Binäre quadratische Formen unter besonderer Berücksichtigung solcher Formen, deren Diskriminante nicht fundamental ist

Diplomarbeit von Jacob Rosenthal im Fach Mathematik

Betreuer: Prof. Dr. Günter Köhler

Mathematisches Institut

Bayerische Julius - Maximilians - Universität Würzburg

Oktober 1994

Inhaltsverzeichnis

Einleitung

In der elementaren Zahlentheorie wird der folgende, von Fermat gefundene Satz bewiesen:

Genau dann ist eine Primzahl $p \in \boldsymbol{N}$ die Summe zweier Quadrate natürlicher Zahlen, wenn $p = 2$ oder $p \equiv 1 \bmod 4$ ist.

Fermat untersuchte also, welche Primzahlen p von der Form $p = x^2 + y^2$ mit $x,y \in \boldsymbol{N}$ sind. Weitergehend fragte er nach Primzahlen der Form $x^2 + 2y^2$ bzw. $x^2 + 3y^2$ und fand ähnliche Sätze. Auch formulierte er den folgenden Satz:

Ist $d \in \boldsymbol{N}$ keine Quadratzahl, so hat die Gleichung $x^2 - dy^2 = 1$ unendlich viele Lösungen in ganzen Zahlen.

Nach diesen Beispielen ist es naheliegend, und geschah auch durch Lagrange und vor allem Gauß, folgende allgemeinere Fragestellung zu behandeln:

Wieviele Lösungen $(x,y) \in \boldsymbol{Z} \times \boldsymbol{Z}$ hat die Gleichung $ax^2 + bxy + cy^2 = n$ mit gegebenen $a,b,c,n \in \boldsymbol{Z}$?

Oder, damit verwandt:

Für welche $n \in \boldsymbol{Z}$ ist die Gleichung $ax^2 + bxy + cy^2 = n$ mit vorgegebenen $a,b,c \in \boldsymbol{Z}$ ganzzahlig lösbar ?

Diese Fragestellungen führen zur allgemeinen Theorie der binären quadratischen Formen, von der hier ein grundlegender und klassischer Teil behandelt wird. Die Behandlung ist insofern elementar, als keine komplexe Analysis verwendet wird.

Eine binäre quadratische Form ist ein Polynom über $\boldsymbol{Z}$ in zwei Unbestimmten x,y über $\boldsymbol{Z}$ der Bauart $f(x,y) = ax^2 + bxy + cy^2$. Gibt es zu $n \in \boldsymbol{Z}$ $u,v \in \boldsymbol{Z}$ mit $f(u,v) = n$, so sagt man: „ Die Form f stellt die Zahl n dar. “ $D(f) = b^2 - 4ac$ heißt Diskriminante von f . Binäre quadratische Formen, von nun an einfach „Formen“ genannt, werden nach ihrer Diskriminante klassifiziert. Der Gang der Untersuchung ist im Groben dieser:

Im ersten Kapitel werden diverse Begriffe eingeführt und einige einfache, aber fundamentale Sachverhalte bewiesen. Zu gegebener Form f und gegebener ganzer Zahl n wird auf der Menge aller Lösungen von $f(x,y) = n$, d.h. auf $\{(u,v) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid f(u,v) = n\}$ eine Äquivalenzrelation definiert. Die Anzahl der Äquivalenzklassen wird als „Darstellungszahl R(n,f) von n durch f ” bezeichnet. Obiger Fragestellung zufolge ist man an Aussagen über R(n,f) interessiert. Ferner wird mittels einer Äquivalenztransformation ein Äquivalenzbegriff für Formen eingeführt. Es stellt sich heraus, daß für Formen f , g gilt: Aus $f \sim g$ folgt $D(f) = D(g)$ und $R(n,f) = R(n,g)$ für jedes $n \in \boldsymbol{Z}$. Das erste bedeutet, daß man jeder Äquivalenzklasse von Formen eine Diskriminante zuordnen und so von der „Klassenzahl h(D) von Formen der Diskriminante D“ sprechen kann. Das zweite besagt, daß man sich hinsichtlich der Bestimmung der Darstellungszahlen auf gewisse „typische Vertreter“ der Formenklassen beschränken kann.

Im zweiten Kapitel werden für jede Diskriminante D solche Vertreter der Formenklassen konstruiert und dadurch insbesondere die Endlichkeit der Klassenzahl h(D) gezeigt.

Durch gewisse der oben erwähnten Äquivalenztransformationen wird eine Form f in sich selbst übergeführt. Diese bilden die „Automorphismengruppe U(f) von f.“ Im dritten Kapitel wird die Struktur dieser Gruppe geklärt. Dabei wird die Lösungsmenge der „Pellschen Gleichung“ $x^2 - Dy^2 = 4$, D eine Diskriminante, bestimmt, was ein wichtiger Spezialfall der Ausgangsfragestellung und daher auch für sich selbst von Interesse ist.

Die Kapitel 4 und 8 bringen große Schritte in Richtung auf die Berechnung von R(n,f). Es werden nämlich explizite Ausdrücke hergeleitet für die „Gesamtdarstellungszahl G(n,D) von n durch Formen der Diskriminante D.“ Dabei handelt es sich um die Summe der Darstellungszahlen $R(n, f_i)$, wobei f_i ein Repräsentantensystem der Formenklassen zur Diskriminante D durchläuft. Insbesondere läßt sich G(n,D) effizient berechnen.

Als unmittelbare Folge ergeben sich im fünften Kapitel Sätze, die die durchschnittliche Gesamtdarstellungszahl einer natürlichen Zahl durch Formen der Diskriminante D, d.h. die Größe $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} G(n, D)$ als identisch erweisen mit dem Wert einer gewissen L-Reihe an der Stelle 1.

In diesem Sinne einen Schritt weiter gehen die Sätze im sechsten Kapitel, die Formeln liefern für die durchschnittliche Darstellungszahl einer natürlichen Zahl durch eine gegebene Form f , d.h. für $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} R(n, f)$.

Durch Verbindung der Resultate der Kapitel 5 und 6 wird schließlich in Kapitel 7 eine Formel für die Anzahl h(D) der Formenklassen der Diskriminante D erreicht.

Es ist festzuhalten, daß zwar einige schöne und tiefliegende Sachverhalte bewiesen werden (wie eben skizziert), die mit dem Ausgangsproblem in engem Zusammenhang stehen, aber das Problem selbst nicht in voller Allgemeinheit gelöst, d.h. keine Formel für R(n,f) erreicht wird. Man hat allerdings Formeln für G(n,D) und h(D), und das reicht aus, um R(n,f) in gewissen Fällen zu berechnen, denn für manche Diskriminanten D ist h(D) = 1.

Grundlage der vorliegenden Arbeit ist eine Vorlesung von G. Köhler, die dem §8 „Binäre quadratische Formen“ des Buches „Zetafunktionen und quadratische Körper“ von D.B. Zagier folgt. Diese Vorlesung wurde ausgearbeitet. Zagier beweist von den Sätzen der Kapitel 4, 5 und 6 nur jeweils den ersten, grundlegenden, denn er beschränkt sich dort auf fundamentale, d.h. quadratfreie Diskriminanten. Die Beweisargumente lassen sich aber verallgemeinern, und Zagier gibt in Aufgaben dazu Anleitung. Diese Verallgemeinerungen wurden durchgeführt und ergaben die weiteren Sätze der genannten Kapitel. Der im achten Kapitel gegebene allgemeine Satz über die Gesamtdarstellungszahl läßt sich jedoch auf diesem Wege nicht erreichen. Sein Beweis wurde einem Zeitschriftenartikel entnommen und stark ausgearbeitet. Schließlich erfordern Formen mit quadratischer Diskriminante eine eigene, wenngleich einfache Behandlung und werden von Zagier deshalb nicht berücksichtigt. Dieser Randfall wurde überall ergänzt.

1 Grundlegendes über binäre quadratische Formen

Definition 1.1 *Ein Polynom* $f(x,y) = ax^2 + bxy + cy^2$ *in zwei Unbestimmten x,y mit ganzzahligen Koeffizienten a,b,c heißt* ***binäre quadratische Form*** *(kurz:* ***Form****).*

Allgemein ist eine „Form“ ein Polynom, bei dem alle Monome denselben Grad haben. Dieser gemeinsame Grad ist hier 2, daher „quadratische Form.“ Und die Anzahl der auftretenden Unbestimmten ist ebenfalls 2, daher „binäre quadratische Form.“

In Matrixschreibweise ist $ax^2 + bxy + cy^2 = (x \;\; y)\begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$.

Jeder Form entspricht daher umkehrbar eindeutig eine Matrix aus $\boldsymbol{Q}^{2\,x\,2}$ dieser Bauart. Ferner entspricht jeder Form ihr Koeffiziententripel aus $\boldsymbol{Z}^3$. Manchmal ist es günstiger, mit der zugehörigen Matrix oder dem zugehörigen Zahlentripel zu operieren anstatt mit der Form selbst. Der Zusammenhang wird ausgedrückt durch die Schreibweise

$$f(x,y) = ax^2 + bxy + cy^2 \cong (a\,,\,b\,,\,c) \cong \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \quad (\text{lies } \cong \text{ „entspricht“}) .$$

Ist $f \cong Q = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}$ eine Form und $L = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \in \boldsymbol{Z}^{2\,x\,2}$, so beschreibt $Q' = LQL^T$

$$= \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}\begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} a\alpha^2 + b\alpha\gamma + c\gamma^2 & a\alpha\beta + \frac{1}{2}b(\alpha\delta + \beta\gamma) + c\gamma\delta \\ a\alpha\beta + \frac{1}{2}b(\alpha\delta + \beta\gamma) + c\gamma\delta & a\beta^2 + b\beta\delta + c\delta^2 \end{pmatrix} =$$

$\begin{pmatrix} a' & \frac{1}{2}b' \\ \frac{1}{2}b' & c' \end{pmatrix}$ wieder eine Form f ’.

Dabei ist $f'(x\,,\,y) = (x\;\;y)Q'(x\;\;y)^T = (x\;\;y)LQL^T(x\;\;y)^T = ((x\;\;y)L)\,Q\,((x\;\;y)L)^T = f((x\;\;y)L) = f(\alpha x + \beta y\,,\,\gamma x + \delta y)$. Ferner gilt für die Koeffizienten a’, b’, c’ von f ’ :

$a' = a\alpha^2 + b\alpha\gamma + c\gamma^2 = f(\alpha,\gamma)$,

$b' = 2a\alpha\beta + b(\alpha\delta + \beta\gamma) + 2c\gamma\delta = f(\alpha,\gamma) + f(\beta,\delta) - f(\alpha - \beta\,,\,\gamma - \delta)$,

$c' = a\beta^2 + b\beta\delta + c\delta^2 = f(\beta,\delta)$.

Man schreibt statt $Q' = LQL^T$ auch einfach $f' = Lf$ und sagt: „Die Matrix L überführt die Form f in die Form f ’. “ oder: „Die Form f geht durch Transformation mit der Matrix L in die Form f ’ über.“

Definition 1.2 *Zwei Formen f und f ’ heißen* ***äquivalent*** *(in Zeichen:* $f \sim f'$*) , wenn eine Matrix* $L \in SL_2(\boldsymbol{Z})$ *existiert mit* $f' = Lf$.

Weil $SL_2(\boldsymbol{Z})$ bzgl. der Matrizenmultiplikation eine Gruppe ist, ist durch diese Erklärung in der Tat eine Äquivalenzrelation auf der Menge aller Formen gegeben. Äquivalente Formen verhalten sich hinsichtlich der Darstellung ganzer Zahlen gleich:

Proposition 1.3 *Es sei $f' = Lf$ mit $L \in SL_2(\boldsymbol{Z})$ und $n \in \boldsymbol{Z}$. Durch $(x\ y) \to (x\ y)L^{-1}$ ist eine Bijektion von $\{(x,y) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid f(x,y) = n\}$ auf $\{(x,y) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid f'(x,y) = n\}$ gegeben. Insbesondere ist die Anzahl der Lösungen von $f(x,y) = n$ gleich der Anzahl der Lösungen von $f'(x,y) = n$.*

Beweis: Es gelte $f \cong Q$ und $f' \cong Q'$ mit $Q, Q' \in \boldsymbol{Q}^{2 \times 2}$. Nach Voraussetzung ist $Q' = LQL^T$. $f(x,y) = n \Leftrightarrow (x\ y)Q(x\ y)^T = n \Leftrightarrow (x\ y)\,L^{-1}Q'(L^T)^{-1}\,(x\ y)^T = n \Leftrightarrow (x\ y)L^{-1}\,Q'\,((x\ y)L^{-1})^T = n \Leftrightarrow f'((x\ y)L^{-1}) = n$. Also ist durch die obige Zuordnungsvorschrift eine Abbildung von der Lösungsmenge von $f(x,y) = n$ in die Lösungsmenge von $f'(x,y) = n$ gegeben. Aus $(x\ y)L^{-1} = (x'\ y')L^{-1}$ folgt $(x\ y) = (x'\ y')$ durch Multiplikation mit L, somit ist die Abbildung injektiv. Sie ist auch surjektiv, denn aus $f'(x,y) = n$ folgt $f((x\ y)L) = n$ durch eine analoge Umformung wie eben. ♦

Definition 1.4 *Es sei f eine Form. Ein $A \in SL_2(\boldsymbol{Z})$ mit $Af = f$ heißt **Automorphismus** von f. Die Menge aller Automorphismen von f wird mit **U(f)** bezeichnet.*

U(f) ist ersichtlich eine Untergruppe von $SL_2(\boldsymbol{Z})$. Ist $A \in U(f)$ und $n \in \boldsymbol{Z}$, so ist durch $(x\ y) \to (x\ y)A^{-1}$ eine Bijektion der Lösungsmenge von $f(x,y) = n$ auf sich gegeben.

Definition 1.5 *Zwei Lösungen (x_1, y_1) und (x_2, y_2) von $f(x,y) = n$ heißen **äquivalent**, falls ein $A \in U(f)$ existiert mit $(x_2, y_2) = (x_1, y_1)A^{-1}$. Die Anzahl der so auf der Lösungsmenge von $f(x,y) = n$ entstehenden Äquivalenzklassen heißt **Darstellungszahl R(n,f)** der ganzen Zahl n durch die Form f.*

Weil U(f) eine Gruppe ist, ist durch diese Erklärung in der Tat eine Äquivalenzrelation auf der Lösungsmenge von $f(x,y) = n$ gegeben. Der Sinn dieser Begriffsbildung liegt darin, daß R(n,f), wie sich herausstellen wird, stets endlich ist, während die Anzahl der Lösungen von $f(x,y) = n$ unendlich sein kann.

Proposition 1.6 *Sind f und f' äquivalente Formen, so sind U(f) und U(f') konjugierte Untergruppen von $SL_2(\boldsymbol{Z})$, und es ist $R(n,f) = R(n,f')$ für jedes $n \in \boldsymbol{Z}$.*

Beweis: Es gelte $f \cong Q$, $f' \cong Q'$ und $Q' = LQL^T$ mit $L \in SL_2(\boldsymbol{Z})$. Dann gilt: $A \in U(f') \Leftrightarrow AQ'A^T = Q' \Leftrightarrow L^{-1}AQ'A^T(L^T)^{-1} = Q \Leftrightarrow (L^{-1}AL)\,Q\,(L^{-1}AL)^T = Q \Leftrightarrow L^{-1}AL \in U(f) \Leftrightarrow A \in L\,U(f)\,L^{-1}$. Also $U(f') = LU(f)L^{-1}$: U(f) und U(f') sind konjugiert.

Durch $(x\ y) \to (x\ y)L^{-1}$ ist eine Bijektion von der Lösungsmenge von $f(x,y) = n$ auf die Lösungsmenge von $f'(x,y) = n$ gegeben. Es seien (x_1, y_1) und (x_2, y_2) äquivalente Lösungen von $f(x,y) = n$, d.h. es gibt ein $A \in U(f)$ mit $(x_2, y_2) = (x_1, y_1)A^{-1}$. Wie

eben gezeigt wurde, ist $A = L^{-1}A'L$ für ein $A' \in U(f')$. $(x_2 , y_2) = (x_1 , y_1)L^{-1}(A')^{-1}L$ ist gleichwertig zu $(x_2 , y_2)L^{-1} = (x_1 , y_1)L^{-1}(A')^{-1}$, also sind $(x_1 , y_1)L^{-1}$ und $(x_2 , y_2)L^{-1}$ äquivalente Lösungen von $f'(x,y) = n$. Die Argumentation läßt sich umkehren, denn f und f ' sind austauschbar. Es folgt $R(n,f) = R(n,f')$. ♦

Man kann also jeder Äquivalenzklasse von Formen die allen ihren Elementen gemeinsame Darstellungszahl von n zuordnen. Da man an R(n,f) interessiert ist, ist der oben eingeführte Äquivalenzbegriff für Formen somit sachgemäß.

Definition 1.7 *Zu einer Form* $f \cong (a , b , c) \cong Q \in \boldsymbol{Q}^{2 x 2}$ *heißt die ganze Zahl* $\boldsymbol{D(f)} = b^2 - 4ac = -4\,Det(Q)$ *die* ***Diskriminante*** *von f .*

Offenbar ist $D(f) \equiv 0 \bmod 4$ oder $D(f) \equiv 1 \bmod 4$. Äquivalente Formen haben wegen $Det(LQL^T) = Det(Q)$ für $L \in SL_2(\boldsymbol{Z})$ dieselbe Diskriminante. Man kann also jeder Äquivalenzklasse von Formen die allen ihren Elementen gemeinsame Diskriminante zuordnen und so von der „Anzahl der Formenklassen der Diskriminante D“ sprechen.

Definition 1.8 *Es sei* $D \in \boldsymbol{Z}$ *mit* $D \equiv 0 \bmod 4$ *bzw.* $D \equiv 1 \bmod 4$. $f_D(x,y) = x^2 - \frac{D}{4} y^2$ *bzw.* $f_D(x,y) = x^2 + xy + \frac{1-D}{4} y^2$ *heißt die* ***Grundform*** *zur Diskriminante D .*

Offenbar ist $f_D(x,y)$ eine Form der Diskriminante D. Als Diskriminanten von Formen treten also genau die ganzen Zahlen auf, die $\equiv 0 \bmod 4$ oder $\equiv 1 \bmod 4$ sind.

Es sei $f(x,y) = ax^2 + bxy + cy^2$ eine Form und $t = ggT(a,b,c)$ der größte gemeinsame Teiler der Koeffizienten von f . Offenbar teilt t jede von f dargestellte ganze Zahl, und t^2 teilt die Diskriminante von f. Ist $f' = Lf$ mit $L \in \boldsymbol{Z}^{2 x 2}$, so teilt t alle Koeffizienten von f ', denn diese lassen sich durch f ausdrücken. Äquivalente Formen haben daher denselben Koeffizienten-ggT, und man kann jeder Äquivalenzklasse von Formen diesen allen ihren Elementen gemeinsamen ggT zuordnen.

Definition 1.9 *Eine Form mit teilerfremden Koeffizienten heißt* ***primitiv****.*

Zu jeder Diskriminante gibt es eine primitive Form, etwa die Grundform.

Lemma 1.10 *Es sei D eine Diskriminante und* $t \in \boldsymbol{N}$ *mit* $t^2 \mid D$ *. Es gibt ebensoviele Formenklassen mit Diskriminante D und Koeffizienten-ggT t wie Formenklassen mit Diskriminante* D / t^2 *und Koeffizienten-ggT 1.*

Beweis: Ist $f(x,y) = ax^2 + bxy + cy^2$ eine Form der Diskriminante D mit Koeffizienten-ggT t, so ist $\frac{1}{t} f(x,y) = \frac{a}{t} x^2 + \frac{b}{t} xy + \frac{c}{t} y^2$ eine primitive Form der Diskriminante $\frac{D}{t^2}$.

Genau dann ist f ' zu f äquivalent, wenn f '/ t äquivalent ist zu f / t , denn $Q' = LQL^T$ ist gleichwertig mit $\frac{1}{t}Q' = L\frac{1}{t}QL^T$. Durch Klasse von f $\rightarrow$ Klasse von $\frac{1}{t}$f ist somit eine injektive Abbildung gegeben von der Menge aller Formenklassen mit Diskriminante D und Koeffizienten-ggT t in die Menge aller Formenklassen mit Diskriminante $\frac{D}{t^2}$ und Koeffizienten-ggT 1. Die Abbildung ist auch surjektiv, denn ist g eine primitive Form der Diskriminante $\frac{D}{t^2}$, so ist tg eine Form der Diskriminante D mit Koeffizienten-ggT t . ♦

Man beachte, daß in der Situation des Lemmas D / t^2 keine Diskriminante zu sein braucht, denn diese Zahl kann durchaus $\equiv 2 \bmod 4$ oder $\equiv 3 \bmod 4$ sein. Genau dann gibt es keine Form der Diskriminante D mit Koeffizienten-ggT t.

Definition 1.11 *Eine Form f heißt* ***positiv definit*** *bzw.* ***negativ definit*** *, wenn für alle* $(x,y) \in \boldsymbol{Z} \times \boldsymbol{Z}$, $(x,y) \neq (0,0)$, *gilt* $f(x,y) > 0$ *bzw.* $f(x,y) < 0$.

Proposition 1.12 *Eine Form mit negativer Diskriminante ist entweder positiv definit oder negativ definit, je nachdem, ob der erste Koeffizient positiv oder negativ ist. Und eine Formenklasse negativer Diskriminante besteht entweder nur aus positiv definiten oder nur aus negativ definiten Formen.*

Beweis: Sei $f(x,y) = ax^2 + bxy + cy^2$ eine Form mit $D(f) = b^2 - 4ac < 0$. Es ist $a \neq 0$.

Sei $a > 0$. Dann ist $f(x,y) = ax^2 + bxy + cy^2 = \left(\sqrt{a}\,x + \frac{b}{2\sqrt{a}}y\right)^2 - \frac{b^2y^2}{4a} + \frac{4acy^2}{4a} = \left(\sqrt{a}\,x + \frac{b}{2\sqrt{a}}y\right)^2 - \frac{D(f)}{4a}y^2 > 0$ für alle $x,y \in \boldsymbol{Z}$ mit $(x,y) \neq (0,0)$. f ist positiv definit.

Sei $a < 0$. In diesem Fall ist $f(x,y) = -\left(\sqrt{-a}\,x - \frac{b}{2\sqrt{-a}}y\right)^2 + \frac{b^2y^2}{-4a} + \frac{4acy^2}{4a} = -\left(\sqrt{-a}\,x - \frac{b}{2\sqrt{-a}}y\right)^2 - \frac{D(f)}{4a}y^2 < 0$ für alle $x,y \in \boldsymbol{Z}$, $(x,y) \neq (0,0)$. f ist negativ definit

Sei $f' = a'x^2 + b'xy + c'y^2 = Lf$ mit $L = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \in SL_2(\boldsymbol{Z})$. Dann ist $aa' = a\,f(\alpha,\gamma) = a(a\alpha^2 + b\alpha\gamma + c\gamma^2) = (a\alpha)^2 + b\gamma(a\alpha) + ac\gamma^2 = \left(a\alpha + \frac{1}{2}b\gamma\right)^2 - \frac{1}{4}D(f)\gamma^2 > 0$, d.h. a und a' sind entweder beide positiv oder beide negativ. Also sind f und f ' entweder beide positiv definit oder beide negativ definit. ♦

Man sieht leicht, daß durch Klasse von f $\rightarrow$ Klasse von $-f$ eine Bijektion gegeben ist von der Menge aller positiv definiten auf die Menge aller negativ definiten Formenklassen einer vorgegebenen negativen Diskriminante D. Dabei werden primitive Klassen auf

primitive Klassen abgebildet. Es gibt also ebensoviele (primitive) positiv definite wie (primitive) negativ definite Formenklassen der Diskriminante D.

Definition 1.13 *Es sei D eine positive bzw. negative Diskriminante. Die Anzahl der primitiven bzw. der primitiven und positiv definiten Formenklassen der Diskriminante D heißt* ***Klassenzahl h(D)*** *von D .*

Da es zu jeder Diskriminante D eine primitive bzw. primitive und positiv definite Form gibt (beispielsweise die Grundform), ist h(D) > 0 . h(D) wird sich als endlich herausstellen. h(0) bleibt zunächst undefiniert. Wenn man naheliegenderweise h(D) = 0 setzt für D $\in$ $\boldsymbol{Z}$ D keine Diskriminante, so erhält man unter Verwendung von Lemma 1.10, falls D $\neq$ 0 eine Diskriminante ist:

Gesamtzahl der Formenklassen zur Diskriminante D = $\sum_{t\in N,\, t^2|D}$ Anzahl der Formenklassen mit Koeffizienten-ggT t zu D = $\sum_{t^2|D}$ Anzahl der primitiven Formenklassen zu $\frac{D}{t^2}$ = $\sum_{t^2|D} h(\frac{D}{t^2})$ für D > 0 bzw. $2\sum_{t^2|D} h(\frac{D}{t^2})$ für D < 0 .

Damit ist die Gesamtklassenzahl zurückgeführt auf die primitive Klassenzahl h(D).

Definition 1.14 *Es sei D $\neq$ 0 eine Diskriminante und n $\neq$ 0 eine ganze Zahl. Die Menge* $\{f_1 , f_2 , \ldots , f_{h(D)}\}$ *sei für D > 0 ein Repräsentantensystem der primitiven Formenklassen zur Diskriminante D, für D < 0 und n > 0 ein Repräsentantensystem der primitiven und positiv definiten Formenklassen zur Diskriminante D, für D < 0 und n < 0 ein Repräsentantensystem der primitiven und negativ definiten Formenklassen zur Diskriminante D. Dann heißt* ***GP(n,D)*** = $\sum_{i=1}^{h(D)} R(n , f_i)$ *die* ***Gesamtdarstellungszahl*** *von n durch primitive Formen der Diskriminante D .*

Da für D < 0 die negativ definiten Formen ein positives n und die positiv definiten Formen ein negatives n nicht darstellen können, ist GP(n,D) einfach die Summe der Darstellungszahlen von n durch primitive Formenklassen der Diskriminante D. GP(0,D) und GP(n,0) bleiben zunächst undefiniert.

Es bezeichne G(n,D) die Gesamtdarstellungszahl von n durch beliebige Formen der Diskriminante D, d.h. die Summe der Darstellungszahlen von n durch beliebige Formenklassen der Diskriminante D. G(n,D) läßt sich auf GP(n,D) zurückspielen durch

$$G(n,D) = \sum_{t\in N,\, t^2|D} GP(\frac{n}{t} , \frac{D}{t^2}) .$$

Erstens nämlich korrespondieren nach Lemma 1.10 die Formenklassen mit Koeffizienten-ggT t zur Diskriminante D durch Klasse von f $\rightarrow$ Klasse von f / t bijektiv mit den primitiven Formenklassen zur Diskriminante D / t^2 , und zweitens ist in dieser Situation

$R(n,f) = R(n / t , f / t)$, denn offenbar besitzen f und f / t dieselben Automorphismen, und $f(x,y) = n$ und $f(x,y) / t = n / t$ haben dieselben Lösungen.

Es sind jetzt alle im Rahmen dieser Untersuchung wichtigen Begriffe eingeführt. Man beachte, daß bisher weder für die Klassenzahl h(D) noch für die Darstellungszahl R(n,f) die Endlichkeit gewährleistet ist. Davon werden aber die angegebenen Identitäten bei geeigneter Interpretation nicht berührt. Weiterhin war überall zu sehen, daß man sich bei der Untersuchung von Formen auf primitive Formen beschränken kann.

Lemma 1.15 *Genau die Formen, deren Diskriminante eine Quadratzahl ist, zerfallen im Polynomring $\boldsymbol{Z}[x,y]$ in Linearfaktoren.*

Beweis: Es sei $f(x,y) = ax^2 + bxy + cy^2$ eine Form der Diskriminante m^2 , $m \in \boldsymbol{N}_0$. Ist $a = 0$, so zerfällt $f(x,y) = (bx + cy)y$ offenbar in Linearfaktoren. Sei $a \neq 0$. Dann ist $f(x,y) = \frac{1}{a}(ax + \frac{b+m}{2}y)(ax + \frac{b-m}{2}y)$. Es ist $b^2 \equiv m^2 \bmod 4$, also $b \equiv m \bmod 2$, so daß $\frac{b+m}{2}$ und $\frac{b-m}{2}$ ganze Zahlen sind. Ihr Produkt ist wegen $ac = \frac{b^2 - m^2}{4}$ ein Vielfaches von a . Daher gibt es ganze Zahlen r,s so, daß r ein Teiler von $\frac{b+m}{2}$, s ein Teiler von $\frac{b-m}{2}$ und $a = rs$ ist. Also ist $f(x,y) = (sx + \frac{b+m}{2r}y)(rx + \frac{b-m}{2s}y)$ in $\boldsymbol{Z}[x,y]$.

Es sei nun umgekehrt f(x,y) eine Form, die in $\boldsymbol{Z}[x,y]$ in Linearfaktoren zerfällt, d.h. $f(x,y) = (rx + sy + t)(ux + vy + w)$ mit $r,s,t,u,v,w \in \boldsymbol{Z}$. Es ist $tw = 0$,d.h. $t = 0$ oder $w = 0$. Sei etwa $w = 0$. Daraus folgt weiter $tu = tv = 0$, also $t = 0$ oder $u = v = 0$. Im letzteren Fall ist f die Nullform und hat somit die Diskriminante 0. Im ersteren Fall ist $f(x,y) = (rx + sy)(ux + vy) = rux^2 + (rv + su)xy + svy^2$ und hat also die Diskriminante $(rv + su)^2 - 4rsuv = (rv - su)^2$. In jedem Fall ist D(f) ein Quadrat. ♦

Formen mit quadratischer Diskriminante spielen also eine Sonderrolle, was im weiteren Verlauf entsprechende Fallunterscheidungen erfordern wird.

2 Die Endlichkeit der Klassenzahl

Satz 2.1 *Es sei D eine Diskriminante, aber keine Quadratzahl. Dann gibt es nur endlich viele Äquivalenzklassen von Formen der Diskriminante D.*

Beweis: $f(x,y) = ax^2 + bxy + cy^2$ sei eine Form der Diskriminante D. f stellt von Null verschiedene ganze Zahlen dar. Unter allen diesen Zahlen sei a' eine dem Betrage nach minimale. Seien $\alpha,\gamma \in \boldsymbol{Z}$ mit $a' = f(\alpha,\gamma)$. $f(\frac{\alpha}{ggT(\alpha,\gamma)} , \frac{\gamma}{ggT(\alpha,\gamma)}) = \frac{a'}{(ggT(\alpha,\gamma))^2}$, also $ggT(\alpha,\gamma) = 1$ wegen der Minimalität von a'. Es gibt $\beta,\delta \in \boldsymbol{Z}$ mit $\alpha\delta - \beta\gamma = 1$.

Sei $f'(x,y) = a'x^2 + b'xy + c'y^2$ definiert durch $f' = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} f$. f ' ist zu f äquivalent und sein erster Koeffizient $a' = f(\alpha,\gamma)$ ist das a' von eben. Man dividiert b' mit Rest durch 2a' und erhält $b' = 2a'n + b''$ mit $n,b'' \in \boldsymbol{Z}$ und $-|a'| < b'' \leq |a'|$.

Sei $f''(x,y) = a''x^2 + b''xy + c''y^2$ definiert durch $f'' = \begin{pmatrix} 1 & 0 \\ -n & 1 \end{pmatrix} f'$. f '' ist zu f äquivalent und sein zweiter Koeffizient $b'' = 2a'(-n) + b'$ ist das b'' von eben. $a'' = f''(1,0) = a'$, also $|b''| \leq |a''|$. $c'' = f''(-n,1)$ wird von f ' dargestellt, also auch von f , und ist nicht Null, denn $D = (b'')^2 - 4a''c''$ ist kein Quadrat. Wegen der Minimalität von a' ist somit $|a'| \leq |c''|$ und insgesamt gilt $|b''| \leq |a''| \leq |c''|$.

Resultat: Jede Form der Diskriminante D ist äquivalent zu einer Form $f \cong (a,b,c)$ mit $|b| \leq |a| \leq |c|$. Dann ist $|D| = |b^2 - 4ac| \geq 4|a||c| - |b|^2 \geq 4|a||a| - |a|^2 = 3|a|^2$, also $|a| \leq \sqrt{\frac{|D|}{3}}$. Es existieren offenbar nur endlich viele Tripel (a,b,c) ganzer Zahlen, die den Bedingungen $|a| \leq \sqrt{\frac{|D|}{3}}$, $|b| \leq |a|$, $c = \frac{b^2 - D}{4a}$ genügen, also nur endlich viele Formen der Diskriminante D, deren Koeffizienten die obigen Ungleichungen erfüllen. ♦

Satz 2.2 *Es sei $m \in \boldsymbol{N}$. Dann gibt es genau m Äquivalenzklassen von Formen der Diskriminante m^2. Sie werden repräsentiert durch die Formen $ax^2 + mxy$ mit $a \in \{1, \ldots, m\}$.*

Beweis: Es sei f eine primitive Form der Diskriminante m^2 . Nach 1.15 zerfällt f in lineare Faktoren: $f(x,y) = (rx + sy)(ux + vy)$ mit $r,s,u,v \in \boldsymbol{Z}$, $ggT(r,s) = ggT(u,v) = 1$. Dann ist $m^2 = (rv - su)^2$, also $|rv - su| = m$. Es gibt $\alpha,\gamma \in \boldsymbol{Z}$ mit $\alpha u + \gamma v = -1$.

Es sei $f' = \begin{pmatrix} \alpha & \gamma \\ v & -u \end{pmatrix} f$. $f'(x,y) = f(\alpha x + vy , \gamma x - uy) = (r(\alpha x + vy) + s(\gamma x - uy))$ $(u(\alpha x + vy) + v(\gamma x - uy)) = ((r\alpha + s\gamma)x + (rv - su)y)(-x) = -(r\alpha + s\gamma)x^2 - (rv - su)xy$.

f ' ist also von der Bauart $ax^2 \pm mxy$. Dabei ist $ggT(a,m) = 1$, denn äquivalente Formen haben denselben Koeffizienten-ggT. Es sei $f'(x,y) = ax^2 - mxy$. Es gibt $\alpha,\gamma \in \boldsymbol{Z}$ mit

$\alpha a - \gamma m = 1$, und setzt man $f'' = \begin{pmatrix} \alpha & \gamma \\ m & a \end{pmatrix} f'$, so ist $f''(x,y) = f'(\alpha x + my , \gamma x + ay)$ $= (\alpha x + my)(a(\alpha x + my) - m(\gamma x + ay)) = (\alpha x + my)x = \alpha x^2 + mxy$ mit $ggT(\alpha,m) = 1$. Man darf also annehmen, daß $f'(x,y) = ax^2 + mxy$ ist.

Es gibt ein $\gamma \in \boldsymbol{Z}$ so, daß $a + \gamma m \in \{1, ... ,m\}$ ist. Setzt man $g = \begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} f'$, so ist $g(x,y) = f'(x, \gamma x + y) = (ax + m(\gamma x + y))x = (a + \gamma m)x^2 + mxy$. $ggT(a + \gamma m, m) = 1$.

Resultat: Jede primitive Form der Diskriminante m^2 ist äquivalent zu einer der Formen $ax^2 + mxy$ mit $a \in \{1, ... ,m\}$ und $ggT(a,m) = 1$. Es gibt, wenn φ die Eulersche Phi-Funktion bezeichnet, genau $\varphi(m)$ derartige Formen. Diese sind paarweise inäquivalent:

Sei $f(x,y) = ax^2 + mxy$ äquivalent zu $f'(x,y) = a'x^2 + mxy$ durch $f' = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} f$.

Dann ist $a' = a\alpha^2 + m\alpha\gamma$, $m = 2a\alpha\beta + m(\alpha\delta + \beta\gamma)$, $0 = a\beta^2 + m\beta\delta$ und $\alpha\delta - \beta\gamma = 1$. Wäre $\beta \neq 0$, so wäre $a\beta + m\delta = 0$, also $a = -\frac{m\delta}{\beta}$ und $m = -2m\alpha\delta + m\alpha\delta + m\beta\gamma$ $= -m$, was falsch ist. Folglich ist $\beta = 0$, also $\alpha = \delta = \pm 1$ und daher $a' = a \pm m\gamma$, d.h. $a' \equiv a \bmod m$. Das war zu zeigen.

Es gibt somit genau $\varphi(m)$ primitive Formenklassen der Diskriminante m^2 : $h(m^2) = \varphi(m)$ (siehe Definition 1.13). Die Gesamtzahl der Formenklassen zur Diskriminante m^2 ist $\sum_{t \in N, t^2 \mid m^2} h(\frac{m^2}{t^2}) = \sum_{t \mid m} h(\left(\frac{m}{t}\right)^2) = \sum_{t \mid m} \varphi(\frac{m}{t}) = \sum_{t \mid m} \varphi(t) = m$ nach einem bekannten Satz über die Phi-Funktion. Es gibt also genau m Formenklassen der Diskriminante m^2. Sie werden repräsentiert durch die Formen $ax^2 + mxy$ mit $a \in \{1, ... , m\}$, denn diese sind nach der eben durchgeführten Überlegung paarweise inäquivalent. ♦

Satz 2.3 *Es gibt unendlich viele Äquivalenzklassen von Formen der Diskriminante 0. Sie werden repräsentiert durch die Formen* ax^2 , $a \in \boldsymbol{Z}$.

Beweis: Für jedes $a \in \boldsymbol{Z}$ ist ax^2 eine Form der Diskriminante 0. Weil äquivalente Formen denselben Koeffizienten-ggT haben, sind ax^2 und $a'x^2$ für $a' \neq \pm a$ nicht äquivalent. Für $a \in \boldsymbol{N}$ stellt ax^2 positive Zahlen dar, $-ax^2$ hingegen nicht, und deshalb sind ax^2 und $-ax^2$ nicht äquivalent.

Es sei f eine primitive Form der Diskriminante 0. Nach 1.15 ist $f(x,y) = (rx + sy)(ux + vy)$ mit $r,s,u,v \in \boldsymbol{Z}$, $ggT(r,s) = ggT(u,v) = 1$. Es ist $(rv - su)^2 = 0$, also $rv = su$. r ist ein Teiler von su, also ein Teiler von u, und u teilt umgekehrt rv, also r. Somit ist $u = \pm r$ und ebenso $v = \pm s$. Es gilt entweder $u = r$, $v = s$ oder $u = -r$, $v = -s$. Folglich ist $f(x,y) = \pm(rx + sy)^2$.

Es gibt $k,l \in \boldsymbol{Z}$ mit $kr + ls = 1$. Sei $f' \cong (a',b',c')$ gegeben durch $f' = \begin{pmatrix} k+s & l-r \\ -s & r \end{pmatrix} f$.

Dann ist f ' zu f äquivalent und es gilt $a' = f(k + s , l - r) = \pm(kr + rs + ls - rs)^2 = \pm 1$,

$c' = f(-s, r) = 0$ und $b' = 0$ wegen $D(f') = 0$. Somit ist f entweder zu x^2 oder zu $-x^2$ äquivalent, so daß es genau zwei primitive Formenklassen der Diskriminante 0 gibt. Mithilfe von Lemma 1.10 folgt weiter, daß für jedes $a \in \boldsymbol{N}$ genau zwei Formenklassen mit Koeffizienten-ggT a zur Diskriminante 0 existieren. Diese werden offenbar durch ax^2 und $-ax^2$ repräsentiert. Nur genau die Nullform hat Koeffizienten-ggT 0. ♦

Wie das obige Repräsentantensystem zeigt, stellt eine Form der Diskriminante 0 nur nicht negative oder nur nicht positive Zahlen dar, ist also „positiv semidefinit“ oder „negativ semidefinit“. Die Situation ist ähnlich wie bei negativen Diskriminanten, und daher ist es sinnvoll, die Klassenzahl h(0) zu definieren als die Anzahl der primitiven, positiv semidefiniten Formenklassen der Diskriminante 0. Damit ist $h(0) = 1$.

Die Sätze 2.2 und 2.3 liefern für quadratische Diskriminanten die Klassenzahl und ein Repräsentantensystem der Klassen. Diese Aufgaben sind für nicht quadratische Diskriminanten viel schwieriger zu lösen. Kapitel 7 bringt als Abschluß tiefer Untersuchungen Formeln für die Klassenzahl in diesem Fall. Die Reduktionstheorie quadratischer Formen, die hier nicht behandelt wird, liefert effiziente Algorithmen zur Gewinnung eines Repräsentantensystems der Formenklassen einer gegebenen Diskriminante. Die Formen, die den in Satz 2.1 gefundenen Bedingungen genügen, sind zu festem D zwar einfach aufzustellen, bilden jedoch im allgemeinen kein solches System, denn sie sind nicht notwendig paarweise inäquivalent.

3 Die Automorphismengruppe

Es soll die Automorphismengruppe U(f) einer gegebenen Form f ermittelt werden. Da f und $t\,f$ für jedes $t \in \boldsymbol{N}$ dieselben Automorphismen haben, reicht es aus, primitive Formen zu betrachten.

Satz 3.1 *Es sei* $f(x,y) = ax^2 + bxy + cy^2$ *eine primitive Form der Diskriminante D. Dann ist durch* $(t,u) \to \begin{pmatrix} \frac{1}{2}(t-bu) & au \\ -cu & \frac{1}{2}(t+bu) \end{pmatrix}$ *eine Bijektion von der Lösungsmenge der sog. Pellschen Gleichung* $t^2 - Du^2 = 4$ *auf die Automorphismengruppe von f gegeben. Mit der Vereinbarung* $(t_1, u_1) * (t_2, u_2) = \left(\frac{t_1t_2 + Du_1u_2}{2}, \frac{t_1u_2 + t_2u_1}{2} \right)$ *wird diese Bijektion ein Gruppenisomorphismus. Für die Struktur von U(f) gilt:*

$U(f) \cong \boldsymbol{Z}/2\boldsymbol{Z}$ *für* $D < -4$, $U(f) \cong \boldsymbol{Z}/4\boldsymbol{Z}$ *für* $D = -4$, $U(f) \cong \boldsymbol{Z}/6\boldsymbol{Z}$ *für* $D = -3$, $U(f) \cong (\boldsymbol{Z}/2\boldsymbol{Z}) \times \boldsymbol{Z}$ *für* $D = 0$, $U(f) \cong \boldsymbol{Z}/2\boldsymbol{Z}$ *für quadratisches* $D > 0$ *und* $U(f) \cong (\boldsymbol{Z}/2\boldsymbol{Z}) \times \boldsymbol{Z}$ *für nichtquadratisches* $D > 0$.

Die Struktur von U(f) hängt also nur von der Diskriminante von f ab, und U(f) ist endlich für negative und unendlich für positive nichtquadratische Diskriminanten.

Beweis: Es sei zunächst D kein Quadrat.

I.) Es sei $A = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$ ein Automorphismus von f. Dann ist $a = a\alpha^2 + b\alpha\gamma + c\gamma^2$, $b = 2a\alpha\beta + b(\alpha\delta + \beta\gamma) + 2c\gamma\delta$, $c = a\beta^2 + b\beta\delta + c\delta^2$ und $\alpha\delta - \beta\gamma = 1$.

$2(a\alpha\beta + b\beta\gamma + c\gamma\delta) = 2b\beta\gamma + b - b(\alpha\delta + \beta\gamma) = 2b\beta\gamma + b(\alpha\delta - \beta\gamma) - b(\alpha\delta + \beta\gamma) = 0$ also $a\alpha\beta + b\beta\gamma + c\gamma\delta = 0$.

$\beta a = \beta(a\alpha^2 + b\alpha\gamma + c\gamma^2) = \alpha(a\alpha\beta + b\beta\gamma) + c\beta\gamma^2 = -c\alpha\gamma\delta + c\beta\gamma^2 = c\gamma(\beta\gamma - \alpha\delta) = -c\gamma$, also $\frac{\gamma}{a} = -\frac{\beta}{c}$. Man beachte, daß weder a noch c Null sein können, denn D ist kein Quadrat. $(\alpha - \delta)c = \alpha(a\beta^2 + b\beta\delta + c\delta^2) - \delta c = \beta(a\alpha\beta + b\alpha\delta) + c\alpha\delta^2 - c\delta = \beta(a\alpha\beta + b\alpha\delta) + c\beta\gamma\delta = \beta(a\alpha\beta + c\gamma\delta) + b\alpha\beta\delta = -b\beta^2\gamma + b\alpha\beta\delta = b\beta$, und somit $-\frac{\beta}{c} = \frac{\delta - \alpha}{b} = \frac{\gamma}{a}$ für $b \neq 0$ und $\alpha = \delta$, $\frac{\gamma}{a} = -\frac{\beta}{c}$ für $b = 0$. In jedem Fall ist $\frac{\gamma}{a} \in \boldsymbol{Z}$, denn f ist primitiv, d.h. es ist $ggT(a,b,c) = 1$. Setzt man $u = \frac{\gamma}{a}$ und $t = \alpha + \delta$, so ist $\begin{pmatrix} \frac{1}{2}(t-bu) & au \\ -cu & \frac{1}{2}(t+bu) \end{pmatrix} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$ und $\frac{1}{4}(t^2 - b^2u^2) + acu^2 = \alpha\delta - \beta\gamma = 1$ also $t^2 - Du^2 = 4$. A ist von der im Satz angegebenen Form.

II.) Es seien nun umgekehrt $t,u \in \boldsymbol{Z}$ mit $t^2 - Du^2 = 4$. Setze $\alpha = \frac{1}{2}(t - bu)$, $\beta = -cu$, $\gamma = au$, $\delta = \frac{1}{2}(t + bu)$. $\beta,\gamma \in \boldsymbol{Z}$ ist klar. Ist $D \equiv 0 \bmod 4$, so sind b und t gerade und daher $\alpha,\delta \in \boldsymbol{Z}$. Ist $D \equiv 1 \bmod 4$, so ist b ungerade und $t \equiv u \bmod 2$, also wieder $\alpha,\delta \in \boldsymbol{Z}$. $\alpha\delta - \beta\gamma = \frac{1}{4}(t^2 - b^2u^2) + acu^2 = \frac{1}{4}(t^2 - Du^2) = 1$, folglich ist $\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \in SL_2(\boldsymbol{Z})$.

$a\alpha^2 + b\alpha\gamma + c\gamma^2 = \frac{1}{4}a(t^2 - 2btu + b^2u^2) + \frac{1}{2}b(atu - abu^2) + ca^2u^2 = \frac{1}{4}at^2 - \frac{1}{4}ab^2u^2 + ca^2u^2 = \frac{1}{4}a(t^2 - Du^2) = a$.

$2a\alpha\beta + b(\alpha\delta + \beta\gamma) + 2c\gamma\delta = a(bcu^2 - ctu) + b(\frac{1}{4}(t^2 - b^2u^2) - acu^2) + c(atu + abu^2)$ $= abcu^2 + \frac{1}{4}b(t^2 - b^2u^2) [= \frac{1}{4}b(t^2 - Du^2) [= b$.

$a\beta^2 + b\beta\delta + c\delta^2 = ac^2u^2 - bcu\frac{1}{2}(t + bu) + \frac{1}{4}c(t^2 + 2btu + b^2u^2) = ac^2u^2 - \frac{1}{4}cb^2u^2 + \frac{1}{4}ct^2 = \frac{1}{4}c(t^2 - Du^2) = c$.

$\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(t - bu) & au \\ -cu & \frac{1}{2}(t + bu) \end{pmatrix}$ ist also ein Automorphismus von f. Zu verschiedenen Lösungen der Pellschen Gleichung gehören verschiedene Automorphismen, denn $u = \frac{\gamma}{a}$ und $t = 2\alpha + bu = 2\alpha + \frac{b}{a}\gamma$. Somit ist durch die im Satz angegebene Zuordnung in der Tat eine Bijektion von der Lösungsmenge der Pellschen Gleichung auf U(f) gegeben.

III.) Es seien (t_1,u_1) und (t_2,u_2) Lösungen von $t^2 - Du^2 = 4$. Die zugehörigen Automorphismen von f sind $\begin{pmatrix} \frac{1}{2}(t_1 - bu_1) & au_1 \\ -cu_1 & \frac{1}{2}(t_1 + bu_1) \end{pmatrix}$ und $\begin{pmatrix} \frac{1}{2}(t_2 - bu_2) & au_2 \\ -cu_2 & \frac{1}{2}(t_2 + bu_2) \end{pmatrix}$. Deren Produkt $\begin{pmatrix} \frac{1}{4}(t_1t_2 + Du_1u_2 - bt_1u_2 - bt_2u_1) & \frac{1}{2}a(t_1u_2 + t_2u_1) \\ -\frac{1}{2}c(t_1u_2 + t_2u_1) & \frac{1}{4}(t_1t_2 + Du_1u_2 + bt_1u_2 + bt_2u_1) \end{pmatrix}$ ist auch aus U(f), also gehört dazu eine Lösung $(t,u) \in \boldsymbol{Z} \times \boldsymbol{Z}$ der Pellschen Gleichung. Andererseits aber korrespondiert mit dieser Matrix bzgl. der im Satz angegebenen Zuordnung das Paar $(\frac{t_1t_2 + Du_1u_2}{2}, \frac{t_1u_2 + t_2u_1}{2}) \in \boldsymbol{Q} \times \boldsymbol{Q}$. Weil durch die Zuordnungsvorschrift ersichtlich auch eine injektive Abbildung von $\boldsymbol{Q} \times \boldsymbol{Q}$ in $\boldsymbol{Q}^{2 \times 2}$ gegeben ist, ergibt sich daraus $(t,u) = (\frac{t_1t_2 + Du_1u_2}{2}, \frac{t_1u_2 + t_2u_1}{2})$. Durch $(t_1,u_1) * (t_2,u_2) = (\frac{t_1t_2 + Du_1u_2}{2}, \frac{t_1u_2 + t_2u_1}{2})$ ist also eine zweistellige Operation auf der Lösungsmenge der Pellschen Gleichung gegeben. Die Zuordnung ist bzgl. dieser Verknüpfung und der Matrizenmultiplikation homomorph und wird so zu einem Gruppenisomorphismus. U(f) ist abelsch, denn $*$ ist offenbar kommutativ. Da die Lösungsmenge von $t^2 - Du^2 = 4$ samt der Verknüpfung $*$ nur von D abhängt, haben alle primitiven Formen der Diskriminante D isomorphe Automorphismengruppen.

IV.) Um die Struktur von U(f) für die einzelnen Diskriminanten zu klären, soll nun die Lösungsmenge der Pellschen Gleichung untersucht werden. Da die auf dieser Menge erklärte Multiplikation wenig handlich ist, wird ein weiterer Gruppenisomorphismus eingeführt. Betrachte $\psi : \{(t,u) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid t^2 - Du^2 = 4\} \to \boldsymbol{C}$ durch $\psi((t,u)) = \frac{t+u\sqrt{D}}{2}$. Weil D kein Quadrat ist, ist $\sqrt{D}$ irrational. Daher ist ψ injektiv und durchweg ungleich 0.

$\psi((t_1,u_1) * (t_2,u_2)) = \psi((\frac{t_1t_2 + Du_1u_2}{2}, \frac{t_1u_2 + t_2u_1}{2})) = \frac{1}{4}(t_1t_2 + Du_1u_2 + (t_1u_2 + t_2u_1)\sqrt{D})$

$= \frac{t_1 + u_1\sqrt{D}}{2} \cdot \frac{t_2 + u_2\sqrt{D}}{2} = \psi((t_1,u_1)) \cdot \psi((t_2,u_2))$.

Somit ist ψ ein injektiver Homomorphismus von $(\{(t,u) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid t^2 - Du^2 = 4\}, *)$ in $(\boldsymbol{C}\backslash\{0\}, \cdot)$. Setze $E_D = \text{Bild } \psi = \{ \frac{t+u\sqrt{D}}{2} \mid t,u \in \boldsymbol{Z}, t^2 - Du^2 = 4\}$. $(E_D, \cdot)$ ist eine zu U(f) isomorphe Untergruppe von $(\boldsymbol{C}\backslash\{0\}, \cdot)$. Damit liegt U(f) in einer Gestalt vor, die für viele Zwecke günstiger ist: Statt ganzzahliger 2 x 2 - Matrizen werden komplexe Zahlen multipliziert.

V.) Es sei $D < -4$. Dann ist $-Du^2 > 4$ für $u \neq 0$, also hat $t^2 - Du^2 = 4$ nur die trivialen Lösungen (2,0) und (–2,0). $E_D = \{1, -1\}$ ist die Menge der zweiten Einheitswurzeln. Somit ist U(f) isomorph zu $\boldsymbol{Z} / 2\boldsymbol{Z}$.

Es sei $D = -4$. $t^2 + 4u^2 = 4$ hat genau die Lösungen (2,0), (–2,0), (0,1) und (0, –1). $E_D = \{1, -1, i, -i\}$ ist die Menge der vierten Einheitswurzeln im Komplexen, folglich ist U(f) isomorph zu $\boldsymbol{Z} / 4\boldsymbol{Z}$.

Es sei $D = -3$. $t^2 + 3u^2 = 4$ hat genau die Lösungen (2,0), (–2,0), (1,1), (1, –1), (–1,1) und (–1, –1). $E_D = \{1, -1, \frac{1+i\sqrt{3}}{2}, \frac{1-i\sqrt{3}}{2}, \frac{-1+i\sqrt{3}}{2}, \frac{-1-i\sqrt{3}}{2}\}$ ist die Menge der komplexen sechsten Einheitswurzeln, also ist U(f) isomorph zu $\boldsymbol{Z} / 6\boldsymbol{Z}$.

Es sei nun $D > 0$. Dann ist $\sqrt{D}$ reell und daher E_D eine Untergruppe von $(\boldsymbol{R}\backslash\{0\}, \cdot)$. Da die Pellsche Gleichung stets die trivialen Lösungen hat (dem entspricht, daß jede Form die Automorphismen $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ und $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ besitzt), ist $\{1, -1\}$ eine Untergruppe von E_D.

Auch $E_D^+ = \{\varepsilon \in E_D \mid \varepsilon > 0\}$ ist eine Untergruppe von E_D, und offenbar ist E_D das innere direkte Produkt dieser beiden Untergruppen: $E_D \cong \{1, -1\} \times E_D^+$. Es wäre nun denkbar, daß $E_D^+ = \{1\}$ ist, d.h. daß die Pellsche Gleichung nur die trivialen Lösungen besitzt. Daß dies nicht der Fall ist, wird erst in Kapitel 6 gezeigt werden und muß an dieser Stelle einfach zur Kenntnis genommen werden.

Es sei $\varepsilon = \frac{t+u\sqrt{D}}{2} \in E_D^+$. Annahme: $t \leq 0$. Dann ist $u > 0$. $\varepsilon^{-1} = \frac{2}{t+u\sqrt{D}} = \frac{2(t-u\sqrt{D})}{t^2 - Du^2} = \frac{t-u\sqrt{D}}{2} < 0$, also $\varepsilon^{-1} \notin E_D^+$: Widerspruch. Folglich ist $t > 0$.

Weiter gilt: $\varepsilon > 1 \Leftrightarrow \varepsilon > \varepsilon^{-1} \Leftrightarrow \frac{t+u\sqrt{D}}{2} > \frac{t-u\sqrt{D}}{2} \Leftrightarrow u > -u \Leftrightarrow u > 0$.

$t,u \in \boldsymbol{N}$ impliziert $\frac{t+u\sqrt{D}}{2} \geq \frac{1+\sqrt{5}}{2} > \frac{3}{2}$. Also: $\varepsilon > 1 \Leftrightarrow t,u > 0 \Leftrightarrow \varepsilon > \frac{3}{2}$.

Seien nun $\varepsilon_1, \varepsilon_2 \in E_D$ mit $1 < \varepsilon_1 < \varepsilon_2$. Dann ist $\frac{\varepsilon_2}{\varepsilon_1} > 1$ und $\frac{\varepsilon_2}{\varepsilon_1} \in E_D$, also $\frac{\varepsilon_2}{\varepsilon_1} > \frac{3}{2}$

Es folgt $\varepsilon_2 - \varepsilon_1 > \frac{1}{2}\varepsilon_1 > \frac{3}{4}$. Die Menge $\{\varepsilon \in E_D \mid \varepsilon > 1\}$ ist nicht leer, weil E_D^+ nicht trivial ist, und besitzt keinen Häufungspunkt in $\boldsymbol{R}$, weil irgend zwei verschiedene ihrer Elemente mindestens den Abstand $\frac{3}{4}$ haben. Sie hat also ein kleinstes Element ε_D. Dieses wird als **Grundeinheit** zur Diskriminante D bezeichnet. Ist $\varepsilon \in E_D$ mit $\varepsilon > 1$, so gibt es ein $n \in \boldsymbol{N}$ mit $\varepsilon_D^n \leq \varepsilon < \varepsilon_D^{n+1}$, also $1 \leq \frac{\varepsilon}{\varepsilon_D^n} < \varepsilon_D$, also $\varepsilon_D^n = \varepsilon$ wegen der Minimalität von ε_D. Daher ist $E_D^+ = \{\varepsilon_D^n \mid n \in \boldsymbol{Z}\}$ und folglich $E_D \cong \{1, -1\} \times \{\varepsilon_D^n \mid n \in \boldsymbol{Z}\} \cong (\boldsymbol{Z} / 2\boldsymbol{Z}) \times \boldsymbol{Z}$, also ist U(f) isomorph zu $(\boldsymbol{Z} / 2\boldsymbol{Z}) \times \boldsymbol{Z}$. Damit ist der Satz für nichtquadratische Diskriminanten (bis auf eine in Kapitel 6 zu schließende Lücke) bewiesen.

VI.) Nun sei $D = m^2$ mit $m \in \boldsymbol{N}$. Dann ist f nach Satz 2.2 äquivalent zu einer Form der Bauart $ax^2 + mxy$ mit $a \in \{1, \ldots, m\}$. Es sei $\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$ ein Automorphismus dieser Form. Dann ist $a = a\alpha^2 + m\alpha\gamma$, $m = 2a\alpha\beta + m(\alpha\delta + \beta\gamma)$, $0 = a\beta^2 + m\beta\delta$ und $1 = \alpha\delta - \beta\gamma$. Annahme: $\beta \neq 0$. Dann ist $a\beta + m\delta = 0$, also $a = -\frac{m\delta}{\beta}$. $m = 2a\alpha\beta + m(\alpha\delta + \beta\gamma) = -2m\alpha\delta + m\alpha\delta + m\beta\gamma = -m$: Widerspruch. Somit ist $\beta = 0$, also $\alpha = \delta = \pm 1$. Weiter folgt $a = a \pm m\gamma$, d.h. $\gamma = 0$. $ax^2 + mxy$ besitzt also nur die trivialen Automorphismen.

Da die Automorphismengruppen äquivalenter Formen isomorph sind, hat auch f nur diese Automorphismen. Andererseits hat die Pellsche Gleichung $t^2 - m^2u^2 = 4$ auch nur die trivialen Lösungen: $(t - mu)(t + mu) = 4$ impliziert $t \pm mu \in \{-4, -2, -1, 1, 2, 4\}$. Annahme: $t - mu = \pm 4$. Dann ist $t \equiv mu \bmod 4$, also $t + mu \equiv 0 \bmod 2$ und somit $(t - mu)(t + mu) \equiv 0 \bmod 8$: Widerspruch. Zu demselben Widerspruch führt die Annahme $t + mu = \pm 4$. Also $t + mu = t - mu = \pm 2$. In jedem Fall ist $2mu = 0$, also $u = 0$ und folglich $t = \pm 2$. Damit ist der Satz für positive quadratische Diskriminanten bewiesen.

VII.) Schließlich sei $D = 0$.

1. Fall: $a \neq 0$, $c \neq 0$. Dann gehen die Beweisteile I.), II.) und III.) glatt durch, so daß die Automorphismengruppe von f in der angegebenen Weise isomorph ist zur Lösungsmenge der Pellschen Gleichung. Die Struktur von U(f) ist dieselbe wie im nachfolgenden Fall, da f nach Satz 2.3 zu x^2 oder zu $-x^2$ äquivalent ist.

2. Fall: $c = 0$. Dann ist auch $b = 0$ und somit $f(x,y) = \pm x^2$. Es sei $A = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \in U(f)$. Dann ist $\pm 1 = \pm\alpha^2$, $0 = \pm 2\alpha\beta$, $0 = \pm\beta^2$ und $1 = \alpha\delta - \beta\gamma$, gleichwertig $\beta = 0$ und

$\alpha = \delta = \pm 1$, d.h. $U(f) = \{\pm\begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} \mid \gamma \in \boldsymbol{Z}\} \cong \{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\} \times \{\begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} \mid \gamma \in \boldsymbol{Z}\}$ $\cong (\boldsymbol{Z} / 2\boldsymbol{Z}) \times \boldsymbol{Z}$. Die Pellsche Gleichung $t^2 - 0u^2 = 4$ besitzt $\{(2,u) , (-2,u) \mid u \in \boldsymbol{Z}\}$ als Lösungsmenge. Es gilt $(2,u) \rightarrow \begin{pmatrix} 1 & \pm u \\ 0 & 1 \end{pmatrix}$ und $(-2,u) \rightarrow \begin{pmatrix} -1 & \pm u \\ 0 & -1 \end{pmatrix}$ bzgl. der im Satz angegebenen Zuordnung, womit die bijektive Korrespondenz zwischen der Automorphismengruppe von f und der Lösungsmenge der Pellschen Gleichung erwiesen ist. Teil III.) geht auch hier glatt durch, so daß man die Gruppenisomorphie erhält.

3. Fall: $a = 0$. Dieser Fall ist zum vorhergehenden analog. ♦

Interessiert man sich also für die Automorphismengruppe einer gegebenen, o.B.d.A. primitiven Form f, so kann man aus Satz 3.1 unmittelbar die Struktur von U(f) ablesen. Will man darüber hinaus U(f) explizit bestimmen, so hat man die Gleichung $t^2 - Du^2 = 4$ vollständig zu lösen, wobei D die Diskriminante von f ist. Das ist für negatives und für quadratisches D einfach und im Beweis des Satzes schon geschehen. Schwieriger ist die Situation, falls D positiv und kein Quadrat ist. In diesem Fall hat die Pellsche Gleichung positive Lösungen, d.h. Lösungen $(t,u) \in \boldsymbol{N} \times \boldsymbol{N}$ (das ist noch zu zeigen). Unter diesen sei (t_D , u_D) die kleinste, d.h. diejenige mit minimalem t (und damit auch mit minimalem u). Zur Bestimmung dieser Minimallösung gibt es effiziente Algorithmen. Aus Beweisteil V.) geht nun hervor: $\frac{t_D + u_D\sqrt{D}}{2}$ ist die Grundeinheit ε_D zur Diskriminante D, und mit der Vereinbarung $\varepsilon_n = \frac{t_n + u_n\sqrt{D}}{2} = \varepsilon_D^n$ für $n \in \boldsymbol{N}$ sind (t_n , u_n) , $n \in \boldsymbol{N}$, alle positiven Lösungen der Pellschen Gleichung. t_n und u_n lassen sich wegen $\varepsilon_n = \varepsilon_D \varepsilon_{n-1}$ rekursiv berechnen durch $t_n = \frac{t_D t_{n-1} + D u_D u_{n-1}}{2}$ und $u_n = \frac{t_D u_{n-1} + t_{n-1} u_D}{2}$.

Sie lassen sich aber auch explizit in Abhängigkeit von n ausdrücken: $t_n = \frac{t_n + u_n\sqrt{D}}{2}$ $+ \frac{t_n - u_n\sqrt{D}}{2} = \varepsilon_n + \varepsilon_n^{-1} = \varepsilon_D^n + (\varepsilon_D^{-1})^n = \left(\frac{t_D + u_D\sqrt{D}}{2}\right)^n + \left(\frac{t_D - u_D\sqrt{D}}{2}\right)^n$ und $u_n = \frac{\varepsilon_n - \varepsilon_n^{-1}}{\sqrt{D}} = \frac{1}{\sqrt{D}}\left(\frac{t_D + u_D\sqrt{D}}{2}\right)^n - \frac{1}{\sqrt{D}}\left(\frac{t_D - u_D\sqrt{D}}{2}\right)^n$. So erhält man eine Parameterdarstellung der Lösungsmenge der Pellschen Gleichung und damit von U(f).

4 Die Gesamtdarstellungszahl (I)

Das vielleicht wichtigste und jedenfalls schwierigste der in dieser Arbeit behandelten Probleme ist die Gewinnung expliziter Formeln für die Gesamtdarstellungszahl GP(n,D) der ganzen Zahl n durch primitive Formen der Diskriminante D. Insbesondere läßt sich diese Zahl dann leicht berechnen. Wegen des Zusammenhangs $G(n,D) = \sum_{t \in N, t^2 | D} GP(\frac{n}{t}, \frac{D}{t^2})$ (siehe Definition 1.14) hat man damit auch die Gesamtdarstellungszahl G(n,D) von n durch beliebige Formen der Diskriminante D. Bevor das Problem in Angriff genommen werden kann, müssen allerdings diverse Hilfsmittel bereitgestellt werden. Der Stoff wurde wegen seines Umfangs auf zwei Kapitel verteilt.

Es seien $x,y \in \boldsymbol{Z}$ und $\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \in SL_2(\boldsymbol{Z})$. Dann lassen sich aus $\alpha x + \beta y$ und $\gamma x + \delta y$ dieselben ganzen Zahlen unter Verwendung ganzzahliger Koeffizienten linear kombinieren wie aus x und y, denn $k(\alpha x + \beta y) + l(\gamma x + \delta y) = (k\alpha + l\gamma)x + (k\beta + l\delta)y$ und umgekehrt ist $kx + ly = (k\delta - l\gamma)(\alpha x + \beta y) + (l\alpha - k\beta)(\gamma x + \delta y)$. Daher ist insbesondere $ggT(x,y) = ggT(\alpha x + \beta y , \gamma x + \delta y)$.

Es sei f eine Form und n eine ganze Zahl. In Definition 1.5 wurde auf der Lösungsmenge von $f(x,y) = n$ eine Äquivalenzrelation eingeführt. Die eben angestellte Überlegung zeigt, daß äquivalente Lösungen denselben Komponenten-ggT besitzen. Jeder Lösungsklasse läßt sich also der allen ihren Elementen gemeinsame Komponenten-ggT zuordnen.

Definition 4.1 *Die ganze Zahl n heißt durch die Form f* ***eigentlich*** *dargestellt, wenn teilerfremde ganze Zahlen x,y existieren mit $f(x,y) = n$. Die Anzahl der Äquivalenzklassen von Lösungen von $f(x,y) = n$, deren Komponenten teilerfremd sind, heißt* ***eigentliche Darstellungszahl R*(n,f)*** *von n durch f .*

Sind f und f ' äquivalente Formen, so ist nach Proposition 1.6 $R(n,f) = R(n,f')$ für jedes $n \in \boldsymbol{Z}$. Zum Beweis dieser Aussage wurde eine relationstreue Bijektion von der Lösungsmenge von $f(x,y) = n$ auf die Lösungsmenge von $f'(x,y) = n$ angegeben. Man sieht leicht, daß diese Bijektion den Komponenten-ggT nicht verändert, d.h. insbesondere eigentliche Lösungen auf eigentliche Lösungen abbildet. Folglich gilt $R^*(n,f) = R^*(n,f')$ für jedes $n \in \boldsymbol{Z}$, und man kann jeder Äquivalenzklasse von Formen die allen ihren Elementen gemeinsame eigentliche Darstellungszahl der ganzen Zahl n zuordnen.

Definition 4.2 *Es sei $D \neq 0$ eine Diskriminante und $n \neq 0$ eine ganze Zahl. Die Summe der eigentlichen Darstellungszahlen von n durch primitive Formenklassen der Diskriminante D heißt* ***eigentliche Gesamtdarstellungszahl GP*(n,D)*** *von n durch primitive Formen der Diskriminante D. Die Summe der eigentlichen Darstellungszahlen von n durch beliebige Formenklassen der Diskriminante D heißt eigentliche Gesamtdarstellungszahl* ***G*(n,D)*** *von n durch Formen der Diskriminante D.*

$GP(\cdot ,D)$ läßt sich auf $GP^*(\cdot ,D)$ und $G(\cdot ,D)$ läßt sich auf $G^*(\cdot ,D)$ zurückführen gemäß

Lemma 4.3 *Es sei $D \neq 0$ eine Diskriminante und $n \neq 0$ eine ganze Zahl. Dann gilt:* $GP(n,D) = \sum_{t \in N,\, t^2 \mid n} GP^*(\frac{n}{t^2}, D)$ *und* $G(n,D) = \sum_{t \in N,\, t^2 \mid n} G^*(\frac{n}{t^2}, D)$.

Beweis: Sei f eine Form und $f(x,y) = n$ mit $ggT(x,y) = t$. Dann ist $f(x/t , y/t) = n/t^2$ und falls auch $f(x',y') = n$ ist, so sind offenbar (x,y) und (x',y') genau dann äquivalent, wenn $(x/t , y/t)$ und $(x'/t , y'/t)$ es sind (siehe Definition 1.5). Man erhält auf diese Weise eine bijektive Korrespondenz zwischen den Lösungsklassen von $f(x,y) = n$ mit Komponenten-ggT t und den Lösungsklassen von $f(x,y) = n/t^2$ mit Komponenten-ggT 1. Also ist $R(n,f) = \sum_{t^2 \mid n} R^*(\frac{n}{t^2}, f)$. Durchläuft nun f_i ein Repräsentantensystem der primitiven Formenklassen der Diskriminante D, so ist $GP(n,D) = \sum_i R(n, f_i) = \sum_i \sum_{t^2 \mid n} R^*(\frac{n}{t^2}, f_i)$

$= \sum_{t^2 \mid n} \sum_i R^*(\frac{n}{t^2}, f_i) = \sum_{t^2 \mid n} GP^*(\frac{n}{t^2}, D)$. Die zweite Behauptung beweist man analog. ♦

Das folgende Lemma ist nun der erste Schritt in Richtung auf die Bestimmung der Gesamtdarstellungszahl. Ihm kommt eine zentrale Bedeutung zu, denn sämtliche Sätze dieses und des achten Kapitels beruhen letztlich auf der Untersuchung und Explizierung der hier angegebenen Identität.

Lemma 4.4 *Es sei $D \neq 0$ eine Diskriminante und $n \neq 0$ eine ganze Zahl. Dann ist* $G^*(n,D) = /\{1 \leq b \leq 2/n// b^2 \equiv D \bmod 4n\}/$.

Beweis:

I.) Es geht los mit einer allgemeinen abzähltheoretischen Überlegung.

Es sei G eine (multiplikativ geschriebene) Gruppe und X eine Menge. Eine **Operation** von G auf X ist eine Abbildung von $G \times X$ in X, $(g,x) \to gx$, mit $g(hx) = (g \cdot h)x$ für alle $g,h \in G$ und alle $x \in X$ und mit $1x = x$ für alle $x \in X$. Liegt eine Operation von G auf X vor, so sagt man: G **operiert** auf X.

Ist das der Fall, so ist auf X eine Äquivalenzrelation gegeben durch: $x \sim x'$ genau dann, wenn ein $g \in G$ existiert mit $gx = x'$. Die Äquivalenzklasse $Gx = \{gx \mid g \in G\}$ eines festen $x \in X$ heißt die **Bahn** von x. Die Menge aller Bahnen von G auf X wird mit X / G bezeichnet. Zu festem $x \in X$ ist die Menge $G_x = \{g \in G \mid gx = x\}$ offenbar eine Untergruppe von G. Sie wird als der **Stabilisator** von x bezeichnet.

Es seien nun X und Y Mengen, und die Gruppe G operiere sowohl auf X als auch auf Y. Dann operiert G vermöge $g(x,y) = (gx,gy)$ auch auf $X \times Y$. Es sei S eine Teilmenge von $X \times Y$, die unter der Operation von G invariant bleibt, d.h. die die Vereinigung gewisser Bahnen von G auf $X \times Y$ ist. Dann operiert G auch auf S. Es sollen die Bahnen von G auf S in besonderer Weise abgezählt werden.

Es seien (x,y) und $(x',y') \in S$ mit $(x,y) \sim (x',y')$. Dann gibt es ein $g \in G$ mit $(x',y') = g(x,y) = (gx,gy)$, also gilt insbesondere $x' = gx$. Liegen also (x,y) und (x',y') in dersel-

ben Bahn von G auf S, so liegen x und x' in derselben Bahn von G auf X. Somit ist durch Bahn von $(x,y) \in S \rightarrow$ Bahn von $x \in X$ eine Abbildung von S / G in X / G gegeben. Zu vorgegebenem $x \in X$ lautet nun die Frage: Wie oft kommt die Bahn von x als Bild bei dieser Abbildung vor, auf wieviele Bahnen verteilt sich die Menge $\{(x,y) \in S \mid y \in Y\}$? $(x,y) \sim (x,y') \Leftrightarrow gx = x , gy = y'$ für ein $g \in G \Leftrightarrow gy = y'$ für ein $g \in G_x$. Setze $Y_x = \{y \in Y \mid (x,y) \in S\}$. Dann operiert G_x auf Y_x , und die Anzahl der Bahnen von G_x auf Y_x ist die Antwort auf die Frage. Somit gilt, wenn R(X/G) ein Repräsentantensystem der Bahnen von G auf X ist: $| S / G | = \sum_{x \in R(X/G)} |Y_x / G_x|$. Vertauscht man in dieser Argumentation die Rollen von X und Y, so erhält man $| S / G | = \sum_{y \in R(Y/G)} |X_y / G_y|$.

II.) Es sei jetzt speziell $G = SL_2(\boldsymbol{Z})$, X die Menge aller Formen der Diskriminante D und $Y = \{(x,y) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid ggT(x,y) = 1\}$. G operiert auf X gemäß 1.2 durch $(L,f) \rightarrow Lf$ und auf Y durch $(L , (x,y)) \rightarrow (x,y)L^{-1}$. Die Bahnen von G auf X sind gerade die Formenklassen zur Diskriminante D. Es sei $S = \{(f , (x,y)) \in X \times Y \mid f(x,y) = n\}$ die Menge aller eigentlichen Darstellungen von n durch Formen der Diskriminante D. Nach Proposition 1.3 operiert G auf S, und nach Abschnitt I.) ist somit $| S / G | = \sum_{f \in R(X/G)} |Y_f / G_f|$.

Zu $f \in X$ ist der Stabilisator G_f die Automorphismengruppe U(f) von f . $Y_f = \{(x,y) \in Y \mid f(x,y) = n\}$ ist die Menge aller eigentlichen Darstellungen von n durch f . Die Operation von G_f auf Y_f ist also die Operation von U(f) auf den eigentlichen Lösungen von $f(x,y) = n$. Daher ist $| Y_f / G_f | = R^*(n,f)$ und $| S / G | = \sum_{f \in R(X/G)} R^*(n,f) = G^*(n,D)$.

Andererseits ist $| S / G | = \sum_{(x,y) \in R(Y/G)} |X_{(x,y)} / G_{(x,y)}|$. Bei der Operation von G auf Y gibt es nur eine einzige Bahn: Zu teilerfremden $x,y \in \boldsymbol{Z}$ gibt es $\alpha,\beta \in \boldsymbol{Z}$ mit $\alpha x + \beta y = 1$, also $(x \;\; y)\begin{pmatrix} \alpha & -y \\ \beta & x \end{pmatrix} = (1 \;\; 0)$, so daß unter der Operation von G jedes Element von Y zu (1,0) äquivalent ist. $G_{(1,0)} = \{\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \in G \mid (1 \;\; 0)\begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix} = (1 \;\; 0)\} = \{\begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \mid \beta \in \boldsymbol{Z}\}$

$X_{(1,0)} = \{f \in X \mid f(1,0) = n\} = \{nx^2 + bxy + \frac{b^2 - D}{4n}y^2 \mid b \in \boldsymbol{Z} , b^2 \equiv D \bmod 4n\}$. Wegen $\begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix}\begin{pmatrix} n & \frac{b}{2} \\ \frac{b}{2} & \frac{b^2 - D}{4n} \end{pmatrix}\begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} n & \frac{b + 2n\beta}{2} \\ \frac{b + 2n\beta}{2} & n\beta^2 + b\beta + \frac{b^2 - D}{4n} \end{pmatrix}$ operiert $G_{(1,0)}$ auf $X_{(1,0)}$ dergestalt, daß die zu b gehörige Form $(n , b , \frac{b^2 - D}{4n})$ durch $\begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix}$ in die zu $b + 2n\beta$ gehörige Form übergeführt wird. Da $\beta \in \boldsymbol{Z}$ frei wählbar ist, sind zwei Formen aus $X_{(1,0)}$ genau dann in derselben Bahn bzgl. der Operation von $G_{(1,0)}$, wenn ihre mittleren Koeffizienten kongruent modulo 2n sind. Es folgt

$| S / G | = | X_{(1,0)} / G_{(1,0)} | = |\{1 \le b \le 2|n| \mid b^2 \equiv D \bmod 4n\}|$.

Setzt man die beiden Resultate über $|S / G|$ zusammen, so erhält man die Behauptung. ♦

Insbesondere ist $G^*(n,D)$ endlich und daher auch $G(n,D)$ nach Lemma 4.3. Erst recht sind also $GP^*(n,D)$ und $GP(n,D)$ endlich. Weiterhin sieht man, daß $G^*(-n,D) = G^*(n,D)$ ist und daher auch $G(-n,D) = G(n,D)$ nach Lemma 4.3.

Lemma 4.5 *Es sei $D \neq 0$ eine Diskriminante. $G^*(\cdot,D)$ und $G(\cdot,D)$ sind multiplikativ, d.h. für alle teilerfremden $m,n \in \boldsymbol{Z} \setminus \{0\}$ gilt $G^*(mn,D) = G^*(m,D)G^*(n,D)$ und ebenso $G(mn,D) = G(m,D)G(n,D)$.*

Beweis:

Die Aussage wird zuerst für $G^*(\cdot,D)$ bewiesen. Es seien $m,n \in \boldsymbol{Z} \setminus \{0\}$ mit $ggT(m,n) = 1$. Es sei $b \in \{1, \dots, 2|m|\}$ mit $b^2 \equiv D \bmod 4m$ und $c \in \{1, \dots, 2|n|\}$ mit $c^2 \equiv D \bmod 4n$. Dann ist $b^2 \equiv c^2 \bmod 4$, also $b \equiv c \bmod 2$, also $\frac{1}{2}(c-b) \in \boldsymbol{Z}$. Nach dem chinesischen Restsatz gibt es in der Menge $\{0, 1, \dots, |mn|-1\}$ genau eine Zahl h mit $h \equiv 0 \bmod m$ und $h \equiv \frac{1}{2}(c-b) \bmod n$, also $h = mx = ny + \frac{1}{2}(c-b)$ für gewisse $x,y \in \boldsymbol{Z}$.

Setze $a = 2mx + b = 2ny + c$. $0 \leq h = mx \leq |m|(|n|-1)$, also $b \leq 2mx + b \leq 2|mn|$. Somit ist $a \in \{1, \dots, 2|mn|\}$. Ferner ist $a^2 = 4m^2x^2 + 4bmx + b^2 \equiv b^2 \equiv D \bmod 4m$ und $a^2 = 4n^2y^2 + 4cny + c^2 \equiv c^2 \equiv D \bmod 4n$. Wegen der Teilerfremdheit von m und n folgt daraus $a^2 \equiv D \bmod 4mn$. Auf diese Weise erhält man eine Abbildung von der Menge $\{(b,c) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid 1 \leq b \leq 2|m|, 1 \leq c \leq 2|n|, b^2 \equiv D \bmod 4m, c^2 \equiv D \bmod 4n\}$ in die Menge $\{a \in \boldsymbol{Z} \mid 1 \leq a \leq 2|mn|, a^2 \equiv D \bmod 4mn\}$.

Diese Abbildung ist surjektiv wegen der Möglichkeit und injektiv wegen der Eindeutigkeit der Division mit Rest in $\boldsymbol{Z}$: Zu $a \in \{1, \dots, 2|mn|\}$ mit $a^2 \equiv D \bmod 4mn$ gibt es genau ein $b \in \{1, \dots, 2|m|\}$ mit $a = 2mx + b$ für ein $x \in \boldsymbol{Z}$ und ebenso genau ein $c \in \{1, \dots, 2|n|\}$ mit $a = 2ny + c$ für ein $y \in \boldsymbol{Z}$, und weiter ist $b^2 = (a - 2mx)^2 \equiv a^2 \equiv D \bmod 4m$ und $c^2 = (a - 2ny)^2 \equiv a^2 \equiv D \bmod 4n$. Die Bijektivität der Abbildung bedeutet 4.4 aber gerade $G^*(mn,D) = G^*(m,D)G^*(n,D)$.

Nun zu $G(\cdot,D)$. Die Folge $(q(n))_{n=1}^{\infty}$, definiert durch $q(n) = 1$, falls n ein Quadrat ist und $q(n) = 0$, falls n kein Quadrat ist, ist offenbar multiplikativ. Auch $(G^*(n,D))_{n=1}^{\infty}$ ist multiplikativ. Also ist nach einem bekannten Satz auch die Faltung $((q * G^*)(n))_{n=1}^{\infty}$ der beiden Folgen multiplikativ. Aber nach Lemma 4.3 ist

$$(q * G^*)(n) = \sum_{m|n} q(m) G^*\left(\frac{n}{m}, D\right) = \sum_{m^2|n} G^*\left(\frac{n}{m^2}, D\right) = G(n,D)$$

und damit ist die Behauptung für positive Zahlen klar. Wegen $G(-n,D) = G(n,D)$ ist sie für alle ganzen Zahlen richtig. ♦

Nun werden einige wesentliche Sachverhalte (ohne Beweis) über sogenannte Restklassencharaktere mitgeteilt, die dem §5 von Zagiers Buch entnommen sind. Diese Restklassencharaktere werden im folgenden eine Schlüsselrolle spielen.

Definition *Es sei $n \in \boldsymbol{N}$. Eine Abbildung $\chi : \boldsymbol{Z} \to \boldsymbol{C}$ mit den Eigenschaften (i) $\chi(a) = 0$ genau dann, wenn a und n nicht teilerfremd sind , (ii) $\chi(ab) = \chi(a)\chi(b)$ und (iii) Aus $a \equiv b \bmod n$ folgt $\chi(a) = \chi(b)$ für alle $a,b \in \boldsymbol{Z}$ heißt **Restklassencharakter** modulo n.*

χ läßt sich auffassen als ein Homomorphismus von $((\boldsymbol{Z}/n\boldsymbol{Z})^* , \cdot)$ in $(\boldsymbol{C}\backslash\{0\} , \cdot)$, der auf ganz $\boldsymbol{Z}/n\boldsymbol{Z}$ dadurch ausgedehnt wird, daß er für die weiteren Elemente von $\boldsymbol{Z}/n\boldsymbol{Z}$ gleich Null gesetzt wird. Seine von Null verschiedenen Werte sind daher $\varphi(n)$ te Einheitswurzeln. χ heißt **reell**, wenn χ nur reelle Werte, d.h. Werte aus $\{0, 1, -1\}$ annimmt.

Beispiele (i) $\chi_o : \boldsymbol{Z} \to \{0,1\}$ durch $\chi_o(a) = \begin{cases} 1 & \text{falls } ggT(a, n) = 1 \\ 0 & \text{falls } ggT(a, n) > 1 \end{cases}$ ist offenbar ein reeller Charakter modulo n. Er wird als **Hauptcharakter** modulo n bezeichnet.

(ii) Es sei $p \in \boldsymbol{N}$ prim und ungerade. Das **Legendre-Symbol** $\left(\frac{-}{p}\right) : \boldsymbol{Z} \to \{0, 1, -1\}$ via

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{falls a ein Vielfaches von p ist} \\ 1 & \text{falls a quadratischer Rest modulo p ist} \\ -1 & \text{falls a quadratischer Nichtrest modulo p ist} \end{cases},$$

das aus der elementaren Zahlentheorie bekannt ist, ist ein reeller Charakter modulo p.

(iii) Das Legendre-Symbol läßt sich verallgemeinern: Es sei $b \in \boldsymbol{N}$ ungerade. Dann ist $b = p_1 \ldots p_k$ mit ungeraden Primzahlen $p_1 , \ldots , p_k , k \geq 0$. Das **Jacobi-Symbol** $\left(\frac{-}{b}\right) : \boldsymbol{Z} \to \{0, 1, -1\}$ durch $\left(\frac{a}{b}\right) = \left(\frac{a}{p_1}\right) \ldots \left(\frac{a}{p_k}\right)$ ist ein reeller Restklassencharakter modulo b. Ist b prim, so ist $\left(\frac{-}{b}\right)$ das Legendre-Symbol.

Definition *Eine Diskriminante D heißt **Fundamentaldiskriminante**, falls entweder $D \equiv 0 \bmod 4$, $\frac{D}{4} \equiv 2$ oder $3 \bmod 4$, $\frac{D}{4}$ quadratfrei oder $D \equiv 1 \bmod 4$, D quadratfrei.*

Gleichwertig damit ist, daß man kein Quadrat (außer 1) von D abdividieren kann, ohne daß die Eigenschaft, Diskriminante zu sein, verlorengeht. Zentral ist für uns nun der folgende

Satz *Es sei D eine Fundamentaldiskriminante. Dann ist durch*

$$\chi_D(1) = 1 , \chi_D(-1) = \operatorname{sgn} D , \chi_D(2) = \begin{cases} 0 & \text{falls } D \equiv 0 \bmod 4 \\ 1 & \text{falls } D \equiv 1 \bmod 8 \\ -1 & \text{falls } D \equiv 5 \bmod 8 \end{cases} , \chi_D(p) = \left(\frac{D}{p}\right) \text{ für } p$$

ungerade Primzahl und $\chi_D(ab) = \chi_D(a)\chi_D(b)$ für alle $a,b \in \boldsymbol{Z}$ ein reeller Restklassencharakter modulo $|D|$ der Periode $|D|$ gegeben.

„Periode |D|“ besagt: Für $m \in \boldsymbol{Z}$ ist genau dann $\chi_D(a + m) = \chi_D(a)$ für alle $a \in \boldsymbol{Z}$, wenn m ein Vielfaches von |D| ist.

Es sei $b \in \boldsymbol{N}$ ungerade, also $b = p_1 \ldots p_k$ mit ungeraden Primzahlen $p_1, \ldots, p_k$, $k \geq 0$. Dann hat man $\chi_D(b) = \chi_D(p_1) \ldots \chi_D(p_k) = \left(\frac{D}{p_1}\right) \ldots \left(\frac{D}{p_k}\right) = \left(\frac{D}{b}\right)$ (Jacobi-Symbol). Das Jacobi-Symbol aber läßt sich unter Verwendung folgender Regeln schnell berechnen: (i) $\left(\frac{a}{b}\right) = 0$ genau dann, wenn $ggT(a,b) > 1$ (ii) $\left(\frac{aa'}{b}\right) = \left(\frac{a}{b}\right)\left(\frac{a'}{b}\right)$ (iii) $a \equiv a'$ mod b impliziert $\left(\frac{a}{b}\right) = \left(\frac{a'}{b}\right)$ für alle $a,a' \in \boldsymbol{Z}$ (diese drei Eigenschaften besagen gerade, daß das Jacobi-Symbol $\left(\frac{\ }{b}\right)$ ein Restklassencharakter modulo b ist) (iv) $\left(\frac{a}{b}\right) = \left(\frac{b}{a}\right)(-1)^{\frac{a-1}{2}\frac{b-1}{2}}$ für ungerades $a \in \boldsymbol{N}$ (Reziprozitätsgesetz) (v) $\left(\frac{1}{b}\right) = 1$, $\left(\frac{2}{b}\right) = (-1)^{\frac{b^2-1}{8}}$, $\left(\frac{-1}{b}\right) = (-1)^{\frac{b-1}{2}}$.

Weiterhin ist $\chi_D(2^\alpha b) = (\chi_D(2))^\alpha \chi_D(b)$ für $\alpha \in \boldsymbol{N}_0$ und $\chi_D(-2^\alpha b) = \operatorname{sgn} D \cdot \chi_D(2^\alpha b)$. Also läßt sich $\chi_D(a)$ für jedes $a \in \boldsymbol{Z}$ schnell berechnen.

Die Vorbereitungen sind nun endlich so weit gediehen, daß eine erste, grundlegende Formel für die Gesamtdarstellungszahl bewiesen werden kann.

Satz 4.6 *Es sei D eine Fundamentaldiskriminante und $n \neq 0$ eine ganze Zahl. Dann ist* $G(n,D) = GP(n,D) = \sum_{m \in N,\, m \mid n} \chi_D(m)$.

Beweis:

I.) Es sei $f \cong (a,b,c)$ eine Form der Diskriminante D mit Koeffizienten-ggT t. Dann ist t^2 ein Teiler von D, und da D fundamental ist, folgt daraus $t = 1$ oder $t = 2$. Das letztere ist aber unmöglich: Dann wäre nämlich einerseits $D \equiv 0$ mod 4, d.h. $\frac{1}{4}D \equiv 2$ oder 3 mod 4, und andererseits wäre $\frac{1}{2}f$ eine Form der Diskriminante $\frac{1}{4}D$, also $\frac{1}{4}D \equiv 0$ oder 1 mod 4. Somit ist $t = 1$, d.h. sämtliche Formen der Diskriminante D sind primitiv. Insbesondere ist $G(n,D) = GP(n,D)$ und $G^*(n,D) = GP^*(n,D)$ für jedes $n \in \boldsymbol{Z} \setminus \{0\}$.

Die Folge $(\chi_D(n))_{n=1}^\infty$ ist multiplikativ, ebenso offenbar die durch $I(n) = 1$ für alle $n \in \boldsymbol{N}$ definierte Folge. Also ist die Faltung $((\chi_D * I)(n))_{n=1}^\infty = (\sum_{m|n} \chi_D(m))_{n=1}^\infty$ ebenfalls multiplikativ. Außerdem gilt $\sum_{m|-n} \chi_D(m) = \sum_{m|n} \chi_D(m)$. Andererseits hat auch die linke Seite der behaupteten Identität diese Eigenschaften: Es ist ja $G(\cdot,D)$ multiplikativ nach Lemma 4.5, und nach Lemma 4.4 gilt $G(-n,D) = G(n,D)$. Also reicht es aus, den Satz für

Primzahlpotenzen $n = p^k$ zu beweisen. Dabei wird die wesentliche Arbeit darin bestehen, $G^*(\cdot,D)$ für Primzahlpotenzen durch Explizierung von Lemma 4.4 zu berechnen.

II.) Es sei p eine ungerade Primzahl und $k \in \boldsymbol{N}$. $\{1 \le a \le 2p^k \mid a^2 \equiv D \bmod 4p^k\}$ wird abgebildet auf $\{1 \le b \le p^k \mid b^2 \equiv D \bmod p^k\}$ durch die Vorschrift $a \to a$, falls $a \le p^k$ und $a \to a - p^k$, falls $a > p^k$. Diese Abbildung ist bijektiv: Ist $b \in \{1, \dots, p^k\}$ mit $b^2 \equiv D \bmod p^k$, so ist auch $(b + p^k)^2 \equiv D \bmod p^k$. p^k ist ungerade, also ist genau eine der Zahlen b, $b + p^k \equiv D \bmod 2$, deren Quadrat ist dann $\equiv D \bmod 4$ und daher $\equiv D \bmod 4p^k$, so daß genau eine der Zahlen $b, b + p^k$ in der Urbildmenge liegt. Nach Lemma 4.4 ist somit $G^*(p^k,D) = |\{1 \le b \le p^k \mid b^2 \equiv D \bmod p^k\}|$. Insbesondere ist $G^*(p,D) = 1 + \left(\frac{D}{p}\right)$.

1. Fall: p ist kein Teiler von D. Ist dann $b \in \{1, \dots, p^k\}$ mit $b^2 \equiv D \bmod p^k$, so ist p auch zu 2b teilerfremd. Also gibt es genau ein $x \in \{0, \dots, p-1\}$ mit $2bx \equiv \frac{D - b^2}{p^k} \bmod p$. p teilt $2bx - \frac{D - b^2}{p^k}$, also ist p^{k+1} ein Teiler von $2bxp^k - D + b^2 + x^2p^{2k}$, d.h. es ist $(b + xp^k)^2 \equiv D \bmod p^{k+1}$. $0 \le x \le p - 1$ impliziert $1 \le b + xp^k \le p^{k+1}$. Durch die Vorschrift $b \to b + xp^k$ ist somit eine Abbildung von $\{1 \le b \le p^k \mid b^2 \equiv D \bmod p^k\}$ in $\{1 \le a \le p^{k+1} \mid a^2 \equiv D \bmod p^{k+1}\}$ gegeben. Diese Abbildung ist bijektiv: Zu gegebenem $a \in \{1, \dots, p^{k+1}\}$ gibt es genau ein $b \in \{1, \dots, p^k\}$ mit $a = b + xp^k$ für ein $x \in \boldsymbol{Z}$, dann ist $x \in \{0, \dots, p-1\}$, und aus $a^2 \equiv D \bmod p^{k+1}$ folgt $b^2 = (a - xp^k)^2 \equiv a^2 \equiv D \bmod p^k$ Es ist also $G^*(p^{k+1},D) = G^*(p^k,D)$ und daher $G^*(p^k,D) = 1 + \left(\frac{D}{p}\right)$ für jedes $k \in \boldsymbol{N}$.

2. Fall: p teilt D. Dann gibt es zu $k \ge 2$ kein $b \in \boldsymbol{Z}$ mit $b^2 \equiv D \bmod p^k$. Denn daraus würde folgen, daß p ein Teiler von b ist, also p^2 ein Teiler von b^2 und damit auch von D, was wegen der Fundamentalität von D unmöglich ist: D bzw. $\frac{1}{4}D$ ist ja quadratfrei. Also ist $G^*(p,D) = 1$ und $G^*(p^k,D) = 0$ für $k \ge 2$.

Damit ist $G^*(\cdot,D)$ für die Potenzen ungerader Primzahlen ausgerechnet.

III.) Dasselbe soll nun in ganz ähnlicher Art und Weise für die Primzahl 2 geschehen. Nach Lemma 4.4 ist $G^*(2^k,D) = |\{1 \le b \le 2^{k+1} \mid b^2 \equiv D \bmod 2^{k+2}\}|$.

1. Fall: $D \equiv 1 \bmod 4$. Dann ist $G^*(2,D) = |\{1 \le b \le 4 \mid b^2 \equiv D \bmod 8\}| = 2$ bzw. 0 für $D \equiv 1 \bmod 8$ bzw. $D \equiv 5 \bmod 8$. Sei $k \in \boldsymbol{N}$ und $b \in \{1, \dots, 2^{k+1}\}$ mit $b^2 \equiv D \bmod 2^{k+2}$. Dann ist b ungerade, weil D ungerade ist. Folglich gibt es genau ein $x \in \{0,1\}$ so, daß $\frac{b^2 - D}{2^{k+2}} + bx$ gerade ist. Dann ist 2^{k+3} ein Teiler von $b^2 - D + 2^{k+2}bx + 2^{2k+2}x^2$, d.h. es

ist $(b + 2^{k+1}x)^2 \equiv D \bmod 2^{k+3}$. Dabei ist $1 \le b + 2^{k+1}x \le 2^{k+2}$. Durch die Vorschrift $b \to b + 2^{k+1}x$ erhält man also eine Abbildung von $\{1 \le b \le 2^{k+1} \mid b^2 \equiv D \bmod 2^{k+2}\}$ in $\{1 \le a \le 2^{k+2} \mid a^2 \equiv D \bmod 2^{k+3}\}$. Diese Abbildung ist bijektiv: Zu $a \in \{1, \dots, 2^{k+2}\}$ gibt es genau ein $b \in \{1, \dots, 2^{k+1}\}$ mit $a = b + 2^{k+1}x$ für ein $x \in \boldsymbol{Z}$. Dann ist $x \in \{0,1\}$, und aus $a^2 \equiv D \bmod 2^{k+3}$ folgt $b^2 = (a - 2^{k+1}x)^2 \equiv a^2 \equiv D \bmod 2^{k+2}$. Somit ist $G^*(2^k,D) = G^*(2^{k+1},D)$, und daher ist für jedes $k \in \boldsymbol{N}$ $G^*(2^k,D) = 2$ für $D \equiv 1 \bmod 8$ und $G^*(2^k,D) = 0$ für $D \equiv 5 \bmod 8$.

2. Fall: $D \equiv 0 \bmod 4$. Dann ist $G^*(2,D) = |\{1 \le b \le 4 \mid b^2 \equiv D \bmod 8\}| = 1$ und $G^*(4,D) = |\{1 \le b \le 8 \mid b^2 \equiv D \bmod 16\}| = 0$, denn als Fundamentaldiskriminante ist D modulo 16 kongruent zu 8 oder 12. Sei $k \in \boldsymbol{N}$ und $a \in \{1, \dots, 2^{k+2}\}$ mit $a^2 \equiv D \bmod 2^{k+3}$ Dann ist $a = b + 2^{k+1}x$ für ein $b \in \{1, \dots, 2^{k+1}\}$ und ein $x \in \{0,1\}$. Dabei ist $b^2 = (a - 2^{k+1}x)^2 \equiv a^2 \equiv D \bmod 2^{k+2}$. Ist also $G^*(2^{k+1},D)$ positiv, so auch $G^*(2^k,D)$. Daher folgt aus $G^*(4,D) = 0$, daß $G^*(2^k,D) = 0$ ist für alle $k \ge 2$.

IV.) Nach Lemma 4.3 ist $G(n,D) = \sum_{t^2 \mid n} G^*(\frac{n}{t^2}, D)$ für jedes $n \ne 0$, also hat man: $G(p^k,D) = G^*(p^k,D) + G^*(p^{k-2},D) + \dots + G^*(p^3,D) + G^*(p,D)$ für ungerades $k \in \boldsymbol{N}$ und $G(p^k,D) = G^*(p^k,D) + G^*(p^{k-2},D) + \dots + G^*(p^2,D) + G^*(1,D)$ für gerades k, falls p eine Primzahl ist. Daraus erhält man nun die Behauptung für Primzahlpotenzen:

1. Fall: $\chi_D(p) = 1$, d.h. $\left(\frac{D}{p}\right) = 1$ für ungerades p bzw. $D \equiv 1 \bmod 8$ für $p = 2$. Dann ist $G^*(p^i,D) = 2$ für jedes $i \in \boldsymbol{N}$. Es ist $G^*(1,D) = 1$ nach Lemma 4.4. Man erhält $G(p^k,D) = k + 1 = \sum_{i=0}^{k} 1 = \sum_{i=0}^{k} (\chi_D(p))^i = \sum_{i=0}^{k} \chi_D(p^i) = \sum_{m \mid p^k} \chi_D(m)$.

2. Fall: $\chi_D(p) = -1$, d.h. $\left(\frac{D}{p}\right) = -1$ für ungerades p bzw. $D \equiv 5 \bmod 8$ für $p = 2$. Dann ist $G^*(p^i,D) = 0$ für jedes $i \in \boldsymbol{N}$, also $G(p^k,D) = 0$ für ungerades k, $G(p^k,D) = 1$ für gerades k. Wieder ist $G(p^k,D) = \sum_{i=0}^{k} (\chi_D(p))^i = \sum_{m \mid p^k} \chi_D(m)$.

3. Fall: $\chi_D(p) = 0$, d.h. $\left(\frac{D}{p}\right) = 0$ für ungerades p bzw. $D \equiv 0 \bmod 4$ für $p = 2$. Dann ist $G^*(p,D) = 1$ und $G^*(p^i,D) = 0$ für jedes $i \ge 2$, also $G(p^k,D) = 1 = \sum_{i=0}^{k} (\chi_D(p))^i = \sum_{m \mid p^k} \chi_D(m)$ für jedes $k \in \boldsymbol{N}$. ♦

Die in der Beweisidee von Satz 4.6 liegenden Möglichkeiten sind noch nicht ausgeschöpft, vielmehr läßt sich mit nur geringen Modifikationen auch eine Aussage für beliebige

Diskriminanten beweisen, die genauso aussieht. Bevor das geschieht, muß der Zusammenhang zwischen beliebigen und fundamentalen Diskriminanten geklärt werden.

Lemma 4.7 *Jede Diskriminante $D \neq 0$ hat genau eine Darstellung der Form $D = D_0 r^2$ wobei D_0 eine Fundamentaldiskriminante und $r \in \boldsymbol{N}$ ist.*

Beweis: Es sei $s \in \boldsymbol{N}$ maximal mit $s^2 \mid D$. Wenn s ungerade ist, ist $D \equiv 1 \bmod 4$. Setze in diesem Fall $r = s$ und $D_0 = D/r^2$. D_0 ist quadratfrei wegen der Maximalität von s, und es ist $D_0 \equiv 1 \bmod 4$, denn das gilt für D und r^2 . Ist s gerade und $D/s^2 \equiv 1 \bmod 4$, dann setze wieder $r = s$ und $D_0 = D/r^2$ mit denselben Konsequenzen wie eben. Ist s gerade und $D/s^2 \equiv 2$ oder $3 \bmod 4$, so setze $r = s/2$ und $D_0 = D/r^2$. Dann ist $D_0 \equiv 0 \bmod 4$ und $D_0/4$ ist quadratfrei und $\equiv 2$ oder $3 \bmod 4$. In jedem Fall ist $D = D_0 r^2$ mit $r \in \boldsymbol{N}$ und D_0 Fundamentaldiskriminante.

Es sei auch $D = D_1 t^2$ eine solche Darstellung. Dann ist $D_1 = D_0 \frac{r^2}{t^2}$. Es seien $u,v \in \boldsymbol{N}$ teilerfremd mit $\frac{u^2}{v^2} = \frac{r^2}{t^2}$. Dann ist v^2 ein Teiler von D_0 . Im Falle $D_0 \equiv 1 \bmod 4$ folgt daraus sogleich $v = 1$. Im Falle $D_0 \equiv 0 \bmod 4$ folgt $v = 1$ oder $v = 2$. Letzteres ist aber unmöglich, denn dann wäre u ungerade, also $u^2 \equiv 1 \bmod 4$, also $D_1 = D_0 \frac{u^2}{v^2} = \frac{D_0}{4} u^2 \equiv D_0/4 \equiv 2$ oder $3 \bmod 4$, was im Widerspruch zu der Tatsache steht, daß D_1 eine Diskriminante ist. In jedem Fall ist also $v = 1$, d.h. t ist ein Teiler von r. Analog zeigt man, daß auch umgekehrt r ein Teiler von t ist. Es folgt $t = r$ und $D_1 = D_0$.♦

Es sei $D = D_0 r^2$ wie eben. Zur Fundamentaldiskriminante D_0 gehört ja ein gewisser reeller Restklassencharakter modulo $|D_0|$, der mit χ_{D_0} bezeichnet wurde. Offenbar ist dann durch $\chi_D(m) = \left\{ \begin{matrix} \chi_{D_0}(m) & \text{falls } ggT(m,D) = 1 \\ 0 & \text{falls } ggT(m,D) > 1 \end{matrix} \right\} = \left\{ \begin{matrix} \chi_{D_0}(m) & \text{falls } ggT(m,r) = 1 \\ 0 & \text{falls } ggT(m,r) > 1 \end{matrix} \right\}$ für $m \in \boldsymbol{Z}$ ein reeller Restklassencharakter modulo $|D|$ gegeben. χ_D wird als der durch χ_{D_0} **induzierte** Charakter modulo $|D|$ bezeichnet.

Satz 4.8 *Es sei $D \neq 0$ eine Diskriminante, $D = D_0 r^2$ mit D_0 Fundamentaldiskriminante und $r \in \boldsymbol{N}$, und es sei $n \neq 0$ eine zu r teilerfremde ganze Zahl.*

Dann ist $G(n,D) = GP(n,D) = \sum_{m \in N,\, m \mid n} \chi_D(m)$.

Der **Beweis** unterscheidet sich nur wenig von dem Beweis von Satz 4.6 :

I.) Es sei $f \cong (a,b,c)$ eine Form der Diskriminante D mit Koeffizienten-ggT s . f stelle n dar. Dann gilt einerseits $s \mid n$, also $ggT(r,s) = 1$, und andererseits $s^2 \mid D_0 r^2$. Daraus folgt

$s^2 \mid D_0$ und somit ist $s = 1$ oder $s = 2$. Letzteres ist aber unmöglich: Dann wäre nämlich $D_0 \equiv 0 \bmod 4$ und r ungerade, also $r^2 \equiv 1 \bmod 4$, so daß $\frac{1}{2}f$ eine Form der Diskriminante $\frac{1}{4}D = \frac{1}{4}D_0 r^2 \equiv 2$ oder $3 \bmod 4$ wäre, was im Widerspruch dazu steht, daß Diskriminanten $\equiv 0$ oder $1 \bmod 4$ sind. Es ist also auf jeden Fall $s = 1$, d.h. n wird nur durch primitive Formen der Diskriminante D dargestellt. Daher ist wie in Satz 4.6. $G(n,D) = GP(n,D)$ sowie $G^*(n,D) = GP^*(n,D)$, aber hier nur für zu r teilerfremdes n. Und genau wie bei Satz 4.6 braucht die Behauptung nur für Primzahlpotenzen bewiesen zu werden, wobei man hier aber nur zu r teilerfremde Primzahlen zu berücksichtigen hat.

II.) stimmt im wesentlichen mit II.) in 4.6 überein, nur muß man zusätzlich voraussetzen, daß p kein Teiler von r ist. Dies wird im 2. Fall benötigt: Aus $p^2 \mid D$ folgt dann $p^2 \mid D_0$, was unmöglich ist.

III.) stimmt im wesentlichen mit III.) in 4.6 überein, man muß aber beachten, daß die Primzahl 2 nur im Falle der Ungeradheit von r zu untersuchen ist. Dies wird im 2. Fall benötigt: Dann ist $r^2 \equiv 1 \bmod 4$, also $\frac{1}{4}D = \frac{1}{4}D_0 r^2 \equiv \frac{1}{4}D_0 \equiv 2$ oder $3 \bmod 4$, und somit ist $D \equiv 8$ oder $12 \bmod 16$, so daß $b^2 \equiv D \bmod 16$ unlösbar ist.

IV.) ist wesentlich identisch mit IV.) in 4.6, man muß aber zusätzlich $ggT(p,r) = 1$ voraussetzen. Dies benötigt man für die Identität $\chi_D(p) = \left(\frac{D}{p}\right)$ für ungerades p:

$\chi_D(p) = \chi_{D_0}(p) = \left(\frac{D_0}{p}\right) = \left(\frac{D_0}{p}\right)\left(\frac{r}{p}\right)^2 = \left(\frac{D}{p}\right)$. Für ungerades r muß auch $p = 2$ betrachtet werden. Dann ist $r^2 \equiv 1 \bmod 8$, also $D \equiv D_0 \bmod 8$, und $\chi_D(2) = \chi_{D_0}(2)$, so daß tatsächlich $\chi_D(2) = 1$ bzw. -1 bzw. 0 gleichwertig ist mit $D \equiv 1 \bmod 8$ bzw. $D \equiv 5 \bmod 8$ bzw. $D \equiv 0 \bmod 4$. ♦

4.6 ist in 4.8 als Spezialfall r = 1 enthalten. Man beachte, daß in der Situation von Satz 4.8 $G(n,D) = G(n,D_0)$ ist, denn es gilt $\chi_D(m) = \chi_{D_0}(m)$ für $ggT(m,r) = 1$.

Von der Bedingung $ggT(n,r) = 1$ kann man sich bei dieser Art der Beweisführung nicht freimachen. Der in Kapitel 8 gegebene völlig allgemeine Satz über die Gesamtdarstellungszahl wird auf andere Weise bewiesen, wobei aber Satz 4.6 benutzt wird.

5 Die mittlere Gesamtdarstellungszahl

In diesem kurzen Kapitel geht es darum, aus den Sätzen 4.6 und 4.8 Formeln herzuleiten für die durchschnittliche (im Sinne des arithmetischen Mittels) Gesamtdarstellungszahl einer natürlichen Zahl durch primitive Formen einer festen Diskriminante D.

Lemma 5.1 *Es sei χ ein Restklassencharakter modulo r , aber nicht der Hauptcharakter. Dann ist* $\sum_{m \bmod r} \chi(m) = 0$. *(m durchlaufe ein Vertretersystem der Restklassen modulo r)*

Beweis:

Weil χ nicht der Hauptcharakter modulo r ist, gibt es ein $n \in \boldsymbol{Z}$ mit $ggT(n,r) = 1$ und $\chi(n) \neq 1$. $(1 - \chi(n)) \sum_{m \bmod r} \chi(m) = \sum_{m \bmod r} \chi(m) - \sum_{m \bmod r} \chi(mn) = \sum_{m \bmod r} \chi(m) - \sum_{m \bmod r} \chi(m) = 0$ denn mit m durchläuft auch mn ein Repräsentantensystem der Restklassen modulo r. Wegen $1 - \chi(n) \neq 0$ folgt daraus $\sum_{m \bmod r} \chi(m) = 0$. ♦

Definition 5.2 *Es sei χ ein Restklassencharakter. $L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$ mit $s \in \boldsymbol{C}$ heißt die zu χ gehörige **L-Reihe** .*

Im §6 des Buchs von Zagier wird gezeigt, daß $L(s,\chi)$ für alle $s \in \boldsymbol{C}$ mit $Re(s) > 1$ bzw. $Re(s) > 0$ konvergiert, je nachdem, ob χ ein Hauptcharakter ist (d.h. nur die Werte 0 und 1 annimmt) oder nicht. Dieses allgemeine Resultat wird hier aber nicht benötigt.

Von nun an sei χ ein reeller Charakter modulo r und $s \in \boldsymbol{R}$. Weil $\sum_{n=1}^{\infty} \frac{1}{n^s}$ für $s > 1$ konvergiert und $|\chi(n)| \leq 1$ ist, ist $L(s,\chi)$ für $s > 1$ absolut konvergent. Was geschieht an der Stelle $s = 1$?

$\sum_{p \text{ prim}} \frac{1}{p}$ divergiert, also auch $\sum_{\substack{p \text{ prim} \\ ggT(p,r)=1}} \frac{1}{p}$ und erst recht $\sum_{\substack{n \in N \\ ggT(n,r)=1}} \frac{1}{n} = L(1, \chi)$, wenn χ der Hauptcharakter modulo r ist. Ist χ nicht der Hauptcharakter, so erhält man für den Rest der Reihe $L(1, \chi)$ die Abschätzung: $\left| \sum_{n=m+1}^{m+k} \frac{\chi(n)}{n} \right| = \left| \sum_{n=m+1}^{m+r[k/r]} \frac{\chi(n)}{n} + \sum_{n=m+r[k/r]+1}^{m+k} \frac{\chi(n)}{n} \right|$

$\leq \left| \sum_{n=1}^{[k/r]} \sum_{i=m+(n-1)r+1}^{m+nr} \frac{\chi(i)}{i} \right| + \left| \sum_{n=m+r[k/r]+1}^{m+k} \frac{\chi(n)}{n} \right| \leq \sum_{n=1}^{[k/r]} \frac{r}{2} \left(\frac{1}{m+(n-1)r+1} - \frac{1}{m+nr} \right)$

$+ \frac{r}{2m}$ (denn nach Lemma 5.1 wird χ auf einem Vertretersystem modulo r ebenso oft 1

wie -1, und daher läßt sich $\left|\sum_{i=m+(n-1)r+1}^{m+nr}\frac{\chi(i)}{i}\right|$ durch $\frac{r}{2}\left(\frac{1}{m+(n-1)r+1}-\frac{1}{m+nr}\right)$ nach oben abschätzen. Aus demselben Grund ist $\left|\sum_{n=m+r[k/r]+1}^{m+k}\frac{\chi(n)}{n}\right| \le \frac{r}{2m}$.)

$$\le \frac{r}{2}\sum_{n=1}^{[k/r]}\left(\frac{1}{m+(n-1)r}-\frac{1}{m+nr}\right)+\frac{r}{2m} = \frac{r}{2}\left(\frac{1}{m}-\frac{1}{m+r[k/r]}\right)+\frac{r}{2m} \le \frac{r}{m}.$$

Der Rest der Reihe wird also mit wachsendem m unabhängig von k dem Betrage nach beliebig klein. $L(1,\chi)$ konvergiert, wenn χ nicht der Hauptcharakter ist.

Satz 5.3 *Ist D eine Fundamentaldiskriminante, so ist* $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} GP(n,D) = L(1,\chi_D)$.

Die linke Seite dieser Identität kann offenbar gedeutet werden als das arithmetische Mittel der Zahlen GP(n,D) , $n \in \boldsymbol{N}$.

Beweis:

$$\frac{1}{N}\sum_{n=1}^{N} GP(n,D) = \frac{1}{N}\sum_{n=1}^{N}\sum_{m|n}\chi_D(m) \quad \text{(nach Satz 4.6)} \quad = \frac{1}{N}\sum_{m=1}^{N}\left[\frac{N}{m}\right]\chi_D(m) \quad \text{und nun}$$

1. Fall: $D = 1$. Dann ist $\chi_D(m) = 1$ für jedes $m \in \boldsymbol{N}$ und folglich $\frac{1}{N}\sum_{n=1}^{N} GP(n,D) =$

$$\frac{1}{N}\sum_{m=1}^{N}\left[\frac{N}{m}\right] = \frac{1}{N}\sum_{m=1}^{N}\left(\frac{N}{m}+O(1)\right) = \sum_{m=1}^{N}\frac{1}{m}+O(1) \xrightarrow[N\to\infty]{} \infty = L(1,\chi_1).$$

Der Satz gilt für $D = 1$ also nur „im uneigentlichen Sinne."

2. Fall: $D \neq 1$. Dann ist D kein Quadrat, also gibt es (sogar unendlich viele) Primzahlen p mit $\chi_D(p) = \left(\frac{D}{p}\right) = -1$ (siehe Satz 10.21 bei Ischebeck): χ_D ist nicht der Hauptcharakter.

Man erhält $\frac{1}{N}\sum_{n=1}^{N} GP(n,D) = \frac{1}{N}\sum_{m<\sqrt{N}}\left[\frac{N}{m}\right]\chi_D(m) + \frac{1}{N}\sum_{m\ge\sqrt{N}}\left[\frac{N}{m}\right]\chi_D(m) =$

$$\frac{1}{N}\sum_{m<\sqrt{N}}\left(\frac{N}{m}+O(1)\right)\chi_D(m) + \frac{1}{N}\sum_{k\le\sqrt{N}}\sum_{\substack{m\ge\sqrt{N}\\ km\le N}}\chi_D(m) = \sum_{m<\sqrt{N}}\frac{1}{m}\chi_D(m) + \frac{1}{N}O(\sqrt{N}) +$$

$$\frac{1}{N}\sum_{k\le\sqrt{N}}O(1) \quad \text{(denn } \sum_{m \bmod |D|}\chi_D(m) = 0 \text{ und daher } \left|\sum_{\substack{m\ge\sqrt{N}\\ km\le N}}\chi_D(m)\right| \le |D| \text{ für alle k,N)}$$

$$= \sum_{m<\sqrt{N}}\frac{\chi_D(m)}{m} + O\left(\frac{1}{\sqrt{N}}\right) \xrightarrow[N\to\infty]{} \sum_{m=1}^{\infty}\frac{\chi_D(m)}{m} = L(1,\chi_D).$$

Der Satz gilt hier im eigentlichen Sinne, denn $L(1,\chi_D)$ konvergiert. ♦

Satz 5.4 *Es sei $D \neq 0$ eine Diskriminante, $D = D_0 r^2$ mit D_0 Fundamentaldiskriminante und $r \in \boldsymbol{N}$. Dann gilt:* $\lim_{N\to\infty} \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} GP(n,D) = L(1,\chi_D)$.

Die Aussage bedeutet, daß Satz 5.3 für beliebige Diskriminanten richtig bleibt, wenn der Mittelwert nur für zu r teilerfremde natürliche Zahlen gebildet wird. Denn in {1, ... ,N} gibt es $\varphi(r)\frac{N}{r} + O(1)$ zu r teilerfremde Zahlen. 5.3 ist als Spezialfall $r = 1$ enthalten.

Beweis: $\frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} GP(n,D) = \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} \sum_{m|n} \chi_D(m)$ (nach Satz 4.8) $= \frac{r}{N\varphi(r)} \sum_{\substack{m=1 \\ ggT(m,r)=1}}^{N} \sum_{\substack{km\le N \\ ggT(k,r)=1}} \chi_D(m)$. $ggT(m,r) = 1$ kann in den Summationsbedingungen des letzten Ausdrucks weggelassen werden, denn ist m zu r nicht teilerfremd, so ist $\chi_D(m) = 0$.

1. Fall: D ist ein Quadrat. Dann ist $D_0 = 1$, also $\chi_{D_0}(m) = 1$ für alle $m \in \boldsymbol{Z}$, so daß χ_D der Hauptcharakter modulo |D| ist. $\frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} GP(n,D) = \frac{r}{N\varphi(r)} \sum_{\substack{m=1 \\ ggT(m,r)=1}}^{N} \sum_{\substack{km\le N \\ ggT(k,r)=1}} 1$

$= \frac{r}{N\varphi(r)} \sum_{\substack{m=1 \\ ggT(m,r)=1}}^{N} (\varphi(r)\frac{N}{mr} + O(1)) = \sum_{\substack{m=1 \\ ggT(m,r)=1}}^{N} \frac{1}{m} + O(1) \xrightarrow[N\to\infty]{} \infty = L(1,\chi_D)$.

Der Satz gilt für quadratische Diskriminanten nur „im uneigentlichen Sinne."

2. Fall: D ist kein Quadrat. Zunächst soll gezeigt werden, daß dann χ_D nicht der Hauptcharakter modulo |D| ist. Es ist $D_0 \neq 1$ und daher ist jedenfalls χ_{D_0} kein Hauptcharakter. Es gibt also eine Primzahl p mit $\chi_{D_0}(p) = -1$. s sei das Produkt aller von p verschiedenen Primteiler von r. Annahme: Es gibt eine Primzahl q, die sowohl D als auch $p + sD_0$ teilt. q teilt D_0 oder r. Im ersten Fall folgt $q = p$. Im zweiten Fall ist $q = p$ oder q ein Teiler von s, aber auch daraus folgt $q = p$. Es ist also jedenfalls $q = p$, also ist p ein Teiler von D_0 und daher $\chi_{D_0}(p) = 0$: Widerspruch. D und $p + sD_0$ sind teilerfremd. Somit ist $\chi_D(p + sD_0) = \chi_{D_0}(p + sD_0) = \chi_{D_0}(p) = -1$. χ_D ist nicht der Hauptcharakter. $\frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} GP(n,D) = \frac{r}{N\varphi(r)} \sum_{m<\sqrt{N}} \sum_{\substack{km\le N \\ ggT(k,r)=1}} \chi_D(m) +$

$\frac{r}{N\varphi(r)} \sum_{\substack{k\le\sqrt{N} \\ ggT(k,r)=1}} \sum_{\substack{m\ge\sqrt{N} \\ km\le N}} \chi_D(m) = \frac{r}{N\varphi(r)} \sum_{m<\sqrt{N}} (\varphi(r)\frac{N}{mr} + O(1))\chi_D(m) +$

$\frac{r}{N\varphi(r)} \sum_{\substack{k\le\sqrt{N} \\ ggT(k,r)=1}} O(1)$ (denn $\sum_{m \bmod D} \chi_D(m) = 0$) $= \sum_{m<\sqrt{N}} \frac{\chi_D(m)}{m} + O(\frac{1}{\sqrt{N}}) \xrightarrow[N\to\infty]{}$

$L(1, \chi_D)$ im eigentlichen Sinne. ♦

6 Die mittlere Darstellungszahl

In diesem Kapitel geht es darum, Formeln zu finden für die durchschnittliche Darstellungszahl einer natürlichen Zahl durch eine gegebene Form f , also das, was im letzten Kapitel für die Gesamtdarstellungszahl geleistet wurde, nun für die einzelnen Darstellungszahlen zu erreichen. Da man R(n,f) anders als G(n,D) nicht angeben kann, ist das viel schwerer.

Satz 6.1 *Es sei D kein Quadrat und f eine primitive, für D < 0 außerdem positiv definite, Form der Diskriminante D. Dann ist* $\lim\limits_{N\to\infty} \frac{1}{N}\sum\limits_{n=1}^{N} R(n,f) = \left\{\begin{array}{ll} \frac{2\pi}{w\sqrt{-D}} & \text{für } D<0 \\ \frac{\log\varepsilon_D}{\sqrt{D}} & \text{für } D>0 \end{array}\right\}$.

Dabei ist w = 6 für D = –3 , w = 4 für D = –4 und w = 2 für D < –4 . ε_D ist die Grundeinheit zu D gemäß Satz 3.1.

Die linke Seite der Identität kann offenbar gedeutet werden als das arithmetische Mittel der Zahlen R(n,f) , $n \in \boldsymbol{N}$. Die rechte Seite zeigt, daß diese Größe nur von der Diskriminante von f abhängt.

Beweis: Es sei $f(x,y) = ax^2 + bxy + cy^2$.

I.) Für negatives D ist die Sache relativ einfach. U(f) operiert auf $\boldsymbol{Z} \times \boldsymbol{Z} \setminus \{(0,0)\}$ durch $(A, (x,y)) \to (x,y)A^{-1}$. Der Stabilisator $U(f)_{(x,y)} = \{A \in U(f) \mid (x,y)A^{-1} = (x,y)\}$ eines Zahlenpaares $(x,y) \neq (0,0)$ ist eine Untergruppe von U(f) , die $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ nicht enthält.

Ist $D < -4$, so ist U(f) isomorph zu $\boldsymbol{Z} / 2\boldsymbol{Z}$; ist $D = -4$, so ist U(f) isomorph zu $\boldsymbol{Z} / 4\boldsymbol{Z}$. Jedenfalls ist $U(f)_{(x,y)} = \{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\}$, denn alle anderen Untergruppen von U(f) enthalten $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. Im Falle $D = -3$ ist $U(f) = \{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -\frac{1}{2}(b-1) & a \\ -c & \frac{1}{2}(b+1) \end{pmatrix}, \begin{pmatrix} \frac{1}{2}(b-1) & -a \\ c & -\frac{1}{2}(b+1) \end{pmatrix}, \begin{pmatrix} \frac{1}{2}(b+1) & -a \\ c & -\frac{1}{2}(b-1) \end{pmatrix}, \begin{pmatrix} -\frac{1}{2}(b+1) & a \\ -c & \frac{1}{2}(b-1) \end{pmatrix}\}$. $\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\}$ und $\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -\frac{1}{2}(b+1) & a \\ -c & \frac{1}{2}(b-1) \end{pmatrix}, \begin{pmatrix} \frac{1}{2}(b-1) & -a \\ c & -\frac{1}{2}(b+1) \end{pmatrix}\}$ sind die einzigen Untergruppen von U(f) , die $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ nicht enthalten. Annahme: $(x,y)\begin{pmatrix} -\frac{1}{2}(b+1) & a \\ -c & \frac{1}{2}(b-1) \end{pmatrix} = (x,y)$.

Dann ist $ax + \frac{1}{2}(b-1)y = y$ und $-\frac{1}{2}(b+1)x - cy = x$, also $y = -\frac{b+3}{2c}x$.

Einsetzen in die erste Gleichung ergibt $4acx - (b^2 + 2b - 3)x = -2(b + 3)x$, d.h. $12x = 0$ und somit $x = y = 0$: Widerspruch. Daraus folgt $U(f)_{(x,y)} = \{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\}$.

Der Stabilisator von $(x,y) \neq (0,0)$ ist also in jedem Fall trivial. Daher haben alle Bahnen der Operation die Länge $w = |U(f)|$, und es ist $R(n,f) = \frac{1}{w} |\{(x,y) \in \boldsymbol{Z}^2 \mid f(x,y) = n\}|$ für $n \in \boldsymbol{Z}\backslash\{0\}$. Weiter hat man $\sum_{n=1}^{N} R(n,f) = \frac{1}{w} |\{(x,y) \in \boldsymbol{Z}^2 \mid (x,y) \neq (0,0) , f(x,y) \leq N\}|$ für jedes $N \in \boldsymbol{N}$, denn f ist positiv definit.

a ist positiv. $ax^2 + bxy + cy^2 \leq N$ ist gleichwertig mit $a(x + \frac{b}{2a} y)^2 + \frac{-D}{4a} y^2 \leq N$. Die Menge aller Punkte $(x,y) \in \boldsymbol{R}^2$, die dieser Ungleichung genügen, ist eine Ellipsenfläche Ihr Flächeninhalt ist bis auf einen Fehler von der Größenordnung ihres Umfangs gleich der Anzahl der in ihr befindlichen Gitterpunkte, d.h. gleich $|\{(x,y) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid f(x,y) \leq N\}|$.

Durch $(x,y) \rightarrow (u,v) = (x + \frac{b}{2a} y , y)$ ist eine Koordinatentransformation im $\boldsymbol{R}^2$ gegeben. Sie ist inhaltstreu, denn die Ableitungsmatrix hat die Determinante 1. Die Ungleichung $au^2 + \frac{-D}{4a} v^2 \leq N$ beschreibt eine achsenparallele Ellipsenfläche mit den Halbachsenlängen $\sqrt{\frac{N}{a}}$ und $\sqrt{\frac{4aN}{-D}}$. Ihr Flächeninhalt ist somit $\frac{2N\pi}{\sqrt{-D}}$, ihr Umfang ist von der Größenordnung $O(\sqrt{N})$. Es folgt die Behauptung für negatives D :

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} R(n,f) = \lim_{N \to \infty} \frac{1}{N} \frac{1}{w} \left(\frac{2N\pi}{\sqrt{-D}} + O(\sqrt{N})\right) = \frac{2\pi}{w\sqrt{-D}} .$$

II.) Es sei jetzt D positiv und kein Quadrat. Mit $\theta = -\frac{b - \sqrt{D}}{2a}$ und $\theta' = -\frac{b + \sqrt{D}}{2a}$ ist $f(x,y) = a(x - \theta y)(x - \theta' y)$. U(f) operiert auf der Lösungsmenge von $f(x,y) = n$ durch $(A , (x,y)) = (\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}, (x,y)) \rightarrow (x', y') = (x,y)A^{-1} = (\delta x - \beta y , \alpha y - \gamma x)$.

$(U(f) , \cdot)$ ist isomorph zu $(E_D , \cdot)$ mit $E_D = \{\frac{t + u\sqrt{D}}{2} \mid t,u \in \boldsymbol{Z} , t^2 - Du^2 = 4\}$ (siehe Schritt IV.) in Satz 3.1) . Der Matrix $A = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$ entspricht dabei $\varepsilon = \frac{t + u\sqrt{D}}{2}$ mit $\alpha = \frac{1}{2}(t - bu)$, $\beta = -cu$, $\gamma = au$, $\delta = \frac{1}{2}(t + bu)$. Betrachte nun zu einer Lösung (x,y) von $f(x,y) = n$, $n \in \boldsymbol{N}$, die Größe $\frac{x - \theta' y}{x - \theta y}$. Mit (x', y') wie eben hängt $\frac{x' - \theta' y'}{x' - \theta y'}$

mit $\frac{x-\theta' y}{x-\theta y}$ so zusammen: $\varepsilon(x'-\theta'y') = \frac{t+u\sqrt{D}}{2}(\delta x-\beta y+\frac{b+\sqrt{D}}{2a}(\alpha y-\gamma x)) = \frac{t+u\sqrt{D}}{2}((\delta-\frac{b+\sqrt{D}}{2a}\gamma)x+(\frac{b+\sqrt{D}}{2a}\alpha-\beta)y) = \frac{t+u\sqrt{D}}{2}((\frac{t+bu}{2}-\frac{b+\sqrt{D}}{2a}au)x+(\frac{b+\sqrt{D}}{2a}\frac{t-bu}{2}+cu)y) = \frac{t+u\sqrt{D}}{2}(\frac{t-u\sqrt{D}}{2}x+\frac{bt+t\sqrt{D}-bu\sqrt{D}-Du}{4a}y) = \frac{t+u\sqrt{D}}{2}\frac{t-u\sqrt{D}}{2}(x+\frac{b+\sqrt{D}}{2a}y) = x-\theta'y$. Durch eine analoge Rechnung (oder auch durch Konjugation der Gleichung $\varepsilon(x'-\theta'y') = x-\theta'y$ im quadratischen Zahlkörper $\boldsymbol{Q}(\sqrt{D})$) erhält man $\varepsilon^{-1}(x'-\theta y') = x-\theta y$. Somit ist $\frac{x-\theta' y}{x-\theta y} = \varepsilon^2\frac{x'-\theta' y'}{x'-\theta y'}$.

III.) Nach Schritt V.) in Satz 3.1 ist entweder $E_D = \{1,-1\}$ oder $E_D = \{\pm\varepsilon_D^m \mid m\in\boldsymbol{Z}\}$. Es sei zunächst das letztere der Fall. Außerdem sei $a>0$ vorausgesetzt; den Fall $a<0$ kann man analog behandeln. Nach Schritt II.) ist $\frac{x'-\theta' y'}{x'-\theta y'} = \frac{x-\theta' y}{x-\theta y}\varepsilon_D^{2m}$ für ein $m\in\boldsymbol{Z}$. Wegen $a(x-\theta y)(x-\theta'y) = n > 0$ und $a>0$ ist auch $\frac{x-\theta' y}{x-\theta y} > 0$. Daher gibt es genau ein $m\in\boldsymbol{Z}$ mit $1 < \frac{x'-\theta' y'}{x'-\theta y'} \le \varepsilon_D^2$. Jede Lösung (x,y) von $f(x,y) = n$ ist also zu einer Lösung (x', y') mit $1 < \frac{x'-\theta' y'}{x'-\theta y'} \le \varepsilon_D^2$ äquivalent. Diese ist eindeutig bestimmt, wenn man außerdem $x'-\theta y' > 0$ fordert, denn offenbar führt jede Transformation der Lösung mit einer von $\pm\begin{pmatrix}1&0\\0&1\end{pmatrix}$ verschiedenen Matrix wieder aus dem gewünschten Intervall heraus, während Transformation mit $\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$ das Vorzeichen von $x'-\theta y'$ ändert. Somit ist $R(n,f) = |\{(x,y)\in\boldsymbol{Z}\times\boldsymbol{Z} \mid f(x,y)=n,\ x-\theta y>0,\ 1<\frac{x-\theta' y}{x-\theta y}\le\varepsilon_D^2\}|$ und $\sum_{n=1}^{N} R(n,f) = |\{(x,y)\in\boldsymbol{Z}\times\boldsymbol{Z} \mid 0<f(x,y)\le N,\ x-\theta y>0,\ 1<\frac{x-\theta' y}{x-\theta y}\le\varepsilon_D^2\}|$. Diese Bedingungen lassen sich vereinfachen. Wegen $a>0$ ist $\theta>\theta'$, und daher folgt aus $\frac{x-\theta' y}{x-\theta y} > 1$ und $x-\theta y>0$, daß $y>0$ ist. Ferner folgt aus $\frac{x-\theta' y}{x-\theta y} \le \varepsilon_D^2$ durch Auflösung nach x die Ungleichung $x \ge \frac{\varepsilon_D^2\theta-\theta'}{\varepsilon_D^2-1}y$, wieder mit $x-\theta y>0$. Man erhält also die drei Ungleichungen $f(x,y)\le N$, $y>0$ und $x\ge\frac{\varepsilon_D^2\theta-\theta'}{\varepsilon_D^2-1}y$.

Nun seien umgekehrt diese vorausgesetzt. Wegen $\theta > \theta'$ folgt $x \geq \frac{\varepsilon_D^2\theta - \theta'}{\varepsilon_D^2 - 1} y > \theta y$, also $x - \theta y > 0$. Es ist $x - \theta y < x - \theta' y$ und daher $\frac{x - \theta' y}{x - \theta y} > 1$ sowie $x - \theta' y > 0$. Daraus folgt $f(x,y) = a(x - \theta y)(x - \theta' y) > 0$. Endlich folgt aus $x \geq \frac{\varepsilon_D^2\theta - \theta'}{\varepsilon_D^2 - 1} y$ noch $\frac{x - \theta' y}{x - \theta y} \leq \varepsilon_D^2$. Also $\sum_{n=1}^{N} R(n,f) = |\{(x,y) \in \boldsymbol{Z}^2 \mid f(x,y) \leq N, y > 0, x \geq \frac{\varepsilon_D^2\theta - \theta'}{\varepsilon_D^2 - 1} y\}|$.

IV.) Die Gleichung $f(x,y) = ax^2 + bxy + cy^2 = a(x + \frac{b}{2a} y)^2 - \frac{D}{4a} y^2 = N$ beschreibt eine Hyperbel im $\boldsymbol{R}^2$, denn a und $\frac{D}{4a}$ sind positiv. Durch $y = 0$ ist die x-Achse charakterisiert, und durch $x = \frac{\varepsilon_D^2\theta - \theta'}{\varepsilon_D^2 - 1} y$ ist eine zweite Ursprungsgerade gegeben. Der Flächeninhalt des von diesen drei Kurven in der oberen Halbebene begrenzten Gebietes ist bis auf einen Fehler von der Größenordnung des Gebietsumfangs gleich der Anzahl der darin befindlichen Gitterpunkte, d.h. gleich $\sum_{n=1}^{N} R(n,f)$.

Durch $(u,v) \to (x,y) = (\frac{\theta v - \theta' u}{\theta - \theta'}, \frac{v - u}{\theta - \theta'})$ ist eine Koordinatentransformation im $\boldsymbol{R}^2$ gegeben, denn $x = \frac{\theta v - \theta' u}{\theta - \theta'}$, $y = \frac{v - u}{\theta - \theta'}$ ist äquivalent mit $u = x - \theta y$, $v = x - \theta' y$ und die Ableitungsmatrix $\frac{1}{\theta - \theta'} \begin{pmatrix} -\theta' & \theta \\ -1 & 1 \end{pmatrix}$ hat die Determinante $\frac{1}{\theta - \theta'} = \frac{a}{\sqrt{D}} \neq 0$. $a(x - \theta y)(x - \theta' y) \leq N$ ist äquivalent mit $auv \leq N$, $y > 0$ bedeutet wegen $\theta > \theta'$ gerade $v > u$, und drittens gilt: $(\varepsilon_D^2 - 1)x \geq (\varepsilon_D^2\theta - \theta')y \Leftrightarrow (\varepsilon_D^2 - 1)(\theta v - \theta' u) \geq (\varepsilon_D^2\theta - \theta')(v - u) \Leftrightarrow \varepsilon_D^2(\theta - \theta')u \geq (\theta - \theta')v \Leftrightarrow v \leq \varepsilon_D^2 u$. Zu berechnen ist somit $\int_{\substack{u < v \leq \varepsilon_D^2 u \\ auv \leq N}} \frac{a}{\sqrt{D}} d(u,v)$.

Aus $u < v \leq \varepsilon_D^2 u$ folgt insbesondere $u < \varepsilon_D^2 u$, also $u > 0$. $auv \leq N$ und $0 < u < v$ implizieren $au^2 < N$, also $u < \sqrt{\frac{N}{a}}$. Ferner folgt $v \leq \frac{N}{au}$. Die Integrationsbedingungen lauten somit $0 < u < \sqrt{\frac{N}{a}}$ und $u < v \leq \min(\varepsilon_D^2 u, \frac{N}{au})$. $\varepsilon_D^2 u \leq \frac{N}{au}$ ist gleichwertig mit $u \leq \sqrt{\frac{N}{\varepsilon_D^2 a}}$. Setzt man $u_1 = \sqrt{\frac{N}{\varepsilon_D^2 a}}$ und $u_2 = \sqrt{\frac{N}{a}}$, so ist das Integral gleich

$$\int_0^{u_1}\int_u^{\varepsilon_D^2 u}\frac{a}{\sqrt{D}}\,dv\,du \;+\; \int_{u_1}^{u_2}\int_u^{\frac{N}{au}}\frac{a}{\sqrt{D}}\,dv\,du \;=\; \frac{a}{\sqrt{D}}\Big(\int_0^{u_1}(\varepsilon_D^2-1)u\,du \;+\; \int_{u_1}^{u_2}\Big(\frac{N}{au}-u\Big)du\Big) \;=$$

$$\frac{a}{\sqrt{D}}\Big((\varepsilon_D^2-1)\frac{u_1^2}{2} \;+\; \frac{N}{a}(\log u_2 - \log u_1) \;-\; (\varepsilon_D^2-1)\frac{u_1^2}{2}\Big) \;=\; \frac{N\log\varepsilon_D}{\sqrt{D}}\,.$$

Dies ist der gesuchte Flächeninhalt. Nun muß die Größenordnung des Gebietsumfangs ermittelt werden. Der Hyperbelbogen $v = \frac{N}{au}$, $u > 0$, schneidet die Gerade $v = \varepsilon_D^2 u$ an der Stelle u_1 und die Gerade $v = u$ an der Stelle u_2. Seine Länge zwischen den beiden Schnittpunkten ist gleich $\int_{u_1}^{u_2}\sqrt{1+\frac{N^2}{a^2u^4}}\,du$, liegt also zwischen $\int_{u_1}^{u_2}\sqrt{\frac{N^2}{a^2u^4}}\,du = \int_{u_1}^{u_2}\frac{N}{au^2}\,du$

$= \frac{N}{a}\Big(\frac{1}{u_1}-\frac{1}{u_2}\Big) = \frac{(\varepsilon_D-1)\sqrt{N}}{\sqrt{a}}$ und $\int_{u_1}^{u_2}\sqrt{\frac{N^2}{a^2u^4}+\frac{N^2}{a^2u^4}}\,du = \frac{(\varepsilon_D-1)\sqrt{2N}}{\sqrt{a}}$.

Die Strecke von $(0,0)$ nach $(u_1, \varepsilon_D^2 u_1)$ hat die Länge $\sqrt{u_1^2+\varepsilon_D^4 u_1^2} = \sqrt{\frac{N}{a}}\sqrt{\varepsilon_D^2+\frac{1}{\varepsilon_D^2}}$,

die Strecke von $(0,0)$ nach (u_2, u_2) hat die Länge $\sqrt{2u_2^2} = \sqrt{\frac{N}{a}}\sqrt{2}$.

Die obige Koordinatentransformation ist linear und von N unabhängig. Der Gebietsumfang ist also von der Größenordnung $\sqrt{N}$. Es folgt die Behauptung für positive Diskriminanten:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}R(n,f) \;=\; \lim_{N\to\infty}\frac{1}{N}\Big(\frac{N\log\varepsilon_D}{\sqrt{D}} + O(\sqrt{N})\Big) \;=\; \frac{\log\varepsilon_D}{\sqrt{D}}\,.$$

V.) Es ist noch zu zeigen, daß der Fall $E_D = \{1, -1\}$ nicht eintreten kann, d.h. daß U(f) nicht trivial ist (vgl. Schritt V.) in Satz 3.1). Dazu darf angenommen werden, daß f die Grundform f_D ist (siehe Definition 1.8), denn primitive Formen derselben Diskriminante haben nach Satz 3.1, Schritt III.) isomorphe Automorphismengruppen.

Annahme: $U(f) = \{\pm\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\}$. Dann ist $R(n,f) = |\{(x,y)\in \boldsymbol{Z}^2 \mid f(x,y)=n\,,\, x-\theta y>0\}|$ für jedes $n\in\boldsymbol{N}$. Es ist $D = D_0 r^2$ mit D_0 Fundamentaldiskriminante und $r\in\boldsymbol{N}$.

$$\sum_{\substack{n=1\\ ggT(n,r)=1}}^{N}R(n,f) \;=\; |\{(x,y)\in \boldsymbol{Z}\times\boldsymbol{Z} \mid 0< f(x,y)\le N\,,\, ggT(f(x,y),r)=1\,,\, x-\theta y>0\}|\,.$$

Es sei $y\in\boldsymbol{N}$ fest gewählt. Wieviele $x\in\boldsymbol{Z}$ gibt es mit $f(x,y)\le N$ und $x-\theta y>0$? $f(x,y)\le N$ ist wegen $a=1$ gleichwertig mit $(x+\frac{1}{2}by)^2 \le N+\frac{1}{4}Dy^2$, also mit $-\frac{1}{2}(by+\sqrt{4N+Dy^2}) \le x \le -\frac{1}{2}(by-\sqrt{4N+Dy^2})$. $\theta y > -\frac{1}{2}(by+\sqrt{4N+Dy^2})$, d.h. $f(x,y)\le N$, $x-\theta y>0$ ist äquivalent zu $-\frac{1}{2}(b-\sqrt{D})y < x \le -\frac{1}{2}(by-\sqrt{4N+Dy^2})$.

Für $\Delta \in \boldsymbol{R}$, $\Delta \geq 0$ ist $\sqrt{1+\Delta} \geq 1+\frac{1}{3}\Delta$ genau dann, wenn $\Delta \leq 3$ ist, und $\frac{4N}{Dy^2} \leq 3$ ist gleichbedeutend mit $y \geq \sqrt{\frac{4N}{3D}}$. Daher gilt für jedes feste $y \geq \sqrt{\frac{4N}{3D}}$:

$$-\tfrac{1}{2}(by-\sqrt{4N+Dy^2}) + \tfrac{1}{2}(b-\sqrt{D})y = \frac{\sqrt{D}y}{2}\left(\sqrt{1+\frac{4N}{Dy^2}}-1\right) \geq \frac{\sqrt{D}y}{2}\frac{4N}{3Dy^2} = \frac{2N}{3\sqrt{D}y}$$

. Es gibt also zumindest $[\frac{2N}{3\sqrt{D}y}]$ ganze Zahlen x mit $f(x,y) \leq N$ und $x - \theta y > 0$. Wegen $\theta > \theta'$ ist dann auch $x - \theta' y > 0$, also $f(x,y) = (x-\theta y)(x-\theta' y) > 0$.

Jetzt ist noch die Bedingung $ggT(f(x,y), r) = 1$ zu berücksichtigen. Es sei p ein Primteiler von r. Für $p \neq 2$ gilt: p teilt nicht f(x,y) $\Leftrightarrow$ p teilt nicht $(2x + by)^2 - Dy^2$ $\Leftrightarrow$ p teilt nicht $2x + by$. Da 2 modulo p invertierbar ist, ist bei vorgegebenem y für x genau ein Rest modulo p „verboten“, wenn f(x,y) nicht durch p teilbar sein soll. Dasselbe gilt für $p = 2$. Denn dann ist $D \equiv 0 \bmod 4$, also $b = 0$ nach Definition 1.8 , und folglich gilt: 2 teilt nicht f(x,y) $\Leftrightarrow$ 2 teilt nicht $x^2 + cy^2$ $\Leftrightarrow$ 2 teilt nicht $x + cy$.

Es seien nun p_1 , ... , p_k die verschiedenen Primteiler von r, und $s = p_1 \dots p_k$. Bei vorgegebenem y ist f(x,y) genau dann zu r teilerfremd, wenn x einen bestimmten Rest modulo p_i nicht läßt für $i = 1, \dots , k$. Nach dem chinesischen Restsatz kommt unter s aufeinanderfolgenden ganzen Zahlen jede „Restekombination“ modulo p_1 , ... , p_k genau einmal vor. Bei gegebenem y sind von den s möglichen Restekombinationen für x genau $(p_1 - 1) \dots (p_k - 1) = \varphi(s)$ Stück erlaubt, wenn f(x,y) zu r teilerfremd sein soll. Unter m aufeinanderfolgenden Zahlen haben mindestens $[\frac{m}{s}]\varphi(s)$ eine erlaubte Restekombination. Zu vorgegebenem $y \geq \sqrt{\frac{4N}{3D}}$ gibt es somit wenigstens $[\frac{2N}{3\sqrt{D}sy}]\varphi(s)$ ganze Zahlen x mit $0 < f(x,y) \leq N$, $ggT(f(x,y), r) = 1$ und $x - \theta y > 0$. $\frac{2N}{3\sqrt{D}sy} \geq 1$ ist gleichwertig mit $y \leq \frac{2N}{3\sqrt{D}s}$. Setze $y_1 = [\sqrt{\frac{4N}{3D}}] + 1 = O(\sqrt{N})$ und $y_2 = [\frac{2N}{3\sqrt{D}s}] = O(N)$.

$$\sum_{y=y_1}^{y_2}[\frac{2N}{3\sqrt{D}sy}]\varphi(s) = \sum_{y=y_1}^{y_2}(\frac{2N}{3\sqrt{D}sy}+O(1))\varphi(s) = \frac{2N\varphi(s)}{3\sqrt{D}s}\sum_{y=y_1}^{y_2}\frac{1}{y} + \varphi(s)\sum_{y=y_1}^{y_2}O(1) =$$

$O(N)(O(\log N) - O(\log \sqrt{N})) + O(N) - O(\sqrt{N}) = O(N \log N)$.

Es folgt $\lim_{N\to\infty} \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} R(n,f) \geq \lim_{N\to\infty} \frac{r}{N\varphi(r)} O(N \log N) = \lim_{N\to\infty} O(\log N) = \infty$

und daher erst recht $\lim_{N\to\infty} \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} GP(n,D) = \infty$ im Widerspruch zu Satz 5.4.

Die Annahme ist falsch. U(f) ist nicht trivial. Damit ist Satz 3.1 vollständig bewiesen. ♦

Satz 6.1 läßt sich leicht auf imprimitive Formen verallgemeinern:

Satz 6.2 *Es sei D kein Quadrat und f eine Form der Diskriminante D mit Koeffizienten-ggT s . Für $D < 0$ sei f außerdem positiv definit. Dann gilt:*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} R(n,f) \;=\; \left\{ \begin{array}{ll} \dfrac{2\pi}{w\sqrt{-D}} & \text{für } D<0 \\ \dfrac{\log \varepsilon_{D/s^2}}{\sqrt{D}} & \text{für } D>0 \end{array} \right\}.$$

Dabei ist $w=6$ für $D/s^2 = -3$, $w=4$ für $D/s^2 = -4$ und $w=2$ für $D/s^2 < -4$.

Beweis:

f stellt nur durch s teilbare ganze Zahlen dar, und für solche gilt R(n,f) = R(n/s, f/s) , denn die Gleichungen f(x,y) = n und f(x,y)/s = n/s haben dieselben Lösungen, und f und f /s haben dieselben Automorphismen. f /s ist eine primitive (und für D < 0 auch positiv definite) Form der Diskriminante D/s^2 , die mit D kein Quadrat ist, und genügt somit den Bedingungen von Satz 6.1.

$$\frac{1}{N}\sum_{n=1}^{N} R(n,f) \quad\le\quad \frac{1}{s[N/s]}\sum_{n=1}^{[N/s]} R(ns,f) \quad=\quad \frac{1}{s}\frac{1}{[N/s]}\sum_{n=1}^{[N/s]} R(n,f/s) \quad \xrightarrow[N\to\infty]{}$$

$$\left\{ \begin{array}{lll} \dfrac{1}{s}\dfrac{2\pi}{w\sqrt{-D/s^2}} & = & \dfrac{2\pi}{w\sqrt{-D}} \qquad \text{für } D<0 \\ \dfrac{1}{s}\dfrac{\log \varepsilon_{D/s^2}}{\sqrt{D/s^2}} & = & \dfrac{\log \varepsilon_{D/s^2}}{\sqrt{D}} \qquad \text{für } D>0 \end{array} \right\}.$$

Andererseits ist $\frac{1}{N}\sum_{n=1}^{N} R(n,f) \;\ge\; \frac{1}{s}\frac{1}{[N/s]+1}\sum_{n=1}^{[N/s]} R(ns,f) \;=\; \frac{1}{s}\frac{1}{[N/s]+1}\sum_{n=1}^{[N/s]+1} R(n,f/s)$

$- \;\dfrac{R([N/s]+1,f/s)}{s([N/s]+1)}$, und man hat nun für $N \to \infty$ dieselben Grenzwerte wie eben,

denn $\dfrac{R([N/s]+1,f/s)}{s([N/s]+1)} \;=\; \dfrac{O(\sqrt{N})}{O(N)}$ wie aus dem Beweis von 6.1 hervorgeht: Dort ist R(N,f) offenbar von der Größenordnung der Länge des Ellipsen- bzw. Hyperbelbogens. ♦

Satz 6.3 *Es sei D kein Quadrat und $D = D_0 r^2$ mit D_0 Fundamentaldiskriminante, $r \in \mathbf{N}$. Es sei f eine primitive, für $D < 0$ außerdem positiv definite, Form der Diskriminante D.*

Dann gilt: $$\lim_{N\to\infty} \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} R(n,f) \;=\; \left\{ \begin{array}{ll} \dfrac{2\pi}{w\sqrt{-D}} & \text{für } D<0 \\ \dfrac{\log \varepsilon_{D}}{\sqrt{D}} & \text{für } D>0 \end{array} \right\}.$$

Dabei ist $w=6$ für $D = -3$, $w=4$ für $D = -4$ und $w=2$ für $D < -4$.

Das bedeutet, daß Satz 6.1 gültig bleibt, wenn man den Mittelwert nur für die zu r teilerfremden natürlichen Zahlen bildet, denn in {1, ... , N} gibt es $\frac{N}{r}\varphi(r) + O(1)$ solche Zahlen.

Beweis: Der Beweis von Satz 6.1 ist derart zu modifizieren, daß man nicht mehr alle Gitterpunkte (x,y) in den jeweiligen Gebieten zählt, sondern nur noch diejenigen mit $ggT(f(x,y) , r) = 1$. Die folgenden Überlegungen ähneln daher Schritt V.) in Satz 6.1.

Es sei $f(x,y) = ax^2 + bxy + cy^2$ und p ein Primteiler von r. Weil f primitiv ist, sind a und c nicht beide durch p teilbar. Es sei p kein Teiler von a. Für $p \neq 2$ gilt: p teilt nicht f(x,y) $\Leftrightarrow$ p teilt nicht $(2ax + by)^2 - Dy^2$ $\Leftrightarrow$ p teilt nicht 2ax + by . Da 2a modulo p invertierbar ist, ist bei vorgegebenem y für x genau ein Rest modulo p „verboten“, wenn f(x,y) nicht durch p teilbar sein soll. Dasselbe gilt für $p = 2$: Dann ist $D \equiv 0 \bmod 4$, also $b \equiv 0 \bmod 2$ und somit gilt: 2 teilt nicht f(x,y) $\Leftrightarrow$ 2 teilt nicht $ax^2 + cy^2$ $\Leftrightarrow$ 2 teilt nicht x + cy. Wenn p kein Teiler von c ist, so ist analog bei vorgegebenem x für y genau ein Rest modulo p verboten. Daher sind von den p^2 möglichen Restekombinationen modulo p für x und y genau $p(p-1)$ erlaubt, wenn f(x,y) nicht durch p teilbar sein soll.

Es sei $r = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ mit $p_1 < \dots < p_k$ prim und $\alpha_1 , \dots , \alpha_k \in \boldsymbol{N}$. Sei $s = p_1 \dots p_k$. Betrachte das Gitterquadrat $Q = \{(x,y) \in \boldsymbol{Z} \times \boldsymbol{Z} \mid x_0 < x \leq x_0 + s , \; y_0 < y \leq y_0 + s\}$ der Seitenlänge s ($x_0 , y_0 \in \boldsymbol{R}$ fest). Wenn man die Reste modulo $p_1 , \dots , p_k$ sowohl für x als auch für y irgendwie vorschreibt, so gibt es nach dem chinesischen Restsatz in Q genau ein Element, dessen Komponenten die verlangten Reste lassen. Von den s^2 Elementen von Q erfüllen also nach der obigen Überlegung genau $p_1(p_1 - 1) \dots p_k(p_k - 1) = s\varphi(s)$ Stück die Bedingung $ggT(f(x,y) , r) = 1$.

Ein Gitterquadrat der Seitenlänge r ist die disjunkte Vereinigung von $(r/s)^2$ Quadraten der Seitenlänge s und enthält somit genau $(r/s)^2 \, s \, \varphi(s) = p_1^{2\alpha_1 - 1} \dots p_k^{2\alpha_k - 1}(p_1 - 1) \dots (p_k - 1) = r\varphi(r)$ Elemente (x,y) mit $ggT(f(x,y) , r) = 1$. Ein Gebiet des Flächeninhalts F in der Ebene setzt sich, bis auf einen Fehler von der Größenordnung seines Umfangs, aus genau F/r^2 disjunkten Quadraten der Seitenlänge r zusammen, enthält also, bis auf einen Fehler der genannten Art, genau $F \, \varphi(r)/r$ Gitterpunkte (x,y) mit $ggT(f(x,y) , r) = 1$. Es folgt:

$$\lim_{N \to \infty} \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} R(n,f) = \lim_{N \to \infty} \frac{r}{N\varphi(r)} \frac{1}{w} \left(\frac{2\pi N}{\sqrt{-D}} \frac{\varphi(r)}{r} + O(\sqrt{N}) \right) = \frac{2\pi}{w\sqrt{-D}} \quad \text{bzw.}$$

$$\lim_{N \to \infty} \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} R(n,f) = \lim_{N \to \infty} \frac{r}{N\varphi(r)} \left(\frac{N \log \varepsilon_D}{\sqrt{D}} \frac{\varphi(r)}{r} + O(\sqrt{N}) \right) = \frac{\log \varepsilon_D}{\sqrt{D}} \quad \text{für}$$

$D < 0$ bzw. $D > 0$. ♦

In Satz 6.3 ist die Voraussetzung der Primitivität von f wesentlich. Denn Schritt I.) in Satz 4.8 zeigt, daß eine imprimitive Form f der Diskriminante D keine zu r teilerfremden Zahlen darstellt, d.h. es ist $R(n,f) = 0$ für $ggT(n,r) = 1$.

Der Vollständigkeit halber sollen noch Formen mit quadratischer Diskriminante untersucht werden.

Ergänzung 6.4 *Es sei f eine Form der Diskriminante r^2 mit $r \in \boldsymbol{N}$. Dann gilt:* $\lim\limits_{N\to\infty} \frac{1}{N}\sum\limits_{n=1}^{N} R(n,f) = \infty$. *Für primitives f ist auch* $\lim\limits_{N\to\infty} \frac{r}{N\varphi(r)} \sum\limits_{\substack{n=1 \\ ggT(n,r)=1}}^{N} R(n,f) = \infty$.

Beweis: Nach Satz 2.2 repräsentieren die Formen $ax^2 + rxy$, $a \in \{1, \dots , r\}$, die Formenklassen der Diskriminante r^2 . Man darf also annehmen, daß f eine dieser Formen ist. Nach Satz 3.1 ist U(f) trivial. Daher ist $R(n,f) = |\{(x,y) \in \boldsymbol{Z}^2 \mid ax^2 + rxy = n\ , x > 0\}|$ für $n \in \boldsymbol{N}$, also $\sum\limits_{n=1}^{N} R(n,f) = |\{(x,y) \in \boldsymbol{Z}^2 \mid 0 < ax^2 + rxy \le N\ , x > 0\}|$.

Für positives x ist $0 < ax^2 + rxy \le N$ gleichwertig mit $-\frac{ax}{r} < y \le \frac{N}{rx} - \frac{ax}{r}$.

Zu festem $x > 0$ gibt es also $[\frac{N}{rx}]$ ganze Zahlen y mit $0 < ax^2 + rxy \le N$. Es folgt:

$$\frac{1}{N}\sum_{n=1}^{N} R(n,f) = \frac{1}{N}\sum_{x=1}^{[N/r]} [\frac{N}{rx}] = \frac{1}{N}\sum_{x=1}^{[N/r]} (\frac{N}{rx} + O(1)) = \frac{1}{r}\sum_{x=1}^{[N/r]} \frac{1}{x} + O(1) \xrightarrow[N\to\infty]{} \infty.$$

Ist f primitiv, d.h. $ggT(a,r) = 1$, so ist $ax^2 + rxy$ genau dann zu r teilerfremd, wenn x zu r teilerfremd ist. Betrachtet man also $\frac{r}{N\varphi(r)} \sum\limits_{\substack{n=1 \\ ggT(n,r)=1}}^{N} R(n,f)$, so hat man in den Überlegungen nur überall die Bedingung $ggT(x,r) = 1$ hinzuzunehmen. ♦

7 Die Klassenzahl

Durch Vergleich der Sätze 5.4 und 6.3 können nun Formeln für die Anzahl h(D) der primitiven bzw. primitiven und positiv definiten Formenklassen der Diskriminante D gewonnen werden.

Satz 7.1 *Es sei D eine Diskriminante. Dann gilt für die Klassenzahl h(D):*

$$h(D) \;=\; \begin{cases} \dfrac{w\sqrt{-D}}{2\pi} L(1,\chi_D) & \text{für } D < 0 \\ 1 & \text{für } D = 0 \\ \varphi(\sqrt{D}) & \text{für quadratisches } D > 0 \\ \dfrac{\sqrt{D}}{\log \varepsilon_D} L(1,\chi_D) & \text{für nichtquadratisches } D > 0 \end{cases} .$$

Dabei ist $w = 6$ *für* $D = -3$, $w = 4$ *für* $D = -4$ *und* $w = 2$ *für* $D < -4$.

Beweis:

Es sei D kein Quadrat und $D = D_0 r^2$ mit D_0 Fundamentaldiskriminante und $r \in \boldsymbol{N}$. $\{f_1, \ldots, f_{h(D)}\}$ sei ein Repräsentantensystem der primitiven, für negatives D außerdem positiv definiten, Formenklassen der Diskriminante D. Nach Satz 5.4 gilt: $L(1,\chi_D) = \lim\limits_{N\to\infty} \dfrac{r}{N\varphi(r)} \sum\limits_{\substack{n=1 \\ ggT(n,r)=1}}^{N} GP(n,D) = \lim\limits_{N\to\infty} \dfrac{r}{N\varphi(r)} \sum\limits_{\substack{n=1 \\ ggT(n,r)=1}}^{N} \sum\limits_{i=1}^{h(D)} R(n,f_i)$ (nach Definition 1.14)

$$= \sum_{i=1}^{h(D)} \lim_{N\to\infty} \frac{r}{N\varphi(r)} \sum_{\substack{n=1 \\ ggT(n,r)=1}}^{N} R(n,f_i) = \begin{cases} h(D)\dfrac{2\pi}{w\sqrt{-D}} & \text{für } D < 0 \\ h(D)\dfrac{\log \varepsilon_D}{\sqrt{D}} & \text{für } D > 0 \end{cases} \quad \text{nach Satz 6.3.}$$

Das ist die Behauptung für nichtquadratische Diskriminanten. Nach Satz 2.3 ist $h(0) = 1$ und nach Satz 2.2 ist $h(m^2) = \varphi(m)$ für $m \in \boldsymbol{N}$. ♦

Die obigen Klassenzahlformeln ermöglichen die numerische Berechnung von h(D), denn man weiß ja, daß h(D) ganzzahlig ist, und braucht deshalb nur ein hinreichend großes Anfangsstück der Reihe $L(1,\chi_D)$ zu nehmen. Mithilfe sog. Gaußscher Summen lassen sich die Werte $L(1,\chi_D)$ aber auch ausrechnen, so daß man h(D) geschlossen angeben kann. Dies wird hier nicht durchgeführt. Nachfolgend sind die Klassenzahlen betragsmäßig kleiner Diskriminanten aufgeführt.

D	−31	−28	−27	−24	−23	−20	−19	−16	−15	−12	−11	−8	−7	−4	−3
h(D)	3	1	1	2	3	2	1	1	2	1	1	1	1	1	1

D	0	1	4	5	8	9	12	13	16	17	20	21	24	25	28	29
h(D)	1	1	1	1	1	2	2	1	2	1	3	2	2	4	2	1

Das Ausgangsproblem war die Berechnung von R(n,f) für eine gegebene Form f der Diskriminante D und eine gegebene ganze Zahl n. Hat f den Koeffizienten-ggT t $\in \boldsymbol{N}$, so ist R(n,f) = 0, falls t kein Teiler von n ist, und R(n,f) = R(n/t , f /t) ansonsten. Man darf also annehmen, daß f primitiv ist. Ist D = 0 oder n = 0, so ist die Sache einfach (siehe die Ergänzungen 8.2 und 8.3). Es gelte also D ≠ 0 und n ≠ 0. Für D < 0 sei außerdem f positiv definit für positives und negativ definit für negatives n. Man berechne GP(n,D) mit Satz 8.1. Ist GP(n,D) = 0, so ist erst recht R(n,f) = 0. Ist h(D) = 1, was bei betragsmäßig kleinem D häufig der Fall sein wird, so ist R(n,f) = GP(n,D).

Man kann also mithilfe der hier bewiesenen Sätze R(n,f) in vielen Fällen berechnen. Allgemein ist das jedoch nur durch Algorithmen möglich, die Gegenstand der Reduktionstheorie sind. Eine geschlossene Formel für R(n,f) existiert nicht.

Die Gesamtzahl der Lösungen von f(x,y) = n ist einfach |U(f)| · R(n,f), wie aus dem Beweis von Satz 6.1 hervorgeht.

8 Die Gesamtdarstellungszahl (II)

Satz 8.1 *Es sei $D \neq 0$ eine Diskriminante, $D = D_0 r^2$ mit D_0 Fundamentaldiskriminante und $r \in \boldsymbol{N}$, und es sei $n \neq 0$ eine ganze Zahl. Dann gilt:*

Ist $ggT(n,r^2)$ kein Quadrat, so ist $GP(n,D) = 0$. Ist $ggT(n,r^2) = s^2$ mit $s \in \boldsymbol{N}$, $n = n's^2$ und $D = D's^2$, so ist $GP(n,D) = s \prod_{p \text{ prim}, p|s} (1 - \frac{\chi_{D'}(p)}{p}) \sum_{m \in N, m|n'} \chi_{D'}(m)$.

Der Satz liefert GP(n,D) für jede Diskriminante $D \neq 0$ und jede ganze Zahl $n \neq 0$. Das in der Formulierung auftauchende D' ist eine Diskriminante: Es ist $r^2 = s^2 t^2$ für ein $t \in \boldsymbol{N}$, somit $D = D_0 s^2 t^2 = D's^2$, also $D' = D_0 t^2 \equiv 0$ oder $1 \bmod 4$. $\chi_{D'}$ ist der von χ_{D_0} induzierte reelle Charakter modulo $|D'|$. Satz 4.8 ist als Spezialfall enthalten, aus $ggT(n,r) = 1$ folgt ja $ggT(n,r^2) = 1$, also $s = 1$, $n' = n$, $D' = D$.

Beweis: (F. Hirzebruch, D.B. Zagier, Inventiones Mathematicae 36 (1976), Seite 69–70, Proposition 2)

I.) Das erste Ziel ist, für die Größe $G^*(n,D) = G^*(n,D_0r^2)$ eine „simultane Multiplikativität“ in n und r nachzuweisen: Wenn $\{p_i \mid i \in \boldsymbol{N}\}$ die Menge aller Primzahlen ist und $n = \prod_{i=1}^{\infty} p_i^{\alpha_i}$ bzw. $r = \prod_{i=1}^{\infty} p_i^{\beta_i}$ die Primfaktorenzerlegung von $n \in \boldsymbol{N}$ bzw. r ist, dann gilt

$$G^*(n,D_0r^2) = G^*(\prod_{i=1}^{\infty} p_i^{\alpha_i}, D_0 \prod_{i=1}^{\infty} p_i^{2\beta_i}) = \prod_{i=1}^{\infty} G^*(p_i^{\alpha_i}, D_0 p_i^{2\beta_i}) .$$

Betrachte zum Beweis dieser Aussage irgendeine Primzahl p. p^β, $\beta \in \boldsymbol{N}_0$, sei die höchste Potenz, in der r von p geteilt wird. Dann ist $G^*(p^\alpha,D_0r^2) = G^*(p^\alpha, D_0p^{2\beta})$ für alle $\alpha \in \boldsymbol{N}_0$.

1. Fall: p ist ungerade. Es ist $G^*(p^\alpha,D) = |\{1 \le b \le p^\alpha \mid b^2 \equiv D \bmod p^\alpha\}|$ nach Teil II.) in Satz 4.6. Es ist $r = p^\beta t$ für ein $t \in \boldsymbol{N}$ mit $ggT(t , p^\alpha) = 1$. t ist modulo p^α multiplikativ invertierbar, also gibt es zu $b \in \{1, \dots , p^\alpha\}$ mit $b^2 \equiv D \bmod p^\alpha$ genau ein $a \in \{1, \dots , p^\alpha\}$ mit $at \equiv b \bmod p^\alpha$, folglich $a^2 t^2 \equiv b^2 \equiv D_0 p^{2\beta} t^2 \bmod p^\alpha$, also $a^2 \equiv D_0 p^{2\beta} \bmod p^\alpha$. Auf diese Weise erhält man eine Abbildung von der Menge $\{1 \le b \le p^\alpha \mid b^2 \equiv D \bmod p^\alpha\}$ in $\{1 \le a \le p^\alpha \mid a^2 \equiv D_0 p^{2\beta} \bmod p^\alpha\}$. Man sieht leicht, daß diese Abbildung bijektiv ist. Daraus folgt $G^*(p^\alpha,D) = G^*(p^\alpha , D_0 p^{2\beta})$.

2. Fall: $p = 2$. Nach Lemma 4.4 gilt $G^*(2^\alpha,D) = |\{1 \le b \le 2^{\alpha+1} \mid b^2 \equiv D \bmod 2^{\alpha+2}\}|$. Es ist $r = 2^\beta t$ für ein $t \in \boldsymbol{N}$ mit $ggT(t , 2^{\alpha+2}) = 1$. t ist modulo $2^{\alpha+2}$ invertierbar, daher existiert zu $b \in \{1, \dots , 2^{\alpha+1}\}$ mit $b^2 \equiv D \bmod 2^{\alpha+2}$ genau ein $c \in \{1, \dots , 2^{\alpha+2}\}$ mit $ct \equiv b \bmod 2^{\alpha+2}$, also $c^2 t^2 \equiv b^2 \equiv D_0 2^{2\beta} t^2 \bmod 2^{\alpha+2}$, also $c^2 \equiv D_0 2^{2\beta} \bmod 2^{\alpha+2}$. Für $c \le 2^{\alpha+1}$ setze $a = c$, für $c > 2^{\alpha+1}$ setze $a = c - 2^{\alpha+1}$. Es ist $(c - 2^{\alpha+1})^2 \equiv c^2 \bmod 2^{\alpha+2}$. Man erhält derart eine Abbildung von $\{1 \le b \le 2^{\alpha+1} \mid b^2 \equiv D \bmod 2^{\alpha+2}\}$ in

$\{1 \le a \le 2^{\alpha+1} \mid a^2 \equiv D_0 2^{2\beta} \bmod 2^{\alpha+2}\}$. Die Abbildung ist bijektiv: Ist $a \in \{1, \dots, 2^{\alpha+1}\}$ mit $a^2 \equiv D_0 2^{2\beta} \bmod 2^{\alpha+2}$ vorgegeben, so gilt $(at)^2 \equiv ((a + 2^{\alpha+1})\, t\,)^2 \equiv D \bmod 2^{\alpha+2}$, und die Zahlen at , $(a + 2^{\alpha+1})t$ sind zwar kongruent modulo $2^{\alpha+1}$, nicht aber modulo $2^{\alpha+2}$ denn t ist ungerade. Also ist genau eine der beiden Zahlen kongruent modulo $2^{\alpha+2}$ zu einem $b \in \{1, \dots, 2^{\alpha+1}\}$. Somit ist $G^*(2^\alpha, D) = G^*(2^\alpha, D_0 2^{2\beta})$.

Nun: $G^*(n, D_0 r^2) = G^*(\prod_{i=1}^{\infty} p_i^{\alpha_i}, D_0 r^2) = \prod_{i=1}^{\infty} G^*(p_i^{\alpha_i}, D_0 r^2) = \prod_{i=1}^{\infty} G^*(p_i^{\alpha_i}, D_0 p_i^{2\beta_i})$

denn nach Lemma 4.5 ist $G^*(\cdot, D)$ multiplikativ.

II.) Diese simultane Multiplikativität überträgt sich auf $G(n, D_0 r^2)$ und $GP(n, D_0 r^2)$.

$G(n, D_0 r^2) = G(\prod_{i=1}^{\infty} p_i^{\alpha_i}, D_0 r^2) = \prod_{i=1}^{\infty} G(p_i^{\alpha_i}, D_0 r^2)$ (nach Lemma 4.5) =

$\prod_{i=0}^{\infty} \sum_{t^2 \mid p_i^{\alpha_i}} G^*(\frac{p_i^{\alpha_i}}{t^2}, D_0 r^2)$ (nach Lemma 4.3) $= \prod_{i=0}^{\infty} \sum_{t^2 \mid p_i^{\alpha_i}} G^*(\frac{p_i^{\alpha_i}}{t^2}, D_0 p_i^{2\beta_i})$ (nach I.)) =

$\prod_{i=1}^{\infty} G(p_i^{\alpha_i}, D_0 p_i^{2\beta_i})$ (nach Lemma 4.3).

Für $GP(n, D_0 r^2)$ ist die Sache schwieriger. Es sei zunächst allgemeiner $n \in \boldsymbol{Z} \setminus \{0\}$. Es gilt

$G(n, D_0 r^2) = \sum_{t^2 \mid D_0 r^2} GP(\frac{n}{t}, \frac{D_0 r^2}{t^2})$ (siehe Definition 1.14) $= \sum_{t \mid ggT(n,r)} GP(\frac{n}{t}, \frac{D_0 r^2}{t^2})$,

denn $\frac{D_0 r^2}{t^2}$ ist nur genau dann eine Diskriminante, wenn t ein Teiler von r ist, und $\frac{n}{t}$ ist nur genau dann eine ganze Zahl, wenn t ein Teiler von n ist.

Es sei $\mu : \boldsymbol{N} \to \{0, 1, -1\}$ die Möbius-Funktion, d.h. $\mu(t) = 0$, falls t nicht quadratfrei ist, und $\mu(t) = (-1)^k$, falls t quadratfrei und k die Anzahl der Primteiler von t ist.

$\sum_{t \mid ggT(n,r)} \mu(t)\, G(\frac{n}{t}, D_0 \frac{r^2}{t^2}) = \sum_{t \mid n,t \mid r} \mu(t) \sum_{u \mid ggT(n/t,\, r/t)} GP(\frac{n}{tu}, D_0 \frac{r^2}{t^2 u^2}) =$

$\sum_{t \mid n,t \mid r} \sum_{u \mid n/t,\, u \mid r/t} \mu(t)\, GP(\frac{n}{tu}, D_0 \frac{r^2}{t^2 u^2}) = \sum_{d \mid n,\, d \mid r} (\sum_{tu = d} \mu(t))\, GP(\frac{n}{d}, D_0 \frac{r^2}{d^2}) =$

$\sum_{d \mid ggT(n,r)} (\sum_{t \mid d} \mu(t))\, GP(\frac{n}{d}, D_0 \frac{r^2}{d^2}) = GP(n, D_0 r^2)$, denn nach einem bekannten Satz ist

$\sum_{t \mid d} \mu(t) = 0$ für $d > 1$. Also ist $GP(n, D_0 r^2) = \sum_{t \mid ggT(n,r)} \mu(t)\, G(\frac{n}{t}, D_0 \frac{r^2}{t^2})$ für jedes $n \neq 0$.
Das ist eine Art Möbiussche Umkehrformel.

Aus ihr folgt zunächst $GP(-n, D) = GP(n, D)$ denn nach der dem Lemma 4.4 folgenden Bemerkung ist $G(-n, D) = G(n, D)$.

Betrachte nun wieder $n = \prod_{i=1}^{\infty} p_i^{\alpha_i} \in \boldsymbol{N}$, $r = \prod_{i=1}^{\infty} p_i^{\beta_i}$. Es seien $p_1, \dots, p_m$, $m \in \boldsymbol{N}_0$, diejenigen Primzahlen, die sowohl n als auch r teilen.

$$\prod_{i=1}^{\infty} GP(p_i^{\alpha_i}, D_0 p_i^{2\beta_i}) = \prod_{i=1}^{\infty} \sum_{t \mid ggT(p_i^{\alpha_i}, p_i^{\beta_i})} \mu(t)\, G(\frac{p_i^{\alpha_i}}{t}, D_0 \frac{p_i^{2\beta_i}}{t^2}) =$$

$$\prod_{i=1}^{m} (G(p_i^{\alpha_i}, D_0 p_i^{2\beta_i}) - G(p_i^{\alpha_i - 1}, D_0 p_i^{2(\beta_i - 1)})) \prod_{i=m+1}^{\infty} G(p_i^{\alpha_i}, D_0 p_i^{2\beta_i}) =$$

$$\sum_{h:\{1,\dots,m\} \to \{0,1\}} (-1)^{h(1)+\dots+h(m)}\, G(\frac{n}{p_1^{h(1)} \cdots p_m^{h(m)}}, D_0 \frac{r^2}{p_1^{2h(1)} \cdots p_m^{2h(m)}})$$

(denn $G(n,D_0r^2)$ ist simultan multiplikativ) $= \sum_{t \mid ggT(n,r)} \mu(t)\, G(\frac{n}{t}, D_0 \frac{r^2}{t^2}) = GP(n,D_0r^2)$.

Damit ist die simultane Multiplikativität in n und r auch für $GP(n,D_0r^2)$ gezeigt.

III.) Um $GP(n,D_0r^2)$ für jedes $r \in \boldsymbol{N}$ und jedes $n \in \boldsymbol{Z}\backslash\{0\}$ ausrechnen zu können, braucht man also nur $GP(p^{\alpha}, D_0 p^{2\beta})$ für jede Primzahl p und alle $\alpha,\beta \in \boldsymbol{N}_0$ zu kennen. Um diese Größen zu ermitteln, werden erzeugende Funktionen eingeführt:

Es seien x,y unabhängige Unbestimmte über $\boldsymbol{R}$ oder $\boldsymbol{C}$ und p eine Primzahl. Setze

$$FP_{D_0,p}(x,y) = \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} GP(p^{\alpha}, D_0 p^{2\beta})\, x^{\alpha} y^{\beta} \quad \text{und}$$

$$F^*_{D_0,p}(x,y) = \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} G^*(p^{\alpha}, D_0 p^{2\beta})\, x^{\alpha} y^{\beta}.$$

Der zentrale Schritt im Beweis ist die Berechnung von $FP_{D_0,p}(x,y)$. Das ist nicht einfach und geschieht dadurch, daß $F^*_{D_0,p}(x,y)$ berechnet und zu $FP_{D_0,p}(x,y)$ in Beziehung gesetzt wird. Man könnte mit diesen Potenzreihen ganz formal operieren. Das will ich nicht tun, sondern sie als Objekte der Analysis auffassen. Zunächst werden daher Konvergenzüberlegungen angestellt.

Betrachte für festes $\alpha \in \boldsymbol{N}_0$ die Reihe $\sum_{\beta=0}^{\infty} GP(p^{\alpha}, D_0 p^{2\beta})\, y^{\beta}$. Nach Lemma 4.4 ist $G^*(p^{\alpha}, D_0 p^{2\beta}) = |\{1 \le b \le 2p^{\alpha} \mid b^2 \equiv D_0 p^{2\beta} \bmod 4p^{\alpha}\}| \le 2p^{\alpha}$ und somit

$$GP(p^{\alpha}, D_0 p^{2\beta}) \le G(p^{\alpha}, D_0 p^{2\beta}) = \sum_{t^2 \mid p^{\alpha}} G^*(\frac{p^{\alpha}}{t^2}, D_0 p^{2\beta}) \le \sum_{t^2 \mid p^{\alpha}} 2p^{\alpha} \le (\alpha+2)p^{\alpha}.$$

$\sqrt[\beta]{(\alpha+2)p^{\alpha}} \xrightarrow[\beta \to \infty]{} 1$, also ist die Reihe $\sum_{\beta=0}^{\infty} GP(p^{\alpha}, D_0 p^{2\beta})\, y^{\beta}$ für $|y| < 1$ absolut konvergent. Es sei sogar $|y| < 1/2$.

Dann ist $|\sum_{\beta=0}^{\infty} GP(p^\alpha, D_0 p^{2\beta})\, y^\beta| \;\le\; \sum_{\beta=0}^{\infty} (\alpha+2)p^\alpha (1/2)^\beta \;=\; 2(\alpha+2)p^\alpha$. Es folgt

$$\limsup_{\alpha=0}^{\infty} \left|\sum_{\beta=0}^{\infty} GP(p^\alpha, D_0 p^{2\beta})\, y^\beta\right|^{\frac{1}{\alpha}} \;\le\; \lim_{\alpha\to\infty} \sqrt[\alpha]{2(\alpha+2)p^\alpha} \;=\; p\,.$$

Somit ist die Doppelreihe $\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} GP(p^\alpha, D_0 p^{2\beta}) y^\beta\, x^\alpha$ zumindest für $|x|, |y| < 1/p$ absolut konvergent. Für solche Argumente darf man daher ohne Bedenken die Summationsreihenfolge ändern, Produkte mit anderen derartigen Reihen durch Ausmultiplizieren bilden etc. Dasselbe gilt ersichtlich für $\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} G^*(p^\alpha, D_0 p^{2\beta}) y^\beta\, x^\alpha$.

IV.) Es soll die Funktion $F^*_{D_0,\,p}(x,y)$ ausgerechnet werden. Zu diesem Zweck müssen die Größen $G^*(p^\alpha, D_0 p^{2\beta}) = |\{1 \le b \le 2p^\alpha \mid b^2 \equiv D_0 p^{2\beta} \bmod 4p^\alpha\}|$ untersucht werden.

Es sei $\alpha \ge 2$, $\beta \ge 1$ und $k \in \boldsymbol{Z}$. Dann ist durch $b \to bp + 2kp^{\alpha-1}$ eine Bijektion von $\{1 \le b \le 2p^{\alpha-2} \mid b^2 \equiv D_0 p^{2(\beta-1)} \bmod 4p^{\alpha-2}\}$ auf $\{2kp^{\alpha-1}+1 \le a \le 2(k+1)p^{\alpha-1} \mid a^2 \equiv D_0 p^{2\beta} \bmod 4p^\alpha\}$ gegeben. Denn erstens ist $1 \le b \le 2p^{\alpha-2}$ äquivalent zu $2kp^{\alpha-1}+1 \le bp + 2kp^{\alpha-1} \le 2kp^{\alpha-1} + 2p^{\alpha-1}$. Zweitens ist für $a \in \{2kp^{\alpha-1}+1, \ldots, 2(k+1)p^{\alpha-1}\}$ mit $a^2 \equiv D_0 p^{2\beta} \bmod 4p^\alpha$ p ein Teiler von a, so daß a bei Division durch $2p^{\alpha-1}$ einen durch p teilbaren Rest läßt und es genau ein $b \in \{1, \ldots, 2p^{\alpha-2}\}$ mit $a = bp + 2kp^{\alpha-1}$ gibt. Drittens gilt: $b^2 \equiv D_0 p^{2(\beta-1)} \bmod 4p^{\alpha-2} \;\Leftrightarrow\; 4p^\alpha \mid (b^2p^2 - D_0 p^{2\beta} + 4bkp^\alpha + 4k^2p^{2(\alpha-1)})$ $\Leftrightarrow\; (bp + 2kp^{\alpha-1})^2 \equiv D_0 p^{2\beta} \bmod 4p^\alpha$. Nun ist $\{1 \le b \le 2p^\alpha \mid b^2 \equiv D_0 p^{2\beta} \bmod 4p^\alpha\} = \bigcup_{k=0}^{p-1} \{2kp^{\alpha-1}+1 \le a \le 2(k+1)p^{\alpha-1} \mid a^2 \equiv D_0 p^{2\beta} \bmod 4p^\alpha\}$ disjunkte Vereinigung, also ist

$G^*(p^\alpha, D_0 p^{2\beta}) = p\, G^*(p^{\alpha-2}, D_0 p^{2(\beta-1)})$ für $\alpha \ge 2$, $\beta \ge 1$.

Nun sollen die Randfälle $\alpha = 0$, $\alpha = 1$ und $\beta = 0$ untersucht werden.

$G^*(1, D_0 p^{2\beta}) = |\{1 \le b \le 2 \mid b^2 \equiv D_0 p^{2\beta} \bmod 4\}| = 1$. Nach Schritt II.) in Satz 4.6 ist

$G^*(p, D_0 p^{2\beta}) = |\{1 \le b \le p \mid b^2 \equiv D_0 p^{2\beta} \bmod p\}| = 1 + \left(\frac{D_0 p^{2\beta}}{p}\right) = 1$ für $p \ne 2$, $\beta \ne 0$.

$G^*(2, D_0 2^{2\beta}) = |\{1 \le b \le 4 \mid b^2 \equiv D_0 2^{2\beta} \bmod 8\}| = 1$ für $\beta \ne 0$.

$$G^*(p^\alpha, D_0) = \sum_{t^2 \mid p^\alpha} G^*\left(\frac{p^\alpha}{t^2}, D_0\right) - \sum_{t^2 \mid p^{\alpha-2}} G^*\left(\frac{p^{\alpha-2}}{t^2}, D_0\right) = G(p^\alpha, D_0) - G(p^{\alpha-2}, D_0)$$

$$= \sum_{m \mid p^\alpha} \chi_{D_0}(m) - \sum_{m \mid p^{\alpha-2}} \chi_{D_0}(m) = \chi_{D_0}(p^\alpha) + \chi_{D_0}(p^{\alpha-1}) \text{ für } \alpha \ge 1 \text{ , unter Verwendung}$$

von Lemma 4.3 und Satz 4.6.

V.) $F^*_{D_0,p}(x,y) = \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} G^*(p^\alpha, D_0p^{2\beta})\, x^\alpha y^\beta = \sum_{\alpha=2}^{\infty}\sum_{\beta=1}^{\infty} G^*(p^\alpha, D_0p^{2\beta})\, x^\alpha y^\beta +$

$\sum_{\beta=0}^{\infty} G^*(1, D_0p^{2\beta})\, y^\beta + \sum_{\beta=1}^{\infty} G^*(p, D_0p^{2\beta})\, x\, y^\beta + \sum_{\alpha=1}^{\infty} G^*(p^\alpha, D_0)\, x^\alpha =$

$\sum_{\alpha=2}^{\infty}\sum_{\beta=1}^{\infty} p\, G^*(p^{\alpha-2}, D_0p^{2(\beta-1)})\, x^\alpha y^\beta + \sum_{\beta=0}^{\infty} y^\beta + \sum_{\beta=1}^{\infty} x\, y^\beta + \sum_{\alpha=1}^{\infty}(\chi_{D_0}(p^\alpha) + \chi_{D_0}(p^{\alpha-1}))\, x^\alpha$

$= px^2y \sum_{\alpha=2}^{\infty}\sum_{\beta=1}^{\infty} G^*(p^{\alpha-2}, D_0p^{2(\beta-1)})\, x^{\alpha-2}y^{\beta-1} + \frac{1}{1-y} + \frac{xy}{1-y} + \sum_{\alpha=1}^{\infty}(\chi_{D_0}(p))^\alpha\, x^\alpha +$

$\sum_{\alpha=1}^{\infty}(\chi_{D_0}(p))^{\alpha-1}x^\alpha = px^2y\, F^*_{D_0,p}(x,y) + \frac{1+xy}{1-y} + \frac{\chi_{D_0}(p)\, x}{1-\chi_{D_0}(p)\, x} + \frac{x}{1-\chi_{D_0}(p)\, x} =$

$px^2y\, F^*_{D_0,p}(x,y) + \frac{1+xy}{1-y} + \frac{(1+\chi_{D_0}(p))\, x}{1-\chi_{D_0}(p)\, x}$ für $|x|, |y| < 1/p$. Somit ist

$$F^*_{D_0,p}(x,y) = \frac{(1+xy)(1-\chi_{D_0}(p)x) + (1+\chi_{D_0}(p))\,x(1-y)}{(1-y)(1-\chi_{D_0}(p)x)(1-px^2y)} \quad \text{und schließlich}$$

$$F^*_{D_0,p}(x,y) = \frac{(1+x)(1-\chi_{D_0}(p)xy)}{(1-y)(1-px^2y)(1-\chi_{D_0}(p)x)} \quad \text{für } |x|, |y| < 1/p .$$

Die Funktionen $FP_{D_0,p}(x,y)$ und $F^*_{D_0,p}(x,y)$ hängen wie folgt zusammen:

$\frac{1}{1-xy} FP_{D_0,p}(x,y) = \sum_{\gamma=0}^{\infty} x^\gamma y^\gamma \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} GP(p^\alpha, D_0p^{2\beta})\, x^\alpha y^\beta =$

$\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\sum_{\gamma=0}^{\infty} GP(p^\alpha, D_0p^{2\beta})\, x^{\alpha+\gamma}y^{\beta+\gamma} = \sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}\Big(\sum_{\substack{\alpha+\gamma=\mu\\ \beta+\gamma=\nu}} GP(p^\alpha, D_0p^{2\beta})\Big)\, x^\mu y^\nu =$

$\sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}\Big(\sum_{\gamma=0}^{\min(\mu,\nu)} GP(p^{\mu-\gamma}, D_0p^{2(\nu-\gamma)})\Big)\, x^\mu y^\nu = \sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}\Big(\sum_{t \mid ggT(p^\mu, p^\nu)} GP(\frac{p^\mu}{t}, D_0\frac{p^{2\nu}}{t^2})\Big)\, x^\mu y^\nu =$

$\sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}\Big(\sum_{t^2 \mid p^\mu} G^*(\frac{p^\mu}{t^2}, D_0p^{2\nu})\Big)\, x^\mu y^\nu$ (denn nach Schritt II.) gilt $\sum_{t \mid ggT(n,r)} GP(\frac{n}{t}, \frac{D_0r^2}{t^2}) =$

$G(n,D_0r^2)$, und nach Lemma 4.3 ist $G(n,D_0r^2) = \sum_{t^2 \mid n} G^*(\frac{n}{t^2}, D_0r^2)$ für $n \neq 0$) =

$\sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}\Big(\sum_{0 \le \gamma \le \frac{\mu}{2}} G^*(p^{\mu-2\gamma}, D_0p^{2\nu})\Big)\, x^\mu y^\nu = \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\sum_{\gamma=0}^{\infty} G^*(p^\alpha, D_0p^{2\beta})\, x^{\alpha+2\gamma}y^\beta =$

$\sum_{\gamma=0}^{\infty} x^{2\gamma} \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} G^*(p^\alpha, D_0p^{2\beta})\, x^\alpha y^\beta = \frac{1}{1-x^2} F^*_{D_0,p}(x,y)$. Man erhält

$$FP_{D_0,p}(x,y) = \frac{(1-xy)(1-\chi_{D_0}(p)xy)}{(1-x)(1-y)(1-px^2y)(1-\chi_{D_0}(p)x)} \quad \text{für } |x|, |y| < 1/p .$$

Diese Identität ist der Angelpunkt des Beweises. Entwickelt man die rechte Seite in eine Potenzreihe in x und y, so kann man durch Koeffizientenvergleich die gesuchten Größen $GP(p^\alpha, D_0\, p^{2\beta})$ bestimmen.

VI.) $\frac{1}{1-x}\,\frac{1}{1-y} = \sum_{\kappa=0}^{\infty}\sum_{\lambda=0}^{\infty} x^{\kappa}y^{\lambda}$ für $|x|, |y| < 1$. $\frac{1}{1-px^2y}\,\frac{1}{1-\chi_{D_0}(p)x} =$

$\sum_{\mu=0}^{\infty}(px^2y)^{\mu}\sum_{\upsilon=0}^{\infty}(\chi_{D_0}(p)x)^{\nu} = \sum_{\mu=0}^{\infty}\sum_{\nu=2\mu}^{\infty} p^{\mu}(\chi_{D_0}(p))^{\nu-2\mu}x^{\nu}y^{\mu}$ für $|x|, |y| < 1/p$.

$\sum_{\kappa=0}^{\infty}\sum_{\lambda=0}^{\infty} x^{\kappa}y^{\lambda}\sum_{\mu=0}^{\infty}\sum_{\nu=2\mu}^{\infty} p^{\mu}(\chi_{D_0}(p))^{\nu-2\mu}x^{\nu}y^{\mu} = \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\Big(\sum_{\substack{\kappa+\nu=\alpha\\ \lambda+\mu=\beta\\ \nu\ge 2\mu}} p^{\mu}(\chi_{D_0}(p))^{\nu-2\mu}\Big)x^{\alpha}y^{\beta} =$

$\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\Big(\sum_{\mu=0}^{\beta}\sum_{\nu=0}^{\alpha-2\mu} p^{\mu}(\chi_{D_0}(p))^{\nu}\Big)x^{\alpha}y^{\beta}$ für $|x|, |y| < 1/p$. Nun:

$\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} GP(p^{\alpha}, D_0p^{2\beta})x^{\alpha}y^{\beta} = FP_{D_0,p}(x,y) = \frac{(1-xy)(1-\chi_{D_0}(p)xy)}{(1-x)(1-y)(1-px^2y)(1-\chi_{D_0}(p)x)}$

$= (1-xy)(1-\chi_{D_0}(p)xy)\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\Big(\sum_{\mu=0}^{\beta}\sum_{\nu=0}^{\alpha-2\mu} p^{\mu}(\chi_{D_0}(p))^{\nu}\Big)x^{\alpha}y^{\beta} =$

$(1 - (1+\chi_{D_0}(p))xy + \chi_{D_0}(p)x^2y^2)\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\Big(\sum_{\mu=0}^{\beta}\sum_{\nu=0}^{\alpha-2\mu} p^{\mu}(\chi_{D_0}(p))^{\nu}\Big)x^{\alpha}y^{\beta} = \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\Big($

$\sum_{\mu=0}^{\beta}\sum_{\nu=0}^{\alpha-2\mu} p^{\mu}(\chi_{D_0}(p))^{\nu} - \sum_{\mu=0}^{\beta-1}\sum_{\nu=0}^{\alpha-2\mu-1} p^{\mu}(\chi_{D_0}(p))^{\nu} - \sum_{\mu=0}^{\beta-1}\sum_{\nu=1}^{\alpha-2\mu} p^{\mu}(\chi_{D_0}(p))^{\nu} +$

$\sum_{\mu=0}^{\beta-2}\sum_{\nu=1}^{\alpha-2\mu-1} p^{\mu}(\chi_{D_0}(p))^{\nu}\Big)x^{\alpha}y^{\beta} = \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\Big(\sum_{\nu\le\alpha-2\beta} p^{\beta}(\chi_{D_0}(p))^{\nu} + \sum_{\substack{\nu\le\beta-1\\ 2\nu\le\alpha}} p^{\nu}(\chi_{D_0}(p))^{\alpha-2\nu} -$

$\sum_{\nu\le\alpha-2\beta+1} p^{\beta-1}(\chi_{D_0}(p))^{\nu+1} - \sum_{\substack{\nu\le\beta-2\\ 2\nu\le\alpha-1}} p^{\nu}(\chi_{D_0}(p))^{\alpha-2\nu}\Big)x^{\alpha}y^{\beta}$ für $|x|, |y| < 1/p$, also

$GP(p^{\alpha}, D_0p^{2\beta}) = \sum_{\nu\le\alpha-2\beta} p^{\beta}(\chi_{D_0}(p))^{\nu} + \sum_{\substack{\nu\le\beta-1\\ 2\nu\le\alpha}} p^{\nu}(\chi_{D_0}(p))^{\alpha-2\nu} - \sum_{\nu\le\alpha-2\beta+1} p^{\beta-1}(\chi_{D_0}(p))^{\nu+1}$

$- \sum_{\substack{\nu\le\beta-2\\ 2\nu\le\alpha-1}} p^{\nu}(\chi_{D_0}(p))^{\alpha-2\nu}$ für alle $\alpha,\beta \in \boldsymbol{N}_0$. Dabei laufen alle Summen von Null an.

1. Fall: $\alpha < 2\beta$, α gerade. Dann sind die erste und die dritte Summe leer, die zweite Summe läuft bis $\alpha/2$ und die vierte bis $\alpha/2 - 1$, so daß sich diese beiden Summen bis auf den Term $p^{\alpha/2}$ wegheben. $GP(p^{\alpha}, D_0p^{2\beta}) = p^{\alpha/2}$.

2. Fall: $\alpha < 2\beta - 1$, α ungerade. Die erste und die dritte Summe sind leer, die zweite und die vierte heben sich weg, denn beide laufen bis $(\alpha - 1)/2$. $GP(p^{\alpha}, D_0p^{2\beta}) = 0$.

3. Fall: $\alpha = 2\beta - 1$. Die erste Summe ist leer, die dritte ist gleich $-p^{\beta-1}\chi_{D_0}(p)$. Die zweite Summe läuft bis $\beta - 1$, die vierte bis $\beta - 2$, so daß die beiden sich wegheben bis auf den Term $p^{\beta-1}\chi_{D_0}(p)$. $GP(p^{\alpha}, D_0p^{2\beta}) = 0$.

4. Fall: $\beta = 0$. Die dritte Summe ist nicht vorhanden, die zweite und die vierte Summe sind leer. $GP(p^{\alpha}, D_0p^{2\beta}) = \sum_{\nu=0}^{\alpha}(\chi_{D_0}(p))^{\nu}$.

5. Fall: $\alpha \geq 2\beta > 0$. Die zweite Summe läuft bis $\beta - 1$, die vierte bis $\beta - 2$, sie heben sich also weg bis auf den Term $p^{\beta-1}(\chi_{D_0}(p))^{\alpha-2(\beta-1)}$.

$$GP(p^\alpha, D_0 p^{2\beta}) = \sum_{\nu=0}^{\alpha-2\beta} p^\beta(\chi_{D_0}(p))^\nu + p^{\beta-1}(\chi_{D_0}(p))^{\alpha-2(\beta-1)} - \sum_{\nu=0}^{\alpha-2\beta+1} p^{\beta-1}(\chi_{D_0}(p))^{\nu+1} =$$

$$(p^\beta - \chi_{D_0}(p)p^{\beta-1}) \sum_{\nu=0}^{\alpha-2\beta}(\chi_{D_0}(p))^\nu = p^\beta(1 - \frac{\chi_{D_0}(p)}{p}) \sum_{\nu=0}^{\alpha-2\beta}(\chi_{D_0}(p))^\nu .$$

Resultat: $$GP(p^\alpha, D_0 p^{2\beta}) = \begin{cases} p^{\alpha/2} & \text{für } \alpha < 2\beta, \alpha \text{ gerade} \\ 0 & \text{für } \alpha < 2\beta, \alpha \text{ ungerade} \\ \sum_{\nu=0}^{\alpha}(\chi_{D_0}(p))^\nu & \text{für } \beta = 0 \\ p^\beta(1-\frac{\chi_{D_0}(p)}{p}) \sum_{\nu=0}^{\alpha-2\beta}(\chi_{D_0}(p))^\nu & \text{für } \alpha \geq 2\beta > 0 \end{cases} .$$

VII.) Der Satz kann nun bewiesen werden. Wie vorhin sei $n = \prod_{i=1}^{\infty} p_i^{\alpha_i}$ und $r = \prod_{i=1}^{\infty} p_i^{\beta_i}$.

Nach Abschnitt II.) ist $GP(n, D_0 r^2) = \prod_{i=1}^{\infty} GP(p_i^{\alpha_i}, D_0 p_i^{2\beta_i})$.

Ist $ggT(n,r^2)$ kein Quadrat, so gibt es ein $i \in \boldsymbol{N}$ so, daß $\min(\alpha_i , 2\beta_i)$ ungerade ist, d.h. α_i ist kleiner als $2\beta_i$ und ungerade. Es folgt $GP(p_i^{\alpha_i}, D_0 p_i^{2\beta_i}) = 0$ und $GP(n, D_0 r^2) = 0$.

Es sei $ggT(n,r^2) = s^2$, $s \in \boldsymbol{N}$. Dann ist $GP(n, D_0 r^2) = \prod_{\alpha_i < 2\beta_i} p_i^{\alpha_i/2} \prod_{\beta_i=0}^{\alpha_i} \sum_{\nu=0} (\chi_{D_0}(p_i))^\nu$

$$\prod_{\alpha_i \geq 2\beta_i > 0} p_i^{\beta_i}(1 - \frac{\chi_{D_0}(p_i)}{p_i}) \sum_{\nu=0}^{\alpha_i-2\beta_i}(\chi_{D_0}(p_i))^\nu = s \prod_{\alpha_i \geq 2\beta_i > 0}(1 - \frac{\chi_{D_0}(p_i)}{p_i}) \prod_{\alpha_i \geq 2\beta_i} \sum_{\nu=0}^{\alpha_i-2\beta_i}(\chi_{D_0}(p_i))^\nu .$$

Es sei $n = n's^2$ und $D_0 r^2 = D's^2$. $\prod_{\alpha_i \geq 2\beta_i} \sum_{\nu=0}^{\alpha_i-2\beta_i}(\chi_{D_0}(p_i))^\nu = \prod_{\alpha_i > 2\beta_i} \sum_{\nu=0}^{\alpha_i-2\beta_i}(\chi_{D_0}(p_i))^\nu =$

$\prod_{p_i | n'} \sum_{\nu=0}^{\alpha_i-2\beta_i} \chi_{D_0}(p_i^\nu) = \prod_{p_i | n'} \sum_{m | p_i^{\alpha_i-2\beta_i}} \chi_{D_0}(m) = \prod_{p_i | n'} (\chi_{D_0} * I)(p_i^{\alpha_i-2\beta_i})$ (vgl. Abschnitt I.) in Satz 4.6) $= (\chi_{D_0} * I)(\prod_{p_i | n'} p_i^{\alpha_i-2\beta_i}) = (\chi_{D_0} * I)(n') = \sum_{m | n'} \chi_{D_0}(m) = \sum_{m | n'} \chi_{D'}(m)$,

denn ist m ein Teiler von n', so ist m zu r^2/s^2 teilerfremd, so daß $\chi_{D_0}(m) = \chi_{D'}(m)$ ist. $\alpha_i \geq 2\beta_i > 0$ ist gleichwertig damit, daß p_i ein Teiler von s, aber kein Teiler von r^2/s^2 ist. Wenn eine Primzahl p ein Teiler von r^2 / s^2 ist, so ist $\chi_{D'}(p) = 0$, andernfalls hat man $\chi_{D'}(p) = \chi_{D_0}(p)$. Also ist $\prod_{\alpha_i \geq 2\beta_i > 0}(1 - \frac{\chi_{D_0}(p_i)}{p_i}) = \prod_{p_i | s}(1 - \frac{\chi_{D'}(p_i)}{p_i})$ und insgesamt

$GP(n,D_0r^2) = s \prod_{\substack{p \text{ prim} \\ p | s}}(1 - \frac{\chi_{D'}(p)}{p}) \sum_{m | n'} \chi_{D'}(m)$. Damit ist der Satz für positives n bewiesen, und wegen $GP(-n, D_0 r^2) = GP(n,D_0r^2)$ ist er auch für negatives n richtig. ♦

Der Satz liefert GP(n,D) für jede ganze Zahl $n \neq 0$ und jede Diskriminante $D \neq 0$. Die Randfälle $D = 0$ und $n = 0$ sind einfach.

GP(n,0) sei für $n \neq 0$ die Summe der Darstellungszahlen von n durch primitive Formenklassen der Diskriminante 0. Somit ist $GP(n,0) = R(n,x^2) + R(n,-x^2)$ nach Satz 2.3. Für $m \in \boldsymbol{N}$ ist $GP(m^2,0) = R(m^2,x^2)$ und $GP(-m^2,0) = R(-m^2,-x^2)$. Die Lösungsmenge von $x^2 = m^2$ ist $\{(m,y), (-m,y) \mid y \in \boldsymbol{Z}\}$. Nach Schritt VII.) in Satz 3.1 ist $\{\begin{pmatrix}1 & \gamma\\ 0 & 1\end{pmatrix}, \begin{pmatrix}-1 & \gamma\\ 0 & -1\end{pmatrix} \mid \gamma \in \boldsymbol{Z}\}$ die Automorphismengruppe von x^2. Sie operiert auf der Lösungsmenge wie folgt: $\begin{pmatrix}1 & 0\\ \gamma & 1\end{pmatrix}\begin{pmatrix}m\\ y\end{pmatrix} = \begin{pmatrix}m\\ \gamma m + y\end{pmatrix}$, $\begin{pmatrix}-1 & 0\\ \gamma & -1\end{pmatrix}\begin{pmatrix}m\\ y\end{pmatrix} = \begin{pmatrix}-m\\ \gamma m - y\end{pmatrix}$, $\begin{pmatrix}1 & 0\\ \gamma & 1\end{pmatrix}\begin{pmatrix}-m\\ y\end{pmatrix} = \begin{pmatrix}-m\\ -\gamma m + y\end{pmatrix}$ und $\begin{pmatrix}-1 & 0\\ \gamma & -1\end{pmatrix}\begin{pmatrix}-m\\ y\end{pmatrix} = \begin{pmatrix}m\\ -\gamma m - y\end{pmatrix}$. Daran sieht man, daß die Menge $\{\begin{pmatrix}m\\ 1\end{pmatrix}, \ldots, \begin{pmatrix}m\\ m\end{pmatrix}\}$ die Äquivalenzklassen von Lösungen repräsentiert. Also ist $R(m^2, x^2) = m$ und ebenso $R(-m^2, -x^2) = m$. Man erhält die

Ergänzung 8.2 *Es ist $GP(\pm m^2, 0) = m$ und $GP(n,0) = 0$ sonst $(m,n \neq 0)$.*

GP(0,D) sei die Summe der Darstellungszahlen von 0 durch primitive Formenklassen der Diskriminante D, wobei aber die stets mögliche triviale Darstellung der 0 durch f(0,0) unberücksichtigt bleiben möge. Zählt man diese Darstellung auch in R(0,f) nicht mit, wenn f irgendeine Form ist, so ist $R(0,f) = \sum_{t^2 \mid 0} R^*(0,f) = \sum_{t=1}^{\infty} R^*(0,f) = \infty$ bzw. 0, je nachdem, ob 0 durch f eigentlich dargestellt wird oder nicht (siehe den Beweis von 4.3). Also ist auch $GP(0,D) = \infty$ bzw. 0, je nachdem, ob 0 durch eine primitive Form der Diskriminante D nichttrivial dargestellt wird oder nicht.

Es gelte $ax^2 + bxy + cy^2 = 0$ mit $(x,y) \neq (0,0)$. Im Falle $a = 0$ ist $D = b^2 - 4ac$ ein Quadrat. Ansonsten hat man $ax^2 + bxy + cy^2 = 0 \Leftrightarrow (x + \frac{b}{2a}y)^2 = \frac{D}{4a^2}y^2 \Leftrightarrow \pm(2ax + by) = y\sqrt{D}$, d.h. $\sqrt{D}$ ist rational und D ist wieder ein Quadrat. Nur Formen mit quadratischer Diskriminante können also 0 nichttrivial darstellen, und jede solche Form tut das offenbar auch nach Lemma 1.15. Man erhält die

Ergänzung 8.3 *Es ist $GP(0,D) = \infty$ falls D ein Quadrat ist, und $GP(0,D) = 0$ sonst.*

Literaturverzeichnis

Grundlage der gesamten Arbeit ist das Werk

Don Bernard Zagier, *Zetafunktionen und quadratische Körper*, Springer-Verlag, Berlin und Heidelberg 1981.

Die ohne Beweis verwendeten elementar-zahlentheoretischen Tatsachen findet man z.B. in

Friedrich Ischebeck, *Einladung zur Zahlentheorie*, BI-Wissenschaftsverlag, Mannheim 1992.

Die historischen Anmerkungen in der Einleitung stammen aus

Winfried Scharlau und Hans Opolka, *Von Fermat bis Minkowski*, Springer-Verlag, Berlin und Heidelberg 1980.

Der Beweis von Satz 8.1 ist eine Ausarbeitung von

Friedrich Hirzebruch und Don Bernard Zagier, „Hilbert Modular Surfaces and Modular Forms of Nebentypus“, *Inventiones Mathematicae* 36, 1976, Seite 69–70, Proposition 2.

Symbolverzeichnis

$\boldsymbol{N}$ Menge der natürlichen (positiven ganzen) Zahlen

$\boldsymbol{N}_0$ Menge der natürlichen Zahlen einschließlich Null

$\boldsymbol{Z}$ Menge der ganzen Zahlen

$\boldsymbol{Q}$ Menge der rationalen Zahlen

$\boldsymbol{R}$ Menge der reellen Zahlen

$\boldsymbol{C}$ Menge der komplexen Zahlen

$\boldsymbol{Z}$ / n$\boldsymbol{Z}$ Restklassengruppe $(\boldsymbol{Z}, +) / (n\boldsymbol{Z}, +)$

$(\boldsymbol{Z}$ / n$\boldsymbol{Z})^*$ Einheitengruppe des Restklassenringes $(\boldsymbol{Z}, +, \cdot) / (n\boldsymbol{Z}, +, \cdot)$

$SL_2(\boldsymbol{Z})$ spezielle lineare Gruppe der 2 x 2 - Matrizen über $\boldsymbol{Z}$ (Gruppe der 2 x 2 – Matrizen mit ganzzahligen Einträgen und Determinante 1)

A x B cartesisches Produkt von Mengen bzw. äußeres direktes Produkt von Gruppen

$\cong$ „ist isomorph zu“ bei Gruppen bzw. „entspricht“ bei Formen

$|\cdot|$ Betrag einer Zahl bzw. Anzahl der Elemente einer Menge

Det(Q) Determinante der Matrix Q

Q^T Transponierte der Matrix Q

D(f) Diskriminante der Form f

U(f) Automorphismengruppe der Form f

R(n,f) Darstellungszahl der ganzen Zahl n durch die Form f

R*(n,f) eigentliche Darstellungszahl der ganzen Zahl n durch die Form f

f_D Grundform zur Diskriminante D

ε_D Grundeinheit zur Diskriminante D

χ_D reeller Charakter zur Diskriminante D

E_D $\{\frac{1}{2}(t + u\sqrt{D}) \mid t,u \in \boldsymbol{Z}, t^2 - Du^2 = 4\}$

h(D) Klassenzahl der Diskriminante D

GP(n,D) Gesamtdarstellungszahl von n durch primitive Formen der Diskriminante D

G(n,D) Gesamtdarstellungszahl von n durch beliebige Formen der Diskriminante D

GP*(n,D) eigentliche Gesamtdarstellungszahl von n durch primitive Formen zu D

G*(n,D) eigentliche Gesamtdarstellungszahl von n durch beliebige Formen zu D

a | n a teilt n in $\boldsymbol{Z}$

ggT größter gemeinsamer Teiler ganzer Zahlen, stets nicht negativ gedacht

sgn Vorzeichenfunktion: $\operatorname{sgn}(n) = 1, 0, -1$ für $n > 0, n = 0, n < 0$

φ Eulersche Phi-Funktion: $\varphi(n) = |\{a \in \{1, \ldots, n\} \mid \operatorname{ggT}(a,n) = 1\}|$ für $n \in \boldsymbol{N}$

I Eins-Funktion: $I(n) = 1$ für $n \in \boldsymbol{N}$

$\left(\frac{a}{n}\right)$ Legendre- bzw. Jacobi-Symbol

$*$ Faltungsoperator bei Folgen: $(F * G)(n) = \sum_{m|n} F(m)G(\frac{n}{m})$ für $n \in \boldsymbol{N}$

$F(n) = O(G(n))$ Landausche O-Notation für Folgen reeller Zahlen: Die Größenordnung von F ist höchstens so groß wie die von G, d.h. die Quotientenfolge $(F(n) / G(n))_{n=1}^{\infty}$ ist beschränkt.

$L(\cdot, \chi)$ zum Charakter χ gehörige L-Reihe