PRIMZERLEGUNG IN IMAGINÄR-QUADRATISCHEN ZAHLRINGEN UND SUMMEN VON ZWEI QUADRATEN

Klaus POMMERENING, Mainz

Inhaltsverzeichnis

Zusammenfassung

Die zahlentheoretische Frage, welche ganzen Zahlen man als Summe von zwei Quadraten darstellen kann, läßt sich am elegantesten durch einen Ausflug in die komplexen Zahlen beantworten. Man verwendet dazu die ganzen Gaußschen Zahlen, eine „komplexe“ Verallgemeinerung der gewöhnlichen ganzen Zahlen. Der theoretische Hintergrund handelt von imaginär-quadratischen Zahlringen, ihren Einheiten und Primelementen. Als Nebenprodukt fällt eine Charakterisierung derjenigen imaginär-quadratischen Ringe ab, in denen stets eine Primfaktorzerlegung möglich ist.

Einleitung

Von den natürlichen Zahlen lassen sich einige als Summe von zwei Quadraten ganzer Zahlen schreiben:

$$1 = 1^2 + 0^2, 2 = 1^2 + 1^2, 4 = 2^2 + 0^2, 5 = 2^2 + 1^2, 8 = 2^2 + 2^2, \ldots,$$

andere dagegen nicht: $3, 6, 7, \ldots$. In vielen Fällen sieht man das mit einer ganz elementaren Beobachtung:

Hilfssatz 1 *Sei $n \in \mathbb{N}$ Summe von zwei Quadraten. Dann ist $n \equiv 0, 1$ oder $2 \pmod 4$.*

Beweis. Jedes Quadrat ist $\equiv 0$ oder $1 \pmod 4$. $\diamondsuit$

Korollar 1 *Ist $n \in \mathbb{N}$, $n \equiv 3 \pmod 4$, so ist n nicht Summe von zwei Quadraten.*

Eine allgemeine Charakterisierung der Zahlen mit einer Darstellung als Summe zweier Quadrate ist mit etwas Geduld zu finden, aber nicht leicht zu beweisen.

Ein sehr eleganter Zugang zu diesem Problem geht über eine Erweiterung der Zahlentheorie auf komplexe Zahlen. Gegenstand dieser Erweiterung sind die Ringe

$$\mathbb{Z}[\sqrt{d}\,] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\} \quad \text{mit } d \in \mathbb{Z};$$

dabei kann man o.B.d.A. annehmen, daß d kein Quadrat ist, weil sonst der triviale Fall $\mathbb{Z}[\sqrt{d}\,] = \mathbb{Z}$ eintritt. Viele Zahlentheorie- oder Algebra-Lehrbücher enthalten einen Beweis, daß $\mathbb{Z}[i]$ (das ist der Fall $d = -1$) euklidisch, also erst recht faktoriell ist. Außerdem erscheinen oft $\mathbb{Z}[\sqrt{-5}\,]$ oder $\mathbb{Z}[\sqrt{10}\,]$ als Beispiele für Ringe ohne Primzerlegung.

Die Ringe $\mathbb{Z}[\sqrt{d}\,]$ mit $d < 0$, also die imaginär-quadratischen Zahlringe unter ihnen, sollen im folgenden so weit systematisch behandelt werden, wie das mit ganz elementaren Methoden möglich ist. Dabei ergibt sich eine vollständige Bestimmung, welche dieser Ringe euklidisch oder faktoriell sind. Wir wollen uns hier nicht darum kümmern, warum Algebraiker oder Zahlentheoretiker einwenden, daß $\mathbb{Z}[\sqrt{d}\,]$ oft gar nicht der „richtige“ quadratische Zahlring ist, den man untersuchen sollte. (Die „richtigen“ Ringe in diesem Sinne sind die ganzen Abschlüsse in ihrem Quotientenkörper.) Von besonderer Bedeutung, auch für die angewandte Mathematik, ist der Ring $\mathbb{Z}[i]$ der „ganzen Gaußschen Zahlen“.

Elementare Kenntnisse über Kongruenzen werden vorausgesetzt. Im Anhang wird angegeben, was von den folgenden Ausführungen für die Quadratsummen-Zerlegung wirklich gebraucht wird.

1 Imaginär-quadratische Zahlringe

Sei $d \in \mathbb{Z}$, $d < 0$. Die Wurzel $\sqrt{d}$ sei stets als $\sqrt{d} = i \cdot \sqrt{|d|}$, gewählt, also mit positivem Imaginärteil. Die Menge

$$\mathbb{Q}(\sqrt{d}\,) = \{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\}$$

ist ein Körper, und zwar ein Teilkörper des Körpers $\mathbb{C}$ der komplexen Zahlen; das rechnet jeder ohne Probleme nach, wenn er daran denkt, daß

$$\frac{1}{a+b\sqrt{d}} = \frac{a-b\sqrt{d}}{a^2-db^2}\,,$$

falls $a+b\sqrt{d} \neq 0$. Die Darstellung $a+b\sqrt{d}$ der Elemente von $\mathbb{Q}(\sqrt{d}\,)$ ist eindeutig (anders ausgedrückt: 1 und $\sqrt{d}$ sind über $\mathbb{Q}$ linear unabhängig): Ist $a'+b'\sqrt{d} = a''+b''\sqrt{d}$, so $a+b\sqrt{d} = 0$ mit $a = a'-a''$, $b = b'-b''$, also $a^2 = db^2$. Da $d < 0$, geht das nur, wenn $a = b = 0$, also $a' = a''$ und $b' = b''$.

Die Formeln für die Addition und Multiplikation in $\mathbb{Q}(\sqrt{d}\,)$ (sie stecken in der eben erwähnten Rechnung „ohne Probleme“) zeigen, daß $\mathbb{Z}[\sqrt{d}\,]$ ein Unterring von $\mathbb{Q}(\sqrt{d}\,)$ ist. Außerdem ist $\mathbb{Z}[\sqrt{d}\,]$ ein Rechtecksgitter in der komplexen Ebene, siehe Abbildung 1.

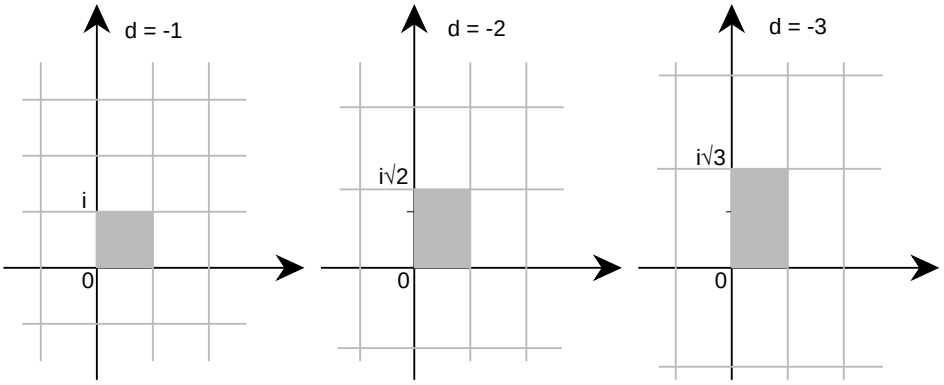

Abbildung 1: Imaginär-quadratische Zahlringe als Gitter

Ein wichtiges Hilfsmittel sind die folgenden beiden Funktionen:

a) Die *Konjugationsabbildung* $\mathbb{Q}(\sqrt{d}\,) \longrightarrow \mathbb{Q}(\sqrt{d}\,)$, $z \mapsto \bar{z}$; für $z = a+b\sqrt{d}$ mit $a, b \in \mathbb{Q}$ ist $\bar{z} = a-b\sqrt{d}$.

b) Die *Norm* $N : \mathbb{Q}(\sqrt{d}\,) \longrightarrow \mathbb{Q}$, $N(a+b\sqrt{d}) = a^2-db^2$, wenn $a, b \in \mathbb{Q}$.

Bemerkung. Da $d < 0$, ist $z \mapsto \bar{z}$ die gewöhnliche komplexe Konjugation, und $N(z) = |z|^2$ mit dem gewöhnlichen Betrag auf $\mathbb{C}$. Für $z, w \in \mathbb{Q}(\sqrt{d}\,)$ gilt

insbesondere: $\overline{zw} = \bar{z}\bar{w}$, $N(z) = z\bar{z}$, $N(zw) = N(z)N(w)$ („Multiplikativität der Norm“), $N(z) = 0 \Longleftrightarrow z = 0$.

Die Frage, ob eine natürliche Zahl n als Summe von zwei Quadraten ganzer Zahlen darstellbar ist, kann man mit Hilfe des Falls $d = -1$, also des Ringes $\mathbb{Z}[i]$ der ganzen Gaußschen Zahlen, nun so ausdrücken:

Satz 1 *Eine natürliche Zahl ist genau dann Summe von zwei Quadraten, wenn sie als Norm einer ganzen Gaußschen Zahl auftritt.*

Eine analoge Bemerkung gilt natürlich allgemeiner für die Darstellung $n = x^2 - dy^2$ mit $d < 0$ und den Ring $\mathbb{Z}[\sqrt{d}\,]$. Da die Norm multiplikativ ist, folgt sofort:

Korollar 2 *Sind m und $n \in \mathbb{N}$ jeweils als Summe zweier Quadrate darstellbar, so auch ihr Produkt $m \cdot n$.*

Das kann man auch schnell, aber weniger einleuchtend, direkt sehen; die explizite Formel ist nämlich

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2.$$

2 Teilbarkeit

Die weitere Untersuchung, welche natürlichen Zahlen sich als Summen von zwei Quadraten schreiben lassen, hängt mit der Teilbarkeitslehre im Ring der ganzen Gaußschen Zahlen zusammen. In diesem Abschnitt wird Teilbarkeit im allgemeinen Rahmen der Ringtheorie behandelt, und es werden euklidische und faktorielle Ringe eingeführt.

2.1 Integritätsringe

Ein **Integritätsring** ist ein kommutativer Ring R mit 1 ohne Nullteiler; dabei ist $1 \neq 0$ verlangt. Genauer: Auf der Menge R sind zwei zweistellige Verknüpfungen

$$+, \cdot : R \times R \longrightarrow R$$

gegeben mit den Eigenschaften:

a) *Assoziativität:* $a + (b + c) = (a + b) + c$ und $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ für alle $a, b, c \in R$.

b) *Kommutativität:* $a + b = b + a$ und $a \cdot b = b \cdot a$ für alle $a, b, \in R$.

c) *Distributivität:* $a \cdot (b + c) = a \cdot b + a \cdot c$ für alle $a, b, c \in R$.

d) *Nullelement:* Es gibt $0 \in R$ mit $a + 0 = a$ für alle $a \in R$.

e) *additives Inverses:* Zu jedem $a \in R$ gibt es ein $b \in R$ mit $a + b = 0$.

f) *Einselement:* Es gibt $1 \in R \setminus \{0\}$ mit $a \cdot 1 = a$ für alle $a \in R \setminus \{0\}$.

g) *Nullteilerfreiheit:* Ist $a \cdot b = 0$ für $a, b \in R$, so $a = 0$ oder $b = 0$.

Elementare Folgerungen aus dieser Definition werde ich ohne viel Aufhebens benützen; hier einige Beispiele:

1.) Das additive Inverse zu $a \in R$ ist eindeutig bestimmt und wird mit $-a$ bezeichnet; $a - b$ wird als $a + (-b)$ definiert.

2.) Null- und Einselement sind eindeutig bestimmt; es gilt $0 \cdot a = 0$ für alle $a \in R$.

3.) *Kürzungsregel:* Ist $a \cdot c = b \cdot c$ für $a, b \in R$ und $c \in R \setminus \{0\}$, so $a = b$.

Der Punkt für die Multiplikation wird oft weggelassen, also $ab := a \cdot b$. Die Definition für **Teiler** liegt nahe:

$$b|a :\Longleftrightarrow \text{es gibt ein } c \in R \text{ mit } a = bc.$$

Teiler von 1 heißen **Einheiten**; dies sind genau die Elemente von R, die

ein multiplikatives Inverses besitzen. Die Menge der Einheiten wird mit $R^\times$ bezeichnet und bildet eine Gruppe.

Ein **Körper** K ist ein Integritätsring mit $K^\times = K \setminus \{0\}$, mit anderen Worten ein Integritätsring, in dem jedes Element $\neq 0$ ein multiplikatives Inverses besitzt. (Dieses ist dann wieder eindeutig bestimmt.)

2.2 Euklidische Ringe

Definition. (i) Eine **euklidische Norm** auf einem Integritätsring R ist eine Funktion

$$\varphi : R \longrightarrow \mathbb{N}_0 \cup \{-\infty\}$$

mit den Eigenschaften: Für beliebige $a \in R$, $b \in R \setminus \{0\}$

(EN 1) gilt $\varphi(ab) \geq \varphi(a)$ („schwache Multiplikativität“),

(EN 2) gibt es ein $q \in R$ mit $\varphi(a - qb) < \varphi(b)$ („Division mit Rest“).

(ii) Ein **euklidischer Ring** ist ein Integritätsring, der mindestens eine euklidische Norm besitzt.

Beispiele.

1. $R = \mathbb{Z}$, $\varphi(a) = |a|$.
2. $R = K[X]$, der Polynomring über dem Körper K, mit $\varphi(f) = \operatorname{Grad} f$ (wobei $\operatorname{Grad} 0 = -\infty$ vereinbart wird).
3. Der triviale euklidische Ring: $R = K$ ein Körper, $\varphi(0) = -\infty$, $\varphi(a) = 0$ für $a \in K^\times$.

Die Divisionsmöglichkeit in den ersten beiden Ringen wurde durch (EN 2) abstrakt gefaßt: Zu allen $a, b \in R$, $b \neq 0$, finden sich ein „Quotient“ q und ein „Rest“ r mit $a = qb + r$, so daß der Rest r eine echt kleinere Norm hat: $\varphi(r) < \varphi(b)$.

Weitere wichtige Beispiele für euklidische Ringe werden wir bald kennenlernen. In der Regel gibt es auf einem Integritätsring *keine* euklidische Norm; wir werden noch sehen, wie man dies (oft) beweisen kann, und auch viele Beispiele dafür finden.

Hier noch ein paar einfache Aussagen über die euklidische Norm:

Hilfssatz 2 *Sei $\varphi : R \longrightarrow \mathbb{N}_0 \cup \{-\infty\}$ eine euklidische Norm. Dann gilt:*

(i) *Für $b \in R$, $b \neq 0$, ist $\varphi(b) > \varphi(0)$; d. h., 0 ist eindeutige Minimalstelle.*

(ii) *Ist $b \in R$ und $\varphi(b) > \varphi(0)$ minimal, so ist $b \in R^\times$ (d. h. Einheit).*

(iii) *Ist $a|b$, $b\not|a$, $b \neq 0$ (d. h., a echter Teiler von b), so $\varphi(a) < \varphi(b)$.*

Beweis. (i) Sei o.B.d.A. $b \in R\backslash\{0\}$ mit minimalem $\varphi(b)$ gewählt. Die Division von 0 durch b gibt $0 = qb + r$ mit $\varphi(r) < \varphi(b)$. Da $\varphi(b)$ minimal war, bleibt nur $r = 0$ übrig, also $\varphi(0) < \varphi(b)$.

(ii) Die Division $1 = qb+r$ ergibt $\varphi(r) < \varphi(b)$, also $r = 0$, $1 = qb$, $b \in R^\times$.

(iii) Sei $a = qb + r$ mit $\varphi(r) < \varphi(b)$. Da $b\not|a$, ist $r \neq 0$. Da $b = ac$ mit $c \in R$, $c \neq 0$, $c \notin R^\times$, folgt $r = a - qb = a(1 - qc)$ mit $1 - qc \neq 0$, also $\varphi(r) \geq \varphi(a)$ und erst recht $\varphi(b) > \varphi(a)$. $\diamondsuit$

Beispiele. Der Leser sollte sich die Aussagen des Hilfssatzes an den Beispielen $R = \mathbb{Z}$ und $R = K[X]$ verdeutlichen.

2.3 Größte gemeinsame Teiler

In einem euklidischen Ring funktioniert die Teilbarkeitslehre in vielen wesentlichen Punkten genauso, wie man es vom Ring $\mathbb{Z}$ der ganzen Zahlen gewöhnt ist. Bei der Definition des größten gemeinsamen Teilers ist es allerdings sinnvoll, auf die Eindeutigkeit zu verzichten:

Definition. Sei R ein Integritätsring, $a, b \in R$. Dann heißt $d \in R$ **ein größter gemeinsamer Teiler** von a und b, wenn

(G 1) $d|a$, $d|b$,

(G 2) ist $c \in R$ ein Teiler von a und b, so $c|d$.

Wie vieldeutig ist diese Definition? Sind d und d' größe gemeinsame Teiler von a und b, so $d|d'$ und $d'|d$, also $d' = cd$ und $d = c'd'$ mit $c, c' \in R$. Es folgt $d = c'd' = c'cd$. Wir müssen zwei Fälle unterscheiden:

1.) $d = 0$. Dann folgt $a = b = 0$, $d' = d$, und d ist eindeutig bestimmt.

2.) $d \neq 0$. Dann ergibt die Kürzungsregel $c'c = 1$, c Einheit.

Hilfssatz 3 *Zwei größte gemeinsame Teiler gehen stets durch Multiplikation mit einer Einheit auseinander hervor.*

Der Begriff des Integritätsrings ist so allgemein gefaßt, daß es längst nicht immer größte gemeinsame Teiler gibt. Jedoch:

Satz 2 *Sei R ein euklidischer Ring. Dann gibt es zu je zwei Elementen $a, b \in R$ stets einen größten gemeinsamen Teiler. Jeder solche ist Linearkombination von a und b.*

Beweis. Sei $M = Ra + Rb = \{xa + yb \,|\, x, y \in R\}$ die Menge aller Linearkombinationen von a und b. Sei $d \in M$ mit minimalem $\varphi(d) > \varphi(0)$ gewählt. Für $u \in M$ wird dividiert: $u = qd + r$ mit $\varphi(r) < \varphi(d)$. Es ist $r = u - qd \in M$, und wegen der Minimalität folgt $r = 0$. Also $u \in Rd$, und $M = Rd$. Insbesondere sind a und $b \in Rd$, also Vielfache von d. Also ist d gemeinsamer Teiler, und $d = xa + yb$ mit $x, y \in R$. Wegen dieser Linearkombination muß jeder gemeinsame Teiler von a und b auch d teilen. Also ist d größter gemeinsamer Teiler. Ist d' ein anderer, so $d' = cd = cxa + cyb$ mit $c \in R^\times$. $\diamondsuit$

2.4 Irreduzible Elemente und Primelemente

Das nächste Thema ist die Primzerlegung, analog zur Zerlegung von ganzen Zahlen in Produkte von Primzahlen. Primzahlen haben verschiedene Eigenschaften, aus denen man im allgemeinen Rahmen abstrakte Definitionen machen kann:

Definition. Sei R ein Integritätsring und $p \in R$. Dann heißt p

(i) **irreduzibel**, wenn aus $p = ab$ mit $a, b \in R$ stets folgt, daß a oder b Einheit ist,

(ii) **prim** oder **Primelement**, wenn aus $p|ab$ mit $a, b \in R$ stets folgt, daß $p|a$ oder $p|b$,

und wenn $p \neq 0$ und keine Einheit ist.

Beispiel. Ist $R = \mathbb{Z}$, so sind die irreduziblen Elemente und die Primelemente genau die Zahlen $\pm p$ mit einer Primzahl p.

Das Gegenteil von irreduzibel ist reduzibel.

Hilfssatz 4 *Sei R ein Integritätsring.*

(i) *Sei $p \in R$ prim. Dann ist p irreduzibel.*

(ii) *Seien $p_1, \ldots, p_m \in R$ prim, $a = p_1 \cdots p_m$. Sei außerdem $a = q_1 \cdots q_n$ mit irreduziblen $q_1, \ldots, q_n \in R$. Dann ist $n = m$ und bei geeigneter Numerierung $q_i = c_i p_i$ mit $c_i \in R^\times$. („Primzerlegungen sind eindeutig.“)*

Beweis. (i) Sei $p = ab$. Dann ist $p|ab$, also etwa $p|a$, $a = pc$, $p = pcb$, $cb = 1$, b Einheit.

(ii) Induktion über m: Es gilt $p_1|q_1 \cdots q_n$. Die Definition von „prim“ (durch Induktion erweitert) ergibt, daß p_1 einen Faktor teilen muß, also etwa (bei geeigneter Numerierung) $p_1|q_1$, $q_1 = c_1 p_1$. Da q_1 irreduzibel ist, muß c_1 Einheit sein. Abdividieren eines Faktors ergibt $p_2 \cdots p_m = c_1 q_2 \cdots q_n$. Falls

$m = 1$, folgt $c_1 q_2 \cdots q_n = 1$, also notwendig $n = 1$. Sonst wird auf das restliche Produkt die Induktionsannahme angewendet. $\diamondsuit$

Beispiele für den Unterschied zwischen „prim“ und „irreduzibel“ werden wir noch kennenlernen. Im Moment wird jedoch bewiesen:

Satz 3 *Sei R ein euklidischer Ring und $p \in R$. Genau dann ist p prim, wenn es irreduzibel ist.*

Beweis. „prim $\Longrightarrow$ irreduzibel“ gilt nach Hilfssatz 4 allgemein. Für die Umkehrung sei nun $p \in R$ irreduzibel, und sei $p|ab$ mit $a, b \in R$. Sei o.B.d.A. $p \not| a$ (sonst sind wir fertig); zu zeigen ist $p|b$. Daß p irreduzibel ist, bedeutet, daß alle Teiler die Gestalt $c \in R^\times$ oder cp mit $c \in R^\times$ haben. Für einen größten gemeinsamen Teiler von p und a bleiben da nur die Einheiten übrig. Nach Satz 2 gibt es also $x, y \in R$ mit $1 = ax + py$. Daraus folgt $b = abx + pby$; da $p|ab$, folgt $p|(abx + pby) = b$. $\diamondsuit$

Dieser Satz ist im Fall $R = \mathbb{Z}$ ein Spezialfall des bekannten „Lemmas von Euklid“.

2.5 Primzerlegung in euklidischen Ringen

Satz 4 *Sei R ein euklidischer Ring und $a \in R$; a sei nicht 0 und keine Einheit. Dann gibt es Primelemente $p_1, \ldots, p_m \in R$ mit*

$$a = p_1 \cdots p_m.$$

Bemerkung. Die Eindeutigkeit der Zerlegung folgt aus Hilfssatz 4.

Beweis. Wegen Satz 3 ist nur zu zeigen, daß a sich in irreduzible Elemente zerlegen läßt. Der Beweis wird indirekt geführt. Falls es Elemente a gibt, die

eine solche Zerlegung nicht haben, gibt es auch eines, das unter diesen einen minimalen Wert $\varphi(a)$ hat. Irreduzibel kann dieses a nicht sein, sonst hätten wir eine Zerlegung mit $m = 1$. Also ist $a = b \cdot c$ mit echten Teilern b und c, also $\varphi(b), \varphi(c) < \varphi(a)$. Wegen der Minimalität von $\varphi(a)$ gibt es Zerlegungen $b = p_1 \cdots p_k$ und $c = q_1 \cdots q_l$ mit irreduziblen p_i, q_j, und a hat doch die Zerlegung $a = p_1 \cdots p_k q_1 \cdots q_l$ in irreduzible Elemente. $\diamondsuit$

Beispiele.

1. Für $R = \mathbb{Z}$ haben wir die gewöhliche Primzahl-Zerlegung, wobei ein eventuelles negatives Vorzeichen einer der Primzahlen zugeschlagen wird.

2. Für $R = K[X]$ haben wir also auch eine Zerlegung in Primelemente. Wie sehen diese aus? Die Antwort hängt davon ab, welcher Körper K ist. In jedem Fall sind lineare Polynome $aX + b$ mit $a \in K^\times$ irreduzibel, also prim. Ist $K = \mathbb{C}$, so sind alle irreduziblen Polynome linear. Das folgt aus dem Fundamentalsatz der Algebra. Jedes Polynom ist also in ein Produkt von Linearfaktoren aufspaltbar. Ist $K = \mathbb{Q}$, so ist keine vollständige Aufzählung aller irreduziblen Polynome bekannt. Beispiele sind $X^2 - 2$, $X^2 + 1$ (Beweis leicht), $X^{p-1} + \cdots + X + 1$ mit p prim (Beweis nicht ganz einfach).

Definition. Ein **faktorieller** Ring (oder ZPE-Ring – für „Zerlegung in Primelemente“) ist ein Integritätsring, in dem sich jedes Element $a \in R \backslash (R^\times \cup \{0\})$ als Produkt von Primelementen schreiben läßt.

Insbesondere ist in einem faktoriellen Ring jedes irreduzible Element prim. Satz 4 läßt sich nun auch so formulieren:

Korollar 3 *Jeder euklidische Ring ist faktoriell.*

3 Primzerlegung in imaginär-quadratischen Zahlringen

Wie lassen sich diese allgemeinen Konzepte auf imaginär-quadratische Zahlringe anwenden?

3.1 Normeuklidische Ringe

Wir wollen zunächst ganz direkt untersuchen, ob die Norm N auf dem imaginär-quadratischen Zahlring $\mathbb{Z}[\sqrt{d}\,]$ eine euklidische Norm ist.

Satz 5 *Sei $d \in \mathbb{Z}$, $d < 0$. Dann sind äquivalent:*

(i) *N ist eine euklidische Norm auf $\mathbb{Z}[\sqrt{d}\,]$.*

(ii) *Für jedes $z \in \mathbb{Q}(\sqrt{d}\,)$ gibt es ein $q \in \mathbb{Z}[\sqrt{d}\,]$ mit $N(z-q) < 1$.*

Beweis. „(ii) $\Longrightarrow$ (i)“: Seien $z, w \in \mathbb{Z}[\sqrt{d}\,]$, $w \neq 0$. Dann ist $N(w) \neq 0$, also $N(zw) = N(z)N(w) \geq N(z)$, also (EN 1) erfüllt.

Sei nun $q \in \mathbb{Z}[\sqrt{d}\,]$ mit $N(\frac{z}{w} - q) < 1$ gewählt. Dann ist

$$N(z - qw) = N(w) \cdot N(\frac{z}{w} - q) < N(w).$$

„(i) $\Longrightarrow$ (ii)“: Sei $z = \frac{x}{y}$ mit $x, y \in \mathbb{Z}[\sqrt{d}\,]$, $y \neq 0$; man kann für $z = a + b\sqrt{d}$ mit $a, b \in \mathbb{Q}$ etwa einen Hauptnenner $y \in \mathbb{N}$ mit $ya, yb \in \mathbb{Z}$ wählen. Wählt man nun nach (EN 2) ein $q \in \mathbb{Z}[\sqrt{d}\,]$ mit $N(x - qy) < N(y)$, so ist $N(y) \cdot N(z - q) = N(x - qy) < N(y)$, also $N(z - q) < 1$. $\diamondsuit$

Dieses Kriterium läßt sich sehr einfach geometrisch deuten: Da $N(z-q)$ der Betrag der komplexen Zahl $z - q$ zum Quadrat ist, ist bei festem q die Menge der z mit $N(z - q) < 1$ der offene Kreis um q mit Radius 1. Die Bedingung (ii) in Satz 5 besagt also: Legt man um jedes $q \in \mathbb{Z}[\sqrt{d}\,]$ den offenen

Kreis vom Radius 1, so überdecken diese Kreise die ganze Ebene. Da die Figur sich sowohl waagerecht als auch senkrecht periodisch wiederholt, reicht es statt dessen zu sagen: Die offenen Kreise um 0, 1, $\sqrt{d}$ und $1+\sqrt{d}$ vom Radius 1 überdecken das ganze (in Abbildung 1 getönte) „Fundamentalrechteck“

$$\{\xi + i\eta \mid 0 \le \xi \le 1,\ 0 \le \eta \le \sqrt{d}\}.$$

Korollar 4 *Genau dann ist N eine euklidische Norm auf $\mathbb{Z}[\sqrt{d}]$, wenn die offenen Kreise vom Radius 1 um die komplexen Zahlen 0, 1, $\sqrt{d}$ und $1+\sqrt{d}$ das Fundamentalrechteck überdecken.*

Man sieht sofort, daß dies für $d = -1, -2$ erfüllt ist, für $d = -3$ dagegen nicht, und erst recht nicht für $d < -3$, siehe Abbildung 2.

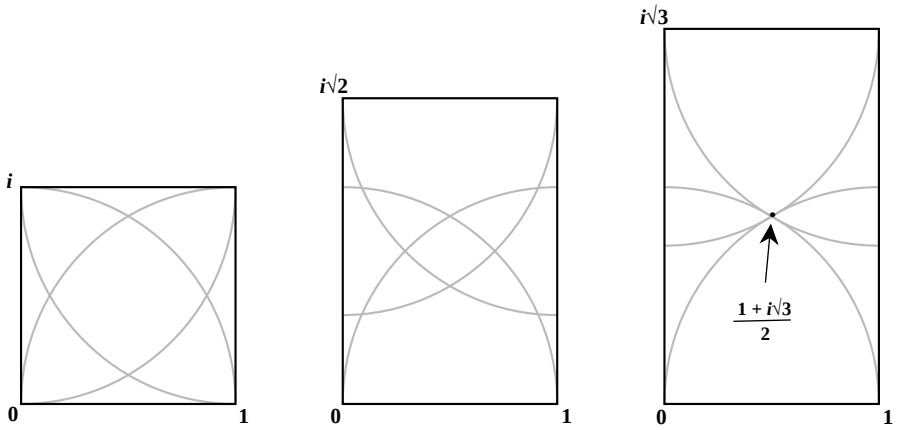

Abbildung 2: Das geometrische Kriterium für Norm-Euklidizität

Der Fall $d = -3$ ist optisch nicht leicht zu erkennen. Rechnet man aber die Entfernung des Mittelpunkts des Fundamentalrechtecks, $\frac{1}{2} + \frac{i}{2}\sqrt{3}$, von 0 aus, erhält man 1; wegen der Symmetrie ist die Entfernung zu den anderen Ecken auch 1. Der Mittelpunkt wird also von der Überdeckung ausgelassen. Allgemein ist die Bedingung in Korollar 4 äquivalent zu

$$\left|\frac{1+i\sqrt{d}}{2}\right| < 1.$$

Definition. Der imaginär-quadratische Zahlring $\mathbb{Z}[\sqrt{d}\,]$ $(d < 0)$ heißt **Norm-euklidisch**, wenn seine Norm N eine euklidische Norm ist.

Korollar 5 *Der imaginär-quadratische Zahlring $\mathbb{Z}[\sqrt{d}\,]$ $(d < 0)$ ist genau dann Norm-euklidisch, wenn $d = -1$ oder -2.*

3.2 Einheiten und Primzahlen in imaginär-quadratischen Zahlringen

Hilfssatz 5 *Sei $d \in \mathbb{Z}$, $d < 0$. Dann gilt:*

(i) *Ein Element $z \in \mathbb{Z}[\sqrt{d}\,]$ ist genau dann Einheit, wenn $N(z) = 1$.*

(ii) *Ist $N(q) = p$ für $q \in \mathbb{Z}[\sqrt{d}\,]$ mit einer Primzahl $p \in \mathbb{Z}$, so ist q in $\mathbb{Z}[\sqrt{d}\,]$ irreduzibel.*

(iii) *Sei $p \in \mathbb{Z}$ eine Primzahl. Dann ist p in $\mathbb{Z}[\sqrt{d}\,]$ genau dann reduzibel, wenn es ein $z \in \mathbb{Z}[\sqrt{d}\,]$ gibt mit $N(z) = p$.*

Beweis. (i) Ist $zw = 1$, so $N(z)N(w) = 1$, also $N(z) = N(w) = 1$. Ist umgekehrt $N(z) = 1$, so $z\bar{z} = 1$, also z Einheit.

(ii) Zunächst ist q weder 0 noch eine Einheit. Sei nun $q = wz$ mit $w, z \in \mathbb{Z}[\sqrt{d}\,]$. Dann ist $p = N(q) = N(w)N(z)$, also $N(w) = 1$ oder $N(z) = 1$, also w oder z Einheit.

(iii) Sei p reduzibel, $p = zw$ mit $z, w \in \mathbb{Z}[\sqrt{d}\,]$, beide $\neq 0$ und keine Einheiten. Dann ist $p^2 = N(p) = N(z)N(w)$; da $N(z), N(w) \neq 1$, bleibt nur die Möglichkeit $N(z) = N(w) = p$.

Sei umgekehrt $z \in \mathbb{Z}[\sqrt{d}\,]$ mit $N(z) = p$ gegeben. Dann ist $z\bar{z} = p$, und $z, \bar{z}$ sind weder 0 noch Einheiten. Also ist p reduzibel. $\diamondsuit$

Beispiel. Die Einheiten in $\mathbb{Z}[i]$ sind die Elemente $z = x + iy$ mit $x^2 + y^2 = 1$. Diese Gleichung ist nur mit $x = \pm 1$, $y = 0$ oder $x = 0$, $y = \pm 1$ erfüllbar. Also ist $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$.

Hilfssatz 6 *Sei $d \in \mathbb{Z}$, $d < 0$. Dann ist 2 in $\mathbb{Z}[\sqrt{d}\,]$ nicht prim.*

Beweis. Es gilt

$$\begin{cases} 2|d = \sqrt{d} \cdot \sqrt{d}, & \text{wenn } d \text{ gerade,} \\ 2|d - 1 = (1 + \sqrt{d})(-1 + \sqrt{d}), & \text{wenn } d \text{ ungerade,} \end{cases}$$

aber 2 teilt keinen der Faktoren. Denn $2|a + b\sqrt{d} \iff 2|a, b$ in $\mathbb{Z}$. $\diamondsuit$

Satz 6 *Sei $d \in \mathbb{Z}$, $d < 0$. Es gelte eine der folgenden Bedingungen:*

(i) *Die Gleichung $a^2 - db^2 = 2$ hat keine Lösung $a, b \in \mathbb{Z}$.*

(ii) *2 ist in $\mathbb{Z}[\sqrt{d}\,]$ irreduzibel.*

Dann ist $\mathbb{Z}[\sqrt{d}\,]$ nicht faktoriell.

Beweis. „(i) $\Longrightarrow$ (ii)“ folgt aus Hilfssatz 5(iii).

„(ii) $\Longrightarrow$ Behauptung“ folgt aus Hilfssatz 6, da 2 dann irreduzibel, aber nicht prim ist. $\diamondsuit$

Korollar 6 *Für $d \leq -3$ ist $\mathbb{Z}[\sqrt{d}\,]$ nicht faktoriell.*

Beweis. $a^2 - db^2 = 2$ ist unmöglich: Für $|b| \geq 1$ ist $a^2 - db^2 \geq -d \geq 3$; für $b = 0$ ist $a^2 - db^2 = a^2 \neq 2$. $\diamondsuit$

3.3 Euklidische imaginär-quadratische Ringe

Für $d \leq -3$ ist $\mathbb{Z}[\sqrt{d}\,]$ also nicht Norm-euklidisch, siehe Korollar 5, nicht einmal faktoriell, siehe Korollar 6. Ein solcher Ring kann also auch nicht auf

andere Weise euklidisch sein. Die Ergebnisse über Euklidizität und Primzerlegung in imaginär-quadratischen Zahlringen sagen also zusammengefaßt:

Hauptsatz 1 *Sei $d \in \mathbb{Z}$, $d < 0$. Dann sind folgende Aussagen äquivalent:*

(i) *Der Ring $\mathbb{Z}[\sqrt{d}\,]$ ist Norm-euklidisch.*

(ii) *Der Ring $\mathbb{Z}[\sqrt{d}\,]$ ist euklidisch.*

(iii) *Der Ring $\mathbb{Z}[\sqrt{d}\,]$ ist faktoriell.*

(iv) *$d = -1$ oder -2.*

3.4 Gaußsche Primzahlen

Im Ring $\mathbb{Z}[i]$ der ganzen Gaußschen Zahlen ist 2 nicht prim nach Hilfssatz 6; wir kennen sogar die Zerlegung $2 = (1+i)(1-i)$. Da $N(1 \pm i) = 2$, sind $1 \pm i$ in $\mathbb{Z}[i]$ nach Hilfssatz 5 (iii) irreduzibel. Trivialerweise ist $2 = 1 + 1 = N(1 + i)$ auch Summe von zwei Quadraten.

Satz 7 *Ist $p \equiv 3 \pmod 4$ eine Primzahl in $\mathbb{Z}$, so ist p in $\mathbb{Z}[i]$ irreduzibel, also prim.*

Beweis. Das folgt aus dem Korollar 1 zusammen mit Hilfssatz 5 (iii). $\diamondsuit$

Etwas schwieriger ist der übrige Fall $p \equiv 1 \pmod 4$ zu behandeln. Zunächst einige Vorbereitungen.

Hilfssatz 7 [Wilson/Lagrange] *Für jede Primzahl $p \in \mathbb{Z}$ gilt $(p-1)! \equiv -1 \pmod p$.*

Beweis. Sei o.B.d.A. $p \geq 5$. Zu jedem $x \in \{1, \ldots, p-1\}$ gibt es ein $x' \in \{1, \ldots, p-1\}$ mit $xx' \equiv 1 \pmod p$. Dabei ist $x = x' \iff x^2 \equiv 1 \iff x = 1$ oder $p - 1$. Die Zahlen $2, \ldots, p-2$ zerfallen also in $\frac{p-3}{2}$ solche Paare x, x'. Daher ist $2 \cdot 3 \cdots (p-2) \equiv 1$, und $(p-1)! \equiv 1 \cdot (p-1) \equiv -1$. $\diamondsuit$

Hilfssatz 8 *Ist $p \in \mathbb{N}$ prim, $p \geq 3$. Dann sind äquivalent:*

(i) $p \equiv 1 \pmod 4$

(ii) *Es gibt ein $x \in \mathbb{Z}$ mit $x^2 \equiv -1 \pmod p$.*

Beweis. „(i) $\Longrightarrow$ (ii)“: Sei $p = 2m + 1$, m gerade. Dann ist $p - 1 \equiv -1$, $p - 2 \equiv -2$, …, $m + 1 \equiv -m \pmod p$. Daraus folgt mit Hilfssatz 7

$$-1 \equiv (p-1)! \equiv 1 \cdot 2 \cdots m \cdot (-m) \cdot \cdots \cdot (-2) \cdot (-1) \equiv [1 \cdot 2 \cdots m]^2.$$

Also ist $x = m!$ die Lösung.

„(ii) $\Longrightarrow$ (i)“: Ist $-1 \equiv x^2$, so nach dem Satz von Fermat

$$(-1)^{p-1/2} \equiv x^{p-1} \equiv 1 \pmod p.$$

Also ist $\frac{p-1}{2}$ gerade. $\diamondsuit$

Satz 8 *Sei $p \in \mathbb{N}$ eine Primzahl mit $p \equiv 1 \pmod 4$. Dann ist p in $\mathbb{Z}[i]$ nicht prim, sondern das Produkt $p = q\bar{q}$ zweier Primelemente $q, \bar{q} \in \mathbb{Z}[i]$, die sich nicht nur um eine Einheit unterscheiden.*

Beweis. Nach Hilfssatz 8 ist -1 ein Quadrat $\bmod p$, also gibt es ein $x \in \mathbb{Z}$ mit $1 \leq x \leq p - 1$ und

$$p | x^2 + 1 = (x + i)(x - i).$$

Da x in $\mathbb{Z}$ nicht durch p teilbar ist, ist $x \pm i$ in $\mathbb{Z}[i]$ nicht durch p teilbar. Also ist p in $\mathbb{Z}[i]$ nicht prim. Nach Hilfssatz 5 (iii) ist $p = N(q) = q\bar{q}$ für ein $q \in \mathbb{Z}[i]$; nach Hilfssatz 5 (ii) sind q und $\bar{q}$ prim.

Ferner unterscheiden sich q und $\bar{q}$ nicht nur um eine Einheit; alle Einheiten sind ja ± 1 und $\pm i$. Setzt man nämlich $q = x + iy$ mit $x, y \in \mathbb{Z}$ und $q = c\bar{q}$ mit einer Einheit c an, so gelangt man jeweils zu einer der Bedingungen $x = 0$, $y = 0$ oder $x = y$; diese können aber alle nicht erfüllt sein, sonst wäre $p = y^2$ oder $p = x^2$ oder $p = 2x^2$. $\diamondsuit$

Korollar 7 [Girard/Fermat/Euler] *Jede Primzahl $p \in \mathbb{N}$ mit $p \equiv 1 \pmod 4$ ist als Summe von zwei Quadraten darstellbar.*

Gibt es noch andere Primelemente in $\mathbb{Z}[i]$? Sei q ein solches, und $N(q) = p_1 \cdots p_r$ die Primzerlegung in $\mathbb{Z}$. Dann hat $N(q) = q\bar{q}$ auch in $\mathbb{Z}[i]$ mindestens r Primfaktoren. Also ist $r \leq 2$. Ist $r = 1$, so ist $p = N(q)$ prim in $\mathbb{Z}$, aber nicht in $\mathbb{Z}[i]$, also $p = 2$ oder $p \equiv 1 \pmod 4$. Ist $r = 2$, so ist $q\bar{q} = p_1 p_2$, und p_1, p_2 sind auch in $\mathbb{Z}[i]$ prim. Also ist $q = cp_\nu$ mit einer Einheit c und $\nu = 1$ oder 2.

Es gibt also keine weiteren Primelemente in $\mathbb{Z}[i]$:

Hauptsatz 2 *Die Primelemente von $\mathbb{Z}[i]$ sind genau die folgenden:*

(i) $1 + i$.

(ii) *Die Primzahlen $p \equiv 3 \pmod 4$.*

(iii) *Zu jeder Primzahl $p \equiv 1 \pmod 4$ ein Paar $q, \bar{q}$ mit $p = q\bar{q}$.*

Dazu kommen alle Zahlen, die daraus durch Multiplikation mit -1, i oder $-i$ entstehen.

Eine bildliche Vorstellung von der Menge der „Gaußschen Primzahlen“ gibt Abbildung 3, die in erweiterter Form als Titelbild dieser Tagung verwendet wurde.

3.5 Summen von zwei Quadraten

Die Ergebnisse über die Darstellung von natürlichen Zahlen als Summe von zwei Quadraten werden wie folgt zusammengefaßt:

Hauptsatz 3 *Sei $n \in \mathbb{N}$, $n = r^2 \cdot s$ mit quadratfreiem s. Genau dann hat n eine Darstellung als Summe von zwei Quadraten, wenn s keinen Primfaktor $p \equiv 3 \pmod 4$ enthält.*

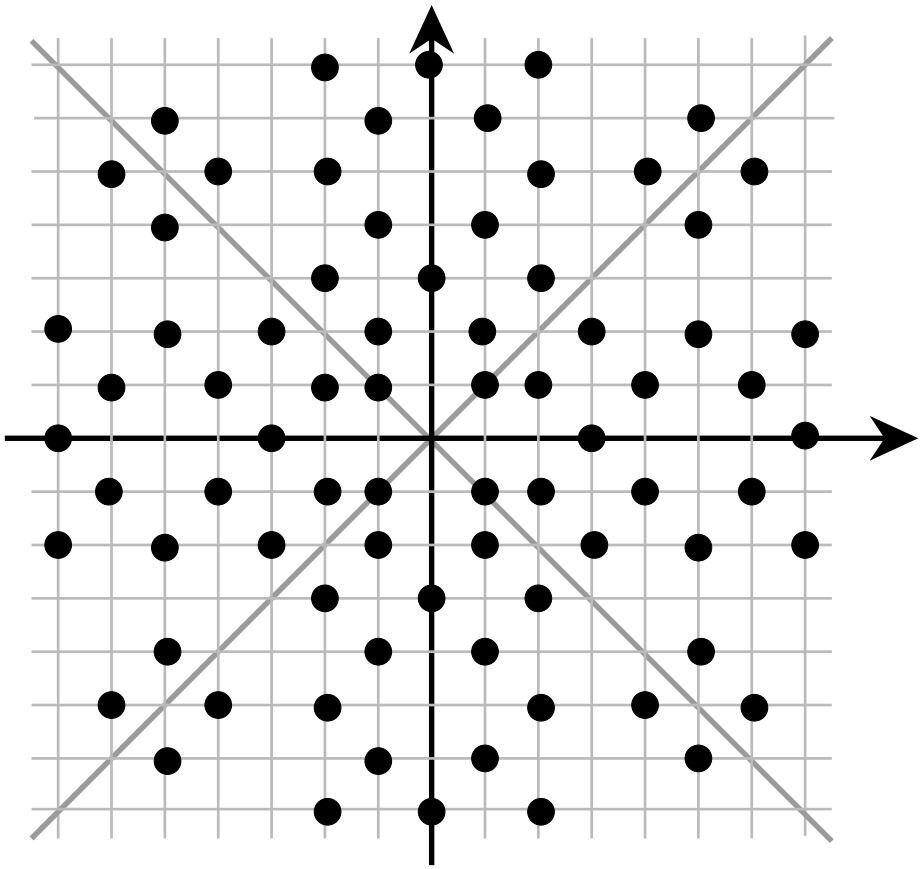

Abbildung 3: Die Gaußschen Primzahlen

Beweis. Daß jede solche Zahl Summe von zwei Quadraten ist, ist einfach: s ist es, weil alle seine Primfaktoren es sind, und ist $s = x^2 + y^2$, so $n = (rx)^2 + (ry)^2$.

Sei umgekehrt $n = x^2 + y^2$ Summe zweier Quadrate. Sei o.B.d.A. $s > 1$. Sei p ein ungerader Primteiler von s. Sei d der größte gemeinsame Teiler von x und y, und $x = du$, $y = dv$ mit teilerfremden u, v. Es folgt $d^2 \cdot (u^2+v^2) = r^2 \cdot s$, also, weil s quadratfrei ist, $d^2 | r^2$, und $p | s | u^2+v^2$. Da u und v teilerfremd sind, können sie nicht beide Vielfache von p sein. Sei also etwa $p \not| u$. Sei $ut \equiv 1 \pmod{p}$. Aus $u^2 + v^2 \equiv 0$ folgt dann $1 + (tv)^2 \equiv 0$. Also ist -1 ein Quadrat mod$p$ und somit $p \equiv 1 \pmod{4}$ nach Hilfssatz 8, was zu zeigen war. $\diamondsuit$

Der Beweis war nicht konstruktiv; der Beweis von Korollar 7 gibt nämlich keinen Algorithmus zur Quadratsummen-Darstellung oder, äquivalent, zur Gaußschen Primzerlegung von $p \equiv 1 \pmod{4}$. Ein einfacher, wenn auch nicht besonders effizienter Algorithmus ist: Probiere x von $\lfloor \sqrt{p} \rfloor$ abwärts bis 1, so lange bis $p - x^2$ ein Quadrat ist. Mit diesem Algorithmus wurde auch Abbildung 3 erzeugt.

Anhang:

Der schnelle Weg zur Quadratsummen-Zerlegung

Vieles von den vorhergehenden Ausführungen braucht man nicht, wenn man nur an dem Satz über die Quadratsummen-Zerlegung (Hauptsatz 3) interessiert ist. Daher wird hier der vorgestellte Beweis auf sein Minimum reduziert.

Zunächst wird das elementare Rechnen mit Kongruenzen benötigt; dazu

der Satz von FERMAT:

$$p \text{ prim}, p \not| x \Longrightarrow x^{p-1} \equiv 1 \pmod{p}.$$

Insbesondere gibt es dann ein y mit $xy \equiv 1 \pmod{p}$. Ferner braucht man die Äquivalenz:

$$x^2 \equiv 1 \pmod{p} \Longleftrightarrow x \equiv \pm 1 \pmod{p}.$$

Dazu kommen noch die Hilfssätze 7 und 8.

Die elementaren Aussagen von Hilfssatz 1 und Korollar 1 aus der Einleitung werden natürlich benötigt.

Dann braucht man den Ring der ganzen Gaußschen Zahlen

$$\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$$

als Unterring von $\mathbb{C}$ und als Quadratgitter in der Ebene. Statt der Norm N reicht es, den komplexen Betrag $|\bullet|$ zu verwenden. Damit formuliert man Satz 1 und Korollar 2, wobei man statt „Norm“ einfach „Betragsquadrat“ sagt.

Von den allgemeinen Begriffen im Zusammenhang mit der Teilbarkeit braucht man (speziell für die ganzen Gaußschen Zahlen) „Teiler“, „größter gemeinsamer Teiler“, „irreduzibel“ und „prim“.

Aus Satz 5 und seinem Beweis extrahiert man die Aussage: *Zu $a, b \in \mathbb{Z}[i]$, $b \neq 0$, gibt es ein $q \in \mathbb{Z}[i]$ mit $|a - qb| < |b|$.* Zum Beweis kann man mit Hilfe des ersten Bildes in der Abbildung 2 argumentieren, daß es q gibt mit $|\frac{a}{b} - q| < 1$.

Dann braucht man Hilfssatz 3 und Satz 2 speziell für $\mathbb{Z}[i]$, wobei im Beweis $|\bullet|$ statt φ verwendet wird. Damit kann man dann eine abgespeckte Version von Satz 3 beweisen: *Ist $p \in \mathbb{Z}[i]$ irreduzibel, so auch prim.*

Aus Hilfssatz 5 benötigt man nur die Aussage: *Ist eine Primzahl $p \in \mathbb{Z}$ als Element von $\mathbb{Z}[i]$ reduzibel, so gibt es ein $z \in \mathbb{Z}[i]$ mit $p = |z|^2$.* Damit kann

man dann die „Minimalform“ von Satz 8 beweisen: *Sei $p \in \mathbb{Z}$ eine Primzahl mit $p \equiv 1 \pmod 4$. Dann gibt es ein $q \in \mathbb{Z}[i]$ mit $p = |q|^2$.* Korollar 7 folgt daraus unmittelbar, und der Beweis von Hauptsatz 3 ist dann auch gesichert.

Literatur

Allgemeine Einführung in die komplexen Zahlen:

H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, A. Prestel, R. Remmert: *Zahlen.* Springer-Verlag.

Einführungen in die Zahlentheorie mit elementarer Behandlung des Quadratsummen-Problems:

D. M. Burton: *Elementary Number Theory.* Allyn and Bacon Inc., Boston 1976.

H. Lüneburg: *Kleine Fibel der Arithmetik.* BI-Wissenschaftsverlag, Mannheim 1987.

Quadratische Zahlringe (einschließlich der Anwendung auf die Quadratsummen-Zerlegung) werden auf einem höheren Niveau behandelt in:

H. Lüneburg: *Vorlesungen über Zahlentheorie.* Birkhäuser-Verlag, Basel 1978.