PELL'S EQUATION

by

ROY LOUIS NEFF

B. S., Bethany College, 1964

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

Approved by:

R L Yates

Major Professor

TABLE OF CONTENTS

INTRODUCTION

The second degree Diophantine equation of the form $x^2-Dy^2 = N$, with the qualification that D and N be integers, is referred to as Pell's equation. The attachment of Pell's name to this equation form, however, was due to an error on Euler's part rather than to Pell's contribution to the solution of the equation. The early Greeks and Hindus considered special cases of this general equation form and, specifically, the Hindus were able to solve the particular form $x^2-Dy^2 = 1$.

Fermat was the first to deal systematically with the general equation form but he chose not to publish specific proofs. There remains only an outline of Fermat's proof that there exists an infinite number of solutions to the equation $x^2-Dy^2 = 1$. Lagrange published the first proof of the existence of a solution in this specific case, using the theory of continued fractions. Wallis and Lord Brouncker also found a solution and published it in 1658. Prior to this, Euler had shown that there are infinitely many solutions if there is one. It was Euler and Lagrange whose contributions furthered the development of the solution of the general equation form for $|N| > 1$.

Application of the solutions of Pell's equation is seen most obviously in finding integral solutions of the general

quadratic form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

in which a, b, c, d, e, f are integers. Considering the left hand side of the equality first as a polynomial in x, the discriminant

$$(by + d)^2 - 4a(cy^2 + ey + f)$$

must be a perfect square if one is to get integral solutions. Consider this discriminant now as a polynomial in y:

$$(b^2 - 4ac)y^2 + (2bd - 4ae)y + d^2 - 4af.$$

Let this polynomial in y be called z^2 and, for purposes of algebraic simplification, set

$$b^2 - 4ac = p,\ 2bd - 4ae = q,\ d^2 - 4af = r.$$

One now has the situation

$$py^2 + qy + r = z^2\ ,$$

or

$$py^2 + qy + r - z^2 = 0.$$

Considering the discriminant of the latter equation and the original intent of this investigation, one finds that $q^2 - 4p(r - z^2)$ must be a perfect square if one is to obtain the desired integral solutions to the original equation. If this discriminant is called w^2, it reduces to $q^2 - 4p(r - z^2) = w^2$, or the Pell equation $w^2 - 4pz^2 = q^2 - 4pr$.

If this last equation is solvable, one can then present rational solutions to the original equation and, hopefully, find integral solutions among this list of rational solutions.

A second utilization of solutions of Pell's equation arises in finding units of the multiplicative integral domain $R\left[\sqrt{D}\right]$, where D is a square free integer. One finds that if $D \equiv 2$ or $3 \pmod 4$, the units are given by the solutions of

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

and if $D \equiv 1 \pmod 4$, the units are the integers of the form $\frac{x+y\sqrt{D}}{2}$, where $x+y\sqrt{D}$ is a solution of $x^2 - Dy^2 = \pm 4$.

A third area of application of the solutions of the general Pell equation arises in finding convergents of a continued fraction. If it is required that $D > 0$ and $|N| < \sqrt{D}$, with N and D being integers and D not a perfect square, then all positive solutions of $x^2 - Dy^2 = N$ are such that x/y is a convergent of the continued fraction expansion of $\sqrt{D}$.

The consideration which follows will be broken down into three main parts: First, where $|N| = 1$ specifically; second, where $|N| = 4$ specifically; third, where $|N| > 0$ in general. It should be noted that the notational forms p,q and $p+q\sqrt{D}$ for solutions of the Pell equation are equivalent and will be used interchangeably throughout the paper.

THE CASE $|N| = 1$

In the case where $|N| = 1$, consideration will be given first to the specific form $x^2-Dy^2=1$ of the Pell equation. A considerable amount of restriction can be done at the outset of this discussion. If $D=-1$, one is left with the equation $x^2+y^2=1$, which has only the trivial integral solutions $(\pm 1, 0)$ and $(0, \pm 1)$. If $D < -1$, obviously one is left with $(\pm 1, 0)$ as the only possible integral solutions to the resulting equation. Finally, if D is a perfect square, one is left with the equation $x^2-(\sqrt{Dy})^2=1$, or, what is the same:

$$x^2-K^2=1 ,$$

where $K=\sqrt{Dy}$. One notes, however, that the only two integral perfect squares which differ by one are 1 and 0. It follows, then, that the only possible integral solutions to this equation are $(\pm 1, 0)$. For the duration of the discussion of this first case, therefore, it will be assumed that D is a positive integer which is not a perfect square.

Continuing now with these restrictions on the equation

$$x^2-Dy^2=1 , \qquad (1)$$

it is sought to establish the existence of solutions of equation (1) other than the trivial ones $(\pm 1, 0)$.

THEOREM 1. There exist positive integers p and q such that the absolute value of the real quantity $p-q\sqrt{D}$ is less than any arbitrarily small positive quantity E and, consequently, less than $1/q$.

Proof. Choose an integer t such that $tE > 1$ and let q take on successively the integral values from 0 to t. For each such choice of value of q, assign to p the least integral value greater then $q\sqrt{D}$, i.e., $p_i = q_i\sqrt{D} + d_i$, where $0 \leq d_i \leq 1$ for all $0 \leq i \leq t$. The quantity $p-q\sqrt{D}$, then, lies between 0 and 1 for all such choices of p and q since

$$p_i - q_i\sqrt{D} = q_i\sqrt{D} + d_i - q_i\sqrt{D} = d_i ,$$

and it has been specified that $0 \leq d_i \leq 1$ for all $0 \leq i \leq t$. Also it is noted that for no two separate choices of p and q, say p_i, q_i and p_j, q_j with $i \neq j$, are the quantities $p_i - q_i\sqrt{D}$ and $p_j - q_j\sqrt{D}$ equal. This inequality follows by assuming that two such quantities are equal and establishing a contradiction. If, for $i \neq j$, $p_i - q_i\sqrt{D} = p_j - q_j\sqrt{D}$, then

$$(p_i - p_j) = (q_i - q_j)\sqrt{D} .$$

p_i and p_j, however, are integral so that their difference will be integral. This equality, then, with the existing restrictions on D, can hold only if $q_i = q_j$. However, this is a

contradiction of the choice of q's, and the sought after inequality has been established.

A division of the unit interval from 0 to 1 into t subintervals is made next, with each of the subintervals having length 1/t. Since there are (t+1) of the above quantities of form $p-q\sqrt{D}$ and each such quantity has a distinct value between 0 and 1, it follows that two of these quantities lie in the same interval of length 1/t. Call these two quantities $p_i-q_i\sqrt{D}$ and $p_j-q_j\sqrt{D}$ such that $p_i \neq p_j$ and $q_i \neq q_j$. Then, since these quantities are distinct, their difference can be taken such that

$$(p_i-q_i\sqrt{D})-(p_j-q_j\sqrt{D}) = (p_i-p_j)-(q_i-q_j)\sqrt{D}$$

is positive. It is noted also that this difference is of the original form, $p-q\sqrt{D}$ with p and q positive, and has absolute value less than 1/t and, therefore, less than E. Since the q's take on the integral values from 0 to t, it follows that the absolute value of (q_i-q_j) is less than or equal to t. Therefore, it follows that the absolute value of $(p_i-p_j)-(q_i-q_j)\sqrt{D}$ is less than the absolute value of $\frac{1}{(q_i-q_j)}$ and the proof of the theorem is complete.

Repetition of the argument utilized in the preceding proof guarantees the existence of infinitely many integer

pairs p,q fulfilling the requirement of that theorem. The existence of one integer pair p,q satisfying the conditions of Theorem 1 has been demonstrated. Choose a positive constant E_1 such that

$$E > |p-q\sqrt{D}| > E_1 .$$

One can then repeat the steps outlined in the proof of Theorem 1 and discover an integer pair p_1,q_1 such that

$$|p_1-q_1\sqrt{D}| < E_1 ,$$

which shows that

$$|p_1-q_1\sqrt{D}| < E .$$

This new integer pair p_1,q_1, then, satisfies the conditions of Theorem 1. By choosing yet another positive constant E_2 such that $E_2 < |p_1-q_1\sqrt{D}|$ and repeating the steps in the above proof, one is able to determine a third integer pair p_2,q_2 which satisfied the conditions of Theorem 1. Clearly, then, an interminable iterative process has been defined which will generate an infinite set of integer pairs p_i,q_i such that $|p_i-q_i\sqrt{D}| < E$ and $|p_i-q_i\sqrt{D}| < 1/q_i$.

THEOREM 2. There exists an integer K such that

$$p^2-Dq^2 = K \qquad (2)$$

for an infinite number of integer pairs p,q.

Proof. Choose the integer pair p,q as in the iterative process previously described in the discussion following the proof of Theorem 1. Then,

$$|p+q\sqrt{D}| = |p-q\sqrt{D} + q\sqrt{D} + q\sqrt{D}|$$

$$= |p-q\sqrt{D} + 2q\sqrt{D}| .$$

By the triangle inequality,

$$|p+q\sqrt{D}| \leq |p-q\sqrt{D}| + |2q\sqrt{D}| .$$

It was noted earlier that $|p-q\sqrt{D}| < 1/q$ and, consequently, $|p-q\sqrt{D}| < |1/q|$ so that the obvious substitution only strengthens the inequality and shows that

$$|p + q\sqrt{D}| < |1/q| + |2q\sqrt{D}| . \qquad (3)$$

Multiplication of inequality (3) by $|p-q\sqrt{D}| < |1/q|$ appropriately term by term shows that

$$|p + q\sqrt{D}|\ |p-q\sqrt{D}| = |p^2-Dq^2| < |1/q|\left[|1/q| + |2q\sqrt{D}|\right] .$$

It follows that

$$|p^2-Dq^2| < |1/q^2| + |2\sqrt{D}| = 1/q^2 + 2\sqrt{D}$$

and, since $0 < 1/q \leq 1$ implies $1/q^2 \leq 1$, that

$$|p^2-Dq^2| < 1 + 2\sqrt{D} .$$

It has been shown previously that there is an infinite set of integer pairs p,q satisfying the conditions herein required,

and it is now shown that $|p^2-Dq^2| < 1 + 2\sqrt{D}$ is true for any and, therefore, all of these integer pairs. It is noted, however, that there are only a finite number of positive integers less than $(1 + 2\sqrt{D})$ and that the quantity $|p^2-Dq^2|$ is always integral in value. It follows, then, that the quantity $|p^2-Dq^2|$ takes on at least one integral value less than $(1 + 2\sqrt{D})$ an infinite number of times. Let that integral value be called K and the proof of the theorem is complete.

THEOREM 3. The equation

$$x^2-Dy^2 = 1 \tag{1}$$

has at least one integral solution in which $y \neq 0$.

Proof. Considering the infinite set of solutions which have been established previously for equation (2) in the proof of Theorem 2, one divides these solutions into K^2 different congruence classes, putting two integer pairs p_1,q_1 and p_2,q_2 in the same class if and only if $p_1 \equiv p_2 \pmod K$ and $q_1 \equiv q_2 \pmod K$. It follows that some class contains an infinite number of these integer pairs. Consider this class which contains an infinite number of such pairs and choose the two pairs p_1,q_1 and p_2,q_2 from this class such that $p_1 \neq \pm p_2$ and $q_1 \neq \pm q_2$. Let $x = \frac{p_1p_2-Dq_1q_2}{K}$ and

$y = \frac{p_1q_2 - p_2q_1}{K}$. Direct algebraic verification shows that

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

$$x^2 - Dy^2 = \frac{(p_1p_2 - Dq_1q_2)^2}{K^2} - \frac{D(p_1q_2 - p_2q_1)^2}{K^2}$$

$$= \frac{1}{K^2}(p_1^2p_2^2 - 2Dp_1p_2q_1q_2 + D^2q_1^2q_2^2 - Dp_1^2q_2^2 + 2Dp_1p_2q_1q_2 - Dp_2^2q_1^2)$$

$$= \frac{1}{K^2}(p_1^2p_2^2 - Dp_1^2q_2^2 - Dp_2^2q_1^2 + D^2q_1^2q_2^2)$$

$$= \frac{1}{K^2}\left[p_1^2(p_2^2 - Dq_2^2) - Dq_1^2(p_2^2 - Dq_2^2)\right]$$

$$= \frac{1}{K^2}(p_1^2 - Dq_1^2)(p_2^2 - Dq_2^2)$$

$$= \frac{1}{K^2}(K)\ (K)$$

$$= 1 .$$

Utilizing rules of multiplication of congruences and recalling that $p_1 \equiv p_2 \pmod{K}$ and $q_1 \equiv q_2 \pmod{K}$, one sees that $p_1q_2 \equiv p_2q_1 \pmod{K}$. This means that $p_1q_2 - p_2q_1 = mK$, where m is an integer. However,

$$y = \frac{p_1q_2 - p_2q_1}{K} = \frac{mK}{K} = m.$$

Thus, y is an integer.

To show that $y \neq 0$, one can assume $y = 0$ and exhibit a contradiction. If $y = 0$, then

$$y = \frac{p_1 q_2 - p_2 q_1}{K} = 0$$

implies that $p_1 q_2 = p_2 q_1$. Solving this equality for p_1, one finds that $p_1 = \frac{p_2 q_1}{q_2}$. Note that $y = 0$ also implies that $x = \pm 1$. This in turn shows that $p_1 p_2 - D q_1 q_2 = \pm K$. Making the appropriate substitution shows that

$$\pm K = \left[\frac{p_2 q_1}{q_2}\right] p_2 - D q_1 q_2$$

$$= \frac{q_1}{q_2}(p_2^2 - D q_2^2).$$

However, $p_2^2 - D q_2^2 = K$. Making the appropriate substitution, it follows that $\frac{q_1}{q_2} = \pm 1$ which shows that $q_1 = \pm q_2$. This in turn shows that $p_1 = \pm p_2$. However, this is a contradiction of the choice of p_1, p_2, q_1, q_2 and it has therefore been shown that $y \neq 0$.

It remains only to be shown that x is an integer. Again recalling that $p_1 \equiv p_2 \pmod K$ and $q_1 \equiv q_2 \pmod K$ and utilizing rules for multiplication of congruences, one finds that $p_1 p_1 \equiv p_1 p_2 \pmod K$ and $q_1 q_1 \equiv q_1 q_2 \pmod K$. However,

$q_1^2 \equiv q_1 q_2 \pmod K$ implies that $-Dq_1^2 \equiv -Dq_1 q_2 \pmod K$.

Adding congruences, one sees that

$$p_1 p_2 - Dq_1 q_2 \equiv p_1^2 - Dq_1^2 \pmod K.$$

However, since p_1, q_1 is a solution of equation (2), one sees that $p_1 p_2 - Dq_1 q_2 \equiv K \equiv 0 \pmod K$. This means that $p_1 p_2 - Dq_1 q_2 = nK$ for some integer n. Noting that $x = \frac{p_1 p_2 - Dq_1 q_2}{K} = \frac{nK}{K} = n$, it has been shown that x is an integer. This completes the proof of Theorem 3.

Having established the existence of at least one solution to equation (1), the next step will be to verify the existence of infinitely many solutions for that equation.

THEOREM 4. If q,r and s,t are any non-trivial solutions, excluding only the cases where $q = s$ and $r = -t$ or $q = -s$ and $r = t$, of the equation

$$x^2 - Dy^2 = 1 , \qquad (1)$$

then a non-trivial solution for that same equation and different from both of those used to establish it is given by

$$(qs + Drt),\ (qt + sr)\ .$$

Proof. Since q,r and s,t are solutions for equation (1), $q^2-Dr^2 = 1$ and $s^2-Dt^2 = 1$. Therefore,

$$(q^2-Dr^2)(s^2-Dt^2) = 1\ .$$

Multiplying out this product, one finds that

$$(qs)^2 + (Drt)^2-D\left[(qt)^2 + (sr)^2\right] = 1\ . \qquad (4)$$

Next, add and subtract the quantity 2Dqrst on the left hand side of equation (4). Rearranging terms, it follows that

$$(qs)^2 + 2Dqrst + (Drt)^2-D\left[(qt)^2 + 2qrst + (sr)^2\right] = 1$$

or,

$$(qs + Drt)^2-D(qt + sr)^2 = 1.$$

Thus, it has been shown that the suggested quantities do satisfy equation (1).

Next it will be shown that this newly established solution for equation (1) is in fact different from those solutions used to produce it. This is accomplished by assuming the contrary and exhibiting a contradiction which must then result. Assume that $(qs+Drt) = q$ and $(qt+sr) = r$. Algebraic manipulation in the second of these assumed equalities shows that

$$r(1-s) = qt$$

or

$$r = \frac{qt}{1-s} . \tag{5}$$

Substituting equation (5) into the first of the assumed equalities above, one finds that

$$qs + \frac{Dqt^2}{1-s} = q .$$

Simplifying,

$$s + \frac{Dt^2}{1-s} = 1$$

or

$$s-s^2 + Dt^2 = 1-s .$$

Thus,

$$s^2-Dt^2 = 2s-1$$

or, since $s^2-Dt^2 = 1$, $s = 1$. This, then, implies that $t = 0$, or that s,t was the trivial solution to equation (1). However, this contradicts the hypothesis and it has been established that this new solution is different from the solution q,r. A similar argument shows that it is also different from s,t.

That this new solution is non-trivial follows by noting that the only situation which can produce the trivial solution is that situation which Theorem 4 specifically excludes. This is shown by assuming the new solution to be trivial and exhibiting the implications of such an assumption. Assume that $qs + Drt = 1$ and $qt + rs = 0$. The second equality demands

that $q = \frac{-sr}{t}$. Substituting this into the first equality, one finds that $\frac{-s^2r}{t} + Drt = 1$, or $s^2 - Dt^2 = \frac{-t}{r}$. However, $s^2 - Dt^2 = 1$. Thus, $\frac{-t}{r} = 1$, or $-t = r$. This means also that $q = s$, referring to the second assumed equality above. A similar consideration when $qs + Drt = -1$ shows that $r = t$ and $q = -s$. These, then, are the only possible combinations which will produce a trivial new solution when the previously outlined procedure is used. This completes the proof of the theorem.

Theorem 3 has established the existence of at least one non-trivial solution to equation (1). One can take that solution x,y and, utilizing the procedure of Theorem 4, establish a new non-trivial solution $(x^2 + Dy^2),(2xy)$ for equation (1). This new solution will be different from the solution x,y. One can then take the two solutions and generate a third, distinct from the first two. One restriction is necessary in the choice of known solutions used to generate new solutions. That restriction is the same one listed in Theorem 4; namely, the choice of a pair of solutions such that either $x_i = x_j$ and $y_i = -y_j$ or $x_i = -x_j$ and $y_i = y_j$ is specifically ruled out in the generation of the new solution

$$(x_ix_j + Dy_iy_j),(x_iy_j + x_jy_i) .$$

One is always able to produce a different solution by choosing to pair the last solution generated either with itself or a previously generated solution, as the situation demands, to give the desired new solution. Clearly, then, this process can be repeated indefinitely, thus demonstrating the existence of infinitely many solutions to equation (1).

The following corollary to Theorem 4 is given, although it offers essentially the same result as that theorem, since that result is given in a slightly different form.

COROLLARY 4-1. If g,h and p,q are solutions to the equation

$$x^2-Dy^2 = 1 , \tag{1}$$

then so also are the integers s and t defined by the equation

$$(g + h\sqrt{D})(p + q\sqrt{D}) = s + t\sqrt{D} . \tag{6}$$

Proof. By the definition of the integers s and t in the corollary, the following equality is also valid:

$$(g-h\sqrt{D})(p-q\sqrt{D}) = s-t\sqrt{D} . \tag{7}$$

Multiplication of equations (6) and (7) gives the new equality $(g^2-Dh^2)(p^2-Dq^2) = s^2-Dt^2$. One notes that the number pairs g,h and p,q are solutions to equation (1) and, making the appropriate substitutions, finds that $(1)(1) = 1 = s^2-Dt^2$, so that the integer pair s,t is a solution of equation (1). This completes the proof of the corollary.

Corollary 4-1 can now be generalized to state that for any solution p,q of equation (1) the integers r and s defined by

$$(p + q\sqrt{D})^n = (r + s\sqrt{D}) \tag{8}$$

also make up a solution for equation (1), provided that n is integral and positive. This is true simply by repeated application of the corollary.

Algebraic manipulation shows that if x,y is a solution of equation (1), then

$$\frac{1}{x + y\sqrt{D}} = \left[\frac{x-y\sqrt{D}}{x-y\sqrt{D}}\right]\left[\frac{1}{x + y\sqrt{D}}\right]$$

$$= \frac{x-y\sqrt{D}}{x^2-y^2D}$$

$$= x-y\sqrt{D}$$

or,

$$\frac{1}{x + y\sqrt{D}} = x-y\sqrt{D}\ . \tag{9}$$

One is thus able to extend the generalization of Corollary 4-1 which was mentioned in the previous paragraph to negative integral values of n. If $n = 0$, one is left with the trivial solution 1,0 and the generalization of Corollary 4-1 is complete.

A solution of the equation $x^2-Dy^2 = 1$ is called positive if both $x > 0$ and $y > 0$. The positive solutions of this

equation are ordered by the size of the x value in each case. Ordering solutions by the value of the x term involved is no compromise as one can see by noting that if x_1, y_1 and x_2, y_2 are two solutions with $x_1 > x_2$, then $x_1^2 - Dy_1^2 = 1 = x_2^2 - Dy_2^2$. Subtracting x_1^2 from the left side of the equality and x_2^2 from the right side and noting that since $x_1 > x_2 > 0$, then $x_1^2 > x_2^2$, one finds that $-Dy_1^2 < -Dy_2^2$. Dividing the inequality by -D, it follows that $y_1^2 > y_2^2$ and, taking positive square roots, that $y_1 > y_2$. Thus, ordering positive solutions by the comparative sizes of the x values involved is a valid procedure.

THEOREM 5[1]. If a,b is a positive solution for the equation

$$x^2 - Dy^2 = 1 \tag{1}$$

such that b is the smallest positive integral value possible for y, and if $x_o = 1$, $y_o = 0$, then the recursion relations

$$x_n = ax_{n-1} + bDy_{n-1} \tag{10}$$

---

[1]This theorem and proof taken from an article by S.T. Parker in the American Mathematical Monthly, Volume 54, 1947, pp. 97 - 100.

and

$$y_n = bx_{n-1} + ay_{n-1} \tag{11}$$

give all the positive solutions for equation (1).

Proof. That the recursion relations (10) and (11) actually produce solutions for equation (1) is shown by mathematical induction. For $n = 1$, the solution a,b is produced. Assume that the relations are valid for $n = k$. Then

$$x_{k+1} = ax_k + bDy_k$$

and

$$y_{k+1} = bx_k + ay_k .$$

It follows that

$$(x_{k+1})^2 - D(y_{k+1})^2 = (ax_k + bDy_k)^2 - D(bx_k + ay_k)^2$$

$$= a^2x_k^2 + 2abDx_k y_k + b^2D^2y_k^2 - b^2Dx_k^2 - 2abDx_k y_k - a^2Dy_k^2 .$$

Simplification shows that

$$(x_{k+1})^2 - D(y_{k+1})^2 = a^2x_k^2 - a^2Dy_k^2 + b^2D^2y_k^2 - b^2Dx_k^2$$

$$= a^2(x_k^2 - Dy_k^2) - Db^2(x_k^2 - Dy_k^2)$$

$$= (a^2 - Db^2)(x_k^2 - Dy_k^2)$$

$$= (1)(1)$$

$$= 1 ,$$

so that one sees that the recursion relations (10) and (11) do give solutions for equation (1).

To show that these recursion relations give all the positive solutions for equation (1), assume that they do not and it will be possible to establish a contradiction. Assume, then, that there exist positive integer pairs x_i, y_i satisfying equation (1) and not obtainable from the recursion relations (10) and (11). Therefore there must be one pair x_m, y_m from this set of solutions not obtainable from relations (10) and (11) for which y_m is the least. Since $x_o = 1$, $y_o = 0$ is the initiating pair for the set obtainable from the relations, it follows that $y_m > b$. Then

$$x_m^2 = Dy_m^2 + 1 = y_m^2\left[D + \frac{1}{y_m^2}\right] < y_m^2\left[D + \frac{1}{b^2}\right] = y_m^2\frac{a^2}{b^2} .$$

Therefore, $x_m < \frac{a}{b} y_m$. Suppose that $x_m \leq \frac{a-1}{b} y_m$. This would yield

$$x_m^2 - Dy_m^2 \leq \left[\frac{(a-1)^2}{b^2} - D\right] y_m^2 = \frac{2-2a}{b^2} y_m^2 < 0 ,$$

since $D > 1$ and $b \geq 1$ and, therefore, $a > 1$.

This contradiction leads to the double inequality

$$\frac{a-1}{b} y_m < x_m < \frac{a}{b} y_m . \qquad (12)$$

Algebraic manipulation of the relations (10) and (11) shows that

$$x_{n-1} = ax_n - bDy_n \tag{13}$$

and

$$y_{n-1} = -bx_n + ay_n . \tag{14}$$

On replacing x_n, y_n in relations (13) and (14) by x_m, y_m, one obtains a new pair x_{m-1}, y_{m-1} which satisfies equation (1). Moreover, considering the inequality (12) and the relations (13) and (14), one finds that

$$y_{m-1} > -b\left[\frac{a}{b}\right] y_m + ay_m = 0$$

and

$$y_{m-1} < -b\left[\frac{a-1}{b}\right] y_m + ay_m = y_m .$$

Thus, $y_m > y_{m-1} > 0$ and there exists an integer pair x_{m-1}, y_{m-1} with a positive y_{m-1} less than y_m.

If x_{m-1}, y_{m-1} is a pair given by the recursion relations (10) and (11), then so is x_m, y_m as is seen by applying the relations (13) and (14). Therefore, the pair x_{m-1}, y_{m-1} cannot be in the set of solutions produced by the recursion relations. The contradiction $y_m > y_{m-1} > 0$, then, implies that Theorem 5 is true.

COROLLARY 5-1. If a,b is the minimal positive solution of the equation

$$x^2-Dy^2 = 1 , \tag{1}$$

then a general solution is given by the set of all x,y satisfying

$$(x + y\sqrt{D}) = \pm(a + b\sqrt{D})^n \tag{15}$$

where n can be any integral value, positive, negative, or zero.

Proof. In the remarks following Corollary 4-1 it was established that equation (15) truly does furnish solutions for equation (1) for all integral values of n, positive, negative, or zero.

If a,b, and n are positive so that

$$x + y\sqrt{D} = (a + b\sqrt{D})^n > 1 ,$$

then

$$-x-y\sqrt{D} = -(a + b\sqrt{D})^n < 1 ,$$

$$x-y\sqrt{D} = (a + b\sqrt{D})^{-n} < 1 ,$$

and

$$-x + y\sqrt{D} = -(a + b\sqrt{D})^{-n} < 1 .$$

Thus, it can be shown that equation (15) gives all the solutions to equation (1), with $y \neq 0$, by showing that every solution of equation (1) with both x and y positive satisfies equation (15) with $n > 0$.

Let $a + b\sqrt{D} = A$, the minimal positive solution of equation (1), so that any positive solution x,y of equation

(1) is such that $x + y\sqrt{D} \geq A$, since A is minimal. Then there exists an $n > 0$ such that

$$A^n \leq x + y\sqrt{D} < A^{n+1}.$$

It follows that

$$1 \leq (x + y\sqrt{D})A^{-n} = (x + y\sqrt{D})(a + b\sqrt{D})^{-n}$$

$$= (x + y\sqrt{D})(a-b\sqrt{D})^n < A,$$

since the inequality has been divided by the positive quantity A^n. However, Corollary 4-1 indicates that

$$1 \leq (x + y\sqrt{D})(a-b\sqrt{D})^n < A$$

is a contradiction unless $(x + y\sqrt{D})(a-b\sqrt{D})^n = 1$. It follows, then, that $(x + y\sqrt{D}) = (a + b\sqrt{D})^n$ and the proof is complete.

In considering the situation where $N = -1$, one finds a similarity to the case in which $N = 1$ in that all solutions of the equation $x^2-Dy^2 = -1$ can be expressed in terms of a single solution. However, there is a basic difference in the two situations in that when $N = -1$, the equation is not always solvable. This is true specifically for $D = 3$.

THEOREM 6. Let D be a positive nonsquare integer. Then if the equation

$$x^2-Dy^2 = -1 \tag{16}$$

is solvable and if $g + h\sqrt{D}$ is the minimal positive solution

of equation (16), the general solution is given by the set of all x,y satisfying

$$x + y\sqrt{D} = \pm(g + h\sqrt{D})^{2n+1}, n = 0, \pm 1, \pm 2, \ldots . \quad (17)$$

The following lemma is stated and proven to facilitate the proof of Theorem 6.

LEMMA 1. Let $A = a + b\sqrt{D}$ be the minimal positive solution of equation (1) and let $g + h\sqrt{D}$ be the minimal positive solution of equation (16), then

$$A = a + b\sqrt{D} = (g + h\sqrt{D})^2 .$$

Proof. Since

$$(g + h\sqrt{D})^2 = g^2 + 2gh\sqrt{D} + h^2D$$

$$= (g^2 + h^2D) + 2gh\sqrt{D}$$

and

$$(g^2 + h^2D)^2 - D(2gh)^2 = g^4 + 2g^2h^2D + h^4D^2 - 4g^2h^2D$$

$$= g^4 - 2g^2h^2D + h^4D^2$$

$$= (g^2 - h^2D)^2$$

$$= (-1)^2$$

$$= 1,$$

one can see that $(g + h\sqrt{D})^2$ is a solution for equation (1). Therefore, due to the minimality of A,

$$1 < a + b\sqrt{D} \leq (g + h\sqrt{D})^2 .$$

Since $(g+h\sqrt{D})^{-1} = (-g + h\sqrt{D})$, it follows that

$$-g + h\sqrt{D} < (a + b\sqrt{D})(-g + h\sqrt{D}) \leq g + h\sqrt{D}$$

or

$$-g + h\sqrt{D} < -ag + bhD + (ah-bg)\sqrt{D} \leq g + h\sqrt{D} .$$

Considering just the middle term of this inequality for a moment, the following algebraic manipulation shows that it is a solution for equation (16):

$$\begin{aligned}(-ag + bhD)^2-(-gb + ah)^2D &= a^2g^2-2agbhD + b^2h^2D^2-g^2b^2D + 2agbhD-a^2h^2D \\ &= a^2g^2-g^2b^2D-a^2h^2D + b^2h^2D^2 \\ &= g^2(a^2-b^2D)-Dh^2(a^2-Db^2) \\ &= (g^2-Dh^2)(a^2-b^2D) \\ &= (g^2-Dh^2) \\ &= -1 .\end{aligned}$$

It follows, then, that $(-gb + ah)$ is not equal to zero. For notational simplicity, let $p = -ag + bhD$ and $q = -gb + ah$. One notes that if a number lies between the minimal positive solution $g + h\sqrt{D}$ of equation (16) and the reciprocal of the minimal positive solution $-g + h\sqrt{D}$, then the reciprocal of

that number must also lie between the minimal positive solution and its reciprocal. It follows that either

$$1 < p + q\sqrt{D} \leq g + h\sqrt{D}$$

or

$$1 < -p + q\sqrt{D} < g + h\sqrt{D} .$$

However, g,h is the minimal positive solution of equation (16) so it must be that $p + q\sqrt{D} = g + h\sqrt{D}$. This implies that

$$\frac{a + b\sqrt{D}}{g + h\sqrt{D}} = g + h\sqrt{D}$$

or

$$a + b\sqrt{D} = (g + h\sqrt{D})^2$$

and the proof of the lemma is complete.

For the proof of Theorem 6, one chooses any solution x,y of equation (16) such that $x,y > 0$. Such a choice is possible since, as in the proof of Corollary 5-1, if a,b, and n are positive so that

$$x + y\sqrt{D} = (a + b\sqrt{D})^n > 1 ,$$

then

$$-x-y\sqrt{D} = -(a + b\sqrt{D})^n < 1 ,$$

$$x-y\sqrt{D} = (a + b\sqrt{D})^{-n} < 1 ,$$

and

$$-x + y\sqrt{D} = -(a + b\sqrt{D})^{-n} < 1.$$

Thus one can find an n such that

$$A^n = (g+h\sqrt{D})^{2n} \leq (x+y\sqrt{D}) < (g+h\sqrt{D})^{2n+2} = A^{n+1}.$$

Dividing throughout the inequality by A^n, one finds that

$$1 \leq (x + y\sqrt{D})A^{-n} < A = (g + h\sqrt{D})^2.$$

Dividing throughout the inequality by $(g + h\sqrt{D})$, it becomes

$$(-g + h\sqrt{D}) \leq s + t\sqrt{D} < g + h\sqrt{D}, \qquad (18)$$

where s,t is a solution of equation (1).

That s,t is a solution of equation (1) is verified by the following demonstration:

$$s + t\sqrt{D} = (x + y\sqrt{D})(a + b\sqrt{D})^{-n}(-g + h\sqrt{D}). \qquad (19)$$

Taking the product $(x + y\sqrt{D})(-g + h\sqrt{D})$, both elements of which are solutions of equation (16), one shows that their product is a solution of equation (1) by the following algebraic manipulation:

$$(x+y\sqrt{D})(-g+h\sqrt{D}) = (-xg+yhD) + (-gy+xh)\sqrt{D}.$$

Substituting into the form x^2-Dy^2, one finds that

$$(-xg + yhD)^2-(-gy + xh)^2D$$

$$= x^2g^2-2xyghD + y^2h^2D^2-g^2y^2D + 2xyghD-x^2h^2D$$

$$= x^2g^2 - x^2h^2D - g^2y^2D + y^2h^2D^2$$

$$= x^2(g^2 - Dh^2) - y^2D(g^2 - Dh^2)$$

$$= (x^2 - Dy^2)(g^2 - Dh^2)$$

$$= (-1)(-1)$$

$$= 1,$$

so that the product $(x + y\sqrt{D})(-g + h\sqrt{D})$ does produce an element which is a solution of equation (1). One notes that $(a + b\sqrt{D})^{-n}$ also is a solution of equation (1) and, by the previously presented generalization of Corollary 4-1, the overall product (19) is seen to be a solution of equation (1). This completes the demonstration that s,t is a solution of equation (1).

Return now to the inequality (18) and consider just the left hand term $-g + h\sqrt{D}$. This quantity is the inverse of the positive quantity $g + h\sqrt{D}$, which is greater than one. Therefore, $-g + h\sqrt{D}$ is positive but less than one. It follows that $(-g + h\sqrt{D})^2$ is also positive and less than $(-g + h\sqrt{D})$. Also, since $g + h\sqrt{D}$ is greater than one, it follows that $(g + h\sqrt{D})^2$ is greater than $(g + h\sqrt{D})$. Making the appropriate substitutions in inequality (18), one finds that

$$(-g + h\sqrt{D})^2 < s + t\sqrt{D} < (g + h\sqrt{D})^2$$

or

$$A^{-1} < s + t\sqrt{D} < A .$$

However, since A is the minimal positive solution of equation (1), it follows that $s + t\sqrt{D} = 1$. Substituting this relationship into equation (19), one finds that

$$1 = (x + y\sqrt{D})(a + b\sqrt{D})^{-n}(-g + h\sqrt{D}),$$

or, multiplying by $(A^n)(g + h\sqrt{D})$, that

$$\begin{aligned}(x + y\sqrt{D}) &= (g + h\sqrt{D})(A^n) \\ &= (g + h\sqrt{D})(g + h\sqrt{D})^{2n} \\ &= (g + h\sqrt{D})^{2n+1} .\end{aligned}$$

In the proof of Lemma 1 it was established that $(g + h\sqrt{D})^2$ is a solution of equation (1). The generalization of Corollary 4-1 then shows that

$$\left[(g + h\sqrt{D})^2\right]^n = (g + h\sqrt{D})^{2n}$$

is also a solution of equation (1) for all integral n. Let

$$(g + h\sqrt{D})^{2n} = g_1 + h_1\sqrt{D} .$$

Then algebraic manipulation shows that the product

$$(g_1 + h_1\sqrt{D})(g + h\sqrt{D}) = (g + h\sqrt{D})^{2n+1}$$

is a solution of equation (16):

$$(g_1 + h_1\sqrt{D})(g + h\sqrt{D}) = (g_1 g + h_1 hD) + (g_1 h + g h_1)\sqrt{D}$$

and

$$(g_1 g + h_1 hD)^2 - D(g_1 h + g h_1)^2$$

$$= g_1^2 g^2 + 2 g_1 g h_1 hD + h_1^2 h^2 D^2 - g_1^2 h^2 D - 2 g_1 g h_1 hD - g^2 h_1^2 D$$

$$= g_1^2 (g^2 - Dh^2) - Dh_1^2 (g^2 - Dh^2)$$

$$= (g_1^2 - Dh_1^2)(g^2 - Dh^2)$$

$$= (1)(-1)$$

$$= -1.$$

This completes the proof of the theorem.

As an example of the situation where equation (16) can be solved, one can look at the equation for $D = 5$. In this case the minimal positive solution is $(2 + \sqrt{5})$. Substituting this quantity into equation (17) for $n = 1,2$ gives the additional solutions for equation (16) of $(38 + 17\sqrt{5})$ and $(682 + 305\sqrt{5})$.

An example has been given for which equation (16) is not solvable, thus ruling out the possibility of universal solvability. Also, a general solution has been established for those situations in which equation (16) is solvable. Thus, the discussion of the situation where $N = -1$ is complete and, likewise, the discussion of the first case where $|N| = 1$ is complete.

THE CASE $|N| = 4$

Two theorems are stated and proven in the consideration of the case where $|N| = 4$, the first dealing with $N = 4$ and the second dealing with $N = -4$. Attention is called to the close similarity between these two theorems and Corollary 5-1 and Theorem 6.

THEOREM 7. If D is a positive nonsquare integer and if $e + f\sqrt{D}$ is the minimal positive solution of the equation

$$x^2 - Dy^2 = 4, \tag{20}$$

then the general solution to equation (20) is given by the set of all x,y satisfying

$$x + y\sqrt{D} = \pm 2\left[\frac{e + f\sqrt{D}}{2}\right]^n, \quad n = 0, \pm 1, \pm 2, \ldots . \tag{21}$$

Proof. That equation (20) is always solvable follows directly from noting that one need only double a solution of $x^2 - Dy^2 = 1$ to produce a solution for the equation $x^2 - Dy^2 = 4$ and that the equation $x^2 - Dy^2 = 1$ has been shown to be solvable in all cases in Theorem 3. It does not follow, however, that merely doubling the solutions of equation (1) will give all of the solutions of equation (20).

If $x_2 + y_2\sqrt{D}$ and $x_3 + y_3\sqrt{D}$ are any two solutions of equation (20), then the number pair x_1, y_1 described by

their product in the following form is also an acceptable solution to equation (20):

$$2\left[\frac{x_2 + y_2\sqrt{D}}{2}\right]\left[\frac{x_3 + y_3\sqrt{D}}{2}\right] = x_1 + y_1\sqrt{D} \ . \qquad (22)$$

This statement is verified by showing that $x_1 + y_1\sqrt{D}$ describes an integer pair and that x_1, y_1 actually satisfies equation (20).

First, one sees that $x_2^2 - Dy_2^2 = 4$ and $x_3^2 - Dy_3^2 = 4$ implies that $x_2^2 \equiv Dy_2^2 \pmod 2$ and $x_3^2 \equiv Dy_3^2 \pmod 2$, which in turn implies that $x_2 \equiv Dy_2 \pmod 2$ and $x_3 \equiv Dy_3 \pmod 2$. The last two congruence relations hold true since congruence (mod 2) is simply a check on the agreement of the parity of two elements and one notes that the square of an integer is even or odd as the integer itself is even or odd. Multiplying out the product

$$2\left[\frac{x_2 + y_2\sqrt{D}}{2}\right]\left[\frac{x_3 + y_3\sqrt{D}}{2}\right] = x_1 + y_1\sqrt{D} \ ,$$

one finds that

$$x_1 = \frac{x_2x_3 + y_2y_3D}{2}$$

or that $2x_1 = x_2x_3 + y_2y_3D$. Substituting and adding appropriate congruences (mod 2), one finds that

$$2x_1 = x_2x_3 + y_2y_3D \equiv D^2y_2y_3 + Dy_2y_3 \equiv D(D+1)y_2y_3 \equiv 0 \pmod 2.$$

The last congruence in the preceding series is valid since one or the other of the two consecutive integers D and (D+1) must be even, thus making the entire product even and congruent to zero (mod 2).

Similarily, $y_1 = \dfrac{x_2y_3 + x_3y_2}{2}$ or $2y_1 = x_2y_3 + x_3y_2$ and

$$2y_1 = x_2y_3 + x_3y_2 \equiv Dy_2y_3 + Dy_2y_3 \equiv 2Dy_2y_3 \equiv 0 \pmod 2.$$

Thus, both x_1 and y_1 as defined in equation (22) are integers.

Consider the product

$$x_1^2 - Dy_1^2 = (x_1 - y_1\sqrt{D})(x_1 + y_1\sqrt{D})$$

$$= \left[2\left[\frac{x_2 - y_2\sqrt{D}}{2}\right]\left[\frac{x_3 - y_3\sqrt{D}}{2}\right]\right]\left[2\left[\frac{x_2 + y_2\sqrt{D}}{2}\right]\left[\frac{x_3 + y_3\sqrt{D}}{2}\right]\right]$$

$$= 4\left[\frac{x_2^2 - Dy_2^2}{4}\right]\left[\frac{x_3^2 - Dy_3^2}{4}\right]$$

$$= 4\left[\frac{4}{4}\right]\left[\frac{4}{4}\right]$$

$$= 4 .$$

From this one sees that the integer pair x_1, y_1 as defined by equation (22) is actually a solution of equation (20). The generalization is now obvious. The equation

$$x + y\sqrt{D} = \pm 2\left[\frac{e + f\sqrt{D}}{2}\right]^n \qquad (21)$$

defines a set of solutions for equation (20), one for each integer n. It remains only to show that equation (21) exhibits all solutions for equation (20) to complete the proof of this theorem.

Referring again to the restrictions as offered in the proof of Corollary 5-1, one recalls that if a,b, and n are positive so that

$$x + y\sqrt{D} = (a + b\sqrt{D})^n > 1 ,$$

then

$$-x - y\sqrt{D} = -(a + b\sqrt{D})^n < 1 ,$$

$$x - y\sqrt{D} = (a + b\sqrt{D})^{-n} < 1 ,$$

and

$$-x + y\sqrt{D} = -(a + b\sqrt{D})^{-n} < 1 .$$

Thus, one can choose any solution $x + y\sqrt{D}$ of equation (20) such that $x > 0$ and $y > 0$, and will have shown that all solutions of equation (20) are of the form

$$\pm 2\left[\frac{e + f\sqrt{D}}{2}\right]^n$$

when one has shown that all solutions $x + y\sqrt{D}$ such that $x > 0$ and $y > 0$ are of that form.

There exists an $n > 0$ such that

$$2\left[\frac{e + f\sqrt{D}}{2}\right]^n \leq x + y\sqrt{D} < 2\left[\frac{e + f\sqrt{D}}{2}\right]^{n+1} \qquad (23)$$

for the positive solution $x + y\sqrt{D}$, due to the minimality of $e + f\sqrt{D}$. Multiply (23) throughout by the positive quantity

$$\left[\frac{e + f\sqrt{D}}{2}\right]^{-n} = \left[\frac{e-f\sqrt{D}}{2}\right]^n .$$

The inequality

$$2 \leq (x + y\sqrt{D})\left[\frac{e-f\sqrt{D}}{2}\right]^n < e + f\sqrt{D} \qquad (24)$$

results. However, it has been previously established that all quantities of the form (21) are solutions of equation (20).

One notes, therefore, that $\left[\frac{e-f\sqrt{D}}{2}\right]^n$ is of the form

$(1/2)(t + u\sqrt{D})$ for some t,u, an integral solution of

equation (20). Thus, the center term in the inequality (24) is of the form (22), which indicates that it is an integral solution of equation (20). However, since $e + f\sqrt{D}$ is the minimal positive solution of equation (20), it must be that

$$(x + y\sqrt{D}) \left[\frac{e-f\sqrt{D}}{2}\right]^n = 2 . \tag{25}$$

Multiplying (25) by the quantity

$$\left[\frac{e+f\sqrt{D}}{2}\right]^n ,$$

it follows that

$$x + y\sqrt{D} = 2\left[\frac{e+f\sqrt{D}}{2}\right]^n .$$

This completes the proof of Theorem 7.

THEOREM 8. If the equation

$$x^2 - Dy^2 = -4 \tag{26}$$

is solvable and if its minimal positive solution is $u+v\sqrt{D}$, then a general solution for equation (26) is given by the set of all x,y satisfying

$$x+y\sqrt{D} = \pm 2\left[\frac{u+v\sqrt{D}}{2}\right]^{2n+1}, \quad n=0,\pm 1,\pm 2,\ldots . \tag{27}$$

The following lemma is stated and proven to facilitate the proof of the theorem.

LEMMA 2. If $e + f\sqrt{D}$ is the minimal positive solution of equation (20) and $u + v\sqrt{D}$ is the minimal positive solution of equation (26), then

$$e + f\sqrt{D} = 2\left[\frac{u + v\sqrt{D}}{2}\right]^2 . \tag{28}$$

Proof. A consideration of parity entirely analogous to the argument used in the proof of Theorem 7 assures one that the quantity $2\left[\frac{u + v\sqrt{D}}{2}\right]^2$ is integral. Direct substitution shows that

$$\frac{u^2 + Dv^2}{2}^2 - D(uv)^2 = \frac{u^4 + 2Du^2v^2 + v^4D^2 - 4Du^2v^2}{4}$$

$$= \frac{(u^2 - Dv^2)^2}{4}$$

$$= \frac{(-4)(-4)}{4}$$

$$= 4 .$$

Thus, $2\left[\frac{u + v\sqrt{D}}{2}\right]^2$ is an integral solution of equation (20). It follows that

$$1 < e + f\sqrt{D} \leq 2\left[\frac{u + v\sqrt{D}}{2}\right]^2, \qquad (29)$$

due to the minimality of $(e + f\sqrt{D})$. Multiplying through (29) by the quantity $\frac{-u + v\sqrt{D}}{2}$, one sees that

$$\frac{-u + v\sqrt{D}}{2} < (e + f\sqrt{D})\left[\frac{-u + v\sqrt{D}}{2}\right] \leq u + v\sqrt{D}. \qquad (30)$$

An argument similar to the one used in the proof of Theorem 7 indicates that the product in the center of the inequality (30) is an integral solution of equation (26). Since

$\frac{-u + v\sqrt{D}}{2}$ is positive and $(u + v\sqrt{D})$ is the minimal positive solution of equation (26), it follows that

$$(e + f\sqrt{D})\left[\frac{-u + v\sqrt{D}}{2}\right] = u + v\sqrt{D}$$

must be true. However, this implies that

$$e + f\sqrt{D} = 2\left[\frac{u + v\sqrt{D}}{2}\right]^2$$

and the proof of the lemma is complete.

Continuing with the proof of Theorem 6, if $x + y\sqrt{D}$ is any positive solution of equation (26), it is possible to

find an n such that

$$2\left[\frac{e+f\sqrt{D}}{2}\right]^{n} \leq x+y\sqrt{D} < 2\left[\frac{e+f\sqrt{D}}{2}\right]^{n+1} .$$

It follows that

$$2 \leq \left[\frac{e+f\sqrt{D}}{2}\right]^{-n}(x+y\sqrt{D}) < e+f\sqrt{D} ,$$

$$2 \leq \left[\frac{e+f\sqrt{D}}{2}\right]^{-n}(x+y\sqrt{D}) < 2\left[\frac{u+v\sqrt{D}}{2}\right]^{2} ,$$

$$2\left[\frac{-u+v\sqrt{D}}{2}\right] < \left[\frac{e+f\sqrt{D}}{2}\right]^{-n}(x+y\sqrt{D})\left[\frac{-u+v\sqrt{D}}{2}\right] < 2\left[\frac{u+v\sqrt{D}}{2}\right] .$$

Since $0 < \frac{-u+v\sqrt{D}}{2} < 1$ and $1 < \frac{u+v\sqrt{D}}{2}$,

$$2\left[\frac{-u+v\sqrt{D}}{2}\right]^{2} < \left[\frac{e+f\sqrt{D}}{2}\right]^{-n}(x+y\sqrt{D})\left[\frac{-u+v\sqrt{D}}{2}\right] < 2\left[\frac{u+v\sqrt{D}}{2}\right]^{2} .$$

$(e+f\sqrt{D}) = 2\left[\frac{u+v\sqrt{D}}{2}\right]^{2}$ implies that

$$e-f\sqrt{D} < \left[\frac{e+f\sqrt{D}}{2}\right]^{-n}(x+y\sqrt{D})\left[\frac{-u+v\sqrt{D}}{2}\right] < e+f\sqrt{D} . \quad (31)$$

Again utilizing an argument similar to the one used in the proof of Theorem 7, one sees that the center product in the inequality (31) is an integral solution for equation (20). However, (31) shows that this integral solution value is positive, since it is larger than the inverse of the minimal positive solution of equation (20), and that it is less than the minimal positive solution of that equation. The only possibility, then, is that

$$\left[\frac{e+f\sqrt{D}}{2}\right]^{-n}(x+y\sqrt{D})\left[\frac{-u+v\sqrt{D}}{2}\right]=2\ .$$

Multiplying through by $\left[\frac{u+v\sqrt{D}}{2}\right]\left[\frac{e+f\sqrt{D}}{2}\right]^{n}$, one sees that

$$x+y\sqrt{D}=2\left[\frac{e+f\sqrt{D}}{2}\right]^{n}\left[\frac{u+v\sqrt{D}}{2}\right].$$

However, $\frac{e+f\sqrt{D}}{2}=\left[\frac{u+v\sqrt{D}}{2}\right]^{2}$ so that

$$x+y\sqrt{D}=2\left[\frac{u+v\sqrt{D}}{2}\right]^{2n+1}\ .$$

Lemma 2 shows that $\frac{e+f\sqrt{D}}{2}=\left[\frac{u+v\sqrt{D}}{2}\right]^{2}$ or that

$2\left[\frac{e+f\sqrt{D}}{2}\right]^n = 2\left[\frac{u+v\sqrt{D}}{2}\right]^{2n}$. However, Theorem 7 has established that this is a valid solution for equation (20) for all integers n. Let $2\left[\frac{u+v\sqrt{D}}{2}\right]^{2n} = u_1 + v_1\sqrt{D}$. Then direct algebraic manipulation shows that the solution formed by the product

$$(u_1 + v_1\sqrt{D})\left[\frac{u+v\sqrt{D}}{2}\right] = 2\left[\frac{u+v\sqrt{D}}{2}\right]^{2n+1}$$

satisfies equation (26):

$$(u_1 + v_1\sqrt{D})\,\frac{u+v\sqrt{D}}{2} = \left[\frac{u_1 u + v_1 vD}{2}\right] + \left[\frac{u_1 v + uv_1}{2}\right]\sqrt{D}$$

and

$$\left[\frac{u_1 u + v_1 vD}{2}\right]^2 - D\left[\frac{u_1 v + uv_1}{2}\right]^2$$

$$= \frac{u_1^2 u^2 + 2u_1 uv_1 vD + v_1^2 v^2 D^2}{4} - \frac{u_1^2 v^2 D + 2u_1 uv_1 vD + u^2 v_1^2 D}{4}$$

$$= \frac{u_1^2(u^2-Dv^2)-v_1^2D(u^2-v^2D)}{4}$$

$$= \frac{(u_1^2-Dv_1^2)(u^2-Dv^2)}{4}$$

$$= \frac{(4)(-4)}{4}$$

$$= -4 .$$

A consideration of parity analogous to the argument used in the proof of Theorem 7 shows that quantities of the form $2\left[\frac{u+v\sqrt{D}}{2}\right]^{2n+1}$ are integral. This completes the proof of Theorem 8 and the consideration of the case $|N| = 4$ as well.

THE CASE $|N| > 0$

The third and final case to be considered is that situation in general where $N \neq 0$. One should note that the general equation in this case is not always solvable and that the following theorem is based on the assumption of solvability.

THEOREM 9. If j,k is a solution of equation (1), and if p,q is a solution of the general equation

$$x^2-Dy^2 = N \ , \tag{32}$$

where $D > 0$ and nonsquare, then the integer pair s,t defined by

$$s+t\sqrt{D} = (j+k\sqrt{D})(p+q\sqrt{D}) \tag{33}$$

is a solution for equation (32).

Proof. The proof is simply a matter of algebraic manipulation to obtain verification of the statement of the theorem.

$$(s+t\sqrt{D}) = (j+k\sqrt{D})(p+q\sqrt{D})$$

implies that

$$s+t\sqrt{D} = (jp+Dqk)+(jq+pk)\sqrt{D}$$

or that $s=(jp+Dqk)$ and $t=(jq+pk)$. Substituting these values into the form x^2-Dy^2, one finds that

$$\begin{aligned}(jp+qkD)^2-D(jq+pk)^2 &= j^2p^2+2Djkpq+q^2k^2D^2-Dj^2q^2-2Djkpq-Dp^2k^2 \\ &= j^2(p^2-Dq^2)-Dk^2(p^2-Dq^2) \\ &= (j^2-Dk^2)(p^2-Dq^2) \\ &= (1)(N) \\ &= N \ .\end{aligned}$$

This completes the proof of the theorem.

COROLLARY 9-1. Under the conditions stated in Theorem 9, if equation (32) has one solution, it has infinitely many solutions.

Proof. In Theorem 9 it was shown that the product of a solution of equation (32) and a solution of equation (1) produces a solution for equation (32). Also, in Corollary 5-1 it was shown that there exist infinitely many solutions of equation (1). It is therefore possible to form infinitely many different products involving unique solutions of equation (1) and the single solution of equation (32) which is postulated by the theorem. This infinite set of products would, therefore, contain infinitely many solutions of equation (32). That these are different solutions follows directly by assuming two solutions to be equal and noting the implications. If $x_1 + y_1\sqrt{D}$ and $x_2 + y_2\sqrt{D}$ are solutions of equation (1) and $p + q\sqrt{D}$ is a solution of equation (32) such that

$$(x_1 + y_1\sqrt{D})(p + q\sqrt{D}) = (x_2 + y_2\sqrt{D})(p + q\sqrt{D}),$$

then, multiplying through by $(p + q\sqrt{D})^{-1}$, one sees that

$$x_1 + y_1\sqrt{D} = x_2 + y_2\sqrt{D}.$$

Thus, $x_1 = x_2$ and $y_1 = y_2$. This completes the proof of the corollary.

Although the above process assures the existence of and shows how to find infinitely many solutions of equation (32) whenever one such solution exists, it by no means offers

a method of obtaining all solutions for equation (32). A specific example will be sufficient to demonstrate this fact: 7,0 and $9+4\sqrt{2}$ are two solutions of the equation $x^2-2y^2=49$ but neither can be obtained by multiplying the other by a solution of $x^2-2y^2=1$.

Referring to the equation $(j+k\sqrt{D})(p+q\sqrt{D})=s+t\sqrt{D}$ as explained in Theorem 9, the two integer pair solutions p,q and s,t of equation (32) are in the same class, or belong to the same class. That is to say, two integer pair solutions of equation (32) are in the same class if and only if one integer pair can be obtained from the other by multiplication by a solution of equation (1).

The following theorem establishes a finite test for solvability of equation (32). This is done by establishing bounds for the smallest element of each solution class, with the ordering being based on the magnitude of the x term of the solution pair x,y. One is able to make the restriction that $x > 0$ by noting that the two solutions $p+q\sqrt{D}$ and $-p-q\sqrt{D}$ are in the same class.

THEOREM 10. If the equation

$$x^2-Dy^2=N \tag{32}$$

is solvable, it has a solution s,t with

$$0 < s < \sqrt{\frac{Ba+1}{2} \cdot N}\,, \tag{34}$$

where $A = a + b\sqrt{D}$ is the minimal positive solution of equation (1) and $B = \frac{A}{A-1}$. If there is more than one class of solutions to equation (32), each solution class contains an element for which the inequality (34) holds.

Proof. If $p + q\sqrt{D}$ is the minimal positive solution of equation (32), then the conditions of the theorem are satisfied. Otherwise, given any solution $p + q\sqrt{D}$ of equation (32) with $p > 0$ and $p + q\sqrt{D}$ non-minimal, a solution $s + t\sqrt{D}$ of that same equation is sought such that

$$(x + y\sqrt{D})(p + q\sqrt{D}) = s + t\sqrt{D}, \tag{35}$$

with $x + y\sqrt{D}$ a solution of equation (1) and $0 < s < p$. The proof is broken down into two parts: (1) $N > 0$ and (2) $N < 0$.

If $N > 0$, let $A = a + b\sqrt{D}$ again be the minimal positive solution of equation (1) and, referring to equation (35), if $q > 0$, choose $x + y\sqrt{D}$ to be $A^{-1} = a - b\sqrt{D}$ while, if $q < 0$, choose $x + y\sqrt{D}$ to be $A = a + b\sqrt{D}$. It follows, then, that $s = pa - b|q|D$. Rearranging terms, one finds that

$$s = p\left[a - b\sqrt{D}\left[\frac{|q|\sqrt{D}}{p}\right]\right]. \tag{36}$$

By observing that

$$\frac{|q|\sqrt{D}}{p} = \sqrt{\frac{Dq^2}{p}} = \sqrt{\frac{p^2-N}{p^2}} = \sqrt{1-\frac{N}{p^2}}$$

and further rearranging terms, one sees that

$$s = p\left[a-b\sqrt{D} + b\sqrt{D}\left[1-\sqrt{1-\frac{N}{p^2}}\right]\right]. \qquad (37)$$

Since $N > 0$ and $p^2-q^2D = N$, then $1-\frac{q^2D}{p^2} = \frac{N}{p^2}$. This implies that $0 < \frac{N}{p^2} < 1$.

Note, for a general K such that $0 < K < 1$, that

$$0 < 1-\sqrt{1-K} = \frac{K}{1+\sqrt{1-K}} < \frac{K}{2-K} \quad . \qquad (38)$$

The last inequality is valid since, for $0 < K < 1$, $\sqrt{1-K} > 1-K$ implies $1+\sqrt{1-K} > 1+(1-K) = 2-K$. Substituting $\frac{N}{p^2}$ for K in inequality (38) and then making the appropriate substitutions in equation (37), one finds that

$$0 < p\left[a-b\sqrt{D}+b\sqrt{D}\left[1-\sqrt{1-\frac{N}{p^2}}\right]\right] < p\left[a-b\sqrt{D}+b\sqrt{D}\left[\frac{\frac{N}{p^2}}{2-\frac{N}{p^2}}\right]\right] \quad .$$

Simplification shows that

$$0 < s < p\left[A^{-1} + b\sqrt{D}\ \left[\frac{N}{2p^2-N}\right]\right] \quad . \tag{39}$$

Note that $b\sqrt{D} = \sqrt{a^2-1} < a$ and make the appropriate substitution in (39). One then sees that

$$0 < s < p\left[A^{-1} + b\sqrt{D}\ \left[\frac{N}{2p^2-N}\right]\right] < p\left[A^{-1} + a\left[\frac{N}{2p^2-N}\right]\right]$$

or that

$$0 < s < p\left[A^{-1} + a\left[\frac{N}{2p^2-N}\right]\right] \quad . \tag{40}$$

Inequality (40) shows that $s < p$ will hold, if it is true that $\left[A^{-1} + a\left[\frac{N}{2p^2-N}\right]\right] < 1$. Algebraic manipulation of this last inequality shows that

$$2p^2-N+A(aN) < A(2p^2-N),$$

$$2p^2(1-A) < N(1-A-Aa) .$$

Since $(1-A) < 0$,

$$2p^2 > N\left[\frac{1-A-Aa}{1-A}\right] .$$

Thus

$$p^2 > \frac{N}{2}\left[\frac{1-A}{1-A} - \frac{Aa}{1-A}\right] ,$$

$$p^2 > \left[1+\frac{Aa}{A-1}\right]\frac{N}{2} .$$

Noting that $B = \frac{A}{A-1}$,

$$p^2 > (1+Ba)\frac{N}{2}$$

$$p > \sqrt{\frac{Ba+1}{2} \cdot N} . \qquad (41)$$

Since p,q was any solution of equation (32) such that $p > 0$, it is immediately apparent that if $s < \sqrt{\frac{Ba+1}{2} \cdot N}$ does not hold, the steps in the above consideration can be repeated using s,t instead of p,q, thus getting a solution with the x value less than s. Since all of the values obtained in this fashion are integral and positive, a finite number of repitions of the above process establishes inequality (34). This completes the proof for $N > 0$.

In the case where $N < 0$, one makes the same initial choices of values involved as in the proof of the case where $N > 0$ down to the point where it was found that $s = pa - b|q|D$. Algebraic manipulation shows that

$$s = |q|\sqrt{D}\left[\frac{pa}{|q|\sqrt{D}} - b\sqrt{D}\right]$$

$$= |q|\sqrt{D}\left[a - b\sqrt{D} + a\left[-1 + \frac{p}{q\sqrt{D}}\right]\right]$$

$$= |q|\sqrt{D}\left[a - b\sqrt{D} + a\left[\sqrt{\frac{p^2}{q^2D}} - 1\right]\right]$$

$$= |q|\sqrt{D}\left[a - b\sqrt{D} + a\left[\sqrt{\frac{q^2D+N}{q^2D}} - 1\right]\right]$$

$$= |q|\sqrt{D}\left[a - b\sqrt{D} + a\left[\sqrt{1+\frac{N}{q^2D}} - 1\right]\right]$$

$$= |q|\sqrt{D}\left[a - b\sqrt{D} - a\left[1 - \sqrt{1-\frac{-N}{q^2D}}\right]\right].$$

Noting here that $0 < \frac{-N}{q^2D} < 1$, one can see, as in the proof of the case where $N > 0$, that

$$s < |q|\sqrt{D}\left[a - b\sqrt{D} - a\left[\frac{-\frac{N}{q^2D}}{2 - \frac{-N}{q^2D}}\right]\right]$$

$$s < |q|\sqrt{D}\left[A^{-1} + a\left[\frac{\frac{N}{q^2D}}{2 + \frac{N}{q^2D}}\right]\right]$$

$$s < |q|\sqrt{D}\left[A^{-1} + a\left[\frac{N}{2q^2D + N}\right]\right]$$

$$s < |q|\sqrt{D}\left[A^{-1} + a\left[\frac{N}{2p^2 - N}\right]\right] .$$

If $s < p$ is to be valid, $|q|\sqrt{D}\left[A^{-1} + a\left[\frac{N}{2p^2-N}\right]\right] \leq p$ must also hold true. Since $N < 0$, the quantity $\frac{p}{|q|\sqrt{D}}$ must be less than one. Thus $\left[A^{-1} + a\left[\frac{N}{2p^2-N}\right]\right] \leq 1$ must be true if p is to be greater than s. Algebraic manipulation of this last inequality shows that

$$(2p^2 - N + AaN) < A(2p^2 - N)$$

$$2p^2(1-A) < N(1-A-Aa) .$$

Since $(1-A) < 0$,

$$p^2 > \frac{N}{2}\left[\frac{1-A-Aa}{1-A}\right]$$

$$p^2 > \left[1+\frac{Aa}{A-1}\right]\frac{N}{2}$$

$$p > \sqrt{\frac{Ba+1}{2} \cdot N} \ .$$

As in the proof for $N > 0$, the capacity for repitition of this process coupled with the fact that only positive integral solutions are obtained insures the establishment of inequality (34) after a finite number of repititions. This completes the proof of the theorem.

Theorem 10 has reduced the question of the solvability of the equation $x^2-Dy^2 = N$ to a finite consideration. After determining the minimal positive solution of $x^2-Dy^2 = 1$, one need only consider the numbers of the form $\frac{(s^2-N)}{D}$ for s in the interval established in Theorem 10 to see if any of these numbers are perfect squares. If there are two or more acceptable values of s in the interval, it is easily determined whether the solutions are in the same class.

As an example, if $D = 2$, the minimal positive solution of $x^2-2y^2 = 1$ is 3,2. The condition of Theorem 10 is satisfied if $0 < s < \frac{3}{2}\sqrt{N}$. Since $N = s^2-2t^2 < s^2$, one needs only to

investigate the integers between $\sqrt{N}$ and $\frac{3}{2}\sqrt{N}$. Obviously, this greatly diminishes the task of answering the question of the solvability of the equation $x^2-2y^2=N$.

BIBLIOGRAPHY

1. Carmichael, Robert D. The Theory of Numbers and Diophantine Analysis. New York: Dover Publications, Inc., 1959.

2. Dickson, Leonard Eugene. Modern Elementary Theory of Numbers. Chicago: The University of Chicago Press, 1939.

3. Dickson, Leonard Eugene. Studies in the Theory of Numbers. Chicago: The University of Chicago Press, 1930.

4. Griffin, Harriet. Elementary Theory of Numbers. New York: McGraw-Hill Book Company, Inc., 1954.

5. Jones, Burton W. The Theory of Numbers. New York: Holt, Rinehart and Winston, 1961.

6. LeVeque, William Judson. Topics in Number Theory, Volumes I and II. Reading, Mass.: Addison - Wesley Publishing Company, Inc., 1956.

7. Parker, S. T. "A Pair of Recursion Relations", American Mathematical Monthly, Volume 54, 1947, pp. 97 - 100.

ACKNOWLEDGMENT

The author wishes to express his appreciation to Dr. R. L. Yates, who has given generously of his time in criticism and editorial assistance during the writing of this paper.

PELL'S EQUATION

by

ROY LOUIS NEFF

B. S., Bethany College, 1964

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1966

The purpose of this paper is to investigate the conditions under which Pell's equation is solvable and, if it is solvable, to define a general solution form. Applications of solutions will be given without detail.

Pell's equation is a second degree Diophantine equation of the form $x^2-Dy^2 = N$, with the restriction that D and N be integers and that D not be a square. This equation form is so named as a result of an historical error rather than as a result of Pell's contribution to its solution.

The paper is divided into three sections. In the first section the specific case where $|N| = 1$ is dealt with. The situations where $N = 1$ and where $N = -1$ are considered separately within this section. In the second section the case where $|N| = 4$ is treated, again with the situations $N = 4$ and $N = -4$ being considered separately. Finally, in the third section the case $|N| > 0$ in general is considered, without benefit of either of the previous restrictions, but nevertheless encompassing both of those cases as sub-cases of the overall situation.