An Introduction to Continued Fractions

A. J. van der Poorten

---

1. Notice that if

$$\begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} c_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix}, \quad n = 0, 1, 2, \ldots$$

then

$$\frac{p_n}{q_n} = c_0 + \cfrac{1}{c_1 + \cfrac{1}{c_2 + \ddots + \cfrac{1}{c_n}}} = [c_0,\ c_1,\ c_2,\ \ldots,\ c_n].$$

This is easy to see by induction: First observe that

$$\begin{pmatrix} 0 & 1 \\ 1 & -c \end{pmatrix}\begin{pmatrix} c & 0 \\ 1 & 0 \end{pmatrix} = I$$

and that

$$\begin{pmatrix} 0 & 1 \\ 1 & -c_0 \end{pmatrix}\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} q_n & q_{n-1} \\ p_n - c_0 q_n & p_{n-1} - c_0 q_{n-1} \end{pmatrix}.$$

But

$$\frac{q_n}{p_n - c_0 q_n} = \frac{1}{\frac{p_n}{q_n} - c_0} = [c_1;\ c_2,\ \ldots,\ c_n]$$

Our remark sets up a correspondence $\leftrightarrow$ between certain products of 2×2 matrices and continued fractions, which we shall exploit below. Of course this correspondence has an immediate geometric interpretation (cf Stark [1], Chap. 7). However, we shall obtain the usual properties of the continued fraction algorithm directly from the formalism rather than from the geometry. For example, we read that

$$\boxed{\begin{cases} p_{n+1} &= c_{n+1}p_n + p_{n-1} \\ q_{n+1} &= c_{n+1}q_n + q_{n-1} \end{cases} \qquad n = 0, 1, 2, \ldots}$$

and by taking the determinant of the matrix product

$$\boxed{p_n q_{n-1} - p_{n-1} q_n = (-1)^{n+1}, \qquad n = 0, 1, 2, \ldots}$$

From this critical formula we readily obtain by induction

$$\frac{p_n}{q_n} = c_0 + \frac{1}{q_0 q_1} - \frac{1}{q_1 q_2} + \cdots + \frac{(-1)^{n-1}}{q_{n-1} q_n}\ .$$

It follows that given c_0 in $\mathbf{Z}$ and $c_1, c_2, \ldots$ positive integers, the continued fraction $[c_0;\ c_1,\ c_2,\ \ldots]$ converges to a real number α satisfying

$$\alpha - \frac{p_n}{q_n} = (-1)^n \left(\frac{1}{q_n q_{n+1}} - \frac{1}{q_{n+1} q_{n+2}} + \cdots \right)\ .$$

Hence

$$\boxed{|q_n \alpha - p_n| < \frac{1}{q_{n+1}}\ .}$$

As usual, denote by $[x]$ the greatest integer less than or equal to the real number x. Set

$$\alpha_0 = \alpha \quad \text{and} \quad \alpha_n = [c_n;\ c_{n+1},\ \dots], \qquad n = 0, 1, 2, \dots$$

Then

$$c_n = [\alpha_n] \quad \text{and} \quad \alpha_{n+1} = (\alpha_n - c_n)^{-1}$$

because, evidently, $0 < [0;\ c_{n+1},\ c_{n+2},\ \dots] < 1$. This describes the usual regular continued fraction expansion for α and the algorithm yielding it.

Notice that the correspondences

$$\alpha \leftrightarrow \begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \quad \text{and} \quad \alpha_{n+1} \leftrightarrow \begin{pmatrix} c_{n+1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_{n+2} & 1 \\ 1 & 0 \end{pmatrix} \cdots$$

imply, on multiplying the corresponding matrices

$$\boxed{\alpha = \frac{\alpha_{n+1} p_n + p_{n-1}}{\alpha_{n+1} q_n + q_{n-1}} \quad \text{and} \quad \alpha_{n+1} = -\frac{\alpha q_{n-1} - p_{n-1}}{\alpha q_n - p_n}}$$

The latter expression together with $\alpha_{n+1} > 1$ entails that the sequence

$$\{q_n \alpha - p_n\}, \qquad n = 0, 1, 2, \dots$$

alternates in sign, a fact that we had noticed earlier.

Next we consider the familiar 'best approximation' property. Let $q_{n-1} < q < q_n$ with q in $\mathbf{Z}$, and take p in $\mathbf{Z}$. Then there are integers a and b so that

$$p = a p_n + b p_{n-1}, \qquad q = a q_n + b q_{n-1}$$

implying that a and b have opposite signs. Thus

$$q\alpha - p = a(q_n \alpha - p_n) + b(q_{n-1}\alpha - p_{n-1})$$

yields

$$|q\alpha - p| = |p q_{n-1} - q p_{n-1}||q_n \alpha - p_n| + |p q_n - q p_n||q_{n-1}\alpha - p_{n-1}| \ .$$

So for all p in $\mathbf{Z}$, $q_{n-1} < q < q_n$ entails

$$|q\alpha - p| > |q_n \alpha - p_n| \ .$$

A fortiori, since the sequence $\{|q_n\alpha - p_n|\}$, $n = 0, 1, \dots$ is monotonic decreasing, we see that the *convergents* p_n/q_n of α yield locally best approximation to α in that $0 < q < q_n$ implies

$$|q\alpha - p| > |q_n \alpha - p_n|, \qquad n = 1, 2, \dots \ .$$

A more delicate investigation of our argument will show that this inequality continues to hold for a certain range of q greater than q_n.

Given a rational p/q, thus p, q in $\mathbf{Z}$ with $q > 0$, the continued fraction algorithm is just the Euclidean algorithm:

$$\begin{aligned} p &= c_0 q + r_0 & 0 \le r_0 < q \\ q &= c_1 r_0 + r_1 & 0 \le r_1 < r_0 \\ &\vdots & \vdots \\ r_{n-2} &= c_n r_{n-1} \end{aligned}$$

Then

$$\frac{p}{q} = [c_0;\ c_1,\ \ldots,\ c_n\ ] = \frac{p_n}{q_n}$$

and we verify that the continued fraction of α terminates if and only if α is rational. Moreover suppose that the greatest common divisor (p, q) is $\pm d$. Then an earlier identity implies

$$\boxed{q_{n-1}p - p_{n-1}q = \pm d}\,,$$

making explicitly the linear combination of p and q yielding their greatest common divisor. We may also see just why the c_i are called *partial quotients*.

We conclude with the following important fact: *If p, q are realtively prime and $|q\alpha - p| < 1/(2q)$ then p/q is a convergent of α.* To see this take integers r, s with $0 < s < q$ and notice that

$$\begin{aligned} 1 \le |qr - ps| &= |s(q\alpha - p) - q(s\alpha - r)| \\ &\le s|q\alpha - p| + q|s\alpha - r| \le \frac{s}{2q} + q|s\alpha - r| \end{aligned}$$

It follows that

$$|q\alpha - p| < |s\alpha - r|$$

since certainly

$$q|s\alpha - r| \ge 1 - \frac{s}{2q} > \frac{1}{2}\,.$$

Hence p/q yields a locally best approximation. But above we might have noted that $q_{n-1} < q < q_n$ also implies

$$|q\alpha - p| > |q_{n-1}\alpha - p_{n-1}|\,,$$

entailing that only a convergent can yield a locally best approximation.

Of course, the fact that the continued fraction algorithm yields all and only locally best approximations is its critical property. And it is the failure of this property to apply to higher dimensional analogues of the algorithm that renders their theory so much more difficult.

2. We turn now to various formal properties of continued fractions which are immediate consequences of the correspondence:

$$\begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} c_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix}$$

$$\leftrightarrow [c_0,\ c_1,\ c_2,\ \ldots,\ c_n] = \frac{p_n}{q_n}, \quad n = 0, 1, 2, \ldots\ .$$

Firstly observe that each matrix in the product is symmetric. Hence taking the transpose yields

$$\begin{pmatrix} c_n & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & q_n \\ p_{n-1} & q_{n-1} \end{pmatrix}\ .$$

and we read that

$$\boxed{\begin{cases} [c_n;\ c_{n-1},\ \ldots,\ c_0] &= \dfrac{p_n}{p_{n-1}}\ , \\ [c_n;\ c_{n-1},\ \ldots,\ c_1] &= \dfrac{q_n}{q_{n-1}}\ . \end{cases}}$$

Both formulas are interesting in that they report a pair of consecutive denominators, respectively numerators, contain the total prior 'history' of the continued fraction expansion.

The following beautiful application of the first of these formulas has been reported elsewhere [2] but warrants repetition here. It is the delightful proof of H. J. S. Smith [3] that a prime p is a sum of two squares if and only if $p \equiv 1 \pmod 4$ or $p = 2$. The case $p = 2$ is trivial, and the only if part of the claim is plain since the square of an integer is $\equiv 1 \text{ or } 0 \pmod 4$.

Now take p prime, $p \equiv 1 \pmod 4$, and consider the set

$$H = \left\{ \frac{p}{2},\ \frac{p}{3},\ \dots,\ \frac{p}{(p-1)/2} \right\}$$

noting that $\#H = \frac{1}{2}(p-1)-1$ is odd. We take H to consist of terminating continued fractions

$$\frac{p}{q} = [c_0;\ c_1,\ \dots,\ c_n], \quad q = 2, 3, \dots, \frac{1}{2}(p-1)$$

rather than, so to speak, the rationals we display. We have to digress to remark that there is a possible ambiguity in that if $c_n > 1$

$$[c_0;\ c_1,\ \dots,\ c_n] = [c_0;\ c_1,\ \dots,\ c_n - 1,\ 1]$$

But we resolve that problem by insisting that always $c_n \geq 2$. Now consider the map R which reverses terminating continued fractions

$$R\ :\ [c_0;\ c_1,\ \dots,\ c_n] \ \longrightarrow\ [c_n;\ c_{n-1},\ \dots,\ c_0].$$

Allow R to act on H. Since for p/q in H, $p/q > 2$ we have $c_0 \geq 2$. By convention also $c_n \geq 2$. Moreover certainly $c_n \leq q < \frac{1}{2}p$, and similarly $c_0 < \frac{1}{2}p$ in each case. But reversing a continued fraction preserves the numerator of the convergent. Thus R maps the continued fractions of H to continued fractions representing rationals β with numerator p and

$$2 < \beta \leq \frac{1}{2}p.$$

R is plainly one-one and thus R permutes the elements of H. But of course R^2 is just the identity map. So R is an involution of the set H with H containing an odd number of elements. Hence a member of H is invariant under R : it has a symmetric continued fraction.

We pause to consider symmetric continued fractions. For convenience set

$$\overrightarrow{w} = c_0 c_1 \dots c_n.$$

Then a symmetric continued fractionhas the shape $[\overrightarrow{w}, \overleftarrow{w}]$ or $[\overrightarrow{w}, c, \overleftarrow{w}]$ according as it is of even or odd length. In the first case the continued fraction corresponds to the matrix product

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} p_n & q_n \\ p_{n-1} & q_{n-1} \end{pmatrix} = \begin{pmatrix} p_n^2 + p_{n-1}^2 & p_n q_n + p_{n-1} q_{n-1} \\ p_n q_n + p_{n-1} q_{n-1} & q_n^2 + q_{n-1}^2 \end{pmatrix}$$

so

$$[\overrightarrow{w}] = \frac{p_n}{q_n}\ , \qquad [\overrightarrow{w}\,\overleftarrow{w}] = \frac{p_n^2 + p_{n-1}^2}{p_n q_n + p_{n-1} q_{n-1}}\ .$$

In the second case

$$\begin{aligned} [\overrightarrow{w}, c, \overleftarrow{w}] \leftrightarrow & \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} c & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_n & q_n \\ p_{n-1} & q_{n-1} \end{pmatrix} \\ = & \begin{pmatrix} p_n(cp_n + 2p_{n-1}) & cp_n q_n + p_n q_{n-1} + p_{n-1} q_n \\ cp_n q_n + p_n q_{n-1} + p_{n-1} q_n & q_n(cq_n + 2q_{n-1}) \end{pmatrix} \\ & \leftrightarrow \frac{p_n(cp_n + 2p_{n-1})}{cp_n q_n + p_n q_{n-1} + p_{n-1} q_n} \end{aligned}$$

Of course, in the second case the numerator is evidently composite whilst the numerators of elements of H are prime. Hence the first case applies and we read that

$$p = p_n^2 + p_{n-1}^2 \ .$$

We might note that in so reading we have used the fact that the formula

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^{n+1}$$

implies that convergents p_n/q_n are always presented in their 'lowest terms'; that is, so that their numerator and denominator are relatively prime.

In any event we have displayed the prime $p \equiv 1 \pmod 4$ as a sum of squares, proving our allegation and giving a congenial demonstration of the benefits of applying the correspondence between continued fractions and certain matrix products.

Next we note a further formula. We have

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} x - \frac{q_{n-1}}{q_n} & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n x - \frac{(p_n q_{n-1} - p_{n-1} q_n)}{q_n} & p_n \\ q_n x & q_n \end{pmatrix}$$

Hence

$$[\overrightarrow{w}] = [c_0;\ c_1,\ \dots,\ c_n] = \frac{p_n}{q_n}$$

implies that

$$\left[\overrightarrow{w},\ x - \frac{q_{n-1}}{q_n}\right] = \frac{p_n}{q_n} + \frac{(-1)^n}{x q_n^2} \ .$$

Indeed, we may go on to observe that

$$\begin{aligned} x - \frac{q_{n-1}}{q_n} &= x - 1 + \frac{q_n - q_{n-1}}{q_n} \\ \frac{q_n}{q_n - q_{n-1}} &= 1 + \frac{q_{n-1}}{q_n - q_{n-1}} \\ \frac{q_n - q_{n-1}}{q_{n-1}} &= -1 + \frac{q_n}{q_{n-1}} \\ \frac{q_{n-1}}{q_n} &= 0 + \frac{q_{n-1}}{q_n} \\ \frac{q_n}{q_{n-1}} &= [c_n;\ c_{n-1},\ \dots,\ c_1] \end{aligned}$$

This suggests it is convenient to change our notation by using $\overrightarrow{w}$ to denote the word

$$\overrightarrow{w} = c_1 c_2 \dots c_n \ .$$

Then our formula is

$$\frac{p_n}{q_n} + \frac{(-1)^n}{x q_n^2} = \left[c_0;\ \overrightarrow{w},\ x - \frac{q_{n-1}}{q_n}\right] = [c_0;\ \overrightarrow{w},\ x-1,\ 1,\ -1,\ 0,\ \overleftarrow{w}] \ .$$

Note that this is a 'formal' formula. Certain of the entries, say $x-1$ and -1 and 0 may not be admissible in a regular continued fraction expansion. However, a zero is never a problem: for

$$\begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a+b & 1 \\ 1 & 0 \end{pmatrix}$$

implies that

$$[\dots,\ a,\ 0,\ b,\ \dots] = [\dots,\ a+b,\ \dots] \ .$$

It seems a pity to eschew an application (see [2] for references and other details). Consider the sum

$$1+\frac{1}{2}+\frac{1}{8}+\frac{1}{128}+\cdots=2\sum_{n=0}^{\infty}2^{-2^n}$$

and observe that if some partial sum

$$2\left(2^{-1}+2^{-2}+\cdots+2^{-2^m}\right)=\frac{p_n}{q_n}$$

then $q=2^{2^m-1}$, and the next term is

$$2^{-2^{m+1}+1}=\frac{1}{2q_n^2}\ .$$

Set $\overrightarrow{p}=1,\ 1,\ -1,\ 0$ and denote by $S_{\vec{p}}$ the operator

$$S_{\vec{p}}\ :\ \overrightarrow{w}\quad\rightarrow\quad\overrightarrow{w}\,\overrightarrow{p}\,\overleftarrow{w}\ .$$

Our remarks yields the formula

$$2\sum_{n=0}^{\infty}2^{-2^n}=\left[1;\ S_{\vec{p}}^{\infty}\ \cdot\right]\ .$$

with $\cdot$ denoting the empty word. We have just

$$\left[1;\ S_{\vec{p}}^{\infty}\ \cdot\right]=[1;\ \overrightarrow{p}\,\overrightarrow{p}\,\overleftarrow{p}\,\overrightarrow{p}\,\overrightarrow{p}\,\overleftarrow{p}\,\overleftarrow{p}\,\overrightarrow{p}\,\overrightarrow{p}\,\overrightarrow{p}\,\overleftarrow{p}\,\overleftarrow{p}\ \ \ldots]\ .$$

The arrow sequence is the celebrated paperfolding sequence: again see [2]. We now need some nontrivial tidying to eliminate inadmissible -1 and 0 entries. It may constitute a useful challenge to the reader to obtain

$$2\sum_{n=0}^{\infty}2^{-2^n}=\left[1;\ 1,\ 1,\ 1,\ S_{1,1,1,1,2,1}^{\infty}(2,1,1,1)\right],$$

contrary to the allegation of [2], observation 2.13, but in accordance with the correction [4]. The example is remarkable in two respects. Firstly, it is extraordinary that we can obtain a formula explicitly giving the expansion. Secondly, it is noteworthy that the only partial quotients that appear are 1 and 2. These properties are shared by the uncountably many numbers

$$1+2\sum_{n=1}^{\infty}\pm2^{-2^n}\ ;$$

that is, regardless of the sign chosen at each term.

3. PERIODIC CONTINUED FRACTIONS

3.1 In the sequel D is a positive integer and not a square. Suppose we know that (the so-called) Pell's equation

$$X^2-DY^2=\pm1$$

has a solution (x,y) in positive integers. We shall apply this knowledge to prove that the continued fraction expansion of $\sqrt{D}$ is periodic and to uncover some of the structure of the period. Set

$$M=\begin{pmatrix}x & Dy\\ y & x\end{pmatrix}\qquad\text{and}\qquad J=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

We have, by hypothesis

$$\det M = x^2 - Dy^2 = \pm 1 \ .$$

Applying the Euclidean algorithm to the rows of the matrix MJ we obtain the unique expansion

$$MJ = \begin{pmatrix} Dy & x \\ x & y \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}$$

with positive integers $a_0, a_1, \ldots, a_n$. It is evident that

$$a_0 = \left[\frac{Dy}{x} \right] = \left[\frac{x}{y} \right] = \left[\sqrt{D} \right]$$

and because the matrix MJ is symmetric, the word $a_0 a_1 a_2 \ldots a_n$ is a palindrome (example: a palindrome is 'never even'). Hence

$$a_0 = a_n = \left[\sqrt{D} \right] \ .$$

Now consider the periodic continued fraction

$$\delta = [a_0;\ \overline{a_1,\ a_2,\ \ldots,\ a_{n-1},\ 2a_0}\ ] \ .$$

This is just to say that

$$\delta = [a_0;\ a_1,\ a_2,\ \ldots,\ a_{n-1},\ 2a_0 + (\delta - a_0)]$$

which corresponds to

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_0 + \delta & 1 \\ 1 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \delta & 1 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} Dy & x \\ x & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \delta & 1 \end{pmatrix} = \begin{pmatrix} Dy + \delta x & x \\ x + \delta y & y \end{pmatrix} \leftrightarrow \delta$$

So

$$\delta = \frac{Dy + \delta x}{x + \delta y} \qquad \text{or} \qquad \delta^2 = D \ .$$

This shows that if the Diophantine equation

$$X^2 - DY^2 = \pm 1$$

has a nontrivial solution then $\sqrt{D}$ has a periodic continued fraction

$$\sqrt{D} = [a_0;\ \overline{a_1,\ a_2,\ \ldots,\ a_{n-1},\ 2a_0}\ ] \ .$$

with $a_1 a_2 \ldots a_{n-1}$ a palindrome. Moreover, every nontrivial solution (x, y) of the equation is given by a convergent

$$\frac{x}{y} = [a_0;\ a_1,\ \ldots,\ a_{n-1}] \ .$$

We may have

$$X^2 - DY^2 = -1$$

only if the primitive period of $\sqrt{D}$ is of odd length.

It is of course easy to prove that every periodic continued fraction represents a quadratic irrational. The converse, Lagrange's theorem, is a little more difficult. Assuming, as we have, that Pell's equation has solutions we know that $\sqrt{D}$ has a periodic expansion. The general quadratic irrational is of the shape

$$\frac{a+b\sqrt{D}}{c}$$

for integers a, b, c, so that

$$A = \begin{pmatrix} a & b \\ c & 0 \end{pmatrix}$$

is nonsingular. Then our approach would lead us to attempt to prove Lagrange's theorem by showing that a product

$$A \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} \overline{\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2a_0 & 1 \\ 1 & 0 \end{pmatrix}} \cdots$$

yields a periodic product

$$\begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \overline{\begin{pmatrix} c_k & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix}} \begin{pmatrix} c_k & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_k & 1 \\ 1 & 0 \end{pmatrix} \cdots$$

obtained by commutating A through the given product until it disappears in the $\cdots$ on the right. The appropriate theory exists. A particularly elegant version is that of Raney [5]: It is shown that moving A through the product can be represented by a finite-state transducer acting on R-L words; that is, words on two symbols R and L. General theory of finite automata shows that finite-state transduction preserves the nature of the transduced sequence: for example that it is generated by a finite p-automaton. In particular, periodic sequences are transduced to periodic sequences. We will return to these notions below. For the present we endeavour to retrieve somewhat less well-known properties of the continued fraction expansion of $\sqrt{D}$.

3.2 Suppose then that $(x, y) \in \mathbf{Z}^2$ satisfies the equation

$$x^2 - Dy^2 = \pm 2$$

Just as solutions of Pell's equation occur at the end of the period of $\sqrt{D}$, we shall find that solutions of the present equation occur at the middle of the period. The precise sense of this remark should become clear below. Set

$$M = \begin{pmatrix} x & Dy \\ y & x \end{pmatrix}$$

and consider the matrix

$$M^2 J = MJM^T = \begin{pmatrix} x & Dy \\ y & x \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x & y \\ Dy & x \end{pmatrix} .$$

There are two cases, according as D is or is not even.

(a) If $2 \mid D$ then $2 \mid x$. We dismantle M to obtain

$$M = \begin{pmatrix} x & Dy \\ y & x \end{pmatrix} = \begin{pmatrix} x & \frac{1}{2}Dy \\ y & \frac{1}{2}x \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} x & x' \\ y & y' \end{pmatrix} \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix} J \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

with

$$c_m = \left[\frac{x}{2y} \right], \qquad x' = \frac{1}{2}Dy - c_m x, \qquad y' = \frac{1}{2}x - c_m y .$$

Now notice that

$$\begin{pmatrix}1&0\\0&2\end{pmatrix} J \begin{pmatrix}1&0\\0&2\end{pmatrix} = \begin{pmatrix}2&0\\0&2\end{pmatrix} J = 2J.$$

Moreover, the matrix

$$M' = \begin{pmatrix}x & x'\\ y & y'\end{pmatrix}$$

corresponds to some continued fraction, say

$$M' \leftrightarrow [c_0;\ c_1,\ \ldots,\ c_{m-1}]\ .$$

We have

$$M^2J = MJ^3M^T = 2M' \begin{pmatrix}c_m&1\\1&0\end{pmatrix}\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}c_m&1\\1&0\end{pmatrix} M'^T$$

and thus $\frac{1}{2}M^2J$ corresponds to the continued fraction

$$[c_0;\ c_1,\ \ldots,\ c_{m-1},\ 2c_m,\ c_{m-1},\ \ldots,\ c_1,\ c_0]$$

We conclude the argument after describing the second case.

(b) If $2 \not| D$ then $2 \mid Dy - x$ and $2 \mid x - y$. Much as above we obtain

$$\begin{aligned} M &= \begin{pmatrix}x & Dy-x\\ y & x-y\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix} = \begin{pmatrix}x & \frac{1}{2}(Dy-x)\\ y & \frac{1}{2}(x-y)\end{pmatrix}\begin{pmatrix}1&0\\0&2\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}\\ &= \begin{pmatrix}x & x'\\ y & y'\end{pmatrix}\begin{pmatrix}c_m&1\\1&0\end{pmatrix} J \begin{pmatrix}1&0\\0&2\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}\end{aligned}$$

with

$$c_m = \left[\frac{x-y}{2y}\right], \qquad x' = \frac{1}{2}(Dy-x) - c_m x, \qquad y' = \frac{1}{2}(x-y) - c_m y\ .$$

It is convenient at this point to introduce the notation

$$R = \begin{pmatrix}1&1\\0&1\end{pmatrix}, \qquad L = \begin{pmatrix}1&0\\1&1\end{pmatrix}.$$

and to note that

$$\begin{pmatrix}c&1\\1&0\end{pmatrix} = R^cJ = JL^c\ .$$

Below we need

$$\begin{pmatrix}1&0\\0&2\end{pmatrix} RLJ \begin{pmatrix}1&0\\0&2\end{pmatrix} = R\begin{pmatrix}2&0\\0&2\end{pmatrix} J = 2RJ = 2JL.$$

With the notation above we find that

$$M^2J = 2M'R^{c_m}J^2RJ^3L^{c_m}M'^T, \qquad R^{c_m+1}JL^{c_m} = R^{2c_m+1}J$$

so the unimodular matrix $\frac{1}{2}MJ^2$ corresponds to the continued fraction

$$[c_0;\ c_1,\ \ldots,\ c_{m-1},\ 2c_m+1,\ c_{m-1},\ \ldots,\ c_1,\ c_0]\ .$$

In both cases we have

$$M^2J = \begin{pmatrix}x&Dy\\y&x\end{pmatrix}\begin{pmatrix}Dy&x\\x&y\end{pmatrix} = 2\begin{pmatrix}DY&X\\X&Y\end{pmatrix}$$

with

$$2X = x^2 + Dy^2 \qquad \text{and} \qquad 2Y = 2xy$$

both divisible by 2. Hence

$$\frac{1}{2}M^2J = \begin{pmatrix} DY & X \\ X & Y \end{pmatrix}$$

is a unimodular matrix exactly of the shape of the matrix we discussed in our analysis of the continued fraction of $\sqrt{D}$. Plainly we have found that if

$$x^2 - Dy^2 = \pm 2$$

and

$$\frac{x}{y} = [c_0;\ c_1,\ \ldots,\ c_{m-1}]$$

then

(1) $2|D$ $\qquad \sqrt{D} = [c_0;\ c_1,\ \ldots,\ c_{m-1},\ 2c_m,\ c_{m-1},\ \ldots,\ c_1,\ c_0]$

$$\text{with } \ c_m = \left[\frac{x}{2y}\right];$$

(2) $2 \not| D$ $\qquad \sqrt{D} = [c_0;\ c_1,\ \ldots,\ c_{m-1},\ 2c_m + 1,\ c_{m-1},\ \ldots,\ c_1,\ c_0]$

$$\text{with } \ c_m = \left[\frac{x-y}{2y}\right].$$

Two simple examples:

- $156^2 - 46 \cdot 23^2 = 2, \quad \frac{156}{23} = [6;\ 1,\ 3,\ 1,\ 1,\ 2], \quad \left[\frac{156}{2 \cdot 23}\right] = 3.$

  and

  $$\sqrt{46} = [6;\ \overline{1,\ 3,\ 1,\ 1,\ 2,\ \underline{6},\ 2,\ 1,\ 1,\ 3,\ 1,\ 12}\ ]$$

- $13^2 - 19 \cdot 3^2 = -2, \quad \frac{13}{3} = [4;\ 2,\ 1], \quad \left[\frac{13-3}{2 \cdot 3}\right] = 1.$

  and

  $$\sqrt{19} = [4;\ \overline{2,\ 1,\ \underline{3},\ 1,\ 2,\ 8}\ ]$$

3.2 Next suppose that $(x, y) \in \mathbf{Z}$ satisfies the equation

$$x^2 - Dy^2 = \pm 4\ .$$

Of course we take x, y relatively prime; hence both x, y and D are odd. Set

$$M = \begin{pmatrix} x & Dy \\ y & x \end{pmatrix}$$

In this case it is congenial to consider the matrix M^3J with the expectation that solutions of the equation will occur about a third of the way along the period of $\sqrt{D}$, and driven by the foresight that

$$M^3J = 8\begin{pmatrix} DY & X \\ X & Y \end{pmatrix} \quad \text{with } 8X = x(x^2 + 3Dy^2) \text{ and } 8Y = y(3x^2 + Dy^2)\ .$$

It is easy to see that both X, Y are integers since

$$x^2 + 3Dy^2 = (x^2 - Dy^2) + 4Dy^2 \equiv 0 \pmod 8$$
$$3x^2 + Dy^2 = 4x^2 - (x^2 - Dy^2) \equiv 0 \pmod 8$$

The unimodular matrix $\frac{1}{8}M^3J$ then yields the period of $\sqrt{D}$ in the manner already described. For the finer detail we proceed by noting that $D \equiv 1 \pmod 4$ so that both $Dy + x$ and $x + y$ or $Dy - x$ and $x - y$ are divisible by 4. Writing $M^3J = MMJM^T$ we dismantle the first, and the third M to obtain

$$M = \begin{pmatrix} x & \frac{1}{4}(Dy \pm x) \\ y & \frac{1}{4}(x \pm y) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} R^{\mp 1} = M' R^c A'^2 R^{\mp 1} \ .$$

As before

$$M' = \begin{pmatrix} x & x' \\ y & y' \end{pmatrix}$$

is the matrix corresponding to the convergent x/y, so

$$c = \left[\frac{Dy \pm x}{4x} \right] = \left[\frac{x \pm y}{4y} \right] \ .$$

For convenience we have set

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A' = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \ ;$$

we note that $AA' = A'A = 2I$. To take advantage of this last observation we dismantle the middle M as

$$M = P^{\pm 1} A \begin{pmatrix} \frac{1}{2}(x \mp y) & \frac{1}{4}(Dy \mp 2x + y) \\ y & \frac{1}{2}(x \mp y) \end{pmatrix} A' R^{\mp 1} \ .$$

On the left we notice that

$$A'^2 R^{\mp 1} R^{\pm 1} A = 2A' \ ,$$

and on the right, recalling that $R^T = L$, $RJ = JL$, $JA' = AJ$, we find

$$A' R^{\pm 1} J L^{\mp 1} A'^2 = 2JA' = 2AJ \ .$$

But

$$A' \begin{pmatrix} \frac{1}{2}(x \mp y) & \frac{1}{4}(Dy \mp 2x + y) \\ y & \frac{1}{2}(x \mp y) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(x \mp y) & \frac{1}{8}(Dy \mp 2x + y) \\ 2y & \frac{1}{2}(x \mp y) \end{pmatrix} A' \ .$$

In justification, we remark that, plainly

$$Dy \mp x \equiv \pm(x \mp y) \pmod 8 \ .$$

Finally we have obtained

$$M^3 J = 8 M' R^c \begin{pmatrix} \frac{1}{8}(Dy \mp 2x + y) & \frac{1}{2}(x \mp y) \\ \frac{1}{2}(x \mp y) & 2y \end{pmatrix} L^c M'^T$$

thus represent $\frac{1}{8}M^3J$ as a product of unimodular matrices which can readily be made to corresponds to a continued fraction.

We have done too much explicit computation already; so we give only a simple example:

$$25^2 - 69 \cdot 3^2 = 4 \ .$$

Hence

$$M = \begin{pmatrix} 25 & 69 \cdot 3 \\ 3 & 25 \end{pmatrix} = \begin{pmatrix} 25 & 8 \\ 3 & 1 \end{pmatrix} R^2 A'^2 R^{-1}, \quad c = 2$$

and

$$\frac{1}{8}M^3J = \begin{pmatrix}25 & 8\\ 3 & 1\end{pmatrix}\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}\begin{pmatrix}20 & 11\\ 11 & 6\end{pmatrix}\begin{pmatrix}1 & 0\\ 2 & 1\end{pmatrix}\begin{pmatrix}25 & 3\\ 8 & 1\end{pmatrix}$$
$$= \begin{pmatrix}8 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}3 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}2 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}4 & 1\\ 1 & 0\end{pmatrix}$$
$$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}2 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}3 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}8 & 1\\ 1 & 0\end{pmatrix}$$
$$\leftrightarrow [8;\ 3,\ 3,\ 1,\ 4,\ 1,\ 3,\ 3,\ 8]\ .$$

Of course

$$\frac{25}{3} = [8;\ 3]\ ,$$

the next 3 is a highcough about a third of the way along the period and the [1, 4, 1] has appeared a little mysteriously. We will understand it better from what follows.

3.4 Suppose that (x, y), with x, y relatively prime, satisfies the Diophantine equation

$$x^2 - Dy^2 = \pm k$$

with $k < \sqrt{D}$. Then x/y is a convergent of $\sqrt{D}$ because either

$$\left|y\sqrt{D} - x\right| < \frac{1}{2y} \qquad \text{or} \qquad \left|x\sqrt{D} - y\right| < \frac{1}{2x}\ .$$

It will be convenience in the sequel to notice that

$$\begin{pmatrix}c_0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}c_1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}c_2 & 1\\ 1 & 0\end{pmatrix}\cdots = R^{c_0}J\cdot JL^{c_1}\cdot R^{c_2}J\cdots = R^{c_0}L^{c_1}R^{c_2}L^{c_3}\cdots$$

because $J^2 = I$. Hence continued fractions correspond to words on the symbols R and L; in brief, to R-L words. We exploited this clumsily in the preceding section.

As before, set

$$M = \begin{pmatrix}x & Dy\\ y & x\end{pmatrix}\ .$$

Operating the Euclidean algorithm on the rows of M we obtain a decomposition

$$M = M_1E_1$$

with M_1 an R-L word and E_1 the *remnant*, or *ejected portion*. Other than (perhaps) for its last entry the continued fraction corresponding to M_1 is that of x/y, and E_1 is a matrix with $|\det E_1| = k$.

Now consider the matrix M^n, $n = 1, 2, \ldots$. On defining

$$E_sM_s = M_{s+1}E_{s+1}, \qquad s = 1, 2, \ldots$$

we readily obtain by induction on n

$$M^n = M_1M_2\cdots M_n\ \cdot\ E_nE_{n-1}\cdots E_1\ , \qquad n = 1, 2, \ldots\ .$$

We claim that the R-L word $M_1M_2\cdots M_n$ corresponds, except (perhaps) for its last symbol, to a convergent of $\sqrt{D}$ which is n times as far along the expansion of $\sqrt{D}$ as is x/y. To see this, consider the direct decomposition

$$M^n = M'_nE'_n\ .$$

It is enough to establish the claim for the R-L word M'_n, since M'_n is just $M_1 M_2 \cdots M_n$ together with a further R-L word that may be removable from the left of $E_n E_{n-1} \cdots E_1$. But suppose

$$M^n = \begin{pmatrix} x & Dy \\ y & x \end{pmatrix}^n = \begin{pmatrix} x_n & Dy_n \\ y_n & x_n \end{pmatrix} = \begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_{m+1} & 1 \\ 1 & 0 \end{pmatrix} E'_n$$

where the second row of E'_n is no longer dominated by its first row. Then the continued fraction (note that we omit the last partial quotient)

$$[c_0;\ c_1,\ \ldots,\ c_m] = \frac{p}{q}$$

yields both a convergent of x_n/y_n and Dy_n/x_n.

The following remark apply for any matrix of the shape

$$\begin{pmatrix} x & Dy \\ y & x \end{pmatrix}$$

with x and y nonnegative integers, so we drop the suffix n. If m is even we have

$$\frac{p}{q} < \frac{x}{y} \quad \text{and} \quad \frac{p}{q} < \frac{Dy}{x} \quad \text{so} \quad \frac{p^2}{q^2} < D \quad \text{implying} \quad \frac{p}{q} < \sqrt{D}\,.$$

Similarly if m is odd we obtain

$$\frac{p}{q} > \sqrt{D}\,.$$

Accordingly, if $x^2 - Dy^2 > 0$ we find that

$$\frac{p}{q} < \sqrt{D} < \frac{x}{y} \quad \text{or} \quad \frac{p}{q} > \sqrt{D} > \frac{Dy}{x}\,,$$

and if $x^2 - Dy^2 < 0$ then

$$\frac{p}{q} < \sqrt{D} < \frac{Dy}{x} \quad \text{or} \quad \frac{p}{q} > \sqrt{D} > \frac{x}{y}\,,$$

Since p/q is a convergent of x/y and Dy/x, we see in each case that p/q is also a convergent of $\sqrt{D}$, as claimed.

Finally, x_n/y_n and Dy_n/x_n converge to $\sqrt{D}$ so their common convergents are also close to $\sqrt{D}$. To see this, note that the minimal equation for M is

$$M^2 - 2xM + (x^2 - Dy^2)I = 0\,,$$

whence the recurrence relation

$$M^{n+2} = 2xM^{n+1} \mp kM^n, \qquad n = 0, 1, 2, \ldots\,.$$

It is elementary linear algebra to verify that

$$\begin{aligned} 2x_n &= \left(x + y\sqrt{D}\right)^n + \left(x - y\sqrt{D}\right)^n\,, \\ 2y_n\sqrt{D} &= \left(x + y\sqrt{D}\right)^n - \left(x - y\sqrt{D}\right)^n\,. \end{aligned}$$

so, approximately

$$\frac{x_n}{y_n} \approx \sqrt{D}\left(1 + 2(\pm k)^n \left(x + y\sqrt{D}\right)^{-2n}\right)$$

with a similar estimate on the opposite side of $\sqrt{D}$ for Dy_n/x_n. We have more to say on these matters in section 4.

3.5 A by-product of our remarks in the following: If

$$x^2 - Dy^2 = t$$

then there are convergents p/q of $\sqrt{D}$ which are also convergents of x/y. Thus a solution $x, y > 0$ of the given Diophantine equation necessarily arises from a solution of

$$p^2 - Dq^2 = k$$

with $|k| < \sqrt{D}$, by way of a chain

$$x_i y_{i-1} - x_{i-1} y_i = \pm 1, \qquad i = 1, 2, \ldots, m$$

with $x_0 = \pm p$, $y_0 = \pm q$ and $x_m = \pm x$, $y_m = \pm y$. More to the point, this observation yields a practical algorithm: Suppose

$$x'y - xy' = \pm 1 \quad \text{and} \quad x'^2 - Dy'^2 = t' \ .$$

We claim that there is a choice for (x', y') so that $|t'| < |t|$. This claim yields the above-mentioned algorithms since after, say, m steps we will have $|t^{(m)}| < \sqrt{D}$ and then $x^{(m)}/y^{(m)}$ must be a convergent of $\sqrt{D}$. Indeed, noting that the condition $x'y - y'x = \pm 1$ does not define (x', y') uniquely because

$$(x' + sx)y - (y' + sy)x = \pm 1$$

for any $s \in \mathbf{Z}$. On taking determinants of

$$\begin{pmatrix} x & Dy \\ y & x \end{pmatrix} \begin{pmatrix} x' & -Dy' \\ -y' & x \end{pmatrix} = \begin{pmatrix} xx' - Dyy' & D(x'y - xy') \\ x'y - xy' & xx' - Dyy' \end{pmatrix}$$

we obtain

$$(xx' - Dyy')^2 - D = tt' \ .$$

Since

$$x(x' + sx) - Dy(y' + sy) = (xx' - Dyy') + st, \qquad s \in \mathbf{Z} \ ,$$

there is no loss of generality in supposing that

$$|xx' - Dyy'| < \frac{1}{2}|t| \ .$$

And since $|t| > \sqrt{D}$, we have

$$|t'| < |t| \ .$$

It suffices to look for squares Q so that $\sqrt{Q} < \frac{1}{2}|t|$ and $tt' = Q - D$. In particular Q is such that $t \mid (Q - D)$. If the given equation has a solution, we obtain one or more values of t' and new equations

$$x'^2 - Dy'^2 = t'$$

with $|t'| < |t|$. Repeating the argument if necessary we eventually obtain equations

$$x^{(m)^2} - Dy^{(m)^2} = t^{(m)}$$

with $\left|t^{(m)}\right| < \sqrt{D}$. These equations can be solved and the solutions traced back to yield possible solutions (x, y) in $\mathbf{Z}$.

3.6 Of course one obtains the continued fraction expansion of $\sqrt{D}$ from

$$\sqrt{D} = \frac{\sqrt{D}+P_0}{Q_0} = c_0 + \frac{\sqrt{D}-c_0Q_0+P_0}{Q_0}, \quad P_0 = 0,\ Q_0 = 1,\ c_0 = \left[\sqrt{D}\right]$$

$$\vdots$$

$$\frac{\sqrt{D}+P_{m-1}}{Q_{m-1}} = c_{m-1} + \frac{\sqrt{D}-c_{m-1}Q_{m-1}+P_{m-1}}{Q_{m-1}}$$

$$\frac{\sqrt{D}+P_m}{Q_m} = c_m + \frac{\sqrt{D}-c_mQ_m+P_m}{Q_m}$$

$$\vdots$$

$$\text{with } Q_{m-1}Q_m = D - P_m^2,\ \ c_m = \left[\frac{\sqrt{D}+P_m}{Q_m}\right],$$

$$P_m = c_{m-1}Q_{m-1} - P_{m-1}\ .$$

It is easy to confirm, say by induction on m, that the P_m are nonnegative integers and the Q_m positive integers. The theory we sketched at the beginning of this section, and more, can be recovered by observing that if p_m/q_m denotes the convergents of $\sqrt{D}$ then

$$p_m^2 - Dq_m^2 = (-1)^{m+1}Q_{m+1}, \qquad m = 0, 1, 2, \ldots\ .$$

The remarks of this subsection are readily generalised to arbitrary real quadratic irrationals. We will not go into these well-known matters here. However, an example will yield useful data for use below:

Example : Set $\alpha = \sqrt{46}$. Then

$$\begin{aligned}
\alpha &= 6 + \alpha - 6\\
\frac{\alpha+6}{10} &= 1 + \frac{\alpha-4}{10}\\
\frac{\alpha+4}{3} &= 3 + \frac{\alpha-5}{3}\\
\frac{\alpha+5}{7} &= 1 + \frac{\alpha-2}{7}\\
\frac{\alpha+2}{6} &= 1 + \frac{\alpha-4}{6}\\
\frac{\alpha+4}{5} &= 2 + \frac{\alpha-6}{5}\\
\frac{\alpha+6}{2} &= 6 + \frac{\alpha-6}{2}
\end{aligned}$$

$\vdots$ and now an evident symmetry yields, eventually

$$\alpha + 6 = 12 + \alpha - 6$$

and we restart the period, having obtained:

$$\sqrt{46} = [6;\ \overline{1,\ 3,\ 1,\ 1,\ 2,\ 6,\ 2,\ 1,\ 1,\ 3,\ 1,\ 12}\ ]\ .$$

Moreover

n	c_n	p_n	q_n	
		0	1	
		1	0	
0	6	6	1	$6^2 - 46 \cdot 1^2 = -10$
1	1	7	1	$7^2 - 46 \cdot 1^2 = 3$
2	3	27	4	$27^2 - 46 \cdot 4^2 = -7$
3	1	34	5	$34^2 - 46 \cdot 5^2 = 6$
4	1	61	9	$61^2 - 46 \cdot 9^2 = -5$
5	2	156	23	$156^2 - 46 \cdot 23^2 = 2$
6	6	·	·	·
⋮	⋮	⋮	⋮	⋮

4. Multiplying a continued fraction (by a rational)

4.1 Given

$$\alpha = [c_0;\ c_1,\ c_2,\ \ldots]$$

and integers a, b, c, d with $ad - bc \neq 0$, it is plain that the continued fraction of

$$\beta = \frac{a\alpha + b}{c\alpha + d}$$

corresponds to the matrix product

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots .$$

But we must transform this product to one corresponding to a continued fraction. An easy case is

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots = \begin{pmatrix} -c_0 - 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 - 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots$$

Indeed, say

$$\frac{36}{17} = [2;\ 8,\ 2] \quad \text{and} \quad -\frac{36}{17} = [-3;\ 1,\ 7,\ 2]$$

whilst $53/8 = [6;\ 1,\ 1,\ 1,\ 2]$ so

$$-\frac{53}{8} = [-7;\ 1,\ 0,\ 1,\ 1,\ 2] = [-7;\ 2,\ 1,\ 2]\ .$$

In general, there is no such simple formula and we must develop a systematic approach to the problem of multiplying an R-L word

$$R^{c_0} L^{c_1} R^{c_2} L^{c_3} \cdots$$

on the left by some given matrix and then commuting that matrix through the word until, so to speak, the matrix disappears in the $\cdots$ on the right, leaving some R-L word on its left. There are two steps to our approach. The first is to recognise that we need only consider multiplication by matrices of a restricted shape. The second is to construct simple tables (formulas) to facilitate the commuting process to which we have alluded.

We will deal only with matrices with just nonnegative integer entries, unless we say explicitly that we are not. We will see that this loses no generality. Given a

matrix N, we apply the Euclidean algorithm to its rows until the matrix is *row-balanced*: neither row dominates the other. We now apply the Euclidean algorithm to the columns of the row-balanced matrix until we have a matrix that is also *colum-balanced*: neither column dominates the other. We obtain

$$N = W_1 M W_2$$

with M *doubly-balanced* and both W_1, W_2 finite R-L words. Plainly if we can multiply by M (thus commute M through R-L words) then we can multiply by N. Hence it suffices to learn to multiply by doubly-balanced matrices. Going further, we may restrict ourselves to double-balanced matrices N with $\det N = p > 0$, p prime. Nevertheless, the remarks below apply somewhat more generally. Indeed suppose only that N is row-balanced with $\det N > 0$, and

$$N = \begin{pmatrix} a & b \\ c & d \end{pmatrix} .$$

Then $a - c > 0$ and $d - b > 0$ whilst

$$\det N = (a-c)(d-b) + c(d-b) + b(a-c) .$$

Each quantity on the right is nonnegative. It follows that there are only finitely many possibilities for the entries of N, thus that there are only finitely many row-balanced matrices of given determinant and *a fortiori* only finitely many such doubly-balanced matrices. Moreover, if W is an R-L word with sufficiently many letters (counted according to multiplicity) then NW has entries too large for it to be row-balanced: so an R-L word can be withdrawn from its left.

Example 1 : $\det N = 3$.

	$A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$	$B = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$	$A' = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$
A	$R : R^3$ $L^3 : L$	$LR : R$	$L^2R : RL^2$
B	$L : LR$		$R : RL$
A'	$R^2L : LR^2$	$RL : L$	$R^3 : R$ $L : L^3$

The table lists the three double-balanced matrices of determinant 3 and their transitions (commuting properties). In detail, and in a quite different format:

$$\begin{array}{ccc} AR = R^3A & BL = LRA & A'L = L^3A' \\ AL^2R = RB & BR = RLA' & A'RL = LB \\ AL^2R = RL^2A' & & A'R^2L = LR^2A \\ AL^3 = LA & & A'R^3 = RA' \end{array}$$

Happily, only a few transitions need to be computed directly. In particular, double-balance is preserved under the transpose, and for example

$$AR = R^3A \qquad \xrightarrow{T} \qquad LA = AL^3 .$$

Further, in general: If

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} , \quad \text{set } M' = JMJ = \begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

Then

$$ALR = RB \quad \overset{'}{\longrightarrow} \quad A'RL = LB$$
$$A'RL = LB \quad \overset{T}{\longrightarrow} \quad RLA' = BR$$

In this manner the transitions of A yield the entire transition table in the present example. Note that the *double-transpose* $'$ preserves both row- and column-balance.

Example 2 : A yet simpler case permits us to illustrate the allegation concluding section 2. Here $\det N = 2$.

	$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$	$A' = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$
	$R : R^2$	$LR : RL$
A	$L^2 : L$	
		$R^2 : R$
A'	$RL : LR$	$L : L^2$

In [2], observation 2.12 we see that

$$\sum_{h=0}^{\infty} 2^{-2^h} = [0;\ 1,\ 4,\ S_{4,6}^{\infty}(2,4)]$$
$$= [0;\ 1,\ 4,\ 2,\ 4,\ 4,\ 6,\ 4,\ 2,\ 4,\ 6,\ 2,\ 4,\ 6,\ 4,\ 4,\ 2,\ 4,\ 6,\ 2,\ 4,\ \ldots]$$

To obtain the continued fraction of

$$2\sum_{h=0}^{\infty} 2^{-2^h}$$

we set our work as follows (noticing that we are multiplying by A):

```
   ;1  4     2    4      4         6      4      2   4
   LR  R²  RL  LR  R²  RL  L²  LR  R²  R²  RL  L²  LR  RL  L²  LR  R²  ...
A  A'  A'  A    A'  A'  A   A   A'  A'  A'  A   A   A'  A   A   A'  A   ...
   RL  R   LR  RL  R  LR  L   RL  R   R   LR  L   RL  LR  L   RL  R   ...
   1;1 1   1   2 1   1   11  1     11    2   11  1   1   2 1 1   11
```

This provides a nice instance of *finite-state transducer* (the *states* being A and A') transforming a given R-L sequence to a new sequence corresponding to the continued fraction:

$$[1;\ 1,\ 1,\ 1,\ 1,\ 2,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 2,\ 1,\ 1,\ 1,\ 1,\ 2,\ 1,\ 1,\ 1,\ 1,\ \ldots]$$

which, with a little more work, turns out to be

$$[1;\ 1,\ 1,\ 1,\ S_{1,1,1,1,2,1}^{\infty}(2,1,1,1)]\ .$$

Our tabulation of the transduction should be read as

$$ALR = RLA',\ \ A'R^2 = RA',\ \ A'RL = LRA,\ \ ALR = RLA',\ \ \ldots$$

The examples do not lie: In general there are exactly p doubly-balanced matrices of determinant p, and these matrices are readily computed. Indeed suppose that

$$N = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is row-balanced with $\det N > 0$. Set

$$\frac{d-b}{a-c} = [a_0;\ a_1,\ \ldots,\ a_m]$$

and denote by W_N the R-L word corresponding to $[a_0;\ a_1,\ \ldots,\ a_{m-1},\ a_m - 1]$. Then NW_N has the special shape

$$\begin{pmatrix} s+g & t \\ s & t+g \end{pmatrix}$$

with $g = \gcd(d-b, a-c)$. It follows that

$$\det N = \det NW_N = g(g+s+t)\ .$$

Hence if $\det N = p$ we must have $g = 1$ and $s + t = p - 1$, yielding p possibilities in all. Moreover, the set of doubly-balanced matrices of given determinant is in one-one correspondence with the matrices of that determinant with special shape above. To see this, one verifies that if N is row-balanced and NW has special shape then $W = W_N$, whilst conversely if N_1, N_2 are doubly-balanced and $N_1 W_{N_1} = N_2 W_{N_2}$ then $W_{N_1} = W_{N_2}$ so $N_1 = N_2$. Finally, every row-balanced matrix, and *a fortiori* each of special shape, yields a doubly-balanced matrix by decomposing it by the Euclidean algorithm on its columns.

Example 3 : The doubly-balanced matrices of determinant 7 are

$$\begin{aligned} A &= \begin{pmatrix} 7 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 7 & 0 \\ 6 & 1 \end{pmatrix} L^{-6},\ \ B = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 5 & 2 \end{pmatrix} L^{-2}, \\ C &= \begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 4 & 3 \end{pmatrix} L^{-1},\ \ D = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}, \\ C' &= \begin{pmatrix} 3 & 1 \\ 2 & 3 \end{pmatrix},\ \ B' = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix},\ \ A' = \begin{pmatrix} 1 & 0 \\ 0 & 7 \end{pmatrix}\ . \end{aligned}$$

Alternatively, given A, the other matrices may be obtained from

$$\begin{aligned} ALR &= R^3 B \qquad & AL^2R &= R^2 C' \qquad & AL^3R &= RD \\ AL^4R &= RLC \qquad & AL^5R &= RL^2 B' \qquad & AL^6R &= RL^6 A' \end{aligned}$$

a few extra computations, together with the transformations ^{T} and $'$ yield the entire transition table.

With these remarks, we are in position to obtain the continued fraction of

$$\beta = \frac{a\alpha + b}{c\alpha + d}$$

given that of α. It may be helpful to remark that for $\alpha < 0$ one may change the multiplier to

$$\begin{pmatrix} -a & b \\ -c & d \end{pmatrix}$$

and deal with $-\alpha > 0$. Moreover, one may compute $-\beta$ using

$$\begin{pmatrix} -a & -b \\ c & d \end{pmatrix}$$

and later obtain β. Further, for any $k \neq 0$,

$$\begin{pmatrix} ka & kb \\ kc & kd \end{pmatrix}$$

yields the same β, so we may always suppose that $\gcd(a,b,c,d)=1$, and by taking $k=-1$ if necessary, that at most two of a, b, c, or d are negative. Replacing α by α^{-1} changes the multiplier to

$$\begin{pmatrix} b & a \\ d & c \end{pmatrix} .$$

allowing one to presume that the multiplier has positive determinant. In any event, we may decompose the multiplier so that it becomes a product of matrices R, L, R^{-1} and L^{-1} and doubly-balanced matrices with prime determinant. If one wished one could construct transition tables for the case of composite determinant, but this may be avoided. Raney [5], from which we have translated all the remarks of this section, provides a detailed algorithm to transform an arbitrary multiplier to one of doubly-balanced shape.

4.2 In section 3.5 we allege that given any part of the continued fraction of $\sqrt{D}$ we may obtain more of the period by consider powers of a corresponding matrix.

Example 4 : $D=46, \quad 7^2-46\cdot 1^2=3.$

$$M=\begin{pmatrix} 7 & 46 \\ 1 & 7 \end{pmatrix}=\begin{pmatrix} 7 & 6 \\ 1 & 1 \end{pmatrix} RA'R = R^6L\cdot RA'R\,,$$

where

$$[6;\ 1]=\frac{7}{1} \quad \text{and} \quad A'=\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} .$$

Recall that we write $M=M_1E_1$ so $M_1=R^6LR$, $E_1=A'R$ and then sequentially compute

$$E_sM_s=M_{s+1}E_{s+1}, \qquad s=1,2,3,\ldots$$

obtaining

$$M^n=M_1M_2\cdots M_n \ \cdot\ E_nE_{n-1}\cdot E_1\ .$$

We employ the transition table of Example 1, and the setting-out of Example 2:

		$A'R$	R^3	R^3	RL	R	R^2L	R	L^2R	R^3	R^2L	LR
			A'	A'	B	A'	A	AL	A'	$A'R$	A	B
R^6	L	R	R	R	L	RL	LR^2	R^3	RL^2	R	LR^2	R

6	1		3		1	1	2		6	2	1	1	3
	M_1				M_2			M_3			M_4		M_5

and continuing

	L	R	R	R	LR	R	R^3	R^3	R^3	RL	R	R^2L	R
B	A	A	A	A	B	A'	A'	$A'R^2$	A'	B	$A'R$	A	AL
LR^3	LR	R^3	R^3	R^3	R	RL	R	R	R	L	RL	LR^2	R^3

	1			12			1		3			1	1	2
			M_6					M_7				M_8		M_9

Noting that $M_3=M_9=LR^5$ and $E_3=E_9=AL$ we verify periodicity. Our computation proves that

$$\sqrt{46}=[6;\ 1,\ 3,\ 1,\ 1,\ 2,\ \overline{6,\ 2,\ 1,\ 1,\ 3,\ 1,\ 12,\ 1,\ 3,\ 1,\ 1,\ 2}\ \ldots];$$

note that the final entry 5 we obtain is not necessarily complete, and in this case is in face an imcomplete quotient.

Example 5 : In section 3.3 we found that

$$\sqrt{69} = [8;\ \overline{3,\ 3,\ 1,\ 4,\ 1,\ 3,\ 3,\ 16}\ ]$$

with $[8;\ 3] = 25/3$ and $25^2 - 69 \cdot 3^2 = 4$. Then with

$$M = \begin{pmatrix} 25 & 69 \cdot 3 \\ 3 & 25 \end{pmatrix}$$

we showed that

$$\frac{1}{8} M^3 J \leftrightarrow [8;\ 3,\ 3,\ 1,\ 4,\ 1,\ 3,\ 3,\ 8]\ .$$

We recall that $M^3 J = M^2 J M^T$ (because $M^T = M'$). But

$$M = R^8 L^3 \cdot R^2 \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} R^{-1}\ .$$

We employ the transition table of Example 2, and multiply twice by

$$A' = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

	R^2	R^2	R^2	RL	L^2	R	R
$A'R^{-1}$	A'	A'	A'	A	A	A	AA'^2R^{-1}
$R^8L^3R^2A'^2R^{-1}$	R	R	R	LR	L	R^2	R^2

	R^2	RL	R	LR	R^2
A'	A'	A	A	A'	$A'RAA'^2R^{-1}$
$R^3LRLR^4AA'^2R^{-1}$	R	LR	R^2	RL	R

We read off that

$$M^2 = R^8 L^3 R^2 \cdot RLR^4 L \cdot RA'RAA'^2 R^{-1}\ .$$

But

$$JM^T = J \cdot L^{-1} A'^2 L^2 R^3 L^8 = R^{-1} A^2 R^2 L^3 R^8 J\ .$$

The mess we have left resolves itself in that

$$A'RAA'^2R^{-1} \cdot R^{-1}A^2 = 8I \quad (AA' = 2I,\ RA' = A'R^2),$$

so, indeed, $\frac{1}{8}M^3J$ yields what it should. As suggested back in section 3.3, if we have $x^2 - Dy^2 = \pm 4$ then the continued fraction of x/y yields (the first) third of the period of the continued fraction of $\sqrt{D}$ and, by symmetry, the last third. The middle third of the period is, up to slight perturbations at its edges, obtained by multiplying (or dividing, that being much the same thing) the given third by 4.

Example 6 : In applying these ideas a little care is required. In particular, if the portion of the continued fraction yielding the matrix we called M in section 3 is of odd length, we need to keep track of the J. For example, noting

$$164^2 - 61 \cdot 21^2 = -5, \qquad \frac{164}{21} = [7;\ 1,\ 4,\ 3,\ 10]\ ,$$

we write

$$M = \begin{pmatrix} 164 & 61 \cdot 21 \\ 21 & 164 \end{pmatrix} = R^7 L R^4 L^3 R J \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} = R^7 L R^4 L^3 R \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} L^4 J\ .$$

The transition table for determinant 5 is

	$A = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$	$B = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$	$C = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$	$B' = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$	$A' = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$
A	$R : R^5, L^5 : L$	$LR : R^2$	$L^2R : R$	$L^3R : RL$	$L^4R : RL^4$
B	$L^2 : LR$			$R : R$	$LR : RL^3$
C	$L : LR^2$				$R : RL^2$
B'	$RL : LR^3$	$L : L$			$R^2 : RL$
A'	$R^4L : LR^4$	$R^3L : LR$	$R^2L : L$	$RL : L^2$	$R^5 : R, L : L^5$

Write, for example

	L^{10}	LR	L^2	L^2R	R	RL	R^2	L^2	RL	L	LR
AL^4	A	B	A	C	A'	B'	A'	A'	B'	B	$A'R^3$
$J : L^7RL^4R^3L$					$R^2L^2RL^3R^4$						$LR^{14}LR^3$
$R^7LR^4L^3R$	L^2	R^2	LR	R	RL^2	L^2	RL	L^{10}	L^2	L	RL^3

7 1 4 3 1 2 2 1 3 4 1 14 1

discovering that

$$\sqrt{61} = [7;\ 1,\ 4,\ 3,\ 1,\ 2,\ 2,\ 1,\ 3,\ 4,\ 1,\ 14,\ 1,\ \ldots] .$$

We need not go on to see periodicity explicitly since we know the end of the period to be signalled by the large partial quotient $2[\sqrt{D}]$; in this case, 14.

4.3 Here and there we have used some extraneous but informative terminology. Namely, it is evident that multiplication transforms *R-L* sequences to *R-L* sequences. Moreover, by concentrating on doubly-balanced multipliers, we see that we can split the multiplication into steps. At each step we have a *state* (the doubly-balanced multiplier), part of the sequence is transformed, and we move to some new state. But this is just a *finite-state* (there being only finitely many doubly-balanced matrices of the given determinant) *transduction* of *R-L* sequences, familiar from the theory of finite automata. And from that theory we learn immediately taht certain properties, in particular periodicity, are preserved under finite-state transduction.

5. The object of these notes has been to mention various facts concerning continued fractions in a setting that aided a description of Raney's algorithm that, it is to be hoped, minimal discourages genaralisation to higher-dimensional analogues.

In particular, the following generalisation of the work on quadratic irrationals suggests itself: Given a basis $\theta_1, \theta_2, \ldots, \theta_r$ of a number field $\mathbf{K}/\mathbf{Q}$ viewed as an

r-dimensional $\mathbf{Q}$-vector space, each element α of $\mathbf{K}$ has a given matrix representation, namely the $r \times r$ matrix (with respect to the given basis) of the $\mathbf{Q}$-linear transformation

$$\mathbf{K} \longrightarrow \mathbf{K} \quad : \quad \beta \longrightarrow \alpha\beta \; .$$

For example, with $\mathbf{K} = \mathbf{Q}\left(\sqrt{D}\right)$ and $\theta_1 = 1$, $\theta_2 = \sqrt{D}$, the matrix of $x + y\sqrt{D}$ is

$$M = \begin{pmatrix} x & Dy \\ y & x \end{pmatrix} \; .$$

Thus ibn the r-dimensional case on might hope to achieve the following: Given $\alpha = x_1\theta_1 + \cdots + x_r\theta_r$ yielding a 'good approximation' of the basis $\theta_1, \ldots, \theta_r$ (say, α has small norm relative to the x_i in $\mathbf{Z}$):

(i) Obtain the matrix M corresponding to α, as above.

(ii) Decompose $M = M_1E_1$ with M_1 unimodular, and 'maximal' in some sense, and hence compute

$$M^n = M_1M_2 \cdots M_n \; \cdot \; E_nE_{n-1} \cdots E_1 \; .$$

One might expect that in some sense $M_1M_2 \cdots M_n$ corresponds to some 'later' or better approximation than α. More generally, given n not necessarily different 'good' approximations one might hope that the product of their corresponding matrices yields new approximations. However it is not as clear just how one should define and effect the suggested decomposition of M.

References

[1] Stark, H. M. *An Introduction to Number Theory*, MIT Press, 1978.

[2] F. M. Dekking, M. Mendès France and A. J. van der Poorten, *Folds!* The Mathematical Intelligencer, **4** (1982), 130-138.

[3] H. J. S. Smith, *De compositione numerorum primorum formae* $4\lambda+1$ *ex duobus quadratis*, J. reine angew. Math. (Crelle) **50** (1855).

[4] L. Auteurs, *Letter to the Editor*, The Mathematical Intelligencer, **5** (2) (1983), 5.

[5] G. N. Raney, *On Continued Fractions and Finite Automata*, Math. Ann. **206** (1973), 265-283.