12. Units

Suppose that the components of $\alpha \in \mathbb{Z}[\sqrt{d}]$ are positive. Then $\alpha > 1$. The four elements of $\mathbb{Z}[\sqrt{d}]$ with components equal to the components of α up to sign are

$$\alpha, \qquad \bar{\alpha}, \qquad -\bar{\alpha}, \qquad \text{and} \qquad -\alpha.$$

If in addition $N(\alpha) = 1$ then

$$\alpha\bar{\alpha} = 1,$$

so that the numbers above are

$$\alpha, \qquad \frac{1}{\alpha}, \qquad -\frac{1}{\alpha}, \qquad \text{and} \qquad -\alpha.$$

Note that the first is bigger than 1, the second is smaller than one and bigger than zero, the last is smaller than -1 and the third is between -1 and 0. In particular these numbers are arranged largest to smallest and the signs of the components of α correspond to the size of α. In particular, solutions with positive components correspond to elements of $\mathbb{Z}[\sqrt{d}]$ bigger than one.

Definition-Lemma 12.1. *The solutions α to*

$$N(\alpha) = 1,$$

are a subgroup G_1 of the non-zero real numbers under multiplication.

Proof. $1 \in G_1$ (and in fact we already proved that G_1 contains a non-trivial element) so that G_1 is non-empty. If α and $\beta \in G_1$ then

$$\begin{aligned} N(\alpha\beta) &= N(\alpha)N(\beta) \\ &= 1, \end{aligned}$$

so that $\alpha\beta \in G_1$. Similarly

$$\begin{aligned} N(\bar{\alpha}) &= N(\alpha) \\ &= 1, \end{aligned}$$

so that

$$\frac{1}{\alpha} = \bar{\alpha} \in G_1. \qquad \square$$

Definition 12.2. *We call $\delta \in G_1$ the **fundamental solution** if the components of δ are positive and δ is minimal with these properties.*

Theorem 12.3. *The map*

$$\phi\colon \mathbb{Z} \times \mathbb{Z}_2 \longrightarrow G_1 \qquad \textit{given by} \qquad (n, e) \longrightarrow e\delta^n,$$

where $\mathbb{Z}_2 = \{\pm 1\}$, is an isomorphism of groups.

Proof. It is clear that ϕ is a group homomorphism and the kernel is trivial.

We have already seen that if $\alpha \in G_1$, $\alpha \neq 1$, then one of $\pm\alpha^{\pm 1}$ has all of its components positive. Hence it suffices to show that if $\alpha \in G_1$ and all of the components of α are positive then $\alpha = \delta^n$ for some natural number n.

Since $\alpha \in G_1$ has positive components it follows that $\delta \leq \alpha$ by minimality of δ. There is a unique natural number n such that

$$\delta^n \leq \alpha < \delta^{n+1}.$$

Let

$$\beta = \frac{\alpha}{\delta^n} \in G_1.$$

Then $1 \leq \beta < \delta$ and so $\beta = 1$ by minimality of δ. But then $\alpha = \delta^n$. $\square$

We also want to consider three other special cases of Pell's equation. Let $E(k)$ denote all solutions to

$$x^2 - dy^2 = k.$$

Proposition 12.4. *Suppose that $E(-1)$ is non-empty. Let $\gamma \in E(-1)$ be minimal with positive coefficients.*

Then $\delta = \gamma^2$ and the elements of $E(-1)$ are $\pm\gamma\delta^n$, $n \in \mathbb{Z}$. In fact $G_{-1} = E(1) \cup E(-1)$ is a subgroup of the group of all non-zero real numbers, $G_1 = E(1)$ is a normal subgroup of index 2 and $E(-1)$ is the other coset.

Proof. We have

$$\begin{aligned} N(\gamma^2) &= N(\gamma)N(\gamma) \\ &= 1, \end{aligned}$$

so that $\gamma^2 \in G_1$. As γ^2 has positive coefficients it follows that $1 < \delta \leq \gamma^2$, by minimality of δ. As

$$\frac{1}{\gamma} = -\bar{\gamma},$$

it follows that

$$\gamma^{-1} < -\delta\bar{\gamma} \leq \gamma.$$

Let $\beta = -\delta\bar{\gamma}$. Then

$$\begin{aligned} N(\beta) &= N(-1)N(\delta)N(\gamma) \\ &= -1, \end{aligned}$$

so that $\beta \in E(-1)$. In particular $\beta \neq 1$.

There are two possibilities for where β lies:

$$\gamma^{-1} < \beta \leq 1 \qquad \text{and} \qquad 1 < \beta \leq \gamma.$$

The former inequality implies that

$$1 \geq \beta^{-1} < \gamma.$$

By minimality of γ this cannot happen and in the latter inequality we must have equality, so that $\beta = \gamma$ and so $\delta = \gamma^2$.

Note that G_{-1} is non-empty and closed under multiplication and inverses. Therefore it is a subgroup of the group of non-zero real numbers under multiplication. If $\alpha \in E(-1)$ then

$$\begin{aligned} N(\gamma\alpha) &= N(\gamma)N(\alpha) \\ &= (-1)^2 \\ &= 1. \end{aligned}$$

Thus $\gamma\alpha \in E(1) = G_1$ and so $\gamma\alpha = \pm\delta^n$. But then $\alpha = \pm\gamma\delta^n$. □

In fact the map

$$\phi\colon \mathbb{Z} \times \mathbb{Z}_2 \longrightarrow G_{-1} \qquad \text{given by} \qquad (n, e) \longrightarrow e\gamma^n,$$

where $\mathbb{Z}_2 = \{ \pm 1 \}$, is an isomorphism of groups.

Note that $E(4)$ is non-empty. Indeed, if $\beta \in E(1)$

$$\begin{aligned} N(2\beta) &= N(2)N(\beta) \\ &= 4. \end{aligned}$$

However, not every element of $E(4)$ has to be of this form. For example,

$$3^2 - 5 \cdot 1^2 = 4,$$

so that

$$3 + \sqrt{5} \in \mathbb{Z}[\sqrt{5}],$$

and yet the components are odd.

Proposition 12.5. *If $\zeta \in E(4)$ is minimal subject to $\zeta > 1$ then*

$$E(4) = \{ \pm 2 \left(\frac{\zeta}{2} \right)^n \mid n \in \mathbb{Z} \}.$$

If $E(-4)$ is non-empty and $\eta \in E(-4)$ is minimal subject to $\eta > 1$ then $\zeta = \eta^2$ and

$$E(-4) = \{ \eta \left(\frac{\zeta}{2} \right)^n \mid n \in \mathbb{Z} \}.$$

Proof. Suppose that

$$x^2 - dy^2 = 4.$$

As d is square-free, it follows that x and y have the same parity (odd versus even). Thus each solution looks like $a(1 + \sqrt{d}) \mod 2$, where

$a = 0$ or 1. If there is a solution with $a = 1$ then d is odd. Anyway, if we are given α and $\beta \in E(4)$ then we may write

$$\alpha = a(1 + \sqrt{d}) \qquad \text{and} \qquad \beta = b(1 + \sqrt{d}) \mod 2,$$

where $ab = 0$ or 1. In this case

$$\alpha\beta = ab(d + 1 + 2\sqrt{d}) \mod 2.$$

By what we just said, $ab(d + 1) \equiv 0 \mod 2$. Thus

$$\frac{\alpha\beta}{2} = 2 \cdot \frac{\alpha}{2} \cdot \frac{\beta}{2} \in \mathbb{Z}[\sqrt{d}]$$

and

$$\begin{aligned} N(\frac{\alpha\beta}{2}) &= \frac{1}{4} N(\alpha)N(\beta) \\ &= 4, \end{aligned}$$

so that

$$\frac{\alpha\beta}{2} \in E(4).$$

Thus the RHS of

$$E(4) = \{ \pm 2 \left(\frac{\zeta}{2} \right)^n \mid n \in \mathbb{Z} \}$$

is indeed a subset of the LHS.

The rest of the proof follows the same line of argument as (12.3) and (12.4). □

Definition-Lemma 12.6. *Define a relation $\sim$ on elements of $E(k)$ by the rule $\alpha \sim \beta$ if there is an element $\gamma \in E(1)$ such that $\alpha = \beta\gamma$.*

*Then $\sim$ is an equivalence relation and the equivalence classes are called **classes**.*

Proof. Easy check. □

For example, the equation

$$x^2 - 2y^2 = 49,$$

has solutions 7 and $9 + 4\sqrt{2}$ and one can easily see that these solutions belong to different classes. On the other hand there are only finitely many classes in general, and determining all of the classes is easy once we know the fundamental solution δ.

Theorem 12.7. *Every class has a representative $\alpha = u + v\sqrt{d}$ with*

$$\sqrt{k} < u \leq \sqrt{\Delta k},$$

where

$$\Delta = \frac{1}{2}\left(1 + \frac{\delta}{\delta - 1}x_1\right)$$

and $\delta = x_1 + y_1\sqrt{d}$ *is the fundamental solution.*

In particular there are only finitely many classes.

Proof. Given $\alpha_1 = u_1 + v_1\sqrt{d} \in E(k)$ we want to find $\alpha = u + v\sqrt{d} \in E(k)$ in the same class such that

$$\sqrt{k} < u \leq \sqrt{\Delta k}.$$

As α_1 and $-\alpha_1$ belong to the same class we may assume that $u_1 > 0$. In this case it simply suffices to make sure that if

$$u_1 > \sqrt{\Delta k},$$

then we may find α such that $0 < u < u_1$.

Thus we want to find α such that

$$u+v\sqrt{d} = (x+y\sqrt{d})(u_1+v_1\sqrt{d}) \qquad 0 < u < u_1 \qquad \text{and} \qquad x^2-dy^2 = 1.$$

If $v_1 > 0$ then let

$$x + y\sqrt{d} = \delta^{-1} = x_1 - y_1\sqrt{d}$$

but if $v_1 < 0$ then let

$$x + y\sqrt{d} = \delta.$$

Either way we have

$$\begin{aligned} u &= u_1x_1 - y_1|v_1|d \\ &= u_1\left(x_1 - y_1\sqrt{d}\frac{|v_1|\sqrt{d}}{u_1}\right) \\ &= u_1\left[x_1 - y_1\sqrt{d} + y_1\sqrt{d}\left(1 - \sqrt{1 - \frac{k}{u_1^2}}\right)\right]. \end{aligned}$$

Note that for $0 < t < 1$ we have

$$\begin{aligned} 0 &< 1 - \sqrt{1-t} \\ &= \frac{t}{1 + \sqrt{1-t}} \\ &< \frac{t}{2-t}. \end{aligned}$$

Hence

$$0 < u < u_1\left(\delta^{-1} + \frac{y_1\sqrt{d}k}{2u_1^2 - k}\right).$$

It is not too hard to check that if

$$u_1 > \sqrt{\frac{\delta' y_1 \sqrt{d} + 1}{2} k} \qquad \text{where} \qquad \delta' = \frac{\delta}{\delta - 1},$$

then the coefficient of u_1 is less than one, so that $u < u_1$. Since

$$y_1 \sqrt{d} = \sqrt{x_1^2 - 1} < x_1,$$

it follows that if $u > \sqrt{\Delta k}$ then we can find a smaller solution.

As we have bounded u and u is an integer, it follows that there are only finitely many choices for u and so only finitely many classes. $\square$

In theory (12.7) gives a way to write down all classes. In practice the solutions can be quite large. For the equation

$$x^2 - 61y^2 = 1$$

the fundamental solution has

$$x_1 = 1,766,319,049.$$

We will see later an efficient way to find solutions.

Linear equations and Pell's equations are practically the only equations with infinitely many integral solutions. No curve of genus at least two has infinitely many rational solutions; this is a famous theorem due to Faltings. No curve of genus one has infinitely many integral solutions; this is a famous theorem due to Siegel. If a curve of genus zero has infinitely many solutions then it has a parametrisation of the form

$$x = \frac{A(t)}{C^n(t)} \qquad \text{and} \qquad y = \frac{B(t)}{C^n(t)},$$

where A, B and C are polynomials in t with rational coefficients and $C(t)$ is either linear or a quadratic equation which assumes both positive and negative values, as does the polynomial $t^2 - d$ of the Pell equation.

Note however that there is a big distrinction between the Pell equation and linear equations. The equation

$$ax + by + c = 0$$

assymptotically has solutions which grow faster than X, in the range $0 < x < X$. Pell's equation has solutions which grow like $\log X$.