2

p-adic numbers

Most of the familiar properties of the ordinary absolute value on the real or complex fields are consequences of the following three:

(i) $|r| \geq 0$, with equality precisely for $r = 0$.
(ii) $|rs| = |r||s|$.
(iii) $|r+s| \leq |r| + |s|$.

A real-valued function $|.|$ on a field k is said to be a *valuation* if it satisfies (i), (ii) (iii). Since $(-1)^2 = 1$, properties (i)-(iii) imply that $|-1| = 1$, $|-r| = |r|$ (all r).

The rational field $\mathbb{Q}$ has other valuations than the absolute value. Let p be a fixed prime. Any rational $r \neq 0$ can be put in the shape

$$r = p^\rho u/v, \ \rho \in \mathbb{Z}, \ u, \ v \in \mathbb{Z}, \ p \nmid u, \ p \nmid v.$$

We define

$$|r|_p = p^{-\rho}$$

and

$$|0|_p = 0.$$

This definition clearly satisfies (i), (ii) above. Let

$$s = p^\sigma m/n \qquad m, \ n \in \mathbb{Z}, \ p \nmid m, \ p \nmid n,$$

so

$$|s|_p = p^{-\sigma},$$

where without loss of generality

$$\sigma \geq \rho, \text{ i.e. } |s|_p \leq |r|_p.$$

Then

$$r + s = p^\rho(un + p^{\sigma-\rho}mv)/vn.$$

Here $p \nmid vn$. The numerator $un + p^{\sigma-\rho}mv$ is an integer, but, at least for for $\rho = \sigma$, it may be divisible by p. Hence

$$|r+s|_p \leq p^{-\rho},$$

that is

(iii*) $|r+s|_p \leq \max\{|r|_p, \ |s|_p\}$.

Clearly (iii*) implies (iii), so $| \ |_p$ is a valuation. We call it the *p-adic valuation.* The inequality (iii*) is called the ***ultrametric inequality***, since (iii), the *triangle inequality*, expresses the fact that $|r-s|$ is a metric. A valuation which satisfies the ultrametric inequality is said to be *non-archimedean.*

We can transfer familiar terminology from the ordinary absolute value to the p-adic case. For example, we say that a sequence $\{a_n\}$, $n = 1, 2, \ldots$ is a *fundamental sequence* if for any $\varepsilon > 0$ there is an $n_0(\varepsilon)$ such that

$$|a_m - a_n|_p < \varepsilon \qquad \text{whenever} \qquad m, n \geq n_0(\varepsilon).$$

The sequence $\{a_n\}$ *converges* to b if

$$|a_n - b|_p < \varepsilon \qquad (\text{all } n \geq n_0(\varepsilon)).$$

For example let

$$p = 5$$

and consider the sequence

$$\{a_n\}: \quad 3, \quad 33, \quad 333, \quad 3333, \quad \ldots .$$

Then

$$a_m \equiv a_n \qquad \mod 5^n \qquad (m \geq n)$$

i.e.

$$|a_m - a_n|_5 \leq 5^{-n} \qquad (m \geq n).$$

Hence $\{a_n\}$ is a fundamental sequence. Indeed it is a convergent sequence, since

$$3a_n = 99\ldots99 \equiv -1(5^n),$$

i.e.

$$|3a_n + 1|_5 \leq 5^{-n}$$

and so

$$a_n \quad \to -1/3$$

5-adically.

As the above example shows, the main difficulties with the p-adic valuation are psychological: something is p-adically small if it is divisible by a high power of p. Not every p-adic fundamental sequence is convergent. Let us take $p = 5$ again. Then we construct a sequence of $a_n \in \mathbb{Z}$ such that

$$a_n^2 + 1 \equiv 0 \ (5^n)$$

and

$$a_{n+1} \equiv a_n \ (5^n).$$

We start with $a_1 = 2$. Suppose that we already have a_n for some n and put $a_{n+1} = a_n + b5^n$, where $b \in \mathbb{Z}$ is to be determined. We require

$$(a_n + b5^n)^2 + 1 \equiv 0 \ (5^{n+1}),$$

that is

$$2a_n b + c \equiv 0 \ (5), \tag{*}$$

where we already have

$$c = (a_n^2 + 1)/5^n \in \mathbb{Z}.$$

Clearly $5 \nmid a_n$ and so we can solve the congruence (*) for the unknown b.

The sequence $\{a_n\}$ just constructed is a 5-adic fundamental sequence since

$$|a_m - a_n|_5 \le 5^{-n} \qquad (m \ge n).$$

Suppose, if possible, that a_n tends 5-adically to some $e \in \mathbb{Q}$. Then

$$a_n^2 + 1 \to e^2 + 1.$$

On the other hand, by our construction,

$$a_n^2 + 1 \to 0.$$

Hence $e^2 + 1 = 0$; a contradiction.

Just as the real numbers are constructed by completing the rationals with respect to the ordinary absolute value, so the rationals can be completed with respect to $| \ |_p$ to give the field $\mathbb{Q}_p$ of *p-adic numbers.* In fact the process can be simplified because $| \ |_p$ is non-archimedean. For the reader who is unfamiliar with this way of constructing the reals, we sketch a construction of $\mathbb{Q}_p$ at the end of this section.

We say that a field K is *complete* with respect to a valuation $|.|$ if every fundamental sequence is convergent. A field K with valuation $||.||$ is said to be the *completion* of the field k with valuation $|.|$ if there is an injection

$$\lambda : k \to K$$

which preserves the valuation:

$$||\lambda a|| = |a| \qquad (a \in k)$$

and such that

(i) K is complete with respect to $||.||$
(ii) K is the closure of λk with respect to the topology induced by $||.||$ (K is not "too large").

The completion always exists and is unique (up to a unique isomorphism). We henceforth identify k with λk and $|.|$ with $||.||$, so regard k as a subfield of K.

We now discuss the structure of the p-adic field $\mathbb{Q}_p$ with its valuation $| \ |_p$.

We note that

$$|a + b|_p = |a|_p \qquad \text{if} \qquad |b|_p < |a|_p.$$

For by (iii*) $|a + b|_p \le |a|_p$ and, since $a = (a + b) + (-b)$, we have a contradiction if $|a + b|_p < |a|_p$. It follows that the set of values taken by $| \ |_p$ on $\mathbb{Q}_p$ is precisely the same as the set for $\mathbb{Q}$. Indeed if $\alpha \in \mathbb{Q}_p$, $\alpha \ne 0$ then by (ii) of the definition of the completion, there is an $a \in \mathbb{Q}$ with $|a - \alpha|_p < |\alpha|_p$, so $|\alpha|_p = |a|_p$.

The set of $\alpha \in \mathbb{Q}_p$ with $|\alpha| \le 1$ is called the set of *p-adic integers* $\mathbb{Z}_p$. Because $| \ |_p$ is non-archimedean, $\mathbb{Z}_p$ is a ring:

$$|\alpha|_p, \ |\beta|_p \le 1 \Rightarrow |\alpha\beta|_p \le 1, \ |\alpha + \beta|_p \le 1.$$

A rational number b is in $\mathbb{Z}_p$ precisely when it has the form $b = u/v$, where $u, v \in \mathbb{Z}$, $p \nmid v$.

The numbers $\varepsilon \in \mathbb{Q}_p$ with $|\varepsilon| = 1$ are the *p-adic units.* From what was said about the values taken by $|.|_p$ on $\mathbb{Q}_p$, every $\beta \ne 0$ in $\mathbb{Q}_p$ is of the shape $\beta = p^n \varepsilon$, where $n \in \mathbb{Z}$ and ε is a unit. The units are just the elements ε of $\mathbb{Q}_p$ such that $\varepsilon \in \mathbb{Z}_p$, $\varepsilon^{-1} \in \mathbb{Z}_p$.

As we have already noted, elementary analysis continues to hold in $\mathbb{Q}_p$, but can be simpler; as the following lemma shows.

Lemma 1. *In $\mathbb{Q}_p$ the series $\sum_0^\infty \beta_n$ converges if and only if $\beta_n \to 0$.*

Proof. By saying that the sum converges, we mean, of course, that the partial sums $\sum_0^N$ tend to a limit.

That convergence implies $\beta_n \to 0$ is true even in real analysis. To prove the opposite implication, we note that

$$| \sum_0^N - \sum_0^M |_p = | \sum_{M+1}^N \beta_n |_p$$
$$\leq \max_{M<n\leq N} |\beta_n|_p$$

by an obvious extension of the ultrametric inequality (iii*) to several summands. Hence $\left\{ \sum_0^N \beta_n \right\}$ is a fundamental sequence, so tending to a limit by the completeness of $\mathbb{Q}_p$.

We are now in a position to give an explicit description of $\mathbb{Z}_p$. We write

$$\mathcal{A} = \{0, 1, \ldots, p-1\}.$$

Lemma 2. *The elements of $\mathbb{Z}_p$ are precisely the sums*

$$\alpha = \sum_0^\infty a_n p^n,$$

where

$$a_n \in \mathcal{A} \qquad (\text{all } n).$$

Proof. By the preceeding lemma, the infinite sum converges, and its value is clearly in $\mathbb{Z}_p$.

Now let $\alpha \in \mathbb{Z}_p$ be given. There is a $b \in \mathbb{Q}$ such that $|b - \alpha|_p < 1$, and it is easy to prove that there is precisely one $a_0 \in \mathcal{A}$ such that $|a_0 - b|_p < 1$. Then

$$\alpha = a_0 + p\alpha_1$$

where $|\alpha_1| \leq 1$, i.e. $\alpha_1 \in \mathbb{Z}_p$. Proceeding inductively, we get

$$\alpha = a_0 + a_1 p + \ldots + a_N p^N + \alpha_N p^{N+1}$$

with $\alpha_N \in \mathbb{Z}_p$.

For the final result we must distinguish between $p = 2$ and $p \neq 2$.

Lemma 3 ($p \neq 2$). *Let $\alpha \in \mathbb{Q}_p$ be a unit. A necessary and sufficient condition that $\alpha = \beta^2$ for some $\beta \in \mathbb{Q}_p$ in that there is some $\gamma \in \mathbb{Q}_p$ with*

$$|\alpha - \gamma^2|_p < 1.$$

Proof. Necessity is obvious. We have already in effect given a proof in the special case $p = 5$, $\alpha = -1$. That in the general case is similar: one constructs inductively $\beta_1 = \gamma$, β_2, $\beta_3, \ldots$ such that

$$|\beta_n^2 - \alpha| \leq p^{-n}$$
$$|\beta_{n+1} - \beta_n| \leq p^{-n}$$

If we already have β_n, we take $\beta_{n+1} = \beta_n + \delta$, so

$$\beta_{n+1}^2 = \beta_n^2 + 2\beta_n\delta + \delta^2$$

and it is enough to take

$$\delta = (\alpha - \beta_n^2)/2\beta_n.$$

This lemma ceases to hold for $p = 2$ (consider $\alpha = 5$, $\beta = 1$). We have

Lemma 4 ($p = 2$). *Let $\alpha \in \mathbb{Q}_2$ be a unit. A necessary and sufficient condition that $\alpha = \beta^2$ for some $\beta \in \mathbb{Q}_2$ is that $|\alpha - 1| \leq 2^{-3}$.*

Proof. Here again, the necessity is obvious. For sufficiency we construct a sequence $\beta_1 = 1$, β_2, $\beta_3, \ldots$ as in the previous proof. The details are left to the reader.

We conclude this section by the promised sketch of the construction of $\mathbb{Q}_p$.

Denote by $\mathfrak{F}$ the set of fundamental sequences $\{a_n\}$ for $|\ |_p$, where $a_n \in \mathbb{Q}$. Then $\mathfrak{F}$ is a ring under componentwise addition and multiplication.

$$\{a_n\} + \{b_n\} = \{a_n + b_n\} : \{a_n\}\{b_n\} = \{a_n b_n\}.$$

A sequence $\{a_n\}$ is a null sequence if $a_n \to 0$ (p-adically). The set $\mathfrak{N}$ of null-sequences is clearly an ideal in $\mathfrak{F}$.

Let $\{a_n\} \in \mathfrak{F}$ but $\{a_n\} \notin \mathfrak{N}$. Then it is easy to see that there is at least one N such that $|a_N - a_n| < |a_N|_p$ for all $n > N$. Then $|a_n|_p = |a_N|_p$ for all $n \geq N$. We write $|\{a_n\}|_p = |a_N|_p$. If $a_n \neq 0$ for all n, it is now easy to deduce that $\{a_n^{-1}\} \in \mathfrak{F}$.

We show that $\mathfrak{N}$ is a maximal ideal in $\mathfrak{F}$. For, if not, let $\mathfrak{M}$ be a strictly bigger ideal than $\mathfrak{N}$. It must contain an $\{a_n\} \notin \mathfrak{N}$. Then only finitely many of the a_n can be 0, and replacing them by (say) 1 merely adds an element of $\mathfrak{N}$. Hence we can suppose that $a_n \neq 0$ for all n. Then $\{a_n^{-1}\} \in \mathfrak{F}$, and so $\{a_n^{-1}\}\{a_n\} \in \mathfrak{M}$. Hence we should have $\mathfrak{M} = \mathfrak{F}$, a contradiction. We conclude that $\mathfrak{N}$ is maximal, and thus $\mathfrak{F}/\mathfrak{N}$ is a field.

The field $\mathbb{Q}$ is mapped into $\mathfrak{F}/\mathfrak{N}$ by

$$r \to \{r\} \in \mathfrak{F}.$$

The function $|\{a_n\}|$ on $\mathfrak{F}$ induces a function on $\mathfrak{F}/\mathfrak{N}$ which is easily seen to be a valuation and to coincide with $|\ |_p$ on the image of $\mathbb{Q}$.

Finally, it is not difficult to check that $\mathfrak{F}/\mathfrak{N}$ is itself complete by a diagonal argument on a sequence of elements of $\mathfrak{F}$.

§2. Exercises

1. For each of the sets of p, m, r given, either find an $x \in \mathbf{Z}$ such that

$$|r - x|_p \leq p^{-m},$$

or show that no such x exists.

(i) $p = 257$, $r = 1/2$, $m = 1$;
(ii) $p = 3$, $r = 7/8$, $m = 2$;
(iii) $p = 3$, $r = 7/8$, $m = 7$;
(iv) $p = 3$, $r = 5/6$, $m = 9$;
(v) $p = 5$, $r = 1/4$, $m = 4$.

2. Construct further examples along the lines of Exercise 1 until the whole business seems trivial.

3. For given p, m, r either find an $x \in \mathbf{Z}$ such that

$$|r - x^2|_p \leq p^{-m}$$

or show that no such x exists.

(i) $p = 5$, $r = -1$, $m = 4$;
(ii) $p = 5$, $r = 10$, $m = 3$;
(iii) $p = 13$, $r = -4$, $m = 3$;
(iv) $p = 2$, $r = -7$, $m = 6$;
(v) $p = 7$, $r = -14$, $m = 4$;
(vi) $p = 7$, $r = 6$, $m = 3$;
(vii) $p = 7$, $r = 1/2$, $m = 3$.

4. As Exercise 2.

5. Let $p > 0$ be prime, $p \equiv 2$ (3). For any integer a, $p \not| a$, show that there is an $x \in \mathbf{Z}_p$ with $x^3 = a$.

3

The local-global principle for conics

We have seen that the theory of curves of genus 0 over $\mathbf{Q}$ turns on deciding whether a given conic has a rational point.

We use homogeneous co-ordinates. A conic C defined over $\mathbf{Q}$ is given by an equation

$$F(\mathbf{X}) = \sum f_{ij} X_i X_j = 0$$

where $\mathbf{X} = (X_1, X_2, X_3)$,

$$f_{ij} = f_{ji} \in \mathbf{Q}$$

and the quadratic form F (recall a *form* is a homogeneous polynomial) is nonsingular, i.e.

$$\det(f_{ij}) \neq 0.$$

In our initial discussion we noted that, apart from reality considerations, we could disprove the existence of rational points by congruence considerations. These we now replace by reference to p-adic numbers.

A criterion for the existence of a rational point on a conic was given by Legendre. It was left to Hasse to give it the following succinct formulation.

Theorem 1. *A necessary and sufficient condition for the existence of a rational point on a conic C defined over $\mathbf{Q}$ is that there is a point defined over the real field $\mathbf{R}$ and over $\mathbf{Q}_p$ for every prime p.*

Necessity is trivial. We shall prove sufficiency, but it will require some time and preparation. First we introduce some conventional terminology.

The real field $\mathbf{R}$ is somewhat analogous to the $\mathbf{Q}_p$ and is conventionally denoted by $\mathbf{Q}_\infty$. When we write $\mathbf{Q}_p$ we will not include $p = \infty$ unless we explicitly say so. The fields $\mathbf{Q}_p$ (including $p = \infty$) are called the localizations of $\mathbf{Q}$. In contrast, $\mathbf{Q}$ is called the global field. We say that something is true "everywhere locally" if it is true for all $\mathbf{Q}_p$ (including ∞). In this lingo the theorem becomes "A necessary and sufficient condition for the existence of a global point on a conic is that there should be a point everywhere locally".

The local-global theorem for conics implies a local-global theorem for curves of genus 0 but some care must be taken in the formulation ["point" must be interpreted as "place"]. We do not pursue this further.

In the rest of this section we transform the theorem into a shape better suited for attack[1].

A transformation

$$T:\ X_i = \sum_i t_{ij} Y_j$$

with

$$t_{ij} \in \mathbf{Q}, \qquad \det(t_{ij}) \neq 0$$

takes the quadratic form $F(\mathbf{X})$ into a quadratic form $G(\mathbf{Y})$, say. Then T takes points defined over $\mathbf{Q}$ on $F(\mathbf{X}) = 0$ into points defined over $\mathbf{Q}$ on $G(\mathbf{Y}) = 0$ and, similarly, the inverse T^{-1} takes points on $G(\mathbf{Y}) = 0$ to points on $F(\mathbf{X}) = 0$. Likewise for points defined over $\mathbf{Q}_p$ for each p (including ∞). Hence the theorem holds for $F(\mathbf{X}) = 0$ if and only if it holds for $G(\mathbf{Y}) = 0$.

By suitable choice of transformation T we thus need consider only "diagonal" forms

$$F(\mathbf{X}) = f_1 X_1^2 + f_2 X_2^2 + f_3 X_3^2.$$

By substitutions $X_j \to t_j X_j$ ($t_j \in \mathbf{Q}$) we may suppose without loss of generality that the

$$f_j \in \mathbf{Z}$$

are square free.

If f_1, f_2, f_3 have a prime factor p in common, we replace $F(\mathbf{X})$ by $p^{-1}F(\mathbf{X})$. If two of the f_j, say f_1, f_2 have a prime p in common but $p \not| f_3$, we replace X_3 by pX_3 and then divide F by p. Both of these transformations reduce the absolute value of the integer $f_1 f_2 f_3$. After a finite number of steps we are reduced to the case when $f_1 f_2 f_3$ is square free. We have thus proved the

Metalemma 1. *To prove the Theorem, it is enough to prove it for conics*

$$F(\mathbf{X}) = f_1 X_1^2 + f_2 X_2^2 + f_3 X_3^2 = 0,$$

where $f_j \in \mathbf{Z}$ and $f_1 f_2 f_3$ is square free.

The next stage is to draw conclusions from the hypothesis that a conic as described in the Metalemma has points everwhere locally. There is a point defined over $\mathbf{Q}_p$ when there is a vector $\mathbf{a} = (a_1, a_2, a_3) \neq (0,0,0)$ with $a_j \in \mathbf{Q}_p$ such that $F(\mathbf{a}) = 0$. By multiplying the a_j by an element of $\mathbf{Q}_p$ we may suppose without loss of generality that

$$\max |a_j|_p = 1. \qquad (*)$$

For our later purposes we have to consider several cases.

First case. $p \neq 2$, $p \mid f_1 f_2 f_3$. Without loss of generality $p \mid f_1$, so $p \not| f_2$, $p \not| f_3$. Then $|f_1 a_1^2|_p < 1$. Suppose, if possible that $|a_2|_p < 1$. Then

$$|f_3 a_3^2|_p = |f_1 a_1^2 + f_2 a_2^2|_p < 1$$

and $|a_3|_p < 1$. Now

$$|f_1 a_1^2|_p = |f_2 a_2^2 + f_3 a_3^2|_p \leq p^{-2}$$

and so $|a_1|_p < 1$ since f_1 is square free. This contradicts the normalization (*), and so $|a_2|_p = |a_3|_p = 1$. But now

$$|f_2 a_2^2 + f_3 a_3^2|_p < 1.$$

On dividing by the unit a_2, we deduce that there is some $r_p \in \mathbf{Z}$ such that

$$f_2 + r_p^2 f_3 \equiv 0\ (p).$$

Second case. $p = 2$, $2 \not| f_1 f_2 f_3$. It is easy to see that precisely two of the a_j are units, say a_2 and a_3. Now $a^2 \equiv 1$ or 0 (4) for $a \in \mathbf{Z}$; and so

$$f_2 + f_3 \equiv 0\ (4).$$

Third case. $p = 2$, $2 \mid f_1 f_2 f_3$, say $2 \mid f_1$. Now $|a_2|_2 = |a_3|_3 = 1$. Now $a^2 \equiv 1$ (8) for $a \in \mathbf{Z}$, $2 \not| a$; and so

$$f_2 + f_3 \equiv 0\ (8)$$

or

$$f_1 + f_2 + f_3 \equiv 0\ (8)$$

[1] The details of the proof of Theorem 1 will not be required for the treatment of elliptic curves. The reader who is interested only in the latter should omit the rest of this § and also omit §§4,5.

according as $|a_1|_2 < 1$ or $|a_1|_2 = 1$.

In the next two sections, we show that the conditions just derived are sufficient to ensure the existence of a global point on $F(\mathbf{X}) = 0$.

§3. Exercises

1. (i) Let $p > 2$ be prime and let $b, c \in \mathbb{Z}$, $p \not| b$. Show that $bx^2 + c$ takes precisely $\frac{1}{2}(p+1)$ distinct values p for $x \in \mathbb{Z}$. (ii) Suppose that, further, $a \in \mathbb{Z}$, $p \not| a$. Show that there are $x, y \in \mathbb{Z}$ such that $bx^2 + c \equiv ay^2$ (p).

2. Let $a, b, c \in \mathbb{Z}_p$, $|a|_p = |b|_p = |c|_p = 1$ where p is prime, $p > 2$. Show that there are $x, y \in \mathbb{Z}_p$ such that $bx^2 + c = ay^2$.

3. Let $p > 2$ be prime, $a_{ij} \in \mathbb{Z}$ $(1 \le i, j \le 3)$, $a_{ji} = a_{ij}$ and let $d = \det(a_{ij})$. Suppose that $p \not| d$. Show that there are $x_1, x_2, x_3 \in \mathbb{Z}$, not all divisible by p, such that $\sum_{i,j} a_{ij} x_i x_j \equiv 0$ (p).

4. Let $a, b, c \in \mathbb{Z}$, $2 \not| abc$. Show that a necessary and sufficient condition that the only solution in $\mathbb{Q}_2$ of $ax^2 + by^2 + cz^2 = 0$ is the trivial one is that $a \equiv b \equiv c$ (4).

5. For each of the following sets of a, b, c find the set of primes p (including ∞) for which the only solution of $ax^2 + by^2 + cz^2 = 0$ in $\mathbb{Q}_p$ is the trivial one:

(i) $(a, b, c) = (1, 1, -2)$
(ii) $(a, b, c) = (1, 1, -3)$
(iii) $(a, b, c) = (1, 1, 1)$
(iv) $(a, b, c) = (14, -15, 33)$

6. Do you observe anything about the parity of the number N of primes (including ∞) for which there is insolubility? If not, construct similar exercises and solve them until the penny drops.

7.(i) Prove your observation in (6) in the special case $a = 1$, $b = -r$, $c = -s$, where r, s are distinct primes > 2.
[Hint. Quadratic reciprocity]
(ii) [Difficult]. Prove your observation for all $a, b, c \in \mathbb{Z}$.

4

Geometry of numbers

At this stage we require a tool from the Geometry of Numbers, which we shall develop from scratch.

A generalization of the pigeon-hole principle (Schubfachprinzip) says that if we have N things to file in H holes and $N > mH$ for an integer m, then at least one of the holes will contain $\ge (m+1)$ things. We start with a continuous analogue.

Let $\mathbb{R}^n$ denote the vector space of real n-tuples $\mathbf{r} = (r_1, \dots, r_n)$. It contains the group $\mathbb{Z}^n$ of $\mathbf{r}$ for which $r_j \in \mathbb{Z}$ (all j). By the volume $V(\mathcal{S})$ of a set $\mathcal{S} \subset \mathbb{R}^n$ we shall mean its Lebesgue measure, but in the applications we will be concerned only with very simple-minded $\mathcal{S}$.

Lemma 1. *Let $m > 0$ be an integer and let $\mathcal{S} \subset \mathbb{R}^n$ with*

$$V(\mathcal{S}) > m.$$

Then there are $m + 1$ distinct points $\mathbf{s}_0, \dots, \mathbf{s}_m$ of $\mathcal{S}$ such that

$$\mathbf{s}_i - \mathbf{s}_j \in \mathbb{Z}^n \qquad (0 \le i, j \le m).$$

Proof. Let $\mathcal{W} \subset \mathbb{R}^n$ be the "unit cube" of points $\mathbf{w}$ with

$$0 \le w_j < 1 \qquad (1 \le j \le n).$$

Then every $\mathbf{x} \in \mathbb{R}^n$ is uniquely of the shape

$$\mathbf{x} = \mathbf{w} + \mathbf{z},$$

where $\mathbf{z} \in \mathbb{Z}^n$. Let $\psi(\mathbf{x})$ be the characteristic function of $\mathcal{S}$ ($= 1$ if $\mathbf{x} \in \mathcal{S}$,

$= 0$ otherwise). Then

$$m < V(\mathcal{S}) = \int_{\mathbf{R}^n} \psi(\mathbf{x})d\mathbf{x}$$

$$= \int_{\mathcal{W}} \left(\sum_{\mathbf{z}\in\mathbf{Z}^n} \psi(\mathbf{w}+\mathbf{z}) \right) d\mathbf{w}.$$

Since $V(\mathcal{W}) = 1$, there must be some $\mathbf{w}_0 \in \mathcal{W}$ such that

$$\sum_{\mathbf{z}\in\mathbf{Z}^n} \psi(\mathbf{w}_0 + \mathbf{z}) > m,$$

$$\text{so} \quad \geq m+1.$$

We may now take for the $\mathbf{s}_j$ the $\mathbf{w}_0 + \mathbf{z}$ for which $\psi(\mathbf{w}_0 + \mathbf{z}) > 0$.

The set $\mathcal{S}$ is said to be *symmetric* (about the origin) if $-\mathbf{x} \in \mathcal{S}$ whenever $\mathbf{x} \in \mathcal{S}$. It is *convex* if whenever $\mathbf{x}$, $\mathbf{y} \in \mathcal{S}$, then the whole line-segment

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \mathcal{S} \qquad (0 \leq \lambda \leq 1)$$

joining them is in $\mathcal{S}$. In particular, the mid-point $\frac{1}{2}(\mathbf{x}+\mathbf{y})$ is in $\mathcal{S}$.

Theorem 1. *Let Λ be a subgroup of $\mathbf{Z}^n$ of index m. Let $\mathcal{C} \subset \mathbf{R}^n$ be a symmetric convex set of volume*

$$V(\mathcal{C}) > 2^n m.$$

Then $\mathcal{C}$ and Λ have a common point other than $\mathbf{0} = (0, \ldots, 0)$.

Proof. Let $\mathcal{S} = \frac{1}{2}\mathcal{C}$ be the set of points $\frac{1}{2}\mathbf{c}$, $\mathbf{c} \in \mathcal{C}$. Then

$$V(\frac{1}{2}\mathcal{C}) = 2^{-n}V(\mathcal{C}) > m.$$

By Lemma 1, there are $m+1$ distinct points $\mathbf{c}_0, \ldots, \mathbf{c}_m \in \mathcal{C}$ such that

$$\frac{1}{2}\mathbf{c}_i - \frac{1}{2}\mathbf{c}_j \in \mathbf{Z}^n \qquad (0 \leq i,\ j \leq m).$$

There are $m+1$ points

$$\frac{1}{2}\mathbf{c}_i - \frac{1}{2}\mathbf{c}_0 \qquad (0 \leq i \leq m)$$

and m cosets of $\mathbf{Z}^n$ modulo Λ. By the pigeon hole principle, two must be in the same coset, that is there are i, j with $i \neq j$ such that

$$\frac{1}{2}\mathbf{c}_i - \frac{1}{2}\mathbf{c}_j \in \Lambda.$$

Now $-\mathbf{c}_j \in \mathcal{C}$ by symmetry; and so

$$\frac{1}{2}\mathbf{c}_i - \frac{1}{2}\mathbf{c}_j = \frac{1}{2}\mathbf{c}_i + \frac{1}{2}(-\mathbf{c}_j) \in \mathcal{C}$$

by convexity.

Note. Lemma 1 and Theorem 1 with $m = 1$ are due to Blichfeldt and Minkowski respectively. The generalizations to $m > 1$ are by van der Corput.

As a foretaste of the flavour of the application in the next section, we give

Lemma 2. *Let N be a positive integer. Suppose that there is an $l \in \mathbf{Z}$ such that*

$$l^2 \equiv -1\ (N).$$

Then $N = u^2 + v^2$ for some u, $v \in \mathbf{Z}$.

Proof. We take $n = 2$ and denote the co-ordinates by x, y. For $\mathcal{C}$ we take the open disc

$$x^2 + y^2 < 2m$$

of volume (= area)

$$V(\mathcal{C}) = 2\pi m > 2^2 m.$$

The subgroup Λ of $\mathbf{Z}^2$ is given by

$$x, y \in \mathbf{Z}, \qquad y \equiv lx\ \ (m).$$

It is clearly of index m. Hence by the Theorem there is

$$(0,0) \neq (u,v) \in \Lambda \cap \mathcal{C}.$$

Then

$$0 < u^2 + v^2 < 2m$$

and

$$u^2 + v^2 \equiv u^2(1 + l^2) \equiv 0\ \ (m).$$

Hence $u^2 + v^2 = m$, as required.

We note, in passing, that the condition of the lemma is certainly satisfied for primes p with $p \equiv 1\ (4)$.

§4. Exercises

1. Let $m \in \mathbf{Z}$, $m > 1$ and suppose that there is some $f \in \mathbf{Z}$ such that $f^2 + f + 1 \equiv 0\ (m)$. Show that $m = u^2 + uv + v^2$ for some u, $v \in \mathbf{Z}$.

2. Find a prime $p > 0$ for which there is an $f \in \mathbf{Z}$ such that

$$1 + 5f^2 \equiv 0\ (p)$$

but p is not of the shape $u^2 + 5v^2$ $(u,\ v \in \mathbf{Z})$.

5

Local-global principle. Conclusion of proof

We now complete the proof of the local-global principle for conics using the theorem of the last section. We recall that we had reduced the proof to that for

$$f_1X_1^2 + f_2X_2^2 + f_3X_3^2 = 0$$

where f_1, f_2, $f_3 \in \mathbf{Z}$ and $f_1f_2f_3$ is square free. We assume that there are points everywhere locally and we showed that this implied certain congruences to primes p dividing $2f_1f_2f_3$.

We first define a subgroup Λ of $\mathbf{Z}^3$ by imposing congruence conditions on the components of $\mathbf{x} = (x_1, x_2, x_3)$.

First case. $p \neq 2$, $p \not| f_1f_2f_3$, say $p \mid f_1$. We saw (end of §3) that then there is an $r_p \in \mathbf{Z}$ and that

$$f_2 + r_p^2 f_3 \equiv 0 \ \ (p).$$

We impose the condition

$$x_3 \equiv r_p x_2 \ \ (p).$$

Then

$$\begin{aligned} F(\mathbf{x}) &= f_1x_1^2 + f_2x_2^2 + f_3x_3^2 \\ &\equiv (f_2 + r_p^2 f_3)x_2^2 \\ &\equiv 0 \ \ (p). \end{aligned}$$

Second case. $p = 2$, $2 \not| f_1f_2f_3$. Then without loss of genérality

$$f_2 + f_3 \equiv 0 \ \ (4).$$

We impose the conditions

$$\left.\begin{aligned} x_1 &\equiv 0 \quad (2) \\ x_2 &\equiv x_3 \ \ (2) \end{aligned}\right\},$$

which imply

$$F(\mathbf{x}) \equiv 0 \ \ (4).$$

Third case. $p = 2$, $2 \mid f_1f_2f_3$, say $2 \mid f_1$. Then

$$s^2 f_1 + f_2 + f_3 \equiv 0 \ \ (8),$$

where $s = 0$ or 1. We impose the conditions

$$\left.\begin{aligned} x_2 &\equiv x_3 \quad (4) \\ x_1 &\equiv \int x_3 \ \ (2) \end{aligned}\right\},$$

which imply

$$F(\mathbf{x}) \equiv 0 \ \ (8).$$

To sum up. The group Λ is of index m (say) $= 4|f_1f_2f_3|$ in $\mathbf{Z}^3$, where throughout this section $|\ |$ is the absolute value. Further,

$$F(\mathbf{x}) \equiv 0 \qquad (4\,|f_1f_2f_3|)$$

for $\mathbf{x} \in \Lambda$.

We apply the theorem of the previous section to Λ and the convex symmetric set

$$\mathcal{C} : |f_1|x_1^2 + |f_2|x_2^2 + |f_3|x_3^2 < 4|f_1f_2f_3|.$$

School geometry shows that

$$\begin{aligned} V(\mathcal{C}) &= (\pi/3).2^3.|4f_1f_2f_3| \\ &> 2^3|4f_1f_2f_3| \\ &= m. \end{aligned}$$

Hence there is an $\mathbf{c} \neq \mathbf{0}$ in $\Lambda \cap \mathcal{C}$. For this $\mathbf{x}$ we have

$$F(\mathbf{x}) \equiv 0 \ \ (4|f_1f_2f_3|)$$

and

$$\begin{aligned} |F(\mathbf{x})| &\leq |f_1|x_1^2 + |f_2|x_2^2 + |f_3|x_3^2 \\ &< 4|f_1f_2f_3|; \end{aligned}$$

so

$$F(\mathbf{x}) = 0,$$

as required.

We conclude with some remarks.

Remark 1. We have not merely shown that there is a solution of $F(\mathbf{x}) = 0$, but we have found that there is one in a certain ellipsoid. This facilitates the search in explicitly given cases.

Remark 2. We have made no use of the condition of solubility in $\mathbb{Q}_p$ for $p \not| 2f_1f_2f_3$. In fact this condition tells us nothing [cf. §3, Exercises 2, 3]. It is left to the reader to check that for any f_1, f_2, f_3 and p with $p \not| 2f_1f_2f_3$ there is always a point defined over $\mathbb{Q}_p$ on

$$f_1X_1^2 + f_2X_2^2 + f_3X_3^2 = 0.$$

Remark 3. We have also nowhere used that there is local solubility for $\mathbb{Q}_\infty = \mathbb{R}$.

Hence solubility at $\mathbb{Q}_\infty$ is implied by solubility at all the $\mathbb{Q}_p$ $(p \neq \infty)$. This phenomenon is connected with quadratic reciprocity. In fact for any conic over $\mathbb{Q}$, the number of p (including ∞) for which there is not a point over $\mathbb{Q}_p$ is always even [cf. §3, Exercises 6,7]. See a book on quadratic forms (such as the author's).

§5. Exercises

1. Let

$$F(X,Y,Z) = 5X^2 + 3Y^2 + 8Z^2 + 6(YZ + ZX + XY).$$

Find rational integers x, y, z not all divisible by 13, such that

$$F(x,y,z) \equiv 0 \pmod{13^2}.$$

[*Hint.* cf. Hensel's Lemma 2 of §10.]

2. Let

$$F(X,Y,Z) = 7X^2 + 3Y^2 - 2Z^2 + 4YZ + 6ZX + 2XY.$$

Find rational integers x, y, z not all divisible by 17 such that

$$F(x,y,z) \equiv 0 \pmod{17^3}.$$

6

Cubic curves

In this section we consider curves given by

$$\mathcal{C} : F(\mathbf{X}) = F(X_1, X_2, X_3) = 0,$$

where F is a homogeneous cubic form. The case of interest is when the ground field is the rationals $\mathbb{Q}$, but our initial remarks apply to any ground field.

A point $\mathbf{x}$ on $\mathcal{C}$ is said to be *singular* when

$$\frac{\partial F}{\partial X_j}(\mathbf{x}) = 0 \qquad (j = 1,2,3).$$

If we choose co-ordinates so that $\mathbf{x} = (0,0,1)$, this is equivalent to F not containing terms in X_3^3, $X_1X_3^2$, $X_2X_3^2$.

A singular point counts with multiplicity at least 2 as an intersection with a line. More precisely, if $\mathbf{a}$, $\mathbf{b}$ are two points on the line, the general point on it is

$$\lambda\mathbf{a} + \mu\mathbf{b},$$

where the numbers λ, μ are not both 0. The intersections with $\mathcal{C}$ are given by

$$F(\lambda\mathbf{a} + \mu\mathbf{b}) = 0, \qquad (*)$$

a homogeneous cubic in λ, μ. What is claimed is that if one of the intersections is a singular point of $\mathcal{C}$ then the corresponding ratio $\lambda : \mu$ occurs as a multiple root of (*). An easy way to check this is to take $\mathbf{b} = \mathbf{x}$.

Suppose that $\mathcal{C}$ has two distinct singular points $\mathbf{x}$, $\mathbf{y}$. The line joining them cuts $\mathcal{C}$ at both $\mathbf{x}$, $\mathbf{y}$ with multiplicity ≥ 2. This can happen only if $F(\lambda\mathbf{x} + \mu\mathbf{y})$ vanishes identically, i.e. if $\mathcal{C}$ contains the whole line. If we suppose, as we shall, that $\mathcal{C}$ is *irreducible* (i.e. that F does not factorize), this cannot happen. An irreducible cubic curve has at most one singular point.

Now take the ground field to be $\mathbf{Q}$. If there is a singular point over the algebraic closure $\overline{\mathbf{Q}}$, there is at most one. By Galois theory[2] it must be defined over $\mathbf{Q}$. Hence, as we have already seen in §1, $\mathcal{C}$ is birationally equivalent over $\mathbf{Q}$ to the line.

From now on we restrict attention to ***non-singular*** cubic curves, i.e. those which have no singular points over $\mathbf{Q}$. Let $\mathbf{a}$, $\mathbf{b}$ be rational points on $\mathcal{C}$. The line joining them meets $\mathcal{C}$ in a third point, in general distinct: it is also rational since it is given by a cubic equation, two of whose roots are rational. This process was used already by Diophantos to find new unobvious points from known obvious ones. The variant in which one takes the third point of intersection with the curve of the tangent at a rational point was, according to Weil, first noted by Newton. An older generation of mathematicians refer to these as the "chord and tangent processes".

In general, starting from one rational point $\mathbf{a}$ on $\mathcal{C}$ one obtains infinitely many by the chord and tangent processes. If this is not the case, $\mathbf{a}$ is said to be *exceptional*. For example we have

Lemma 1. *Let $a \geq 1$ be a cubic-free integer and let*

$$\mathcal{C}:\ X^3 + Y^3 - aZ^3 = 0.$$

The point $(1, -1, 0)$ is exceptional. For $a = 1$ the points $(0, 1, 1)$, $(1, 0, 1)$ are also exceptional. For $a = 2$ the point $(1, 1, 1)$ is exceptional. No other rational point is exceptional.

Proof. We first show that the given points are indeed exceptional. The tangent at $(1, -1, 0)$ is $X + Y = 0$, which meets $\mathcal{C}$ only at $(1, -1, 0)$. The other cases for $a = 1$ are similar. The tangent at $(1, 1, 1)$ for $a = 2$ is $X + Y - 2Z = 0$, which meets $\mathcal{C}$ again only at $(1, -1, 0)$.

Let $\mathbf{x} = (x, y, z)$ be a rational point other than those named. We may suppose that x, y, z are integers without common factor. The equation for $\mathcal{C}$ implies that then x, y, z are coprime in pairs.

[2] *For the cognoscenti.* If the ground field is not perfect, the conclusion does not necessarily hold. See *Note* at end of §9.

Let $\mathbf{x}_1 = (x_1, y_1, z_1)$ be the third point of intersection, where again x_1, y_1, z_1 are integers without common factor. It may be verified[3] that

$$x_1 : y_1 : z_1 = x(x^3 + 2y^3) : -y(2x^3 + y^3) : z(x^3 - y^3)$$

Let d be the greatest common divisor of the three terms on the right hand side. If a prime p divides both x and d it must also divide y, a contradiction. Hence d divides $x^3 + 2y^3$ and $2x^3 + y^3$. It thus divides $3x^3$ and $3y^3$, so $d = 1$ or 3. Hence

$$z_1 = \pm z(x^3 - y^3) \qquad \text{or} \qquad z_1 = \pm z(x^3 - y^3)/3.$$

In either case, it is readily verified that $|z_1| > |z|$ except for the $\mathbf{x}$ listed in the enunciation. By repeating the tangent process we thus get a sequence of points $\mathbf{x}$, $\mathbf{x}_1$, $\mathbf{x}_2, \ldots$ with

$$|z| < |z_1| < |z_2| < \ldots .$$

Hence the $\mathbf{x}_j$ are distinct, and $\mathbf{x}$ is not exceptional.

§6. Exercises

1. (i) Show that the cubic curve

$$Y^2Z = X^3 + AXZ^2 + BZ^3$$

is non-singular provided that

$$4A^3 + 27B^2 \neq 0.$$

(ii) If $4A^3 + 27B^2 = 0$, find a singularity and decide whether it is a cusp or a double point with distinct tangents.

2. (i) Let

$$F(\mathbf{x}) = a_1X_1^3 + a_2X_2^3 + a_3X_3^3 + dX_1X_2X_3,$$

where

$$a_1a_2a_3 \neq 0.$$

Show that $F(\mathbf{x}) = 0$ is non-singular provided that

$$27a_1a_2a_3 + d^3 \neq 0.$$

(ii) If $a_1 = a_2 = a_3 = 1$, $d = -3$, show that any point (x_1, x_2, x_3) with $a_1^3 = x_2^3 = x_3^3 = x_1x_2x_3 = 1$ is a singularity.

[3] This is essentially a special case of elegant formulae of Desboves for the chord and tangent processes. See Exercises and Formulary.

10

Reduction

The philosophy is to approach the rational field $\mathbb{Q}$ through the local fields $\mathbb{Q}_p$ and, similarly, to approach the $\mathbb{Q}_p$ through the finite fields $\mathbb{F}_p$ by reduction modulo p. We do no more than is required for the applications.

The mod p map $\mathbb{Z}_p \to \mathbb{F}_p$ is denoted by a bar $a \to \overline{a}$. This is extended to the corresponding 2-dimensional projective planes V, $\overline{V}$ as follows. Let (a_1, a_2, a_3) be projective co-ordinates of a point $\mathbf{a}$ of V. By multiplying a_1, a_2, a_3 by the same element of $\mathbb{Q}_p$, we have without loss of generality

$$\max\{|a_1|, |a_2|, |a_3|\} = 1,$$

where $|| = ||_p$. Then $(\overline{a_1}, \overline{a_2}, \overline{a_3})$ are the co-ordinates of a well-defined point $\overline{\mathbf{a}}$ of $\overline{V}$.

In a similar way, we define the reduction $\overline{\mathbf{l}}$ of a line

$$\mathbf{l}: \ l_1X_1 + l_2X_2 + l_3X_3 = 0.$$

If the point $\mathbf{a}$ lies on the line $\mathbf{l}$, then clearly $\overline{\mathbf{a}}$ lies on $\overline{\mathbf{l}}$.

We need only the least sophisticated of the many ways of reducing a cubic curve

$$\mathcal{C}: \ F(\mathbf{X}) = 0$$

defined over $\mathbb{Q}_p$. Here

$$F(\mathbf{X}) = \sum_{i \le j \le k} f_{ijk} X_i X_j X_k \in \mathbb{Q}_p[\mathbf{X}]$$

where the $f_{ijk} \in \mathbb{Q}_p$ are not all 0 and without loss of generality

$$\max_{i,j,k} |f_{ijk}| = 1.$$

Then

$$\overline{F}(\mathbf{X}) = \sum_{i \le j \le k} \overline{f}_{ijk} X_i X_j X_k | \in \mathbb{F}_p[\mathbf{X}]$$

is not the zero polynomial, and defines the reduced curve

$$\overline{\mathcal{C}}: \ \overline{F}(\mathbf{X}) = 0$$

over $\mathbb{F}_p$. It may, of course, be reducible[8].

If a point $\mathbf{a}$ lies on $\mathcal{C}$, then clearly $\overline{\mathbf{a}}$ lies on $\overline{\mathcal{C}}$. There is a weak converse

Lemma 1. *Let $\overline{\mathbf{b}}$ be a nonsingular point of $\overline{\mathcal{C}}$. Then there is an $\mathbf{a}$ on $\mathcal{C}$ such that $\overline{\mathbf{a}} = \overline{\mathbf{b}}$.*

Note. The notation $\overline{\mathbf{b}}$ is intended to denote a point defined over $\mathbb{F}_p$ not necessarily derived from a $\mathbf{b}$. We say that $\overline{\mathbf{b}}$ *lifts* to $\mathbf{a}$. It is easy to see by examples that a singular point on $\overline{\mathcal{C}}$ may or may not lift to a point of $\mathcal{C}$ (cf. Exercises).

We construct $\mathbf{a}$ by successive approximation à la Newton. The generic term for such constructions in p-adic analysis is Hensel's Lemma.

Lemma 2. *Let $G(T) \in \mathbb{Z}_p[T]$ and let $t_0 \in \mathbb{Z}_p$ be such that*

$$|G(t_0)| < 1, \qquad |G'(t_0)| = 1,$$

where G' is the formal derivative of G. Then there is a $t \in \mathbb{Z}_p$ such that

$$G(t) = 0 \qquad |t - t_0| \le G(t_0).$$

Assuming the truth of Hensel's Lemma for the moment, we complete the proof of the Lemma. Since $\overline{\mathbf{b}}$ is nonsingular on $\overline{\mathcal{C}}$, we may suppose that

$$\frac{\partial \overline{F}}{\partial X_1}(\overline{\mathbf{b}}) \neq 0.$$

Pick any $b_j \in \mathbb{Z}_p$ such that $\overline{\mathbf{b}} = (\overline{b}_1, \dots, \overline{b}_n)$. Then the conditions of Hensel's Lemma apply to

$$G(T) = F(T, b_2, \dots, b_n), \qquad t_0 = b_1.$$

Put $\mathbf{a} = (t, b_2, \dots, b_n)$, where t is provided by Hensel. Clearly $F(\mathbf{a}) = 0$, $\overline{\mathbf{a}} = \overline{\mathbf{b}}$, so $\mathbf{a}$ does what is required.

It remains to prove the Hensel's Lemma. Let U be an indeterminate.

[8] In the sense that $\overline{F}(\mathbf{X})$ factorizes. There is an unfortunate clash of meanings between "reduced" (mod p) and "reducible".

Then

$$G(T+U) = G(T) + UG_1(T) + U^2G_2(T) + \dots$$

where $G_j \in \mathbb{Z}_p[T]$ and $G_1 = G'$. Now define

$$u = -G(t_0)/G'(t_0),$$

so

$$G(t_0+u) = u^2G_2(t_0) + u^3G_3(t_0) + \dots .$$

Hence

$$|G(t_1)| \leq |G(t_0)|^2,$$

where

$$t_1 = t_0 + u.$$

Clearly

$$|G'(t_1)| = |G'(t_0)| = 1.$$

We may therefore iterate the process and get a fundamental sequence t_i $(t \geq 0)$. The limit t clearly does what is required.

We shall also need information about the behaviour of the intersection of a line and a cubic curve under reduction. From what we have already proved, if $\mathbf{l}$ meets $\mathcal{C}$ in $\mathbf{a}$, then $\bar{\mathbf{l}}$ meets $\bar{\mathcal{C}}$ in $\bar{\mathbf{a}}$. But suppose that $\mathbf{l}$ meets $\mathcal{C}$ in $\mathbf{a}$, $\mathbf{b}$ with $\mathbf{a} \neq \mathbf{b}$: if $\bar{\mathbf{a}} = \bar{\mathbf{b}}$, can we be sure that it has multiplicity ≥ 2 in the intersection?

The following lemma confirms expectations.

Lemma 3. *Suppose that the line* $\mathbf{l}$ *meets the cubic curve* $\mathcal{C}$ *in* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, *multiple points of intersection being given with their multiplicities. Then either*

(I) *the entire line* $\bar{\mathbf{l}}$ *is in* $\bar{\mathcal{C}}$ *or*

(II) $\bar{\mathbf{l}}$ *meets* $\bar{\mathcal{C}}$ *in* $\bar{\mathbf{a}}$, $\bar{\mathbf{b}}$, $\bar{\mathbf{c}}$, *multiple points occuring with the correct multiplicities.*

Proof. We have without loss of generality

$$l_3 = 1 = \max(|l_1|, |l_2|, |l_3|).$$

Consider

$$G(X_1, X_2) = F(X_1, X_2, -l_1X_1 - l_2X_2)$$
$$= \mathbb{Z}_p[X_1, X_2].$$

Its reduction is

$$\overline{G}(X_1, X_2) = \overline{F}(X_1, X_2 - \bar{l}_1X_1 - \bar{l}_2X_2).$$

If $\overline{G}(X_1, X_2) = 0$, we have case (I) of the Lemma, so we may suppose that

$$\overline{G}(X_1, X_2) \neq 0.$$

We normalize the coefficients of $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ so that

$$\max(|a_1|, |a_2|, |a_3|) = 1.$$

Since $\mathbf{la} = 0$, it follows that

$$(\bar{a}_1, \bar{a}_2) \neq (0, 0)$$

etc.

By hypothesis, there is some $\lambda \in \mathbb{Q}_p$ such that

$$G(X_1, X_2) = \lambda(a_2X_1 - a_1X_2)(b_2X_1 - b_1X_2)(c_2X_1 - c_1X_2)$$
$$= \lambda H(X_1, X_2).$$

Now

$$\overline{H}(X_1, X_2) = (\bar{a}_2X_1 - \bar{a}_1X_2)(\bar{b}_2X_1 - \bar{b}_1X_2)(\bar{c}_2X_1 - \bar{c}_1X_2)$$
$$\neq 0.$$

Hence $\overline{G}$, $\overline{H}$ differ only by a scalar multiple, which is what we needed to prove.

§10. Exercises

1. (i) Let $\mathcal{C}$ be the curve $Y^2 = X^3 + p$ over $\mathbb{Q}_p$. Show that the point $(0, 0)$ on the mod p curve does not lift to a point of $\mathcal{C}$.

(ii) Find an example of an elliptic curve $\mathcal{C}$ over $\mathbb{Q}_p$ such that the mod p curve has a cusp which is the reduction of a point on $\mathcal{C}$.

2. Find examples of curves $\mathcal{C}$ over $\mathbb{Q}_p$ such that the mod p curve has a double point with distinct tangents which (i) lifts, (ii) does not lift, to $\mathcal{C}$.

One opening gambit is to observe that

$$(x+y+z)+x=(x+z)+(y+z).]$$

3. Let $\mathcal{C}: Y^2 = X^3+AX+B$ and suppose that $\mathbf{x}_1$, $\mathbf{x}_2$ are independent generic points. Let $\mathbf{x}_3 = \mathbf{x}_1+\mathbf{x}_2$, $\mathbf{x}_4 = \mathbf{x}_1-\mathbf{x}_2$. Show that

$$(x_1-x_2)^2(x_1+x_2+x_3) = (y_1-y_2)^2$$
$$(x_1-x_2)^2(x_1+x_2+x_4) = (y_1+y_2)^2.$$

Deduce that $x_1+x_2+x_3$, $x_1+x_2+x_4$ are roots of an equation

$$(x_1-x_2)^2T^2+uT+v=0,$$

where u, v are *polynomials* in x_1, x_2.

Deduce that a similar result holds for x_3, x_4.

4. (Required in text.) Let $G(\mathbf{X}) \in \mathbb{Q}[\mathbf{X}]$ be a nonsingular quadratic form in $\mathbf{X} = (X, Y, Z)$ and suppose that there is an $\mathbf{x} = (x, y, z) \neq (0,0,0)$ such that $G(\mathbf{x}) = 0$. Show that there are linear forms $L(\mathbf{X})$, $M(\mathbf{X})$, $N(\mathbf{X}) \in \mathbb{Q}[\mathbf{X}]$ and a $d \in \mathbb{Q}^*$ such that

$$G(\mathbf{X}) = L(\mathbf{X})M(\mathbf{X}) + dN(\mathbf{X})^2.$$

[*Hints.*

(i) Without loss of generality $\mathbf{x} = (1,0,0)$.

(ii) After a linear transformation on Y, Z, we may suppose $G(X) = XY +$ form in Y, Z.

(iii) Complete the square with respect to Z.]

5. Let $\hat{h}$ be the canonical height on some curve $\mathcal{C}$ and suppose that there are representatives of all classes of $\mathfrak{G}/2\mathfrak{G}$ in $\hat{h}(\mathbf{x}) \leq t$ for some t. Show that $\mathfrak{G}$ is generated by the $\mathbf{a} \in \mathfrak{G}$ with $\hat{h}(\mathbf{a}) \leq t$.

18

Local-global for genus 1

Our attention now moves from elliptic curves to curves of genus 1 in general. In this section we give a couple of examples to show that there is no local-global principle for rational points on curves of genus 1. Subsequently, we shall give a structure to the "obstruction" to a local-global principle, namely the Tate-Shafarevich group.

The two examples we shall discuss are

$$3X^3+4Y^3+5Z^3=0, \tag{1}$$

due to Selmer, and

$$X^4-17=2Y^2, \tag{2}$$

due (independently) to Lind and Reichardt. The techniques we have developed so far enable us to disprove the existence of rational points. We have not, however, developed techniques to show that there are solutions everwhere locally. This is because we have left a fairly highbrow discussion of curves of genus 1 over finite fields until the end (§25). The reader may, of course, verify for any given p that there is a point defined over $\mathbb{Q}_p$ but this can never disprove the existence of some $P > 10^{10}$ (say) such that (1) or (2) has no solution in $\mathbb{Q}_P$. We shall assume without present proof that a curve of genus 1 over a finite field $\mathbb{F}_p$ always has a point defined over $\mathbb{F}_p$ (§25, Theorem 2). If, therefore, a curve such as (1) or (2) reduced mod p is still of genus 1, then there is a point mod p which can, by Lemma 1 of §10, be lifted to a point defined over $\mathbb{Q}_p$.

Assuming this[22], the only $\mathbb{Q}_p$ to be considered for (1) are $p = 2, 3, 5$ and the only ones for (2) are $p = 2, 17$. It may confidently be left to the reader to confirm that there are points for these p.

The disproof of rational points on (1) uses

Lemma 1. *Let a, b, c be distinct integers > 1 and suppose that $d = abc$ is cube free. Suppose that there are u, v, $w \in \mathbb{Z}$ not all 0 such that*

$$au^3 + bv^3 + cw^3 = 0.$$

Then there are x, y, $z \in \mathbb{Z}$ with $z \neq 0$ such that

$$x^3 + y^3 + dz^3 = 0.$$

Proof. Let $\rho^3 = 1$, $\rho \neq 1$ and put

$$\xi = au^3 + \rho bv^3 + \rho^2 cw^3$$
$$\eta = au^3 + \rho^2 bv^3 + \rho cw^3.$$

Then

$$\xi + \eta = 3au^3$$
$$\rho\xi + \rho^2\eta = 3cw^3$$
$$\rho^2\xi + \rho\eta = 3bv^3$$

and so

$$\xi^3 + \eta^3 + d\zeta^3 = 0, \qquad \zeta = -3uvw.$$

Now the two points $(\xi, \rho\eta, \zeta)$, $(\eta, \rho^2\xi, \zeta)$ are conjugate over $\mathbb{Q}$. Hence the line joining them meets $X^3 + Y^3 + dZ^3 = 0$ in a point defined over $\mathbb{Q}$ and distinct from $(1, -1, 0)$.

Lemma 2. *The only point defined over $\mathbb{Q}$ on*

$$X^3 + Y^3 + 60Z^3 = 0$$

is $(1, -1, 0)$.

Proof. There is no torsion, e.g. by the discussion of exceptional points on cubic curves (§6, Lemma 1). The curve is birationally equivalent over $\mathbb{Q}$ to

$$Y^2 = X^3 - 2^4.3^3.60^2,$$

[22] **Often, including for (1), (2), this can be proved elementarily. See e.g. A. Weil, Numbers of solutions of equations in finite fields. *Bull. Amer. Math. Soc.* 55 (1949), 497–508 (= *Collected Papers* I, 395–410.)**

for which $\mathfrak{G}/2\mathfrak{G}$ is trivial by the proof at the end of the section on the weak theorem (§15, Second example). It follows from the Finite Basis Theorem that $\mathfrak{G}$ is trivial.

Theorem 1. *There are no rational points on* (1).

Proof. The last two lemmas.

The preceding proof used the theory of algebraic numbers. The next proof works entirely in the rationals.

Theorem 2. *There are no rational points on* (2).

Proof. If not, suppose (x, y) is on (2). Let $x = a/c$ as a fraction in its lowest terms. Then

$$a^4 - 17c^4 = 2b^2, \qquad \gcd(a,c) = \gcd(b,c) = \gcd(a,b) = 1.$$

Putting

$$A = a^2, \qquad C = c^2$$

we have

$$A^2 - 17C^2 = 2b^2.$$

This equation is soluble everwhere locally, so globally, and in fact

$$5^2 - 17.1^2 = 2.2^2.$$

Now

$$(5A + 17C + 4b)(5A + 17C - 4b) = 17(A + 5C)^2.$$

If there is a common odd prime divisor of the two factors on the left hand side, it divides $5A + 17C$ and $A + 5C$, so divides $8A$ and $8C$: a contradiction. The two factors on the left hand side have the same sign, which for $A = a^2$, $C = c^2$ must be positive. Hence for integers u, v there is one of two possibilities

	First Case	Second Case
$5a^2 + 17c^2 \pm 4b =$	$17u^2$	$34u^2$
$5a^2 + 17c^2 \mp 4b =$	v^2	$2v^2$
$a^2 + 5c^2 \ =$	uv	$2uv$

In the first case

$$10a^2 + 34c^2 = 17u^2 + v^2$$
$$a^2 + 5c^2 = uv.$$

We show that this is impossible in $\mathbf{Q}_{17}$. Write $|| = ||_{17}$. By homogenei

$$\max(|a|, |c|, |u|, |v|) = 1.$$

Since 10 is a quadratic non residue mod 17, we have

$$|a| < 1, \qquad |v| < 1.$$

The second equation gives

$$|c| < 1.$$

Finally, the first equation gives

$$|u| < 1.$$

Contradiction.

The second case gives

$$5a^2 + 17c^2 = 17u^2 + v^2$$
$$a^2 + 5c^2 = 2uv.$$

The proof that this is impossible in $\mathbf{Q}_{17}$ is similar.

§18. Exercises

1. [Uses algebraic number theory.] Supply the details of the followi alternative proof of Theorem 2.

(i) The field $\mathbf{Q}(\sqrt{17})$ has class number 1. A basis of integers is 1 $\frac{1}{2}(1+\sqrt{17})$. A fundamental unit is $4+\sqrt{17}$ of norm -1. Th prime 2 splits into $(5 \pm \sqrt{17})/2$.

(ii) Suppose $a^4 - 17c^4 = 2b^2$ with a, b, $c \in \mathbf{Z}$, $\gcd(a,c) = 1$. Then a, c are odd and

$$\frac{a^2 \pm c^2\sqrt{17}}{2}$$

are coprime.

(iii)

$$\frac{a^2+\sqrt{17}c^2}{2} = \left(\frac{5 \pm \sqrt{17}}{2}\right)\eta\mu^2$$

for some unit η and some integer μ.

(iv) $\eta > 0$ in both real embeddings. Hence η is a square and so can b absorbed in μ^2.

(v) Put $\eta = 1$, $\mu = (u + v\sqrt{17})/2$ in and equate terms independent $\sqrt{17}$. Then $4a^2 = 5(u^2+17v^2) \pm 34uv$, which is impossible in $\mathbf{Q}$ (and in $\mathbf{Q}_{17}$).

19

Elements of Galois cohomology

In the next section we have occasion to consider two curves which are both defined over $\mathbf{Q}$ and which are birationally equivalent over $\overline{\mathbf{Q}}$. Here we consider a simpler case and then set up some general machinery.

The conic

$$\mathcal{A}: \quad X_1^2 + X_2^2 = 3$$

has no rational point and so is not equivalent over $\mathbf{Q}$ to the line (coordinate Y, no equation). They are, however, equivalent over $\mathbf{Q}(\sqrt{3})$, for example by the equations

$$y = (x_1 - \sqrt{3})/x_2$$

$$x_1 = \frac{\sqrt{3}(1-y^2)}{y^2+1}, \qquad x_2 = \frac{-2\sqrt{3}y}{y^2+1}.$$

Let y be **transcendental**, so (x_1, x_2) **is a generic point of** $\mathcal{A}$.

The Galois group $\mathrm{Gal}(\mathbf{Q}(\sqrt{3})/\mathbf{Q})$ can be made to act in two different ways on $\mathbf{Q}(\sqrt{3}, y) = \mathbf{Q}(\sqrt{3}, x_1, x_2)$. We can either make it act trivially on y or we can make it act trivially on (x_1, x_2).

In the first case, the non-trivial element of the Galois group induces the automorphism

$$x_1 \to -x_1, \qquad x_2 \to -x_2$$

of $\mathcal{A}$. In the second case, it induces the automorphism

$$y \to -1/y$$

of the line.