Pell's method

Stefano Scacco[a]

[a] *Università degli Studi di Roma La Sapienza and INFN Section of Roma 1, Piazzale Aldo Moro 5, 00185 Roma, Italy*

Motivation

These notes were born from a simple necessity: consolidating scattered theorems and results regarding quadratic Diophantine equations into one cohesive place. It is a brilliant subject, and what follows is just a brief, humble glimpse into a much larger mathematical world. I have intentionally stripped away all citations and references. These notes rely entirely on their own self-consistency, padded only by a few introductory historical facts you can easily search yourself. As long as the logical steps are sound and the proofs are complete, no bibliography is required.
Math is self-evident.

Contents

1 Introduction

First, let's introduce the following Diophantine equation

$$x^2 - Ay^2 = 1 \tag{1.1}$$

where the unknowns are x, y, which are positive integers. We also set A to be integer. If $A = B^2$ is a perfect square, the equation becomes trivial, because it can be factorized

$$x^2 - Ay^2 = (x - By)(x + By) = 1 \tag{1.2}$$

which requires, individually, $x - By = 1$ and $x + By = 1$, which can only be solved if $y = 0$. In general, the equation

$$x^2 - Ay^2 = N \tag{1.3}$$

is always trivially solved when $A = B^2$, by factorizing N. Take all the possible two-factor expressions $N = n_1 n_2$, and impose

$$\begin{cases} x - By = n_1 \\ x + By = n_2 \end{cases} \quad \text{or} \quad \begin{cases} x - By = n_2 \\ x + By = n_1 \end{cases} \tag{1.4}$$

depending on the sign of B, and repeat for all possible factors of n_1. Because the number of factors is finite, the number of solutions of this trivial case will be finite.

Now let us look at the following Diophantine equation

$$x^2 - Ay^2 = 1 \tag{1.5}$$

where this time A is not a perfect square. This means that $\sqrt{A}$ is an irrational number. This is called *Pell's equation*.

This Diophantine equation is one of the most famous equations in the history of mathematics. Just to highlight its importance, the case $A = 2$ was studied in India and Greece as early as 400 BC: the Pythagoreans found the solutions to be strictly related to the irrationality of $\sqrt{2}$, as x/y tend to be good approximants of it. This was known also to the Indian mathematician Baudhayana, who lived around the same time.

Later, the case $A = 3$ was allegedly studied by Archimedes. Moreover, his famous cattle problem can be studied with a generalized version of Pell's equation. In 250 BC, Diophantus presented solutions for a few cases of the generalized version. During the European Middle Age, the Indians developed the first method to always obtain solutions to this equation, called the *chakravala method*.

Fermat studied and solved this equation. Euler, reading a revised translation by John Pell of a manuscript describing the solution of the equation, misattributed the solution to Pell himself, and so he named the equation after him. Finally, the general theory of Pell's equation was developed by Lagrange in the 18th century.

2 Solution to Pell's equation

Having discussed its history, let us now approach the solution to this equation. Define the set of numbers

$$\mathbb{Q}(A) = \{p + q\sqrt{A} | p, q \in \mathbb{Q}\} \tag{2.1}$$

Closed under the operation of summation

$$(p + q\sqrt{A}) + (r + s\sqrt{A}) = (p + r) + (q + s)\sqrt{A} \tag{2.2}$$

and multiplication

$$(p + q\sqrt{A})(r + s\sqrt{A}) = (pr + qsA) + (ps + qr)\sqrt{A} \tag{2.3}$$

clearly, $\mathbb{Q}(A)$ is a *ring*. The inverse under addition is always contained. Moreover, we could define the inverse under multiplication

$$(p + q\sqrt{A})^{-1} = \frac{p - q\sqrt{A}}{p^2 - q^2 A} \tag{2.4}$$

which, of course, does not work for the number zero. The existence of the inverse under multiplication implies $\mathbb{Q}(A)$ not only a ring, but a *field*. Notice that if $p^2 - q^2 A = 0$, then either $p = q = 0$, or $q \neq 0$ and then $A = p^2/q^2$, which means $\sqrt{A} = p/q$, contradicting the fact that it must be irrational. So, $p^2 - q^2 A = 0$ iff $p = q = 0$, hence can be thought of as a norm: for any $u = p + q\sqrt{A}$, define its norm on the field $\mathbb{Q}(A)$ as $N(u) = p^2 - q^2 A$.

Introduce the set

$$\mathbb{Z}(A) = \{z_1 + z_2\sqrt{A} | z_1, z_2 \in \mathbb{Z}\} \tag{2.5}$$

which is a ring, but not a field, because we lose the property of the inverse. However, some numbers do still have an inverse in the ring. In particular, those elements $u = p + q\sqrt{A}$ so that $N(u) = 1$: for them we can define

$$\bar{u} := u^{-1} = p - q\sqrt{A}, \qquad \rightarrow \qquad u\bar{u} = N(u) = 1 \tag{2.6}$$

Observe that, taking $v = r + s\sqrt{A}$

$$\bar{u}\bar{v} = (p - q\sqrt{A})(r - s\sqrt{A}) = (pr + qsA) - (ps + qr)\sqrt{A} = \overline{uv} \tag{2.7}$$

Which allows us to prove the crucial property

$$\begin{aligned} N(uv) &= (pr + qsA)^2 - (ps + qr)^2 A = \\ &= p^2r^2 + q^2s^2A^2 - p^2s^2 - q^2r^2 = \\ &= \left(p^2 - q^2A\right)\left(r^2 - s^2A\right) = \\ &= N(u)\,N(v) \end{aligned} \tag{2.8}$$

an equivalent, one line proof of the previous statement is

$$N(uv) = uv\overline{uv} = uv\bar{u}\bar{v} = u\bar{u}v\bar{v} = N(u)\,N(v) \tag{2.9}$$

Now, observe that solving Pell's equation $x^2 - Ay^2 = 1$ is equivalent to asking for the element $u \in \mathbb{Z}(A)$, $u = x + y\sqrt{A}$, such that $N(u) = 1$. Moreover, define the *generalized Pell's equation*

$$x^2 - Ay^2 = k \tag{2.10}$$

This is solved by those element u of the ring $\mathbb{Z}(A)$ such that $N(u) = k$.

Theorem 1: *Suppose the pair (a, b) solves Pell's equation 1.5, and the pair (x, y) solves 2.10. Then, denoting $u = a + b\sqrt{A}$ and $v = x - y\sqrt{A}$, and*

$$uv = (ax + byA) + (bx + ay)\sqrt{A} = x' + y'\sqrt{A} \tag{2.11}$$

then, the pair of integers (x', y') is also a solution of 2.10.

Proof: this comes from the property of the norm. We have $N(u) = 1$ and $N(v) = k$, hence $N(uv) = N(u)\,N(v) = k$, hence uv solves generalized Pell's equation. *End of proof.*

Theorem 2: *Suppose that a pair of integers (a, b) solves Pell's equation. Consider a real transformation of the plane $(x, y) \rightarrow (x', y')$*

$$x' = ax + bAy, \tag{2.12}$$

$$y' = bx + ay \tag{2.13}$$

This transformation maps solutions to 2.10 to other solutions of 2.10.

Proof: This follows trivially by simply interpreting the coefficients in the proof of Theorem 1. *End of proof.*

That is why the solution $(\pm 1, 0)$ of Pell's equation 1.5 is called *trivial*: it is equivalent to the identity map, which maps each solution onto itself (or onto minus itself).

Let us now make an example. The equation

$$x^2 - 2y^2 = 1 \tag{2.14}$$

has a non-trivial solution $(3, 2)$. Then, the transformation

$$x' = 3x + 4y, \tag{2.15}$$

$$y' = 2x + 3y \tag{2.16}$$

produces more solutions of

$$x^2 - 2y^2 = k \tag{2.17}$$

provided we already know one solution. For example, setting $k=-1$, a non-trivial solution is $(1,1)$. Then, we map $(1,1)\to(7,5)\to(41,29)\to(239,169)\to\dots$. Observe that

$$\frac{239}{169}-\sqrt{2}=-0.0000123 \tag{2.18}$$

Apparently, (x,y) become better and better approximants of $\sqrt{2}$.
Of course, it all hinges on the fact that

$$x^2-Ay^2=1 \tag{2.19}$$

always has a non-trivial solution, for every non-square A. Let us prove this rigorously.

Lemma 1: *Let α be an irrational number. Then, for every integer t the inequality*

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{tq} \tag{2.20}$$

always has a solution (p,q) for q integer and $1\le q\le t$.

Proof: Let $[\alpha]$ be the integer part and $\{\alpha\}\in[0,1)$ be the fractional part of a real number α, so that $\alpha=[\alpha]+\{\alpha\}$. Divide the unit interval $[0,1)$ into $1/t$ length intervals

$$\left[0,\frac{1}{t}\right),\left[\frac{1}{t},\frac{2}{t}\right),\dots,\left[\frac{t-1}{t},1\right) \tag{2.21}$$

Consider $t+1$ numbers $\{\alpha k\}$, $k=1,2,\dots,t+1$. by Pigeonhole Principle there will be at least two of them, say $\{\alpha k_1\}$ and $\{\alpha k_2\}$ (wlog $k_1>k_2$) that belong to the same interval, hence

$$|\{\alpha k_1\}-\{\alpha k_2\}|=|\alpha k_1-[\alpha k_1]-\big(\alpha k_2-[\alpha k_2]\big)\,|<\frac{1}{t} \tag{2.22}$$

define $q=k_1-k_2$ and $p=[\alpha k_1]-[\alpha k_2]$. Then, we get

$$|q\alpha-p|<\frac{1}{q},\qquad\Longrightarrow\qquad\left|\alpha-\frac{p}{q}\right|<\frac{1}{tq} \tag{2.23}$$

End of proof.

Corollary 1 (Dirichlet): *Let α be an irrational number. Then, the inequality*

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{q^2} \tag{2.24}$$

has infinitely many integer solutions (p,q).

Proof: From Lemma 1, because $t\ge q$, then

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{tq}\le\frac{1}{q^2} \tag{2.25}$$

so, every (p,q) that solves the first inequality, solves the second. But, as t grows more and more, new solutions will emerge[1][2]. *End of proof.*

Lemma 2: *For some integer $k<2\sqrt{A}+1$ the equation*

$$x^2-Ay^2=k \tag{2.26}$$

has infinitely many solutions.

Proof: Let (p,q) be a solution to 2.24, with $\alpha=\sqrt{A}$. Then, by Corollary 1:

$$\left|p^2-q^2A\right|=q^2\left|\sqrt{A}-\frac{p}{q}\right|\left|\sqrt{A}+\frac{p}{q}\right|<q^2\frac{1}{q^2}\left|\sqrt{A}+\frac{p}{q}\right| \tag{2.27}$$

[1]The complete reasoning is the following. The proof of Lemma 1 forces you to change your solution at some point, as t will grow so that $|\{\alpha k_1\}-\{\alpha k_2\}|<1/t$ will not be true anymore, and you will have to find a new solution, whose existence is guaranteed. This substitution can be done infinitely many times.

[2]Moreover, you can see that the fraction p/q becomes a better and better approximant of the irrational number $\sqrt{A}$, as q increases. We will discuss this property later.

and, of course, $p/q < \sqrt{A}+1$, hence

$$\left|p^2 - q^2 A\right| < \left|\sqrt{A} + \frac{p}{q}\right| < 2\sqrt{A}+1 \tag{2.28}$$

therefore this norm can only take finitely many integer values between $-(2\sqrt{A}+1)$ and $2\sqrt{A}+1$. Because, by Corollary 1, the inequality has infinitely many solutions, then there must be infinitely many pairs (p,q) such that

$$\left|p^2 - q^2 A\right| < 2\sqrt{A}+1 \tag{2.29}$$

and because the set of values of this norm is finite, by Pigeonhole Principle, there must exist some k such that $-(2\sqrt{A}+1) < k < 2\sqrt{A}+1$ and

$$p^2 - q^2 A = k \tag{2.30}$$

for infinitely many (p,q). *End of proof.*

Lemma 3: *There exists a non zero integer k and two positive integers $0 \leq a, b < |k|$ such that 2.10 has infinitely many solutions (x,y) such that*

$$x \equiv a \ (mod\ |k|), \qquad y \equiv b \ (mod\ |k|). \tag{2.31}$$

Proof: Let k be such that infinitely many solutions to 2.10 exist (whose existence is guaranteed by Lemma 2). Then, for any solution (x,y), we have $x \equiv i \pmod{|k|}$ and $y \equiv j \pmod{|k|}$, with $0 \leq i, j < |k|$. There are only k^2 possibilities for the pair (i,j), and infinitely many solutions. Again, by Pigeonhole Principle, there must be a pair (a,b) of residue for infinitely many solutions (x,y). *End of proof.*

Theorem 3: *For every positive integer A that is not the square of a whole number, Pell's equation 1.5 has a non-trivial integer solution $(a,b) \neq (\pm 1, 0)$.*

Proof: Let k be such that $x^2 - Ay^2 = k$ has infinitely many solutions, whose existence is guaranteed by Lemma 2. Let (x_1, y_1) and (x_2, y_2) be two distinct solutions of $x^2 - Ay^2 = k$. Then for $u_1 = x_1 + y_1\sqrt{A}$ and $u_2 = x_2 + y_2\sqrt{A}$ we have $N(u_1) = N(u_2) = k$. Consider $v = u_1\bar{u}_2 = a + b\sqrt{A}$ for some a, b. Then

$$N(v) = N(u_1\bar{u}_2) = \frac{N(u_1)}{N(u_2)} = 1 \tag{2.32}$$

which means that (a,b) is at least a rational solution to Pell's equation 1.5. Now, we have to check that among all the possible (a,b) built as such, there exists at least one that is integer and non-trivial. Calculate a, b

$$v = (x_1 + y_1\sqrt{A})\frac{(x_2 - y_2\sqrt{A})}{k} = \frac{x_1x_2 - y_1y_2A}{k} + \frac{x_1y_2 - x_2y_1}{k}\sqrt{A} \tag{2.33}$$

and so

$$a = \frac{x_1x_2 - y_1y_2A}{k}, \qquad b = \frac{x_1y_2 - x_2y_1}{k} \tag{2.34}$$

Let (x_1, y_1) and (x_2, y_2) be distinct solutions of 2.10 such that

$$x_1 \equiv x_2 \pmod{|k|}, \qquad y_1 \equiv y_2 \pmod{|k|} \tag{2.35}$$

whose existence is guaranteed by Lemma 3. Then

$$x_1x_2 - y_1y_2A \equiv x_1^2 - Ay_1^2 \equiv 0 \pmod{|k|}, \qquad x_1y_2 - x_2y_1 \equiv x_1y_1 - x_1y_1 = 0 \pmod{|k|} \tag{2.36}$$

which means that (a,b) must be integers. Finally, suppose by contradiction that $(a,b) = (\pm 1, 0)$. Then, $v = u_1\bar{u}_2 = 1 + 0\sqrt{A} = 1$. By uniqueness of the inverse, this is only possible if $u_1 = u_2$. Except the two solutions were distinct. Hence, $(a,b) \neq (\pm 1, 0)$. *End of proof.*

Lemma 4: *Let (x,y) be a non-trivial integer solution to Pell's equation 1.5, and $u = x + y\sqrt{A}$. Then*

- *If $x > 0$ and $y > 0$, then $u > 1$.*
- *If $x > 0$ and $y < 0$, then $0 < u < 1$.*
- *If $x < 0$ and $y > 0$, then $-1 < u < 0$.*
- *If $x < 0$ and $y < 0$, then $u < -1$.*

Proof: For every non-trivial integer solution to Pell's equation 1.5 (x, y), then $(x, -y)$ is also a solution. This means that for every $u \in \mathbb{Z}(A)$, then $\bar{u} \in \mathbb{Z}(A)$. Suppose $x > 0$ and $y > 0$. Then, $u > 0$. Clearly, $u = x + y\sqrt{A} > x - y\sqrt{A} = \bar{u}$. Moreover, $u\bar{u} = x^2 - Ay^2 = 1$, which forces $\bar{u} > 0$, $u > 1$ and $\bar{u} < 1$. This proves the first two statements. Analogously, $(-x, y)$ is also a solution, so the last two statements follow. *End of proof.*

Theorem 4: *Let $k \neq 1$ be an integer and A be a positive integer, and not the square of a whole number. Suppose that 2.10 has at least one solution. Then, it has infinite distinct solutions.*

Proof: Let (x_0, y_0) be the non-trivial solution to Pell's equation 1.5 guaranteed by Theorem 3. Let (x_1, y_1) be the solution of 2.10. Define $u_0 = x_0 + y_0\sqrt{A}$ and $u_1 = x_1 + y_1\sqrt{A}$. Define the succession

$$u_n = u_0 u_{n-1} = x_n + y_n\sqrt{A} \tag{2.37}$$

Then, by Theorem 1, (x_n, y_n) is a solution to 2.10 as well. Moreover, for every non-trivial pair (x_0, y_0), because of Lemma 4, $u_0 \neq 1$. Thus, the solutions $\{u_1, \dots, u_n, \dots\}$ are all distinct. We have exhibited infinitely many distinct solutions to 2.10. *End of proof.*

Definition 1: *Let (a, b) be a non-trivial solution to Pell's equation 1.5 such that $a > 0$, $b > 0$. We say that this solution is* fundamental *if $u = a + b\sqrt{A}$ takes the smalles possible value.*

Note that the fundamental solution is unique. If $a + b\sqrt{A} = a' + b'\sqrt{A}$, then either $\sqrt{A}$ is rational, or $b - b' = a - a' = 0$. Also, by Lemma 4, $u > 1$.

Theorem 5: *Let (x_1, y_1) be the fundamental solution to Pell's equation 1.5 and $u = x_1 + y_1\sqrt{A}$. Let*

$$u^n = (x_1 + y_1\sqrt{A})^n := x_n + y_n\sqrt{A}, \qquad n \in \mathbb{N} \tag{2.38}$$

Then, $(\pm x_n, \pm y_n)$ is the complete set of solutions to Pell's equation.

Proof: The trivial solution $(1, 0)$ is obtained for $n = 0$. Let (x, y) be an arbitrary, non-trivial solution to Pell's equation 1.5, according to Theorem 3. Because $(\pm x, \pm y)$ will be a solution, we choose without loss of generality $x > 0$ and $y > 0$. All we need to show is that there exists $n \in \mathbb{N}^*$ so that $v = x + y\sqrt{A} = u^n$. Suppose the contrary. By Lemma 4, $v > 1$ and $u > 1$. This means that the sequence $1, u, u^2, \dots, u^n, \dots$ is growing and diverging. Hence, there must exist some n such that $u^n < v < u^{n+1}$. Dividing by u^n

$$1 < v(u^n)^{-1} < u \tag{2.39}$$

where $v(u^n)^{-1} := \bar{x} + \bar{y}\sqrt{A} \in \mathbb{Q}(A)$, for some $\bar{x}, \bar{y} \in \mathbb{Q}$. Observe that $(u^n)^{-1} = (u^{-1})^n = (x_1 - y_1\sqrt{A})^n$. This means $(u^n)^{-1} \in \mathbb{Z}(A)$, so $v(u^n)^{-1} \in \mathbb{Z}(A)$ and thus $\bar{x}, \bar{y} \in \mathbb{Z}$. Moreover

$$N(v(u^n)^{-1}) = \frac{N(v)}{N(u)^n} = 1 \tag{2.40}$$

This means that $(\bar{x}, \bar{y})$ is a solution to Pell's equation 1.5. Finally, because $v(u^n)^{-1} > 1$, we must have $\bar{x} > 0$ and $\bar{y} > 0$ by Lemma 4. But $v(u^n)^{-1} < u$ contradicts the definition of u as fundamental solution. Then, v cannot exist, and we exhibited every solution of Pell's equation. *End of proof.*

Theorem 6: *The complete set of solutions to Pell's equation 1.5 forms an abelian group under multiplication, called* Pell's orbit*.*

Proof: Let

$$\mathbb{Z}_1(A) = \{u \in \mathbb{Z}(A) | N(u) = 1\} \tag{2.41}$$

be the set of solutions to Pell's equation. Let (x_0, y_0) be the fundamental solution of the equation, and $\varepsilon = x_0 + y_0\sqrt{A}$, so that $(\pm x_0, \pm y_0)$ correspond to $\pm\varepsilon, \pm\varepsilon^{-1}$. By Theorem 5, we have

$$\mathbb{Z}_1(A) = \{\pm\varepsilon^n, n \in \mathbb{Z}\} = \{\dots, \pm\varepsilon^{-n}, \pm\varepsilon^{-n-1}, \dots, \pm\varepsilon^{-2}, \pm\varepsilon^{-1}, \pm 1, \pm\varepsilon, \pm\varepsilon^2, \cdots \pm\varepsilon^{n-1}, \pm\varepsilon^n, \dots\} \tag{2.42}$$

which is close under multiplication, and it has a neutral elements, the inverse and it is associative[3]. Moreover, multiplication is commutative, making $(\mathbb{Z}_1(A), *)$ abelian. *End of proof.*

Theorem 7: *(x_0, y_0), with $x_0 > 0$ and $y_0 > 0$ is the fundamental solution to Pell's equation 1.5 if and only if y_0 is minimal among all integer solutions with positive components.*

[3]We leave the proof of these properties as an exercise for the reader.

Proof: Select the orbit of all solutions $\mathbb{Z}_1(A)$, and call $\varepsilon = x_0 + y_0\sqrt{A}$ the fundamental unit, so that (x_0, y_0) is the fundamental solution. Then, by Theorem 5, every other integer solutions with positive components are of the kind

$$\varepsilon^n = (x_0 + y_0\sqrt{A})^n = \sum_{k \text{ even}} \binom{n}{k} x_0^{n-k}(y_0\sqrt{A})^k + \left[\sum_{k \text{ odd}} \binom{n}{k} x_0^{n-k} y_0^k (\sqrt{A})^{k-1}\right]\sqrt{A} := x_n + y_n\sqrt{A} \tag{2.43}$$

where

$$y_n = \sum_{k \text{ odd}} \binom{n}{k} x_0^{n-k} y_0^k (\sqrt{A})^{k-1} \tag{2.44}$$

is the sum of positive terms. Already setting $k = 1$, we have $x_0, \sqrt{A} > 1$ and

$$y_n > n x_0^n y_0 \sqrt{A} > y_0 \tag{2.45}$$

which proves that every other solution has $y_n > y_0$. *End of proof.*

3 Theory on continued fractions

Now we have determined the algorithm to find all solutions to Pell's equation. But we are left with the problem of determining the fundamental solution. Rather than brute forcing it, there is a quite beautiful link between the fundamental solution of the equation and the continued fraction representation of $\sqrt{A}$. Let us set the notation for this chapter

$$\frac{13}{5} = 2 + \frac{1}{1+\frac{1}{1+\frac{1}{2}}} = [2,1,1,2] = 2 + \frac{1}{1+}\frac{1}{1+}\frac{1}{2} \tag{3.1}$$

Now, we formalize it. How do we expand a rational number into its simple continued fraction? We need Euclid's algorithm for the *greatest common divisor* (gcd).

Euclid's algorithm: *Take two numbers p, q. The algorithm proceeds as follows:*

- *If $p < q$, swap p and q.*
- *Divide p by q and call a the quotient and r the remainder. If $r = 0$, then q is the gcd.*
- *If $r \neq 0$, set $p = q$ and $q = r$ and repeat.*

As an example, take $p = 43$ and $q = 19$. Then, the succession of quotients and remainders according to the algorithm is

$$43/19 = (2,5) \quad \to \quad 19/5 = (3,4) \quad \to \quad 5/4 = (1,1) \quad \to \quad 4/1 = (4,0) \tag{3.2}$$

Then, the continued fraction of p/q is just the list of quotients

$$\frac{43}{19} = 2 + \frac{1}{3+\frac{1}{1+\frac{1}{4}}} = [2,3,1,4] \tag{3.3}$$

We can now give the generalized version of Euclid's algorithm, working for any real number x.

Definition 2: *Let $x \in \mathbb{R}$. The* continued fraction expansion *of x is $[a_0, a_1, a_2, a_3, \dots] = a_0 + \frac{1}{a_1+}\frac{1}{a_2+}\dots$ where $a_0 \in \mathbb{Z}$ and $a_j \in \mathbb{N}$ for $j \geq 1$, such that:*

- *$a_0 = [x]$ is the integer part, rounded down (floor function). Call the remainder fractional part $\{x\} = r_1$, with $0 \leq r_1 < 1$.*
- *For $j \geq 0$, inductively set*

$$a_{j+1} = \begin{cases} [1/r_j], & r_j \neq 0 \\ 0, & r_j = 0 \end{cases} \tag{3.4}$$

  so that $r_{j+1} = a_j - r_j \in [0,1)$.

If $r_j = 0, \forall j > n$, the continued fraction expansion is called finite*, and it is represented as a finite sequence $[a_0, a_1, \dots, a_n]$.*

Observe that Euclid's algorithm terminates if and only if the resulting continued fraction is finite. Also, note that in this notation, we have

$$[a_0, \dots, a_n] = [a_0, a_1, \dots, a_m, [a_{m+1}, \dots, a_n]], \qquad 0 \leq m \leq n \tag{3.5}$$

Definition 3: *For $0 \leq m \leq n$, we call $[a_0, \dots, a_m]$ the* m-th *convergent to $[a_0, \dots, a_n]$.*

Theorem 8: *A number is rational if and only if its continued fraction expansion is finite.*

Proof: If a number is rational, then Euclid's algorithm always terminates. This is because at each step $a > b > r$ is required, and $b > r$: the successions of integers b_i is then always strictly decreasing, until eventually $b_n = 1$. Instead, if we have a finite continued fraction $[q_1, \dots, q_n]$, by induction

$$[a_0, \dots, a_n] = [a_0, [a_1, \dots, a_n]] = a_0 + \frac{1}{[a_1, \dots, a_n]} = \frac{a_0[a_1, \dots, a_n] + 1}{[a_1, \dots, a_n]} \tag{3.6}$$

so, if $[a_1, \dots, a_n]$ is a rational number, $[a_0, \dots, a_n]$ is. But $[a_n] = a_n$, which is rational, so we are done. *End of proof.*

We remark that there are at least two ways[4] to write down any rational number as a continued fraction: if $a_n \geq 2$

$$[a_0, \ldots, a_n] = [a_0, \ldots, a_n - 1, 1] \tag{3.7}$$

and if $a_n = 1$

$$[a_0, \ldots, 1] = [a_0, \ldots, a_{n-1} + 1] \tag{3.8}$$

depending on a_n, we may add or subtract a convergent.

Theorem 9: *Let* $p_0 = a_0, p_1 = a_1 a_0 + 1, q_0 = 1, q_1 = a_1$ *and*

$$p_n = a_n p_{n-1} + p_{n-2} \qquad n \geq 2 \tag{3.9}$$

$$q_n = a_n q_{n-1} + q_{n-2} \qquad n \geq 2 \tag{3.10}$$

Then

$$[a_0, \ldots, a_n] = \frac{p_n}{q_n}. \tag{3.11}$$

Proof: We go by induction. The base cases are

$$[a_0] = a_0 = \frac{p_0}{q_0}, \qquad [a_0, a_1] = \frac{a_0 a_1 + 1}{a_1} = \frac{p_1}{q_1} \tag{3.12}$$

Inductive reasoning at the m-th step

$$[a_0, \ldots, a_{m-1}, a_m] = \frac{p_m}{q_m} = \frac{a_m p_{m-1} + p_{m-2}}{a_m q_{m-1} + q_{m-2}} \tag{3.13}$$

implies at the $m+1$-th step

$$\begin{aligned}
[a_0, \ldots, a_{m-1}, a_m, a_{m+1}] &= \left[a_0, \ldots, a_{m-1}, a_m + \frac{1}{a_{m+1}}\right] = \\
&= \frac{(a_m + 1/a_{m+1}) p_{m-1} + p_{m-2}}{(a_m + 1/a_{m+1}) q_{m-1} + q_{m-2}} = \\
&= \frac{a_{m+1}\left(a_m p_{m-1} + p_{m-2}\right) + p_{m-1}}{a_{m+1}\left(a_m q_{m-1} + q_{m-2}\right) + q_{m-1}} = \\
&= \frac{a_{m+1} p_m + p_{m-1}}{a_{m+1} q_m + q_{m-1}} = \\
&= \frac{p_{m+1}}{q_{m+1}}
\end{aligned} \tag{3.14}$$

and the theorem is proven. *End of proof.*

Theorem 10: *The numbers* p_n *and* q_n *satisfy*

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}. \tag{3.15}$$

Proof: Again, by induction: substitute the rules for p_n, q_n.

$$p_n q_{n-1} - p_{n-1} q_n = (a_n p_{n-1} + p_{n-2}) q_{n-1} - p_{n-1}(a_n q_{n-1} + q_{n-2}) = -(p_{n-1} q_{n-2} - p_{n-2} q_{n-1}) \tag{3.16}$$

and the first steps at $n = 1$ gives

$$p_1 q_0 - p_0 q_1 = a_1 a_0 + 1 - a_1 a_0 = 1 \tag{3.17}$$

which completes the proof. *End of proof.*

Corollary 1: p_n *and* q_n *are coprimes, so each convergent is in its lowest terms.*

Proof: If p_n and q_n have d as their gcd, then $p_n q_{n-1} - p_{n-1} q_n$ would be a multiple of d. But by Theorem 10 $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$, which forces $d = 1$. *End of proof.*

Definition 4: *We call* $a'_m = [a_m, a_{m+1}, \ldots, a_n]$ *the* m-th complete quotient *of the continued fraction* $[a_0, \ldots, a_n]$.

[4] Exactly two ways, indeed. This fact will not be used, so we do not need to prove it.

This allows us to have, from Theorem 9

$$[a_0, \ldots, a_n] = a'_0 = \frac{a'_1 a_0 + 1}{a'_1} = \frac{a'_n p_{n-1} + p_{n-2}}{a'_n q_{n-1} + q_{n-2}} \tag{3.18}$$

Theorem 11: *The following statements are true:*

- *If* $a_n \neq 1$*, then* $a_m = [a'_m]$ *for* $0 \leq m \leq n-1$.
- *If* $a_n = 1$*, then* $a_m = [a'_m]$ *for* $0 \leq m \leq n-2$.
- *If* $a_n = 1$*, then* $a_{n-1} = a'_{n-1} - 1 = [a'_{n-1}] - 1$.
- $a_n = a'_n = [a'_n]$.

Proof: By induction. Trivially, $a_0 = a'_0 = [a'_0]$. Then

$$a'_m = a_m + \frac{1}{a'_{m+1}} \tag{3.19}$$

Because $a_0, \ldots, a_n$ are all non-negative integers, $a'_{m+1} \geq 1$. In particular, $a'_{m+1} = 1$ only when $m = n-1$ and $a_n = 1$. In such case, $a'_{n-1} = a_{n-1} + 1$, which is also an integer, so $a_{n-1} = a'_{n-1} - 1 = [a'_{n-1}] - 1$. In every other case, $a'_{m+1} > 1$, so

$$a_m < a'_m < a_m + 1, \qquad \Longrightarrow \qquad a_m = [a'_m] \tag{3.20}$$

And, of course, for $m = n$, a'_n is an integer, so $a_n = a'_n = [a'_n]$. *End of proof.*

Theorem 12: *Take two continued fractions* $[a_0, \ldots, a_n]$*, with* $a_n > 1$*, and* $[b_0, \ldots, b_l]$*, with* $b_l > 1$*. If*

$$[a_0, \ldots, a_n] = [b_0, \ldots, b_l] \tag{3.21}$$

then the two fractions are identical, $n = l$ *and* $a_m = b_m$ *for* $0 \leq m \leq n$.

Proof: Suppose, without loss of generality, that $n \geq l$. Let us proceed by induction. Trivially, if the value of the two fractions is the same, then $a_0 = b_0$ and $[a_0, a'_1] = [b_0, b'_1]$. This implies

$$a_0 + \frac{1}{a'_1} = b_0 + \frac{1}{b'_1} \tag{3.22}$$

which means $a'_1 = b'_1$. From Theorem 11, this implies $a_1 = b_1$. Then, suppose the first m coefficients are the same

$$[a_0, a_1, \ldots, a_{m-1}, a'_m] = [b_0, b_1, \ldots, b_{m-1}, b'_m] \tag{3.23}$$

this means that every convergent p_i, q_i is the same up to $m-1$. Then, we can use 3.18

$$\frac{a'_m p_{m-1} + p_{m-2}}{a'_m q_{m-1} + q_{m-2}} = \frac{b'_m p_{m-1} + p_{m-2}}{b'_m q_{m-1} + q_{m-2}} \tag{3.24}$$

which reduces to

$$(a'_m - b'_m)\,(p_{m-1} q_{m-2} - p_{m-2} q_{m-1}) = (a'_m - b'_m)(-1)^m \tag{3.25}$$

where we used Theorem 10. So, $a'_m = b'_m$, which by Theorem 12 implies $a_m = b_m$. By induction principle, we have shown that $a_m = b_m$ for $0 \leq m \leq n$. Now, if $l > n$, then

$$\frac{p_n}{q_n} = [a_0, \ldots a_n] = [b_0, \ldots b_n, b'_{n+1}] = \frac{b'_{n+1} p_n + p_{n-1}}{b'_{n+1} q_n + q_{n-1}} \tag{3.26}$$

which reduces to

$$p_n q_{n-1} - p_{n-1} q_n = 0 \tag{3.27}$$

which contradicts Theorem 10. So, $n = l$, and the theorem is proven. *End of proof.*

Theorem 13: *Let* x *be a continued fraction and* $x_m = p_m/q_m$ *the* m*-th convergent. Then:*

- x_{2m} *is a strictly increasing succession, while* x_{2m+1} *is a strictly decreasing succession.*
- $x_{2m} < x < x_{2m+1}$ *for every* m.

Proof: Calculate, using Theorems 10 and 11

$$\begin{aligned} x_m - x_{m-2} &= \frac{p_m}{q_m} - \frac{p_{m-2}}{q_{m-2}} = \frac{p_m}{q_m} - \frac{p_{m-1}}{q_{m-1}} + \frac{p_{m-1}}{q_{m-1}} - \frac{p_{m-2}}{q_{m-2}} = \\ &= \frac{(-1)^{m-1}}{q_m q_{m-1}} + \frac{(-1)^{m-2}}{q_{m-1}q_{m-2}} = \frac{(-1)^m(q_m - q_{m-2})}{q_m q_{m-1} q_{m-2}} = \\ &= \frac{(-1)^m a_m q_{m-1}}{q_m q_{m-1} q_{m-2}} = \frac{(-1)^m a_m}{q_m q_{m-2}} \end{aligned} \tag{3.28}$$

and a_m, q_m, q_{m-2} are strictly positive, so the sign is decided by the parity of m: it is positive (negative) if m is even (odd). Similarly

$$x_m - x_{m-1} = \frac{p_m}{q_m} - \frac{p_{m-1}}{q_{m-1}} = \frac{(-1)^{m-1}}{q_m q_{m-1}} \tag{3.29}$$

which is positive (negative) if m is odd (even). Finally, x is the last convergent, the n-th. If n is odd (even), it is smaller (greater) than any other odd (even) convergent, but it is greater (smaller) than x_{n-1}, because $n-1$ is even (odd), and because x_{n-1} is the last even (odd) convergent, it must be greater (smaller) than any other even (odd) convergent. This tells us that the picture with the convergents of x is

$$x_0 < x_2 < x_4 < \cdots < x < \cdots < x_5 < x_3 < x_1 \tag{3.30}$$

Hence, $x_{2m} < x < x_{2m+1}$ is always verified. *End of proof.*

Definition 5: *The rational number p/q is the* best approximation *to a real number x if the distance from p/q to x on the real line is less than the distance from any other rational number to x (with denominator less than or equal to q).*

Theorem 14: *The denominators of the convergents satisfy $q_n \geq n$, with inequality becoming strict when $n > 3$.*

Proof: $q_0 = 1 > 0$, $q_1 = a_1 \geq 1$. For $n \geq 2$

$$q_n = a_n q_{n-1} + q_{n-2} \geq q_{n-1} + 1 \tag{3.31}$$

and inductively, we see $q_n \geq n$. If $n > 3$

$$q_n \geq q_{n-1} + q_{n-2} > q_{n-1} + 1 \geq n \tag{3.32}$$

then $q_n > n$. *End of proof.*

Theorem 15: *Every continued fraction can be written as an alternating sum in the following manner*

$$[a_0, \ldots, a_n] = a_0 + \frac{1}{q_1 q_0} - \frac{1}{q_2 q_1} + \cdots + \frac{(-1)^{n-1}}{q_n q_{n-1}}. \tag{3.33}$$

Proof: Just write the n-th convergent

$$\frac{p_n}{q_n} = \frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} + \frac{p_{n-1}}{q_{n-1}} - \frac{p_{n-2}}{q_{n-2}} + \frac{p_{n-2}}{q_{n-2}} + \cdots + \frac{p_1}{q_1} - \frac{p_0}{q_0} + \frac{p_0}{q_0} \tag{3.34}$$

Apply Theorem 13

$$\frac{p_n}{q_n} = [a_0, \ldots, a_n] = a_0 - \frac{1}{q_1 q_0} + \frac{1}{q_2 q_1} + \cdots + \frac{(-1)^n}{q_n q_{n-1}}. \tag{3.35}$$

and we're done. *End of proof.*

Theorem 16: *For any number x with convergent p_m/q_m*

$$\left| x - \frac{p_m}{q_m} \right| \leq \frac{1}{q_m q_{m+1}} \tag{3.36}$$

Proof: By using complete quotients, 3.18 and Theorem 13

$$x - \frac{p_m}{q_m} = \frac{a'_{m+1} p_m + p_{m-1}}{a'_{m+1} q_m + q_{m-1}} - \frac{p_m}{q_m} = \frac{-p_m q_{m-1} + p_{m-1} q_m}{q_m (a'_{m+1} q_m + q_{m-1})} = \frac{(-1)^m}{q_m (a'_{m+1} q_m + q_{m-1})} = \frac{(-1)^m}{q_m q'_{m+1}} \tag{3.37}$$

where $q'_{m+1} = a'_{m+1}q_m + q_{m-1}$. By Theorem 11, $a'_{m+1} > a_m$, so $q'_{m+1} > q_{m+1}$, proving the theorem. *End of proof.*

Theorem 17: *The m-th convergent p_m/q_m to x is the best approximation to x with denominator smaller or equal to q_m.*

Proof: Because $q_{m-1} < q_m$, then p_{m-1}/q_{m-1} and p_m/q_m are two convergents with denominator less of equal to q_m. By Theorem 13, x is sandwiched between them. So, suppose p/q is the best approximation to x with denominator less or equal to q_m. Then, it has to lie between p_{m-1}/q_{m-1} and p_m/q_m. Then, we have two inequalities. The first one

$$\left|\frac{p}{q} - \frac{p_{m-1}}{q_{m-1}}\right| = \left|\frac{pq_{m-1} - p_{m-1}q}{qq_{m-1}}\right| \geq \frac{1}{qq_{m-1}} \tag{3.38}$$

because, by Corollary 1, p_{m-1} and q_{m-1} are coprimes. And so are p, q, by definition of best approximation. This means that $pq_{m-1} - p_{m-1}q = 0$ is not possible, and so $pq_{m-1} - p_{m-1}q \geq 1$. The second one

$$\left|\frac{p}{q} - \frac{p_{m-1}}{q_{m-1}}\right| \leq \left|\frac{p_m}{q_m} - \frac{p_{m-1}}{q_{m-1}}\right| = \frac{1}{q_m q_{m-1}} \tag{3.39}$$

again by Theorem 10. Therefore, we have

$$\frac{1}{qq_{m-1}} \leq \left|\frac{p}{q} - \frac{p_{m-1}}{q_{m-1}}\right| \leq \frac{1}{q_m q_{m-1}} \tag{3.40}$$

but we also know that $q \leq q_m$. This implies that for the inequality to hold, $p/q = p_m/q_m$. The theorem is proven: the convergents p_m/q_m are best approximations of the number x. *End of proof.*

We mention en passant that, although all convergents are the best approximations, not all the best approximations are necessarily convergents, but this fact is not needed, so we will not exhibit any proof of that.

Let us now enter the realm of infinite continued fractions. All of the theorems stay the same, with the exception of the only missing part being the proof of convergence of an infinite continued fraction into a real number x. Let us show just that.

Theorem 18: *Let $\{a_i\}$ be a sequence of strictly positive integers, $i \in \mathbb{N}$, and let $x_n = [a_0, a_1, \ldots, a_n]$. Then, the limit*

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} [a_0, a_1, \ldots, a_n] \tag{3.41}$$

is well defined and finite.

Proof: From Theorem 13, we know that the even (odd) convergents x_{2n} (x_{2n-1}) are an increasing (decreasing) succession, that is limited by the odd (even) convergents. This means that the succession of even (odd) convergents has a well defined limit. Again, from Theorem 13 and Theorem 14

$$x_{2n-1} - x_{2n} = \frac{p_{2n-1}}{q_{2n-1}} - \frac{p_{2n}}{q_{2n}} = \frac{1}{q_{2n}q_{2n-1}} \leq \frac{1}{2n(2n-1)} \tag{3.42}$$

which goes to zero as $n \to \infty$, proving that the even and odd convergents approach the same limit x. *End of proof.*

Now that we know that the infinite continued fraction converges. All of the other theorems stay the same, as inductive reasoning and convergence of the continued fraction by Theorem 18 are all it takes to generalize to the infinite case every theorem we have proven so far. One last theorem on approximations and convergents is needed to be proven.

Theorem 19 (Legendre): *If x is a real number and p, q are positive integers such that*

$$\left|x - \frac{p}{q}\right| < \frac{1}{2q^2} \tag{3.43}$$

then p/q is a convergent of x.

Proof (Hardy): Assume, without loss of generality, that $x > p/q$. Then, define the real number θ

$$x - \frac{p}{q} = \frac{\theta}{q^2} \tag{3.44}$$

with $0 < \theta < 1/2$. Write the finite continued fraction of p/q

$$\frac{p}{q} = [a_0, a_1, \ldots, a_n] \tag{3.45}$$

Because each rational number has two distinct continued fraction representations, choose the one with even n. So $p/q = p_n/q_n$ is also the n-th convergent. Set

$$\omega = \frac{1}{\theta} - \frac{q_{n-1}}{q_n}, \qquad \Longrightarrow \qquad \theta = \frac{q_n}{q_{n-1} + \omega q_n} \tag{3.46}$$

where $\omega > 1$ because $1/\theta > 2$ and $q_{n-1} \leq q_n$. Thus, by Theorem 10, using the fact that n is even

$$\begin{aligned} x = \frac{p}{q} + \frac{\theta}{q^2} = \frac{p_n}{q_n} + \frac{1}{q_n(q_{n-1} + \omega q_n)} = \\ = \frac{(p_n q_{n-1} + 1) + \omega p_n q_n}{q_n(q_{n-1} + \omega q_n)} = \frac{p_{n-1} q_n + \omega p_n q_n}{q_n(q_{n-1} + \omega q_n)} = \frac{\omega p_n + \omega p_{n-1}}{\omega q_n + q_{n-1}} \end{aligned} \tag{3.47}$$

which means that, by 3.18, $\omega = a'_{n+1}$. This implies directly that $p/q = p_n/q_n$ must be a convergent for x. *End of proof.*

And now, we introduce periodic continued fractions, which are the fulcrum of the algorithm for the fundamental solution to Pell's equation.

Definition 6: *A* periodic continued fraction *is an infinite continued fraction in which* $a_l = a_{l+k}$ *for a fixed positive* k *and for all* $l \geq L$. *The set of partial quotients*

$$a_L, a_{L+1}, \ldots, a_{L+k-1} \tag{3.48}$$

is called the period *and the continued fraction may be written*

$$[a_0, \ldots, a_{L-1}, \overline{a_L, a_{L+1}, \ldots, a_{L+k-1}}]. \tag{3.49}$$

If $L = 0$, *then the fraction is called* purely periodic.

Theorem 20: *A periodic continued fraction is a quadratic surd,* i.e. *an irrational root of a quadratic equation with whole coefficients.*

Proof: With the notation in Definition 6, we have

$$a'_L = [a_L, a_{L+1} \ldots, a_{L+k-1}, a_L, a_{L+1} \ldots] = [a_L, a_{L+1} \ldots, a_{L+k-1}, a'_L] \tag{3.50}$$

if we call p''/q'' and p'/q' the last two convergents of $[a_L, a_{L+1} \ldots, a_{L+k-1}]$, we have the quadratic equation for a'_L

$$a'_L = \frac{p' a'_L + p''}{q' a'_L + q''}, \qquad \Longrightarrow \qquad q'(a'_L)^2 + (q'' - p')a'_L - p'' = 0 \tag{3.51}$$

But we also have that

$$x = \frac{a'_L p_{L-1} + p_{L-2}}{a'_L q_{L-1} + q_{L-2}}, \qquad \text{inverting} \qquad a'_L = \frac{p_{L-2} - q_{L-2} x}{q_{L-1} x - p_{L-1}} \tag{3.52}$$

and substituting into the equation for a'_L, we find a quadratic equation for x. There is no point in writing it, the goal was to show it is quadratic. *End of proof.*

To prove the inverse, unfortunately, we need a bit more theory on quadratic surds.

3.1 Lagrange's theorem

The two roots of the quadratic equation

$$ax^2 + bx + c = 0 \tag{3.53}$$

with integer coefficients a, b, c, are

$$x = \frac{P + \sqrt{A}}{Q}, \qquad \bar{x} = \frac{P - \sqrt{A}}{Q} \tag{3.54}$$

where $P = -b$, $Q = 2a$, $A = b^2 - 4ac$ are all integers. Assume $Q > 0$ and A is not a perfect square. We shall refer to $\bar{x}$ as the *conjugate* of x. Finally, note that P, Q, A and a, b, c are in 1:1 correspondence, hence for every quadratic equation there is a quadratic surd, and viceversa.

Definition 7: *A quadratic surd x is said to be* reduced *if $x > 1$ and its conjugate $\bar{x}$ lies between -1 and 0.*

A few properties of reduced quadratic surds are in order:

- Because $x > 1$ and $\bar{x} > 1$, $x + \bar{x} = 2A/Q > 0$. But $Q > 0$ was assumed, so $P > 0$.
- Because $\bar{x} < 0$, then $P < \sqrt{A}$. Because $x > 1$, $P + \sqrt{A} > Q$. Because $\bar{x} > -1$, then $\sqrt{A} - P < Q$. So, $0 < P < \sqrt{A}$ and $\sqrt{A} - P < Q < \sqrt{A} + P < 2\sqrt{A}$.
- For a given A that is not a perfect square, there is only a finite number of possible reduced quadratic surds, because $P < \sqrt{A}$ and $Q < 2\sqrt{A}$ are limited positive integers.
- For a given A that is not a perfect square, there always exists at least one reduced quadratic surd. Just set $Q = 1$ and choose any P smaller than $\sqrt{A}$ ($P = 1$ is guaranteed because $A > 1$).

Theorem 21: *For a reduced quadratic surd*

$$x = \frac{P + \sqrt{A}}{Q} \tag{3.55}$$

Then, the first complete quotient a_1' is still a reduced quadratic surd of the form

$$a_1' = \frac{P_1 + \sqrt{A}}{Q_1} \tag{3.56}$$

Moreover, $y = -1/\bar{x}$ is a reduced quadratic surd.

Proof: With the notation introduced in the previous chapter

$$x = a_0' = a_0 + \frac{1}{a_1'} \tag{3.57}$$

substitute in the starting equation

$$a\left(a_0 + \frac{1}{a_1'}\right)^2 + b\left(a_0 + \frac{1}{a_1'}\right) + c = 0 \tag{3.58}$$

which reduces to

$$(aa_0^2 + ba_0 + c)(a_1')^2 + (2aa_0 + b)a_1' + a = 0 \tag{3.59}$$

solving, we get

$$a_1' = \frac{P_1 + \sqrt{A_1}}{Q_1}, \qquad \bar{a}_1' = \frac{P_1 - \sqrt{A_1}}{Q_1} \tag{3.60}$$

where $P_1 = -(2aa_0 + b)$, $Q_1 = 2(aa_0^2 + ba_0 + c)$. As for the irrational part

$$A_1 = (2aa_0 + b)^2 - 4a(aa_0^2 + ba_0 + c) = b^2 - 4ac = A \tag{3.61}$$

We just need to show that a_1' is reduced. Clearly, $a_1' > 1$ is guaranteed, as it is the residual of the integral part of x, which is not an irrational number. Now, inverting the definition of a_1'

$$a_1' = \frac{1}{x - a_0} = \frac{Q[(P - a_0Q) - \sqrt{A}]}{(P - a_0Q)^2 - A} \tag{3.62}$$

with this, it is easy to check that

$$\bar{a}_1' = \overline{(x - a_0)^{-1}} = \frac{Q[(P - a_0Q) + \sqrt{A}]}{(P - a_0Q)^2 - A} = \frac{1}{\bar{x} - a_0} \tag{3.63}$$

which means, because $-1 < \bar{x} < 0$ and $a_0 \geq 1$

$$-\frac{1}{\bar{a}_1'} = a_0 - x > 1 \tag{3.64}$$

This implies that $-1 < \bar{a}_1' < 0$, making a_1' a reduced quadratic surd.
Finally, move to y, which clearly has $y > 1$. Moreover, $\bar{y} = -1/x$, so $-1 < \bar{y} < 0$, hence y is a reduced quadratic irrational. *End of proof.*

Theorem 22: *If x is a reduced quadratic surd, its continued fraction is purely periodic, namely*

$$x = [\overline{a_0, a_1, \ldots, a_L}]. \tag{3.65}$$

Proof: By Theorem 8, because x is not rational, its continued fraction must be infinite. By Theorem 21, the algorithm is stable under reduced quadratic surds, which means that every m-th complete quotient a'_m must be a reduced quadratic surd, for the same A, for every m. Namely, we have

$$x = a_0 + \frac{1}{a'_1}, \qquad a'_m = a_m + \frac{1}{a'_{m+1}} \tag{3.66}$$

for every $m \geq 1$. We have an infinite succession of reduced quadratic surds $a'_1, a'_2, \ldots, a'_{m-1}a'_m, \ldots$ with the same A. However, we know there must be only a finite number of reduced quadratic surds, so by Pigeonhole principle, there must be two identical complete quotients. Call a'_l the first repeated quotient, hence $a'_l = a'_k$, with $0 \leq k < l$.
Now, we prove that the continued fraction repeats after l. This is because of Theorem 11

$$a'_l = a'_k, \qquad \Longrightarrow \qquad a_l = a_k \tag{3.67}$$

Furthermore

$$a'_l = a_l + \frac{1}{a'_{l+1}} = a'_k = a_k + \frac{1}{a'_{k+1}} \tag{3.68}$$

which implies $a'_{l+1} = a'_{k+1}$, which implies $a_{l+1} = a_{k+1}$, and so on. Thus, the continued fraction must be periodic, namely

$$x = [a_0, \ldots, a_{k-1}, \overline{a_k, a_{k+1}, \ldots, a_{l-1}}]. \tag{3.69}$$

Finally, we prove that a_0 must be repeated. Suppose that a_0 is not repeated, hence $k > 0$. Then, $a'_l = a'_k$ implies $\bar{a}'_l = \bar{a}'_k$, which implies

$$b'_l = -\frac{1}{\bar{a}'_l} = -\frac{1}{\bar{a}'_k} = b'_k \tag{3.70}$$

Moreover, we have

$$a'_{k-1} = a_{k-1} + \frac{1}{a'_k}, \qquad a'_{l-1} = a_{l-1} + \frac{1}{a'_l} \tag{3.71}$$

take the conjugates

$$\bar{a}'_{k-1} = a_{k-1} + \frac{1}{\bar{a}'_k}, \qquad \bar{a}'_{l-1} = a_{l-1} + \frac{1}{\bar{a}'_l} \tag{3.72}$$

hence

$$-\frac{1}{\bar{a}'_k} = a_{k-1} - \bar{a}'_{k-1}, \qquad -\frac{1}{\bar{a}'_l} = a_{l-1} - \bar{a}'_{l-1} \tag{3.73}$$

or, equivalently

$$b'_k = a_{k-1} + \frac{1}{b'_{k-1}}, \qquad b'_l = a_{l-1} + \frac{1}{b'_{l-1}} \tag{3.74}$$

where, of course, because a'_l, a'_k are reduced, $b'_k > 1$ and $b'_k > 1$, which means that $0 < 1/b'_{k-1} < 1$ and $0 < 1/b'_{l-1} < 1$. Because $b'_k = b'_l$, we are left with

$$a_{k-1} = [b'_k] = [b'_l] = a_{l-1} \tag{3.75}$$

which means a_{l-1} is a repeated coefficient. This contradicts the fact that a_l was the first repeated coefficient. As a result, $k = 0$, and a_0 is repeated, making the sequence purely periodic

$$x = [\overline{a_0, a_1, \ldots, a_L}] \tag{3.76}$$

End of proof.

Continued Fraction Theorem (Lagrange): *Any quadratic surd x has a periodic continued fraction expansion. Namely*

$$x = [a_0, \ldots, a_{k-1}, \overline{a_k, a_{k+1}, \ldots a_{l-1}}]. \tag{3.77}$$

Proof: The idea of the proof is to show that every quadratic surd's continued fraction expansion will reach a complete quotient that is reduced. From Definition 4, we have

$$x = \frac{a'_{n+1}p_n + p_{n-1}}{a'_{n+1}q_n + q_{n-1}} \tag{3.78}$$

with $a_{n+1} > 1$ and quadratic surd. We will now show that for some n, $-1 < \bar{a}_{n+1} < 0$. Take the conjugate

$$\bar{x} = \frac{\bar{a}'_{n+1} p_n + p_{n-1}}{\bar{a}'_{n+1} q_n + q_{n-1}} \tag{3.79}$$

and solve for $\bar{a}_{n+1}$

$$\bar{a}'_{n+1} = -\frac{\bar{x} q_{n-1} - p_{n-1}}{\bar{x} q_n - p_n} = -\frac{q_{n-1}}{q_n}\left(\frac{\bar{x} - x_{n-1}}{\bar{x} - x_n}\right) \tag{3.80}$$

we know that $0 < q_{n-1} \leq q_n$. Moreover, by Theorem 13, if n is odd, we have $x_{n-1} < x < x_n$. This means that

$$\frac{\bar{x} - x_{n-1}}{\bar{x} - x_n} < \frac{\bar{x} - x}{\bar{x} - x} = 1 \tag{3.81}$$

Finally, because by Theorem 18

$$\lim_{n\to\infty} \frac{\bar{x} - x_{n-1}}{\bar{x} - x_n} = \frac{\bar{x} - x}{\bar{x} - x} = 1 \tag{3.82}$$

eventually, that fraction will be positive. This implies $-1 < \bar{a}'_{n+1} < 0$, and so a_{n+1} is reduced. By Theorem 22, a'_{n+1} has a purely periodic continued fraction, completing the proof. *End of proof.*

3.2 Fundamental Pell's solution

In this chapter, we will explore how to finally derive the fundamental solution to any Pell's equation, by exploiting properties of the continued fraction of irrational numbers, which thanks to Lagrange's theorem are periodic.

Theorem 23: *If $x = \sqrt{A}$, then its continued fraction takes the following form*[5]

$$\sqrt{A} = [a_0, \overline{a_1, \ldots, a_{l-1}, 2a_0}]. \tag{3.83}$$

Proof: Because $\sqrt{A}$ is not a reduced quadratic surd, it is not purely periodic. The period will start once we reach a reduced quadratic surd. Start with

$$\sqrt{A} = a_0 + \frac{1}{a'_1} \tag{3.84}$$

where we calculate

$$a'_1 = \frac{1}{\sqrt{A} - a_0} = \frac{a_0 + \sqrt{A}}{A - a_0^2} \tag{3.85}$$

Clearly, $a'_1 > 1$. Notice that its conjugate

$$\bar{a}'_1 = -\frac{1}{\sqrt{A} + a_0} \tag{3.86}$$

because $\sqrt{A} > a_0 \geq 1$, we have $-1 < \bar{a}'_1 < 0$, which means that a'_1 is reduced and the periodic part starts at a_1. Finally, suppose the period is l. Then $a_{l+1} = a_1$. This means

$$a'_l = a_l + \frac{1}{a'_1} = a_l + \sqrt{A} - a_0 = (a_l - a_0) + \sqrt{A} \tag{3.87}$$

take its conjugate

$$\bar{a}'_l = (a_l - a_0) - \sqrt{A} \tag{3.88}$$

which, because we know a'_l is reduced quadratic surd, is such that $-1 < \bar{a}'_l < 0$. This means that

$$a_0 + \sqrt{A} - 1 < \bar{a}'_l < a_0 + \sqrt{A} \tag{3.89}$$

because $\sqrt{A} - a_0 < 1$, we get

$$2a_0 < \bar{a}'_l < 2a_0 + 1 \tag{3.90}$$

and finally, by Theorem 11

$$a_l = [a'_l] = [\bar{a}'_l] = 2a_0 \tag{3.91}$$

which completes the proof. *End of proof.*

Theorem 24: *Let (p, q) be a solution to Pell's equation 1.5*

$$p^2 - Aq^2 = 1 \tag{3.92}$$

[5] Moreover, the sequence $a_1, \ldots a_{l-1}$ can be shown to be palindromic. This fact, however, is not needed.

with A not a perfect square, such that $p > 0, q > 0$. Then, p/q must be a convergent of A.

Proof: Let us calculate directly

$$\left|\sqrt{A} - \frac{p}{q}\right| = \frac{\left|p - q\sqrt{A}\right|}{q} = \frac{\left|p^2 - Aq^2\right|}{q(p + \sqrt{A})} = \frac{1}{q(p + \sqrt{A})} \tag{3.93}$$

However, because $p^2 = Aq^2 + 1$, then $p > q\sqrt{A} > q$, because $A > 1$. Then

$$\left|\sqrt{A} - \frac{p}{q}\right| = \frac{1}{q(p + \sqrt{A})} < \frac{1}{2q^2} \tag{3.94}$$

By Theorem 19, this implies p/q must be a convergent of $\sqrt{A}$. *End of proof.*

Lemma 5: *Let*

$$a'_{k+1} = \frac{P_{k+1} + \sqrt{A}}{Q_{k+1}} \tag{3.95}$$

be the $k+1$-th complete quotient for some continued fraction $[a_0, \dots]$. Then

$$p_k^2 - Aq_k^2 = (-1)^{k-1} Q_{k+1}. \tag{3.96}$$

Proof: By definition of complete quotient, we have

$$\sqrt{A} = \frac{a'_{k+1} p_k + p_{k-1}}{a'_{k+1} q_k + q_{k-1}} \tag{3.97}$$

which, after substituting a'_{k+1}, gives rise to the following equation

$$Aq_k + (q_{k-1}Q_{k+1} + P_{k+1})\sqrt{A} = (p_{k-1}Q_{k+1} + p_k P_{k+1}) + p_k\sqrt{A} \tag{3.98}$$

Both RHS and LHS belong to $\mathbb{Z}(A)$. So, they are equal if and only if

$$p_k = q_{k-1}Q_{k+1} + P_{k+1},$$
$$Aq_k = p_{k-1}Q_{k+1} + p_k P_{k+1}$$

multiply the first equation by p_k, the second equation by q_k, and subtract the two, making use of Theorem 10

$$p_k^2 - Aq_k^2 = (p_k q_{k-1} - p_{k-1}q_k)Q_{k+1} = (-1)^{k-1}Q_{k+1} \tag{3.99}$$

which ends the proof. *End of proof.*

Theorem 25: *Let $\sqrt{A} = [a_0, \overline{a_1, \dots, a_{l-1}, 2a_0}]$ be an irrational number, and p_n/q_n its n-th convergent. Let (x, y) be integer solutions to Pell's equation 1.5. Then, the fundamental solution to Pell's equation is:*

- *(p_{l-1}, q_{l-1}) if l is even.*
- *(p_{2l-1}, q_{2l-1}) if l is odd.*

Proof: First, we check that these are solutions. From Lemma 5, setting $k = l - 1$

$$(p_{l-1})^2 - A(q_{l-1})^2 = (-1)^l Q_l \tag{3.100}$$

From Theorem 23, we have

$$a'_l = a_0 + \sqrt{A} \tag{3.101}$$

which means that $Q_l = 1$. Thus

$$(p_{l-1})^2 - A(q_{l-1})^2 = (-1)^l \tag{3.102}$$

If l is even, we just checked that (p_{l-1}, q_{l-1}) is a solution. For l odd, we found a solution to $x^2 - Ay^2 = -1$, which is not Pell's equation. However, by periodicity, we also have

$$a'_l = a'_{2l} = a'_{3l} = \dots = a'_{kl} = a_0 + \sqrt{A} \tag{3.103}$$

Selecting a'_{2l} and repeating the argument, we get directly

$$(p_{2l-1})^2 - A(q_{2l-1})^2 = (-1)^{2l} = 1 \tag{3.104}$$

hence, (p_{2l-1}, q_{2l-1}) is a solution for every l, specifically in the case l is odd, this works.
Now, we need to prove that these solutions are fundamental. First, we observe that $p_{l-1}, p_{2l-1}, q_{l-1}, q_{2l-1} > 0$. Then, by Theorem 24, we know that every possible solution of Pell's equation (p_n, q_n) must be such that p_n/q_n is a convergent of $\sqrt{A}$. Moreover, during calculation of the periodic coefficients, we know that the complete quotients a'_{k+1} must always be reduced quadratic surds, by Theorem 22. Now, suppose that $Q_{k+1} = 1$, for some k. Then, because a'_{k+1} is a reduced quadratic surd, $-1 < \bar{a}'_{k+1} = P_{k+1} - \sqrt{A} < 0$, which implies that $P_{k+1} = a_0$. This forces $a'_{k+1} = a_0 + \sqrt{A}$, hence $a_{k+1} = 2a_0$, which forces l to divide $k+1$ by Theorem 23 (a_{k+1} must be the last coefficient of the period). So, out of every $Q_{k+1} = 1$, which produces candidate solutions to Pell's equation, $k = l - 1$ is the first possible one that works for l even, and $k = 2l - 1$ is the first possible one that works for l odd.
Finally, because $q_k \leq q_{k+1}$, this proves that the first possible solution of Pell's equation also has the smallest possible value of $y = q_{l-1}(q_{2l-1})$, which by Theorem 14 of section 2 proves that this is, indeed, the fundamental solution. *End of proof.*

4 Solution to generalized Pell's equation

In the last two subsections, we have catalogued all of the solutions to Pell's equation, and we have exhibited infinitely many integer solutions to the general Pell's equation

$$x^2 - Ay^2 = k \tag{4.1}$$

where A is not a perfect square, as long as one solution is provided. We are now going to catalogue all possible solutions to this general equation, and spell out the final algorithm.

Theorem 26: *Let $\sqrt{A}$ be an irrational number. Fix $\varepsilon = x_0 + y_0\sqrt{A}$ as the fundamental unit in Pell's orbit $\mathbb{Z}_1(A)$, so (x_0, y_0) is the fundamental solution to Pell's equation 1.5. Then, for every integer n:*

- *If $n = 0$, the equation $x^2 - Ay^2 = n$ has no solution.*
- *If $n \neq 0$, let U be the following set*

$$U := \{\bar{x} + \bar{y}\sqrt{A} \,|\, \bar{x}^2 - A\bar{y}^2 = n\} \tag{4.2}$$

with

$$|\bar{x}| \leq \frac{\sqrt{|n|}}{2}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right), \qquad |\bar{y}| \leq \frac{\sqrt{|n|}}{2\sqrt{A}}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right). \tag{4.3}$$

Then, every integral solution (x, y) of $x^2 - Ay^2 = n$ can be written introducing $u = x + y\sqrt{A}$ where $u \in U * \mathbb{Z}_1(A)$ and

$$\begin{aligned} U * \mathbb{Z}_1(A) &= \left\{uu_0 \,|\, (u, u_0) \in U \times \mathbb{Z}_1(A)\right\} = \\ &= \left\{\pm(\bar{x} + \bar{y}\sqrt{A})\varepsilon^k \,|\, k \in \mathbb{Z},\ \bar{x}^2 - A\bar{y}^2 = n\right\} \end{aligned} \tag{4.4}$$

and the limits set in 4.3. I will call the element uu_0 a Pell's orbit multiple *of u.*

- *If $n > 0$, the bound on $\bar{y}$ is sharper*

$$|\bar{y}| \leq \frac{\sqrt{n}}{2\sqrt{A}}\left(\sqrt{\varepsilon} - \frac{1}{\sqrt{\varepsilon}}\right) < \frac{\sqrt{n\varepsilon}}{2\sqrt{A}} \tag{4.5}$$

Proof: In the case $n = 0$, there can be no integral solutions. If there existed a solution (p, q), then we could write

$$\sqrt{A} = \frac{p}{q} \tag{4.6}$$

except $\sqrt{A}$ is irrational, so that cannot be.
For $(x, y) \in \mathbb{Z}^2 \setminus (0, 0)$, introduce

$$L(x + y\sqrt{A}) = (\log|x + y\sqrt{A}|, \log|x - y\sqrt{A}|) \in \mathbb{R}^2 \tag{4.7}$$

Defining $\alpha = x_1 + y_1\sqrt{A}$ and $\beta = x_2 + y_2\sqrt{A}$, we can easily show

$$L(\alpha\beta) = (\log|\tilde{x} + \tilde{y}\sqrt{A}|, \log|\tilde{x} - \tilde{y}\sqrt{A}|) = L(\alpha) + L(\beta)$$

$$\tilde{x} = x_1x_2 + y_1y_2A, \qquad \tilde{y} = x_1y_2 + y_1x_2$$

In particular, $L(\alpha^k) = kL(\alpha)$, $k \in \mathbb{Z}$. Because $\varepsilon^{-1} = x_0 - y_0\sqrt{A}$, then

$$L(\varepsilon) = \left(\log(\varepsilon), \log\left(\varepsilon^{-1}\right)\right) = \log(\varepsilon)(1, -1) \tag{4.8}$$

because $\varepsilon > 1$ (it is non-trivial by definition), then $L(\varepsilon)$ is independent from the vector $(1, 1)$. This means that $\{(1, 1), L(\varepsilon)\}$ is a basis of $\mathbb{R}^2$. For every solution $x^2 - Ay^2 = n$, we have

$$L(x + y\sqrt{A}) = c_1(1, 1) + c_2L(\varepsilon) \tag{4.9}$$

for some c_1, c_2 real numbers. In coordinates

$$(\log|x + y\sqrt{A}|, \log|x - y\sqrt{A}|) = (c_1 + c_2\log(\varepsilon), c_1 - c_2\log(\varepsilon)) \tag{4.10}$$

solving for c_1

$$c_1 = \frac{\log|x + y\sqrt{A}| + \log|x - y\sqrt{A}|}{2} = \frac{\log|n|}{2} \tag{4.11}$$

Let $k \in \mathbb{Z}$ minimize $|c_2 - k|$, so that $\delta = c_2 - k$ is such that $|\delta| \leq 1/2$. Then

$$L(x + y\sqrt{A}) = \frac{\log|n|}{2}(1, 1) + kL(\varepsilon) + \delta L(\varepsilon) \tag{4.12}$$

where $kL(\varepsilon) = L(\varepsilon^k)$. Subtract it from both sides, use the property of the logarithm and define $\bar{x} + \bar{y}\sqrt{A} = \varepsilon^{-k}(x + y\sqrt{A})$, which has integer coefficients, because we already know that it solves $x^2 - Ay^2 = n$ from Theorem 1.

$$L(\bar{x} + \bar{y}\sqrt{A}) = \frac{\log|n|}{2}(1, 1) + \delta L(\varepsilon) \tag{4.13}$$

Again, in coordinates

$$(\log|\bar{x} + \bar{y}\sqrt{A}|, \log|\bar{x} - \bar{y}\sqrt{A}|) = \left(\frac{\log|n|}{2} + \delta\log(\varepsilon), \frac{\log|n|}{2} - \delta\log(\varepsilon)\right) \tag{4.14}$$

Suppose $\delta \geq 0$ without loss of generality. Then

$$\begin{aligned}\log|\bar{x} + \bar{y}\sqrt{A}| &= \frac{\log|n|}{2} + \delta\log(\varepsilon) \leq \frac{\log|n|}{2} + \frac{\log(\varepsilon)}{2} = \log\sqrt{|n|\varepsilon} \\ \log|\bar{x} - \bar{y}\sqrt{A}| &= \frac{\log|n|}{2} - \delta\log(\varepsilon) \leq \frac{\log|n|}{2} = \log\sqrt{|n|}\end{aligned} \tag{4.15}$$

Define

$$s = |\bar{x} + \bar{y}\sqrt{A}|, \qquad \frac{|n|}{s} = |\bar{x} - \bar{y}\sqrt{A}| \tag{4.16}$$

because the product of the two numbers makes $|x^2 - Ay^2| = |n|$. Observe that because $s > |n|/s$, then $s \geq \sqrt{|n|}$. On the other hand, the bound above sets $s \leq \sqrt{|n|\varepsilon}$. So

$$\sqrt{|n|} < s < \sqrt{|n|\varepsilon} \tag{4.17}$$

On the other hand, if $s > \sqrt{|n|}$, the function $f(s) = s + |n|/s$ is growing. From there, we have our bounds by setting $s = \sqrt{|n|\varepsilon}$

$$\begin{aligned}|\bar{x}| &= \frac{|(\bar{x} + \bar{y}\sqrt{A}) + (\bar{x} - \bar{y}\sqrt{A})|}{2} \leq \frac{|\bar{x} + \bar{y}\sqrt{A}| + |\bar{x} - \bar{y}\sqrt{A}|}{2} = \\ &= \frac{1}{2}\left(s + \frac{|n|}{s}\right) \leq \frac{1}{2}\left(\sqrt{|n|\varepsilon} + \frac{|n|}{\sqrt{|n|\varepsilon}}\right) = \frac{\sqrt{|n|}}{2}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right)\end{aligned} \tag{4.18}$$

$$\begin{aligned}|\bar{y}| &= \frac{|(\bar{x} + \bar{y}\sqrt{A}) - (\bar{x} - \bar{y}\sqrt{A})|}{2\sqrt{A}} \leq \frac{|\bar{x} + \bar{y}\sqrt{A}| + |\bar{x} - \bar{y}\sqrt{A}|}{2\sqrt{A}} = \\ &= \frac{1}{2\sqrt{A}}\left(s + \frac{|n|}{s}\right) \leq \frac{1}{2\sqrt{A}}\left(\sqrt{|n|\varepsilon} + \frac{|n|}{\sqrt{|n|\varepsilon}}\right) = \frac{\sqrt{|n|}}{2\sqrt{A}}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right)\end{aligned} \tag{4.19}$$

Finally, if $n > 0$, we have that $\bar{x}+\bar{y}\sqrt{A}$ and $\bar{x}-\bar{y}\sqrt{A}$ are both positive, because their product is $\bar{x}^2-A\bar{y}^2 = n > 0$. So, $\bar{x} + \bar{y}\sqrt{A} = s$ and $\bar{x} - \bar{y}\sqrt{A} = n/s$. This means

$$|\bar{y}| = \frac{|(\bar{x} + \bar{y}\sqrt{A}) - (\bar{x} - \bar{y}\sqrt{A})|}{2\sqrt{A}} = \frac{1}{2\sqrt{A}}\left|s - \frac{n}{s}\right| \tag{4.20}$$

Moreover, $\sqrt{n} \leq s \leq \sqrt{n\varepsilon}$ and $\sqrt{n/\varepsilon} \leq n/s \leq \sqrt{n}$. So, they both lie in the interval $[\sqrt{n/\varepsilon}, \sqrt{n\varepsilon}]$. So, at most their difference is the length of that interval

$$|\bar{y}| = \frac{1}{2\sqrt{A}}\left|s - \frac{n}{s}\right| \leq \frac{\sqrt{n}}{2\sqrt{A}}\left(\sqrt{\varepsilon} - \frac{1}{\sqrt{\varepsilon}}\right) < \frac{\sqrt{n\varepsilon}}{2\sqrt{A}} \tag{4.21}$$

This terminates the proof. *End of proof.*

Notice that if $n > 0$, the bounds satisfy the generalized Pell's equation. Namely if

$$\bar{x} = \frac{\sqrt{n}}{2}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right), \qquad \bar{y} = \frac{\sqrt{n}}{2\sqrt{A}}\left(\sqrt{\varepsilon} - \frac{1}{\sqrt{\varepsilon}}\right) \tag{4.22}$$

then

$$\bar{x}^2 - A\bar{y}^2 = \frac{n}{4}\left[\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right)^2 - \left(\sqrt{\varepsilon} - \frac{1}{\sqrt{\varepsilon}}\right)^2\right] = n \tag{4.23}$$

Now, in general $\bar{x}, \bar{y}$ are not integers. But, sometimes, they are, and they also could be the only solution well within the bounds. This can be useful.

Corollary 1: *For a generalized Pell equation $x^2Ay^2 = n$ with $n \neq 0$, there is a finite set of solutions U such that every solution is obtained multiplying U by Pell's orbit.*

Proof: By Theorem 26, each solution is a Pell's orbit multiple of some element inside a set U, which we defined as

$$U := \{\bar{x} + \bar{y}\sqrt{A} \,|\, \bar{x}^2 - A\bar{y}^2 = n\} \tag{4.24}$$

with

$$|\bar{x}| \leq \frac{\sqrt{|n|}}{2}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right), \qquad |\bar{y}| \leq \frac{\sqrt{|n|}}{2\sqrt{A}}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right) \tag{4.25}$$

Because bounds are finite numbers, there is only a finite choice of $\bar{x}, \bar{y}$, hence U is finite. Specifically, oserve that, trivially, if U is empty, there is no solution to the generalized Pell's equation. *End of proof.*

4.1 Complete Pell's algorithm

Here is the full algorithm needed to find the complete set of solutions of the generalized Pell's equation

$$x^2 - Ay^2 = n \tag{4.26}$$

If $n = 0$, and A is not a perfect square, there is no solution. If A is a perfect square, LHS factorizes, and solutions are finite and described by a linear Diophantine equation. Focus only on the case $n \neq 0$ and A not a perfect square.

Algorithm for $n = 1$

- *Eliminate sign degeneration by focusing only on positive integer solutions. Remember that for every* (x, y) *solution,* $(\pm x, \pm y)$ *are also solutions.*
- *Calculate the continued fraction of* $\sqrt{A} = [a_0, \overline{a_1, \ldots, a_{l-1}, 2a_0}]$. *Call* p_n/q_n *the* n*-th convergent, and define*

$$(x_1, y_1) = \begin{cases} (p_{l-1}, q_{l-1}), & \text{if } l \text{ is even} \\ (p_{2l-1}, q_{2l-1}), & \text{if } l \text{ is odd} \end{cases} \tag{4.27}$$

- *Set* $(x_1, y_1) := (x_0, y_0)$*, and apply recursively the following linear transformation*

$$x_{k+1} = x_0 x_k + y_0 y_k A, \tag{4.28}$$

$$y_{k+1} = y_0 x_k + x_0 y_k \tag{4.29}$$

  The set (x_k, y_k) *is the complete set of solutions.*

- *Add back the sign degeneration (optionally):* $(x_k, y_k) \to (\pm x_k, \pm y_k)$.

Algorithm for $n \neq 1$

- *Eliminate sign degeneration by focusing only on positive integer solutions. Remember that for every* (x, y) *solution,* $(\pm x, \pm y)$ *are also solutions.*
- *Calculate the continued fraction of* $\sqrt{A} = [a_0, \overline{a_1, \ldots, a_{l-1}, 2a_0}]$. *Call* p_n/q_n *the* n*-th convergent, and define*

$$(x_0, y_0) = \begin{cases} (p_{l-1}, q_{l-1}), & \text{if } l \text{ is even} \\ (p_{2l-1}, q_{2l-1}), & \text{if } l \text{ is odd} \end{cases} \tag{4.30}$$

  Introduce $\varepsilon = x_0 + y_0\sqrt{A}$*, called the* fundamental unit.

- *Calculate the following bounds*

$$\text{if } n < 0, \qquad \bar{x} \leq \frac{\sqrt{|n|}}{2}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right), \qquad \bar{y} \leq \frac{\sqrt{|n|}}{2\sqrt{A}}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right) \tag{4.31}$$

$$\text{if } n > 0, \qquad \bar{x} \leq \frac{\sqrt{|n|}}{2}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right), \qquad \bar{y} \leq \frac{\sqrt{|n|}}{2\sqrt{A}}\left(\sqrt{\varepsilon} - \frac{1}{\sqrt{\varepsilon}}\right) \tag{4.32}$$

- *Brute force all the positive integer solutions to* $x^2 - Ay^2 = n$*, with* $x < \bar{x}$*, and* $y < \bar{y}$*. Call* U *the finite set of solutions found this way. If* U *is empty,* $x^2 - Ay^2 = n$ *has no integer solutions.*
- *Choose any solution* $(x, y) \in U$*, set* $(x_1, y_1) := (x, y)$*, and apply recursively the following linear transformation*

$$x_{k+1} = x_0 x_k + y_0 y_k A, \tag{4.33}$$

$$y_{k+1} = y_0 x_k + x_0 y_k \tag{4.34}$$

  The set (x_k, y_k) *is an infinite set of solutions.*

- *Remove* (x, y) *from* U*. If any other* (x_k, y_k) *happens to be in* U*, remove it.*
- *Repeat the previous two points until* U *is empty.*

- *Add back the sign degeneration (optionally):* $(x_k, y_k) \to (\pm x_k, \pm y_k)$.

Let us make an example. Set $A = 6$, $n = 3$. The equation is

$$x^2 - 6y^2 = 3 \tag{4.35}$$

Evaluate the continued fraction: $\sqrt{6} = [2, \overline{2, 4}]$. The period is equal to 2, so the fundamental solution is generated from the second convergent $x_0/y_0 = [2, 2] = 5/2$. Hence $\varepsilon = 5 + 2\sqrt{6} = (\sqrt{3} + \sqrt{2})^2$. In fact $(5, 2)$ solves $x^2 - 6y^2 = 1$.

Calculate the bounds

$$\bar{x} = \frac{\sqrt{3}}{2}\left(\sqrt{3} + \sqrt{2} + \frac{1}{\sqrt{3} + \sqrt{2}}\right) = 3, \qquad \bar{y} = \frac{\sqrt{3}}{2\sqrt{6}}\left(\sqrt{3} + \sqrt{2} - \frac{1}{\sqrt{3} + \sqrt{2}}\right) = 1 \tag{4.36}$$

so, we brute force all possible values $(x, y) = (1, 1), (2, 1), (3, 1)$ into $x^2 - 6y^2 = 3$. $(3, 1)$ works. Apply the linear transformation

$$x_{k+1} = 5x_k + 12y_k, \tag{4.37}$$

$$y_{k+1} = 2x_k + 5y_k \tag{4.38}$$

So, the complete set of positive integer solutions to $x^2 - 6y^2 = 3$ is

$$(3, 1) \to (27, 11) \to (267, 109) \to (2643, 1079) \to \ldots \tag{4.39}$$

And, by adding back sign degeneration

$$(\pm 3, \pm 1) \to (\pm 27, \pm 11) \to (\pm 267, \pm 109) \to (\pm 2643, \pm 1079) \to \ldots \tag{4.40}$$