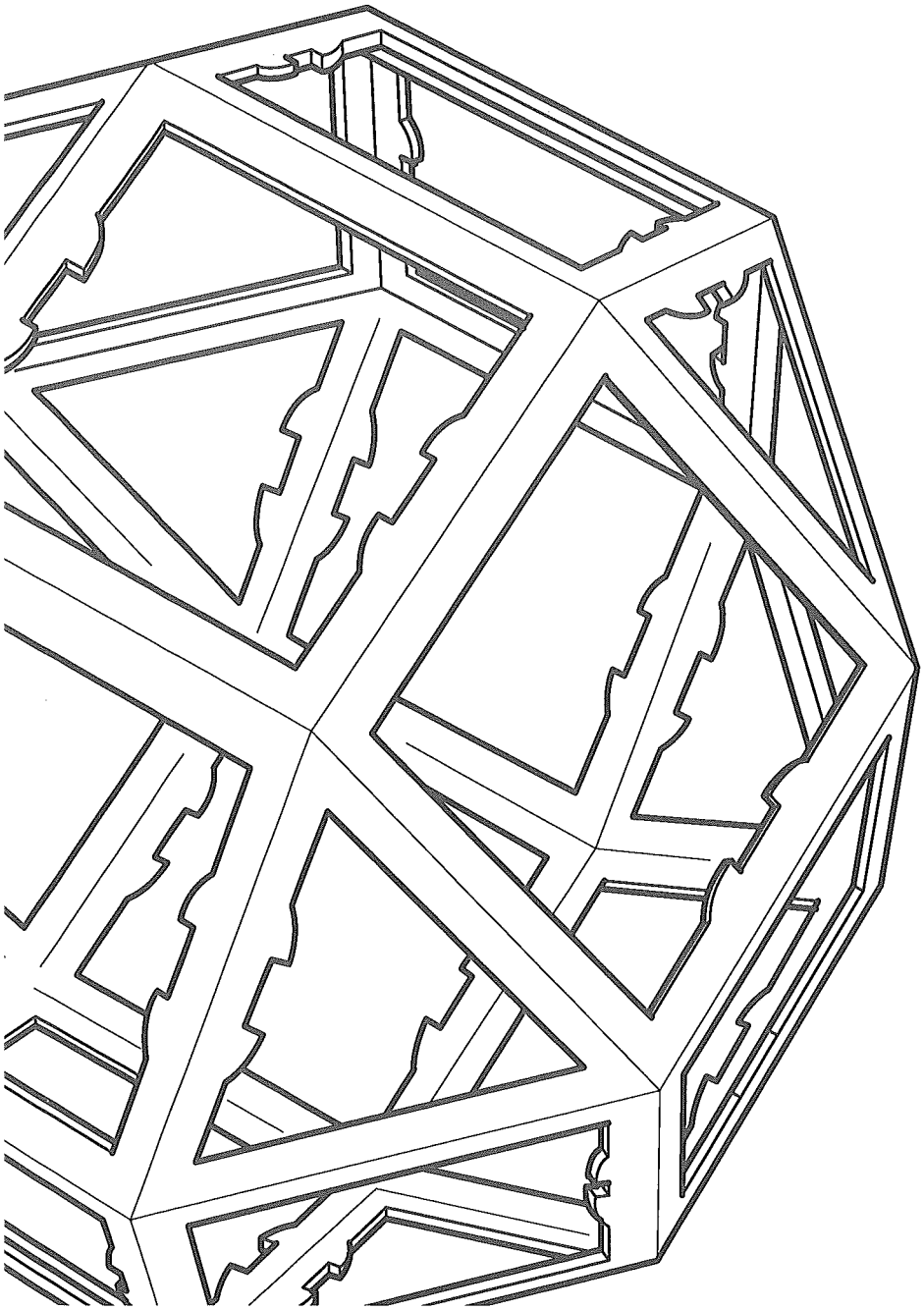
Department of Mathematics and Statistics
College of Engineering

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Summer Research Project

Solving Pell's Equation with Continued Fractions

By Jesse Unger



09

Solving Pell's Equation with Continued Fractions

Jesse Unger

Supervisor: Dr Rua Murray
University of Canterbury

February 5, 2009

Abstract

In this report we will use continued fractions to solve Pell's equation

$$x^2 - Dy^2 = 1$$

We explore some of the properties of simple continued fractions, discuss the relationship between reduced quadratic irrationals and purely periodic simple continued fractions and then give the solution to Pell's and the negative Pell equation. We close by summarizing the entire process in the PQa algorithm which also shows us how to solve some Pell-like equations.

1 Introduction

For more than a millennium, mathematicians have been intrigued with the equation

$$x^2 - Dy^2 = 1$$

where x, y, D are all natural numbers. The equation has been cropping up in places since the time of Archimedes [1, Page 249] and is now named after the 17^{th} century mathematician John Pell and referred to as Pell's equation. This is not because John Pell had a great deal to do with the equation, but because in the 18^{th} century Euler called it the Pell equation, due to mistaking a solution method given by William Brouncker as Pell's work. [1, Page 248]

Among the better ways to solve Pell's equation is the use of continued fractions. Continued fractions have also intrigued mathematicians for centuries and have been worked on by mathematicians such as Lagrange and Euler. [4, Page 30] Despite this, continued fractions can just be seen as a fraction that may contain another fraction, that may contain another, and so on. They behave the same as normal fractions and throughout this report will commonly be manipulated as such.

This report first focuses on continued fractions, exploring some of the basic concepts that may be found in most textbooks about continued fractions (for example, the convergents of a simple continued fraction). We then go on to Section 3 which describes the relationship between reduced quadratic irrationals and purely periodic simple continued fractions. This section follows Chapter 4 of C.D.Olds' book quite closely [4, Page 7], with some of the proofs being shortened and improved on. The solution to Pell's equation is given and then we diverge from [4] to discuss some variants of the Pell equation and give an algorithm that summarizes the entire process, enabling us to quickly solve Pell's equation by hand.

2 Continued Fractions

General Continued Fractions and Simple Continued Fractions

A *general continued fraction* is any expression of the form

$$a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \cfrac{b_4}{\ddots}}}}$$

where the a_n and b_n are independent variables, called the *partial denominators* and the *partial numerators* respectively, and a_0 is the integer part of the general continued fraction. In his book *Continued Fractions*, C.D.Olds [4, Page 7] explains that the partial denominators and partial numerators may be real or complex numbers, and there may be a finite or infinite number of them.

This report focuses on *simple continued fractions*, a subset of general continued fractions of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}} \tag{1}$$

The above notation at times can be too cumbersome and so an abbreviated form of $[a_0; a_1, a_2, a_3, \ldots]$ is preferred and is equivalent to equation (1).

The a_n are referred to as the *partial quotients* or *quotients* of the simple continued fraction. Unlike general continued fractions, they are all integers, and, with the exception of a_0, must be positive. Furthermore, the simple continued fraction is called a finite simple continued fraction if the sequence a_n is finite, and an infinite simple continued fraction if the sequence is infinite.

The Floor and the Fractional Part of a Number

In this report one of the most common operations that we will use on a number is the *floor* function.[7] When applied to a number, the floor function returns the greatest integer that is no greater than the number itself. For example, the floor of 5 is 5, the floor of 2.48 is 2, and the floor of -8.3 is -9. The notation for the floor of x is $\lfloor x \rfloor$

Another important concept in this report is the *fractional part* of a number.[8] This is defined as the number minus its floor, and is denoted as $\{x\}$. That is

$$\{x\} = x - \lfloor x \rfloor. \tag{2}$$

Intuitively we can see that for all x,

$$0 \leq \{x\} < 1. \tag{3}$$

For example, $\{5\} = 0$, $\{2.48\} = 0.48$, and $\{-8.3\} = 0.7$.

It is important that we recognize that dividing a number up into its floor and its fractional part is the only way that we can divide it into two parts with one part an integer and the other part equal or greater than 0 and still less than 1.

Lemma 2.1. *If $\alpha = a + x = b + y$ where $a, b \in \mathbb{Z}$ and $0 \leq x, y < 1$ then $a = b$ and $x = y$.*

Proof. If

$$a + x = b + y$$

then

$$a - b = y - x.$$

Note that $a - b \in \mathbb{Z}$ and $y - x \in (-1, 1)$. Because 0 is the only integer in the interval $(-1, 1)$ it follows that $a = b$ and $x = y$. □

Representation of Numbers

Simple continued fractions are another way to represent real numbers. For example, the number $\frac{54}{19}$ can be represented in any of the following ways:

$$2 + \cfrac{1}{1 + \cfrac{1}{5 + \cfrac{1}{3}}} = [2; 1, 5, 3] = \frac{54}{19} = 2.8421\ldots$$

This simple continued fraction representation can be checked by simplifying the expression starting from the bottom-right.

Such a simple continued fraction is calculated by first obtaining the floor and the fractional part of $\frac{54}{19}$. This fractional part is then expressed as the reciprocal of the reciprocal of the fractional part. Because the reciprocal of the fractional part is greater than 1, it may then be divided into its floor and (a new) fractional part. This fractional part is then treated the same as the last and the process is repeated twice more until there are no more fractional parts. Symbolically the whole process looks like

$$\frac{54}{19} = 2 + \frac{1}{\frac{19}{16}} = 2 + \cfrac{1}{1 + \cfrac{1}{\frac{16}{3}}} = 2 + \cfrac{1}{1 + \cfrac{1}{5 + \cfrac{1}{3}}} = [2; 1, 5, 3].$$

The previous method seems unique and leads us naturally to the following theorem, a rework of [4, Thm 1.1]:

Theorem 2.2. *Every rational number has exactly two finite simple continued fraction expansions, and every finite simple continued fraction expansion represents a rational number.*

Proof. Let $\alpha \in \mathbb{Q}$ and divide it into its floor and fractional part.

If the fractional part of α is 0, α is an integer and one simple continued fraction expansion is $[\alpha]$ and a second is $[\alpha - 1; 1]$. We can see that these both satisfy the conditions of a simple continued fraction as the quotients are all integers and both are equal to α. These are the only two expansions. There are no more with only one quotient as α only has one value. There are no more with two quotients because if the second quotient is not 1 then the expansion consists of a fraction that is less than 1 and α would have had a fractional part. Finally, for the same

reason there are no expansions with three or more quotients. Thus there are only two ways to express an integer as a simple continued fraction:

$$[\alpha] = \alpha \qquad \text{or} \qquad [\alpha - 1; 1] = (\alpha - 1) + \frac{1}{1}. \tag{4}$$

If the fractional part of α is not 0 then define the 1^{st} *residue* of alpha, and in general, of any continued fraction, as the reciprocal of its fractional part. Then

$$\alpha = a_0 + \frac{1}{r_1}$$

where $a_0 \in \mathbb{Z}$ and is the floor of α, and r_1 is the 1^{st} residue of α. Note that $1 < r_1$ because of equation (3).

Now if the k^{th} residue is not an integer it has a fractional part and we can define the $(k+1)^{\text{th}}$ residue recursively by the relationship

$$r_k = a_k + \frac{1}{r_{k+1}} \tag{5}$$

where $a_k = \lfloor r_k \rfloor$ and $r_{k+1} = \dfrac{1}{\{r_k\}}$.

Now if $1 < r_k$ then $a_k \in \mathbb{N}$, and of course $1 < r_{k+1}$. Because $1 < r_1$ it follows by induction that $1 < r_n$ for each r_n which is defined.

It can easily be seen that if α is rational and not an integer then r_1 is rational, and also that if r_n is rational and not an integer then r_{n+1} is rational. Now let $\frac{b}{c} = r_k$ where r_k is not an integer and $b, c \in \mathbb{N}$ with their only common factor being 1. That is, $\dfrac{b}{c}$ is in its lowest terms. Because r_k is not an integer $1 < c$. Then from equations (2), (3) and (5)

$$0 < r_k - a_k = \frac{b}{c} - a_k = \frac{b - a_k c}{c} = \frac{d}{c} = \frac{1}{r_{k+1}} < 1$$

where $0 < b - a_k c = d \in \mathbb{Z}$. Then

$$1 < r_{k+1} = \frac{c}{d}$$

and so

$$d < c.$$

Now because r_k has a denominator of c and r_{k+1} has a denominator of d each subsequent rational residue has a smaller integer denominator when reduced to its lowest terms. So then the denominators of the residues form a decreasing sequence of integers that are all greater than zero. Eventually one of the denominators must be 1 and then the residue is an integer.

Let the first residue that is an integer be r_n. Then the recurrence relationship given in equation (5) will not hold as there is no fractional part of r_n. Instead let

$$r_n = a_n.$$

Notice that again we have $a_k = \lfloor r_k \rfloor$.

At this point we have n different equations that look like

$$\begin{aligned} \alpha &= a_0 + \frac{1}{r_1} \\ &\vdots \\ r_{n-1} &= a_{n-1} + \frac{1}{r_n} \\ r_n &= a_n. \end{aligned}$$

Combining these into one equation generates a simple continued fraction:

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}} = [a_0; a_1, \ldots, a_n].$$

To establish uniqueness, assume that

$$\alpha = [a_0; a_1, \ldots, a_n] = [b_0; b_1, \ldots, b_m]$$

where each $b_i \in \mathbb{N}$. Also assume without loss of generality that neither a_n or b_m are 1.*

Now $0 < [0; b_m] < 1$. Also, because $1 < b_{m-1} + [0; b_m]$ it follows that $0 < [0; b_{m-1}, b_m] < 1$. One can prove by induction that

$$[0; b_j, b_{j+1}, \ldots, b_m] \in (0, 1)$$

for $j = 1 \ldots m$.

Because

$$a_0 + [0; a_1, \ldots, a_n] = b_0 + [0; b_1, \ldots, b_m]$$

by Lemma 2.1, $a_0 = b_0$ and

$$a_1 + [0; a_2, \ldots, a_n] = b_1 + [0; b_2, \ldots, b_m].$$

Continuing on in this way one finds that $a_0 = b_0$, $a_1 = b_1$, ..., $a_n = b_m$ and also that $n = m$. So if the last partial quotient is not 1 then the simple continued fraction expansion is unique.

However, as stated earlier, an integer can be represented in two ways, given by equations (4). So α can be represented one of two ways as a simple continued fraction;

$$\alpha = [a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_n - 1, 1].$$

Proving the second part of the theorem is trivial; it is evident that any finite simple continued fraction represents a rational number as the simple continued fraction can just be simplified from the lower right-hand corner upwards to generate the rational number it represents. □

The previous theorem has a counterpart that follows on naturally, found in [4, Page 52].

Theorem 2.3. *Every irrational number has a infinite simple continued fraction expansion.*

*If $a_n = 1$ then $r_{n-1} = a_{n-1} + 1 \in \mathbb{N}$ and so we can replace $[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_{n-1}, 1]$ with $[a_0; a_1, \ldots, a_{n-1} + 1]$. Likewise for b_m.

Proof. Let $\alpha \in \mathbb{R}$; $\alpha \notin \mathbb{Q}$. Even though α is irrational, the residues are defined the same as in Theorem 2.2. In fact, because α is irrational its fractional part is not 0 and so there must be a 1^{st} residue. Now it is obvious that r_1 is irrational, else α equals the sum of two rational numbers, and would not be irrational. Furthermore, for every irrational r_n, r_{n+1} is irrational, because of equation (5). So all the residues are irrational. Thus there will never be a residue such that $r_k = a_k$ and so the irrational number has an infinite simple continued fraction expansion.

Now from the recursion formula of equation (5) an infinite number of equations are produced that look like

$$\begin{aligned} \alpha &= a_0 + \frac{1}{r_1} \\ r_1 &= a_1 + \frac{1}{r_2} \\ r_2 &= a_2 + \frac{1}{r_3} \\ &\vdots \end{aligned}$$

We can combine these into one equation and generate an infinite simple continued fraction:

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}} = [a_0; a_1, a_2, a_3, \ldots].$$

□

Convergents

For these infinite continued fraction expansions, and even the finite expansions, we can crop the expansion to a finite number of quotients. If we crop the (finite or infinite) expansion $\alpha = [a_0; a_1, \ldots, a_n, a_{n+1}, \ldots]$ to the n^{th} quotient we get the rational number $[a_0; a_1, \ldots, a_n]$. This is called the n^{th} *convergent* of α. Notice for α with a finite number of convergents, the last convergent is equal to α.

For example, the 0^{th} convergent of α is $[a_0] = a_0$. It is more common to see convergents divided into their numerators and denominators so we use p_n to denote the numerator and q_n to denote the denominator of the n^{th} convergent when it is in its lowest terms. Thus

$$\begin{aligned} p_0 &= a_0 \\ q_0 &= 1. \end{aligned}$$

The 1^{st} convergent is

$$[a_0; a_1] = a_0 + \frac{1}{a_1} = \frac{a_1 a_0 + 1}{a_1}$$

and so

$$\begin{aligned} p_1 &= a_1 a_0 + 1 \\ q_1 &= a_1. \end{aligned}$$

Continuing in this fashion, the 2^{nd} convergent is

$$[a_0; a_1, a_2] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2}} = \frac{a_2 a_1 a_0 + a_2 + a_0}{a_2 a_1 + 1}$$

and so

$$p_2 = a_2 a_1 a_0 + a_2 + a_0 = a_2 p_1 + p_0 \tag{6}$$
$$q_2 = \quad a_2 a_1 + 1 \quad = a_2 q_1 + q_0. \tag{7}$$

This can be generalized in the following theorem, similar to [4, Thm 1.3].

Theorem 2.4. *The numerator and denominator of the n^{th} convergent of a real number $\alpha = [a_0; a_1, \ldots]$ where $0 < a_i \in \mathbb{R}$ is given by*

$$\begin{aligned} p_n &= a_n p_{n-1} + p_{n-2} \\ q_n &= a_n q_{n-1} + q_{n-2} \end{aligned} \tag{8}$$

so that the n^{th} convergent is

$$\frac{p_n}{q_n} = [a_0; a_1, \ldots, a_n]. \tag{9}$$

Proof. From equations (6) and (7) we can see that the theorem holds for $n = 2$, for all $0 < a_0, a_1 \in \mathbb{R}$. Let us now assume that the theorem holds for $n = 1 \ldots k$ and then we will prove that it holds for $n = k + 1$ and so the result follows by strong induction. Specifically, assume

$$\begin{aligned} p_k &= a_k p_{k-1} + p_{k-2} \\ q_k &= a_k q_{k-1} + q_{k-2} \end{aligned}$$

and

$$\frac{p_k}{q_k} = [a_0; a_1, \ldots, a_k]. \tag{10}$$

Now note that

$$\begin{aligned} [a_0; a_1, \ldots, a_k, a_{k+1}] &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{\left(a_k + \cfrac{1}{a_{k+1}}\right)}}} \\ &= \left[a_0; a_1, \ldots, \left(a_k + \frac{1}{a_{k+1}}\right)\right]. \end{aligned} \tag{11}$$

Because $0 < a_k + \dfrac{1}{a_{k+1}} \in \mathbb{R}$,the only difference between the $(k+1)^{\text{th}}$ convergent and the k^{th} convergent is that $a_k + \dfrac{1}{a_{k+1}}$ is in place of a_k. So if we replace a_k in equation (10) we get

an expression for $(k+1)^{\text{th}}$ convergent. This is valid because it is evident from the recursion relationships of the numerators and denominators of the convergents, given from the induction assumption, that changing the value of a_k to $a_k + \dfrac{1}{a_{k+1}}$ will have no effect on the value of $p_{n-1}, p_{n-2}, \ldots$ or $q_{n-1}, q_{n-2}, \ldots$. From equations (10) and (11);

$$\begin{aligned}
[a_0; a_1, \ldots, a_k, a_{k+1}] &= \frac{\left(a_k + \dfrac{1}{a_{k+1}}\right) p_{k-1} + p_{k-2}}{\left(a_k + \dfrac{1}{a_{k+1}}\right) q_{k-1} + q_{k-2}} \\
&= \frac{a_{k+1}a_k p_{k-1} + p_{k-1} + a_{k+1}p_{k-2}}{a_{k+1}a_k q_{k-1} + q_{k-1} + a_{k+1}q_{k-2}} \\
&= \frac{a_{k+1}p_k + p_{k-1}}{a_{k+1}q_k + q_{k-1}} \\
&= \frac{p_{k+1}}{q_{k+1}}
\end{aligned}$$

and the theorem is proved by the induction principle. □

It is common practice to define $p_{-2} = 0, p_{-1} = 1, q_{-2} = 1$ and $q_{-1} = 0$. Then the previous theorem can apply to $n = 0$ and $n = 1$.

Because in this section there are no requirements that a_n are integers, we can use the idea of residues to prove this corollary to the previous theorem; an expanded form of [3, Thm 5].

Corollary 2.5. *If r_n is the n^{th} residue of α and the numerators and denominators of the convergents of α are defined as in Theorem 2.4 then*

$$\alpha = \frac{r_n p_{n-1} + p_{n-2}}{r_n q_{n-1} + q_{n-2}}\,. \tag{12}$$

Proof. When assembling a simple continued fraction from the recursion relationships given by equation (5), if we stop at the n^{th} equation we will end up with the equation

$$\begin{aligned}
\alpha &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_{n-1} + \cfrac{1}{r_n}}}} \\
&= [a_0; a_1, \ldots, a_{n-1}, r_n].
\end{aligned}$$

While this may not be a simple continued fraction because r_n may not be an integer, all the terms of the continued fraction are real, so Theorem 2.4 still applies. So in equation (9) we can replace a_n with r_n to make a modified form of the last convergent, and thus

$$\alpha = \frac{r_n p_{n-1} + p_{n-2}}{r_n q_{n-1} + q_{n-2}}\,.$$

□

The formulae derived in Theorem 2.4 lead us to a very important relation that is central in solving Pell's equation.

Theorem 2.6 ([4, Thm 1.4]). *If the numerators and denominators of the convergents of a continued fraction are defined as in Theorem 2.4 then*

$$p_{n+1}q_n - p_nq_{n+1} = (-1)^n. \tag{13}$$

Proof. Using the extended definitions of p_{-2}, p_{-1}, q_{-2} and q_{-1} we can establish the relation true for $n = -2$.

$$p_{-1}q_{-2} - p_{-2}q_{-1} = 1 \cdot 1 - 0 \cdot 0 = 1 = (-1)^{-2}.$$

Now assume the relation true for $n = k$. That is

$$p_{k+1}q_k - p_kq_{k+1} = (-1)^k. \tag{14}$$

Then consider when $n = k + 1$.

$$p_{k+2}q_{k+1} - p_{k+1}q_{k+2} = (a_{k+2}p_{k+1} + p_k)q_{k+1} - p_{k+1}(a_{k+2}q_{k+1} + q_k)$$

from equations (8)

$$\begin{aligned} &= p_kq_{k+1} - p_{k+1}q_k \\ &= (-1)(p_{k+1}q_k - p_kq_{k+1}) \\ &= (-1)^{k+1} \end{aligned}$$

by the induction assumption of equation (14). So the theorem is proved by induction. □

In a simple continued fraction the a_n are all integers and it is easy to see from the recurrence relations given in equations (8) that all the p_n and q_n will be integers. They also have an important property described by a corollary of Theorem 2.6.

Corollary 2.7 ([4, Cor 1.5]). *When defined by equations* (8)*, the numerators of each convergent of a simple continued fraction share no common factors with their corresponding denominator other than 1, and so the convergents are in their lowest terms.*

Proof. Any common factor of p_n and q_n could be factored out of the left-hand side of equation (13) and so must also be a factor of the right-hand side. However, the only factors of the right-hand side are -1 and 1. Thus they share no common factors other than 1, and the convergents are in their lowest terms. □

3 Reduced Quadratic Irrationals and Purely Periodic Simple Continued Fractions

A quadratic irrational is an irrational number in the form $F + G\sqrt{M}$ where $F, G \in \mathbb{Q}$ and $M \in \mathbb{N}$ and is not a perfect square. We go on to show in Lemma 3.3 that this means it is the solution to a quadratic equation with integer coefficients, but first we will prove that it is irrational with the following theorem, a modified version of [2].†

Theorem 3.1. *$\sqrt{M}$ is irrational if $M \in \mathbb{N}$ and is not a perfect square.*

†Proof originally by Richard Dedekind.

Proof. If $M \in \mathbb{N}$ and is not a perfect square then $\exists n \in \mathbb{N}$ such that $n < \sqrt{M} < n+1$ and so,

$$n^2 < M < (n+1)^2. \tag{15}$$

If $\sqrt{M}$ is rational then it can be expressed as $\sqrt{M} = \frac{a}{b}$ where $a, b \in \mathbb{N}$ or as

$$M = \frac{a^2}{b^2} = \frac{c^2}{d^2}$$

where the latter is just the former except in its lowest terms (again $c, d \in \mathbb{N}$). So

$$c^2 = Md^2. \tag{16}$$

Now from equations (15) and (16) we get

$$\begin{array}{rrrl} n^2d^2 < & Md^2 & < & (n+1)^2d^2 \\ n^2d^2 < & c^2 & < & (n+1)^2d^2 \\ nd < & c & < & (n+1)d \\ 0 < & c-nd & < & d \\ 0 < & f & < & d \end{array} \tag{17}$$

where $c - nd = f \in \mathbb{N}$.

Again from equations (15) and (16) we get

$$\begin{array}{rrrl} n^2c^2 < & Mc^2 & < & (n+1)^2c^2 \\ n^2c^2 < & M^2d^2 & < & (n+1)^2c^2 \\ nc < & Md & < & (n+1)c \\ 0 < & Md-nc & < & c \\ 0 < & g & < & c \end{array} \tag{18}$$

where $Md - nc = g \in \mathbb{N}$.

Then

$$\begin{aligned} g^2 - Mf^2 &= M^2d^2 - 2Mncd + n^2c^2 - M(c^2 - 2ncd + n^2d^2) \\ &= (M^2d^2 - Mc^2) - (2Mncd - 2Mncd) + (n^2c^2 - Mn^2d^2) \\ &= 0. \end{aligned}$$

So

$$M = \frac{g^2}{f^2}\,. \tag{19}$$

But because of equations (17) and (18) in conjunction with equation (19) above, $\frac{c^2}{d^2}$ was not in its lowest terms, contrary to assumption. Hence there is a contradiction and $\sqrt{M}$ must be irrational. □

It is obvious that adding a rational number to an irrational number will produce an irrational number; if the result were rational then an irrational number could be produced by subtracting a rational from a rational, which is clearly impossible. Furthermore, multiplying an irrational by a rational produces an irrational. Else an irrational could be produced by dividing a rational by a rational, which we know only produces rational numbers. Thus, numbers of the form $F + G\sqrt{M}$ as described previously are also irrational. There is another condition that holds that we will state in a theorem:

Theorem 3.2 ([4, Page 96]). *If $a+b\sqrt{M} = c+d\sqrt{M}$ where $a, b, c, d \in \mathbb{Q}$, $M \in \mathbb{N}$ and is not a perfect square, then $a = c$ and $b = d$.*

Proof. If the above conditions are true and $d - b \neq 0$ then

$$\sqrt{M} = \frac{a-c}{d-b}$$

and is rational, contrary to Theorem 3.1. So $d - b = 0$ and hence $a = c$ also. □

To go on to talking about reduced quadratic irrationals we must first discuss quadratic irrationals in general, and prove three crucial lemmas.

Lemma 3.3. *All quadratic irrationals $\alpha = F + G\sqrt{M}$ where $F, G \in \mathbb{Q}$ and $M \in \mathbb{N}$ and is not a perfect square, solve a quadratic equation with integer coefficients and also have a conjugate of the form $\alpha' = F - G\sqrt{M}$ which satisfies the same quadratic equation.*

Proof. The quadratic equation

$$\begin{aligned} 0 &= \left(x - (F + G\sqrt{M})\right)\left(x - (F - G\sqrt{M})\right) \\ &= x^2 - 2Fx + F^2 - G^2M \end{aligned}$$

has roots α and α', and has rational coefficients. Now let H be the common denominator of $2F$ and $F^2 - G^2M$. That is, let H be the least number that makes both $2FH$ and $F^2H - G^2MH$ integers. Then α and α' are roots of the quadratic equation

$$Hx^2 - 2FHx + F^2H - G^2MH = 0$$

which has integer coefficients. □

From now on, let the symbol $'$ be the symbol for conjugate. So α' means the conjugate of α and $(\alpha + \beta)'$ means the conjugate of the sum of α and β. With this enhanced notation we can easily prove the second lemma.

Lemma 3.4. *When applying one of the following operations;*

i) addition

ii) subtraction

iii) multiplication

iv) division

between two quadratic irrationals involving the same integer as the subject of the square root, conjugating the quadratic irrationals before the operation is equivalent to conjugating the result after the operation.

Proof. Let $\alpha = F + G\sqrt{M}$ be a quadratic irrational and $\beta = H + I\sqrt{M}$ be another. So α and β have the same integer under the square root. Now consider each operation:

i)

$$\begin{aligned} \alpha' + \beta' &= (F + G\sqrt{M})' + (H + I\sqrt{M})' \\ &= F - G\sqrt{M} + H - I\sqrt{M} \\ &= (F + H) - (G + I)\sqrt{M} \\ &= (F + G\sqrt{M} + H + I\sqrt{M})' \\ &= (\alpha + \beta)' \end{aligned}$$

ii)

$$\begin{aligned}\alpha' - \beta' &= (F + G\sqrt{M})' - (H + I\sqrt{M})' \\ &= F - G\sqrt{M} - H + I\sqrt{M} \\ &= (F - H) - (G - I)\sqrt{M} \\ &= (F + G\sqrt{M} - H - I\sqrt{M})' \\ &= (\alpha - \beta)'\end{aligned}$$

iii)

$$\begin{aligned}\alpha'\beta' &= (F - G\sqrt{M})(H - I\sqrt{M}) \\ &= (FH + GIM) - (FI + HG)\sqrt{M} \\ &= (FH + GIM + FI\sqrt{M} + HG\sqrt{M})' \\ &= \Big((F + G\sqrt{M})(H + I\sqrt{M})\Big) \\ &= (\alpha\beta)'\end{aligned}$$

iv)

$$\begin{aligned}\frac{\alpha'}{\beta'} &= \frac{(F - G\sqrt{M})}{(H - I\sqrt{M})} \times \frac{(H + I\sqrt{M})}{(H + I\sqrt{M})} \\ &= \frac{(FH - GIM) - (HG - FI)\sqrt{M}}{H^2 - I^2M} \\ &= \left(\frac{(FH - GIM) + (HG - FI)\sqrt{M}}{H^2 - I^2M}\right)' \\ &= \left(\frac{(F + G\sqrt{M})}{(H + I\sqrt{M})} \times \frac{(H - I\sqrt{M})}{(H - I\sqrt{M})}\right)' \\ &= \left(\frac{\alpha}{\beta}\right)'\end{aligned}$$

Hence Lemma 3.4 is proved for all four operations. □

It is useful expressing a quadratic irrational in the form $F + G\sqrt{M}$ where $F, G \in \mathbb{Q}$ and $M \in \mathbb{N}$ and is not a perfect square, as it is simple and easy to manipulate as shown in the previous proofs. However, as we go on it will be more useful for us to express quadratic irrationals in the form

$$\frac{A \pm \sqrt{D}}{B}$$

where $A \in \mathbb{Z}; B, D \in \mathbb{N}$ and D is not a perfect square. We can show that this is equivalent to the previous form.

Lemma 3.5. *Every expression in the form $F \pm G\sqrt{M}$ where $F, G \in \mathbb{Q}$; $M \in \mathbb{N}$ and is not a perfect square and $0 < G$[‡], has an equivalent expression in the form*

$$\frac{A \pm \sqrt{D}}{B}$$

[‡]Notice the slight change: From now on we must specify if the multiple of the root is added or subtracted as in the new form there is no coefficient in front of the root that is able to be positive or negative.

where $A \in \mathbb{Z}$; $B, D \in \mathbb{N}$ and D is not a perfect square.

Proof. If $F, G \in \mathbb{Q}$ then they can be expressed $F = \frac{f}{h}$; $G = \frac{g}{h}$ where h is the lowest common denominator of F and G, and $f \in \mathbb{Z}$; $g, h \in \mathbb{N}$. Then

$$\begin{aligned} F \pm G\sqrt{M} &= \frac{f \pm g\sqrt{M}}{h} \\ &= \frac{f \pm \sqrt{g^2 M}}{h} \\ &= \frac{A \pm \sqrt{D}}{B} \end{aligned}$$

where A,B and D satisfy the previously stated conditions and the two forms are equivalent. □

Using this new notation we can easily show the solutions to a quadratic equation with integer coefficients. The equation $ax^2 + bx + c = 0$ where $a, b, c \in \mathbb{Z}; 0 < a$ has two solutions,

$$\begin{aligned} \alpha = \quad & \frac{-b + \sqrt{b^2 - 4ac}}{2a} = \frac{A + \sqrt{D}}{B} \\ \alpha' = \quad & \frac{-b - \sqrt{b^2 - 4ac}}{2a} = \frac{A - \sqrt{D}}{B} \end{aligned}$$

where

$$\begin{aligned} A = \quad & -b & \in \mathbb{Z} \\ 0 < B = \quad & 2a & \in \mathbb{N} \\ D = \quad & b^2 - 4ac & \in \mathbb{N} \end{aligned}$$

and assume for now that $0 < D$ and is not a perfect square, otherwise α would not be a quadratic irrational; it would either be complex or rational.

Reduced Quadratic Irrationals

Thus far we have had much to say about quadratic irrationals, but not much to say about those that are reduced. A *reduced quadratic irrational* is one whose value is greater than one and whose conjugate's value is greater than negative one but less than zero. [4, Page 101] In symbolic notation, α is reduced if $1 < \alpha$ and $-1 < \alpha' < 0$.

If α is reduced and in regular form $\frac{A + \sqrt{D}}{B}$ then $\alpha' = \frac{A - \sqrt{D}}{B}$ and we can deduce

$$\begin{aligned} 1 &< \frac{A + \sqrt{D}}{B} \\ B &< A + \sqrt{D} \end{aligned} \tag{20}$$

and

$$\begin{aligned} -1 &< \frac{A - \sqrt{D}}{B} \\ -B &< A - \sqrt{D} \\ \sqrt{D} - A &< B. \end{aligned} \tag{21}$$

Also because $0 < B$ we can deduce

$$\begin{aligned} \frac{A-\sqrt{D}}{B} &< 0 \\ A-\sqrt{D} &< 0 \\ A &< \sqrt{D} \end{aligned} \tag{22}$$

and

$$\begin{aligned} -1+1 &< \alpha' + \alpha \\ 0 &< \frac{2A}{B} \\ 0 &< A. \end{aligned} \tag{23}$$

From equations (22) and (23) we find

$$0 < A < \sqrt{D} \tag{24}$$

and from equations (20), (21) and (24) we conclude

$$0 < \sqrt{D} - A < B < \sqrt{D} + A < 2\sqrt{D}. \tag{25}$$

This set of inequalities will allow us to state our next theorem.

Theorem 3.6 ([4, Page 102]). *There are only a finite number of reduced quadratic irrationals associated with any given D.*

Proof. We know from equations (24) and (25) that for α to be a reduced quadratic irrational it is necessary that A must be between 0 and $\sqrt{D}$ and B must be between 0 and $2\sqrt{D}$. Furthermore, both A and B must be integers. Thus, if we fix D then there are a finite number of pairs of integers which meet this criteria. Thus there are only a finite number of potential candidates of A and B that make α reduced when D is fixed. □

The next theorem will start to relate the idea of reduced quadratic irrationals to that of simple continued fractions:

Theorem 3.7 ([4, Page 102]). *If α_n is reduced and $\alpha_n = a_n + \dfrac{1}{\alpha_{n+1}}$ where a_n is the floor of α_n then α_{n+1} is reduced and has the same integer as the subject of the square root.*

Proof. First we show α_{n+1} is reduced.

$$\begin{aligned} 0 < \quad & \alpha_n - a_n \quad < 1 \\ 0 < \quad & \frac{1}{\alpha_{n+1}} \quad < 1 \end{aligned}$$

so

$$1 < \alpha_{n+1}. \tag{26}$$

Furthermore, from Theorem 3.4

$$\begin{aligned} (\alpha_n - a_n)' &= \left(\frac{1}{\alpha_{n+1}}\right)' \\ \alpha_n' - a_n &= \frac{1}{\alpha_{n+1}'} \end{aligned}$$

then because $1 < a_n$ and $-1 < \alpha'_n < 0$

$$1 < a_n - \alpha'_n = \frac{-1}{\alpha'_{n+1}}$$

so

$$-1 < \alpha'_{n+1} < 0. \tag{27}$$

Equations (26) and (27) fulfill the requirements for α_{n+1} being reduced.

Now we show the subject of the square root is the same for α_{n+1} as it is for α_n. Let

$$\alpha_n = \frac{-b + \sqrt{b^2 - 4\,a\,c}}{2\,a} = \frac{A_n + \sqrt{D}}{B_n}\,.$$

So

$$\begin{aligned}
0 &= a\,\alpha_n^2 + b\,\alpha_n + c \\
&= a\left(a_n + \frac{1}{\alpha_{n+1}}\right)^2 + b\left(a_n + \frac{1}{\alpha_{n+1}}\right) + c \\
&= a\,a_n^2 + \frac{2\,a\,a_n}{\alpha_{n+1}} + \frac{a}{\alpha_{n+1}^2} + b\,a_n + \frac{b}{\alpha_{n+1}} + c \\
&= (a\,a_n^2 + b\,a_n + c)\,\alpha_{n+1}^2 + (2\,a\,a_n + b)\,\alpha_{n+1} + a.
\end{aligned} \tag{28}$$

Notice that the coefficients of this last quadratic equation are integers. Thus when we solve it for the positive root $\alpha_{n+1} = \dfrac{A_{n+1} + \sqrt{D_{n+1}}}{B_{n+1}}$ we will get $A_{n+1}, B_{n+1}, D_{n+1}$ all integers. Now solve for D_{n+1}:

$$\begin{aligned}
D_{n+1} &= (2\,a\,a_n + b)^2 - 4(a\,a_n^2 + b\,a_n + c)a \\
&= 4\,a^2a_n^2 + 4\,a\,b\,a_n + b^2 - 4\,a^2a_n^2 - 4\,a\,b\,a_n - 4\,a\,c \\
&= b^2 - 4\,a\,c \\
&= D.
\end{aligned}$$

Thus the theorem is proved. □

Finally we come to the focal point of this section. The following theorem is shorter rework of [4, Thm 4.2] and is a key to solving Pell's equation.

Theorem 3.8. *If α is a reduced quadratic irrational then its simple continued fraction expansion is purely periodic.*

Proof. Theorem 2.3 proves that the simple continued fraction of α is infinite. However, Theorem 3.7 implies that every residue of α is reduced, and Theorem 3.7 implies that there are a finite number of these reduced quadratic irrationals. As a consequence of these two seemingly opposite statements, at some point there occurs some residue, r_k, that is a repetition of a previous residue, r_j.

Considering the simple continued fraction expansion of α, because a_j and a_k are the largest integers less than r_j and r_k respectively and $r_j = r_k$, it follows that $a_j = a_k$ and

$$\begin{aligned}
r_j &= r_k \\
a_j + \frac{1}{r_{j+1}} &= a_k + \frac{1}{r_{k+1}} \\
r_{j+1} &= r_{k+1}.
\end{aligned}$$

Furthermore, the same reasoning can be applied to show that $r_{j+2} = r_{k+2}$, $r_{j+3} = r_{k+3}$ and so on.

Now because $r_{n-1} = a_{n-1} + \dfrac{1}{r_n}$ we can manipulate two expressions about r_{j-1} and r_{k-1} side by side to eventually show that that they are equal.

$$r_{j-1} = a_{j-1} + \frac{1}{r_j} \qquad r_{k-1} = a_{k-1} + \frac{1}{r_k}$$

$$r'_{j-1} = a_{j-1} + \frac{1}{r'_j} \qquad r'_{k-1} = a_{k-1} + \frac{1}{r'_k}$$

from Theorem 3.4. Because $r_j = r_k$ it follows that $r'_j = r'_k$, and

$$r'_{j-1} - a_{j-1} = r'_{k-1} - a_{k-1}$$

so

$$a_{j-1} - r'_{j-1} = a_{k-1} - r'_{k-1}.$$

Because r_{j-1} and r_{k-1} are reduced, it follows from Lemma 2.1 that $a_{j-1} = a_{k-1}$ and $r'_{j-1} = r'_{k-1}$ and thus, $r_{j-1} = r_{k-1}$. As before, the same method shows that $r_{j-2} = r_{k-2}$, $r_{j-3} = r_{k-3}$ and so on, up to $r_1 = r_{k-j+1}$ and $\alpha = r_{k-j}$.

Let m be the value where r_m is the *first* residue where the value equals α. Then $r_i = r_{m+i}$ for all $i \in \mathbb{N}$. Furthermore, taking the unique integer a_n for α and each of r_n we get $a_0 = a_m$ and $a_i = a_{m+i}$ for all $i \in \mathbb{N}$.

Thus $\alpha = [\overline{a_0; a_1, ..., a_{m-1}}]$[§] and so is purely periodic. □

The variable m is known as the length of the period and will be quite important in solving Pell's equation.

4 Square Roots and Pell's Equation

Square roots of natural numbers are not reduced quadratic irrationals because their conjugates are never between -1 and 0. Thus they are never purely periodic. However, they do have a special form.

Theorem 4.1 ([4, Page 112]). *Simple continued fractions of square roots take the form*

$$\sqrt{D} = [a_0; \overline{a_1, a_2, \ldots, a_{m-1}, 2a_0}]$$

when $D \in \mathbb{N}$ and is not a perfect square.

Proof. If $D \in \mathbb{N}$ and is not a perfect square then $1 < \sqrt{D}$. This means its conjugate $-\sqrt{D} < -1$ and so $\sqrt{D}$ is not reduced. However, if a_0 is the greatest integer less than $\sqrt{D}$ then

$$1 < a_0 + \sqrt{D}$$

and its conjugate

$$-1 < a_0 - \sqrt{D} < 0.$$

[§]The overhead bar is the standard mathematical notation used when the content below it repeats forever.

So $a_0 + \sqrt{D}$ is reduced and by Theorem 3.8 its simple continued fraction representation is purely periodic.

So

$$a_0 + \sqrt{D} = 2a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_{m-1} + \cfrac{1}{2a_0 + \cfrac{1}{\ddots}}}}}$$

and

$$\sqrt{D} = [a_0; \overline{a_1, a_2, \ldots, a_{m-1}, 2a_0}].$$

□

Before we go on to solving Pell's equation we will find the simple continued fraction representation of $\sqrt{7}$ and show that it takes the form above.

First note that $a_0 = \lfloor \sqrt{7} \rfloor = 2$. So $\sqrt{7} = 2 + \dfrac{1}{r_1}$ and

$$\begin{aligned} r_1 &= \frac{1}{\sqrt{7}-2} \cdot \frac{\sqrt{7}+2}{\sqrt{7}+2} \\ &= \frac{\sqrt{7}+2}{3}. \end{aligned}$$

Now $a_1 = \lfloor r_1 \rfloor = \lfloor \dfrac{\sqrt{7}+2}{3} \rfloor = 1$. So $r_1 = \dfrac{\sqrt{7}+2}{3} = 1 + \dfrac{1}{r_2}$ and

$$\begin{aligned} r_2 &= \frac{1}{\left(\dfrac{\sqrt{7}+2}{3} - 1\right)} \\ &= \frac{3}{\sqrt{7}-1} \cdot \frac{\sqrt{7}+1}{\sqrt{7}+1} \\ &= \frac{\sqrt{7}+1}{2}. \end{aligned}$$

Now $a_2 = \lfloor r_2 \rfloor = \lfloor \dfrac{\sqrt{7}+1}{2} \rfloor = 1$. So $r_2 = \dfrac{\sqrt{7}+1}{2} = 1 + \dfrac{1}{r_3}$ and

$$\begin{aligned} r_3 &= \frac{1}{\left(\dfrac{\sqrt{7}+1}{2} - 1\right)} \\ &= \frac{2}{\sqrt{7}-1} \cdot \frac{\sqrt{7}+1}{\sqrt{7}+1} \\ &= \frac{\sqrt{7}+1}{3}. \end{aligned}$$

Now $a_3 = \lfloor r_3 \rfloor = \lfloor \frac{\sqrt{7}+1}{3} \rfloor = 1$. So $r_3 = \frac{\sqrt{7}+1}{3} = 1 + \frac{1}{r_4}$ and

$$\begin{aligned} r_4 &= \frac{1}{\left(\frac{\sqrt{7}+1}{3} - 1\right)} \\ &= \frac{3}{\sqrt{7}-2} \cdot \frac{\sqrt{7}+2}{\sqrt{7}+2} \\ &= \sqrt{7}+2. \end{aligned}$$

Now $a_4 = \lfloor r_4 \rfloor = \lfloor \sqrt{7}+2 \rfloor = 4$. So $r_4 = \sqrt{7}+2 = 4 + \frac{1}{r_5}$ and

$$\begin{aligned} r_5 &= \frac{1}{\sqrt{7}-2} \\ &= r_1. \end{aligned}$$

As a consequence, $a_5 = a_1, a_6 = a_2, \ldots$ Thus the simple continued fraction representation is

$$\sqrt{7} = [2; \overline{1, 1, 1, 4}] \tag{29}$$

which is consistent with Theorem 4.1.

Pell's Equation

Finally we come to solving Pell's equation $x^2 - Dy^2 = 1$ for $D \in \mathbb{N}$. To do this we will exploit the fact given in Theorem 4.1 that square roots have a particular type of simple continued fraction expansion. Along the way we will also solve the negative Pell equation; a Pell equation with the right-hand side negative one instead of positive one.

Theorem 4.2 ([4, Page 114]). *Let $D \in \mathbb{N}$ and not be a perfect square, so $\sqrt{D} = \left[a_0; \overline{a_1, a_2, \ldots, a_{m-1}, 2a_0}\right]$. Also let p_n and q_n be defined as in Theorem 2.4.*

If the length of the period, m, is even then $(x, y) = (p_{m-1}, q_{m-1})$ solves the Pell equation $x^2 - Dy^2 = 1$ for integers. If the length of the period, m, is odd then $(x, y) = (p_{m-1}, q_{m-1})$ solves the negative Pell equation $x^2 - Dy^2 = -1$ for integers and $(x, y) = (p_{2m-1}, q_{2m-1})$ solves the Pell equation $x^2 - Dy^2 = 1$ for integers.

Proof. Because $\sqrt{D} = \left[a_0; \overline{a_1, a_2, \ldots, a_{m-1}, 2a_0}\right]$ it follows that

$$\sqrt{D} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_{m-1} + \cfrac{1}{a_0 + \sqrt{D}}}}} \tag{30}$$

Now from Theorem 2.5 we get

$$\begin{aligned} \sqrt{D} &= \frac{(a_0 + \sqrt{D})p_{m-1} + p_{m-2}}{(a_0 + \sqrt{D})q_{m-1} + q_{m-2}} \\ (a_0 + \sqrt{D})q_{m-1}\sqrt{D} + q_{m-2}\sqrt{D} &= (a_0 + \sqrt{D})p_{m-1} + p_{m-2} \\ q_{m-1}D + (a_0 q_{m-1} + q_{m-2})\sqrt{D} &= a_0 p_{m-1} + p_{m-2} + p_{m-1}\sqrt{D} \end{aligned}$$

and from Theorem 3.2 it follows that

$$q_{m-1}D = a_0p_{m-1} + p_{m-2} \qquad \text{and} \qquad p_{m-1} = a_0q_{m-1} + q_{m-2}$$

so

$$p_{m-2} = q_{m-1}D - a_0p_{m-1} \qquad \text{and} \qquad q_{m-2} = p_{m-1} - a_0q_{m-1}. \tag{31}$$

We can adjust the formula given in Theorem 2.6 by letting $n = m-2$ to show

$$\begin{aligned}(-1)^{m-2} &= p_{m-1}q_{m-2} - p_{m-2}q_{m-1} \\ &= p_{m-1}(p_{m-1} - a_0q_{m-1}) - (q_{m-1}D - a_0p_{m-1})q_{m-1}\end{aligned}$$

from equations (31). So

$$p_{m-1}^2 - Dq_{m-1}^2 = (-1)^m. \tag{32}$$

So when m is even $(x, y) = (p_{m-1}, q_{m-1})$ solves the Pell equation, and when m is odd it solves the negative Pell equation.

Note that when setting up this proof in equation (30) we did not need to stop at the end of the first period. Instead, we could have stopped at the end of any period. If we stopped at the end of the second period, equation (30) would look like

$$\sqrt{D} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_{2m-1} + \cfrac{1}{a_0 + \sqrt{D}}}}}$$

From this equation the previous logic can be carried out the same with the only difference being that $m-1$ is replaced by $2m-1$ and $m-2$ is replaced by $2m-2$. The new version of equation (32) will then look like

$$\begin{aligned}p_{2m-1}^2 - Dq_{2m-1}^2 &= (-1)^{2m} \\ &= 1.\end{aligned} \tag{33}$$

Thus when m is odd and it is not sufficient to stay in the first period to solve the Pell equation, $(x, y) = (p_{2m-1}, q_{2m-1})$ will give a solution in integers. □

That is not to say that equation (33) cannot be solved when m is even. The equation is just as valid for even m as it is for odd m. In fact, the general form of equation (33) for the k^{th} period is

$$p_{km-1}^2 - Dq_{km-1}^2 = (-1)^{km}.$$

for $k \in \mathbb{N}$.[¶] This shows when m is even, $(x, y) = (p_{km-1}, q_{km-1})$ will solve Pell's equation for all k, and when m is odd, $(x, y) = (p_{km-1}, q_{km-1})$ will solve Pell's equation for all even k, and will solve the negative Pell equation for all odd k.

A direct consequence of this is that if there is a solution for Pell's equation given by Theorem 4.2, then there are an infinite number of solutions of Pell's equation. Furthermore,

[¶]The process to get this equation is identical to the previous process but instead stopping at the end of the k^{th} period.

if there is a solution for the negative Pell equation given by Theorem 4.2, then there are an infinite number of solutions of the negative Pell equation.

Later we will go on to discuss how to obtain even more solutions of the Pell equation and its variants. For now though, let us consider an example, and solve the Pell equation $x^2 - 7y^2 = 1$.

We know from equation (29) that $\sqrt{7} = [2; \overline{1, 1, 1, 4}]$. We can see that the length of the period is 4 so Theorem 4.2 tells us that the numerator and the denominator of the 3rd convergent will solve the Pell equation for $D = 7$. The numerators and denominators of the first 7 convergents are calculated from Theorem 2.4 and are as follows:

n	-2	-1	0	1	2	3	4	5	6	7
a_n			2	1	1	1	4	1	1	1
p_n	0	1	2	3	5	8	37	45	82	127
q_n	1	0	1	1	2	3	14	17	31	48

We can see that $p_3 = 8$ and $q_3 = 3$. This indeed is a solution to Pell's equation with $D = 7$ as $8^2 - 7 \times 3^2 = 1$. Furthermore, calculating into the second period shows that $p_7 = 127$ and $q_7 = 48$. A quick check with a calculator shows it is true that $127^2 - 7 \times 48^2 = 1$. If more solutions are required this is easily extended into the 3rd period or further.

When there are no solutions

This report has now answered its main question: How to solve Pell's equation for D that is not a perfect square. Throughout all our explorations we have always maintained this for D and have never explained why. It is only right, then, to give an explanation to what happens when D *is* a perfect square.

If D is a perfect square then there exists some $d \in \mathbb{N}$ such that $D = d^2$. Then

$$\begin{aligned} 1 &= x^2 - Dy^2 \\ &= x^2 - d^2y^2 \\ &= x^2 - (dy)^2 \\ &= x^2 - z^2 \end{aligned}$$

where $z = dy \in \mathbb{N}$, and so

$$1 = (x + z)(x - z).$$

Because the only possible factors of 1 are -1 and 1, this leaves us with one of two options. The first is that

$$x + z = x - z = 1$$

and $x = 1$ while $z = 0 \notin \mathbb{N}$, and so this is not a solution the Pell equation.

The second is that

$$x + z = x - z = -1$$

and $x = -1 \notin \mathbb{N}$ and also $z = 0 \notin \mathbb{N}$, and so this too is not a solution to the Pell equation. Thus there is no way to obtain an integer solution to the Pell equation when D is a perfect square.

However, Pell's equation is still quite gracious; there are solutions for all D that are not square. The negative Pell equation is less merciful; there are many D that do not have integer

solutions for the equation. For example, it can be easily shown that when $D = 7$ the equation $x^2 - Dy^2 = -1$ does not have any integer solutions. ‖

If $x^2 - 7y^2 = -1$ is true for some integers x and y, then one of the following 4 cases is true.

i) x and y are both even

Let $x = 2u; y = 2v$ for $u, v \in \mathbb{N}$. Then

$$\begin{aligned} x^2 - 7y^2 &= 4u^2 - 28v^2 \\ &= 4(u^2 - 7v^2). \end{aligned}$$

Now because $u^2 - 7v^2 \in \mathbb{Z}$, $x^2 - 7y^2$ cannot equal -1 and so x and y cannot be both even.

ii) x and y are both odd

Let $x = 2u - 1; y = 2v - 1$ for $u, v \in \mathbb{N}$. Then

$$\begin{aligned} x^2 - 7y^2 &= (2u-1)^2 - 7(2v-1)^2 \\ &= 4u^2 - 4u + 1 - 7(4v^2 - 4v + 1) \\ &= 2(2u^2 - 2u - 14v^2 + 14v - 3). \end{aligned}$$

Because $2u^2 - 2u - 14v^2 + 14v - 3 \in \mathbb{Z}$, $x^2 - 7y^2$ cannot equal -1 and so x and y cannot be both odd.

iii) x is even and y is odd

Let $x = 2u; y = 2v - 1$ for $u, v \in \mathbb{N}$. Then

$$\begin{aligned} x^2 - 7y^2 &= 4u^2 - 7(4v^2 - 4v + 1) \\ &= 4(u^2 - 7v^2 + 7v - 1) - 3. \end{aligned}$$

Now let $w = u^2 - 7v^2 + 7v - 1 \in \mathbb{Z}$. Assume that there is a solution where x is even and y is odd. Then

$$4w - 3 = -1$$

and $w = \frac{1}{2} \notin \mathbb{Z}$. So there is a contradiction; thus x cannot be even and y be odd.

iv) x is odd and y is even

Let $x = 2u - 1; y = 2v$ for $u, v \in \mathbb{N}$. Then

$$\begin{aligned} x^2 - 7y^2 &= 4u^2 - 4u + 1 - 28v^2 \\ &= 4(u^2 - u - 7v^2) + 1. \end{aligned}$$

Now let $w = u^2 - u - 7v^2 \in \mathbb{Z}$. Assume that there is a solution where x is odd and y is even. Then

$$4w + 1 = -1$$

and $w = \frac{-1}{2} \notin \mathbb{Z}$. Again there is a contradiction and x cannot be odd and y be even.

Because all 4 cases are not possible, $x^2 - 7y^2 = -1$ has no integer solutions.

‖Similar to the proof given in the appendix of [4] as to why $x^2 - 3y^2 = -1$ has no integer solutions.

Other Solutions

It has been shown that when one solution of Pell's equation is found from the simple continued fraction method, an infinite number of solutions can be generated. In 628 AD the Indian mathematician Brahmagupta showed a general form of this without the need for continued fractions.[6, Pages 72-73] The result is as follows:

Theorem 4.3. *If* $x_1^2 - Dy_1^2 = N_1$ *and* $x_2^2 - Dy_2^2 = N_2$ *then* $(x_1x_2 + Dy_1y_2)^2 - D(x_1y_2 + x_2y_1)^2 = N_1N_2$.

Proof.

$$\begin{aligned} N_1N_2 &= (x_1^2 - Dy_1^2)(x_2^2 - Dy_2^2) \\ &= x_1^2x_2^2 + D^2y_1^2y_2^2 - Dx_1^2y_2^2 - Dx_2^2y_1^2 \\ &= x_1^2x_2^2 + Dx_1x_2y_1y_2 + D^2y_1^2y_2^2 - Dx_1^2y_2^2 - Dx_1x_2y_1y_2 - Dx_2^2y_1^2 \\ &= (x_1x_2 + Dy_1y_2)^2 - D(x_1y_2 + x_2y_1)^2. \end{aligned}$$

□

Sometimes this in itself is enough to solve the Pell equation, without knowing any prior solutions, but the method is not as straightforward as the simple continued fraction method. For more information, see [6, Page 75].

5 Summary of Method; The PQa Algorithm

The PQa algorithm** is an algorithm that automates the process on page 17 to work out the simple continued fraction representations of quadratic irrationals. The algorithm presented here is a simplified version that applies only to square roots, however the full version can be found online. [5, Page 4]

It is desired to calculate the simple continued fraction representation of $\sqrt{D}$, where $0 < D \in \mathbb{N}$ and is not a perfect square. Define $A_0 = 0$ and $B_0 = 1$. To obtain the simple continued fraction, recursively define

$$A_n = a_{n-1}B_{n-1} - A_{n-1} \tag{34}$$

$$B_n = \frac{D - A_n^2}{B_{n-1}}. \tag{35}$$

It then follows that

$$a_n = \lfloor \frac{\sqrt{D} + A_n}{B_n} \rfloor = \lfloor \frac{a_0 + A_n}{B_n} \rfloor. \tag{36}$$

This is especially effective if calculated in a table with the p_n and q_n. The table when $D = 7$ for n up to 8 is as follows;

**Traditionally, the variables used to represent a quadratic irrational are $\frac{P + \sqrt{D}}{Q}$. The algorithm is used to find the quotients of the simple continued fraction representation; the a_n. Hence the name 'PQa'. I have used the variables A and B instead thoughout this report to avoid confusion with the p_n and q_n. It would make more sense to call this the ABa algorithm; however, more importance was placed on the official name than the correlation with the variables used.

n	-2	-1	0	1	2	3	4	5	6	7	8
A_n			0	2	1	1	2	2	1	1	2
B_n			1	3	2	3	1	3	2	3	1
a_n			2	1	1	1	4	1	1	1	4
p_n	0	1	2	3	5	8	37	45	82	127	590
q_n	1	0	1	1	2	3	14	17	31	48	223

Notice from the example on page 17 that $r_i = \dfrac{\sqrt{D}+A_i}{B_i}$. By carefully examining the equations (34), (35) and (36) we see this is because they are the same process. However, the PQa algorithm is much faster to use. It also allows us to state the last theorem of this report. This theorem is stated in [5] without a proof; the proof presented here is original.

Theorem 5.1. *Let $D \in \mathbb{N}$ and not be a perfect square, and $\sqrt{D} = [a_0; a_1, a_2, \ldots]$. If p_n and q_n are defined as in Theorem 2.4 and A_n and B_n are defined as in equations* (34) *and* (35) *with $A_0 = 0$ and $B_0 = 1$ then*

$$p_n^2 - Dq_n^2 = (-1)^{n+1}B_{n+1}.$$

Proof. Consider when $n = 0$:

$$\begin{aligned}\text{L.H.S.} &= p_0^2 - Dq_0^2 \\ &= a_0^2 - D \\ &= (-1)(D - a_0^2) \\ &= \text{R.H.S.}\end{aligned}$$

Now assume that the statement holds for $1 \ldots k = n$. Specifically, assume that

$$p_{k-1}^2 - Dq_{k-1}^2 = (-1)^k B_k \tag{37}$$

and

$$p_k^2 - Dq_k^2 = (-1)^{k+1}B_{k+1}. \tag{38}$$

Before we go on let us prove a lemma that

$$p_k p_{k-1} - Dq_k q_{k-1} = (-1)^k A_{k+1}. \tag{39}$$

For $k = 0$, L.H.S. $= a_0 - 0 = A_1 =$ R.H.S. Now assume that the lemma is true for $k = j$. That is,

$$p_j p_{j-1} - Dq_j\, q_{j-1} = (-1)^j A_{j+1}. \tag{40}$$

Now consider when $k = j + 1$:

$$\begin{aligned}\text{L.H.S.} &= p_{j+1}p_j - Dq_{j+1}q_j \\ &= a_{j+1}p_j^2 + p_j p_{j-1} - D\left(a_{j+1}q_j^2 + q_j q_{j-1}\right) \\ &= a_{j+1}\left(p_j^2 - Dq_j^2\right) + \left(p_j p_{j-1} - Dq_j q_{j-1}\right) \\ &= a_{j+1}\left(p_j^2 - Dq_j^2\right) + (-1)^j A_{j+1}\end{aligned}$$

from the induction assumption of equation (40);

$$= a_{j+1}(-1)^{j+1}B_{j+1} - (-1)^{j+1}A_{j+1}$$

from the induction assumption of equation (38);

$$= (-1)^{j+1}A_{j+2} = \text{R.H.S.}$$

from equation (34). Thus, the lemma holds if the assumption of equation (38) holds.

Getting back to the original hypothesis, consider when $n = k + 1$:

$$\begin{aligned}
\text{L.H.S.} &= p_{k+1}^2 - Dq_{k+1}^2 \\
&= (a_{k+1}p_k + p_{k-1}) - D\left(a_{k+1}q_k + q_{k-1}\right) \\
&= a_{k+1}^2p_k^2 + 2a_{k+1}p_kp_{k-1} + p_{k-1}^2 - D\left(a_{k+1}^2q_k^2 + 2a_{k+1}q_kq_{k-1} + q_{k-1}^2\right) \\
&= a_{k+1}^2\left(p_k^2 - Dq_k^2\right) + \left(p_{k-1}^2 - Dq_{k-1}^2\right) + 2a_{k+1}\left(p_kp_{k-1} - Dq_kq_{k-1}\right) \\
&= a_{k+1}^2(-1)^{k+1}B_{k+1} + \left(p_{k-1}^2 - Dq_{k-1}^2\right) + 2a_{k+1}\left(p_kp_{k-1} - Dq_kq_{k-1}\right)
\end{aligned}$$

from equation (38);

$$= a_{k+1}^2(-1)^{k+1}B_{k+1} + (-1)^kB_k + 2a_{k+1}\left(p_kp_{k-1} - Dq_kq_{k-1}\right)$$

from equation (37);

$$= a_{k+1}^2(-1)^{k+1}B_{k+1} - (-1)^{k+1}B_k + 2a_{k+1}(-1)^kA_{k+1}$$

from equation (39);

$$= (-1)^{k+1}\left(a_{k+1}^2B_{k+1} - \frac{D - A_{k+1}^2}{B_{k+1}} - 2a_{k+1}A_{k+1}\right)$$

from equation (35);

$$\begin{aligned}
&= (-1)^{k+2}\frac{D - a_{k+1}^2B_{k+1}^2 - A_{k+1}^2 + 2a_{k+1}A_{k+1}B_{k+1}}{B_{k+1}} \\
&= (-1)^{k+2}\frac{D - \left(a_{k+1}B_{k+1} - A_{k+1}\right)^2}{B_{k+1}} \\
&= (-1)^{k+2}\frac{D - A_{k+2}^2}{B_{k+1}}
\end{aligned}$$

from equation (34);

$$= (-1)^{k+2}B_{k+2} = \text{R.H.S.}$$

from equation (35). Thus, the theorem is proved by strong induction. □

This theorem is one of the first steps to solving Pell's equation with a general right-hand side. This is much more complicated than the regular Pell's equation, but more information can be found in [5].

6 Conclusion

Over the course of this report we have investigated simple continued fractions and explored an application of solving Pell's equation. Furthermore, a practical method of solving Pell's equation quickly was given in the PQa algorithm. Brahmagupta once said "A person solving the equation $x^2 - 92y^2 = 1$ within a year is a mathematician." [6, Page 73] [1, Page 252] Using the methods in this report I managed to solve it using only a pen and paper in less than 7 minutes. [††] A solution is $x = 1151$ and $y = 120$. A brute force method by hand would take much longer than this and to find a solution within a week would be extremely doubtful. Brahmagupta seemed to think it would take more than a year. Thus, this calculation of less than 7 minutes demonstrates the efficiency of using simple continued fractions to solve Pell's equation.

As stated earlier, this report could lead on to a study of Pell's equation with a generalised right-hand side. But for now we have accomplished our main purpose of solving Pell's equation.

References

[1] Albert H. Beiler. *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* Dover Publications, Inc., New York, NY, 1966.

[2] A. Bogomolny. *Square root of 2 is irrational.* From Interactive Mathematics Miscellany and Puzzles. http://www.cut-the-knot.org/proofs/sq_root.shtml, Accessed 23 January 2009.

[3] A. Ya. Khinchin. *Continued Fractions.* Dover Publications, Inc., Mineola, N.Y., third edition, 1997.

[4] C. D. Olds. *Continued Fractions.* New Mathematical Library. The Mathematical Association of America, Washington, 1963.

[5] John P. Robertson. Solving the generalised Pell equation. Accessible at http://www.jpr2718.org/Pell.pdf., 2004.

[6] John Stillwell. *Mathematics and its History.* Springer-Verlag New York, Inc., New York, NY, second edition, 2002.

[7] Eric W. Weisstein. *Floor Function.* From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/FloorFunction.html, Accessed 03 February 2009.

[8] Eric W. Weisstein. *Fractional Part.* From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/FractionalPart.html, Accessed 03 February 2009.

[††] Furthermore, I managed to solve the notorious $D = 61$ case only using one sheet of paper and a pen in less than 30 minutes, to deduce that $x = 1,776,319,049$ and $y = 226,153,980$