MA426 Elliptic curves[1]
Syllabus/Summary

Lecture course by Miles Reid

Thank you for your interest in this course. Please let me have corrections or suggestions for improving it in future years.

Contents

Introduction

The course is about $C : y^2 = x^3 + ax + b$ viewed as a curve in the (x, y)-plane, with a, b thought of as fixed. $\Delta = 4a^3 + 27b^2$ is the discriminant. We assume $\Delta \neq 0$ as part of the definition of elliptic curve. Over $\mathbb{Q}$, it is a Diophantine problem, and the Mordell–Weil theorem (Chapter 4) gives the answer $C(\mathbb{Q})$ = f.g. Abelian group. We first have to set up the group law in geometric form (Chapter 3) for this to make sense.

Over $\mathbb{C}$ the theory of elliptic functions gives rise to the quotient torus $C = \mathbb{C}/L$, where L is a lattice, and embedding C in 2-space (plus a point at infinity) is the problem of elliptic functions (Chapter 2).

[1]First latex draft by Stuart Price, April 2000

Figure 0.1: Pictures over $\mathbb{R}$ and over $\mathbb{C}$.

Nature of the course

Synthesis of geometry, algebra, analysis, number theory and algebraic geometry. The course needs some background information from Galois theory, algebraic number theory, geometry etc., but I will spend some time on the background if students need it.

1 Background: rational curves and the function theory of $\mathbb{P}^1$

A rational curve is $\mathbb{C}$ or $\mathbb{C} \cup \{\infty\} = \mathbb{P}^1_{\mathbb{C}} = S^2$. It occurs in math whenever you say that a problem is *rationally solvable*.

1.1

Figure 1.1: Picture of $g = 0$ (rational curve), $g = 1$ (elliptic curve), $g \geq 2$. (The course doesn't really say anything about $g \geq 2$.)

1.2 Reminder from complex analysis

Holomorphic and meromorphic functions on a domain in $\mathbb{C}$, poles of order k and their principal part.

1.3 Reminder: Cauchy's theorem and Laurent expansion

U is a domain in $\mathbb{C}$ and $z_0 \in U$. Then f is holomorphic on $U \setminus \{z_0\} \Rightarrow f$ has a Laurent expansion.

Corollary 1.4 (Removable singularities) *If f is bounded on U then it extends to a holomorphic function at z_0.*

1.5 Liouville's theorem

Basically the same result as 1.3, but "at infinity". Set $w = 1/z$ for the coordinate at infinity.

1. f holomorphic on $\mathbb{C}$ (entire) and of bounded growth (that is, $|f(z)| < \text{const} \cdot |z|^k$) $\Rightarrow$ f is a polynomial of degree $\leq k$.

2. f holomorphic on $\mathbb{C}$ and bounded $\Rightarrow$ f constant (particular case $k = 0$).

3. if f is meromorphic on $\mathbb{C}$ and of bounded growth at infinity (that is, $|f(z)| < \text{const} \cdot |z|^k$ for $|z| >$ some R) then we can treat it as meromorphic at infinity, with pole of order $\leq k$.

1.6 Riemann sphere: $\mathbb{C} \cup \{\infty\} = \mathbb{P}^1_{\mathbb{C}} = S^2$

Covered by two pieces $\mathbb{C}$ with coordinates z and $w = 1/z$.

1.7 Global meromorphic functions

The field of meromorphic functions (an object of analysis) equals the field of rational functions (an object of algebra):

$$M(\mathbb{P}^1_{\mathbb{C}}) = \mathbb{C}(z) = \text{rational function field} = \{p(z)/q(z) \mid p, q \in \mathbb{C}[z]\}.$$

1.8 How many rational functions?

$f(z) = \text{const} \prod (z - \alpha_i)^{m_i}$, with $m_i > 0$ corresponding to zeros, $m_i < 0$ to poles, and the order of zero or pole at infinity determined by $\sum m_i = 0$. Then

1. f is determined up to nonzero scalar multiple by its zeros and poles;

2. if $D = \sum m_i P_i$ is specified number of poles and $\deg D = \sum m_i$, then $\mathcal{L}(D) = \{$ rational functions with poles $\leq D\}$ is a vector space of dimension $1 + \deg D$. (Here $\mathcal{L}(D)$ is a particular case of the Riemann–Roch space of a divisor D, and the formula for $\dim \mathcal{L}(D)$ is a particular case of the Riemann–Roch theorem.)

1.9 Automorphisms of $\mathbb{P}^1_{\mathbb{C}}$

Automorphisms of $\mathbb{P}^1_{\mathbb{C}}$ are defined as 1-to-1 holomorphic maps (with holomorphic inverse), and are given by fractional linear transformations

$$z \mapsto (az + b)/(cz + d) \quad \text{for} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{PGL}(2, \mathbb{C})$$

that is, 2×2 matrixes with determinant $\neq 0$, modulo scalars.

1.10 Cross ratio of 4 points on $\mathbb{P}^1_{\mathbb{C}}$

The map $z \mapsto \frac{b-a}{c-b} \times \frac{c-z}{z-a}$ sends $a \mapsto \infty, b \mapsto 1, c \mapsto 0$, and $z \mapsto$ cross-ratio.

1.11 Effect of permutation

The effect of permuting the 4 points is to take the cross-ratio to

$$x, \quad \frac{1}{x}, \quad 1-x, \quad \frac{1}{1-x}, \quad \frac{x-1}{x} \quad \text{or} \quad \frac{x}{x-1}.$$

The symmetric group S_4 acts via the quotient group $S_4 \to S_3$, and the invariant cross-ratio is

$$j(x) = \frac{4}{27} \cdot \frac{(x^2-x+1)^3}{x^2(1-x)^2}.$$

Any rational function of x invariant under the cross-ratio group is a function of $j(x)$. [*** This proof is not examinable.**]

1.12 Degree of a map $f\colon \mathbb{P}^1_{\mathbb{C}} \to \mathbb{P}^1_{\mathbb{C}}$

The degree of f is defined as the number of inverse images of a general point. Reminder from Galois theory: $K \subset L$ a field extension. The degree of the extension is $\deg L/K = [L:K]$ is the dimension of L as a vector space over K.

Theorem 1.13 *The degree of a map f equals the degree of the field extension $\mathbb{C}(f(z)) \subset \mathbb{C}(z)$.*

1.14 Lüroth's theorem

Theorem 1.15 *Every intermediate field between $\mathbb{C}$ and $\mathbb{C}(z)$ is $\mathbb{C}(f(z))$ for some rational function $f(z)$.*

See 1.18

Figure 1.2: Pencil of lines through a point $O \in C$.

1.16 Geometry of conics and rational functions on $\mathbb{P}^1_{\mathbb{C}}$

A nonsingular conic $C \subset \mathbb{P}^2_{\mathbb{C}}$ is isomorphic to $\mathbb{P}^1_{\mathbb{C}}$

Take $P \in C$, let L, M be a basis of the pencil of lines through P, and set $f = L/M$ restricted to C. This gives a rational function on the conic with one zero and one pole. Basically the same as stuff on zeros and poles of meromorphic functions.

1.17 Preview of Chapter 3

For plane cubics, the picture would involve quite different geometry (line through 2 points determine a third) and quite different function theory (need to allow 2 poles before you get a nonconstant function).

1.18 Appendix: Lüroth's theorem

This section is not examinable. It's a tricksy bit of algebra, but completely elementary.

Let $\mathbb{C} \subset \mathbb{C}(x)$ be a purely transcendental extension.

Lemma 1.19 *For any nonconstant $t \in \mathbb{C}(x)$, write $t = p(x)/q(x)$ in coprime form. Then $p(x) - tq(x) \in \mathbb{C}(t)[x]$ is irreducible, and hence it is the minimal polynomial of x over $\mathbb{C}(t)$, and $[\mathbb{C}(x) : \mathbb{C}(t)] = \max \deg(p, q) = n$.*

Proof $p(x) - tq(x)$ is certainly irreducible in $\mathbb{C}[t, x]$, since it is linear in t, and p, q have no common factor. Therefore it is irreducible in $\mathbb{C}(t)[x]$ by Gauss' lemma. QED

Theorem 1.20 (Lüroth's theorem) *Let $\mathbb{C} \subset K \subset \mathbb{C}(x)$ be an intermediate field extension with $\mathbb{C} \neq K$. Then K contains a nonconstant rational functions, so that x is algebraic over K; let*

$$f_0(z) = z^n + a_1 z^{n-1} + \cdots + a_i z^{n-i} + \cdots + a_n \in K[z]$$

be the minimal polynomial of x over K.

For some i, suppose that a_i is nonconstant. Then x is algebraic over $\mathbb{C}(a_i)$ of the same degree n, so that $K = \mathbb{C}(a_i)$.

Proof Clear denominators in $f_0(z)$ to get

$$f(x,z) = b_0(x)z^n + b_1(x)z^{n-1} + \cdots + b_i(x)z^{n-i} + \cdots + b_n(x) \in \mathbb{C}[x,z]$$

where the b_i are polynomials without a common factor. Write m for the degree of f in x, so that $m = \max \deg b_i$.

Write $a_i = b_i/b_0 = g(x)/h(x)$, where $g(x)$ and $h(x)$ have no common factor (and obviously, $\deg g(x), h(x) \leq m$). Consider the polynomial

$$g(z) - a_i h(z) = g(z) - \frac{g(x)}{h(x)} h(z) \in K[z].$$

This vanishes on substituting x for z; therefore by definition of the minimal polynomial, it is divisible by $f_0(z)$ in $K[z]$. By Gauss's lemma, also

$$f(x,z) \;\Big|\; h(x)g(z) - g(x)h(z) \in \mathbb{C}[x,z].$$

Now, however, the right-hand side must have degree $= m$ in x, and by symmetry also in z. Hence $n = \deg f$ in z gives $n \leq m$

$$h(x)g(z) - g(x)h(z) = q(z) \cdot f(x,z)$$

with $q(z)$ independent of x. But because g, h are coprime polynomials,

$$g(z) - a_i h(z)$$

doesn't have a nonconstant factor $q(z)$, so by Gauss' lemma again, neither does $h(x)g(z) - g(x)h(z)$, and $q(z) = \text{const}$.

Therefore $f(x,z)$ has degree $m = n$ in x and z. QED

2 Elliptic functions

2.0 Aim

$L \subset \mathbb{C}$ a lattice (see Figure 2.1). An elliptic function for L is defined as a doubly periodic meromorphic function on $\mathbb{C}$. That's the same thing as a

Figure 2.1: Lattice $L = \mathbb{Z}w_1 \oplus \mathbb{Z}w_2$

meromorphic function on the complex torus $\mathbb{C}/L = C$. We ask the question "how many?" in the style of 1.8. We find also that there are enough elliptic functions to embed $C \setminus \{0\}$ into $\mathbb{C}^2$, or C into $\mathbb{C}^2$ union one point at infinity. Thus the torus C is a cubic curve in the complex plane.

2.1 Definitions

A lattice is a discrete subgroup $L \subset \mathbb{C}$.

Theorem 2.2 *It has rank* ≤ 2*, and if* $= 2$ *then* $L = \mathbb{Z}w_1 \oplus \mathbb{Z}w_2$*, with* $w_1, w_2 \in \mathbb{C}$ *linearly independent over* $\mathbb{R}$*.*

Proof (To prove, we have to see that rank ≥ 3, or 2 $\mathbb{R}$-linearly dependent elements contradicts discrete.) □

We usually write $\tau = w_1/w_2$ and assume $\operatorname{Im}\tau > 0$, and say that w_1, w_2 is an oriented basis. Similarity of lattices: $L \sim aL$ where a is a nonzero complex number. (This is Euclidean similarity: scale by $|a|$ and rotate by $\arg a$.) Then up to similarity, $L = \mathbb{Z}\tau \oplus \mathbb{Z}1$ with w in the upper halfplane. Unit cell or fundamental parallelogram of L has vertexes $0, 1, \tau, 1 + \tau$ (see Figure 2.2).

Figure 2.2: Fundamental parallelogram

2.3 Special lattices

1. Real, that is complex conjugate $\overline{L} = L$. There are two solutions: the Rectangular lattice and the Rhombic or Centred Rectangular lattice.

2. Lattice with extra symmetry. There are two, which appear throughout the subject:

   Square $L_i = \mathbb{Z}i \oplus \mathbb{Z}$, having a rotation by $\pi/4$ or complex multiplication by i, and

   Equilateral Triangular $L_w = \mathbb{Z}w \oplus \mathbb{Z}$, where w is the primitive cube root of unity $w = (-1 + \sqrt{-3})/2$. This has 6-fold rotation by $\pi/3$ or complex multiplication by the 6th root of unity $-w^2$.

   Preview: these two special cases correspond to the elliptic curves

   - $y^2 = x^3 + ax$ with symmetry $(x, y) \mapsto (-x, iy)$ of order 4 and
   - $y^2 = x^3 + b$ with symmetry $(x, y) \mapsto (wx, -y)$ of order 6.

   [The topics below not included in this year's course:]

3. Complex multiplication (general imaginary quadratic lattices, having multiplications $L \to L$ with image a sublattice of finite index) and

4. Degeneration of a lattice of rank 2 to a lattice of rank 1, obtained by taking $\lim w \to +i\infty$.

2.4 Sums over lattice

Write $\sum'$ for sum taken over all nonzero $w \in L$. The Eisenstein series are defined by $G_k(L) = \sum' \frac{1}{w^k}$. Proof that this is absolutely convergent for any $k > 2$. Obviously homogeneous of degree k, that is $G_k(aL) = G_k(L)/a^k$ for nonzero $a \in \mathbb{C}$.

2.5 Special values

If $aL = L$ then the homogeneity implies that $G_k(L)$ can only be nonzero if $a^k = 1$. Therefore

- $G_k(L) = 0$ if k is odd (because $L = -L$ for all L);
- $G_k(L_i) = 0$ unless $4 \mid k$ for the square lattice L_i;
- $G_k(L_w) = 0$ unless $6 \mid k$ for the equilateral triangular lattice L_w.

2.6 How many elliptic functions?

Elliptic function is defined as a function $f\colon \mathbb{C} \to \mathbb{C}$ that is meromorphic everywhere, and satisfies $f(z + w) = f(z)$ for all $w \in L$. Zeroes and poles of f are isolated (this is part of the definition of meromorphic), and determined by what happens in the unit cell, therefore f has only finitely many zeros and poles. It is harmless to assume that these do not fall on the boundary lines of the unit cell: if they do, we just "move the goalposts."

2.7 Restrictions on f

NOTATION Suppose f has zeros of order m_i at $z = \alpha_i$ and poles of order n_j at $z = \beta_j$.

If $z = \beta$ is a pole of order n, write out the principal part

$$f = \sum_{-n \le i \le 0} b_i (z - \beta)^i.$$

By Liouville's theorem, f is determined up to an additive constant by its principal part at all its poles. (Because f, g same poles and same principal parts $\Rightarrow f - g$ has no poles, and is holomorphic on whole plane and periodic, therefore bounded. So $f - g = \text{const.}$)

This implies that (if we fix β_j and n_j), elliptic functions with poles only at $z = \beta_j$ of order $\leq n_j$ form a vector space of dimension $\leq 1 + \sum n_j$ (compare 1.8, ii). In fact, Theorem 2.11 gives two different proofs that it has dimension $\leq \sum n_j$.

2.8 Contour integration

Lemma 2.9 *If h is an elliptic function then*

$$\int_\Gamma h\, dz = 0$$

(contour integral around Γ the perimeter of the unit parallelogram).

The point is just that going around the contour Γ, each side is cancelled by its opposite side.

2.10 Restriction on zeros and poles of elliptic functions

Theorem 2.11 (Main restriction) *In the notation of 2.7,*

(I) $\sum$ *residues* $= 0$*;*

(II) $\sum m_i = \sum n_j$ *(so number of zeros = number of poles);*

(III) the difference $\sum m_i\alpha_i - \sum n_j\beta_j$ *is in* L*.*

2.12 Order of an elliptic function

In terms of Theorem 2.11, (II) we define order $f = \sum m_i = \sum n_j$. If f is an elliptic function then f and $f - c$ have exactly the same poles for any $c \in \mathbb{C}$, so that order f is the number of zeros in the unit cell of $f = c$, counted with multiplicities. Picture: an elliptic function represents $\mathbb{C}/L$ as a d-sheeted cover of $\mathbb{P}^1_{\mathbb{C}}$.

Order f = order $(af + b)/(cz + d)$.

2.13

If the zeros of $f - c$ with multiplicities are $\{\alpha_i, m_i\}(c)$ then $\sum m_i\alpha_i$ modulo L is constant: By 2.11, (III), it is equal to the sum of poles $\sum n_j\beta_j$.

2.14

When $f - c$ has some zero α of multiplicity $m \geq 2$, then the function $f(z-\alpha)$ is not a local isomorphism: it maps a disc around α to a disc around c by $(z - \alpha) \mapsto (z - \alpha)^m$ times a unit holomorphic function (see Figure 2.3). We

Figure 2.3: Ramified cover

say that f is ramified at $z = \alpha$ with order m. The set of ramification points (away from the poles) is determined as the set of zeros of the derivative f', and is therefore a finite set modulo L. (At a pole $z = \beta$, we say that f is ramified if the pole has order $n \geq 2$; each pole of order n of f is a pole of order $n + 1$ of f'.)

2.15 Elliptic function of order 2

There is no elliptic function of order 1 (because by 2.11, (I), it would have residue zero at its alleged pole of order 1). An elliptic function f of order 2 has either one pole of order 2, or 2 poles of order 1, and for any c, the two zeros of $f - c$ add to a constant. Thus up to a translation in $\mathbb{C}$, f is an even function, that is, $f(-z) = f(z)$.

2.16 Weierstrass $\wp$ function

We'd like to write $f(z) = \sum \frac{1}{(z-w)^2}$, where the sum runs over all $w \in L$. For $v \in L$, the sum $f(z + v)$ appears to be formally just a permutation of $f(z)$, so the sum seems to describe a function with periodic lattice L. But this argument is nonsense: the series is not absolutely convergent, so permuting the terms is illegal.

Instead, note first that $f_k(z) = \sum \frac{1}{(z-w)^k}$ for $k \geq 3$ is obviously kosher (for any bounded domain $D \subset \mathbb{C}$, the sum is a finite number of terms that may

have poles in D, plus a series that is absolutely convergent and uniformly convergent in D). And f_k is manifestly an elliptic function.

Now define

$$\wp = \frac{1}{z^2} + \sum{}'\Big[\frac{1}{(z-w)^2} - \frac{1}{w^2}\Big], \quad \text{where } \sum{}' \text{ runs over all nonzero } w \in L.$$

This is once again absolutely convergent, and uniformly convergent on any bounded domain (calculate). It's no longer obviously periodic. But clearly $\wp'(z) = -2f_3(z)$ is periodic, so $\wp(z+v) - \wp(z) = c$ is constant. Also, $\wp(z)$ is an even function, and so zero at the halfperiods $w/2$, and we conclude $c = 0$, and $\wp$ is periodic.

2.17 Taylor series of $\wp(z)$ at $z = 0$

$$\wp(z) - 1/z^2 = \sum (2k+1)G_{2k+2}z^{2k} \quad \text{summed over } k \geq 1.$$

here G_k are the Eisenstein series defined in 2.4.

Subtracting the $\sum' \frac{1}{w^2}$ is not well defined (it depends on the order of summation). But it is an amazing trick normalising $\wp$ and giving it nice expansion coefficients around its pole $z = 0$.

2.18 Differential equation satisfied by $\wp$

$$\wp'^2 = 4\wp^3 - g_2\wp - g_3,$$

where $g_2 = 60G_4$ and $g_3 = 140G_6$. Thus $(\wp, \wp')$ is a map $C := \mathbb{C}/L \to \mathbb{C}^2 \cup \{\infty\}$, and the image is the plane cubic curve $y^2 = 4x^3 - g_2x - g_3$.

2.19 $\wp$ is even

$\wp$ is an even function by construction. It is an elliptic function of order 2 (see 2.15). It maps $\mathbb{C}/L$ 2-to-1 to $\mathbb{P}^1_{\mathbb{C}}$, identifying z and $-z$. It has a pole of order 2 at 0, and 3 other ramification points of order 2 at the 3 halfperiods. These are the zeros of $\wp'$ (see 2.12). The zeros of $\wp$ are not distinguished. (The normalisation was to do with killing the constant term in the Taylor series around $z = 0$, or the x^2 term in the equation. It arranges for the 3 roots of $4x^3 - g_2x - g_3$ to add to zero; they are the values of $\wp$ at the halfperiods.)

2.20 All elliptic functions

Theorem 2.21 *All elliptic functions for L form a field, equal to $\mathbb{C}(\wp, \wp') = \mathbb{C}(X)[Y]/(Y^2 - 4X^3 + g_2X + g_3)$. The even functions are the subfield $\mathbb{C}(\wp)$.*

2.22 Proof that $(\wp, \wp')$ embeds $\mathbb{C}/L$ isomorphically

At each point of $\mathbb{C}/L$ a local parameter is given by one of $\wp - c$ (at a general point) or $\wp'$ (at a halfperiod) or $\wp'/\wp$ (at 0).

3 Geometry of plane cubics

Aim Derivation of the normal form $C : y^2 = x^3 + ax + b$ starting from a general cubic curve over K and a point. Tate's formulas to deal with the case of char 2 or 3. The points $C(K)$ form a group, where the group law is defined geometrically, and can be written out as explicit rational functions. Relation of the group law to the function theory of C.

3.1 Fields

K is a field about which we want to assume as little as possible at first. The characteristic p of K is determined as the smallest natural number (if any) such that $1 + 1 + \cdots + 1 = 0$ in K (p summands). Every field K either has characteristic 0 and contains a copy of the rational number field $\mathbb{Q}$, or has characteristic a prime p and contains a copy of the field with p elements, $\mathbb{F}_p = \mathbb{Z}/(p)$. K^* is the multiplicative group of nonzero elements of K.

The projective plane is defined by $\mathbb{P}^2_K = (K^3 \setminus \{0\})/K^*$, that is, it is the set of equivalence classes of nonzero $(x, y, z) \in K^3$ modulo nonzero scalar multiple, or equivalently, the set of ratios $(x : y : z)$, or 1-dimensional vector subspaces of K^3. Each equivalence class with $z \neq 0$ has a preferred representative $(x/z, y/z, 1)$, so $\mathbb{P}^2_K$ contains the (x, y) plane K^2 as a big subset. The complement is the set of ratios $(x : y : 0)$, the line at infinity.

For most purposes, you don't need to know too much about $\mathbb{P}^2$, because we work with something like $C : (y^2 = x^3 + ax + b) \subset K^2$, and that has only one point $(0 : 1 : 0)$ at infinity, where $x, y \to \infty$, but $x/y \to 0$.

3.2 Homogeneous form in 2 variables

A *form* is a homogeneous polynomial. The space of forms of degree d in 2 variables u, v is based by $u^d, u^{d-1}v, \ldots, uv^{d-1}, v^d$, that is,

$$\big\{u^i v^j \text{ with } i, j \geq 0 \text{ and } i + j = d\big\}.$$

$F(u, v) = a_0u^d + a_1u^{d-1}v + \cdots + a_dv^d$. You can pass from forms of degree d in u, v to polynomials in u of degree $\leq d$ by setting $v = 1$:

$$F(u, v,) \mapsto f(u) = F(u, 1) = a_0u^d + a_1u^{d-1} + \cdots + a_1u + a_d$$

and back by

$$f(u) \mapsto F(u, v) = v^d f(u/v).$$

Obviously F is identically divisible by v^i if and only if the first i coefficients vanish $a_0 = a_1 = \cdots = a_{i-1}$, and then f has degree $\leq d - i$. In other words, if $\deg f < d$, we can view it as having a zero of multiplicity $d - \deg f$ at infinity of $\mathbb{P}^1$. We say that $F(u, v)$ *splits* as a product of linear forms if it can be written

$$F(u, v) = \prod L_i(u, v)^{m_i}$$

with $L_i(u, v)$ linear forms, not proportional, and $\sum m_i = d$. Over an algebraically closed field, this always happens. All of this is just a trivial device to allow us to say that F has exactly d zeros, counted with multiplicities and including points at infinity.

3.3 Line intersect curve in $\mathbb{P}^2_K$

A plane curve $C \subset \mathbb{P}^2_K$ of degree d is defined by a form $F_d(x, y, z)$. To say that F_d is *homogeneous* means that $F_d(ax, ay, az) = a^d F(x, y, z)$, so that the condition $F_d(x, y, z) = 0$ depends only on the ratio $(x : y : z) \in \mathbb{P}^2_K$. If L is a line, we can count the intersection points of C with L (see Figure 3.1) by factoring F restricted to L. We just parametrise L by (u, v), and write

Figure 3.1: Line meets world

$F_{|L} = F_d(u, v)$. If all the roots are in K, we get $\prod L_i(u, v)^{m_i}$ as in 3.2. Most lines can be written $z = ax + by$, so the parametrisation is just $x = u, y = v, z = au + bv$, and $F_{|L}$ is the form obtained by substituting these into F.

If $d = 3$ and $C \subset \mathbb{P}^2_K$ is the cubic defined by $F = 0$, then $C \cap L = 3$ points. (See Figure 3.2) If two points $P, Q \in C$ are given, they determine a line PQ, and hence the third point R.

Definition 3.4 *A cubic $C \subset \mathbb{P}^2_K$ given by $F = 0$ is* nonsingular *if for every point $P \in C$ there is a unique line L so that $F_{|L}$ has a double root at P. This*

Figure 3.2: Get to the point!

is equivalent to

$$\left(\frac{\partial f}{\partial x}(P), \frac{\partial f}{\partial y}(P), \frac{\partial f}{\partial z}(P)\right) \neq (0,0,0).$$

Assume C is nonsingular. Then for any $P, Q \in C$ there is a third point $R = P * Q \in C$ so that $C \cap L = \{P, Q, R\}$. (See Figure 3.3: what happens if $P = Q$ or $P = Q = R$, tangent line and flex line etc.) Note that

- This map is well defined (without exception)
- If $P, Q \in C(K)$ (that is, their coordinates are in the field K) then also $R \in C(K)$.

Figure 3.3: Tangent and flexes

3.5 Group law

C nonsingular cubic, and $O \in C$ a given point. In 3.3 we had

$$(P, Q) \mapsto P * Q = R = \text{3rd point of intersection of line } PQ.$$

The group law is obtained by reflecting $P * Q$ in O:

$$P, Q \mapsto P * Q \mapsto O * (P * Q) = (P + Q).$$

In other words, first join P, Q by a straight line and take R to be the 3rd point of intersection. Then join O, R to give 3rd point $P + Q$. You verify 0 and inverse. (Abelian is obvious.)

3.6 Cubics through 8 points

Lemma 3.7 *C_1, C_2 are cubics, and suppose $C_1 \cap C_2 = \{P_1, \ldots, P_9\}$. Then any cubic D through $P_1, \ldots, P_8$ also passes through P_9.*

Proof By plane geometry; see for example [4], Chaps. 1–2. Compare: suppose C_1 is nonsingular. Then D/C_2 is a rational function of C_1 having a unique possible pole P_9. □

3.8 Proof of associativity "in general"

To prove $(P + Q) + R = P + (Q + R)$, write down 2 triples of lines $L_1L_2L_3$ and $M_1M_2M_3$ whose 9 points of intersection with C are

$$P, Q, P * Q; \quad P + Q, R, (P + Q) * R; \quad 0, Q + R, Q * R$$

and

$$O, P * Q, P + Q; \quad P, Q + R, P * (Q + R); \quad Q, R, Q * R$$

By Lemma 3.7, $(P + Q) * R = P * (Q + R)$, hence associative. The argument as given assumes that the intersection points are all distinct. I don't finish it.

3.9 Divisors, linear equivalence and group law

Define

$$\operatorname{Div} C = \text{free Abelian group on } C = \left\{ \sum n_i P_i \;\middle|\; P_i \in C, n_i \in \mathbb{Z} \right\}$$
$$\operatorname{Div}^0 C = \text{divisors of degree } 0 = \left\{ \text{ditto} \;\middle|\; \sum n_i = 0 \right\}$$

(finite sums). For a line L, write $\operatorname{div} L = L \cap C = P + Q + R$ as in 3.3. Define *linear equivalence* on $\operatorname{Div} C$ by saying $\operatorname{div} L - \operatorname{div} L' \overset{\text{lin}}{\sim} 0$ for any two lines. In other words, let $\operatorname{Div}^{\text{lin}}$ be the subgroup of $\operatorname{Div}^0 C$ generated by all the differences $\operatorname{div} L - \operatorname{div} L'$ for all lines L, L' of $\mathbb{P}^2_K$ and say $D_1 - D_2 \overset{\text{lin}}{\sim} 0$ or $D_1 \overset{\text{lin}}{\sim} D_2$ if $D_1 - D_2 \in \operatorname{Div}^{\text{lin}}$.

Finally, define $C^{(0)} = (\operatorname{Div}^0 C)/D^{\text{lin}}$, the group of divisor classes of degree 0 modulo linear equivalence. The point is that all of these are manifestly groups.

Next, if we fix $O \in C$, we get a map $C \to C^{(0)}$ by $P \mapsto P - O$. By the construction of 3.5, this map is surjective.

PROOF If $D = \sum n_i P_i$ has degree 0, then $D + O$ has degree 1. Whenever D has a negative term, say $P_1 + O - P_2$, I can find lines L, L' with $\operatorname{div} L = P_1 + O + R$ with $R = P_1 * O$, and $\operatorname{div} L' = R + P_2 + Q$ with $Q = P_2 * R$, so $P_1 + O - P_2 \overset{\text{lin}}{\sim} Q$, which reduces the negative terms in D. By induction, $D + O \overset{\text{lin}}{\sim} Q$. □

A restatement of associative is to say that $C \mapsto C^{(0)}$ is injective. If not, there would be $P \neq Q$ is C with $P \overset{\text{lin}}{\sim} Q$. Then there would be a rational function (with numerator and denominator a product of lines) with zero P and pole Q. I break off the proof at this point (as before, this implies that C would be isomorphic to $\mathbb{P}^1$, which is not the case for a nonsingular cubic, but I would need a bit more algebraic geometry to conclude this honestly).

REMARK Addition in the group law corresponds to zeros and poles of rational functions: given O, we have $P + Q = R$ is the group law if and only if $P + Q \overset{\text{lin}}{\sim} O + R$, which happens if and only if there is a rational function (ratio of two lines) with zeros $P+Q$ and poles $O+R$. Thus the group law on a plane cubic is determined by adding zeros and poles of rational functions. Compare Theorem 2.11, (III), where addition in $C = \mathbb{C}/L$ is determined by adding zeros and poles of elliptic functions. Since C/L can be embedded into $\mathbb{P}^2_{\mathbb{C}}$ by elliptic functions, the two group laws coincide.

3.10 Normal form

In characteristic $\neq 2, 3$, the normal form of an elliptic curve is

$$C : y^2 = 4x^3 - g_2 x - g_3 \qquad \text{or} \qquad y^2 = x^3 + ax + b. \tag{$*$}$$

For C in form $(*)$, the point at infinity $O = (0, 1, 0)$ is a flex. Conversely, if C has a flex (with coordinates in K), we can make a linear change of coordinates to put the equation of C in the form $(*)$.

3.11 Finding a flex

There are 2 quite different approaches: over an algebraically closed field K with $\operatorname{char} K \neq 2, 3$, we can find a flex by using the Hessian determinant $H = \det |\partial^2 F / \partial x_i \partial x_j|$. This is a homogeneous cubic and $H \cap C \neq \emptyset$ if K is algebraically closed (in fact it always consists of 9 distinct points). In char $\neq 2, 3$, a point of $H \cap C$ is a flex of C. This gives the existence of a flex.

On the other hand, if $O \in C$ is given, we can re-embed C into $\mathbb{P}^2_K$ to make O into a flex. In fact, choose linear coordinates so that $O = (1, 0, 0)$, $(Z = 0)$ is the tangent line to C at O and meets C at P, $(X = 0)$ is the tangent line at P, and $(Y = 0)$ is any line through O. Set $x = X/Z$ and $y = Y/Z$. Then the equation of C is

$$xy^2 + (ax + b)y = cx^2 + dx + e.$$

Now multiply though by x and write out the right hand side in terms of $\eta = xy$, giving

$$\eta^2 + (ax + b)\eta = cx^3 + dx^2 + ex.$$

In other words, the coordinate change $(x, y) \mapsto (x, \eta = xy)$ takes C into a new cubic curve that has $O = (0, 1, 0)$ as a flex. Note however that this is not a linear map of the ambient space $\mathbb{P}^2_K$.

3.12 The discriminant

The discriminant of $y^2 = x^3 + ax + b$ is $-(4a^3 + 27b^2)$. You can get it as the resultant of $f = x^3 + ax + b$ and $f' = 3x^2 + a$. In fact f and f' have a common root if and only $xf, f, x^2 f', xf', f'$ are linearly dependent in the space of quartics, which gives the 5×5 determinant

$$\begin{vmatrix} 1 & 0 & a & b & 0 \\ 0 & 1 & 0 & a & b \\ 3 & 0 & a & 0 & 0 \\ 0 & 3 & 0 & a & 0 \\ 0 & 0 & 3 & 0 & a \end{vmatrix}$$

Or you can get it by setting $x^3 + ax + b = (x - e_1)(x - e_2)(x - e_3)$ and calculating $[(e_1 - e_2)(e_2 - e_3)(e_3 - e_1)]^2$ by the rules for symmetric functions.

3.13 Appendix: Tate's formulas

In char 2 or 3 the Weierstrass normal form of 3.10 cannot be used, and we use Tate's formulas instead. More generally, suppose that E is defined over a ring R in which 2 and 3 are maybe not invertible (such as $\mathbb{Z}$). Tate's formulas are just a way of carefully keeping track of the powers of 2 and 3 involved in changing to the standard normal form. We start from

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{3.1}$$

with $a_i \in R$. Multiply this by 4 and rewrite in terms of $y' = 2y + a_1x + a_3$ in order to complete the square:

$$(y')^2 = 4x^3 + b_2x^2 + 2b_4x + b_6, \tag{3.2}$$

where

$$b_2 = 4a_2 + a_1^2, \quad b_4 = a_1a_3 + 2a_4, \quad b_6 = 4a_6 + a_3^2. \tag{3.3}$$

Now if you do $b_2b_6 - b_4^2$, the terms in $a_1^2a_3^2$ cancel, giving

$$4b_8 = b_2b_6 - b_4^2 = 4b_8 \quad \text{where} \quad b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2. \tag{3.4}$$

Now to "complete the cube" to get rid of b_2 in (3.2), we should write it in terms of $x + \frac{1}{12}b_2$. First multiply by 2^43^6 and write it in terms of $x'' = 36x + 3b_2$, and $y'' = 108y'$. We get:

$$E : (y'')^2 = (x'')^3 - 27c_4x'' - 54c_6, \tag{3.5}$$

where

$$c_4 = b_2^2 - 24b_4, \quad c_6 = -b_2^3 + 36b_2b_4 - 216b_6. \tag{3.6}$$

Now if you do $c_4^3 - c_6^2$, the terms in b_2^6 cancel, giving

$$c_4^3 - c_6^2 = 1728\Delta \quad \text{where} \quad \Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6. \tag{3.7}$$

3.14 Addition and duplication laws

In Tate form, the addition law takes the form (set $P_i = (x_i, y_i)$ for $i = 1, 2$ or $\emptyset$)

$$x(P_1 + P_2) = m^2 + a_1 m - a_2 - x_1 - x_2, \quad \text{where} \quad m = \frac{y_2 - y_1}{x_2 - x_1}$$

(that assumes $x_1 \neq x_2$) and

$$x(2P) = \frac{x^4 - b_4 x^2 - 2b_6 x - b_8}{4x^3 + b_2 x^2 + 2b_4 x + b_6}.$$

3.15 Discriminant in Tate form

If E is a cubic in Tate form, reducing it to 3.10 by completing the square and cube simplifies some things, but introduces powers of 2 and 3 into the discriminant. For example, $y^2 + y = x^3 - x^2$ has Tate discriminant -11, and is nonsingular over any field of characteristic $\neq 11$. But to get it in the form 3.10, we have to do

$$v = 2^3 \cdot 3^3 (y + \frac{1}{2}), \quad u = 2^2 \cdot 3^2 (x - \frac{1}{3}), \quad v^2 = u^3 - 2^4 \cdot 3^3 u + 2^4 \cdot 3^3 \cdot 19,$$

giving $\Delta = 3^{12} \cdot 2^8 \cdot (-11)$.

3.16 Facts:

1. (3.1) defines a nonsingular curve if and only if $\Delta \neq 0$. We already knew this if $\operatorname{char} k \neq 2, 3$. The point of the whole rigmarole is to get the same result in characteristic 2 and 3.

   Set $j = \frac{c_4^3}{\Delta}$. Then over any algebraically closed field K:

2. There exists a curve (3.1) with any given $j \in K$.

3. Two curves with equations (3.1) are isomorphic if and only if they have the same j.

4 Mordell–Weil theorem

4.1 Idea of descent

Fermat's method of "infinite descent" can sometimes be used to prove that a Diophantine problem has *no* nontrivial solution. The idea is: suppose a solution exists; after some choices, show that the solution comes from a smaller solution. This sometimes gives a contradiction; but here we use it together with the idea of height to show that all solutions can be generated from a finite subset.

4.2 Example from [UAG], Ex. 2.12

4.3 Example: case $n = 4$ of Fermat's last theorem

See [Knapp], p. 81, [Silverman, Friendly introduction], Chaps. 27 and 32 or Question D.2 of assignment.

Fermat: $u^4 + v^4 = w^2$ has no nontrivial solutions; apply the formula for the solution of Pythagorean triples successively. We get (with finitely many other choices) $u = r^4 - s^4$, $v = 2rst$ and $w = r^2 + s^2$, where $r^4 + s^4 = t^2$. Here any nontrivial solution comes from a strictly smaller nontrivial solution, which is a contradiction. These formulas are actually the duplication formula on the elliptic curve $y^2 = x^3 - 4x$ in mild disguise.

4.4 Division by 2: the split case

We work with the split case $C : y^2 = (x - e_1)(x - e_2)(x - e_3)$ defined over K.

Criterion 4.5 *A point $P_2 = (x_2, y_2)$ in $C(K)$ with $y_2 \neq 0$ is in $2C(K)$ if and only if each of*

$$x_2 - e_1, \quad x_2 - e_2, \quad x_2 - e_3 \quad \textit{is a perfect square in } K.$$

The question is to find a line $y = mx + d$ through P_2 touching C at some point P_1 (see Figure 4.1). That is: what values m, d give rise to identity of cubic polynomials

$$(x - e_1)(x - e_2)(x - e_3) - (mx + d)^2 \equiv (x - x_1)^2(x - x_2)?$$

Figure 4.1: A touching scene

By setting $x = e_i$ in this, one sees that this is equivalent to

$$x_2 - e_i = f_i^2 \quad \text{with} \quad f_i = \frac{me_i + d}{e_i - x_1},$$

where x_1, m, d are given in terms of the square roots f_i by

$$\begin{aligned} m &= f_1 + f_2 + f_3, \\ x_1 &= f_1 f_2 + f_1 f_3 + f_2 f_3 + x_2 \\ d &= -f_1 f_2 f_3 - x_2(f_1 + f_2 + f_3). \end{aligned}$$

Lemma 4.6 *The map $C(K)/2C(K) \to K^*/(K^*)^2$ that takes $P = (x, y)$ with $y \neq 0$ into $x - e_1$ and $(e_1, 0)$ into $(e_1 - e_2)(e_1 - e_3)$ modulo squares is group homormorphism. Similarly for e_2, e_3.*

Interpretation:

$$C(K)/2C(K) \to 2 \text{ copies of } [K^*/(K^*)^2]$$

is injective.

4.7 Perfect square at each prime

Criterion 4.5 works for any field K of char $\neq 2$. We want to go further, to get $C(\mathbb{Q})/2C(\mathbb{Q})$ finite. This uses a special property of the rational field $\mathbb{Q}$, derived from the UFD property of $\mathbb{Z}$. Namely, by unique factorisation in $\mathbb{Z}$, any nonzero element $q \in \mathbb{Q}$ is plus or minus a product of prime powers:

$$q = \pm \prod p^{a_p} \quad \text{with} \quad a_p \in \mathbb{Z} \quad \text{(finite product)};$$

and q is a perfect square if and only if the sign is $+1$ and each a_p is even. That is, the group of rationals modulo squares

$$\mathbb{Q}^*/(\mathbb{Q}^*)^2 \cong \prod \mathbb{Z}/2 \quad \text{taken over sign, and each } p.$$

Proposition 4.8 *$C(\mathbb{Q})/2C(\mathbb{Q})$ is a finite group.*

PROOF For $P = (x, y)$ in $C(\mathbb{Q})$, to determine if P in $2C(\mathbb{Q})$, we need only ask $x - e_i > 0$ and is exactly divisible by an even power of p for each p among the finite set of factors of $e_1 - e_2, e_2 - e_3$ and $e_3 - e_1$. (Argue separately on numerator and denominator: the 3 elements $x - e_i$ have the same powers of p in denominator, which must be even. Any prime factor not dividing $e_1 - e_2, e_2 - e_3$ and $e_3 - e_1$ can divide at most one numerator, and again, it must divide to an even power.) Together with Crieterion 4.5 and Lemma 4.6, this proves the result. □

4.9 Reiteration of the idea of descent

Fermat's construction in 4.3 is 2-division on the elliptic curve $y^2 = x^3 - 4x$. The logic there was: start from any solution, make a small number of choices (e.g., order or signs of x, y), apply the formula for Pythagorean squares, after a couple of steps, find a smaller solution, hence a contradiction. This logic also forms the basis for the proof of Mordell–Weil:

1. start from any solution $P \in \mathbb{C}(\mathbb{Q})$;
2. up to a finite number of obstructions given by $C(\mathbb{Q})/2C(\mathbb{Q})$, we can divide P by 2 in the group law;
3. it is intuitively clear that this makes P "smaller";
4. after a finite number of steps, P is "small", and we can find all small solutions explicitly.

4.10 Height

To define the height of $x \in \mathbb{Q}$, write it as a fraction in reduced form $x = m/n$ and set $H(x) = \max(|m|, |n|)$. The point is that there are only finitely many

x of bounded height (because $H(x) \le K$ gives $m, n = -K, \dots, K$ so at most $(2K+1)^2$ possibilities).

Is H any kind of homomorphism? Not very:

$$H(x_1x_2) \le H(x_1)H(x_2),$$

because $x_1x_2 = m_1m_2/n_1n_2$, but it could be much less if there is cancellation. Similarly

$$H(x_1 + x_2) \le 2H(x_1) \cdot H(x_2) \quad \text{and} \quad H(x)^2 = (H(x))^2.$$

It is traditional to set $h(x) = \log H(x)$, which translates relations that are multiplicative in nature to additive. For example, the above become

$$h(x_1x_2) \le h(x_1) + h(x_2), \quad h(x_1 + x_2) \le \log 2 + h(x_1) + h(x_2)$$
$$\text{and} \quad h(x)^2 = 2h(x).$$

4.11 Plan of proof of the Mordell–Weil theorem

$C : y^2 = (x - e_1)(x - e_2)(x - e_3)$ with $e_i \in \mathbb{Z}$. Height H and h as in 4.10. For $P = (x, y) \in C(\mathbb{Q})$, we set $H(P) = H(x)$. The MW theorem is a formal consequence of Proposition 4.8 and the following 3 statements:

Lemma 4.12 *(i) $h(x) < K$ gives only finitely many possibilities for x. (OK)*

(ii) Fix $P_0 \in C(\mathbb{Q})$. Then there exists constant k_0 such that

$$h(P + P_0) \le 3h(P) + k_0 \text{ for all } P \in C(\mathbb{Q}).$$

In other words, $P + P_0$ is a "quadratic" function of P.

(iii) There exists a constant k such that

$$h(2P) \ge 4h(P) - k \quad \text{for all } P \in C(\mathbb{Q}).$$

In other words, $2P$ is a quartic function of P and *there is not too much cancellation.*

Assume these for the moment and prove MW. First pick a finite set $P_1, \dots, P_n$ in $C(\mathbb{Q})$ that covers every coset of $2C(\mathbb{Q})$. Set k_i as in Lemma 4.12, (ii) for P_i and $k' = \max k_i$. Then $h(P + P_i) \le 3h(P) + k'$. Choose k as in Lemma 4.12, (iii), and by Lemma 4.12, (i), let $\{Q_1, \dots, Q_m\}$ be the finite set of Q with $h(Q) \le k + k'$. It is easy to prove that $\{P_1, \dots, P_n, Q_1, \dots, Q_m\}$ generate $C(\mathbb{Q})$. Because if $h(P) > k + k'$ then $P + P_i$ is 2-divisible, so $= 2P'$, and you check that $h(P') < \frac{3}{4}h(P)$, etc.

4.13 Proof of Lemma 4.12, (iii)

Figure 4.2: Off on a tangent

Duplication formula: The tangent line at the point $P_1 = (x_1, y_1)$ has slope $m = \frac{3x_1^2+a}{2y_1}$ (see Figure 4.2); we can assume that $y_1 \neq 0$. The construction of $2P$ gives

$$(x - x_1)^2(x - x_2) \equiv (mx + c)^2 - x^3 - ax - b,$$

(identity of cubic polynomials) and from the coefficient of x^2 we get

$$x_2 = \left[\frac{3x_1^2 + a}{2y_1}\right]^2 - 2x_1 = \frac{x^4 - 2ax^2 - 8bx + a^2}{(4(x^3 + ax + b)}.$$

From this, if $x_1 = p/q$ (reduced), we get $x_2 = F(p, q)/4G(p, q)$ where

$$\begin{aligned} F(p, q) &= p^4 - 2ap^2q^2 - 8bpq^3 + a^2q^4, \\ G(p, q) &= q(p^3 + apq^2 + bq^3). \end{aligned}$$

Then

$$RF + SG = 4dq^6 \quad \text{and similarly} \quad R'F + S'G = 4dp^6 \tag{4.1}$$

for some polynomials R, S, R', S' of degree 2 in p, q (see Assessment D.3). This implies that any common factor of $F(p, q), G(p, q)$ divides $4d$, so in fact not much cancellation happens. But since R, S, R', S' are of degree 2, it follows from (4.1) that

$$\max(F(p, q), G(p, q)) \geq \text{const} \max(p^4, q^4).$$

This proves Lemma 4.12, (iii). Part (ii) is easier, and this completes the proof of MW in the split case $y^2 = (x - e_1)(x - e_2)(x - e_3)$.

4.14 Proof in nonsplit case

— this section is not examinable —

We used the assumption that $x^3 + ax + b$ splits as $(x - e_1)(x - e_2)(x - e_3)$ essentially in the proof that $C(\mathbb{Q})/2C(\mathbb{Q})$ is finite. If $x^3 + ax + b$ is not split over $\mathbb{Q}$, there is a finite extension field $\mathbb{Q} \subset K$ over which it splits with $e_i \in K$. The idea is to replace $\mathbb{Q}$ by K. This is a reduction argument, the logic of which is:

(a) construct the extension $\mathbb{Q} \subset K$ as the splitting field,

(b) prove whatever we need about C/K, in this case that $C(K)/C(2K)$ is finite,

(c) prove that the result over K implies that over $\mathbb{Q}$.

Here (a) is already done. (b) involves algebraic number theory: we can set up a ring $A \subset K$, which is like the ring of integers with a finite number of primes made invertible, such that A is a UFD, whose units we control, and $C(K)/2C(K)$ involves only the question of the exponents of finitely many primes. Nothing hard here, but it needs 10 lectures in algebraic number theory to make sense.

For (c), we need to prove that $C(\mathbb{Q})/2C(\mathbb{Q}) \to C(K)/2C(K)$ has finite kernel. We need a couple of lectures' worth of Galois theory, specifically, the theory of quartic equations. The Galois group of $K/\mathbb{Q}$ is the subgroup of permutations of e_1, e_2, e_3 that extend to symmetries (field automorphisms) of K.

Proposition 4.15 *There is an injective map*

$$\ker[C(\mathbb{Q})/2C(\mathbb{Q}) \to C(K)/2C(K)] \hookrightarrow \mathrm{Maps}[\mathrm{Gal}(K/Q) \to 2\text{-}\textit{torsion of}\, C(K)]$$

Since $\mathrm{Gal}(K/\mathbb{Q})$ is a subgroup of the symmetric group S_3, the right-hand side is a set with at most 4^6 elements, and this does (c) for us.

Sketch proof Start from an element $P \in C(\mathbb{Q})$ whose 2-division we want to study. The equation $2P_1 = P$ has 4 solutions in some extension field of $\mathbb{Q}$. First ask if $P = 2P_1$ for $P_1 \in C(2K)$. Over K, finding P_1 is a Galois problem with group $\mathbb{Z}/2 \oplus \mathbb{Z}/2 =$ 2-torsion of $C(K)$: if P_1 is any solution then so is $P_1 + Q$ for any 2-torsion point $Q \in C(K)$. Over $\mathbb{Q}$, however, it is a general quartic problem. If $P = 2P_1$ with $P_1 \in C(K)$, and $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$ is a symmetry of $K/\mathbb{Q}$ then $\sigma(P_1)$ is some other solution to $P = 2\sigma(P_1)$ so $P_1 - \sigma(P_1) \in \mathbb{Z}/2 \oplus \mathbb{Z}/2 =$ 2-torsion of $C(K)$. If $\sigma(P_1) = P_1$ for all $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$ then $P_1 \in C(\mathbb{Q})$.

Remark The map in the proposition is not a group homomorphism, but a 1-cocycle, parametrising extension groups of $\mathrm{Gal}(K/Q)$ by $\mathbb{Z}/2 \oplus \mathbb{Z}/2$. Recall that the Galois theory of the quartic depends on the normal subgroup

$$\mathbb{Z}/2 \oplus \mathbb{Z}/2 = \ker[S_4 \to S_3] \lhd S_4.$$

Everything here could be made explicit and elementary theory of equations, or could be handled by a simple appeal to Galois cohomology. A simple analogy is the Galois extension corresponding to an irreducible equation $x^r = a$. If the roots of unity are present, this is a problem with Galois group $\mathbb{Z}/r$; if not, its Galois group is an extension of $\mathbb{Z}/r$ by $(\mathbb{Z}/r)^*$.

4.16 Torsion subgroup

Here $C : y^2 = x^3 + ax + b$ with $a, b \in \mathbb{Z}$, $4a^3 + 27b^2 = \Delta \neq 0$.

Theorem 4.17 (Lutz–Nagell) *Suppose $P = (x, y) \in C(\mathbb{Q})$ is of finite order in the group law, that is, $nP = 0$. Then*

(i) x, y in $\mathbb{Z}$

(ii) $y = 0$ or y divides Δ.

The proof of (i) is devious (and not examinable): for every prime p, we show that x, y has no p in denominator. It is clear that if p appears in either the denominator of x or y then it appears to power $2m$ in that of x and $3m$ in that of y, and a calculation shows that pP has bigger denominator. If P is torsion, this is a contradiction. (The business about powers of p in the denominator is a formal analog of $x = \wp$ and $y = \wp'$ having poles of order 2 and 3. The proof amounts to considering P in a neighbourhood of the point at infinity, and treating it p-adically.)

Proof of (i) $\Longrightarrow$ (ii) If P is of finite order then so is $P_2 = 2P = (x_2, y_2)$. But by the usual duplication rule,

$$x_2 = \frac{(3x^2 + a)^2}{4y} - 2x$$

which implies that

$$y \mid 3x^2 + a, \text{ and } y \mid x^3 + ax + b.$$

Therefore y divides Δ. □

The Lutz–Nagell theorem gives a simple algorithmic way of determining the torsion subgroup of $C(\mathbb{Q})$: just take $y = 0$ or any divisor of Δ and ask for x (among the divisors of $b - y^2$). You can do that at once by a brute force computer program. In fact y^2 divides Δ, and the statement already holds without assuming the whole of Weierstrass normal form. Moreover, there are very sophisticated computer routines around that calculate anything in this chapter.

5 Modular forms and modular elliptic curves

Taniyama–Shimura–Weil and Fermat's last theorem

Theorem 5.1 *Every elliptic curve over $\mathbb{Q}$ is modular.*

This was a conjecture of Taniyama, Shimura 1950s and Weil 1960s, proved by Wiles, Taylor, Diamond and co. in 1990s.

What is modular? What good does it do? How can you prove the theorem? This chapter discusses what the statement is about. The material here is just a colloquial presentation, and the answers to the above questions may not be wholly satisfying. It would take the content of at least 4 graduate courses to do justice to this material. It will take most of the chapter just to say what a modular elliptic curve is. Roughly it means two things:

1. In complex analysis, C comes from modular forms, that is, functions on the upper halfspace $\mathcal{H}$ with special symmetry.

2. In arithmetic, for each prime p, consider the equation of C as a congruence modulo p, and count the number of solutions, that is, the number $\#(C(\mathbb{F}_p))$ of points of the elliptic curve modulo p with coordinates in the finite field $\mathbb{F}_p = \mathbb{Z}/p$. Modular is the statement that the totality of all $\#(C(\mathbb{F}_p))$ satisfy "lots of crazy relations".

Theorem 5.2 (Fermat's last theorem) *$a^n + b^n = c^n$ has no integer solution with $abc \neq 0$ for $n \geq 3$.*

The cases $n = 3, 4$ are known. It's enough to do primes $p \geq 5$. The idea is to work by contradiction: suppose a, b, c is a nontrivial solution of $a^p + b^p = c^p$. Write down the Frey elliptic curve

$$y^2 = x(x - a^p)(x + b^p). \tag{5.1}$$

This has discriminant $\Delta = 2^{-8} \cdot (abc)^{2p}$. But it has conductor (discussed below) $N = \prod q$ the product of prime factors of a, b, c (with power 1). Serre and Ribet proved that (5.1) is not modular (discussed below). On the other hand, Wiles and co. proved that every elliptic curve over $\mathbb{Q}$ is modular. This is a contradiction. The only way out is that no such a, b, c exists. □

5.3 The upper halfplane $\mathcal{H}$ and the action of $\mathrm{SL}(2,\mathbb{Z})$

Define

$$\mathcal{H} = \big\{\tau = x + iy \;\big|\; \operatorname{Im}\tau = y > 0\big\}.$$

$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C})$ acts on $\mathbb{P}^1_{\mathbb{C}}$ by the fractional linear transformation $z \mapsto \frac{az+b}{cz+d}$. If $g \in \mathrm{SL}(2,\mathbb{R})$, it obviously preserves the real line $\mathbb{P}^1_{\mathbb{R}}$, and it is easy to calculate imaginary parts and see that g takes $\mathcal{H}$ to itself. [In fact $\mathrm{SL}(2,\mathbb{R})$ is the group of all holomorphic automorphisms of $\mathcal{H}$, or the group of all isometries of $\mathcal{H}$ with its hyperbolic metric.] The subgroup $\mathrm{SL}(2,\mathbb{Z})$ acts as a discrete group on $\mathcal{H}$, and has the fundamental domain

$$D = \big\{\tau \;\big|\; |\operatorname{Re}\tau| \le 1/2,\ |\tau| \ge 1\big\}.$$

See Figure 5.1. Here the matrix $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is the translation $\tau \mapsto \tau + 1$,

Figure 5.1: Fundamental Domain

that glues the two sides of D, and $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is the inversion $\tau \mapsto -1/\tau$, glueing the boundary halfarc from complex w to i to the other halfarc from $-w^2$ to i (see Figure 5.1). *Fundamental domain* means that anything in $\mathcal{H}$ is taken into D by some element of the group, unique except for the identifications along the boundary. See Assessment E.1 for the proof that S, T generate $\mathrm{SL}(2,\mathbb{Z})$ and D is the fundamental domain. If $\Gamma \in \mathrm{SL}(2,\mathbb{Z})$ is a subgroup of finite index and $g_1, \ldots, g_k$ coset representatives of Γ then $g_1(D) \cup \cdots \cup g_k(D)$ is a fundamental domain of Γ.

Definition 5.4 *A* cusp *for* Γ *is a point of the closure of a fundamental domain at* $+i\infty$ *or at a point of the real line, necessarily rational.*

Figure 5.2: Glued boundaries

We usually identify cusps for Γ taken into one another by the action of Γ, that is, in the same orbit of Γ acting on $\mathbb{P}^1_{\mathbb{Q}} \subset \mathbb{P}^1_{\mathbb{R}}$. For $\mathrm{SL}(2,\mathbb{Z})$, every point of $\mathbb{P}^1_{\mathbb{Q}}$ (that is, rational points on the real line, plus the single point $+i\infty$ at infinity) is a cusp, but they only form one orbit under $\mathrm{SL}(2,\mathbb{Z})$.

5.5 Definition of modular form for $\mathrm{SL}(2,\mathbb{Z})$ or for Γ

The definition has three parts:

0. holomorphic function on $\mathcal{H}$ (meromorphic would also make sense),
1. symmetry under Γ, and
2. behaviour at the cusps (holomorphic, or meromorphic or holomorphic and prescribed zero).

More precisely, we name a weight $2k$, and define a *modular form* of weight $2k$ for Γ to be a holomorphic function on $\mathcal{H}$ such that

1. modularity:

$$f(g(\tau)) = f(\frac{a\tau+b}{c\tau+d}) = f(\tau)(c\tau+d)^{2k} \quad \text{for all } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

   Note that the differential $\frac{\mathrm{d}g(\tau)}{\mathrm{d}\tau} = (c\tau+d)^{-2}$, so that the modularity condition on f says that $f \cdot (\mathrm{d}\tau)^k$ is a Γ-invariant k times differential form on $\mathcal{H}$.

2. holomorphic at infinity: for simplicity, discuss first only the cusp $+i\infty$, and assume that Γ contains the translation T. Then f is a function of $q = \exp(2\pi i\tau)$; that is, it is a holomorphic function on the punctured unit disk in the q plane (see Figure 5.3).

Figure 5.3: Filling in the punctured q disk

So as in 1.3, by complex analysis it has a Laurent expansion $f = \sum a_n q^n$ with some coefficients a_n. We say that f is *holomorphic at infinity* if $a_n = 0$ for all $n < 0$, that is, f is a holomorphic function of q. We say f is a *cusp form* if also $a_0 = 0$, that is, f is holomorphic and zero at the cusp. (It also makes sense to allow f to be a meromorphic function at infinity with pole of given order. But we forbid f to have an essential singularity.)

In the more general case, there are several cusps. Each can be shifted to $+i\infty$ by a g element of $\mathrm{SL}(2,\mathbb{Z})$ (possibly not in Γ), and the conjugate subgroup $g\Gamma g^{-1}$ contains a translation $T^N : z \mapsto z+N$ for some N (the *width* of the cusp), and you say more or less the same with $q_N = \exp(2\pi i N\tau)$.

5.6 Eisenstein series G_{2k}

In 2.5 we defined the lattice sums $G_{2k}(L) = \sum' \frac{1}{w^{2k}}$. This is a function of the lattice L only, with the homogeneity $G_{2k}(aL) = a^{-2k}G_{2k}(L)$. Set $L_\tau = \mathbb{Z}\tau + \mathbb{Z}\cdot 1$ and $G_{2k}(\tau) = G_{2k}(L_\tau)$. This is now a function on the upper halfplane. Changing basis in L_τ by $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ doesn't change the lattice, so $L_\tau = \mathbb{Z}(a\tau+b) + \mathbb{Z}(c\tau+d)$, and clearly

$$(c\tau+d)^{-1}L_\tau = \mathbb{Z}g(\tau) + \mathbb{Z}\cdot 1.$$

Thus G_{2k} satisfies $G_{2k}(g(\tau)) = (c\tau + d)^{2k} G_{2k}(\tau)$. Notice the play between the 4 sets

$$\begin{array}{ccc} \big\{\text{lattice + oriented basis}\big\} & \longrightarrow & \big\{\text{lattice}\big\} \\ \big\downarrow & & \big\downarrow \\ \big\{\text{lattice } L_\tau = \mathbb{Z}g(\tau) + \mathbb{Z}\cdot 1\big\} & \longrightarrow & \big\{\text{lattice}\big\}/\text{sim} \end{array}$$

The top left-hand set is $\big\{w_1, w_2 \in \mathbb{C} \;\big|\; \operatorname{Im} w_1/w_2 > 0\big\}$; the left vertical arrow is $(w_1, w_2) \mapsto \tau = w_1/w_2 \in \mathcal{H}$. On the right, the same modulo similarity.

G_{2k} is a function of a lattice L, but introducing a basis then dividing by similarity to get L_τ, we find that lattices up to similarity is $\mathcal{H}$ modulo $\mathrm{SL}(2, \mathbb{Z})$.

Now

$$G_{2k}(\tau) = 2\zeta(2k) + \frac{2(2\pi i)^{2k}}{(2k-1)!} \sum \sigma_{2k-1}(n) q^n. \tag{5.2}$$

where ζ is the Riemann zeta function and σ is the sum of powers of divisors of n, that is $\sigma_l = \sum_{d|n} d^l$. (See Section 5.9 for these formulas.) To prove (5.2), write

$$\begin{aligned} \pi \cot \pi\tau &= i\pi \frac{q+1}{q-1} \\ &= i\pi - \frac{2\pi i}{1-q} = i\pi - (2\pi i) \sum q^d, \end{aligned} \tag{5.3}$$

where as usual $q = \exp(2\pi i \tau)$. Now it is "well known" that

$$\pi \cot \pi\tau = \frac{1}{\tau} + \sum_{m \geq 1} \Big(\frac{1}{\tau + m} + \frac{1}{\tau - m} \Big). \tag{5.4}$$

Differentiate k times to get

$$\sum_{m \in \mathbb{Z}} \frac{1}{(\tau + m)^{2k}} = (2\pi i)^{2k} \sum_{d \geq 1} d^{2k-1} q^d. \tag{5.5}$$

Now take the sum $G_{2k}(\tau) = \sum' \frac{1}{(n\tau + m)^{2k}}$ and break up into terms with $n = 0$ and $n \neq 0$. Those with $n = 0$ give

$$\sum_{m \neq 0} \frac{1}{m^{2k}} = 2\zeta(2k). \tag{5.6}$$

Those with $n \neq 0$ give

$$2 \sum_{n \geq 1} \sum_{m \in \mathbb{Z}} \frac{1}{(n\tau + m)^{2k}} \,. \tag{5.7}$$

Substitute from (5.5) for the inner term to get the double sum

$$\frac{2(2\pi i)^{2k}}{(2k-1)!} \sum_{d \geq 1} \sum_{a \geq 1} d^{2k-1} q^{da}.$$

5.7 Geometry of modular forms for $\mathrm{SL}(2, \mathbb{Z})$

$\mathcal{H}/\mathrm{SL}(2, \mathbb{Z})$ is a Riemann surface of genus 0, that is, the sphere $\mathbb{P}^1_{\mathbb{C}} = S^2$. You glue the sides of D and the boundary circle as in Figure 5.2, and fill in the q disk as in Figure 5.3. The map $\mathcal{H} \to S^2$ has ramification of order 2 at i and 3 at w, and of course logarithmic ramification at $\tau = +i\infty$, corresponding to $q = 0$.

Modular forms for $\mathrm{SL}(2, \mathbb{Z})$ correspond to k times differential forms on $\mathbb{P}^1_{\mathbb{C}}$ with poles of order $\leq k$ at the cusp $q = 0$, and of order $\leq [k/2]$ at i and $[2k/3]$ at w, where [ ] denotes integral part. It is not hard to see that the dimension of these is

$$\dim M_2 k = \begin{cases} [k/6] + 1 & \text{if } k \not\equiv 1 \bmod 6, \\ [k/6] & \text{if } k \equiv 1 \bmod 6. \end{cases}$$

The idea is the same as Riemann–Roch on $\mathbb{P}^1_{\mathbb{C}}$ of 1.8: k times differential forms contribute $-2k$ to the degree. k times the pole at infinity contributes k, and $k/2$, $2k/3$ at the finite ramification points gives $k/6$ minus what you lose in the fractional part. This is fun and not hard, but I don't have time to explain properly.

Theorem 5.8 *The ring of modular forms (summed over all weights $2k$) is generated by G_4 and G_6.*

In other words, the vector space of all modular forms of weight $2k$ has basis made up of $G_4^a G_6^b$ for all $4a + 6b = 2k$.

5.9 Table of formulas

$$G_{2k}(\tau) = 2\zeta(2k) + \frac{2(2\pi i)^{2k}}{(2k-1)!} \sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n, \quad \text{where} \quad \sigma_l(n) = \sum_{d \,|\, n} d^l. \tag{5.8}$$

We need the values

$$\zeta(2) = \frac{\pi^2}{6}, \quad \zeta(4) = \frac{\pi^4}{90} \quad \text{and} \quad \zeta(6) = \frac{\pi^6}{3^3 \cdot 5 \cdot 7}\,. \tag{5.9}$$

(More generally, note that $\frac{e^x+1}{e^x-1}$ is an odd function of x, so that we can define the *Bernoulli numbers* as the coefficients of the power series expansion

$$\frac{e^x+1}{e^x-1} \times \frac{x}{2} = \frac{x}{e^x-1} + \frac{x}{2} = 1 + \sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^{2k}}{(2k)!} B_k \tag{5.10}$$

Then $\zeta(2k) = \frac{2^{2k-1}\pi^{2k}}{(2k)!} B_k$. It's easy to find B_1, B_2, B_3, etc., by taking the first few terms in the expansion.)

Thus

$$G_{2k}(\tau) = 2\zeta(2k) + \frac{2(2\pi i)^{2k}}{(2k-1)!} \sum_{d=1}^{\infty} \frac{d^{k-1} q^d}{1-q^d}, \tag{5.11}$$

In writing out the Weierstrass equation $(\wp')^2 = 4\wp^3 - g_2\wp - g_3$ for the elliptic function $\wp$ we used the scaling factors

$$g_2 = 60 G_4, \quad g_3 = 140 G_6 \tag{5.12}$$

and defined the discriminant Δ by

$$\Delta(\tau) = g_2(\tau)^3 - 27 g_3(\tau)^2 = 60^3 G_4(\tau) - 27 \cdot 140^2 G_6(\tau) \tag{5.13}$$

It can be shown that

$$\Delta(\tau) = (2\pi)^{12} q \prod_{n=1}^{\infty} (1-q^n)^{24}. \tag{5.14}$$

The j function is

$$j(\tau) = 1728 \frac{g_2(\tau)^3}{\Delta(\tau)}. \tag{5.15}$$

$$G_4(\tau) = \pi^4 \Big(\frac{1}{45} + 2q + \frac{2}{3!}(1+2^4) q^2 + \frac{2}{5!}(1+3^4) q^3 + \cdots \Big) \tag{5.16}$$

$$G_6(\tau) = \pi^6 \Big(\frac{2}{3^3 \cdot 5 \cdot 7} - 2q + \frac{2}{3!}(1+2^6) q^2 + \frac{2}{5!}(1+3^6) q^3 + \cdots \Big) \tag{5.17}$$

$$j(\tau) = \frac{1}{q} + 744 + 196884 q + \cdots \tag{5.18}$$

5.10 Hecke subgroups

The most useful subgroups of finite index $\Gamma \subset \mathrm{SL}(2,\mathbb{Z})$ are

$$\begin{aligned}\Gamma(N) &= \left\{ g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) \;\middle|\; g \equiv \text{ identity} \bmod N \right\} \\ \Gamma_0(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) \;\middle|\; c \equiv 0 \bmod N \right\}.\end{aligned}$$

Here $\Gamma(N) = \ker\big[\mathrm{SL}(2,\mathbb{Z}) \to \mathrm{SL}(2,\mathbb{Z}/N)\big]$. In other words, if we make $\mathrm{SL}(2,\mathbb{Z})$ act on $\mathbb{Z}/N \oplus \mathbb{Z}/N$ in the obvious way, then $\Gamma(N)$ is the subgroup of matrixes that act trivially, $\Gamma_0(N)$ the subgroup of elements that take the second summand to itself.

Recall Euler's phi function

$$\varphi(N) = N \prod \left(1 - \frac{1}{p}\right) \quad \text{product over prime factors of } N.$$

The index of $\Gamma(N)$ in $\mathrm{SL}(2,\mathbb{Z})$ equals the number of bases of $\mathbb{Z}/N \oplus \mathbb{Z}/N$ divided by $\varphi(N)$, which can be calculated. It's obviously a lot simpler for a prime $N = p$. Then the number of bases is $(p^2-1)(p^2-p)$ and the index of $\Gamma(p)$ equals $(p^2-p)(p+1)$.

The index of $\Gamma_0(N)$ in $\mathrm{SL}(2,\mathbb{Z})$ equals $N\prod(1+\frac{1}{p})$, where the product runs over all $p \mid N$. For a prime $N = p$, this index is $p+1$.

The cusps of $\Gamma_0(N) =$ the orbits of $\Gamma_0(N)$ on $\mathbb{P}^1_{\mathbb{Q}}$. The cusps correspond to the cyclic subgroups of $\mathbb{Z}/N$ generated by the first entry of a primitive vector of $\mathbb{Z}/N + \mathbb{Z}/N$, so are in 1-to-1 correspondence with divisors of N.

For a prime $N = p$ there are just two cusps, 0 and $+i\infty$. To see the fundamental domain of $\Gamma_0(p)$. It's easier to do the conjugate of $\Gamma_0(p)$ by S, which is

$$S\Gamma_0(p)S = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) \;\middle|\; b \equiv 0 \bmod N \right\}.$$

This has coset representatives S and $1, T, T^2, \ldots, T^{p-1}$, and its fundamental domain is D, its translates $T(D), \ldots, T^{p-1}(D)$ and inversion $S(D)$. Its cusps are $+i\infty$ of width p and 0 of width 1.

5.11 Modular forms of weight 2 for $\Gamma_0(N)$

Write $M_2(\Gamma_0(N))$ for the vector space of modular forms of weight 2 for $\Gamma_0(N)$ and $S_2(\Gamma_0(N))$ for the cusp forms (that is, modular forms vanishing at the

cusps, with no constant term in their q-expansion). The modularity condition on f just says that $f\,\mathrm{d}\tau$ is an invariant differential form; in view of ($q = \exp(2\pi i\tau)$ so $\frac{\mathrm{d}q}{q} = 2\pi i\,\mathrm{d}\tau$), the cusp condition just says that $f(\tau)\,\mathrm{d}\tau = (2\pi i)^{-1}\frac{f(q)}{q}\,\mathrm{d}q$ is a holomorphic differential. Thus $S_2(\Gamma_0(N))$ equals the space of holomorphic differentials on the Riemann surface $X_0(N) =$ completion of $H/\Gamma_0(N)$ (completed by adding the cusps). By the general theory of Riemann surfaces, the dimension of this space equals the genus of $X_0(N)$. This genus can be calculated from the Euler number, which can be calculated exactly as for the index and set of cusps in 5.7. Obviously the 2-fold and 3-fold ramification of $H \to X_0(N)$ at the orbits of i and w will intervene. For a prime $N = p$ it turns out that

$$\dim S_2(\Gamma_0(p)) = \begin{cases} n-1 & \text{if } p = 12n+1 \\ n & \text{if } p = 12n+5 \text{ or } 12n+7 \\ n+1 & \text{if } p = 12n+11. \end{cases}$$

Thus for

$$\begin{aligned} p &= 5, 7, 13 \text{ we get } g = 0 \\ p &= 11, 17, 19 \text{ we get } g = 1, \\ p &= 23, 29, 31, 37 \text{ we get } g = 2, \ldots \end{aligned}$$

$X_0(11), X_0(17), X_0(19)$ are elliptic curves, and are our first examples of the modular elliptic curves we've come so far to define.

5.12 Example of cusp forms

Recall from Section 5.9 that we own a cusp form for $\mathrm{SL}(2,\mathbb{Z})$ of weight 12, namely

$$\Delta(\tau) = 60^3 G_4(\tau) - 27 \cdot 140^2 G_6(\tau) = (2\pi)^{12} q \prod (1-q^n)^{24}.$$

If we take

$$\eta(\tau) = \exp\Big(\frac{\pi i \tau}{12}\Big) \prod_{n=1}^{\infty} (1-q^n)$$

so that $\Delta = (2\pi)^{12}\eta^{24}$, then η is not itself a modular form, but Assessment E.5 gave simple functional equations for $\eta(\tau+1)$ and $\eta(-1/\tau)$ that are "almost modular," and make it the mother of many cusp forms:

$$(\eta(\tau)^p/\eta(p\tau))^2 \in S_2(\Gamma_0(p)) \quad \text{for } p \equiv 11 \bmod 12.$$

5.13 Modular curves $X_0(N)$

The completed quotients $X_0(N)$ = completion of $H/\Gamma_0(N)$ are algebraic curves defined over $\mathbb{Q}$ with rational points given by the cusps. An elliptic curve C over $\mathbb{Q}$ (with origin $O \in C$) is *modular* if there is a surjective map $X_0(N) \to C$ that is a morphism of algebraic curves defined over $\mathbb{Q}$ (taking a cusp to O).

If $X_0(N)$ has genus = 1 then $X_0(N)$ itself is a modular elliptic curve. Modular elliptic curves can be predicted from $f \in S_2(X_0(N))$ in terms of more stuff on modular forms that I don't have time to explain: given

1. f is an eigenform of all the Hecke operators (that is, has a bit more symmetry deriving from matrixes $\mathrm{SL}(2,\mathbb{Q})$ with some divisors of N in the denominators)

2. f is a newform (that is, is orthogonal w.r.t. the natural inner product to all the forms in $S_2(X_0(N'))$ with $N' \mid N$)

then f defines a quotient map to a modular elliptic curve $X_0(N) \to E_f$.

5.14 Number of points of E over $\mathbb{F}_p$

If E is an elliptic curve over $\mathbb{Q}$, there is a way of writing it in Tate form with integer coefficients, and giving minimal discriminant Δ. For example $y^2 + y = x^3 - x^2$ has $\Delta = 11$. Then modulo every good prime (not dividing Δ),

$$E_p = \text{ the curve over } \mathbb{F}_p \text{ defined by the same equation}$$

is a nonsingular elliptic curve. We count its points over $\mathbb{F}_p$, or equivalently, the number of solutions to the equation of E viewed as a congruence modulo p. (I always include the point at infinity in this calculation.)

If $E : y^2 = x^3 + ax + b = f(x)$, you expect to get about $1 + p$ solutions. For there are $1 + p$ values of x (including ∞), and for each finite value either $f(x) = 0$, and you get one solution, or $f(x) \neq 0$, in which case there is probability $\frac{1}{2}$ that it is a quadratic residue (q.r.), when you get two values for y. There are a number of cases in which you can do all this exactly by baby number theory: for example, if the r.h.s. is $x^3 + b$ and $p \equiv 2 \bmod 3$ then x^3 just runs through all values mod p in a 1-to-1 way, and the number of solutions is exactly $1 + p$. Similarly for $y^2 = x(x^2 + a)$ and $p \equiv 1 \bmod 4$, because each pair $\pm x$ contains one q.r. and one nonresidue. (However, all these cases are complex multiplication, so not typical.)

Theorem 5.15 (Hasse–Weil estimate)

$$\#(E(\mathbb{F}_p) = 1 - a_p + p,$$

where $|a_p| < 2\sqrt{p}$.

We think of $1, a_p$ and p as corresponding to $H^0(E(\mathbb{C})), H^1(E(\mathbb{C})), H^2(E(\mathbb{C}))$ respectively. The Hasse–Weil estimate was vastly generalised (to an analogous result for an arbitrary algebraic variety over any finite field) by Weil, Grothendieck and Deligne. The L function of E is

$$L(E, s) = \prod_p \text{ (local factor at } p)$$

where the local factor is

$$\frac{1}{1 - a_p p^{-s} + p p^{-2s}} \quad \text{if } p \text{ is good, or} \quad \frac{1}{1 - a_p p^{-s}} \quad \text{if } p \mid \Delta.$$

You can expand this out as a Dirichlet series $\sum_{n \geq 1} a_n n^{-s}$, where the coefficient a_p is the same, and a_n is an elementary combination of a_p for $p \mid n$. By elementary convergence of products, the product and the Dirichlet series converge for $\operatorname{Re} s > 2$.

5.16 L function of E

$L(E, s)$ is what analytic number theorists do to make a generating function for the data $\#(E(\mathbb{F}_p)$ at each p. Compare the Euler product for the Euler–Riemann zeta function

$$\zeta(s) = \prod \frac{1}{(1 - p^{-s})} = \sum n^{-s},$$

at the most basic level, the pole of $\zeta(s)$ at $n = 1$ says that there are infinitely many primes. Similar things for the L functions in Dirichlet's proof of primes in arithmetic progressions. The L function $L(E, s)$ of an elliptic curve is made out of the finite congruences mod p, but (at least conjecturally) contains information about E over $\mathbb{Q}$. For example, the main difficult problem after Wiles and co.'s solution of Taniyama–Shimura–Weil is the Birch–Swinerton-Dyer conjecture that $L(E, s)$ extends analytically, and has pole at $s = 1$ of order equal to the rank of the Mordell–Weil group of E. This is a very precise

quantitative statement to the effect that if $E(\mathbb{Q})$ has large rank, then $E(\mathbb{F}_p)$ tends to be consistently a little bigger than $1 + p$.

However, the only way anyone knows of showing that an L function defined by a Dirichlet series has analytic extension is to prove functional equations saying $L(2 - s) =$ closely related to $L(s)$. Why should such functional equations happen?

5.17 Langlands correspondence

The following is a purely formal way of going between Dirichlet series and Fourier series:

$$\sum_{n \geq 1} a_n n^{-s} \longleftrightarrow \sum_{n \geq 1} a_n q^n$$

where $q = \exp(2\pi i \tau)$. This is a kind of Fourier transform up the imaginary axis, called Mellin transform.

The left-hand side is where L functions of elliptic curves $L(E, s)$ live. The right-hand side is where the q-expansion of a cusp form f in $S_2(\Gamma_0(N))$ lives. Everything fits together. An elliptic curve has a conductor $N = \prod p^{\varepsilon}$, where the product runs through primes p dividing the discriminant Δ and $\varepsilon = 1$ or 2 depending on the nature of the "bad reduction" modulo p (in other words, how singular the curve E_p over $\mathbb{F}_p$ is).

Eichler and Shimura proved that if f is a cusp form for $\Gamma_0(N)$ that is an eigenform of all the Hecke operators, and is newform, then the curve $X_0(N) =$ completion of $H/\Gamma_0(N)$ has a surjective map to an elliptic curve E_f defined over $\mathbb{Q}$, and $L(E_f, s)$ (and its twists by characters) is the Dirichlet series corresponding to the q-expansions of the cusp form f. Conversely, Weil proved that if the L-function of an elliptic curve over $\mathbb{Q}$ with character N have functional equations (and also a few of its twists $L_\chi(s)$ by characters), then the corresponding Fourier series are q-expansions at cusps of a modular form for $\Gamma_0(N)$.

The two equivalent definitions of modular elliptic curve are here:

1. uniformised by certain very good modular forms for $\Gamma_0(N)$;
2. $E/\mathbb{Q}$ with conductor N whose L function $L(E, s)$ (and L functions with characters) has enough functional equations.

The Taniyama–Shimura–Weil conjecture is that every elliptic curve over $\mathbb{Q}$ is modular. This was proved by Wiles and Taylor–Wiles in 1995–96, under

some extra assumption, but sufficient to imply FLT. It has since been proved in its entirety by Conrad, Diamond, Taylor.

Serre and Ribet proved that the Frey curve $(*)$ is not modular. Since the conductor N of $(*)$ is small, there are few suitable modular forms around, and we can look up each of them and show that Eichler–Shimura does not produce $(*)$ from them.

References

[1] P. Du Val, Elliptic functions and elliptic curves, CUP 1973

[2] A W Knapp, Elliptic curves, Princeton 1992

[3] H McKean and V Moll, Elliptic curves, CUP 1997

[4] M Reid, Undergraduate algebraic geometry, CUP (Chapters 1–2 only)

[5] J-P Serre, A course of arithmetic, Springer (Chap. VII only)

[6] J Silverman, A friendly introduction to number theory, Prentice-Hall (for premodule reading)

[7] J Silverman, The arithmetic of elliptic curves, Springer (Advanced and detailed)

[8] J Silverman and J Tate, Rational points on elliptic curves, Springer

[9] I Stuart and D Tall, Algebraic number theory and Fermat's last theorem (Second edition of Algebraic number theory, Prentice–Hall)

[10] E.T. Whittaker and Watson, A course of modern analysis, CUP 1927 (reissued)