Notes on continued fractions

1. Chapter 49: The Topsy-turvy world of continued fractions

First, let's go back, way back, to the Euclidean algorithm. Let's say for 23 and 5. If we run this through we get

$$\begin{aligned} 23 &= 4 * 5 + 3 \\ 5 &= 1 * 3 + 2 \\ 3 &= 1 * 2 + 1 \\ 2 &= 2 * 1. \end{aligned}$$

Now if we divide through appropriately we instead get...

$$\begin{aligned} \frac{23}{5} &= 4 + \frac{3}{5} \\ \frac{5}{3} &= 1 + \frac{2}{3} \\ \frac{3}{2} &= 1 + \frac{1}{2} \end{aligned}$$

and the last one just stays $2 = 2$. But notice that the last fraction we reach in each line becomes the reciprocal of the next line's fraction. So we can stack these up nd get

$$\frac{23}{5} = 4 + \frac{1}{5/3} = 4 + \frac{1}{1 + \frac{2}{3}} = 4 + \frac{1}{1 + \frac{1}{1 + \frac{1}{2}}}.$$

This looks rather peculiar. We call this a continued fraction expansion of 23/5. We could do this with any pair of integers, expanding them off in this way. We could even start with a smaller number in the numerator (we'd just treat the first digit as 0 then). Even if the gcd of our two integers is not 1, when we reduce the fractions it will become 1, so we can still do that.

Let's include a bit of notation to make this more readable and less typographically intensive. We write

$$23/5 = [4; 1, 1, 2].$$

To signify the continued fraction with these associated digits. (Note that Silverman does not use the semicolon. I prefer it as it makes it clear whether there is a zeroth digit or not.)

Observation 1: This continued fraction process works for any rational number and since the euclidean algorithm is finite, so will the continued fraction eventually stop. Moreover the reverse is true, if we could write out a number as $[a_0; a_1, a_2, \ldots, a_n]$ then it should be rational.

Observation 2: For the last digit in 23/5 we could have chosen 2 or $1 + 1/1$, namely $23/5 = [4; 1, 1, 2] = [4; 1, 1, 1]$. This ambiguity always exists for rational numbers: we can always get $[\ldots, n + 1] = [\ldots, n, 1]$. So there are always 2 continued fraction expansions.

(Note: As the grad students saw, for instance, there is more than one kind of Euclidean algorithm, which gives rise to more than one kind of continued fraction expansion but we'll be sticking to this one for now.)

Observation 3: Why did this procedure work? We started with a number at least as large as 1 (except maybe in the "zeroth" step), we wrote it as its floor plus some number that was in the interval $[0, 1)$. Then for this last number, we took it's reciprocal, giving us a number larger than 1 and starting the whole process over again.

This suggests that we could run through the whole process with a number other than a rational number...

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

Huh, this seems to suggest that $\sqrt{2} = [1; 2, 2, 2, \dots]$, but what does that really mean? And is it useful?

Well, we're going to show a couple things about these strange continued fraction expansions: First, if you want to understand how to approximate an irrational by rationals, looking at this continued fraction expansion is really helpful. Second, if you want to understand square roots of integers (or more generally quadratic irrationals), continued fractions are a GREAT way to do that too. Third, as suggested by the previous two, it turns out that continued fractions help solve Pell's equation with relative ease.

Well, let's bring in some definitions. For an infinite continued fraction $[a_0; a_1, a_2, \dots]$ we let p_n/q_n (the nth convergent) denote the rational number given by the finite truncation $[a_0; a_1, a_2, \dots, a_n]$. We say that $x = [a_0; a_1, a_2, \dots]$ if $\lim_{n\to\infty} p_n/q_n = x$. (This isn't terribly different from our notion of convergence for base 10 expansions!)

Moreover we can for any x define the continued fraction digits of x in the following way: $a_0 = \lfloor x \rfloor$, $x_0 = x - a_0$, and then iteratively define $a_i = \lfloor 1/x_{i-1} \rfloor$ and $x_i = 1/x_{i-1} - \lfloor 1/x_{i-1} \rfloor$. so with these definitions

$$x = a_0 + x_0 = a_0 + \frac{1}{1/x_0} = a_0 + \frac{1}{a_1 + x_1} = \dots$$

or alternately

$$x = [a_0 + x_0;] = [a_0; a_1 + x_1] = [a_0; a_1, a_2 + x_2].$$

We stop this creation of digits if ever $x_i = 0$.

So the big thing we will be looking to prove first is that (a) all infinite continued fractions like this converge, and (b) if we start from a given x, this process yields an infinite continued fraction which converges to x (and not to some other point).

To do this, let's introduce a little more notation, namely a linear fractional transformation defined by matrices:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} x = \frac{ax+b}{cx+d}.$$

If the determinant of the matrix on the left is non-zero then this transformation is in fact a bijection (on the whole real line plus the point at infinity). Why care about this? Here's some things that should look familiar...

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} x = x + a \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} x = \frac{1}{x} \qquad \begin{pmatrix} 0 & 1 \\ 1 & a \end{pmatrix} x = \frac{1}{a+x}.$$

Fact This is in fact a group action. By this I mean that if you have two matrices M_1 and M_2 then $M_1(M_2 x) = (M_1 M_2)x$, i.e. you can choose whether you want to manipulate matrices first or apply them to the point in order.

Lemma Let $a_0 \in \mathbb{Z}$, $a_i \in \mathbb{N}$ for $1 \le i \le n$. Then

$$\begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix}.$$

where p_{n-1}/q_{n-1} and p_n/q_n are $[a_0; a_1, a_2, \dots, a_{n-1}]$ and $[a_0; a_1, a_2, \dots, a_n]$ respectively.

Observation The left hand side of this equation, seen as a matrix action, is just "appending" a partial continued fraction expansion to a point. Very simply we have that $\begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} x_n = x_{n-1}$. More generally we get that

$$x = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} x_n = \frac{p_{n-1}x_n + p_n}{q_{n-1}x_n + q_n}.$$

Proof Let the matrix on the right hand side of the equation be $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. Note that $\det(M) = (-1)^n$ since each of the matrices on the left except the first has determinant -1.

We will assume that $a_0 \geq 0$ for simplicity.

We have $M0 = B/D$. But at the same time if we apply M piece by piece to 0 we get that it should equal

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{a_n + 0}}} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{a_n}}}.$$

So $B/D = p_n/q_n$. Since M has determinant ± 1, B and D cannot have any common factors and thus are in lowest terms, but might be negative. However, if we look at how M was made, all its terms are positive so $B = p_n$ and $D = q_n$.

To see that the left column is also what we want, just note that by the work we've already done, we have

$$\begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} = \begin{pmatrix} * & p_{n-1} \\ * & q_{n-1} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}$$

But comparing this with M we see that $A = p_{n-1}$ and $C = q_{n-1}$.

Observation As an immediate consequence, we get that $p_{n-1}q_n - p_n q_{n-1} = (-1)^n$.

Lemma For a sequence $a_0 \in \mathbb{Z}$, $a_i \in \mathbb{N}$ for $i \geq 1$, the sequence of convergents may be defined recursively by $p_{-1} = 1$, $q_{-1} = 0$, $p_0 = a_0$, $q_0 = 1$ and then $p_n = a_n p_{n-1} + p_{n-2}$ and $q_n = a_n q_{n-1} + q_{n-2}$.

Proof (Sketch) This follows by applying the previous lemma and noting that

$$\begin{pmatrix} p_{n-2} & p_{n-1} \\ q_{n-2} & q_{n-1} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix}.$$

Note This is why I like bringing in matrix actions. If you read the chapter in Silverman, this formula takes a bit of work: it's an induction argument and it's not clear exactly why it is true, but with matrix actions it is almost trivial. This also explains why continued fractions show up a lot in places where one is studying matrices in $SL_2(\mathbb{Z})$.

Observation since all the q_n's, $n \geq 0$, are positive, we see that the sequence of q_n's is strictly increasing for $n \geq 1$.

Lemma Any infinite continued fraction expansion converges (provided all the $a_n > 0$, $n \geq 1$).

Proof We have to prove that p_n/q_n converges to something as $n \to \infty$. Consider the difference

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

This is alternating and strictly decreasing in absolute value for large enough n (since $q_n > q_{n-1}$). Thus the sum of these terms converges by the alternating series test, that is, the sum

$$\left(\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}}\right) + \left(\frac{p_{n-1}}{q_{n-1}} - \frac{p_{n-2}}{q_{n-2}}\right) + \cdots + \left(\frac{p_2}{q_2} - \frac{p_1}{q_1}\right) = \frac{p_n}{q_n} - \frac{p_1}{q_1}$$

converges, and thus p_n/q_n converges.

But we want more than this, we want that the continued fraction expansion of x actually converges to x itself, not something else.

Lemma We have for any x with at last n continued fraction digits that

$$x - \frac{p_n}{q_n} = \frac{(-1)^n x_n}{q_n(q_n + q_{n-1}x_n)}.$$

Since $x_n \in [0,1)$ and $q_{n-1} \leq q_n$, this implies that

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2}$$

and thus the infinite continued fraction expansion for x does converge to x.

Proof This is again quite direct.

$$x - \frac{p_n}{q_n} = \frac{p_{n-1}x_n + p_n}{q_{n-1}x_n + q_n} - \frac{p_n}{q_n} = \frac{q_n(p_{n-1}x_n + p_n) - p_n(q_{n-1}x_n + q_n)}{q_n(q_n + q_{n-1}x_n)}$$

The numerator simplifies down to the desired thing.

Note If we rearrange our last inequality we get $|q_n x - p_n| < 1/q_n$, thus all the continued fraction convergents give solutions to the diophantine inequality we studied earlier! This gives the first indication that studying these continued fractions is a good way to solve Pell's equation.

We can in fact do even better:

Lemma We have

$$\frac{1}{(a_{n+1}+2)q_n^2} < \left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{a_{n+1}q_n^2}$$

Proof By the previous lemma, we have

$$\left| x - \frac{p_n}{q_n} \right| = \frac{1}{q_n^2\left(\frac{1}{x_n} + \frac{q_{n-1}}{q_n}\right)} = \frac{1}{q_n^2\left(a_{n+1} + x_{n+1} + \frac{q_{n-1}}{q_n}\right)}$$

The result follows by noting that both $x_{n+1} \in [0,1)$ and $q_{n-1}/q_n \in [0,1]$.

Note This tells us something even more interesting. If we want really good rational approximations, then we should hope that x has very large digits in its continued fraction expansion. If the digits are arbitrarily large then we can actually find solutions to $|x - p/q| < c/q^2$ for any constant c.

This also suggests, although does not directly prove, that if the digits are small, then it should be hard to get too good of an approximation to x by rationals.

2. Chapter 47.5: a digression on best approximation

Let's show that continued fractions really are as powerful at approximating irrationals as we claim. Let's call a fraction p/q in lowest terms a best approximation (of the second kind) to an irrational number x if $|qx - p| < |q'x - p'|$ for all other rational numbers p'/q' with $0 < q' \leq q$.

Theorem Every best approximation to a point x is a convergent p_n/q_n for some n.

Proof In the interest of time, let's assume that $x \in [0,1)$ and let's skip checking any other x's as well as any best approximation equal to an integer.

So we may assume that p/q is a best approximation and is in the interval $(0,1)$. Recall that the convergents start with $p_0/q_0 = 0$ and then bounce back and forth on either side of x, so

$$0 = \frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \cdots \leq x \leq \ldots \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1} < 1.$$

To prove that p/q is a convegent, let us instead assume that p/q is NOT a convergent and derive a contradiction.

It's possible that $p/q > p_1/q_1$, but in this case, we have

$$\left| x - \frac{p}{q} \right| > \left| \frac{p_1}{q_1} - \frac{p}{q} \right| \geq \frac{1}{qq_1},$$

so $|qx - p| > 1/q_1 = 1/a_1$, where a_1 is the first CF digit of x. But in order for a_1 to be the first digit of x, we would need to have that $|x| < 1/a_1$, thus $0/1$ is a better approximation than p/q even though $1 \leq q$, so p/q cannot be a best approximation.

Otherwise, we should be able to find a k so that p/q is between p_{k-1}/q_{k-1} and p_{k+1}/q_{k+1}. So we have

$$\frac{1}{q_k q_{k-1}} = \left|\frac{p_k}{q_k} - \frac{p_{k-1}}{q_{k-1}}\right| > \left|\frac{p}{q} - \frac{p_{k-1}}{q_{k-1}}\right| \geq \frac{1}{q q_{k-1}}.$$

By comparing the two sides of this we see that $q > q_k$.

But at the same time we have that

$$|qx - p| = q\left|x - \frac{p}{q}\right| \geq q\left|\frac{p_{k+1}}{q_{k+1}} - \frac{p}{q}\right| \geq q\frac{1}{q_{k+1}q} = \frac{1}{q_{k+1}},$$

and

$$|q_k x - p_k| = q_k\left|x - \frac{p_k}{q_k}\right| \leq q_k \cdot \frac{1}{q_k q_{k+1}} = \frac{1}{q_{k+1}}.$$

But together this implies that $|qx - p| \geq |q_k x - p_k|$ while $q > q_k$, which implies that it cannot be a best approximation.

Theorem Every convergent is a best approximation. The sole exception is the case when $x = a_0 + 1/2$ and $p_0/q_0 = a_0$.

Note we're just going to show that p_n/q_n for $n \geq 1$ is always a best approximation.

Proof As part of a previous proof we showed that

$$\left|x - \frac{p_n}{q_n}\right| = \frac{1}{q_n^2\left(a_{n+1} + x_{n+1} + \frac{q_{n-1}}{q_n}\right)}.$$

So, taking advantage of the fact that $q_{n+1} = a_{n+1}q_n + q_{n-1}$ we have that

$$|q_n x - p_n| = \frac{1}{q_{n+1} + q_n x_{n+1}}.$$

We also have that

$$q_{n+2} + q_{n+1}x_{n+2} \geq q_{n+2} = a_{n+2}q_{n+1} + q_n > q_{n+1} + q_n x_{n+1}$$

since $a_{n+2} \geq 1$ and $x_{n+1} \in [0, 1$. Thus, the terms $|q_n x - p_n|$ are strictly decreasing as n increases.

Suppose that p_n/q_n, $n \geq 1$ is not a best approximation. Let p/q be the rational number that minimizes $|qx - p|$ with $q \leq q_n$. This is a best approximation and thus is a convergent p_k/q_k with $n \geq k \geq 0$. But the terms $|q_k x - p_k|$ are decreasing with k, so k has to equal n and thus p_n/q_n is a best approximation.

Theorem let x be an irrational number. Then x is badly approximable—that is, there exists a constant C such that

$$\left|x - \frac{p}{q}\right| \geq \frac{C}{q^2}$$

for all rationals p/q—if and only if the CF digits of x are bounded.

Proof. Since we had shown that $|x - p_n/q_n| < 1/a_{n+1}q_n^2$, it is clear that if the digits are unbounded then x cannot be badly approximable.

So suppose the digits are bounded, say $a_n < A$ for all n.

Note that $q_n = a_n q_{n-1} + q_{n-2} \leq a_n q_{n-1} + q_{n-1} < (A+1)q_{n-1}$, so the convergent denominators do not grow too quickly.

Now consider an arbitrary rational number p/q. Let n be given by $q_{n-1} \leq q < q_n$. Note that $q/q_n \geq q_{n-1}/q_n > 1/(A+1)$

Now, we have

$$\left|x - \frac{p}{q}\right| = \frac{1}{q}|qx - p| \geq \frac{1}{q}|q_n x - p_n|$$

since p_n/q_n is a best rational approximation. But then we have by our work previously that

$$\left|x - \frac{p}{q}\right| \geq \frac{1}{q} \cdot \frac{1}{q_n(a_{n+1} + x_{n+1} + \frac{q_{n-1}}{q_n})} \geq \frac{1}{q^2} \cdot \frac{q}{q_n} \cdot \frac{1}{a_{n+1} + 2} \geq \frac{1}{q^2} \cdot \frac{1}{(A+1)(A+2)}$$

and that completes the proof.

3. Chapter 47 and three-quarters: Another digression on periodic continued fractions

Let's start taking a look at square roots and in general quadratic irrationals. By that I mean solutions to $Ax^2 + Bx + C = 0$ with integers A, B, C and $B^2 - 4AC$ not a perfect square. These are also the set of numbers that can be written as $x = (a + \sqrt{b})/c$ with b not a perfect square.

We will also say a CF expansion is eventually periodic if there exists some k so that $a_n = a_{n+k}$ for all sufficiently large n. We will say that a CF expansion is purely periodic if there exists some k so that $a_n = a_{n+k}$ for $n \geq 0$. In particular, eventually periodic or purely periodic expansions are infinite.

Euler's Theorem If x is an irrational number with an eventually periodic continued fraction expansion, then x is a quadratic irrational.

Proof We know there are two distinct integers n, m such that the continued fraction expansion of x from n onwards is the same as the continued fraction expansion of x from m onwards. Thus $x_n = x_m$.

Now recall we have

$$x = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} x_n = \begin{pmatrix} p_{m-1} & p_m \\ q_{m-1} & q_m \end{pmatrix} x_m.$$

Using these last two inequalities together with the fact that $x_n = x_m$ gives

$$\frac{p_{n-1}x_n + p_n}{q_{n-1}x_n + q_n} = \frac{p_{m-1}x_n + p_m}{q_{m-1}x_n + q_m}.$$

Cross-multiplying and simplifying shows that x_n is the root of a non-trivial quadratic polynomial. Thus x_n is a quadratic irrational. Since

$$x = \frac{p_{n-1}x_n + p_n}{q_{n-1}x_n + q_n},$$

we can show that x is a quadratic irrational too by rationalizing the denominator.

Legendre's Theorem If x is a quadratic irrational, then it has an eventually periodic continued fraction expansion.

Proof (sketch) Suppose that $ax^2 + bx + c = 0$ with integer a, b, c and $b^2 - 4ac \neq 0$. Recall that

$$x = \frac{p_{n-1}x_n + p_n}{q_{n-1}x_n + q_n}.$$

If we plug this value of x into $ax^2 + bx + c = 0$, multiply through by $(q_{n-1}x_n + q_n)^2$, and simplify, we can get a formula that looks like $A_n x_n^2 + B_n x_n + C_n = 0$ with

$$\begin{aligned} A_n &= ap_n^2 + bp_nq_n + cq_n^2 \\ B_n &= 2ap_np_{n-1} + b(p_nq_{n-1} + p_{n-1}q_n) + 2cq_nq_{n-1} \\ C_n &= ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2 \end{aligned}$$

This clearly gives that $C_n = A_{n-1}$. Moreover, one can show that $B_n^2 - 4A_nC_n = b^2 - 4ac$.

The idea at this point is to show that the triplet (A_n, B_n, C_n) will eventually repeat and thus x_n, being the solution to the corresponding polynomial, will also repeat.

So make note of the fact that $|x - p_n/q_n| < 1/q_n^2$. Thus $p_n = xq_n + \delta_n/q_n$ where δ_n is some value in the interval $(-1, 1)$.

Thus

$$\begin{aligned} A_n &= a\left(xq_n + \frac{\delta_n}{q_n}\right)^2 + b\left(xq_n + \frac{\delta_n}{q_n}\right)q_n + cq_n^2 \\ &= (ax^2 + bx + c)q_n^2 + 2ax\delta_n + a\frac{\delta_n^2}{q_n^2} + b\delta_n. \end{aligned}$$

But $ax^2 + bx + c = 0$ so $|A_n| = |2ax\delta_n + a\delta_n^2/q_n^2 + b\delta_n| < 2|ax| + |a| + |b|$. Likewise, since $C_n = A_{n-1}$, $|C| < 2|ax| + |a| + |b|$. And since $B_n = \sqrt{b^2 - 4ac + 4A_nC_n}$, we can similarl see that there is a uniform bound for B_n as well. So the sequences (A_n, B_n, C_n) can only take finitely many values despite there being infinitely many n. By the pigeonhole principle, there exist n_1, n_2, n_3, distinct integers, so that

$$(A_{n_1}, B_{n_1}, C_{n_1}) = (A_{n_2}, B_{n_2}, C_{n_2}) = (A_{n_3}, B_{n_3}, C_{n_3}).$$

Therefore, we have that $x_{n_1}, x_{n_2}, x_{n_3}$ are all roots of the polynomial

$$A_{n_1}X^2 + B_{n_1}X + C_{n_1}.$$

This polynomial, however, only has two distinct roots, so we must have two of our iterates—without loss of generality, say x_{n_1} and x_{n_2}—are equal.

Now $x_{n_1} = [0; a_{n_1+1}, a_{n_1+2}, a_{n_1} + 3, \dots]$ and $x_{n_2} = [0; a_{n_2+1}, a_{n_2+2}, a_{n_2} + 3, \dots]$. Thus these being equal implies that the original number x is eventually periodic.

4. Chapter 48...ish

We will call a quadratic irrational x reduced if $x > 1$ and its conjugate x' is in $-1 < x' < 0$. We claim that x is purely periodic if and only if it is reduced. We'll only sketch why this is true though: First off if x is reduced then you can show that all of the quantities $1/x_i$ will also be reduced quadratic irrationals. Then you can show that if $1/x_n = 1/x_m$ for some $n \neq m$, you can use this reduced property to show that $1/x_{n-1} = 1/x_{m-1}$. In this way you can pull back through until you show that $x = 1/x_k$ for some k. And this shows that it is purely periodic.

So let's jump all the way back to Pell's equation. We want to know about $\sqrt{D}$ for some non-square D. Well, $\sqrt{D}$ is not reduced, as it's conjugate is $-\sqrt{D}$ which is less than -1. But on the other hand, $\sqrt{D} + \lfloor\sqrt{D}\rfloor$ IS reduced so must be purely periodic. Since it's zeroth digit is $2\lfloor\sqrt{D}\rfloor$, the last digit of the period must also be $2\lfloor\sqrt{D}\rfloor$. So the CF expansion of $\sqrt{D}$ looks like

$$\sqrt{D} = [a_0; \overline{a_1, a_2, \dots, a_n, 2a_0}].$$

So let's consider something here. Let $p_{n-1}/q_{n-1} = [a_0; a_1, a_2, \dots, a_{n-1}]$ and $p_n/q_n = [a_0; a_1, a_2, \dots, a_n]$. Now we have that

$$x_n = \cfrac{1}{2a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \dots}}} = \frac{1}{a_0 + \sqrt{D}},$$

and thus

$$\sqrt{D} = \frac{p_{n-1}x_n + p_n}{q_{n-1}x_n + q_n} = \frac{p_{n-1} + p_n(a_0 + \sqrt{D})}{q_{n-1} + q_n(a_0 + \sqrt{D})}.$$

Simplifying by cross multiplying (ignoring the middle part of the equality) we get

$$(a_0q_n + q_{n-1})\sqrt{D} + q_nD = p_n\sqrt{D} + (p_{n-1} + p_na_0).$$

Since $\sqrt{D}$ is irrational and everything else is an integer, we get that $a_0q_n + q_{n-1} = p_n$ and $q_nD = p_{n-1} + p_na_0$. Rewriting these we get $q_{n-1} = p_n - a_0q_n$ and $p_{n-1} = Dq_n - p_na_0$.

Now comes the part where we get to wave our magic wand...

Recall that $q_np_{n-1} - p_nq_{n-1} = (-1)^n$. Thus

$$q_n(Dq_n - p_na_0) - p_n(p_n - a_0q_n) = (-1)^n$$

$$p_n^2 - Dq_n^2 = (-1)^{n-1}.$$

This gives us exactly what we want if n is odd and if n is even then we square both sides and get

$$(p_n^2 + q_n^2D)^2 - D(2p_nq_n)^2 = 1.$$

This gives us the following theorem

Theorem 48.4 Write the continued fraction expansion of $\sqrt{D}$ as $\sqrt{D} = [a, \overline{b_1, b_2, b_3, \ldots, b_m}]$ and let $p/q = [a, b_1, b_2, \ldots, b_{m-1}]$. Then the *Smallest* solution in positive integers to Pell's equation $x^2 - Dy^2 = 1$ is given by

$$(x_1, y_1) = \begin{cases} (p, q) & \text{if m is even,} \\ (p^2 + q^2 D, 2pq) & \text{if m is odd.} \end{cases}$$

We've proven all of this but the smallest part (and that would take us a considerable while longer, so we won't).

Here's an example of the theorem in action. $\sqrt{21}$ has a periodic continued fraction equal to $[4; \overline{1, 1, 2, 1, 1, 8}]$. The fifth convergent is given by $55/12 = [4; 1, 1, 2, 1, 1]$. So $(55, 12)$ is the smallest solution to $x^2 - 21y^2 = 1$.

One curiosity here is why does the length of the period matter? This actually comes up in several other places. The reason is because the way we've defined the continued fraction we get matrices with determinant ± 1. If we had a different continued fraction expansion where all the matrices had determinant one, that is if we were in $SL_2(\mathbb{Z})$, we wouldn't have this problem. But this may cause us to lose an important property (such as the connection between convergents and best approximations) elsewhere.