Chapter 2
Continued Fractions, Diophantine Approximation, and Quadratic Rings

The main goal of this chapter is to lay out basic concepts needed in our study in Diophantine Analysis. The first section contains fundamental results pertaining to continued fractions, some without proofs. The Theory of Continued Fractions is not new but it plays a growing role in contemporary mathematics.

Continued fractions have fascinated mankind for centuries if not millennia. The timeless construction of a rectangle obeying the "divine proportion" (the term is in fact from the Renaissance) and the "self-similarity" properties that go along with it are nothing but geometric counterparts of the continued fraction expansion of the golden ratio,

$$\phi \equiv \frac{1+\sqrt{5}}{2} = \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\dots}}}.$$

Geometry was developed in India from the rules for the construction of altars. The Sulva Sutra (a part of the Kalpa Sutra hypothesized to have been written around 800 BC) provides a rule for doubling an area that corresponds to the near-equality:

$$\sqrt{2} = 1 + \frac{1}{3} + \frac{1}{3\times 4} - \frac{1}{3\times 4\times 34} \quad \text{(correct to } 2\cdot 10^{-6}\text{)}.$$

The third and fourth partial sums namely $\frac{17}{12}$ and $\frac{577}{408}$ are respectively the fourth and eight convergents to $\sqrt{2}$.

Accordingly, in the classical Greek world, there is evidence of knowledge of the continued fraction for $\sqrt{2}$ which appears in the works of Theon of Smyrna (discussed in Fowler's reconstruction [74] and in [215]) and possibly of Plato in *Theatetus*, see [49]. As every student knows, Euclid's algorithm is a continued fraction expansion algorithm in disguise, and Archimedes' Cattle Problem (circa 250 BC)

© Springer Science+Business Media New York 2015

T. Andreescu, D. Andrica, *Quadratic Diophantine Equations*,
Developments in Mathematics 40, DOI 10.1007/978-0-387-54109-9_2

most probably presupposes on the part of its author some amount of understanding of quadratic irrationals, Pell's equation, and continued fractions; see [215] for a discussion.

The continued fraction convergent $\pi \approx \frac{355}{113}$ was known to Twu Ch'ung Chi, born in Fan-yang, China in 430 AD. More recently, the Swiss mathematician Lambert proved the 2,000 year conjecture (it already appears in Aristotle) that π is irrational, this thanks to the continued fraction expansion of the tangent function,

$$\tan z = \cfrac{z}{1 - \cfrac{z^2}{3 - \cfrac{z^2}{5 - \dots}}},$$

and Apéry in 1979 gave in "a proof that Euler missed" [176] nonstandard expansions like

$$\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3} = 1 + \cfrac{1}{2\cdot 2 + \cfrac{1^3}{1 + \cfrac{1^3}{2\cdot 6 + \cfrac{2^3}{1 + \cfrac{2^3}{2\cdot 10 + \cfrac{3^3}{1 + \dots}}}}}}$$

from which the irrationality of $\zeta(3)$ eventually derives.

The standard method to prove the irrationality of e^x for nonzero rational x is by obtaining a rational approximation using the differential and integral properties of e^x and the differential properties of $x^n(1-x)^n/n!$, see [88]. Recently, a simple proof by using the theory of continued fractions was given in [154].

The principal references used in this section are [1, 46, 66, 141, 159, 164, 183, 184, 208].

The Section 2.2 presents key results regarding quadratic rings, their units and norms defined in a natural way. Important references for this section are [95, 171, 198].

2.1 Simple Continued Fractions

2.1.1 The Euclidean Algorithm

Given any rational fraction u_0/u_1, in lowest terms so that $\gcd(u_0, u_1) = 1$ and $u_1 > 0$, we apply the Euclidean algorithm (see [21]) to get successively

$$\begin{aligned}
&u_0 = u_1 a_0 + u_2, && 0 < u_2 < u_1 \\
&u_1 = u_2 a_1 + u_3, && 0 < u_3 < u_2 \\
&u_2 = u_3 a_2 + u_4, && 0 < u_4 < u_3 \\
&\cdots && \\
&u_{j-1} = u_j a_{j-1} + u_{j+1}, && 0 < u_{j+1} < u_j \\
&u_j = u_{j+1} a_j. &&
\end{aligned} \tag{2.1.1}$$

If we write ξ_i in place of u_i/u_{i+1} for all values of i with $0 \le i \le j$, then equations (2.1.1) become

$$\xi_i = a_i + \frac{1}{\xi_{i+1}}, \quad 0 \le i \le j-1; \quad \xi_j = a_j. \tag{2.1.2}$$

If we take the first two of these equations, those for which $i = 0$ and $i = 1$, and eliminate ξ_1, we get

$$\xi_0 = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\xi_2}}.$$

In this result we replace ξ_2 by its value from (2.1.2), and then we continue with replacement of $\xi_3, \xi_4, \ldots$, to get

$$\frac{u_0}{u_1} = \xi_0 = a_0 + \cfrac{1}{a_1 + \cfrac{\ddots}{+\cfrac{1}{a_{j-1} + \cfrac{1}{a_j}}}}. \tag{2.1.3}$$

This is a *continued fraction expansion* of ξ_0, or of u_0/u_1. The integers a_i are called the *partial quotients* since they are the quotients in the repeated application of the division algorithm in equations (2.1.1). We presumed that the rational fraction u_0/u_1 had positive denominator u_1, but we cannot make a similar assumption about u_0. Hence a_0 may be positive, negative, or zero. However, since $0 < u_2 < u_1$, we note that the quotient a_1 is positive, and similarly the subsequent quotients $a_2, a_3, \ldots, a_j$ are positive integers. In case $j \ge 1$, that is if the set (2.1.1) contains more than one equation, then $a_j = u_j/u_{j+1}$ and $0 < u_{j+1} < u_j$ imply that $a_j > 1$.

We will use the notation $\langle a_0, a_1, \ldots, a_j \rangle$ to designate the continued fraction in (2.1.3). In general, if $x_0, x_1, \ldots, x_j$ are any real numbers, all positive except perhaps x_0, then we will write

$$\langle x_0, x_1, \ldots, x_j \rangle = x_0 + \cfrac{1}{x_1 + \cfrac{}{\ddots + \cfrac{1}{x_{j-1} + \cfrac{1}{x_j}}}}.$$

Such a finite continued fraction is said to be *simple* if all the x_i are integers. The following notations are often used to simplify the writing:

$$\begin{aligned}\langle x_0, x_1, \ldots, x_j \rangle &= x_0 + \frac{1}{\langle x_1, \ldots, x_j \rangle} \\ &= \left\langle x_0, x_1, \ldots, x_{j-2}, x_{j-1} + \frac{1}{x_j} \right\rangle.\end{aligned}$$

The symbol $[x_0, x_1, \ldots, x_j]$ is sometimes used to represent a continued fraction. We use the notation $\langle x_0, x_1, \ldots, x_j \rangle$ to avoid confusion with the least common multiple and the greatest integer.

2.1.2 Uniqueness

In the last section we saw that such a fraction as $51/22$ can be expanded into a simple continued fraction, $51/22 = \langle 2, 3, 7 \rangle$. It can be verified that $51/22$ can also be expressed as $\langle 2, 3, 6, 1 \rangle$, but it turns out that these are the only two representations of $51/22$. In general, we note that the simple continued fraction expansion (2.1.3) has an alternate form,

$$\frac{u_0}{u_1} = \langle a_0, a_1, \ldots, a_{j-1}, a_j \rangle = \langle a_0, a_1, \ldots, a_{j-2}, a_{j-1}, a_j - 1, 1 \rangle. \tag{2.1.4}$$

The following result [159] establishes that these are the only two simple continued fraction expansions of a fixed rational number.

Theorem 2.1.1. *If $\langle a_0, a_1, \ldots, a_j \rangle = \langle b_0, b_1, \ldots, b_n \rangle$, where these finite continued fractions are simple, and if $a_j > 1$ and $b_n > 1$, then $j = n$ and $a_i = b_i$ for $i = 0, 1, \ldots, n$.*

Proof. We write y_i for the continued fraction $\langle b_i, b_{i+1}, \ldots, b_n \rangle$ and observe that

$$y_i = \langle b_i, b_{i+1}, \ldots, b_n \rangle = b_i + \frac{1}{\langle b_{i+1}, b_{i+2}, \ldots, b_n \rangle} = b_i + \frac{1}{y_{i+1}}. \tag{2.1.5}$$

Thus we have $y_i > b_i$ and $y_i > 1$ for $i = 1, 2, \ldots, n-1$, and $y_n = b_n > 1$. Consequently, $b_i = [y_i]$ for all values of i in the range $0 \le i \le n$. The hypothesis that the continued fractions are equal can be written in the form $y_0 = \xi_0$, where we are using the notation of equation (2.1.3). Now the definition of ξ_i as u_i/u_{i+1} implies that $\xi_{i+1} > 1$ for all values of $i \geq 0$, and so $a_i = [\xi_i]$ for $0 \leq i \leq j$ by equation (2.1.2). It follows from $y_0 = \xi_0$ that, taking integral parts, $b_0 = [y_0] = [\xi_0] = a_0$. By equations (2.1.2) and (2.1.5) we get

$$\frac{1}{\xi_1} = \xi_0 - a_0 = y_0 - b_0 = \frac{1}{y_1}, \quad \xi_1 = y_1, \quad a_1 = [\xi_1] = [y_1] = b_1.$$

This gives us the start of a proof by induction. We now establish that $\xi_i = y_i$ and $a_i = b_i$ imply that $\xi_{i+1} = y_{i+1}$ and $a_{i+1} = b_{i+1}$. To see this, we again use equations (2.1.2) and (2.1.5) to write

$$\frac{1}{\xi_{i+1}} = \xi_i - a_i = y_i - b_i = \frac{1}{y_{i+1}},$$
$$\xi_{i+1} = y_{i+1}, \quad a_{i+1} = [\xi_{i+1}] = [y_{i+1}] = b_{i+1}.$$

It must also follow that the continued fractions have the same length, that is, that $j = n$. For suppose that, say, $j < n$. From the preceding argument we have $\xi_j = y_j$, $a_j = b_j$. But $\xi_j = a_j$ by (2.1.2) and $y_j > b_j$ by (2.1.5), and so we have a contradiction. If we had assumed $j > n$, a symmetrical contradiction would have arisen, and thus j must equal n, and the theorem is proved. $\square$

Theorem 2.1.2. *Any finite simple continued fraction represents a rational number. Conversely, any rational number can be expressed as a finite simple continued fraction, and in exactly two ways.*

2.1.3 Infinite Continued Fractions

Let $a_0, a_1, a_2, \ldots$ be an infinite sequence of integers, all positive except perhaps a_0. We define two sequences of integers $\{h_n\}$ and $\{k_n\}$ inductively as follows:

$$\begin{aligned} &h_{-2} = 0,\ h_{-1} = 1,\ h_i = a_i h_{i-1} + h_{i-2} \text{ for } i \geq 0 \\ &k_{-2} = 1,\ k_{-1} = 0,\ k_i = a_i k_{i-1} + k_{i-2} \text{ for } i \geq 0. \end{aligned} \tag{2.1.6}$$

We note that $k_0 = 1$, $k_1 = a_1 k_0 \geq k_0$, $k_2 > k_1$, $k_3 > k_2$, etc., so that $1 = k_0 \leq k_1 < k_2 < k_3 < \cdots < k_n < \ldots$.

Theorem 2.1.3. *For any positive real number x,*

$$\langle a_0, a_1, \ldots, a_{n-1}, x\rangle = \frac{x h_{n-1} + h_{n-2}}{x k_{n-1} + k_{n-2}}.$$

Theorem 2.1.4. *If we define* $r_n = \langle a_0, a_1, \ldots, a_n \rangle$ *for all integers* $n \geq 0$*, then* $r_n = h_n/k_n$.

Theorem 2.1.5. *The equations*

$$h_i k_{i-1} - h_{i-1} k_i = (-1)^{i-1} \quad \textit{and} \quad r_i - r_{i-1} = \frac{(-1)^{i-1}}{k_i k_{i-1}}$$

hold for $i \geq 1$*. The identities*

$$h_i k_{i-2} - h_{i-2} k_i = (-1)^i a_i \quad \textit{and} \quad r_i - r_{i-2} = \frac{(-1)^i a_i}{k_i k_{i-2}}$$

hold for $i \geq 1$*. The fraction* h_i/k_i *is reduced, that is* $(h_i, k_i) = 1$.

Theorem 2.1.6. *The values* r_n *defined in Theorem 2.1.4 satisfy the infinite chain of inequalities* $r_0 < r_2 < r_4 < r_6 < \cdots < r_7 < r_5 < r_3 < r_1$*. Furthermore,* $\lim_{n\to\infty} r_n$ *exists, and for every* $j \geq 0$*,* $r_{2j} < \lim_{n\to\infty} r_n < r_{2j+1}$.

Proof. The identities of Theorem 2.1.5 for $r_i - r_{i-1}$ and $r_i - r_{i-2}$ imply that $r_{2j} < r_{2j+2}$, $r_{2j-1} > r_{2j+1}$, and $r_{2j} < r_{2j-1}$, because the k_i are positive for $i \geq 0$ and the a_i are positive for $i \geq 1$. Thus we have $r_0 < r_2 < r_4 < \ldots$ and $r_1 > r_3 > r_5 > \ldots$. To prove that $r_{2n} < r_{2j-1}$, we put the previous results together in the form

$$r_{2n} < r_{2n+2j} < r_{2n+2j-1} \leq r_{2j-1}.$$

The sequence $r_0, r_2, r_4, \ldots$ is monotonically increasing and is bounded above by r_1, and so has a limit. Analogously, the sequence $r_1, r_3, r_5, \ldots$ is monotonically decreasing and is bounded below by r_0, and so has a limit. These two limits are equal because, by Theorem 2.1.5, the difference $r_i - r_{i-1}$ tends to zero as i tends to infinity, since the integers k_i are increasing with i. Another way of looking at this to observe that $(r_0, r_1), (r_2, r_3), (r_4, r_5), \ldots$ is a chain of nested intervals defining a real number, namely $\lim_{n\to\infty} r_n$. $\square$

These theorems suggest the following definition.

Definition 2.1.1. An infinite sequence $a_0, a_1, a_2, \ldots$ of integers, all positive except perhaps for a_0, determines an infinite simple continued fraction $\langle a_0, a_1, a_2, \ldots \rangle$. The value of $\langle a_0, a_1, a_2, \ldots \rangle$ is defined to be $\lim_{n\to\infty} \langle a_0, a_1, a_2, \ldots, a_n \rangle$.

This limit, being the same as $\lim_{n\to\infty} r_n$, exists by Theorem 2.1.6. Another way of writing this limit is $\lim_{n\to\infty} h_n/k_n$. The rational number $\langle a_0, a_1, \ldots, a_n \rangle = h_n/k_n = r_n$ is called the nth *convergent* to the infinite continued fraction. We say that the infinite continued fraction converges to the value $\lim_{n\to\infty} r_n$. In the case of a finite simple continued fraction $\langle a_0, a_1, \ldots, a_n \rangle$ we similarly call the number $\langle a_0, a_1, \ldots, a_m \rangle$ the mth *convergent* to $\langle a_0, a_1, \ldots, a_n \rangle$.

Theorem 2.1.7. *The value of any infinite simple continued fraction $\langle a_0, a_1, a_2, \dots \rangle$ is irrational.*

Proof. Writing θ for $\langle a_0, a_1, a_2, \dots \rangle$, we observe by Theorem 2.1.6 that θ lies between r_n and r_{n+1}, so that $0 < |\theta - r_n| < |r_{n+1} - r_n|$. Multiplying by k_n, and making use of the result from Theorem 2.1.5 that $|r_{n+1} - r_n| = (k_n k_{n+1})^{-1}$, we have

$$0 < |k_n\theta - h_n| < \frac{1}{k_{n+1}}.$$

Now suppose that θ were rational, say $\theta = a/b$ with integers a and b, $b > 0$. Then the above inequality would become, upon multiplication by b,

$$0 < |k_n a - h_n b| < \frac{b}{k_{n+1}}.$$

The integers k_n increase with n, so we could choose n sufficiently large so that $b < k_{n+1}$. Then the integer $|k_n a - h_n b|$ would lie between 0 and 1, which is impossible. □

Lemma 2.1.8. *Let $\theta = \langle a_0, a_1, a_2, \dots \rangle$ be a simple continued fraction. Then $a_0 = [\theta]$. Furthermore, if θ_1 denotes $\langle a_1, a_2, a_3, \dots \rangle$, then $\theta = a_0 + 1/\theta_1$.*

Proof. By Theorem 2.1.6 we see that $r_0 < \theta < r_1$, that is $a_0 < \theta < a_0 + 1/a_1$. Now $a_1 \geq 1$, so we have $a_0 < \theta < a_0 + 1$, and hence $a_0 = [\theta]$. Also

$$\begin{aligned}\theta = \lim_{n\to\infty} \langle a_0, a_1, \dots, a_n \rangle &= \lim_{n\to\infty} \left(a_0 + \frac{1}{\langle a_1, \dots, a_n \rangle} \right) \\ &= a_0 + \frac{1}{\lim_{n\to\infty} \langle a_1, \dots, a_n \rangle} = a_0 + \frac{1}{\theta_1}.\end{aligned}$$

□

Theorem 2.1.9. *Two distinct infinite simple continued fractions converge to different values.*

Proof. Let us suppose that $\langle a_0, a_1, a_2, \dots \rangle = \langle b_0, b_1, b_2, \dots \rangle = \theta$. Then by Lemma 2.1.8, $[\theta] = a_0 = b_0$ and

$$\theta = a_0 + \frac{1}{\langle a_1, a_2, \dots \rangle} = b_0 + \frac{1}{\langle b_1, b_2, \dots \rangle}.$$

Hence $\langle a_1, a_2, \dots \rangle = \langle b_1, b_2, \dots \rangle$. Repetition of the argument gives $a_1 = b_1$, and so by induction $a_n = b_n$ for all n. □

2.1.4 Irrational Numbers

We have shown that any infinite simple continued fraction represents an irrational number. Conversely, if we begin with an irrational number ξ, or ξ_0, we can expand it into an infinite simple continued fraction. To do this we define $a_0 = [\xi_0]$, $\xi_1 = 1/(\xi_0 - a_0)$, and next $a_1 = [\xi_1]$, $\xi_2 = 1/(\xi_1 - a_1)$, and so by an inductive definition

$$a_i = [\xi_i], \quad \xi_{i+1} = \frac{1}{\xi_i - a_i}. \tag{2.1.7}$$

The a_i are integers by definition, and the ξ_i are all irrational, since the irrationality of ξ_1 is implied by that of ξ_0, that of ξ_2 by that of ξ_1, and so on. Furthermore, $a_i \geq 1$ for $i \geq 1$ because $a_{i-1} = [\xi_{i-1}]$ and the fact that ξ_{i-1} is irrational implies that

$$a_{i-1} < \xi_{i-1} < 1 + a_{i-1}, \quad 0 < \xi_{i-1} - a_{i-1} < 1,$$
$$\xi_i = \frac{1}{\xi_{i-1} - a_{i-1}} > 1, \quad a_i = [\xi_i] \geq 1.$$

Next we use repeated application of (2.1.7) in the form $\xi_i = a_i + 1/\xi_{i+1}$ to get the chain

$$\begin{aligned} \xi = \xi_0 = a_0 + \frac{1}{\xi_1} &= \langle a_0, \xi_1 \rangle \\ &= \left\langle a_0, a_1 + \frac{1}{\xi_2} \right\rangle = \langle a_0, a_2, \xi_2 \rangle \\ &= \left\langle a_0, a_1, \ldots, a_{m-2}, a_{m-1} + \frac{1}{\xi_m} \right\rangle \\ &= \langle a_0, a_1, \ldots, a_{m-1}, \xi_m \rangle. \end{aligned}$$

This suggests, but does not establish, that ξ is the value of the infinite continued fraction $\langle a_0, a_1, a_2, \ldots \rangle$ determined by the integers a_i.

To prove this we use Theorem 2.1.3 to write

$$\xi = \langle a_0, a_1, \ldots, a_{n-1}, \xi_n \rangle = \frac{\xi_n h_{n-1} + h_{n-2}}{\xi_n k_{n-1} + k_{n-2}} \tag{2.1.8}$$

with the h_i and k_i defined as in (2.1.6). By Theorem 2.1.5 we get

$$\begin{aligned} \xi - r_{n-1} = \xi - \frac{h_{n-1}}{k_{n-1}} &= \frac{\xi_n h_{n-1} + h_{n-2}}{\xi_n k_{n-1} + k_{n-2}} - \frac{h_{n-1}}{k_{n-1}} \\ &= \frac{-(h_{n-1}k_{n-2} - h_{n-2}k_{n-1})}{k_{n-1}(\xi_n k_{n-1} + k_{n-2})} = \frac{(-1)^{n-1}}{k_{n-1}(\xi_n k_{n-1} + k_{n-2})}. \end{aligned} \tag{2.1.9}$$

This fraction tends to zero as n tends to infinity because the integers k_n are increasing with n, and ξ_n is positive. Hence $\xi - r_{n-1}$ tends to zero as n tends to infinity and then, by Definition 2.1.1,

$$\xi = \lim_{n\to\infty} r_n = \lim_{n\to\infty} \langle a_0, a_1, \ldots, a_n \rangle = \langle a_0, a_1, a_2, \ldots \rangle.$$

We summarize the results of the last two sections in the following theorem.

Theorem 2.1.10. *Any irrational number ξ is uniquely expressible, by the procedure that gave equations (2.1.7), as an infinite simple continued fraction $\langle a_0, a_1, a_2, \ldots \rangle$. Conversely, any such continued fraction determined by integers a_i that are positive for all $i > 0$ represents an irrational number, ξ. The finite simple continued fraction $\langle a_0, a_1, \ldots, a_n \rangle$ has the rational value $h_n/k_n = r_n$, and is called the nth convergent to ξ. Equations (2.1.6) relate the h_i and k_i to the a_i. For $n = 0, 2, 4, \ldots$ these convergents form a monotonically sequence with ξ as a limit. Similarly, for $n = 1, 3, 5, \ldots$ the convergents form a monotonically decreasing sequence tending to ξ. The denominators k_n of the convergents are an increasing sequence of positive integers for $n > 0$. Finally, with ξ_i defined by (2.1.7), we have $\langle a_0, a_1, \ldots \rangle = \langle a_0, a_1, \ldots, a_{n-1}, \xi_n \rangle$ and $\xi_n = \langle a_n, a_{n+1}, a_{n+2}, \ldots \rangle$.*

Proof. Only the last equation is new, and it becomes obvious if we apply to ξ_n the process described at the opening of this section. □

Example 1. Let us expand $\sqrt{5}$ as an infinite simple continued fraction.

We see that

$$\sqrt{5} = 2 + (\sqrt{5} - 2) = 2 + 1/(\sqrt{5} + 2)$$

and

$$\sqrt{5} + 2 = 4 + (\sqrt{5} - 2) = 4 + 1/(\sqrt{5} + 2).$$

In view of the repetition of $1/(\sqrt{5} + 2)$, we obtain $\sqrt{5} = \langle 2, 4, 4, 4, \ldots \rangle$.

2.1.5 Approximations to Irrational Numbers

Continuing to use the notation on the preceding sections, we now show that the convergents h_n/k_n form a sequence of "best" rational approximations to the irrational number ξ.

Theorem 2.1.11. *We have for any $n \geq 0$,*

$$\left| \xi - \frac{h_n}{k_n} \right| < \frac{1}{k_n k_{n+1}} \quad \textit{and} \quad |\xi k_n - h_n| < \frac{1}{k_{n+1}}.$$

Proof. The second inequality follows from the first by multiplication by k_n. By (2.1.9) and (2.1.7) we have

$$\left|\xi - \frac{h_n}{k_n}\right| = \frac{1}{k_n(\xi_{n+1}k_n + k_{n-1})} < \frac{1}{k_n(a_{n+1}k_n + k_{n-1})}.$$

Using (2.1.6), we replace $a_{n+1}k_n + k_{n-1}$ by k_{n+1} to obtain the first inequality. □

Theorem 2.1.12. *The convergents h_n/k_n are successively closer to ξ, that is*

$$\left|\xi - \frac{h_n}{k_n}\right| < \left|\xi - \frac{h_{n-1}}{k_{n-1}}\right|.$$

In fact the stronger inequality $|\xi k_n - h_n| < |\xi k_{n-1} - h_{n-1}|$ holds.

Proof. We use $k_{n-1} \leq k_n$ to write

$$\begin{aligned}\left|\xi - \frac{h_n}{k_n}\right| &= \frac{1}{k_n}|\xi k_n - h_n| < \frac{1}{k_n}|\xi k_{n-1} - h_{n-1}| \\ &\leq \frac{1}{k_{n-1}}|\xi k_{n-1} - h_{n-1}| = \left|\xi - \frac{h_{n-1}}{k_{n-1}}\right|.\end{aligned}$$

Now to prove the stronger inequality we observe that $a_n + 1 > \xi_n$ by (2.1.7), and so by (2.1.6), we have

$$\begin{aligned}\xi_n k_{n-1} + k_{n-2} &< (a_n + 1)k_{n-1} + k_{n-2} \\ &= k_n + k_{n-1} \leq a_{n+1}k_n + k_{n-1} = k_{n+1}.\end{aligned}$$

This inequality and (2.1.9) imply that

$$\left|\xi - \frac{h_{n-1}}{k_{n-1}}\right| = \frac{1}{k_{n-1}(\xi_n k_{n-1} + k_{n-2})} > \frac{1}{k_{n-1}k_{n+1}}.$$

We multiply by k_{n-1} and use Theorem 2.1.11 to get

$$|\xi k_{n-1} - h_{n-1}| > \frac{1}{k_{n+1}} > |\xi k_n - h_n|.$$

□

The convergent h_n/k_n is the best approximation to ξ of all the rational fractions with denominator k_n or less. The following theorem states this in a different way. For the proof we refer to [159].

Theorem 2.1.13. *If a/b is a rational number with positive denominator such that $|\xi - a/b| < |\xi - h_n/k_n|$ for some $n \geq 1$, then $b > k_n$. In fact if $|\xi b - a| < |\xi k_n - h_n|$ for some $n \geq 0$, then $b \geq k_{n+1}$.*

Theorem 2.1.14. *Let ξ denote any irrational number. If there is a rational number a/b with $b \geq 1$ such that*

$$\left|\xi - \frac{a}{b}\right| < \frac{1}{2b^2},$$

then a/b equals one of the convergents of the simple continued fraction expansion of ξ.

Theorem 2.1.15. *The nth convergent of $1/x$ is the reciprocal of the $(n-1)$st convergent of x if x is any real number greater than 1.*

2.1.6 Best Possible Approximations

Theorem 2.1.11 provides another method of proving the following well-known result (see [159, p. 302]). If ξ is real and irrational, there are infinitely many distinct rational numbers a/b such that

$$\left|\xi - \frac{a}{b}\right| < \frac{1}{b^2}.$$

Indeed, from Theorem 2.1.11 we can replace k_{n+1} by the smaller integer k_n to get the weaker, but still correct, inequality

$$\left|\xi - \frac{h_n}{k_n}\right| < \frac{1}{k_n^2}.$$

We can also use continued fractions to get different proofs of the following result of Hurwitz [159, pp. 304–305]:

Given an irrational number ξ, there exist infinitely many different rational numbers h/k such that

$$\left|\xi - \frac{h}{k}\right| < \frac{1}{\sqrt{5}k^2}$$

and the constant $\sqrt{5}$ is the best possible. The following auxiliary result is a simple consequence of the sign of the quadratic function.

Lemma 2.1.16. *If x is real, $x > 1$, and $x + x^{-1} < \sqrt{5}$, then $x < \frac{1}{2}(\sqrt{5}+1)$ and $x^{-1} > \frac{1}{2}(\sqrt{5}-1)$.*

Theorem 2.1.17 (Hurwitz). *Given any irrational number ξ, there exist infinitely many rational numbers h/k such that*

$$\left|\xi - \frac{h}{k}\right| < \frac{1}{\sqrt{5}k^2}. \tag{2.1.10}$$

Proof. The idea is to establish that, of every three consecutive convergents of the simple continued fraction expansion of ξ, at least one satisfies the inequality (2.1.10).

Let q_n denote k_n/k_{n-1}. We first prove that

$$q_j + q_j^{-1} < \sqrt{5} \tag{2.1.11}$$

if (2.1.10) is false for both $h/k = h_{j-1}/k_{j-1}$ and $h/k = h_j/k_j$. Suppose (2.1.10) is false for these two values of h/k. We have

$$\left|\xi - \frac{h_{j-1}}{k_{j-1}}\right| + \left|\xi - \frac{h_j}{k_j}\right| \geq \frac{1}{\sqrt{5}k_{j-1}^2} + \frac{1}{\sqrt{5}k_j^2}.$$

But ξ lies between h_{j-1}/k_{j-1} and h_j/k_j and hence we find, using Theorem 2.1.5, that

$$\left|\xi - \frac{h_{j-1}}{k_{j-1}}\right| + \left|\xi - \frac{h_j}{k_j}\right| = \left|\frac{h_{j-1}}{k_{j-1}} - \frac{h_j}{k_j}\right| = \frac{1}{k_{j-1}k_j}.$$

Combining these results we get

$$\frac{k_j}{k_{j-1}} + \frac{k_{j-1}}{k_j} \leq \sqrt{5}.$$

Since the left side is rational we actually have a strict inequality, and (2.1.11) follows.

Now suppose (2.1.10) is false for $h/k = h_i/k_i$, $i = n-1, n, n+1$. We then have (2.1.11) for both $j = n$ and $j = n+1$. By Lemma 2.1.16 we see that $q_n^{-1} > \frac{1}{2}(\sqrt{5}-1)$ and $q_{n+1} < \frac{1}{2}(\sqrt{5}+1)$, and, by (2.1.6) we find $q_{n+1} = a_{n+1} + q_n^{-1}$. This gives us

$$\begin{aligned}\frac{1}{2}(\sqrt{5}+1) > q_{n+1} = a_{n+1} + q_n^{-1} &> a_{n+1} + \frac{1}{2}(\sqrt{5}-1) \\ &\geq 1 + \frac{1}{2}(\sqrt{5}-1) = \frac{1}{2}(\sqrt{5}+1)\end{aligned}$$

and this is a contradiction. □

Theorem 2.1.18. *The constant* $\sqrt{5}$ *in Theorem 2.1.17 is best possible, i.e., Theorem 2.1.17 does not hold if* $\sqrt{5}$ *is replaced by any larger value.*

Proof. It suffices to exhibit an irrational number ξ for which $\sqrt{5}$ is the largest possible constant. Consider the irrational ξ whose continued fraction expansion is $\langle 1, 1, 1, \dots \rangle$. We see that

$$\xi = 1 + \frac{1}{\langle 1, 1, \dots \rangle} = 1 + \frac{1}{\xi}, \quad \xi^2 = \xi + 1, \quad \xi = \frac{1}{2}(\sqrt{5} + 1).$$

Using (2.1.7) we can prove by induction that $\xi_i = (\sqrt{5} + 1)/2$ for all $i \geq 0$, for if $\xi_i = (\sqrt{5} + 1)/2$ then

$$\xi_{i+1} = (\xi_i - a_i)^{-1} = \left(\frac{1}{2}(\sqrt{5} + 1) - 1 \right)^{-1} = \frac{1}{2}(\sqrt{5} + 1).$$

A simple calculation yields $h_0 = k_0 = k_1 = 1$, $h_1 = k_2 = 2$. Equation (2.1.6) becomes $h_i = h_{i-1} + h_{i-2}$, $k_i = k_{i-1} + k_{i-2}$, and so by induction $k_n = h_{n-1}$ for $n \geq 1$. Hence we have

$$\lim_{n \to \infty} \frac{k_{n-1}}{k_n} = \lim_{n \to \infty} \frac{k_{n-1}}{h_{n-1}} = \frac{1}{\xi} = \frac{\sqrt{5} - 1}{2}$$

$$\lim_{n \to \infty} \left(\xi_{n+1} + \frac{k_{n-1}}{k_n} \right) = \frac{\sqrt{5} + 1}{2} + \frac{\sqrt{5} - 1}{2} = \sqrt{5}.$$

If c is any constant exceeding $\sqrt{5}$, then

$$\xi_{n+1} + \frac{k_{n-1}}{k_n} > c$$

holds for only a finite number of values of n. Thus, by (2.1.9),

$$\left| \xi - \frac{h_n}{k_n} \right| = \frac{1}{k_n^2(\xi_{n+1} + k_{n-1}/k_n)} < \frac{1}{ck_n^2}$$

holds for only a finite number of values of n. Thus there are only a finite number of rational numbers h/k satisfying $|\xi - h/k| < 1/(ck^2)$, because any such h/k is one of the convergents to ξ by Theorem 2.1.14. $\square$

2.1.7 Periodic Continued Fractions

An infinite simple continued fraction $\langle a_0, a_1, a_2, \dots \rangle$ is said to be *periodic* if there is an integer n such that $a_r = a_{n+r}$ for all sufficiently large integers r. Thus a periodic continued fraction can be written in the form

$$\begin{aligned}&\langle b_0, b_1, b_2, \ldots, b_j, a_0, a_1, \ldots, a_{n-1}, a_0, a_1, \ldots, a_{n-1}, \ldots\rangle\\&\quad = \langle b_0, b_1, b_2, \ldots, b_j, \overline{a_0, a_1, \ldots, a_{n-1}}\rangle\end{aligned} \tag{2.1.12}$$

where the bar over the $a_0, a_1, \ldots, a_{n-1}$ indicates that this block of integers is repeated indefinitely. For example $\langle\overline{2,3}\rangle$ denotes $\langle 2,3,2,3,2,3,\ldots\rangle$ and its value is easily computed. Writing θ for $\langle\overline{2,3}\rangle$ we have

$$\theta = 2 + \cfrac{1}{3 + \cfrac{1}{\theta}}.$$

This is a quadratic equation in θ, and we discard the negative root to obtain the value $\theta = (3+\sqrt{15})/3$. As a second example consider $\langle 4, 1, \overline{2,3}\rangle$. Calling this ξ, we have $\xi = \langle 4, 1, \theta\rangle$, with θ as above, and so

$$\xi = 4 + (1+\theta^{-1})^{-1} = 4 + \frac{\theta}{\theta+1} = \frac{29+\sqrt{15}}{7}.$$

These two examples illustrate the following result (see [159]).

Theorem 2.1.19. *Any periodic simple continued fraction is a quadratic irrational number, and conversely.*

Proof. Let us write ξ for the periodic continued fraction of (2.1.12) and θ for its purely periodic part,

$$\theta = \langle\overline{a_0, a_1, \ldots, a_{n-1}}\rangle = \langle a_0, a_1, \ldots, a_{n-1}, \theta\rangle.$$

Then equation (2.1.8) gives

$$\theta = \frac{\theta h_{n-1} + h_{n-2}}{\theta k_{n-1} + k_{n-2}}$$

and this is a quadratic equation in θ. Hence θ is either a quadratic irrational number or a rational number, but the latter is ruled out by Theorem 2.1.7. Now ξ can be written in terms on θ,

$$\xi = \langle b_0, b_1, \ldots, b_j, \theta\rangle = \frac{\theta m + m'}{\theta q + q'}$$

where m'/q' and m/q are the last two convergents to $\langle b_0, b_1, \ldots, b_j\rangle$. But θ is of the form $(a+\sqrt{b})/c$, and hence ξ is of similar form because, as with θ, we can rule out the possibility that ξ is rational.

To prove the converse, let us begin with any quadratic irrational ξ, or ξ_0, of the form $\xi = \xi_0 = (a+\sqrt{b})/c$, with integers $a, b, c > 0$, $c \neq 0$. The integer b is not a perfect square since ξ is irrational. We multiply numerator and denominator by $|c|$ to get

$$\xi_0 = \frac{ac + \sqrt{bc^2}}{c^2} \quad \text{or} \quad \xi_0 = \frac{-ac + \sqrt{bc^2}}{-c^2}$$

according as c is positive or negative. Thus we can write ξ in the form

$$\xi_0 = \frac{m_0 + \sqrt{d}}{q_0}$$

where $q_0 | (d - m_0^2), d, m_0$ and q_0 are integers, $q_0 \neq 0$, d not a perfect square. By writing ξ_0 in this form we can get a simple formulation of its continued fraction expansion $\langle a_0, a_1, a_2, \dots \rangle$. We will prove that the equations

$$a_i = [\xi_i], \quad \xi_i = \frac{m_i + \sqrt{d}}{q_i}$$
$$m_{i+1} = a_i q_i - m_i, \quad q_{i+1} = \frac{d - m_{i+1}^2}{q_i} \tag{2.1.13}$$

define infinite sequences of integers m_i, q_i, a_i, and irrationals ξ_i in such a way that equations (2.1.7) hold, and hence we will have the continued fraction expansion of ξ_0.

In the first step, we start with ξ_0, m_0, q_0 as determined above, and we let $a_0 = [\xi_0]$. If ξ_i, m_i, q_i, a_i are known, then we take $m_{i+1} = a_i q_i - m_i$, $q_{i+1} = (d - m_{i+1}^2)/q_i$, $\xi_{i+1} = (m_{i+1} + \sqrt{d})/q_{i+1}$, $a_{i+1} = [\xi_{i+1}]$. That is, (2.1.13) actually does determine sequences ξ_i, m_i, q_i, a_i.

Now we use induction to prove that the m_i and q_i are integers such that $q_i \neq 0$ and $q_i | (d - m_i^2)$.

Next we can verify that

$$\xi_i - a_i = \frac{-a_i q_i + m_i + \sqrt{d}}{q_i} = \frac{\sqrt{d} - m_{i+1}}{q_i} = \frac{d - m_{i+1}^2}{q_i(\sqrt{d} + m_{i+1})}$$
$$= \frac{q_{i+1}}{\sqrt{d} + m_{i+1}} = \frac{1}{\xi_{i+1}}$$

which verifies (2.1.7) and so we have proved that $\xi_0 = \langle a_0, a_1, a_2, \dots \rangle$, with a_i defined by (2.1.13).

Let $\xi_i' = (m_i - \sqrt{d})/q_i$, the conjugate of ξ_i. We get the equation

$$\xi_0' = \frac{\xi_n' h_{n-1} + h_{n-2}}{\xi_n' k_{n-1} + k_{n-2}}$$

by taking conjugates in (2.1.8). Solving for ξ_n' we obtain

$$\xi_n' = -\frac{k_{n-2}}{k_{n-1}}\left(\frac{\xi_0' - h_{n-2}/k_{n-2}}{\xi_0' - h_{n-1}/k_{n-1}}\right).$$

As n tends to infinity, both h_{n-1}/k_{n-1} and h_{n-2}/k_{n-2} tend to ξ_0, which is different from ξ_0', and hence the fraction in parentheses tends to 1. Thus for sufficiently large n, say $n > N$ where N is fixed, the fraction in parentheses is positive, and ξ_n' is negative. But ξ_n is positive for $n \geq 1$ and hence $\xi_n - \xi_n' > 0$ and $n > N$. Applying (2.1.13) we see that this gives $2\sqrt{d}/q_n > 0$ and hence $q_r > 0$ for $n > N$.

It also follows from (2.1.13) that

$$q_n q_{n+1} = d - m_{n+1}^2 \leq d, \quad q_n \leq q_n q_{n+1} \leq d$$
$$m_{n+1}^2 < m_{n+1}^2 + q_n q_{n+1} = d, \quad |m_{n+1}| < \sqrt{d}$$

for $n > N$. Since d is a fixed positive integer we conclude that q_n and m_{n+1} can assume only a fixed number of possible values for $n > N$. Hence the ordered pairs (m_n, q_n) can assume only a fixed number of possible pair values for $n > N$, and so there are distinct integers j and k such that $m_j = m_k$ and $q_j = q_k$. We can suppose we have chosen j and k so that $j < k$. By (2.1.13) this implies that $\xi_j = \xi_k$ and hence that

$$\xi_0 = \langle a_0, a_1, \ldots, a_{j-1}, \overline{a_j, a_{j+1}, \ldots, a_{k-1}}\rangle,$$

and we are done. □

The following result describes the subclass of real quadratic irrationals that have purely periodic continued fraction expansions, that is, expressions of the form $\langle\overline{a_0, a_1, \ldots, a_n}\rangle$ (see [159]).

Theorem 2.1.20. *The continued fraction expansion of the real quadratic irrational number ξ is purely periodic if and only if $\xi > 1$ and $-1 < \xi' < 0$, where ξ' denotes the conjugate of ξ.*

Proof. First we assume that $\xi > 1$ and $-1 < \xi' < 0$. As usual, we write ξ_0 for ξ and take conjugates in (2.1.7) to obtain

$$\frac{1}{\xi_{i+1}'} = \xi_i' - a_i. \tag{2.1.14}$$

Now $a_i \geq 1$ for all i, even for $i = 0$, since $\xi_0 > 1$. Hence if $\xi_i' < 0$, then $1/\xi_{i+1}' < -1$, and we have $-1 < \xi_{i+1}' < 0$. Since $-1 < \xi_0' < 0$, we see, by mathematical induction, that $-1 < \xi_i' < 0$ holds for all $i \geq 0$. Then, since $\xi_i' = a_i + 1/\xi_{i+1}'$ by (2.1.14), we have

$$0 < -\frac{1}{\xi'_{i+1}} - a_i < 1, \quad a_i = \left[-\frac{1}{\xi'_{i+1}}\right].$$

Now ξ is a quadratic irrational, so $\xi_j = \xi_k$ for some integers j and k with $0 < j < k$. Then we have $\xi'_j = \xi'_k$ and

$$a_{j-1} = \left[-\frac{1}{\xi'_j}\right] = \left[-\frac{1}{\xi'_k}\right] = a_{k-1}$$

$$\xi_{j-1} = a_{j-1} + \frac{1}{\xi_j} = a_{k-1} + \frac{1}{\xi_k} = \xi_{k-1}.$$

Thus $\xi_j = \xi_k$ implies $\xi_{j-1} = \xi_{k-1}$. A j-fold iteration of this implication gives us $\xi_0 = \xi_{k-j}$, and we have

$$\xi = \xi_0 = \langle\overline{a_0, a_1, \ldots, a_{k-j-1}}\rangle.$$

To prove the converse, let us assume that ξ is purely periodic, say $\xi = \langle\overline{a_0, a_1, \ldots, a_{n-1}}\rangle$. where $a_0, a_1, \ldots, a_{n-1}$ are positive integers. Then $\xi > a_0 \geq 1$. Also, by (2.1.8) we have

$$\xi = \langle a_0, a_1, \ldots, a_{n-1}, \xi\rangle = \frac{\xi h_{n-1} + h_{n-2}}{\xi k_{n-1} + k_{n-2}}.$$

Thus ξ satisfies the equation

$$f(x) = x^2 k_{n-1} + x(k_{n-2} - h_{n-1}) - h_{n-2} = 0.$$

This quadratic equation has two roots, ξ and its conjugate ξ'. Since $\xi > 1$, we need to only prove that $f(x)$ has a root between -1 and 0 in order to establish that $-1 < \xi' < 0$. We will do this by showing that $f(-1)$ and $f(0)$ have opposite signs. First we observe that $f(0) = -h_{n-2} < 0$ by (2.1.6), since $a_i > 0$ for $i \geq 0$. Next we see that for $n \geq 1$

$$\begin{aligned} f(-1) &= k_{n-1} - k_{n-2} + h_{n-1} - h_{n-2} \\ &= (k_{n-2} + h_{n-2})(a_{n-1} - 1) + k_{n-3} + h_{n-3} \geq k_{n-3} + h_{n-3} > 0. \end{aligned}$$

□

We now turn to the continued fraction expansion of $\sqrt{d}$ for a positive integer d not a perfect square. We get at this by considering the closely related irrational number $\sqrt{d} + [\sqrt{d}]$. This number satisfies the conditions of Theorem 2.1.20, and so its continued fraction is purely periodic,

$$\sqrt{d} + [\sqrt{d}] = \langle\overline{a_0, a_1, \ldots, a_{r-1}}\rangle = \langle a_0, \overline{a_1, \ldots, a_{r-1}, a_0}\rangle. \tag{2.1.15}$$

We can suppose that we have chosen r to be the smallest integer for which $\sqrt{d} + [\sqrt{d}]$ has an expansion of the form (2.1.15). Now we note that $\xi_i = \langle a_i, a_{i+1}, \dots \rangle$ is purely periodic for all values of i, and that $\xi_0 = \xi_r = \xi_{2r} = \dots$. Furthermore, $\xi_1, \xi_2, \dots, \xi_{r-1}$ are all different from ξ_0, since otherwise there would be a shorter period. Thus $\xi_i = \xi_0$ if and only if i is of the form mr.

Now we can start with $\xi_0 = \sqrt{d} + [\sqrt{d}]$, $q_0 = 1$, $m_0 = [\sqrt{d}]$ in (2.1.13) because $1|(d - [\sqrt{d}]^2)$. Then, for all $j \geq 0$, we have

$$\frac{m_{jr} + \sqrt{d}}{q_{jr}} = \xi_{jr} = \xi_0 = \frac{m_0 + \sqrt{d}}{q_0} = [\sqrt{d}] + \sqrt{d}$$

$$m_{jr} - q_{jr}[\sqrt{d}] = (q_{jr} - 1)\sqrt{d} \tag{2.1.16}$$

and hence $q_{jr} = 1$, since the left side is rational and $\sqrt{d}$ is irrational. Moreover $q_i = 1$ for no other values of the subscript i. For $q_i = 1$ implies $\xi_i = m_i + \sqrt{d}$, but ξ_i has a purely periodic expansion so that, by Theorem 2.1.20 we have $-1 < m_i - \sqrt{d} < 0$, $\sqrt{d} - 1 < m_i < \sqrt{d}$, and hence $m_i = [\sqrt{d}]$. Thus $\xi_i = \xi_0$ and i is a multiple of r.

We also establish that $q_i = -1$ does not hold for any i. For $q_i = -1$ implies $\xi_i = -m_i - \sqrt{d}$ by (2.1.13), and by Theorem 2.1.20 we would have $-m_i - \sqrt{d} > 1$ and $-1 < -m_i + \sqrt{d} < 0$. But this implies $\sqrt{d} < m_i < -\sqrt{d} - 1$, which is impossible.

Noting that $a_0 = [\sqrt{d} + [\sqrt{d}]] = 2[\sqrt{d}]$, we can now turn to the case $\xi = \sqrt{d}$. Using (2.1.15) we have

$$\begin{aligned}
\sqrt{d} &= -[\sqrt{d}] + (\sqrt{d} + [\sqrt{d}]) \\
&= -[\sqrt{d}] + \langle 2[\sqrt{d}], \overline{a_1, a_2, \dots, a_{r-1}, a_0} \rangle \\
&= \langle [\sqrt{d}], \overline{a_1, a_2, \dots, a_{r-1}, a_0} \rangle
\end{aligned}$$

with $a_0 = 2[\sqrt{d}]$.

When we apply (2.1.13) to $\sqrt{d} + [\sqrt{d}]$, $q_0 = 1$, $m_0 = [\sqrt{d}]$ we have $a_0 = 2[\sqrt{d}]$, $m_1 = [\sqrt{d}]$, $q_1 = d - [\sqrt{d}]^2$. But we can also apply (2.1.13) to $\sqrt{d}$ with $q_0 = 1$, $m_0 = 0$, and we find $a_0 = [\sqrt{d}]$, $m_1 = [\sqrt{d}]$, $q_1 = d - [\sqrt{d}]^2$. The value of a_0 is different, but the values of m_1, and of q_1, are the same in both cases. Since $\xi_i = (m_i + \sqrt{d})/q_i$ we see that further application of (2.1.13) yields the same values for the a_i, for the m_i, and for the q_i, in both cases. In other words, the expansions of $\sqrt{d} + [\sqrt{d}]$ and $\sqrt{d}$ differ only in the values of a_0 and m_0. Stating our results explicitly for the case $\sqrt{d}$, we have the following theorem.

Theorem 2.1.21. *If the positive integer d is not a perfect square, the simple continued fraction expansion of $\sqrt{d}$ has the form*

$$\sqrt{d} = \langle a_0, \overline{a_1, a_2, \dots, a_{r-1}, 2a_0} \rangle$$

with $a_0 = [\sqrt{d}]$. *Furthermore, with* $\xi_0 = \sqrt{d}$, $q_0 = 1$, $m_0 = 0$, *in equations* (2.1.13), *we have* $q_i = 1$ *if and only if* $r|i$, *and* $q_i = -1$ *holds for no subscript* i. *Here* r *denotes the length of the shortest period in the expansion of* $\sqrt{d}$.

2.2 Units and Norms in Quadratic Rings

2.2.1 Quadratic Rings

Let R be the commutative ring (see [42] and [54])

$$R = \{m + n\sqrt{D} : \ m, n \in \mathbb{Z}\} \tag{2.2.1}$$

where D is a positive that is not a perfect square, endowed with the standard operations induced from the ring of integers $(\mathbb{Z}, +, \cdot)$. An element $\varepsilon \in R$ is called a *unit* in R if it is inversable, that is there exists $\varepsilon_1 \in R$ such that $\varepsilon\varepsilon_1 = \varepsilon_1\varepsilon = 1$. Two elements $\alpha, \beta \in R$ are said to be *divisibility associated* if there exists a unit $\varepsilon \in R$ such that $\alpha = \varepsilon\beta$. We will adopt the notation $\alpha \sim \beta$ to indicate that α and β have the property above. It is not difficult to see that "$\sim$" is an equivalence relation.

If $\mu \in R$, $\mu = a + b\sqrt{D}$, we will denote by $\overline{\mu}$ the element $\overline{\mu} = a - b\sqrt{D}$ and will call it the *conjugate* of μ.

2.2.2 Norms in Quadratic Rings

Let us denote by $N : R \to \mathbb{Z}$ the following function: if $\mu = a + b\sqrt{D}$, then

$$N(\mu) = a^2 - Db^2 = \mu \cdot \overline{\mu}. \tag{2.2.2}$$

Proposition 2.2.1 (N Is Multiplicative). *For all* $\mu_1, \mu_2 \in R$, *the following relation holds:*

$$N(\mu_1\mu_2) = N(\mu_1)N(\mu_2).$$

Proof. If $\mu_1 = m_1 + n_1\sqrt{D}$ and $\mu_2 = m_2 + n_2\sqrt{D}$, then we have

$$\mu_1\mu_2 = (m_1m_2 + Dn_1n_2) + (m_1n_2 + m_2n_1)\sqrt{D}$$

and

$$N(\mu_1\mu_2) = (m_1m_2 + Dn_1n_2)^2 - D(m_1n_2 + m_2n_1)^2$$
$$= m_1^2m_2^2 + D^2n_1^2n_2^2 - Dm_1^2n_2^2 - Dm_2^2n_1^2 = m_1^2(m_2^2 - Dn_2^2) - Dn_1^2(m_2^2 - dn_2^2)$$
$$= (m_1^2 - Dn_1^2)(m_2^2 - Dn_2^2) = N(\mu_1)N(\mu_2).$$

□

Proposition 2.2.2. *An element $\varepsilon \in R$ is an unit in R if and only if $N(\varepsilon) = \pm 1$.*

Proof. If ε is a unit in R, then there exists $\varepsilon_1 \in R$ such that $\varepsilon\varepsilon_1 = 1$. Then from Proposition 2.2.1, $N(\varepsilon)N(\varepsilon_1) = N(1) = 1^2 - D0^2 = 1$. Since $N(\varepsilon)$ and $N(\varepsilon_1)$ are integers, it follows that $N(\varepsilon) = \pm 1$. Conversely, if $N(\varepsilon) = \pm 1$, then (2.2.2) yields $\varepsilon\overline{\varepsilon} = \pm 1$. If $N(\varepsilon) = 1$, then $\varepsilon\overline{\varepsilon} = 1$ and if $N(\varepsilon) = -1$, then $\varepsilon(-\overline{\varepsilon}) = 1$. Both cases show that ε is a unit in R. $\square$

Theorem 2.2.3. *For any integer a, the cardinal number of the set*

$$S = \{\alpha \in R : \ N(\alpha) = a \text{ and } \alpha \not\sim \beta \text{ for all } \beta \in R, \ \beta \neq \alpha\} \tag{2.2.3}$$

is finite and does not exceed a^2.

Proof. If $a = 0$, then the cardinal number of S is 1. We may assume now that a is nonzero. Let $\alpha, \beta \in S$ such that $\alpha \neq \beta$ and $\alpha \equiv \beta (\mathrm{mod}\ a)$. This means that there exists $\gamma \in R$ such that $\alpha - \beta = a\gamma$.

From the definition of the set S it follows that $a = N(\alpha) = N(\beta)$, hence $\alpha - \beta = a\gamma = N(\alpha)\gamma = N(\beta)\gamma$.

Now embed the ring R into the field $\mathbb{Q}(\sqrt{D}) = \{r + s\sqrt{D} : \ r, s \in \mathbb{Q}\}$. Since $N(\alpha) = N(\beta) = a \neq 0$, we have $\alpha, \beta \neq 0$ and

$$\frac{\alpha}{\beta} = \frac{\beta + a\gamma}{\beta} = 1 + \frac{N(\beta)\gamma}{\beta} = 1 + \frac{\beta\overline{\beta}\gamma}{\beta} = 1 + \overline{\beta}\gamma$$

and

$$\frac{\beta}{\alpha} = \frac{\alpha - a\gamma}{\alpha} = 1 - \frac{N(\alpha)\gamma}{\alpha} = 1 - \frac{\alpha\overline{\alpha}\gamma}{\alpha} = 1 - \overline{\alpha}\gamma.$$

The computations above show that

$$\frac{\alpha}{\beta} - \frac{\beta}{\alpha} = (\overline{\beta} - \overline{\alpha})\gamma$$

hence $\dfrac{\alpha}{\beta}, \dfrac{\beta}{\alpha} \in R$ and $\alpha \sim \beta$, in contradiction with the definition of S. It follows that $\alpha \equiv \beta (\mathrm{mod}\ a)$, for all $\alpha, \beta \in S$.

On the other hand, it is not difficult to see that for all b in R there exist positive integers m, n such that $0 \leq m < |a|$, $0 \leq n < |a|$, and $b \equiv m + n\sqrt{D} (\mathrm{mod}\ a)$.

The considerations above show that the mapping

$$S \to \{0, 1, 2, \ldots, |a| - 1\} \times \{0, 1, 2, \ldots, |a| - 1\}$$

given by $\alpha \to (m, n)$, where $0 \leq m, \ n \leq |a| - 1$, $\alpha \equiv m + n\sqrt{D} (\mathrm{mod}\ a)$, is one-to-one.

This means that the set S is finite and its cardinal number is less or equal to a^2. $\square$

Proposition 2.2.4 (The Conjugate Is Multiplicative). *For all $\mu_1, \mu_2 \in R$, the following relation holds:*

$$\overline{\mu_1 \mu_2} = \overline{\mu}_1 \overline{\mu}_2. \tag{2.2.4}$$

Proof. If $\mu_1 = m_1 + n_1\sqrt{D}$ and $\mu_2 = m_2 + n_2\sqrt{D}$, then

$$\mu_1\mu_2 = (m_1m_2 + Dn_1n_2) + (m_1n_2 + m_2n_1)\sqrt{D}$$

and

$$\begin{aligned}\overline{\mu_1\mu_2} &= (m_1m_2 + Dn_1n_2) - (m_1n_2 + m_2n_1)\sqrt{D}\\ &= (m_1 - n_1\sqrt{D})(m_2 - n_2\sqrt{D}) = \overline{\mu}_1\overline{\mu}_2.\end{aligned}$$

□

Remark. Proposition 2.2.4 gives another proof of the fact that N is multiplicative. Indeed, we have

$$N(\mu_1\mu_2) = (\mu_1\mu_2)(\overline{\mu_1\mu_2}) = (\mu_1\mu_2)(\overline{\mu}_1\overline{\mu}_2) = (\mu_1\overline{\mu}_1)(\mu_2\overline{\mu}_2) = N(\mu_1)N(\mu_2).$$