Continued fractions and Pell's equation

Nirjhar Nath
nirjhar@cmi.ac.in

These are my notes of a reading project on Continued fractions and Pell's equation under Dr. Rupam Barman (website) of IIT Guwahati. The book that I used throughout is "An Introduction to the Theory of Numbers" by Ivan Niven, Herbert S. Zuckerman and Hugh L. Montgomery.

Contents

1 Continued fractions

1.1 Finite continued fractions

We shall describe the function

$$\langle x_0, x_1, \ldots, x_j \rangle = x_0 + \cfrac{1}{x_1 + \cfrac{1}{x_2 + \ddots + \cfrac{1}{x_{j-1} + \cfrac{1}{x_j}}}}$$

in $j+1$ variables $x_0, x_1, \ldots, x_j \in \mathbb{R}$ as a *finite continued fraction*, or, when there is no risk of ambiguity, simply as a *continued fraction*. Such a finite continued fraction is called *simple* if all the x_i's are integers. It is obvious that

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

Below we see the simple continued fraction expansion of rational numbers.

1.2 The Euclidean algorithm

Given any rational number u_0/u_1 so that $(u_0, u_1) = 1$ and $u_1 > 0$, by Euclidean algorithm, we have

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

We write $\xi_i = u_i/u_{i+1}$ for all values in the range $0 \leq i \leq j$, the equations (1) become

$$\xi_i = a_i + \frac{1}{\xi_{i+1}}, \quad 0 \leq i \leq j-1; \quad \xi_j = a_j \tag{2}$$

Taking the first two of the equations of (2), i.e. those for which $i = 0$ and $i = 1$, and eliminate ξ_1, we have

$$\xi_0 = a_0 + \frac{1}{a_1 + \frac{1}{\xi_2}}$$

Here we replace ξ_2 by its value from (2) and then continue replacing $\xi_3, \xi_4, \ldots$ to get

$$\frac{u_0}{u_1} = \xi_0 = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_{j-1} + \cfrac{1}{a_j}}}} = \langle a_0, a_1, \ldots, a_j \rangle \tag{3}$$

This is a continued fraction expansion of $\xi_0 = u_0/u_1$. The integers a_i are called the *partial quotients* since they are the quotients in the repeated application of the division algorithm in equations (1).

1.3 Uniqueness

We saw that any such 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 (3) has an alternate form,

$$\frac{u_0}{u_1} = \langle a_0, a_1, \ldots, a_{j-1}, a_j\rangle = \langle a_0, a_1, \ldots, a_{j-1}, a_j - 1, 1\rangle \tag{4}$$

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

Theorem 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 $i = 1, 2, \ldots, n$.*

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

$$y_i = \langle b_i, b_{i+1}, \ldots, b_n\rangle = b_i + \frac{1}{\langle b_{i+1}, b_{i+2}, \ldots, b_n\rangle} = b_i + \frac{1}{y_{i+1}} \tag{5}$$

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, we have $b_i = [y_i]$ for all values of i in the range $0 \le i \le n$. Using the notation of equation (3), we have $y_0 = \xi_0$. We have, $\xi_i = u_i/u_{i+1} > 1$ for all values of $i > 0$ and so $a_i = [\xi_i]$ for $0 \le i \le j$. Now, $b_0 = [y_0] = [\xi_0] = a_0$. By equations (2) and (5), we have

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

We use induction as follows. Assume that $\xi_k = y_k$ and $a_k = b_k$. We use equations (2) and (5) again to write

$$\begin{aligned} &\frac{1}{\xi_{k+1}} = \xi_k - a_k = y_k - b_k = \frac{1}{y_{k+1}} \implies \xi_{k+1} = y_{k+1}, \\ &a_{k+1} = [\xi_{k+1}] = [y_{k+1}] = b_{k+1} \end{aligned}$$

It must also follow that the continued fractions have the same length, i.e., that $j = n$, because if $j < n$ then by equation (2), we have $\xi_j = a_j$ and by equation (5), we have $y_j > b_j$ which contradicts the fact that $\xi_j = y_j, a_j = b_j$. Similar argument holds for $j > n$, and thus $j = n$. □

Theorem 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.*

Proof. The first assertion can be established by induction on the number of terms in the continued fraction, by use of the formula

$$\langle a_0, a_1, \ldots, a_j\rangle = a_0 + \frac{1}{\langle a_1, a_2, \ldots, a_j\rangle} = \frac{a_0(\langle a_1, a_2, \ldots, a_j\rangle) + 1}{\langle a_1, a_2, \ldots, a_j\rangle}$$

The second assertion follows from the development of u_0/u_1 into a finite simple continued fraction in Section 1.2, together with equation (4) and Theorem 1. □

1.4 Infinite continued fractions

Let $a_0, a_1, a_2, \dots$ be an infinite sequence with $a_0 \in \mathbb{Z}$ and $a_1, a_2, \dots \in \mathbb{Z}^+$. We define two sequences of integers $\{h_n\}$ and $\{k_n\}$ inductively as follows:

$$\begin{aligned} h_{-2} = 0, h_{-1} = 1, h_i &= a_i h_{i-1} + h_{i-2} \text{ for } i \geq 0 \\ k_{-2} = 1, k_{-1} = 0, k_i &= a_i k_{i-1} + k_{i-2} \text{ for } i \geq 0 \end{aligned} \tag{6}$$

We note that $k_0 = 1, k_1 = a_1 \geq 1 = k_0, k_2 > k_1, k_3 > k_2$ so that

$$1 = k_0 \leq k_1 < k_2 < k_3 < \cdots < k_n < \cdots$$

Theorem 3. *For any $x \in \mathbb{R}^+$,*

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

Proof. For $n = 0$, we have the equation

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

which is true by equations (6). We have,

$$\langle a_0, x \rangle = a_0 + \frac{1}{x} = \frac{x a_0 + 1}{x} = \frac{x h_0 + h_{-1}}{x k_0 + k_{-1}}$$

i.e., the theorem is true for $n = 1$. We establish the theorem in general by induction. Assuming that the theorem holds for $\langle a_0, a_1, \dots, a_{n-1}, x \rangle$, we have

$$\begin{aligned} \langle a_0, a_1, \dots, a_n, x \rangle &= \left\langle a_0, a_1, \dots, a_{n-1}, a_n + \frac{1}{x} \right\rangle \\ &= \frac{(a_n + 1/x) h_{n-1} + h_{n-2}}{(a_n + 1/x) k_{n-1} + k_{n-2}} \\ &= \frac{x(a_n h_{n-1} + h_{n-2}) + h_{n-1}}{x(a_n k_{n-1} + k_{n-2}) + k_{n-1}} \\ &= \frac{x h_n + h_{n-1}}{x k_n + k_{n-1}} \end{aligned}$$

and hence the theorem is proved. □

Theorem 4. *If $r_n \stackrel{\text{def}}{=} \langle a_0, a_1, \dots, a_n \rangle \ \forall\ n \geq 0$, then $r_n = h_n / k_n$.*

Proof. Using Theorem 3 and equations (6), we have

$$r_n = \langle a_0, a_1, \dots, a_n \rangle = \frac{a_n h_{n-1} + h_{n-2}}{a_n k_{n-1} + k_{n-2}} = \frac{h_n}{k_n}$$

and we are done. □

We call $\langle a_0, a_1, \dots, a_n \rangle = h_n / k_n = r_n$ the n^{th} *convergent* to the infinite continued fraction $\langle a_0, a_1, a_2, \dots \rangle$. In the case of a finite simple continued fraction, we similarly call the number $\langle a_0, a_1, \dots, a_n \rangle$ $(0 \leq n \leq j)$ the n^{th} *convergent* to $\langle a_0, a_1, \dots, a_j \rangle$.

Theorem 5. *The following equations hold for $i \geq 1$:*

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

The fraction h_i/k_i is reduced, i.e., $(h_i, k_i) = 1$.

Proof. The equations (6) imply that $h_{-1}k_{-2} - h_{-2}k_{-1} = 1$. We use induction. Assuming that $h_i k_{i-1} - h_{i-1} k_i = (-1)^{i-1}$ and using equations (6), we have

$$h_{i+1}k_i - h_i k_{i+1} = (a_{i+1}h_i + h_{i-1})k_i - h_i(a_{i+1}k_i + k_{i-1}) = -(h_i k_{i-1} - h_{i-1}k_i) = (-1)^i$$

This proves the first result and dividing this by $k_i k_{i-1}$ and using Theorem 4, we get the second result. The third result is proved below.

$$\begin{aligned} h_i k_{i-2} - h_{i-2} k_i &= (a_i h_{i-1} + h_{i-2})k_{i-2} - h_{i-2}(a_i k_{i-1} + k_{i-2}) \\ &= (h_{i-1}k_{i-2} - h_{i-2}k_{i-1})a_i \\ &= (-1)^{i-2} a_i = (-1)^i a_i \end{aligned}$$

Dividing the third result by $k_i k_{i-2}$ and using Theorem 4, we get the fourth result. Furthermore, the fraction h_i/k_i is reduced since by the first result, any common factor of h_i and k_i is also a factor of $(-1)^{i-1}$. □

Theorem 6. *The even convergents r_{2m} increase strictly with m, while the odd convergents r_{2m+1} decrease strictly, and every odd convergent is greater than any even convergent, i.e., the values r_n satisfy the infinite chain of inequalities*

$$r_0 < r_2 < r_4 < r_6 < \cdots < r_7 < r_5 < r_3 < r_1$$

and every r_{2p} is less than every r_{2q-1}. Furthermore, $\lim_{n\to\infty} r_n$ exists and for every $m \geq 0$,

$$r_{2m} < \lim_{n\to\infty} r_n < r_{2m+1}$$

Proof. Since $a_i > 0$ and $k_i > 0$ for $i \geq 1$ and $i \geq 0$ respectively, thus using the second and fourth results of Theorem 5, we have

$$r_{2m} < r_{2m-1}, r_{2m} < r_{2m+2} \text{ and } r_{2m-1} > r_{2m+1}$$

Using these results, we prove that $r_{2p} < r_{2q-1}$ as follows.

$$r_{2p} < r_{2p+2q} < r_{2p+2q-1} \leq r_{2q-1}$$

Thus, we have proved the desired infinite chain of inequalities.

The sequence $\{r_{2m}\}$ is monotonically increasing and is bounded above by r_1, and so $\lim_{m\to\infty} r_{2m}$ exists. Analogously, $\{r_{2m+1}\}$ is monotonously decreasing and is bounded above by r_0, and so $\lim_{m\to\infty} r_{2m+1}$ also exists. Also, $k_i \geq i \ \forall \ i \geq 1$ since

$$1 = k_0 \leq k_1 < k_2 < k_3 < \cdots < k_n < \cdots$$

and so by the second result of Theorem 5, we have

$$0 \leq r_{2m+1} - r_{2m} = \frac{(-1)^{2m}}{k_{2m+1}k_{2m}} \leq \frac{1}{2m(2m+1)}$$

As $m \to \infty$, $\dfrac{1}{2m(2m+1)} \to 0$ and so by Squeeze theorem, we have

$$\lim_{m\to\infty} r_{2m} = \lim_{m\to\infty} r_{2m+1}$$

and hence $\lim_{n\to\infty} r_n$ exists and $r_{2m} < \lim_{n\to\infty} r_n < r_{2m+1}$ for every $m \geq 0$. □

Definition 1. An infinite sequence $a_0, a_1, a_2, \dots$ with $a_0 \in \mathbb{Z}$ and $a_1, a_2, \dots \in \mathbb{Z}^+$ determines an infinite simple continued fraction $\langle a_0, a_1, a_2, \dots \rangle$ with value

$$\langle a_0, a_1, a_2, \dots \rangle \stackrel{\text{def}}{=} \lim_{n\to\infty} r_n$$

Theorem 7. *The value of any infinite simple continued fraction $\langle a_0, a_1, a_2, \dots \rangle$ is irrational.*

Proof. Writing $\theta = \langle a_0, a_1, a_2, \dots \rangle$, we observe by Theorem 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 5 that $|r_{n+1} - r_n| = 1/k_n k_{n+1}$, we have

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

Now suppose that θ were rational, say $\theta = a/b$ with $a, b \in \mathbb{Z}$, $b > 0$. Then multiplying the above equation by b, we have

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

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 1. *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 6, we see that $r_0 < \theta < r_1$, i.e., $a_0 < \theta < a_0 + 1/a_1$. Since $a_1 \geq 1$, so $a_0 < \theta < a_0 + 1$ and hence $[\theta] = a_0$. Also,

$$\begin{aligned}\theta &= \lim_{n\to\infty} \langle a_0, a_1, \dots, a_n \rangle = \lim_{n\to\infty} \left(a_0 + \frac{1}{\langle a_1, a_2, \dots, a_n \rangle} \right) \\ &= a_0 + \lim_{n\to\infty} \frac{1}{\langle a_1, a_2, \dots, a_n \rangle} = a_0 + \frac{1}{\theta_1}\end{aligned}$$

□

Theorem 8. *Two distinct infinite simple continued fractions converge to different values.*

Proof. Let $\langle a_0, a_1, a_2, \dots \rangle$ and $\langle b_0, b_1, b_2, \dots \rangle = \theta$. Then by Lemma 1, $a_0 = [\theta] = b_0$ 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, $a_n = b_n \ \forall\ n$. □

1.5 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,

$$a_i = [\xi_i], \qquad \xi_{i+1} = \frac{1}{\xi_i - a_i} \tag{7}$$

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

$$a_{i-1} < \xi_{i-1} < a_{i-1} + 1, \qquad 0 < \xi_{i-1} - a_{i-1} < 1,$$

$$\xi_i = \frac{1}{\xi_{i-1} - a_{i-1}} > 1, \qquad a_i = [\xi_i] \geq 1$$

Theorem 9. *With ξ_i as defined in equation (7), we have*

$$\langle a_0, a_1, \dots \rangle = \langle a_0, a_1, a_2, \dots, a_{n-1}, \xi_n \rangle = \xi \;\; \textit{and} \;\; \xi_n = \langle a_n, a_{n+1}, a_{n+2}, \dots \rangle$$

Proof. With repeated application of equation (7) in the form $\xi_i = a_i + 1/\xi_i$, we get

$$\begin{aligned} \xi = \xi_0 = a_0 + \frac{1}{\xi_1} &= \langle a_0, \xi_1 \rangle \\ &= \left\langle a_0, a_1 + \frac{1}{\xi_2} \right\rangle = \langle a_0, a_1, \xi_2 \rangle \\ &= \left\langle a_0, a_1, \dots, a_{n-2}, a_{n-1} + \frac{1}{\xi_n} \right\rangle = \langle a_0, a_1, \dots, a_{n-1}, \xi_n \rangle \end{aligned}$$

Now, to prove that $\xi = \xi_0$ is the value of the infinite continued fraction $\langle a_0, a_1, a_2 \dots \rangle$ determined by the integers a_i, we use Theorem 3 to write

$$\xi = \langle a_0, a_1, \dots, a_{n-1}, \xi_n \rangle = \frac{\xi_n h_{n-1} + h_{n-2}}{\xi_n k_{n-1} + k_{n-2}} \tag{8}$$

with h_i and k_i as defined in equations (6). By Theorem 5, we have

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

As $n \to \infty$, $\dfrac{(-1)^{n-1}}{k_{n-1}(\xi_n k_{n-1} + k_{n-2})} \to 0$ because $\{k_n\}$ is increasing and $\xi_n > 0$. Hence, $\xi - r_{n-1} \to 0$ as $n \to \infty$ and then by Definition 1, we have

$$\xi = \lim_{n\to\infty} r_n = \lim_{n\to\infty} \langle a_0, a_1, \dots, a_n \rangle = \langle a_0, a_1, a_2, \dots \rangle = \langle a_0, a_1, a_2, \dots, a_{n-1}, \xi_n \rangle$$

With repeated application of equation (7) for ξ_n, we get the other equation. □

1.6 Approximations to irrational numbers

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

Theorem 10. *We have for $n \geq 0$,*

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

Proof. Using the result

$$\xi - r_{n-1} = \frac{(-1)^{n-1}}{k_{n-1}(\xi_n k_{n-1} + k_{n-2})}$$

(where $r_n = h_n/k_n$) from the proof of Theorem 9 and also using equation (7), we have

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

Now using equation (6), we get

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

Multiplying this inequality by k_n, we get the second inequality. □

Theorem 11. *The convergents h_n/k_n are successively closer to ξ, i.e.,*

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

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

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

To prove the stronger inequality, we observe that by equation (7), $a_n + 1 > \xi_n$ and therefore by equation (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 along with the inequality

$$\xi - \frac{h_{n-1}}{k_{n-1}} = \frac{(-1)^{n-1}}{k_{n-1}(\xi_n k_{n-1} + k_{n-2})}$$

gives the following inequality

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

Multiplying by k_{n-1} and using Theorem 10, we get

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

□

This means that 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.

Theorem 12. *If a/b is a rational number with $b > 0$ such that*

$$\left|\xi - \frac{a}{b}\right| < \left|\xi - \frac{h_n}{k_n}\right|$$

for some $n \geq 1$, then $b > k_n$. In fact if $|\xi b - a| < |\xi k_n - h_n|$ for $n \geq 0$, then $b \geq k_{n+1}$.

Proof. First we prove that the second part of the theorem implies the first. Suppose that the first part is false so that there is a rational a/b with

$$\left|\xi - \frac{a}{b}\right| < \left|\xi - \frac{h_n}{k_n}\right| \text{ and } b \leq k_n$$

Taking the product of these two inequalities, we get

$$|\xi b - a| < |\xi k_n - h_n|$$

But the second part of the theorem says that this implies $b \geq k_{n+1}$, so we have a contradiction, since $k_n < k_{n+1}$ for $n \geq 1$.

To prove the second part of the theorem we proceed again by indirect argument, assuming that $|\xi b - a| < |\xi k_n - h_n|$ and $b < k_{n+1}$. We consider the following linear equations in x and y,

$$xk_n + yk_{n+1} = b \text{ and } xh_n + yh_{n+1} = a$$

By Theorem 5, the determinants of coefficients is ± 1, and consequently these equations have an integral solution x, y. Also, neither x nor y is zero because if $x = 0$, then $b = yk_{n+1} \implies y > 0$ and $b \geq k_{n+1}$, in contradiction to $b < k_{n+1}$. If $y = 0$, then $a = xh_n$ and $b = xk_n$, and

$$|\xi b - a| = |\xi x k_n - x h_n| = |x||\xi k_n - h_n| \geq |k_n \xi - h_n|$$

and again we have a contradiction.

Next we prove that x and y have opposite signs. If $y < 0$, then $xk_n = b - yk_{n+1}$ shows that $x > 0$. If $y > 0$, then $b < k_{n+1} \implies b < yk_{n+1}$ and hence xk_n is negative, whence $x < 0$. It can be observed from the proof of Theorem 9 that $\xi k_n - h_n$ and $\xi k_{n+1} - h_{n+1}$ have opposite signs and hence $x(\xi k_n - h_n)$ and $y(\xi k_{n+1} - h_{n+1})$ have the same signs. Also from the linear equations defining x and y, we get $\xi b - a = x(\xi k_n - h_n) + y(\xi k_{n+1} - h_{n+1})$. Since the two terms on the right have the same sign, so we have

$$\begin{aligned} |\xi b - a| &= |x(\xi k_n - h_n) + y(\xi k_{n+1} - h_{n+1})| \\ &= |x(\xi k_n - h_n)| + |y(\xi k_{n+1} - h_{n+1})| \\ &> |x(\xi k_n - h_n)| = |x||\xi k_n - h_n| \geq |\xi k_n - h_n| \end{aligned}$$

which is a contradiction. □

Theorem 13. *Let ξ be an 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.

Proof. It suffices to prove the result for the case $(a, b) = 1$. Let the convergents of the simple continued fraction expansion of ξ be h_j/k_j and suppose that a/b is not a convergent. The nested inequality $k_n \leq b < k_{n+1}$ determine an integer n. For this n, the inequality $|\xi b - a| < |\xi k_n - h_n|$ is impossible due to Theorem 12. Therefore,

$$|\xi k_n - h_n| \leq |\xi b - a| < \frac{1}{2b} \implies \left|\xi - \frac{h_n}{k_n}\right| < \frac{1}{2bk_n}$$

Since $a/b \neq h_n/k_n$ and $bh_n - ak_n \notin \mathbb{Z}$, so

$$\frac{1}{bk_n} \leq \frac{|bh_n - ak_n|}{bk_n} = \left|\frac{h_n}{k_n} - \frac{a}{b}\right| \leq \left|\xi - \frac{h_n}{k_n}\right| + \left|\xi - \frac{a}{b}\right| < \frac{1}{2bk_n} + \frac{1}{2b^2}$$

which gives $b < k_n$, a contradiction. □

Theorem 14. *The n^{th} convergent of $1/x$ is the reciprocal of the $(n-1)^{th}$ convergent of x if x is any real number greater than 1.*

Proof. We have, $x = \langle a_0, a_1, \dots \rangle$ and $1/x = \langle 0, a_0, a_1, \dots \rangle$. If h_n/k_n and h'_n/k'_n are the convergents for x and $1/x$ respectively, then using equations (6),

$$h'_0 = 0, h'_1 = 1, k_0 = 1 \text{ and } h'_n = a_{n-1}h'_{n-1} + h'_{n-2}, k_{n-1} = a_{n-1}k_{n-2} + k_{n-3}$$

Also,

$$k'_0 = 1, k'_1 = a_0, h_0 = a_0 \text{ and } k'_n = a_{n-1}k'_{n-1} + k'_{n-2}, h_{n-1} = a_{n-1}h_{n-2} + h_{n-3}$$

The theorem then follows from induction. □

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 r. Thus a periodic continued fraction can be written in the form

$$\langle b_0, b_1, b_2, \dots, b_j, a_0, a_1, a_2, \dots, a_{n-1}, \dots, a_0, a_1, a_2, \dots, a_{n-1}, \dots \rangle$$
$$= \langle b_0, b_1, b_2, \dots, b_j, \overline{a_0, a_1, a_2, \dots, a_{n-1}} \rangle \tag{9}$$

where the bar over $a_0, a_1, a_2, \dots, a_{n-1}$ indicates that this block of integers is repeated indefinitely.

Theorem 15. *Any periodic simple continued fraction is a quadratic irrational number, and conversely.*

Proof. Let us write $\xi = \langle b_0, b_1, b_2, \dots, b_j, \overline{a_0, a_1, a_2, \dots, a_{n-1}} \rangle$ and $\theta = \langle \overline{a_0, a_1, a_2, \dots, a_{n-1}} \rangle$. Thus,

$$\theta = \langle \overline{a_0, a_1, a_2, \dots, a_{n-1}} \rangle = \langle a_0, a_1, a_2, \dots, a_{n-1}, \theta \rangle$$

Then by Theorem 3, we have

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

which is a quadratic equation in θ. Hence θ is either a quadratic irrational number or a rational number, but it cannot be rational due to Theorem 7. Now, ξ can be written in terms of θ as

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

where m'/q' and m/q are the last two convergents to $\langle b_0, b_1, \ldots, b_j \rangle$. But θ is a quadratic irrational, i.e., θ is of the form $\frac{a+\sqrt{b}}{c}$, and hence ξ is of a similar form.

To prove the converse, let us begin with any quadratic irrational ξ or ξ_0, of the form $\xi = \xi_0 = \frac{a+\sqrt{b}}{c}$ with integers $a, b, c, b > 0, c \neq 0$ and b not a perfect square (since ξ is irrational). We multiply the 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 $q_0 \mid d - m_0^2$,d, m_0, q are integers, $q_0 \neq 0$ and d not a perfect square. By writing ξ_0 in this form we can get a simple formulation of continued fraction expansion $\langle a_0, a_1, a_2, \ldots \rangle$. We shall prove that the equations

$$a_i = [\xi_i], \qquad \xi_i = \frac{m_i + \sqrt{d}}{q_i}$$

$$m_{i+1} = a_i q_i - m_i, \qquad q_{i+1} = \frac{d - m_{i+1}^2}{q_i} \tag{10}$$

define infinite sequences of integers m_i, q_i, a_i and irrrationals ξ_i in such a way that equations (7) hold, and hence we will have the continued fraction expansion of ξ_0.

We start with ξ_0, m_0, q_0 as above and let $a_0 = [\xi_0]$. If ξ_i, m_i, q_i, a_i are known, then we take $\xi_{i+1} = \frac{m_{i+1} + \sqrt{d}}{q_{i+1}}, m_{i+1} = a_i q_i - m_i, q_{i+1} = \frac{d - m_{i+1}^2}{q_i}, a_{i+1} = [\xi_{i+1}]$.

Now we use induction to prove that the m_i and q_i are integers such that $q_i \neq 0$ and $q_i \mid d - m_0^2$. This holds for $i = 0$. If it is true at the i^{th} stage, we observe that $m_{i+1} = a_i q_i - m_i$ is an integer. Then the equation

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

implies that q_{i+1} is an integer. Also, $q_{i+1} \neq 0$, because if not, then we would have $d = m_{i+1}^2$, but d is not a perfect square. Finally, we have $q_i = \frac{d - m_{i+1}^2}{q_{i+1}}$, which gives $q_{i+1} \mid d - m_{i+1}^2$. Now we verify that equations (7) hold. We have

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

and hence equations (7) hold and so we have proved that $\xi_0 = \langle a_0, a_1, a_2, \dots \rangle$ with a_i as defined in equation (10).

We denote by ξ_i', the conjugate of ξ_i, i.e.,

$$\xi_i' = \frac{m_i - \sqrt{d}}{q_i}$$

Taking conjugates in equation (8), we get

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

Solving this equation for ξ_n', we have

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

As n tends to infinity, both $h_{n-1}/k_{n-1} = r_{n-1}$ and $h_{n-2}/k_{n-2} = r_{n-2}$ tend to ξ_0, which is different from ξ_0' and hence the fraction in parenthesis 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 $\xi_n - \xi_n' > 0$ for $n > N$. Therefore, using equation (10), we have $2\sqrt{d}/q_n > 0$ and hence $q_n > 0$ for $n > N$. It also follows from equation (10) that

$$q_n q_{n+1} = d - m_{n+1}^2 \leq d, \qquad q_n \leq q_n q_{n+1} \leq d$$

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

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 exist distinct integers j and k such that $m_j = m_k$ and $q_j = q_k$. WLOG, assume $j < k$. Then equations (10) give $\xi_j = \xi_k$ and hence

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

i.e., any quadratic irrational can be written as a periodic simple continued fraction. □

Definition 2. Infinite continued fractions of the form $\langle \overline{a_0, a_1, \dots, a_n} \rangle$ are called *purely periodic* continued fractions.

Theorem 16. *The continued fraction expansion of the 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. Consider an irrational number $\xi = \xi_0$ such that $\xi > 1$ and $-1 < \xi' < 0$. Taking conjugates in equation (7), we get

$$\frac{1}{\xi_{i+1}'} = \xi_i' - a_i \tag{11}$$

Now $a_i \geq 1$ for all $i \geq 0$ (even for $i = 0$ since $\xi_0 > \implies a_0 = [\xi_0] \geq 1$). Since $-1 < \xi_0' < 0$, and if $\xi_i' < 0$, then $1/\xi_{i+1}' < -1$, and we have $-1 < \xi_{i+1}' < 0$. Therefore, by induction hypothesis, $-1 < \xi_i' < 0$ for all $i \geq 0$. Hence, equation (11) gives

$$0 < -\frac{1}{\xi_{i+1}} - a_i < 1 \implies a_i < -\frac{1}{\xi_{i+1}} < a_i + 1 \implies a_i = \left[-\frac{1}{\xi_{i+1}} \right]$$

Now $\xi = \xi_0$ is a quadratic irrational and hence by Theorem 15 has a periodic simple continued fraction expansion, i.e., $\xi_j = \xi_k$ 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

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

i.e., the continued fraction expansion of a quadratic irrational number is purely periodic.

To prove the converse, we assume that ξ is purely periodic, say $\xi = \langle\overline{a_0, a_1, \ldots, a_{n-1}}\rangle$, where a_i's are positive integers. Then $\xi > a_0 \geq 1$ and by equation (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 is a root of the quadratic equation

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

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

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

and hence we are done. □

1.8 Continued fraction expansions of square roots

We want the continued fraction expansion of $\sqrt{d}$ for a positive integer d not a perfect square. We start with the closely related irrational number $\sqrt{d} + [\sqrt{d}] = \xi = \xi_0$, say. Then clearly, $\xi > 1$ and $-1 < \xi' = [\sqrt{d}] - d < 0$ and therefore by Theorem 16, the continued fraction expansion of ξ is purely periodic, say

$$\xi = \sqrt{d} + [\sqrt{d}] = \langle\overline{a_0, a_1, \ldots, a_{r-1}}\rangle = \langle a_0, \overline{a_1, a_2, \ldots, a_{r-1}, a_0}\rangle \tag{12}$$

We can suppose that we have chosen r to be the smallest integer for which ξ has an expansion of the form as in equation (12). We note that $\xi_i = \langle a_i, a_{i+1}, \ldots\rangle$ is purely periodic for all i and that $\xi_r = \xi_{2r} = \cdots$. Also, $\xi_i \neq \xi_0$ for all $i = 1, 2, \ldots, r-1$, because otherwise there would be a shorter period. Therefore, $\xi_i = \xi_0$ if and only if i is of the form jr for some j.

Now we can start with $\xi_0 = \sqrt{d} + [\sqrt{d}], q_0 = 1, m_0 = [\sqrt{d}]$ in equation (10) because $1 \mid (d - [\sqrt{d}])^2$. Thus, for all $j \geq 0$,

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

$$\implies m_{jr} - q_{jr}[\sqrt{d}] = (q_{jr} - 1)\sqrt{d} \tag{13}$$

Since the left hand side of equation (13) is rational, so for the right hand side to be rational, we should have $q_{jr} = 1$. Moreover $q_i = 1$ for no other values of the subscript i. For $q_i = 1$, $\xi_i = m_i + \sqrt{d}$, but ξ_i has a purely periodic expansion and so by Theorem 16, we have

$$-1 < \xi_i' < d \implies -1 < m_i - \sqrt{d} < 0 \implies \sqrt{d} - 1 < m_i < \sqrt{d} \implies m_i = [\sqrt{d}]$$

Now we establish that $q_i = -1$ does not hold for any i. Suppose $q_i = -1$ for some i. Then this implies that $\xi_i = -m_i - \sqrt{d}$ and so by Theorem 16, we have

$$-1 < \xi_i' < d \implies -1 < -m_i + \sqrt{d} < 0 \implies \sqrt{d} < m_i < -\sqrt{d} - 1$$

which is impossible.

Noting that $a_0 = [\xi_0] = \left[\sqrt{d} + [\sqrt{d}]\right] = 2[\sqrt{d}]$, we now turn to the case $\xi = \sqrt{d}$ (don't confuse this ξ with $\xi = \sqrt{d} + [\sqrt{d}]$ in equation (12)). Using equation (12), we have

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

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

Applying equations (10) to $\sqrt{d} + [\sqrt{d}]$ with $q_0 = 1, m_0 = [\sqrt{d}]$, we have

$$a_0 = 2[\sqrt{d}], m_1 = [\sqrt{d}], q_1 = d - [\sqrt{d}]^2$$

But we can also apply equations (10) to $\sqrt{d}$ with $q_0 = 1, m_0 = 0$, to get

$$a_0 = [\sqrt{d}], m_1 = [\sqrt{d}], q_1 = d - [\sqrt{d}]^2$$

We see that though the values of a_0 are different, but the values of m_1 and q_1 are the same in both cases. Since $\xi_i = (m_i + \sqrt{d})/q_i$, we see that further application of equation (10) yields the same values of a_i, m_i, q_i in both the cases, i.e., the expansions of $\sqrt{d} + [\sqrt{d}]$ and $\sqrt{d}$ differ only in the values of a_0 and m_0.

Incidentally we have proved the following theorem.

Theorem 17. *If the positive integer d is not a perfect square, the simple continued fraction expansion of $\sqrt{d}$ has the form*

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

with $a_0 = \sqrt{d}$. Furthermore, with $\xi_0 = \sqrt{d}, q_0 = 1, m_0 = 0$ in equations (10), we have $q_i = 1$ if and only if $r \mid i$ and $q_i = -1$ holds for no subscript i. Here r denotes the length of the shortest period in the expansion of $\sqrt{d}$.

1.9 A numerical example

Example 1: Expand $\sqrt{5}$ as an infinite simple continued fraction.

Solution: To derive the continued fraction expansion of $\sqrt{5}$, subtract the floor, invert what is left, and repeat:

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

and the process will repeat to give

$$\sqrt{5} = \langle 2, 4, 4, 4, \dots \rangle = \langle 2, \overline{4} \rangle$$

□

2 Pell's equation

The equation $x^2 - dy^2 = N$, with given integers d and N and unknowns x and y, is usually called *Pell's equation*. If d is negative, it can have only a finite number of solutions. If d is a perfect square, say $d = a^2$, the equation reduces to $(x - ay)(x + ay) = N$ and again there is only a finite number of solutions. The most interesting case of the equation arises when d is a positive integer not a perfect square. For this case, simple continued fractions are very useful.

We expand $\sqrt{d}$ into a simple continued fraction as in Theorem 17, with convergents $r_n = h_n/k_n$, and with q_n defined by equations (10) with $\xi_0 = \sqrt{d}, q_0 = 1, m_0 = 0$.

Theorem 18. *If d is a positive integer not a perfect square, then*

$$h_n^2 - dk_n^2 = (-1)^{n-1} q_{n+1}$$

for all integers $n \geq -1$.

Proof. Using equations (8) and (10), we have

$$\sqrt{d} = \xi_0 = \frac{\xi_{n+1} h_n + h_{n-1}}{\xi_{n+1} k_n + k_{n-1}} = \frac{\left(\frac{m_{n+1}+\sqrt{d}}{q_{n+1}}\right) h_n + h_{n-1}}{\left(\frac{m_{n+1}+\sqrt{d}}{q_{n+1}}\right) k_n + k_{n-1}} = \frac{(m_{n+1} + \sqrt{d}) h_n + q_{n+1} h_{n-1}}{(m_{n+1} + \sqrt{d}) k_n + q_{n+1} k_{n-1}}$$

This gives

$$(m_{n+1} k_n + q_{n+1} k_{n-1} - h_n)\sqrt{d} = m_{n+1} h_n + q_{n+1} h_{n-1} - dk_n \tag{14}$$

Since the right hand side of equation (14) is rational, so for the right hand side to be rational, we should have

$$m_{n+1} k_n + q_{n+1} k_{n-1} - h_n = 0$$

and hence,

$$m_{n+1} h_n + q_{n+1} h_{n-1} - dk_n = 0$$

Then, from both these equations, we have

$$m_{n+1} = \frac{h_n - q_{n+1} k_{n-1}}{k_n} = \frac{dk_n - q_{n+1} h_{n-1}}{h_n}$$

This gives,

$$h_n^2 - dk_n^2 = (h_n k_{n-1} - h_{n-1} k_n) q_{n+1} = (-1)^{n-1} q_{n+1}$$

using Theorem 5 in the last step, and this equation is true for all integers $n \geq 1$. □

We have the following corollary of Theorem 18.

Corollary 18.1. Taking r as the length of the period of the expansion of $\sqrt{d}$, as in Theorem 17, we have for $n \geq 0$,

$$h_{nr-1}^2 - dk_{nr-1}^2 = (-1)^{nr} q_{nr} = (-1)^{nr}$$

With n even, this gives infinitely many solutions of $x^2 - dy^2 = 1$ in integers, provided d is positive and not a perfect square.

It can be seen that Theorem 18 gives us solutions of Pell's equation for certain values of N. In particular, Corollary 13.1 gives infinitely many solutions of $x^2 - dy^2 = 1$ by the use of even values of nr. Of course if r is even, all values of nr are even. If r is odd, Corollary 13.1 gives infinitely many solutions of $x^2 - dy^2 = -1$ by the use of odd integers $n \geq 1$. Apart from the trivial solutions $x = \pm 1, y = 0$ of $x^2 - dy^2 = 1$, all solutions of $x^2 - dy^2 = N$ fall into sets of four by all combinations of signs $\pm x, \pm y$. Hence it is sufficient to discuss the positive solutions $x > 0, y > 0$.

2.1 Convergents of $\sqrt{d}$ and solutions of Pell's equation

Theorem 19. *Let d be a positive integer not a perfect square, and let the convergents to the continued fraction expansion of $\sqrt{d}$ be $r_n = h_n/k_n$. Let the integer N satisfy $|N| < \sqrt{d}$. Then any positive solution $x = s, y = t$ of the equation $x^2 - dy^2 = N$ with $(s,t) = 1$ satisfies $s = h_n, t = k_n$ for some positive integer n.*

Proof. Let E and M be positive integers such that $(E, M) = 1$ and $E^2 - \rho M^2 = \sigma$, where $\sqrt{\rho}$ is irrational and $0 < \sigma < \sqrt{\rho}$ with $\sigma, \rho \in \mathbb{R}$, not necessarily integers. Then

$$\frac{E}{M} - \sqrt{p} = \frac{\sigma}{M(E + M\sqrt{\rho})}$$

and hence,

$$0 < \frac{E}{M} - \sqrt{\rho} < \frac{\sqrt{\rho}}{M(E + M\sqrt{\rho})} = \frac{1}{M^2\left(\frac{E}{M\sqrt{\rho}} + 1\right)}$$

Also,

$$0 < \frac{E}{M} - \sqrt{\rho} \implies \frac{E}{M\sqrt{\rho}} > 1$$

and therefore,

$$\left|\frac{E}{M} - \sqrt{\rho}\right| < \frac{1}{2M^2}$$

By Theorem 13, E/M is a convergent in the continued fraction expansion of ρ.

If $N > 0$, we take $\sigma = N, \rho = d, E = s, M = t$ and the theorem holds in this case. In $N < 0$, then $t^2 - (1/d)s^2 = -N/d$ and we take $\sigma = -N/d, \rho = 1/d, E = t, M = s$. We find that t/s is a convergent in the expansion of $1/\sqrt{d}$. Then by Theorem 14, s/t is a convergent in the expansion of $\sqrt{d}$. □

As a result of the theorems 17,18 and 19, we have the following theorem.

Theorem 20. *All positive solutions of* $x^2 - dy^2 = \pm 1$ *are to be found among* $x = h_n, y = k_n$*, where* h_n/k_n *are the convergents of the expansion of* $\sqrt{d}$*. If* r *is the period of the expansion of* $\sqrt{d}$ *as in Theorem 17, and if* r *is even, then* $x^2 - dy^2 = -1$ *has no solution, and all positive solutions of* $x^2 - dy^2 = 1$ *are given by* $x = h_{nr-1}, y = k_{nr-1}$ *for* $n = 1, 2, 3, \ldots$*. On the other hand, if* r *is odd, then* $x = h_{nr-1}, y = k_{nr-1}$ *give all positive solutions of* $x^2 - dy^2 = -1$ *by use of* $n = 1, 3, 5, \ldots$ *and all positive solutions of* $x^2 - dy^2 = 1$ *by use of* $n = 2, 4, 6, \ldots$*.*

The sequence of pairs $(h_0, k_0), (h_1, k_1), (h_2, k_2), \ldots$ will include all positive solutions of $x^2 - dy^2 = 1$. Also since $a_0 = [\sqrt{d}] > 0$, so the sequence $h_0, h_1, h_2, \ldots$ is strictly increasing. If (x_1, y_1) is the first solution that appears, then for every other solution (x, y), $x > x_1$ and hence $y > y_1$ also. Having found this least positive solution by means of continued fractions, we can find all the remaining positive solutions by a simpler method, as the following theorem suggests.

Theorem 21. *If* (x_1, y_1) *is the least positive integer solution of* $x^2 - dy^2 = 1$*, where* d *is a positive integer not a perfect square, then all positive integer solutions are given by* (x_n, y_n) *for* $n = 1, 2, 3. \ldots,$ *defined by* $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$*.*

Proof. First we establish that (x_n, y_n) is a solution. Since $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$, so $x_n - y_n\sqrt{d} = (x_1 - y_1\sqrt{d})^n$. Hence we can write

$$\begin{aligned} x_n^2 - dy_n^2 &= (x_n - y_n\sqrt{d})(x_n + y_n\sqrt{d}) \\ &= (x_1 - y_1\sqrt{d})^n(x_1 + y_1\sqrt{d})^n = (x_1^2 - dy_1^2)^n = 1 \end{aligned}$$

Suppose there is a positive integer solution (s, t) that is not in the collection $\{(x_n, y_n)\}$. Since both $x_1 + y_1\sqrt{d}$ and $s + t\sqrt{d}$ are greater than 1, there must be some integer m such that

$$(x_1 + y_1\sqrt{d})^m \leq s + t\sqrt{d} < (x_1 + y_1\sqrt{d})^{m+1}$$

. We cannot have $(x_1 + y_1\sqrt{d})^m = s + t\sqrt{d}$, because this would imply $x_m + y_m\sqrt{d} = s + t\sqrt{d}$ so that $x_m = s$ and $y_m = t$. So we have,

$$(x_1 + y_1\sqrt{d})^m < s + t\sqrt{d} < (x_1 + y_1\sqrt{d})^{m+1}$$

Multiplying this inequality by $(x_1 - y_1\sqrt{d})^m = (x_1 + y_1\sqrt{d})^{-m}$, we get

$$1 < (s + t\sqrt{d})(x_1 - y_1\sqrt{d})^m < x_1 + y_1\sqrt{d}$$

We define integers a and b such that $a + b\sqrt{d} = (s + t\sqrt{d})(x_1 - y_1\sqrt{d})^m$. Then we have

$$a^2 - db^2 = (s^2 - dt^2)(x_1^2 - dy_1^2)^m = 1$$

So, (a, b) is a solution of $x^2 - dy^2 = 1$ such that $1 < a + b\sqrt{d} < x_1 + y_1\sqrt{d}$. But then, $0 < (a + b\sqrt{d})^{-1}$ and hence $0 < a - b\sqrt{d} < 1$. Now we have

$$a = \frac{1}{2}(a + b\sqrt{d}) + \frac{1}{2}(a - b\sqrt{d}) > \frac{1}{2} + 0 > 0$$

and

$$b\sqrt{d} = \frac{1}{2}(a + b\sqrt{d}) - \frac{1}{2}(a - b\sqrt{d} > \frac{1}{2} - \frac{1}{2} = 0$$

Therefore, (a, b) is a positive integer solution. Therefore, $a > x_1, b > y_1$, which contradicts $a + b\sqrt{d} < x_1 + y_1\sqrt{d}$. Therefore, all positive integers solutions are given by (x_n, y_n) for $n = 1, 2, 3, \ldots,$ with x_n and y_n defined as above. □

Theorem 22. *If $x^2 - dy^2 = -1$ is solvable, and (x_1, y_1) is the smallest positive solution. Then (x_2, y_2) defined by $x_2 + y_2\sqrt{d} = (x_1 + y_1\sqrt{d})^2$ is the smallest positive solution of $x^2 - dy^2 = 1$.*

Proof. Assume, to the contrary, that (x_2', y_2') defined by $x_2' + y_2'\sqrt{d} = (x_1 + y_1\sqrt{d})^2$ is a positive integer solution of $x^2 - dy^2 = 1$ and $y_2' < y_2$. Define x_1', y_1' such that

$$x_1' + y_1'\sqrt{d} = \frac{x_2' + y_2'\sqrt{d}}{x_1 + y_1\sqrt{d}} = \frac{(x_2' + y_2'\sqrt{d})(x_1 - y_1\sqrt{d})}{x_1^2 - dy_1^2}$$

Using $x_1^2 - dy_1^2 = -1$, we get

$$x_1' + y_1'\sqrt{d} = (x_2' + y_2'\sqrt{d})(y_1\sqrt{d} - x_1) = (dy_1y_2' - x_1x_2') + (x_2'y_1 - x_1y_2')\sqrt{d}$$

and so $x_1' = dy_1y_2' - x_1x_2', y_1' = x_2'y_1 - x_1y_2'$, which turns out to be a solution of $x^2 - dy^2 = 1$ and smaller than (x_2', y_2'), a contradiction. □

2.2 A numerical example

Example 2: Find the least positive integer solution of $x^2 - 73y^2 = -1$ (if it exists) and of $x^2 - 73y^2 = 1$, given that $\sqrt{73} = \langle 8, \overline{1, 1, 5, 5, 1, 1, 16} \rangle$.

Solution: Since the period of this continued fraction expansion is 7, an odd number, we know from Theorem 20 that the equation $x^2 - 73y^2 = -1$ has solutions. Moreover, the least positive solution is $x = h_6, y = k_6$ from the convergent $r_6 = h_6/k_6$. Using equations (6), we see that the convergents are

$$r_0 = 8/1, r_1 = 9/1, r_2 = 17/2, r_3 = 94/11, r_4 = 487/57, r_5 = 561/68, r_6 = 1068/125$$

Therefore, the least positive integer solution of $x^2 - 73y^2 = -1$ is $x = 1068, y = 125$. To get the least positive integer solution of $x^2 - 73y^2 = 1$, we use Theorem 22 to calculate x and y equating the rational and irrational parts of

$$x + y\sqrt{73} = (1068 + 125\sqrt{73})^2$$

The values of x and y are 2281249 and 267000 respectively, which is the least positive integer solution of $x^2 - 73y^2 = 1$. □