PELL’S EQUATION AND CONTINUED FRACTIONS

I. GREGOR DICK

Abstract. This paper first provides an exposition of the theory of simple continued fractions; it then discusses Pell’s equation \( x^2 - dy^2 = N \) and a method of Lagrange for obtaining its solutions; and finally it presents another method of solution relying on the continued fraction expansion of \( \sqrt{d} \) which is shown to be equivalent to Lagrange’s method.

1. Introduction

Pell’s equation,

\[
x^2 - dy^2 = N,
\]

is a Diophantine equation with \( d \) and \( N \) being nonzero integers. Its study is known to have been conducted as early as 400 B.C., in the context of a variety of geometric and combinatorial problems [4]. In more recent times the discovery of its solutions was proposed as a challenge by Fermat. John Pell contributed little to its analysis, but instead the equation acquired its eponym due to a misattribution by Euler.

The objects of this paper are two-fold: firstly, to describe a method of Lagrange for solving (1.1); and then also to consider an alternative method relying on the theory of continued fractions, which provide a representation of real numbers, and whose properties we begin by exhibiting.

2. Continued Fractions

2.1. Finite Simple Continued Fractions. Let \( n \) be a non-negative integer, and let \( \{a_i\}_{0 \leq i \leq n} \) be a sequence of positive integers, except perhaps \( a_0 \) which may also be zero or negative. Define the finite simple continued fraction \([a_0; a_1, a_2, \ldots, a_n]\) thus:

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

If the sequence \( \{a_n\} \) is allowed to consist of any real numbers, then the corresponding continued fraction is not simple; however, in this paper we will restrict ourselves to the consideration of simple continued fractions, and so the appellation ‘simple’ will often be omitted.

It is clear that the value of a finite simple continued fraction is a rational number. The converse is also true:

Theorem 2.1. Let \( q \in \mathbb{Q} \). Then there exists a finite simple continued fraction \([a_0; a_1, a_2, \ldots, a_n]\) such that \( q = [a_0; a_1, a_2, \ldots, a_n] \). Such a continued fraction is said to be a continued fraction expansion of \( q \).

A proof is given in [1] in which \( q \) is expressed in the form \( a/b \) with \( (a, b) = 1 \), and the Euclidean algorithm is applied to \( a \) and \( b \) to yield the expansion of \( q \).

Date: October 16, 2008.

This project was funded by a Vacation Scholarship from the Carnegie Trust.

The author wishes to express his thanks to Chris Smyth of the University of Edinburgh for his kind supervision of this project.
Definition 2.2. Let \( x = [a_0; a_1, a_2, \ldots, a_n] \) be a finite continued fraction. Then the \( a_i \) are called partial quotients of the continued fraction, and, for \( 0 \leq m \leq n \), \([a_0; a_1, a_2, \ldots, a_m]\) is called the \( m \)-th convergent to \( x \).

The continued fraction expansion of a rational number can also be shown to be unique up to a trivial pairing of expansions.

Theorem 2.3. Let \( A = [a_0; a_1, a_2, \ldots, a_m] \) and \( B = [b_0; b_1, b_2, \ldots, b_n] \) be finite continued fractions with \( m \leq n \) such that \( A = B \). Then either \( m = n \) and \( a_i = b_i \) for all \( 0 \leq i \leq n \), or \( m + 1 = n \) with \( b_m = a_m - 1 \) and \( b_n = 1 \).

Again, a proof of this theorem is given in [1].

2.2. Infinite Simple Continued Fractions. The notation in the previous section can be expanded in the natural fashion to admit infinite simple continued fractions:

\[
[a_0; a_1, a_2, \ldots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots}}.}
\] (2.2)

It is not immediately clear that such a fraction is well-defined. The demonstration of its convergence is extraneous to the purposes of this paper, but is treated of fully in [1], where it is furthermore shown to be irrational. For the present discussion, it is sufficient to consider one particular class of infinite continued fraction.

Definition 2.4. Let \( \xi = [a_0; a_1, a_2, \ldots] \) be an infinite continued fraction such that there exist \( n \geq 0 \) and \( r > 0 \) with the property that \( a_i = a_i + r \) for all \( i \geq n \). Then \( \xi \) is said to be periodic with period \( r \), and is written thus:

\( \xi = [a_0; a_1, a_2, \ldots, a_n, a_{n+1}, \ldots a_{n+r-1}] \).

If \( n = 0 \), \( \xi \) is said to be purely periodic, and can be written as follows:

\( \xi = [a_0, a_1, \ldots, a_{r-1}] \).

The following result is proved at length in [1]:

Theorem 2.5. Let \( d \) be a positive integer not a perfect square. Then the continued fraction expansion of \( \sqrt{d} \) is periodic and is of the form

\( d = [a_0; a_1, a_2, \ldots, a_{r-1}, 2a_0] \),

for some positive integer \( r \). Furthermore, the following iterative scheme yields the partial quotients \( a_i \):

\[
a_i = \lfloor \xi_i \rfloor \\
q_i = m_i + \sqrt{d} \\
m_{i+1} = a_i q_i - m_i \\
q_{i+1} = \frac{d - m_{i+1}^2}{q_i},
\]

where \( a_0 = m_0 = 0 \) and \( q_0 = 1 \).

3. Pell’s Equation

We return now to the study of equation (1.1).
3.1. Initial Observations. We begin by noting that when \( d < 0 \) and when \( d \) is a perfect square, equation (1.1) has only finitely many solutions; in the former case because both summands on the left-hand side become strictly positive, and in the latter case because the left-hand side factorises over the integers. We limit ourselves to considering solutions when \( d > 0 \), not a perfect square.

An exercise in [1] poses the problem of demonstrating that (1.1) has no solutions when \( N = -1 \) and \( d \equiv 3 \pmod{4} \). This can be seen by reducing the equation modulo 4 to give \( x^2 + y^2 \equiv 3 \pmod{4} \). Thus without loss of generality a solution would require \( x^2 \equiv 0 \) (resp.) \( 1 \pmod{4} \) and \( y^2 \equiv 3 \) (resp.) \( 2 \pmod{4} \). However, there are no elements of the ring \( \mathbb{Z}/4\mathbb{Z} \) that square to either 2 or 3, and so the equation has no solution. A similar argument shows that there is no solution when \( d \equiv 0 \pmod{4} \). Indeed, for any \( n \), if there is no element of the ring \( \mathbb{Z}/n\mathbb{Z} \) which squares to \( n - 1 \), then there is no solution when \( d \equiv 0 \pmod{n} \).

A similar problem is posed in [2, question 5]:

**Problem 3.1.** Show that if \( d \) is an odd squarefree integer then a necessary condition for the equation \( x^2 - dy^2 = -1 \) to have an integer solution is that all the primes dividing \( d \) must be of the form \( 4n + 1 \).

**Solution.** Let \( p = 4n + 3 \) be a prime, where \( n \) is a non-negative integer, such that \( p \mid d \). Reducing \( x^2 - dy^2 \equiv -1 \) modulo \( p \) gives \( x^2 \equiv -1 \pmod{p} \), whence \( o_{\mathbb{Z}_p^*}(x) = 4 \), where \( o_{\mathbb{Z}_p^*}(x) \) denotes the order of \( x \) in the multiplicative group \( \mathbb{Z}_p^* \). Then Lagrange’s theorem implies that \( 4 \mid (p - 1) \), which contradicts the definition of \( p \). Hence all the primes dividing \( d \) must be of the form \( 4n + 1 \).

3.2. Lagrange’s Method. Lagrange developed a method for solving (1.1) in the general case, which he describes in [3]. The process is recursive and is rather involved and so is not detailed here for the sake of brevity; however, the terminating case is considered to allow comparison with the method of solution described in section 4.1, infra. The form given here is based on that included in [4].

**Theorem 3.2** (Lagrange). Consider Pell’s equation

\[
\pm E = r^2 - B s^2, \tag{3.1}
\]

where \( B > 0 \) and \( 0 < E < \sqrt{B} \). Construct sequences \( \{E_i\} \), \( \{\epsilon_i\} \) and \( \{\lambda_i\} \) thus:

\[
E_0 = E, \quad E_i E_{i+1} = B - \epsilon_i^2, \quad i \geq 0
\]

\[
\sqrt{B} - E < \epsilon_0 < \sqrt{B}, \quad E \mid (B - \epsilon_0^2)
\]

\[
\epsilon_i = \lambda_i E_{i-1} - \epsilon_{i-1}, \quad i \geq 1
\]

\[
\sqrt{B} + \epsilon_{i-1} \left( \frac{E_{i+1}}{E_i} \right) - 1 < \lambda_i < \sqrt{B} + \epsilon_{i-1} \left( \frac{E_{i+1}}{E_i} \right), \quad i \geq 1.
\]

If no such \( \epsilon_0 \) exists, the equation is insoluble. Otherwise, the sequence \( \{E_i\} \) has a periodic tail beginning after \( E_0 \). Let the length of the periodic part be \( \mu \). Then there exists at least one \( m, 1 \leq m \leq \mu \) such that \( E_m = 1 \). If all terms in the sequence \( \{E_i\} \) are equal to unity, take \( m = 0 \) as a special case. Then for each even (resp. odd) such value of \( m \) when the upper (resp. lower) sign in (3.1) is taken, the equation has a family of solutions given by \( r + s \sqrt{B} = (R + S \sqrt{B})(X + Y \sqrt{B})^n \) for all non-negative integers \( n \), where \( R, S, X \) and \( Y \) are given by

\[
R = \beta_{l_{m-1} + l_{m-2}}, \quad S = l_{m-1}
\]

\[
X = \beta_{l_{\mu-1} + l_{\mu-2}}, \quad Y = l_{\mu-1}
\]

where \( \beta \) is the greatest integer less than \( \sqrt{B} \) and the sequence \( \{l_i\} \) is given by

\[
l_{-1} = 0, \quad l_0 = 1 \quad \text{and} \quad l_i = \lambda_i l_{i-1} + l_{i-2} \text{ for } i \geq 1.
\]

In the case \( m = 0 \), the solution is given by \( (r, s) = (1, 0) \) if the upper sign is taken in (3.1); if the lower sign is taken, no solutions exist. If no such values of \( m \) exist, the equation is insoluble.
In the course of this project, the algorithm described in Theorem 3.2 was implemented, finding the smallest solution to (1.1) for arbitrary values of \(d\) and \(N\).

4. An Alternative to Lagrange’s Method

4.1. Method of Continued Fractions. A method for solving equation (1.1) using continued fractions is given in [5], as follows.

**Theorem 4.1.** Consider the equation
\[
x^2 - dy^2 = 1.
\] (4.1)

Let \(\xi_i\) be the \(i\)-th convergent to \(\sqrt{d}\), and write \(\xi_i = \frac{p_i}{q_i}\) with \(\gcd(p_i, q_i) = 1\) for all integers \(i\). Then there is a least integer \(k\) such that \((x, y) = (p_k, q_k)\) is a solution to (4.1), and the general solution is given by
\[
x + y\sqrt{d} = (p + q\sqrt{d})^n,
\]
for all positive integers \(n\).

**Lemma 4.2.** The equation \(x^2 - dy^2 = -1\) has infinitely many solutions if the length of the periodic part of the continued fraction expansion of \(\sqrt{d}\) is odd, and none if it is even.

In [5] is also given the following more general theorem, accommodating larger values on the right-hand side of (4.1).

**Theorem 4.3.** Let \(\sqrt{d}\) in (1.1) have continued fraction expansion
\[
[a_0; a_1, a_2, \ldots, a_{r-1}, 2a_0]
\]
for some positive integer \(r\), which form is attained by the result of Theorem 2.5. Let \(\xi_i = \frac{p_i}{q_i}\) be defined as in Theorem 4.1. Then, if \(N < \sqrt{d}\), there exists an integer \(k\) with \(0 \leq k \leq r\) such that \((x, y) = (p_k, q_k)\) is a solution to (1.1). Furthermore, if \((x, y) = (r, s)\) is any solution to \(x^2 - dy^2 = 1\), then
\[
\begin{align*}
x &= p_k r \pm dq_k s \\
y &= p_k s \pm q_k r
\end{align*}
\] (4.2)
is also a solution of (1.1).

**Remark 4.4.** The \(k\) in Theorem 4.3 need not be unique. In this case, there is a family of solutions of the form given in (4.2) for each such \(k\).

The following problem from [2, question 6] illustrates an application of Lemma 4.2.

**Problem 4.5.** Let \(d = 4k^2 + k\), where \(k > 1\) is an integer. Show that \(x^2 - dy^2 = -1\) has no integer solutions.

**Solution.** The truth of the proposition can be demonstrated by applying the iterative scheme for the partial quotients \(a_i\) of the continued fraction expansion of \(\sqrt{d}\) given in Theorem 2.5. Observing that \([\sqrt{4k^2 + k}] = 2k\) for all positive integers \(k\), we tabulate \(a_i, m_i\) and \(q_i\) as defined in Theorem 2.5:

<table>
<thead>
<tr>
<th>(i)</th>
<th>(a_i)</th>
<th>(m_i)</th>
<th>(q_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2k</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>2k</td>
<td>(k)</td>
</tr>
<tr>
<td>2</td>
<td>4k</td>
<td>2k</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2k</td>
<td>(k)</td>
</tr>
</tbody>
</table>

Observing that \(a_i, m_i\) and \(q_i\) are determined entirely by \(a_{i-1}, m_{i-1}\) and \(q_{i-1}\) for any \(i \geq 1\), it follows that the period of the continued expansion fraction of \(\sqrt{d}\) is 2, and hence \(x^2 - dy^2 = -1\) has no solutions by Lemma 4.2. \(\square\)
4.2. **Equivalence To Lagrange’s Method.** Although not immediately apparent, this method for solving Pell’s equation can be shown to be equivalent to that of Lagrange by comparing the sequences \( \{E_i\} \), \( \{\lambda_i\} \) and \( \{\epsilon_i\} \) in the latter to \( \{q_i\} \), \( \{a_i\} \) and \( \{m_i\} \) respectively in Theorem 2.5, and by noting that Lagrange’s expressions for the solutions to (3.1) in terms of these sequences corresponds to the formulæ given for the numerators and denominators of successive convergents in, for example, [1].

**References**