5 The Pell equation

5.1 Side and diagonal numbers

In ancient time, only rational numbers were thought of as numbers. Hence the discovery that (1) $\sqrt{2}$ is the length of a hypoteneuse of a right triangle, and (2) $\sqrt{2}$ is irrational (which we proved in Section 3, but the Greeks also had another proof for) was quite perplexing. Hence the ancient Greeks studied the Diophantine equation

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

to try to understand $\sqrt{2}$. They were able to produce a sequence of increasingly large solutions (x_i, y_i). Note that we can rewrite such an equation as

$$\frac{x^2}{y^2} = 2 + \frac{1}{y^2} \iff \frac{x}{y} = \sqrt{2 + \frac{1}{y^2}} \to \sqrt{2}$$

as $y \to \infty$. Hence the solutions (x_i, y_i) provide increasingly good rational approximations to $\sqrt{2}$ as y_i gets large.

We will study the solutions in $\mathbb{Z}$ to the more general **Pell equation**

$$x^2 - ny^2 = 1$$

for any $n \in \mathbb{N}$. Note Pell's equation always has the "trivial" solutions $(\pm 1, 0)$. Further, the case where n is a square is easy:

Exercise 5.1. *If $n \in \mathbb{N}$ is a square, show the only solutions of $x^2 - ny^2 = 1$ are $(\pm 1, 0)$. (Cf. Exercises 5.1.3, 5.1.4.)*

Hence, from now on, we will *assume n is not a square.* Then we know $\sqrt{n}$ is irrational from Section 2.5.

To understand this equation thoroughly over $\mathbb{Z}$, we *need* to work with another number system

$$\mathbb{Z}[\sqrt{n}] = \left\{ a + b\sqrt{n} : a, b \in \mathbb{Z} \right\}.$$

This idea of working with a larger number system than $\mathbb{Z}$ to study problems about $\mathbb{Z}$ is the basis of algebraic number theory. (Note we could also think of each $\mathbb{Z}/m\mathbb{Z}$ a smaller number system than $\mathbb{Z}$, which we used in Chapter 3 to study problems over $\mathbb{Z}$ as well. However because $\mathbb{Z}/m\mathbb{Z}$ is smaller, it can typically only be used to limit the kinds of solutions we might have to an equation over $\mathbb{Z}$, and not actually prove the existence of solutions.) Here we take for granted the existence of $\sqrt{n}$, which as pointed out above, was not always known (or thought) to be a number. The reason we want to work specifically with the **ring**[1] $\mathbb{Z}[\sqrt{n}]$ is of course so we can factor Pell's equation:

$$x^2 - ny^2 = (x + y\sqrt{n})(x - y\sqrt{n}) = 1.$$

If $x = a + b\sqrt{n}$, we say a is the **rational part** and b is the **irrational part** of x. (This is analogous to real and imaginary parts of complex numbers.) Note that two numbers $x, y \in \mathbb{Z}[\sqrt{n}]$

[1] A ring is, roughly, a number system in which you can add, subtract and multiply, but not necessarily divide. We will give the formal definition in Chapter 10. For now, you don't need to know anything about rings; we will just get in the habit of calling some number systems rings to stress that they are somehow similar to $\mathbb{Z}$.

are equal if and only if their rational and irrational parts are equal. The ($\Leftarrow$) direction is obvious. To prove the ($\Rightarrow$) direction, write $x = a_1 + b_1\sqrt{n}$ and $y = a_2 + b_2\sqrt{n}$. Then

$$x = y \iff a_1 - a_2 = (b_2 - b_1)\sqrt{n}.$$

If $b_2 - b_1 \neq 0$, then $\sqrt{n} = \frac{a_1 - a_2}{b_2 - b_1} \in \mathbb{Q}$ which is a contradiction. Hence $b_2 = b_1$, and therefore $a_1 = a_2$ also.

You might wonder about the title of this section. I won't cover it. See the text.

5.2 The equation $x^2 - 2y^2 = 1$

It is simple to determine all rational solutions of $x^2 - 2y^2 = 1$ using the rational slope (Diophantus chord) method. However determining the integer solutions is a different matter. First we observe by trial-and-error that the smallest non-trivial solution is $(3, 2)$.

Exercise 5.2. *Check the following* **composition rule** *holds:*

$$(x_1^2 - 2y_1^2)(x_2^2 - 2y_2^2) = x_3^2 - 2y_3^2$$

where

$$x_3 = x_1x_2 + 2y_1y_2, \quad y_3 = x_1y_2 + y_1x_2.$$

Hence if (x_1, y_1) and (x_2, y_2) are to solutions to $x^2 - 2y^2 = 1$, so is there composition (x_3, y_3), defined as above. We denote

$$(x_3, y_3) = (x_1, y_1) \cdot (x_2, y_2).$$

We will see in the next section that the solutions to $x^2 - 2y^2 = 1$ form a group under this operation. Further, since the definition of composition is symmetric in (x_1, y_1) and (x_2, y_2) this will be an abelian group.

Example. $(x_1, y_1) \cdot (\pm 1, 0) = (\pm x_1, \pm y_1)$.

Example. $(3, 2) \cdot (3, 2) = (9 + 8, 12) = (17, 12)$

Example. $(3, 2)^3 = (3, 2) \cdot (17, 12) = (99, 70)$.

Through composition (the "powers" of $(3, 2)$) we can see that we can get infinitely many solutions, each getting larger. The first three powers give the sequence of approximations

$$\begin{aligned} \frac{3}{2} &= 1.5 \\ \frac{17}{12} &= 1.41\overline{6} \\ \frac{99}{70} &= 1.4\overline{1428757} \\ \approx \sqrt{2} &= 1.4142135623\ldots \end{aligned}$$

Exercise 5.3. *Compute* $(3, 2)^4$. *Use this to obtain a decimal approximation for* $\sqrt{2}$. *To how many digits is it accurate? (Use a calculator/computer.)*

5.3 The group of solutions

This section shows that the solutions of $x^2 - 2y^2 = 1$ form a group under the composition defined above, and this group is generated by $(3, 2)$ and $(-1, 0)$. However it is subsumed in the next section which treats $x^2 - ny^2 = 1$ using norms, so I will not treat the case $n = 2$ separately.

5.4 The general Pell equation and $\mathbb{Z}[\sqrt{n}]$

To put the ideas we will use in context, let us recall some things about complex numbers. Let $z \in \mathbb{C}$. Then we can write $z = x + yi$ where $x, y \in \mathbb{R}$. The complex conjugate of z is $\overline{z} = x - yi$, and

$$z\overline{z} = (x + yi)(x - yi) = x^2 + y^2.$$

Drawing z as a vector in the complex plane, we see that $z\overline{z}$ is the square of the length of this vector, i.e., $z\overline{z} = |z|^2$. Define the norm of z to be

$$N(z) = z\overline{z}.$$

Since complex conjugation respects multiplication,

$$N(z_1 z_2) = z_1 z_2 \overline{z_1 z_2} = z_1 \overline{z}_1 z_2 \overline{z}_2 = N(z_1)N(z_2),$$

i.e, the norm is *multiplicative.*

Similar to the complex case, define the **conjugate** of $\alpha = x + y\sqrt{n} \in \mathbb{Z}[\sqrt{n}]$ to be $\overline{\alpha} = x - y\sqrt{n}$, and the **norm** of α to be

$$N(\alpha) = \alpha\overline{\alpha} = (x + y\sqrt{n})(x - y\sqrt{n}) = x^2 - ny^2.$$

Note $N(\alpha) = N(\overline{\alpha})$.

The following lemma is clear.

Lemma 5.1. *Solutions of $x^2 - ny^2 = 1$ are in 1-1 correspondence with the elements in $\mathbb{Z}[\sqrt{n}]$ of norm 1. The correspondence is given by $(x, y) \leftrightarrow x + y\sqrt{n}$.*

Now we want to know a basic property of norms.

Lemma 5.2. *For $\alpha, \beta \in \mathbb{Z}[\sqrt{n}]$, we have $N(\alpha\beta) = N(\alpha)N(\beta)$, i.e., N is* multiplicative.

Proof. Write $\alpha = x_1 + y_1\sqrt{n}, \beta = x_2 + y_2\sqrt{n}$. Note

$$\overline{\alpha\beta} = x_1x_2 + ny_1y_2 - (x_1y_2 + y_1x_2)\sqrt{n} = (x_1 - y_1\sqrt{n})(x_2 - y_2\sqrt{n}) = \overline{\alpha} \cdot \overline{\beta}.$$

Hence

$$N(\alpha\beta) = \alpha\beta\overline{\alpha\beta} = \alpha\overline{\alpha} \cdot \beta\overline{\beta} = N(\alpha)N(\beta).$$

□

Hence if α and β have norm 1, so does $\alpha\beta$. In light of Lemma 5.1, this says if we have two solutions to $x^2 - ny^2 = 1$, we can compose them to construct a third. Precisely, we can say something stronger.

Corollary 5.3. (Brahmagupta composition rule) *If (x_1, y_1) and (x_2, y_2) are solutions to*

$$x_1^2 - ny_1^2 = a, \;\; x_2^2 - ny^2 = b.$$

Then the composition

$$(x_3, y_3) = (x_1, y_1) \cdot (x_2, y_2) := (x_1x_2 + ny_1y_2, x_1y_2 + y_1x_2)$$

is a solution of

$$x_3^2 - ny_3^2 = ab.$$

Proof. We simply translate the above into a statement about norms. The hypothesis says $N(x_1 + y_1\sqrt{n}) = a$ and $N(x_2 + y_2\sqrt{n}) = b$. Now observe that

$$(x_1 + y_1\sqrt{n})(x_2 + y_2\sqrt{n}) = x_1x_2 + ny_1y_2 + (x_1y_2 + y_1x_2)\sqrt{n} = x_3 + y_3\sqrt{n}.$$

Hence by the multiplicative property of the norm,

$$x_3^2 - ny_3^2 = N(x_3 + y_3\sqrt{n}) = N(x_1 + y_1\sqrt{n})N(x_2 + y_2\sqrt{n}) = ab.$$

□

Note this is much nicer than the straightforward proof given on p. 82.

Aside: this result says that if a and b are of the form $x^2 - ny^2$, so is ab. Hence if we want to ask the question which integers are of the form $x^2 - ny^2$, we should first determine which primes are of the form $x^2 - ny^2$. We will not pursue this now, however we will return to this idea when considering which numbers are sums of squares, or more generally, of the form $x^2 + ny^2$. (The "+" case turns out to be simpler, but still not easy.)

Since $\mathbb{Z}[\sqrt{n}] \subseteq \mathbb{R}$, there is a natural order on $\mathbb{Z}[\sqrt{n}]$. By Lemma 5.1, this gives us a way to order the solutions to Pell's equation. As the case of $n = 2$ suggests, we want to first look for a "smallest" non-trivial solution and try to obtain all other solutions from that.

If (x, y) is a solution to $x^2 - ny^2 = 1$, so are $(\pm x, \pm y)$. So when we say we want a "smallest" solution, we should make a restriction like $x, y > 0$. Thinking in terms of elements of norm 1, note that conjugates and negatives of $\alpha = x + y\sqrt{n}$ give $\pm x \pm y\sqrt{n}$. Hence we want to look for the smallest element of norm 1 such that $x, y > 0$.

Definition 5.4. *The* **fundamental +unit**[2] *ϵ of $\mathbb{Z}[\sqrt{n}]$ is the smallest $\epsilon = x + y\sqrt{n} \in \mathbb{Z}[\sqrt{n}]$ such that $x, y > 0$ and $N(\epsilon) = 1$.*

Lemma 5.5. *The fundamental +unit of $\mathbb{Z}[\sqrt{n}]$ is well defined and always exists.*

Proof. We will show in the next section that there is always some $\epsilon \neq \pm 1$ in $\mathbb{Z}[\sqrt{n}]$ such that $N(\epsilon) = 1$. By possibly taking the negative and/or conjugate of ϵ, we may assume $x, y > 0$. So at least one candidate exists. Since the set of all $a + b\sqrt{n}$ with $a, b \in \mathbb{N}$ is discrete in $\mathbb{R}$, there must be a minimal such ϵ (i.e., such ϵ cannot get arbitrarily close to 1). □

[2]This is not standard terminology. One normally defines the fundamental unit, which can have norm ± 1. If it has norm $+1$, this coincides with our definition; if it has norm -1, its square is what we are calling the fundamental +unit.

Example. *The fundamental +unit of* $\mathbb{Z}[\sqrt{2}]$ *is* $3+2\sqrt{2}$.

Exercise 5.4. *An alternative definition of fundamental unit is the smallest* $\epsilon > 1$ *such that* $N(\epsilon) = 1$. *Prove that this is equivalent to the above definition as follows. Suppose* $\epsilon = x + y\sqrt{n} > 1$ *and* $N(\epsilon) = 1$. *Show (i)* $0 < \overline{\epsilon} < 1$. *Then deduce (ii)* $x, y > 0$.

Theorem 5.6. *Let* $U^+ = \{\alpha \in \mathbb{Z}[\sqrt{n}] : N(\alpha) = 1\}$. *Then* U^+ *is an infinite abelian group under multplication. Furthermore, it is generated by the fundamental unit* ϵ *of* $\mathbb{Z}[\sqrt{n}]$ *and* -1.

Proof. Clearly the identity $1 \in U^+$ and multiplication on U^+ is associative. Note that for any $\alpha \in U^+$, $N(\alpha) = \alpha\overline{\alpha} = 1$ implies that $\overline{\alpha} = \alpha^{-1}$. Since $N(\overline{\alpha}) = 1$ also, we have $\alpha \in U^+$. Also by the multiplicative property of the norm, if $\alpha, \beta \in U^+$ then $N(\alpha\beta) = N(\alpha)N(\beta) = 1$ so $\alpha\beta \in U^+$. This shows U^+ is a group, and it is clearly abelian because multiplication in $\mathbb{Z}[\sqrt{n}]$ is commutative.

Now let ϵ be the fundamental +unit of $\mathbb{Z}[\sqrt{n}]$. Suppose there exists $\alpha \in U^+$ such that $\alpha \neq \pm\epsilon^m$ for any $m \in \mathbb{Z}$. By taking the negative and/or conjugate if need be, we may assume $\alpha > 1$. Since ϵ is minimal and $\epsilon^m \to \infty$ as $m \to \infty$, there must be some $m > 0$ such that $\epsilon^m < \alpha < \epsilon^m + 1$. But then $1 < \alpha\epsilon^{-m} < \epsilon$ and $N(\alpha\epsilon^{-m}) = 1$, contradicting the minimality of ϵ. Hence each $\alpha \in U^+$ is $(\pm)$ a power of ϵ. □

Remark. All $\alpha \in U^+$ are called *units* of $\mathbb{Z}[\sqrt{n}]$, because like ± 1, they are invertible in $\mathbb{Z}[\sqrt{n}]$. The actual definition of the units of $\mathbb{Z}[\sqrt{n}]$ is the set of invertible elements, which is easy to see is precisely the set of elements of norm ± 1.

Hence the solutions of Pell's equation are given by $\pm\epsilon^m$ where $m \in \mathbb{Z}$. Since we know $\overline{\epsilon} = \epsilon^{-1}$, then $\epsilon^{-m} = \overline{\epsilon}^m$. Thus ϵ^m and ϵ^{-m} give essentially the same solutions.

Corollary 5.7. *Suppose* $\epsilon = x_0 + y_0\sqrt{n}$ *is a fundamental +unit of* $\mathbb{Z}[\sqrt{n}]$. *Then all integer solutions to Pell's equation* $x^2 - ny^2 = 1$ *are of the form* $(\pm x, \pm y)$ *where* $x + y\sqrt{n} = (x_0 + y_0\sqrt{n})^m$ *and* $m \geq 0$. *Equivalently, up to sign, all solutions to Pell's equations are given by non-negative powers (in the sense of Brahmagupta composition) of the* **fundamental solution** (x_0, y_0).

Example. *Up to sign, all non-trivial solutions of* $x^2 - 2y^2 = 1$ *are given by* $(x + y\sqrt{2}) = (3 + 2\sqrt{2})^m$ *for* $m > 0$, *i.e., x and y are the rational and irrational parts of* $(3 + 2\sqrt{2})^m$.

The book says little about how to find fundamental solutions (called smallest positive solutions in the text). By rewriting Pell's equation as

$$x^2 = ny^2 + 1$$

it becomes clear that we can find the fundamental solution (or fundamental +unit) by finding the smallest $y > 0$ such that $ny^2 + 1$ is a square. This will give the smallest $x > 0$ which solves $x^2 - ny^2 = 1$, i.e., x and y are simultaneously minimal for this solution, making $x + y\sqrt{n}$ minimal (with $x, y > 0$) among U^+.

Example. *Since* $3 \cdot 1^2 + 1$ *is a square, the smallest positive (fundamental) solution to* $x^2 - 3y^2 = 1$ *is* $(2, 1)$. *Hence the fundamental +unit of* $\mathbb{Z}[\sqrt{3}]$ *is* $2 + \sqrt{3}$. *Up to sign, all solutions are powers of* $(2, 1)$, *e.g.,* $(2, 1)^2 = (7, 4)$ *and* $(2, 1)^3 = (26, 15)$. *This provides the successive approximations* $\frac{2}{1}$, $\frac{7}{4}$, $\frac{26}{15}$ *for* $\sqrt{3}$.

Exercise 5.5. *Find the fundamental solution (x_0, y_0) to $x^2 - 5y^2 = 1$. What is the fundamental $+$unit of $\mathbb{Z}[\sqrt{5}]$? Compute the solutions given by the square and the cube of (x_0, y_0). What rational number decimal approximations to $\sqrt{5}$ do they yield? To how many digits are they accurate? (Use a calculator.)*

Exercise 5.6. *Exercises 5.4.4, 5.4.5.*

5.5 The pigeonhole argument

The simple-minded method for determining fundamental solutions above is only practical for small n. For instance, when $n = 61$, the fundamental solution is

$$(1766319049, 226153980)$$

(Bhaskara II, 12th century; Fermat). In general, one can, for instance, use the classical theory of *continued fractions.* We will not go into this here, but we will prove the existence of a non-trivial solution for all nonsquare n, which is due to Lagrange in 1768. However, we will give a proof due to Dirichlet (ca. 1840). It uses the

Pigeonhole principle. If $m > k$ pigeons go into k boxes, at least one must box must contain more than 1 pigeon (finite version). If infinitely many pigeons go into k boxes, at least one box must contain infinitely many pigeons (infinite version).

Proposition 5.8. (Dirichlet's approximation theorem) *For any nonsquare n and integer $B > 1$, there exist $a, b \in \mathbb{Z}$ such that $0 < b < B$ and*

$$|a - b\sqrt{n}| < \frac{1}{B}.$$

(This says that $\frac{a}{b}$ is close to $\sqrt{n}$.)

Proof. Consider the $B - 1$ irrational numbers

$$\sqrt{n}, 2\sqrt{n}, \ldots, (B-1)\sqrt{n}.$$

For each such $k\sqrt{n}$, let $a_k \in \mathbb{N}$ be such that

$$0 < a_k - k\sqrt{n} < 1.$$

Partition the interval $[0, 1]$ into B subintervals of length $\frac{1}{B}$. Then, of the $B + 1$ numbers

$$0, a_1 - \sqrt{n}, a_2 - \sqrt{n}, \ldots, a_{B-1} - (B-1)\sqrt{n}, 1$$

in $[0, 1]$ two of them must be in the same subinterval of length $\frac{1}{B}$. Hence they are less than distance $\frac{1}{B}$ apart, i.e., their difference satisfies $|a - b\sqrt{n}| < \frac{1}{B}$. Further their irrational parts must be distinct, so we have $-B < b < B$ with $b \neq 0$. If $b > 0$ we are done; if $b < 0$, simply multiply a and b by -1. □

Step 1. Fix $B_1 = B$. Then by above, there exists $|a_1 - b_1\sqrt{n}| < \frac{1}{B} < \frac{1}{b_1}$. Let $B_2 > B_1$ such that $\frac{1}{B_2} < |a_1 - b_1\sqrt{n}|$. Applying Dirichlet's approximation again, we get a new pair (a_2, b_2) of integers such that

$$|a_2 + b_2\sqrt{n}| < \frac{1}{B_2} < \frac{1}{b_2}.$$

Repeating this we see there an *infinite sequence* of integer pairs (a, b) such that $|a - b\sqrt{n}|$ gets smaller and smaller, and

$$|a - b\sqrt{n}| < \frac{1}{b}.$$

for all (a, b). (This is gives a infinite sequence of increasingly good approximations.)

Step 2. Assume (a, b) satisfy $|a - b\sqrt{n}| < \frac{1}{b}$. Note that

$$|a + b\sqrt{n}| \leq |a - b\sqrt{n}| + |2b\sqrt{n}| \leq 1 + 2b\sqrt{n} \leq 3b\sqrt{n}.$$

Then

$$|a^2 - nb^2| = |a + b\sqrt{n}||a - b\sqrt{n}| \leq 3b\sqrt{n}\frac{1}{b} = 3\sqrt{n}.$$

Hence there are infinitely many $a - b\sqrt{n} \in \mathbb{Z}[\sqrt{n}]$ whose norm, in absolute values, is at most $3\sqrt{n}$.

Step 3. By successive applications of the (infinite) pigeonhole principle, we have

(i) infinitely many $a - b\sqrt{n}$ with the same norm N, where $|N| \leq 3\sqrt{n}$ (the norm is always an integer)

(ii) infinitely many $a - b\sqrt{n}$ with norm N and $a \equiv a_0 \bmod N$ for some a_0.

(iii) infinitely many $a - b\sqrt{n}$ with norm N, $a \equiv a_0 \bmod N$, $b \equiv b_0 \bmod N$ for some b_0.

In particular, we have two $a_1 - b_1\sqrt{n}$, $a_2 - b_2\sqrt{n}$ such that they both have norm N, $a_1 \equiv a_2 \bmod N$, $b_1 \equiv b_2 \bmod N$, and $a_1 - b_1\sqrt{n} \neq \pm(a_2 - b_2\sqrt{n})$. (It's possible $N < 0$, and we define mod N for negative N to be the same as mod $|N|$. However, $N \neq 0$ because 0 is the only element of $\mathbb{Z}[\sqrt{n}]$ of norm 0.)

Step 4. Consider

$$a + b\sqrt{n} = \frac{a_1 - b_1\sqrt{n}}{a_2 - b_2\sqrt{n}} = \frac{(a_1 - b_1\sqrt{n})(a_2 - b_2\sqrt{n})}{a_2^2 - nb_2^2} = \frac{a_1a_2 - nb_1b_2}{N} + \frac{a_1b_2 - b_1a_2}{N}\sqrt{n}.$$

Since $a_1 - b_1\sqrt{n} \neq \pm(a_2 - b_2\sqrt{n})$, surely $a + b\sqrt{n} \neq \pm 1$. If we know $a, b \in \mathbb{Z}$, then since

$$N(a + b\sqrt{n}) = N(a_1 - b_1\sqrt{n})N\left((a_2 - b_2\sqrt{n})^{-1}\right) = NN^{-1} = 1,$$

we get that $a + b\sqrt{n}$ is an element of $\mathbb{Z}[\sqrt{n}]$ of norm 1 which is not ± 1.

To show that a is an integer, observe that $N|a_1a_2 - nb_1b_2$ because

$$a_1a_2 - nb_1b_2 \equiv a_1a_1 - nb_1b_1 \equiv a_1^2 - nb_1^2 \equiv 0 \bmod N.$$

The first congruence holds because $a_1 \equiv a_2 \bmod N$ and $b_1 \equiv b_2 \bmod N$. Similarly, b is an integer because

$$a_1b_2 - b_1a_2 \equiv a_1b_1 - b_1a_1 \equiv 0 \bmod N.$$

This proves

Theorem 5.9. *If $n \in \mathbb{N}$ is nonsquare, then $x^2 - ny^2 = 1$ has a nontrivial solution in $\mathbb{Z}$, i.e., a solution besides $(\pm 1, 0)$.*

5.6 *Quadratic forms

Note: This section does not AT ALL follow what is in the text.

The ideas above can be put into a more general context. We say $Q(x, y)$ is a *binary quadratic form* if

$$Q(x, y) = ax^2 + bxy + cy^2.$$

The basic questions are, which numbers k are of the form

$$k = Q(x, y),$$

and for such n, what are the solutions (or at least, how many are there?). We answered the question thoroughly for $Q(x, y) = x^2 - ny^2$ ($n > 0$) and $k = 1$: 1 is always of the form $x^2 - ny^2$—in two ways if n is a square and in infinitely many ways otherwise, and we showed how to determine all solutions.

Assuming n is not a square, if k is of the form $x^2 - ny^2$, then there are infinitely many solutions to

$$x^2 - ny^2 = k,$$

and they are generated from a fundamental solution. The reason is that such solutions correspond to elements of $\mathbb{Z}[\sqrt{n}]$ of norm k. If α has norm k, then so does $\mu\alpha$ for any μ of norm 1, and we showed that there are infinitely many elements of norm 1 in Section 5.4. We will not deal with the question of which k are of the form $x^2 - ny^2$ here, but it was treated in Gauss' *Disquistiones*.

The form $x^2 - ny^2$ is called an **indefinite form** because it takes on positive and negative values. The general theory of indefinite forms is similar, and another interesting example is the case of the form

$$Q(x, y) = x^2 + xy - y^2.$$

Here the solutions to $Q(x, y) = 1$ are given by (F_{2n+1}, F_{2n+2}) where F_n is the n-th Fibonacci number (cf. Exercise 5.8.4; $F_1 = F_2 = 1$). The form $Q(x, y)$ is the norm of the element $x + y\frac{1+\sqrt{5}}{2}$ in

$$\mathbb{Z}[\frac{1+\sqrt{5}}{2}] := \left\{ a + b\frac{1+\sqrt{5}}{2} : a, b \in \mathbb{Z} \right\}.$$

Here, the golden ratio $\frac{1+\sqrt{5}}{2}$ is a fundamental unit for $\mathbb{Z}[\frac{1+\sqrt{5}}{2}]$, but this has norm -1. In fact the solutions are generated by the powers of the fundamental $+$unit, $1 + \frac{1+\sqrt{5}}{2} = \frac{3+\sqrt{5}}{2}$. Hence this gives an interesting way of computing the Fibonacci numbers:

$$\left(\frac{3+\sqrt{5}}{2}\right)^n = F_{2n+1} + F_{2n+2}\frac{1+\sqrt{5}}{2}.$$

In fact, proving this relation (say by induction) is an alternative way of showing Exercise 5.8.4.

Exercise 5.7. *Check that* $\left(\frac{3+\sqrt{5}}{2}\right)^n = F_{2n-1} + F_{2n}\frac{1+\sqrt{5}}{2}$ *holds for* $n = 1, 2, 3$.

Opposed to the indefinite forms, we have the **definite forms**. We say $Q(x, y)$ is **positive definite** (resp. **negative definite**) if $Q(x, y) \geq 0$ (resp $Q(x, y) \leq 0$) for all $x, y \in \mathbb{Z}$. For example,

$x^2 + ny^2$ for $n \in \mathbb{N}$ is a positive definite form. (The negative definite forms are just the negatives of positive definite forms, so it makes sense to study just the positive ones.)

In contrast to the indefinite case, it is clear that if k is of the form

$$x^2 + ny^2 = k$$

there are only finitely many solutions for (x, y). These are in 1-1 correspondence with the elements of norm k in the *imaginary quadratic ring* $\mathbb{Z}[\sqrt{-n}]$. While the point of view of norms is similar to the indefinite case, the definite and indefinite cases have a rather different flavor (with the definite case being the more easy of the two).

In the next chapter, we will study the ring of Gaussian integers $\mathbb{Z}[i]$, with the goal in mind of determining which numbers are the sum of two squares $x^2 + y^2$. Brahmagupta composition, as we remarked earlier, suggests that we can reduce the problem to the question of which *primes* are sums of two squares, for which the pattern becomes much more apparent.

5.7 *The map of primitive vectors

5.8 *Periodicity in the map of $x^2 - ny^2$

The material in these two optional sections is an introduction to Conway's recent (in the last 20 years or so) new insights into a visual approach to binary quadratic forms. While the material is interesting, we will focus on other things in this class. If you are interested in learning about it, I recommend Conway's own (small) book, *The Sensual Quadratic Form*.

5.9 Discussion

Probably worth reading.