Lecture 20: Continued Fractions IV: Applications

William Stein

Math 124 HARVARD UNIVERSITY **Fall 2001**

In this lecture we will learn about two applications of continued fractions. The first is a solution to the computational problem of recognizing a rational number using a computer. The second application is to the following ancient question: Given a positive nonsquare integer d, find *integers* x and y such that $x^2 - dy^2 = 1$.

1 Recognizing Rational Numbers

Suppose that you can compute approximations to a rational number using a computer, and desparately want to know what the rational number is. As Henri Cohen explains in his book *A Course in Computational Algebraic Number Theory*, continued fraction are very helpful.

> Consider the following apparently simple problem. Let $x \in \mathbb{R}$ be given by an approximation (for example a decimal or binary one). Decide if x is a rational number or not. Of course, this question as posed does not really make sense, since an approximation is usually itself a rational number. In practice however the question does make a lot of sense in many different contexts, and we can make it algorithmically more precise. For example, assume that one has an algorithm which allows us to compute x to as many decimal places as one likes (this is usually the case). Then, if one claims that x is (approximately) equal to a rational number p/q, this means that p/q should still be extremely close to x whatever the number of decimals asked for, p and q being fixed. This is still not completely rigorous, but it comes quite close to actual practice, so we will be content with this notion.
>
> Now how does one find p and q if x is indeed a rational number? The standard (and algorithmically excellent) answer is to compute the continued fraction expansion $[a_0, a_1, \ldots]$ of x. The number x is rational if and only if its continued fraction expansion is finite, i.e., if and only if one of the a_i is *infinite*. Since x is only given with the finite precision, x we be considered rational if x has a *very* large partial quotient a_i in its continued fraction expansion.

The following example illustrates Cohen's remarks:

Example 1.1.
```
? x
%13 = 9495/3847
? x*1.0
%14 = 2.4681570054587990642058747075643358461138549519105
? contfrac(x)
%15 = [2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 2]
? contfrac(2.468157005458799064)
%16 = [2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 328210621945, 2, 1, 1, 1, 1, 7]
? contfracpnqn([2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1])
%17 =
[9495 5852]
[3847 2371]
? contfrac(2.4681570054587990642058747075643)
%18 = [2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 1885009518355562936415046, 1, 4]
? \p300
? x*1.0  \\ notice that no repeat is immediately evident in the digits of x
%19 = 2.4681570054587990642058747075643358461138549519105796724720561476475 17..
? \\ in fact, the length of the period of the decimal expansion
  \\ of 1/3847 is 3846 (the order of 10 modulo 3847)!!
```

2 Pell's Equation

In February of 1657, Pierre Fermat issued the following challenge:

> Given a positive integer d, find a positive integer y such that dy^2+1 is a perfect square.

In other words, find a solution to $x^2 - dy^2 = 1$ with $y \in \mathbb{N}$.

Note Fermat's emphasis on *integer* solutions. It is easy to find rational solutions to the equation $x^2 - dy^2 = 1$. Simply divide the relation

$$(r^2+d)^2 - d(2r)^2 = (r^2-d)^2$$

by $(r^2-d)^2$ to arrive at

$$x = \frac{r^2+d}{r^2-d}, \qquad y = \frac{2r}{r^2-d}.$$

Fermat said: "Solutions in fractions, which can be given at once from the merest elements of arithmetic, do not satisfy me."

The equation $x^2 - dy^2 = 1$ is called **Pell's equation**. This is because Euler (in about 1759) accidently called it "Pell's equation" and the name stuck, though Pell (1611–1685) had nothing to do with it.

If d is a perfect square, $d = n^2$, then

$$(x+ny)(x-ny) = x^2 - dy^2 = 1$$

which implies that $x + ny = x - ny = 1$, so

$$x = \frac{x + ny + x - ny}{2} = \frac{1+1}{2} = 1.$$

We will thus always assume that d is not a perfect square. You can read about Pell's equation in Section 0.6 of Kato-Kurokawa-Saito and on pages 107–111 of Davenport. Pell's equation is best understood in terms of units in real quadratic fields.

3 Units in Real Quadratic Fields

Let d be a nonsquare positive integer, and set

$$\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} : a, b \in \mathbb{Q}\}$$
$$\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}.$$

Then $\mathbb{Q}(\sqrt{d})$ is a *real quadratic field* and $\mathbb{Z}[\sqrt{d}]$ is a ring. There is a homomorphism called norm:

$$N : \mathbb{Q}(\sqrt{d})^* \to \mathbb{Q}^*, \qquad N\left(a + b\sqrt{d}\right) = \left(a + b\sqrt{d}\right)\left(a - b\sqrt{d}\right) = a^2 - b^2 d.$$

Definition 3.1. An element $x \in R$ is a *unit* if there exists $y \in R$ such that $xy = 1$.

Proposition 3.2. *The units of $\mathbb{Z}[\sqrt{d}]$ are exactly the elements of norm ± 1 in $\mathbb{Z}[\sqrt{d}]$.*

Proof. Suppose $u \in \mathbb{Z}[\sqrt{d}]$ is a unit. Then

$$1 = N(1) = N(uu^{-1}) = N(u) \cdot N(u^{-1}).$$

Since $N(u), N(u^{-1}) \in \mathbb{Z}$, we have $N(u) = N(u^{-1}) = \pm 1$ □

Thus Fermat's challenge amounts to determing the group U^+ of units in $\mathbb{Z}[\sqrt{d}]$ of the form $a + b\sqrt{d}$ with $a, b \geq 0$.

Theorem 3.3. *The group U^+ is an infinite cyclic group. It is generated by $p_m + q_m\sqrt{d}$, where $\frac{p_m}{q_m}$ is one of the partial convergents of the continued fraction expansion of $\sqrt{d}$. (In fact, if m is the period of the continued fraction of $\sqrt{d}$ then $n = m - 1$ when m is even and $2n - 1$ when m is odd.)*

The theorem implies that *Pell's equation always has a solution*! Warning: the smallest solution is typically shockingly large. For example, the value of x in the smallest solution to $x^2 - 1000099y^2 = 1$ has **1118 digits**.

The following example illustrates how to use Theorem 3.3 to solve Pell's equation when $d = 61$, where the simplest solution is already quite large.

Example 3.4. Suppose $d = 61$. Then

$$\sqrt{d} = [7, \overline{1, 4, 3, 1, 2, 2, 1, 3, 4, 1, 14}],$$

which has odd period $n = 11$. Thus the group U^+ is generated by

$$x = p_{21} = 1766319049$$
$$y = q_{21} = 226153980.$$

That is, we have

$$U^+ = \langle u \rangle = \langle 1766319049 + 226153980\sqrt{61} \rangle,$$

and $x = 1766319049$, $y = 226153980$ gives a solution to $x^2 - dy^2 = 1$. All the other solutions arise from u^n for some n. For example,

$$u^2 = 6239765965720528801 + 798920165762330040\sqrt{61}$$

leads to another solution.

Remark 3.5. To help with your homework, note that if the equation

$$x^2 - dy^2 = n$$

has at least one (nonzero) solution $(x_0, y_0) \in \mathbb{Z} \times \mathbb{Z}$, then it must have infinitely many solutions. This is because if $x_0^2 - dy_0^2 = n$ and u is a generator of the cyclic group U^+, then for any integer i,

$$N(u^i(x_0 + y_0\sqrt{d})) = N(u^i) \cdot N(x_0 + y_0\sqrt{d}) = 1 \cdot n = n,$$

so

$$x_1 + y_1\sqrt{d} = u^i(x_0 + y_0\sqrt{d})$$

provides another solution to $x^2 + dy^2 = n$.

4 Some Proofs

The rest of this lecture is devoted to proving most of Theorem 3.3. We will prove that partial convergents to continued fractions contribute infinitely many solutions to Pell's equation. We will not prove that every solution to Pell's equation is a partial convergent, though this is true.[1]

Fix a positive nonsquare integer d.

Definition 4.1. A quadratic irrational $\alpha = a + b\sqrt{d}$ is *reduced* if $\alpha > 1$ and if the conjugate of α, denoted by α', satisfies $-1 < \alpha' < 0$.

For example, the number $\alpha = 1 + \sqrt{2}$ is reduced.

Definition 4.2. A continued fraction is *purely periodic* if it is of the form $[\overline{a_0, a_1, \ldots, a_n}]$.

The continued fraction $[\overline{2}]$ of $1 + \sqrt{2}$ is purely periodic.

[1] There is a complete proof in Section 13.5 of Burton's *Elementary Number Theory.* It just involves more of the same sort of computations that we've been doing with continued fractions.

Lemma 4.3. *If α is a reduced quadratic irrational, then the continued fraction expansion of α is purely periodic. (The converse is also true, and is easy to prove.)*

Proof. The proof can be found on pages 102–103 of Davenport's book. □

Lemma 4.4. *The continued fraction expansion of $\sqrt{d}$ is of the form*

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

Proof. Let a_0 be the floor of $\sqrt{d}$. Then $\alpha = \sqrt{d} + a_0$ is reduced because $\alpha > 1$ and $\alpha' = -\sqrt{d} + a_0$ satisfies $-1 < \alpha' < 0$. Let $[a_0, a_1, a_2, \ldots]$ be the continued fraction expansion of $\sqrt{d}$. Then the continued fraction expansion of $\sqrt{d}+a_0$ is $[2a_0, a_1, a_2, \ldots]$. By Lemma 4.3, the continued fraction expansion of $\sqrt{d} + a_0$ is purely periodic, so

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

where n is the period. It follows that $a_n = 2a_0$, as claimed. □

The following proposition shows that there are infinitely many solutions to Pell's equation that arise from continued fractions.

Proposition 4.5. *Let p_k/q_k be the partial convergents of the continued fraction expansion of $\sqrt{d}$, and let n be the period of the expansion of $\sqrt{d}$. Then*

$$p_{kn-1}^2 - dq_{kn-1}^2 = (-1)^{kn}$$

for $k = 1, 2, 3, \ldots$.

Proof. [2] By Lemma 4.4, for $k \geq 1$, the continued fraction of $\sqrt{d}$ can be written in the form

$$\sqrt{d} = [a_0, a_1, a_2, \ldots, a_{kn-1}, r_{kn}]$$

where

$$r_{kn} = [2a_0, \overline{a_1, a_2, \ldots, a_n}] = a_0 + \sqrt{d}.$$

Because $\sqrt{d}$ is the last partial convergent of the continued fraction above, we have

$$\sqrt{d} = \frac{r_{kn}p_{kn-1} + p_{kn-2}}{r_{kn}q_{kn-1} + q_{kn-2}}.$$

Upon substituting $r_{kn} = a_0 + \sqrt{d}$ and simplifying, this reduces to

$$\sqrt{d}(a_0 a_{kn-1} + q_{kn-2} - p_{kn-1}) = a_0 p_{kn-1} + p_{kn-2} - dq_{kn-1}.$$

Because the right-hand side is rational and $\sqrt{d}$ is irrational,

$$a_0 a_{kn-1} + q_{kn-2} = p_{kn-1}, \quad \text{and} \quad a_0 p_{kn-1} + p_{kn-2} = dq_{kn-1}.$$

Multiplying the first of these equations by p_{kn-1} and the second by $-q_{kn-1}$, and then adding them, gives

$$p_{kn-1}^2 - dq_{kn-1}^2 = p_{kn-1}q_{kn-2} - q_{kn-1}p_{kn-2}.$$

But

$$p_{kn-1}q_{kn-2} - q_{kn-1}p_{kn-2} = (-1)^{kn-2} = (-1)^{kn},$$

which proves the proposition. □

[2] This proof is from Section 13.5 of Burton's *Elementary Number Theory*.