Pell's Equations

Nikhil Reddy

Given a positive integer n that isn't a perfect square, do there exist solutions to the equation

$$x^2 - ny^2 = 1,$$

and if so, how many? You've probably seen this equation before, and it even has a name: the Pell Equation. In this paper, we explore the Pell Equation and its solutions.

Proving Infinitude of Solutions

First suppose we have a solution (a, b). Is there a way we could generate more solutions from this one? We can, and to do so we use the following identity:

Brahmagupta's Identity:

$$(x^2 - ny^2)(a^2 - nb^2) = (xa + nyb)^2 - n(xb + ya)^2.$$

Proof: We have

$$(x^2 - ny^2)(a^2 - nb^2) = (x + y\sqrt{n})(x - y\sqrt{n})(a + b\sqrt{n})(a - b\sqrt{n}) =$$

$$(xa+nyb+\sqrt{n}(xb+ya))(xa+nyb-\sqrt{n}(xb+ya)) = (xa+nyb)^2-n(xb+ya)^2 \blacksquare$$

Using this identity, we can see that a single solution to the Pell Equation will yield infinitely many solutions. Thus, we need to find which n produce equations that have solutions. In fact, as we will see, all Pell Equation's have solutions.

From now on, we will use $x^2 - dy^2$ for the Pell Equation. To proceed further, we need the following result from Dirichlet:

Lemma 1: Let α be an irrational number. Then the inequality

$$|p - q\alpha| < \frac{1}{q}$$

has infinitely many integer solutions (p, q).

Proof: We won't prove it here, but the idea is to consider the fractional part of α multiplied by $1, \ldots, q-1$, finding the intervals they lie in, and using pigeonhole to finish.

Using this, we first show that there exists some n such that $|n| < 1 + 2\sqrt{d}$ and $x^2 - dy^2 = n$ has infinitely many solutions. We have that

$$x = x - y\sqrt{d} + y\sqrt{d} \leq \left|x - y\sqrt{d}\right| + y\sqrt{d} < \frac{1}{y} + y\sqrt{d} \leq 1 + y\sqrt{d},$$

where the first strict inequality comes from using Dirichlet's result with $\alpha = \sqrt{d}$. Then we have

$$\left|x^2 - dy^2\right| = (x + y\sqrt{d})\left|x - y\sqrt{d}\right| < (1 + y\sqrt{d} + y\sqrt{d})\frac{1}{y} = \frac{1}{y} + 2\sqrt{d} \leq 1 + 2\sqrt{d}.$$

We know there are infinitely many pairs (x, y) for which this inequality is true because of Dirichlet's result, so by the pigeonhole principle, there is some n such that $|n| < 1 + 2\sqrt{d}$ and $x^2 - dy^2 = n$ has infinitely many solutions. ■

Now we reduce the Pell Equation mod $|n|$. Since there infinitely many solutions, by pigeonhole there must be two distinct solutions (x_1, y_1) and (x_2, y_2) that are the same mod n. Write $x_1 = x_2 + ni$ and $y_1 = y_2 + nj$. Then we have

$$\begin{aligned} x_1 + y_1\sqrt{d} &= (x_2 + y_2\sqrt{d}) + n(i + j\sqrt{d}) \\ x_1 - y_1\sqrt{d} &= (x_2 - y_2\sqrt{d}) + n(i - j\sqrt{d}). \end{aligned}$$

Substitute in $n = x_2^2 - dy_2^2 = (x_2 + y_2\sqrt{d})(x_2 - y_2\sqrt{d})$ and factor to get

$$\begin{aligned}x_1 + y_1\sqrt{d} &= (x_2 + y_2\sqrt{d})(1 + (x_2 - y_2\sqrt{d})(i + j\sqrt{d}))\\ x_1 - y_1\sqrt{d} &= (x_2 - y_2\sqrt{d})(1 + (x_2 + y_2\sqrt{d})(i - j\sqrt{d})).\end{aligned}$$

Write $x + y\sqrt{d} = 1 + (x_2 - y_2\sqrt{d})(i + j\sqrt{d})$. Then we have

$$\begin{aligned}x_1 + y_1\sqrt{d} &= (x_2 + y_2\sqrt{d})(x + y\sqrt{d})\\ x_1 - y_1\sqrt{d} &= (x_2 - y_2\sqrt{d})(x - y\sqrt{d}).\end{aligned}$$

Multiplying the equation together yields $n = n(x^2 - dy^2)$, so $x^2 - dy^2 = 1$. We just need to show that (x, y) isn't the trivial solution (1, 0). Suppose it was. That would imply $x_1 = x_2$, but that's a contradiction, so we're done.

Finding All Solutions

So now we've shown that every Pell Equation has a nontrivial solution, but can we find all possible solutions? We can, and the way to generate them is surprisingly simple.

Suppose we have two solutions to a Pell equation, (x, y) and (a, b). Note that

$$\begin{aligned}(x + y\sqrt{d})(a + b\sqrt{d}) &= (xa + ybd) + (xb + ya)\sqrt{d}\\ (x - y\sqrt{d})(a - b\sqrt{d}) &= (xa + ybd) - (xb + ya)\sqrt{d}.\end{aligned}$$

Multiplying the two equations together yields

$$(xa + ybd)^2 - d(xb + ya)^2 = 1,$$

so the set of solutions to Pell equations is closed under multiplication (note this is just Brahmagupta's Identity in disguise). This means that given a solution, we can generate infinitely many more. We just take that solution, (x, y), and to generate another, we take the coefficients of

$$(x + y\sqrt{d})^k$$

for some positive integer k.

In fact, we can show something even stronger:

Theorem: Suppose we have a solution (x_1, y_1) *to a Pell equation, where* x_1 *and* y_1 *are positive integers and* y_1 *is minimal. Then, all positive integer solutions are generated by the coefficients of*

$$(x_1 + y_1\sqrt{d})^k,$$

where k *is a nonnegative integer.*

We need a few lemmas to prove this.

Lemma 2: Suppose $x^2 - dy^2 = 1$ *and* $x + y\sqrt{d} > 1$*. Then* $x \geq 2$ *and* $y \geq 1$*.*

Proof: Note that

$$x + y\sqrt{d} > 1 > x - y\sqrt{d} > 0.$$

Thus $2y\sqrt{d} > 0 \implies y > 0$, and since y is an integer, $y \geq 1$. Then we have $x > y\sqrt{d} \geq \sqrt{d} > 1$, so $x \geq 2$. ■

Lemma 3: If $x^2 - dy^2 = 1$ *and* $a^2 - db^2 = 1$*, where* $a, b, x, y \geq 0$*, then* $a + b\sqrt{d} < x + y\sqrt{d}$ *if and only if* $a < x$ *and* $b < y$*.*

Proof: The if direction is obvious, so suppose $a + b\sqrt{d} < x + y\sqrt{d}$. Reciprocating yields $x - y\sqrt{d} < a - b\sqrt{d}$, and adding these two yields

$$(a + x) + (b - y)\sqrt{d} < (a + x) + (y - b)\sqrt{d}.$$

Subtracting $a + x$, dividing by $\sqrt{d}$, and rearranging yields

$$2b < 2y \implies b < y.$$

Then we have $a^2 = 1 + db^2 < 1 + dy^2 = x^2$, so $a < x$. ■

Now we're ready to prove that all solutions are generated by the minimal solution.

Proof: We already know that any solution generated by $(x_1 + y_1\sqrt{d})^k$ works, so suppose that we have a solution (x, y). We show that $x + y\sqrt{d} = (x_1 + y_1\sqrt{d})^k$ for some k.

Since $x + y\sqrt{d} > 1$, and since $(x_1 + y_1\sqrt{d})^k$ is increasing, we have

$$(x_1 + y_1\sqrt{d})^k \leq x + y\sqrt{d} \leq (x_1 + y_1\sqrt{d})^{k+1}$$

for some k. Dividing through by $(x_1 + y_1\sqrt{d})^k$ yields

$$1 \leq (x + y\sqrt{d})(x_1 + y_1\sqrt{d})^{-k} \leq x_1 + y_1\sqrt{d}.$$

Note that $x_1 + y_1\sqrt{d} = \frac{1}{x_1 - y_1\sqrt{d}} \implies (x_1 + y_1\sqrt{d})^{-k} = (x_1 - y_1\sqrt{d})^k$. Note that $(x_1 - y_1\sqrt{d})^k$ is also a generator of solutions to the Pell equation, since the y part of the solution is just negative. Thus, using Brahmagupta's Identity, we know the middle of the inequality has coefficients which are another solution, so we can write it as

$$1 \leq a + b\sqrt{d} \leq x_1 + y_1\sqrt{d}.$$

Note by the two lemmas we proved that a and b are positive and $b < y_1$, contradicting minimality. Thus, all solutions are generated by $(x_1 + y_1\sqrt{d})^k$. Note that the x coefficients in $(x_1 + y_1\sqrt{d})^k$ are also increasing, so we can substitute minimal y for minimal x if we want to. ■

Showing Existence of Solutions

So now we know that Pell equations have infinitely many solutions, and we even know how to generate all of them. The only issue is finding the minimal solution. How are we supposed to do that. We could try increasing x by 1 and seeing if there is a valid solution for y. However, this isn't always feasible. For example, $d = 61$ has minimal solution $(1766319049, 226153980)$. Instead, we have another method for finding a minimal solution.

Theorem: If (x, y) satisfies $x^2 - dy^2 = n$, with $|n| < \sqrt{d}$, then $\frac{x}{y}$ is a convergent of the continued fraction expansion of $\sqrt{d}$.

Note we are proving a more generalized version, where we replaced 1 with n. Also, this theorem should make sense heuristically, since for a solution to a Pell equation, we can write

$$d = \frac{x^2 - n}{y^2} \implies \sqrt{d} \approx \frac{x}{y}.$$

Proof: A well known theorem about continued fractions is that for a real number α, if x and y are integers with y nonzero and $\left|\frac{x}{y} - \alpha\right| < \frac{1}{2y^2}$, then $\frac{x}{y} = \frac{p}{q}$ for some convergent $\frac{p}{q}$ of α. Taking $\alpha = \sqrt{d}$, assuming the hypothesis of the theorem we have

$$\left|\frac{x}{y} - \sqrt{d}\right| = \left|\frac{x - y\sqrt{d}}{y}\right| = \left|\frac{n}{y(x + y\sqrt{d})}\right| < \frac{\sqrt{d}}{y^2\left(\frac{x}{y} + \sqrt{d}\right)} = \frac{1}{y^2\left(\frac{x}{y\sqrt{d}} + 1\right)}.$$

If this is less than $\frac{1}{2y^2}$, then we're done. Thus, we have

$$y^2\left(\frac{x}{y\sqrt{d}} + 1\right) > 2y^2 \implies x > y\sqrt{d} \implies x^2 - dy^2 > 0,$$

which is evidently true for positive n. For negative, the argument breaks, but we can instead look at $\frac{y}{x}$ as an approximation of $\frac{1}{\sqrt{d}}$. This yields

$$\left|\frac{y}{x} - \frac{1}{\sqrt{d}}\right| < \frac{1}{x^2\left(\frac{y\sqrt{d}}{x} + 1\right)}.$$

It can easily be shown that this is less than $\frac{1}{2x^2}$, so we're done. ■

How does this make finding the minimal solution easier? After all, for something like $d = 61$, it seems like it'd take a long time to reach the correct convergent. However, we actually have a finite search space. Although we won't prove it here, the continued fraction expansion for $\sqrt{d}$ for non square d is eventually periodic, and $x^2 - dy^2$ for that periodic section is also periodic. Thus, one only needs to test the beginning section and the first periodic section of the continued fraction expansion of $\sqrt{d}$ to find the minimal solution.