfeld, http://www.math.tu-clausthal.de/∼mace/
e-mail: elsholtz@math.tu-clausthal.de

Mathematics Subject Classification (2000): 11E25, 11A41

Schlüsselwörter: Two squares theorem, Binary quadratic forms

Der wesentliche Punkt des Beweises ist, dass man hier auf einer geschickt definierten endlichen Menge $\mathcal{S}$ zwei Involutionen definieren kann. Eine Involution auf einer Menge $\mathcal{M}$ ist eine Abbildung $\alpha : \mathcal{M} \to \mathcal{M}$ mit $\alpha^2 = \mathrm{id}$. Die eine der zwei Involutionen ist geeignet, um die Anzahl der Elemente der Menge zu zählen, die andere Involution ist geeignet, um daraus die Existenz eines besonderen Elementes, das die Zerlegung in zwei Quadrate liefert, zu folgern.

Bei obigem Beweis sind eine Reihe Details (insbesondere die Wohldefiniertheit der ersten Abbildung und ihre involutorische Eigenschaft) vom Leser nachzurechnen. Dieser Beweis (bzw. eine auf Heath-Brown zurückgehende Version) wird im Detail im „Buch der Beweise“ von Aigner und Ziegler [1] besprochen.[1] Dort wird zudem auch einer der klassischen Beweise (aufbauend auf einem Abzählargument bei Kongruenzen) samt Einführung in das Rechnen modulo Primzahlen besprochen. Wir geben daher nur eine kurze Erläuterung zu Zagiers Beweis, siehe Abschnitt 4.2.

Satz 1 ist zugleich die Grundlage der vollständigen Klassifikation derjenigen Zahlen, die sich als Summe von zwei Quadraten schreiben lassen.

Satz 2. *Eine natürliche Zahl n kann genau dann als Summe von zwei Quadraten ganzer Zahlen geschrieben werden, wenn für die Primfaktorzerlegung $n = p_1^{\gamma_1} \cdots p_r^{\gamma_r}$ (wobei die p_i paarweise verschiedene Primzahlen seien) gilt: falls $p_i \equiv 3 \bmod 4$, so ist γ_i gerade.*

Hierbei ist die 0 als Quadrat zugelassen, so dass z.B. $9 = 3^2 + 0^2$ erlaubt ist.

Im Abschnitt 2 dieses Aufsatzes werden wir kurz auf die Geschichte eingehen, die klassischen Beweise erwähnen und zeigen, wie man Satz 2 auf Satz 1 zurückführen kann. Das Hauptanliegen dieser Arbeit ist jedoch ein besseres Verständnis für derartig kurze, aber doch etwas mysteriöse Beweise im Stile von Zagier zu erreichen. Dies geschieht auf zweierlei Weisen. Zum einen werden wir einen anderen klassischen Beweis mit der Involutionsidee umformulieren und dadurch ebenfalls zu einer sehr kurzen Formulierung gelangen. Zum anderen werden wir Zagiers Beweis in einem etwas allgemeineren Rahmen betrachten. Dadurch gelingt es zu erklären, wie man auf den Beweis hätte kommen können, und ihn z.B. auf Primzahlen der Form $x^2 + 2y^2$ zu übertragen. Auch wenn die Untersuchungen hierzu etwas umfangreicher werden, so sind die Methoden genauso elementar und ohne Zahlentheorievorkenntnisse zugänglich, wie beim „Buchbeweis“ des Zweiquadratesatzes.

2. Geschichte und klassische Zugänge zum Zweiquadratesatz

Die Geschichte des Zweiquadratesatzes wird in Dickson [5] erläutert, man vergleiche auch Edwards [6]. Diophant von Alexandria kannte bereits Teile der Aussage von Satz 2. (In der Tat interpretierte Jacobi diese so, dass Diophant seiner Meinung nach bereits Satz 2 samt Beweis kannte, siehe Dickson [5], Seite 236.) Allgemein ist man aber der Meinung, dass erst Albert Girard (1595–1632) eine korrekte Formulierung der notwendigen und hinreichenden Bedingungen angab. Kurze Zeit später

[1] Die erste englische Auflage präsentiert Zagiers Version, die zweite englische Auflage und die deutsche Übersetzung präsentieren Heath-Browns Version.

gab Pierre de Fermat (1601–1665) (vermutlich unabhängig von Girard) äquivalente Bedingungen an, und behauptete mehrfach, diese beweisen zu können. Leider ist dieser Beweis nicht überliefert. Die Methoden für einen Beweis hätte er jedenfalls zur Verfügung gehabt. Der erste überlieferte Beweis stammt von Leonhard Euler (1707–1783).

Heute sind zahlreiche verschiedene Beweise der beiden obigen Sätze bekannt. Das Buch von Hardy und Wright [10] gibt fünf verschiedene Beweise von Satz 1. Die meisten bekannten Beweise zeigen zunächst (mehr oder weniger explizit), dass für Primzahlen $p \equiv 1 \bmod 4$ die Kongruenz $x^2 \equiv -1 \bmod p$ lösbar ist. Dies folgt z.B. mit $x = (\frac{p-1}{2})!$ oder mit $x = g^{\frac{p-1}{4}}$, wobei g ein erzeugendes Element der zyklischen Gruppe $(\mathbb{Z}/p\mathbb{Z})^\times$ ist. Auch wenn ein solches x explizit angegeben werden kann, ist das genaue Nachrechnen der Details vom Umfang her bereits der halbe Beweis von Satz 1; zumindest dann, wenn man keine Vorkenntnisse voraussetzt. Viele der Beweise, die die Gaußschen Zahlen $\mathbb{Z}[i]$, den Minkowskischen Gitterpunktsatz, das Schubfachprinzip, die Theorie der Kongruenzen oder Kettenbrüche verwenden, zeigen also letztlich implizit zunächst die Lösbarkeit von $x^2 \equiv -1 \bmod p$. Daher muss $x^2 + 1 = mp$ für eine positive Zahl m lösbar sein. Wäre nun das minimale m mit dieser Eigenschaft größer als 1, so entsteht ein Widerspruch zur Minimalität von m.

Wir zeigen nun, wie man den Beweis des Satzes 2 auf den von Satz 1 zurückführen kann.

Das hinreichende Kriterium von Satz 2 erhält man, indem man in einem ersten Schritt die als Summe zweier Quadrate darstellbaren Primzahlen klassifiziert. Der wichtigste Fall hiervon ist Satz 1 und wird ausführlich in den nächsten Abschnitten besprochen. Wenn wir diesen Satz für den Moment voraussetzen, so folgt die vollständige Klassifikation leicht im nächsten Lemma. In einem zweiten Schritt wird dann gezeigt, dass das Produkt zweier darstellbarer Zahlen selber wieder darstellbar ist.

Lemma 1. *Eine Primzahl p ist genau dann als Summe von zwei Quadraten darstellbar, wenn $p = 2$ oder $p \equiv 1 \bmod 4$ ist.*

Der Fall $p = 2 = 1^2 + 1^2$ ist offensichtlich. Der Fall $p \equiv 1 \bmod 4$ ist gerade der Bestandteil von Satz 1 und wird in den nächsten Abschnitten besprochen. Und dass Primzahlen $p \equiv 3 \bmod 4$ keine Darstellung haben, folgt unmittelbar aus der Beobachtung, dass Quadrate modulo 4 nur die Werte 0 und 1 annehmen: es ist $(2k)^2 = 4k^2$ und $(2k+1)^2 = 4(k^2+k)+1$. □

Lemma 2. *Sind m und n jeweils als Summe zweier Quadrate darstellbar, so auch ihr Produkt mn.*
Ist m als Summe zweier Quadrate darstellbar und ist $r \in \mathbb{N}$, so ist auch mr^2 als Summe zweier Quadrate darstellbar.

Beweis von Lemma 2. Es sei $m = x_1^2 + y_1^2$ und $n = x_2^2 + y_2^2$. Durch Nachrechnen zeigt man, dass $mr^2 = (rx_1)^2 + (ry_1)^2$ und $mn = (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2$ gilt. Die letzte Identität kann durch Multiplikation im Komplexen motiviert werden.

Es ist $m = (x_1 + y_1 i)(x_1 - y_1 i)$ und $n = (x_2 + y_2 i)(x_2 - y_2 i)$. Also ist

$$\begin{aligned} mn &= ((x_1 + y_1 i)(x_2 + y_2 i))((x_1 - y_1 i)(x_2 - y_2 i)) \\ &= (x_1 x_2 - y_1 y_2 + i(x_1 y_2 + x_2 y_1))((x_1 x_2 - y_1 y_2 - i(x_1 y_2 + x_2 y_1)) \\ &= (x_1 x_2 - y_1 y_2)^2 + (x_1 y_2 + x_2 y_1)^2. \quad \square \end{aligned}$$

Lemma 1 und Lemma 2 ergeben zusammen das hinreichende Kriterium aus Satz 2. Wir kommen nun zu dem notwendigen Kriterium von Satz 2.

Lemma 3. *Es sei $n = x^2 + y^2$ und p ein Primteiler von n mit $p \equiv 3 \bmod 4$. Dann sind auch x und y durch p teilbar.*

Da nach dem Lemma also $x = px'$ und $y = py'$ gilt, kann man in $n = x^2 + y^2 = p^2((x')^2 + (y')^2)$ den Faktor p^2 kürzen und das Lemma erneut anwenden. Das Lemma beweist also die Notwendigkeit der Bedingung in Satz 2.

Zum Beweis des Lemmas ist noch der kleine Satz von Fermat als Hilfsmittel für das Rechnen mit Kongruenzen nützlich:

Lemma 4. *Es sei p eine Primzahl und x eine ganze Zahl. Dann gilt:*

$$x^{p-1} \equiv \begin{cases} 0 & \textit{für } x \equiv 0 \bmod p \\ 1 & \textit{für } x \not\equiv 0 \bmod p. \end{cases}$$

Beweis von Lemma 3. Es ist für $p \equiv 3 \bmod 4$

$$(x^2 + y^2)(x^{p-3} - x^{p-5}y^2 + x^{p-7}y^4 \mp \cdots + y^{p-3}) = x^{p-1} + y^{p-1}.$$

Da aber $x^2 + y^2 \equiv 0 \bmod p$ ist, ist auch $x^{p-1} + y^{p-1} \equiv 0 \bmod p$. Da aber nun $p > 2$ ist, muss nach dem kleinen Satz von Fermat $x \equiv y \equiv 0 \bmod p$ gelten. $\square$

Den kleinen Satz von Fermat wiederum sieht man wie folgt:

Beweis von Lemma 4. Durchläuft $a_1, a_2, \ldots, a_{p-1}$ alle von 0 verschiedenen Restklassen modulo p und sei $x \not\equiv 0 \bmod p$, so durchläuft auch $xa_1, xa_2, \ldots, xa_{p-1}$ alle von 0 verschiedenen Restklassen, so dass folgt:

$$\prod_{i=1}^{p-1} a_i \equiv \prod_{i=1}^{p-1} (xa_i) \equiv x^{p-1} \prod_{i=1}^{p-1} a_i \bmod p.$$

Da das Produkt nicht die Nullrestklasse ist, kann man kürzen und es folgt $x^{p-1} \equiv 1 \bmod p$. $\square$

Der kleine Satz von Fermat erlaubt unmittelbar, auch das Inverse einer Restklasse zu definieren. Es sei x eine von 0 verschiedene Restklasse. Dann ist $x^{-1} \bmod p$ die Restklasse, für die $xx^{-1} \equiv 1 \bmod p$ gilt. Das Inverse kann also auch als $x^{p-2} \bmod p$ geschrieben werden.

Insgesamt ist Satz 2 auf Lemma 1 und damit auf Satz 1 zurückgeführt.

3. Ein anderer kurzer Beweis

Einer der fünf Beweise von Fermats Zweiquadratesatz im Buch von Hardy und Wright [10] verwendet ein Abzählargument in Gittern, die als Lösung des p-Damenproblems interpretiert werden können. Als Quelle wird dort Grace [9] zitiert. Der Beweis muss aber älter sein, denn bei Dickson [5] (Seite 245) findet sich ein Hinweis auf Lucas (siehe z.B. Aubry [2]). Auch Pólya [16] behandelt den Beweis ausführlich. Bei diesem Beweis wird bereits vorausgesetzt, dass eine Lösung von $x^2 \equiv -1 \bmod p$ existiert. Wie oben erwähnt ist das aber bereits der halbe Beweis, wenn man keine Vorkenntnisse voraussetzt. Wir geben eine Version an, bei der in einem ersten Teil diese Eigenschaft bewiesen wird und diese Information dann direkt in den zweiten Teil des Beweises einfließt. Hier ist also ein anderer kurzer kombinatorischer Beweis, aufgeschrieben im Stile von Zagiers Beweis.

Die durch

$$a \mapsto \begin{cases} a^{-1} \bmod p & \textit{falls } 2 \le (a^{-1} \bmod p) \le \frac{p-1}{2} \\ -a^{-1} \bmod p & \textit{sonst} \end{cases}$$

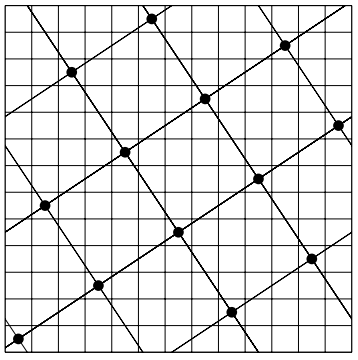
auf der endlichen Menge $\mathcal{S} = \{2 \le a \le \frac{p-1}{2}\}$ definierte Involution hat mindestens einen Fixpunkt z, so dass der Fundamentalbereich des durch

$$S_z = \{(x,\ zx \bmod p), 0 \le x < p\}$$

definierten Gitters ein Quadrat mit Fläche p sein muss, woraus nach Pythagoras der Zweiquadratesatz folgt. □

Hierbei sind die Repräsentanten modulo p als $\{0, 1, 2, \ldots, p-1\}$ gewählt. Dass $|\mathcal{S}|$ ungerade ist, ist wegen $p \equiv 1 \bmod 4$ offensichtlich. Auch die Wohldefiniertheit und die involutorische Eigenschaft (und damit die Existenz des Fixpunktes) ist hier viel einfacher einzusehen. Wegen $a^2 \not\equiv 1 \bmod p$ gibt es keinen Fixpunkt aus der ersten Alternative der Abbildung ($(a+1)(a-1) \equiv 0 \bmod p$ hat für Primzahlen p nur ± 1 als Lösung), so dass also die zweite Alternative einen Fixpunkt mit $z^2 \equiv -1 \bmod p$ liefert.

Wenn (c, cz) ein Gitterpunkt von S_z ist, der am nächsten am Ursprung $(0,0) \in S_z$ liegt, dann sind wegen $z^2 \equiv -1 \bmod p$ auch $(cz, -c)$ und $(cz + c, cz - c)$ Gitterpunkte. Hierbei rechnen wir konsequent modulo p, d.h. gegenüberliegende Seiten sind verbunden. Diese vier Punkte definieren ein Quadrat, in dem kein anderer der p Gitterpunkte liegen kann, denn sonst hätte man einen Gitterpunkt, der näher an einem der vier Eckpunkte liegt, und damit auch einen Punkt, der näher am Ursprung liegt, als der nächstliegende Punkt (c, cz).



Wir berechnen nun die Fläche des Quadrates. Die Fläche A des Quadrates ist ein Vielfaches von p, denn es gilt $A \equiv c^2 + (cz)^2 \equiv 0 \bmod p$. Wir zeigen

nun, dass sogar $A = p$ gilt. Jeder der p Gitterpunkte definiert zugleich den linken unteren Eckpunkt eines Quadrates, so dass es genau p überlappungsfreie kongruente Quadrate gibt. Da die Gesamtfläche p^2 beträgt, ist die Fläche eines Quadrates p, woraus Satz 1 mit Pythagoras folgt. Eine Alternative zu dieser Betrachtung erhält man, indem man das Gitter in der Ebene beliebig weit fortsetzt. Legt man nun einen sehr großen Kreis mit Fläche F in die Ebene, so ist die Anzahl der Gitterpunkte im Kreis asymptotisch $\frac{F}{p}$, denn wenn man $x \bmod p$ festlegt, so ist auch $y \bmod p$ bestimmt. Dies beweist erneut, dass die Fläche des Fundamentalbereiches p ist.

Was all dies besagt: man kann diesen Beweis aus dem Buch von Hardy und Wright [10] ebenso kurz formulieren wie den Beweis der Einleitung nach Zagier. Bezüglich der Einfachheit der Abbildung haben wir im Vergleich einen erheblichen Vorteil. Dafür erhalten wir im Bereich der Gitterargumentation eine Stelle, die erläutert werden muss, wie dies ja auch bei Zagiers Beweis der Fall war. Jeder Leser mag für sich entscheiden, ob obige Argumentation genügend einfach ist. Jedenfalls kann man sich obige Kurzform sicher gut einprägen und die Details dann leicht ausfüllen.

4. Der Beweis nach Liouville, Heath-Brown und Zagier und Verallgemeinerungen

4.1. Einführung

Ein Beweis, der von den in Abschnitt 2 erwähnten klassischen Beweisen deutlich verschieden ist, geht auf ein Abzählargument von Liouville (1809-1882) zurück. Man findet eine Ausarbeitung von Liouvilles Methoden z.B. in Dickson [5], Bachmann [3], Uspensky und Heaslet [19] oder Nathanson [15]. Wer sich diese Darstellungen angesehen hat, wird aber zu schätzen wissen, dass es Heath-Brown gelang, Liouvilles Beweis umzuformulieren. Zagier wiederum gab davon die in der Einleitung zitierte „One-sentence“ -Version an. Aber dabei sind natürlich die Details vom Leser nachzurechnen, der sozusagen den komprimierten Beweis dann wieder dekomprimieren muss.

Kürzlich erschien im American Mathematical Monthly eine Arbeit von Jackson [12], wo eine noch merkwürdigere Abbildung für Primzahlen der Form $x^2 + 2y^2$ verwendet wird. Wir werden die Frage, wie man diese Abbildungen erhalten kann, systematisch angehen und eine befriedigende Lösung erhalten. Außerdem können wir den Ansatz auf folgende Sätze übertragen:

Satz 3. *Es sei p eine Primzahl.*

a) Für $p = 8k + 3$ gibt es eine Lösung von $p = x^2 + 2y^2$ in natürlichen Zahlen.
b) Für $p = 8k + 7$ gibt es eine Lösung von $p = x^2 - 2y^2$ in natürlichen Zahlen.
c) Für $p = 8k + 5$ gibt es eine Lösung $p = x^2 + y^2$ in natürlichen Zahlen. (Ein neuer Beweis!)

Satz 4. *Es sei p eine Primzahl.*

a) Für $p = 12k + 7$ gibt es eine Lösung von $p = 3x^2 + 4y^2$ in natürlichen Zahlen.
b) Für $p = 12k + 11$ gibt es eine Lösung von $p = 3x^2 - 4y^2$ in natürlichen Zahlen.

4.2. Zagiers Beweis

Hier geben wir einige Erläuterungen zu Zagiers Beweis aus dem ersten Abschnitt. Da dieser Beweis in dem Buch von Aigner und Ziegler in allen Details wiedergegeben ist, fassen wir uns hier kurz. Wir sollten aber erwähnen, dass es keine Lösungen mit $y - z = x$ oder $x = 2y$ geben kann, da andernfalls $p = x^2 + 4yz$ nicht prim wäre. Weiterhin werden Lösungen (x, y, z) mit $x < y - z$ auf Lösungen mit $x > 2y$ abgebildet, und umgekehrt. Lösungen mit $y - z < x < 2y$ werden auf Lösungen mit der gleichen Bedingung abgebildet. Fixpunkte der ersten Abbildung liegen also nur in der letzten Menge. Die Fixpunktbedingung liefert hier $x = y$. Da aber $p = x^2 + 4yz$ prim sein soll, muss für einen Fixpunkt $x = y = 1$ gelten; deswegen gibt es genau einen Fixpunkt. Da es also nun genau einen Fixpunkt der ersten Abbildung gibt, muss $|\mathcal{S}|$ ungerade sein, und daher hat jede andere Involution eine ungerade Anzahl Fixpunkte (also mindestens einen). Der Fixpunkt der zweiten Involution liefert dann aber mit $y = z$ die Zerlegung $p = x^2 + 4yz = x^2 + 4y^2$.

Es ist praktisch, für die folgenden Abschnitte noch etwas Notation bereitzustellen. Die erste Abbildung nennen wir $\alpha : \mathcal{S} \to \mathcal{S}$. Diese kann man auch mittels der Matrizen

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & 1 \\ -1 & 1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & -2 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

beschreiben. Zagiers zweite Abbildung nennen wir analog $\beta : \mathcal{S} \to \mathcal{S}$. Wegen $(x, y, z) \mapsto (x, z, y)$ korrespondiert sie zu der Matrix

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

4.3. Wie kann man diese Abbildungen erhalten?

Es ist möglich, die Abbildung α zu konstruieren, wenn man nur davon ausgeht, dass es eine derartige Abbildung mit den gesuchten Eigenschaften überhaupt gibt. Wir suchen also eine Abbildung,

(1) die linear ist, mit ganzzahligen Einträgen in der Matrix, die von p (bzw. $k = \frac{p-1}{4}$) unabhängig sind,
(2) die Lösungen von $p = 4k + 1 = x^2 + 4yz$ auf ebensolche Lösungen abbildet,
(3) die die einfachste Lösung, nämlich $(1, 1, k)$, als einzigen Fixpunkt hat.

Wir werden sehen, dass uns dies eindeutig zu $B = \begin{pmatrix} -1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$ führt. Dazu machen wir einen Ansatz mit unbestimmten Koeffizienten: $B = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$.

Aus der Fixpunkteigenschaft (3) folgt, dass $a+b+ck=1, d+e+fk=1$ und $g+h+ik=k$ gelten muss. Da nun die Matrixeinträge als von k unabhängig vorausgesetzt sind, folgt: $c=f=0$ und $i=1$.

Eigenschaft 2 liefert nun:

$$(x')^2+4y'z' = (ax+by+cz)^2+4(dx+ey+fz)(gx+hy+iz) \stackrel{!}{=} x^2+4yz.$$

Ein Koeffizientenvergleich für x^2, xy und yz liefert

$$\begin{aligned} a^2+4dg &= 1\\ 2ab+4(dh+eg) &= 0\\ 2bc+4(ei+fh) &= 4. \end{aligned}$$

Mit $c=f=0$ folgt $ei=1$, wegen $i=1$ folgt $e=1$ und aus $d+e+fk=1$ folgt $d=0$. Es folgt also auch $a^2=1$. Falls nun $a=1$ wäre, so folgte aus $a+b+ck=1$, dass $b=0$ wäre. Aus $2ab+4(dh+eg)=0$ folgte weiter, dass $g=0$ wäre, und aus $g+h+ik=k$, dass auch $h=0$ gelten müsste. Die Matrix B wäre also die Einheitsmatrix, die sicher mehr als einen Fixpunkt hat. Es muss also $a=-1$ und damit $b=2, g=1$ und schließlich $h=-1$ gelten. Damit ist die Matrix B bestimmt. (Wir haben nicht einmal benötigt, dass B eine Involution sein soll. Dies hätte wegen $B^2=I$, wobei I die Einheitsmatrix bezeichne, eine Reihe weiterer Ansatzpunkte zur Bestimmung der Koeffizienten bereitgestellt.)

Die Abbildung ist in dieser Form allerdings nur für $-x+2y>0$ und $x-y+z>0$ verwendbar. Man kann sich aber leicht helfen, denn die Matrix $X=\begin{pmatrix}-1&0&0\\0&0&1\\0&1&0\end{pmatrix}$ erlaubt einem, das Vorzeichen der Zeilenbedingungen zu vertauschen. Wir erhalten also $A=BX=\begin{pmatrix}1&0&2\\0&0&1\\-1&1&-1\end{pmatrix}$ und $C=XB=\begin{pmatrix}1&-2&0\\1&-1&1\\0&1&0\end{pmatrix}$. Wir erhalten hier also die Zeilenbedingungen $-x+y-z>0$ für A und $x-2y>0$ für C. (Falls $x-2y>0$ gilt, so folgt erst recht $x-y+z>0$). Diese Bedingungen partitionieren also die Lösungsmenge $\mathcal{S}$. Eine Alternative, um A und C zu finden, wäre gewesen, für sehr kleine Primzahlen (nämlich $p=13,17,29$) nachzusehen, welche Lösungen noch nicht von B behandelt werden können. Für $p=13$ muss $(1,3,1)$ auf $(3,1,1)$ abgebildet werden, Für $p=17$, muss $(1,4,1)$ auf $(3,1,2)$ abgebildet werden. Für $p=29$ schließt man zunächst (mit der bereits vorhandenen Information) aus, dass $(1,7,1)$ auf $(5,1,1)$ abgebildet wird. Es folgt, dass $(1,7,1)$ auf $(3,1,5)$ abgebildet wird, was die Matrix A eindeutig bestimmt.

Obwohl wir anfangs nichts von der Dreigliederung der Abbildung wussten, haben wir sie also gleich mit erhalten. Insgesamt ist α also nur stückweise linear, so dass Eigenschaft 1 nicht ganz erfüllt ist. (Heath-Browns Version lässt auch negative Werte der Variablen zu, so dass diese Dreigliederung bei ihm vermieden ist.) Auf diese Weise erhalten wir die einfachste Involution mit den geforderten Eigenschaften.

4.4. Eine konstruktive Version des Beweises

Auch wenn Zagier in seiner Arbeit erwähnt, dass es sich um einen reinen Existenzbeweis handelt, kann man diesen Beweis zu einem konstruktiven Beweis ausbauen. Startet man nämlich mit einer Lösung (x, y, z) und iteriert abwechselnd α und β, so erhält man einen Zykel. Da α und β bijektiv sind, enthält dieser keine Vorperiode. Startet man nun mit $(1, 1, k)$, dem Fixpunkt von α, so wird man nach endlich vielen Schritten wieder zu $(1, 1, k)$ zurückkommen, wobei die vorhergehende Abbildung β gewesen sein muss.

$$(1,1,k) \xrightarrow{\beta} (1,k,1) \xrightarrow{\alpha} (3,1,k-2) \xrightarrow{\beta} \cdots \xrightarrow{\beta} (3,1,k-2) \xrightarrow{\alpha} (1,k,1) \xrightarrow{\beta} (1,1,k).$$

In oben stehender Zeile befinden sich also eine gerade Anzahl von Elementen, die symmetrisch zur Mitte stehen. In der Mitte muss insbesondere ein Fixpunkt sein. Da nun α aber nur einen Fixpunkt hat, handelt es sich um einen Fixpunkt von β, der die Zerlegung $p = x^2+(2y)^2$ liefert. Wendet man diese Idee auf zusammengesetzte Zahlen n an, so kann der Algorithmus durchaus auch von Interesse sein. Sei z.B. $n = p_1p_2$, wobei $p_1 \equiv p_2 \equiv 3 \bmod 4$ verschiedene Primzahlen seien, so liefert obiges Argument, dass ein Zykel, der $(1, 1, \frac{n-1}{4})$ als Fixpunkt von α enthält, einen weiteren Fixpunkt enthalten muss. Da aber nun eine Zahl n dieser Form keine Zerlegung in Summen von zwei Quadraten hat (siehe Satz 2), muss es sich um einen zweiten Fixpunkt von α handeln, der wegen $x = y$ einen Primfaktor von n liefert. Für eine schnelle Berechnung ist aber obiger Algorithmus zu langsam. Eine Beschleunigung wird in Shiu [17] diskutiert. Man vergleiche auch den Artikel von Wagon [21].

4.5. Die Diedergruppe D_6

Die Matrizen haben folgende interessante Eigenschaften.

$$B^2 = I,\ A^3 = C^3 = -I,\ A^6 = C^6 = I,$$

$$A = C^{-1},\ AB = (BX)B = B(XB) = BC = BA^5.$$

Insbesondere sind die 12 Matrizen $A, A^2, \ldots, A^6 = I, BA, \ldots, BA^6 = B$ ein Modell der Diedergruppe D_6.

5. Wie man neue Abbildungen konstruiert

5.1. Ein allgemeiner Ansatz

Man kann sich fragen, ob obige Methode nicht auch auf andere quadratische Formen $p = sx^2 + tyz$ angewendet werden kann. Es ist z.B. bekannt, dass für $p > 2$

$$p \equiv 1, 3 \bmod 8 \Leftrightarrow p = x^2 + 2y^2 \text{ gilt.}$$

Wenn man den Ansatz von Abschnitt 4.3 verallgemeinert, kann man zeigen, dass die Matrix $B = \begin{pmatrix} -1 & 2\frac{m}{n} & 0 \\ 0 & 1 & 0 \\ 4\frac{sm}{tn} & -4\frac{sm^2}{tn^2} & 1 \end{pmatrix}$ Lösungen von $p = sx^2 + tyz$ auf Lösungen derselben Gleichung abbildet und den Fixpunkt $(m, n, k' = \frac{p-sm^2}{tn})$ hat. Hierbei sind $m, n, s, k' \in \mathbb{N}$ und $t \in \mathbb{Z}$. Auch hier gilt $B^2 = I$. Allerdings sind die durch die Zeilen induzierten Grenzbedingungen $-x+2\frac{m}{n}y > 0$ und $4\frac{sm}{tn}x - 4\frac{sm^2}{tn^2}y + z > 0$ dieses Mal nicht so symmetrisch, dass die Anzahl der Lösungen in den beiden Randbereichen gleich groß wäre. Dennoch ist es im Fall $p = x^2 + 2y^2$ und $p = 3x^2 + 4y^2$ möglich, entsprechende Abbildungen zu konstruieren, die allerdings aus mehr als drei Teilen bestehen. Aber auch hier können alle Teile der Abbildung aus B und X erzeugt werden, wobei wie oben $X = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ ist.

Auch wenn dies im folgenden etwas komplizierter zu werden scheint, ist doch die Hauptidee die gleiche wie zuvor. Das Nachrechnen, dass die Abbildung wohldefiniert ist, könnte im Prinzip durch ein automatisches Beweissystem vorgenommen werden.

Wir werden jetzt systematisch untersuchen, welche der Kombinationen von s und t sich zu einem derartigen Beweis eignen. Wir nehmen an, dass auch $A = BX$ wie vorher Lösungen von $p = sx^2 + tyz$ auf ebensolche abbildet und ganzzahlige Einträge hat. Fasst man die Gleichung $p = sx^2 + tyz$ als Quadrik auf, so sieht man, dass es ebenso vernünftig ist, anzunehmen, dass $|\det A| = 1$ gilt; denn sonst würde A einen Teil der Lösungsmenge auf einen anderen Teil der Lösungsmenge mit verschiedener Fläche abbilden, so dass wir im allgemeinen nicht erwarten könnten, dass beide Teile die gleiche Anzahl von Lösungen haben. Analog nehmen wir an, dass die Eigenwerte von A Einheitswurzeln von 1 sind, da sonst die Matrix A unendlich viele verschiedene Potenzen erzeugen würde, wir mithin keine endliche Partitionierung erwarten könnten. Wir betrachten die Eigenwerte λ_i von

$$A = BX \begin{pmatrix} 1 & 0 & 2\frac{m}{n} \\ 0 & 0 & 1 \\ -4\frac{sm}{tn} & 1 & -4\frac{sm^2}{tn^2} \end{pmatrix} =: \begin{pmatrix} 1 & 0 & a \\ 0 & 0 & 1 \\ -c & 1 & -d \end{pmatrix}.$$

Wegen $ac = 2d$ folgt

$$0 = (1-\lambda)(0-\lambda)(-d-\lambda) - (1-\lambda) - (-c)(0-\lambda)a = (\lambda+1)(\lambda^2 + (d-2)\lambda + 1).$$

Daher ist $\lambda_1 = -1$ und $\lambda_{2,3} = -\frac{d-2}{2} \pm \sqrt{\left(\frac{d-2}{2}\right)^2 - 1}$.

Da nun d eine ganze Zahl sein soll und die λ_i den Betrag 1 haben sollen, folgt $d = 0, 1, 2, 3, 4$. Für $d \geq 5$ oder $d \leq -1$, wären $\lambda_{2,3} \in \mathbb{R}$ keine Einheitswurzeln. Eine Abbildung in diesen Fällen würde aus unendlich vielen Komponenten bestehen.

Wir betrachten diese Fälle einzeln. Da wir Primzahlen $p = sx^2 + tyz = sm^2 + tnz$ darstellen wollen, können wir $\mathrm{ggT}(sm, tn) = 1$ annehmen. Für ein festes d kann man wegen $d = \frac{4sm^2}{tn^2}$ recht schnell alle zulässigen Kombinationen für s und t finden.

5.2. $d = 0$

Für $d = 0$ ist $sm = 0$, so dass $p = tnz$ folgt. Dies liefert natürlich keine Darstellung von Primzahlen als Wert einer quadratischen Form.

5.3. $d = 1$

Es ist $d = \frac{4sm^2}{tn^2} = 1$, und $(s,t) = (s,n) = (t,m) = (m,n) = 1$. Es gibt also zwei Möglichkeiten:

a) $s = m = n = 1$, $t = 4$. Dies ist der oben untersuchte Fall von Zagier.
b) $s = m = t = 1$, $n = 2$.
Für $p = x^2 + yz$ wird die Lösung $p = x^2 + y^2 = y^2 + x^2$ zweimal gezählt. Um das alte Argument anwenden zu können, kann man z.B. die Symmetrie aufbrechen, indem man fordert, dass y und z gerade sein müssen. Dies liefert dann die folgende Variante des Beweises:
Die auf der endlichen Menge $\mathcal{S} = \{(x,y,z) \in \mathbb{N} \times 2\mathbb{N} \times 2\mathbb{N} : x^2 + yz = p\}$ definierte Involution

$$(x,y,z) \mapsto \begin{cases} (x+z, z, -2x+y-z) & \text{falls } 2x+z < y \\ (-x+y, y, 2x-y+z) & \text{falls } x < y < 2x+z \\ (x-y, 2x-y+z, y) & \text{falls } y < x \end{cases}$$

hat genau einen Fixpunkt, daher ist $|\mathcal{S}|$ ungerade, und die Involution $(x,y,z) \to (x,z,y)$ hat ebenfalls einen Fixpunkt.

5.4. $d = 2$

5.4.1. Der Fall $p = x^2 + 2yz$

Aus $d = \frac{4sm^2}{tn^2} = 2$ folgt $s = m = n = 1$, $t = 2$. Die Anzahl der Fixpunkte hängt aber von der Restklasse $p \bmod 8$ ab.

a)	Primzahlen	$p \equiv 3 \bmod 8$	induzieren	1	Fixpunkt,
b)	Primzahlen	$p \equiv 7 \bmod 8$	induzieren	2	Fixpunkte,
c)	Primzahlen	$p \equiv 5 \bmod 8$	induzieren	2	Fixpunkte,
d)	Primzahlen	$p \equiv 1 \bmod 8$	induzieren	3	Fixpunkte.

Wir beweisen die Fälle a), b) und c) (vgl. Satz 3). Die Fälle a) und d) wurden auch von Jackson [12] beobachtet. Wir sehen aber genauso wenig wir er, wie man Fall d) beweisen könnte, ohne andere Hilfsmittel über quadratische Formen zu verwenden. Die folgende Darstellung wird leider etwas länger als bisher. Bei dem Fall des Zweiquadratesatzes konnten wir auf das Buch [1] verweisen. Man könnte hier leicht sagen, es gehe genauso, aber vielleicht ist es ehrlicher, wenn wir dem Leser die Details in diesem Fall zeigen, aber im nächsten noch umfangreicheren Fall dann weglassen.

Es sei $\mathcal{S} = \{(x,y,z) \in \mathbb{N}^3 : x^2 + 2yz = p\}$. Die Abbildung $\alpha : \mathcal{S} \to \mathcal{S}$ lautet wie folgt:

$$\alpha = \begin{cases} A = BX & = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & 1 \\ -2 & 1 & -2 \end{pmatrix} \text{ falls } -2x+y-2z>0 \\ E = -XA^2 & = \begin{pmatrix} -3 & 2 & -2 \\ -2 & 2 & -1 \\ 2 & -1 & 2 \end{pmatrix} \begin{cases} \text{falls } -3x+2y-2z>0 \\ \text{und } 2x-y+2z>0 \\ (-2x+2y-z>0 \text{ folgt.})^2 \end{cases} \\ D = -A^2 & = \begin{pmatrix} 3 & -2 & 2 \\ 2 & -1 & 2 \\ -2 & 2 & -1 \end{pmatrix} \begin{cases} \text{falls } 3x-2y+2z>0 \\ \text{und } -2x+2y-z>0 \\ (2x-y+2z>0 \text{ folgt.}) \end{cases} \\ B = XA^3 & = \begin{pmatrix} -1 & 2 & 0 \\ 0 & 1 & 0 \\ 2 & -2 & 1 \end{pmatrix} \begin{cases} \text{falls } -x+2y>0 \\ \text{und } 2x-2y+z>0 \end{cases} \\ C = A^{-1} = XB & = \begin{pmatrix} 1 & -2 & 0 \\ 2 & -2 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{array}{l} \text{falls } x-2y>0, \\ (2x-2y+z>0 \text{ folgt implizit.}) \end{array} \end{cases}$$

Wegen $A^4 = I$ verwenden wir hierbei alle Matrizen der Form $(-1)^{j+1}A^j, (j = 1,2,3)$, und $(-1)^{j+1}XA^j, (j=2, j=3)$. Die Matrix XA hat eine rein negative Zeilenbedingung ($-x-2z>0$), wird also nicht benötigt.

Eine andere Darstellung derselben Abbildung lautet (vgl. Jackson [12]): (x,y,z) wird abgebildet auf

$$\begin{cases} (x-2y, z+2x-2y, y) & \text{falls} & y < \frac{x}{2} \\ (2y-x, y, 2x-2y+z) & \text{falls } \frac{x}{2} & < y < x+\frac{z}{2} \\ (3x-2y+2z, 2x-y+2z, -2x+2y-z) & \text{falls } x+\frac{z}{2} & < y < \frac{3}{2}x+z \\ (-3x+2y-2z, -2x+2y-z, 2x-y+2z) & \text{falls } \frac{3}{2}x+z & < y < 2x+2z \\ (x+2z, z, -2x+y-2z) & \text{falls } 2x+2z & < y. \end{cases}$$

Um die Matrizen und die Gebiete, für die sie gelten, unterscheiden zu können, nennen wir die Teilmengen von $\mathcal{S}$, die zu den Matrizen A, B, C, D und E korrespondieren, $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ bzw. $\mathcal{E}$. Für einen vollständigen Beweis ist zu zeigen, dass

1. $\alpha : \mathcal{S} \to \mathcal{S}$, d.h. dass α Lösungen (x,y,z) von $p = x^2+2yz$ auf (x',y',z') mit $p = x'^2 + 2y'z'$ abbildet,
2. $\alpha^2 = \mathrm{id}$ gilt,
3. die Grenzen ($x-2y=0, 2x-2y+z=0$ etc.) niemals angenommen werden,
4. die Mengen $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ und $\mathcal{E}$ eine Partitionierung der Menge S darstellen,
5. es genau einen Fixpunkt (Teil a) bzw. 2 Fixpunkte (Teile b und c) gibt.

[2] Wegen $-3x+2y-2z>0$ und $x+z>0$

5.4.2. Beweis von Satz 3

1. Da die Abbildung durch $-I, X$ und B erzeugt wird, reicht es, diesen Punkt für die Erzeuger nachzuweisen. Für $-I$ und X ist dies offensichtlich. Für B folgt:
$$(x')^2 + 2y'z' = (-x+2y)^2 + 2(y)(2x-2y+z) = x^2 + 2yz.$$
2. Die Matrix A bildet das Gebiet $\mathcal{A}$ auf das Gebiet $\mathcal{C}$ ab. Wegen $C = A^{-1}$ wird andersherum $\mathcal{C}$ auf $\mathcal{A}$ abgebildet. Der erste Teil der Behauptung folgt wegen $x' - 2y' = (x+2z) - 2z > 0$ und $2x' - 2y' + z' = 2(x+2z) - 2z + (-2x+y-2z) = y > 0$. Der zweite Teil folgt wegen $-2x'+y'-2z' > 0$ mit $x' = x - 2y,\ y = 2x - 2y + z,\ z' = y$. Daher ist $-2x' + y' - 2z' = z > 0$. Für die anderen Matrizen gilt $B^2 = D^2 = E^2 = I$. Die Matrix B bildet $\mathcal{B}$ auf sich selber ab. Das gleiche gilt für $D : \mathcal{D} \to \mathcal{D}$ und $E : \mathcal{E} \to \mathcal{E}$.
3. Wenn wir annehmen, dass eine Lösung auf einem der Ränder liegt, folgt ein Widerspruch wie folgt:
   (a) Für die Grenzen der ersten Zeile, nämlich $x-2y = 0,\ -x+2y = 0,\ 3x-2y+2z = 0,\ -3x+2y-2z = 0$, würde folgen, dass x gerade sein muss, was aber für ungerades p im Widerspruch zu $p = x^2 + 2yz$ steht.
   (b) Wäre $-2x+y-2z = 0$ oder $2x-y+2z = 0$, so wäre auch $p = x^2+2yz = x^2 + 2(2x+2z)z = (x+2z)^2$. Damit wäre p also keine Primzahl.
   (c) Analog für $2x-2y+z = 0$ oder $-2x-2y+z = 0$. Es wäre $p = x^2+2yz = x^2 + 2y(2y-2x) = (x-2y)^2$.
4. Dass die Abbildung in der Tat eine Partition der Lösungsmenge $\mathcal{S}$ darstellt, sieht man gut an der Reihenfolge in der zweiten der obigen Darstellungen.
5. Wir kommen nun zu den Fixpunkten. Hier haben wir die verschiedenen Restklassen $p \bmod 8$ zu unterscheiden. Die Matrizen A und C erzeugen keine Fixpunkte, da sie Lösungen von $\mathcal{A}$ auf $\mathcal{C}$ bzw. umgekehrt abbilden.
   Es sei (x, y, z) ein Fixpunkt von B. Die Fixpunktbedingung lautet:
$$B\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}-x+2y\\y\\2x-2y+z\end{pmatrix} \stackrel{!}{=} \begin{pmatrix}x\\y\\z\end{pmatrix}.$$
   Es ist also $x = y$. Da $p = x^2 + 2yz$ prim sein soll, folgt $x = y = 1$. B erzeugt also genau einen Fixpunkt.
   Für D erhalten wir:
$$D\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}3x-2y+2z\\2x-y+2z\\-2x+2y-z\end{pmatrix} \stackrel{!}{=} \begin{pmatrix}x\\y\\z\end{pmatrix}.$$
   Es gilt also $y = x+z$ und daher $p = x^2+2yz = x^2+2(x+z)z = (x+z)^2+z^2$. Ebenso
$$E\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}-3x+2y-2z\\-2x+2y-z\\2x-y+2z\end{pmatrix} \stackrel{!}{=} \begin{pmatrix}x\\y\\z\end{pmatrix}.$$
   Es folgt $y = 2x + z$ und daher $p = x^2 + 2yz = x^2 + 2(2x+z)z = (x + 2z)^2 - 2z^2$.

Um nun zu sehen, welche dieser Fixpunktbedingungen einen Fixpunkt liefern, überlegt man sich leicht, dass nur die Restklassen $0, 1, 4$ Quadrate modulo 8 sind. Falls $p \equiv 3 \bmod 8$ ist, erzeugen D und E also keinen Fixpunkt. Wie vorher folgt, dass $|\mathcal{S}|$ ungerade ist, und daher β einen Fixpunkt haben muss. Dieser Fixpunkt mit $y = z$ liefert $p = 8k + 3 = x^2 + 2y^2$. (Satz 3a).
Falls $p \equiv 7 \bmod 8$ ist, haben wir wieder den Fixpunkt von B. Das gleiche Argument wie oben zeigt, dass es keinen Fixpunkt von D geben kann. Allerdings kann man jetzt umgekehrt schließen. Da es wegen $p \equiv 7 \bmod 8$ keine Darstellung der Form $p = x^2 + 2y^2$ geben kann, muss es einen Fixpunkt von E geben (sonst würde ja obiges Argument durchgehen). Dies zeigt, dass $p \equiv 7 \bmod 8$ in der Form $p = x^2 - 2y^2$ geschrieben werden kann. (Satz 3b).
Es sei nun $p \equiv 5 \bmod 8$. Es gibt den Fixpunkt von B, es kann aber weder einen Fixpunkt von E geben, noch eine Darstellung der Form $p = x^2 + 2y^2$. Es muss also einen Fixpunkt von D geben, so dass p von der Form $x^2 + y^2$ sein muss. Dies gibt einen neuen Beweis des Zweiquadratesatzes für $p \equiv 5 \bmod 8$ (Satz 3c).
Es sei abschließend $p \equiv 1 \bmod 8$. Wir haben den trivialen Fixpunkt von B und (aufgrund des Zweiquadratesatzes in Zagiers Form) eine ungerade Anzahl von Fixpunkten von D. Um eine Darstellung der Form $p = x^2 + 2y^2$ nachzuweisen, reicht es aus, dass es eine ungerade Anzahl von Fixpunkten von E gibt. Es ist allerdings nicht offensichtlich, wie man dies mittels der hier benutzten Methoden beweisen kann. Immerhin zeigt dies einen Zusammenhang zwischen der Darstellbarkeit als $p = x^2 + 2y^2$ und $p = x^2 - 2y^2$.

5.5. $d = 3$

Hier gilt $d = \frac{4sm^2}{tn^2} = 3$. Erneut führt dies auf zwei Unterfälle.

a) $s = 3$, $m = n = 1$, $t = 4$.
b) $s = 3$, $m = t = 1$, $n = 2$, mit geradem y und z.

Wie oben für $d = 1$ sind beide Fälle äquivalent. Wir behandeln daher nur den ersten. Die quadratische Form $p = 3x^2 + 4yz$ stellt höchstens Primzahlen der Form $p \equiv 3 \bmod 4$ dar. Wir betrachten daher die Fälle $p = 12k + 7$ and $p = 12k + 11$ und gehen wie oben für $d = 2$ vor.

Es ist

$$B = \begin{pmatrix} -1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & -3 & 1 \end{pmatrix}, \quad A = BX = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & 1 \\ -3 & 1 & -3 \end{pmatrix}.$$

Wegen $A^6 = I$ benutzen wir die 9 Matrizen

$$(-1)^{j+1}A^j, (j = 1, \ldots, 5) \text{ und } (-1)^{j+1}XA^j, (j = 2, \ldots, 5).$$

(Wie im vorigen Fall ist die Matrix XA wegen der Zeilenbedingung $-x-2z>0$ nutzlos.) Die Abbildung α ordnet also (x,y,z) folgendes Bild zu:

$$\begin{cases}
(x-2y, 3x-3y+z, y) & \text{falls} & & y < \frac{x}{2} \\
(-x+2y, y, 3x-3y+z) & \text{falls} & \frac{x}{2} & < y < x+\frac{z}{3} \\
(5x-4y+2z, 6x-4y+3z, -3x+3y-z) & \text{falls} & x+\frac{z}{3} & < y < \frac{5}{4}x+\frac{z}{2} \\
(-5x+4y-2z, -3x+3y-z, 6x-4y+3z) & \text{falls} & \frac{5}{4}x+\frac{z}{2} & < y < \frac{3}{2}x+\frac{3}{4}z \\
(7x-4y+4z, 6x-3y+4z, -6x+4y-3z) & \text{falls} & \frac{3}{2}x+\frac{3}{4}z & < y < \frac{7}{4}x+z \\
(-7x+4y-4z, -6x+4y-3z, 6x-3y+4z) & \text{falls} & \frac{7}{4}x+z & < y < 2x+\frac{4}{3}z \\
(5x-2y+4z, 3x-y+3z, -6x+3y-4z) & \text{falls} & 2x+\frac{4}{3}z & < y < \frac{5}{2}x+2z \\
(-5x+2y-4z, -6x+3y-4z, 3x-y+3z) & \text{falls} & \frac{5}{2}x+2z & < y < 3x+3z \\
(x+2z, z, -3x+y-3z) & \text{falls} & 3x+3z & < y.
\end{cases}$$

Um Satz 4a) zu beweisen, reicht es zu zeigen, dass für $p \equiv 7 \bmod 12$ die Abbildung α genau einen Fixpunkt hat. Für $p \equiv 11 \bmod 12$ gibt es einen zweiten Fixpunkt, der die Lösung $3x^2-4y^2$ induziert.

Der komplette Beweis geht analog zu dem von Satz 3. Für die Details vergleiche man das Manuskript [8]. Wir gehen nur kurz auf die Fixpunkteigenschaft ein: Die Matrizen $A, -A^2, -A^4, A^5$ erzeugen keinen Fixpunkt. Die Matrizen $B, A^3, -XA^2, XA^3, -XA^4$ sind aber Involutionen und müssen genauer untersucht werden. Man kann zeigen, dass die Matrizen $A^3, -XA^2$ und XA^3 für Primzahlen $p \equiv 7 \bmod 12$ keinen Fixpunkt zulassen. B hat aber wie zuvor den trivialen Fixpunkt $(1,1,\frac{p-3}{4})$. Dies beweist Satz 4a).

Für Satz 4b) geht man wie in Satz 3 umgekehrt vor: $p \equiv 11 \bmod 12$ kann niemals von der Form $3x^2+4y^2$ sein. Es muss folglich einen weiteren Fixpunkt geben. Eine genauere Analyse zeigt, dass er von den Matrizen $-XA^2$ oder $-XA^4$ kommen muss, die beide als Fixpunkteigenschaft die Zerlegung $p=3x^2-4y^2$ garantieren.

5.6. $d=4$

Hier ist $d=\frac{4sm^2}{tn^2}=4$ und daher $s=t=m=n=1$ und

$$B=\begin{pmatrix} -1 & 2 & 0 \\ 0 & 1 & 0 \\ 4 & -4 & 1 \end{pmatrix}.$$

Aber $A=BX$ erzeugt bereits unendlich viele Potenzen. Es ist nicht auszuschließen, dass man einen Beweis des Zweiquadratesatzes mit einer unendlichen Partitionierung der Lösungsmenge finden kann. Da wir andererseits aber im Fall $s=t=1$ nichts wirklich Neues erwarten, untersuchen wir dies nicht weiter.

Interessanter wäre die Frage, ob man mittels einer unendlichen Partitionierung andere quadratische Formen behandeln kann. Wir haben ja oben vorausgesetzt, dass A endliche Ordnung hat.

6. Danksagung

Ich danke insbesondere Prof. B. Artmann, der mir das Thema nahe brachte. Einige der Resultate wurden bereits im Jahre 1990 gefunden (siehe [7]). Die Verallgemeinerungen fand ich 1995/96. Damals schrieb ich die Ergebnisse zwar auf, beabsichtigte aber nicht, sie zu veröffentlichen. Die Ergebnisse wurden im Frühjahr 1998 im Zahlentheorie Seminar Stuttgart vorgestellt. Die Arbeit wurde aktualisiert, als die Arbeit von Jackson [12] erschien. Der neue Beweis in Abschnitt 3 wurde 2001 gefunden. Ich danke Prof. J. Wolfart und Prof. G. Ziegler, die mich ermutigten, die Arbeit in eine endgültige Form zu bringen und zu veröffentlichen und Prof. L.G. Lucht und den Referenten, die zu einer Verbesserung des Manuskriptes Vorschläge machten. Weiterer Dank geht an die Professoren D.R. Heath-Brown, T. Jackson und K.S. Williams für Kopien ihrer Arbeiten und an Marco Loskamp für das Bild.

Literatur

1. Aigner, M., Ziegler, G.M.: Proofs from THE BOOK, Springer Verlag, Berlin, 1. Auflage 1998, 2. Auflage 2001, deutsche Übersetzung: Das Buch der Beweise, 2002
2. Aubry, A.: Les principes de la géométrie des quinconces. L'Enseignement Mathématique **13**, 187–203 (1911)
3. Bachmann, P.: Niedere Zahlentheorie, Nachdruck von Chelsea Publishing Co., New York, 1968
4. Barbeau, E.J.: Polynomials. Problem Books in Mathematics. Springer-Verlag, New York (korrigierter Nachdruck) 1995
5. Dickson, L.E.: History of the theory of numbers, Band 2. Nachdruck von Chelsea Publishing Co., New York, 1966
6. Edwards, H.M.: A genetic introduction to algebraic number theory. Springer-Verlag, New York, 1996
7. Elsholtz, C.: Primzahlen der Form $4k+1$ sind Summe zweier Quadrate. Mathematik lehren **62**, 58–61 (1994)
8. Elsholtz, C.: The Liouville–Heath-Brown–Zagier proof of the two squares theorem (Preprint 2001/10, Institut für Mathematik, TU Clausthal). Auf der Homepage des Autors erhältlich. http://www.math.tu-clausthal.de/∼mace/papers/papers.html
9. Grace, J.H.: The four square theorem. J. London Math. Soc. **2**, 3–8 (1927)
10. Hardy, G.H., Wright, E.M.: An introduction to the theory of numbers. 5. Auflage. Oxford University Press, New York, 1979
11. Heath-Brown, D.R.: Fermat's two squares theorem, Invariant, 1984, 3–5. (Unregelmässig erscheinende Zeitschrift der Invariant Society (Verein von Mathematikstudenten der Universität Oxford))
12. Jackson, T.: A short proof that every prime $p = 3 \pmod 8$ is of the form $x^2 + 2y^2$. Amer. Math. Monthly **107**, 447 (2000)
13. Jackson, T.: Automorphs and involutions. Tatra Mt. Math. Publ. **20**, 59–63 (2000)
14. Jackson, T.: Direct proofs of some of Euler's results. Number theory (Turku, 1999), 163–166, de Gruyter, Berlin, 2001
15. Nathanson, M.B.: Elementary methods in number theory. Graduate Texts in Mathematics, 195. Springer-Verlag, New York, 2000

16. Pólya, G.: Über die „doppelt-periodischen“ Lösungen des n-Damen Problems. In: Ahrens, W. (ed.) Mathematische Unterhaltungen und Spiele, Teubner, Leipzig, Volume II, 2. Auflage 1918, 364–374
17. Shiu, P.: Involutions associated with sums of two squares. Publ. Inst. Math. (Beograd) (N.S.) **59(73)**, 18–30 (1996)
18. Tikhomirov, V.: Three paths to Mt. Fermat-Euler. Let Lagrange, Zagier, and Minkowski be your guides. Quantum **4**, 5–7 (1994)
19. Uspensky, J.V., Heaslet, M. A.: Elementary Number Theory. McGraw-Hill Book Company, New York, 1939
20. Varouchas, I.: Une démonstration élémentaire du théorème des deux carrés. I.R.E.M. Bull. **6**, 31–39 (1984)
21. Wagon, S.: The Euclidean algorithm strikes again. Amer. Math. Monthly **97**, 125–129 (1990)
22. Wells, D.: Are these the most beautiful? Math. Intelligencer **12**(3), 37–41 (1990)
23. Williams, K.S.: Heath-Brown's elementary proof of the Girard-Fermat theorem. Carleton Coordinates 4–5 (1985)
24. Zagier, D.: A one-sentence proof that every prime $p \equiv 1 \pmod 4$ is a sum of two squares. Amer. Math. Monthly **97**, 144 (1990)