CONTINUED FRACTIONS AND PELL'S EQUATION

SEUNG HYUN YANG

Abstract. In this REU paper, I will use some important characteristics of continued fractions to give the complete set of solutions to Pell's equation. I would like to thank my mentor Avan for introducing and guiding me through this extremely interesting material. I would like to cite Steuding's detailed but slightly flawed book as the main source of learning and Andreescu and Andrica's book as an inspiration for numerous fun experiments I have made this summer.

Contents

1. Continued Fractions 1
2. Solution to Pell's Equation 9
References 12

1. Continued Fractions

This rather long section gives several crucial tools for solving Pell's equation.

Definition 1.1. Let a_0, a_1, a_2, …, a_m be real numbers. Then,

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cfrac{1}{\cdots + \cfrac{1}{a_{m-1} + \cfrac{1}{a_m}}}}}}}$$

is called a **finite continued fraction** and is denoted by $[a_0,a_1,a_2,\ldots,a_m]$. If the chain of fractions does not stop, then it is called an **infinite continued fraction**.

Remark 1.2. From now on, we will use a_n as it is defined here.

Definitions 1.3. (a) For n $\leq$ m, $[a_0,a_1,\ldots,a_n]$ is called n**th convergent** to $[a_0,a_1,a_2, \ldots,a_m]$. (b) Define two sequences of real numbers, (p_n) and (q_n), recursively as follows: (1) $p_{-1} = 1$, $p_0 = a_0$, and $p_n = a_n p_{n-1} + p_{n-2}$; and (2) $q_{-1} = 0$, $q_0 = 1$, and $q_n = a_n q_{n-1} + q_{n-2}$.

Remark 1.4. From now on, we will use p_n and q_n as they are defined here.

Date: AUGUST 22, 2008.

Theorem 1.5. *Let [a_0,a_1,a_2,. . . ,a_m] be a continued fraction. Then, for 0 $\leq n \leq$ m, $\frac{p_n}{q_n}$ = [a_0,a_1,a_2,. . . ,a_n].*

Proof. We proceed by induction. For n = 1,

$$\begin{aligned}[a_0, a_1] &= a_0 + \frac{1}{a_1} \\ &= \frac{a_1a_0 + 1}{a_1} \\ &= \frac{a_1p_0 + p_{-1}}{a_1q_0 + q_{-1}} \\ &= \frac{p_1}{q_1}\end{aligned}$$

as desired. Now, suppose the theorem holds for n.

$$\begin{aligned}[a_0, \ldots, a_{n+1}] &= [a_0, \ldots, a_{n-1}, a_n + \frac{1}{a_{n+1}}] \\ &= \frac{(a_n + \frac{1}{a_{n+1}})p_{n-1} + p_{n-2}}{(a_n + \frac{1}{a_{n+1}})q_{n-1} + q_{n-2}} \\ &= \frac{(a_{n+1}a_n + 1)p_{n-1} + a_{n+1}p_{n-2}}{(a_{n+1}a_n + 1)q_{n-1} + a_{n+1}q_{n-2}} \\ &= \frac{a_{n+1}a_np_{n-1} + a_{n+1}p_{n-2} + p_{n-1}}{a_{n+1}a_nq_{n-1} + a_{n+1}q_{n-2} + q_{n-1}} \\ &= \frac{a_{n+1}(a_np_{n-1} + p_{n-2}) + p_{n-1}}{a_{n+1}(a_nq_{n-1} + q_{n-2}) + q_{n-1}} \\ &= \frac{a_{n+1}p_n + p_{n-1}}{a_{n+1}q_n + q_{n-1}} \\ &= \frac{p_{n+1}}{q_{n+1}}\end{aligned}$$

as desired. □

Remark 1.6. As a result of this theorem, we will refer to $\frac{p_n}{q_n}$ as the nth convergent.

Theorem 1.7. *Suppose that a_0 is an integer, that a_n is a positive integer for each 1 $\leq n \leq m-1$, and that 1 $\leq a_m$. Then, p_n and q_n are (a) integers and (b) coprime for each 1 $\leq n \leq m - 1$.*

Proof. (a) p_{-1}, q_{-1}, p_0, and q_0 are integers by definition. a_n's are also integers for $0 \leq n \leq m - 1$ by the given. Since p_n and q_n $(1 \leq n \leq m - 1)$ are defined as combinations of multiplication and subtraction of said variables (which are integers), they must be integers as well.

(b) We use the following useful lemma:

Lemma 1.8. *For 1 $\leq n \leq m$, following two relations hold:*
(1) $p_nq_{n-1} - p_{n-1}q_n = (-1)^{n-1}$; and
(2) $p_nq_{n-2} - p_{n-2}q_n = (-1)^na_n$.

Proof. (1)

$$\begin{aligned} p_nq_{n-1} - p_{n-1}q_n &= (a_np_{n-1} + p_{n-2})q_{n-1} - p_{n-1}(a_nq_{n-1} + q_{n-2}) \\ &= p_{n-2}q_{n-1} + a_np_{n-1}q_{n-1} - a_np_{n-1}q_{n-1} - p_{n-1}q_{n-2} \\ &= (-1)(p_{n-1}q_{n-2} - p_{n-2}q_{n-1}) \\ &= \ldots \\ &= (-1)^n(p_0q_{-1} - p_{-1}q_0) \\ &= (-1)^n(-1) \\ &= (-1)^{n-1} \end{aligned}$$

as desired.

(2)

$$\begin{aligned} p_nq_{n-2} - p_{n-2}q_n &= (a_np_{n-1} + p_{n-2})q_{n-2} - (a_nq_{n-1} + q_{n-2})p_{n-2} \\ &= a_np_{n-1}q_{n-2} + p_{n-2}q_{n-2} - p_{n-2}q_{n-2} - a_np_{n-2}q_{n-1} \\ &= a_n(p_{n-1}q_{n-2} - p_{n-2}q_{n-1}) \\ &= (1)^{n-2}a_n \\ &= (-1)^na_n \end{aligned}$$

as desired. □

To complete the proof of Theorem 1.7, consider (1) of Lemma 1.8. Since there is a linear combination of p_n and q_n that is equal to ±1, by elementary proposition from number theory, we conclude they are coprime as greatest common divisor between them is at most (equal to) 1. □

We now move on to use continued fractions to approximate real numbers.

Definition 1.9. Let α be a real number. For $n = 0,1,2,\ldots$, define a recursive algorithm as follows: $\alpha_0 = \alpha$, $a_n = \lfloor \alpha_n \rfloor$, and $\alpha_n = a_n + \frac{1}{\alpha_{n+1}}$.

Remark 1.10. Here, $\lfloor \ \rfloor$ is used as a floor function. That means, by the last equation, α_n is positive for all positive n. Observe that $\frac{1}{\alpha_{n+1}} = \alpha_n$ - $\lfloor \alpha_n \rfloor$. Thus, the left side of the equation must be less than 1, and therefore $\alpha_{n+1} > 1$. Then, by definition, a_{n+1} is a positive integer. Thus, we may conclude that $\alpha_n, a_n \geq 1$ for $n \geq 1$.

Observe also that, given positive m, $\alpha = [a_0,a_1,\ldots,a_{m-1},\alpha_m]$. It is called the **$m$th continued fraction of α**.

Remark 1.11. Observe that, if α is rational, then the algorithm above is equivalent to Euclidean algorithm, with α_n, when reduced to a fraction of coprimes, consisting of nth remainder as numerator and $n+1$th remainder as denominator. That means, the continued fraction of a rational number is finite. On the other hand, if α is irrational, then the continued fraction must be infinite simply because any finite continued fraction is rational (and therefore cannot be equal to an irrational number).

Theorem 1.12. *Let α be irrational and $\frac{p_n}{q_n}$ be a convergent to its continued fraction. Then,*

$$\alpha - \frac{p_n}{q_n} = \frac{1}{q_n(\alpha_{n+1}q_n + q_{n-1})}. \tag{1.13}$$

Proof. Let n be a positive number. Then, $\alpha = [a_0,a_1,\ldots,a_n,\alpha_{n+1}]$. Then,

$$\begin{aligned}\alpha - \frac{p_n}{q_n} &= \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}} - \frac{p_n}{q_n} \\ &= \frac{p_{n-1}q_n + \alpha_{n+1}p_nq_n - \alpha_{n+1}p_nq_n - p_nq_{n-1}}{q_n(\alpha_{n+1}q_n + q_{n-1})} \\ &= \frac{p_{n-1}q_n - p_nq_{n-1}}{q_n(\alpha_{n+1}q_n + q_{n-1})} \\ &= \frac{(-1)(p_nq_{n-1} - q_np_{n-1})}{q_n(\alpha_{n+1}q_n + q_{n-1})} \\ &= \frac{(-1)^n}{q_n(\alpha_{n+1}q_n + q_{n-1})}\end{aligned}$$

by the Lemma 1.8. □

Remark 1.14. Observe, for each positive n, $a_n \leq \alpha_n$ by a property of floor function. Also, since q_{-1}, q_0, and a_n $(n > 0)$ are positive integers, the same must be true for q_n $(n > 0)$ by definition. Then,

$$\begin{aligned}\left|\alpha - \frac{p_n}{q_n}\right| &= \frac{1}{q_n(\alpha_{n+1}q_n + q_{n-1})} \\ &< \frac{1}{q_n(a_{n+1}q_n + q_{n-1})} \\ &= \frac{1}{q_nq_{n+1}}.\end{aligned}$$

By definition, $q_n = a_nq_{n-1} + q_{n-2}$. Since $1 \leq a_n$ and $q_{n-2} > 0$, we conclude that q_n is strictly increasing as n increases. Then, we may conclude that the continued fraction of a number converges to that number by the inequality just given here. In other words,

$$\alpha = \lim_{n\to\infty} \tfrac{p_n}{q_n} = [a_0,a_1,a_2,\ldots].$$

Corollary 1.15. *Let α be a real number with convergent $\frac{p_n}{q_n}$. Then,*

$$|q_n\alpha - p_n| < |q_{n-1}\alpha - p_{n-1}|.$$

Proof. By Theorem 1.12,

$$\begin{aligned}\left|\alpha - \frac{p_n}{q_n}\right| &= \frac{1}{q_n(\alpha_{n+1}q_n + q_{n-1})} \\ \Rightarrow |q_n\alpha - p_n| &= \frac{1}{q_n\alpha_{n+1} + q_{n-1}}\end{aligned}$$

Similarly,

$$|q_{n-1}\alpha - p_{n-1}| = \frac{1}{q_{n-1}\alpha_n + q_{n-2}}$$

It now suffices to prove the following inequality:

$$\begin{aligned}
\frac{1}{q_n\alpha_{n+1}+q_{n-1}} &< \frac{1}{q_{n-1}\alpha_n+q_{n-2}} \\
&= \frac{1}{q_{n-1}(a_n+\frac{1}{\alpha_{n+1}})+q_{n-2}} \\
&= \frac{1}{q_{n-1}a_n+\frac{q_{n-1}}{\alpha_{n+1}}+q_{n-2}} \\
&= \frac{\alpha_{n+1}}{q_{n-1}a_n\alpha_{n+1}+q_{n-1}+q_{n-2}\alpha_{n+1}} \\
\Leftrightarrow q_{n-1}a_n\alpha_{n+1}+q_{n-1}+\alpha_{n+1}q_{n-2} &< q_n\alpha_{n+1}^2+q_{n-1}\alpha_{n+1} \\
\Leftrightarrow \alpha_{n+1}(q_{n-1}a_n+q_{n-2})+q_{n-1} &< q_n\alpha_{n+1}^2+q_{n-1}\alpha_{n+1} \\
\Leftrightarrow q_n\alpha_{n+1}+q_{n-1} &< \alpha_{n+1}(q_n\alpha_{n+1}+q_{n-1}) \\
\Leftrightarrow 1 &< \alpha_{n+1}
\end{aligned}$$

which is true by Remark 1.10. □

We are now ready to move onto two extremely beautiful and important approximation theorems involving convergents.

Theorem 1.16. ***(The Law of Best Approximations)*** *Let α be a real number with convergent $\frac{p_n}{q_n}$ and $n \geq 2$. If p,q are integers such that $0 < q \leq q_n$ and $\frac{p}{q} \neq \frac{p_n}{q_n}$, then*

$$|q_n\alpha - p_n| < |q\alpha - p|.$$

Moreover, a reduced fraction $\frac{p'}{q'}$ with $q' \geq q_2$ that satisfies the latter inequality is a convergent.

Proof. By Corollary 1.15, we have already proven the case in which $\frac{p}{q}$ is a convergent. Suppose $q = q_n$. Then, $p \neq p_n$.

$$\begin{aligned}
\left|\frac{p}{q}-\frac{p_n}{q_n}\right| &= \frac{|p-p_n|}{q_n} \\
&\geq \frac{1}{q_n}. \\
\left|\alpha-\frac{p_n}{q_n}\right| &< \frac{1}{q_nq_{n+1}} \\
&< \frac{1}{2q_n}
\end{aligned}$$

because $q_{n+1} \geq 3$ if $n \geq 2$ (q_n is strictly increasing for $n \geq 1$ and $q_1 \geq q_0 = 1$). By the Triangle Inequality, we have

$$\begin{aligned}
\left|\alpha-\frac{p}{q}\right| &\geq \left|\frac{p}{q}-\frac{p_n}{q_n}\right| - \left|\alpha-\frac{p_n}{q_n}\right| \\
&> \frac{1}{q_n}-\frac{1}{2q_n} \\
&= \frac{1}{2q_n} \\
&> \left|\alpha-\frac{p_n}{q_n}\right|
\end{aligned}$$

Multiplying both sides by $q = q_n$, we obtain the desired inequality.

Now suppose $0 < q < q_n$. We may set up following system of two equations with two variables X and Y:

$$p_n X + p_{n-1} Y = p \tag{1.17}$$
$$q_n X + q_{n-1} Y = q \tag{1.18}$$

A series of basic manipulations from high school mathematics yields the following unique solution (x,y):

$$\begin{aligned} x &= \frac{pq_{n-1} - qp_{n-1}}{p_n q_{n-1} - p_{n-1} q_n} \\ y &= \frac{pq_n - qp_n}{p_n q_{n-1} - p_{n-1} q_n} \end{aligned}$$

By Lemma 1.8, the denominators reduce to ± 1:

$$\begin{aligned} x &= \pm(pq_{n-1} - qp_{n-1}) \\ y &= \pm(pq_n - qp_n) \end{aligned}$$

Therefore, x and y are nonzero (otherwise $\frac{p}{q}$ is a convergent) integers. By the equation 1.18, since $q_n > q$, we conclude that x and y have opposite signs. By Lemma 1.8, $\alpha - \frac{p_n}{q_n}$ alternates signs. That is, $\alpha - \frac{p_n}{q_n}$ and $\alpha - \frac{p_{n-1}}{q_{n-1}}$ have opposite signs. This implies that $q_n\alpha - p_n$ and $q_{n-1}\alpha - p_{n-1}$ have opposite signs as well. Therefore, $x(q_n\alpha - p_n)$ and $y(q_{n-1}\alpha - p_{n-1})$ have the same sign. Thus,

$$\begin{aligned} q\alpha - p &= (q_n x + q_{n-1} y)\alpha - (p_n x + p_{n-1} y) \\ &= x(q_n\alpha - p_n) + y(q_{n-1}\alpha - p_{n-1}) \\ \Rightarrow |q\alpha - p| &= |x(q_n\alpha - p_n| + |y(q_{n-1}\alpha - p_{n-1})| \\ \Rightarrow |q\alpha - p| &> |q_{n-1}\alpha - p_{n-1}| \\ &> |q_n\alpha - p_n| \end{aligned}$$

as desired. In particular, this inequality holds for $q_{n-1} < q < q_n$, which, by Induction Principle, implies only convergents satisfy the said inequality. □

Theorem 1.19. *(a) Given any two consecutive convergents to a real number α, there is at least one satisfying*

$$\left|\alpha - \tfrac{p}{q}\right| < \tfrac{1}{2q^2}$$

(b) Any reduced fraction that satisfies the above inequality is a convergent.

Proof. (a) By Lemma 1.8, given any two consecutive convergents, exactly one is greater than or equal to α, and the other is less than or equal to α. Thus,

$$\left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| = \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right| + \left|\frac{p_n}{q_n} - \alpha\right| \tag{1.20}$$

Now, suppose the said inequality is not true for some consecutive convergents $\frac{p_n}{q_n}$ and $\frac{p_{n+1}}{q_{n+1}}$. By Lemma 1.8,

$$\begin{aligned}
\frac{1}{q_nq_{n+1}} &= \left|\frac{p_{n+1}q_n - p_nq_{n+1}}{q_nq_{n+1}}\right| = \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| \\
&= \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right| + \left|\frac{p_n}{q_n} - \alpha\right| \\
&\geq \frac{1}{2q_{n+1}^2} + \frac{1}{2q_n^2} \\
\Rightarrow \frac{1}{q_nq_{n+1}} &\geq \frac{1}{2q_{n+1}^2} + \frac{1}{2q_n^2} = \frac{q_n^2 + q_{n+1}^2}{2q_n^2q_{n+1}^2} \\
\Rightarrow 2q_nq_{n+1} &\geq q_n^2 + q_{n+1}^2 \\
\Rightarrow 0 &\geq (q_n - q_{n+1})^2
\end{aligned}$$

Because q_n is strictly increasing for positive n, this may be true only if $n = 0$ and $q_1 = q_0 = a_1 = 1$. Thus, by contradiction, (a) is true for all positive n, and we only need to check for the case $n = 0$ (the two consecutive convergents being $\frac{p_0}{q_0}$ and $\frac{p_1}{q_1}$):

$$\begin{aligned}
0 < \frac{p_1}{q_1} - \alpha \;&=\; a_1a_0 + 1 - \alpha \;=\; a_0 + 1 - \alpha \;=\; a_0 + 1 - [a_0, 1, a_2, a_3, \dots] \\
&< \; 1 - \cfrac{1}{1 + \cfrac{1}{a_2}} \;=\; 1 - \frac{a_2}{a_2 + 1} \;\leq\; \frac{1}{2}
\end{aligned}$$

which satisfies the statement of the theorem.

(b) Suppose $\frac{p}{q}$ satisfies the said inequality. By the law of best approximations, it suffices to show $\frac{p}{q}$ is a best approximation to α.

Let $\frac{P}{Q}$ be such that $\frac{P}{Q} \neq \frac{p}{q}$ and $|Q\alpha - P| \leq |q\alpha - p| = q\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q}$. Then,

$$\begin{aligned}
\frac{1}{qQ} &\leq \frac{|pQ - Pq|}{qQ} \\
&= \left|\frac{p}{q} - \frac{P}{Q}\right| \\
&\leq \left|\alpha - \frac{p}{q}\right| + \left|\frac{P}{Q} - \alpha\right| \\
&< \frac{1}{2q^2} + \frac{1}{2qQ} \\
&= \frac{q + Q}{2q^2Q} \\
\Rightarrow 1 &< \frac{q + Q}{2q} \\
\Rightarrow 2q &< q + Q \\
\Rightarrow q &< Q
\end{aligned}$$

Thus, $\frac{p}{q}$ is a best approximation. □

We now move on to the discussion of quadratic irrationals. We first present two well-known theorems, Lagrange's and Galois', without proof. Both may be proved with what we have demonstrated so far (plus some basic knowledge of polynomials), and have quite long proofs that would only make this already-too-long paper even longer.

Definitions 1.21. Let α be an irrational number. α is called a **quadratic irrational** if it is a root of integer polynomial of degree two. The other root β is called a **conjugate** of α.

Definitions 1.22. Let $[a_0,a_1,a_2,\ldots]$ be a continued fraction such that $a_n = a_{n+l}$ for all sufficiently large n and a fixed positive integer l. Then, it is **periodic** and l is called a **period**.

Theorem 1.23. ***(Lagrange's Theorem)*** *An irrational number is quadratic irrational if and only if its continued fraction is periodic.*

Definition 1.24. Let α be a quadratic irrational and β be its conjugate. Then, α is **reduced** if $\alpha > 1$ and -1 $< \beta < 0$.

Theorem 1.25. ***(Galois' Theorem)*** *Let α be irrational. Then, α is purely periodic if and only if α is reduced. If $\alpha = [\overline{a_0, a_1, a_2, \ldots, a_{l-1}}]$ and β is its conjugate, then -$\frac{1}{\beta} = [\overline{a_{l-1}, \ldots, a_2, a_1}]$.*

We finally conclude this section with the following simple theorem:

Theorem 1.26. *Let d be a non-square natural number. Then, there exist integers $a_1, a_2, \ldots, a_n$ such that $[\lfloor\sqrt{d}\rfloor,\overline{a_1, a_2, \ldots, a_2, a_1, 2\lfloor\sqrt{d}\rfloor}]$ is the continued fraction of $\sqrt{d}$. Also, $a_1, a_2, \ldots, a_2, a_1$ is a palindrome.*

Proof. Observe that -1 $< \alpha = \lfloor\sqrt{d}\rfloor$ - $\sqrt{d} < 0$ and $1 < \beta = \lfloor\sqrt{d}\rfloor + \sqrt{d}$.

$$\begin{aligned}(x-\alpha)(x-\beta) &= x^2 - (\alpha+\beta)x + \alpha\beta \\ &= x^2 + 2\lfloor\sqrt{d}\rfloor x + \lfloor\sqrt{d}\rfloor^2 - d\end{aligned}$$

which is an integer polynomial of degree two. Thus, α and β are quadratic irrationals and conjugates of each other. More, β is reduced. By Galois' Theorem, β is purely periodic.

$$\begin{aligned}\beta = \sqrt{d} + \lfloor\sqrt{d}\rfloor &= [\overline{2\lfloor\sqrt{d}\rfloor, a_1, a_2, \ldots, a_{l-1}}] \\ &= [2\lfloor\sqrt{d}\rfloor, \overline{a_1, a_2, \ldots, a_{l-1}, 2\lfloor\sqrt{d}\rfloor}] \\ \Rightarrow \sqrt{d} &= [\lfloor\sqrt{d}\rfloor, \overline{a_1, a_2, \ldots, a_{l-1}, 2\lfloor\sqrt{d}\rfloor}]\end{aligned}$$

On the other hand, observe:

$$\begin{aligned}\frac{1}{\sqrt{d} - \lfloor\sqrt{d}\rfloor} &= \frac{1}{-\sqrt{d} + \lfloor\sqrt{d}\rfloor} \\ &= [\overline{a_{l-1}, \ldots, a_2, a_1, 2\lfloor\sqrt{d}\rfloor}]\end{aligned}$$

by Galois' Theorem. That means,

$$\begin{aligned}\sqrt{d} &= [\lfloor\sqrt{d}\rfloor, \frac{1}{\sqrt{d} - \lfloor\sqrt{d}\rfloor}] \\ &= [\lfloor\sqrt{d}\rfloor, \overline{a_{l-1}, \ldots, a_2, a_1, 2\lfloor\sqrt{d}\rfloor}]\end{aligned}$$

Thus, we have obtained the desired palindrome. □

2. Solution to Pell's Equation

Definition 2.1. Let d be a natural number. Then, **Pell's equation** is $X^2 - dY^2 = 1$.

Remarks 2.2. The case in which d is a perfect square is trivial. Let $d = m^2$.

$$\begin{aligned} X^2 - dY^2 &= (X-mY)(X+mY) = 1 \\ \Rightarrow X - mY &= \pm 1 \\ X + mY &= \pm 1 \end{aligned}$$

Solving this linear system of equations yields the only solutions: ($\pm$1,0). Note that these are solutions regardless of d.

Observe that if (x,y) is a solution, then ($-x$,$-y$) and ($\pm x$,$\mp y$) are also solutions. Thus, it suffices to consider only positive solutions.

For convenience, if (x,y) is a solution, then we call $x + y\sqrt{d}$ a solution as well.

Theorem 2.3. *(a) Let d not be a perfect square and $\alpha = \sqrt{d}$. Note that α_n is as defined in Definition 1.9. For nonnegative n, there exist integers Q_n and P_n such that $\alpha_n = \frac{P_n+\sqrt{d}}{Q_n}$ and that $d - P_n^2 \equiv 0$ (mod Q_n). (b) For each $n \geq 2$, we have:*

$$p_{n-1}^2 - dq_{n-1}^2 = (-1)^n Q_n \tag{2.4}$$

Proof. (a) We proceed by induction. For $n = 0$, $\alpha_0 = \sqrt{d}$, $Q_0 = 1$, and $P_0 = 0$. For $n = 1$,

$$\begin{aligned} \alpha_1 &= \frac{1}{\sqrt{d} - \lfloor \sqrt{d} \rfloor} \\ &= \frac{\sqrt{d} - \lfloor \sqrt{d} \rfloor}{d - \lfloor \sqrt{d} \rfloor^2} \\ \Rightarrow Q_1 &= d - \lfloor d \rfloor^2 \text{ and} \\ P_1 &= \lfloor \sqrt{d} \rfloor \end{aligned}$$

Now assume the statement for n.

$$\begin{aligned} \alpha_{n+1} &= \frac{1}{\alpha_n - a_n} \\ &= \frac{1}{\frac{P_n+\sqrt{d}}{Q_n} - a_n} \\ &= \frac{Q_n}{P_n + \sqrt{d} - a_n Q_n} \\ &= \frac{Q_n(P_n - a_n Q_n - \sqrt{d})}{(P_n - a_n Q_n)^2 - d} \\ &:= \frac{P_{n+1} + \sqrt{d}}{Q_{n+1}} \\ \Rightarrow P_{n+1} &= a_n Q_n - P_n, \text{ which is integral, and} \\ Q_{n+1} &= \frac{d - (P_n - a_n Q_n)^2}{Q_n} \\ &= \frac{d - P_n^2}{Q_n} + 2a_n P_n - a_n^2 Q_n, \end{aligned}$$

which is integral because $d - P_n^2 \equiv 0 \pmod{Q_n}$ by assumption. Finally,

$$\begin{aligned} Q_n &= \frac{d-(P_n - a_nQ_n)^2}{Q_{n+1}} \\ &= \frac{d - P_{n+1}}{Q_{n+1}} \\ \Rightarrow d - P_{n+1}^2 &\equiv 0 \pmod{Q_n} \end{aligned}$$

as desired.

(b)

$$\begin{aligned} \sqrt{d} &= \frac{\alpha_n p_{n-1} + p_{n-2}}{\alpha_n q_{n-1} + q_{n-2}} \\ &= \frac{(P_n + \sqrt{d})p_{n-1} + p_{n-2}Q_n}{(P_n + \sqrt{d})q_{n-1} + q_{n-2}Q_n} \\ \Rightarrow \sqrt{d}((P_n + \sqrt{d})q_{n-1} + q_{n-2}Q_n) &= (P_n + \sqrt{d})p_{n-1} + p_{n-2}Q_n \\ \Rightarrow \sqrt{d}(P_nq_{n-1} + Q_nq_{n-2} + \sqrt{d}q_{n-1}) &= p_{n-1}\sqrt{d} + P_np_{n-1} + p_{n-2}Q_n \\ \Rightarrow (P_nq_{n-1} + Q_nq_{n-2})\sqrt{d} + q_{n-1}d &= p_{n-1}\sqrt{d} + P_np_{n-1} + p_{n-2}Q_n \\ \Rightarrow p_{n-1} &= P_nq_{n-1} + q_{n-2}Q_n \text{ and} \\ dq_{n-1} &= P_np_{n-1} + p_{n-2}Q_n \end{aligned}$$

Multiply the first by p_{n-1} and the second by q_{n-1}, and obtain:

$$\begin{aligned} p_{n-1}^2 - dq_{n-1}^2 &= Q_n(p_{n-1}q_{n-2} - p_{n-2}q_{n-1}) \\ p_{n-1}^2 - dq_{n-1}^2 &= (-1)^nQ_n \end{aligned}$$

by Lemma 1.8. □

Definition 2.5. Given Pell's equation with some d, let (x,y) be the positive solution with the minimum x. Such solution is called the **minimal solution**.

We finally give the complete solution to Pell's equation:

Theorem 2.6. *Let variables be as defined in the previous theorem. Let l be the minimal period of the continued fraction of $\sqrt{d}$. (a) The minimal solution to Pell's equation is:*

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

(b) All the solutions (x,y) to Pell's equation are up to sign given by the powers of the minimal solution (x_1,y_1):

$$x + y\sqrt{d} = \pm(x_1 + y_1\sqrt{d})^{\pm n} \ (n = 0,1,2,\dots)$$

Proof. (a) We first establish that all the solutions to Pell's equation $X^2 - dY^2 = 1$ must consist of convergents to $\sqrt{d}$. Let (x,y) be a solution. Observe first that:

$$\begin{aligned} \left|\sqrt{d} - \frac{x}{y}\right| &= \left|\frac{y\sqrt{d} - x}{y}\right| \\ &= \left|\frac{dy^2 - x^2}{y(y\sqrt{d} + x)}\right| \\ &= \frac{1}{y^2(\sqrt{d} + \frac{x}{y})} \\ &< \frac{1}{2y^2} \text{ (since d > 1 and x > y)} \end{aligned}$$

By Theorem 1.19, we conclude (x,y) is a convergent to $\sqrt{d}$.

We now use (2.4) to show that: (1) $Q_n \neq -1$ for every positive integer n; and (2) $Q_n = 1$ if and only if n is a multiple of l. Thus, with regard to (2.4), the minimal solution is obtained by taking the smallest positive odd multiple of l minus 1, as (a) states.

To show (1), observe that $\alpha_1 = \alpha_{n+1}$ if and only if n is a multiple of l because l is minimal. Let $Q_n = 1$. Then, by Theorem 2.3, $\alpha_n = P_n + \sqrt{d}$. Being quadratic irrational, by Lagrange's Theorem, α_n has purely periodic continued fraction expansion. By Galois' Theorem, α_n is reduced:

$$\begin{gathered} -1 < \beta_n = P_n - \sqrt{d} < 0 \\ \Rightarrow \sqrt{d} - 1 < P_n < \sqrt{d} \\ \Rightarrow P_n = \lfloor\sqrt{d}\rfloor \end{gathered}$$

Thus, $\alpha = \lfloor\sqrt{d}\rfloor + \sqrt{d}$. This implies that $\alpha_{n+1} = \alpha_1$, and n is a multiple of l.

On the other hand, if n is a multiple of l, $n = kl$, then:

$$\frac{P_{kl} + \sqrt{d}}{Q_{kl}} = \alpha_{kl} = [0, \overline{a_1, a_2, \ldots, a_{l-1}}] = \sqrt{d} - \lfloor\sqrt{d}\rfloor$$

$$\Rightarrow 1 = \frac{1}{Q_{kl}} \Rightarrow Q_{kl} = 1, \text{ as desired.}$$

For (2), suppose $Q_n = -1$. By Theorem 2.3, $\alpha_n = -P_n - \sqrt{d}$. α_n is quadratic irrational, and thus is reduced by Galois' Theorem:

$$\begin{gathered} -1 < -P_n + \sqrt{d} < 0 \text{ and } 1 < -P_n - \sqrt{d} \\ \Rightarrow \sqrt{d} < P_n < -\sqrt{d} + 1 \Rightarrow \sqrt{d} < \frac{1}{2} \end{gathered}$$

This is contradiction. Thus, $Q_n \neq 1$ for all n, as desired.

(b) We observed before that, given a solution (x,y), signs of x and/or y may be changed without contradicting the statement $x^2 - dy^2 = 1$. We then concluded that it sufficed to find positive solutions. With regard to this, observe that

$$\begin{aligned} \frac{\pm 1}{x + y\sqrt{d}} &= \frac{\pm(x - y\sqrt{d})}{x^2 - dy^2} \\ &= \pm x \mp y\sqrt{d} \ (x^2 - dy^2 = 1) \end{aligned}$$

Thus, we may change the signs of x and y only by taking powers and multiplying by -1. Thus, it now suffices to show this theorem for only positive solutions.

Let $\epsilon := x_1 + y_1\sqrt{d}$ and (x,y) be a positive solution to Pell's equation. Then, there exists n such that $\epsilon^n \leq x + y\sqrt{d} < \epsilon^{n+1}$.

Now, let $X + Y\sqrt{d} = \epsilon^{-n}(x + y\sqrt{d})$. Observe that, due to the definition of ϵ, 1 $\leq X + Y\sqrt{d}$. It suffices to show that it is equal to 1. We do this by contradiction; suppose $\epsilon > X + Y\sqrt{d} > 1$ (the first inequality is due to the definition of epsilon also). By rules of elementary algebra, we get the following conjugation to this quadratic irrational: $X - Y\sqrt{d} = \epsilon^n(x - y\sqrt{d})$. Then, $(X + Y\sqrt{d})(X - Y\sqrt{d}) = X^2 - dY^2 = \epsilon^n\epsilon^{-n}(x - y\sqrt{d})(x + y\sqrt{d}) = x^2 - dy^2 = 1$. We then have:

$$0 < \epsilon^{-n} < (X + Y\sqrt{d})^{-1} = X - Y\sqrt{d} < 1$$

$$\Rightarrow \begin{cases} 2X = (X + Y\sqrt{d}) + (X - Y\sqrt{d}) > 1 + \epsilon^{-1} > 0 \\ 2Y\sqrt{d} \ = (X + Y\sqrt{d}) - (X - Y\sqrt{d}) > 1 - 1 = 0 \end{cases}$$

Thus, (X, Y) is a positive solution with $X + Y\sqrt{d} < x_1 + y_1\sqrt{d}$. That is, $X < x_1$. This is a contradiction to the minimality of (x_1, y_1), and thus our assumption $(X + Y\sqrt{d} > 1)$ is false, as desired. □

Remarks 2.7. In fact, (a) alone suffices to give the complete set of solutions to Pell's equation. The set of (positive) solutions deduced from (a) would be:

$$(x_k, y_k) = \begin{cases} (p_{kl-1}, q_{kl-1}) & \text{if } l \text{ is even} \\ (p_{2kl-1}, q_{2kl-1}) & \text{if } l \text{ is odd} \end{cases}$$

(b) shows that the complete set of solutions to Pell's equation is infinite cyclic group generated by the minimal solution.

References

[1] Jörn Steuding. Diophantine Analysis. Chapman and Hall/CRC. 2005.

[2] Titu Andreescu and Dorin Andrica. An Introduction to Diophantine Equations. Gil Publishing House. 2002.