Journal of Integer Sequences, Vol. 13 (2010),
Article 10.4.3

Pell's Equation without Irrational Numbers

N. J. Wildberger
School of Mathematics and Statistics
UNSW Sydney
Australia
n.wildberger@unsw.edu.au

Abstract

We give a simple way to generate an infinite number of solutions to Pell's equation $x^2 - Dy^2 = 1$, requiring only basic matrix arithmetic and no knowledge of irrational numbers. We illustrate the method for $D = 2, 7$ and 61. Connections to the Stern-Brocot tree and universal geometry are also discussed.

1 Introduction

For D a positive square-free integer, Pell's equation is

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

This is perhaps the most important Diophantine equation. Its history goes back to the ancient Greeks, to Archimedes' *Cattle Problem* (see [5, 10]), to Brahmagupta and Bhaskara, to Fermat and Euler, and it was Lagrange who finally established the main fact:

Theorem 1. *Pell's equation has an infinite number of integer solutions.*

Proofs are found in many books and articles on number theory, see for example [1, 2, 7]. Arguments based on continued fractions for quadratic irrationals are common. The aim of this paper is to give a self-contained, simplified proof which only requires basic knowledge of matrix multiplication and *integer arithmetic*. So you need not know what is meant by an *irrational number*.

We follow classical ideas of Lagrange and Gauss, and also insights due to Conway [4] and others. The use of L and R matrices is not new, see for example [8, 9], and the computation of the fundamental automorphism of a binary quadratic form as described in [3, Section 6.12] also avoids irrationalities—but our approach is simpler and requires a minimum of background knowledge.

We show how the method works explicitly in a few special cases, including the famous historical example of $D = 61$, and mention a palindromic aspect of the solution which also sheds light on the negative Pell equation $x^2 - Dy^2 = -1$. While we generate an infinite number of integer solutions, we do not address the issue of whether or not *all* solutions are obtained by this method.

At the end of the paper we briefly connect our solution to paths in the *Stern-Brocot tree*, and mention the connections with *universal geometry*.

We need just a few basic definitions. A *quadratic form* is an expression in variables x and y of the form

$$Q(x,y) = ax^2 + 2bxy + cy^2 \tag{1}$$

where a, b and c are integers. We write $Q \equiv (a,b,c)$ for short. Note that we are following Gauss' convention in using an even coefficient of xy. For a (column) vector $v = (x,y)^T$ define

$$Q(v) \equiv Q(x,y) = (x,y)\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = v^T A v.$$

The symmetric matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

is the *matrix* of the quadratic form Q. The number $ac - b^2$ is the *determinant* of Q.

The quadratic form Q is *balanced* if $a > 0$ and $c < 0$. For example, the *Pell quadratic form*

$$Q_D(x,y) \equiv x^2 - Dy^2$$

for $D > 0$ is balanced since its matrix is

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -D \end{pmatrix}. \tag{2}$$

2 Equivalence of forms

If Q is a quadratic form with matrix A and M is an invertible integral 2×2 matrix whose inverse also has integer entries, then we may define a new quadratic form Q' by

$$Q'(v) \equiv Q(Mv) = (Mv)^T A (Mv) = v^T \left(M^T A M\right) v$$

with matrix $A' = M^T A M$. We say that Q and Q' are *equivalent* forms, and notice that they take on the same values, since if $Q'(v) = n$ then $Q(Mv) = n$ and conversely if $Q(w) = m$ then $Q'(M^{-1}w) = m$. Since the assumptions on M imply that $\det M = \pm 1$, we must have

$$\det A' = \det A$$

so equivalent forms have the same determinant.

For example, the Pell quadratic form $Q_3 = (1, 0, -3)$ is equivalent to the quadratic form $Q' = (-2, -3, -3)$ since

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & -3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -2 & -3 \\ -3 & -3 \end{pmatrix}.$$

The form Q_3 takes the value 1 at $(2,1)^T$, and Q' takes the same value 1 at $(2,-1)^T$. The relationship between these two solutions is

$$\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}\begin{pmatrix}2\\ -1\end{pmatrix} = \begin{pmatrix}2\\ 1\end{pmatrix}.$$

Note that while Q_3 is balanced, Q' is not.

In practice we will focus on two very simple equivalences, using the following matrices:

$$L \equiv \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix} \qquad \text{and} \qquad R \equiv \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}.$$

For a symmetric matrix

$$A = \begin{pmatrix}a & b\\ b & c\end{pmatrix}$$

we compute

$$L^T AL = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}^T \begin{pmatrix}a & b\\ b & c\end{pmatrix}\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix} = \begin{pmatrix}a+2b+c & b+c\\ b+c & c\end{pmatrix} \tag{3}$$

and

$$R^T AR = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}^T \begin{pmatrix}a & b\\ b & c\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}a & a+b\\ a+b & a+2b+c\end{pmatrix}. \tag{4}$$

Going from A to $L^T AL$ we call a *left step*. Going from A to $R^T AR$ we call a *right step*.

3 Solving Pell's equation

Here is our strategy for generating an infinite number of solutions to Pell's equation $x^2 - Dy^2 = 1$.

1. First observe that $e = (1,0)^T$ is a solution.

2. Perform a *suitable (non-empty) sequence of left and right steps*, beginning with the Pell quadratic form $Q_D = (1, 0, -D)$ with matrix (2), and ending also with $Q_D = (1, 0, -D)$. This yields a matrix equation of the form

$$N^T AN = A \tag{5}$$

   for some invertible matrix N with *strictly positive entries*, actually of the form

$$N = \begin{pmatrix}u & Dv\\ v & u\end{pmatrix}.$$

3. Then observe that $Ne = (u,v)^T$, the first column of N, is also a solution, since

$$(Ne)^T A\,(Ne) = e^T \left(N^T AN\right) e = e^T Ae = 1.$$

4. Iterate to generate an infinite family of solutions, namely $e, Ne, N^2e, \cdots$.

Let us now explain what the *suitable sequence of left and right steps is*, and why we can be sure to return to Q_D after a finite number of such steps. The key is simply to *ensure that at each stage we obtain a balanced form.*

If at some stage we have a balanced form (a, b, c), we compute the *total*

$$T \equiv a + 2b + c.$$

If $T > 0$ we *perform a left step.* If $T < 0$ we *perform a right step.* Then (3) and (4) ensure that the new form (a', b', c') is also balanced, and it is clearly equivalent to (a, b, c).

The case $T = 0$ cannot arise if the determinant of (a, b, c) is $-D$, since the identity

$$-D = ac - b^2 = (a + 2b + c)\, c - (b + c)^2$$

would then imply that D is a square.

To record the step at each stage, we write either

$$(a, b, c)\, L\, (a + 2b + c, b + c, c)$$

or

$$(a, b, c)\, R\, (a, a + b, a + 2b + c)\, .$$

Suppose we have taken a step from the balanced quadratic form (a, b, c) to the balanced quadratic form (a', b', c') . Can we tell whether this step was left or right? In the case of a left step $a' - 2b' + c' = a > 0$ while in the case of a right step $a' - 2b' + c' = c < 0$. So the total of the quadratic form $(a', -b', c')$ determines whether the step was left or right.

If we start with the quadratic form $Q_D = (1, 0, -D)$ with determinant $-D < 0$, and arrive after a certain number of steps to the quadratic form (a, b, c) then we must have $a > 0$, $c < 0$ and

$$ac - b^2 = -D.$$

Since ac is negative, there are only a finite number of integers a, b and c that satisfy this equation. So we must eventually return to a form that we have encountered before, and since we are capable of finding the inverse of any step, we must return first to the initial form, namely Q_D.

At the beginning with Q_D the total is $T = 1 - D < 0$ so the first step must be a right step. Further right steps increase the last entry of $(1, b, c)$, so before we return to Q_D we must have taken at least one left step. So the cumulative matrix N, which is the product of the matrices R and L corresponding to the steps we have taken, has strictly positive entries. This ensures that the sequence of solutions $e, Ne, N^2e, \cdots$ is increasing in each component, and so really does represent an infinite family of solutions.

If

$$N = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

then more generally (5) gives us the identity

$$(\alpha x + \beta y)^2 - D\, (\gamma x + \delta y)^2 = x^2 - Dy^2$$

which allows us to generate new solutions to $Q(x, y) = n$ from existing ones, for any integer n. It also allows us to deduce that N has the form

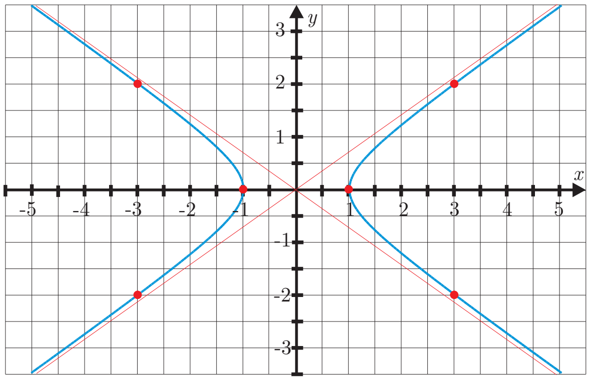
$$N = \begin{pmatrix} u & Dv \\ v & u \end{pmatrix}.$$

4 Example of $D = 2$

The simplest example of Pell's equation is

$$x^2 - 2y^2 = 1.$$

The ancient Greeks knew how to generate solutions. Geometrically we are looking for integer points on a hyperbola, as shown.



The hyperbola $x^2 - 2y^2 = 1$

We begin with the Pell quadratic form $Q_2 = (1, 0, -2)$ with a total of $T = 1 + 2 \times 0 - 2 = -1 < 0$ so we take a right step to get

$$(1, 0, -2)\, R\, (1, 1, -1).$$

This corresponds to the matrix equation

$$R^T \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} R = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The total is now $T = 1 + 2 \times 1 - 1 = 2 > 0$ so we take a left step to get

$$(1, 0, -2)\, R\, (1, 1, -1)\, L\, (2, 0, -1).$$

The total is now $T = 1$ so we take another left step

$$(1, 0, -2)\, R\, (1, 1, -1)\, L\, (2, 0, -1)\, L\, (1, -1, -1).$$

The total is now $T = -2 < 0$ so we take a right step

$$(1, 0, -2)\, R\, (1, 1, -1)\, L\, (2, 0, -1)\, L\, (1, -1, -1)\, R\, (1, 0, -2) \tag{6}$$

to return to our starting quadratic form Q_2. This yields the matrix equation

$$\left(RL^2R\right)^T \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} RL^2R = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}.$$

So

$$N = RL^2R = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix}$$

gives the non-trivial solution $v = Ne = (3,2)^T$ to Pell's equation, and by taking higher powers of N we get further solutions $(17,12)^T$, $(99,70)^T$, $(577,408)^T$ and so on.

More generally the form of N establishes the identity

$$(3x+4y)^2 - 2\,(2x+3y)^2 = x^2 - 2y^2$$

which allows us to generate an infinite number of solutions to $Q_2\,(x,y) = n$ from any initial solution.

5 Example of $D = 7$

The quadratic form $Q_7 = (1,0,-7)$ gives the stepping sequence

$$\begin{gathered}(1,0,-7)\,R\,(1,1,-6)\,R\,(1,2,-3)\,L\,(2,-1,-3)\\ R\,(2,1,-3)\,L\,(1,-2,-3)\,R\,(1,-1,-6)\,R\,(1,0,-7)\,.\end{gathered}$$

This corresponds to the matrix equation

$$\left(R^2LRLR^2\right)^T \begin{pmatrix} 1 & 0 \\ 0 & -7 \end{pmatrix} R^2LRLR^2 = \begin{pmatrix} 1 & 0 \\ 0 & -7 \end{pmatrix}.$$

The first column of

$$N = R^2LRLR^2 = \begin{pmatrix} 8 & 21 \\ 3 & 8 \end{pmatrix}$$

gives the solution $(8,3)^T$, and higher powers give also $(127,48)^T$, $(2024,765)^T$ and so on. Furthermore we have the identity

$$(8x+21y)^2 - 7\,(3x+8y)^2 = x^2 - 7y^2$$

which generates further solutions to any solution of $x^2 - 7y^2 = n$.

6 Palindromic aspect

The palindromic nature of the sequences occurring for N in the above two examples is a general feature, and follows from the fact that the inverse of a left step is

$$(a',b',c')\,L^{-1}\,(a'-2b'+c',b'-c',c')$$

while the inverse of a right step is

$$(a', b', c')\, R^{-1}\, (a', b' - a', a' - 2b' + c')\,.$$

So if we negate the middle b components, the sequence reads the same from the end as from the beginning. To determine when we have arrived at the half-way point we look for symmetry: two adjacent or almost adjacent (separated by one) quadratic forms which differ only in the sign of the middle term.

The central symmetry allows us to identify when the *negative* Pell equation $x^2 - Dy^2 = -1$ is soluble, and give a positive solution for it, if such exists. This will happen when the central part of the sequence looks like

$$(a', b', c')\, (D, 0, -1)\, (a', -b', c')$$

In that case N can be written as $N = MM^T$ with M of the form

$$M = \begin{pmatrix} Dv_1 & u_1 \\ u_1 & v_1 \end{pmatrix}$$

so that its second column gives (u_1, v_1) satisfying $u_1^2 - Dv_1^2 = -1$.

For the $D = 2$ case, (6) shows that

$$M = RL = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

giving $(u_1, v_1) = (1, 1)$. The $D = 7$ case does not have this form, and indeed $x^2 - 7y^2 = -1$ has no solutions, as -1 is not a square mod 7.

We will further exploit this palindromic phenomenon in our last example.

7 Example $D = 61$

The Pell's equation $x^2 - 61y^2 = 1$ was recognized by Fermat as being particularly challenging, and he posed it as a problem to his contemporaries, which was subsequently solved by Euler. However according to [2] the eleventh century Indian mathematician Bhaskara solved this problem. We shorten the presentation by combining powers of L and R when convenient. The required sequence of left and right steps exhibits an additional doubly palindromic aspect, in that each half is itself palindromic in an opposite way, that is with L and R symbols interchanged:

$$\begin{aligned}
&(1, 0, -61)\, R^7\, (1, 7, -12)\, L\, (3, -5, -12)\, R^4\, (3, 7, -4)\, L^3\, (9, -5, -4) \\
&R\, (9, 4, -5)\, L\, (12, -1, -5)\, L\, (5, -6, -5)\, R\, (5, -1, -12)\, R\, (5, 4, -9) \\
&L\, (4, -5, -9)\, R^3\, (4, 7, -3)\, L^4\, (12, -5, -3)\, R\, (12, 7, -1)\, L^7\, (61, 0, -1) \\
&L^7\, (12, -7, -1)\, R\, (12, 5, -3)\, L^4\, (4, -7, -3)\, R^3\, (4, 5, -9)\, L\, (5, -4, -9) \\
&R\, (5, 1, -12)\, R\, (5, 6, -5)\, L\, (12, 1, -5)\, L\, (9, -4, -5)\, R\, (9, 5, -4) \\
&L^3\, (3, -7, -4)\, R^4\, (3, 5, -12)\, L\, (1, -7, -12)\, R^7\, (1, 0, -61)
\end{aligned}$$

Computation of the required matrix N is shortened by the doubly palindromic aspect:

$$\begin{aligned}
N &= R^7 L R^4 L^3 R L^2 R^2 L R^3 L^4 R L^7 L^7 R L^4 R^3 L R^2 L^2 R L^3 R^4 L R^7 \\
&= \begin{pmatrix} 453 & 164 \\ 58 & 21 \end{pmatrix} \begin{pmatrix} 453 & 58 \\ 164 & 21 \end{pmatrix} \begin{pmatrix} 21 & 58 \\ 164 & 453 \end{pmatrix} \begin{pmatrix} 21 & 164 \\ 58 & 453 \end{pmatrix} \\
&= \begin{pmatrix} 1766\,319\,049 & 13\,795\,392\,780 \\ 226\,153\,980 & 1766\,319\,049 \end{pmatrix}
\end{aligned}$$

and its first column gives the smallest non-trivial solution $x = 1766\,319\,049$ and $y = 226\,153\,980$. The occurrence of $(61, 0, -1)$ in the middle of the sequence shows that the negative Pell equation has a solution, and indeed the second column of

$$M = \begin{pmatrix} 453 & 164 \\ 58 & 21 \end{pmatrix} \begin{pmatrix} 453 & 58 \\ 164 & 21 \end{pmatrix} = \begin{pmatrix} 232\,105 & 29\,718 \\ 29\,718 & 3805 \end{pmatrix}$$

satisfies $(29\,718)^2 - 61\,(3805)^2 = -1$.

8 Speeding up the algorithm

The previous example suggests that for large D, it is more efficient to work out how many *successive* R or L steps are needed at each stage. Since

$$R^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad L^m = \begin{pmatrix} 1 & 0 \\ m & 1 \end{pmatrix}$$

it is easy to compute that for arbitrary natural numbers n and m,

$$\begin{aligned}
(a, b, c)\, R^n &= \left(a, b + an, c + 2bn + an^2\right) \\
(a, b, c)\, L^m &= \left(a + 2bm + cm^2, b + cm, c\right).
\end{aligned}$$

So if a right step R is needed, namely if $T = a + 2b + c < 0$, then we find the smallest k so that $c + 2bk + ak^2 > 0$, and choose $n = k - 1$. If a left step L is needed, namely if $T = a+2b+c > 0$, then we find the smallest l so that $a+2bl+cl^2 < 0$, and choose $m = l-1$.

So for example at the first stage of the $D = 61$ case, $(1, 0, -61)$ has a total of $T = -61 < 0$ and since $(-61) + k^2 > 0$ first when $k = 8$, we set $n = 8 - 1 = 7$, giving

$$(1, 0, -61)\, R^7 = (1, 7, -12).$$

Now we know the next step must be L^m for some m, and we compute that $1+2\times 7l-12l^2 < 0$ first when $l = 2$, so that $m = 2 - 1 = 1$, giving

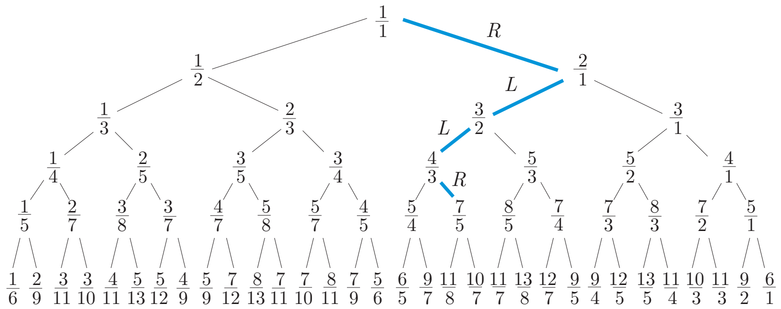
$$(1, 0, -61)\, R^7\, (1, 7, -12)\, L = (3, -5, -12)$$

and so on.

9 Connections with the Stern-Brocot tree

The Stern-Brocot tree (see [6] for a thorough discussion) is generated from the two 'fractions' 0/1 and 1/0 by repeatedly performing the mediant operation. It naturally associates binary sequences in L and R with strictly positive reduced fractions, as shown for the correspondence

$$RL^2R \leftrightarrow \frac{7}{5}.$$



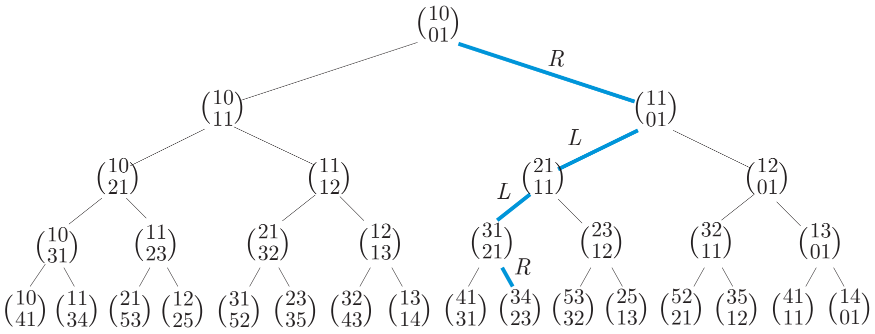
The Stern Brocot tree

The Stern-Brocot tree is closely related to the Euclidean algorithm and the theory of continued fractions, and there is a useful alternate form which contains 2×2 matrices instead of fractions. The matrix corresponding to a fraction a/b has columns which are the two immediate left and right predecessors of a/b in the original tree, and whose columns sum to give the fraction, regarded as a vector. A binary sequence of L's and R's can be directly translated to a matrix by interpreting L and R to be the two matrices we used earlier, and simply performing the matrix multiplication.

So computing N in the Pell's equation solution amounts to following a path in the matrix form of the Stern-Brocot tree. Each step involves adding one of the existing columns to the other, either to the left or to the right. Shown is the case

$$N = RL^2L = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix}$$

which occurred in the solution of $x^2 - 2y^2 = 1$.



Matrix form of the Stern-Brocot tree

10 Geometric interpretation

We have been able to avoid irrationalities because numbers of the form $a + b\sqrt{D}$ behave arithmetically exactly as do matrices of the form

$$\begin{pmatrix} a & Db \\ b & a \end{pmatrix}.$$

This will be familiar to most readers in the imaginary case of $D = -1$, but it holds more generally, see for example [8]. It allows us to consider a quadratic extension of the rationals not only as a subfield of the complex numbers, but also as *a field of rotation/dilation operators* on a two-dimensional (non-Euclidean) geometry *purely over the rational numbers*.

The usual approach to metrical geometry is to consider distance and angle as the fundamental concepts, but this is neither in the original spirit of the ancient Greeks, nor an approach that lends itself easily to generalization. Taking the *quadrance* $x^2 + y^2$ as the fundamental measurement, instead of its square root, together with an algebraic replacement of angle called the *spread* between two lines, allows the development of a Euclidean theory of geometry that extends both to general fields, and also to arbitrary quadratic forms. In particular there is a very rich theory of Euclidean geometry over the rational numbers!

This new *universal geometry* is described in [11] and [12]. Our discussion of Pell's equation thus has a direct *geometric interpretation* (actually this is the context in which it was discovered). So for example the matrix

$$N = \begin{pmatrix} 8 & 21 \\ 3 & 8 \end{pmatrix}$$

in the case of $D = 7$ amounts to a *rotation* in the geometry for which *quadrance* is defined by $Q(x, y) = x^2 - 7y$.

Finally we point out that a quadratic irrational number such as $\sqrt{2}$ manifests itself as a periodic infinite path down the Stern-Brocot tree. A more general irrational number can be defined as a more general infinite path down the Stern-Brocot tree. This approach to the definition of irrationalities has the significant advantages that it involves no ambiguities related to equivalences, and that the initial paths of an irrational give automatically the best rational approximations to it.

References

[1] H. Cohn, *Advanced Number Theory*, Dover, 1980, pp. 110–111.

[2] E. J. Barbeau, *Pell's Equation*, Problem Books in Mathematics. Springer-Verlag, 2003.

[3] J. Buchmann and U. Vollmer, *Binary Quadratic Forms: An Algorithmic Approach,* Algorithms and Computation in Mathematics **20**, Springer, 2007.

[4] J. H. Conway, *The Sensual (Quadratic) Form*, Carus Mathematical Monographs **26**, Mathematical Association of America, 2005.

[5] H. Dorrie, *100 Great Problems of Elementary Mathematics: Their History and Solution*, Dover, 1965.

[6] R. L. Graham. D. E. Knuth and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, Reading, 1994.

[7] H. W. Lenstra Jr., Solving the Pell equation, *Notices Amer. Math. Soc.* **49**, (2002), 182–192.

[8] A. J. van der Poorten, An introduction to continued fractions, in *Diophantine Analysis*, LMS Lecture Note Ser. **109**, Cambridge University Press, 1985, pp. 99–138.

[9] S. Raney, On continued fractions and finite automata, *Math. Annalen* **206** (1973), 265–283.

[10] I. Vardi, Archimedes' cattle problem, *Amer. Math. Monthly* **105** (1998), 305–319.

[11] N. J. Wildberger, *Divine Proportions: Rational Trigonometry to Universal Geometry*, Wild Egg Books, Sydney, 2005.

[12] N. J. Wildberger, One dimensional metrical geometry, *Geometriae Dedicata*, **128** (1), (2007), 145–166.

2010 *Mathematics Subject Classification*: Primary 11D09; Secondary 11E16.
Keywords: Pell's equation, quadratic form, Stern-Brocot tree, universal geometry.

Received December 21 2009; revised version received March 22 2010. Published in *Journal of Integer Sequences*, March 23 2010.

Return to Journal of Integer Sequences home page.