ELLIPTIC CURVES 2025/26. LECTURER: JAMES NEWTON

Contents

These notes are a slightly edited version of those written by Victor Flynn. Please send any typos or queries to newton@maths.ox.ac.uk

The first part of the lecture notes overlaps with the final part of the Preliminary Reading file. References with numbers $0.x$ *refer to that file.*

Section 1. Geometric preliminaries

Affine curves.

Definition 1.1. Let K be a field. We define $\mathbf{A}^n(K) = \{(x_1, \ldots, x_n) : x_1, \ldots x_n \in K\}$, and refer to a point $P \in \mathbf{A}^n(K)$ as a *K-rational* point of the *affine n-space* $\mathbf{A}^n$. We also say that a point $P \in \mathbf{A}^n(K)$ is *defined over* K.

Definition 1.2. An algebraic expression such as a curve, polynomial, rational function, is said to be *defined over* K (or K-rational) if it can be described by an equation with coefficients in K.

Definition 1.3. A (non-constant) polynomial in two variables $f(x, y)$, with coefficients in K, defines an (affine) *curve defined over* K. For any field L with $K \subset L$, the set of L-rational points $\{(a, b) \in \mathbf{A}^2(L) : f(a, b) = 0\}$ on a curve $\mathcal{C}$ with equation $f(x, y)$ is denoted $\mathcal{C}(L)$. The field K is often called the *field of definition* (or the *ground field*).

Example 1.4. Let $\mathcal{C} : f(x, y) = x^2 + y^2 = 0$. This defines an affine curve over $\mathbb{Q}$. Of course, this same curve $\mathcal{C}$ could be regarded as having field of definition (ground field) $\mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{R}, \mathbb{C}$ or indeed any field containing $\mathbb{Q}$. When the field of definition is not stated explicitly, it is taken to be the smallest possible field over which the curve is defined (in this case, $\mathbb{Q}$). The point $(0, 0)$ is $\mathbb{Q}$-rational and it is the only $\mathbb{Q}$-rational point on $\mathcal{C}$, so that $\mathcal{C}(\mathbb{Q}) = \{(0, 0)\}$. It has many $\mathbb{C}$-rational points, for example $(i, 1) \in \mathcal{C}(\mathbb{C})$.

Comment 1.5. Our affine curves are by definition embedded in the plane $\mathbf{A}^2$, and cut out by a single polynomial equation. It is also possible to embed curves in higher dimensional space, as long as the number of 'independent' polynomials defining the curve is one less than the number of variables; for example the 2 equations: $y^2+4x^2-1 = 0, z^2-x^2-x = 0$

define a curve in the variables x, y, z. However, we shall not concern ourselves with that here.

Definition 1.6. The *degree* of a polynomial is the degree of its highest degree monomial. A *homogeneous* polynomial is a polynomial whose terms all have the same degree.

Example 1.7. $f(x,y) = x + y - 8 = 0$ defines a curve of degree 1 (a *linear* curve), $g(x,y) = xy + y^2 - y + 3 = 0$ defines a curve of degree 2 (a *quadratic* curve) and $h(x,y) = x^3 + y^3 + y - 1$ defines a curve of degree 3 (a *cubic* curve). None of these polynomials are homogeneous.

If you try drawing an accurate sketch of, for example, the three curves $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ defined by $x^2 + y^2 = 1, y^2 = x^3, y^2 = x(x-2)^2$, respectively, you will notice distinguishing features. The first curve $\mathcal{C}_1$ appears smooth at all points, and it is easy to see that there is a unique tangent at each point. The curve $\mathcal{C}_2$ has a 'sharp corner' at $(0,0)$, and the third curve $\mathcal{C}_3$ crosses itself at the point $(2,0)$, when there is a plausible choice of two distinct tangents. These sharp corners and crossing points are typified by the fact that both partial derivatives of f vanish, when the curve is written as $f(x,y) = 0$.

Definition 1.8. Let $\mathcal{C} : f(x,y) = 0$ be an (affine) curve defined over a field K and let $P = (x_0, y_0)$ be a point in $\mathcal{C}(\overline{K})$, where $\overline{K}$ is an algebraic closure of K. We say that P is a *singular point* (or a *singularity*) on $\mathcal{C}$ if $\frac{\partial f}{\partial x}(P) = 0$ and $\frac{\partial f}{\partial y}(P) = 0$. Otherwise, P is a *smooth point* (or a *nonsingular point*) on $\mathcal{C}$. A curve $\mathcal{C}$ is called *smooth* (or *nonsingular*) if it does not contain any singular points (the curve is called *singular* if it contains at least one singular point).

Comment 1.9. There is a standard technique for computing all tangents to $\mathcal{C} : f(x,y) = 0$ at a point $P = (x_0, y_0)$, in which we first translate the curve by $(-x_0, -y_0)$ (so that (x_0, y_0) is taken to $(0,0)$), then use the fact that the lowest degree terms dominate near $(0,0)$ and determine the tangent behaviour at $(0,0)$, and then finally translate the curve back to its original position. This gives three steps.

Step 1. Consider $f(x+x_0, y+y_0)$ (this is $f(x,y)$ translated by $(-x_0, -y_0)$) which contains the point $x = y = 0$ and so has no constant term. We can write:

$$f(x + x_0, y + y_0) = R_k(x,y) + R_{k+1}(x,y) + \ldots + R_n(x,y),$$

where $k \geqslant 1$ and where each $R_i(x,y)$ is homogeneous of degree i (for $k \leqslant i \leqslant n$) and $R_k(x,y) \neq 0$.

Step 2. Consider $R_k(x,y)$, which is the lowest degree portion of $f(x + x_0, y + y_0)$, and factorise $R_k(x,y) = L_1(x,y)L_2(x,y)\ldots L_k(x,y)$ over the algebraic closure, where $L_1, \ldots, L_k$ are linear.

Step 3. There are k tangents to $f(x + x_0, y + y_0) = 0$ at $(0,0)$ namely: $L_1(x,y) = 0, \ldots, L_k(x,y) = 0$. So, after reversing the translation of Step 1, there are k tangents to $\mathcal{C} : f(x,y) = 0$ at $P = (x_0, y_0)$, namely:

$$L_1(x - x_0, y - y_0) = 0, \ldots, L_k(x - x_0, y - y_0) = 0.$$

Note that the same tangent may be repeated more than once (e.g. $\mathcal{C} : f(x,y) = y^2 - x^3 = 0$ has 2 tangents at $(0,0)$, namely: $y = 0$ twice, in which case we can say that the tangent $y = 0$ occurs with multiplicity 2).

Comment 1.10. $P = (x_0, y_0)$ is a smooth point on $\mathcal{C}$
$\iff$ $k = 1$ in Step 1
$\iff$ there is only one tangent to $\mathcal{C}$ at P.

When $k \geqslant 2$, the singularity at P is called a *double point* ($k = 2$), *triple point* ($k = 3$), and so on.

Example 1.11. Let $\mathcal{C}_1 : x^2 + y^2 = 1$ (a circle of radius 1 and centre $(0, 0)$). Then we can write: $\mathcal{C}_1 : f(x, y) = x^2 + y^2 - 1 = 0$, and so $\frac{\partial f}{\partial x} = 2x$, $\frac{\partial f}{\partial y} = 2y$. A point (x, y) is a singular point on $\mathcal{C}_1$ exactly when: it lies on $\mathcal{C}_1$ and both partial derivatives are zero, that is, when:

$$(1)\ x^2 + y^2 - 1 = 0, \quad (2)\ 2x = 0, \quad (3)\ 2y = 0.$$

Assuming our ground field does not have characteristic 2, equations (2),(3) force $x = y = 0$, but this does not satisfy (1). We conclude that there are no singular points and that $\mathcal{C}_1$ is smooth.

Example 1.12. Let $\mathcal{C}_2 : y^2 = x^3$, that is: $\mathcal{C}_2 : f(x, y) = y^2 - x^3 = 0$. Then $\frac{\partial f}{\partial x} = -3x^2$, $\frac{\partial f}{\partial y} = 2y$. We can see that the only singular point is $(0, 0)$. For computing tangents at $(0, 0)$, we first take $f(x+0, y+0) = y^2 - x^3 = R_2(x, y) + R_3(x, y)$, where $R_2(x, y) = y^2$ and $R_3(x, y) = -x^3$. Then $R_2(x, y) = y^2 = L_1(x, y)L_2(x, y) = y \cdot y$, so there are two tangents to $\mathcal{C}_2$ at $(0, 0)$, namely: $L_1(x - 0, y - 0) = 0$ and $L_2(x - 0, y - 0) = 0$, that is: $y = 0$ and $y = 0$ (i.e. $y = 0$ with multiplicity 2). A double point singularity where the same tangent line has multiplicity 2 is called a *cusp* (or a *cuspidal singularity*).

Example 1.13. Let $\mathcal{C}_3 : y^2 = x(x - 2)^2$, that is: $\mathcal{C}_3 : f(x, y) = y^2 - x(x - 2)^2 = 0$. The point (x, y) on $\mathcal{C}_3$ is singular when:

$$(1)\ y^2 - x(x-2)^2 = 0, \quad (2)\ \frac{\partial f}{\partial x} = -3x^2 + 8x - 4 = 0, \quad (3)\ \frac{\partial f}{\partial y} = 2y = 0.$$

Assuming our ground field does not have characteristic 2, from (3) we see that $y = 0$, and substituting this into (1) gives: $x(x - 2)^2 = 0$, so that $x = 0$ or 2. Now, $x = 2$ satisfies (2), but $x = 0$ does not, giving $x = 2$ as the only common solution. So, the only possible singular point is $(2, 0)$ (conversely, check that $x = 2, y = 0$ satisfies (1),(2),(3) so that $(2, 0)$ is a singular point). We conclude that $(2, 0)$ is the only singularity on $\mathcal{C}_3$.

For the tangents at $(2, 0)$, first compute $f(x+2, y+0) = y^2 - (x+2)x^2 = y^2 - 2x^2 - x^3 = R_2(x, y) + R_3(x, y)$, where $R_2(x, y) = y^2 - 2x^2$ and $R_3(x, y) = -x^3$. Factorising $R_2(x, y)$ into linear factors gives: $R_2(x, y) = (y + \sqrt{2}x)(y - \sqrt{2}x) = L_1(x, y)L_2(x, y)$. The tangents to the curve $\mathcal{C}_3$ at $(2, 0)$ are then: $L_1(x - 2, y - 0) = 0$ and $L_2(x - 2, y - 0) = 0$, that is: $y = -\sqrt{2}(x - 2)$ and $y = \sqrt{2}(x - 2)$. The point $(2, 0)$ is a double point with two distinct tangents; such a point is called a *node* (or a *nodal singularity*).

Note that the system of equations satisfied by singular points is over-represented, since there are 3 equations and only 2 variables. If you choose a curve 'at random', you would expect the first two of these equations to have only finitely many solutions, and it is rather a fluke if one of these solutions also happens to satisfy the third equation. So, a 'typical' curve will be smooth.

A useful tool, for computing singularities and other purposes, is the idea of the resultant of two polynomials.

Definition 1.14. Let $f(x) = f_m x^m + \ldots + f_0$ and $g(x) = g_n x^n + \ldots + g_0$, where $f_m \neq 0$ and $g_n \neq 0$. The *resultant* of $f(x)$ and $g(x)$, denoted $\mathrm{Res}\,(f(x), g(x))$ or just $\mathrm{Res}(f, g)$, is the determinant of the following $(m+n) \times (m+n)$ matrix:

$$\begin{pmatrix} \overbrace{} & \overbrace{n-1} & & & \overbrace{m+1} & & & \\ 0 & \cdots & 0 & f_m & f_{m-1} & \cdots & f_1 & f_0 \\ 0 & \cdots & f_m & f_{m-1} & f_{m-2} & \cdots & f_0 & 0 \\ \vdots & \iddots & & \iddots & & & & \vdots \\ f_m & \cdots & & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & g_n & g_{n-1} & \cdots & g_1 & g_0 \\ 0 & \cdots & g_n & g_{n-1} & g_{n-2} & \cdots & g_0 & 0 \\ & \iddots & & \iddots & & & & \vdots \\ g_n & \cdots & g_2 & g_1 & g_0 & \cdots & 0 & 0 \end{pmatrix}$$

The following are easy to show.

Lemma 1.15. *Let $f(x), g(x) \in R[x]$ be polynomials of degree m, n, respectively, defined over a commutative ring R.*

(a) *There exist polynomials $p(x) \in R[x]$, of degree at most $n-1$, and $q(x) \in R[x]$, of degree at most $m-1$, such that: $p(x)f(x) + q(x)g(x) = \mathrm{Res}\,(f(x), g(x))$.*

(b) *When R is a field, $\mathrm{Res}\,(f(x), g(x)) = 0 \iff f(x)$ and $g(x)$ have a non-constant common factor.*

Definition 1.16. The *discriminant* of a degree n polynomial $f(x) = f_n x^n + \ldots f_0$ is given by: $\mathrm{Disc}(f) = \mathrm{Res}(f, f')/f_n$.

Comment 1.17. (a) Given a monic polynomial $f(x) \in R[x]$, there exist polynomials $p(x), q(x) \in R[x]$ such that $p(x)f(x) + q(x)f'(x) = \mathrm{Disc}(f)$.

(b) $\mathrm{Disc}(f) = 0 \iff f$ and f' have a common root $\iff f$ has a repeated root. For example, $\mathrm{Disc}(x^3 - 2x^2 + x) = 0$, whereas $\mathrm{Disc}(x^2 + 1) \neq 0$.

Example 1.18. Let $f(x) = ax^2 + bx + c$. Then $\mathrm{Disc}(f) = \mathrm{Res}(f, f')/a$

$$= \mathrm{Res}(ax^2 + bx + c, 2ax + b)/a = \frac{1}{a}\begin{vmatrix} a & b & c \\ 0 & 2a & b \\ 2a & b & 0 \end{vmatrix} = b^2 - 4ac,$$

which is the discriminant you know from school, appearing under the square root sign in the quadratic formula.

Example 1.19. Let $f(x) = x^3 + Ax + B$. Then $\mathrm{Disc}(f) = \mathrm{Res}(f, f')$

$$= \mathrm{Res}(x^3 + Ax + B, 3x^2 + A) = \begin{vmatrix} 0 & 1 & 0 & A & B \\ 1 & 0 & A & B & 0 \\ 0 & 0 & 3 & 0 & A \\ 0 & 3 & 0 & A & 0 \\ 3 & 0 & A & 0 & 0 \end{vmatrix} = 4A^3 + 27B^2.$$

Example 1.20. An application of resultants to singularities is as follows. Consider the curve $\mathcal{C} : y^2 = x^3 + Ax + B$ (i.e. $g(x, y) = x^3 + Ax + B - y^2 = 0$), where $A, B \in K$, a field of characteristic not equal to 2. Suppose (x_0, y_0) is a singular point on $\mathcal{C}$, so that:

$$(1)\ g(x_0, y_0) = 0, \quad (2)\ \frac{\partial g}{\partial x}(x_0, y_0) = 0, \quad (3)\ \frac{\partial g}{\partial y}(x_0, y_0) = 0,$$

giving:

$$(1)\ y_0^2 = x_0^3 + Ax_0 + B, \quad (2)\ 3x_0^2 + A = 0, \quad (3)\ 2y_0 = 0.$$

Since the characteristic of K is not equal to 2, we know that $2 \neq 0$, and so (3) gives $y_0 = 0$. Substituting this into (1) tells us that x_0 is a root of $x^3 + Ax + B$, and (2) tells us that x_0 is a root of its derivative; this is possible exactly when $x^3 + Ax + B$ has a repeated root – in other words, when $\mathrm{Disc}(x^3 + Ax + B) = 0$. We have already seen in Example 1.19 that $\mathrm{Disc}(x^3 + Ax + B) = 4A^3 + 27B^2$.

In summary, the curve $\mathcal{C}$ is smooth if and only if $4A^3 + 27B^2 \neq 0$.

Another basic idea in geometry applies to situations where $f(x, y)$ itself has a proper factorisation, for example: $\mathcal{C} : f(x, y) = x^2 - y^2 = 0$. This is a quadratic curve, but it factors as $(x + y)(x - y) = 0$, and so the graph of $\mathcal{C}$ is just the union of the graphs of the lines $x + y = 0$ and $x - y = 0$. This seems geometrically different from curve such as $x^2 - y^2 + 1 = 0$, which has no such factorisation. This is formalised in the following definition.

Definition 1.21. Let $\mathcal{C} : f(x, y) = 0$ be a curve defined over K, and let L be any field containing K. We say that $\mathcal{C}$ is *irreducible over* L if $f(x, y)$ cannot be expressed as a product of two polynomials, both of degree $\geqslant 1$ and both defined over L (by the word *irreducible* on its own, we mean irreducible over K). For any $\mathcal{C} : f(x, y) = 0$, we can write f uniquely (up to constants and reordering) as a product $f = f_1 f_2 \dots f_n$, where $f_1, \dots, f_n$ are irreducible over L. The curves $\mathcal{C}_1 : f_1(x, y) = 0, \dots, \mathcal{C}_n : f_n(x, y) = 0$ are called the *irreducible components* of $\mathcal{C}$ over L.

Examples 1.22.

(a) $\mathcal{C} : f(x, y) = y^2 - 2x^2 = 0$, defined over $\mathbb{Q}$. This is irreducible (by which we mean irreducible over $\mathbb{Q}$), but it becomes reducible over $\mathbb{C}$, with irreducible components $\mathcal{C}_1 : y = \sqrt{2}x$ and $\mathcal{C}_2 : y = -\sqrt{2}x$.

(b) $\mathcal{C} : f(x, y) = y^4 - x^4 = 0$ is reducible. Its irreducible components (over $\mathbb{Q}$) are: $y - x = 0, y + x = 0, y^2 + x^2 = 0$. The last of these becomes reducible over $\mathbb{C}$, and the irreducible components over $\mathbb{C}$ are: $y - x = 0, y + x = 0, y + ix = 0, y - ix = 0$.

It is also helpful to formalise the relationship between curves such as $x^2 + y^3 - 5 = 0$ and $(x + 1)^2 + y^3 - 5 = 0$, where there are maps from one to the other. In this case, one can map each curve to the other with a linear map, but more generally we consider maps between curves described by rational functions (quotients of polynomials).

Definition 1.23. Let $\mathcal{C} : f(x, y) = 0$ and $\mathcal{C}' : g(x, y) = 0$ be curves over K. A *rational map* $\underline{\phi}$ over L from $\mathcal{C}$ to $\mathcal{C}'$ is a map given by a pair ϕ_1, ϕ_2 of rational functions in x, y, defined over L (i.e. ϕ_1, ϕ_2 are both of the form $\frac{\text{polynomial in } x, y}{\text{polynomial in } x, y}$ and the coefficients of ϕ_1, ϕ_2 are in L), with the property that, given any point $P = (x_0, y_0)$ on $\mathcal{C}$, then the point $(\phi_1(x_0, y_0), \phi_2(x_0, y_0))$ lies on $\mathcal{C}'$ (for all but finitely many points (x_0, y_0) at which the denominators of ϕ_1, ϕ_2 are 0). If there also exists a rational map $\underline{\psi} = (\psi_1(x, y), \psi_2(x, y))$ from $\mathcal{C}'$ to $\mathcal{C}$ such that $\underline{\psi}\ \underline{\phi}$ is the identity on $\mathcal{C}$ and $\underline{\phi}\ \underline{\psi}$ is the identity on $\mathcal{C}'$ then we say that $\underline{\phi}$ is a *birational transformation* over L from $\mathcal{C}$ to $\mathcal{C}'$ and that $\mathcal{C}$ and $\mathcal{C}'$ are *birationally equivalent* over L.

Examples 1.24.

(a) Let $\mathcal{C} : x^4 + y^4 = 1$ (i.e. $f(x, y) = x^4 + y^4 - 1 = 0$) and let $\mathcal{C}' : x^4 + y^2 = 1$ (i.e. $g(x, y) = x^4 + y^2 - 1 = 0$). Define $\underline{\phi} : \mathcal{C} \to \mathcal{C}'$ by $\underline{\phi}(x, y) = (x, y^2)$ (in the notation of

Definition 1.23: $\phi_1(x,y) = x$ and $\phi_2(x,y) = y^2$). This is a rational map from $\mathcal{C}$ to $\mathcal{C}'$ over $\mathbb{Q}$ since, if (x,y) satisfies $\mathcal{C} : x^4 + y^4 = 1$ then $x^4 + (y^2)^2 = 1$ and so (x, y^2) lies on $\mathcal{C}'$. This is a rational map from $\mathcal{C}$ to $\mathcal{C}'$, but it is not a birational transformation, since there is no inverse map ($\underline{\phi}$ is 2-to-1).

(b) Let $\mathcal{C} : x^2 + y^3 - 5 = 0$ and $\mathcal{C}' : (x+1)^2 + y^3 - 5 = 0$. If (x,y) is on $\mathcal{C}$ then $x^2 + y^3 - 5 = 0$ and so $((x-1)+1)^2 + y^3 - 5 = 0$, giving that $(x-1, y)$ lies on $\mathcal{C}'$. The map $\underline{\phi}(x,y) = (\phi_1(x,y), \phi_2(x,y)) = (x-1, y)$ is then a rational map over $\mathbb{Q}$ from $\mathcal{C}$ to $\mathcal{C}'$, and the inverse map is clearly $\underline{\psi}(x,y) = (x+1, y)$. The map $\underline{\phi}$ is a birational transformation from $\mathcal{C}$ to $\mathcal{C}'$ over $\mathbb{Q}$, and so $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{Q}$.

Note that the rational map from $\mathcal{C}$ to $\mathcal{C}'$ is in the opposite direction to the variable replacement which transforms the equations. In the above example, $\underline{\phi}(x,y) = (x-1, y)$ is the map from $\mathcal{C}$ to $\mathcal{C}'$ (in that it maps points on $\mathcal{C}$ to points on $\mathcal{C}'$; for example, the point $(2,1)$ on $\mathcal{C}$ maps to $(1,1)$ on $\mathcal{C}'$), but the variable replacement 'replace x by $x-1$ and y by y' changes the equation for $\mathcal{C}'$ into the equation for $\mathcal{C}$.

(c) Let $\mathcal{C} : x^2 - y^2 = 0$ and $\mathcal{C}' : x^2 + y^2 = 0$. Clearly $\underline{\phi} : \mathcal{C} \to \mathcal{C}'$, defined by $\underline{\phi}(x,y) = (x, iy)$ is a rational map from $\mathcal{C}$ to $\mathcal{C}'$, with inverse $\underline{\psi}(x,y) = (x, -iy)$. This shows that $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{C}$. However, $\mathcal{C}$ and $\mathcal{C}'$ are not birationally equivalent over $\mathbb{Q}$, since any such map would take the infinitely many members of $\mathcal{C}(\mathbb{Q})$ to infinitely many members of $\mathcal{C}'(\mathbb{Q})$, contradicting the fact that $\mathcal{C}'(\mathbb{Q}) = \{(0,0)\}$.

(d) Let $\mathcal{C} : y^2 = x^4 + 3x^2 + 5$ and $\mathcal{C}' : y^2 = 5x^4 + 3x^2 + 1$. Define $\underline{\phi}(x,y) = (\frac{1}{x}, \frac{y}{x^2})$. If (x,y) is a point on $\mathcal{C}$ then $y^2 = x^4 + 3x^2 + 5$ and so $\frac{y^2}{x^4} = 1 + \frac{3}{x^2} + \frac{5}{x^4}$, giving: $\left(\frac{y}{x^2}\right)^2 = 1 + 3\left(\frac{1}{x}\right)^2 + 5\left(\frac{1}{x}\right)^4$, so that $\left(\frac{1}{x}, \frac{y}{x^2}\right)$ is a point on $\mathcal{C}'$. Our map $\underline{\phi}$ is then a rational map (over $\mathbb{Q}$) from $\mathcal{C}$ to $\mathcal{C}'$. The inverse map is $\underline{\psi}(x,y) = (\frac{1}{x}, \frac{y}{x^2})$ (check that $\underline{\psi}\left(\underline{\phi}(x,y)\right) = \underline{\psi}\left(\frac{1}{x}, \frac{y}{x^2}\right) = \left(\frac{1}{1/x}, \frac{y/x^2}{(1/x)^2}\right) = (x,y)$, so that $\underline{\psi}\ \underline{\phi}$ is the identity, as is $\underline{\phi}\ \underline{\psi}$). Hence $\underline{\phi}$ is a birational transformation over $\mathbb{Q}$; the curves $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{Q}$.

(e) Let $\mathcal{C} : x^2 + y^2 = 1$ and $\mathcal{C}' : y = 0$. It might at first seem surprising that a circle should be birationally equivalent to a line, but we can establish the map first by fixing a specific point on $\mathcal{C}$, say $P_0 = (-1, 0)$, and mapping a point on $\mathcal{C}$ to $s = \frac{y}{x+1}$, the slope of the line from P_0 to (x,y) (literally, we are mapping it to $(s,0)$). Define: $\underline{\phi}(x,y) = (\frac{y}{x+1}, 0)$ from $\mathcal{C}$ to $\mathcal{C}'$ (defined everywhere except at the point $(-1,0)$, but this is allowed, since the definition of rational map allows us to have a finite number of points where the map is not defined). For the inverse, note that if the slope is s, then the line through P_0 and (x,y) has equation: $y = s(x+1)$; substituting this into $\mathcal{C}$ gives $x^2 + s^2(x+1)^2 = 1$, and so: $(x+1)(x - 1 + s^2(x+1)) = 0$. When $x \neq -1$, this gives $x = \frac{1-s^2}{1+s^2}$ and $y = s(x+1) = \frac{2s}{1+s^2}$. This suggests that, for the inverse map, we should take: $\underline{\psi}(x,y) = \left(\frac{1-x^2}{1+x^2}, \frac{2x}{1+x^2}\right)$. It is straightforward to check that this is indeed a map from $\mathcal{C}'$ to $\mathcal{C}$ (since $\left(\frac{1-x^2}{1+x^2}\right)^2 + \left(\frac{2x}{1+x^2}\right)^2 = 1$ for any x), that $\underline{\psi}\ \underline{\phi} =$ identity on $\mathcal{C}$ and that $\underline{\phi}\ \underline{\psi} =$ identity on $\mathcal{C}'$. Hence $\mathcal{C}$ and $\mathcal{C}'$ are birationally equivalent over $\mathbb{Q}$.

Definition 1.25. A *parametrisation* of a curve $\mathcal{C}$ is a birational equivalence between $\mathcal{C}$ and a line.

Comment 1.26. The birational transformation in Example 1.24(e) is a parametrisation of the circle $x^2 + y^2 = 1$. Note that a parametrisation is an unusual type of birational

transformation, in that it gives a map to a single variable; in this case, $\left(\frac{1-s^2}{1+s^2}, \frac{2s}{1+s^2}\right)$ gives a description of the points on $\mathcal{C}$ in terms of the parameter s. Since the maps ϕ and ψ are defined over $\mathbb{Q}$, this gives a way of describing all $\mathbb{Q}$-rational points on $\mathcal{C}$, namely: $(x, y) \in \mathcal{C}(\mathbb{Q}) \iff s \in \mathbb{Q}$. For example, $s = 2$ gives $\left(-\frac{3}{5}, \frac{4}{5}\right) \in \mathcal{C}(\mathbb{Q})$. This parametrisation can be used to describe all Pythagorean triples.

The curve $x^2 + y^2 = 1$ is a special case of the following class of curves.

Definition 1.27. A *conic* is a smooth quadratic curve. The general form of the equation is $ax^2 + 2bxy + cy^2 + 2dx + 2fy + g = 0$, satisfying

$$\begin{vmatrix} a & b & d \\ b & c & f \\ d & f & g \end{vmatrix} \neq 0 \quad \text{(which guarantees that the curve is smooth).}$$

A conic is an ellipse, hyperbola or parabola; the name 'conic' refers to the fact that these are the curves which can be obtained by intersecting a plane and a double-cone (two cones with the same axis, placed apex to apex). The parametrisation of the circle given in Example 1.24(e) is a special case of the following result.

Theorem 1.28. *Any conic $\mathcal{C}$ (over K) with a K-rational point is birationally equivalent to a line (i.e. it is parametrisable).*

Proof. We are given that there exists a K-rational point (x_0, y_0) on the curve $\mathcal{C} : f(x, y) = 0$. Let $g(x, y) = f(x + x_0, y + y_0)$. This contains the point $(0, 0)$ so that we can write: $g(x, y) = g_1(x, y) + g_2(x, y)$, where g_1 is homogeneous & linear, and g_2 is homogeneous & quadratic. Hence $g(x, tx) = x\phi_1(t) + x^2\phi_2(t) = 0$. Apart from $x = 0$, we can take $x = -\phi_1(t)/\phi_2(t), y = -t\phi_1(t)/\phi_2(t)$ (with inverse $t = y/x$) as a parametrisation of $g(x, y) = 0$. The parametrisation of $\mathcal{C}$ is then: $x = x_0 - \phi_1(t)/\phi_2(t), y = y_0 - t\phi_1(t)/\phi_2(t)$ (with inverse $t = (y - y_0)/(x - x_0)$). □

Definition 1.29. The curves $\mathcal{C} : f(x, y) = 0$ and $\mathcal{C}' : g(x, y) = 0$ *intersect* at $P = (x_0, y_0)$ if P lies on both of $\mathcal{C}$ and $\mathcal{C}'$ [that is, $f(x_0, y_0) = g(x_0, y_0) = 0$].

Definition 1.30. Suppose the curves $\mathcal{C} : f(x, y) = 0$ and $\mathcal{C}' : g(x, y) = 0$ intersect at $P = (x_0, y_0) \in \mathcal{C}(L)$ (with L a field containing the field of definition of the curve). The curves *intersect with multiplicity* $r > 0$ at P if the dimension of the quotient ring

$$\dim_L L[\![x, y]\!]/(f(x + x_0, y + y_0), g(x + x_0, y + y_0)) = r.$$

The intersection multiplicity is ∞ if and only if $\mathcal{C}$ and $\mathcal{C}'$ have a common irreducible component containing P. We refer to Fulton *Algebraic Curves* for details and proofs of the fundamental properties of the intersection multiplicity. You can also take a look at Part B Algebraic Curves for an approach via resultants.

The proofs of the following two lemmas can be found in the preliminary reading file.

Lemma 1.31. *Consider a curve $\mathcal{C} : f(x, y) = 0$ over K and a line $\mathcal{D}$ parameterised by $x = at + b$, $y = ct + d$, with $a, b, c, d \in K$ and a, c not both zero. Then $\mathcal{C}$ and $\mathcal{D}$ intersect at the points $P = (at_0 + b, ct_0 + d)$ with t_0 a root of the polynomial $F(t) = f(at + b, ct + d)$. If $F(t)$ is identically 0, then $\mathcal{C}$ contains the line $\mathcal{D}$.*

Suppose $t_0 \in \overline{K}$ is a root of $F(t)$ and let $P = (at_0 + b, ct_0 + d)$. Then $\mathcal{C}$ and $\mathcal{D}$ intersect at P with multiplicity equal to the multiplicity of t_0 as a root of $F(t)$.

Lemma 1.32. *Suppose $\mathcal{C}$ and $\mathcal{C}'$ are two curves intersecting at a point $P \in \mathcal{C}(K) \cap \mathcal{C}'(K)$. Suppose moreover that P is a nonsingular point on both curves. Then the intersection multiplicity at P is > 1 if and only if the tangent lines to $\mathcal{C}$ and $\mathcal{C}'$ at P coincide.*

Example 1.33. Let $\mathcal{C} : y^2 = x^3 + 2x + 1$ and $\mathcal{D} : y = x + 1$. On substituting $\mathcal{D}$ into $\mathcal{C}$ we see that the x-coordinate of any point of intersection must satisfy $(x+1)^2 = x^3 + 2x + 1$, and so $x^2(x-1) = 0$, giving only $x = 0, 1$ as possibilities. Substituting $x = 0$ in $\mathcal{D}$ gives $y = 1$; substituting $x = 1$ in $\mathcal{D}$ gives $y = 2$. So, the only possible points of intersection are $(0, 1)$ and $(1, 2)$ [and these do indeed lie on $\mathcal{C}$ and $\mathcal{D}$]. It also follows from Lemma 1.31 that the intersection multiplicities at these points are 2 and 1 respectively.

Comment 1.34. For more complicated examples, we cannot always find the points of intersection by a straightforward substitution of one equation into the other. Given two curves $\mathcal{C} : f(x, y) = 0$ of degree m and $\mathcal{D} : g(x, y) = 0$ of degree n, a systematic approach to finding the points of intersection is possible via resultants. One initially picks one of the variables, y say, and computes the resultant of $f(x, y)$ and $g(x, y)$, regarded as polynomials in y, by writing them as: $f(x, y) = f_m(x)y^m + \ldots + f_0(x)$, and similarly for $g(x, y)$. The matrix in Definition 1.14 will have entries that are polynomials in x, and consideration of the degrees of these polynomials shows that the resultant of $f(x, y)$ and $g(x, y)$ (regarded as polynomials in y) will be a polynomial in x of degree at most mn. Any point of intersection of $\mathcal{C}$ and $\mathcal{D}$ must have x-coordinate which is a root of the at-most-degree-mn polynomial. For each value of x, one can then substitute back into $\mathcal{C}$ and $\mathcal{D}$ to find the corresponding y-coordinates.

Projective curves. There are several respects in which affine space is unsatisfying. Consider, for example, the true statement in affine space: two distinct lines meet at exactly one point, except when parallel. It would be much nicer to have a cleaner statement, in which we remove 'except when parallel'. Intuitively, parallel lines intersect 'at infinity', given that the point of intersection shoots off to infinity as two lines become closer and closer to parallel.

Similarly, consider the affine curves: $\mathcal{C} : y^2 = x^3 + 1$ and $D : y = x + 1$; these meet at the points $(-1, 0), (0, 1), (2, 3)$, each with multiplicity 1. On trying other lines in place of $\mathcal{D}$, one typically finds again that there are 3 points of intersection (when counted with multiplicity). An apparent exception is $\mathcal{D} : x = 0$, which intersects $\mathcal{C}$ only at two points, $(0, 1)$ and $(0, -1)$, and this is true for any vertical line. We seem to have a rule: any line intersects $\mathcal{C}$ at exactly 3 points (counted with multiplicity) except when the line is vertical. Again, we would like a cleaner statement, in which we remove 'except when the line is vertical'. Again, the third point of intersection seems to be 'at infinity'.

Points at infinity are intuitively points (x, y) where there is a denominator of 0. We cannot express this idea using only pairs (x, y), where x, y lie in a field K. A natural approach is to write: $x = X/Z, y = Y/Z$ and identify the point (x, y) with the triple (X, Y, Z). As long as $Z \neq 0$, we can go in the other direction from the triple (X, Y, Z) to (x, y). Note that, for any $k \in K^*$, the triple (kX, kY, kZ) corresponds to $(kX/kZ, kY/kZ) = (X/Z, Y/Z) = (x, y)$, and so we impose a relation, that two triples are regarded as being the same if they are nonzero scalar multiples of each other. Subject to this relation, there is then a $1-1$ correspondence between (x, y) and triples (X, Y, Z) with $Z \neq 0$. On the other hand, the triples (X, Y, Z) with $Z = 0$ do not correspond to any $x, y \in K$, and such triples give us a way of describing formally these new points at infinity.

Definition 1.35. Let K be a field. $\mathbb{P}^n(K) = \{(x_0, \ldots, x_n) : x_0, \ldots, x_n \in K, \text{ not all } 0\}$, subject to the relation that $(x_0, \ldots x_n) = (y_0, \ldots, y_n)$ in $\mathbb{P}^n(K)$ if there exists $r \in K, r \neq 0$, such that $(y_0, \ldots, y_n) = (rx_0, \ldots rx_n)$. $\mathbb{P}^n(K)$ is called *projective* n-space over K.

Example 1.36. $(1, 2, 3) = (3, 6, 9)$ in $\mathbb{P}^2(\mathbb{Q})$. (N.B. $(0, 0, 0) \notin \mathbb{P}^2(\mathbb{Q})$.)

Definition 1.37. A *polynomial in n projective variables* is an $(n + 1)$-variable homogeneous polynomial. A *projective curve* in $\mathbb{P}^2$ is defined by a homogeneous polynomial in 3 variables $F(X, Y, Z) = 0$, for example, $X^3 + Y^3 - Z^3 = 0$.

Definition 1.38. Let $\mathcal{C} : f(x, y) = 0$ be an (affine) curve. The *homogenisation* of $\mathcal{C}$ is the projective curve $F(X, Y, Z) = 0$ of the same degree as $f(x, y)$, with the property that $F(x, y, 1) = f(x, y)$. A point (X_0, Y_0, Z_0) on $F(X, Y, Z) = 0$ with $Z_0 = 0$ is called a *point at infinity* on $\mathcal{C}$. When $Z_0 \neq 0$, the point (X_0, Y_0, Z_0) corresponds to $(X_0/Z_0, Y_0/Z_0)$ on $f(x, y) = 0$.

Example 1.39. Let $\mathcal{C} : y^2 = 4x^2 + 1$, so that $f(x, y) = y^2 - 4x^2 - 1 = 0$. The associated projective curve (the homogenisation) is: $Y^2 = 4X^2 + Z^2$ (so that $F(X, Y, Z) = Y^2 - 4X^2 - Z^2$). The two points at infinity are: $(1, 2, 0)$ and $(1, -2, 0)$.

Example 1.40. For the curve $\mathcal{C} : y^2 = x^3 + 1$, the associated projective curve is $ZY^2 = X^3 + Z^3$. To find the points at infinity (the points where $Z = 0$), substitute $Z = 0$ into the equation, giving $X^3 = 0$ and so $X = 0$. This forces $Y \neq 0$ (since $(0, 0, 0)$ is not allowed as a point in $\mathbb{P}^2$). So, the points at infinity are of the form $(0, Y, 0)$, where $Y \neq 0$. But these are all the same in $\mathbb{P}^2$, since they are scalar multiples of each other; therefore this is exactly one point at infinity, which we can represent by $(0, 1, 0)$, say.

Comment 1.41. Two distinct affine lines $a_1x+b_1y+c_1 = 0$ and $a_2x+b_2y+c_2 = 0$ meet at exactly one point, except when parallel. For example, $x+y+2 = 0$ and $x+y+3 = 0$ do not intersect. For projective lines, the rule is the same, but we can remove the phrase 'except when parallel'. For example, the projective lines $X + Y + 2Z = 0$ and $X + Y + 3Z = 0$ have $(1, -1, 0)$ as the unique point of intersection.

Definition 1.42. A projective curve $F(X, Y, Z) = 0$ has a *singularity* at (X_0, Y_0, Z_0) when:
$F(X_0, Y_0, Z_0) = \frac{\partial F}{\partial X}(X_0, Y_0, Z_0) = \frac{\partial F}{\partial Y}(X_0, Y_0, Z_0) = \frac{\partial F}{\partial Z}(X_0, Y_0, Z_0) = 0.$

Lemma 1.43. *Suppose (X_0, Y_0, Z_0) is a nonsingular point on the projective curve $F(X, Y, Z) = 0$. Then the tangent line at the point (X_0, Y_0, Z_0) has equation*

$$\frac{\partial F}{\partial X}(X_0, Y_0, Z_0)X + \frac{\partial F}{\partial Y}(X_0, Y_0, Z_0)Y + \frac{\partial F}{\partial Z}(X_0, Y_0, Z_0)Z = 0.$$

Proof. Permuting the coordinates and rescaling if necessary, we may assume that $Z_0 = 1$. Then our point lies on the affine curve $F(x, y, 1) = 0$. We compute the tangent line as in Comment 1.9 and get equation

$$\frac{\partial F}{\partial X}(X_0, Y_0, 1)(x - X_0) + \frac{\partial F}{\partial Y}(X_0, Y_0, 1)(y - Y_0) = 0.$$

Homogenising gives the projective tangent line

$$\frac{\partial F}{\partial X}(X_0, Y_0, 1)(X - X_0Z) + \frac{\partial F}{\partial Y}(X_0, Y_0, 1)(Y - Y_0Z) = 0.$$

Finally, we use the fact that

$$\frac{\partial F}{\partial X}(X_0, Y_0, 1)X_0 + \frac{\partial F}{\partial Y}(X_0, Y_0, 1)Y_0 + \frac{\partial F}{\partial Z}(X_0, Y_0, 1) = 0$$

which is a consequence of Euler's identity (which can be checked on monomials):

$$\frac{\partial F}{\partial X}(X, Y, Z)X + \frac{\partial F}{\partial Y}(X, Y, Z)Z + \frac{\partial F}{\partial Z}(X, Y, Z)Z = \deg(F)F(X, Y, Z).$$

□

Comment 1.44. Note that, by multiplying through by denominators, we can take rational maps and birational transformations between projective curves to be of the form:

$$\underline{\phi}(X, Y, Z) = (\phi_1(X, Y, Z), \phi_2(X, Y, Z), \phi_3(X, Y, Z)),$$

where ϕ_1, ϕ_2, ϕ_3 are homogeneous polynomials, rather than rational functions.

Comment 1.45. Suppose that two projective curves $F(X, Y, Z) = 0$ and $G(X, Y, Z) = 0$ have a point of intersection (X_0, Y_0, Z_0). The multiplicity of intersection can always be computed by using some associated affine curve. At least one of X_0, Y_0, Z_0 must be nonzero, since $(0, 0, 0)$ is not allowed in $\mathbb{P}^2$. If $Z_0 \neq 0$ then the multiplicity of intersection is the same as that of $(X_0/Z_0, Y_0/Z_0)$ on the affine curves $F(x, y, 1) = 0$ and $G(x, y, 1) = 0$ (here, $x = X/Z, y = Y/Z$). If $Y_0 \neq 0$ then one can use $F(x, 1, z), G(x, 1, z)$, where $x = X/Y, z = Z/Y$. If $X_0 \neq 0$ then one can use $F(1, y, z), G(1, y, z)$, where $y = Y/X, z = Z/X$.

We can now state one of the basic results in the projective geometry of curves, generalising the fact that two projective lines have a unique point of intersection.

Theorem 1.46. *[Bézout's Theorem] Two projective curves, with no common component, of degrees m, n intersect at precisely mn points, counted with multiplicity.*

Example 1.47. The projective curves $ZY^2 = X^3 + Z^3$ and $X = 0$ intersect at the points $(0, 1, 1), (0, -1, 1), (0, 1, 0)$, each with multiplicity 1.

Elliptic Curves. Curves can be classified according to a property called *genus*, which is invariant under birational equivalence. We shall not go into the technicalities of what precisely is meant by genus, and its properties, which would be an entire lecture course in its own right. The simplest type are curves of genus 0, which can be defined by quadratic and linear equations. Recall from Theorem 1.28 that any conic with a rational point can be parametrised.

Curves of genus 1 are the next natural class of curves to consider; they are, in a sense, the next 'simplest' type of curve after conics. Please don't confuse 'elliptic curves' (which are of genus 1) with ellipses (which are of genus 0). The classical terminology comes from a relationship between cubic curves and elliptic integrals, which were much studied in the 19th century. It can be shown that a curve of genus 1 is not parametrisable. An *elliptic curve* over K is defined to be a nonsingular projective curve of genus 1, defined over K, together with a K-rational point on the curve. It can also be shown that any curve of genus 1 is birationally equivalent over K to a nonsingular projective cubic curve. For the purposes of this lecture course, you can forget about the term 'genus' and will simply take this as the definition of an elliptic curve.

Section 2. The Group Law on an Elliptic Curve

Definition 2.1. An elliptic curve over a field K is a nonsingular projective cubic curve, defined over K, with a specified K-rational point.

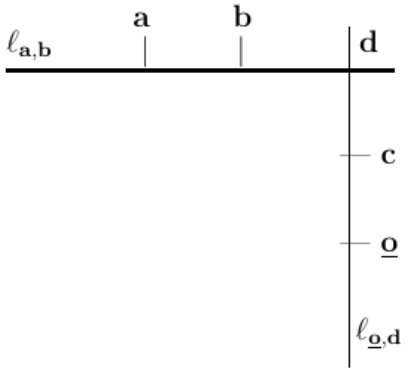
This means that an elliptic curve is defined by a degree 3 homogeneous polynomial, in 3 variables and with coefficients in K.

Remark 2.2. We won't discuss this in the course, but elliptic curves over the complex numbers have a simple description following from the Weierstrass uniformization theorem (see Chapter VI in [2] for more details). This says that for an elliptic curve $\mathcal{C}$ defined over $\mathbb{C}$, we can make an identification

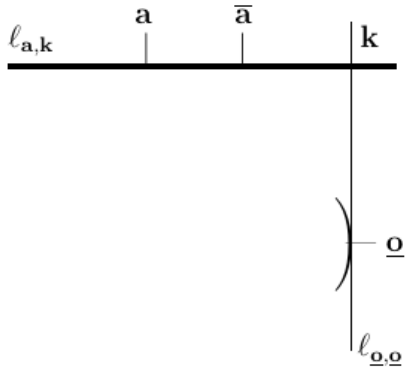
$$\mathcal{C}(\mathbb{C}) \cong \mathbb{C}/\Lambda$$

where Λ is a free rank two abelian group generated by two complex numbers ω_1, ω_2 which are linearly independent over $\mathbb{R}$. The identification sends the specified point of $\mathcal{C}$ to the coset $0 + \Lambda$. We deduce from this that $\mathcal{C}(\mathbb{C})$ has the structure of an abelian group. It turns out that this group law can be defined purely algebraically which is what we are going to do next.

Definition 2.3. Let $\mathcal{C} : F(X, Y, Z) = 0$ be an elliptic curve $/K$ (the notation $/K$ means 'defined over K'; that is, all of the coefficients of $\mathcal{C}$ are in the field K). So, $\mathcal{C}$ is a nonsingular projective cubic curve, with a K-rational point, which we shall denote $\underline{\mathbf{o}}$. For any two points $\mathbf{a}, \mathbf{b}$ on $\mathcal{C}$ (defined over a common extension field L/K), let $\ell_{\mathbf{a},\mathbf{b}}$ denote the line which meets $\mathcal{C}$ at $\mathbf{a}, \mathbf{b}$ (if $\mathbf{a}, \mathbf{b}$ are distinct then $\ell_{\mathbf{a},\mathbf{b}}$ is the unique line through $\mathbf{a}, \mathbf{b}$; if $\mathbf{a} = \mathbf{b}$ then $\ell_{\mathbf{a},\mathbf{b}}$ is the line tangent to $\mathcal{C}$ at $\mathbf{a} = \mathbf{b}$).



Let $\ell_{\mathbf{a},\mathbf{b}}$ denote the line which meets $\mathcal{C}$ at $\mathbf{a}, \mathbf{b}$.
Then $\ell_{\mathbf{a},\mathbf{b}}$ and $\mathcal{C}$ have 3 points of intersection (Bézout).
Let $\mathbf{d}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\mathbf{a},\mathbf{b}}$.
Now, let $\ell_{\underline{\mathbf{o}},\mathbf{d}}$ denote the line which meets $\mathcal{C}$ at $\underline{\mathbf{o}}$ and $\mathbf{d}$.
Let $\mathbf{c}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\underline{\mathbf{o}},\mathbf{d}}$.
Define $\mathbf{a} + \mathbf{b} = \mathbf{c}$.



Let $\ell_{\underline{\mathbf{o}},\underline{\mathbf{o}}}$ be the line tangent to $\mathcal{C}$ at $\underline{\mathbf{o}}$.
Let $\mathbf{k}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\underline{\mathbf{o}},\underline{\mathbf{o}}}$.
Now, let $\ell_{\mathbf{a},\mathbf{k}}$ be the line which meets $\mathcal{C}$ at $\mathbf{a}$ and $\mathbf{k}$.
Let $\overline{\mathbf{a}}$ be the third point of intersection between $\mathcal{C}$ and $\ell_{\mathbf{a},\mathbf{k}}$.
Define $-\mathbf{a}$ to be $\overline{\mathbf{a}}$.

We shall soon show that $\mathbf{a}+\mathbf{b}$ is a commutative group law on the points on $\mathcal{C}$, with identity $\underline{\mathbf{o}}$ and the inverse of $\mathbf{a}$ given by $-\mathbf{a}$. A priori, the various new points we constructed in the above diagrams will be defined over a chosen algebraic closure $\overline{L}$ of the field of definition for the points $\mathbf{a}, \mathbf{b}$. But we will show that these points are all L-rational.

First we need the following technical lemma.

Lemma 2.4. *Let $P_1, \dots, P_8$ be such that no 4 points lie on a line and no 7 points lie on a conic. Then there exists a unique point P_9 which is a 9th point of intersection of any two cubics passing through $P_1, \dots, P_8$.*

Optional Proof. See 0.140.

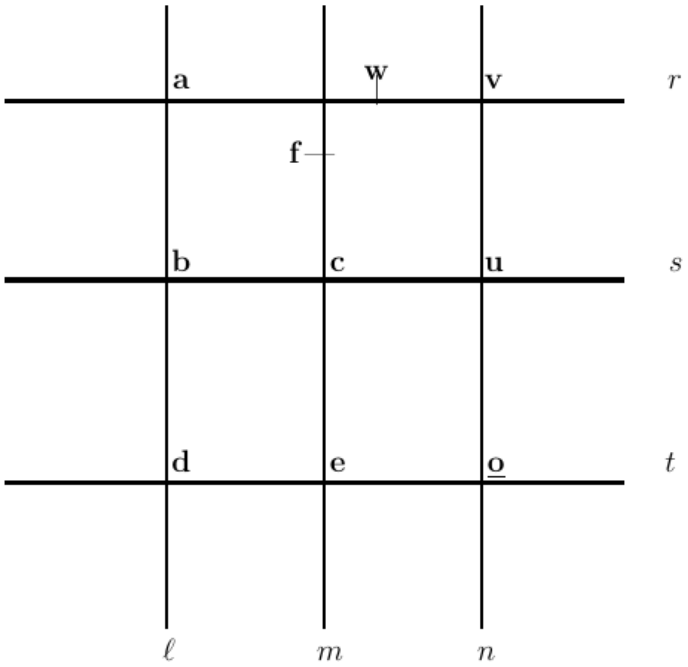
Theorem 2.5. *Let $\mathcal{C}$ be an elliptic curve $/K$, with K-rational point $\underline{\mathbf{o}}$. Then*

$$(\mathbf{a}, \mathbf{b}) \mapsto \mathbf{a}+\mathbf{b},$$

as in Definition 2.3, gives a commutative group law on the points of $\mathcal{C}$, with identity $\underline{\mathbf{o}}$. The inverse of $\mathbf{a}$ is given by the point $-\mathbf{a}$, constructed in Definition 2.3.

Furthermore, the K-rational points $\mathcal{C}(K)$ form a group under this group law, called the Mordell-Weil group[1]*. More generally, for any extension field L/K the L-rational points $\mathcal{C}(L)$ form a group under the group law.*

Proof. It is easy to show commutativity, the fact that $\underline{\mathbf{o}}$ is the identity, and the fact that $-\mathbf{a}$ is the inverse of $\mathbf{a}$. The only difficult problem is associativity. In order to prove associativity, consider the following diagram:



Here, r, s, t, ℓ, m, n are lines. On each line, the labelled points are the points of intersection between $\mathcal{C}$ and that line. From the construction of Definition 2.3, $\mathbf{a}+\mathbf{b}=\mathbf{e}$, and so $(\mathbf{a}+\mathbf{b})+\mathbf{c}$ is the 3rd point of intersection on $\ell_{\underline{\mathbf{o}},\mathbf{f}}$.

Similarly, $\mathbf{b}+\mathbf{c}=\mathbf{v}$, and $\mathbf{a}+(\mathbf{b}+\mathbf{c})$ is the 3rd point of intersection on $\ell_{\underline{\mathbf{o}},\mathbf{w}}$.

[1] Typically this name is reserved for the case where K is a number field.

To show $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$, it is therefore sufficient to show that $\mathbf{f}=\mathbf{w}$. Let $F_1=\ell mn$ and $F_2=rst$, both of which are cubic curves (recall that each line corresponds to a degree one homogeneous polynomials, so their product is a degree three polynomial defining a cubic curve).

Now we observe that $\mathcal{C}$ and F_1 has the following 8 points in common: $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\mathbf{e},\mathbf{u},\mathbf{v},\underline{\mathbf{o}}$ ($\mathbf{f}$ is the 9th common point). $\mathcal{C}$ and F_2 also have same 8 points in common, together with $\mathbf{w}$. From Lemma 2.4, the 9th point of intersection of $\mathcal{C}$ and F_1 must be the same as the 9th point of intersection of $\mathcal{C}$ and F_2; that is, $\mathbf{f}=\mathbf{w}$, as required. Hence, $+$ is a commutative group law.

It remains to show that $\mathcal{C}(K)$ forms a group under $+$. We are given that $\underline{\mathbf{o}}\in\mathcal{C}(K)$. Let $\mathbf{a},\mathbf{b}\in\mathcal{C}(K)$. It is sufficient to show that $\mathbf{a}+\mathbf{b}\in\mathcal{C}(K)$ and that $-\mathbf{a}\in\mathcal{C}(K)$.

Let $\mathbf{a}=(x_1,y_1)$ and $\mathbf{b}=(x_2,y_2)$, where $x_1,y_1,x_2,y_2\in K$. Then the line through $\mathbf{a},\mathbf{b}$ is (in affine form) $\ell_{\mathbf{a},\mathbf{b}}: y=\ell x+m$, where $\ell=\frac{y_1-y_2}{x_1-x_2}\in K$ and $m=\frac{x_1y_2-x_2y_1}{x_1-x_2}\in K$. Substitute $y=\ell x+m$ into the cubic equation for $\mathcal{C}$ to get; $\phi(x)=x^3+c_2x^2+c_1x+c_0=0$, defined over K. Let $\phi(x)=(x-x_1)(x-x_2)(x-x_3)$ be the factorisation of $\phi(x)$. Then x_1,x_2,x_3 are the 3 roots of ϕ and so $x_1+x_2+x_3=-c_2$, giving: $x_3=-c_2-x_1-x_2\in K$ and $y_3=\ell x_3+m\in K$. The line $\ell_{\mathbf{a},\mathbf{b}}$ then meets $\mathcal{C}$ at $\mathbf{a},\mathbf{b},\mathbf{d}=(x_3,y_3)\in\mathcal{C}(K)$. The same argument shows that the line $\ell_{\underline{\mathbf{o}},d}$ through $\underline{\mathbf{o}},\mathbf{d}$ has 3rd point of intersection $\mathbf{c}$ which is also in $\mathcal{C}(K)$. But $\mathbf{c}=\mathbf{a}+\mathbf{b}$ and so we have shown that $\mathbf{a}+\mathbf{b}\in\mathcal{C}(K)$. A similar argument shows that if $\mathbf{a}\in\mathcal{C}(K)$ then $-\mathbf{a}\in\mathcal{C}(K)$. Hence $\mathcal{C}(K)$ is a group, as required. The same argument applies when we replace K by any extension field L/K. □

Aside: It is apparent that, in the above proof, we have dealt with the 'typical' case, where none of our points are repeated (for the proof of associativity), and none are at infinity (for the proof that $\mathcal{C}(K)$ is a group, since the points were written in affine form). It is straightforward to check these special cases; we shall not bother to do so here.

Comment 2.6. When two nonsingular cubics $\mathcal{C}_1,\mathcal{C}_2$ are birationally equivalent over K (under $\phi:\mathcal{C}_1\longrightarrow\mathcal{C}_2$), with $\phi(O_1)=O_2$, it can be shown that ϕ induces a group isomorphism between $\mathcal{C}_1(K)$ and $\mathcal{C}_2(K)$. If ϕ is just a rational map, still sending O_1 to O_2, it induces a group homomorphism.

(*For those who have learned some more algebraic geometry.*) A rational map from a nonsingular curve to a projective variety always extends to a morphism of varieties. In particular, it can be defined at every point of the curve, so the birational transformation ϕ automatically induces a bijection between the sets of K-rational points.

Comment 2.7. By an elliptic curve, we shall always mean a projective curve, but often write the equation in affine form. Note that, whichever way it is written, we are always referring to the projective curve. For example, if we say 'let $\mathcal{C}: y^2=x^3+3$ be an elliptic curve', it should be understood that this is a shorthand notation for the corresponding projective curve $ZY^2=X^3+3Z^3$.

Theorem 2.8. *Let K be a field satisfying* $\mathrm{char}(K)\neq 2,3$ *(recall – this means that $1+1\neq 0$ and $1+1+1\neq 0$). Then any elliptic curve over K is birationally equivalent over K to a curve of the form $y^2=x^3+Ax+B$, with the birational transformation sending the identity $\underline{\mathbf{o}}$ to the point at infinity ($(0:1:0)$ in projective coordinates).*

When $K=\mathbb{Q}$, we can birationally transform any $y^2=$ cubic in x to a curve of the form $y^2=x^3+Ax+B$, with $A,B\in\mathbb{Z}$, using only maps of the form $(x,y)\mapsto(ax+b,cy)$.

Comment 2.9. Let K be a field satisfying $\mathrm{char}(K) \neq 2, 3$, and let $g(x)$ be a quartic polynomial over K with nonzero discriminant. It can be shown that any affine curve

$$\mathcal{D} : y^2 = g(x)$$

with a K-rational point, is birationally equivalent over K to an elliptic curve $\mathcal{C}$ of the form $y^2 = x^3 + Ax + B$ (see p. 35 of [1]). Note that the point at infinity $(0 : 1 : 0)$ in the homogenisation of $\mathcal{D}$ is singular. The affine curve $\mathcal{D}$ is sometimes called an 'affine model' for the elliptic curve $\mathcal{C}$.

Comment 2.10. When $\mathrm{char}(K) \neq 2, 3$, we shall typically take our elliptic curves to have the form

$$\mathcal{E} : y^2 = x^3 + Ax + B, \text{ where } A, B \in K,$$

which should be regarded as shorthand for the projective curve $ZY^2 = X^3 + AXZ^2 + BZ^3$. Sometimes it will be convenient to include an x^2 term. Since $\mathcal{E}$ is nonsingular, we must have $\Delta = 4A^3 + 27B^2 \neq 0$, as was shown in Example 1.20 (note the assumption there that $\mathrm{char}(K) \neq 2$). The notation $\Delta = 4A^3 + 27B^2$ is standard.

It is conventional to choose $\underline{\mathbf{o}} = (0, 1, 0)$, the point at infinity, as the identity (we shall always take $\underline{\mathbf{o}} = (0, 1, 0)$ unless otherwise stated). Note that the line $Z = 0$ meets $\mathcal{E}$ at $\underline{\mathbf{o}}$ three times (such a point is called an *inflexion*). Given a point $\mathbf{a} = (X, Y, Z)$, if we take the line through $\mathbf{a}$ and $\underline{\mathbf{o}} = (0, 1, 0)$ then the third point of intersection is $(X, -Y, Z)$, which must then be $-\mathbf{a}$. In affine form:

$$-(x, y) = (x, -y).$$

This gives an easy rule for finding the inverse of a point, under the group law, namely: the inverse of $\mathbf{a}$ is its reflection in the x-axis.

So, for an elliptic curve $\mathcal{E}$ written in the form $y^2 =$ cubic in x, the points are $\underline{\mathbf{o}}$ (the point at infinity) and the affine points (x, y), and the group law has a simpler description:

Let $\mathbf{d} = (x_3, y_3)$ the 3rd point of intersection of $\mathcal{E}$ and $\ell_{\mathbf{a},\mathbf{b}}$.

Then $\mathbf{a} + \mathbf{b} = (x_3, -y_3)$, the reflection of $\mathbf{d}$ in the x-axis.

We illustrate the group law with the following computation.

Example 2.11. Let $\mathcal{E} : y^2 = x^3 + 1$. Let us compute $\mathbf{a} + \mathbf{b}$, where $\mathbf{a} = (x_1, y_1) = (-1, 0)$ and $\mathbf{b} = (x_2, y_2) = (0, 1)$.

The line through $\mathbf{a}, \mathbf{b}$ is $\ell_{\mathbf{a},\mathbf{b}} : y = x + 1$. Substituting this into $\mathcal{E}$, we see that the x-coordinate of any point of intersection satisfies: $(x + 1)^2 = x^3 + 1$, and so:

$$x^3 - x^2 - 2x = 0. \qquad (*)$$

We are looking for (x_3, y_3), the 3rd point of intersection of $\mathcal{E}$ and $\ell_{\mathbf{a},\mathbf{b}}$. We first find x_3; note that x_1, x_2, x_3 must be the roots of $(*)$.

Method A (for finding x_3). Since the roots of $(*)$ are x_1, x_2, x_3, it follows that $x^3 - x^2 - 2x = (x - x_1)(x - x_2)(x - x_3)$; equating coefficients of x^2 gives that:

$$x_1 + x_2 + x_3 = -(\text{coefficient of } x^2 \text{ in } (*)) = -(-1) = 1,$$

so that $(-1) + 0 + x_3 = 1$, giving $x_3 = 2$.

Method B (for finding x_3). Factorise $(*)$ to give: $x(x + 1)(x - 2)$, whose roots are: $0, -1, 2$. Two of these are the already known $x_1 = -1, x_2 = 0$, and so x_3 must be the remaining root: $x_3 = 2$.

Having found x_3 (by either method), we use the equation of $\ell_{\mathbf{a},\mathbf{b}}$ to compute $y_3 = x_3 + 1 = 3$. In summary: $\mathcal{E}$ and $\ell_{\mathbf{a},\mathbf{b}}$ intersect at: $(-1,0), (0,1), (2,3)$, and so $(-1,0) + (0,1) + (2,3) = \underline{\mathbf{o}}$.

Finally, this gives: $(-1,0) + (0,1) = -(2,3) = (2,-3)$, using the rule that negation is given by reflection in the x-axis.

One can also obtain an explicit general formula for the group law.

Lemma 2.12. *Let $\mathcal{E} : y^2 = x^3 + Ax + B$, where $A, B \in K$, with (as usual) $\underline{\mathbf{o}}$ = the point at infinity. Let $(x_3, y_3) = (x_1, y_1) + (x_2, y_2)$.*

Case 1. *When $x_1 \neq x_2$ then:*

$$x_3 = \frac{x_1x_2^2 + x_1^2x_2 + A(x_1 + x_2) + 2B - 2y_1y_2}{(x_1 - x_2)^2}, \quad y_3 = -\ell x_3 - m,$$

$$\text{where: } \ell = \frac{y_1 - y_2}{x_1 - x_2}, \quad m = \frac{x_1y_2 - x_2y_1}{x_1 - x_2}.$$

Case 2. *When $(x_1, y_1) = (x_2, y_2)$ then $(x_3, y_3) = (x_1, y_1) + (x_1, y_1)$ (which can be written as $2(x_1, y_1)$), and:*

$$x_3 = \frac{x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2}{4y_1^2} = \frac{x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2}{4(x_1^3 + Ax_1 + B)}, \quad y_3 = -\ell x_3 - m,$$

$$\text{where: } \ell = \frac{3x_1^2 + A}{2y_1}, \quad m = \frac{-x_1^3 + Ax_1 + 2B}{2y_1}.$$

Optional Proof See 0.147.

The above formulas give an alternative method for computing the group law, although in practice it often turns out to be easier to compute the group law from first principles, as in Example 2.11.

Comment 2.13. When $\Delta = 4A^3 + 27B^2 \neq 0$, all 3 roots of $x^3 + Ax + B$ are distinct, guaranteeing that $y^2 = x^3 + Ax + B$ has no singularities and is an elliptic curve (if $\text{char}(K) \neq 2$).

When $\Delta = 0$, then this is no longer an elliptic curve and at least two roots of the cubic are repeated: $y^2 = (x - \alpha)^2(x - \beta)$. It is still the case that the set of nonsingular points on $\mathcal{E}$, denoted $\mathcal{E}_{ns}$, forms a group (see pp. 39–41 of [1]). When $\beta \neq \alpha$ the singularity at $(\alpha, 0)$ is a node. When $\beta = \alpha$ the singularity is a cusp. In either case, the curve can be written: $\left(\frac{y}{x-\alpha}\right)^2 = x - \beta$, and so is birationally equivalent to the conic $w^2 = x - \beta$.

Definition 2.14. Let $\mathcal{E}$ be an elliptic curve and let P be a point on $\mathcal{E}$. For any positive integer m, let mP denote $P + \ldots + P$ (m times). We say that P is an *m-torsion* point if $mP = \underline{\mathbf{o}}$. The m-torsion group of $\mathcal{E}$, denoted $\mathcal{E}[m]$, is the set of all m-torsion points (defined over a fixed algebraic closure $\overline{K}$ of the field of definition K).

We also say that P has *order m* (or that P is a *point of order m*) if m is the smallest positive integer for which $mP = \underline{\mathbf{o}}$. When such m exists, P is a *torsion point* (P *has finite order*). If no such m exists, then P is a non-torsion point (P has *infinite order*). The group of all K-rational torsion points on $\mathcal{E}$ is denoted $\mathcal{E}_{\text{tors}}(K)$ (or sometimes $\mathcal{E}(K)_{\text{tors}}$).

Examples 2.15.

(a) Let $\mathcal{E} : y^2 = x^3 - x$, and let $P = (1,0)$ so that $-P = (1,-0) = (1,0) = P$, so that $2P = P + P = P - P = \underline{\mathbf{o}}$. But $1 \cdot P = P \neq \underline{\mathbf{o}}$, and so 2 is the smallest $m > 0$ such that $mP = \underline{\mathbf{o}}$. P has order 2 and $P \in \mathcal{E}_{\text{tors}}(\mathbb{Q})$.

(b) Let $\mathcal{E} : y^2 = x^3 + 1$, and let $P = (0, 1)$. First compute $P + P$. Using $2yy' = 3x^2$ at $(0, 1)$ gives $2 \cdot 1 \cdot y' = 3 \cdot 0^2$ and so the tangent line $\ell_{P,P}$ to $\mathcal{E}$ at P has slope 0 and equation of form $y = 0 \cdot x + m$. But the line goes through $(0, 1)$ and so $m = 1$ and the tangent line is $y = 1$. Substituting $y = 1$ into $y^2 = x^3 + 1$ gives $x^3 = 0$, with roots $0, 0, 0$. So, $\mathcal{E}$ meets $\ell_{P,P}$ at $(0, 1)$ with multiplicity 3, and $(0, 1) + (0, 1) + (0, 1) = \underline{\mathbf{o}}$. Hence: $(0, 1) + (0, 1) = -(0, 1) = (0, -1)$. In summary:

$1 \cdot (0, 1) = (0, 1), \quad 2 \cdot (0, 1) = (0, -1), \quad 3 \cdot (0, 1) = \underline{\mathbf{o}}$.

So $(0, 1)$ has order 3 and $(0, 1) \in \mathcal{E}_{\text{tors}}(\mathbb{Q})$.

When $K = \mathbb{F}_p$, a finite field with p elements, there are of course only finitely many members of $\mathcal{E}(\mathbb{F}_p)$.

Aside: Each of the p possible x-coordinates $0, \ldots, p - 1$ has about a 50% chance of making $x^3 + Ax + B$ a square modulo p. When $x^3 + Ax + B$ is not a square, there are no corresponding y-coordinates. When $x^3 + Ax + B$ is a square, there are at most two corresponding y-coordinates. So, one might expect 'on average' about p affine points, that is, about $p + 1$ points, including the point at infinity.

The following result gives a bound within which the number of points must lie.

Theorem 2.16. *(Hasse). Let $\mathcal{E}$ be an elliptic curve over $\mathbb{F}_p$. Let $N_p = \#\mathcal{E}(\mathbb{F}_p)$ where, as usual, $\mathcal{E}(\mathbb{F}_p)$ should be taken to including $\underline{\mathbf{o}}$ (so that N_p is the number of affine points (x, y) on $\mathcal{E}$ with $x, y \in \mathbb{F}_p$, plus 1, to include the point at infinity $\underline{\mathbf{o}}$). Then:*

$$|N_p - (p + 1)| \leqslant 2\sqrt{p}, \text{ that is, } N_p \in [(p + 1) - 2\sqrt{p}, (p + 1) + 2\sqrt{p}].$$

Similarly, any curve $y^2 = Q(x)$, where $Q(x) = f_4x^4 + \ldots + f_0$ has nonzero discriminant, has at least $p - 1 - 2\sqrt{p}$ affine points.

Proof. See p. 118 of [1] or p. 131 of [2]. □

Example 2.17. Let $\mathcal{E} : y^2 = x^3 + 4x + 1$, defined over $\mathbb{F}_{13}$. Then:

$$\#\mathcal{E}(\mathbb{F}_{13}) \geqslant 13 + 1 - 2\sqrt{13} > 13 + 1 - 2 \cdot 4 = 6, \text{ so that } \#\mathcal{E}(\mathbb{F}_{13}) \geqslant 7.$$

$$\#\mathcal{E}(\mathbb{F}_{13}) \leqslant 13 + 1 + 2\sqrt{13} < 13 + 1 + 2 \cdot 4 = 22, \text{ so that } \#\mathcal{E}(\mathbb{F}_{13}) \leqslant 21.$$

Note that at most 4 of the points on $\mathcal{E}(\mathbb{F}_{13})$ can be $\underline{\mathbf{o}}$ and points of the form $(x, 0)$, so there must exist at least 3 affine points $(x, y) \in \mathcal{E}(\mathbb{F}_{13})$ with $y \neq 0$.

Section 3. The p-adic Numbers $\mathbb{Q}_p$

For $\mathbb{Q}$, let $|\ |_\infty$ denote the standard absolute value (e.g. $|-5|_\infty = |5|_\infty = 5$). Consider the sequence: $x_1 = 1.4, x_2 = 1.41, x_3 = 1.414, \ldots$, where x_n is the largest decimal to n decimal places satisfying $x_n^2 < 2$. Then $|x_m - x_n|_\infty \to 0$ as $m, n \to \infty$, so that the sequence is Cauchy in $\mathbb{Q}, |\ |_\infty$. The sequence x_n cannot be convergent, since if $x_n \to \alpha$ then clearly $\alpha^2 = 2$ and no such α exists in $\mathbb{Q}$. We say that $(\mathbb{Q}, |\ |_\infty)$ is *incomplete* (since not every Cauchy sequence is convergent) and the real numbers $\mathbb{R}$ give the *completion* of $(\mathbb{Q}, |\ |_\infty)$. The absolute value $|\ |_\infty$ is a special case of the following.

Definition 3.1. Let K be a field. A *valuation* on K is a function $|\ | : K \to \mathbb{R}$ satisfying:
(1) $|x| \geqslant 0$ for all $x \in K$, with equality if and only if $x = 0$.
(2) $|xy| = |x| \cdot |y|$ for all $x, y \in K$.
(3) $|x + y| \leqslant |x| + |y|$ for all $x, y \in K$ (the *triangle inequality*).
If a valuation also satisfies the stronger property:
(3)′ $|x + y| \leqslant \max(|x|, |y|)$, for all $x, y \in K$,
then we say that it is a *non-Archimedean valuation*; otherwise it is an *Archimedean valuation*.

For example, $\mathbb{Q}, |\ |_\infty$ (or $\mathbb{R}, |\ |_\infty$) is a valuation. It is Archimedean since, for example, $|1 + 1|_\infty \not\leqslant \max(|1|_\infty, |1|_\infty)$. We shall now introduce another valuation on $\mathbb{Q}$, which gives a different notion of size and distance.

Definition 3.2. Fix a prime p. Let $x = \frac{m}{n} \in \mathbb{Q}$. Write $\frac{m}{n} = p^r \frac{a}{b}$, where $p \nmid a, p \nmid b$. Then the *p-adic valuation* (or *p-adic absolute value* or *p-adic size*) is defined to be:

$$|x|_p = |\frac{m}{n}|_p = p^{-r}$$

so x is 'smaller' the higher the power of p dividing x.
We also define $|0|_p = 0$. For any $x, y \in \mathbb{Q}$, the *p-adic distance* between x and y is defined to be: $d_p(x, y) = |x - y|_p$. (Note that d_p is a metric)

Example 3.3. In $\mathbb{Q}, |\ |_3$, we have: $|\frac{4}{3}|_3 = |3^{-1}\frac{4}{1}|_3 = (3^{-(-1)}) = 3, |9|_3 = |3^2\frac{1}{1}|_3 = 3^{-2} = \frac{1}{9}$, and $|7|_3 = |3^0\frac{7}{1}|_3 = 3^{-0} = 1$.
Also, $d_3(-5, 3) = |-5 - 3|_3 = |-8|_3 = 1, d_3(-5, 19) = |-5 - 19|_3 = |-24|_3 = 3^{-1}$, and $d_3(\frac{1}{2}, \frac{1}{5}) = |\frac{3}{10}|_3 = 3^{-1}$. For integers m, n, $m \not\equiv n \pmod 3 \iff d_3(m, n) = 1$, $m \equiv n \pmod 3 \iff d_3(m, n) \leqslant \frac{1}{3}$, $m \equiv n \pmod{3^2} \iff d_3(m, n) \leqslant \frac{1}{3^2}$, and so on. The integers m, n are 3-adically closer when they are congruent modulo a higher power of 3.

Lemma 3.4. *The function $|\ |_p$ of Definition 3.2 is a non-Archimedean valuation on $\mathbb{Q}$.*

Proof. (1), (2), (3)′ are trivially true when x or $y = 0$. Let $x, y \in \mathbb{Q}$, $x, y \neq 0$, and write $x = p^r \frac{a}{b}, y = p^s \frac{c}{d}$, where $p \nmid a, b, c, d$.

(1) $|x|_p = p^{-r} > 0$.

(2) $|xy|_p = |p^r \frac{a}{b}\ p^s \frac{c}{d}|_p = |p^{r+s}\ \frac{ac}{bd}|_p = p^{-(r+s)}$ (since $p \nmid ac, bd$) $= p^{-r}p^{-s} = |x|_p|y|_p$.

(3)′ Wlog $r \leqslant s$, giving: $|x + y|_p = |p^r \frac{a}{b} + p^s \frac{c}{d}|_p = |p^r \left(\frac{a}{b} + p^{s-r}\frac{c}{d}\right)|_p = |p^r \frac{ad + p^{s-r}bc}{bd}|_p$
$= |p^r \frac{p^k \ell}{bd}|_p$ for some $k \geqslant 0$ and $\ell \in \mathbb{Z}$ with $p \nmid \ell$ (since $ad + p^{s-r}bc \in \mathbb{Z}$)
$= p^{-(r+k)} \leqslant p^{-r} = |x|_p = \max(|x|_p, |y|_p)$. □

Comment 3.5. By induction, $|a_1 + \ldots + a_n|_p \leqslant \max(|a_1|_p, \ldots, |a_n|_p)$. It is also a good exercise to show that $|x|_p \neq |y|_p \implies |x+y|_p = \max(|x|_p, |y|_p)$. We will use this fact repeatedly. Furthermore, if $|a_k|_p > |a_i|_p$ for all i, $1 \leqslant i \leqslant n, i \neq k$, then $|a_1 + \ldots + a_n|_p = \max(|a_1|_p, \ldots, |a_n|_p) = |a_k|_p$.

Definition 3.6. Let $K, |\ |$ be a field with valuation. For $a_n, \ell \in K$, we say that the sequence a_n *converges* to ℓ (denoted $a_n \to \ell$) in $(K, |\ |)$ when $|a_n - \ell| \to 0$ in $(\mathbb{R}, |\ |_\infty)$ as $n \to \infty$.

That is: for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that, $|a_n - \ell| < \epsilon$ for all $n > N$.

Given a sequence $a_n \in K$, if there exists $\ell \in K$ such that $a_n \to \ell$ in $K, |\ |$ then we say that a_n *converges* in $K, |\ |$, or that it is *convergent* in $K, |\ |$. It is *Cauchy* if $|a_m - a_n| \to 0$ in $\mathbb{R}, |\ |_\infty$ as $m, n \to \infty$. That is: for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that, $|a_m - a_n| < \epsilon$ for all $m, n > N$. We say that $K, |\ |$ is *complete* if every Cauchy sequence is convergent.

All of these definitions coincide with the usual definitions for metric spaces, when we equip K with the metric $d(x, y) = |x - y|$.

Examples 3.7.

(a) Let $a_n = 6^n$. Then $|a_n - 0|_3 = |6^n|_3 = 3^{-n} \to 0$ as $n \to \infty$. So $a_n \to 0$ in $\mathbb{Q}, |\ |_3$.

(b) Let $a_1 = 1,\ a_2 = 11,\ a_3 = 111, \ldots$ so that $9a_n = 999\ldots9$ (n times) and $9a_n + 1 = 10^n$. Then $|9a_n - (-1)|_5 = |10^n|_5 = 5^{-n} \to 0$, giving $9a_n \to -1$ in $\mathbb{Q}, |\ |_5$. It follows that $a_n \to -\frac{1}{9}$ in $\mathbb{Q}, |\ |_5$.

(c) Let $x_0 = a_0 = 3$. Then $a_0^2 = 9 \equiv 2 (\text{mod } 7)$, and $|x_0^2 - 2|_7 = |a_0^2 - 2|_7 = |7|_7 = 7^{-1} < 1$. We want to find $a_1 \in \{0, \ldots, 6\}$ such that $(a_0 + a_1 7)^2 \equiv 2 \pmod{7^2}$.

This is satisfied $\iff a_0^2 + 2a_0a_1 7 + a_1^2 7^2 \equiv 2 \pmod{7^2}$

$\iff 6a_1 7 \equiv 2 - 9 = -7 \pmod{7^2} \iff 6a_1 \equiv -1 \pmod 7 \iff a_1 \equiv 1 \pmod 7$,

so we can take $a_1 = 1$.

Let $x_1 = a_0 + a_1 7 = 3 + 1 \times 7 = 10$. Then $x_1^2 = 100 \equiv 2 \pmod{7^2}$ and $|x_1^2 - 2|_7 = 7^{-2}$.

Aside: note how the solvability of the last congruence is affected by $|2a_0|_7 = |f'(a_0)|_7$, where $f(x) = x^2 - 2$. We will see this more generally in the statement of Hensel's lemma.

When we similarly solve for $a_2 \in \{0, \ldots, 6\}$ such that $(a_0 + a_1 7 + a_2 7^2)^2 \equiv 2 \pmod{7^3}$ we find that $a_2 = 2$, giving $x_2 = a_0 + a_1 7 + a_2 7^2 = 3 + 7 + 98 = 108$. Check: $x_2^2 \equiv 2 \pmod{7^3}$ and $|x_2^2 - 2|_7 \leqslant 7^{-3}$.

We can inductively find $x_n = a_0 + a_1 7 + \ldots + a_n 7^n$ such that $x_n^2 \equiv 2 \pmod{7^{n+1}}$, that is, $|x_n^2 - 2|_7 \leqslant 7^{-(n+1)}$. Hence $x_n^2 \to 2$ in $\mathbb{Q}, |\ |_7$.

Intuitively, $(3 + 1 \cdot 7 + 2 \cdot 7^2 + \ldots)^2 = 2$ in $|\ |_7$. The sequence x_n is easily seen to be Cauchy in $\mathbb{Q}, |\ |_7$. The sequence is not convergent since if $x_n \to \alpha$ in $\mathbb{Q}, |\ |_7$ then $\alpha^2 = 2$, which is impossible for $\alpha \in \mathbb{Q}$.

(d) Again, let $a_0 = 3$, but now define $a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)}$, for $n \geqslant 0$, where $f(x) = x^2 - 2$ (the Newton–Raphson formula). Then:

$a_0 = 3,\ a_1 = 3 - \frac{3^2-2}{2 \cdot 3} = \frac{11}{6},\ a_2 = \frac{11}{6} - \frac{(\frac{11}{6})^2 - 2}{2\frac{11}{6}} = \frac{193}{132}$, and so on.

Check that: $|a_0^2 - 2|_7 = |3^2 - 2|_7 \leqslant 7^{-1}$, $|a_1^2 - 2|_7 = |(\frac{11}{6})^2 - 2|_7 = |\frac{49}{36}|_7 \leqslant 7^{-2}$, and that a_n satisfies the same properties as x_n of Example (c), namely: $|a_n^2 - 2|_7 \leqslant 7^{-(n+1)}$ so that $a_n^2 \to 2$ in $\mathbb{Q}, |\ |_7$, again forcing a_n to be Cauchy but not convergent.

The last two examples show that $\mathbb{Q}$ is incomplete with respect to the valuation $|\ |_7$, and indeed $\mathbb{Q}$ is incomplete with respect to any $|\ |_p$. We now define an extension of $\mathbb{Q}$ which performs the same role with respect to $|\ |_p$ that $\mathbb{R}$ performs with respect to $|\ |_\infty$.

Definition 3.8. The set of *p-adic numbers* $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ with respect to the valuation $|\ |_p$, and is the smallest field containing $\mathbb{Q}$ which is complete with respect to $|\ |_p$.

For any $\alpha, \beta \in \mathbb{Q}_p$, we say that $\alpha \equiv \beta \pmod{p^n} \iff |\alpha - \beta|_p \leqslant p^{-n}$ ('α is congruent to β modulo p^n'). A member of $\mathbb{Q}_p$ (a *p-adic number*) x can be written uniquely in the following form (the *p-adic expansion* of x):

$$x = \sum_{n=N}^{\infty} a_n p^n, \text{ where } N \in \mathbb{Z}, a_N \neq 0 \text{ and each } a_n \in \{0, \ldots, p-1\},$$

in which case $|x|_p = p^{-N}$, and the a_n are the *digits* of x. We might use the shorthand notation $a_N \ldots a_0, a_1 a_2 \ldots$ to represent the above sum. Note that, as for decimal expansions in $(\mathbb{R}, |\ |_\infty)$, $x \in \mathbb{Q}$ exactly when the p-adic digits are eventually periodic.

Examples 3.9.

(a) $w = 4 \cdot 5^{-2} + 1 \cdot 5^{-1} + 4 \cdot 5^0 + 1 \cdot 5^1 + 4 \cdot 5^2 + \ldots \in \mathbb{Q}_5$ and $|w|_5 = 5^2$. This can be denoted $414, \overline{14}$.

(b) $\alpha = 3 \cdot 7^0 + 1 \cdot 7^1 + 2 \cdot 7^2 + \ldots \in \mathbb{Q}_7$ from Example 3.7(c) satisfies $\alpha^2 = 2$.

On the other hand, there is no $\beta \in \mathbb{Q}_7$ such that $\beta^2 = 3$ since any such β would satisfy $|\beta|_7^2 = |\beta^2|_7 = |3|_7 = 1$ and so would have 7-adic expansion $\beta = b_0 + b_1 7 + b_2 7^2 + \ldots$ and would satisfy $(b_0 + b_1 7 + b_2 7^2 + \ldots)^2 = 3$. This would give: $b_0^2 \equiv 3 \pmod 7$, which is impossible, since 3 is not a quadratic residue mod 7 (none of $0^2, 1^2, 2^2, 3^2, 4^2, 5^2, 6^2$ are $\equiv 3 \pmod 7$).

(c) In $\mathbb{Q}_5$: $27 = 2 + 5^2 = 2 \cdot 5^0 + 0 \cdot 5^1 + 1 \cdot 5^2 = 2, 01$ (the 5-adic expansion of 27).

(d) Let us find the 5-adic expansion of $-1/4$. We have $|-1/4|_5 = 1$ so that the 5-adic expansion of $-1/4$ must be of the form $\alpha = a_0 + a_1 5 + a_2 5^2 + \ldots$, each $a_i \in \{0, 1, 2, 3, 4\}$ and $a_0 \neq 0$. This satisfies $-1 = 4(a_0 + a_1 5 + a_2 5^2 + \ldots)$ which gives $-1 \equiv 4a_0 \pmod 5$ and so $a_0 = 1$. Then $-1 = 4(1 + a_1 5 + a_2 5^2 + \ldots)$ gives $-5 \equiv 4a_1 5 \pmod{5^2}$, giving $-1 \equiv 4a_1 \pmod 5$, and so $a_1 = 1$. Similarly, we find that $a_2 = 1, a_3 = 1, \ldots$ and we suspect that $-1/4 = 1, \overline{1}$.

Let $\alpha = 1, \overline{1}$. Then $\alpha - 1 = 0, \overline{1} = 5\alpha$, so that $4\alpha = -1$, giving $\alpha = -1/4$, proving that we have the correct 5-adic expansion.

Comment 3.10. The field $\mathbb{Q}$ is often referred to as a *global field* and its completions with respect to valuations, namely $\mathbb{R}$ and $\mathbb{Q}_p$, for any prime p, are its *local fields* (or *localisations*). An equation defined over $\mathbb{Q}$ which has points in $\mathbb{R}$ and every $\mathbb{Q}_p$, but not in $\mathbb{Q}$, is said to *violate the Hasse Principle.*

Definition 3.11. Let K be a field with a non-Archimedean valuation $|\ |$. We say that $x \in K$ is an *integer* (with respect to the valuation) when $|x| \leqslant 1$, and $R = \{x \in K : |x| \leqslant 1\}$ is the *ring of integers* (or *valuation ring*) of K. The set $\mathfrak{m} = \{x \in K : |x| < 1\}$ is the *maximal ideal*, and $k = R/\mathfrak{m}$ is the *residue field*. The *valuation group* is the set $G_K = \{|x| : x \in K^*\}$ under multiplication.

We say that the valuation is *discrete* if there exists $\delta > 0$ such that $1 - \delta < |x| < 1 + \delta \implies |x| = 1$. When the valuation is discrete, there exists an element $\varpi \in \mathfrak{m}$ such that $\mathfrak{m} = (\varpi)$ is principal with generator ϖ. We say that such an element is a *uniformizer* or *prime element* for the valuation.

The ring of integers for $\mathbb{Q}_p$ is often denoted $\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x|_p \leqslant 1\}$. The valuation group $G_{\mathbb{Q}_p} = \{p^r : r \in \mathbb{Z}\} = \{\ldots, p^{-2}, p^{-1}, p^0, p^1, p^2, \ldots\}$, so that $\mathbb{Q}_p$ is discrete, and we

can take p as a prime element (or indeed any element with valuation p^{-1}). The maximal ideal is $\mathfrak{m} = p\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x|_p \leqslant p^{-1}\}$ and the residue field $\mathbb{Z}_p/p\mathbb{Z}_p$ is isomorphic to $\mathbb{F}_p$, the finite field with p elements.

The following result show how, in some respects, analysis is simpler for non-Archimedean valuations.

Theorem 3.12. *Let K be a field, complete with respect to a non-Archimedean valuation $|\ |$, and let x_n be a sequence in K. Then: $x_n \to 0$ in $K \iff \sum x_n$ is convergent in K.*

Proof. Let $S_N = \sum_{n=1}^{N} x_n$.

$\Rightarrow$: Assume that $x_n \to 0$ in K. Then:

$|S_N - S_M| = |x_{M+1} + \ldots + x_N| \leqslant \max(|x_{M+1}|, \ldots, |x_N|) \to 0$ as $M, N \to \infty$.

S_N is Cauchy and so convergent (since K is complete), giving that $\sum x_n$ is convergent.

$\Leftarrow$: Assume that $\sum x_n$ is convergent, that is, $S_N \to \ell$ for some $\ell \in K$. Then:

$$|x_n - 0| = |x_n| = |S_n - S_{n-1}| = |S_n - \ell + \ell - S_{n-1}| \leqslant |S_n - \ell| + |S_{n-1} - \ell| \to 0$$

as $n \to \infty$, so that $x_n \to 0$ in $K, |\ |$. □

For example, $\sum n!$ converges in any $\mathbb{Q}_p$, since $|n!|_p \to 0$ (it is unknown whether the limit of this sequence in any $\mathbb{Q}_p$ is in $\mathbb{Q}$).

The above result applies to $\mathbb{Q}_p$ (since it is non-Archimedean), but not to $\mathbb{R}$ (where, for example, $x_n = \frac{1}{n}$ is a standard counterexample).

Comment 3.13. It is not too hard to check that the rules for finite sums in Comment 3.5 also apply to infinite series. In other words, when $\sum a_n$ converges, $|\sum a_n| \leqslant \max|a_n|$. Furthermore, if there exists a_k such that $|a_k| > |a_i|$ for all $i \neq k$, then $|\sum a_n| = |a_k|$; in particular, it is then impossible for $\sum a_n = 0$.

Aside: Recall Example 3.7(d), where $x_0 = 3$, and $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$, where $f(x) = x^2 - 2$, defined a sequence, which is Cauchy (but not convergent) in $\mathbb{Q}, |\ |_7$, and which is convergent in $\mathbb{Q}_7$ to a root of $f(x)$. The following describes when an initial approximation a_0 gives a solution to $f(x)$.

Theorem 3.14. *(Hensel's Lemma). Let K be a field, complete with respect to a non-Archimedean valuation $|\ |$, with valuation ring $R = \{x \in K : |x| \leqslant 1\}$.*

Let $f(x) \in R[x]$ and let $a_0 \in R$ satisfy:

$$|f(a_0)| < |f'(a_0)|^2 \qquad (*)$$

Then there exists a unique $a \in R$ such that $f(a) = 0$ and $|a - a_0| < |f'(a_0)|$. This solution moreover satisfies $|a - a_0| \leqslant |f(a_0)|/|f'(a_0)|$.

Proof. Define polynomials $f_j(x)$ by

$$f(x + y) = f_0(x) + f_1(x)y + f_2(x)y^2 + \ldots.$$

so that $f_0(x) = f(x)$ and $f_1(x) = f'(x)$. If $f(x) = x^n$, then $f_j(x) = \binom{n}{j} x^{n-j}$. It follows from this that $f_j(x) \in R[x]$ for arbitrary $f(x)$.

Define $b_0 = -f(a_0)/f'(a_0)$. By $(*)$, $|b_0| < 1$. Define $a_1 = a_0 + b_0 = a_0 - f(a_0)/f'(a_0)$. We are going to show that a_1 is a better approximation to a root of $f(x)$ than a_0. We compute:

$$\begin{aligned} |f'(a_1) - f'(a_0)| = |f'(a_0 + b_0) - f'(a_0)| &= |f_1'(a_0)b_0 + f_2'(a_0)b_0^2 + \ldots| \\ &\leqslant |b_0| < |f'(a_0)| \text{ (by } (*)), \end{aligned}$$

so that $|f'(a_1)| = |f'(a_0)|$. Also,

$$|f(a_1)| = |f(a_0 + b_0)| = |f_0(a_0) + f_1(a_0)b_0 + f_2(a_0)b_0^2 + \dots| = |f_2(a_0)b_0^2 + \dots|$$

since $f_0(a_0) + f_1(a_0)b_0 = 0$.

We deduce that

$$|f(a_1)| \leq \max_{j\geqslant 2}|f_j(a_0)||b_0|^j \leqslant |b_0|^2 = \frac{|f(a_0)|^2}{|f'(a_0)|^2} = \rho|f(a_0)| < |f(a_0)|,$$

where $\rho = \frac{|f(a_0)|}{|f'(a_0)|^2} < 1$.

Summarising: $|f'(a_1)| = |f'(a_0)|$ and $|f(a_1)| \leqslant \rho|f(a_0)| < |f(a_0)|$, where

$$\rho = \frac{|f(a_0)|}{|f'(a_0)|^2} < 1.$$

We proceed by iterating this procedure. So, assume we are given $a_0, \dots, a_n \in R$ such that

$$|f'(a_n)| = \dots = |f'(a_1)| = |f'(a_0)|$$

and

$$|f(a_n)| \leqslant \rho|f(a_{n-1})| \leqslant \dots \leqslant \rho^n|f(a_0)|.$$

Define $b_n = -f(a_n)/f'(a_n)$ and $a_{n+1} = a_n + b_n = a_n - f(a_n)/f'(a_n)$.

Then, as in the case $n = 0$, we have $|f'(a_{n+1})| = |f'(a_n)|$ and

$$|f(a_{n+1})| \leqslant |b_n|^2 = \frac{|f(a_n)|^2}{|f'(a_n)|^2} = \frac{|f(a_n)|^2}{|f'(a_0)|^2} \leq \frac{|f(a_0)|}{|f'(a_0)|^2}|f(a_n)| = \rho|f(a_n)| \leqslant \rho^{n+1}|f(a_0)|.$$

In conclusion, we have defined an infinite sequence $(a_n)_{n\geq 0}$ with $|f'(a_n)| = |f'(a_0)|$ and $|f(a_n)| \leq \rho^n|f(a_0)|$ which $\to 0$ as $n \to \infty$.

We also have $|b_n| = |f(a_n)|/|f'(a_n)| = |f(a_n)|/|f'(a_0)| \to 0$, so by Theorem 3.12

$$a_n = a_0 + b_0 + b_1 + \dots + b_n$$

converges to a limit $a \in R$.

By continuity of polynomials, $f(a) = \lim f(a_n) = 0$. Furthermore:

$$|a - a_0| = |\sum b_n| \leqslant \max|b_n| = \max\frac{|f(a_n)|}{|f'(a_n)|} = \max\frac{|f(a_n)|}{|f'(a_0)|} = \frac{|f(a_0)|}{|f'(a_0)|},$$

as required.

For uniqueness, imagine $\hat{a} \neq a$ also satisfied $f(\hat{a}) = 0$ and $|\hat{a} - a_0| < |f'(a_0)|$. Let $\hat{b} = \hat{a} - a \neq 0$. Then

$$0 = f(\hat{a}) - f(a) = f(a + \hat{b}) - f(a) = \hat{b}f_1(a) + \hat{b}^2 f_2(a) + \dots$$

But $|\hat{b}| = |\hat{a} - a_0 + a_0 - a| \leqslant \max(|\hat{a} - a_0|, |a - a_0|) < |f'(a_0)| = |f_1(a_0)| = |f_1(a)|$ (by continuity of $|f'(x)|$).

This gives $|\hat{b}^j f_j(a)| \leqslant |\hat{b}^j| \leqslant |\hat{b}^2| < |\hat{b}f_1(a)|$ (since $|\hat{b}| \neq 0$ & $|\hat{b}| < |f_1(a)|$) for $j \geqslant 2$, so that the leading term of the sum in (3) has valuation strictly greater than the valuations of the other terms, which is inconsistent with the sum being 0. Hence a is unique. □

Example 3.15. Let $f(x) = x^3 - 7$ and $a_0 = 3$. Then $|f(a_0)|_5 = |3^3 - 7|_5 = 5^{-1}$ and $|f'(a_0)|_5 = |3\cdot 3^2|_5 = 1$. So $|f(a_0)|_5 < |f'(a_0)|_5^2$ and by Hensel's Lemma there exists $a \in \mathbb{Z}_5$ such that $f(a) = 0$, that is: $a^3 = 7$.

Corollary 3.16. *Let $\alpha \in \mathbb{Q}_p$ with $|\alpha|_p = 1$. When $p \neq 2$, α is a square in $\mathbb{Q}_p$ iff it is a square modulo p. When $p = 2$, α is a square in $\mathbb{Q}_p$ iff $\alpha \equiv 1$ (mod 8).*

Example 3.17. $23 \in (\mathbb{Q}_7^*)^2$ since $|23|_7 = 1$ and $23 \equiv 2 \equiv 3^2 \pmod 7$. However, $24 \notin (\mathbb{Q}_7^*)^2$ since $|24|_7 = 1$ and $24 \equiv 3 \pmod 7$, which is not a quadratic residue mod 7.

The corollary does not apply to decide the status of 14, but in fact we can see that $14 \notin (\mathbb{Q}_7^*)^2$, since if $14 = \gamma^2$ for some $\gamma \in \mathbb{Q}_7$ then $|\gamma|_7^2 = |\gamma^2|_7 = |14|_7 = 7^{-1}$, contradicting the fact that $|\gamma|_7 = 7^r$ for some $r \in \mathbb{Z}$.

Section 4. The Reduction Map on an Elliptic Curve

Throughout this section, K denotes a complete non-Archimedean field, with valuation ring $R = \{x : |x| \leqslant 1\}$, maximal ideal $\mathfrak{m} = \{x : |x| < 1\}$ and residue field $k = R/\mathfrak{m}$.

Definition 4.1. Then natural mod $\mathfrak{m}$ map $R \to k = R/\mathfrak{m} : r \mapsto r + \mathfrak{m}$, is a surjection and is denoted $a \mapsto \tilde{a}$ (or sometimes $\bar{a}$). For example in $\mathbb{Z}_5$, if $a = 3 + 2 \cdot 5^1 + \ldots$ then $\tilde{a} = 3$; also $\widetilde{17/3} = 2/3 = 2 \cdot 2 = 4$.

Let $a = (a_0, \ldots, a_n) \in \mathbb{P}^n(K)$. We define the reduction map to $\mathbb{P}^n(k)$ as follows.

Step 1. There exists i_0 such that $|a_{i_0}| \geqslant |a_i|$ for $i = 0, \ldots, n$. We replace each a_i by a_i/a_{i_0} (which leaves a unchanged) so that now the largest valuation is 1 (*normalised* form).

Step 2. Define $\tilde{a} = (\tilde{a}_0, \ldots, \tilde{a}_n)$ (easy to check that this is well defined).

In affine space, if $a = (a_1, \ldots, a_n)$ then $\tilde{a} = (\tilde{a}_1, \ldots, \tilde{a}_n)$, provided that all $|a_i| \leqslant 1$.

When $K = \mathbb{Q}_p$, this is just the 'mod p' map, where the coordinates are reduced modulo p.

Example 4.2. In $\mathbb{P}^2(\mathbb{Q}_5)$, let $a = (1/5, 2/15, 2)$. Dividing through by $a_0 = 1/5$ gives $a = (1, 2/3, 10)$ so that $\tilde{a} = (\tilde{1}, \widetilde{2/3}, \widetilde{10}) = (1, 4, 0) \in \mathbb{P}^2(\mathbb{F}_5)$. For $b = (2/3, 25)$ in affine space $A^2(\mathbb{Q}_5)$ (an affine point with no denominators of 5), then $\tilde{b} = (4, 0) \in A^2(\mathbb{F}_5)$.

For the point $P = (1/4, 7/8) \in \mathcal{E}(\mathbb{Q}) \subset \mathcal{E}(\mathbb{Q}_2)$ on the elliptic curve $\mathcal{E} : y^2 = x^3 - x + 1$, we should first write P in projective form: $(1/4, 7/8, 1) = (2/7, 1, 8/7)$ (after dividing through by $7/8$), which reduces modulo 2 to $(0, 1, 0)$, the point at infinity on $\widetilde{\mathcal{E}}(\mathbb{F}_2)$. Clearly any $(x, y) \in \mathcal{E}(\mathbb{Q}_p)$ will reduce mod p to the point at infinity iff $|x|_p > 1$ and $|y|_p > 1$.

Definition 4.3. Let $\mathcal{C} : F(X, Y, Z) = 0$ be a projective curve, defined over K. Let $\{f_i\}$ be the set of all coefficients of $\mathcal{C}$. The curve is unchanged if we multiply all the f_i by a nonzero constant, so after dividing through by f_{i_0} such that $|f_{i_0}| \geqslant |f_i|$ for all i, we can assume that $\max(|f_i|) = 1$.

The reduction of $\mathcal{C}$ mod $\mathfrak{m}$ is then $\widetilde{\mathcal{C}} : \widetilde{F}(X, Y, Z) = 0$, defined over $k = R/\mathfrak{m}$, where every coefficient has been reduced mod $\mathfrak{m}$. When $K = \mathbb{Q}_p$, this is again just a matter of reducing the coefficients mod p.

Clearly, a lies on $\mathcal{C} \implies \tilde{a}$ lies on $\widetilde{\mathcal{C}}$, when we say that a reduces to $\tilde{a}$.

Definition 4.4. Let $b \in \widetilde{\mathcal{C}}(k)$. If there exists $a \in \mathcal{C}(K)$ such that $\tilde{a} = b$, we say that b *lifts* to $\mathcal{C}$ (or that b *lifts* to a point on $\mathcal{C}$).

Example 4.5. Let $\mathcal{E} : ZY^2 = X^3 + pZ^3$, defined over $\mathbb{Q}_p$, and $\widetilde{\mathcal{E}} : ZY^2 = X^3$, defined over $\mathbb{F}_p$. Consider $(0, 0, 1) \in \widetilde{\mathcal{E}}(\mathbb{F}_p)$. Does it lift to a point in $\mathcal{E}(\mathbb{Q}_p)$? Imagine $(X, Y, Z) \in \mathcal{E}(\mathbb{Q}_p)$ reduces mod p to $(0, 0, 1) \in \widetilde{\mathcal{E}}(\mathbb{F}_p)$. Then $p|X, p|Y, p \nmid Z$, that is, $|X|_p < 1, |Y|_p < 1, |Z|_p = 1$. But all p-adic values are of the form: $\ldots, p^{-2}, p^{-1}, p^0, p^1, \ldots$ so that $|X|_p \leqslant p^{-1}, |Y|_p \leqslant p^{-1}$, and $|X^3|_p \leqslant p^{-3}$. Furthermore, $|pZ^3|_p = |p|_p |Z|_p^3 = p^{-1}$.

Since $|X^3|_p \neq |pZ^3|_p$ we must have $|X^3 + pZ^3|_p = \max(|X^3|_p, |pZ^3|_p) = p^{-1}$. But then $|Y^2|_p = |ZY^2|_p = |X^3 + pZ^3|_p = p^{-1}$, a contradiction. We conclude that $(0, 0, 1) \in \widetilde{\mathcal{E}}(\mathbb{F}_p)$ does not lift to a point in $\mathcal{E}(\mathbb{Q}_p)$.

If we had represented the above curves with the affine shorthand: $\mathcal{E} : y^2 = x^3 + p$ and $\widetilde{\mathcal{E}} : y^2 = x^3$, then the above would be expressed by saying that $(0, 0) \in \widetilde{\mathcal{E}}(\mathbb{F}_p)$ does not lift.

On the other hand, the following result shows that we can guarantee lifting a nonsingular point on $\widetilde{\mathcal{E}}$.

Theorem 4.6. *Let $\mathcal{C}$ be defined over K, written so that the coefficients lie in R. Let $\widetilde{\mathcal{C}}$, defined over k, be the reduction of $\mathcal{C}$ modulo $\mathfrak{m}$. Let $b \in \widetilde{\mathcal{C}}(k)$ be a nonsingular point. Then b lifts to $\mathcal{C}$; that is, there exists $a \in \mathcal{C}(K)$ such that $\tilde{a} = b$.*

Proof. Write $\mathcal{C} : F(X_0, X_1, X_2) = 0$ (normalised), so that $\widetilde{\mathcal{C}} : \widetilde{F}(X_0, X_1, X_2) = 0$. Let $b = (b_0, b_1, b_2) \in \widetilde{\mathcal{C}}(k)$ be a nonsingular point. Then at least one of the $\frac{\partial \widetilde{F}}{\partial X_i}(b) \neq 0$; wlog say that $\frac{\partial \widetilde{F}}{\partial X_0}(b) \neq 0$. Let $\alpha_0, \alpha_1, \alpha_2 \in R$ be such that each $\tilde{\alpha}_i = b_i$ under the natural surjection from R to $k = R/\mathfrak{m}$. Then $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ satisfies $\tilde{\alpha} = b$; however, we have no guarantee that α lies on $\mathcal{C}$. We shall construct an adjustment of α which lies on $\mathcal{C}$, and which has the same reduction as α. Let $f(t) = F(t, \alpha_1, \alpha_2)$. Then $\widetilde{f(\alpha_0)} = \widetilde{F}(b) = 0$ so that $|f(\alpha_0)| < 1$. Furthermore, $\widetilde{f'(\alpha_0)} = \frac{\partial \widetilde{F}}{\partial X_0}(\tilde{\alpha}) = \frac{\partial \widetilde{F}}{\partial X_0}(b) \neq 0$, so that $|f'(\alpha_0)| = 1$. By Hensel's Lemma, there exists $a_0 \in R$ such that $f(a_0) = 0$ and $|a_0 - \alpha_0| < 1$, so that $a = (a_0, \alpha_1, \alpha_2)$ is a point on $\mathcal{C}$ and $\tilde{a} = \tilde{\alpha} = b$, as required. □

We wish to see under what circumstances the reduction map is a homomorphism on an elliptic curve.

Theorem 4.7. *Let $\mathcal{C} : F(X_0, X_1, X_2) = 0$ be a cubic curve defined over K, written so that coefficients of F have maximum valuation 1. Suppose the line $\mathcal{L} : L(X_0, X_1, X_2) = 0$ meets $\mathcal{C}$ at a, b, c. Then either:*
(1) $\widetilde{\mathcal{L}} \subset \widetilde{\mathcal{C}}$, that is, $\widetilde{F}(X_0, X_1, X_2) = \widetilde{L}\widetilde{M}$, for some M.
or:
(2) $\widetilde{\mathcal{L}}$ meets $\widetilde{\mathcal{C}}$ precisely at $\tilde{a}, \tilde{b}, \tilde{c}$.

Proof. Let $L : \ell_0 X_0 + \ell_1 X_1 + \ell_2 X_2$, written so that $\max(|\ell_0|, |\ell_1|, |\ell_2|) = 1$, wlog $|\ell_0| = 1$; after dividing through by ℓ_0 (and relabelling $\ell_1/\ell_0, \ell_2/\ell_0$ as ℓ_1, ℓ_2), we can take $\mathcal{L} : X_0 = -\ell_1 X_1 - \ell_2 X_2$, where $\ell_1, \ell_2 \in R$. Write $a = (a_0, a_1, a_2), b = (b_0, b_1, b_2), c = (c_0, c_1, c_2)$ with $\max|a_i| = \max|b_i| = \max|c_i| = 1$. Note that, since a, b, c lie on $\mathcal{L}$, we must then have $\max(|a_1|, |a_2|) = \max(|b_1|, |b_2|) = \max(|c_1|, |c_2|) = 1$.

Now, substitute L into F to get: $G(X_1, X_2) = F(-\ell_1 X_1 - \ell_2 X_2, X_1, X_2) \in R[X_1, X_2]$. Since the points a, b, c lie on both $\mathcal{L}$ and $\mathcal{C}$, the roots of the projective polynomial G are $(a_1, a_2), (b_1, b_2), (c_1, c_2) \in \mathbb{P}^1(K)$, so that:

$G(X_1, X_2) = F(-\ell_1 X_1 - \ell_2 X_2, X_1, X_2) = \lambda(a_2 X_1 - a_1 X_2)(b_2 X_1 - b_1 X_2)(c_2 X_1 - c_1 X_2)$,

for some $\lambda \in R^*$. Now consider $\widetilde{F}(-\tilde{\ell}_1 X_1 - \tilde{\ell}_2 X_2, X_1, X_2)$. If this is 0 then $\widetilde{L}$ is a factor of $\widetilde{F}$, giving case (1). Otherwise, this is a nonzero projective polynomial, defined over k, equal to $\tilde{\lambda}(\tilde{a}_2 X_1 - \tilde{a}_1 X_2)(\tilde{b}_2 X_1 - \tilde{b}_1 X_2)(\tilde{c}_2 X_1 - \tilde{c}_1 X_2)$, with $(\tilde{a}_1, \tilde{a}_2), (\tilde{b}_1, \tilde{b}_2), (\tilde{c}_1, \tilde{c}_2) \in \mathbb{P}^1(k)$ as roots, so that $\tilde{a}, \tilde{b}, \tilde{c}$ lie on $\widetilde{\mathcal{L}}$ and $\widetilde{\mathcal{C}}$. Since $\widetilde{L}$ and $\widetilde{F}$ have no common factor, these must be precisely the points of intersection of $\widetilde{\mathcal{L}}$ and $\widetilde{\mathcal{C}}$. □

When we have an elliptic curve written, not as a general cubic, but birationally transformed to the form $\mathcal{E} : y^2 = x^3 + Ax + B$ with $A, B \in R$ (which, as usual, is shorthand for the projective curve $ZY^2 = X^3 + AXZ^2 + BZ^3$), the reduction $\widetilde{\mathcal{E}}$ will still be of the form $y^2 = x^3 + \ldots$. This cannot contain a line, since any $(y + rx + \ldots)(y - x^2/r + \ldots)$ would have an $x^2 y$ term and so would not give y^2 − cubic in x. For such a curve, only option (2) can apply in the previous theorem. Even though $\mathcal{E}$ is an elliptic curve (and

therefore nonsingular), the reduction $\widetilde{\mathcal{E}}$ might be singular (for example, when $p|\Delta \in \mathbb{Z}$ so that $\widetilde{\Delta} = 0$ in $\mathbb{F}_p$), but even in that case we still have the group $\widetilde{\mathcal{E}}_{ns}(k)$ of nonsingular points (see Comment 2.13). Since the group law is constructed by finding intersections between the curve and lines, and since only option (2) applies, the construction of the group law respects the reduction map, giving the following result.

Corollary 4.8. *Let $\mathcal{E} : y^2 = x^3 + Ax + B$ be an elliptic curve, with $A, B \in R$, with reduction $\widetilde{\mathcal{E}}$. Let $\widetilde{\mathcal{E}}_{ns}(k)$ denote the group of nonsingular points in $\widetilde{\mathcal{E}}(k)$, and let $\mathcal{E}_0(K)$ denote the set of points in $\mathcal{E}(K)$ which reduce to members of $\widetilde{\mathcal{E}}_{ns}(k)$, that is, define: $\mathcal{E}_0(K) = \{P \in \mathcal{E}(K) : \widetilde{P} \in \widetilde{\mathcal{E}}_{ns}(k)\}$. Then the reduction map $P \mapsto \widetilde{P}$ is a homomorphism from $\mathcal{E}_0(K)$ to $\widetilde{\mathcal{E}}_{ns}(k)$.*

Definition 4.9. Let $\mathcal{E}_0(K)$ and $\widetilde{\mathcal{E}}_{ns}(k)$ be as in Corollary 4.8. The *kernel of reduction*, denoted $\mathcal{E}_1(K)$, is the kernel of the reduction map from $\mathcal{E}_0(K)$ to $\widetilde{\mathcal{E}}_{ns}(k)$. That is:

$$\mathcal{E}_1(K) = \{P \in \mathcal{E}(K) : \widetilde{P} = \underline{\mathbf{o}}\},$$

where, as usual, $\underline{\mathbf{o}}$ is the identity element, usually taken to be the point at infinity, in which case

$$\mathcal{E}_1(K) = \{P = (x, y) \in \mathcal{E}(K) : |x| > 1,\ |y| > 1\},$$

since these are the points that map to the point at infinity under the reduction map.

We can summarise what we know so far by the following exact sequence:

$$0 \longrightarrow \mathcal{E}_1(K) \xrightarrow{\ i\ } \mathcal{E}_0(K) \xrightarrow{\ \sim\ } \widetilde{\mathcal{E}}_{ns}(k) \longrightarrow 0,$$

where i is the inclusion map.

We now wish to look more closely at how we can describe the group law inside $\mathcal{E}_1(K)$, the kernel of reduction, for an elliptic curve:

$$\mathcal{E} : y^2 = x^3 + Ax + B, \quad \text{where } A, B \in R.$$

We adopt the usual convention that the identity is $\underline{\mathbf{o}}$, the point at infinity so that, as already observed, $\mathcal{E}_1(K) = \{(x, y) \in \mathcal{E}(K) : |x| > 1,\ |y| > 1\}$. The members of $\mathcal{E}_1(K)$ are in a neighbourhood of $\underline{\mathbf{o}}$, and it is natural to try to describe the group law as a power series. This will be more transparent if we write our equation in a form where the points in the neighbourhood have coordinates with small, rather than large, valuation. We therefore perform the following birational transformation:

$$z = -x/y,\ w = -1/y, \quad \text{with inverse } x = z/w,\ y = -1/w.$$

This transforms $\mathcal{E}$ to:

$$\frac{1}{w^2} = \frac{z^3}{w^3} + A\frac{z}{w} + B,$$

giving the equation

$$\mathcal{E}' : w = f(z, w) = z^3 + Aw^2z + Bw^3.$$

Note that the point at infinity $\underline{\mathbf{o}}$ on $\mathcal{E}$ maps to the point $(0, 0)$ on $\mathcal{E}'$, which we take as our group identity on $\mathcal{E}'$. The condition $|x| > 1, |y| > 1$ corresponds to $|z| < 1, |w| < 1$, so that the kernel of reduction for $\mathcal{E}'$ is:

$$\mathcal{E}'_1(K) = \{(z, w) \in \mathcal{E}'(K) : |z| < 1,\ |w| < 1\}.$$

We now recursively substitute $w = f(z, w)$ into itself. For the first step:

$$w = f(z, w) = f(z, f(z, w)) = z^3 + A(z^3 + Aw^2z + Bw^3)^2z + B(z^3 + Aw^2z + Bw^3)^3$$

$$= z^3 + Az^7 + \dots$$

Inductively define $f_n(z,w)$ by: $f_1(z,w) = f(z,w)$ and $f_{n+1}(z,w) = f_n(z, f(z,w))$. Define

$$w(z) = \lim_{n\to\infty} f_n(z,0) \in \mathbb{Z}[A,B][\![z]\!].$$

The following is then easy to show.

Lemma 4.10. *The power series $w(z) = z^3(1+\dots) \in \mathbb{Z}[A,B][\![z]\!]$ defined above is the unique power series satisfying $w(z) = f(z, w(z))$.*

This means that $(z, w(z))$ satisfies $\mathcal{E}'$. Since we are working in a non-Archimedean field K, we can appeal to the fact (see Theorem 3.12) that a series converges iff its terms converge to 0. When we are in the kernel of reduction $|z| < 1, |w| < 1$, this applies to the above series $w(z)$ (since $A, B \in R$ and so $|A|, |B| \leqslant 1$). Any (z,w) in the kernel of reduction must satisfy $w = w(z)$ (by the uniqueness part of Hensel's lemma), and so is uniquely determined by z, which is called a *local parameter*.

Comment 4.11. We can recover x, y on $\mathcal{E}$ as formal Laurent series:

$$x(z) = \frac{z}{w(z)} = \frac{z}{z^3(1+\dots)} = \frac{1}{z^2} + \dots$$

$$y(z) = -\frac{1}{w(z)} = -\frac{1}{z^3(1+\dots)} = -\frac{1}{z^3} + \dots$$

which gives a formal solution to $\mathcal{E}$.

Let us now perform the addition $(z_1,w_1)+(z_2,w_2)$. As usual, we first write the line $w = \lambda z + \mu$ through the points, given by $\lambda = (w_1 - w_2)/(z_1 - z_2)$ and $\mu = (z_1 w_2 - z_2 w_1)/(z_1 - z_2)$. As long as we are in the kernel of reduction, $w_1 = w(z_1)$ and $w_2 = w(z_2)$, and so:

$$\lambda = \lambda(z_1,z_2) = \frac{w(z_1) - w(z_2)}{z_1 - z_2} = \frac{z_1^3(1+\dots) - z_2^3(1+\dots)}{z_1 - z_2} \in \mathbb{Z}[A,B][\![z_1,z_2]\!],$$

with all terms being of degree $\geqslant 2$, and:

$$\mu = \mu(z_1,z_2) = \frac{z_1 w(z_2) - z_2 w(z_1)}{z_1 - z_2} \in \mathbb{Z}[A,B][\![z_1,z_2]\!].$$

Substituting $w = \lambda z + \mu$ into $\mathcal{E}'$ gives $\lambda z + \mu = z^3 + A(\lambda z + \mu)^2 z + B(\lambda z + \mu)^3$, and so:

$$(1 + A\lambda^2 + B\lambda^3)z^3 + (2A\lambda\mu + 3B\lambda^2\mu)z^2 + \dots = 0.$$

Let $(z_3, w(z_3))$ be the third point of intersection of $\mathcal{E}'$ and the line $w = \lambda z + \mu$, so that z_1, z_2, z_3 are the roots of the above cubic, giving that $z_1 + z_2 + z_3 = -(\text{coeff of } z^2)/(\text{coeff of } z^3)$, so:

$$z_3 = -z_1 - z_2 - \frac{2A\lambda\mu + 3B\lambda^2\mu}{1 + A\lambda^2 + B\lambda^3} \in \mathbb{Z}[A,B][\![z_1,z_2]\!],$$

since the denominator is of the form $1 + \phi(z_1,z_2)$, where $\phi(z_1,z_2)$ has no constant term (and so is an invertible power series, with $1/(1+\phi(z_1,z_2)) = 1 - \phi(z_1,z_2) + \phi(z_1,z_2)^2 + \dots$).

The sum $(z_1,w_1) + (z_2,w_2) + (z_3,w_3) =$ the identity, and so $(z_1,w_1) + (z_2,w_2) = -(z_3,w_3)$. Negation $(x,y) \mapsto (x,-y)$ induces $(z,w) \mapsto (-z,-w)$ (since $z = -x/y, w = -1/y$), so that the z-coordinate of $(z_1,w_1) + (z_2,w_2)$ is given by $F_{\mathcal{E}}(z_1,z_2)$, where:

$$F_{\mathcal{E}}(z_1,z_2) = z_1 + z_2 + (\text{terms of degree } \geqslant 2) \in \mathbb{Z}[A,B][\![z_1,z_2]\!].$$

We summarise this as follows.

Lemma 4.12. *Any point (x, y) on $\mathcal{E}$ ($\leftrightarrow (z, w)$ on $\mathcal{E}'$) in the kernel of reduction (explicitly: $|x| > 1, |y| > 1 \leftrightarrow |z| < 1, |w| < 1$) is uniquely determined by z, with $w = w(z) \in \mathbb{Z}[A, B][\![z]\!]$. The group law is completely described by the above $F_{\mathcal{E}}(z_1, z_2) \in \mathbb{Z}[A, B][\![z_1, z_2]\!]$, which converges to the z-coordinate of the sum of $(z_1, w(z_1))$ and $(z_2, w(z_2))$.*

We have already observed that $F_{\mathcal{E}}(z_1, z_2) = z_1 + z_2 +$ terms of higher degree. The associativity and commutativity properties of the group law on $\mathcal{E}$ also induce the properties:

$$F_{\mathcal{E}}(X, F_{\mathcal{E}}(Y, Z)) = F_{\mathcal{E}}(F_{\mathcal{E}}(X, Y), Z), \quad F_{\mathcal{E}}(X, Y) = F_{\mathcal{E}}(Y, X).$$

Of course, the power series $F_{\mathcal{E}}(z_1, z_2) \in \mathbb{Z}[A, B][\![z_1, z_2]\!]$ can be derived for any $\mathcal{E}$ defined over any ring, regardless of convergence considerations. In the next section, we shall consider power series $F(X, Y)$ which satisfy the above properties, and then apply the results to the special case of $F_{\mathcal{E}}(X, Y)$.

Section 5. Formal Groups

Let R be any ring (by *ring* I shall alway mean a commutative ring with 1).

Definition 5.1. A (one-parameter, commutative) *formal group* defined over R is a power series $F(X,Y) \in R[\![X,Y]\!]$ satisfying:
(1) $F(X,Y) = X + Y +$ terms of degree $\geqslant 2$.
(2) $F(X,F(Y,Z)) = F(F(X,Y),Z)$.
(3) $F(X,Y) = F(Y,X)$.

Example 5.2. The following are all formal groups.
The formal group $F_{\mathcal{E}}(X,Y)$ of an elliptic curve defined over R, as described in Section 4.
The formal additive group $F(X,Y) = \widehat{\mathbb{G}}_a(X,Y) = X + Y$.
The formal multiplicative group $F(X,Y) = \widehat{\mathbb{G}}_m(X,Y) = X + Y + XY$.
Note: the last of these is just XY, but translated one unit to the left: $(1+X)(1+Y)-1$ so that the identity is changed from 1 to 0.

Aside: A formal group does not necessarily induce an actual nontrivial commutative group, since there is no guarantee that the power series will converge for any nonzero X, Y; indeed, our arbitrary ring R may not even come together with any structure (such as a valuation or metric) that provides a definition of convergence. It is merely a power series satisfying properties analogous to associativity and commutativity. The definition appears to be missing properties analogous to the existence of an identity element and inverses. In fact, the following result shows these can be deduced from the given axioms.

Lemma 5.3. *Let $F(X,Y)$ be a formal group over a ring R.*
(1) There is a unique power series $i(T) \in TR[\![T]\!]$ such that $F(T,i(T)) = 0$.
(2) $F(X,0) = X$ and $F(0,Y) = Y$.

Proof. (1) Let $Z_1 = -T \in TR[\![T]\!]$; then the terms of $F(T,Z_1)$ all have degree $\geqslant 2$. Suppose we have $Z_n \in TR[\![T]\!]$ of degree $\leq n$ such that $F(T,Z_n) = a_{n+1}T^{n+1} + \ldots$ has terms all of degree $\geqslant n+1$. Define $Z_{n+1} = Z_n - a_{n+1}T^{n+1}$; then:

$$\begin{aligned} F(T,Z_{n+1}) &= F(T, Z_n - a_{n+1}T^{n+1}) = T + (Z_n - a_{n+1}T^{n+1}) + \ldots \\ &= F(T,Z_n) - a_{n+1}T^{n+1} + \text{ (terms of degree } \geqslant n+2) \\ &= a_{n+1}T^{n+1} - a_{n+1}T^{n+1} + \text{ (terms of degree } \geqslant n+2), \end{aligned}$$

which has terms all of degree $\geqslant n+2$. Moreover Z_{n+1} is the unique polynomial of degree $\leq n+1$ with this property.

This defines a sequence $(Z_n)_{n\geq 1}$ with $Z_{n+1} = Z_n \bmod T^{n+1}$. Letting n tend to infinity, we can define a power series $i(T)$ whose first n terms give Z_n for each n. It satisfies $F(T,i(T)) = 0$ (since $F(T,i(T)) = F(T,Z_n) = 0 \bmod T^{n+1}$ for every n). Furthermore, if $i(T)$ satisfies $F(T,i(T)) = 0$ and we look at the degree n part of $i(T)$ (discarding terms of degree $\geq n+1$), we must get the uniquely determined polynomial Z_n. We deduce that $i(T)$ is unique.

(2) By a similar argument to (1), there exists a unique $j(T) \in TR[\![T]\!]$ such that $F(j(T),i(T)) = 0$. By (1) we can take $j(T) = T$. By associativity $F(F(0,T),i(T)) = F(0,F(T,i(T))) = F(0,0) = 0$, so that we can also take $j(T) = F(0,T)$. Since $j(T)$ is unique, it follows that $F(0,T) = T$. Similarly for $F(T,0) = T$. □

Definition 5.4. Let F, G define formal groups over R. A power series $f(T) \in TR[\![T]\!]$ is a *homomorphism* from F to G if it satisfies $f(F(X,Y)) = G(f(X),f(Y))$. When there

also exists an inverse $g(T) \in TR[\![T]\!]$ (that is: $f(g(T)) = g(f(T)) = T$), then $f(T)$ is an *isomorphism*.

Example 5.5. If $\mathrm{char}(R) = 0$ and $\frac{1}{n} \in R$ for all n, then $f(T) = T - T^2/2 + T^3/3 - \dots$ (i.e. the power series expansion of $\log(1+T)$) is a homomorphism from $\widehat{\mathbb{G}}_m$ to $\widehat{\mathbb{G}}_a$.

Definition 5.6. Let F define a formal group over R. Define the *multiplication by m map* $[m](T) \in R[\![T]\!]$, for $m \in \mathbb{Z}$, inductively by: $[0](T) = 0$, $[m+1](T) = F([m](T), T)$ and $[m-1](T) = F([m](T), i(T))$. This is clearly a homomorphism from F to F, and is of the form: $[m](T) = mT +$ terms of degree $\geqslant 2$.

Lemma 5.7. *Let $a \in R^*$ (that is: $a \in R$ and $a^{-1} \in R$), and let $f(T) \in TR[\![T]\!]$ be of the form $f(T) = aT + \dots$ Then there exists a unique $g(T) \in TR[\![T]\!]$ such that $f(g(T)) = T$. Furthermore, g satisfies $g(f(T)) = T$.*

Proof. We shall construct $g(T) = b_1T + b_2T^2 + \dots$, the limit of $g_1(T) = b_1T$, $g_2(T) = b_1T + b_2T^2, \dots$, first defining $g_1(T) = a^{-1}T$, so that the terms of $f(g_1(T)) - T$ all have degree $\geqslant 2$. Suppose we have $g_n(T)$ of degree n such that $f(g_n(T)) - T = bT^{n+1} + \dots$ and define $g_{n+1}(T) = g_n(T) - a^{-1}bT^{n+1}$. Then

$$f(g_{n+1}(T)) - T = f(g_n(T)) - aa^{-1}bT^{n+1} + (\text{terms of degree} \geqslant n+2) - T,$$

whose terms are all of degree $\geqslant n+2$. The resulting $g(T)$ then satisfies $f(g(T)) = T$ and is unique, since each choice of coefficient was forced.

There similarly exists $h(T) \in R[\![T]\!]$ such that $g(h(T)) = T$, and so $f(g(h(T))) = f(T)$, giving $h(T) = f(T)$. Substituting this into $g(h(T)) = T$ gives $g(f(T)) = T$, as required. □

Aside: The arguments in Lemma 5.3 and Lemma 5.7 can be rewritten as an application of an appropriate version of Hensel's Lemma. We can equip the ring $R[\![T]\!]$ with valuation $|f(T)| = \rho^n$, where ρ is a fixed real number satisfying $0 < \rho < 1$ and n is the degree of the smallest nonzero degree term (for example, $|2T^3 + 5T^4 + \dots| = \rho^3$). Here T takes on a similar role for $R[\![T]\!]$ to that performed by p for $\mathbb{Z}_p$. See Examples 10.10 and 10.11 in https://kconrad.math.uconn.edu/blurbs/gradnumthy/hensel.pdf.

Lemma 5.8. *The homomorphism $[m] : F \to F$ of Definition 5.6 is an isomorphism whenever $m \in R^*$.*

Proof. Since $[m](T) = mT +$ terms of degree $\geqslant 2$, we have from the previous lemma that the homomorphism $[m]$ has an inverse, and so is an isomorphism. □

Aside: You might have wondered in school about the connection between the two properties of log, that it is the integral of $1/x$, and that $\log(ab) = \log(a) + \log(b)$ (a homomorphism from multiplication to addition). One way of seeing the connection is to define $\log(T) = \int v(T)$ (with $\log(1) = 0$), where $v(T) = \frac{1}{T}dT$, and note that (regarding T as a variable and S as a constant) $v(TS) = \frac{1}{TS}d(TS) = v(T)$, that is, v remains invariant under replacing T by TS. Therefore $\log(TS) = \log(T) + f(S)$, where $f(S)$ is a constant; setting $T = 1$ gives $f(S) = \log(S)$. If we were to adjust the multiplicative group, translating by -1, so that the identity is 0: $F(X,Y) = (1+X)(1+Y) - 1 = X + Y + XY$, then $\omega(T) = \frac{1}{1+T}dT = (1 - T + T^2 - \dots)dT$ would have the property that $\omega \circ F(T,S) = \omega(T)$ (and $\int \omega(T)$ would give a homomorphism from $\widehat{\mathbb{G}}_m$ to $\widehat{\mathbb{G}}_a$). It is natural to ask whether ω is unique (up to constants), and how we would construct ω for a general choice of $F(X,Y)$.

Definition 5.9. A differential form on $R[\![T]\!]$ is an expression of the form $\sum_{i=1}^{m} P_i(T)\mathrm{d}Q_i(T)$, where each $P_i(T), Q_i(T) \in R[\![T]\!]$, and these satisfy the natural rules:

$$\mathrm{d}\left(P(T)\right) = P'(T)\mathrm{d}T, \text{ where } P'(T) = \textstyle\sum_{n=1}^{\infty} a_n n T^{n-1}, \text{ for any } P(T) = \displaystyle\sum_{n=0}^{\infty} a_n T^n,$$

$$\mathrm{d}\left(P(T)+Q(T)\right) = \mathrm{d}P(T) + \mathrm{d}Q(T), \quad \mathrm{d}\left(P(T)Q(T)\right) = P(T)\mathrm{d}Q(T) + Q(T)\mathrm{d}P(T).$$

We can see from the first rule that each differential form can be written uniquely as $\omega(T) = P(T)\mathrm{d}T$ with $P(T) \in R[\![T]\!]$.

More formally, the space of *differential forms* on $R[\![T]\!]$ is defined to be the quotient of the free $R[\![T]\!]$-module spanned by the symbols $\{\mathrm{d}f : f \in R[\![T]\!]\}$ by the submodule spanned by $\{\mathrm{d}f - f'\mathrm{d}T : f \in R[\![T]\!]\}$. This is a free $R[\![T]\!]$-module with basis element $\mathrm{d}T$.

An *invariant differential* on a formal group F, defined over R, is a differential form:

$$\omega(T) = P(T)\mathrm{d}T \in R[\![T]\!]\mathrm{d}T, \text{ satisfying } \omega \circ F(T,S) = \omega(T).$$

Note that $\omega \circ F(T,S)$ is the same as $P(F(T,S))\mathrm{d}(F(T,S)) = P(F(T,S))F_X(T,S)\mathrm{d}T$, where $F_X(X,Y)$ denotes the partial derivative of $F(X,Y)$ with respect to X. So, the above condition on ω is equivalent to:

$$\omega(T) = P(T)\mathrm{d}T \in R[\![T]\!]\mathrm{d}T, \text{ satisfying } P\left(F(T,S)\right)F_X(T,S) = P(T).$$

An invariant differential $\omega(T) = P(T)\mathrm{d}T$ is said to be *normalised* if $P(0) = 1$.

Example 5.10. On $\widehat{\mathbb{G}}_a$, the formal group defined by $F(X,Y) = X + Y$, we can take $\omega(T) = \mathrm{d}T$ as a normalised invariant differential. On $\widehat{\mathbb{G}}_m$, the multiplicative formal group defined by $F(X,Y) = X+Y+XY$, we can take $\omega(T) = (1+T)^{-1}\mathrm{d}T = (1-T+T^2-\ldots)\mathrm{d}T$.

Theorem 5.11. *Let F be a formal group over R. There exists a unique normalised invariant differential given by $\omega(T) = F_X(0,T)^{-1}dT \in R[\![T]\!]dT$. Every invariant differential is of the form $a\omega$ for some $a \in R$.*

Proof. Let $P(T) = F_X(0,T)^{-1}$. Note that $F_X(0,T) = 1 + \ldots$ is invertible, so that $P(T)$ is indeed a member of $R[\![T]\!]$. Furthermore, $P(0) = 1$, so that it is normalised.

We need to show that ω is an invariant differential. Recall from Definition 5.9 that this is equivalent to: $P\left(F(T,S)\right)F_X(T,S) = P(T)$ so, in our case, it is sufficient to show:

$$F_X\left(0, F(T,S)\right)^{-1} F_X(T,S) = F_X(0,T)^{-1},$$

which is true iff:

$$F_X\left(0, F(T,S)\right) = F_X(T,S)F_X(0,T).$$

But this last statement is immediate from differentiating $F\left(U, F(T,S)\right) = F\left(F(U,T),S\right)$ (associativity) with respect to U to get: $F_X\left(U, F(T,S)\right) = F_X\left(F(U,T),S\right)F_X(U,T)$ and setting $U = 0$. Hence ω is an invariant differential.

Suppose that $\hat{\omega}(T) = Q(T)\mathrm{d}T \in R[\![T]\!]\mathrm{d}T$ is also an invariant differential, so that $Q(T)$ satisfies $Q\left(F(T,S)\right)F_X(T,S) = Q(T)$. Substituting $T = 0$ gives $Q(S)F_X(0,S) = Q(0)$, so that $Q(S) = Q(0)F_X(0,S)^{-1}$. It follows that $\hat{\omega} = a\omega$, where $a = Q(0)$. □

Corollary 5.12. *Let f be a homomorphism over R from the formal group F to the formal group G. Let ω_F, ω_G be the normalised invariant differentials on F, G, respectively. Then $\omega_G \circ f = f'(0)\ \omega_F$.*

Proof. First, note that $\omega_G \circ f\,(F(T,S)) = \omega_G\,(G(f(T),f(S))) = \omega_G \circ f(T)$, so that $\omega_G \circ f$ is an invariant differential on F. From the previous result, it follows that $\omega_G \circ f = a\,\omega_F$, for some $a \in R$. Since ω_F, ω_G are normalised, $(1+\ldots)\mathrm{d}f(T) = a(1+\ldots)\mathrm{d}T$, and so $(1+\ldots)f'(T)\mathrm{d}T = a(1+\ldots)\mathrm{d}T$; equating constant terms gives $a = f'(0)$, as required. □

Corollary 5.13. *Let F be a formal group over R and let, as usual, $[m](T) \in R[\![T]\!]$ denote the multiplication by m map on F, as in Definition 5.6. Let p be prime. Then there exist $f, g \in R[\![T]\!]$ $(f(T) = T + \ldots)$, such that $[p](T) = pf(T) + g(T^p)$.*

Proof. Let ω be the normalised invariant differential on F. Since $[p](T) = pT + \ldots$, it satisfies $[p]'(0) = p$. Applying the previous result to $[p]$, a homomorphism from F to itself, gives: $\omega \circ [p] = [p]'(0)\omega = p\omega$, and so

$$p\omega(T) = \omega \circ [p](T) = (1+\ldots)\mathrm{d}([p](T)) = (1+\ldots)[p]'(T)\mathrm{d}T.$$

Hence $[p]'(T) \in p\,R[\![T]\!]$. Each term a_nT^n in $[p](T)$ must then satisfy $p|na_n$ in R, and so $p|n$ in $\mathbb{Z}$ or $p|a_n$ in R, as required. □

Definition 5.14. Let $\omega(T) = P(T)\mathrm{d}T = (1 + c_1T + c_2T^2 + \ldots)\mathrm{d}T$ be the normalised invariant differential for the formal group F over R. For the special case when our ring R is a field of characteristic 0, we can define the formal logarithm by: $\log_F(T) = \int \omega(T) = \int P(T)\mathrm{d}T = T + \frac{c_1}{2}T^2 + \frac{c_2}{3}T^3 + \ldots$ and the *formal exponential function* $\exp_F(T)$ as the unique member of $R[\![T]\!]$ satisfying $\log_F(\exp_F(T)) = \exp_F(\log_F(T)) = T$, which exists by Lemma 5.7.

Theorem 5.15. *Let R be a field of characteristic 0; then $\log_F$ (as in the previous definition) is an isomorphism from F to $\widehat{\mathbb{G}}_a$, the additive group $X + Y$.*

Proof. Differentiating $\log_F(F(T,S)) - \log_F(T)$ with respect to T gives:

$P(F(T,S))\,F_X(T,S) - P(T)$ (and this $= 0$, since $\omega(T) = P(T)\mathrm{d}T$ is an invariant differential),

and so $\log_F(F(T,S)) - \log_F(T)$ is a power series purely in S, which we denote $f(S)$; that is: $\log_F(F(T,S)) = \log_F(T) + f(S)$. Putting $T = 0$ forces $f(S) = \log_F(S)$. Hence $\log_F$ is a homomorphism; the inverse is $\exp_F$, and so $\log_F$ is an isomorphism. □

Comment 5.16. Note that our proof of the existence of the invariant differential required no appeal to the commutativity axiom $F(X,Y) = F(Y,X)$. If our formal group F is defined over any integral domain R of characteristic 0 (such as $\mathbb{Z}$ or any $\mathbb{Z}_p$), we can define $\log_F, \exp_F$ over K, the field of fractions of R, and see that $F(X,Y) = \exp_F(\log_F(X) + \log_F(Y))$, which forces F to be commutative. So, at least when F is defined over an integral domain of characteristic 0, we have the somewhat surprising fact that the commutativity axiom is redundant; it can be deduced from: $F(X,Y) = X + Y +$ terms of degree $\geqslant 2$ and associativity. It is possible to construct non-commutative formal groups, but only when defined over unusual rings.

Definition 5.17. Let K be a field, complete with respect to a discrete non-Archimedean valuation, $R = \{x \in K : |x| \leqslant 1\}$ be the valuation ring, $\mathfrak{m} = \{x \in K : |x| < 1\}$ be the maximal ideal, and assume that $k = R/M$ (the residue field) is of characteristic p (for example, $K = \mathbb{Q}_p$, $R = \mathbb{Z}_p$, $\mathfrak{m} = p\mathbb{Z}_p$, $k = \mathbb{F}_p$). Let F be a formal group defined over R. The *group on $\mathfrak{m}$ associated to $F(X,Y)$*, denoted $F(\mathfrak{m})$, is the set $\mathfrak{m}$ together with the group operation: $x \oplus y = F(x,y)$ (which converges for any $x, y \in \mathfrak{m}$). The identity element is 0, and the inverse of x is given by $i(x)$ of Lemma 5.3. Similarly, for any $n \geqslant 1$, define $F(\mathfrak{m}^n)$ to be the set $\mathfrak{m}^n$ with the same group operation.

Comment 5.18. To check that $(F(\mathfrak{m}), \oplus)$ is indeed a group requires checking that various identities between formal power series imply equalities when substituting arguments in $\mathfrak{m}$ for the variables. You can refer to Question 10 on Problem Sheet 3 if you want to see an example of a general result showing that this is valid.

Lemma 5.19. *Let $F, K, R, \mathfrak{m}, k$ (with $\mathrm{char}(k) = p$) be as in Definition 5.17.*
(a) The identity map: $F(\mathfrak{m}^n)/F(\mathfrak{m}^{n+1}), \oplus \to \mathfrak{m}^n/\mathfrak{m}^{n+1}, +$ is an isomorphism.
(b) Every torsion element of $F(\mathfrak{m})$ has order a power of p.

Proof. (a) For any $x, y \in \mathfrak{m}^n$, $x \oplus y = x + y + \ldots \equiv x + y \pmod{\mathfrak{m}^{2n}}$, and so is $\equiv x + y \pmod{\mathfrak{m}^{n+1}}$.

(b) It is sufficient to show there does not exist a point of finite order m for any $m > 1$ with $p \nmid m$ (since any w of order mp^n gives $p^n w$ of order m). But, since $\mathrm{char}(k) = p$, and $p \nmid m$, we have $|m| = 1$ and so $m \in R^*$. By Lemma 5.8, $[m]$ is an isomorphism from $\mathfrak{m}$ to $\mathfrak{m}$, which must then have trivial kernel: $[m]z = 0 \implies z = 0$, as required. □

Theorem 5.20. *Let $F, K, R, \mathfrak{m}, k$ (with $\mathrm{char}(k) = p$) be as in Defn 5.17. Suppose that $z \in F(\mathfrak{m})$ has exact order p^n, for some $n \geqslant 1$, so that $[p^n](z) = 0$, but $[p^{n-1}](z) \neq 0$. Then:*

$$|z| \geqslant |p|^{\frac{1}{p^n - p^{n-1}}}.$$

Proof. If $\mathrm{char}(R) \neq 0$ then $|p| = 0$, so assume that $\mathrm{char}(R) = 0$. We have from Corollary 5.13 that $[p](T) = pf(T) + g(T^p)$ for some $f(T) = T + \ldots \in R[\![T]\!]$ and $g(T) \in R[\![T]\!]$. We shall proceed by induction on n.

Suppose $z \neq 0$, $z \in \mathfrak{m}$ and $[p](z) = 0$. Then $0 = pf(z) + g(z^p) = p(z + \ldots) + g(z^p)$. We cannot have $|pz| > |z^p|$, since then the term pz would have valuation strictly greater than the valuations all other terms. Hence $|pz| \leqslant |z^p| = |z|^p$, and so $|p| \leqslant |z|^{p-1}$, giving $|z| \geqslant |p|^{\frac{1}{p^1 - p^0}}$, proving the result for $n = 1$.

Now, assume the result is true for n, and let $z \in F(\mathfrak{m})$ have order p^{n+1}. Then $[p](z)$ has order p^n, and by the induction hypothesis, $|[p](z)| \geqslant |p|^{\frac{1}{p^n - p^{n-1}}}$. Hence:

$$|p|^{\frac{1}{p^n - p^{n-1}}} \leqslant |[p](z)| = |pf(z) + g(z^p)| \leqslant \max\left(|pz|, |z^p|\right).$$

But $|z| < 1, |p| < 1$, so that $|p|^{\frac{1}{p^n - p^{n-1}}} \geqslant |p| > |pz|$, giving $|p|^{\frac{1}{p^n - p^{n-1}}} \leqslant |z^p|$, and so $|z| \geqslant |p|^{\frac{1}{p^{n+1} - p^n}}$, as required. □

This has immediate consequences for elliptic curves.

Corollary 5.21. *Let $\mathcal{E} : y^2 = x^3 + Ax + B$, be an elliptic curve, where $A, B \in \mathbb{Z}_p$. The kernel $\mathcal{E}_1(\mathbb{Q}_p)$ of the reduction map $\sim : \mathcal{E}_0(\mathbb{Q}_p) \to \widetilde{\mathcal{E}}_{ns}(\mathbb{F}_p)$ has no torsion (apart from $\underline{\mathbf{o}}$). Any $(x, y) \in \mathcal{E}_{\mathrm{tors}}(\mathbb{Q}_p)$ satisfies $|x|_p \leqslant 1, |y|_p \leqslant 1$. When $\widetilde{\mathcal{E}}$ is non-singular, $\mathcal{E}_{\mathrm{tors}}(\mathbb{Q}_p)$ is isomorphic to a subgroup of $\widetilde{\mathcal{E}}(\mathbb{F}_p)$.*

Proof. Let $\underline{\mathbf{o}} \neq (x, y) \in \mathcal{E}(\mathbb{Q}_p)$ be in the kernel of reduction, that is, $|x|_p, |y|_p > 1$. Then, from the equation for $\mathcal{E}$, $|y|_p = |x|_p^{3/2}$ and $|z| = |-x/y|_p = |x|_p^{-1/2} < 1, |w| = |-1/y|_p < 1$. If (x, y) were torsion, then z would be a torsion point in $F_{\mathcal{E}}(\mathfrak{m}) = F_{\mathcal{E}}(p\mathbb{Z}_p)$. By Lemma 5.19(b) it must be of order p^n, and so by Theorem 5.20 must satisfy $1 > |z|_p \geqslant |p|_p^{\frac{1}{p^n - p^{n-1}}}$. Note that, since $|p|_p = p^{-1}$, any p^n apart from 2^1 (so that $p^n - p^{n-1} > 1$) would force $1 > |z|_p > p^{-1}$, contradicting the fact that $|z|_p$ is p^r for some integer r. The only remaining possibility is that (x, y) is of order 2; but then $y = 0$ and x is a root

of $x^3 + Ax + B$; this is incompatible with $|x|_p > 1$ (which makes x^3 have strictly larger valuation than Ax and B). We conclude that x, y cannot be torsion, and that there is no torsion (apart from $\underline{\mathbf{o}}$) in the kernel of reduction.

When $\widetilde{\mathcal{E}}$ is non-singular, $\mathcal{E}_0(\mathbb{Q}_p) = \mathcal{E}(\mathbb{Q}_p)$, $\widetilde{\mathcal{E}}_{ns}(\mathbb{F}_p) = \widetilde{\mathcal{E}}(\mathbb{F}_p)$, and the kernel of the reduction map $\sim : \mathcal{E}(\mathbb{Q}_p) \to \widetilde{\mathcal{E}}(\mathbb{F}_p)$ contains no nontrivial torsion. So it is injective when restricted to $\mathcal{E}_{\text{tors}}(\mathbb{Q}_p)$; hence $\mathcal{E}_{\text{tors}}(\mathbb{Q}_p)$ is isomorphic to a subgroup of $\widetilde{\mathcal{E}}(\mathbb{F}_p)$. $\square$

Section 6. Global Torsion

Aside: We now turn to elliptic curves defined over $\mathbb{Q}$, initially concentrating on the group $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ of points of finite order. Any elliptic curve $\mathcal{E} : y^2 = x^3 + Ax + B$, defined over $\mathbb{Q}$ can be transformed with a map of the form $(x, y) \mapsto (k^2x, k^3y)$ so that $A, B \in \mathbb{Z}$. The following result is a consequence over $\mathbb{Q}$ of the p-adic results of the last section.

Lemma 6.1. *Let $\mathcal{E} : y^2 = x^3 + Ax + B$, where $A, B \in \mathbb{Z}$, be an elliptic curve (so that $\Delta = 4A^3 + 27B^2 \neq 0$). Let p be a prime satisfying: $p \neq 2$ and $p \nmid \Delta$ (such a prime is said to be of good reduction, since $\widetilde{\mathcal{E}}$ mod p is still an elliptic curve over $\mathbb{F}_p$). Then $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ is isomorphic to a subgroup of $\widetilde{\mathcal{E}}(\mathbb{F}_p)$, and so $\#\mathcal{E}_{\text{tors}}(\mathbb{Q}) \mid \#\widetilde{\mathcal{E}}(\mathbb{F}_p)$.*

Proof. Since $\mathbb{Q} \subset \mathbb{Q}_p$, for any p, $\mathcal{E}(\mathbb{Q}) \leqslant \mathcal{E}(\mathbb{Q}_p)$ and $\mathcal{E}_{\text{tors}}(\mathbb{Q}) \leqslant \mathcal{E}_{\text{tors}}(\mathbb{Q}_p)$. Since $p \nmid \Delta$ we have $\widetilde{\Delta} \neq 0$ in $\mathbb{F}_p$; since $\text{char}(\mathbb{F}_p) \neq 2$, this is enough to guarantee that $\widetilde{\mathcal{E}}$ is non-singular, and so $\widetilde{\mathcal{E}}_{ns}(\mathbb{F}_p) = \widetilde{\mathcal{E}}(\mathbb{F}_p)$. By the last result of the previous section (Corollary 5.21), $\mathcal{E}_{\text{tors}}(\mathbb{Q}_p)$ is isomorphic to a subgroup of $\widetilde{\mathcal{E}}(\mathbb{F}_p)$, as must also be $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ (since $\mathcal{E}_{\text{tors}}(\mathbb{Q}) \leqslant \mathcal{E}_{\text{tors}}(\mathbb{Q}_p)$). Lagrange's Theorem then tells us that $\#\mathcal{E}_{\text{tors}}(\mathbb{Q}) \mid \#\widetilde{\mathcal{E}}(\mathbb{F}_p)$. □

Note that, in particular, the above result tells us that $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ is always finite. In practice, we can use reductions modulo finite fields to try to determine $\mathcal{E}_{\text{tors}}(\mathbb{Q})$.

Example 6.2. Let $\mathcal{E} : y^2 = x^3+3$, defined over $\mathbb{Q}$. Then $\Delta = 4A^3+27B^2 = 4{\cdot}0^3+27{\cdot}3^2 = 3^5$. We can choose any prime $p \neq 2, p \nmid \Delta$, that is, $p \neq 2, 3$.

$p = 5$. $\widetilde{\mathcal{E}} : y^2 = x^3+3$, defined over $\mathbb{F}_5$. Then $\widetilde{\mathcal{E}}(\mathbb{F}_5)$ consists of: $\underline{\mathbf{o}}, (1, \pm 2), (2, \pm 1), (3, 0)$, giving 6 points. So $\#\mathcal{E}_{\text{tors}}(\mathbb{Q}) \mid \#\widetilde{\mathcal{E}}(\mathbb{F}_5)$, that is: $\#\mathcal{E}_{\text{tors}}(\mathbb{Q}) \mid 6$.

$p = 7$. $\widetilde{\mathcal{E}} : y^2 = x^3 + 3$, defined over $\mathbb{F}_7$. Then $\widetilde{\mathcal{E}}(\mathbb{F}_7)$ consists of:
$\underline{\mathbf{o}}, (1, \pm 2), (2, \pm 2), (3, \pm 3), (4, \pm 2), (5, \pm 3), (6, \pm 3)$, giving 13 points. So $\#\mathcal{E}_{\text{tors}}(\mathbb{Q}) \mid 13$.

The only possibility is: $\#\mathcal{E}_{\text{tors}}(\mathbb{Q}) = 1$, and so $\mathcal{E}_{\text{tors}}(\mathbb{Q}) = \{\underline{\mathbf{o}}\}$. Note that $(1, 2) \in \mathcal{E}(\mathbb{Q})$, but we know that $(1, 2)$ is not of finite order, so that $(1, 2), 2(1, 2), 3(1, 2), \ldots$ are all distinct, and can conclude that $\mathcal{E}(\mathbb{Q})$ is infinite.

Note that, if we are given (for example) $\mathcal{F} : y^2 = x^3 + \frac{3}{5^6}$, we can apply $(x, y) \mapsto (5^2x, 5^3y)$ [with inverse $(x, y) \mapsto (\frac{x}{5^2}, \frac{y}{5^3})$] to transform $\mathcal{F}$ to $\mathcal{E}$ and so deduce that $\mathcal{F}_{\text{tors}}(\mathbb{Q}) = \{\underline{\mathbf{o}}\}$ also.

Aside: Another consequence of the p-adic results of the last section is the integrality of the coordinates of any torsion point.

Lemma 6.3. *Let $(x_1, y_1) \neq \underline{\mathbf{o}}$ be a $\mathbb{Q}$-rational torsion point on $\mathcal{E} : y^2 = x^3 + Ax + B$, where $A, B \in \mathbb{Z}$. Then $x_1, y_1 \in \mathbb{Z}$.*

Proof. For any prime p, we have $A, B \in \mathbb{Z} \subset \mathbb{Z}_p$. Furthermore, $(x_1, y_1) \in \mathcal{E}_{\text{tors}}(\mathbb{Q}) \subset \mathcal{E}_{\text{tors}}(\mathbb{Q}_p)$. By the last result of the previous section (Corollary 5.21) we know that $|x_1|_p \leqslant 1, |y_1|_p \leqslant 1$. In summary: $x_1, y_1 \in \mathbb{Q}$ and $x_1, y_1 \in \mathbb{Z}_p$ for all primes p.

Imagine that $x_1 \notin \mathbb{Z}$, that is, $x_1 = \frac{m}{n}$, where $m, n \in \mathbb{Z}$, $\gcd(m, n) = 1$, $n \neq \pm 1$. Then some prime p must divide n (and not divide m), giving $|x_1|_p = |\frac{m}{n}|_p = p^r$ (for some $r > 0$), which is > 1. This contradicts $x \in \mathbb{Z}_p$, and so we conclude that $x_1 \in \mathbb{Z}$. Similarly $y_1 \in \mathbb{Z}$. □

For example, this tells us immediately that the point $(\frac{1}{4}, \frac{7}{8})$ is of infinite order on the elliptic curve $\mathcal{E} : y^2 = x^3 - x + 1$,

Aside: Reduction to finite fields usually works well enough in practice, but there is the potential problem that it might leave us with $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ undetermined. For example, suppose

that, after trying several primes, we repeatedly find that $3 \mid \#\widetilde{\mathcal{E}}(\mathbb{F}_p)$*, but a search has not found a point of order* 3*. In that case, the group* $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ *would be unresolved. It would be nice to have a finite search area within which the members of* $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ *must lie. This is provided by the following result.*

Theorem 6.4. *(Nagell-Lutz). Let* $\underline{\mathbf{o}} \neq (x_1, y_1) \in \mathcal{E}_{\text{tors}}(\mathbb{Q})$*, where* $\mathcal{E} : y^2 = x^3 + Ax + B$*, and* $A, B \in \mathbb{Z}$*. Then* $x_1, y_1 \in \mathbb{Z}$ *and either* $y_1 = 0$ *or* $y_1^2 \mid \Delta$*, where* $\Delta = 4A^3 + 27B^2$*.*

Proof. From the last lemma, $x_1, y_1 \in \mathbb{Z}$. If $y_1 = 0$ then the result is satisfied; otherwise, (x_1, y_1) is not 2-torsion and we can consider $(x_2, y_2) = 2(x_1, y_1)$, with $(x_2, y_2) \neq \underline{\mathbf{o}}$, and so $x_2, y_2 \in \mathbb{Q}$. But (x_2, y_2) is also a torsion point, so $x_2, y_2 \in \mathbb{Z}$. The line tangent to $\mathcal{E}$ at (x_1, y_1) has slope $\lambda = (3x_1^2 + A)/(2y_1)$; as usual, substituting $y = \lambda x + \mu$ into $\mathcal{E}$ gives $(\lambda x + \mu)^2 = x^3 + Ax + B$ and so $x^3 - \lambda^2 x^2 + \ldots = 0$, giving $x_1 + x_1 + x_2 = -(\text{coeff of } x^2)/(\text{coeff of } x^3) = \lambda^2$, that is:

$$x_2 = \left(\frac{3x_1^2 + A}{2y_1}\right)^2 - 2x_1 \in \mathbb{Z}.$$

Now, we know $x_1, x_2 \in \mathbb{Z}$ and so $\left(\frac{3x_1^2+A}{2y_1}\right)^2 \in \mathbb{Z}$. It follows that $4y_1^2 \mid (3x_1^2 + A)^2$ and so $y_1^2 \mid (3x_1^2 + A)^2$. Also, $y_1^2 = x_1^3 + Ax_1 + B$ and so trivially $y_1^2 \mid (x_1^3 + Ax_1 + B)$. Applying Euclid's Algorithm to $(3x^2 + A)^2$ and $x^3 + Ax + B$ gives the identity

$$\phi_1(x)\psi_1(x) + \phi_2(x)\psi_2(x) = 4A^3 + 27B^2,$$

where $\phi_1(x) = 3x^2 + 4A$, $\psi_1(x) = (3x^2 + A)^2$, $\phi_2(x) = -27(x^3 + Ax - B)$, $\psi_2(x) = x^3 + Ax + B$. Since $y_1^2 \mid \psi_1(x_1)$ and $y_1^2 \mid \psi_2(x_1)$ we must have $y_1^2 \mid (\phi_1(x_1)\psi_1(x_1) + \phi_2(x_1)\psi_2(x_1)) = \Delta$, as required. □

Example 6.5. Let $\mathcal{E} : y^2 = x^3 + 3x + 1$. Then $\Delta = 4 \cdot 3^3 + 27 \cdot 1^2 = 135 = 5 \cdot 3^3$. If $(x, y) \in \mathcal{E}_{\text{tors}}(\mathbb{Q})$, $(x, y) \neq \underline{\mathbf{o}}$, then $x, y \in \mathbb{Z}$ and either $y = 0$ or $y^2 \mid 5 \cdot 3^3$, giving only $y = 0, \pm 1, \pm 3$ as possibilities.

Case $y = \pm 1$. From $\mathcal{E}$, $(\pm 1)^2 = x^3 + 3x + 1$ and so $x(x^2 + 3) = 0$. The only solution in $\mathbb{Z}$ is $x = 0$, giving $(0, \pm 1)$ as the only possibilities.

Case $y = \pm 3$. In this case, $x \in \mathbb{Z}$ satisfies $(\pm 3)^2 = x^3 + 3x + 1$ and so $x^3 + 3x - 8 = 0$. Let $f(x) = x^3 + 3x - 8$. Any integer root x of $f(x)$ must satisfy $x|(\text{constant term}) = (-8)$, giving $x = \pm 1, \pm 2, \pm 4, \pm 8$ as the only possibilities. When we substitute these, we find that $f(1), f(-1), \ldots, f(-8)$ are all nonzero, so there are no points on $\mathcal{E}$ with $x \in \mathbb{Z}$ and $y = \pm 3$.

Case $y = 0$. In this case, $x \in \mathbb{Z}$ satisfies $0 = x^3 + 3x + 1$, and we only need to check $x = \pm 1$. neither of which are roots of $x^3 + 3x + 1$. So, there are no points on $\mathcal{E}$ with $x \in \mathbb{Z}$ and $y = 0$.

In summary, $\underline{\mathbf{o}}, (0, 1), (0, -1)$ are the only possible torsion points. Is $(0, 1) \in \mathcal{E}_{\text{tors}}(\mathbb{Q})$? If it were then so would be $2(0, 1)$. But $2(0, 1) = (0, 1) + (0, 1) = (\frac{9}{4}, -\frac{35}{8})$; the coordinates are not in $\mathbb{Z}$ and so this is not a torsion point. Hence $(0, 1)$ must have infinite order. The same must be true for $(0, -1)$, since it is the inverse of $(0, 1)$. Conclusion: $\mathcal{E}_{\text{tors}}(\mathbb{Q}) = \{\underline{\mathbf{o}}\}$.

The previous method of reductions modulo finite fields is usually quicker in practice, but the Nagell-Lutz method is an effective procedure.

Comment 6.6. It was merely to ease the algebra in previous sections that we used only the form $y^2 = x^3 + Ax + B$, and all of the previous arguments apply equally well to any

elliptic curve $\mathcal{E} : y^2 = x^3 + ax^2 + bx + c$, where $a, b, c \in \mathbb{Z}$, with Δ now taken to be the discriminant of $x^3 + ax^2 + bx + c$, which has the formula:

$$\Delta = 4a^3c + 27c^2 + 4b^3 - a^2b^2 - 18abc.$$

So, it remains true that, for any prime $p \nmid 2\Delta$, $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ is isomorphic to a subgroup of $\widetilde{\mathcal{E}}(\mathbb{F}_p)$, that $\#\mathcal{E}_{\text{tors}}(\mathbb{Q}) \mid \#\widetilde{\mathcal{E}}(\mathbb{F}_p)$, and that any $(x, y) \in \mathcal{E}_{\text{tors}}(\mathbb{Q})$ ($(x, y) \neq \underline{\mathbf{o}}$) satisfies $x, y \in \mathbb{Z}$, with $y = 0$ or $y^2 \mid \Delta$.

Section 7. A 2-isogeny on an Elliptic Curve

(In the following, we shall use upper case letters $X, Y, \ldots$ for variables, and lower case letters $x, y, \ldots$ for a point (x, y).)

Suppose that $\mathcal{E}$ is an elliptic curve over $\mathbb{Q}$, together with a $\mathbb{Q}$-rational point of order 2: $(x_0, 0)$. After a birational transformation $(x, y) \mapsto (x - x_0, y)$ (inverse $(x, y) \mapsto (x + x_0, y)$) we can assume that $(0, 0) \in \mathcal{E}(\mathbb{Q})$, so that $Y^2 =$ cubic in X, with no constant term. As usual, after mappings of the form $(x, y) \mapsto (k^2 x, k^3 y)$, we can assume that the coefficients are in $\mathbb{Z}$. So, our elliptic curve can be taken to have the form

$$\mathcal{C} : Y^2 = X(X^2 + aX + b), \quad a, b \in \mathbb{Z},\ b(a^2 - 4b) \neq 0,$$

the last condition ensuring that the curve is non-singular. The point $(0, 0)$ is of order 2 on $\mathcal{C}$.

Let $P = (x, y)$ be a point on $\mathcal{C}$, and let $P_1 = (x, y) + (0, 0) = (x_1, y_1)$. Define $T_{(0,0)}$ by:

$$T_{(0,0)} : \mathcal{C} \to \mathcal{C} : (x, y) \mapsto (x, y) + (0, 0) = (x_1, y_1).$$

That is, $P \mapsto P + (0, 0)$. What are x_1, y_1 in terms of x, y?

When $(x, y) = (0, 0)$, then $T_{(0,0)} : (0, 0) \mapsto \underline{\mathbf{o}}$, since $(0, 0)$ is of order 2. When $x \neq 0$, we first find the line through $(0, 0)$ and (x, y), which is: $Y = \frac{y}{x} X$. Substituting this into $\mathcal{C}$ gives:

$$\begin{aligned}
\left(\frac{y}{x}\right)^2 X^2 &= X(X^2 + aX + b) \\
y^2 X^2 &= x^2 X^3 + a x^2 X^2 + b x^2 X \\
x(x^2 + ax + b) X^2 &= x^2 X^3 + a x^2 X^2 + b x^2 X \text{ [since } (x, y) \text{ is on } \mathcal{C}] \\
0 &= x X^3 - (x^2 + b) X^2 + bxX, \text{ [since } x \neq 0]
\end{aligned}$$

and so $X(X - x)(xX - b) = 0$. The roots of this cubic are: $X = 0, X = x, X = b/x$. The line $Y = \frac{y}{x} X$ and $\mathcal{C}$ intersect at:

$$(0, 0), (x, y) \text{ and } \left(\frac{b}{x}, \frac{by}{x^2}\right) \text{ (since } X = \tfrac{b}{x} \text{ gives } Y = \tfrac{y}{x}\tfrac{b}{x} = \tfrac{by}{x^2})$$

and so $(x, y) + (0, 0) = \left(\frac{b}{x}, -\frac{by}{x^2}\right) = (x_1, y_1)$, where $x_1 = \frac{b}{x},\ y_1 = -\frac{by}{x^2}$.

We want to construct a 2-to-1 map ϕ from $\mathcal{C}$ to another curve $\mathcal{D}$ such that $\phi(P + (0, 0)) = \phi(P)$ for any P. We want expressions in x, y, call them $\lambda(x, y),\ \mu(x, y)$, such that $P = (x, y)$ and $P + (0, 0) = (x_1, y_1)$ map to the same (λ, μ). Natural attempts are: $x + x_1 = x + \frac{b}{x}$ and $y + y_1 = y - \frac{by}{x^2}$. It turns out to be more convenient to choose $x + x_1 + a$ instead of $x + x_1$.

$$\text{Define: } \lambda = x + x_1 + a = x + \frac{b}{x} + a = \frac{x(x^2 + ax + b)}{x^2} = \frac{y^2}{x^2} = \left(\frac{y}{x}\right)^2.$$

$$\text{Define: } \mu = y + y_1 = y - \frac{by}{x^2}.$$

Both λ, μ are invariant under $T_{(0,0)}$. We have a map from $\mathcal{C}$, given by $(x, y) \mapsto (\lambda, \mu) = \left(\left(\frac{y}{x}\right)^2, y - \frac{by}{x^2}\right)$, which we shall call ϕ. We want to find the new curve $\mathcal{D}$ which this map is to, that is, we want the equation satisfied by λ and μ. Try:

$$\mu^2 = \left(y - \frac{by}{x^2}\right)^2 = \left(\frac{y}{x}\left(x - \frac{b}{x}\right)\right)^2 = \left(\frac{y}{x}\right)^2 \left(x - \frac{b}{x}\right)^2 = \lambda\left(x^2 - 2b + \frac{b^2}{x^2}\right)$$

$$= \lambda\left(x^2+2b+\frac{b^2}{x^2}-4b\right) = \lambda\left(\left(x+\frac{b}{x}\right)^2-4b\right) = \lambda\left((\lambda-a)^2-4b\right) = \lambda(\lambda^2-2a\lambda+a^2-4b).$$

So (λ,μ) is a point on the curve $\mathcal{D}: V^2 = U(U^2+a_1U+b_1)$, where $a_1 = -2a$ and $b_1 = a^2-4b$. Our map ϕ is a rational map (but not a birational transformation, since it is 2-to-1). It is easy to check that it is a homomorphism, with kernel $\{\underline{\mathbf{o}},(0,0)\}$; such a map ϕ is a 2-*isogeny* on $\mathcal{C}$.

We can apply the same process to $\mathcal{D}$, taking $(u,v)\mapsto\left(\left(\frac{v}{u}\right)^2, v-\frac{b_1v}{u^2}\right)$ from $\mathcal{D}$ to the curve $Y^2 = X(X^2-2a_1X+a_1^2-4b_1)$, which is the same as $Y^2 = X(X^2+4aX+16b)$ (since $-2(-2a)=4a$ and $a_1^2-4b_1 = (-2a)^2-4(a^2-4b) = 16b$), that is:

$$\frac{Y^2}{64} = \frac{X}{4}\left(\frac{X^2}{16}+\frac{4aX}{16}+\frac{16b}{16}\right) = \frac{X}{4}\left(\frac{X^2}{16}+\frac{aX}{4}+b\right),$$

and so $\left(\frac{Y}{8}\right)^2 = \frac{X}{4}\left(\left(\frac{X}{4}\right)^2+a\left(\frac{X}{4}\right)+b\right)$. So, the map $\hat{\phi}:(u,v)\mapsto\left(\frac{1}{4}\left(\frac{v}{u}\right)^2, \frac{1}{8}\left(v-\frac{b_1v}{u^2}\right)\right)$ is a map from $\mathcal{D}$ back to $\mathcal{C}$ (the *dual isogeny*). The properties are the same as for ϕ, namely: $\hat{\phi}$ is a homomorphism with kernel $\{\underline{\mathbf{o}},(0,0)\}$.

Note also that, if we let $\alpha_1 = \frac{-a+\sqrt{a^2-4b}}{2}$, $\alpha_2 = \frac{-a-\sqrt{a^2-4b}}{2}$ denote the roots of X^2+aX+b, then $\phi((\alpha_1,0)) = \phi((\alpha_2,0)) = (0,0)$, and so the kernel of $\hat{\phi}\circ\phi$ consists precisely of the 2-torsion of $\mathcal{C}$, namely: $\{\underline{\mathbf{o}},(0,0),(\alpha_1,0),(\alpha_2,0)\}$. Indeed, it is easy to show that $\hat{\phi}\circ\phi$ is the multiplication by 2 map on $\mathcal{C}$. We summarise as follows.

Lemma 7.1. *Let $\mathcal{C}: Y^2 = X(X^2+aX+b)$, where $a,b\in\mathbb{Z}, b\neq 0, a^2-4b\neq 0$, and let $\mathcal{D}: V^2 = U(U^2+a_1U+b_1)$, where $a_1=-2a$ and $b_1 = a^2-4b$.*

$$\textit{Define}\quad \phi:\mathcal{C}\longrightarrow\mathcal{D}\quad \textit{by}\quad \phi(x,y) = \left(\left(\frac{y}{x}\right)^2, y-\frac{by}{x^2}\right).$$

$$\textit{Define}\quad \hat{\phi}:\mathcal{D}\longrightarrow\mathcal{C}\quad \textit{by}\quad \hat{\phi}(u,v) = \left(\frac{1}{4}\left(\frac{v}{u}\right)^2, \frac{1}{8}\left(v-\frac{b_1v}{u^2}\right)\right).$$

Then the 2-isogenies $\phi,\hat{\phi}$ are 2-to-1 homomorphisms, each with kernel $\{\underline{\mathbf{o}},(0,0)\}$. Since $\phi,\hat{\phi}$ are defined over $\mathbb{Q}$, we also have $\phi:\mathcal{C}(\mathbb{Q})\to\mathcal{D}(\mathbb{Q})$ and $\hat{\phi}:\mathcal{D}(\mathbb{Q})\to\mathcal{C}(\mathbb{Q})$. The compositions $\hat{\phi}\circ\phi$ and $\phi\circ\hat{\phi}$ are the multiplication by 2 maps $[2]$ on $\mathcal{C}$ and $\mathcal{D}$, respectively.

We shall concentrate for the moment on $\phi:\mathcal{C}\to\mathcal{D}$. Note that we can formally invert $(u,v) = \phi(x,y) = \left(\left(\frac{y}{x}\right)^2, y-\frac{by}{x^2}\right)$, as follows. Since $u = \left(\frac{y}{x}\right)^2$, we have $\frac{y}{x} = \pm u^{1/2}$. For the moment, say $\frac{y}{x} = u^{1/2}$. We also have

$$u^{-1/2}v = \frac{x}{y}\left(y-\frac{by}{x^2}\right) = x-\frac{b}{x},$$
$$u = \left(\frac{y}{x}\right)^2 = \frac{y^2}{x^2} = \frac{x(x^2+ax+b)}{x^2} = x+a+\frac{b}{x},$$

and so: $u^{-1/2}v+u = 2x+a$. Solving for x,y then gives the following preimages.

Lemma 7.2. *Let $\mathcal{C}, \mathcal{D}, \phi$ be as in Lemma 7.1, and let (u, v) be a point on $\mathcal{D}$ with $u \neq 0$. Let*

$$x_1 = \left(u + u^{-1/2}v - a\right)/2, \quad y_1 = u^{1/2}x_1 = u^{1/2}\left(u + u^{-1/2}v - a\right)/2,$$
$$x_2 = \left(u - u^{-1/2}v - a\right)/2, \quad y_2 = -u^{1/2}x_2 = -u^{1/2}\left(u - u^{-1/2}v - a\right)/2.$$

Then $\phi(x_1, y_1) = \phi(x_2, y_2) = (u, v)$.

We shall shortly make use of these to define helpful maps on $\mathcal{C}(\mathbb{Q})$ and $\mathcal{D}(\mathbb{Q})$. First, we recall the notation $\mathbb{Q}^*$ and $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ (see also Example 0.30(b)). As usual, let $\mathbb{Q}^*$ denote the group of nonzero members of $\mathbb{Q}$ under multiplication, so that $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ is $\mathbb{Q}^*$ modulo squares. For example, $\frac{12}{49} = 3$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ since $\frac{12}{49} = 3\frac{4}{49} = 3\left(\frac{2}{7}\right)^2 = 3$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$. Note that any member of $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ can be written uniquely as a square free integer (that is, as an integer not divisible by any square except 1).

Aside: Our main aim here is to show the Weak Mordell-Weil Theorem, that $\mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q})$ is finite, which we shall achieve by showing that $\mathcal{D}(\mathbb{Q})/\phi(\mathcal{C}(\mathbb{Q}))$ and $\mathcal{C}(\mathbb{Q})/\hat{\phi}(\mathcal{D}(\mathbb{Q}))$ are finite, and then using the fact that $\hat{\phi} \circ \phi = [2]$.

From now on, we denote $\mathcal{C}(\mathbb{Q})$ by $\mathcal{G}$ and $\mathcal{D}(\mathbb{Q})$ by $\mathcal{H}$ (both groups under addition $+$ given by the group law on elliptic curves, with identity $\underline{\mathbf{o}}$).

Lemma 7.3. *Let $(u, v) \in \mathcal{H}$. Then:*

$$(u, v) \in \phi(\mathcal{G}) \iff u \in (\mathbb{Q}^*)^2 \text{ or } (u = 0 \text{ and } a^2 - 4b \in (\mathbb{Q}^*)^2).$$

Proof. **Case 1** $u \neq 0$. From the expressions in Lemma 7.2 for $(x_1, y_1), (x_1, y_1)$ such that $\phi(x_1, y_1) = \phi(x_2, y_2) = (u, v)$, which are in terms of $u, v, u^{1/2}$, we see that:

$$(u, v) \in \phi(\mathcal{G}) \iff u^{1/2} \in \mathbb{Q} \iff u \in (\mathbb{Q}^*)^2.$$

Case 2 $u = 0$. The expressions in Lemma 7.2 do not apply here, since they include $u^{-1/2}$. But we know that $\phi(\alpha_1, 0) = \phi(\alpha_2, 0) = (0, 0)$, where

$$\alpha_1 = \frac{-a + \sqrt{a^2 - 4b}}{2}, \alpha_2 = \frac{-a - \sqrt{a^2 - 4b}}{2}$$

denote the roots of $X^2 + aX + b$. Hence:

$$(0, 0) \in \phi(\mathcal{G}) \iff \alpha_1 \text{ or } \alpha_2 \in \mathbb{Q} \iff a^2 - 4b \in (\mathbb{Q}^*)^2$$

as required. □

This suggests the following map on $\mathcal{H}$.

Definition 7.4. Define the map $q : \mathcal{H} \to \mathbb{Q}^*/(\mathbb{Q}^*)^2$ by:

$$q(u, v) = \begin{cases} u & \text{when } u \neq 0 \\ b_1 = a^2 - 4b & \text{when } u = 0. \end{cases}$$

We also define $q(\underline{\mathbf{o}}) = 1$.

Note that we can equivalently define $q(u, v)$ to be d such that the preimages of (u, v) under ϕ are defined over $\mathbb{Q}(\sqrt{d})$.

Lemma 7.5. *The map $q : \mathcal{H} \to \mathbb{Q}^*/(\mathbb{Q}^*)^2$ of Definition 7.4 is a homomorphism with kernel $\phi(\mathcal{G})$ (so that the induced map $q : \mathcal{H}/\phi(\mathcal{G}) \to \mathbb{Q}^*/(\mathbb{Q}^*)^2$ is an injective homomorphism).*

Proof. We only show that $q(P+Q) = q(P)q(Q)$ in the typical case when none of $P, Q, P+Q$ are $(0,0)$ or $\mathbf{o}$.

Let $(u_1, v_1), (u_2, v_2), (u_3, v_3)$ be 3 points on $\mathcal{H} = \mathcal{D}(\mathbb{Q})$ which sum to $\underline{\mathbf{o}}$, (so that $(u_1, v_1) + (u_2, v_2) = (u_3, -v_3)$). Then these are the 3 points of intersection between $\mathcal{D}$ and some line defined over $\mathbb{Q}$: $V = \ell U + m$, say.

Substituting $V = \ell U + m$ into $\mathcal{D}$ gives: $U(U^2 + a_1U + b_1) - (\ell U + m)^2$, whose 3 roots must be u_1, u_2, u_3. So

$$U(U^2 + a_1U + b_1) - (\ell U + m)^2 = (U - u_1)(U - u_2)(U - u_3).$$

Equating constant terms gives: $u_1u_2u_3 = m^2 = 1$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$, and so $u_1u_2 = 1/u_3 = u_3$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$. (Note $u_1u_2u_3 \neq 0$, by our assumption.)

Therefore, by the definition of q we have:

$$q\left((u_1, v_1)\right) q\left((u_2, v_2)\right) = q\left((u_3, -v_3)\right) = q\left((u_1, v_1) + (u_2, v_2)\right),$$

so that q is a homomorphism.

The fact that $\ker q = \phi(\mathcal{G})$ is an immediate consequence of Lemma 7.3. □

Lemma 7.6. *The map $q : \mathcal{H} \to \mathbb{Q}^*/(\mathbb{Q}^*)^2$ of Definition 7.4 has finite image. Moreover, if $r \in \mathbb{Q}^*/(\mathbb{Q}^*)^2$ is written as a square free integer, then $r \in \operatorname{im} q \implies r | b_1$.*

Under q, $\mathcal{H}/\phi(\mathcal{G})$ is isomorphic to the subgroup of $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ consisting of all square free integers $r|b_1$ such that there is are solutions $\ell, m, n \in \mathbb{Z}$, not all 0, with $\gcd(\ell, m) = 1$ to the equation:

$$W_r : r\ell^4 + a_1\ell^2m^2 + (b_1/r)m^4 = n^2.$$

When this is satisfied, there is a point $(u, v) \in \mathcal{H}$ such that $q(u, v) = r$, satisfying $u = r\left(\frac{\ell}{m}\right)^2$.

Proof. Let $r \in \mathbb{Q}^*/(\mathbb{Q}^*)^2, r \in \operatorname{im} q, r \in \mathbb{Z}, r$ square free. We want to prove that $r|b_1$. Suppose $r = q(u, v)$, where $(u, v) \in \mathcal{D}(\mathbb{Q})$, which must exist since $r \in \operatorname{im} q$. Then: $r = q(u, v) = u = u^2 + a_1u + b_1$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ (since $u(u^2 + a_1u + b_1) = v^2$). So, $r, u, u^2 + a_1u + b_1$ are all the same modulo squares, which means we can write:

$$u^2 + a_1u + b_1 = rs^2 \text{ and } u = rt^2 \text{ for some } s, t \in \mathbb{Q}.$$

Hence: $(rt^2)^2 + a_1(rt^2) + b_1 = rs^2$. Let $t = \ell/m$, where $\ell, m \in \mathbb{Z}$ and $\gcd(\ell, m) = 1$. Then: $r^2\ell^4/m^4 + a_1r\ell^2/m^2 + b_1 = rs^2$, and so: $r^2\ell^4 + a_1r\ell^2m^2 + b_1m^4 = r(m^2s)^2$. Now, $a_1, b_1, r, \ell, m \in \mathbb{Z}$, so the LHS of this last equation is in $\mathbb{Z}$, and so the RHS is also in $\mathbb{Z}$; that is: $r(m^2s)^2 \in \mathbb{Z}$. Since r is square free, we must therefore have $m^2s \in \mathbb{Z}$. Define: $n = m^2s \in \mathbb{Z}$. Then our equation becomes:

$$r^2\ell^4 + a_1r\ell^2m^2 + b_1m^4 = rn^2, \text{ for some } \ell, m, n \in \mathbb{Z} \text{ with } \gcd(\ell, m) = 1 \qquad (*)$$

(from which we have W_r in the statement of the lemma, after dividing both side by r). We want to show that $r|b_1$, and we know that r is square free. It is sufficient to show, for any prime p, that $p|r \Rightarrow p|b_1$.

Suppose, for a contradiction, that $p|r$ and $p \nmid b_1$, for some prime p. Then

$$p | r^2\ell^4, a_1r\ell^2m^2, \text{ and } rn^2$$

and so by $(*)$, $p|b_1m^4$, which in turn gives: $p|m$ (since $p \nmid b_1$). Hence, since now $p|r$ and $p|m$, we have: $p^2|r^2\ell^4, a_1r\ell^2m^2, b_1m^4$, and so by $(*)$, $p^2|rn^2$, which in turn gives: $p|n$ (since r is square free). Hence, since now $p|r, m, n$, we have: $p^3|a_1r\ell^2m^2, b_1m^4, rn^2$, and so by $(*)$, $p^3|r^2\ell^4$, which in turn gives: $p|\ell$ (since r is square free). This is a contradiction,

since $p|\ell$ and $p|m$ but $\gcd(\ell, m) = 1$. We deduce that $p|r \Rightarrow p|b_1$ for any prime p, so $r|b_1$ as desired.

We finally note that if r satisfies W_r, then $\left(r(\ell/m)^2\right)^2 + a_1 r(\ell/m)^2 + b_1 = r(n/m^2)^2$, so

$$r(\ell/m)^2 \left(\left(r(\ell/m)^2\right)^2 + a_1 r(\ell/m)^2 + b_1 \right) = (r\ell n/m^3)^2.$$

This tells us that $(u, v) = (r(\ell/m)^2, r\ell n/m^3)$ is in $\mathcal{H}$; we have $q(u, v) = r$, which gives $r \in \text{im } q$. □

Comment 7.7. If we similarly define $\hat{q} : \mathcal{G} \to \mathbb{Q}^*/(\mathbb{Q}^*)^2$ by:

$$\hat{q}(x, y) = \begin{cases} x & \text{when } x \neq 0 \\ b = a_1^2 - 4b_1 & \text{when } x = 0, \end{cases}$$

and $\hat{q}(\underline{\mathbf{o}}) = 1$, then, by the same argument, $\hat{q}$ has finite image. If $r \in \mathbb{Q}^*/(\mathbb{Q}^*)^2$ is written as a square free integer, then $r \in \text{im } \hat{q} \implies r|b$. Under $\hat{q}$, $\mathcal{G}/\hat{\phi}(\mathcal{H})$ is isomorphic to the subgroup of $\mathbb{Q}^*/(\mathbb{Q}^*)^2$ consisting of all square free integers $r|b$ such that

$$\widehat{W}_r : r\ell^4 + a\ell^2 m^2 + (b/r)m^4 = n^2, \text{ for some } \ell, m, n \in \mathbb{Z}, \text{ not all 0, with } \gcd(\ell, m) = 1.$$

When $\widehat{W}_r$ is satisfied, there is a point $(x, y) \in \mathcal{G}$ such that $q(x, y) = r$, satisfying $x = r\left(\frac{\ell}{m}\right)^2$.

Since $\mathcal{H}/\phi(\mathcal{G})$ and $\mathcal{G}/\hat{\phi}(\mathcal{H})$ have been shown to be isomorphic to finite groups, we can immediately deduce one of our main goals.

Theorem 7.8. *Both $\mathcal{G}/\hat{\phi}(\mathcal{H})$ and $\mathcal{H}/\phi(\mathcal{G})$ are finite.*

Corollary 7.9. *(The Weak Mordell-Weil Theorem, for an elliptic curve $\mathcal{C}$ which has a rational point of order 2). $\mathcal{G}/2\mathcal{G} = \mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q})$ is finite.*

Proof. We know from Theorem 7.8 that $\mathcal{G}/\hat{\phi}(\mathcal{H})$ and $\mathcal{H}/\phi(\mathcal{G})$ are finite, so let $\mathcal{G}/\hat{\phi}(\mathcal{H}) = \{g_1, \ldots, g_k\}$ and $\mathcal{H}/\phi(\mathcal{G}) = \{h_1, \ldots, h_\ell\}$. Let $g \in \mathcal{G}$. We can write g as:

$$\begin{aligned} g &= g_i + \hat{\phi}(h), \text{ for some } g_i \in \{g_1, \ldots, g_k\}, \ h \in \mathcal{H} \\ &= g_i + \hat{\phi}\left(h_j + \phi(g')\right), \text{ for some } h_j \in \{h_1, \ldots, h_\ell\}, \ g' \in \mathcal{G} \\ &= g_i + \hat{\phi}(h_j) + \hat{\phi}(\phi(g')) \ \ (\text{since } \hat{\phi} \text{ is a homomorphism}) \\ &= g_i + \hat{\phi}(h_j) + 2g' \ \ (\text{since } \hat{\phi} \circ \phi = [2]) \\ &= g_i + \hat{\phi}(h_j) \quad \text{in } \mathcal{G}/2\mathcal{G}. \end{aligned}$$

Hence $\mathcal{G}/2\mathcal{G}$ is a subset of $\{g_i + \hat{\phi}(h_j) : 1 \leqslant i \leqslant k, \ 1 \leqslant j \leqslant \ell\}$, which is finite, and so $\mathcal{G}/2\mathcal{G}$ is finite. □

The above proves the Weak Mordell-Weil Theorem, that $\mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q})$ is finite, for the case when $\mathcal{C} : Y^2 = X(X^2 + aX + b)$ has a $\mathbb{Q}$-rational point of order 2. In fact, the same result can be proved for any elliptic curve $\mathcal{E} : Y^2 = F(X)$, regardless of whether it has a $\mathbb{Q}$-rational point of order 2 (see Chapter VIII of [2]), giving:

Theorem 7.10. *(The Weak Mordell-Weil Theorem). Let $\mathcal{E}$ be any elliptic curve over $\mathbb{Q}$. Then $\mathcal{E}(\mathbb{Q})/2\mathcal{E}(\mathbb{Q})$ is finite.*

The proof of the more general version is in a similar spirit, but requires some algebraic number theory, working in the number field $\mathbb{Q}(\alpha)$, where α is a root of $F(X)$.

Comment 7.11. A *Boolean* group is defined to be a group such that $g * g$ is the identity, for any element g. A finite Boolean group, generated by the independent elements $g_1, \ldots, g_n$, has 2^n elements. Given any Abelian group G, the quotient group $G/2G$ is always Boolean. When $G/2G$ is finite, $\#G/2G$ is always a power of 2 and is isomorphic to $C_2 \times \ldots \times C_2$.

Suppose we are give an elliptic curve of the form $\mathcal{C} : Y^2 = X(X^2 + aX + b)$, and we derive the associated objects already described, namely $\mathcal{D} : V^2 = U(U^2 + a_1U + b_1)$, where $a_1 = -2a, b_1 = a^2 - 4b$, with $\mathcal{G} = \mathcal{C}(\mathbb{Q}), \mathcal{H} = \mathcal{D}(\mathbb{Q})$, $\phi : \mathcal{G} \to \mathcal{H}$, $\hat{\phi} : \mathcal{H} \to \mathcal{G}$, $q : \mathcal{H}/\phi(\mathcal{G}) \to \mathbb{Q}^*/(\mathbb{Q}^*)^2$, $\hat{q} : \mathcal{G}/\hat{\phi}(\mathcal{H}) \to \mathbb{Q}^*/(\mathbb{Q}^*)^2$. Then the above results and their proofs give a method for trying to compute $\mathcal{G}/2\mathcal{G}$.

Step 1. Try to find $\mathcal{H}/\phi(\mathcal{G})$ by finding all square free integers $r|b_1$ satisfying W_r.

Step 2. Try to find $\mathcal{G}/\hat{\phi}(\mathcal{H})$ by finding all square free integers $r|b$ satisfying $\widehat{W_r}$.

Step 3. Combine $\mathcal{G}/\hat{\phi}(\mathcal{H})$ and $\hat{\phi}\left(\mathcal{H}/\phi(\mathcal{G})\right)$ to generate $\mathcal{G}/2\mathcal{G}$.

Example 7.12. Let $\mathcal{C} : Y^2 = X(X^2 - X + 6)$. Then $\mathcal{G}/2\mathcal{G} = \mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q}) \cong C_2 \times C_2$.

Proof. Here, $a = -1, b = 6$ and so $a_1 = -2a = 2,\ b_1 = a^2 - 4b = -23$, giving $\mathcal{D} : V^2 = U(U^2 + 2U - 23)$. The isogeny $\phi : \mathcal{C} \to \mathcal{D}$ is given by

$$\phi(x, y) = \left(\left(\frac{y}{x}\right)^2, y - \frac{by}{x^2}\right) = \left(\left(\frac{y}{x}\right)^2, y - \frac{6y}{x^2}\right)$$

The isogeny $\hat{\phi} : \mathcal{D} \longrightarrow \mathcal{C}$ is given by

$$\hat{\phi}(u, v) = \left(\frac{1}{4}\left(\frac{v}{u}\right)^2, \frac{1}{8}\left(v - \frac{b_1 v}{u^2}\right)\right) = \left(\frac{1}{4}\left(\frac{v}{u}\right)^2, \frac{1}{8}\left(v + \frac{23v}{u^2}\right)\right)$$

Step 1. Find $\mathcal{H}/\phi(\mathcal{G})$. We need to consider $r|b_1 = -23, r \in \mathbb{Z}$, r square free, that is, $r = \pm 1, \pm 23$, and $q(\underline{\mathbf{o}}) = 1,\ q(0, 0) = b_1 = -23$, so that: $\{1, -23\} \leqslant \text{im } q \leqslant \{\pm 1, \pm 23\}$. Note that $-1 \in \text{im } q \iff 23 \in \text{im } q$, and so it is only necessary to check one member of the coset $\{-1, 23\}$.

Choose $r = -1$. Then equation W_r, $r\ell^4 + a_1\ell^2m^2 + (b_1/r)m^4 = n^2$ becomes:

$$W_{-1} : -\ell^4 + 2\ell^2m^2 + 23m^4 = n^2, \quad \text{for some } \ell, m, n \in \mathbb{Z}, \text{ not all 0, with } \gcd(\ell, m) = 1.$$

On completing the square, we obtain:

$$-(\ell^2 - m^2)^2 + 24m^4 = n^2. \qquad (1)$$

This gives $-(\ell^2 - m^2)^2 \equiv n^2 \pmod 3$.

Imagine $3 \nmid (\ell^2 - m^2)$; then $\ell^2 - m^2$ would have an inverse α mod 3, and so $-1 \equiv (\alpha n)^2 \pmod 3$, contradicting the fact that -1 is not a quadratic residue mod 3.

We deduce that $3|(\ell^2 - m^2)$, and so $3|n$ (since $3|n^2$), giving that $3^2|(\ell^2 - m^2)^2$ and $3^2|n^2$. Then, from (1), $3^2|24m^4$, and so $3|m^4$ and hence $3|m$.

But combining $3|m$ with $3|(\ell^2 - m^2)$ gives $3|\ell^2$, so that $3|\ell$. We have shown that $3|\ell$ and $3|m$, contradicting $\gcd(\ell, m) = 1$. Hence there are no solutions to W_{-1}, giving that $-1 \notin \text{im } q$ (indeed, we have shown that there are no solutions $(\ell, m, n) \neq (0, 0, 0)$ in $\mathbb{Q}_3$).

This gives $\text{im } q = \{1, -23\}$ and $\mathcal{H}/\phi(\mathcal{G}) = \{\underline{\mathbf{o}}, (0, 0)\} = \langle (0, 0) \rangle \cong C_2$.

Step 2. Find $\mathcal{G}/\hat{\phi}(\mathcal{H})$. We need to consider $r|b = 6, r \in \mathbb{Z}$, r square free, that is, $r = \pm 1, \pm 2, \pm 3, \pm 6$. Also, $\hat{q}(\underline{\mathbf{o}}) = 1,\ \hat{q}(2, 4) = 2,\ \hat{q}(3, -6) = 3,\ \hat{q}(0, 0) = b = 6$, so that $\{1, 2, 3, 6\} \leqslant \text{im } \hat{q} \leqslant \{\pm 1, \pm 2, \pm 3, \pm 6\}$. Note that $-1 \in \text{im } \hat{q} \iff -2 \in \text{im } \hat{q} \iff -3 \in \text{im } \hat{q} \iff -6 \in \text{im } \hat{q}$, and so it is only necessary to check one member of the coset $\{-1, -2, -3, -6\}$.

Choose $r=-1$. Then $\widehat{W}_{-1}$, $r\ell^4+a\ell^2m^2+(b/r)m^4=n^2$ becomes:

$$\widehat{W}_{-1}: -\ell^4-\ell^2m^2-6m^4=n^2, \quad \text{for some } \ell,m,n\in\mathbb{Z}, \text{ not all 0, with } \gcd(\ell,m)=1.$$

For any $\ell,m,n\in\mathbb{Z}$, $\ell^4,\ell^2m^2,6m^4\geqslant 0$, so $-\ell^4-\ell^2m^2-6m^4\leqslant 0$, and

$$\text{LHS}=-\ell^4-\ell^2m^2-6m^4=0 \iff \ell^4=\ell^2m^2=6m^4=0 \iff \ell=m=0.$$

Also, RHS $=n^2\geqslant 0$ and $n^2=0 \iff n=0$. Both sides are equal $\iff$ both sides are $0 \iff \ell=m=n=0$, but we require ℓ,m,n to be not all 0. Hence there are no solutions to $\widehat{W}_{-1}$, giving that $-1\notin \text{im }\hat{q}$ (indeed, we have shown that there are no solutions $(\ell,m,n)\neq(0,0,0)$ in $\mathbb{R}$).

We conclude that $\text{im }\hat{q}=\{1,2,3,6\}$ and $\mathcal{G}/\hat{\phi}(\mathcal{H})=\{\underline{\mathbf{o}},(0,0),(2,4),(3,-6)\}=\langle(0,0),(2,4)\rangle$.

Step 3. Find $\mathcal{G}/2\mathcal{G}$. This is generated by $\mathcal{G}/\hat{\phi}(\mathcal{H})=\{\underline{\mathbf{o}},(0,0),(2,4),(3,-6)\}=\langle(0,0),(2,4)\rangle$, together with $\hat{\phi}\left(\mathcal{H}/\phi(\mathcal{G})\right)=\{\hat{\phi}(\underline{\mathbf{o}}),\hat{\phi}(0,0)\}=\{\underline{\mathbf{o}}\}$, which gives nothing new that wasn't already in $\mathcal{G}/\hat{\phi}(\mathcal{H})$. Therefore, $\mathcal{G}/2\mathcal{G}=\{\underline{\mathbf{o}},(0,0),(2,4),(3,-6)\}=\langle(0,0),(2,4)\rangle\cong C_2\times C_2$, as required. Note that $(0,0),(2,4)$ are independent in $\mathcal{G}/\hat{\phi}(\mathcal{H})$ and so are independent in $\mathcal{G}/2\mathcal{G}$ (since $2\mathcal{G}=\hat{\phi}(\phi(\mathcal{G}))\leq\hat{\phi}(\mathcal{H})$). □

Comment 7.13. The equations

$$W_r: r\ell^4+a_1\ell^2m^2+(b_1/r)m^4=n^2,$$
$$\widehat{W}_r: r\ell^4+a\ell^2m^2+(b/r)m^4=n^2,$$

(which can also be expressed as: $rX^4+a_1X^2+b_1/r=Y^2$ and $rX^4+aX^2+b/r=Y^2$, for $X,Y\in\mathbb{Q}$) are called *homogeneous spaces*. Finding $\mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q})$, as in the last example, comes down to deciding, for each $r|b_1$, whether W_r has a solution $\ell,m,n\in\mathbb{Z}$, not all 0, with $\gcd(\ell,m)=1$, and for each $r|b$, whether $\widehat{W}_r$ has such a solution.

In the last example, it turned out that each $W_r,\widehat{W}_r$ either had a solution ℓ,m,n, or we were able to show such a solution was impossible with a modulo-power-of-p argument (a p-adic argument) or that it was impossible in $\mathbb{R}$. That is, each $W_r,\widehat{W}_r$ either had a point or it was impossible in $\mathbb{R}$ or some $\mathbb{Q}_p$.

This doesn't always happen. It is possible in some examples for W_r or $\widehat{W}_r$ to have solutions in $\mathbb{R}$ and every $\mathbb{Q}_p$, but not in $\mathbb{Q}$ (that is, for there to be a violation of the Hasse Principle). For example, consider $\mathcal{C}: Y^2=X^3+17X$. Here, $a=0,b=17$, so that $a_1=0,b_1=-68$, giving $\mathcal{D}: Y^2=X^3-68X$. When computing $\mathcal{H}/\phi(\mathcal{G})$, we consider $r|b_1=-68$ and so $r=\pm1,\pm2,\pm17,\pm34$. For the case $r=2$, the homogeneous space $r\ell^4+a_1\ell^2m^2+(b_1/r)m^4=n^2$ becomes $2\ell^4-34m^4=n^2$. Note that the equation forces n to be even; setting $n=2k$ and dividing both sides by 2 gives the slightly simpler form: $\ell^4-17m^4=2k^2$. As shown on Problem Sheet 3, this has no solutions $k,\ell,m\in\mathbb{Z}$ (not all 0, $\gcd(\ell,m)=1$), and so $2\notin\text{im }q$, even though there exist solutions in $\mathbb{R}$ and every $\mathbb{Q}_p$ (and so proving $2\notin\text{im }q$ requires an argument different to those in the last example). Instances of such W_r (or $\widehat{W}_r$) correspond to members of a structure known as the *Tate–Shafarevich group*, $Ш(\mathcal{C}/\mathbb{Q})$.

Comment 7.14. There is another approach to the Weak Mordell-Weil Theorem, using Galois cohomology. Recall that the map

$$q: \mathcal{D}(\mathbb{Q})/\phi(\mathcal{C}(\mathbb{Q}))\to\mathbb{Q}^*/(\mathbb{Q}^*)^2$$

is given by $q(Q) = d$, where $\mathbb{Q}(\sqrt{d})$ is the field over which the pre-images P, P' of Q under ϕ are defined. Since $\ker \phi = \{\underline{\mathbf{o}}, (0,0)\}$, we must have $P' = P + (0,0)$. Furthermore, if $\mathrm{Gal}(\mathbb{Q}(\sqrt{d})/\mathbb{Q}) = \langle\sigma\rangle$ has order two (i.e. d is not a square), then $P' = \sigma(P)$.

So, we have a group homomorphism

$$c_Q : \mathrm{Gal}(\mathbb{Q}(\sqrt{d})/\mathbb{Q}) \to \ker \phi$$

given by sending σ to $\sigma(P) - P$. It has the property that, for any member of $\{P, P'\}$, the effect of applying $\gamma \in \mathrm{Gal}(\mathbb{Q}(\sqrt{d})/\mathbb{Q})$ is the same as adding $c_Q(\gamma)$. We then have a map $Q \mapsto c_Q$ which takes a member of $\mathcal{D}(\mathbb{Q})/\phi(\mathcal{C}(\mathbb{Q}))$ to a homomorphism between a Galois group and $\ker \phi$.

As we have seen, there are two main elements required to prove the Weak Mordell-Weil Theorem: showing that q is a homomorphism and that im q is finite. We deal with them both in turn. For showing that q is a homomorphism, suppose that $q(Q_1) = d_1$ and $q(Q_2) = d_2$. Then, by definition, P_1, P_1' (such that $\phi(P_1) = \phi(P_1') = Q_1$) are defined over $\mathbb{Q}(\sqrt{d_1})$, and P_2, P_2' (such that $\phi(P_2) = \phi(P_2') = Q_2$) are defined over $\mathbb{Q}(\sqrt{d_2})$. Since ϕ is a homomorphism,

$$\phi(P_1 + P_2) = Q_1 + Q_2$$

and $P_1 + P_2$ is defined over $\mathbb{Q}(\sqrt{d_1}, \sqrt{d_2})$. But $\sqrt{d_1} \mapsto -\sqrt{d_1}$, $\sqrt{d_2} \mapsto -\sqrt{d_2}$ has the same effect as adding $(0,0)$ to each of P_1, P_2 and so leaves $P_1 + P_2$ unchanged. This means that $P_1 + P_2$ is in fact defined over $\mathbb{Q}(\sqrt{d_1 d_2})$. Hence $q(Q_1 + Q_2) = d_1 d_2 = q(Q_1)q(Q_2)$, giving that q is a homomorphism (without needing to work explicitly with the group law).

For the finiteness of im q, let $q(Q) = d$, a square free integer, and imagine that an odd prime p of good reduction is a factor of d. By the definition of q, there are P, P', defined over $\mathbb{Q}(\sqrt{d})$ such that $\phi(P) = \phi(P') = Q$. But, on reduction modulo $\sqrt{p}$, conjugation $\sqrt{d} \mapsto -\sqrt{d}$ has no effect modulo $\sqrt{p}$. This shows that $P' - P$ is in the kernel of the reduction map. On the other hand, we know that $P' - P$ is a 2-torsion point. So it follows from Lemma 5.19 that $P' = P$ which is a contradiction. Hence d has only primes dividing the discriminant as factors, and so has only finitely many possibilities. We note in passing that we can regard each c_Q as a homomorphism

$$c_Q : \mathrm{Gal}(L/\mathbb{Q}) \to \ker \phi$$

where L is the composite of the quadratic fields $\mathbb{Q}(\sqrt{p})$ with $p = 2$ or a prime of bad reduction.

This approach is cleaner, and more amenable to generalisation, since it does not require getting our hands dirty with explicit group law manipulations. On the other hand, it is often worth a more from-first-principles proof (as given previously), as it provides us with an explicit method for trying to compute $\mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q})$.

Section 8. The Mordell-Weil Theorem

When $\mathcal{E}$ is an elliptic curve over $\mathbb{Q}$, we've seen that $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ and $\mathcal{E}(\mathbb{Q})/2\mathcal{E}(\mathbb{Q})$ are finite. But $\mathcal{E}(\mathbb{Q})$ may sometimes be infinite (if $P \in \mathcal{E}(\mathbb{Q})$ and $P \notin \mathcal{E}_{\text{tors}}(\mathbb{Q})$ then P is of infinite order and so $\mathcal{E}(\mathbb{Q})$ is infinite). We shall show that $\mathcal{E}(\mathbb{Q})$ (whether finite or infinite) is always finitely generated. That is, we aim to show that, for any elliptic curve $\mathcal{E}$, there exists finite number of elements $P_1, \ldots, P_k \in \mathcal{E}(\mathbb{Q})$ such that every $P \in \mathcal{E}(\mathbb{Q})$ can be written as:

$$P = m_1 P_1 + \ldots + m_k P_k, \quad m_1, \ldots, m_k \in \mathbb{Z}.$$

This will be achieved via height functions; we first describe the general properties of a height function on a general Abelian group.

Definition 8.1. Let A be an Abelian group with group operation $+$.

We say that $h : A \longrightarrow \mathbb{R}$ is a *height function* if it satisfies:

(1) For any $Q \in A$, there exists $C_1 = C_1(Q)$ such that $h(P+Q) \leq 2h(P) + C_1$ for all $P \in A$.

(2) There exists C_2, independent of P, such that $h(2P) \geq 4h(P) - C_2$ for all $P \in A$.

(3) For any C_3, the set $\{P \in A : h(P) \leq C_3\}$ is finite.

Theorem 8.2. *Let A be an Abelian group which has a height function h, and suppose that $A/2A$ is finite. Then A is finitely generated.*

Proof. We are given that $A/2A$ is finite, so let $A/2A = S = \{Q_1, \ldots Q_r\} \subset A$. Let P be any element of A. Then $P = Q_{i_1}$ in $A/2A$ for some $Q_{i_1} \in S$ and so we can write: $P = 2P_1 + Q_{i_1}$, for some $P_1 \in A$. Inductively, continue to write: $P_1 = 2P_2 + Q_{i_2}, P_2 = 2P_3 + Q_{i_3}, \ldots$, where each $P_j \in A$ and each $Q_{i_j} \in S$. Now:

$$h(P_j) \leq \frac{1}{4}\left(h(2P_j) + C_2\right) \overset{\text{by (2)}}{=} \frac{1}{4}\left(h(P_{j-1} - Q_{i_j}) + C_2\right) \overset{\text{by (1)}}{\leq} \frac{1}{4}\left(2h(P_{j-1}) + C_1' + C_2\right),$$

where $C_1' = \max\{C_1(-Q) : Q \in S\}$.

So, if $h(P_{j-1}) > (C_1' + C_2)/2$ then:

$$h(P_j) < \frac{1}{4}\left(2h(P_{j-1}) + 2h(P_{j-1})\right) = h(P_{j-1}).$$

Imagine that $h(P) > (C_1' + C_2)/2$ and $h(P_j) > (C_1' + C_2)/2$ for all j. Then the sequence $h(P), h(P_1), h(P_2), \ldots$ would be strictly decreasing, giving infinitely many distinct members of A with height $\leq h(P)$, which would contradict (3). This contradiction shows that there must exist an n such that $h(P_n) \leq (C_1' + C_2)/2$. So, we can write:

$$P = 2P_1 + Q_{i_1} = 2(2P_2 + Q_{i_2}) + Q_{i_1} = \ldots,$$

and after n steps P will be written as a linear combination of P_n and members of S.

Let $T = \{Q \in A : h(Q) \leq (C_1' + C_2)/2\}$. We have shown (since $P_n \in T$) that any $P \in A$ is a linear combination of members of $S \cup T$. Furthermore, T is finite, by (3). In conclusion: A is generated by the finite set $S \cup T$, and so is finitely generated. $\square$

A height function on $\mathcal{E}(\mathbb{Q})$ can be obtained as follows.

Lemma 8.3. *Let $\mathcal{E}$ be an elliptic curve, defined over $\mathbb{Q}$. Define $h_x : \mathcal{E}(\mathbb{Q}) \to \mathbb{R}$ by:*

$$h_x\left((x,y)\right) = \log \max\left(|a|, |b|\right), \quad \textit{where } x = \frac{a}{b},\ a, b \in \mathbb{Z},\ \gcd(a,b) = 1,$$

and define $h_x(\underline{\mathbf{o}}) = 0$. *Then* h_x *is a height function on* $\mathcal{E}(\mathbb{Q})$. *Indeed, there exists a constant* C, *independent of* P, Q, *such that* $|h_x(P+Q)+h_x(P-Q)-2h_x(P)-2h_x(Q)| \leq C$, *for all* $P, Q \in \mathcal{E}(\mathbb{Q})$, *from which properties (1),(2) can be deduced (property (3) is trivially true).*

For the proof (optional) see, for example, p. 201 of [2].

Aside: The proof uses the explicit group law; for example, $x' = a'/b'$, *the* x*-coordinate of* $2P = 2(x, y)$ *is given by (quartic in* x*)/(cubic in* x*), and so* $\max(|a'|, |b'|)$ *is 'approximately'* $\max(|a|, |b|)^4$, *giving that* $\log \max(|a'|, |b'|)$ *is 'approximately'* $4 \log \max(|a|, |b|)$, *that is* $h_x(2P)$ *is 'approximately'* $4h_x(P)$. *It is only necessary to control the amount of cancellation occurring, when writing the* x*-coordinate of* $2P$ *in lowest terms.*

Theorem 8.4. *(The Mordell-Weil Theorem). Let* $\mathcal{E}$ *be any elliptic curve over* $\mathbb{Q}$. *Then* $\mathcal{E}(\mathbb{Q})$ *is finitely generated.*

Proof. This follows immediately from Theorem 7.10, Theorem 8.2 and Lemma 8.3. □

Comment 8.5. This means that we know what $\mathcal{E}(\mathbb{Q})$ looks like:

$$\mathcal{E}(\mathbb{Q}) \cong \mathcal{E}_{\text{tors}}(\mathbb{Q}) \times \mathbb{Z}^r, \text{ for some } r \geqslant 0, r \in \mathbb{Z}.$$

The number r is called the *rank* of $\mathcal{E}(\mathbb{Q})$ (or just the rank of $\mathcal{E}$). Clearly:

$$\mathcal{E}(\mathbb{Q}) \text{ has finitely many points } \iff \text{ rank } (\mathcal{E}(\mathbb{Q})) = 0.$$

To solve $\mathcal{E}(\mathbb{Q})$, we want to know: $\mathcal{E}_{\text{tors}}(\mathbb{Q})$ and r (the rank). Note that:

$$\mathcal{E}(\mathbb{Q})/2\mathcal{E}(\mathbb{Q}) \cong \mathcal{E}_{\text{tors}}(\mathbb{Q})/2\mathcal{E}_{\text{tors}}(\mathbb{Q}) \times (\mathbb{Z}/2\mathbb{Z})^r,$$

so that:

$$\mathcal{E}(\mathbb{Q})/2\mathcal{E}(\mathbb{Q}) \cong \mathcal{E}(\mathbb{Q})[2] \times C_2^r,$$

where $\mathcal{E}(\mathbb{Q})[2]$ denotes the 2-torsion subgroup of $\mathcal{E}(\mathbb{Q})$ (see Comment 0.40).

Example 8.6. Let $\mathcal{C} : Y^2 = X(X^2 - X + 6)$. In Example 7.12, we found that $\mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q}) \cong C_2 \times C_2$. Also,

$$\mathcal{C}(\mathbb{C})[2] = \{\underline{\mathbf{o}}\} \cup \{\text{points of order 2}\} = \{\underline{\mathbf{o}}, (0,0), \left(\frac{1+\sqrt{-23}}{2}, 0\right), \left(\frac{1-\sqrt{-23}}{2}, 0\right)\},$$

so that $\mathcal{C}(\mathbb{Q})[2] = \{\underline{\mathbf{o}}, (0,0)\} \cong C_2$. Since $\mathcal{C}(\mathbb{Q})/2\mathcal{C}(\mathbb{Q}) \cong \mathcal{C}(\mathbb{Q})[2] \times C_2^r$, we deduce that $C_2 \times C_2 \cong C_2 \times C_2^r$ and so the rank $r = 1$ ($\mathcal{C}(\mathbb{Q})$ is infinite, but is generated by $\mathcal{C}_{\text{tors}}(\mathbb{Q})$ and one element of infinite order).

Section 9. Factorising integers using elliptic curves

Public key cryptography (see also ASO Number Theory).
Public keys allow message to be encoded (not decoded). Suppose A wants to send the integer X to B safely; we assume that everything transmitted can be intercepted.

Step 1. B (in private) takes 2 large prime numbers p, q (usually about 250 digits) and multiplies them together to give $N = pq$, chooses an exponent d, and publicises N, d to the world.

Step 2. A (in private) computes $Y \equiv X^d \pmod N$ and sends the message Y to B.

Step 3. B privately computes $\phi(N) = \phi(p)\phi(q) = (p-1)(q-1)$ and also computes (by Euclid's Algorithm) e such that $de \equiv 1 \pmod{\phi(N)}$. Note that:

$$Y^e \equiv (X^d)^e \equiv X^{de} = X^{1+k\phi(N)} \text{ (for some } k \in \mathbb{Z}) \ \equiv X(X^{\phi(N)})^k \equiv X,$$

since $X^{\phi(N)} \equiv 1 \pmod N$ by Euler's Theorem, provided that X, N are coprime. Assuming $X < N$, this decodes the message.

Note that computing $X^d \pmod N$ (and $Y^e \pmod N$) is fast even when d is large, by writing d in base 2 as $d = 2^{k_1} + \ldots + 2^{k_m}$ $(k_1 < \ldots < k_m)$. One then obtains $X^{2^0} \equiv X$, $X^{2^1} \equiv (X^{2^0})^2$, $X^{2^2} \equiv (X^{2^1})^2, \ldots, X^{2^{k_m}}$, by k_m squaring operations, after which:

$$X^d \equiv X^{2^{k_1}} X^{2^{k_2}} \ldots X^{2^{k_m}} \pmod N,$$

which takes roughly $\log d$ operations.

Anyone wishing to crack the code must be able to compute $\phi(N)$, which requires finding p, q from $N = pq$. A naive (and very slow) approach is trial division: checking for each $c = 2, \ldots, [\sqrt{N}\,]$ whether $c|N$.

Pollard's $p-1$ factorisation method. Much better is Pollard's $p-1$ method. One chooses base a and exponent $k =$ product of powers of small primes. Compute $a^k \pmod N$ (as usual, after first writing k in binary), and then $\gcd(a^k - 1, N)$ using Euclid's Algorithm. If there exists prime $p|N$ such that $p-1|k$ ($k = (p-1)s$, say) then:

$$a^k \equiv \left(a^{p-1}\right)^s \equiv 1^s \equiv 1 \pmod p \text{ (by Fermat)},$$

provided that $p \nmid a$. This gives $p|(a^k - 1)$ and so $p|\gcd(a^k - 1, N)$. Unless we have bad luck, $\gcd(a^k - 1, N) \neq N$, and so $\gcd(a^k - 1, N)$ will be a proper factor of N.

Example 9.1. A four-letter word $L_1L_2L_3L_4$ has been divided into two pairs: L_1L_2 and L_3L_4. Each of these pairs has been converted into an integer (of at most 4 digits) via the standard map: $A \mapsto 01, B \mapsto 02, \ldots, Z \mapsto 26$. These integers have been encoded by taking each to the power of $d = 6587$, modulo $N = 10123$. The encoded message reads:

$$4268,\ 5744.$$

We shall factorise N by applying Pollard's "$p-1$" method, using base 2 and exponent 52, and then use the factorisation of N to decode the message.

Write 52 as a sum of powers of 2: $52 = 4+16+32$. First compute (modulo $N = 10123$): $2^1 \equiv 2$, $2^2 \equiv (2^1)^2 \equiv 4$, $2^4 \equiv (2^2)^2 \equiv 16$, $2^8 \equiv (2^4)^2 \equiv 256$, $2^{16} \equiv (2^8)^2 \equiv 4798$, $2^{32} \equiv (2^{16})^2 \equiv 4798^2 \equiv 1102$ (where each of these was obtained be squaring the previous one, and reducing modulo N). Since $52 = 4 + 16 + 32$, we have: $2^{52} \equiv 2^4 2^{16} 2^{32} \equiv 16 \cdot 4798 \cdot 1102 \equiv 5907 \cdot 1102 \equiv 425$ modulo N, so that $2^{52} - 1 \equiv 424$ modulo N.

Now, compute $\gcd(424, N)$ by Euclid's Algorithm:
$10123 = 23 \cdot 424 + 371$; $424 = 1 \cdot 371 + 53$; $371 = 7 \cdot 53 + 0$.

So, 53 is a factor of N. Compute $10123/53 = 191$, giving the factorisation $N = 10123 = 53 \cdot 191$.

Since $N = 53 \cdot 191$, we have $\phi(N) = 52 \cdot 190 = 9880$. Compute the gcd of $\phi(N) = 9880$ and $d = 6587$ we see:

$\left(\begin{smallmatrix}1 & 0 \\ 0 & 1\end{smallmatrix} \middle| \begin{smallmatrix}9880 \\ 6587\end{smallmatrix}\right) \to^{R_1 - R_2} \left(\begin{smallmatrix}1 & -1 \\ 0 & 1\end{smallmatrix} \middle| \begin{smallmatrix}3293 \\ 6587\end{smallmatrix}\right) \to^{R_2 - 2R_1} \left(\begin{smallmatrix}1 & -1 \\ -2 & 3\end{smallmatrix} \middle| \begin{smallmatrix}3293 \\ 1\end{smallmatrix}\right) \to^{R_1 - 3293R_2} \left(\begin{smallmatrix}* & * \\ -2 & 3\end{smallmatrix} \middle| \begin{smallmatrix}0 \\ 1\end{smallmatrix}\right)$,

where the $*$ entries need not be computed. This gives us, all in the same computation, that $\gcd(9880, 6587) = 1$, and the bottom row of the last matrix gives $\gcd(9880, 6587)$ as a linear combination of $9880, 6587$, namely: $1 = -2 \cdot 9880 + 3 \cdot 6587$. Hence $3 \cdot 6587 \equiv 1 \pmod{9880}$, that is, 3 is the inverse of 6587 modulo $\phi(N) = 9880$.

The decoding operation is therefore $Y \mapsto Y^3 \bmod N$. Computing $4268^3 = 4268^2 \cdot 4268 \equiv 4547 \cdot 4268 \equiv 805$ (modulo $N = 10123$). Also: $5744^3 = 5744^2 \cdot 5744 \equiv 2679 \cdot 5744 \equiv 1216$ (modulo $N = 10123$). The decoded message is therefore: 0805, 1216; that is: HELP.

The exponent k is typically chosen to be a product of powers of the first r primes, for some r. Pollard's $p-1$ Method is fast when there exists at least one prime $p|N$ such that $p - 1 = \#\mathbb{F}_p^*$ is only divisible by small primes, so that $\text{order}(a)|\#\mathbb{F}_p^*|k$.

When Pollard's $p-1$ method is slow for some N, we can replace 'powers of an integer base a' with multiples kP of a point P on an elliptic curve $\mathcal{E}$.

We hope that, there exists prime $p|N$ such that $\#\widetilde{\mathcal{E}}(\mathbb{F}_p)|k$, which would guarantee that $kP = \underline{\mathbf{o}}$ (the point at infinity) mod p; that is to say, a denominator divisible by p, in which case, taking the gcd of the denominator and N will reveal the factor p. This will be fast if there exists $p|N$ such that $\#\widetilde{\mathcal{E}}(\mathbb{F}_p)$ is only divisible by small primes. Each new choice of elliptic curve gives a new chance of this happening.

The Elliptic Curve Method (ECM) for attempting to factor an integer N is as follows. Choose an elliptic curve $\mathcal{E}$ mod N, some point P on $\mathcal{E}$, and some choice of k (normally a product of powers of small primes). Attempt to compute kP (mod N) and hope that, in performing one of the additions $kP = k_1P + k_2P$, a denominator will have gcd with N that is a nontrivial factor of N ($\neq 1$ and $\neq N$). See Section XI.2 in [2] (only in the 2nd edition) for more details.

Example 9.2. Let $N = 10123$, as in Example 9.1. We shall factorise N by applying the Elliptic Curve Method, using the curve $\mathcal{E} : Y^2 = X^3 + 5X - 5$ and $4P$, where $P = (1, 1)$.

The line tangent to $\mathcal{E}$ at $P = (1, 1)$ has slope y' given by $2yy' = 3x^2 + 5$, with $x = 1, y = 1$; that is, the slope is $8/2 = 4$. This tangent line also goes through $(1, 1)$ and so has equation: $Y = 4X - 3$. The x-coordinate of $2P$ is therefore $4^2 - (1 + 1) = 14$, and the y-coordinate is: $-(4 \cdot 14 - 3) = -53 \equiv 10070$, so that $Q = 2P = (14, 10070)$ (modulo $N = 10123$). We now wish to double the point $Q = 2P$, and so again the first step is to find the line tangent to $\mathcal{E}$ at Q. This has slope y' given by $2 \cdot 10070 \cdot y' = 3 \cdot 14^2 + 5$, and so we need to compute $(3 \cdot 14^2 + 5)/(2 \cdot 10070)$ (modulo $N = 10123$), for which the first step is to find the inverse of $2 \cdot 10070 \equiv 10017$ (modulo $N = 10123$). Using Euclid's Algorithm:

$10123 = 1 \cdot 10017 + 106$; $10017 = 94 \cdot 106 + 53$; $106 = 2 \cdot 53 + 0$.

So, we cannot find the inverse of 10017 (modulo $N = 10123$), and this step has given us our factor 53 of N. As in the previous example, compute $10123/53 = 191$, giving the factorisation $N = 10123 = 53 \cdot 191$.

References

[1] J.W.S. Cassels. *Lectures on Elliptic Curves*. LMS–ST **24**. Cambridge University Press, Cambridge, 1991.
[2] J.H. Silverman. *The Arithmetic of Elliptic Curves*, 2nd edition. GTM **106**. Springer-Verlag, 2009.