Brigham Young University
BYU ScholarsArchive

Theses and Dissertations

2009-07-08

Some Congruence Properties of Pell's Equation

Nathan C. Priddis
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

 Part of the Mathematics Commons

BYU ScholarsArchive Citation

Priddis, Nathan C., "Some Congruence Properties of Pell's Equation" (2009). *Theses and Dissertations*. 1798.
https://scholarsarchive.byu.edu/etd/1798

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

SOME CONGRUENCE PROPERTIES OF PELL'S EQUATION

by

Nathan C. Priddis

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Mathematics

Brigham Young University

August 2009

Copyright © 2009 Nathan C. Priddis

All Rights Reserved

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Nathan C. Priddis

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Jasbir S. Chahal, Chair
Date	Wayne Barrett
Date	David Cardon

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Nathan C. Priddis in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date	Jasbir S. Chahal Chair, Graduate Committee
Accepted for the Department	Tyler Jarvis Department Chair
Accepted for the College	Thomas Sederberg, Associate Dean College of Physical and Mathematical Sciences

ABSTRACT

SOME CONGRUENCE PROPERTIES OF PELL'S EQUATION

Nathan C. Priddis

Department of Mathematics

Master of Science

In this thesis I will outline the impact of Pell's equation on various branches of number theory, as well as some of the history. I will also discuss some recently discovered properties of the solutions of Pell's equation.

ACKNOWLEDGMENTS

Thanks go to my wife Amy for her help and patience through the long process of writing and editing this thesis, also to my advisor, Dr. Chahal for his insights and several crucial ideas that led to these results. Finally I would like to thank my father, Robert Priddis, as well as Emily Manwaring and my graduate committee for valuable feedback.

Contents

List of Tables

1 Introduction

The term *Pell's equation* is used to refer to equations of the form

$$x^2 - Ny^2 = 1. \tag{1}$$

with $N > 1$. There is perhaps no other equation that has influenced the development of number theory as much as Pell's equation.

As with many Diophantine equations, we are particularly interested in integer solutions to (1). From the Diophantine point of view, among all conics the hyperbola, namely the solutions to (1), is the only non-trivial case to study. For example, the circle $x^2 + y^2 = 1$ has only four integer solutions. The parabola, $y = ax^2$ has an integer solution for each integer, x.

Geometrically, the set of solutions to (1) (including non-integer solutions) has two connected components (see Fig.1). The integer solutions on the component with $x > 0$ form an infinite cyclic group (see Section 3), from which the solutions on the other component may be obtained by changing the sign for x. Thus it suffices to concentrate on the solutions with $x > 0$.

Of course if N is square, then $Ny^2 + 1$ is not square unless $y = 0$. So there is only the trivial solution to (1) if N is a square. Furthermore, if N has square factors, a change of variables will reduce the equation to an equivalent equation of the same form, with square-free N. So in general, when talking about Pell's equation, we will only consider N to be a square-free integer. We will let H denote the set of all solutions to (1) for a fixed N over some field.

One of the reasons that Pell's equation is so important, is that its solution set provides an example of many important objects in number theory. Among other things, Pell's equation is a Diophantine equation, and the set of solutions is an example of both an algebraic group, and an algebraic torus. Furthermore, it was the study of

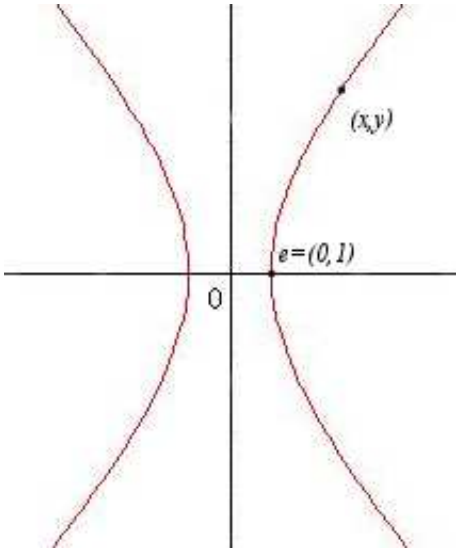

Figure 1: Pell's equation defines a hyperbola.

Pell's equation that led to Dirichlet's Unit Theorem. We will consider each of these ideas separately, and show how Pell's equation is related to each idea.

1.1 Diophantine Equations

A Diophantine equation is a polynomial equation

$$f(x_1, x_2, \ldots, x_n) = 0 \tag{2}$$

in the several variables $x_1, x_2, \ldots, x_n$ with integer coefficients. In number theory, we want solutions of (2) also in integers, or sometimes in the rational numbers $\mathbb{Q}$.

Suppose we have several Diophantine equations, or more generally, a set of polynomials

$$f_j(x_1, x_2, \ldots, x_n) = 0$$

with $1 \leq j \leq m$ for some positive integer m with rational coefficients. If we allow the solutions to be contained in an algebraically closed field L, then the set of solutions

forms an algebraic variety V defined over $\mathbb{Q}$. If K is a subfield of L, we call the points $(x_1, x_2, \ldots, x_n) \in V$ with $x_j \in K$ the set of K-rational points, or simply the set of K-points of V, and denote it by $V(K)$. In number theory we are interested in the case when K is a number field, in particular, $K = \mathbb{Q}$. The set of integer points on V is a subset of the lattice $\mathbb{Z}^n$, and is a lattice in its own right when the integer points on V form a group.

It is clear that Pell's equation is a Diophantine equation. The set H is a variety and the set of integer solutions forms a lattice in H.

1.2 Algebraic Groups

The study of the variety becomes more interesting when it has the structure of a group. It is a fact that H is an algebraic group. this will become apparent in Section 3. We call a variety V an algebraic group defined over $\mathbb{Q}$, when its identity e has rational coordinates, and the inverse map $x \to x^{-1}$ and multiplication map $(x, y) \to xy$ are given by polynomials in the coordinates of x and y with coefficients in $\mathbb{Q}$. (Unless otherwise stated, all algebraic groups are assumed to be defined over $\mathbb{Q}$.)

The simplest example of an algebraic group is SL_n, the group of $n \times n$ matrices $\big(x_{ij}\big)$. The group is a hypersurface defined by the single equation

$$\det\big(x_{ij}\big) - 1 = 0$$

which is clearly a polynomial equation in the n^2 variables x_{ij}. Recall that each entry of the adjoint matrix $\operatorname{adj} x$ is a polynomial in the variables x_{ij}, being the determinant of an $(n-1) \times (n-1)$ matrix obtained from x by deleting its i-th row and j-th column. Recall further that

$$x^{-1} = \frac{1}{\det(x)} \cdot \operatorname{adj} x = \operatorname{adj} x.$$

Hence, each entry of x^{-1} is a polynomial in the variables x_{ij}. Clearly each entry of the product xy is a polynomial in x_{ij} and y_{ij}. Finally, the identity of the group SL_n, namely, the identity matrix I, is in $\mathbb{Q}^{n^2}$. Thus SL_n is an algebraic group defined over $\mathbb{Q}$.

Another example is the group GL_n of invertible $n \times n$ matrices. This group is an algebraic variety in $n^2 + 1$ variables x_{ij} and z defined by the single equation

$$z \cdot \det(x_{ij}) = 1. \tag{3}$$

Equation (3) is equivalent to saying $\det(x_{ij}) \neq 0$. Similar to SL_n, we see that for $x \in \mathrm{GL}_n$, we have

$$x^{-1} = \frac{1}{\det(x)} \cdot \operatorname{adj} x = z \cdot \operatorname{adj} x.$$

The rest of the argument showing that GL_n is an algebraic group is much the same as with SL_n. We will also see in Section 3 that the set of integer solutions to (1) forms an algebraic group.

It is a standard fact that any algebraic group can be realized as a subgroup of GL_n for some $n \geq 1$. In fact, if (x, y) is a solution to (1), then the map

$$(x, y) \mapsto \begin{pmatrix} x & y \\ Ny & x \end{pmatrix}$$

is an injective group homomorphism into the group $\mathrm{SL}_2(\mathbb{Z})$, a subgroup of $\mathrm{GL}_2(\mathbb{Q})$. This fact will become apparent after we have defined the group operation on G. In fact, the requirement that (x, y) is a solution to (1) is equivalent to saying that

$$\det \begin{pmatrix} x & y \\ Ny & x \end{pmatrix} = 1.$$

1.3 Algebraic Tori

An algebraic group T is called a *torus* if (as an algebraic group) T is isomorphic to a group of diagonal matrices. We say T *splits* over an extension K of $\mathbb{Q}$ if this isomorphism is given by polynomials with coefficients in K.

We have seen that H can be realized as a subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Let

$$g = \begin{pmatrix} \sqrt{N} & 1 \\ N & -\sqrt{N} \end{pmatrix}.$$

Consider the map

$$\begin{pmatrix} x & y \\ Ny & x \end{pmatrix} \mapsto g \begin{pmatrix} x & y \\ Ny & x \end{pmatrix} g^{-1} = \begin{pmatrix} x + y\sqrt{N} & 0 \\ 0 & x - y\sqrt{N} \end{pmatrix}.$$

This map gives us a bijection from H into the group of invertible diagonal matrices with entries in $\mathbb{Q}(\sqrt{N})$. From here, it can easily be checked that this defines an isomorphism of algebraic groups. We leave the details as an exercise.

1.4 Dirichlet's Unit Theorem

When studying certain fields, we are interested in a set called the *ring of integers*, and the *group of units* of this ring. Dirichlet's Unit Theorem tells us exactly the structure of the group of units for any number field.

In order to understand Dirichlet's Unit Theorem, we must give a little more background in number fields.

A *number field* K is a subfield of $\mathbb{C}$ such that $\dim_{\mathbb{Q}}(K)$ is finite. For example, if $N > 1$ is a square-free integer the set $\mathbb{Q}(\sqrt{N}) = \left\{ r + s\sqrt{N} | r, s \in \mathbb{Q} \right\}$ is a field with $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{N}) = 2$.

The set of integers, $\mathbb{Z}$ forms an important subset of the set of rational numbers,

$\mathbb{Q}$. We can define a similar notion for any number field. We call the set

$$\mathscr{O}_K = \{\alpha \in K | \alpha \text{ is a root of a monic polynomial with coefficients in } \mathbb{Z}\}$$

the *ring of integers* of K. It is a well-known fact that $\mathscr{O}_K$ is a subring of K.

The *group of units* of a ring A is the set

$$A^\times = \{a \in A | ab = 1 \text{ for some } b \in A\}.$$

In other words, the group of units is the set of invertible integers. For example, if we take the ring of $n \times n$ matrices, then the group of units is the group GL_n. Dirichlet, a German mathematician, gave a complete description of the group of units of $\mathscr{O}_K$ as follows:

Theorem 1 (Dirichlet's Unit Theorem)**.** *The group of units $\mathscr{O}_K^\times \cong W_K \times \mathbb{Z}^r$, where r is determined by K and W_K is the set of roots of unity contained in K.*

Consider the quadratic field $\mathbb{Q}(\sqrt{N})$. For this field, $r = 1$, so we have

$$\mathscr{O}_{\mathbb{Q}(\sqrt{N})}^\times \cong \{\pm 1\} \times \mathbb{Z}.$$

Furthermore, if $N \equiv 2, 3 \pmod 4$, then the group of units of $\mathscr{O}_{\mathbb{Q}(\sqrt{N})}$ is essentially the set of solutions to (1). In other words, the set of integer solutions to (1) is isomorphic to $\{\pm 1\} \times \mathbb{Z}$. We explain this in more detail in Appendix B. A proof Dirichlet's Unit Theorem for quadratic fields is also given there.

2 History

Pell's equation has a long and interesting history. We find the first reference in a problem given by Archimedes in 200 B.C. Mathematicians in India had an interesting

solution to Pell's equation dating around the 7th century A.D. We will discuss both of these pieces of history, as well as the modern treatment of Pell's equation beginning with Fermat, and culminating in Lagrange's treatise on Pell's equation.

2.1 The Cattle Problem

Pell's equation dates back to as early as 200 B.C. Gotthold Lessing, a German literary figure, discovered a manuscript written by Archimedes in the Wolfenbüttel Library in 1773. On the manuscript, Archimedes presented a problem to the mathematicians in Alexandria which has come to be known as the "cattle problem." The following is a translation of the problem by Ivor Thomas [5, p. 203-207]:

> If thou art diligent and wise, O stranger, compute the number of cattle of the Sun, who once upon a time grazed on the fields of the Thrinacian isle of Sicily, divided into four herds of different colours, one milk white, another glossy black, the third yellow and the last dappled. In each herd were bulls, mighty in number according to these proportions: Understand, stranger, that the white bulls were equal to half and a third of the black together with the whole of the yellow, while the black were equal to the fourth part of the dappled and the fifth, together with, once more, the whole of the yellow. Observe further that the remaining bulls, the dappled, were equal to a sixth part of the white and a seventh, together with all the yellow. These were the proportions of the cows: The white were precisely equal to the third part and a fourth of the whole herd of black; while the black were equal to the fourth part once more of the dappled and with it the fifth part, when all, including the bulls, went to pasture together. Now the dappled in four parts were equal in number to a fifth part and a sixth of the yellow herd. Finally the yellow were in number equal to a

> sixth part and a seventh of the white herd. If thou canst accurately tell, O stranger, the number of cattle of the Sun, giving separately the number of well-fed bulls and again the number of females according to each colour, thou wouldst not be called unskilled or ignorant of numbers, but not yet shalt thou be numbered among the wise. But come, understand also all these conditions regarding the cows of the Sun. When the white bulls mingled their number with the black, they stood firm, equal in depth and breadth, and the plains of Thrinacia, streching far in all ways, were filled with their multitude. Again, when the yellow and the dappled bulls were gathered into one herd they stood in such a manner that their number, beginning from one, grew slowly greater till it completed a triangular figure, there being no bulls of other colours in their midst nor none of them lacking. If thou art able, O stranger, to find out all these things and gather them together in your mind, giving all the relations, thou shalt depart crowned with glory and knowing that thou hast been adjudged perfect in this species of wisdom.

To solve this problem, we need to use Pell's equation, as we will demonstrate. Heath also gives a solution in [4], and Weil discusses how the problem leads to the Pell equation in [7]. If we let X, Y, Z, W denote the numbers of white, black, yellow, and dappled bulls, resp., and x, y, z, w represent the numbers of white, black, yellow, and dappled cows, resp., then the first half of the epigram gives us the following seven equations in eight variables:

$$X = (\tfrac{1}{2} + \tfrac{1}{3})Y + Z$$

$$Y = (\tfrac{1}{4} + \tfrac{1}{5})W + Z$$

$$W = (\tfrac{1}{6} + \tfrac{1}{7})X + Z$$

$$x = (\tfrac{1}{3} + \tfrac{1}{4})(Y + y)$$
$$y = (\tfrac{1}{4} + \tfrac{1}{5})(W + w)$$
$$w = (\tfrac{1}{5} + \tfrac{1}{6})(Z + z)$$
$$z = (\tfrac{1}{6} + \tfrac{1}{7})(X + x)$$

The solution to the first three equations can found using elementary linear algebra techniques. Up to a scaling factor the solution is

$$X = 2226t, \quad Y = 1602t, \quad Z = 891t, \quad W = 1580t.$$

When we substitute this solution into the last four equations, we get four equations in five unknowns:

$$x = \frac{7}{12} + \frac{1869}{2}t$$
$$y = \frac{9}{20} + 711t$$
$$w = \frac{11}{30} + \frac{3267}{10}t$$
$$z = \frac{13}{42} + 689t$$

We require a solution in integers. The smallest value for t that yields such a solution is 4657, which happens to be prime. The solution up to a scaling factor is

$$X = 10,366,482 \cdot n; \qquad x = 7,206,360 \cdot n$$
$$Y = 7,460,514 \cdot n; \qquad y = 4,893,246 \cdot n$$
$$Z = 4,149,387 \cdot n; \qquad z = 5,439,213 \cdot n$$
$$W = 7,358,060 \cdot n; \qquad w = 3,515,820 \cdot n.$$

The second half of the problem gets more difficult. In essence, it says $X + Y$ is a perfect square, and $Z + W$ is a triangular number. That is,

$$X + Y = p^2 \tag{4}$$

$$Z + W = \frac{q(q+1)}{2} \tag{5}$$

for some integers p and q. Combining our previous solution and (4), we can write

$$X + Y = 17826996 \cdot n = 2^2 \cdot 3 \cdot 11 \cdot 29 \cdot 4657 \cdot n = p^2.$$

It is clear that (4) will be satisfied if we find an integer U, such that

$$n = 3 \cdot 11 \cdot 29 \cdot 4657 \cdot U^2.$$

Let $A = 3 \cdot 11 \cdot 29 \cdot 4657$.

Adding together Z and W, we get

$$Z + W = 115074473 \cdot n = 7 \cdot 353 \cdot 4657 \cdot n.$$

Let $B = 7 \cdot 353 \cdot 4657$. It can be easily verified from (5) that if we multiply a triangular number by 8, and then add 1, the result is a perfect square. Therefore, we can write $8(Z + W) + 1 = 8Bn + 1 = V^2$ for some integer V, and the two equations (4) and (5) become

$$V^2 - 8ABU^2 = 1.$$

We can make one further simplification, since $8AB$ is not square-free. We notice $8AB = 2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 29 \cdot 353 \cdot 4657^2$. Thus we put $u = 2 \cdot 4657 \cdot U$, and $v = V$ to obtain

$$v^2 - 4729494u^2 = 1 \tag{6}$$

So the cattle problem is reduced to finding the integer solutions to equation (1), with $N = 4,729,494$.* The "smallest" non-trivial integer solution to (6) is

$$v = 109931986732829734979866232821433543901088049$$

$$u = 50549485234315033074477819735540408986340.$$

In this solution, the first number has 45 digits and the second number has 41 digits.

However, to solve the cattle problem, because of our simplification, we require that the second coordinate be divisible by 9314. As we will later see, all solutions of (6) can be found from the solution given above. However, it has been shown in [3] that the number of cattle required to solve the problem is an integer with 206,545 digits. The complete solution was not found out until 1965. (For more on this solution, see [8] and [4]. The number representing the total number of cattle is found in the Unpublished Mathematical Tables of *Mathematics of Computation*.)

2.2 Cakravāla

The Indian school of mathematics, particularly the work of Brahmagupta and Bhāskara had an interesting method for calculating integer solutions to (1). The works of Brahmagupta, dating back to the seventh century, contain a method for combining certain solutions to form new solutions. This method was known as the *bhāvanā*. This method could also be used in some cases for finding solutions to (1), given some other information. These solutions were used, among other things, for approximating $\sqrt{N}$ with rational numbers. Indeed, if (x, y) is a solution to (1) with $x > 0$ and $y > 0$,

*Some authors will reduce the problem to Pell's equation with $N = 410286423278424$. In that case, they have not removed the square-free factors.

then we have $\left(\frac{x}{y}\right)^2 - N = \frac{1}{y^2}$, and so

$$\frac{x}{y} - \sqrt{N} = \frac{1}{y(x + y\sqrt{N})} < \frac{1}{2y^2}.$$

The inequality follows from the fact that $\frac{x}{y} > \sqrt{N} > 1$, and so $x + y\sqrt{N} > 2y$. So for large values of y, the approximation is quite good; however, it requires that we know a solution to (1).

In order to understand the *bhāvanā*, we need to consider equations of the form

$$x^2 - Ny^2 = m \tag{7}$$

where m is a non-zero (positive or negative) integer. Notice that the Pell equation is a special case of (7). In order to make the following discussion easier to follow, we shall adopt the notation of André Weil [7]. We will denote by the triple $(x, y; m)$ a solution to (7). We call x the *major root*, y the *minor root*, and m the *additive*.

The *bhāvanā* is based on *Brahmagupta's identity*

$$(x^2 - Ny^2)(z^2 - Nw^2) = (xz \pm Nyw)^2 - N(xw \pm yz)^2.$$

The verification of this identity is a straightforward calculation. Given two solutions $(x, y; m_1)$ and $(z, w; m_2)$, then the *bhāvanā* is

$$(x, y; m_1) * (z, w; m_2) = (xz + Nyw, xw + yz; m_1 m_2). \tag{8}$$

The *bhāvanā* always yields a solution to (7) with additive $m_1 m_2$, because of Brahmagupta's identity. If $m_1 = 1$, then the *bhāvanā* yields another solution, different from the first, with additive m_2. Continued use of the *bhāvanā*, will then yield infinitely many solutions to (7) with additive m_2. Similary if we know a single solution

$(p, q; 1)$ to (1), using the *bhāvanā*, we get a new solution $(p^2 + Nq^2, 2pq; 1)$, with first coordinate strictly larger than p. If we continue to use the *bhāvanā*, we see that we actually obtain infinitely many solutions to (1). This method already enabled Brahmagupta to solve the Pell equation in many cases. For example, if one composes $(x, y; m)$ with itself, then one obtains $(X, Y; M)$, with $M = m^2$. Now if X/m, and Y/m are integers, then we have a solution to the Pell equation $(X/m, Y/m; 1)$.

A more complete method for solving (1) was developed in the twelfth century and is found in the works of Bhaskara, and a less well-known mathematician known as Jayadeva. This method is known as the *cakravāla*. The basic idea of the method is this: given a solution $(p, q; m)$, we construct another solution $(x, y; M)$ with $M = mm'$ so that m' is "small." Then the composition $(p, q; m) * (x, y; M)$ yields the solution $(X, Y; m^2m')$. The construction will ensure that X/m and Y/m are integers, and so we get a new solution $(X/m, Y/m; m')$. The process is then continued in this way, each time getting a new additive, m'', m''', etc. until the additive becomes equal to 1.

More specifically, we begin with a solution $(p_i, q_i; m_i)$. We may assume that q_i and m_i are relatively prime to each other. If this were not the case, then there would exist a square-free integer $d \neq 1$ such that $q_i = qd$, and $m_i = md$, and

$$p_i^2 = m_i + Nq_i^2 = md + Nq^2d^2$$

so that d would divide p_i making d^2 divide m_i. Then we would have the solution

$$\left(\frac{p_i}{d}, \frac{q_i}{d}; \frac{m_i}{d^2}\right).$$

Next, we construct a triple $(x_i, 1; x_i^2 - N)$, by choosing x_i so that $p_i + q_i x_i$[†] is a multiple of m_i. In order to find such a number, the Indian mathematicians would refer

[†]The reader should notice that this is the minor root of the subsequent composition.

to a method they called the *kuttaka*–better known today as the Chinese remainder theorem. Further, they would insist that we choose x_i in such a way so that $|N - x_i^2|$ is as small as possible. We will discuss these requirements in more detail after we explain the method. Using the *bhāvanā* with this newly constructed solution gives us

$$(p_i, q_i; m_i) * (x_i, 1; x_i^2 - N) = (p_i x_i + N q_i, p_i + q_i x_i; m_i(x_i^2 - N)).$$

Put $X_i = p_i x_i + N q_i$ and $Y_i = p_i + q_i x_i$. Now we observe that

$$\begin{aligned} q_i^2(x_i^2 - N) &= q_i^2 x_i^2 - p_i^2 + m_i \\ &= (q_i x_i + p_i)(q_i x_i - p_i) + m_i. \end{aligned}$$

The first equality follows from (7), since $p_i^2 - m_i = N q_i^2$. Now m_i divides $Y_i = q_i x_i + p_i$ by construction, and m_i is relatively prime to q_i, so m_i divides $x_i^2 - N$. Furthermore, $X_i^2 = m_i(x_i^2 - N) + N Y_i^2$ from the composition. Thus m_i^2 divides X_i^2, and m_i divides X_i.

Now we put $p_{i+1} = \frac{X_i}{|m_i|}$, $q_{i+1} = \frac{Y_i}{|m_i|}$, and $m_{i+1} = \frac{(x_i^2 - N)}{m_i}$, and we have a new solution $(p_{i+1}, q_{i+1}; m_{i+1})$. We continue this process until we reach $m_{i+k} = 1$ for some positive integer k.[‡]

Nowhere in the writings of the Indian mathematicians do we find proof that the process should yield a solution to (1). However, their ability to find solutions in hard cases surely gave them confidence in the generality of the method. The reason for calling the method *cakravāla*, which means "circular", is that the sequence of additives repeats itself (see [7]).

We began by taking a known solution $(p_i, q_i; m_i)$, so to begin the algorithm, we need to find a solution $(p_0, q_0; m_0)$. In practice, the Indian mathematicians would

[‡]In practice, they would stop when they reached ± 1, ± 2, or ± 4, because of Brahmagupta's work with the *bhāvanā*. For more details, see [6].

choose $p_0 > 0$, such that p_0^2 is the closest square to N, above or below. Then we clearly have the triple $(p_0, 1; m_0)$ with $m_0 = p_0^2 - N$.

We make a few final remarks here before leaving this topic. In his book [7, p. 23], Weil notices that we actually don't need to use *kuttaka* to define the x_i. In defining x_0, we simply choose $x_0 \equiv -p_0 \pmod{|m_0|}$ (but we remember to choose x_0 in the congruence class so that $|N - x_0^2|$ is as small as possible). Now having chosen x_i, we can simply choose x_{i+1}, so that $x_{i+1} \equiv -x_i \pmod{m_{i+1}}$. Indeed, from the composition, we have

$$\begin{aligned} Y_{i+1} &= p_{i+1} - x_i q_{i+1} \\ &= \frac{p_i x_i + N q_i - p_i x_i - q_i x_i^2}{|m_i|} \\ &= \frac{q_i(N - x_i^2)}{|m_i|} \\ &= \pm q_i m_{i+1} \end{aligned}$$

so clearly Y_{i+1} is divisible by m_{i+1} with this choice of x_{i+1}. Also, if we choose x_i so that $x_i < \sqrt{N} < x_i + |m_i|$, and we assume $|m_i| < 2\sqrt{N}$, then we have

$$2\sqrt{N} < x_i + \sqrt{N} + |m_i|.$$

If $x_i + \sqrt{N} < 0$, this would imply that $2\sqrt{N} < |m_i|$, contrary to assumption. So if $|m_i| < 2\sqrt{N}$, then $x_i + \sqrt{N} > 0$ and

$$0 < N - x_i^2 = (\sqrt{N} - x_i)(\sqrt{N} + x_i) < 2|m_i|\sqrt{N}$$

and therefore $|m_{i+1}| = \frac{N - x_i^2}{|m_i|} < 2\sqrt{N}$. If there is a better choice (remember the prescription said that $|N - x_i^2|$ should be as small as possible), i.e. if $y_i \equiv x_i \pmod{|m_i|}$ with $|N - y_i^2| < |N - x_i^2|$, then using y_i instead of x_i would yield m_{i+1} still with

$|m_{i+1}| < 2\sqrt{N}$. Hence, as Weil notes, the integers $m_0, m_1, m_2, \ldots$ are bounded, and they must repeat themselves.

Finally, let us remark that if we begin with $p_0 < \sqrt{N} < p_0 + 1$, and always choose x_i so that $x_i < \sqrt{N} < x_i + |m_i|$, then the numbers $\frac{p_i}{q_i}$ correspond to the convergents of the simple continued fraction for $\sqrt{N}$. (We will not prove this fact here. For an outline of the proof, see [6, p. 29-30].)

Example. In order to demonstrate how effective the *cakravāla* can be, we consider an example. Let $N = 41$. The closest square to 41 is 36, so we put

$$p_0 = 6, \quad q_0 = 1, \quad m_0 = -5.$$

Now we choose $x_0 \equiv -6 \pmod 5$, so that $x_0 < \sqrt{41} < x_0 + 5$. That is $x_0 = 4$. Using the *bhāvanā* we get

$$(6, 1; -5) * (4, 1; -25) = (65, 10; 125).$$

Upon division by 5 and -5, resp., we have

$$p_1 = 13, \quad q_1 = 2, \quad m_1 = 5.$$

Using the recommendation of Weil, we now need to find $x_1 \equiv -x_0 \pmod 5$, with $x_1 < \sqrt{41} < x_1 + 5$. Recall $x_0 = -4$, so the desired value is $x_1 = 6$. Using the *bhāvanā*,

$$(13, 2; 5) * (6, 1; -5) = (160, 25; -25).$$

Upon dividing by 5, we get

$$p_2 = 32, \quad q_2 = 5, \quad m_2 = -1.$$

Continuing in this manner, we obtain the following solutions:

$$(397, 62; 5), \quad (826, 129; -5)$$

and finally

$$(2049, 320; 1). \tag{9}$$

Since the additive is 1, we have a solution to Pell's equation, $(2049, 320)$. If we were to continue with the calculations, we would see the following "circular" pattern of additives,

$$-5, 5, -1, 5, -5, 1, -5, 5, -1, 5, -5, 1, \ldots.$$

Finally, we notice that in calculating (9), we could have stopped using the *cakravāla* after obtaining $(32, 5; -1)$, for we can use the *bhāvanā* to calculate

$$(32, 5; -1) * (32, 5; -1) = (2049, 320; 1).$$

2.3 More Recent History

In the seventeenth century, the French Mathematician, Fermat, challenged several English mathematicians, in particular, Wallis and Brouncker, to solve a few problems, one of which was to solve (1) in integers. Fermat could not have known of Archimedes' cattle problem, nor of the previous works of the Indian mathematicians. In [7, p. 81] Weil remarks, "What would have been Fermat's astonishment, if some missionary, just back from India, had told him that this problem had been successfully tackled there by native mathematicians almost six centuries earlier!"

Wallis and Brouncker, first sent him a solution in rational numbers, which proved to be quite simple. They later contrived a method yielding a solution in integers that was much the same as the *cakravāla* and the later treatment by Lagrange in continued

fractions. In the method they devised, we assume the existence of a solution (x, y) and then write $x = ym + z$. Upon substitution into (1), the ordered pair (y, z) becomes a solution to a related equation

$$Ay^2 - 2Byz - Cz^2 = \pm 1.$$

The number m was taken to be the largest integer smaller than the positive root of (1), and it was assumed then that z was less that y. The procedure was then continued, each time reducing the equation to another one with a smaller solution until one reached an equation with an obvious solution, usually $(1, 0)$.[§] The subsitutions were then traced back to obtain the desired solution (x, y).

This method was enough to find solutions in all of the cases that Brouncker and Wallis tried, including $N = 61$ and $N = 109$, in which cases the first solutions are $(1766319049, 226153980)$ and $(158070671986249, 15140424455100)$, resp.

This method is essentially what Fermat called the "method of descent." They apparently did not see the need to check any further that this method would yield a solution for all choices of N. [¶]

In 1730, Euler wrote a letter to Goldbach pointing out an error in the latter's claim concerning a certain problem regarding triangular numbers. In the letter, Euler noted that the problem in question reduced to an equation of the form $x^2 - 8y^2 = 1$. Euler wrote, "Such problems have been agitated between Wallis and Fermat ... and the Englishman Pell devised for them a peculiar method described in Wallis's works" [7, p. 174].

In [7], Weil notes that this was a mistake of Euler's, since Pell had nothing to do with the problem at all. In [2, p.352], Euler again attributed the method devised by Wallis and Brouncker to Pell. Because of Euler's mistake, (1) is called Pell's equation

[§]This method is explained more fully in [2, p. 351-60].

[¶]This method will in fact give a solution in every case. see [7]

even today.

In 1768, Lagrange wrote the first published proof of the fact that Pell's equation always yields a solution in integers. The proof relied on the concept of continued fractions. He also gave a definitive treatment of Pell's equation in three papers presented in 1768, 1769, and 1770, resp. to the Berlin Academy along with several results about continued fractions. This was before he received a copy of Euler's *Elements of Algebra*. When he received a copy, he conceived a plan to have it translated into French, and to add to it an improved exposition of his three papers presented earlier. We will give Lagrange's proof in section 3.2.

3 Group Structure

We now consider the set of integer solutions to (1) with $x > 0$. In other words consider the set

$$G = \left\{(x, y) \in \mathbb{Z}^2 | x^2 - Ny^2 = 1, \text{ and } x > 0\right\}.$$

We will show that this set forms a group, and we will give the group structure.

The set G constitutes "half" of the solutions to Pell's equation. Notice that if $(x, y) \in G$, then $(-x, y)$ is also a solution to (1). It turns out that G is a group using the binary operation defined by the *bhāvanā*. In other words, given two elements $(x_1, y_1), (x_2, y_2) \in G$, we define

$$(x_1, y_1) * (x_2, y_2) = (x_1x_2 + Ny_1y_2, x_1y_2 + x_2y_1). \tag{10}$$

Recall Brahmagupta's identity, which states

$$(x^2 - Ny^2)(z^2 - Nw^2) = (xz \pm Nyw)^2 - N(xw \pm yz)^2.$$

This identity shows that G is closed under the binary operation. The identity element of G is $(1, 0)$, and if $(x, y) \in G$, then the inverse, $(x, y)^{-1}$, is the element $(x, -y)$. We leave it as an exercise to check these facts, and that (10) defines an associative operation. (Notice that this is actually an algebraic group.) Regarding the structure of G, we have the following theorem:

Theorem 2. *G is an infinite cyclic group.*

In other words, every solution to (1) is generated by a "least" solution. As we mentioned previously, Lagrange was the first to provide a proof for this fact. The proof requires some knowledge of continued fractions, so before we can give Lagrange's proof, we will provide more background.

3.1 Continued Fractions

In this section, we recall the definition and some properties of continued fractions, which we will need. We will end with a proof that the integer solutions to (1) are found by computing the continued fraction expansion for $\sqrt{N}$. A more detailed exposition of continued fractions can be found among other places in [2] and [3] (see also notes to Chapter 10 in [3]). Much of what we explain here will be modeled after [3].

A *finite continued fraction* is an expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_K}}}}. \tag{11}$$

More generally, we call an expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} \tag{12}$$

a *continued fraction*. As this presentation is cumbersome, we shall henceforth write

$[a_0, a_1, a_2, \ldots, a_K]$ to represent (11), and $[a_0, a_1, a_2, \ldots]$ for (12). Furthermore, if a_0 is an integer and a_i is a positive integer for $i \geq 1$, then we say (11) is a *simple finite continued fraction* and (12) is a *simple continued fraction*.

In either case, we call the numbers $a_0, a_1, a_2, \ldots$ the *partial convergents*.

The following properties of finite continued fractions are clear from the definition:

$$\begin{aligned}
[a_0] &= a_0 \\
[a_0, a_1] &= a_0 + \frac{1}{a_1} = \frac{a_0 a_1 + 1}{a_1} \\
[a_0, a_1, \ldots, a_K] &= a_0 + \frac{1}{[a_1, a_2, \ldots, a_K]} \\
[a_0, a_1, \ldots, a_K] &= [a_0, a_1, \ldots, a_{K-1} + \frac{1}{a_K}]
\end{aligned}$$

Given a continued fraction (finite or infinite), we say $[a_0, a_1, \ldots, a_n]$ is the *n-th convergent* to $[a_0, a_1, a_2 \ldots]$ in the infinite case or to $[a_0, a_1, a_2, \ldots, a_K]$ if $0 \leq n \leq K$ in the finite case. Let $a'_n = [a_n, a_{n+1}, a_{n+2}, \ldots]$ in the infinite case or $a'_n = [a_n, a_{n+1}, a_{n+2}, \ldots, a_K]$ in the finite case. We say a'_n is the *n-th complete quotient*.

For ease in working with continued fractions, we also define the following sequences:

$$\begin{aligned}
&p_1 = a_0 && q_1 = 1 \\
&p_2 = a_1 p_1 + 1 && q_2 = a_1 q_1 \\
&p_{n+1} = a_n p_n + p_{n-1} && q_{n+1} = a_n q_n + q_{n-1}, \text{ for } n \geq 2.
\end{aligned} \tag{13}$$

These sequences are key to working with continued fraction. They also have some unique properties. The next few theorems will outline some of those properties.

Theorem 3. *If p_n and q_n are defined as above, then*

$$[a_0, a_1, a_2, \ldots, a_n] = \frac{p_{n+1}}{q_{n+1}}.$$

Proof. We prove this by induction on n. It is true, if $n = 0$, for $[a_0] = a_0 = \frac{p_1}{q_1}$.

Now suppose the statement is true for $[a_0, a_1, a_2, \ldots, a_m]$ whenever $m \leq n$, i.e.

$$[a_0, a_1, \ldots, a_m] = \frac{p_{m+1}}{q_{m+1}} = \frac{p_m a_m + p_{m-1}}{q_m a_m + q_{m-1}}.$$

Now we can write

$$\begin{aligned}
[a_0, a_1, \ldots, a_{n+1}] &= [a_0, a_1, \ldots, a_n + \frac{1}{a_{n+1}}] \\
&= \frac{p_n \left(a_n + \frac{1}{a_{n+1}}\right) + p_{n-1}}{q_n \left(a_n + \frac{1}{a_{n+1}}\right) + q_{n-1}} \\
&= \frac{a_{n+1}a_n p_n + p_n + a_{n+1}p_{n-1}}{a_{n+1}a_n q_n + q_n + a_{n+1}q_{n-1}} \\
&= \frac{a_{n+1}(a_n p_n + p_{n-1}) + p_n}{a_{n+1}(a_n q_n + q_{n-1}) + q_n} \\
&= \frac{a_{n+1}p_{n+1} + p_n}{a_{n+1}q_{n+1} + q_n} \\
&= \frac{p_{n+2}}{q_{n+2}}
\end{aligned}$$

□

Because of the previous theorem, we will also say that the numbers $\frac{p_{n+1}}{q_{n+1}}$ are convergents.

Theorem 4. *The numbers p_n and q_n also satisfy*

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^n$$

for $n \geq 2$, and

$$p_n q_{n-2} - p_{n-2} q_n = (-1)^{n-1} a_{n-1}$$

for $n \geq 3$.

Proof. The first statement is proven by induction on n. For $n = 2$, we have

$$p_2 q_1 - p_1 q_2 = (p_1 a_1 + 1) - p_1 a_1 q_1 = (-1)^2.$$

Now for $n \geq 2$, we have

$$\begin{aligned} p_{n+1}q_n - q_{n+1}p_n &= (a_n p_n + p_{n-1})q_n - p_n(a_n q_n + q_{n-1}) \\ &= p_{n-1}q_n - p_n q_{n-1} \\ &= -(p_n q_{n-1} - p_{n-1}q_n) \\ &= (-1)^{n+1} \end{aligned}$$

This proves the first statement. For the second statement, we have

$$\begin{aligned} p_{n+1}q_{n-1} - q_{n+1}p_{n-1} &= (a_n p_n + p_{n-1})q_{n-1} - p_{n-1}(a_n q_n + q_{n-1}) \\ &= a_n p_n q_{n-1} - a_n p_{n-1} q_n \\ &= a_n(p_n q_{n-1} - p_{n-1}q_n) \\ &= a_n(-1)^n. \end{aligned}$$

□

Because of this theorem, we can write

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{(-1)^{n-1}a_{n-1}}{q_n q_{n-2}}.$$

In particular, if n is even, then $\frac{p_n}{q_n} < \frac{p_{n-2}}{q_{n-2}}$, and if n is odd, $\frac{p_n}{q_n} > \frac{p_{n-2}}{q_{n-2}}$. In other words, we have two sequences of convergents, one increasing, and the other decreasing. We also know that

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^n}{q_n q_{n-1}}.$$

In particular $\frac{p_n}{q_n} > \frac{p_{n-1}}{q_{n-1}}$ whenever n is even.

Theorem 5. *If p_n and q_n are as above, then $q_n \geq q_{n-1}$ and $q_n \geq n-1$, whenever $n \geq 2$.*

Proof. We prove both statements by induction on n. Recall $q_1 = 1$, and so $q_2 = a_1 \geq$

$1 = q_1$. Now for $n > 2$, we have

$$\begin{aligned} q_{n+1} &= a_n q_n + q_{n-1} \\ &\geq q_n + q_{n-1} \\ &\geq q_n + 1 \\ &\geq (n-1) + 1 \end{aligned}$$

The last two inequalities follow from the induction hypothesis. □

Now suppose that $[a_0, a_1, a_2, \ldots]$ is an infinite simple continued fraction. From the previous theorems, we see that we have two sequences of convergents

$$\left\{\frac{p_{2n}}{q_{2n}}\right\} \text{ and } \left\{\frac{p_{2n+1}}{q_{2n+1}}\right\}.$$

The first is a strictly decreasing sequence, and the second is a strictly increasing sequence. We have also seen that every term in the first sequence is larger than every term in the second sequence. From this we know that both sequences converge. But we also know that

$$\left|\frac{p_{2n}}{q_{2n}} - \frac{p_{2n-1}}{q_{2n-1}}\right| = \frac{1}{q_{2n}q_{2n-1}} \leq \frac{1}{(2n-1)(2n-2)}.$$

Thus as $n \to \infty$, we see that $[a_0, a_1, a_2, \ldots, a_n]$ converges to some real number, x. We say that x is the *value* of the simple continued fraction $[a_0, a_1, a_2, \ldots]$. In other words, every infinite simple continued fraction represents some real number. We would like to know if every real number can be expressed as a simple continued fraction, and whether this expression is unique. We will deal with these questions in the next several theorems. First, however, we need to prove the following fact, which will be useful in proving several important results.

Theorem 6. *A finite simple continued fraction represented by an odd number of convergents can be represented by an even number, and visa versa.*

Proof. Given $[a_0, a_1, \ldots, a_n]$, with $a_n \geq 2$, we see that

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

Otherwise (if $a_n = 1$), then

$$[a_0, a_1, \ldots, a_{n-1}, 1] = [a_0, a_1, \ldots, a_{n-1} + 1]. \qquad \square$$

For example, the fraction $\frac{2}{3}$ can be expressed as $[0, 1, 1, 1]$, or as $[0, 1, 2]$.

$$\frac{2}{3} = 0 + \frac{1}{1 + \frac{1}{2}} = 0 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1}}}$$

We will now describe an algorithm for expressing any real number as a continued fraction. Let x be a real number. Let a_0 be the integer with $a_0 \leq x < a_0 + 1$, i.e. let a_0 be the largest number less than or equal to x. Put $x = a_0 + x_0$. If $x_0 \neq 0$, then $0 < x_0 < 1$, and so we define $a_1' = \frac{1}{x_1}$.

Now given a_i', let a_i be the largest integer smaller than or equal to a_i', i.e. $a_i \leq a_i' < a_i + 1$. Clearly $a_i \geq 1$. Again put $a_i' = a_i + x_i$. If $x_i \neq 0$, then $0 < x_i < 1$, so we define $a_{i+1}' = \frac{1}{x_i}$. Thus $x = [a_0'] = [a_0, a_1'] = \ldots$. Now this algorithm continues until $x_K = 0$ for some K. If we reach a value K such that $x_K = 0$, then $x = [a_0, a_1, \ldots, a_K]$.

Now suppose that x is rational. Let $x = \frac{p}{q}$ in lowest terms. By applying the algorithm we can write $\frac{p}{q} = a_0 + x_0$, or in other words

$$p = qa_0 + qx_0.$$

Apparently $r_1 = qx_0$ is an integer, and $0 \leq x_0 < 1$, so $0 \leq r_1 < q$. So now if $x_0 \neq 0$

(and therefore also $r_1 \neq 0$), we put $a_1' = \frac{1}{x_0} = \frac{q}{r_1}$. Again, we put $a_1' = a_1 + x_1$, or in other words

$$q = a_1 r_1 + x_1 r_1.$$

Again we see that $r_2 = x_1 r_1$ is an integer and $0 \leq r_2 < r_1 < q$. Continuing in this manner, we get a system of equations

$$p = qa_0 + r_1$$

$$q = r_1 a_1 + r_2$$

$$r_1 = a_2 r_2 + r_3$$

$$\cdots .$$

We recognize this as the Euclidean algorithm. We have a decreasing sequence of integers $q > r_1 > r_2 > \cdots \geq 0$, or $r_{K+1} = 0$ for some K. Let r_{K+1} be the first of the sequence with this property. We had $r_{K+1} = r_K x_K$, and therefore $x_K = 0$. So we have

$$\frac{p}{q} = [a_0, a_1, \ldots, a_K].$$

Because of this algorithm, we see that every rational number can be expressed as a finite simple continued fraction. The next theorem will be useful in proving some uniqueness conditions on simple continued fractions.

Theorem 7. *If a_n' is the n-th complete quotient of a finite simple continued fraction $[a_0, a_1, \ldots, a_K]$, then $a_n < a_n' < a_n + 1$, except $a_{K-1} = a_{K-1}' - 1$ if $a_K = 1$. Also if a_n' is the n-th complete quotient of an infinite simple continued fraction, then $a_n < a_n' < a_n + 1$.*

Proof. If $x = [a_0, a_1, \ldots, a_K]$, then $a_n' = [a_n, a_{n+1}, \ldots, a_K]$. If $K = 0$, then $a_0 = a_0'$. If on the other hand, $K > 0$, then $a_n' = a_n + \frac{1}{a_{n+1}'}$, for $0 \leq n \leq K - 1$, and $a_{n+1}' > 1$

whenever $0 \leq n \leq K-1$, except $a'_{n+1} = 1$, when $n = K-1$, and $a_K = 1$. Hence $a_n < a'_n < a_n + 1$ for $0 \leq n \leq K-1$, except in the case $a_K = 1$.

In the infinite case, suppose $x = [a_0, a_1, a_2, \dots]$. Then

$$\begin{aligned} a'_n &= [a_n, a_{n+1}, a_{n+2}, \dots] \\ &= \lim_{N\to\infty} [a_n, a_{n+1}, a_{n+2}, \dots, a_N] \\ &= a_n + \frac{1}{\lim_{N\to\infty}[a_{n+1}, a_{n+2}, \dots, a_N]} \\ &= a_n + \frac{1}{a'_{n+1}} \end{aligned}$$

In particular, $x = a'_0 = a_0 + \frac{1}{a'_1}$. Thus $a'_n > a_n \geq 1$ whenever $n \geq 1$. So $0 < \frac{1}{a'_{n+1}} < 1$. □

For any real number, regarding the n-th partial quotient, we can see that $x = a'_0$. Furthermore,

$$\begin{aligned} x &= a_0 + \frac{1}{a'_1} \\ &= \frac{a_0 a'_1 + 1}{a'_1} \\ &= \frac{a'_1 p_1 + 1}{a'_1 q_1} \end{aligned} \tag{14}$$

Continuing in this manner, suppose $x = \frac{a'_i p_i + p_{i-1}}{a'_i q_i + q_{i-1}}$. Then we see

$$\begin{aligned} x &= \frac{\left(a_i + \frac{1}{a'_{i+1}}\right) p_i + p_{i-1}}{\left(a_i + \frac{1}{a'_{i+1}}\right) q_i + q_{i-1}} \\ &= \frac{a'_{i+1}(a_i p_i + p_{i-1}) + p_i}{a'_{i+1}(a_i q_i + q_{i-1}) + q_i} \\ &= \frac{a'_{i+1} p_{i+1} + p_i}{a'_{i+1} q_{i+1} + q_i}. \end{aligned} \tag{15}$$

We will use this fact in the next theorem, which shows the uniqueness of the simple

continued fraction representation of any real number x.

Theorem 8. *Each irrational number x can be expressed in exactly one way as a simple continued fraction.*

Proof. If we do the continued fraction algorithm, the process will not terminate, as it will for finite continued fractions. So if x is irrational, we can write

$$x = [a_0, a_1'] = [a_0, a_1, \ldots, a_n']$$

where $a_{n+1}' = a_{n+1} + \frac{1}{a_{n+2}'} > a_{n+1}$. So from (15) we can write

$$x = \frac{a_{n+1}' p_{n+1} + p_n}{a_{n+1}' q_{n+1} + q_n}.$$

And now

$$\begin{aligned} x - \frac{p_n}{q_n} &= \frac{p_n q_{n+1} - p_{n+1} q_n}{q_{n+1}(a_{n+1}' q_{n+1} + q_n)} \\ &= \frac{1}{q_{n+1} q_{n+2}} \\ &\leq \frac{1}{n(n+1)} \end{aligned}$$

Now as $n \to \infty$, we see that $\left| x - \frac{p_n}{q_n} \right| \to 0$. In particular,

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

Thus the algorithm gives us a simple continued fraction whose value is x. By the previous theorem, this representation must be unique. □

Now we deal with the uniqueness of finite simple continued fractions. We will see that the representation of rational numbers by finite simple continued fractions is almost unique.

Theorem 9. *Every rational number can be expressed in exactly two ways as a finite simple continued fraction.*

Proof. From Theorem 6, we know that there are two representations to any finite simple continued fraction. One ends with 1, and the other does not. Furthermore, we saw that every rational number can be expressed as a finite simple continued fraction. So suppose $x = [a_0, a_1, \ldots, a_N] = [b_0, b_1, \ldots, b_M]$ with $a_N > 1$ and $b_M > 1$. From Theorem 7, $a_0 = b_0$. Now suppose that the first n partial quotients are identical, i.e. $a_i = b_i$ whenever $0 \leq i \leq n-1$. Then we have $[a_0, a_1, \ldots, a_{n-1}, a'_n] = [b_0, b_1, \ldots, b_{n-1}, b'_n]$. If $n = 1$, then we have $a_0 + \frac{1}{a'_1} = a_0 + \frac{1}{b'_1}$, and so $a'_1 = b'_1$ and therefore $a_1 = b_1$. If $n > 1$, then

$$\frac{a'_n p_n + p_{n-1}}{a'_n q_n + q_{n-1}} = \frac{b'_n p_n + p_{n-1}}{b'_n q_n + q_{n-1}}$$
$$\text{i.e. } (a'_n - b'_n)(p_n q_{n-1} - p_{n-1} q_n) = 0.$$

Therefore $a'_n - b'_n = 0$, and thus $a_n = b_n$. Now suppose that $N \leq M$. We have shown that $a_n = b_n$ whenever $0 \leq n \leq N$. If $M > N$, then

$$\frac{p_{N+1}}{q_{N+1}} = [a_0, a_1, \ldots, a_N] = [a_0, a_1, \ldots, a_N, b_{N+1} \ldots, b_M] = \frac{b'_n p_n + p_{n-1}}{b'_n q_n + q_{n-1}}$$

or in other words, $p_{N+1} q_N - p_N q_{N+1} = 0$, which is a contradiction. □

Having established the uniqueness of representations of simple continued fractions, and the fact that every real number has a representation as a simple continued fraction, we want to show that the integer solutions to (1) are found among the convergents to the simple continued fraction representation of $\sqrt{N}$. This fact is a result of the following technical lemma and the following theorem.

Lemma 10. *If*

$$x = \frac{P\omega + R}{Q\omega + S}$$

where $\omega > 1$, and P, Q, R, S are integers satisfying $PS - QR = \pm 1$ and $Q > S > 0$, then $\frac{R}{S}$ and $\frac{P}{Q}$ are consecutive convergents to the simple continued fraction expansion for x.

Proof. We write $\frac{P}{Q}$ as a simple continued fraction $[a_0, a_1, \ldots, a_n] = \frac{p_{n+1}}{q_{n+1}}$. We know $PS - QR = \pm 1 = (-1)^{n-1}$, since we can choose n even or odd at our leisure. We know that $\gcd(P, Q) = 1$, and $Q > 0$. But we also know that $\gcd(p_{n+1}, q_{n+1}) = 1$ and $q_{n+1} > 0$, so we have $P = p_{n+1}$ and $Q = q_{n+1}$. Thus

$$p_{n+1}S - q_{n+1}R = (-1)^{n-1} = p_{n+1}q_n - q_{n+1}p_n.$$

Now since $\gcd(p_n, q_n) = 1$, we see that $q_{n+1}|(S - q_n)$. But now we have $q_{n+1} = Q > S > 0$, and also $q_{n+1} \geq q_n$, so in particular, $|S - q_n| < q_{n+1}$. Therefore, $S - q_n = 0$, and so $S = q_n$, $R = p_n$. Now if we develop ω as a simple continued fraction, we must have

$$x = [a_0, a_1, \ldots, a_n, \omega]$$

Now if we expand the coninued fraction expansion for ω, we get $\omega = [a_n, a_{n+1}, a_{n+2}, \ldots]$, and the fractions $\frac{p_{n+1}}{q_{n+1}}$ and $\frac{p_n}{q_n}$ are consecutive convergents for x. □

Theorem 11. *If*

$$\left| \frac{p}{q} - x \right| < \frac{1}{2q^2},$$

then $\frac{p}{q}$ is a convergent in the simple continued fraction, whose value is x.

Proof. We express $\frac{p}{q}$ as a simple continued fraction $[a_0, a_1, \ldots, a_n]$, and we can choose

n even or odd as we please, so that

$$\frac{p}{q} - x = \frac{(-1)^{n-1}\theta}{q^2}$$

where $0 < \theta < \frac{1}{2}$. Now we put

$$X = \frac{\omega p_{n+1} + p_n}{\omega q_{n+1} + q_n}$$

where $\frac{p_{n+1}}{q_{n+1}}$ and $\frac{p_n}{q_n}$ are consecutive convergents to $\frac{p}{q}$. (To do this simple put $\omega = \frac{p_n - xq_n}{xq_{n+1} - p_{n+1}}$.)

Now

$$\frac{(-1)^{n-1}\theta}{q_{n+1}^2} = \frac{p_n}{q_n} - x = \frac{p_{n+1}q_n - p_n q_{n+1}}{q_{n+1}(\omega q_{n+1} + q_n)} = \frac{(-1)^{n-1}}{q_{n+1}(\omega q_{n+1} + q_n)}$$

so

$$\frac{\theta}{q_{n+1}} = \frac{1}{\omega q_{n+1} + q_n},$$

and

$$\omega = \frac{1}{\theta} - \frac{q_n}{q_{n+1}} > 1.$$

Therefore by the previous theorem, $\frac{p_n}{q_n}$ and $\frac{p_{n+1}}{q_{n+1}}$ are consecutive convergents to x. But $\frac{p}{q} = \frac{p_{n+1}}{q_{n+1}}$. □

We saw earlier that any integer solution (p, q) to (1) satisfies

$$\left| \frac{p}{q} - \sqrt{N} \right| < \frac{1}{2q^2}.$$

By the preceding theorem, $\frac{p}{q}$ is a convergent to the simple continued fraction expansion of $\sqrt{N}$.

3.2 Lagrange's Proof

In the solution to the cattle problem, we saw how difficult the solutions to (1) can be to find. It is true, however, that an integer solution always exists. In fact the set of integer solutions is an infinite cyclic group. If N is not square free, then we can make a change of variables, to reduce to the case that N is square-free.

If (x, y) is an integer solution to (1), then $x^2 - Ny^2 = 1$. In Section 2.2, we saw that $|\frac{x}{y} - \sqrt{N}| < \frac{1}{2y^2}$. Therefore, by Theorem 11, if a solution to (1) exists, it is to be found among the convergents of the simple continued fraction whose value is $\sqrt{N}$.[‖]

Here we give Lagrange's proof of the existence of a non-trivial solution to (1), no matter what N is, as well as his proof that all non-trivial integer solutions to (1) can be found from a single solution.

Lagrange's proof. In view of the preceding comments, we define the following quantities:

$$\begin{array}{lll} p_0 = 1, & q_0 = 0, & \mu_0 < \sqrt{N} \\ p_1 = \mu_0, & q_1 = 1, & \mu_1 < \frac{p_0 - \sqrt{N}q_0}{q_1\sqrt{N} - p_1} = \frac{1}{\sqrt{N} - \mu_0} \\ p_{i+1} = \mu_i p_i + p_{i-1} & q_{i+1} = \mu_i q_i + q_{i-1}, & \mu_{i+1} \leq \frac{p_i - q_i\sqrt{N}}{q_{i+1}\sqrt{N} - p_{i+1}} \end{array} \tag{16}$$

for all $i \geq 1$. In defining μ_i we take the greatest integer smaller than the above-named quantity, i.e. $\mu_{i+1} \leq \frac{p_i - q_i\sqrt{N}}{q_{i+1}\sqrt{N} - p_{i+1}} < \mu_{i+1} + 1$. Notice the similarity to (13).

[‖]Lagrange's proof was actually more general than the proof we give here. The method in the proof is the same, though. In [2], Lagrange gives a method for minimizing the quantity $Ax^2 + Bxy + Cy^2$ with integers x and y.

Now if we put $a_i = \frac{p_{i-1}-\sqrt{N}q_{i-1}}{\sqrt{N}q_i - p_i}$ for $i \geq 1$, then we can calculate

$$\begin{aligned} a_{i+1} &= \frac{p_i - \sqrt{N}q_i}{\sqrt{N}q_{i+1} - p_{i+1}} \\ &= \frac{p_i - \sqrt{N}q_i}{\sqrt{N}(\mu_i q_i + q_{i-1}) - (\mu_i p_i + p_{i-1})} \\ &= \frac{p_i - \sqrt{N}q_i}{(\sqrt{N}q_{i-1} - p_{i-1}) + \mu_i(\sqrt{N}q_i - p_i)} \\ &= \frac{1}{\frac{\sqrt{N}q_{i-1}-p_{i-1}}{p_i - \sqrt{N}q_i} + \mu_i \frac{\sqrt{N}q_i - p_i}{p_i - \sqrt{N}q_i}} \\ &= \frac{1}{a_i - \mu_i} \end{aligned}$$

Because of this, we see that in defining the numbers p_i, q_i, μ_i, we were merely doing the continued fraction algorithm. In other words, the numbers μ_i are the quotients in the simple continued fraction expansion of $\sqrt{N}$, i.e. $[\mu_0, \mu_1, \mu_2, \dots]$ is the simple continued fraction whose value is $\sqrt{N}$. Furthermore, the fractions $\frac{p_i}{q_i}$ are the convergents to $\sqrt{N}$.

Now we have $\mu_i < a_i$. In fact, we have chosen μ_i so that $a_i < \mu_i + 1$, so that $a_{i+1} > 1$ for all $i \geq 1$. It is also true that $a_1 > 1$ by definition. This shows that $p_{i-1} - \sqrt{N}q_{i-1}$ and $p_i - \sqrt{N}q_i$ have different signs, for each choice of i. This fact will be useful later.

In order to facilitate calculation of a_i, we multiply numerator and denominator of a_i by $p_i + \sqrt{N}q_i$. Recall from Theorem 4 that $p_i q_{i-1} - p_{i-1}q_i = (-1)^i$, so in the

numerators, we have the following:

$$\begin{aligned}(p_0 - \sqrt{N}q_0)(p_1 + \sqrt{N}q_1) &= p_0p_1 - Nq_0q_1 + \sqrt{N}(p_0q_1 - p_1q_0) \\ &= \sqrt{N} + \mu_0 \\ (p_{i-1} - \sqrt{N}q_{i-1})(p_i + \sqrt{N}q_i) &= p_{i-1}p_i - Nq_{i-1}q_i + \sqrt{N}(p_{i-1}q_i - p_iq_{i-1}) \\ &= p_ip_{i-1} - Nq_iq_{i-1} + (-1)^{i+1}\sqrt{N}, \quad \text{for } i \geq 2.\end{aligned}$$

In the denominators we have

$$\begin{aligned}(\sqrt{N}q_0 - p_0)(p_0 + \sqrt{N}q_0) &= Nq_0^2 - p_0^2 = -1 \\ (\sqrt{N}q_i - p_i)(p_i + \sqrt{N}q_i) &= Nq_i^2 - p_1^2, \quad \text{for } i \geq 2. \end{aligned} \tag{17}$$

Now we define

$$\begin{aligned}P_0 &= 1 & Q_0 &= 0 \\ P_1 &= \mu_0^2 - N & Q_1 &= \mu_0 \\ P_i &= p_i^2 - Nq_i^2 & Q_i &= p_ip_{i-1} - Nq_iq_{i-1}.\end{aligned} \tag{18}$$

So now we have

$$\begin{aligned}a_i &= \frac{Q_i + (-1)^{i-1}\sqrt{N}}{-P_i} \\ \mu_0 &< \frac{Q_0 - \sqrt{N}}{-P_0} < \mu_0 + 1 \\ \mu_i &< \frac{Q_i + (-1)^{i+1}\sqrt{N}}{-P_i} < \mu_i + 1.\end{aligned} \tag{19}$$

We can write Q_{i+1} in terms of P_i and Q_i in the following way:

$$\begin{aligned} Q_1 &= \mu_0 P_0 + Q_0, \quad \text{and} \\ Q_{i+1} &= p_i p_{i+1} - N q_i q_{i+1} \\ &= p_i(\mu_i p_i + p_{i-1}) - N q_i(\mu_i q_i + q_{i-1}) \\ &= \mu_i(p_i^2 - N q_i^2) + (p_i p_{i-1} - N q_i q_{i-1}) \\ &= \mu_i P_i + Q_i. \end{aligned} \tag{20}$$

Similarly,

$$\begin{aligned} P_{i+1} &= p_{i+1}^2 - N q_{i+1}^2 \\ &= (\mu_i p_i + p_{i-1})^2 - N(\mu_i q_i + q_{i-1})^2 \\ &= \mu_i^2 P_i - 2\mu_i Q_i + P_{i-1}. \end{aligned}$$

So we can calculate the sequences $P_0, P_1, P_2, \ldots$ and $Q_0, Q_1, \ldots$ independent of p_i and q_i. Futhermore,

$$\begin{aligned} Q_1^2 - P_1 &= (\mu_0^2 - (p_1^2 - N q_1^2)) = N, \quad \text{and} \\ Q_{i+1}^2 - P_i P_{i+1} &= (\mu_i P_i + Q_i)^2 - P_i(\mu_i^2 P_i + 2\mu_i Q_i + P_{i-1}) \\ &= Q_i^2 - P_{i-1} P_i. \end{aligned}$$

So by induction, we have $Q_{i+1}^2 - P_i P_{i+1} = N$ for $i \geq 1$. From the definition of P_i, we see that P_i and P_{i+1} have different signs for $i \geq 0$, beginning with $P_0 = 1$ and $P_1 = \mu_0^2 - N < 0$. Since the P_i are integers and $Q_{i+1}^2 - P_i P_{i+1} = N$, we see that $|P_i| \leq N$ and $|Q_i| \leq \sqrt{N}$ for $i \geq 0$.

Now since P_i and Q_i are bounded sequences of integers, some pair P_λ, Q_λ must be

repeated in the sequence, i.e. there is some positive integer ρ such that

$$P_\lambda = P_{\lambda+\rho}, \quad Q_\lambda = Q_{\lambda+\rho}.$$

From (19), we must also have $\mu_\lambda = \mu_{\lambda+\rho}$, and hence

$$Q_{\lambda+1} = \mu_\lambda P_\lambda + Q_\lambda = Q_{\lambda+\rho+1} \quad \text{and}$$
$$P_{\lambda+1} = \frac{N - Q_{\lambda+1}^2}{-P_\lambda} = P_{\lambda+\rho+1}.$$

Continuing in this manner, we see that $P_i = P_{i+\rho}$ and $Q_i = Q_{i+\rho}$ whenever $i \geq \lambda$. In other words, the sequences are periodic after λ.

Recall $p_i - \sqrt{N}q_i$ and $p_{i+1} - \sqrt{N}q_{i+1}$ have different signs. From (17) and (18) we see that the terms in the sequence $P_0, P_1, P_2, \ldots$ also have alternating signs. Now we have $p_i + \sqrt{N}q_i > 0$ for all i and

$$p_{i+1} + \sqrt{N}q_{i+1} = \mu_i(p_i + \sqrt{N}q_i) + p_{i-1} + \sqrt{N}q_{i-1}$$

for all $i \geq 1$. Solving for μ_i we see that

$$\mu_i = \frac{p_{i+1} + \sqrt{N}q_{i+1}}{p_i + \sqrt{N}q_i} - \frac{p_{i-1} + \sqrt{N}q_{i-1}}{p_i + \sqrt{N}q_i}.$$

Now $\mu_i > 0$ is an integer for all i. Also $\frac{p_{i-1}+\sqrt{N}q_{i-1}}{p_i+\sqrt{N}q_i} > 0$. Hence $\frac{p_{i+1}+\sqrt{N}q_{i+1}}{p_i+\sqrt{N}q_i} > 1$ for $i \geq 1$. Therefore $\frac{p_i+\sqrt{N}q_i}{p_{i+1}+\sqrt{N}q_{i+1}} < 1$ for $i \geq 1$, and $\frac{p_0+\sqrt{N}q_0}{p_1+\sqrt{N}q_1} = \frac{1}{\mu_0+\sqrt{N}} < 1$. So in particular, we have

$$\mu_i < \frac{p_{i+1} + \sqrt{N}q_{i+1}}{p_i + \sqrt{N}q_i} < \mu_i + 1, \tag{21}$$

for $i \geq 1$. Furthermore,

$$\frac{p_{i+1} + \sqrt{N}q_{i+1}}{p_i + \sqrt{N}q_i}\left(\frac{p_i - \sqrt{N}q_i}{p_i - \sqrt{N}q_i}\right) = \frac{p_{i+1}p_i - Nq_{i+1}q_i + \sqrt{N}(p_iq_{i+1} - p_{i+1}q_i)}{p_i^2 - Nq_i^2}$$
$$= \frac{Q_{i+1} + (-1)^{i+1}\sqrt{N}}{P_i},$$

thus transforming (21) into

$$\mu_i < \frac{Q_{i+1} + (-1)^{i-1}\sqrt{N}}{P_i} < \mu_i + 1. \tag{22}$$

Recall $P_i = P_{i+\rho}$, $Q_i = Q_{i+\rho}$, whenever $i \geq \lambda$. From our preceding calculations, we have $Q_\lambda^2 - P_{\lambda-1}P_\lambda = N$. Therefore $P_{\lambda-1} = P_{\lambda+\rho-1}$. So from (22) we have $\mu_{\lambda-1} = \mu_{\lambda+\rho-1}$. From these two facts and (20) we have $Q_{\lambda+1} - \mu_\lambda P_\lambda = Q_\lambda$. Therefore, $Q_{\lambda-1} = Q_{\lambda+\rho-1}$. Proceeding in this manner, we see that in fact $P_i = P_{i+\rho}$ for all $i \geq 0$. Now since $P_0 = 1$, we see that $P_i = 1$ for infinitely many choices of i. In fact, we have $P_\rho = 1$, which gives us $p_\rho^2 - Nq_\rho^2 = 1$, with $p_\rho \neq 1$ and $q_\rho \neq 0$. □

This completes the proof of the existence of a non-trivial solution of (1). Now we need to show that every solution is generated by a "least" solution. To see this first notice that if (x_1, y_1) and (x_2, y_2) are solutions of (1), with $0 < x_1 < x_2$, and $y_1 > 0$, $y_2 > 0$, then we have $x_2^2 - Ny_2^2 = x_1^2 - Ny_1^2$, and therefore $N(y_2^2 - y_1^2) = x_2^2 - x_1^2 > 0$, so that $y_2 > y_1$. So there exists a solution (t_1, u_1) with $t_1 > 0$,$u_1 > 0$, such that given any other solution (x, y), we have $t_1 < x$ and $u_1 < y$.

We can construct new solutions from known solutions in the following way: Let (x_1, y_1), (x_2, y_2) be two solutions. We can compose the two solutions to obtain

$$(x_1x_2 + Ny_1y_2, x_1y_2 + x_2y_1).$$

From Brahmagupta's identity, we can see that the result of this composition is also

a solution. Now we construct the solutions (t_i, u_i) recursively by the following:

$$\begin{aligned} t_2 &= t_1^2 + Nu_1^2 & u_2 &= 2t_1u_1 \\ t_{i+1} &= t_1t_i + Nu_1u_i & u_{i+1} &= t_1u_i + t_iu_1. \end{aligned}$$

Clearly $t_i < t_{i+1}$ and therefore $u_i < u_{i+1}$. We can also write

$$\begin{aligned} t_2 \pm u_2\sqrt{N} &= (t_1 \pm u_1\sqrt{N})^2 \quad \text{and} \\ t_i \pm u_i\sqrt{N} &= (t_1 \pm u_1\sqrt{N})^i. \end{aligned}$$

Because of the ambiguity of sign, we have

$$\begin{aligned} t_i &= \frac{(t_1 + u_1\sqrt{N})^i + (t_1 - u_1\sqrt{N})^i}{2} \\ u_i &= \frac{(t_1 + u_1\sqrt{N})^i - (t_1 - u_1\sqrt{N})^i}{2\sqrt{N}}. \end{aligned}$$

We now give Lagrange's proof that any integer solution to (1) is of the form $(\pm t_i, \pm u_i)$.

Proof. We first notice that (x, y) is a solution to 1 if and only if $(x, \pm y)$, and $(-x, \pm y)$ are all solutions as well. So it suffices to know those solutions with $x > 0$ and $y > 0$.

Now suppose that there were another solution (θ, ν), with $t_i < \theta < t_{i+1}$ for some i. Then we also have $u_i < \nu < \nu_{i+1}$. We can construct another solution (t, u) to (1) by $t = \theta t_{i+1} - N\nu u_{i+1}$ and $u = \theta u_{i+1} - t_{i+1}\nu$. The fact that (t, u) is a solution follows from Brahmagupta's identity. We can write

$$\begin{aligned} \theta - \nu\sqrt{N} &= \frac{1}{\theta - \nu\sqrt{N}} \\ t_{i+1} - \sqrt{N}u_{i+1} &= \frac{1}{t_{i+1} + u_{i+1}\sqrt{N}}. \end{aligned}$$

Substituting, we have

$$\begin{aligned} u &= u_{i+1}\left(\nu\sqrt{N} + \frac{1}{\theta + \nu\sqrt{N}}\right) - \nu\left(\sqrt{N}u_{i+1} + \frac{1}{t_{i+1} + u_{i+1}\sqrt{N}}\right) \\ &= \frac{u_{i+1}}{\theta + \nu\sqrt{N}} - \frac{\nu}{t_{i+1} + u_{i+1}\sqrt{N}}. \end{aligned}$$

Likewise we can express $t_i u_{i+1} - t_{i+1}u_i$ in the form

$$t_i u_{i+1} - t_{i+1}u_i = \frac{u_{i+1}}{t_i + u_i\sqrt{N}} - \frac{u_i}{t_{i+1} + u_{i+1}\sqrt{N}}.$$

And since $\theta > t_i$ and $\nu > u_i$, we also have

$$u = \frac{u_{i+1}}{\theta + \nu\sqrt{N}} - \frac{\nu}{t_{i+1} + u_{i+1}\sqrt{N}} < \frac{u_{i+1}}{t_i + u_i\sqrt{N}} - \frac{u_i}{t_{i+1} + u_{i+1}\sqrt{N}} = t_i u_{i+1} - t_{i+1}u_i.$$

But remember,

$$\begin{aligned} t_i &= \frac{(t_1 + u_1\sqrt{N})^i + (t_1 - u_1\sqrt{N})^i}{2} \\ u_i &= \frac{(t_1 + u_1\sqrt{N})^i - (t_1 - u_1\sqrt{N})^i}{2\sqrt{N}}, \end{aligned}$$

so

$$\begin{aligned} t_i u_{i+1} - t_{i+1}u_i &= \frac{(t_1 + u_1\sqrt{N})^{i+1}(t_1 - u_1\sqrt{N})^i + (t_1 - u_1\sqrt{N})^{i+1}(t_1 + u_1\sqrt{N})^i}{2\sqrt{N}} \\ &= \frac{(t_1 + u_1\sqrt{N}) - (t_1 - u_1\sqrt{N})}{2\sqrt{N}} \\ &= \frac{2u_1\sqrt{N}}{2\sqrt{N}} \\ &= u_1. \end{aligned}$$

The second equality follows from the fact that

$$(t_1 + u_1\sqrt{N})^i(t_1 - u_1\sqrt{N})^i = (t_1^2 - Nu_1^2)^i = 1.$$

So we get a solution (t, u) with $u < u_1$ contradicting the minimality of u_1. □

Remark. If we consider Lagrange's proof, that all solutions are generated by the least solution, we can use modern algebraic techniques to obtain the same result. That is to say, the solution (t, u) from which we desired to derive a contradiction is obtained by multiplying (θ, ν) and $(t_{i+1}, u_{i+1})^{-1} = (t_{i+1}, -u_{i+1})$.

Example. We give an example here of finding an integer solution to (1) using Lagrange's method, again with $N = 41$ (see the example following Section 2.2). A command is also given in Appendix A.2 for computing the smallest integer solution using Maple Software. Using (16) we can compute the convergents to the continued fraction whose value is $\sqrt{41}$.

$$\begin{array}{llll}
p_0 = 1, & q_0 = 0, & \mu_0 = 6 & \\
p_1 = 6, & q_1 = 1, & \mu_1 < \frac{1}{\sqrt{41}-6} \approx 2.48 & \text{so } \mu_1 = 2 \\
p_2 = 13, & q_2 = 2, & \mu_2 < \frac{6-\sqrt{41}}{2\sqrt{41}-13} \approx 2.08 & \text{so } \mu_2 = 2 \\
p_3 = 32, & q_3 = 5, & \mu_3 < \frac{13-2\sqrt{41}}{5\sqrt{41}-32} \approx 12.4 & \text{so } \mu_3 = 12 \\
p_4 = 397, & q_4 = 62, & \mu_4 < \frac{32-5\sqrt{41}}{62\sqrt{41}-397} \approx 2.48 & \text{so } \mu_4 = 2 \\
p_5 = 826, & q_5 = 129, & \mu_5 < \frac{397-62\sqrt{41}}{129\sqrt{41}-826} \approx 2.08 & \text{so } \mu_5 = 2 \\
p_6 = 2049, & q_6 = 320, & \mu_6 < \frac{826-129\sqrt{41}}{320\sqrt{41}-2049} \approx 2.48 & \text{so } \mu_6 = 12 \\
\text{etc...} & & &
\end{array}$$

Now we compute from (18)

$$\begin{aligned} P_0 &= 1, & Q_0 &= 0 \\ P_1 &= -5, & Q_1 &= 6 \\ P_2 &= 5, & Q_2 &= -4 \\ P_3 &= -1, & Q_3 &= 6 \\ P_4 &= 5, & Q_4 &= -6 \\ P_5 &= -5, & Q_5 &= 4 \\ P_6 &= 1, & Q_6 &= -6 \end{aligned}$$

Now we can stop there, because $P_6 = 1$. So we have the solution $(2049, 320)$ to (1).

4 Main Result

In this last section, we come to the main theme of this work. Before we can state our results, we need some background.

For an integer $m > 1$, the reduction modulo m map

$$\mathrm{red}_m : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$$

is the map that sends an integer a to its remainder r $(0 \leq r < m)$ on division by m. We shall denote the image of red_m by putting bar on the elements of $\mathbb{Z}$. For example, if $m = 5$, then $\mathrm{red}_5(16) = \bar{1}$, and $\mathrm{red}_5(-22) = \bar{3}$. The map red_m is a ring homomorphism. Furthermore, if p is prime, then the set of remainders upon division by p, $\mathbb{Z}/p\mathbb{Z}$ is a field. All that we need to check is that every element has an inverse. The rest of the axioms defining a field are easily verified. To see that every element in $\mathbb{Z}/p\mathbb{Z}$ has an inverse, notice that is $0 < a < p$, then $\gcd(a, p) = 1$. By the Euclidean algorithm, we can write $px + ay = 1$ for some integers x and y. Upon reduction

modulo p of this equation, we see that

$$\bar{a}\bar{y} = 1$$

and $\bar{y} \in \mathbb{Z}/p\mathbb{Z}$.

The map red_m induces a ring homomorphism on the ring of matrices also denoted by

$$\mathrm{red}_m : M_2(\mathbb{Z}) \to M_2(\mathbb{Z}/m\mathbb{Z}),$$

where $M_2(\mathbb{Z})$ and $M_2(\mathbb{Z}/m\mathbb{Z})$ are the rings of 2×2 matrices with entries in $\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z}$, resp. For example, it is not difficult to see that

$$\mathrm{red}_5 \begin{pmatrix} 6 & 5 \\ 23 & -2 \end{pmatrix} = \begin{pmatrix} \bar{1} & \bar{0} \\ \bar{3} & \bar{3} \end{pmatrix}.$$

Similarly, given the map $G \to \mathrm{SL}_2(\mathbb{Z})$, red_m also induces a group homomorphism

$$\mathrm{red}_m : G \to \mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z}).$$

The image of this map is clearly a finite group. Moreover, since G is cyclic, $\mathrm{red}_m(G)$ is a finite cyclic group. We denote by $g_N(m)$ the order of the image of G under the reduction mod m homomorphism. In the following discussion, since each element in the image $\mathrm{red}_m(G)$ is of the form

$$\begin{pmatrix} \bar{x} & \bar{y} \\ \bar{N}\bar{y} & \bar{x} \end{pmatrix},$$

we will denote these elements simply by $(\bar{x}, \bar{y})$. The group law defining G as a group is equivalent to the corresponding matrix multiplication, so when we multiply elements

of $\mathrm{red}_m(G)$, we will simply use (10).

The following theorems and the corollary sum up our main result.

Theorem 12. *Let p be a prime with $p \neq 2$.*
If $p|N$, then $g_N(p)|2p$. On the other hand suppose $p \nmid N$.

1. *If $\bar{N}$ is a square as an element of $\mathbb{F}_p$, then $g_N(p)|(p-1)$.*

2. *If $\bar{N}$ is not a square as an element of $\mathbb{F}_p$, then $g_N(p)|(p+1)$.*

Finally, $g_N(2) = 1$ or 2.

Theorem 13. *Let $m = p_1^{\epsilon_1} p_2^{\epsilon_2} \dots p_r^{\epsilon_r}$, where the p_i are distinct primes and ϵ_i positive integers. Then*

$$g_N(m) \mid \left(p_1^{\epsilon_1 - 1} g_N(p_1) p_2^{\epsilon_2 - 1} g_N(p_2) \dots p_r^{\epsilon_r - 1} g_N(p_r)\right).$$

Furthermore,

$$\mathrm{lcm}(g_N(p_1), g_N(p_2), \dots, g_N(p_r)) \mid g_N(m).$$

Corollary 14. *If m is as in the theorem, and $\left(\frac{N}{p_i}\right) = 1$ for all i, then $g_N(m) \mid \phi(N)$, where $\phi(N)$ is the Euler phi-function.*

Here we will give an example of how $g_N(m)$ can be computed.

Example. Consider $N = 7$, i.e. $x^2 - 7y^2 = 1$. The first integer solution to this equation, and therefore a generator of G, is $(8, 3)$. Now we want to consider $\mathrm{red}_5(G)$. We will use bar notation to indicate elements of $\mathbb{Z}/5\mathbb{Z}$. Upon reduction, the generator becomes $(\bar{3}, \bar{3})$. Now we know that $\mathrm{red}_5(G)$ is cyclic, so to find the order of the group, it suffices to multiply the generator until we get back to the identity, $(\bar{1}, \bar{0})$. If we multiply $(\bar{3}, \bar{3}) * (\bar{3}, \bar{3})$ using 10, we get $(\bar{72}, \bar{18})$. Now we notice $72 \equiv 2 \pmod 5$, and $18 \equiv 3 \pmod 5$. So we have the next solution $(\bar{2}, \bar{3})$. Continuing in this way, we

compute the following:

$$(\bar{3},\bar{3}) * (\bar{3},\bar{3}) = (\bar{2},\bar{3})$$

$$(\bar{3},\bar{3}) * (\bar{2},\bar{3}) = (\bar{4},\bar{0})$$

$$(\bar{3},\bar{3}) * (\bar{4},\bar{0}) = (\bar{2},\bar{2})$$

$$(\bar{3},\bar{3}) * (\bar{2},\bar{2}) = (\bar{3},\bar{2})$$

$$(\bar{3},\bar{3}) * (\bar{3},\bar{2}) = (\bar{1},\bar{0})$$

So we see that the order of $\mathrm{red}_5(G)$ is 6, i.e. $g_7(5) = 6$. From our theorem, we expected $g_7(5)|5+1$.

Without further conditions on N and m these results (Theorems 12 and 13) cannot be any stronger. Consider the following examples.

Example. Let $N = 11$. The generator for the group of solutions to (1) is $(10, 3)$. We can compute $g_N(m)$ for different values of m. For example,

$$g_N(3) = 1,\ g_N(5) = 4,\ g_N(7) = 3,\ g_N(13) = 7.$$

So we see that $g_N(3)$ is as small as possible, whereas $g_N(5) = \phi(5)$. Also,

$$g_N(5^3 \cdot 7 \cdot 13^2) = 5^2 \cdot 13 \cdot g_N(5) \cdot g_N(7) \cdot g_N(13) = 27300.$$

So the order of the group with $N = 11$ and $m = 147875$ has attained the upper bound prescribed by Theorem 13. On the other hand, let $N = 17$. The generator for G is $(33, 8)$, and

$$g_N(3) = 4,\ g_N(5) = 6,\ g_N(7) = 8,$$

so we have $g_N(p) = p + 1$ for these three primes. However, we have

$$g_N(3^2 \cdot 5 \cdot 7) = \text{lcm}(4, 6, 8) = 24.$$

In this example, the order of the group attains the lower bound.

4.1 Proof of Main Result

For the proof we need the following techincal lemma. **

Lemma 15. *If $(x, y) \in G$, then for any positive integer n, $(x, y)^n = (x_n, y_n)$, where*

$$x_n = \binom{n}{0}x^n + \binom{n}{2}Nx^{n-2}y^2 + \cdots + \binom{n}{2k}N^k x^{n-2k}y^{2k} + \cdots + \binom{n}{n}N^{\frac{n}{2}}y^n$$
$$y_n = \binom{n}{1}x^{n-1}y + \binom{n}{3}Nx^{n-3}y^3 + \cdots + \binom{n}{2k+1}N^k x^{n-2k-1}y^{2k+1} + \cdots + \binom{n}{n-1}N^{\frac{n-2}{2}}xy^{n-1}$$

if n is even, and

$$x_n = \binom{n}{0}x^n + \binom{n}{2}Nx^{n-2}y^2 + \cdots + \binom{n}{2k}N^k x^{n-2k}y^{2k} + \cdots + \binom{n}{n-1}N^{\frac{n-1}{2}}xy^{n-1}$$
$$y_n = \binom{n}{1}x^{n-1}y + \binom{n}{3}Nx^{n-3}y^3 + \cdots + \binom{n}{2k+1}N^k x^{n-2k-1}y^{2k+1} + \cdots + \binom{n}{n}N^{\frac{n-1}{2}}y^n$$

if n is odd.

Proof. We proceed by induction. The theorem is obviously true for $n = 1$. The

**This formula is also given by Lagrange, but with a different proof

inductive step is as follows. If n is even, then from (10)

$$\begin{aligned}
x_{n+1} &= xx_n + Nyy_n \\
&= x\left(\binom{n}{0}x^n + \binom{n}{2}Nx^{n-2}y^2 + \cdots + \binom{n}{2k}N^k x^{n-2k}y^{2k} + \cdots + \binom{n}{n}N^{\frac{n}{2}}y^n\right) \\
&\quad + Ny\left(\binom{n}{1}x^{n-1}y + \binom{n}{3}Nx^{n-3}y^3 + \cdots + \binom{n}{2k+1}N^k x^{n-2k-1}y^{2k+1} + \right. \\
&\quad \left. \cdots + \binom{n}{n-1}N^{\frac{n-2}{2}}xy^{n-1}\right) \\
&= x^{n+1} + \left(\binom{n}{1} + \binom{n}{2}\right)Nx^{n-1}y^2 + \cdots + \left(\binom{n}{2k-1} + \binom{n}{2k}\right)N^k x^{n-2k+1}y^{2k} + \\
&\quad \cdots + \left(\binom{n}{n-1} + \binom{n}{n}\right)N^{\frac{n}{2}}xy^n
\end{aligned}$$

and

$$\begin{aligned}
y_{n+1} &= x_n y + xy_n \\
&= \left(\binom{n}{0}x^n + \binom{n}{2}Nx^{n-2}y^2 + \cdots + \binom{n}{2k}N^k x^{n-2k}y^{2k} + \cdots + \binom{n}{n}N^{\frac{n}{2}}y^n\right)y \\
&\quad + x\left(\binom{n}{1}x^{n-1}y + \binom{n}{3}Nx^{n-3}y^3 + \cdots + \binom{n}{2k+1}N^k x^{n-2k-1}y^{2k+1} + \right. \\
&\quad \left. \cdots + \binom{n}{n-1}N^{\frac{n-2}{2}}xy^{n-1}\right) \\
&= \left(\binom{n}{0} + \binom{n}{1}\right)x^n y + \cdots + \left(\binom{n}{2k} + \binom{n}{2k+1}\right)N^k x^{n-2k}y^{2k+1} + \\
&\quad \cdots + \binom{n}{n}N^{\frac{n}{2}}y^{n+1}.
\end{aligned}$$

Now the binomial theorem tells us that $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$, so we have

$$\begin{aligned} x_{n+1} &= \binom{n+1}{0}x^{n+1} + \binom{n+1}{2}Nx^{n-1}y^2 + \cdots + \binom{n+1}{2k}N^k x^{n+1-2k}y^{2k} + \\ &\quad \cdots + \binom{n+1}{n}N^{\frac{n}{2}}xy^n \\ y_{n+1} &= \binom{n+1}{1}x^n y + \binom{n+1}{3}Nx^{n-2}y^3 + \cdots + \binom{n+1}{2k+1}N^k x^{(n+1)-2k-1}y^{2k+1} + \\ &\quad \cdots + \binom{n+1}{n+1}N^{\frac{n}{2}}y^{n+1} \end{aligned}$$

as desired. A similar calculation shows the case when n is odd. The only difference being that the pure y term shows up in x_{n+1} instead of y_{n+1}. □

Proof of Theorem 12. Now we want to consider the group $\mathrm{red}_p(G)$, where p is a prime, $p \neq 2$. If $x \in \mathbb{Z}$, then we denote by $\bar{x} \in \mathbb{Z}/p\mathbb{Z}$ the image of the canonical map $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$, namely reduction mod p. When p is a prime, $\mathbb{Z}/p\mathbb{Z}$ is the finite field $\mathbb{F}_p$. We will also denote by $(\bar{x}, \bar{y})$ the elements of $\mathrm{red}_p(G)$. Recall that for $\bar{x} \in \mathbb{F}_p$, $\bar{x}^p = \bar{x}$. This property of the Frobenius map will be important in the following discussion.

If $p|N$, then after reduction mod p, equation (1) becomes $\bar{x}^2 = 1$, so $\bar{x} = \pm 1$.

If $\bar{x} = 1$ and $\bar{y} = 0$, then $(\bar{x}, \bar{y}) = (1, 0)$, which is the identity of the group. If $\bar{x} = 1$ and $\bar{y} \neq 0$, then since p is odd,

$$(\bar{x}, \bar{y})^p = \left(\bar{x}^p + \cdots + \binom{p}{p-1}\bar{N}^{\frac{p-1}{2}}\bar{x}\bar{y}^{p-1}, \binom{p}{1}\bar{x}^{p-1}\bar{y} + \cdots + \bar{N}^{\frac{p-1}{2}}\bar{y}^p\right). \tag{23}$$

Because $p|\binom{p}{k}$ if $0 < k < p$, equation (23) shows that $(\bar{x}, \bar{y}) = (1, 0)$.

On the other hand, if $\bar{x} = -1$ and $\bar{y} = 0$, we have $(\bar{x}, \bar{y})^2 = (1, 0)$, whereas for $\bar{y} \neq 0$, by (23) again, $(\bar{x}, \bar{y})^p = (-1, 0)$. Therefore, $(\bar{x}, \bar{y})^{2p} = (1, 0)$.

Clearly the generator of $\mathrm{red}_p(G)$ falls into one of the categories listed above.

We now consider the case when p is not a factor of N. To fix notation, we recall some standard facts from basic number theory that can be found, for example, in

[1, p. 36]. Consider the map $\psi : \mathbb{F}_p^\times \to \mathbb{F}_p^\times$ defined by $\psi(x) = \bar{x}^{\frac{p-1}{2}}$, $\mathbb{F}_p^\times$ being the multiplicative group of non-zero elements of the finite field $\mathbb{F}_p$. The image of the map is $\{\pm 1\}$. Now if $\bar{x}$ is a square in $\mathbb{F}_p^\times$, then $\bar{x} = t^2$, and $\psi(\bar{x}) = (t^2)^{\frac{p-1}{2}} = 1$. So we have $(\mathbb{F}_p^\times)^2 \subset \ker\psi$. Also there are some elements $\bar{y} \in \mathbb{F}_p^\times$ such that $\psi(\bar{y}) = -1$ otherwise $\bar{y}^{\frac{p-1}{2}} - 1$ has more roots than its degree. Therefore $\ker\psi \subsetneq \mathbb{F}_p^\times$, and $[\mathbb{F}_p^\times : \ker\psi] \geq 2$. Now

$$2 = [\mathbb{F}_p^\times : (\mathbb{F}_p^\times)^2] = [\mathbb{F}_p^\times : \ker\psi][\ker\psi : (\mathbb{F}_p^\times)^2],$$

showing that $(\mathbb{F}_p^\times)^2 = \ker\psi$.

To summarize, if $\bar{x}$ is a square then $\bar{x}^{\frac{p-1}{2}} = 1$, and if $\bar{x}$ is not a square, then $\bar{x}^{\frac{p-1}{2}} = -1$.

Suppose $\bar{N}$ is a square. Then since p is odd, Lemma 15 gives us

$$\begin{aligned}(\bar{x},\bar{y})^p &= \left(\bar{x}^p + \cdots + \binom{p}{p-1}\bar{N}^{\frac{p-1}{2}}\bar{x}\bar{y}^{p-1}, \binom{p}{1}\bar{x}^{p-1}\bar{y} + \cdots + \bar{N}^{\frac{p-1}{2}}y^p\right)\\ &= (\bar{x}^p, \bar{N}^{\frac{p-1}{2}}\bar{y})\\ &= (\bar{x},\bar{y})\end{aligned}$$

The second equality follows because $p|\binom{p+1}{k}$ whenever k is not equal to $0, 1, p$ or $p+1$. This shows that $(\bar{x},\bar{y})^{p-1} = (1,0)$. Therefore $g_N(p)|(p-1)$.

Now suppose $\bar{N}$ is not a square. Then we have $\bar{N}^{\frac{p-1}{2}} = -1$, and so $\bar{N}^{\frac{p+1}{2}} = -\bar{N}$. Since $p+1$ is even Lemma 15 shows that

$$\begin{aligned}(x,y)^{p+1} &= \left(\bar{x}^{p+1} + \cdots + \bar{N}^{\frac{p+1}{2}}\bar{y}^{p+1}, \binom{p+1}{1}\bar{x}^p\bar{y} + \cdots + \binom{p+1}{p}\bar{N}^{\frac{p-1}{2}}\bar{x}\bar{y}^p\right)\\ &= \left(\bar{x}^2 - \bar{N}\bar{y}^2, \bar{x}\bar{y} - \bar{x}\bar{y}\right)\\ &= (1,0).\end{aligned}$$

Since this is true for any $(\bar{x},\bar{y}) \in \mathrm{red}_p(G)$, we have $g_N(p)|(p+1)$.

Now suppose $p = 2$. From (1), it is easy to see that $\mathrm{red}_2(G)$ is one of three possibilities. These possibilities are $\{(1,0)\}$, $\{(1,0),(0,1)\}$, and $\{(1,0),(1,1)\}$ (the second only if $2 \mid N$, and the third only if $2 \nmid N$). □

We also have the immediate corollary.

Corollary 16. *Let $p \neq 2$ be a prime number. If (x,y) is a solution to the Pell equation (1), then p does not divide x unless*

1. $\left(\frac{N}{p}\right) = 1$ *if* $p \equiv 1 \pmod 4$ *and*
2. $\left(\frac{N}{p}\right) = -1$ *if* $p \equiv 3 \pmod 4$.

Here $\left(\frac{N}{p}\right)$ denotes the Legendre symbol, i.e. $\left(\frac{N}{p}\right) = \bar{N}^{\frac{p-1}{2}} = \pm 1$ (as an element of $\mathbb{F}_p$).

Proof. If $p|N$, then $g_N(p)|2p$, but $4 \nmid 2p$.

On the other hand suppose $p \nmid N$. If (1) has a solution (x,y) with $p|x$, then $\mathrm{red}_p(G)$ has an element of the form $(0,\bar{y})$. But $(0,\bar{y})^2 = (Ny^2, 0) = (-1,0)$, since $\bar{x}^2 - \bar{N}\bar{y}^2 = 1$. So our previous calculations show that $|(0,\bar{y})| = 4$. But if $\mathrm{red}_p(G)$ contains an element of order 4, then $4|g_N(p)$, which by our theorem is only possible under the conditions stated. □

Remark. It is not difficult to see that elements of the form $(0,\bar{y})$ are the only elements of order four in $\mathrm{red}_p(G)$. Suppose $(\bar{x},\bar{y})$ has order four. Then

$$\begin{aligned}(\bar{x},\bar{y})^4 &= (\bar{x}^4 + 6\bar{N}\bar{x}^2\bar{y}^2 + \bar{N}^2\bar{y}^4, 4\bar{x}\bar{y}(\bar{x}^2 + \bar{N}\bar{y}^2)) \\ &= (1,0)\end{aligned}$$

Since we are working in a field, the second coordinate is zero only if one of the factors is zero. If $\bar{x} \neq 0$, then either $\bar{y} = 0$ or $(\bar{x}^2 + \bar{N}\bar{y}^2) = 0$. If the first, then $\bar{x}^2 = 1$, and so

$(\bar{x}, \bar{y})$ has order at most two. If the latter, then since $(\bar{x}, \bar{y})^2 = (\bar{x}^2 + \bar{N}\bar{y}^2, 2\bar{x}\bar{y})$, our proof of the corollary shows that the order of $(\bar{x}, \bar{y})^2$ is four, so $(\bar{x}, \bar{y})$ has order 8.

Theorem 12 provides a complete classification of the groups $\mathrm{red}_p(G)$, where p is a prime. However, if we consider $\mathrm{red}_m(G)$, where $m > 1$ is composite, we also get a finite cyclic group. The order of $\mathrm{red}_m(G)$ is given by Theorem 13. The proof of Theorem 13 will follow from the following two propositions.

Proposition 17. *Let $m = p^k$, where k is a positive integer. Then $g_N(m) | p^{k-1} g_N(p)$.*

Remark. Notice the similarity of the function $g_N(m)$ for a fixed N to the Euler ϕ-function $\phi(m)$:

$$\phi(p^k) = p^{k-1}\phi(p).$$

Proof of Proposition. If $k = 1$, the statement is trivial. So we may assume $k \geq 2$. Consider the map $\varphi : \mathrm{red}_{p^k}(G) \to \mathrm{red}_{p^{k-1}}(G)$ induced by $\mathrm{red}_{p^{k-1}}$. In other words, reduce the entries of elements of $\mathrm{red}_{p^k}(G)$, by p^{k-1}. This is a surjective map, and so we have

$$\mathrm{red}_{p^k}(G)/\ker\varphi \cong \mathrm{red}_{p^{k-1}}(G),$$

or in other words, $g_N(p^k) = g_N(p^{k-1}) \cdot |\ker\varphi|$. We will use bar notation to denote elements of $\mathbb{Z}_{p^k}$, i.e. if $x \in \mathbb{Z}$, then $\bar{x}$ is its image in $\mathbb{Z}_{p^k}$.

We need to know the order of $\ker\varphi$, so let $(\bar{x}, \bar{y}) \in \ker\varphi$. Then $(\bar{x}, \bar{y})$ has the form $(1 + \bar{n}_1\bar{p}^{k-1}, \bar{n}_2\bar{p}^{k-1})$, where n_1 and n_2 are integers. But we know that

$$\bar{x}^2 - \bar{N}\bar{y}^2 = 1 + 2\bar{n}_1\bar{p}^{k-1} + \bar{n}_1^2\bar{p}^{2k-1} - \bar{N}\bar{n}_2^2\bar{p}^{2k-1} = 1. \tag{24}$$

Since $k \geq 2$, $2(k-1) \geq k$, and so $p^k | p^{2k-1}$. From (24) we have $\bar{n}_1 = 0$. Therefore $(\bar{x}, \bar{y})$ has the form $(1, \bar{n}_2\bar{p}^{k-1})$.

Now for $p \neq 2$ we have

$$(\bar{x}, \bar{y})^p = \left(1 + \binom{p}{2}\bar{N}(\bar{n}_2\bar{p}^{k-1})^2 \cdots + \binom{p}{p-1}\bar{N}^{\frac{p-1}{2}}(\bar{n}_2\bar{p}^{k-1})^{p-1}\,,\right.$$
$$\left.\binom{p}{1}\bar{n}_2\bar{p}^{k-1} + \cdots + \bar{N}^{\frac{p-1}{2}}(\bar{n}_2\bar{p}^{k-1})^p\right)$$
$$= (1, 0).$$

The second equality holds because a power of p^k shows up in every term excepting the first term of the first entry.

If $p = 2$, then we have

$$(\bar{x}, \bar{y})^2 = \big(1 + \bar{N}(\bar{n}_2(2^{k-1}))^2, 2\bar{n}_2(2^{k-1})\big)$$
$$= (1, 0).$$

Thus we see that $|\ker\varphi|$ divides p. By induction then, $g_N(p^k)|p^{k-1}g_N(p)$. □

Proposition 18. *Suppose $m = qr$, with $\gcd(q, r) = 1$. Then $g_N(m)|\big(g_N(q) \cdot g_N(r)\big)$.*

Proof. The Chinese remainder theorem gives an isomorphism

$$\mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/r\mathbb{Z}$$

where the map in the first (resp. second) is just reduction mod q (resp. r). Let φ_q denote the map into the first coordinate, and φ_r the map into the second coordinate, and consider the map

$$\psi : \mathrm{red}_m(G) \to \mathrm{red}_q(G) \times \mathrm{red}_r(G)$$

given by $(\bar{x}, \bar{y}) \mapsto \Big(\big(\varphi_q(\bar{x}), \varphi_q(\bar{y})\big), \big(\varphi_r(\bar{x}), \varphi_r(\bar{y})\big)\Big)$. Now $(\bar{x}, \bar{y}) \in \ker\psi$ iff $\varphi_q(\bar{x}) = \varphi_r(\bar{x}) = 1$ and $\varphi_q(\bar{y}) = \varphi_r(\bar{y}) = 0$. But since the Chinese Remainder Theorem gave

us an isomorphism of rings, this happens exactly when $\bar{x} = 1$, and $\bar{y} = 0$. Therefore ψ is injective, and we see that $g_N(m) | \big(g_N(q) \cdot g_N(r)\big)$. □

Proof of Theorem 13. The proof follows from the previous two propositions. □

Corollary 19. *If m is as in the theorem, and $\left(\frac{N}{p_i}\right) = 1$ for all i, then $g_N(m) \mid \phi(m)$, where $\phi(m)$ is the Euler phi-function.*

Remark. The previous corollary does not hold generally if $\left(\frac{N}{p_i}\right) = -1$ for any p_i. Consider the following example:

Example. Suppose $N = 13$. The generator for G with this choice of N is $(649, 180)$. It is not difficult to check that $13 \equiv 6 \pmod 7$ is not square and $13 \pmod{19}$ are not square in their respective fields. We have $g_{13}(7) = 8$, and $g_{13}(19) = 20$, and $g_{13}(133) = 40$. However, $\phi(133) = 6 \cdot 18 = 108$.

Appendices

A Tables and Maple Code

We have seen how difficult the generator for G can be to compute. Calculating $g_N(m)$ can also require a great deal of work. It is best done using a computer, especially for large values of N. In this appendix, I will give some tables displaying $g_N(m)$, and the Maple code that I have written to compute $g_N(m)$, as well. The code and tables are provided merely to assist in the computation of $g_N(m)$.

A.1 Tables

I have included ten tables listing values of $g_N(m)$. Table 1 shows $g_N(p)$ for the first 25 primes. There, I have taken the first square-free integers less than or equal to 51 for N. Table 2 shows $g_N(p^k)$ with $1 \leq k \leq 3$ for the first several primes. Again, I have taken the first sqare-free integers less than or equal to 51. Tables 3-10 show the values of $g_N(m)$ for the integers $2 \leq m \leq 100$. In these last tables, we have taken all of the square-free integers N with $2 \leq N \leq 9$.

$g_N(p)$	$p=2$	3	5	7	11	13	17	19	23	29	31	37	41	43	47	53	59	61	67	71	73	79	83	89	97
$N=2$	1	4	6	3	12	14	8	20	11	10	15	38	5	44	23	54	20	62	68	35	36	13	84	44	48
3	2	6	3	8	10	12	18	5	11	15	32	36	14	11	23	9	58	60	34	7	36	80	82	90	16
5	1	4	10	8	5	14	6	3	8	7	5	38	20	44	16	18	29	10	68	35	74	13	28	22	98
6	1	6	4	8	3	7	18	18	11	28	32	19	42	42	23	52	30	31	66	35	36	80	7	10	12
7	2	2	6	7	12	14	3	18	12	28	15	36	21	44	23	52	58	62	68	18	37	40	82	45	49
10	1	1	10	8	12	3	18	4	24	30	15	6	20	21	48	13	20	62	11	35	74	39	41	44	98
11	2	1	4	3	22	7	18	6	24	15	32	36	42	42	48	52	30	31	34	24	74	39	82	44	24
13	1	1	2	8	4	26	8	20	11	14	32	38	14	7	16	13	4	5	68	24	74	13	28	30	98
14	1	4	4	7	10	12	9	20	12	6	3	38	7	42	23	54	60	20	66	18	37	10	28	45	7
15	2	3	10	6	10	7	16	20	24	30	8	19	21	7	48	13	58	12	33	5	74	40	42	45	98
17	1	4	6	8	4	6	34	9	24	30	32	38	42	21	23	13	29	62	3	72	74	16	41	44	98
19	2	2	4	8	3	1	4	38	8	15	3	19	42	22	48	27	58	60	66	35	9	39	42	90	98
21	1	1	4	14	4	14	16	10	8	5	16	3	40	14	23	9	29	62	22	24	74	26	41	88	98
22	1	2	3	1	22	12	18	5	24	28	32	19	42	22	16	27	58	60	66	24	74	3	21	44	48
23	2	4	2	3	10	12	9	18	23	28	16	38	10	42	6	54	12	62	22	36	9	13	82	9	49
26	1	4	1	8	5	26	4	9	11	30	32	18	42	44	48	18	29	62	11	72	74	39	41	90	98
29	1	4	1	1	4	2	6	20	11	58	32	38	14	44	16	26	29	62	11	35	74	80	41	30	98
30	1	6	5	6	4	12	16	9	3	14	32	12	42	11	24	27	60	62	17	35	37	80	82	30	49
31	2	2	4	1	10	2	9	4	11	10	31	38	5	42	24	54	60	62	68	36	37	39	82	45	48
33	1	6	6	8	11	7	16	20	4	28	30	9	40	44	3	54	15	31	66	9	37	80	41	30	48
34	1	2	4	4	10	14	17	20	6	28	8	36	21	44	23	54	12	20	68	3	37	8	12	44	49
35	2	4	5	14	6	3	16	18	22	28	5	38	42	21	12	54	58	31	33	8	72	80	84	18	32
37	1	1	6	3	5	14	18	20	24	6	32	74	20	44	23	26	60	62	33	7	4	80	41	30	98
38	1	1	3	8	10	12	8	38	8	28	5	4	42	42	48	52	30	31	34	35	6	39	82	90	98
39	1	3	4	6	12	26	3	18	11	15	10	38	40	22	16	27	60	30	22	72	74	10	84	88	98
41	1	4	2	8	12	14	6	20	11	30	15	18	82	21	48	54	29	15	68	24	36	80	41	90	98
42	1	3	6	14	10	4	16	9	24	14	16	19	40	44	23	13	10	60	68	72	74	78	14	88	98
43	2	2	3	3	6	12	8	18	8	15	32	19	10	86	48	52	1	31	17	35	74	80	21	30	24
46	1	2	4	3	12	2	3	20	1	30	4	12	20	44	24	52	58	12	68	36	12	13	28	45	49

Table 1: $g_N(p)$ for the first 25 primes

$g_N(m)$	$m=2$	4	8	3	9	27	5	25	125	7	49	343	11	121	1331	13	169	2197
$N=2$	1	2	4	4	12	36	6	30	150	3	21	147	12	132	1452	14	14	182
3	2	4	4	6	18	54	3	15	75	8	56	392	10	110	1210	12	156	2028
5	1	1	2	4	4	12	10	50	250	8	56	392	5	55	605	14	182	2366
6	1	2	4	6	6	18	4	20	100	8	8	56	3	33	363	7	91	1183
7	2	4	4	2	6	18	6	30	150	7	49	343	12	132	1452	14	182	2366
10	1	2	4	1	3	9	10	50	250	8	56	392	12	132	1452	3	39	507
11	2	4	4	1	3	9	4	20	100	3	21	147	22	242	2662	7	91	1183
13	1	1	2	1	1	3	2	10	50	8	56	392	4	44	484	26	338	4394
14	1	2	2	4	12	36	4	20	100	7	49	343	10	110	1210	12	156	2028
15	2	4	4	3	3	9	10	50	250	6	42	294	10	110	1210	7	91	1183
17	1	1	1	4	12	36	6	30	150	8	56	392	4	44	484	6	78	1014
19	2	4	4	2	6	18	4	20	100	8	56	392	3	33	363	1	13	169
21	1	2	2	1	3	9	4	20	100	14	98	686	4	44	484	14	182	2366
22	1	2	4	2	6	18	3	15	75	1	7	49	22	242	2662	12	156	2028
23	2	4	4	4	12	36	2	10	50	3	3	21	10	110	1210	12	156	2028
26	1	2	4	4	12	36	1	5	25	8	56	392	5	55	605	26	338	4394
29	1	1	2	4	4	4	1	5	25	1	7	49	4	4	44	2	26	338
30	1	2	4	6	18	54	5	25	125	6	42	294	4	44	484	12	156	2028
31	2	4	4	2	6	18	4	20	100	1	7	49	10	110	1210	2	26	338
33	1	2	2	6	6	18	6	30	150	8	56	392	11	121	1331	7	91	1183
34	1	2	4	2	6	18	4	20	100	4	28	196	10	110	1210	14	182	2366
35	2	4	4	4	12	36	5	25	125	14	98	686	6	66	726	3	39	507
37	1	1	2	1	3	9	6	30	150	3	3	21	5	55	605	14	182	2366
38	1	2	4	1	3	9	3	3	15	8	56	392	10	110	1210	12	156	2028
39	1	1	2	3	9	27	4	4	20	6	6	42	12	132	1452	26	338	4394
41	1	1	1	4	12	36	2	10	50	8	56	392	12	132	1452	14	182	2366
42	1	2	4	3	3	3	6	6	30	14	98	686	10	110	1210	4	52	676
43	2	4	4	2	2	6	3	15	75	3	21	147	6	66	726	12	156	2028
46	1	2	2	2	6	18	4	20	100	3	21	147	12	132	1452	2	26	338
47	2	4	4	4	12	36	6	30	150	2	14	98	10	110	1210	14	182	2366
51	2	4	4	6	6	18	4	4	20	1	7	49	6	66	726	12	156	2028

Table 2: $g_N(p)$ for the first several prime powers

$g_N(m)$	$m=2$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$N=2$	1	4	2	6	4	3	4	12	6	12	4	14	3	12	8	8	12	20	6	12	12	11	4	30
3	2	6	4	3	6	8	4	18	6	10	12	12	8	6	8	18	18	5	12	24	10	11	12	15
5	1	4	1	10	4	8	2	4	10	5	4	14	8	20	4	6	4	3	10	8	5	8	4	50
6	1	6	2	4	6	8	4	6	4	3	6	7	8	12	8	18	6	18	4	24	3	11	12	20
7	2	2	4	6	2	7	4	6	6	12	4	14	14	6	4	3	6	18	12	14	12	12	4	30
10	1	1	2	10	1	8	4	3	10	12	2	3	8	10	8	18	3	4	10	8	12	24	4	50
11	2	1	4	4	2	3	4	3	4	22	4	7	6	4	8	18	6	6	4	3	22	24	4	20
13	1	1	1	2	1	8	2	1	2	4	1	26	8	2	4	8	1	20	2	8	4	11	2	10
14	1	4	2	4	4	7	2	12	4	10	4	12	7	4	4	9	12	20	4	28	10	12	4	20
15	2	3	4	10	6	6	4	3	10	10	12	7	6	30	4	16	6	20	20	6	10	24	12	50
17	1	4	1	6	4	8	1	12	6	4	4	6	8	12	2	34	12	9	6	8	4	24	4	30
19	2	2	4	4	2	8	4	6	4	3	4	1	8	4	8	4	6	38	4	8	6	8	4	20
21	1	1	2	4	1	14	2	3	4	4	2	14	14	4	4	16	3	10	4	14	4	8	2	20
22	1	2	2	3	2	1	4	6	3	22	2	12	1	6	8	18	6	5	6	2	22	24	4	15
23	2	4	4	2	4	3	4	12	2	10	4	12	6	4	4	9	12	18	4	12	10	23	4	10
26	1	4	2	1	4	8	4	12	1	5	4	26	8	4	8	4	12	9	2	8	5	11	4	5
29	1	4	1	1	4	1	2	4	1	4	4	2	1	4	4	6	4	20	1	4	4	11	4	5
30	1	6	2	5	6	6	4	18	5	4	6	12	6	30	8	16	18	9	10	6	4	3	12	25
31	2	2	4	4	2	1	4	6	4	10	4	2	2	4	4	9	6	4	4	2	10	11	4	20
33	1	6	2	6	6	8	2	6	6	11	6	7	8	6	4	16	6	20	6	24	11	4	6	30
34	1	2	2	4	2	4	4	6	4	10	2	14	4	4	8	17	6	20	4	4	10	6	4	20
35	2	4	4	5	4	14	4	12	10	6	4	3	14	20	8	16	12	18	20	28	6	22	4	25
37	1	1	1	6	1	3	2	3	6	5	1	14	3	6	4	18	3	20	6	3	5	24	2	30
38	1	1	2	3	1	8	4	3	3	10	2	12	8	3	8	8	3	38	6	8	10	8	4	3
39	1	3	1	4	3	6	2	9	4	12	3	26	6	12	4	3	9	18	4	6	12	11	6	4
41	1	4	1	2	4	8	1	12	2	12	4	14	8	4	1	6	12	20	2	8	12	11	4	10
42	1	3	2	6	3	14	4	3	6	10	6	4	14	6	8	16	3	9	6	42	10	24	12	6
43	2	2	4	3	2	3	4	2	6	6	4	12	6	6	8	8	2	18	12	6	6	8	4	15
46	1	2	2	4	2	3	2	6	4	12	2	2	3	4	4	3	6	20	4	6	12	1	2	20
47	2	4	4	6	4	2	4	12	6	10	4	14	2	12	4	8	12	6	12	4	10	11	4	30
51	2	6	4	4	6	1	4	6	4	6	12	12	2	12	8	34	6	10	4	6	6	24	12	4

Table 3: $2 \leq m \leq 25$, and $N \leq 51$

$g_N(m)$	$m=2$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$N=53$	1	4	1	2	4	1	2	4	2	5	4	1	1	4	4	4	4	20	2	4	5	8	4	2
55	1	2	1	10	2	8	2	6	10	11	2	12	8	10	4	16	6	9	10	8	11	22	2	50
57	1	3	2	1	3	6	2	9	1	12	6	14	6	3	4	9	9	38	2	6	12	24	6	5
58	1	1	2	6	1	3	4	1	6	1	2	2	3	6	8	18	1	9	6	3	1	11	4	30
59	2	2	4	4	2	8	4	6	4	10	4	7	8	4	8	8	6	10	4	8	10	1	4	20
61	1	1	1	2	1	8	2	1	2	4	1	1	8	2	4	6	1	3	2	8	4	8	2	10
62	1	4	2	6	4	4	2	4	6	12	4	12	4	12	2	9	4	18	6	4	12	11	4	6
65	1	4	1	10	4	3	1	12	10	12	4	26	3	20	1	18	12	20	10	12	12	8	4	50
66	1	6	1	4	6	8	1	18	4	22	6	4	8	12	2	8	18	18	4	24	22	24	6	20
67	2	2	4	3	2	3	4	2	6	10	4	1	6	6	8	1	2	10	12	6	10	24	4	15
69	1	2	2	4	2	8	2	2	4	5	2	1	8	4	2	16	2	20	4	8	5	23	2	4
70	1	2	2	1	2	14	4	6	1	5	2	7	14	2	8	16	6	20	2	14	5	22	4	5
71	2	4	4	4	4	1	4	12	4	10	4	14	2	4	4	9	12	20	4	4	10	11	4	20
73	1	1	1	2	1	8	1	3	2	12	1	14	8	2	2	18	3	9	2	8	12	11	1	2
74	1	4	2	2	4	3	4	4	2	12	4	6	3	4	8	18	4	9	2	12	12	24	4	10
77	1	4	2	1	4	7	2	4	1	22	4	4	7	4	2	16	4	3	2	28	22	22	4	5
78	1	2	2	6	2	6	4	6	6	5	2	13	6	6	8	18	6	20	6	6	5	11	4	30
79	2	2	4	4	2	3	4	2	4	12	4	12	6	4	4	9	2	20	4	6	12	3	4	20
82	1	1	2	6	1	8	4	1	6	5	2	6	8	6	8	18	1	9	6	8	5	11	4	6
83	2	1	4	3	2	8	4	1	6	3	4	7	8	3	8	8	2	18	12	8	6	24	4	15
85	1	1	1	10	1	1	2	1	10	4	1	14	1	10	4	34	1	3	10	1	4	11	2	50
86	1	1	2	4	1	3	4	3	4	2	2	7	3	4	8	1	3	10	4	3	2	8	4	20
87	2	1	4	6	2	4	4	3	6	6	4	12	4	6	4	16	6	3	12	4	6	11	4	30
89	1	4	1	1	4	8	1	12	1	5	4	14	8	4	2	8	12	20	1	8	5	24	4	1
91	2	2	4	2	2	14	4	6	2	1	4	13	14	2	8	18	6	20	4	14	2	24	4	10
93	1	1	2	1	1	2	2	1	1	5	2	14	2	1	4	16	1	6	2	2	5	11	2	5
94	1	2	2	4	2	4	2	6	4	4	2	12	4	4	2	8	6	4	4	4	4	11	2	20
95	1	4	2	10	4	6	2	12	10	6	4	4	6	20	4	18	12	19	10	12	6	22	4	50
97	1	1	1	6	1	8	1	3	6	5	1	14	8	6	2	18	3	20	6	8	5	8	1	30

Table 4: $2 \leq m \leq 25$, and $51 < N < 100$

$g_N(m)$	$m=26$	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
$N=2$	14	36	6	10	12	15	16	12	8	6	12	38	20	28	12	5	12	44	12	12	11	23	8	21	30
3	12	54	8	15	6	32	16	30	18	24	36	36	10	12	12	14	24	11	20	18	22	23	24	56	30
5	14	12	8	7	20	5	8	20	6	40	4	38	3	28	10	20	8	44	5	20	8	16	4	56	50
6	7	18	8	28	12	32	16	6	18	8	6	19	18	42	4	42	24	42	6	12	11	23	24	8	20
7	14	18	28	28	6	15	4	12	6	42	12	36	18	14	12	21	14	44	12	6	12	23	4	49	30
10	3	9	8	30	10	15	16	12	18	40	6	6	4	3	20	20	8	21	12	30	24	48	8	56	50
11	14	9	12	15	4	32	16	22	18	12	12	36	6	7	4	42	6	42	44	12	24	48	8	21	20
13	26	3	8	14	2	32	8	4	8	8	1	38	20	26	2	14	8	7	4	2	11	16	4	56	10
14	12	36	14	6	4	3	8	20	9	28	12	38	20	12	4	7	28	42	10	12	12	23	4	49	20
15	14	9	12	30	30	8	8	30	16	30	12	19	20	21	20	21	6	7	20	30	24	48	12	42	50
17	6	36	8	30	12	32	4	4	34	24	12	38	9	12	6	42	8	21	4	12	24	23	4	56	30
19	2	18	8	15	4	3	16	6	4	8	12	19	38	2	4	42	8	22	12	12	8	48	8	56	20
21	14	9	14	5	4	16	8	4	16	28	6	3	10	14	4	40	14	14	4	12	8	23	4	98	20
22	12	18	2	28	6	32	16	22	18	3	6	19	5	12	12	42	2	22	22	6	24	16	8	7	15
23	12	36	12	28	4	16	4	20	18	6	12	38	18	12	4	10	12	42	20	12	46	6	4	3	10
26	26	36	8	30	4	32	16	20	4	8	12	18	9	52	4	42	8	44	10	12	11	48	8	56	5
29	2	4	1	58	4	32	8	4	6	1	4	38	20	4	2	14	4	44	4	4	11	16	4	7	5
30	12	54	6	14	30	32	16	12	16	30	18	12	9	12	20	42	6	11	4	90	3	24	24	42	25
31	2	18	4	10	4	31	4	10	18	4	12	38	4	2	4	5	2	42	20	12	22	24	4	7	20
33	7	18	8	28	6	30	8	66	16	24	6	9	20	42	6	40	24	44	22	6	4	3	12	56	30
34	14	18	4	28	4	8	16	10	17	4	6	36	20	14	4	21	4	44	10	12	6	23	8	28	20
35	6	36	28	28	20	5	16	12	16	70	12	38	18	12	20	42	28	21	12	60	22	12	8	98	50
37	14	9	3	6	6	32	8	5	18	6	3	74	20	14	6	20	3	44	5	6	24	23	4	3	30
38	12	9	8	28	3	5	16	10	8	24	6	4	38	12	12	42	8	42	10	3	8	48	8	56	3
39	26	27	6	15	12	10	8	12	3	12	9	38	18	78	4	40	6	22	12	36	11	16	12	6	4
41	14	36	8	30	4	15	1	12	6	8	12	18	20	28	2	82	8	21	12	12	11	48	4	56	10
42	4	3	14	14	6	16	16	30	16	42	6	19	9	12	12	40	42	44	10	6	24	23	24	98	6
43	12	6	12	15	6	32	16	6	8	3	4	19	18	12	12	10	6	86	12	6	8	48	8	21	30
46	2	18	6	30	4	4	8	12	3	12	6	12	20	2	4	20	6	44	12	12	1	24	4	21	20
47	14	36	4	30	12	15	4	20	8	6	12	12	6	28	12	7	4	42	20	12	22	47	4	14	30
51	12	18	4	28	12	15	16	6	34	4	12	19	10	12	4	20	6	11	12	12	24	23	24	7	4

Table 5: $26 \leq m \leq 50$, and $N \leq 51$

$g_N(m)$	$m=26$	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
$N=53$	1	12	1	14	4	32	8	20	4	2	4	6	20	4	2	14	4	7	5	4	8	23	4	7	2
55	12	18	8	15	10	8	8	22	16	40	6	38	9	12	10	7	8	44	11	30	22	46	4	56	50
57	14	27	6	28	3	8	8	12	9	6	18	38	38	42	2	40	6	6	12	9	24	48	12	42	5
58	2	3	6	58	6	32	16	1	18	6	2	9	9	2	12	42	3	21	2	6	11	48	8	21	30
59	14	18	8	28	4	15	16	10	8	8	12	19	10	14	4	10	8	14	20	12	2	23	8	56	20
61	1	3	8	10	2	32	8	4	6	8	1	38	3	1	2	5	8	44	4	2	8	23	4	56	10
62	12	12	4	28	12	31	4	12	9	12	4	12	18	12	6	20	4	44	12	12	11	24	4	28	6
65	26	36	3	14	20	32	2	12	18	30	12	3	20	52	10	42	12	4	12	60	8	23	4	21	50
66	4	54	8	15	12	15	4	66	8	8	18	19	18	12	4	10	24	6	22	36	24	48	6	56	20
67	2	2	12	28	6	15	16	10	2	3	4	36	10	2	12	42	6	42	20	6	24	16	8	21	30
69	1	6	8	10	4	10	4	10	16	8	2	19	20	2	4	14	8	44	10	4	23	8	2	56	4
70	7	18	14	30	2	15	16	10	16	14	6	36	20	14	4	42	14	22	10	6	22	24	8	98	5
71	14	36	4	4	4	15	4	20	18	4	12	36	20	28	4	21	4	44	20	12	22	23	4	7	20
73	14	9	8	30	2	32	4	12	18	8	3	18	9	14	2	20	8	44	12	6	11	48	2	8	2
74	6	4	6	7	4	32	16	12	18	6	4	74	9	12	4	20	12	1	12	4	24	23	8	3	10
77	4	4	14	10	4	32	4	44	16	7	4	3	3	4	2	40	28	11	22	4	22	16	4	49	5
78	13	18	6	28	6	10	16	10	18	6	6	18	20	26	12	8	6	42	10	6	11	12	8	42	30
79	12	6	12	30	4	16	4	12	18	12	4	38	20	12	4	7	6	42	12	4	6	23	4	21	20
82	6	3	8	7	6	15	16	5	18	24	2	38	9	6	12	82	8	44	10	6	11	16	8	56	6
83	14	3	8	28	6	32	16	3	8	24	4	36	18	7	12	4	8	14	12	3	24	23	8	56	30
85	14	1	1	10	10	32	8	4	34	10	1	3	3	14	10	2	1	44	4	10	11	16	4	7	50
86	7	9	6	28	4	32	16	2	1	12	6	36	10	7	4	20	3	86	2	12	8	48	8	21	20
87	12	9	4	58	6	30	8	6	16	12	12	19	6	12	12	40	4	21	12	6	22	48	4	28	30
89	14	36	8	10	4	32	4	20	8	8	12	38	20	28	1	42	8	44	5	12	24	23	4	8	1
91	26	18	28	28	2	16	16	2	18	14	12	38	20	26	4	40	14	11	4	6	24	3	8	98	10
93	14	3	2	4	1	62	8	5	16	2	2	38	6	14	2	7	2	22	10	1	11	16	4	14	5
94	12	18	4	28	4	15	4	4	8	4	6	38	4	12	4	3	4	44	4	12	11	47	2	28	20
95	4	36	6	15	20	15	8	12	18	30	12	36	19	4	10	7	12	14	6	60	22	46	4	42	50
97	14	9	8	10	6	5	4	5	18	24	3	38	20	14	6	42	8	21	5	6	8	23	2	56	30

Table 6: $26 \leq m \leq 50$, and $51 < N < 100$

$g_N(m)$	$m=51$	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
$N=2$	8	14	54	36	12	12	20	10	20	12	62	15	12	32	42	12	68	8	44	6	35	12	36	38	60
3	18	12	9	54	30	8	30	30	58	12	60	32	72	32	12	30	34	36	66	24	7	36	36	36	30
5	12	14	18	12	10	8	12	7	29	20	10	5	8	16	70	20	68	6	8	40	35	4	74	38	100
6	18	14	52	18	12	8	18	28	30	12	31	32	24	32	28	6	66	18	66	8	35	12	36	19	60
7	6	28	52	18	12	28	18	28	58	12	62	30	42	8	42	12	68	12	12	42	18	12	37	36	30
10	18	6	13	9	60	8	4	30	20	10	62	15	24	32	30	12	11	18	24	40	35	12	74	6	50
11	18	28	52	18	44	12	6	30	30	4	31	32	3	32	28	22	34	36	24	12	24	12	74	36	20
13	8	26	13	3	4	8	20	14	4	2	5	32	8	16	26	4	68	8	11	8	24	2	74	38	10
14	36	12	54	36	20	14	20	6	60	4	20	3	84	16	12	20	66	18	12	28	18	12	37	38	20
15	48	28	13	18	10	12	60	30	58	60	12	8	6	16	70	30	33	16	24	30	5	12	74	38	150
17	68	6	13	36	12	8	36	30	29	12	62	32	24	8	6	4	3	34	24	24	72	12	74	38	60
19	4	4	27	18	12	8	38	30	58	4	60	6	24	32	4	6	66	4	8	8	35	12	9	38	20
21	16	14	9	9	4	14	10	5	29	4	62	16	42	16	28	4	22	16	8	28	24	6	74	3	20
22	18	12	27	18	66	4	10	28	58	6	60	32	6	32	12	22	66	18	24	3	24	12	74	19	30
23	36	12	54	36	10	12	36	28	12	4	62	16	12	8	12	20	22	36	92	6	36	12	9	38	20
26	4	26	18	36	5	8	36	30	29	4	62	32	24	32	26	20	11	4	44	8	72	12	74	18	20
29	12	2	26	4	4	2	20	58	29	4	62	32	4	16	2	4	11	6	44	1	35	4	74	38	20
30	48	12	27	54	20	12	18	14	60	30	62	32	18	32	60	12	17	16	6	30	35	36	37	12	150
31	18	4	54	18	20	4	4	10	60	4	62	62	6	4	4	10	68	36	22	4	36	12	37	38	20
33	48	14	54	18	66	8	60	28	15	6	31	30	24	16	42	66	66	16	12	24	9	6	37	9	30
34	34	14	54	18	20	4	20	28	12	4	20	8	12	32	28	10	68	34	6	4	3	12	37	36	20
35	16	12	54	36	30	28	36	28	58	20	31	10	84	32	15	12	33	16	44	70	8	12	72	38	100
37	18	14	26	9	30	6	20	6	60	6	62	32	3	16	42	5	33	18	24	6	7	6	4	74	30
38	8	12	52	9	30	8	38	28	30	6	31	5	24	32	12	10	34	8	8	24	35	12	6	4	3
39	3	26	27	27	12	6	18	15	60	12	30	10	18	16	52	12	22	3	33	12	72	18	74	38	12
41	12	14	54	36	12	8	20	30	29	4	15	15	24	1	14	12	68	6	44	8	24	12	36	18	20
42	48	4	13	3	30	28	9	14	10	6	60	16	42	32	12	30	68	16	24	42	72	12	74	19	6
43	8	12	52	6	6	12	18	30	1	12	31	32	6	32	12	6	17	8	8	6	35	4	74	38	30
46	6	2	52	18	12	6	20	30	58	4	12	4	6	16	4	12	68	6	2	12	36	6	12	12	20
47	8	28	52	36	30	4	12	30	60	12	60	30	12	4	42	20	22	8	44	6	18	12	37	12	60
51	102	12	27	18	12	4	30	28	58	12	31	30	6	32	12	6	17	68	24	4	72	12	74	38	12

Table 7: $51 \leq m \leq 75$, and $N \leq 51$

$g_N(m)$	$m=51$	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
$N=53$	4	1	106	12	10	2	20	14	29	4	62	32	4	16	2	20	68	4	8	2	8	4	74	6	4
55	16	12	54	18	110	8	18	15	6	10	31	8	24	16	60	22	66	16	22	40	36	6	8	38	50
57	9	14	52	27	12	6	114	28	29	6	15	8	18	16	14	12	34	18	24	6	35	18	36	38	15
58	18	2	18	3	6	12	9	58	60	6	10	32	3	32	6	1	68	18	11	6	7	4	74	9	30
59	8	28	4	18	20	8	10	28	118	4	31	30	24	32	28	10	66	8	2	8	72	12	74	38	20
61	6	1	6	3	4	8	3	10	4	2	122	32	8	16	2	4	68	6	8	8	8	2	12	38	10
62	36	12	52	12	12	4	36	28	58	12	60	31	4	8	12	12	66	18	44	12	9	4	37	12	12
65	36	26	54	36	60	3	20	14	60	20	30	32	12	4	130	12	33	18	8	30	72	12	9	3	100
66	24	4	52	54	44	8	18	15	58	12	12	15	72	8	4	66	17	8	24	8	8	18	74	19	60
67	2	4	27	2	30	12	10	28	15	12	31	30	6	32	3	10	134	4	24	6	8	4	36	36	30
69	16	2	52	6	20	8	20	10	5	4	31	10	8	8	4	10	68	16	46	8	6	2	6	19	4
70	16	14	52	18	5	28	20	30	60	2	30	15	42	32	7	10	17	16	22	14	72	12	72	36	10
71	36	28	54	36	20	4	20	4	2	4	62	30	12	8	28	20	66	36	44	4	71	12	18	36	20
73	18	14	54	9	12	8	9	30	60	2	30	32	24	8	14	12	3	18	11	8	35	3	146	18	2
74	36	6	54	4	12	12	36	7	29	4	5	32	12	32	6	12	68	18	24	6	7	4	18	74	20
77	16	4	13	4	22	14	12	10	20	4	20	32	28	8	4	44	22	16	44	7	10	4	24	3	20
78	18	26	4	18	30	12	20	28	29	6	31	10	6	32	78	10	68	18	22	6	36	12	37	18	30
79	18	12	6	6	12	12	20	30	58	4	62	16	6	4	12	12	68	36	6	12	35	4	18	38	20
82	18	6	13	3	30	8	9	7	60	6	62	15	8	32	6	5	33	18	11	24	72	4	18	38	6
83	8	28	27	6	3	8	18	28	15	12	60	32	8	32	21	6	66	8	24	24	35	4	74	36	15
85	34	14	18	1	20	2	3	10	29	10	62	32	1	16	70	4	68	34	11	10	8	2	12	3	50
86	1	14	27	9	4	12	10	28	58	4	20	32	3	32	28	2	66	2	8	12	35	12	74	36	20
87	16	12	54	18	6	4	3	58	58	12	31	30	12	16	12	6	68	16	11	12	35	12	74	38	30
89	8	14	2	36	5	8	20	10	12	4	62	32	24	8	14	20	33	8	24	8	35	12	9	38	4
91	18	52	52	18	2	28	20	28	20	4	31	16	42	32	26	2	3	36	24	14	70	12	72	38	10
93	16	14	52	3	5	2	6	4	20	2	62	62	2	16	14	5	22	16	11	2	24	2	74	38	5
94	8	12	54	18	4	4	4	28	58	4	2	15	12	8	12	4	68	8	22	4	36	6	37	38	20
95	36	4	52	36	30	6	76	15	29	20	5	15	12	16	20	12	68	18	44	30	35	12	74	36	100
97	18	14	26	9	30	8	20	10	20	6	30	5	24	8	42	5	68	18	8	24	72	3	36	38	30

Table 8: $51 \leq m \leq 75$, and $51 < N < 100$

$g_N(m)$	$m=76$	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
$N=2$	20	12	28	13	24	108	5	84	12	24	44	20	12	44	12	42	22	60	23	60	16	48	21	12	30
3	20	40	12	80	24	162	14	82	24	18	22	30	20	90	18	24	44	96	46	15	48	16	56	90	60
5	3	40	28	13	20	36	20	28	8	30	44	28	10	22	20	56	8	20	16	30	8	98	56	20	50
6	18	24	42	80	8	54	42	7	24	36	42	84	12	10	12	56	22	96	23	36	48	12	8	6	20
7	36	84	14	40	12	54	42	82	28	6	44	28	12	45	6	14	12	30	46	18	4	49	98	12	60
10	4	24	3	39	40	27	20	41	8	90	21	30	12	44	30	24	24	15	48	20	16	98	56	12	50
11	12	66	14	39	8	27	42	82	12	36	42	15	44	44	12	21	24	32	48	12	16	24	42	66	20
13	20	8	26	13	4	9	14	28	8	8	7	14	4	30	2	104	11	32	16	20	8	98	56	4	10
14	20	70	12	10	4	108	7	28	28	36	42	12	10	45	12	84	12	12	23	20	8	7	49	60	20
15	20	30	42	40	20	27	42	42	12	80	14	30	20	45	30	42	24	24	48	20	24	98	42	30	100
17	9	8	12	16	6	108	42	41	8	102	21	60	4	44	12	24	24	32	23	18	4	98	56	12	30
19	76	24	2	39	8	54	42	42	8	4	22	30	12	90	12	8	8	6	48	76	16	98	56	6	20
21	10	28	14	26	4	27	40	41	14	16	14	5	4	88	12	14	8	16	23	20	8	98	98	12	20
22	10	22	12	3	24	54	42	21	2	18	22	28	44	44	6	12	24	32	16	15	16	48	7	66	30
23	36	30	12	13	4	108	10	82	12	18	42	28	20	9	12	12	92	16	6	18	4	49	6	60	20
26	18	40	52	39	8	108	42	41	8	4	44	60	20	90	12	104	22	32	48	9	16	98	56	60	10
29	20	4	4	80	4	4	14	41	4	6	44	116	4	30	4	2	11	32	16	20	8	98	7	4	5
30	18	12	12	80	40	162	42	82	6	80	11	42	4	30	90	12	6	96	24	45	48	49	42	36	50
31	4	10	2	39	4	54	10	82	4	36	42	10	20	45	12	2	44	62	24	4	4	48	14	30	20
33	20	88	42	80	12	54	40	41	24	48	44	84	22	30	6	56	4	30	3	60	24	48	56	66	30
34	20	20	14	8	8	54	21	12	4	68	44	28	20	44	12	28	6	8	23	20	16	49	28	30	20
35	36	42	12	80	40	108	42	84	28	80	42	28	12	18	60	42	44	20	12	90	16	32	98	12	100
37	20	15	14	80	12	27	20	41	3	18	44	6	10	30	6	42	24	32	23	60	8	98	3	15	30
38	38	40	12	39	24	27	42	82	8	24	42	28	20	90	3	24	8	5	48	114	16	98	56	30	6
39	18	12	78	10	4	81	40	84	6	12	22	15	12	88	36	78	11	30	16	36	24	98	6	36	4
41	20	24	28	80	2	108	82	41	8	6	21	60	12	90	12	56	11	60	48	20	4	98	56	12	10
42	18	70	12	78	24	9	40	14	42	48	44	42	20	88	6	28	24	48	23	18	48	98	98	30	6
43	36	6	12	80	24	18	10	21	12	24	86	30	12	30	6	12	8	32	48	18	16	24	42	6	60
46	20	12	2	13	4	54	20	28	6	12	44	30	12	45	12	6	2	4	24	20	8	49	21	12	20
47	12	10	28	20	12	108	14	12	4	24	42	60	20	11	12	14	44	60	94	6	4	3	14	60	60
51	20	6	12	39	8	54	20	82	12	68	22	84	12	90	12	12	24	30	46	20	48	98	14	6	4

Table 9: $76 \leq m \leq 100$, and $N \leq 51$

$g_N(m)$	$m=76$	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
$N=53$	20	5	4	80	4	36	14	28	4	4	7	28	10	44	4	1	8	32	23	20	8	16	7	20	2
55	9	88	12	39	20	54	7	84	8	80	44	30	22	4	30	24	22	8	46	90	8	98	56	66	50
57	38	12	42	40	4	81	40	28	6	9	6	84	12	88	9	42	24	24	48	38	24	98	42	36	10
58	18	3	2	80	24	9	42	28	6	18	21	58	4	30	6	6	22	32	48	18	16	98	21	1	30
59	20	40	14	80	8	54	10	82	8	8	14	28	20	90	12	56	4	30	46	20	16	98	56	30	20
61	3	8	1	80	4	9	5	41	8	6	44	10	4	10	2	8	8	32	23	6	8	16	56	4	10
62	18	12	12	39	6	36	20	84	4	18	44	28	12	15	12	12	22	124	24	18	4	48	28	12	6
65	20	12	52	39	10	108	42	41	12	90	4	28	12	10	60	78	8	32	23	20	4	48	21	12	50
66	18	88	12	80	4	162	10	21	24	8	6	30	22	18	36	8	24	30	48	36	12	48	56	198	20
67	20	30	2	13	24	6	42	21	12	3	42	28	20	44	6	3	24	30	16	30	16	98	42	10	60
69	20	40	2	16	4	18	14	41	8	16	44	10	10	8	4	8	46	10	8	20	4	7	56	10	4
70	20	70	14	80	8	54	42	82	14	16	22	30	20	30	6	14	22	30	24	20	16	32	98	30	10
71	20	10	28	40	4	108	42	84	4	36	44	4	20	22	12	14	44	60	46	20	4	49	14	60	20
73	9	24	14	13	2	27	20	84	8	18	44	30	12	1	6	56	11	32	48	18	4	48	8	12	2
74	18	12	12	80	8	12	20	84	12	18	1	28	12	90	4	6	24	32	23	18	16	98	3	12	10
77	6	154	4	8	2	12	40	41	28	16	11	20	22	5	4	28	22	32	16	3	4	49	49	44	10
78	20	30	26	80	24	54	8	41	6	18	42	28	20	88	6	78	22	10	12	60	16	49	42	30	30
79	20	12	12	79	4	18	14	84	12	36	42	30	12	11	4	12	12	16	46	20	4	24	42	12	20
82	18	40	6	80	24	9	82	84	8	18	44	7	20	90	6	24	22	15	16	18	16	98	56	5	6
83	36	24	14	13	24	9	4	166	8	24	14	28	12	90	6	56	24	32	46	18	16	14	56	3	60
85	3	4	14	80	20	3	2	4	1	170	44	10	4	44	10	14	11	32	16	30	8	8	7	4	50
86	10	6	7	80	8	27	20	82	6	4	86	28	4	90	12	21	8	32	48	20	16	16	21	6	20
87	12	12	12	78	12	27	40	82	4	48	42	58	12	88	6	12	44	30	48	6	8	14	28	6	60
89	20	40	28	39	2	108	42	28	8	8	44	20	5	178	12	56	24	32	23	20	4	48	8	60	1
91	20	14	26	80	8	54	40	84	28	18	22	28	4	88	6	182	24	16	6	20	16	96	98	6	20
93	6	10	14	5	4	9	7	41	2	16	22	4	10	88	1	14	22	62	16	6	8	16	14	5	10
94	4	4	12	10	4	54	3	82	4	8	44	28	4	44	12	12	22	30	47	4	4	24	28	12	20
95	38	6	4	3	20	108	7	82	12	90	14	60	6	45	60	12	22	60	46	190	8	96	42	12	50
97	20	40	14	39	6	27	42	84	8	18	21	10	5	44	6	56	8	5	23	60	4	194	56	15	30

Table 10: $76 \leq m \leq 100$, and $51 < N < 100$

A.2 Maple Code

Following are a list of commands, which can be useful for computing solutions of Pell's equation, and $g_N(m)$. These commands have been written for Maple software version 12.02.

To begin we need the following packages:

```
restart; with(numtheory)
```

The command `NGen` takes an integer (preferably square free) and returns the generator of G as a list.

```
NGen := proc (N::integer)
local cf, z, x, y, j, test, i;
cf := cfrac(sqrt(N));
x := nthnumer(cf, 1);
y := nthdenom(cf, 1);
test := false; i := 1;
while test = false do
if x^2-N*y^2 = 1
then test := true
else i := i+1;
cf := cfrac(sqrt(N), i);
x := nthnumer(cf, i);
y := nthdenom(cf, i)
end if
end do;
return [x, y]
end proc:
```

The command `NMult` accepts as input two lists (these should be solutions to (1))

and an integer (N). It multiplies the two lists according to the binary operation defined for G, and returns the product as a list. The command `mMult` accepts as input two lists (these should be solutions to (1)) and an integer (N) and an integer (m). It performs the multiplication mod m, and then returns the product as a list.

```
   NMult := proc (x1::list, x2::list, N)
return [x1[1]*x2[1]+N*x1[2]*x2[2], x1[1]*x2[2]+x1[2]*x2[1]]
end proc:
   mMult := proc (x1::list, x2::list, N::integer, m::integer)
return [mod(x1[1]*x2[1]+N*x1[2]*x2[2], m), mod(x1[1]*x2[2]+x1[2]*x2[1],
m)]
end proc:
```

The command `mGroup` accepts as input two integers. The first is N, and the second is the integer m, which will be used in reduction mod m. It returns a list of group elements (each presented as a list with two elements), $g_N(m)$, N, and m.

```
   mGroup := proc (N::integer, m::integer)
local g, group, h;
group := [];
g := NGen(N);
h := mMult(g, [1, 0], N, m);
while not h = [1, 0] do
group := [op(group), h];
h := mMult(h, g, N, m)
end do;
group := [op(group), h];
return [group, nops(group), N, m]
end proc:
```

The command `GN` accepts the same input as the previous command. It returns

the order of the group $\mathrm{red}_m(G)$.

```
    GN := proc (N::integer, m::integer)
local g, k, h;
g := NGen(N);
h := mMult(g, [1, 0], N, m);
k := 1; while not h = [1, 0] do
h := mMult(h, g, N, m)
k:= k+1
end do;
return k
end proc:
```

The command `mGroup` can take some time to run, because of the inherent difficulty in finding the generator for the group G. If the generator is known, then `mGroupGen` will accept two integers (N and m) and a generator as input, and build the group generated by this element. It performs a check that the alleged generator is in fact a solution of (1). This command can save on time if the generator for G is difficult to compute. It returns the same as the previous command.

```
   mGroupGen := proc (N::integer, m::integer, gen::list)
local group, h;
group := [];
if gen[1]^2-N*gen[2]^2 = 1
then h := mMult(gen, [1, 0], N, m);
while not h = [1, 0] do
group := [op(group), h];
h := mMult(h, gen, N, m)
end do;
group := [op(group), h];
```

```
return [group, nops(group), N, m]
else print("The input should be a solution to Pell's equation")
end if
end proc:
```

B Dirichlet's Unit Theorem

In Lagrange's proof we saw that the study of the Pell equation naturally leads to the study of quadratic field extensions of $\mathbb{Q}$. Consider the quadratic field extention $\mathbb{Q}(\sqrt{N}) = \left\{a + b\sqrt{N} | a, b \in \mathbb{Q}\right\}$. There is an injective map from G to the set $I = \left\{x + y\sqrt{N} | x, y \in \mathbb{Z}\right\}$. In fact, if we restrict the image to the set $\left\{x + y\sqrt{N} \in I | x^2 - Ny^2 = \pm 1\right\}$, then the map is actually a group homomorphism. Because of this natural relation, we can use Dirichlet's Unit Theorem to describe the group structure of G.

Before we can state Dirichlet's Unit Theorem, we will need more background in quadratic fields. Then we will state and prove Dirichlet's Unit Theorem for quadratic fields of the form $\mathbb{Q}(\sqrt{N})$ where $N > 1$ is a square-free integer.

It is not hard to check that $\mathbb{Q}(\sqrt{N})$ is a field. In fact, if $\alpha = a + b\sqrt{N}$, then

$$\alpha^{-1} = \frac{a - b\sqrt{N}}{a^2 - b^2 N}.$$

Now if $\alpha = a + b\sqrt{N}$ is an element of our field, $\mathbb{Q}(\sqrt{N})$, then we define the *conjugate* of α to be

$$\bar{\alpha} = a - b\sqrt{N}.$$

There is an associated norm $\mathrm{N} : \mathbb{Q}(\sqrt{N}) \to \mathbb{Q}$ defined by

$$\mathrm{N}(\alpha) = \alpha\bar{\alpha} = a^2 - Nb^2.$$

We also want to define the trace function $\mathrm{Tr} : \mathbb{Q}(\sqrt{N}) \to \mathbb{Q}$ by

$$\mathrm{Tr}(\alpha) = \alpha + \bar{\alpha} = 2a.$$

Now we put $A = \mathbb{Z}[\omega] = \{a + b\omega | a, b \in \mathbb{Z}\}$, where ω is defined to be

$$\omega = \begin{cases} \sqrt{N} & \text{if } N \equiv 2, 3 \pmod 4 \\ \frac{1+\sqrt{N}}{2} & \text{if } N \equiv 1 \pmod 4 \end{cases}.$$

It is straightforward to show that A is a ring. Also, it is clear that if $N \equiv 2, 3 \pmod 4$, then for $\alpha \in A$, $\mathrm{N}(\alpha)$ and $\mathrm{Tr}(\alpha)$ are integers. If $N \equiv 1 \pmod 4$, and $\alpha \in A$, we can write $\alpha = a + b\omega = \left(a + \frac{1}{2}b\right) + \frac{1}{2}\sqrt{N}$. Therefore $\bar{\alpha} = \left(a + \frac{1}{2}b\right) - \frac{1}{2}\sqrt{N} = a + b\bar{\omega}$. Hence

$$\mathrm{N}(\alpha) = \alpha\bar{\alpha} = a^2 + b(\omega + \bar{\omega}) + b^2\omega\bar{\omega}$$
$$\mathrm{Tr}(\alpha) = \alpha + \bar{\alpha} = 2a + b(\omega + \bar{\omega}).$$

Now,

$$\omega\bar{\omega} = \tfrac{1}{4} - \tfrac{1}{4}N = \tfrac{1}{4}(N - 1)$$

which is an integer, since $N \equiv 1 \pmod 4$. Also,

$$\omega + \bar{\omega} = (\tfrac{1}{2} + \tfrac{1}{2}) + (\tfrac{1}{2} - \tfrac{1}{2})\omega = 1.$$

So no matter how we choose N, we have $\alpha \in A$ satisfying the polynomial

$$x^2 - \mathrm{Tr}(\alpha)x + \mathrm{N}(\alpha) = 0,$$

which is a monic polynomial with integer coefficients. We call A the *ring of integers* of the field $\mathbb{Q}(\sqrt{N})$.

Definition 20. For any ring R with unity, the *group of units*, $R^\times$ is the set of invertible elements of R, i.e.

$$R^\times = \{u \in R | uv = 1 \text{ for some } v \in R\}$$

It is straightforward to show that $R^\times$ forms a group under the operation multiplication. The identity of R is clearly a unit. Further, if r is a unit, then r^{-1} is clearly a unit, and given two units r, and s, it is clear that $(rs)^{-1} = r^{-1}s^{-1}$, so the product rs is also a unit.

We have seen that our norm function restricts to $\mathrm{N} : A \to \mathbb{Z}$. Now if α is a unit of A there exists $\beta \in A$, such that $\alpha\beta = 1$. Now our norm function is multiplicative, so we have $\mathrm{N}(\alpha)\,\mathrm{N}(\beta) = \mathrm{N}(\alpha\beta) = \mathrm{N}(1) = 1$. Since $\mathrm{N}(\alpha)$ is an integer, $\mathrm{N}(\alpha) = \pm 1$.

Conversely, if $\mathrm{N}(\alpha) = 1$, we have $\alpha\bar{\alpha} = 1$, and if $\mathrm{N}(\alpha) = -1$, then $\alpha(-\bar{\alpha}) = 1$. In either case α is a unit. Thus we see that $A^\times = \{\alpha \in A | \,\mathrm{N}(\alpha) = \pm 1\}$.

We need to know the structure of $A^\times$. The structure is given by Dirichlet's unit theorem. We will give a proof of Dirichlet's Theorem for quadratic fields. This generalizes to arbitrary number fields, for which proof, we refer to [1].

Theorem 21 (Dirichlet)**.** *The group of units $A^\times$ is isomorphic to $\{\pm 1\} \times \mathbb{Z}$.*

We will need a few lemmas, before we can prove this theorem.

Lemma 22. *For any constant $c > 0$, there are only finitely many elements $\alpha \in A$, such that*

$$|\alpha| \leq c, \text{ and } |\bar{\alpha}| \leq c.$$

Proof. For any element $\alpha \in A$, we can write $\alpha = a + b\omega$ where $a, b \in \mathbb{Z}$. Similarly we

can write $\bar{\alpha} = a + b\bar{\omega}$. So we define the matrix

$$P = \begin{pmatrix} 1 & \omega \\ 1 & \bar{\omega} \end{pmatrix}, \quad x = \begin{pmatrix} a \\ b \end{pmatrix}$$

Then we see that $Px = \begin{pmatrix} \alpha \\ \bar{\alpha} \end{pmatrix}$. Clearly P is invertible. The inverse is

$$P^{-1} = \frac{1}{\sqrt{N}} \begin{pmatrix} \bar{\omega} & -\omega \\ -1 & 1 \end{pmatrix},$$

and so we have $x = P^{-1} \begin{pmatrix} \alpha \\ \bar{\alpha} \end{pmatrix}$, or in other words, $\sqrt{N}a = \alpha\bar{\omega} - \bar{\alpha}\omega$, and $\sqrt{N}b = \bar{\alpha} - \alpha$. Hence, if $|\alpha| \leq c$ and $|\bar{\alpha}| \leq c$, then

$$\begin{aligned} \sqrt{N}|a| &\leq |\alpha||\bar{\omega}| + |\bar{\alpha}||\omega| \\ &\leq 2c\omega \\ \sqrt{N}|b| &\leq |\bar{\alpha}| + |\alpha| \\ &\leq 2c. \end{aligned}$$

Now we have a bound for a and b, independent of our choice of α. Since a and b are necessarily integers, there are only finitely such α. □

Definition 23. We say a subgroup $\Gamma \subset \mathbb{R}$ is *discrete* if for every $c > 0$, there are only finitely many elements $g \in \Gamma$ such that $|g| < c$.

Lemma 24. *Any non-trivial discrete subroup $\Gamma \subset \mathbb{R}$ is isomorphic to the infinite cyclic group, $\mathbb{Z}$.*

Proof. Since Γ is discrete and non-trivial, we can find a smallest positive element

$g \in \Gamma$. Let Γ' be the infinite cyclic group generated by g. Clearly $\Gamma' \subset \Gamma$. We need to show the reverse, i.e. $\Gamma \subset \Gamma'$. We can write any element $\gamma \in \Gamma$ as $\gamma = ng + r$, with $0 \leq r < g$. Since Γ is a group, we see that $r \in \Gamma$. Since g is the smallest positive element, we must have $r = 0$, i.e. $\gamma = ng$. □

Lemma 25. *Given two real numbers a_1, a_2, let $a = |a_1| + |a_2|$. If $t > 1$ is an integer, then there exist non-zero integers x_1, x_2, such that if $y = a_1x_1 + a_2x_2$,*

$$\max\{|x_1|, |x_2|\} \leq t \text{ and } |y| \leq 2at^{-1}$$

Proof. Let h be an integer with $t^2 \leq h < (t+1)^2$. Subdivide the interval $[-at, at]$ into h equal parts. There are $(t+1)^2$ pairs of integers x_1, x_2 such that $x_i \in \{0, 1, \ldots, t\}$. There must be two pairs (ξ_1, ξ_2), and (η_1, η_2) with $a_1\xi_1 + a_2\xi_2$ and $a_1\eta_1 + a_2\eta_2$ in the same subinterval. So now $|y| = |a_1(\xi_1 - \eta_1) + a_2(\xi_2 - \eta_2)| \leq \frac{2at}{h} \leq 2at^{-1}$ □

Lemma 26. *There is a constant $c > 0$ depending only on N, such that given an integer $t > 1$, there is a non-zero algebraic integer α with*

$$c^{-1}t^{-1} \leq |\alpha| \leq ct^{-1} \text{ and } c^{-1}t \leq |\bar{\alpha}| \leq ct$$

Proof. By the previous lemma, there are integers x_1, x_2 such that $|x_i| \leq t$, and $|x_1 + x_2\omega| \leq 2at^{-1}$, where $a = |1| + |\omega|$. Let $\alpha = x_1 + x_2\omega$. Clearly

$$|\alpha| \leq 2at^{-1} \text{ and } |\bar{\alpha}| = |x_1 + x_2\bar{\omega}| \leq c_1t$$

where $c_1 = \max\{1, \bar{\omega}\}$. Let $c = \max\{2a, c_1\}$. Then we have

$$1 \leq \mathrm{N}(\alpha) = |\alpha||\bar{\alpha}|$$

So from our previous calculations, we have $|\bar{\alpha}| \geq c^{-1}t$, and $|\alpha| \geq c^{-1}t^{-1}$. □

Now we can prove the theorem.

Proof of Dirichlet's Unit Theorem. Define the homomorphism

$$\lambda : A^\times \to \mathbb{R}$$

by $\lambda(\alpha) = \log|\alpha|$. Here we think of $\mathbb{R}$ as an additive group. To see that this is indeed a homomorphism of groups, suppose $\alpha, \beta \in A^\times$. Then we have

$$\lambda(\alpha\beta) = \log|\alpha\beta| = \log|\alpha| + \log|\beta|.$$

Now consider $\ker\lambda$. If $\epsilon \in \ker\lambda$, then $\log|\epsilon| = 0$, i.e. $|\epsilon| = 1$. But recall $A^\times = \{\epsilon \in A | \operatorname{N}\epsilon = \pm 1\}$. Therefore $|\epsilon\bar{\epsilon}| = |\operatorname{N}(\epsilon)| = 1$, and so $|\bar{\epsilon}| = 1$ as well. So by lemma 22, we see that $\ker\lambda$ is finite. Since the kernel is a subgroup, every element in the kernel has finite order.

Conversely, suppose that $u \in A^\times$ has finite order, i.e. there is an integer k, such that $u^k = 1$. Then we have $|u^k| = |u|^k = 1$, and so $|u| = 1$. So we see that $u \in \ker\lambda$. So the kernel is exactly the set of elements of finite order. Since $\mathbb{Q}(\sqrt{N}) \subset \mathbb{C}$ every element of the kernel is a root of unity. Furthermore, $\mathbb{Q}(\sqrt{N})$ is a real extension of $\mathbb{Q}$, therefore it contains exactly two roots of unity, ± 1. Clearly $\{\pm 1\} \subset A^\times$, therefore $\ker\lambda = \{\pm 1\}$.

We will now show that $\lambda(A^\times)$ is a discrete subgroup of $\mathbb{R}$. To this end, suppose we are given $c > 0$. We must show that there are only finitely many $\alpha \in A^\times$ with the property that $|\log|\alpha|| < c$, i.e.

$$e^{-c} \leq |\alpha| \leq e^c.$$

Because $\operatorname{N}(\alpha) = 1$, we know that $|\bar{\alpha}| = |\alpha|^{-1}$. Hence $e^{-c} \leq |\bar{\alpha}| \leq e^c$. Lemma 22 again shows that there are finitely many such α. So now we know that $\lambda(A^\times)$ is a discrete

subgroup of $\mathbb{R}$. We need to show that the image is non-trivial.

To show that $\lambda(A^\times)$ is non-trivial, we will show that there exists an element $\epsilon \in A^\times$ with $|\epsilon| > 1$.

Let c be as in Lemma 26. We can choose an integer $M > c^2$. Let $t_1 > 1$ be an integer, and let $t_{i+1} = Mt_i$ for each $i \geq 1$. Then for each t_i, the previous lemma shows that there is an algebraic integer α_1, such that

$$|\alpha_{i+1}| \leq ct_{i+1}^{-1} = c(Mt)^{-1} < c^{-1}t_i^{-1} \leq |\alpha_i|$$
$$|\bar{\alpha}_{i+1}| \geq c^{-1}t_{i+1} = c^{-1}Mt_i > ct_i \geq |\bar{\alpha}_i|$$

Also we have $|\,\mathrm{N}(\alpha_i)| \leq (ct_i^{-1})(ct_i) = c^2$. By a previous theorem there are only finitely many non-associate α such that $|\,\mathrm{N}(\alpha)| \leq c^2$, hence $\alpha_\nu = \epsilon\alpha_\mu$ for $\mu > \nu$ and some unit ϵ. So we have

$$|\epsilon| = \left|\frac{\alpha_\mu}{\alpha_\nu}\right| < 1 \text{ and } |\bar{\epsilon}| = \left|\frac{\bar{\alpha}_\mu}{\bar{\alpha}_\nu}\right| > 1$$

□

Now in order to see that G is cyclic, we notice that there is a group homomorphism $G \to A^\times$ defined by sending $(x, y) \to x + y\sqrt{N}$. The kernel of this map is the set $\{(1, 0)\}$, hence the map is injective.

References

[1] J.S. Chahal. *Topics in Number Theory*. Plenum Press, 1988.

[2] Leonhard Euler. *Elements of Algebra, with additions of M. de LaGrange*. Longman, Reeves, Hurst and Co, 3rd edition, 1822.

[3] G.H. Hardy and E.M. Wright. *An Introduction to the Theory of Numbers*. Clarendon Press, 1960.

[4] T.L. Heath, editor. *The Works of Archimedes*. Dover Publications, 2002.

[5] Ivor Thomas. *Selections Illustrating the History of Greek Mathematics: with an english translation by Ivor Thomas*, volume 2. Harvard University Press, 1941.

[6] V.S. Varadarajan. *Algebra in Ancient and Modern Times*. American Mathematical Society, 1998.

[7] Andre Weil. *Number Theory: An approach through history*. Birkhauser, 1906.

[8] H.C. Williams, R.A. German, and C.R. Zarnke. Solution of the cattle problem of archimedes. *Mathematics of Computation*, 19:671–674, 1965.