4.1 Reduction theory

Let $Q(x,y) = ax^2 + bxy + cy^2$ be a binary quadratic form ($a, b, c \in \mathbb{Z}$). The **discriminant** of Q is $\Delta = \Delta_Q = b^2 - 4ac$. This is a fundamental invariant of the form Q.

Exercise 4.1. *Show there is a binary quadratic form of discriminant $\Delta \in \mathbb{Z}$ if and only if $\Delta \equiv 0, 1 \bmod 4$. Consequently, any integer $\equiv 0, 1 \bmod 4$ is called a* **discriminant**.

We say two forms $ax^2 + bxy + cy^2$ and $Ax^2 + Bxy + Cy^2$ are **equivalent** if there is an invertible change of variables

$$x' = rx + sy,\ y' = tx + uy,\ r, s, t, u \in \mathbb{Z}$$

such that

$$a(x')^2 + bx'y' + c(y')^2 = Ax^2 + Bxy + Cy^2.$$

Note that the change of variables being invertible means the matrix

$$\begin{pmatrix} r & s \\ t & u \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z}).$$

In fact, in terms of matrices, we can write the above change of variables as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} r & s \\ t & u \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Going further, observe that

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + bxy + cy^2,$$

so we can think of the quadratic form $ax^2 + bxy + cy^2$ as being the symmetric matrix $\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$. It is easy to see that two quadratic forms $ax^2 + bxy + cy^2$ and $Ax^2 + Bxy + Cy^2$ are equivalent if and only if

$$\begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} = \tau^T \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \tau \tag{4.1}$$

for some $\tau \in \mathrm{GL}_2(\mathbb{Z})$.

Note that the discriminant of $ax^2 + bxy + cy^2$ is $-4\mathrm{disc}\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$.

Lemma 4.1.1. *Two equivalent forms have the same discriminant.*

Proof. Just take the matrix discriminant of Equation (4.1) and use the fact that any element of $\mathrm{GL}_2(\mathbb{Z})$ has discriminant ± 1. □

What this means then is that $\mathrm{GL}_2(\mathbb{Z})$ acts on the space $\mathcal{F}_\Delta$ of binary quadratic forms of discriminant Δ for any Δ. The importance of equivalent forms is in the following.

We say $Q(x,y) \in \mathcal{F}_\Delta$ **represents** an integer n if $Q(x,y) = n$ for some $x, y \in \mathbb{Z}$.

Lemma 4.1.2. *Two equivalent forms represent the same integers.*

Proof. This is obvious from the definition of equivalence—just make the change of variables! □

To prove some basic results, it will be helpful to have more refined notions of equivalence and representations of integers.

Definition 4.1.3. *We say two forms $ax^2+bxy+cy^2$ and $Ax^2+Bxy+Cy^2$ are* **properly equivalent** *if they satisfy Equation (4.1) for some $\tau \in \mathrm{SL}_2(\mathbb{Z})$. In this case we will write $ax^2 + bxy + cy^2 \sim Ax^2 + Bxy + Cy^2$.*

Recall $\mathrm{SL}_2(\mathbb{Z})$ means 2×2 integer matrices of determinant 1, so $\mathrm{GL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z}) \cup \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathrm{SL}_2(\mathbb{Z})$ since $\mathrm{GL}_2(\mathbb{Z})$ consists of matrices of determinant ± 1 In other words the quotient $\mathrm{GL}_2(\mathbb{Z})/\mathrm{SL}_2(\mathbb{Z})$ consists of two cosets. We can take $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ as a set of representatives for these cosets.

Clearly proper equivalence implies equivalence but the converse is not true. In fact, the notion of proper equivalence turns out to give a nicer theory as we will see below.*

Example 4.1.4. *Using the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ we see $ax^2 + bxy + cy^2$ is always equivalent to $ax^2 - bxy + cy^2$. Sometimes they are properly equivalent (e.g., $2x^2 \pm 2xy + 3y^2$—see exercise below) and sometimes they are not (e.g., $3x^2 \pm 2xy + 5y^2$—see exercise below).*

Exercise 4.2. *Determine the discriminants of $Q_1(x, y) = 2x^2 + 2xy + 3y^2$, $Q_2(x, y) = 2x^2 - 2xy + 3y^2$, $Q_3(x, y) = 3x^2 + 2xy + 5y^2$ and $Q_4(x, y) = 3x^2 - 2xy + 5y^2$. Show Q_1 and Q_2 are properly equivalent but Q_3 and Q_4 are not. (If you have trouble, see the theorem below.)*

Exercise 4.3. *Fix $a, b, c \in \mathbb{Z}$ and let $Q_1(x, y) = ax^2 + bxy + cy^2$ and $Q_2(x, y) = cx^2 + bxy + ay^2$.*

(i) Show Q_1 and Q_2 are equivalent.

(ii) If $b = 0$ show Q_1 and Q_2 are properly equivalent.

(iii) Can you find a, b, c so that Q_1 and Q_2 are equivalent by not properly equivalent? (One often calls this improper equivalence*.)*

There are three types of binary quadratic forms $Q(x, y) = ax^2 + bxy + cy^2$ based on the sign of the discriminant $\Delta = b^2 - 4ac$:

1) If $\Delta = 0$, then Q factors into two linear forms and we say Q is **degenerate**. Otherwise Q is **nondegenerate**. If $\Delta = b^2 - 4ac = 0$, then $Q(x, y) = (\sqrt{a}x + \sqrt{c}y)^2$ and it is easy to see what numbers Q represents. Hence, one may just consider nondegenerate forms.

2) If $\Delta < 0$, then $Q(x, y)$ has no real roots. In other words, considering x, y real, the graph of $z = Q(x, y)$ in $\mathbb{R}^3$ never crosses the $z = 0$ plane. Hence $Q(x, y)$ takes on either only positive values or negative values (and zero if $x = y = 0$). Accordingly we say, Q is either a **positive definite** form (e.g., $x^2 + y^2$) or a **negative definite** form (e.g., $-x^2 - y^2$). Note positive definite implies $a, c > 0$ and negative definite implies $a, c < 0$. (This is not if and only if: $x^2 - 100000xy + y^2$ is not positive definite.) Since $-Q$ will be positive definite whenever Q is negative definite, it suffices to study the positive definite case.

*For this reason, many authors use the term "equivalence of forms" to mean proper equivalence. If you consult other references, take note of this.

3) If $\Delta > 0$, then $Q(x, y)$ has a real root and $Q(x, y)$ takes on positive and negative values for $x, y \in \mathbb{Z}$. In this case, we say Q is an **indefinite** form. Note whenever a and c have different sign (or one is 0), Q must be indefinite. The theory of indefinite forms is similar to the theory of (positive) definite forms, but there are some technical differences which make it more complicated. For simplicity, as well as the fact that positive definite forms tend to be of more interest, we will restrict our study to positive definite forms, though we will make some comments about what happens for indefinite forms along the way.

Hence, **from now on, we assume all our forms are positive definite** (in particular have discriminant $\Delta < 0$) unless otherwise stated.

Definition 4.1.5. *Let $Q(x, y) = ax^2+bxy+cy^2$ be a (positive definite) form. We say Q is* **reduced** *if*

$$|b| \leq a \leq c$$

and $b \geq 0$ if $a = c$ or $a = |b|$.

Lagrange introduced the notion of reduced forms, and the point is the following.

Theorem 4.1.6. *Any (positive definite) form Q is properly equivalent to a unique reduced form.*

Proof. First we will show Q is properly equivalent to a reduced form $ax^2 + bxy + cy^2$. Suppose $|b|$ is minimal such that there is a form $R(x, y) = ax^2 + bxy + cy^2$ with $Q \sim R$. If $|b| > a$, then there exists $m \in \mathbb{Z}$ such that $|2am + b| < |b|$. But this implies

$$R'(x, y) = R(x + my, y) = ax^2 + (2am + b)xy + (am^2 + bm + c)y^2 \sim R(x, y) = ax^2 + bxy + cy^2,$$

so $R' \sim Q$ has smaller xy coefficient than $ax^2 + bxy + cy^2$, contradicting the choice of R. Hence $|b| \leq a$ and similarly $|b| \leq c$. If necessary, we may replace $R(x, y)$ with $R(y, -x) = cx^2 - bxy + ay^2$ to assume $|b| \leq a \leq c$.

We also need to show we can take $b \geq 0$ if $a = c$ or $a = |b|$. If $a = c$, then the xy-coefficient of either $R(x, y)$ or $R(y, -x)$ is nonnegative, so we may assume $b \geq 0$. Similarly if $b = -a$, then $R(x + y, y) = ax^2 + ax + cy^2$, so again we may assume $b \geq 0$ (in fact $b > 0$ since $a > 0$). This shows Q is properly equivalent to some reduced form $R(x, y) = ax^2 + bxy + cy^2$.

Now we show that this R is unique. Suppose not, so $Q \sim S$ where $S = dx^2 + exy + fy^2$ is also reduced. Interchanging R and S if necessary, we may assume $a \geq d$. Recall $R \sim S$ means we can write

$$S(x, y) = R(rx + sy, tx + uy) = a(rx + sy)^2 + b(rx + sy)(tx + uy) + c(tx + uy)^2$$

with $r, s, t, u \in Z$ such that $ru - st = 1$.

Since S clearly represents d and, we know $R \sim S$, we know R represents d. Thus

$$d = ax_0^2 + bx_0y_0 + cy_0^2 \geq a(x_0^2 + y_0^2) + bx_0y_0 \geq a(x_0^2 + y_0^2) - a|x_0y_0| \geq a|x_0y_0|$$

for some $x_0, y_0 \in \mathbb{Z}$. Since $d \leq a$ we must either have $x_0y_0 = 0$ or $|x_0y_0| = 1$. We will finish the proof in three cases.

First suppose $y_0 = 0$. Then $d = ax_0^2$ together with $d \leq a$ means $x_0^2 = 1$ and $d = a$. Then the x^2-coefficient of $S(x, y)$ is

$$ar^2 + brt + ct^2 = R(r, t) = d = a.$$

Observe the minimum nonzero value of a reduced form $R(x,y)$ is obtained precisely when $(x,y) = (\pm 1, 0)$, so we must have $r = \pm 1, t = 0$.[†] (Hence the minimum positive value of a reduced form is the x^2-coefficient, in this case a.) Further $ru - st = ru = 1$ implies $u = r^{-1}$. Then the xy-coefficient of $S(x,y)$ is

$$2ars + bru = 2ars + b = b \pm 2as = e.$$

Since S is reduced, we have $|e| \leq d = a$, but the only was this can happen is if $s = 0$ (which means $R = S$) or if $b = a$ and $s = \pm 1$, which means $e = -d$, which we have excluded from our definition of reduced. This shows uniqueness when $y_0 = 0$.

Next suppose $x_0 = 0$. Similar to the above, we see $d = c$ and looking at the x^2-coefficient of $S(x,y)$ one can conclude $r = 0, s = t = \pm 1$ as the second smallest minimum nonzero value of a reduced form $R(x,y)$ is c and is obtained precisely when $(x,y) = (0, \pm 1)$[‡]—see Remark below. This means the xy-coefficient of $S(x,y)$ must be $b + 2cu$. Then $|e| = |b + 2cu| \leq a \leq c$ means either $u = 0$ (so $R = S$) or $b = a = c$ and $u = -1$ in which case $S(x,y) = ax^2 - axy + ay^2$, which is not reduced.

Finally suppose $|x_0 y_0| = 1$. The above inequalities for d say $a \geq d \geq a|x_0 y_0|$ so $a = d$. The rest follows like the $y_0 = 0$ case. □

Remark. It should be fairly obvious that for a reduced form $R(x,y) = ax^2 + bxy + cy^2$ the minimum nonzero value is obtained is a, which happen precisely when $(x,y) = (\pm 1, 0)$ (assuming $a < c$). It may be less clear that the second smallest nonzero value obtained is c, but both of these assertions follow from the simple exercise that $R(x,y) \geq (a - |b| + c)\min(x^2, y^2)$. If you want, you can work this out on your own, but I'm not assigning it as homework.

The above theorem tells us that if we want to study positive definite forms, it suffices to consider reduced forms.

4.2 The mass formula

One of the most important early discoveries about quadratic forms is that they are better studied collectively than individually. Precisely, we make the following

Definition 4.2.1. *Let Δ be a discriminant. The* **form class group** *$\mathcal{C}l(\Delta)$ of discriminant Δ is the set of proper equivalence classes of forms of discriminant Δ, i.e., $\mathcal{C}l(\Delta) = \mathcal{F}_\Delta / \sim$.*

We will later see how to define a group structure on this, justifying the name. From the last section, we know we can take a set of representatives for $\mathcal{C}l(\Delta)$ to be the set of reduced forms of discriminant Δ.

Proposition 4.2.2. *For any discriminant Δ, the number $h(\Delta) := |\mathcal{C}l(\Delta)| < \infty$.*

Proof. Note $h(\Delta)$ is the number of reduced forms of discriminant Δ. If $ax^2 + bxy + cy^2$ is reduced of discriminant Δ, then $|b| \leq a \leq c$ so $4b^2 \leq 4ac = b^2 + \Delta$, i.e., $3b^2 \leq |\Delta|$. In other words, there are only finitely many choices for b. Each choice for b determines ac, and the product ac determines a finite number of choices for a and c. □

[†] If $a = c$, the minimum nonzero value of the form is also obtained at $(0, \pm 1)$, which corresponds to $r = 0, t = \pm 1$. Though we technically omit this case here, we can actually absorb this situation into our argument for the subsequent $x_0 = 0$ case.

[‡] Again, technically if $a = c$, this actually gives the minimum nonzero value of the form, but this does not affect our argument.

Using the notion of the class group, we can get a formula for the number of representations of n by a quadratic form of discriminant Δ. It will be convenient to consider proper representations.

Definition 4.2.3. *We say $Q(x,y) \in \mathcal{F}_\Delta$* **properly represents** *n if $n = Q(x,y)$ for some $x, y \in \mathbb{Z}$ with $\gcd(x,y) = 1$. In this case, the solution (x,y) is called a* **proper representation** *of n by Q.*

Example 4.2.4. *Let $Q(x,y) = x^2 + y^2$. Then Q represents $4 = 2^2 + 0^2 = Q(2,0)$, but it does not properly represent 4 since $\gcd(2,0) = 2$ and $(\pm 2, 0)$ and $(0, \pm 2)$ are the only solutions to $Q(x,y) = 4$.*

On the other hand even though 25 has an improper representation *by Q, namely $25 = 5^2 + 0^2 = Q(5,0)$ and $\gcd(5,0) = 5$, 25 also has a* proper representation *by Q: $25 = 3^2 + 4^2 = Q(3,4)$ and $\gcd(3,4) = 1$. Hence we say $Q(x,y)$ properly represents 25.*

Lemma 4.2.5. *$Q(x,y)$ represents n if and only if $Q(x,y)$ properly represents m for some $m|n$ such that $\frac{n}{m}$ is a square.*

Proof. ($\Rightarrow$) Suppose $Q(x,y)$ represents n. If Q properly represents n, we can just take $m = n$ and we are done. If not, then $Q(x,y) = n$ for some x, y with $\gcd(x,y) = d > 1$. Then $Q(x/d, y/d) = Q(x,y)/d^2$ is a proper representation of $m = n/d^2$.

($\Leftarrow$). Suppose $Q(x,y)$ properly represents m where $n = d^2 m$. Then $Q(dx, dy) = d^2 Q(x,y) = d^2 m = n$. □

In other words, understanding what numbers are properly represented by Q tells us which numbers are represented by Q, since the latter numbers are just squares times the former numbers. Let $r_Q(n)$ (resp. $R_Q(n)$) denote the number of proper representations (resp. number of representations) of Q by n.

Theorem 4.2.6. (Dirichlet's mass formula, first version) *Let $d > 1$ be squarefree and set $\Delta = -4d$. Let $Q_1, Q_2, \ldots, Q_h$ be a set of representatives for the form class group $\mathcal{C}l(\Delta)$. Then*

$$r_{Q_1}(n) + r_{Q_2}(n) + \cdots + r_{Q_h}(n) = 2 \prod_{p|n} \left(1 + \left(\frac{\Delta}{p}\right)\right)$$

where p runs over prime divisors of $n > 0$ and $\gcd(n, \Delta) = 1$.

There are many different versions of the statement of this result, but the above one is the most applicable to the forms $x^2 + dy^2$ (with d squarefree). The proof of the mass formula is quite elementary, and we will omit it now, but John Paul will present a proof of the following version later this semester.

Theorem 4.2.7. (Dirichlet's mass formula, second version) *Let Δ be the discriminant of an imaginary quadratic field, $Q_1, Q_2, \ldots, Q_h$ a set of representatives for $\mathcal{C}l(\Delta)$ and $n > 0$ such that $\gcd(n, \Delta) = 1$. Then*

$$R_{Q_1}(n) + R_{Q_2}(n) + \cdots + R_{Q_h}(n) = w \sum_{k|n} \left(\frac{\Delta}{k}\right),$$

where w is the number of units in the ring of integers of $\mathbb{Q}(\sqrt{\Delta})$.

(Recall $\left(\frac{\Delta}{1}\right) = 1$ for any Δ.)

The first version of the mass formula immediately gives

Corollary 4.2.8. *Suppose $\Delta = -4d$ with $d > 1$ squarefree. Then $n > 0$ relatively prime to Δ is properly represented by some form of discriminant Δ if and only if $\left(\frac{\Delta}{p}\right) = 1$ for all primes $p|n$.*

The point is that while it is difficult to study the numbers represented by an *individual* quadratic form in general, it is relatively easy to understand which numbers are represented by *some* quadratic form of discriminant Δ, for fixed Δ. However, if the class number $h(\Delta) = 1$, then the above formulas are in fact formulas for a specific $r_Q(n)$ (resp. $R_Q(n)$).

Example 4.2.9. *Consider $Q(x, y) = x^2+y^2$. This has discriminant $\Delta = -4$, which is the discriminant of the quadratic field $\mathbb{Q}(i)$. If $ax^2+bxy+cy^2$ is a reduced form of discriminant -4, then $3b^2 \leq 4$ (see proof of Proposition 4.2.2), so $b = 0$ or $b = \pm 1$. Clearly $b = \pm 1$ makes $b^2 - 4ac = 1 - 4ac = -4$ unsolvable, so $b = 0$. Then we must have $ac = 1$ so $a = c = 1$ (since we are just working with positive definite forms). In particular the class number $h(-4) = 1$, and $\{Q\}$ is a set of representatives for $\mathcal{C}l(\Delta)$.*

Then the second version of Dirichlet's mass formula reads

$$R_Q(n) = 4 \sum_{k|n} \left(\frac{\Delta}{k}\right),$$

for n odd. If $n = p$ is prime, then for $p \equiv 1 \bmod 4$ we have

$$R_Q(p) = 4 \left\{ \left(\frac{\Delta}{1}\right) + \left(\frac{\Delta}{p}\right) \right\} = 4(1 + 1) = 8$$

and if $p \equiv 3 \bmod 4$ we have

$$R_Q(p) = 4 \left\{ \left(\frac{\Delta}{1}\right) + \left(\frac{\Delta}{p}\right) \right\} = 4(1 - 1) = 0.$$

In other words, $x^2 + y^2$ represents an odd prime p if and only if $p \equiv 1 \bmod 4$. So Dirichlet's mass formula gives Fermat's two square theorem as a special case.

Moreover, it tells us two more things about $x^2 + y^2$. If $p = x_0^2 + y_0^2$ is odd, then $x_0 \neq y_0$ so $(\pm x_0, \pm y_0)$ and $(\pm y_0, \pm x_0)$ are also solutions to $Q(x, y) = p$. This accounts for 8 solutions. Since $R_Q(p) = 8$, this means up to sign and interchanging x and y, $p = x_0^2 + y_0^2$ is the only way to write p as a sum of 2 squares.

Recall Brahmagupta's composition tells us the product of two numbers of the form $x^2 + y^2$ is again of the form $x^2 + y^2$. Since $2 = 1^2 + 1^2$ and $p^2 = p^2 + 0^2$ for any p, we know that n is of the form $x^2 + y^2$ if any $p|n$ with $p \equiv 3 \bmod 4$ occurs to an even power in the prime factorization of n.

In fact, these are the only n represented by $x^2 + y^2$, and we can actually prove this for n odd using the mass formula. Indeed, suppose n is odd and not of the above form, i.e., there is a prime $p \equiv 3 \bmod 4$ dividing n which occurs to an odd power in the prime factorization of n. Let D_1 be the set of divisors k of n such that p occurs to an even power in the prime factorization of k, and let $D_2 = \{pk : k \in D_1\}$. Then $D_1 \cup D_2$ are the divisors of n and D_1 and D_2 are disjoint. So

$$R_Q(n) = 4 \left\{ \sum_{k \in D_1} \left(\frac{\Delta}{k}\right) + \sum_{pk \in D_2} \left(\frac{\Delta}{pk}\right) \right\} = 4 \sum_{k \in D_1} \left\{ \left(\frac{\Delta}{k}\right) + \left(\frac{\Delta}{p}\right)\left(\frac{\Delta}{k}\right) \right\} = 0$$

since $\left(\frac{\Delta}{p}\right) = -1$.

Exercise 4.4. *Determine the reduced forms of discriminant Δ for $\Delta = -3, -8, -12, -20, -24$. In particular, determine $h(\Delta)$ for these Δ.*

Exercise 4.5. *Use Dirichlet's mass formula and Brahmagupta's composition law to determine to which odd numbers are of the form $x^2 + 2y^2$.*

Exercise 4.6. *Use Dirichlet's mass formula, Brahmagupta's composition law and (i) to determine to which numbers prime to 6 are of the form $x^2 + 3y^2$.*

4.3 The form class group

The idea behind Gauss's composition of binary quadratic forms comes from Brahmagupta composition, which says the product of two numbers of the form $x^2 + dy^2$ is again of the form $x^2 + dy^2$. Precisely, Brahmagupta's composition law is

$$(x_1^2 + dy_1^2)(x_2^2 + dy_2^2) = (x_1x_2 - dy_1y_2)^2 + d(x_1y_2 + x_2y_1)^2 = X^2 + dY^2$$

where $X = x_1x_2 - dy_1y_2$, $Y = x_1y_2 + x_2y_1$. Gauss's composition says that if Q_1 and Q_2 are quadratic forms of discriminant Δ, then there is a form Q_3 of the same discriminant such that

$$Q_1(x_1, y_1)Q_2(x_2, y_2) = Q_3(X, Y)$$

where X, Y are some (homogeneous) quadratic expressions in x_1, y_1, x_2 and y_2. In other words, the product of a number of represented by Q_1 with a number represented by Q_2 is represented by Q_3. We will write this composition as

$$Q_1 \circ Q_2 = Q_3$$

and this will make $\mathcal{C}l(\Delta)$ into a finite abelian group.

However, the explicit determination of X, Y and the coefficients of Q_3 in Gauss's composition is rather complicated and we will not describe it explicitly. Instead, we will approach Gauss composition via ideals. But to get a feeling of how this composition can be done without ideal, we will briefly explain Dirichlet's approach to Gauss composition.

Two forms $Q_1(x, y) = a_1x^2 + b_1xy + c_2y^2$ and $Q_2(x, y) = a_2x^2 + b_2xy + c_2y^2$ of discriminant Δ are called **united** if $\gcd(a_1, a_2, \frac{b_1+b_2}{2}) = 1$. If they are united, there exist $B, C \in \mathbb{Z}$ such that $a_1x^2 + b_1xy + c_2y^2 \sim a_1x^2 + Bxy + a_2Cy^2$ and $a_2x^2 + b_2xy + c_2y^2 \sim a_2x^2 + Bxy + a_1Cy^2$. Then the **Dirichlet composition** is defined to be

$$Q_1 \circ Q_2 = a_1a_2x^2 + Bxy + Cy^2.$$

To see that this follows our notion of what composition should be, observe

$$(a_1x^2 + Bxy + a_2Cy^2)(a_2x^2 + Bxy + a_1Cy^2) = a_1a_2X^2 + BXY + CY^2$$

where $X = x_1x_2 - Cy_1y_2$ and $Y = a_1x_1y_2 + a_2x_2y_1 + By_1y_2$. One can check that the latter form has discriminant Δ. Note that Dirichlet composition does not define composition of any two forms of discriminant Δ (only united forms), but it is enough to define a composition (or multiplication) law on the proper equivalences classes $\mathcal{C}l(\Delta)$.

Now we will present the approach to Gauss's composition via ideals. For simplicity we will work with forms whose discriminant Δ is the discriminant of a quadratic field. We say Δ is a **fundamental discriminant** if $\Delta = \Delta_K$ for some quadratic field K.

Exercise 4.7. *Let Δ be the discriminant of $Q(x,y) = ax^2 + bxy + cy^2$. Show if Δ is a fundamental discriminant, then Q is* **primitive**, *i.e.,* $\gcd(a,b,c) = 1$.

We remark that in working with quadratic forms, one often restricts to primitive forms, since any form is just a multiple of a primitive form.

From now on, for the rest of this section **we assume $\Delta < 0$ is a fundamental discriminant**. Let $K = \mathbb{Q}(\sqrt{\Delta})$ be the imaginary quadratic field of discriminant Δ.

Definition 4.3.1. *Let $\mathcal{I}$ be an ideal of $\mathcal{O}_K$ with ordered $\mathbb{Z}$-basis $\{\alpha, \beta\}$. Then the* **quadratic form associated to $\mathcal{I}$** *is*

$$Q_{\mathcal{I}}(x,y) = N(\alpha x + \beta y)/N(\mathcal{I}) = ax^2 + bxy + cy^2.$$

Here the first norm is the norm of the element $\alpha x + \beta y \in \mathcal{O}_K$, and the second is of course the ideal norm. Explicitly we have

$$N(\alpha x + \beta y) = N(\alpha)x^2 + Tr(\alpha\overline{\beta})xy + N(\beta)y^2$$

so $a = N(\alpha)/N(\mathcal{I})$, $b = Tr(\alpha\overline{\beta})/N(\mathcal{I})$ and $c = N(\beta)/N(\mathcal{I})$ in the definition above. Technically, the form $Q_{\mathcal{I}}$ depends upon the choice of an ordered $\mathbb{Z}$-basis for $\mathcal{I}$, but it is not too difficult to see that a different basis will lead to a properly equivalent form. Further $Q_{\mathcal{I}}$ has discriminant Δ.

Example 4.3.2. *Let $\Delta = -4$ so $K = \mathbb{Q}(i)$. Consider the ideals $\mathcal{I} = \langle 1, i\rangle = \mathbb{Z}[i]$ and $\mathcal{J} = \langle 1+i, 1-i\rangle = (1+i)$ of $\mathcal{O}_K$. Then*

$$Q_{\mathcal{I}}(x,y) = N(x+iy)/N(\mathcal{I}) = N(x+iy) = x^2 + y^2$$

and

$$Q_{\mathcal{J}}(x,y) = N((1+i)x + (1-i)y)/N(\mathcal{J}) = (2x^2 + Tr(2i) + 2y^2)/2 = x^2 + y^2.$$

So we see different (but equivalent) ideals may lead to the same form.

Exercise 4.8. *Let $\Delta = -20$ so $K = \mathbb{Q}(\sqrt{-5})$ and consider the ideals $\mathcal{I} = \langle 2, 1+\sqrt{-5}\rangle$ and $\mathcal{J} = \langle 3, 1+\sqrt{-5}\rangle$. Compute $\mathbb{Q}_{\mathcal{I}}$ and $\mathbb{Q}_{\mathcal{J}}$. Check they have discriminant Δ. Are they properly equivalent?*

Definition 4.3.3. *Let $Q(x,y) = ax^2 + bxy + cy^2$ be a form of discriminant Δ. The* **ideal of $\mathcal{O}_K$ associated to Q** *is*

$$\mathcal{I}_Q = (a, \frac{b - \sqrt{\Delta}}{2}).$$

Lemma 4.3.4. *For any form Q, $Q_{\mathcal{I}_Q} = Q$. In other words, if we take the ideal $\mathcal{I}_Q = (a, \frac{b-\sqrt{\Delta}}{2})$ associated to $Q(x,y) = ax^2 + bxy + cy^2$, then the form $N(ax + \frac{b-\sqrt{\Delta}}{2}y)/N(\mathcal{I}_Q)$ associated to $\mathcal{I}_Q$ equals Q.*

This lemma says the association $\mathcal{I} \mapsto Q_{\mathcal{I}}$ is a right inverse to $Q \mapsto \mathcal{I}_Q$. The proof is elementary. However the converse is not true in general. What we can say is the following

Lemma 4.3.5. *Let $\mathcal{I}$ be an ideal of $\mathcal{O}_K$ and let $Q_{\mathcal{I}}$ be the associated form. If $\mathcal{J} = \mathcal{I}_{Q_{\mathcal{I}}}$ is the ideal associated to $Q_{\mathcal{I}}$, then $\mathcal{J} \sim \mathcal{I}$.*

Then we can *define* multiplication of forms $Q_{\mathcal{I}}$ and $Q_{\mathcal{J}}$ associated to ideals by $Q_{\mathcal{I}} \circ Q_{\mathcal{J}} = Q_{\mathcal{I}\mathcal{J}}$. Upon showing the map $\mathcal{C}l_K \to \mathcal{C}l(\Delta)$ induced by $\mathcal{I} \mapsto Q_{\mathcal{I}}$ is surjective, this defines a multiplication on the form class group $\mathcal{C}l(\Delta)$. Precisely, one gets

Theorem 4.3.6. *We have an isomorphism*

$$\begin{aligned} \mathcal{C}l_K &\simeq \mathcal{C}l(\Delta) \\ \mathcal{I} &\rightarrowtail Q_{\mathcal{I}} \\ \mathcal{I}_Q &\leftarrowtail Q. \end{aligned}$$

Proofs may be found in [Cohn]. The proof are not difficult, but we omit them in the interest of time.

Exercise 4.9. *Show $x^2 + \frac{-\Delta}{4}y^2$ is the identity of $\mathcal{C}l(\Delta)$ if $\Delta \equiv 0 \bmod 4$ and $x^2 - xy + \frac{1-\Delta}{4}y^2$ is the identity of $\mathcal{C}l(\Delta)$ if $\Delta \equiv 1 \bmod 4$.*

Exercise 4.10. *Show $Q_2(x,y) = ax^2 - bxy + cy^2$ is the inverse of $Q_1(x,y) = ax^2 + bxy + cy^2$. We know Q_1 and Q_2 are always equivalent by a transformation of determinant -1, namely $(x,y) \mapsto (x,-y)$. Deduce that $Q_1 \sim Q_2$ if and only if Q_1 has order 2 in $\mathcal{C}l(\Delta_{Q_1})$.*

Exercise 4.11. *We say $Q(x,y) = ax^2 + bxy + cy^2$ is* **ambiguous** *if $a|b$. Show if Q is ambiguous, then Q has order 1 or 2 in $\mathcal{C}l(\Delta_Q)$.*

In fact it can be shown that Q has order ≤ 2 in the form class group if and only if $Q \sim Q'$ for some ambiguous form Q'. In this case, the reduced form in the proper equivalence class of Q is either ambiguous (so $b = a$ or $b = 0$) or of the form $ax^2 + bxy + ay^2$.

Exercise 4.12. *Determine all reduced forms of discriminant $\Delta = -84$. Use this to deduce $\mathbb{Q}(\sqrt{-21})$ has class group isomorphic to $V_4 = C_2 \times C_2$.*

4.4 Genus theory

As in the previous section, let K be an imaginary quadratic field of discriminant Δ. Dirichlet's mass formula tells us which numbers are represented by *some* form in $\mathcal{F}_\Delta$, but it doesn't tell us which numbers are represented by a specific form of discriminant Δ. The problem of distinguishing between forms (or rather equivalence classes of forms) of discriminant Δ is not at all a simple problem in general, however there is a simple approach which gives a complete solution when the class group is isomorphic to C_2^r.

To motivate genus theory, let's consider our favorite example.

Example 4.4.1. *Let $\Delta = -20$ so $K = \mathbb{Q}(\sqrt{-5})$. The reduced forms of discriminant Δ are $Q_1(x,y) = x^2 + 5y^2$ and $Q_2(x,y) = 2x^2 + 2xy + 3y^2$. Hence $\mathcal{C}l_K \simeq \mathcal{C}l(\Delta)$ has order 2, so must be isomorphic to C_2. From the exercises in the previous section, Q_1 is the identity and Q_2 has order 2 in $\mathcal{C}l(\Delta)$.*

First let us determine the primes represented by Q_1 and Q_2. By the mass formula (first version) we know

$$R_{Q_1}(p) + R_{Q_2}(p) = r_{Q_1}(p) + r_{Q_2}(p) = 2(1 + \left(\frac{\Delta}{p}\right)) = \begin{cases} 4 & p \equiv 1,3,7,9 \bmod 20 \\ 0 & p \equiv 11,13,17,19 \bmod 20 \end{cases}$$

for $p \nmid \Delta$. (Note for p prime, $R_Q(p) = r_Q(p)$ for any form Q.) Note that if $R_{Q_1}(p) > 0$ for $p \nmid 20$, then $R_{Q_1}(p) \geq 4$ because if (x, y) is a solution to $x^2 + 5y^2 = p$, then so are $(\pm x, \pm y)$, which gives 4 different solutions as long as $p \neq 5$. In other words, any $p \equiv 1, 3, 7, 9 \bmod 20$ is represented either by Q_1 or Q_2, but not both.

However, looking at what numbers relatively prime to 20 are of the form $x^2 + 5y^2$ mod 20 we see 3 and 7 are not possible. Similarly, simple computations show that 1 and 9 are not of the form $2x^2 + 2xy + 3y^2$ mod 20. (In fact, it suffices to observe $x^2 + 5y^2$ does not represent $3 \bmod 4$ and $2x^2 + 2xy + 3y^2$ does not represent $1 \bmod 4$.) Hence we have

$$p \text{ is represented by } \begin{cases} Q_1 & \iff p = 5 \text{ or } p \equiv 1, 9 \bmod 20 \\ Q_2 & \iff p = 2 \text{ or } p \equiv 3, 7 \bmod 20 \end{cases}$$

Now we can ask what integers $n > 0$ are represented by Q_1. By the mass formula (first form), we know if Q_1 represents n, then n cannot be divisible by any prime p such that $\left(\frac{\Delta}{p}\right) = -1$, i.e., any $p \equiv 11, 13, 17, 19 \bmod 20$. Write $n = \prod p_i^{e_i} \cdot \prod q_j^{f_j}$ where each p_i is represented by Q_1 and each q_j is represented by Q_2. Gauss's composition says n is represented by $\prod_i Q_1^{e_i} \cdot \prod_j Q_2^{f_j}$ (where the multiplication here denotes Gauss composition). In other words, n is represented by Q_1 if $\sum f_j$ is even and n is represented by Q_2 if $\sum f_j$ is odd.

We would like to say the above statement about which n's are represented by Q_1 and which are represented by Q_2 is if and only if. Note that Q_1 represents $0, 1, 2 \bmod 4$ and $0, 1, 4 \bmod 5$, where as Q_2 represents $0, 2, 3 \bmod 4$ and $0, 2, 3 \bmod 5$. The only overlap here are the numbers $\equiv 0, 2 \bmod 4$ and $\equiv 0 \bmod 5$. Hence Q_1 and Q_2 cannot *represent the same numbers, except possibly for numbers divisible by 10. The case where n is not prime to the discriminant is more subtle, and we will not prove this, but it turns out Q_1 and Q_2 never represent the same numbers, so the above characterization of numbers represented by Q_1 (or Q_2) is if and only if.*

Genus theory allow us to generalize the above example to separate (at least partially) representations by different forms of discriminant Δ.

Definition 4.4.2. *Let $Q_1, Q_2 \in \mathcal{F}_\Delta$. We say Q_1 and Q_2 are in the same* **genus** *if they represent the same values mod Δ. The* **principal genus** *is the genus containing the identity of the form class group.*

It is a theorem that Q_1 and Q_2 are in the same genus if and only if Q_1 and Q_2 represent the same values mod m for every m. What is more important for us however, is that forms in different genera (the plural of genus) represent disjoint sets of numbers in $(\mathbb{Z}/\Delta\mathbb{Z})^\times$. This is the content of the following proposition.

Proposition 4.4.3. *Regard $\chi_\Delta = \left(\frac{\Delta}{\cdot}\right)$ as a real character of $(\mathbb{Z}/\Delta\mathbb{Z})^\times$. Let $H = \ker \chi_\Delta$. Let Q_0 (resp. Q) be in the prinicipal genus (resp. any genus) of $\mathcal{F}_\Delta$ and H_0 (resp. H_Q) denote the set of values in $(\mathbb{Z}/\Delta\mathbb{Z})^\times$ represented by Q_0 (resp Q). Then H_0 is a subgroup of H and H_Q is a coset of H_0 in H.*

(Note that H being the kernel of a group homomorphism, is a subgroup of $(\mathbb{Z}/\Delta\mathbb{Z})^\times$.)

For instance, in the above example, with $Q_0 = x^2 + 5y^2$ and $Q = 2x^2 + 2xy + 3y^2$, we have $H = \{1, 3, 7, 9\} \subseteq (\mathbb{Z}/20\mathbb{Z})^\times$, $H_0 = \{1, 9\}$ and $H_Q = \{3, 7\} = 3\{1, 9\}$.

Proof. Let $n \in (\mathbb{Z}/\Delta\mathbb{Z})^\times$. If n is represented by a form of discriminant Δ, we have $n \in H = \ker \chi_\Delta$ by Dirichlet's mass formula. To see that H_0 is a subgroup of H, observe it must be closed under multiplication by Gauss composition. To show it is closed under inversion, note by Exercise 4.9, we can assume $Q_0 = x^2 - \frac{\Delta}{4}y^2$ or $x^2 + xy + \frac{1-\Delta}{4}y^2$. Using either Brahmagupta or Dirichlet composition, it is straightforward to explicitly check H_0 is closed under inverses (and is nonempty—it contains 1).

It follows from Gauss composition that H_Q must be a translate of H_0 in H. □

Since the cosets of H_0 in H are disjoint, the integers n relatively prime to Δ can only be represented by forms in a single genus of $\mathcal{F}_\Delta$. In particular, if we want to determine which numbers are of the form $x^2 + dy^2$ (say relatively prime to $\Delta = -4d$), we can at least say n are represented by some form in the principal genus. In particular, if, up to equivalence, $x^2 + dy^2$ is the only form in the principal genus, we can say exactly which primes are represented by $x^2 + dy^2$ by (i) the mass formula and (ii) considerations mod Δ. In this case, we say Δ has **one class per genus** (see exercise below).

Exercise 4.13. *Let $Q_1, \ldots, Q_h$ denote representatives for $\mathcal{C}l(\Delta)$. Using Gauss composition, show the number of Q_i in a given genus is the same for each genus.*

Exercise 4.14. *Pick representatives Q_1, Q_2 for $\mathcal{C}l(-24)$. Determine what values Q_1 and Q_2 represent mod 24. Using this with Dirichlet's mass formula, determine all primes represented by Q_1 and all primes represented by Q_2. In particular, you should get a determination of all primes of the form $x^2 + 6y^2$.*

Now of course it's natural to ask, for which discriminants Δ do we have one class per genus? It's clearly true if the class number $h(\Delta) = 1$. We know there are only 9 fundamental discriminants $\Delta < 0$ with class number 1 (Gauss's class number problem), and this was easy to determine. Conversely, it is still an unsettled problem (also posed by Gauss) for which Δ have one class per genus. It is conjectured that there are precisely 65 fundamental discriminants (and 101 arbitrary negative discriminants) $\Delta < 0$ with one class per genus. It is not too difficult to show the following.

Theorem 4.4.4. *The principal genus of $\mathcal{C}l(\Delta)$ consists of the subgroup of squares of $\mathcal{C}l(\Delta)$.*

Corollary 4.4.5. *Δ has one class per genus if and only if $\mathcal{C}l(\Delta) \simeq C_2^r$ for some $r \geq 0$.*

We remark that for a specific r, we can compute all imaginary quadratic fields with class group C_2^r. There shouldn't be any for large enough r, and this is the most difficult part.

In the case of one class per genus, one can always determine the primes of the form $x^2 + dy^2$ by simple congruence conditions. However, at least conjecturally, this only happens finitely many times (for negative Δ). In the rest of the cases, the determination of primes of the form $x^2 + dy^2$ is more complicated.

Example 4.4.6. *Consider $Q_0 = x^2 + 14y^2$. This has discriminant $\Delta = -56$ and corresponds to the field $K = \mathbb{Q}(\sqrt{-14})$. There are 3 other reduced forms of discriminant -56, given by $Q_1 = 2x^2 + 7y^2$, $Q_2 = 3x^2 + 2xy + 5y^2$ and $Q_3 = 3x^2 - 2xy + 5y^2$. It is easy to check the (form) class group $\mathcal{C}l(\Delta) \simeq C_4$ (see exercise below). One can show $p = x^2 + 14y^2$ if and only if (i) $\left(\frac{-14}{p}\right) = 1$, and (ii) $(x^2+1)^2 \equiv 8 \bmod p$ has a solution. See [Cox]. We will discuss this briefly in the next chapter.*

Exercise 4.15. *Check that $Q_2 \circ Q_2 \sim Q_3 \circ Q_3 \sim Q_1$. Conclude $\mathcal{C}l(-56) \simeq C_4$.*