Pell's Equation
History, Algorithms, and Open Problems

JAMES MARSHALL REBER
December 1, 2017

Abstract

In this talk, we discuss finding integral solutions to the Pell equation, which is a Diophantine equation of the form $x^2 - Dy^2 = 1$ for integral solutions $x, y \in \mathbb{Z}$ and $D > 0$ a square-free integer. In particular, we focus on finding fundamental solutions for the Pell equation, which in turn will give us all solutions for the Pell equation.

History

Diophantine equations were of great interest to the Greeks. In particular, Archimedes (287 - 212 BC) had a fascination with finding solutions to a particular Diophantine equation, which we now know is a particular Pell equation. The problem has to do with calculating the number of cattle of particular colors in a herd which belonged to the sun God, the poem itself which outlines this problem is 40 lines. Letting W, B, D, Y denote white, black, dappled, and yellow cows, and w, b, d, y denote their respective colors but for cows. Then for example, one of these restriction was

$$W = \left(\frac{1}{2} + \frac{1}{3}\right) b + y.$$

This was not the hard part; one could solve this system of restrictions in order to find a particular number of cows and cattle using linear algebra. The real kicker of this problem was what followed, and in particular Archimedes said that if one could solve this, then one could consider themselves "... most skilled in numbers." [4] Archimedes then said that the number of white bulls added to the number of black pulls must be a square number, and the nummber of dappled

cows added to the number of yellow bulls must be a triangular number. This could eventually be reformatted into

$$u^2 - (606)(7766)^2 v^2 = 1$$

where we have $D + y = v(v+1)/2$ and $u = 2v + 1$. Many tried to solve this problem, using a variety of different techniques, but what was really required in order to solve this was an efficient algorithm to find a Fundamental solution to the Pell equation; that is, the smallest integer tuple (u, v) so that they satisfy the equation. This was not an easy task, and it wasn't until Amthor in 1880 discovered the way to find it that the problem was finally satisfied (Amthor established that the number would be extremely large and thus nearly impossible to calculate by hand; it wasn't until the advent of computers that we actually had the exact answer).

The key issues with the Pell equation is two fold; do solutions exist to the equation, and if so how do we find these solutions? The existence issue wouldn't be realized for long after Archimedes, but finding a solutions was of much interest to Indian mathematicians in 628 B.C.. Brahmagupta (598 - 670 AD) discovered the identity of composition of forms (referred to at the time as samsa). Using this, Brahmagupta found an ad hoc way to discover fundamental solutions to the Pell equation. Later, Bhaskara II found a more cyclic algorithm to find solutions to solve the Pell equation, which resembles the modern PQA algorithm. [5]

The history of the Pell equation fell silent until Fermat posed the question,

"Given any number whatever that is not a square, there are also given an infinite number of squares such that, if the square is multiplied into the given number and unity is added to the product, the result is a square."

Brounecker (1620-1684) found an algorithmic solution method to this problem, but Euler would later mistakenly attribute this to Pell (hence the name). Thus, we see that Brounecker, Euler, Brahmagupta, and Brahska II all found similar methodologies to finding these fundamental solutions, but none of them were concerned about whether or not this would always succeed in finding a solution for all $D > 0$ which are not square. This was later rectified by Lagrange, although Weil claims that Fermat had an unpublished solution. [1]

Using Algebraic Number Theory, the algorithms built by Euler, Lagrange, et al. were improved upon and made more efficient, and it evolved into what we now refer to as the PQA algorithm (or sometimes as the continued fraction method). We now explore the specifics of what the Pell equation entails.

Existence and Finding a Fundamental Solution.

Lagrange was able to notice two main facts about solutions to the Pell equation that allowed him to generate a good algorithm for finding all solutions. First, he

noticed that the set of solutions is a cyclic group generated by something called a fundamental unit, and second he noticed that this fundamental unit is related to the continued fraction expansion of the quadratic surd; or, in other words, if the equation is of the form $x^2 - dy^2 = 1$, then the fundamental solution is related to to the continued fraction expansion of $\sqrt{d}$.

To start, we define the norm mapping, which we'll be using throughout.

Definition 1. If K is an extension of a field F, and K/F is separable, then for any $\alpha \in K$ we have

$$N_{K/F}(\alpha) = \prod_{i=1}^{n} \sigma_i(\alpha),$$

where the σ_i are all of the isomorphisms of K into the algebraic closure $\bar{F}$ of F fixing the elements of F.

Notice that if $d > 0$ is a square-free integer, we have that $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$ is a Galois extension. Moreover, $\mathrm{Gal}(\mathbb{Q}(\sqrt{d})/\mathbb{Q}) = \{\mathrm{Id}, \sigma\}$ where $\sigma : \sqrt{d} \mapsto -\sqrt{d}$; i.e. the conjugate mapping. We can then simplify our definition of the norm mapping to our more specific case.

Definition 2. If $d > 0$ is a square-free integer, $K = \mathbb{Q}(\sqrt{d})$, and $F = \mathbb{Q}$, then we define the norm mapping to be

$$N_{K/F}(\alpha) = \alpha\bar{\alpha}.$$

There is a special feature of the norm that allows us to determine the units within our field. We need a special feature of the norm mapping in order to establish this, though.

Remark. *Throughout, we'll be using N to denote the norm mapping instead of $N_{K/F}$.*

Lemma 1. *The norm mapping is multiplicative; that is, for $a, b \in K$ defined as prior, we have $N(ab) = N(a)N(b)$.*

Proof. This readily follows by the multiplicativity of conjugates. Notice $N(ab) = ab\bar{a}\bar{b} = ab\bar{a}\bar{b} = a\bar{a}b\bar{b} = N(a)N(b)$. ■

Using this, we can determine the units using the norm mapping.

Corollary 1. *We have that $a \in K$ is a unit if and only if $N(a) = \pm 1$.*

Now, we go back to the Pell equation. Recall that it is an equation of the form $x^2 - dy^2 = 1$. We can factor the left hand side in the order $\mathbb{Z}[\sqrt{d}]$ to be $(x - \sqrt{d}y)(x + \sqrt{d}y) = 1$. Let $\alpha = x + \sqrt{d}y$. Then, using the norm, we can rewrite this as $N(\alpha) = 1$. Thus, we can see that the norm is related to the units of the order $\mathbb{Z}[\sqrt{d}]$. Define $G = \{\alpha \in K : N(\alpha) = 1\}$. We'll observe that this is an abelian group.

Lemma 2. *G as defined prior is an Abelian group.*

Proof. We need to go through the five requirements for an Abelian group. It's closed; that is, if $a, b \in G$ we have $ab \in G$, since $N(a) = N(b) = 1$ and so $N(a)N(b) = N(ab) = 1$. Next, we notice that there is an identity; we clearly have $N(1) = 1$. Inverses follows since $a\bar{a} = N(a) = 1$, so dividing both sides gives us $\bar{a} = 1/a$. We then see that $\bar{a} \in \mathbb{Z}[\sqrt{d}]$ and $N(\bar{a}) = 1$. Associativity follows, since $N(a)(N(b)N(c)) = (N(a)N(b))N(c)$ clearly. Finally, commutativity follows since $N(ab) = N(a)N(b) = N(b)N(a) = N(ba)$. ■

So we see that the set of solutions for the Pell equation forms a group. How do we know that this group is nontrivial though? Moreover, can we show that this group is cyclic and therefore generated by a cyclic element? If it is a nontrivial group, what is the size of the group (or, informally, how many equations are there)? Finally, if the group is cyclic, how do we find the generator? These questions (suggestive as they may be) did not follow in this order throughout history; as discussed, the second and third question were answered before Lagrange was able to really answer the first. We will go through them as stated.

Lemma 3. *The equation $x^2 - dy^2 = 1$, where $d > 1$ and square-free, has at least one solution where $y \neq 0$.*

Proof. To prove this, we will need a lemma that we'll state without proof.

Lemma 4. *There are infinitely many solutions of the equation $x^2 - dy^2 = k$, $d, k \in \mathbb{Z}$ and d square-free, where $x, y \in \mathbb{Z}_{>0}$ for some k with $|k| < 1 + 2\sqrt{d}$.*

We utilize the proof by LeVeque [7]. Using this lemma, there is an integer for which one of the two equations $N(a) = \pm k$ has infinitely many solutions a in $\mathbb{Z}[\sqrt{d}]$. Since there are only finitely many residue classes mod k in $\mathbb{Z}[\sqrt{d}]$, some residue classes must contain at least three of these solutions. Let $N(a_1) = N(a_2) = \pm k$ and $a_1 \equiv a_2 \pmod{k}$, but that $a_1 \neq \pm a_2$. Then $a_1\bar{a}_2 = a_2\bar{a}_2 \equiv 0 \pmod{k}$, so that $b = a_1\bar{a}_2/k$ is an element of $\mathbb{Z}[\sqrt{d}]$; that is, it has integral components. Since

$$N(b) = b\bar{b} = \frac{a_1\bar{a}_2 \cdot \bar{a}_1 a_2}{k^2} = \frac{N(a_1)N(a_2)}{k^2} = 1,$$

b yields a solution of the equation. If the second component of b were 0, then $N(b) = 1$ would imply that $b = \pm 1$, whence

$$a_1\bar{a}_2 = \pm k = \pm a_1\bar{a}_1$$

$$\bar{a}_2 = \pm \bar{a}_1$$

$$a_2 = \pm a_1,$$

contrary to hypothesis. ■

With the existence of at least one solution, we are led to the next question – are these solutions generated by a single element? It turns out yes, and it's generated by what we call the fundamental solution; that is, the smallest such solution.

Lemma 5. *Let $d > 0$ be a square-free integer, and assume that there is a solution to $x^2 - dy^2 = 1$. Then there exists a unique $f \in G$ so that $f > 1$ and for all $a \in G$, we have $a = \pm f^n$ for some $n \in \mathbb{Z}$.*

Proof. This follows the proof stated in LeVeque [7]. Let $a \in G$. We notice that $a, 1/a, -a$ and $-1/a$ give four solutions to Pell's equation, differing only by signs of x and y. So we need to show that every $a > 1$ with $N(a) = 1$ is of the form $a = f^n$ where f is the fundamental solution; i.e. the smallest such solution.

Since $a > 1$ and we chose f to be minimal, we have $a \geq f$. Hence, there is a positive integer n such that $f^n \leq a < f^{n+1}$. Now notice that $a/f^n = a\bar{f}^n$ is in $\mathbb{Z}[\sqrt{d}]$, and $N(a/f^n) = 1$. Thus, $a/f^n = b$ gives an integral solution of Pell's equation. From the definition of n, we have $1 \leq b < f$ and by how we defined f we cannot have $1 < b < f$. Hence, $b = 1$ and $a = f^n$, as we required. ∎

Remark. *This result should seem very similar to Dirichlet's Unit theorem. This is because it is, in fact, a special case of it. However, this was proven before Dirichlet's Unit theorem, and so we provide the proof without using this result.*

So we have that the group has two generators, -1 and f where f is the fundamental solution, and f is of infinite order, that is, there are infinitely many solutions. We now have the final question; how do we find such a solution? This is where the method of continued fractions comes into play. In order to discuss the method of continued fractions, we must make some preliminary definitions.

Definition 3. Given any irrational number x, define the sequence of irrational numbers recursively by

$$x_0 = x, \quad x_{k+1} = \frac{1}{x_k - \lfloor x_k \rfloor} \text{ for } k = 0, 1, 2, \ldots$$

in terms of the floor function. The continued fraction of x is the infinite nested fraction

$$a_0 + \frac{1}{a_1 + 1\frac{1}{a_2 + \cdots}} \text{ where } a_k = \lfloor x_k \rfloor \text{ is an integer.}$$

We really only want to deal with finitely many terms, so this leads us to another definition.

Definition 4. Denote the nth convergent as the quantity which is obtained by the first n terms of the continued fraction

$$\{a_0; a_1, a_2, \ldots, a_{n-1}\} = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\cdots + \frac{1}{a_{n-1}}}}}.$$

Lagrange was able to use these continued fractions to establish his theorem.

Theorem 1. *(Lagrange) Fix $d > 0$ square-free.*
(a) The continued fraction of $\sqrt{d}$ is in the form

$$\sqrt{d} = \{a_0; \overline{a_1, a_2, \ldots, a_{h-1}, 2a_0}\}$$

where the bar means that the sequence repeats indefinitely.
(b) Write the hth convergent of the continued fraction above as the rational number

$$\{a_0; a_1, a_2, \ldots, a_{h-1}\} = \frac{u_h}{v_h}.$$

Then u_h and v_h are positive integers which satisfy the relation $u_h^2 - dv_h^2 = (-1)^h$.

Moreover, using this we get how to find the fundamental solution.

Corollary 2. *Let h denote the smallest period of the continued fraction sequence. Then we have that the fundamental solution f is*

$$f = \begin{cases} u_h + v_h\sqrt{d}, \text{ if } h \equiv 0 \pmod 2 \\ u_{2h} + v_{2h}\sqrt{d} = (u_h + v_h\sqrt{d})^2, \text{ if } h \equiv 1 \pmod 2. \end{cases}$$

Instead of proving these results, we'll show an example of it in action.

Example 1. Let $d = 7$. Then we have the sequence $\{2; 1, 1, 1, 4, 1, 1, 1, 4, \ldots\}$. Moreover, we see $h = 4$ (cutting it off before 4), and so we calculate $2 + \frac{1}{1+\frac{1}{1+1}} = \frac{8}{3}$; or, in other words, $u_h = 8$, $v_h = 3$, and our fundamental solution is $8 + 3\sqrt{7}$.

Final Thoughts and Open Problems

From prior, we saw that in the case $x^2 - dy^2 = 1$, we have infinitely many solutions and we have a "good" algorithm in order to find all of these solutions. What about the equation $x^2 - dy^2 = k$ for arbitrary $k \in \mathbb{Z}$? It turns out that Lagrange's method gives us a good algorithm for also finding out if there are solutions in the case $k = -1$, and if there are solutions Lagrange's theorem also gives us a way to find all of the solutions. This case is dubbed the negative Pell equation. In fact, Lagrange's continued fraction method can be used to generate all solutions to the Pell equation for arbitrary k, although it is not always guaranteed to have a solution (such as in the case where $k = -1$).

Thus, Pell's equation shows us that we have a "good" algorithm for finding a generator for the units of the order $\mathcal{O} = \mathbb{Z}[\sqrt{d}] \subseteq K = \mathbb{Q}(\sqrt{d})$. One then wonders is there always a good algorithm for finding the generator for the units of any arbitrary order? It turns out that currently there is no such algorithm. It is, in fact, the "holy grail" of algebraic number theory to find such an algorithm.

There is still yet more to the Pell equation. This short presentation has really only skimmed the topic as a whole. For more details on Pell's equation, I recommend reading *"Fundamentals of Number Theory"* [7] and *"Solving the Pell equation"* [5]. LeVeque gives a good introduction to the topic, while Jacobson and Williams really go into depth on the topic as a whole.

References

1. Hendrick W. Lenstra, Jr. *"Solving the Pell equation"* MSRI Publications, 44, 2008.

2. Jarrod A. Cunningham, Nancy Ho, Karen Lostritto, Jon A. Middleton, Nikia T. Thomas. *"On Large Rational Solutions of Cubic Thue Equations: What Thue Did to Pell"* Rose-Hulman Undergraduate Mathematics Journal: Vol. 7: Iss. 2, Article 6, 2006.

3. Keith Conrad. *"Dirichlet's Unit Theorem"* Publisher: Author.

4. Mansfield Merriman. *"The Cattle Problem of Archimedes"* Popular Science Monthly, 67: 660-665, 1905.

5. Michael J. Williams and Hugh C. Williams. *"Solving the Pell Equation"* Springer, 2009.

6. Seung Hyun Yang. *"Continued Fractions and Pell's Equation"* University of Chicago REU Papers, 2008.

7. William J. LeVeque. *Fundamentals of Number Theory*. Addison-Wesley Publishing Co., Reading, Mass-London-Amsterdam, 1977.