# UMichigan MATH 679 – Elliptic Curves

Based on lectures by Andrew Snowden

Notes taken by Sachin Kumar

Univeristy of Waterloo, Faculty of Mathematics

Winter 2024

These notes were taken while attending Andrew Snowden's lectures on MATH 679 at the Univeristy of Michigan. This course is an introduction to elliptic curves, focusing on the arithmetic side of the theory. It develops the theory of elliptic curves over various types of fields and in particular over algebraic number fields. The purpose of this course is to prove Mazur's theorem on torsion in elliptic curves over the rational numbers. Much of the course is devoted to developing background material on elliptic curves and their moduli.

*Prerequisites: Commutative Algebra, algebraic number theory, and algebraic geometry*

# Contents

# 0 Overview

Here is the link for the lecture videos by Andrew Snowden for MATH 679.

In this lecture, I give an introduction to Mazur's theorem, including a sketch of the proof. The purpose of this course is to develop these ideas in more detail. I also talk a little about some related results: Serre's uniformity theorem and Merel's uniform boundedness theorem.

## Mazur's theorem

Consider the following problem:

**Problem.** Let $f \in \mathbb{Q}[x, y]$ be a polynomial. Describe the set of points $(x, y) \in \mathbb{Q}^2$ such that $f(x, y) = 0$.

This can be phrased equivalently as:

**Problem.** Let $C/\mathbb{Q}$ be an algebraic curve (connected, smooth, projective). Describe the set $C(\mathbb{Q})$.

This is an extremely fundamental problem, and cases of it have been considered for thousands of years. Much about this problem has been discovered in the last century. The first thing to mention, probably, is the fundamental trichotomy depending on the genus of $C$:

(i) Genus 0. There are two possibilities: either $C$ has no rational points, or $C$ is isomorphic to $\mathbb{P}^1$, in which case it has infinitely many rational points, and these points form a 1-parameter algebraic family.

(ii) Genus 1. If $C$ is non-empty then $C(\mathbb{Q})$ has the structure of a finitely generated commutative group. The hard part of this statement (the finite generation) is Mordell's theorem, from 1922. (First conjectured by Poincaré.)

(iii) Genus $\geq 2$. the set $C(\mathbb{Q})$ is finite. This is Faltings' theorem (MR0718935), proved in 1983 and first conjectured by Mordell in 1922.

Given these results, one can start to ask more quantitative questions. For example:

**Question.** How many rational points can a genus 2 curve have?

It is conjectured that there is an absolute bound, i.e., there exists a number $N$ such that if $C/\mathbb{Q}$ is any genus 2 curve then $\#C(\mathbb{Q}) \leq N$. This has not been proved, however. So far, the record for number of rational points seems to be 642, found by Michael Stoll in 2008. It is worth mentioning here a recent result of Manjul Bhargava (arXiv:1308.0395): most genus 2 curves have no rational points. The situation in higher genus is similar.

In genus 1, one should not simply ask "how many points" but "what is the structure of the group of points." Suppose $C$ is a genus 1 curve with a point. Then, according to Mordell's theorem, we have a decomposition $C(\mathbb{Q}) = C(\mathbb{Q})_{\text{tors}} \times \mathbb{Z}^r$, where $C(\mathbb{Q})_{\text{tors}}$ is a finite abelian group (the torsion subgroup) and $r \geq 0$ is an integer, called the rank. So, one would like to know the possibilities for $r$ and $C(\mathbb{Q})_{\text{tors}}$.

Very little is known about the rank. For instance, it is unknown if it can be arbitrarily large. The current record is $r \geq 28$, found by Noam Elkies in 2006.

The situation is much better for the torsion subgroup; in fact, this is exactly what Mazur's theorem describes:

**Theorem** (Mazur, 1977, MR488287)**.** $C(\mathbb{Q})_{\text{tors}}$ is isomorphic to one of the following 15 groups:

(i) $C(\mathbb{Q})_{\text{tors}}$ is isomorphic to one of the following 15 groups:

(ii) $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ with $n = 2, 4, 6, 8$.

Furthermore, each of these groups does occur. One can view this theorem as a first step to a quantitative answer to the original problem.

## Overview of the proof

The goal of this course is to prove Mazur's theorem. I'll now give an overview of the proof, which I'll break into three steps. I must thank Jacob Tsimerman here, as he and I came to this organization while reading Mazur's paper together.

**Step 1:** A criterion for the non-existence of $N$-torsion.

The hard part of the proof is to show that an elliptic curve over $\mathbb{Q}$ cannot have an $N$-torsion point, for $N$ a prime $> 7$. Let $Y_1(N)$ be the set of isomorphism classes of pairs $(E, P)$ where $E/\mathbf{C}$ is an elliptic curve and $P \in E$ is a point of exact order $N$. Then $Y_1(N)$ is actually (the set of complex points of) an algebraic curve defined over the rational numbers. Furthermore, the set of rational points on $Y_1(N)$ is exactly what you'd expect: it consists of those $(E, P)$ for which $E$ is defined over $\mathbb{Q}$ and $P \in E(\mathbb{Q})$. Thus Mazur's theorem essential amounts to showing that $Y_1(N)(\mathbb{Q})$ is empty for $N > 7$. The proof will appeal to the dual nature of $Y_1(N)$: it can be thought of as a single geometric object, or as a set of geometric objects.

We'll need some slight variants of $Y_1(N)$. Let $Y_0(N)$ be the set of pairs $(E, G)$ where $E/\mathbf{C}$ is an elliptic curve and $G \subset E$ is a cyclic subgroup of order $N$. This is also an algebraic curve over $\mathbb{Q}$. There is a natural map $Y_1(N) \to Y_0(N)$ (take the subgroup generated by the point). The curve $Y_0(N)$ is affine: it's missing two points, which are labeled 0 and $\infty$, and called the cusps. The compactified curve is denoted $X_0(N)$. We can now state the criterion:

**Theorem** (Theorem A)**.** Suppose $N > 7$ and there exists an abelian variety $A/\mathbb{Q}$ and a map of varieties $f\colon X_0(N) \to A$ (defined over $\mathbb{Q}$) such that the following conditions hold:

(i) $A$ has good reduction away from $N$.

(ii) $f(0) \neq f(\infty)$.

(iii) $A(\mathbb{Q})$ has rank 0.

Then no elliptic curve defined over $\mathbb{Q}$ has a point of order $N$.

*Proof.* We just offer a sketch here. Suppose $E/\mathbb{Q}$ is an elliptic curve which has a point of order $N$. Let $x \in X_0(N)(\mathbb{Q})$ be the resulting rational point. We first remark that $X_0(N)$ naturally extends to a scheme over $\mathbb{Z}[1/N]$ (or even all of $\mathbb{Z}$) and $x$ extends to a section over this base as well. By studying the reduction of $E$ mod 3, and using the fact that 3 is small compared to $N$, one finds that $E$ must have bad reduction at 3. This means that $x$ must reduce to either 0 or $\infty$ mod 3, and in fact it must be $\infty$. To see this, one must be familiar with the modular interpretations of the two cusps, which we will cover later in the course.

Now for the key step: the difference $f(x) - f(\infty)$ is an element of $A(\mathbb{Z}[1/N])$ which reduces to 0 in $A(\mathbb{F}F_3)$. However, $f(x) - f(\infty)$ is a torsion point (since $A$ has rank 0), and the reduction map is injective on torsion. We conclude that $f(x) = f(\infty)$. It follows from this, and the assumption that $f(\infty) \neq f(0)$, that if $p$ is any prime of bad reduction for $E$ then $x$ reduces to $\infty$ mod $p$.

Now, consider $E[N]$ as a 2-dimensional representation of the absolute Galois group $G_{\mathbb{Q}}$ over the finite field $\mathbb{Z}/N\mathbb{Z}$. Since $E$ has an $N$-torsion point, this representation contains the trivial representation $\mathbb{Z}/N\mathbb{Z}$ as a sub. The Weil pairing implies that the quotient is $\mu_N$. The modular interpretation of the cusp $\infty$ shows that the resulting extension is actually split at all the bad primes. A number-theoretic argument then shows that the extension is split globally, i.e., $E[N]$ is isomorphic to $\mathbb{Z}/N\mathbb{Z} \oplus \mu_N$. One can apply the same argument to $E/\mu_N$ to see that its $N$-torsion is split; continuing in this way, one finds that the $N$-adic Tate module of $E$ is reducible, which cannot happen. This contradiction completes the proof. □

**Step 2:** A criterion for rank 0.

To apply Theorem A, we must find the abelian variety $A$ and verify the conditions of the theorem. The hardest of these is the rank 0 condition. We now give a criterion for an abelian variety to have rank 0. This may look like a general criterion, but the hypotheses are actually very restrictive; it will apply to the case of interest, however.

**Theorem** (Theorem B)**.** Let $A/\mathbb{Q}$ be an abelian variety and let $N$ and $p$ be distinct prime numbers, with $N$ odd. Suppose the following conditions hold:

(i) $A$ has good reduction away from $N$.

(ii) $A$ has completely toric reduction at $N$.

(iii) The Jordan–Holder constituents of $A[p](\overline{\mathbb{Q}})$ are 1-dimensional, and either trivial or cyclotomic.

Then $A(\mathbb{Q})$ has rank 0.

*Proof.* Again, just a sketch. Let $\mathcal{A}/\mathbb{Z}$ be the Néron model of $A$. One first shows that the group scheme $\mathcal{A}[p^n]$ is built of very simple pieces: it has a filtration such that the successive quotients are each one of four very specific group schemes. Computing explicitly with these specific group schemes, one shows that the order of $\mathrm{H}^1_{\mathrm{fppf}}(\mathrm{Spec}(\mathbb{Z}), \mathcal{A}[p^n])$ is bounded independent of $n$. This implies that the inverse limit over $n$ of these cohomology groups is finite, which completes the proof, as $A(\mathbb{Q})$ injects into the inverse limit. This proof is closely related to the proof of the Mordell–Weil theorem, which I'll talk about some. □

**Step 3:** Completion of the proof.

We now wish to prove Mazur's theorem by applying the above criteria. But first we must find the abelian variety $A$. Every curve has a Jacobian, a universal abelian variety to which it maps (given a point). Thus we are more or less forced to try to find $A$ as a quotient of the Jacobian $J_0(N)$ of $X_0(N)$.

Using the modular interpretation of $X_0(N)$, one constructs certain Hecke operators $T_p$ on $J_0(N)$. These generate a commutative ring of operators, called the Hecke algebra. We'll find $A$ by defining an explicit ideal in the Hecke algebra (closely related to the Eisenstein ideal appearing in the title of Mazur's paper), and forming the corresponding quotient of $J_0(N)$. We'll then go through each of the hypotheses in the two criteria and verify that $A$ satisfies them.

In fact, this argument will only end up working for $N > 13$, so auxiliary arguments are needed for $N = 11, 13$. For $N = 11$, the result was first established in 1939 by Billing–Mahler. For $N = 13$, it was established by Mazur–Tate in 1973.

## Plan of the course

The course will be divided into three parts, corresponding to the three steps above (though out of order!).

**Part I.** Elliptic curves and abelian varieties

(i) Theory over fields. I will give very few proofs here. I'll assume you're either familiar with this, or can do outside reading to learn it.

(ii) Group schemes. I won't assume you know much at all here, and I'll attempt to prove nearly everything we'll need.

(iii) Theory in mixed characteristic, including Néron models (though not the proof of their existence).

(iv) Jacobians.

(v) The culmination of Part I will be the proof of Theorem B.

**Part II.** Moduli of elliptic curves

(i) Modular curves, over $\mathbf{C}$, $\mathbb{Q}$, and $\mathbb{Z}$.

(ii) Modular forms and Hecke operators.

(iii) The Eichler–Shimura theorem, and the Galois representation attached to a modular form.

(iv) The culmination of Part II will be the proof of Theorem A.

**Part III.** Proof of Mazur's theorem

(i) The Eisenstein ideal and the Eisenstein quotient of $J_0(N)$.

(ii) The special fiber at $N$ of $J_0(N)$.

(iii) Ogg's theorem on the order of $[0] - [\infty]$ in $J_0(N)$.

(iv) Application of Theorems A and B.

(v) Auxiliary results (Mazur–Tate, etc).

## Related results

To end this lecture, I'll discuss two families of results that generalize Mazur's theorem. Unfortunately, we probably won't have time in this course to discuss these results further.

### Merel's theorem

You might be wondering: does a version of Mazur's theorem exist over general number fields? The answer is yes! For an integer $d \geq 1$, let $S(d)$ be the set of prime numbers $p$ for which there exists a number field $K/\mathbb{Q}$ of degree $\leq d$ and an elliptic curve $E/K$ such that $E$ has a $K$-point of order $p$. Mazur's theorem is that $S(1) = \{2, 3, 5, 7\}$. Kamienny proved (in 1992, MR1172689) that $S(2) = \{2, 3, 5, 7, 11, 13\}$. Mazur and Kamienny then conjectured that $S(d)$ is always finite (the Uniform Boundedness Conjecture, or UBC), which was proven by Merel:

**Theorem** (Merel, 1996, MR1369424)**.** The set $S(d)$ is finite. In fact, if $p \in S(d)$ then $p \leq d^{3d^2}$.

In 2003, Parent computed $S(3)$ and found it to be the same as $S(2)$ (MR2142238, see also MR1779891). The sets $S(4)$, $S(5)$, and $S(6)$ have also been computed, as I learned from Maarten's comment below. See his slides for more information (his website has other relevant notes and slides).

The review of Merel's paper by Darmon, linked above, contains a thorough overview of the UBC.

### Serre's uniformity theorem

Let $E/\mathbb{Q}$ be an elliptic curve, and let $N$ be a prime. Then $E[N](\overline{\mathbb{Q}})$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z})^2$, and carries an action of the absolute Galois group $G_{\mathbb{Q}}$. We can therefore regard it as a representation $\rho_{E,N} \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. Serre proved the following:

**Theorem** (Serre, 1972, MR387283)**.** Assume $E$ does not have complex multiplication. Then there exists a number $N_0(E)$ such that $\rho_{E,N}$ is surjective for all $N > N_0(E)$.

Serre posed the following question:

**Question** (Serre's uniformity problem)**.** Can $N_0(E)$ be taken independent of $E$? Precisely, does there exist a number $N_0$ such that $\rho_{E,N}$ is surjective whenever $E$ is non-CM and $N > N_0$?

It is thought that the answer to this question is yes, and that one can even take $N_0 = 37$.

If $\rho_{E,N}$ is not surjective, then its image is a proper subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, and thus contained in a maximal proper subgroup. One can therefore attack Serre's question by proving, for each maximal proper subgroup $G$ of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, that the image of $\rho_{E,N}$ is not contained in $G$ (for $N$ large enough). It's not difficult to enumerate the maximal proper subgroups:

(i) The Borel subgroup (upper-triangular matrices).

(ii) The normalizer of the split Cartan (monomial matrices).

(iii) The normalizer of the non-split Cartan.

(iv) Exceptional subgroups (those having projective image $A_4$, $S_4$, or $A_5$).

Serre himself dealt with the exceptional case: the image of $\rho_{E,N}$ is not contained in an exceptional subgroup if $N > 7$ (check this!). (In this and what follows, $E$ is non-CM.)

Mazur's theorem that we have been discussing above is close to handling the Borel, but doesn't quite: it shows that the image of $\rho_{E,N}$ is not contained in the group of matrices of the form $\begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix}$ for $N > 7$. However, Mazur extended his results (MR482230) and handled the Borel case: he showed that the image of $\rho_{E,N}$ is not contained in a Borel for $N > 37$.

In 2009, Bilu and Parent (MR2753610) handled the split Cartan case: the image of $\rho_{E,N}$ is not contained in the normalizer of the split Cartan for $N > N_0$, for some constant $N_0$. I took a very brief look at their paper and wasn't able to see if $N_0$ can be made explicit or not.

The non-split Cartan case is still open!

# 1 Elliptic curves and abelian varieties

## 1.1 Elliptic curves and abelian varieties over fields

This section covers the basics of elliptic curves. I begin with a brief review of algebraic curves. I then define elliptic curves, and talk about their group structure and defining equations. Following this is the theory of isogenies, including the important fact that "degree" is quadratic. Next is the complex theory: elliptic curves are one-dimensional tori. Finally, I talk about the Tate module and Weil pairing.

A good reference for this section is Silverman's "The arithmetic of elliptic curves" (MR0817210), especially chapters II, III, and VI.

### 1.1.1 Review of curves - Separability

**Divisors**

Let $k$ be a field. Let $C/k$ be a (smooth, projective, connected) algebraic curve. A divisor on $C$ is a formal sum $\sum_{x \in C} n_x[x]$, where $n_x$ is an integer and all but finitely many of the $n_x$ are 0. The set of divisors forms a group, $\mathrm{Div}(C)$.

We define the degree of $x \in C$ to be the the degree of the extension $k(x)/k$, where $k(x)$ is the residue field at $x$. The degree of a divisor $D = \sum n_x[x]$, denoted $\deg(D)$, is $\sum n_x \deg(x)$. We thus have a homomorphism $\deg\colon \mathrm{Div}(X) \to \mathbb{Z}$. We let $\mathrm{Div}^0(X)$ be its kernel.

We write $D \geq 0$ if $n_x \geq 0$ for all $x$; such divisors are said to be effective. We write $D \geq D'$ to mean $D - D' \geq 0$.

Let $f$ be a non-constant function on $C$. We define the divisor of $f$, denoted $\mathrm{div}(f)$, as $\sum v_x(f)[x]$, where $v_x(f)$ denotes the valuation of $f$ at $x$, i.e., the order of zero or pole of $f$ at $x$. An important theorem states that $\mathrm{div}(f)$ has degree 0; in other words, the number of zeros of $f$ is equal to the number of poles of $f$, when multiplicities are taken into account. A divisor of the form $\mathrm{div}(f)$ is called principal. The set of principal divisors forms a group $\mathrm{PDiv}(X)$.

The divisor class group of $X$, denoted $\mathrm{Cl}(X)$, is the quotient $\mathrm{Div}(X)/\mathrm{PDiv}(X)$. Since principal divisors have degree 0, it makes sense to speak of the degree of a divisor class, and we have a subgroup $\mathrm{Cl}^0(X)$.

Let $f\colon X \to Y$ be a map of curves. Given a divisor $D = \sum n_x[x]$ on $X$, we let $f_*(D)$ be the divisor $\sum n_x[f(x)]$ on $Y$. Given a divisor $D = \sum n_y[y]$ on $Y$, we let $f^*(D)$ be the divisor $\sum_{y \in Y} \sum_{f(x)=y} e(x|y) n_y[x]$ on $X$, where $e(x|y)$ is the ramification index. Both $f_*$ and $f^*$ are homomorphisms and preserve principal divisors. Furthermore, we have $f_*(f^*(D)) = \deg(f)D$.

**Riemann-Roch**

Let $D$ be a divisor on $X$. Define $\mathcal{L}(D)$ to be the set of functions $f$ on $X$ with $\mathrm{div}(f) \geq -D$. (This includes the zero function.) For example, if $D = [x]$ then $\mathcal{L}(D)$ consists of functions which have at worst a simple pole at $x$ and are holomorphic everywhere else. It's easy to see that $\mathcal{L}(D)$ is a $k$-vector space. Note that if $\deg(D) < 0$ then $\mathcal{L}(D) = 0$, since the divisor of a function has degree 0. We let $\ell(D)$ be the dimension of $\mathcal{L}(D)$.

**Theorem** (Riemann–Roch)**.** We have $\ell(D) - \ell(K - D) = \deg(D) - g + 1$, where $g$ is the genus of $C$ and $K$ is the so-called canonical divisor, which has degree $2g - 2$.

**Corollary.** If $\deg(D) > 2g - 2$ then $\ell(D) = \deg(D) - g + 1$.

*Proof.* With this hypothesis, $\deg(K - D) < 0$, and so $\ell(K - D) = 0$. □

**Special Case:** if $C$ has genus 1 then $\ell(D) = \deg(D)$ for $\deg(D) > 0$.

#### Separability

Let $f\colon X \to Y$ be a non-constant map of curves. We then have an extension of function fields $k(X)/k(Y)$. Field theory implies that there is a maximal intermeidate field $K$ such that $k(X)/K$ is purely inseparable and $K/k(Y)$ is separable. Going back to geometry, this means that we can factor $f$ as $X \xrightarrow{g} X' \xrightarrow{h} Y$, where $g$ is purely inseparable and $h$ is separable. We define the separable degree of $f$ to be the degree of $h$, and the inseparable degree of $f$ to be the degree of $g$.

Suppose $k$ has characteristic $p$ and $X$ is given by the equation $f(x, y) = 0$. Let $f^{(p)}(x, y)$ be the polynomial obtained by raising all the coefficients of $f$ to the $p$th power, and let $X^{(p)}$ be the curve defined by $f^{(p)}(x, y) = 0$. If $f(x, y) = 0$ then $f(x, y)^p = 0$. But since we're in characteristic $p$, raising to the $p$th power is a ring homomorphism, and so $f(x, y)^p = f^{(p)}(x^p, y^p)$. It follows that $(x, y) \mapsto (x^p, y^p)$ defines a map of curves $F_p\colon X \to X^{(p)}$. This map is called the Frobenius map, and is purely inseparable. We can similarly define a Frobenius map for powers of $p$.

The Frobenius map is essentially the only example of a purely inseparable map: a map $X \to Y$ factors as $X \to X^{(q)} \to Y$, where the first map is $F_q$ and the second map is separable. Of course, $q$ is in the inseparable degree of $f$.

In characteristic 0, all maps are separable 0.

### 1.1.2 Elliptic Curves

**Definition.** An elliptic curve is a pair $(E, 0)$ where $E$ is a genus 1 curve over $k$ and is a $k$-point of $E$.

#### Group Law

**Proposition.** The map $E(k) \to \mathrm{Cl}^0(E)$ given by $x \mapsto [x] - [0]$ is an isomorphism.

*Proof.* Suppose $D$ is a degree 0 divisors on $E$. Then $\ell(D + [0]) = 1$ by Riemann–Roch. Let $f \in \mathcal{L}(D + [0])$ be non-constant. Then necessarily $\mathrm{div}(f) = -D - [0] + [x]$ for some $x \in E$ of degree 1, i.e., $x \in E(k)$. Thus $D = [x] - [0]$ in $\mathrm{Cl}(E)$. This proves surjectivity. Injectivity is left to the reader. □

Since $\mathrm{Cl}^0(E)$ is a group, the above isomorphism allows us to define a group structure on $E(k)$. In fact, $E$ itself is a group variety, that is, the group law on $E(k)$ is induced from a map of varieties $E \times E \to E$.

#### Equations

The space $\mathcal{L}([0])$ obviously contains the constant functions. By Riemann–Roch, $\ell([0]) = 1$, and so $\mathcal{L}([0])$ consists exactly of the constant functions. We have $\ell(2[0]) = 2$, and so $\mathcal{L}(2[0])$ contains a non-constant function; call it $x$. We have $\ell(3[0]) = 3$, and so $\mathcal{L}(3[0])$ contains a function which does not belong to the span of 1 and $x$; call it $y$. We have $\ell(4[0]) = 4$, but we know what the new function is: it's just $x^2$. Similarly, we have $\ell(5[0]) = 5$, but we know what the new function is: it's $xy$. Finally, we have $\ell(6[0]) = 6$. But we know two new functions: $x^3$ and $y^2$. We therefore have 7 functions in $\mathcal{L}(6[0])$, namely, 1, $x$, $y$, $x^2$, $xy$, $x^3$, and $y^2$. It follows that there is a linear dependence:

$$a_1 y^2 + a_2 x^3 + a_3 xy + a_4 x^2 + a_5 y + a_6 x + a_7 = 0$$

This equation defines a plane curve $E'$, and we have a natural map $E \to E'$ taking a point $p \in E$ to $(x(p), y(p))$. One can show that this map is an isomorphism of projective curves. Thus every elliptic curve is given by an equation of the above form.

Assume now that $k$ is not of characteristic 2 or 3. Then using a simple change of variables we can eliminate many of the above terms and reach an equation of the form

$$y^2 = x^3 + ax + b$$

Let us call this equation (or the curve it defines) $E_{a,b}$. By putting $y = u^{-3}y_1$ and $x = u^{-2}x_1$, for some non-zero $u \in k$, we see that $E_{a,b}$ is isomorphic to $E_{u^4a,u^6b}$. In fact, these are the only isomorphisms between these curves.

#### Discriminant and $j$-invariant

We have shown that every elliptic curve is of the form $E_{a,b}$, but are all $E_{a,b}$ elliptic curves? The answer is no: some of them are singular. In fact, $E_{a,b}$ is singular if and only if the discriminant $\Delta = -16(4a^3 + 27b^2)$ vanishes. If $\Delta$ is non-zero then $E_{a,b}$ is an elliptic curve.

We have therefore shown that the set of isomorphism classes of elliptic curves over $k$ is naturally in bijection with the set of pairs $(a, b) \in k^2$ with $\Delta \neq 0$, modulo the equivalence $(a, b) \sim (u^4a, u^6b)$ for $u \in k^\times$.

Define $j = -1728(4a)^3/\Delta$. The constants here are not so important. It's clear that $j$ is invariant under the equivalence relation. Thus $j$ is an invariant of elliptic curves: it is called the $j$-invariant. Using what we have stated above, it is easy to see that if $k$ is algebraically closed then two elliptic curves are isomorphic if and only if they have the same $j$-invariant. This is not true, in general, if $k$ is not closed.

### 1.1.3 Isogenies

#### Definition and examples

**Definition.** An isogeny $f\colon E_1 \to E_2$ is a non-constant map of curves with $f(0) = 0$.

One can show that any isogeny is a group homomorphism. We let $\mathrm{Hom}(E_1, E_2)$ denote the set of isogenies, together with the zero morphism. This is a group, via the group law on either $E_2$. In fact, it is a free abelian group of finite rank. We write $\mathrm{End}(E)$ for $\mathrm{Hom}(E, E)$. This is a ring, where multiplication is composition.

**Example.** The multiplication-by-$n$ map, denote $[n]\colon E \to E$, is an isogeny. We regard $\mathbb{Z}$ as a subring of $\mathrm{End}(E)$ by $n \mapsto [n]$.

**Example.** If $k$ has characteristic $p$, then the Frobenius map $F_q\colon E \to E^{(q)}$ is an isogeny.

#### Basic results

**Proposition.** Let $f\colon E_1 \to E_2$ be an isogeny of separable degree $n$ and inseparable degree $m$.

(i) For any $y \in E_2(\overline{k})$, the set $f^{-1}(y) \subset E_1(\overline{k})$ has $n$ elements.

(ii) For any $y \in E_2$ and any $x \in E_1$ lying over $y$, the ramification index $e(x|y)$ is equal to $m$.

(iii) The map $f$ is everywhere unramified if and only if $m = 0$; this is automatic in characteristic 0.

**Proposition.** Let $E_1$ and $E_2$ be elliptic curves and let $\omega_1$ and $\omega_2$ be non-zero global differentials on them.

(i) An isogeny $f$ is separable if and only if $f^*(\omega_2)$ is non-zero.

(ii) For an isogeny $f$ define $\alpha(f) \in k$ by $f^*(\omega_2) = \alpha(f)\omega_1$. Then $\alpha\colon \mathrm{Hom}(E_1, E_2) \to k$ is a group homomorphism.

(iii) If $E_1 = E_2$ and $\omega_1 = \omega_2$ then $\alpha$ is a ring homomorphism.

**Corollary.** The isogeny $[n]\colon E \to E$ is separable if and only if $n$ is prime to the characteristic of $k$.

### Dual isogeny

**Proposition.** Let $f\colon E_1 \to E_2$ be an isogeny. Then there exists an isogeny $f^\vee\colon E_2 \to E_1$, called the dual isogeny, such that the following diagram commutes.

$$\begin{tikzcd}
E_2(k) \arrow[r, equal] \arrow[d, "f^\vee"'] & \mathrm{Cl}^0(E_2) \arrow[d, "f^*"] \\
E_1(k) \arrow[r, equal] & \mathrm{Cl}^0(E_1)
\end{tikzcd}$$

In fact, this diagram continues to commute if we pass to extensions of $k$, and this uniquely specifies $f^\vee$. With a little more sophistication, one can construct $f^\vee$ using this diagram. We have the following two important facts: (1) $f^\vee f = [\deg f]$; and (2) $(f+g)^\vee = f^\vee + g^\vee$. These can be deduced easily from the above characterization of $f^\vee$ and properties of $f^*$.

### The quadratic nature of degree

Let $E_1$ and $E_2$ be elliptic curves and let $\Lambda = \mathrm{Hom}(E_1, E_2)$, a finite free $\mathbb{Z}$-module. For $f, g \in \Lambda$, define an element $\langle f, g\rangle$ of $\frac{1}{2}\mathbb{Z}$ by

$$2\langle f, g\rangle = \deg(f+g) - \deg(f) - \deg(g)$$

Using the identity $\deg(f) = f^\vee f$, we find

$$2\langle f, g\rangle = f^\vee g + g^\vee f$$

It follows that $\langle,\rangle$ is bilinear. The above expression shows that $\langle f, f\rangle = \deg(f)$, which shows that $\langle,\rangle$ is positive definite. It also shows that deg is a quadratic function.

An important corollary of this discussion is that $\deg([n]) = n^2$. This follows from the quadratic nature of deg and the obvious fact that $\deg([1]) = 1$.

### 1.1.4 Elliptic curves over the complex numbers

### Uniformization

Let $E$ be an elliptic curve over $\mathbb{C}$, and identify $E$ with its complex points. Then $E$ is a genus 1 surface. Its universal cover is therefore the complex numbers, and, in fact, the covering map $\pi : \mathbb{C} \to E$ is a group homomorphism. The kernel of $\pi$ is a lattice $\Lambda$, i.e., a subgroup of $\mathbb{C}$ such that the map $\Lambda \otimes \mathbb{R} \to \mathbb{C}$ is an isomorphism. We note that $\Lambda$ is naturally identified with the homology $\mathrm{H}_1(E, \mathbb{Z})$. Note that if $f$ is a meromorphic function on $E$ then $f \circ \pi$ is a doubly-periodic meromorphic function on $\mathbb{C}$, that is, it is invariant by translation by any element of $\Lambda$.

Now suppose that $\Lambda$ is a lattice in $\mathbb{C}$, and let $E = \mathbb{C}/\Lambda$ (and $0 = \pi(0)$). Then $E$ is a genus 1 Riemann surface. Riemann's theory implies that $E$ is an algebraic curve. To prove this, one must construct meromorphic functions on $E$, which amounts to constructing meromorphic functions on $\mathbb{C}$ with period lattice $\Lambda$. The basic example of such a function is the Weierstrass $\wp$ function, defined by

$$\wp(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda^*} \frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}$$

where here $\Lambda^* = \Lambda \setminus \{0\}$. The derivative $\wp'(z)$ is also doubly periodic, and satisfies an equation of the form

$$\wp'(z)^2 = 4\wp(z)^3 + a\wp(z) + b$$

Thus $z \mapsto (\wp(z), \wp'(z))$ defines an isomorphism of $E$ with a plane algebraic curve.

### Isogenies

Suppose $E_1 = \mathbb{C}/\Lambda_1$ and $E_2 = \mathbb{C}/\Lambda_2$. Then one can show that $\mathrm{Hom}(E_1, E_2)$ is naturally in bijection with the set of complex numbers $\alpha \in \mathbb{C}$ such that $\alpha\Lambda_1 \subset \Lambda_2$. The map corresponding to $\alpha$ is an isogeny if and only if $\alpha \neq 0$, and an isomorphism if and only if $\alpha\Lambda_1 = \Lambda_2$. We can thus say that the set of isomorphism classes of elliptic curves over $\mathbb{C}$ is naturally in bijection with the set of lattices in $\mathbb{C}$ modulo scaling (aka, homothety).

### Complex multiplication

Let $\Lambda$ be a lattice. Scaling $\Lambda$, we can assume it is generated by 1 and some complex number $\tau$. Let's example $\mathrm{End}(E)$. Suppose $\alpha\Lambda \subset \Lambda$. Then $\alpha \cdot 1 \in \Lambda$, and so $\alpha = a + b\tau$ for integers $a$ and $b$. Similarly, $\alpha\tau \in \Lambda$, and so $\alpha\tau = c + d\tau$ for integers $c$ and $d$. If $b = 0$ then $\alpha \in \mathbb{Z}$, which is not so interesting. But if $b \neq 0$ then, combining the equations, we find

$$b\tau^2 + (a-d)\tau - c = 0$$

which shows that $\tau$, and thus $\alpha$ as well, belongs to a quadratic subfield $K$ (necessarily imaginary) of $\mathbb{C}$. In fact, $\alpha$ must belong to an order of $K$, which shows that $\mathrm{End}(E)$ is an order in a quadratic imaginary field. We thus find that $\mathrm{End}(E)$ is either $\mathbb{Z}$ or an order in a quadratic imaginary field. In the latter case, we say that $E$ has complex multiplication (CM).

As an example, suppose $\Lambda$ is generated by 1 and $i$. Then multiplication by $i$ induces an endomorphism $[i]$ of $E$. In fact, $E$ is given by the equation $y^2 = x^3 + x$, and the map $[i]$ is given by $(x, y) \mapsto (-x, iy)$.

## 1.1.5 The Tate module and Weil pairing

### The Tate module of an elliptic curve

Let $E/k$ be an elliptic curve and let $n$ be an integer coprime to the characteristic of $k$. From what we have said above, it follows that $[n]$ is a separable map of degree $n^2$. Thus all fibers of the map $n \colon E(\overline{k}) \to E(\overline{k})$ have cardinality $n^2$; in other words, $E[n](\overline{k})$ has cardinality $n^2$, where $E[n]$ is the kernel of the isogeny $[n]$ (regarded as a subscheme of $E$). If $n$ is prime then this implies that $E[n](\overline{k}) = (\mathbb{Z}/n\mathbb{Z})^2$. The same conclusion holds for composite $n$ using an inductive argument involving all divisors of $n$.

Let $\ell$ be a prime number different from the characteristic. The $\ell$-adic Tate module of $E$, denoted $T_\ell(E)$, is the inverse limit of the groups $E[l^n](\overline{k})$, where the transition maps are multiplication by $\ell$. Explicitly, an element of $T_\ell(E)$ is a sequence $(x_0, x_1, \ldots)$ of $\overline{k}$-points of $E$, where $x_0 = 0$ and $\ell x_i = x_{i-1}$ for $i > 0$. The results of the previous paragraph imply that $T_\ell(E)$ is isomorphic to $\mathbb{Z}_\ell^2$.

If $k = \mathbb{C}$ and $E = \mathbb{C}/\Lambda$ then the $n$-torsion of $E$ is $\frac{1}{n}\Lambda/\Lambda$. It follows that $T_\ell(E)$ is naturally isomorphic to $\Lambda \otimes \mathbb{Z}_\ell$. Thus $T_\ell(E)$ is very close to just being $\Lambda$. Over a general field, one should think of $T_\ell(E)$ as the best possible replacement for the lattice $\Lambda$.

An extremely important property of the Tate module which cannot be seen in the complex picture is its Galois action. If $k$ is not algebraically closed then the $n$-torsion of $E$ will typically not be defined over $k$, and so the absolute Galois group $G_k = \mathrm{Gal}(\overline{k}/k)$ will move the $n$-torsion points around. This carries through the inverse limit, and so there is an action of $G_k$ on $T_\ell(E)$. Picking a basis for $T_\ell(E)$, this action can be thought of as a homomorphism $\rho\colon G_k \to \mathrm{GL}_2(\mathbb{Z}_\ell)$, i.e., an $\ell$-adic representation of the Galois group. This perspective has proved to be very useful.

### The Tate module of the multiplicative group

The multiplicative group, denoted $\mathbb{G}_m$ is the algebraic group which represents the functor $R \mapsto R^\times$ (where $R$ is a $k$-algebra). As a scheme, it is simply $\mathbb{A}^1 \setminus \{0\}$, i.e., $\mathrm{Spec}(k[t, t^{-1}])$.

The construction of the Tate module in the previous section can be applied equally well to $\mathbb{G}_m$. If $n$ is prime to the characteristic then the $n$-torsion $\mathbb{G}_m[n]$ is just the group of $n$th roots of unity; its $\overline{k}$-points is isomorphic to $\mathbb{Z}/n\mathbb{Z}$. It follows that $T_\ell(\mathbb{G}_m)$ is isomorphic to $\mathbb{Z}_\ell$ as a group. Of course, it also carries a Galois action, which can be recorded as a homomorphism $\chi\colon G_k \to \mathrm{GL}_1(\mathbb{Z}_\ell) = \mathbb{Z}_\ell^\times$. This homomorphism is called the cyclotomic character, and describes how the Galois group acts on roots of unity.

A common notation, which we will use, is to write $\mathbb{Z}_\ell(1)$ for $T_\ell(\mathbb{G}_m)$. The idea is that the underlying group is $\mathbb{Z}_\ell$ and the (1) records that the Galois group is acting through the first power of the cyclotomic character.

### Weil Pairing

In what follows, I'll abbreviate $E[n](\overline{k})$ to $E[n]$ and write $\mu_n$ for the $n$th roots of unity in $\overline{k}$.

**Proposition.** Let $E/k$ be an elliptic curve and let $n$ be prime to the characteristic. Then there exists a pairing $e_n\colon E[n] \times E[n] \to \mu_n$ satisfying the following:

(i) Bilinear: $e_n(x+y, z) = e_n(x,z)e_n(y,z)$. (Note: the group law on $E[n]$ is typically written additively, while that on $\mu_n$ is written multiplicatively.)

(ii) Alternating: $e_n(x,x) = 1$. This implies $e_n(x,y) = -e_n(y,x)$, but is stronger if $n$ is even.

(iii) Non-degenerate: if $e_n(x,y) = 1$ for all $y \in E[n]$ then $x = 0$.

(iv) Galois equivariant: $e_n(\sigma x, \sigma y) = \sigma e_n(x,y)$ for $\sigma \in G_k$.

(v) Compatibility: if $x \in E[nm]$ and $y \in E[n]$ then $e_{nm}(x,y) = e_n(mx, y)$.

The compatibility condition allows us to take the inverse limit of the $e_{\ell^n}$ to obtain a pairing on the Tate module

$$\langle , \rangle \colon T_\ell(E) \times T_\ell(E) \to \mathbb{Z}_\ell(1)$$

The properties of the Weil pairing imply that the above pairing induces an isomorphism $\bigwedge^2(T_\ell E) \cong \mathbb{Z}_\ell(1)$. In other words, if one regards the Tate module as a two dimensional Galois representation $\rho$, the Weil pairing implies that $\det(\rho) = \chi$, the cyclotomic character.
The Weil pairing has another important compatibility property:

**Proposition.** Let $f\colon E_1 \to E_2$ be an isogeny, let $x \in E_1[n]$, and let $y \in E_2[n]$. Then $e_n(f(x), y) = e_n(x, f^\vee(y))$, where $f^\vee$ is the dual isogeny.

From this, we can deduce the following useful result:

**Proposition.** Let $f\colon E \to E$ be an isogeny. Then $\deg(f) = \det(f \mid T_\ell E)$.

*Proof.* Suppose $x, y \in T_\ell(E)$. Then $\langle f(x), f(y)\rangle = \det(f)\langle x, y\rangle$. This is essentially the definition of the determinant! On the other hand, we have $\langle f(x), f(y)\rangle = \langle f^\vee(f(x)), y\rangle$, and $f^\vee f = [\deg f]$. This completes the proof. □

### 1.1.6 Elliptic curves over finite fields

A good reference for this section is Chapter V of Silverman's "The arithmetic of elliptic curves" (MR0817210).

**Point counting**

Let $E$ be an elliptic curve over the finite field $\mathbb{F}_q$. Then $E^{(q)} = E$, and so the Frobenius map $F_q$ maps $E$ to itself. A point $x$ of $E(\overline{\mathbb{F}}_q)$ belongs to $E(\mathbb{F}_q)$ if and only if it is fixed by $F_q$ (since this is equivalent to it being Galois invariant). Thus $E(\mathbb{F}_q)$ is the set of $\overline{\mathbb{F}}_q$-points of the kernel of the endomorphism $1 - F_q$. This endomorphism is separable: indeed, if $\omega$ is a differential on $E$ then $F_q^*(\omega) = 0$, and so $(1 - F_q)^*\omega = \omega$ is non-zero. We have thus proved the following proposition:

**Proposition.** $\#E(\mathbb{F}_q) = \deg(1 - F_q)$.

Recall that we have defined a positive definite bilinear pairing $\langle , \rangle$ on $\mathrm{End}(E)$, and that $\langle f, f\rangle = \deg(f)$. Appealing to the Cauchy–Schwartz inequality, we find $\langle 1, -F_q\rangle^2 \leq \deg(q)\deg(F_q) = q$, and so $\langle 1, -F_q\rangle \leq \sqrt{q}$. But, by definition,

$$2\langle 1, -F_q\rangle = \deg(1 - F_q) - \deg(1) - \deg(F_q)$$

and so we have the following theorem

**Theorem** (Hasse Bound)**.** $|\#E(\mathbb{F}_q) - q - 1| \leq 2\sqrt{q}$.

In other words, we can write $\#E(\mathbb{F}_q)$ as $q + 1 - a$, where $a$ is an error term of size at most $2\sqrt{q}$. We have $a = \langle 1, F_q\rangle$ by the above. We also have the following interpretation of $a$:

**Proposition.** We have $a = \mathrm{tr}(F_q \mid T_\ell E)$.

*Proof.* This is formal: if $A$ is any $2 \times 2$ matrix, then

$$\mathrm{tr}(A) = 1 + \det(A) - \det(1 - A)$$

Applying this to the matrix of $F_q$ on $T_\ell E$, the result follows. □

A Weil number (with respect to $q$) of weight $w$ is an algebraic number with the property that any complex embedding of it has absolute value $q^{w/2}$.

**Theorem** (Riemann Hypothesis)**.** The eigenvalues of $F_q$ on $T_\ell E$ are Weil numbers of weight 1.

*Proof.* The characteristic polynomial of $F_q$ on $T_\ell E$ is $T^2 - aT + q$. The eigenvalues are the roots of this polynomial, i.e., $(a \pm \sqrt{a^2 - 4q})/2$. The Hasse bound shows that $a^2 - 4q \leq 0$, and so the absolute value of this algebraic number (or its complex conjugate) is $\sqrt{q}$. This completes the proof. □

The zeta function of a variety $X/\mathbb{F}_q$ is defined by

$$Z_X(T) = \exp\left(\sum_{r=1}^{\infty} \#X(\mathbb{F}_q)\frac{T^r}{r}\right)$$

**Theorem** (Rationality of the zeta function)**.** We have $Z_E(T) = \frac{1-aT+qT^2}{(1-T)(1-qT)}$.

*Proof.* The above results show that

$$\#E(\mathbb{F}_{q^r}) = q^r + 1 - \operatorname{tr}(F_{q^r} \mid T_\ell E)$$

Let $\alpha$ and $\beta$ be the eigenvalues of $F_q$ on $T_\ell E$. Since $F_{q^r}$ is just $F_q^r$, the eigenvalues of $F_{q^r}$ on $T_\ell(E)$ are $\alpha^r$ and $\beta^r$. We thus see that

$$\#E(\mathbb{F}_{q^r}) = q^r + 1 - \alpha^r - \beta^r$$

We now have

$$\sum_{r=1}^{\infty} \#E(\mathbb{F}_{q^r}) \frac{T^r}{r} = -\log(1-T) - \log(1-qT) + \log(1-\alpha T) + \log(1-\beta T)$$

and so

$$Z_E(T) = \frac{(1-\alpha T)(1-\beta T)}{(1-T)(1-qT)}$$

from which the result easily follows. □

**Corollary.** $\#E(\mathbb{F}_{q^r})$ is determined, for any $r$, from $\#E(\mathbb{F}_q)$.

Suppose that $f \colon E_1 \to E_2$ is an isogeny. Then $f$ induces a map $T_\ell(E_1) \to T_\ell(E_2)$ which commutes with Frobenius. Since the kernel of $f$ is finite, the map it induces on Tate modules has finite index image; in particular, it induces an isomorphism after tensoring with $\mathbb{Q}_\ell$. It follows that the eigenvalues of Frobenius on the two Tate modules agree, and so:

**Theorem.** If $E_1$ and $E_2$ are isogenous then $\#E_1(\mathbb{F}_q) = \#E_2(\mathbb{F}_q)$.

In fact, the converse to this theorem is also true, as shown by Tate.

### Ordinary and supersingular curves

Let $E$ be an elliptic curve over a field $k$ of characteristic $p$. Then the map $[p] \colon E \to E$ is not separable and has degree $p^2$. It follows that the separable degree of $[p]$ is either $p$ or 1. In the first case, $E$ is called ordinary, and in the second case, supersingular. The following result follows immediately from the definitions, and earlier results:

**Proposition.** If $E$ is ordinary then $E[p](\overline{k}) \cong \mathbb{Z}/p\mathbb{Z}$. If $E$ is supersingular then $E[p](\overline{k}) = 0$.

We will revisit the ordinary/supesingular dichotomy after discussing group schemes. For now, we prove just one more result.

**Proposition.** If $E$ is supersingular then $j(E) \in \mathbb{F}_{p^2}$.

*Proof.* Suppose $E$ is supersingular. Then $[p]$ is completely inseparable, and thus factors as $E \to E^{(p^2)} \to E$, where the first map is the Frobenius $F_{p^2}$ and the second map is an isomorphism (since it has degree 1). Since $j(E^{(p^2)})$ is equal to $F_{p^2}(j(E))$ and $j$ is an isomorphism invariant, we see that $j(E) = F_{p^2}(j(E))$, from which the result follows. □

**Corollary.** Assume $k$ algebraically closed. Then there are only finitely many supersingular elliptic curves over $k$, and they can all be defined over $\mathbb{F}_{p^2}$.

*Proof.* An elliptic curve over an algebraically closed field descends to the field of its $j$-invariant, which gives the final statement. The finiteness statement follows immediately from this. □

### 1.1.7 Abelian varieties (analytic theory)

This section covers abelian varieties over the complex numbers from the analytic point of view. A good reference for this section is the first chapter of Mumford's "Abelian varieties" (MR0282985).

**Definition and relation to elliptic curves**

**Definition.** An abelian variety is a complete connected group variety (over some field).

**Example.** An elliptic curve is a one-dimensional abelian variety.

**Proposition.** Every one-dimensional abelian variety is an elliptic curve.

*Proof.* Let $A$ be a one-dimensional abelian variety. We must show that $A$ has genus 1. Pick a non-zero cotangent vector to $A$ at the identity. The group law on $A$ allows us to translate this vector uniquely to any other point, and so we can find a nowhere vanishing holomorphic 1-form on $A$. This provides an isomorphism $\Omega^1_A \cong \mathcal{O}_A$, and so $\mathrm{H}^0(A, \Omega^1_A)$ is one-dimensional. □

For the rest of this section we work over the complex numbers.

**Compact complex Lie groups**

Let $A$ be an abelian variety. Then $A(\mathbb{C})$ is a connected compact complex Lie group. We begin by investigating such groups. Thus let $X$ be such a group. Define $V$ to be the tangent space to $X$ at the identity (the Lie algebra). Let $g = \dim(X)$. Recall that there is a holomorphic map $\exp\colon V \to X$. We have the following results:

(i) $X$ is commutative. Reason: the map $\mathrm{Ad}\colon X \to \mathrm{End}(V)$ is holomorphic, and therefore constant, since $X$ is compact and $\mathrm{End}(V)$ is a vector space. Since Ad assumes the value 1, this is the only value it assumes. It follows that $X$ acts trivially on $\mathrm{End}(V)$, and so $V$ is a commutative Lie algebra. The result follows.

(ii) exp is a homomorphism. Reason: this follows from commutativity.

(iii) exp is surjective. Reason: the image of exp contains an open subset of $X$, since exp is a local homeomorphism. The image of exp is also a subgroup of $X$. Thus the image is an open subgroup $U$. The quotient $X/U$ is discrete, since $U$ is open, and connected, since $X$ is, and is therefore a point. Thus $X = U$.

(iv) $M = \ker(\exp)$ is a lattice in $V$, and thus isomorphic to $\mathbb{Z}^{2g}$. Reason: since exp is a local homeomorphism, $M$ is discrete. Since $X = V/M$ is compact, $M$ is cocompact.

(v) $X$ is a torus, i.e., isomorphic to a product of circles. Reason: clear from $X = V/M$.

(vi) The $n$-torsion $X[n]$ is isomorphic to $(\mathbb{Z}/n\mathbb{Z})^{2g}$. Reason: $X[n]$ is isomorphic to $\frac{1}{n}M/M$ by the exponential map.

(vii) $\mathrm{H}^i(X, \mathbb{Z})$ is naturally isomorphic to $\mathrm{Hom}(\bigwedge^i(M), \mathbb{Z})$. Reason: a simple application of the Künneth formula shows that if $T$ is any torus then cup product induces an isomorphism $\bigwedge^i(\mathrm{H}^1(T, \mathbb{Z})) \to \mathrm{H}^i(T, \mathbb{Z})$. For our torus $X$, we have $\mathrm{H}_1(X, \mathbb{Z}) = M$, and the result follows.

**Line bundles on complex tori**

Let $X = V/M$, as above. Define $\mathrm{Pic}(X)$ (the Picard group of $X$) to be the set of isomorphism classes of line bundles on $X$. This is a group under tensor product. Define $\mathrm{Pic}^0(X)$ to be the subgroup consisting of those bundles which are topologically trivial, and define $\mathrm{NS}(X)$ (the

Néron–Severi group) to be the quotient $\mathrm{Pic}(X)/\mathrm{Pic}^0(X)$. We are now going to describe how to compute these groups in terms of $V$ and $M$.

A Riemann form on $V$ (with respect to $M$) is a Hermitian form $H$ such that $E = \operatorname{im} H$ takes integer values when restricted to $M$. (Note: some people include positive definite in their definition of Riemann form; we do not.) Let $\mathcal{R}$ be the set of Riemann forms, which forms a group under addition. Let $\mathcal{P}$ be the set of pairs $(H, \alpha)$, where $H \in \mathcal{R}$ and $\alpha \colon M \to U(1)$ is a function satisfying $\alpha(x+y) = e^{i\pi E(x,y)}\alpha(x)\alpha(y)$. (Here $U(1)$ is the set of complex numbers of absolute value 1.) We give $\mathcal{P}$ the structure of a group by $(H_1, \alpha_1)(H_2, \alpha_2) = (H_1 + H_2, \alpha_1\alpha_2)$. Let $\mathcal{P}^0$ be the group of homomorphisms $M \to U(1)$, regarded as the subgroup of $\mathcal{P}$ with $H = 0$.

**Theorem** (Appell–Humbert)**.** We have an isomorphism $\mathrm{Pic}(X) \cong \mathcal{P}$, which induces isomorphisms $\mathrm{Pic}^0(X) \cong \mathcal{P}^0$ and $\mathrm{NS}(X) \cong \mathcal{R}$.

Some remarks on the theorem:

(i) Let $\pi \colon V \to X$ be the quotient map. If $L$ is a line bundle on $X$ then $\pi^*(L)$ is the trivial line bundle on $V$, since all line bundles on $V$ are trivial. Furthermore, $\pi^*(L)$ is $M$-equivariant, and $L$ can be recovered as the quotient of $\pi^*(L)$ by $M$. Thus to prove the theorem, it suffices to understand the $M$-equivariant structures on the trivial line bundle over $V$.

(ii) Let $(H, \alpha) \in \mathcal{P}$. Define an action of $M$ on $V \times \mathbb{C}$ by

$$\lambda \cdot (v, z) = (v + \lambda, \alpha(\lambda)e^{\pi H(v,\lambda) + \pi H(\lambda,\lambda)/2} z)$$

This gives the trivial bundle on $V$ an $M$-equivariance. We let $L(H, \alpha)$ be the quotient, a line bundle on $X$. The isomorphism $\mathcal{P} \to \mathrm{Pic}(X)$ is $(H, \alpha) \mapsto L(H, \alpha)$. The main content of the theorem is to show that the equivariances we just constructed are all of them.

(iii) There is a bijection between Hermitian forms $H$ on $V$ and alternating real forms $E$ satisfying $E(ix, iy) = E(x, y)$. The correspondence takes $H$ to $E = \operatorname{im} H$, and $E$ to $H(x, y) = E(ix, y) + iE(x, y)$. Thus a Riemann form $H$ is determined by the associated alternating pairing on $M$.

(iv) Let $(H, \alpha) \in \mathcal{P}$, and let $E = \operatorname{im} H$. Then $E$ defines an element of $\mathrm{Hom}(\bigwedge^2(M), \mathbb{Z})$. But we have previously identified this group with $\mathrm{H}^2(X, \mathbb{Z})$. In fact, $E$, regarded as an element of $\mathrm{H}^2$, is the Chern class $c_1(L(H, \alpha))$. We thus see that $L(H, \alpha)$ is topologically trivial if and only if $E = 0$, which is the same as $H = 0$. This gives the isomorphic $\mathrm{Pic}^0(X) \cong \mathcal{P}^0$.

Let $x \in X$ and let $t_x \colon X \to X$ be the translation-by-$x$ map, i.e., $t_x(y) = x + y$. Given a line bundle $L$ on $X$, we get a new line bundle $t_x^*(L)$ on $X$. We thus get an action of $X$ on $\mathrm{Pic}(X)$, with $x$ acting by $t_x^*$. The following proposition describes this action in terms of the Appell–Humbert theorem.

**Proposition.** We have an isomorphism $t_x^* L(H, \alpha) \cong L(H, \alpha \cdot e^{2\pi i E(x,-)})$.

A few remarks:

(i) First, we note that $\lambda \mapsto e^{2\pi i E(x,\lambda)}$ makes sense as a function on $M$, since $E$ takes integral values on $M$.

(ii) The line bundle $L(H, \alpha)$ is translation invariant (i.e., isomorphic to its pullbacks by $t_x^*$) if and only if $H = 0$. Indeed, it is clear that if $H = 0$ then $L(H, \alpha)$ is translation invariant. Conversely, if $L(H, \alpha)$ is translation invariant then $e^{2\pi i E(x,\lambda)} = 1$ for all $x \in V$ and all $\lambda \in M$, from which it easily follows that $E = 0$, and so $H = 0$ as well. We can therefore characterize $\mathrm{Pic}^0(X)$ as the group of translation invariant line bundles on $X$.

(iii) Let $L$ be a line bundle on $X$. Then $x \mapsto t_x^*(L) \otimes L^*$ defines a group homomorphism $\phi_L \colon X \to \mathrm{Pic}^0(X)$. Indeed, taking $L = L(H, \alpha)$, we see that $t_x^*(L) \otimes L^*$ is equal to $L(0, e^{2\pi i E(x,-)})$. It follows that, in fact, $\phi_L$ depends only on $c_1(L)$.

### Sections of line bundles

A $\theta$-function on $V$ with respect to $(H, \alpha) \in \mathcal{P}$ is a holomorphic function $\theta \colon V \to \mathbb{C}$ satisfying the functional equation

$$\theta(v + \lambda) = \alpha(\lambda) e^{\pi H(v,\lambda) + \pi H(\lambda,\lambda)/2}$$

Given a section $s$ of $L(H, \alpha)$ over $X$, we obtain a section $\pi^*(s)$ of $\pi^*(L(H, \alpha))$ over $V$. Identifying $\pi^*(L(H, \alpha))$ with the trivial bundle, $\pi^*(s)$ becomes a function on $V$, and the equivariance condition is exactly the above functional equation. We therefore find:

**Proposition.** The space $\Gamma(X, L(H, \alpha))$ is canonically identified with the space of $\theta$-functions for $(H, \alpha)$.

Suppose that $H$ is degenerate, and let $V_0$ be its kernel (i.e., $x \in V_0$ if $H(x, -) = 0$). Then $V_0$ is also the kernel of $E$, and since $E$ takes integral values on $M$, it follows that $M_0 = V_0 \cap M$ is a lattice in $V_0$. Let $\theta$ be a $\theta$-function, and $u$ a large element of $V_0$. Write $u = \lambda + \epsilon$ with $\lambda \in M_0$ and $\epsilon$ in some fundamental domain. Then for any $v \in V$ we have

$$|\theta(v + u)| = |\theta(v + \epsilon)|$$

since $H(\lambda, -) = 0$. It follows that $u \mapsto \theta(v + u)$ is a bounded holomorphic function on $V_0$, and therefore constant. Thus $\theta$ factors through $V/V_0$. In particular, $L(H, \alpha)$ is not ample.

Now suppose that $H(w, w) < 0$ for some $w \in V$. Let $t$ be a large complex number and write $tw = \lambda + \epsilon$, similar to the above. Then

$$|\theta(v + tw)| = |\theta(v + \epsilon)| |e^{\pi \mathrm{Re}(H(v+\epsilon,\lambda)) + \pi H(\lambda,\lambda)/2}$$

The quantity $H(\lambda, \lambda)$ is dominant, and very negative. We thus see that $|\theta(v + tw)| \to 0$ as $|t| \to \infty$, which implies $\theta(v + tw)$ is 0 as a function of $t$. Thus $\theta(v) = 0$ for all $v$, and so 0 is the only $\theta$-function.

We have thus shown that if $H$ is not positive definite then $L(H, \alpha)$ is not ample. The converse holds as well:

**Theorem** (Lefschetz)**.** The bundle $L(H, \alpha)$ is ample if and only if $H$ is positive definite.

Some remarks:

(i) This theorem shows that $X$ is a projective variety is and only if there exists a positive definite Riemann form on $V$.

(ii) In fact, one can show that if $X$ is algebraic then it is necessarily projective, and so $X$ is algebraic if and only if it has a positive definite Riemann form.

(iii) One can show that if $H$ is positive definite then $L(H, \alpha)^{\otimes n}$ is very ample for all $n \geq 3$.

(iv) Suppose $E$ is the elliptic curve given by $\mathbb{C}/\langle 1, \tau \rangle$. Then $H(x, y) = \frac{x\overline{y}}{|\mathrm{Im}(\tau)|}$ is a positive definite Riemann form on $\mathbb{C}$. This recovers the fact that all one-dimensional complex tori are algebraic.

(v) Most complex tori of higher dimension do not possess even a non-zero Riemann form, and so most are not algebraic.

### Maps of tori

A map of complex tori $X \to Y$ is a holomorphic group homomorphism. In fact, any holomorphic map taking 0 to 0 is a group homomorphism. W write $\mathrm{Hom}(X, Y)$ for the group of maps. An isogeny is a map of tori which is surjective and has finite kernel. The degree of the isogeny is the cardinality of the kernel.

**Example.** Multiplication-by-$n$, denoted $[n]$, is an isogeny of degree $n^{2g}$.

### The dual torus

Let $X = V/M$ be a complex torus. Let $\overline{V}^*$ be the vector space of conjugate-linear functions $V \to \mathbb{C}$, and let $M^\vee \subset \overline{V}^*$ be the set of such functions $f$ for which $\mathrm{im}\, f(M) \subset \mathbb{Z}$. Then $M^\vee$ is a lattice in $\overline{V}^*$, and we define $X^\vee = \overline{V}^*/M^\vee$. We call $X^\vee$ the dual torus of $X$. Note that we have a natural isomorphism $(X^\vee)^\vee = X$.

Formation of the dual torus is clearly a functor: if $f \colon X \to Y$ is a map of tori then there is a natural map $f^\vee \colon Y^\vee \to X^\vee$. If $f$ is an isogeny, then so is $f^\vee$, and they have the same degree. Even better:

**Proposition.** If $f$ is an isogeny then $\ker(f)$ and $\ker(f^\vee)$ are canonically dual (in the sense of finite abelian groups).

*Proof.* Write $X = V_1/M_1$ and $Y = V_2/M_2$, and let $g \colon V_1 \to V_2$ be the linear map inducing. Then $\ker(f) = g^{-1}(M_2)/M_1$, while $\ker(f^\vee) = (\overline{g}^*)^{-1}(M_1^\vee)/M_2^\vee$. If $x \in \ker(f)$ and $y \in \ker(f^\vee)$ then $\langle g(x), y\rangle$ is a rational number (since $g(x) \in M_2$ and $y$ is in a lattice containing $M_2^\vee$ with finite index), and is well-defined up to integers. We thus have a pairing $\ker(f) \times \ker(f^\vee) \to \mathbf{Q}/\mathbb{Z}$ with $n = \deg(f)$, which puts the two groups in duality. □

Applying this in the case where $X = Y$ and $f = [n]$, we see that $X[n]$ and $X^\vee[n]$ are in duality. This gives us a canonical pairing $X[n] \times X^\vee[n] \to \mathbb{Z}/n\mathbb{Z} \cong \mu_n$, which is called the Weil pairing.

**Proposition.** We have a natural isomorphism of groups $X^\vee = \mathrm{Pic}^0(X)$.

*Proof.* The map $\overline{V}^* \to \mathcal{P}^0$ which takes $f \in \overline{V}^*$ to the map $\lambda \mapsto e^{2\pi i\, \mathrm{im}(f(\lambda))}$ is easily seen to be a surjective homomorphism with kernel $M^\vee$. It thus descends to an isomorphism $X^\vee \to \mathrm{Pic}^0(X)$. □

Let $H$ be a Riemann form on $V$. Then $v \mapsto H(V, -)$ defines an isomorphism of complex vector spaces $V \to \overline{V}^*$, and carries $M$ into $M^\vee$. It thus defines a map $\phi_H \colon X \to X^\vee$ of complex tori. This map is an isogeny if and only if $H$ is non-degenerate. Identifying $X^\vee$ with $\mathrm{Pic}^0(X)$, $\phi_H$ coincides with $\phi_L$, where $L = L(H, \alpha)$ for any $\alpha$. A polarization of $X$ is a map of the form $\phi_H$ (or $\phi_L$) with $H$ positive-definite (or $L$ ample). A principal polarization is a polarization of degree 1. We thus see that $X$ admits a polarization if and only if it is algebraic.

The next 4 sections covers the algebraic theory of abelian varieties over arbitrary fields. I begin with the basic results such as commutativity and the structure of torsion. Then I discuss the dual abelian variety. Next I prove the weak Mordell–Weil theorem, as the same ideas will be important for us later on. Last is Poincaré irreducibility, and its interpretation in terms of the isogeny category. A good reference for today is Mumford's "Abelian varieties" (MR0282985) or Milne's notes.

### 1.1.8 General facts about abelian varieties

Fix a field $k$. Many of the results about abelian varieties over $\mathbb{C}$ continue to hold over $k$. However, the proofs are quite different and more complicated. We give some indications as to how the theory is developed, but omit most of the arguments.

**Commutativity**

We begin by explaining the most basic fact: commutativity. One can establish this using an argument similar to the one we used in the complex case. We present a different argument here, which provides a more general result. It is based on the following general fact:

**Lemma** (Rigidity lemma)**.** Let $X$ be a complete variety, let $Y$ and $Z$ be arbitrary varieties, and let $f\colon X \times Y \to Z$ be a map of varieties. Suppose there exists $x_0 \in X$ and $y_0 \in Y$ such that the restriction of $f$ to each of $X \times \{y_0\}$ and $\{x_0\} \times Y$ is constant. Then $f$ is constant.

**Corollary.** Let $X$ and $Y$ be abelian varieties and let $f\colon X \to Y$ be any map of varieties such that $f(0) = 0$. Then $f$ is a morphism of abelian varieties, i.e., it respects the group structure.

*Proof.* Consider the map $h\colon X \times X \to Y$ given by $(x, y) \mapsto f(x + y) - f(x) - f(y)$. Then $h(x, 0) = h(0, x) = 0$ for all $x \in X$, and so by the Rigidity Lemma $h = 0$, i.e., $f$ is a homomorphism. □

**Corollary.** An abelian variety is commutative.

*Proof.* The map $x \mapsto -x$ takes 0 to 0 and is therefore a homomorphism, which implies commutativity. □

**Theorem of the cube**

**Theorem** (Theorem of the cube)**.** Let $X$, $Y$, and $Z$ be varieties, with $X$ and $Y$ complete, and let $x_0 \in X$, $y_0 \in Y$, and $z_0 \in Z$ be points. Let $L$ be a line bundle on $X \times Y \times Z$, and suppose the restrictions of $L$ to each of $X \times Y \times \{z_0\}$, $X \times \{y_0\} \times Z$, and $\{x_0\} \times Y \times Z$ are trivial. Then $L$ is trivial.

**Remark.** This can be thought of as a version of the rigidity lemma for maps to the stack $\mathrm{B}\mathbb{G}_m$.

**Corollary.** Let $A$ be an abelian variety. Let $p_i\colon A \times A \times A \to A$ denote the projection map, and let $p_{ij} = p_i + p_j$ and $p_{123} = p_1 + p_2 + p_3$. Let $L$ be a line bundle on $A$. Then the line bundle

$$p_{123}^*L \otimes p_{12}^*L^{-1} \otimes p_{13}^*L^{-1} \otimes p_{23}^*L^{-1} \otimes p_1^*L \otimes p_2^*L \otimes p_3^*L$$

on $A \times A \times A$ is trivial.

*Proof.* This follows immediately from the theorem of the cube. For example, if we restrict to $A \times A \times \{0\}$ then $p_{123}^*L = p_{12}^*L$, $p_{13}^*L = p_1^*L$, and $p_3^*L = 1$, so all factors cancel. □

**Corollary.** Let $A$ be an abelian variety, let $X$ be any variety, let $f, g, h\colon X \to A$ be maps, and let $L$ be a line bundle on $A$. Then the line bundle

$$(f + g + h)^*L \otimes (f + g)^*L^{-1} \otimes (f + h)^*L^{-1} \otimes (g + h)^*L^{-1} \otimes f^*L \otimes g^*L \otimes h^*L$$

on $X$ is trivial.

*Proof.* This follows from the previous corollary by considering the map $X \to A \times A \times A$ given by $(f, g, h)$. □

**Structure of torsion**

**Proposition.** Let $L$ be a line bundle on an abelian variety $A$. Then

$$[n]^*L = L^{(n^2+n)/2} \otimes [-1]^*L^{(n^2-n)/2}$$

In particular, if $L$ is symmetric ($[-1]^*L = L$) then $[n]^*L = L^{n^2}$, while if $L$ is anti-symmetric ($[-1]^*L = L^{-1}$) then $[n]^*L = L^n$.

*Proof.* Applying the previous corollary to the maps $[n]$, $[1]$, and $[-1]$, we find

$$[n]^*L \otimes [n+1]^*L^{-1} \otimes [n-1]^*L^{-1} \otimes [n]^*L \otimes L \otimes [-1]^*L$$

is trivial. In other words,

$$[n+1]^*L = [n]^*L^2 \otimes [n-1]^*L^{-1} \otimes L \otimes [-1]^*L$$

The result now follows by induction. □

**Proposition.** The map $[n]$ is an isogeny, i.e., it is surjective with finite kernel.

*Proof.* One can show that abelian varieties are projective. Let $L$ be an ample line bundle on $A$. Replacing $L$ by $L \otimes [-1]^*L$, we can assume $L$ is symmetric. Since $[n]^*L = L^{n^2}$, it is ample. However, the restriction of this to the $n$-torsion is obviously trivial. Since the $n$-torsion is a complete variety on which the trivial bundle is ample, it must be finite. This implies that $[n]$ is surjective, by reasoning with dimension. □

**Proposition.** The degree of $[n]$ is $n^{2g}$.

*Proof.* Let $f\colon X \to Y$ be a finite map of complete varieties of degree $d$. If $D_1, \ldots, D_n$ are divisors on $Y$, where $n = \dim(X) = \dim(Y)$, then there is an equality of intersection numbers:

$$(f^*D_1 \cdots f^*D_n) = d(D_1 \cdots D_n)$$

Now, let $D$ be an ample divisor such that $[-1]^*D$ is linearly equivalent to $D$ (e.g., the divisor associated to the line bundle used above). Then $[n]^*D$ is linearly equivalent to $n^2D$. We thus find

$$\deg([n])(D \cdots D) = ((n^2D) \cdots (n^2D)) = n^{2g}(D \cdots D)$$

Since $D$ is ample, $(D \cdots D) \neq 0$, and thus $\deg([n]) = n^{2g}$. □

One can show that $[n]\colon A \to A$ induces multiplication by $n$ on the tangent space. This shows that $[n]$ is separable if and only if $n$ is prime to the characteristic. Combined with the above (and the usual induction argument), we see that:

**Corollary.** If $n$ is prime to the characteristic, then $A[n](\overline{k})$ is isomorphic to $(\mathbb{Z}/n\mathbb{Z})^{2g}$.

**Corollary.** If $\ell$ is a prime different from the characteristic then $T_\ell A$ is isomorphic to $\mathbb{Z}_\ell^{2g}$.

Since $[p]$ is not separable, $A[p](\overline{k})$ must have fewer than $p^{2g}$ points. In fact, when we study group schemes we will see that it can have at most $p^g$ points.

### Theorem of the square

**Theorem** (Theorem of the square)**.** Let $L$ be a line bundle on an abelian variety $A$, and let $x$ and $y$ be two points of $A$. Then

$$t_{x+y}^*L \otimes L = t_x^*L \otimes t_y^*L$$

Here $t_x$ denotes translation by $x$.

*Proof.* Apply the $f$, $g$, $h$ proposition with $f = x$ (constant map), $g = y$, and $h = \mathrm{id}$ □

Define $\mathrm{Pic}(A)$ to be the set of isomorphism classes of line bundles on $A$. For a line bundle $L$, let $\phi_L\colon A(k) \to \mathrm{Pic}(A)$ be the map $\phi_L(x) = t_x^*L \otimes L^{-1}$. The theorem of the square exactly states that $\phi_L$ is a group homomorphism.

### 1.1.9 The dual variety

Over the complex numbers, we can write an abelian variety $A$ as $V/M$, where $V$ is a complex vector space and $M$ is lattice. We defined the dual abelian variety $A^\vee$ as $\overline{V}^*/M^\vee$. We would like to be able to define the dual variety over any field, but this definition obviously does not carry over. The key idea is to reinterpret $A^\vee$ in terms of line bundles.

Recall that over $\mathbb{C}$ we showed that the set $A^\vee$ was canonically in bijection with the set $\mathrm{Pic}^0(A)$. Furthermore, although our definition of $\mathrm{Pic}^0(A)$ was originally topological (and does not generalize to other fields), we characterized $\mathrm{Pic}^0(A)$ as the translation invariant line bundles (which does generalize to other fields). We therefore have a possible method of defining the dual.

#### Definition of the dual

Let $k$ be an arbitrary field, and let $A$ be an abelian variety over $k$. We defined $\mathrm{Pic}(A)$ above to be the set of isomorphism classes of line bundles on $A$. We now define $\mathrm{Pic}^0(A)$ to be the subgroup consisting of those line bundles $L$ which are translation invariant, i.e., which satisfy $t_x^*(L) \cong L$ for all $x \in A$. Motivated by the complex case, we want to define $A^\vee$ to be an abelian variety with point-set $\mathrm{Pic}^0(A)$. However, it is not good enough to just define the points of a variety over a field: we must define its functor of points.

For a variety $T$, let $F(T)$ be the of isomorphism classes of line bundles $L$ on $A \times T$ satisfying the following two conditions: (a) for all $t \in T$, the restriction of $L$ to $A \times \{t\}$ belongs to $\mathrm{Pic}^0(A)$; and (b) the restriction of $L$ to $\{0\} \times T$ is trivial. Thus $F(k) = \mathrm{Pic}^0(A)$. We define the dual abelian variety $A^\vee$ to be the variety that represents $F$, if it exists. If it does exist, then it automatically comes with a universal bundle $\mathcal{P}$ on $A \times A^\vee$, which is called the Poincaré bundle.

#### Construction of the dual

Let $L$ be an ample bundle on $A$. We then have the map $\phi_L \colon A \to \mathrm{Pic}^0(A)$. (The theorem of the square implies the image is in $\mathrm{Pic}^0$.) Over the complex numbers, we saw that this map was an isogeny of tori. In general, one can prove the it is surjective, and has finite kernel $K(L)$. In fact, one can give $K(L)$ a natural scheme structure. This suggests that $A^\vee$ should be the quotient $A/K(L)$, and one can show that this is indeed the case.

#### Another approach to the dual

Let $L$ be in $\mathrm{Pic}^0(A)$. Then, by definition, $t_x^*(L)$ and $L$ are isomorphic for all $x \in A$. Choose an isomorphism $\phi_x$. Then $\phi_y t_y^*(\phi_x)$ and $\phi_{x+y}$ are two isomorphisms $t_{x+y}^*(L) \to L$, and thus differ by an element $\alpha_{x,y}$ of $\mathrm{Aut}(L) = \mathbb{G}_m$. It is obvious that $\alpha$ is a 2-cocycle of $A$ with coefficients in $\mathbb{G}_m$, and thus defines a central extension $\mathcal{G}(L)$ of $A$ by $\mathbb{G}_m$. In fact, $\mathcal{G}(L)$ is a commutative group.

Here is a different construction of $\mathcal{G}(L)$. One can show that $L$ being translation invariant is equivalent to $p_1^*L \otimes p_2^*L$ being isomorphic to $m^*L$, where $m$ is the multiplication map $A \times A \to A$ and $p_i$ are the projection maps. The fiber at $(x, y)$ of this isomorphism is an identification $L_x \otimes L_y \to L_{x+y}$. In other words, there is a natural map $L \times L \to L$ (identifying $L$ with its total space) over the multiplication map on $A$. The group $\mathcal{G}(L)$ is then just $L$ minus its zero section, with this multiplication.

We have thus constructed a map $\mathcal{G} \colon \mathrm{Pic}^0(A) \to \mathrm{Ext}^1(A, \mathbb{G}_m)$, where $\mathrm{Ext}^1$ is taken in the category of commutative group varieties. Serre showed (MR0103191, Chapter VII, Section 3) that this map is an isomorphism. Forming $\mathrm{Ext}^1$ in the category of fppf sheaves allows one to recover the functor of points of $A^\vee$.

### 1.1.10 The Mordell–Weil theorem

**Theorem** (Mordell–Weil theorem)**.** Let $A$ be an abelian variety over the number field $K$. Then $A(K)$ is a finitely generated abelian group.

The proof of this theorem usually proceeds in two steps: first, one shows that $A(K)/nA(K)$ is a finite group (the so-called weak Mordell–Weil theorem), and then one uses height functions to deduce the theorem. We will only discuss the proof of the first step. A complete proof, in the case of elliptic curves, is given in Chapter VII of Silverman's "The arithmetic of elliptic curves" (MR0817210). Consider the exact sequence

$$0 \to A[n](\overline{K}) \to A(\overline{K}) \xrightarrow{n} A(\overline{K}) \to 0.$$

Taking Galois cohomology, one obtains an exact sequence

$$0 \to A(K)/nA(K) \to \mathrm{H}^1(G_K, A[n](\overline{K})) \to \mathrm{H}^1(G_K, A(\overline{K}))[n] \to 0$$

This is called the Kummer sequence, and is very important. To show that $A(K)/nA(K)$ is finite, it suffices to show that the middle cohomology group is finite. This is not quite true; however, one can show that the image of the first map only hits classes which are unramified outside a fixed finite set of places $S$, and so it's enough to establish finiteness for such classes, which is true. (The set $S$ can be taken to be the set of places of bad reduction for $A$, together with those places above $n$.)

Let $L/K$ be a finite Galois extension containing all the $n$-torsion of $A$, and enlarge $S$ so that $L/K$ is unramified outside $S$. Then one has the inflation–restriction sequence:

$$0 \to \mathrm{H}^1(\mathrm{Gal}(L/K), A[n](\overline{K})) \to \mathrm{H}^1(G_{K,S}, A[n](\overline{K})) \to \mathrm{H}^1(G_{L,S}, A[n](\overline{K}))$$

and so to prove finiteness of the middle group it suffices to prove finiteness of the outer groups. Finiteness of the group on the left comes for free, since $\mathrm{Gal}(L/K)$ and $A[n](\overline{K})$ are both finite. Since $G_{L,S}$ acts trivially on $A[n](\overline{K})$, the right group is just $\mathrm{Hom}(G_{L,S}, \mathbb{Z}/n\mathbb{Z})^{2g}$. Giving a map $G_{L,S} \to \mathbb{Z}/n\mathbb{Z}$ is (almost) the same as giving a $\mathbb{Z}/n\mathbb{Z}$ extension of $L$ unramified outside of $S$. Since there are only finitely many such extensions unramified, the finiteness result follows.

### 1.1.11 Structure of the isogeny category

#### Poincaré reducibility

**Theorem** (Poincaré reducibility)**.** Let $A$ be an abelian variety, and let $B$ be an abelian subvariety. Then there exists an abelian subvariety $C$ such that $B \cap C$ is finite and $B \times C \to A$ is an isogeny.

*Proof.* Choosing polarizations on $A$ and $A/B$ to identify them with their duals, the dual to the quotient map $A \to A/B$ is a map $A/B \to A$. We let $C$ be its image. The properties are easy to verify □

We say that an abelian variety $A$ is simple if the only abelian subvarieties are 0 and $A$.

**Corollary.** Every abelian variety is isogenous to a product of simple varieties.

#### The isogeny category

Define a category Isog as follows. The objects are abelian varieties. For two abelian varieties $A$ and $B$, we put $\mathrm{Hom}_{\mathrm{Isog}}(A, B) = \mathrm{Hom}(A, B) \otimes \mathbb{Q}$. One can show that if $f\colon A \to B$ is an isogeny then there exists an isogeny $g\colon B \to A$ such that $gf = [n]$, for some $n$; it follows that $\frac{1}{n}g$ is the inverse to $f$ in Isog. Thus isogenies become isomorphisms in Isog.

It is not difficult to see that Isog is in fact an abelian category. The simple objects of this category are exactly the simple abelian varieties. Poincaré's theorem shows that Isog is semi-simple as an abelian category.

From this formalism, and general facts about abelian varieties, we deduce two results:

(i) The decomposition (up to isogeny) into a product of simple abelian varieties is unique (up to isogeny). (Reason: in any semi-simple abelian category, the decomposition into simples is unique up to isomorphism.)

(ii) If $A$ is a simple abelian variety then $\mathrm{End}(A) \otimes \mathbb{Q}$ is a division algebra over $\mathbb{Q}$. (Reason: if $A$ is a simple object in an abelian category and $\mathrm{End}(A)$ contains a field $k$, then it is a division algebra over $k$.)

## 1.2 Group schemes, over fields and DVRs, including Raynaud's theorem

This is the section on group schemes. I begin by introducing the idea of a group object in a category. I then define what a group scheme is, and explain the connection to Hopf algebras. This is followed by several important examples. For the rest of the section, I focus on finite commutative group schemes over fields, and cover most of the basic facts (existence of quotients, classification in the étale case, the connected-étale sequence, etc.). A good reference for today is Tate's article "Finite flat group schemes" in the book "Modular forms and Fermat's last theorem" (MR1638478).

### 1.2.1 Group objects in a category

**Group objects**

Let $\mathcal{C}$ be a category with all finite products; denote the final object by pt. A group object in $\mathcal{C}$ is a tuple $(G, m, i, e)$, where:

(i) $G$ is an object of $\mathcal{C}$

(ii) $m$ is a map $G \times G \to G$, the multiplication map

(iii) $i$ is a map $G \to G$, the inversion map; and

(iv) $e$ is a map $\mathrm{pt} \to G$, the identity section (or identity element),

such that the usual axioms of group theory hold:

(i) Associativity: the following diagram commutes:

$$\begin{tikzcd}
G \times G \times G \arrow[r, "m\times\mathrm{id}"] \arrow[d, "\mathrm{id}\times m"'] & G \times G \arrow[d, "m"] \\
G \times G \arrow[r, "m"] & G
\end{tikzcd}$$

(ii) Identity element: the following composition is equal to the identity:

$$\begin{tikzcd}
G \arrow[r, equal] & G \times \mathrm{pt} \arrow[r, "\mathrm{id}\times e"] & G \times G \arrow[r, "m"] & G
\end{tikzcd}$$

And similarly if $\mathrm{id} \times e$ is changed to $e \times \mathrm{id}$.

(iii) Inverses: the following diagram commutes:

$$\begin{tikzcd}
G \arrow[r, "\mathrm{id}\times i"] \arrow[d] & G \times G \arrow[d, "m"] \\
\mathrm{pt} \arrow[r, "e"] & G
\end{tikzcd}$$

And similarly if $\mathrm{id} \times i$ is changed to $i \times \mathrm{id}$.

We say that a group object $G$ is commutative if the following diagram commutes:

$$\begin{tikzcd}
G \times G \arrow[r, "\tau"] \arrow[d, "m"'] & G \times G \arrow[d, "m"] \\
G \arrow[r, equal] & G
\end{tikzcd}$$

Here $\tau$ is the switching-of-factors map. Suppose $G$ and $H$ are group objects. A homomorphism from $G$ to $H$ is a morphism $G \to H$ in $\mathcal{C}$ such that all the relevant diagrams commute. In this way, there is a category of group objects in $\mathcal{C}$.

**Example.** An abelian variety is, by definition, a group object in the category of complete varieties.

**Functor of points**

Let $X$ be an object of $\mathcal{C}$. For an object $Y$ of $\mathcal{C}$, let $h_X(Y) = \mathrm{Hom}_{\mathcal{C}}(Y, X)$. Then $h_X$ defines a contravariant functor from $\mathcal{C}$ to the category of sets. Yoneda's lemma says that $X$ is determined from $h_X$, in a precise sense.

Suppose $G$ is a group object of $\mathcal{C}$. One then verifies that $h_G(Y)$ inherits the structure of a group; for example, the multiplication map $h_G(Y) \times h_G(Y) \to h_G(Y)$ is induced from $m$. Furthermore, if $f \colon Y \to Y'$ is a morphism in $\mathcal{C}$, then the induced map $f^* \colon h_G(Y') \to h_G(Y)$ is a group homomorphism. Conversely, if $G$ is an object of $\mathcal{C}$ such that each $h_G(Y)$ is endowed with the structure of a group and each $f^*$ is a group homomorphism then $G$ naturally has the structure of a group object of $\mathcal{C}$. In other words, giving $G$ a group structure is the same as lifting $h_G$ to a functor from $\mathcal{C}$ to the category of groups. Said yet again, the Yoneda embedding is an equivalence between group objects in $\mathcal{C}$ and group objects in the functor category $\mathrm{Fun}(\mathcal{C}, \mathrm{Set})$ which are representable. This point of view allows one to define group objects even if $\mathcal{C}$ doesn't have finite products.

The group object $G$ is commutative if and only if $h_G(Y)$ is commutative for all $Y$.

**Kernels and cokernels**

The category of group objects in $\mathcal{C}$ has a zero object 1, namely the object pt endowed with its unique group structure. One therefore has a definition for kernel and cokernel of a map of group objects, namely fiber (co)product with the zero object (if it exists).

Let $f \colon G \to H$ be a homomorphism of group objects. Since $\ker(f)$ is defined by what maps to it look like, one has a good description of its functor of points: $(\ker f)(T)$ is equal to the kernel of the map $f \colon G(T) \to H(T)$. In contrast, $\mathrm{coker}(f)$ is defined by what maps out of it look like, and so its functor of points does not admit an easy description in general. In particular, it is not true that $(\mathrm{coker}\, f)(T)$ is the cokernel of $f \colon G(T) \to H(T)$.

### 1.2.2 Group schemes

**The connection with Hopf algebras**

Fix a field $k$. The category of affine schemes over $k$ is anti-equivalent to the category of $k$-algebras. One therefore finds that an affine group scheme $G$ over $k$ correspond to a $k$-algebra $A$ equipped with all the structure of a group object, but with the arrows going in the opposite direction:

(i) The multiplication map $G \times G \to G$ turns into a comultiplication map $A \to A \otimes A$.

(ii) The inversion map $G \to G$ turns into the antipode map $A \to A$.

(iii) The identity section pt $\to G$ turns into the counit $A \to k$.

One can furthermore translate the group axioms: for example, associativity of $G$ means the comultiplication on $A$ is coassociative.

A (commutative) $k$-algebra $A$ equipped with a comultiplication, counit, and antipode satisfying the necessary axioms is called a (commutative) Hopf algebra. It is revealing to think of a Hopf algebra not as an algebra with comultiplication, counit, and antipode, but as a vector space with multiplication, unit, comultiplication, counit, and antipode. In this way, the data becomes completely symmetric with respecting to flipping all the arrows.

### Examples

In what follows, we write $T = \mathrm{Spec}(R)$ for a test scheme.

- *The additive group.* Let $\mathbb{G}_a = \mathrm{Spec}(k[t])$. We have $\mathrm{Hom}(T, \mathbb{G}_a) = R$. Regarding $R$ as an additive group, this shows that $\mathbb{G}_a$ naturally has the structure of a commutative group scheme. It is called the additive group. The comultiplication on $k[t]$ is given by $t \mapsto t \otimes 1 + 1 \otimes t$.

- *The multiplicative group.* Let $\mathbb{G}_m = \mathrm{Spec}(k[t, t^{-1}])$. We have $\mathrm{Hom}(T, \mathbb{G}_m) = R^\times$, the group of units in $R$. Again, this shows that $\mathbb{G}_m$ naturally has the structure of a commutative group scheme. The comultiplication is given by $t \mapsto t \otimes t$.

- *The constant group.* Let $\Gamma_0$ be an ordinary group. Let $\Gamma$ be the disjoint union of $\mathrm{Spec}(k)$'s indexed by $\Gamma_0$. We have $\mathrm{Hom}(T, \Gamma) = \mathrm{Hom}(\pi_0(T), \Gamma_0)$, which is a group; therefore $\Gamma$ is a group scheme, which we call the constant group scheme on $\Gamma_0$. In fact, $\Gamma = \mathrm{Spec}(A)$, where $A$ is the ring of functions $\Gamma_0 \to k$. We can identify $A \otimes A$ with the ring of functions $\Gamma_0 \times \Gamma_0 \to k$, and then comultiplication takes a function $f$ to the function $(x, y) \mapsto f(xy)$. In the future, we do not notationally distinguish between $\Gamma$ and $\Gamma_0$.

- *Roots of unity.* Let $\mu_n = \mathrm{Spec}(k[t]/(t^n - 1))$. We have that $\mathrm{Hom}(T, \mu_n)$ equal to the set of elements $x \in R$ such that $x^n = 1$. This is obviously a commutative group under multiplication, and so $\mu_n$ is a commutative group scheme. It is the kernel of the multiplication-by-$n$ map on $\mathbb{G}_m$.

- *The group scheme $\alpha_p$.* Assume $k$ has characteristic $p$. Let $\alpha_p = \mathrm{Spec}(k[t]/(t^p))$. The set $\mathrm{Hom}(T, \alpha_p)$ is identified with the set of elements $x \in R$ which satisfy $x^p = 0$. Since $k$ has characteristic $p$, this is a group under addition. It follows that $\alpha_p$ is a commutative group scheme.

**Remark.** The schemes $\alpha_p$ and $\mu_p$ are isomorphic as schemes, but not as group schemes.

### Quotients

We are chiefly interested in finite commutative group schemes over $k$. Note that finite schemes are always affine, so such group schemes are described by finite dimensional commutative and cocommutative Hopf algebras. Examples include the constant group scheme on a finite group, $\mu_n$, and $\alpha_p$. We define the order of such a group scheme $G$, denoted $\#G$, to be the dimension of its Hopf algebra.

We state without proof the following theorem, first proved by Grothendieck.

**Theorem.** Let $G$ be a finite commutative group scheme over $k$ and let $H$ be a closed subgroup.

(i) Then the quotient $G/H$ exists, and is a finite group scheme over $k$.

(ii) The functor $h_{G/H}$ is the quotient of $h_G$ by $h_H$ in the category of sheaves (on the big fppf site over $k$). In other words, $h_{G/H}$ is the sheafification of the presheaf $T \mapsto G(T)/H(T)$.

(iii) We have $\#(G/H) = \#G/\#H$.

Proving part 1 is not difficult: it simply amounts to showing that kernels exist in the category of Hopf algebras, which can be checked explicitly. Parts 2 and 3 are more difficult.

The theorem shows that the category of finite commutative group schemes over $k$ is an abelian category.

**The étale case**

We now study the case where $G$ is étale over $k$. Recall that a finite dimensional $k$-algebra is étale if and only if it is a product of separable extensions of $k$; when $k$ has characteristic 0, this is equivalent to being reduced.

Let $A$ be an étale $k$-algebra and let $k^s$ be the separable closure of $k$. Then $A \otimes k^s$ is a finite product of copies of $k^s$ indexed by some set $I$. The Galois group $G_k$ naturally permutes the set $I$. We have thus defined a functor.

$$\Phi\colon \{\text{finite étale } k\text{-algebra}\} \to \{\text{finite } G_k\text{-sets}\}$$

We note that $\Phi(A) = X(k^s)$, where $X = \operatorname{Spec}(A)$.

Now let $I$ be a finite $G_k$-set. Let $\overline{A} = \prod_{i \in I} k^s$. Then $G_k$ naturally acts on $\overline{A}$, through its action on both $I$ and $k^s$. Let $A$ be the invariant subalgebra. One easily sees that $A$ is a finite dimensional algebra and étale over $k$. We thus have a functor

$$\Psi\colon \{\text{finite } G_k\text{-sets}\} \to \{\text{finite étale } k\text{-algebras}\}$$

We have the following basic result:

**Theorem.** The functors $\Phi$ and $\Psi$ are naturally quasi-inverse.

**Corollary.** The functor

$$\{\text{finite étale schemes over } k\} \to \{\text{finite } G_k\text{-sets}\}$$

given by $X \mapsto X(k^s)$ is an equivalence.

**Corollary.** The functor

$$\{\text{finite étale commutative group schemes over } k\} \to \{\text{finite } G_k\text{-modules}\}$$

given by $G \mapsto G(k^s)$ is an equivalence.

We thus see that the study of étale group schemes is equivalent to the study of Galois representations.

**The connected-étale sequence**

Let $G = \operatorname{Spec}(A)$ be a finite commutative group scheme over $k$. Write $A = \prod A_i$ with each $A_i$ local. There is a unique index, denoted 0, such that the counit of $A$ factors through $A_0$. Let $G^\circ = \operatorname{Spec}(A_0)$, a connected closed subscheme of $G$. Since $G^\circ$ has a $k$-point, it is geometrically connected, and so $G^\circ \times G^\circ$ is still connected; it follows that multiplication maps $G^\circ \times G^\circ$ into $G^\circ$, from which one easily sees that $G^\circ$ is a subgroup of $G$. We call $G^\circ$ the identity component of $G$.

Let $A_{\mathrm{et}}$ be the maximal étale subalgebra of $A$ (fix!). Concretely, $A_{\mathrm{et}} = \prod (A_i)_{\mathrm{et}}$, where $(A_i)_{\mathrm{et}}$ is the separable closure of $k$ in $A_i$. Put $G^{\mathrm{et}} = \operatorname{Spec}(A^{\mathrm{et}})$. Formation of $A^{\mathrm{et}}$ respects tensor products, and so if $G$ is a group scheme then so is $G^{\mathrm{et}}$, and the natural map $G \to G^{\mathrm{et}}$ is a homomorphism. The universal property of $A^{\mathrm{et}}$ implies the following: a map from $G$ to an étale group scheme factors uniquely through $G^{\mathrm{et}}$. Note that the natural map $G(\overline{k}) \to G^{\mathrm{et}}(\overline{k})$ is an isomorphism.

The tensor product $A \otimes_{A_{\mathrm{et}}} k$ (where the map $A_{\mathrm{et}} \to k$ is the counit) is the universal quotient of $A$ in which the idempotent defining $A_0$ is identified with 1, and is therefore equal to $A_0$. In other words, $G^\circ$ is the fiber product of $G$ with the trivial group over $G^{\mathrm{et}}$. We have thus proved the sequence

$$0 \to G^\circ \to G \to G^{\mathrm{et}} \to 0$$

is exact. This sequence is called the connected-étale sequence.

Suppose now that $k$ is perfect, i.e., every finite extension of $k$ is separable. Then the separable closure of $k$ in $A_i$ coincides with the algebraic closure, and maps isomorphically onto the residue field of $A_i$. It follows that the map $G^{\mathrm{red}} \to G^{\mathrm{et}}$ is an isomorphism of schemes. Since $k$ is perfect, a product of reduced schemes is still reduced, and so $G^{\mathrm{red}}$ is a closed subgroup of $G$. We have thus shown that, in this case, the connected-étale sequence splits. Furthermore, since there are no non-zero maps from an étale group scheme to a connected group scheme, the splitting is canonical. In other words, we have a canonical decomposition $G = G^\circ \times G^{\mathrm{et}}$.

**Example.** We now give an example where the connected-étale sequence does not split. This is not the most elementary example, but it is a natural one. Let $X$ be some moduli scheme of elliptic curves over $\mathbb{F}_p$ (say $X_0(N)$), and let $\mathcal{E} \to X$ be the universal family of elliptic curves. Let $k$ be the function field of $X$ (which is not perfect), and let $E/k$ be the generic fiber of $\mathcal{E}$. Then $E$ is not defined over the $\overline{\mathbb{F}}_p$: indeed, the function $j\colon X \to \mathbb{P}^1$ is not constant, and so $j \in k$, which is the $j$-invariant of $E$, is transcendental over $\mathbb{F}_p$. In particular, $E$ is ordinary. Let $G_n = E[p^n]$, a finite commutative group scheme over $k$. Since $E$ is ordinary, $G_n(\overline{k})$ is non-zero, and so $G_n^{\mathrm{et}}$ is non-trivial. Since $G_n$ is not étale, $G_n^\circ$ must also be non-trivial. We claim that the connected-étale sequence for $G_n$ is non-split, for $n$ large enough. Indeed, suppose to the contrary it split for all $n$. Then we have a decomposition $G_\infty = G_\infty^{\mathrm{et}} \times G_\infty^\circ$ of $p$-divisible groups. It follows that $\mathrm{End}(G_\infty) = \mathbb{Z}_p \oplus \mathbb{Z}_p$. Thus, by the Tate conjecture (a theorem in this case), $\mathrm{End}(E)$ has rank two over $\mathbb{Z}$, and thus $E$ has CM. But this implies $E$ is defined over $\overline{\mathbb{F}}_p$, a contradiction.

### Order invertible implies étale

Let $G = \mathrm{Spec}(A)$ be a finite connected commutative group scheme, so $A$ is a local ring. Let $I \subset A$ be the kernel of the counit map. Then $A = k \oplus I$, where $k$ is the span of the unit. We let $\pi\colon A \to I/I^2$ be the projection map, which is easily seen to be a derivation. Let $x_1, \ldots, x_n$ be elements of $I$ mapping to a basis for $I/I^2$. Define $D_i\colon A \to A$ to be the composition

$$A \to A \otimes A \to A \otimes I/I^2 \to A$$

where the first map is comultiplication, the second is $\mathrm{id} \otimes \pi$, and the third is induced from the map $I/I^2 \to k$ taking $x_i$ to 1 and $x_j$ to 0 for $i \neq j$. This is easily seen to be a derivation.

**Proposition.** Suppose

(i) $k$ has characteristic 0. Then the natural map $\varphi\colon k[x_i] \to A$ is an isomorphism.

(ii) $k$ has characteristic $p$ and $x_i^p = 0$ for all $i$. Then the natural map $\varphi\colon k[x_i]/(x_i^p) \to A$ is an isomorphism.

*Proof.* Clearly, $\varphi$ is surjective. In each case, one has $\varphi \frac{\partial}{\partial x_i} = D_i \varphi$, since both derivations agree on the $x_i$. This implies $\ker(\varphi)$ is stable by $\frac{\partial}{\partial x_i}$, which, by induction on degree, implies that it is either 0 or the unit ideal. Since it is not the unit ideal, we conclude that $\varphi$ is injective. □

**Corollary.** If $k$ has characteristic 0 then $G$ is trivial.

*Proof.* he proposition shows that $G$ is isomorphic to affine $n$-space for some $n$. By finiteness, $n = 0$, which establishes the corollary. □

**Corollary.** If $k$ has characteristic $p$ then $\#G$ is a power of $p$.

*Proof.* Let $G_1$ be the kernel of the Frobenius map $F_p \colon G \to G^{(p)}$, which is a group homomorphism, and let $G_2 = G/G_1$. The proposition shows that $G_1$ is isomorphic to $p^n$, where $n = \dim(I/I^2)$. The result now follows from induction, since $\#G = (\#G_1)(\#G_2)$. □

**Corollary.** Suppose $G$ is a finite commutative group scheme such that $\#G$ is invertible in $k$. Then $G$ is étale.

Now, I will discuss the second part of group schemes. In the first, I discuss Cartier duality and give the basic examples. In the second, I go deeper into the theory over finite fields, giving the classification of height 1 group schemes, and using it to classify the simple group schemes over an algebraically closed field. I also briefly discuss Dieudonné modules. In the third section, I apply the theory of group schemes to the of study abelian varieties. First, I relate abelian variety duality and Cartier duality. Then I characterize ordinary and supersingular elliptic curves using their $p$-torison. Finally, I give a tight bound on the $p$-torsion of an abelian variety.

### 1.2.3 Cartier duality

Let $G = \mathrm{Spec}(A)$ be a finite commutative group scheme over the field $k$. Let $A^*$ be the $k$-linear dual of $A$. Since the Hopf algebra axioms are completely symmetric with respect to reversing arrows, $A^*$ is still a Hopf algebra, and, of course, both commutative and cocommutative. We let $G^\vee = \mathrm{Spec}(A^*)$, a finite commutative group scheme over $k$. We call $G^\vee$ the Cartier dual of $G$. Obviously, $G$ and $G^\vee$ have the same order and the natural map $(G^\vee)^\vee \to G$ is an isomorphism.

We now describe the functor of points of $G^\vee$. Let $R$ be a $k$-algebra. Giving a $k$-algebra homomorphism $A^* \to R$ is the same as giving an $R$-algebra homomorphism $A^*_R \to R$, where $(-)_R = - \otimes R$. Taking $R$-linear duals, this is the same as giving a morphism $R$-coalgebras $R \to A_R$ of $R$-coalgebras. This, in turn, is the same as giving an element $x$ of $A_R$ such that $\Delta(x) = x \otimes x$ and $\eta(x) = 1$, where $\Delta$ is the comultiplication and $\eta$ is the counit on $A_R$. (The bijection takes a map $R \to A_R$ to the image of $1 \in R$. The condition on $x$ is exactly the condition needed to make the map one of coalgebras.) We note that one of the Hopf algebra axioms is $m(\mathrm{id} \otimes i)\Delta = \eta$, and so $xi(x) = 1$, i.e., $x$ is a unit of $A_R$. But giving a unit of $A_R$ is the same as giving a map $R[t, t^{-1}] \to A_R$, and the condition $\Delta(x) = x \otimes x$ exactly makes this map one of Hopf algebras! We have thus shown:

**Proposition.** There is a natural bijection $G^\vee(R) = \mathrm{Hom}(G_R, (\mathbb{G}_m)_R)$. Equivalently, $G^\vee = \underline{\mathrm{Hom}}(G, \mathbb{G}_m)$ as sheaves on the big fppf site.

In words: the $R$-points of $G^\vee$ are the characters of $G$ defined over $R$.

**Example.** Suppose $G = \mathbb{Z}/r\mathbb{Z}$. Then $A = \prod_{i \in \mathbb{Z}/r\mathbb{Z}} ke_i$, with multiplication $e_i e_j = \delta_{ij} e_i$ and comultiplication $\Delta(e_n) = \sum_{i+j=n} e_i e_j$. Let $e_i^*$ be the dual basis of $A^*$. The element $\Delta(e_n^*)$ is the linear functional $A \otimes A \to A \to k$, where the first map is multiplication and the second is $e_n^*$. Given the formula for multiplication, we see that $\Delta(e_n^*)(e_i \otimes e_j)$ is 1 if $i = j = n$ and 0 otherwise. Thus $\Delta(e_n^*) = e_n^* \otimes e_n^*$. The product $e_i^* e_j^*$ is the linear functional $A \to A \otimes A \to k$, where the first map is comultiplication and the second is $e_i * \otimes e_j^*$. We thus see that $(e_i^* e_j^*)(e_n)$ is 1 if $i + j = n$ and 0 otherwise. Thus $e_i^* e_j^* = e_{i+j}^*$. It follows that $e_i^* \mapsto t^i$ defines an isomorphism of Hopf algebras $A^* \to k[t]/(t^r - 1)$. Thus $(\mathbb{Z}/r\mathbb{Z})^\vee = \mu_r$. (This can be seen more conceptually using the description of the functor of points of $G^\vee$.) Of course, this gives $\mu_r^\vee = \mathbb{Z}/r\mathbb{Z}$ as well.

**Example.** Let's now consider the case $G = \alpha_p$, where $k$ has characteristic $p$. So $A = k[t]/(t^p)$. Let $e_i = t^i$ for $0 \le i < p$. We have $e_i e_j = e_{i+j}$ for $i + j < n$ and $e_i e_j = 0$ for $i + j \ge n$. We have $\Delta(t) = t \otimes 1 + 1 \otimes t$. Since $\Delta$ is a ring homomorphism,

$$\Delta(e_n) = (t \otimes 1 + 1 \otimes t)^n = \sum_{i+j=n} \binom{n}{i} t^i \otimes t^j$$

Now, let $e_i^*$ be the dual basis of $A^*$. Then $\Delta(e_n^*)(e_i \otimes e_j)$ is equal to 1 if $i + j = n$ and 0 otherwise. Thus $\Delta(e_n^*) = \sum_{i+j=n} e_i^* \otimes e_j^*$. We have that $(e_i^* e_j^*)(e_n)$ is equal to $\binom{n}{i}$ if $n = i + j$ and 0 otherwise; thus $e_i^* e_j^* = \binom{i+j}{i} e_{i+j}^*$ if $i + j < n$, and 0 otherwise. It follows that $e_i^* \mapsto t^i/i!$ defines an isomorphism of Hopf algebras $A^* \to A$. Thus $\alpha_p^\vee = \alpha_p$.

We have already seen two fundamental types of finite commutative group schemes: the local (i.e., connected) ones, and the étale ones. Cartier duality allows us to further refine these types by considering the type of the dual as well. We thus have local–local, local–étale, étale–local, and étale–étale group schemes (the first refers to the type of $G$, the second to $G^\vee$). Examples (in characteristic $p$) of these are $\alpha_p$, $\mu_p$, $\mathbb{Z}/p\mathbb{Z}$, and $\mathbb{Z}/\ell\mathbb{Z}$ (for $\ell \neq p$). In characteristic 0, we only have étale–étale. Over a perfect field, every finite group scheme canonically decomposes into a product $G_{ll} \times G_{le} \times G_{el} \times G_{ee}$, where $G_{ll}$ is local–local, etc.

### 1.2.4 Group schemes over finite fields

#### Frobenius and Verschiebung

Let $G = \mathrm{Spec}(A)$. We have already seen the Frobenius map $F_p \colon G \to G^{(p)}$, though we have not carefully defined it. Let $\sigma$ be the $p$th power map (on any ring of characteristic $p$), the so-called absolute Frobenius. Then $\sigma(\alpha x) = \sigma(\alpha)\sigma(x)$ for $\alpha \in k$ and $x \in A$, and so $\sigma \colon A \to A$ is not a homomorphism of $k$-algebras. Let $A^{(p)} = k \otimes_{k,\sigma} A$, i.e., $\alpha \otimes x = 1 \otimes \alpha^p x$ in $A^{(p)}$. Then the map $A^{(p)} \to A$ given by $\alpha \otimes x \mapsto \alpha x^p$ is a well-defined map of $k$-algebras; this is $F_p$. We define $F_q$, for $q = p^r$, by using $\sigma^r$ in place of $\sigma$.

**Proposition.** *$G$ is étale if and only if $F_p$ is an isomorphism, and connected if and only if $F_q = 0$ for some $q = p^r$.*

*Proof.* Suppose $G = \mathrm{Spec}(A)$ is connected. Then clearly $F_q = 0$ on $A$ for some $A$, since the maximal ideal of $A$ is nilpotent. If $F_p$ is an isomorphism on $G$ then $F_p$ is an isomorphism on $G^\circ$ as well, and so $G^\circ = 0$, i.e., $G$ is étale. Now, $F_q$ induces an isomorphism $G(\overline{k}) \to G^{(q)}(\overline{k})$. Thus if $F_q = 0$ for some $q$ then $G(\overline{k}) = 0$, i.e., $G^{\mathrm{et}} = 0$, and so $G$ is connected. □

The dual of the Frobenius map on $G^\vee$ is a map $V_p \colon G^{(p)} \to G$, called the Verschiebung map. It is a homomorphism, and one can show $F_p V_p = V_p F_p = [p]$. Obviously, $V_p$ is an isomorphism if and only if $G^\vee$ is étale and $V_q = 0$ for some $q$ if and only if $G^\vee = 0$.

#### Classification in height 1

Let $G = \mathrm{Spec}(A)$ be a finite commutative connected group scheme over $k$, which we assume to have characteristic $p$. Write $L(G)$ for the Lie algebra of $G$, which is naturally the dual of the $k$-vector space $I/I^2$, where $I$ is the maximal ideal of $A$.

We say that a derivation $D \colon A \to A$ is invariant if $\Delta D = (D \otimes 1)\Delta$, where $\Delta$ is comultiplication. Given $v \in L(G)$, thought of as an element of $(I/I^2)^*$, let $D_v$ be the composition,

$$A \to A \otimes A \to A \otimes I/I^2 \to A$$

where the first map is $\Delta$, the second is $\mathrm{id} \otimes \pi$, where $\pi$ is projection onto $I/I^2$ (as discussed previously), and the final map is $\mathrm{id} \otimes v$. Then one easily verifies that $v \mapsto D_v$ is an isomorphism of $L(G)$ with the space of invariant derivations.

Let $D$ be a derivation of $A$. Let $D^n$ be the $n$-fold iterate of $D$ on $A$. Since $D^n(xy)$ is computed using the binomial theorem, it follows that $D^p$ satisfies the Liebniz rule, and is therefore a derivation. It's easy to verify that if $D$ is invariant then so is $D^p$, and so $D \mapsto D^p$ induces a map $L(G) \to L(G)$ which we denote by $F$. Note that $F(av) = a^p F(v)$ for $a \in k$.

We define an $F$-module over $k$ to be a $k$-vector space $L$ equipped with an additive map $F$ satisfying $F(av) = a^p F(v)$ for $a \in k$ and $v \in V$. Thus $L(G)$ is an example of an $F$-module.

**Example.** Suppose $G = \alpha_p = \mathrm{Spec}(k[t]/(t^p))$. One easily verifies that $D = \frac{d}{dt}$ is an invariant differential and spans $L(G)$. We have $D^p = 0$ since $D^p(t) = 0$ and $t$ generates. Thus $L(G) = k$, with $F = 0$.

**Example.** Suppose $G = \mu_p = \mathrm{Spec}(k[t]/(t^p - 1))$. One easily verifies that $D = t\frac{d}{dt}$ is an invariant differential and spans $L(G)$. We have $D^p = D$ since $D^p(t) = t = D(t)$. Thus $L(G) = k$, with $F$ being the $p$th power map.

It is clear that $G$ is not determined from $L(G)$, for two reasons: (1) in the étale case, $L(G) = 0$; and (2) a non-isomorphism of groups can induce an isomorphism on tangent spaces, e.g., the map $\mu_{p^2} \to \mu_p$. We say that $G$ has height 1 if it is connected and killed by Frobenius. This hypothesis eliminate the two obvious obstructions just mentioned. In fact:

**Theorem.** The functor $G \mapsto L(G)$ is an equivalence between the category finite commutative height 1 group schemes over $k$ and the category of finite dimensional $F$-modules.

*Proof.* Let $L$ be a finite dimensional $F$-module. Let $A$ be the quotient of $\mathrm{Sym}(L)$ by the ideal generated by $x^p - F(x)$ for $x \in L$. Then $A$ is obviously a finite dimensional $k$-algebra. One verifies that $x \mapsto x \otimes 1 + 1 \otimes x$ for $x \in L$ descends to a comultiplication on $A$, and that $A$ is naturally a Hopf algebra. The inverse functor takes $L$ to $\mathrm{Spec}(A^*)$, the Cartier dual of $\mathrm{Spec}(A)$. For details, see Mumford's "Abelian varieties" (MR0282985), section 14. □

### Consequences of classification

**Theorem.** Suppose $k$ is algebraically closed. Then $L(\alpha_p)$ and $L(\mu_p)$ are the only simple objects in the category of finite dimensional $F$-modules (up to isomorphism).

*Proof.* Let $L$ be an $F$-module. If there exists $x \in L$ non-zero such that $F(x) = 0$ then $kx$ is a non-trivial submodule of $L$ isomorphic to $L(\alpha_p)$. Now suppose $F(x) \neq 0$ for all $x \neq 0$; we must show $L$ contains $L(\mu_p)$.

Let $e_1, \dots, e_n$ be a basis of $L$. Identify an element $x = \sum a_i e_i$ of $L$ with the vector $v = (a_i)$. Write $F(e_i) = \sum_j C_{ij} e_j$ with $C_{ij} \in k$, and let $C$ be the matrix $(C_{ij})$. Then if $x$ corresponds to the vector $v = (a_i)$, we see that $F(x)$ corresponds to the vector $Cv^p$, where $v^p = (a_i^p)$. From this we see that $C$ is invertible: indeed, if $Cv = 0$ then $F(x) = 0$, where $x$ corresponds to the vector $v^{1/p}$ (which exists since $k$ is closed).

We thus see that $F$-fixed vectors of $L$ correspond to solutions to the equation $v^p = C^{-1}v$. Let $R = k[x_i]/(x_i^p = \sum_j C_{ij}^{-1} x_j)$. Then $F$-fixed vectors exactly coincide with $k$-points of $\mathrm{Spec}(R)$. Note that $\Omega^1_{R/k} = 0$ and $R$ has dimension $p^r$ over $k$. It follows that $\mathrm{Spec}(R)$ is finite étale over $k$, and thus has exactly $p^r$ $k$-points, since $k$ is closed. We have therefore shown that $\dim_{\mathbb{F}_p}(L^{F=1}) = \dim_k(L)$. It is easy to see that the natural map $L^{F=1} \otimes_{\mathbb{F}_p} k \to L$ is injective (apply $F$ to a hypothetical minimal linear dependence), which implies that $L \cong L(\mu_p)^{\oplus n}$. □

**Theorem.** Suppose $k$ is algebraically closed. Then the simple finite commutative group schemes over $k$ are $\mathbb{Z}/\ell\mathbb{Z}$ ($\ell \neq p$ prime), $\mathbb{Z}/p\mathbb{Z}$, $\mu_p$, and $\alpha_p$

*Proof.* A simple group scheme is either connected or étale. The simple étale group schemes are obviously $\mathbb{Z}/\ell\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z}$. A simple connected group scheme is killed by Frobenius, and therefore of height 1, and therefore $\mu_p$ or $\alpha_p$. □

**Corollary.** Let $G$ be a finite commutative group scheme of order $n$. Then $[n] = 0$ on $G$. In particular, for any $x \in G(R)$ we have $nx = 0$.

*Proof.* This can be verified over $\overline{k}$. If it is true for the outer groups in a short exact sequence, then it's true for the middle group. It therefore suffices to verify the case of simple group schemes, which follows easily from the classification. □

**Remark.** An $F$-isomodule is an $F$-module with $F$ injective. Then the category of $F$-isomodules is equivalent to the category of groups $G$ such that $G_{\overline{k}}$ is isomorphic to $\mu_p^n$ for some $n$. But, by Cartier duality, this category is equivalent ot the category of groups $G$ such that $G_{\overline{k}}$ is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^n$, which is the same to say that $G$ is étale and killed by $p$. But we know that this category is equivalent to the category of $\mathbb{F}_p$-representations of $G_k$. We find that we have an equivalence of categories

$$\{F\text{-isomodules}\} \cong \{\mathbb{F}_p\text{-representations of } G_k\}$$

This equivalence can be described explicitly using the "kernel object" $k^s$, which has a compatible Galois action and $F$-module structure. Precisely, an $F$-isomodule $M$ is taken to $(M \otimes k^s)^{F=1}$ while a Galois representation $V$ is taken to $(V \otimes k^s)^{G_k}$. Fontaine generalized this to a description of the category of $\mathbb{Z}_p[G_k]$-modules.

### Dieudonné theory

It would of course be desirable to remove the height 1 restriction in the above classification of group schemes. This is exactly what Dieudonné theory does, assuming $k$ is perfect. Let $W = W(k)$ be the Witt vectors of $k$. If $k = \mathbb{F}_q$, which is the most common case, $W$ is the ring of integers in the unramified extension of $\mathbb{Q}_p$ with residue field $k$. The absolute Frobenius on $k$ induces an automorphism $\varphi$ of $W$. A Dieudonné module is a $W$-module $D$ equipped with two additive maps $F, V\colon D \to D$ satisfying $F(ax) = \varphi(a)x$, $V(ax) = \varphi^{-1}(a)x$, and $FV = VF = p$. The main theorem is then:

**Theorem.** Suppose $k$ is perfect. The category of finite commutative group schemes over $k$ of $p$-power order is equivalent to the category of Dieudonné modules of finite length over $W$.

Write $D(G)$ for the Dieudonné module associated to $G$. The functor $D$ has several nice properties in addition to being an equivalence:

(i) $D$ is an exact functor.

(ii) The group $G$ is killed by $p^n$ if and only if $D(G)$ is.

(iii) The order of $G$ is equal to $p^r$, where $r$ is the length of $D(G)$ as a $W$-module.

(iv) $G$ is connected if and only if $F$ is nilpotent on $D(G)$, and étale if and only if $F$ is an isomorphism on $D(G)$.

(v) $D(G^\vee)$ is naturally the dual of $D(G)$, where the dual of a Dieudonné module $M$ is the $W$-module $\mathrm{Hom}_W(M, K/W)$ with $F$ and $V$ defined by $(Ff)(x) = \varphi(f(Vx))$ and $(Vf)(x) = \varphi^{-1}(f(Fx))$. Here $K$ is the field of fractions of $W$.

(vi) If $G$ has height 1 then $D(G)^\vee = L(G)$ (and $V = 0$).

### 1.2.5 Applications to abelian varieties

**Duality of abelian varieties revisited**

We previously showed that if $f\colon X \to Y$ is an isogeny of complex tori then $\ker(f)$ and $\ker(f^\vee)$ are naturally Pontryagin dual groups. We now generalize this to arbitrary fields:

**Proposition.** Let $f\colon A \to B$ be an isogeny of abelian varieties. Then $\ker(f^\vee)$ is naturally Cartier dual to $\ker(f)^\vee$.

*Proof.* Put $G = \ker(f)$. Applying $\underline{\mathrm{Hom}}(-, \mathbb{G}_m)$ to the short exact sequence of fppf sheaves,

$$0 \to G \to A \to B \to 0$$

we obtain a long exact sequence

$$0 \to \underline{\mathrm{Hom}}(B, \mathbb{G}_m) \to \underline{\mathrm{Hom}}(A, \mathbb{G}_m) \to \underline{\mathrm{Hom}}(G, \mathbb{G}_m) \to \underline{\mathrm{Ext}}^1(B, \mathbb{G}_m) \to \underline{\mathrm{Ext}}^1(A, \mathbb{G}_m)$$

There are no maps from an abelian variety to $\mathbb{G}_m$ (since abelian varieties are proper and $\mathbb{G}_m$ is affine), so the first two groups vanish. We've seen that $\underline{\mathrm{Hom}}(-, \mathbb{G}_m)$ is Cartier duality for finite commutative group schemes and $\underline{\mathrm{Ext}}^1(-, \mathbb{G}_m)$ is duality for abelian varieties. Thus $G^\vee = \ker(f^\vee)$. A more elementary proof is given in section 15 of Mumford's "Abelian varieties" (MR0282985). □

**Corollary.** Let $A$ be an abelian variety. Then $A[n]$ and $A^\vee[n]$ are Cartier dual. In particular, there is a canonical pairing $A[n] \times A^\vee[n] \to \mu_n$, the Weil pairing.

**The p-torsion of an elliptic curve**

Let $E$ be an elliptic curve over $k$, which we assume to be algebraically closed of characteristic $p$. Then $G = E[p]$ is a finite commutative group scheme over $k$ of order $p^2$. We know a lot about such group schemes, so it's reasonable to think we could describe $G$ fairly precisely.

We know two things right off the bat. First, $G$ is not étale. And second, since $E$ is self-dual (as an abelian variety), $G$ is self-dual (in the sense of Cartier duality).

First suppose that $E$ is ordinary. Then $G(k) \neq 0$, and so $G^{\mathrm{et}}$ is non-zero. Thus $G = G^\circ \times G^{\mathrm{et}}$, and both factors have order $p$. By the classification of étale groups, $G^{\mathrm{et}} = \mathbb{Z}/p\mathbb{Z}$. Since $G$ is self-dual, $G^\circ$ is necessarily the dual of $G^{\mathrm{et}}$, so $G^\circ = \mu_p$. We thus find $G = \mu_p \times \mathbb{Z}/p\mathbb{Z}$.

Now suppose that $E$ is supersingular. Then $G(k) = 0$, and so $G^{\mathrm{et}} = 0$. It follows that $G$ is local, and thus local–local since it is self-dual. Since the only simple local–local group is $\alpha_p$, we must have an extension of the form

$$0 \to \alpha_p \to G \to \alpha_p \to 0$$

This extension cannot be split, for then $G = \alpha_p \oplus \alpha_p$, which has a two-dimensional tangent space, but the tangent space of $G$ agrees with that of $E$, and has dimension 1. We also cannot have $G = \alpha_{p^2}$, since this group is not self-dual (it has $V = 0$ but $F \neq 0$); of course, we cannot have $G = \alpha_{p^2}^\vee$ either. In fact, up to isomorphism, there are only four self-extensions of $\alpha_p$ (as can easily be seen using Dieudonné theory), and $G$ is the one we haven't named! One can describe $G$ as the sum of $\alpha_{p^2}$ and its dual in $\mathrm{Ext}^1(\alpha_p)$, and one can also explicitly describe the Dieudonné module $D(G)$.

**The p-torsion of an abelian variety**

Let $A$ be an abelian variety over $k$, assumed to be of characteristic $p$. Write $A[p] = G_1 \oplus G_2 \oplus G_3$ where $G_1$ is étale, $G_2$ is local–étale, and $G_3$ is local–local. Write $\#G_1 = p^r$, $\#G_2 = p^s$, and $\#G_3 = p^t$.

**Proposition.** The numbers $r$, $s$, and $t$ are isogeny invariant.

*Proof.* Decompose $A[p^n]$ as $G_{1,n} \oplus G_{2,n} \oplus G_{3,n}$ as above. Note that $A[p^n]$ is a successive extension of $A[p]$'s; it follows that $G_{i,n}$ is a successive extension of $G_i$'s. In particular, $\#G_{1,n} = p^{nr}$, $\#G_{2,n} = p^{ns}$, and $\#G_{3,n} = p^{nt}$.

Now suppose that $A \to A'$ is an isogeny of degree $d$. Then, in the obvious notation, the kernel of $G_{1,n} \to G'_{1,n}$ has order at most $d$. Clearly, for $n \gg 0$, this is only possible if $r \leq r'$. Since "isogenous" is an equivalence relation, there exists an isogeny $A' \to A$ and so $r' \leq r$ as well. Thus $r = r'$. The equality of the other invariants is similar. □

**Proposition.** We have $r = s$ and $t = 2g - 2r$, where $g = \dim(A)$.

*Proof.* By duality, we have $r(A) = s(A^\vee)$. But $A$ and $A^\vee$ are isogenous (via a polarization), and so $r = s$. The formula for $t$ follows, since $A[p]$ has order $p^{2g}$. □

**Corollary.** We have $A[p](\overline{k}) = (\mathbb{Z}/p\mathbb{Z})^r$ with $r \leq g$.

*Proof.* Since $r = s$, we have $2r = r + s \leq 2g$, and so $r \leq g$. □

**The Dieudonné module as a p-adic Tate module**

Let $A$ be an abelian variety of dimension $g$ over $k$, a perfect field of characteristic $p$. Then $T_p(A)$, the $p$-adic Tate module of $A$, has rank at most $g$, and could even be 0; it is therefore very much unlike the $\ell$-adic Tate modules of $A$. Define the Dieudonné module of $A$, denoted $D(A)$, to be the inverse limit of those of the $A[p^n]$. Then $D(A)$ is a free $W$-module of rank $2g$ equipped with a semi-linear map $F$, and thus looks more like the $\ell$-adic Tate module. (Note: $V$ is not needed since $VF = p$.)

Now suppose $k = \mathbb{F}_q$ with $q = p^r$. Let $F' = F^r$. Then $F'$ is a $W$-linear automorphism of $D(A)$. This looks even more like the $\ell$-adic Tate module! In fact, the analogy is very good: the eigenvalues of $F'$ are the same as the eigenvalues of Frobenius on the $\ell$-adic Tate module.

### 1.2.6 Finite flat group schemes

In the final section of group schemes, we will devote it to Raynaud's theorem: in mixed characteristic and low ramification, a group scheme is determined by its generic fiber. The proof proceeds by first reducing the statement to a special class of group schemes, the Raynaud $F$-module schemes. Next, these schemes are classified; this is the meat of the argument. (I didn't have time to prove the important congruence relations, but they are covered in the notes here.) Finally, one verifies the theorem for Raynaud $F$-module schemes using the classification result. A good reference for this is Tate's article "Finite flat group schemes" in the book "Modular forms and Fermat's last theorem" (MR1638478), especially section 4.

Up until now, we have considered finite commutative group schemes over a field. We now work over a more general base scheme $S$, which we assume to be affine and noetherian; write $S = \operatorname{Spec}(R)$. To have an analogous theory, we only consider flat group schemes. A finite flat $R$-module is projective, and so the coordinate rings of our group schemes will be projective $R$-modules.

Much of what we have previously done carries over:

(i) Finite flat group schemes over $S$ correspond to Hopf algebras over $R$ which are finitely generated and projective.

(ii) We define the order of a finite flat group scheme as the rank of the corresponding Hopf algebra. In general, this is a locally constant function on $S$, but if $S$ is connected it can be treated as a single number.

(iii) Quotients work: if $G$ and $H$ are finite and flat and $H$ is a closed subgroup of $G$ then $G/H$ exists and is finite and flat of the expected order.

(iv) The classification of étale group schemes is similar: assuming $S$ is connected, the category of finite flat commutative étale group schemes is equivalent to the category of finite $\pi_1(S, s)$-modules, where $s$ is a geometric point of $S$. If $R$ is the ring of integers in a finite extension $K$ of $\mathbb{Q}_p$, then $\pi_1(S, s)$ is the Galois group of the maximal unramified extension of $K$.

(v) Assume $R$ is henselian local (e.g., complete local). Then one can treat the connected–étale sequence in the same manner. If $G = \mathrm{Spec}(A)$ is a group scheme, then $A$ is semi-local and thus decomposes as $A = \prod A_i$. Once again, $G^0 = \mathrm{Spec}(A_0)$, where the counit factors through $A_0$. One then has $G^{\mathrm{et}} = G/G^0$.

(vi) If the order of $G$ is invertible on $S$ then $G$ is étale.

(vii) Cartier duality works in the same way.

### 1.2.7 Raynaud's theorem

#### Statement of theorem

Let $K/\mathbb{Q}_p$ be a finite extension, let $R$ be its ring of integers in $K$, let $k$ be the residue field of $R$, and let $e$ be the ramification index of $K/\mathbb{Q}_p$. A prolongation of a finite group scheme $G_0/K$ is a finite flat group scheme $G/R$ equipped with an isomorphism $G_K \to G_0$.

**Theorem** (Raynaud)**.** Suppose $e < p - 1$. Let $G_0$ be a finite commutative group scheme over $K$. Then any two prolongations of $G_0$ to $R$ are isomorphic.

**Remark.** This theorem is clearly not true without the hypothesis on $e$: indeed, if $K$ has the $p$th roots of unity then $\mu_p$ and $\mathbb{Z}/p\mathbb{Z}$ are isomorphic over $K$, and thus both prolongations of $\mathbb{Z}/p\mathbb{Z}$, but are not isomorphic over $R$ (as one is étale and one is connected).

We say that a group scheme $G_0$ over $K$ has property UP (unique prolongation) if any two prolongations of $G_0$ to $R$ are isomorphic. Raynaud's theorem says UP holds for all $G_0$ if $e < p - 1$.

The proof of Raynaud's theorem can be divided into three steps. First we show that if $G_0$ is an extension and UP holds for the sub and quotient then it holds for $G_0$. This means it suffices to prove UP for simple groups. We then classify the simple groups and their prolongations. This is the most involved step. Finally, we check by hand that UP holds when $e < p - 1$.

#### prolongations

Let $G_0 = \mathrm{Spec}(A_0)$ be a finite commutative group scheme over $K$. Prolongations of $G_0$ correspond to finite $R$-subalgebras $A$ of $A_0$ which are closed under comultiplication and span $A_0$ over $K$ (these conditions imply that $A$ is closed under the antipode). We partially order prolongations using inclusion on rings.

**Proposition.** Two prolongations have an inf and a sup.

*Proof.* If $A$ and $A'$ are the rings of two prolongations then $AA'$ is clearly closed under comultiplication, and thus a prolongation greater than each; it is clearly the unique minimal one, and thus the inf. Sups follow from Cartier duality. □

**Proposition.** If $G_0$ has a prolongation then it has a maximal one $G^+$ and a minimal one $G^-$.

*Proof.* Since $A_0$ is a finite étale $K$-algebra, it has a maximal order. It follows that any ascending chain of prolongations stabilizes, and so there exists a maximal prolongation. The minimal one follows from Cartier duality. □

We thus see that UP holds for $G_0$ if and only if the natural map $G^+ \to G^-$ is an isomorphism (or, in terms of rings, $A^+ = A^-$). In particular, one can check UP by passing to an extension of $K$. Suppose now that

$$0 \to G_0' \to G_0 \to G_0'' \to 0$$

is a short exact sequence of group schemes and that $G_0$ admits a prolongation $G$. Then the scheme-theoretic closure $G'$ of $G_0'$ in $G$ is a prolongation of $G_0'$, and the quotient $G'' = G/G'$ is a prolongation of $G_0''$. Furthermore, if $H$ is a second prolongation of $G_0$ which is less than $G$ (i.e., there is a map $G \to H$), then clearly $H'$ is less than $G'$ and $H''$ is less than $G''$, and the following diagram commutes

$$\begin{tikzcd}
0 \arrow[r] & G' \arrow[r] \arrow[d] & G \arrow[r] \arrow[d] & G'' \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & H' \arrow[r] & H \arrow[r] & H'' \arrow[r] & 0
\end{tikzcd}$$

Thus, if the maps $G' \to H'$ and $G'' \to H''$ are isomorphisms, so is $G \to H$. This proves the following:

**Proposition.** In the above situation, if $G_0'$ and $G_0''$ satisfy UP then so does $G_0$.

### Raynaud F-module schemes

Let $G_0/K$ be a simple group scheme of $p$-power order, and let $V = G_0(\overline{K})$, and irreducible $\mathbb{F}_p$-representation of the Galois group. Then $\mathbb{F} = \mathrm{End}_{G_K}(V)$ is a finite extension of $\mathbb{F}_p$, and we can regard $V$ as an absolutely irreducible $\mathbb{F}$-linear representation of $G_K$.

Suppose now that $k$ is algebraically closed, e.g., pass to $K^{\mathrm{un}}$. Then $G_K$ is an extension of a tame part $I^t$, which is abelian, by a wild part $I^w$, which is pro-$p$. Since $I^w$ is pro-$p$, it must fix a non-zero vector in $V$; thus $V^{I^w}$ is non-zero. But $I^w$ is normal in $G_K$, and so $V^{I^w}$ is a subrepresentation of $V$. Since $V$ is simple, this implies $V = V^{I^w}$, i.e., $I^w$ acts trivially. Thus the action of $G_K$ on $V$ factors through $I^t$. Since $V$ (regarded as an $\mathbb{F}$-vector space) is an absolutely irreducible representation of an abelian group, it necessarily has dimension 1, i.e., $\dim_{\mathbb{F}}(V) = 1$.

An $\mathbb{F}$-module scheme (over $R$ or $K$) is a group scheme $G$ equipped with a ring homomorphism $\mathbb{F} \to \mathrm{End}(G)$. If $G = \mathrm{Spec}(A)$, we write $[t]$ for the map $A \to A$ induced by $t \in \mathbb{F}$. A Raynaud $\mathbb{F}$-module scheme $G$ is an $\mathbb{F}$-module scheme of the same order as $\mathbb{F}$; thus it is an $\mathbb{F}$-module scheme such that $G(\overline{K})$ is one-dimensional over $\mathbb{F}$. The above discussion proves the following:

**Proposition.** Suppose $k$ is algebraically closed and $G_0/K$ is simple of $p$-power order. Then $G_0$ is canonically a Raynaud $\mathbb{F}$-module scheme (for some $\mathbb{F}$).

**Proposition** (No hypothesis on $k$)**.** Suppose UP holds for every Raynaud $\mathbb{F}$-module scheme over $K^{\mathrm{un}}$. Then UP holds for all finite order group schemes over $K$.

*Proof.* Let $G_0/K$ be given. We can check UP for $G_0$ over $K^{\mathrm{un}}$, so we can assume $k$ is closed. We have shown that UP can be deduced if it is known for the outer groups in an extension. It thus suffices to treat the case where $G_0$ is simple. If $G_0$ is $p$-power order then it is a Raynaud $\mathbb{F}$-module scheme, and satisfies UP by hypothesis. If $G_0$ is prime-to-$p$ then UP is automatic, as any prolongation is étale. □

Suppose now that $G_0$ is a Raynaud $\mathbb{F}$-module scheme over $K$. The action of $\mathbb{F}$ need not extend to an arbitrary prolongation of $G_0$. However, it necessarily extends to the maximal and minimal prolongations since they are unique. (For $t \in F^\times$, the map $[t]$ of $A_0$ is an automorphism of Hopf algebras, and therefore must carry $A^+$ into itself.) Thus $G^+$ and $G^-$ are Raynaud $\mathbb{F}$-module schemes, and the map $G^+ \to G^-$ respects the $\mathbb{F}$-structure.

### 1.2.8 Analysis of Raynaud $F$-module schemes

**Set-up**

We now analyze Raynaud $\mathbb{F}$-module schemes. We fix the finite field $\mathbb{F}$, and write $q = p^r$ for its order. We assume that $k$ contains the $q-1$ roots of unity (i.e., $\mathbb{F}$ embeds into $k$).

A character $\mathbb{F}^\times \to R^\times$ is fundamental if the composite map $\mathbb{F}^\times \to k^\times$ extends to an embedding of fields. If $\chi$ is a fundamental character, then any other one is of the form $\chi^{p^k}$ for some $k$. Enumerate the fundamental characters as $(\chi_i)_{i \in \mathcal{I}}$, for some index set $\mathcal{I}$. For $i \in I$, define $i+1$ by $\chi_{i+1} = \chi_i^p$. Then $\mathcal{I}$ is a torsor for $\mathbb{Z}/r\mathbb{Z}$.

An arbitrary character $\mu \colon \mathbb{F}^\times \to R^\times$ can be expressed as a product $\prod_{i \in I} \chi_i^{a_i}$ with $a_i \in \mathbb{Z}$. This product is unique if we impose the conditions that $0 \le a_i \le p-1$ and not all the $a_i$ are equal to 0. We write $\mu(i)$ in place of $a_i$. Note that if $\mu$ is the trivial character then $\mu(i) = p-1$ for all $i$.

**Initial analysis**

Let $G = \operatorname{Spec}(A)$ be a Raynaud $\mathbb{F}$-module scheme over $R$. Then $G_{\overline{K}}$ is isomorphic to the constant group scheme $\mathbb{F}$ over $\overline{K}$. We fix an isomorphism of $A_{\overline{K}}$ with the algebra of functions $\mathbb{F} \to \overline{K}$. For a character $\mu \colon \mathbb{F}^\times \to K^\times$, we let $\epsilon_\mu$ be the function $\mathbb{F} \to K$ extending $\mu$ and with $\epsilon_\mu(0) = 0$. Thus $\epsilon_\mu$ is an element of $A_{\overline{K}}$. We put $\epsilon_i = \epsilon_{\chi_i}$.

Let $I$ be the augmentation ideal of $A$. Since $\mathbb{F}^\times$ is a finite group whose order is invertible in $R$ and all characters of $\mathbb{F}^\times$ are defined over $R$ (since $k$ contains the $q-1$ roots of unity), we can decompose $I$ as a sum $\bigoplus_\mu I_\mu$, where the sum is over the characters $\mu$ of $\mathbb{F}^\times$, and $I_\mu$ is the $R$-submodule of $I$ consisting of elements $x$ such that $[t]x = \mu(t)x$ for all $t \in \mathbb{F}^\times$. Clearly, $I_\mu \otimes_R \overline{K}$ is spanned by $\epsilon_\mu$, and so $I_\mu$ is a free $R$-module of rank 1.

For each $i \in \mathcal{I}$, choose a non-zero element $X_i$ of $I_{\chi_i}$. Then $X_i = c_i \epsilon_i$ for some $c_i \in \overline{K}^\times$. Since $X_i^p$ clearly belongs to $I_{\chi_{i+1}}$, we have $X_i^p = \delta_i X_{i+1}$ for some $\delta_i \in R$. As $\epsilon_i^p = \epsilon_{i+1}$, we find $\delta_i = c_i^p / c_{i+1}$. For a character $\mu = \prod \chi_i^{\mu(i)}$ of $F$, let $X^\mu$ be the product $\prod X_i^{\mu(i)}$, an element of $I_\mu$. Note that what we might call $\epsilon^\mu$, namely $\prod \epsilon_i^{\mu(i)}$, is simply $\epsilon_\mu$.

Let $G^\vee = \operatorname{Spec}(B)$ be the Cartier dual of $G$, so that $B$ is the $R$-linear dual of $A$. Then $B_{\overline{K}}$ is naturally identified with the group algebra $\overline{K}[\mathbb{F}]$. We write $\{t\}$ for the element of $\overline{K}[\mathbb{F}]$ corresponding to $t \in F$. Note that $\{t\}\{s\} = \{t+s\}$ (multiplication in the group algebra), but $[t]\{s\} = \{ts\}$ (the $\mathbb{F}$-module structure). Note also that the pairing $\langle , \rangle \colon A_{\overline{K}} \times B_{\overline{K}} \to \overline{K}$ is given by evaluating functions on elements: $\langle f, \{t\} \rangle = f(t)$.

Let $J$ be the augmentation ideal of $B$. We again have a decomposition $J = \bigoplus_\mu J_\mu$. For $\mu$ non-trivial, the space $J_\mu \otimes_R \overline{K}$ is spanned by the element

$$e_\mu = \frac{1}{q-1} \sum_{t \in F^\times} \mu^{-1}(t)\{t\}$$

The space $J_1 \otimes_R \overline{K}$, on the other hand, is spanned by

$$e_1 = -1 + \frac{1}{q-1} \sum_{t \in F^\times} \{t\}$$

The vector spaces $I_{\overline{K}}$ and $J_{\overline{K}}$ are canonically dual. The bases $\epsilon_\mu$ and $e_\mu$ are dual bases, that is, $\langle \epsilon_\mu, e_\nu \rangle = \delta_{\mu,\nu}$. We write $e_i$ for $e_{\chi_i}$.

Let $Y_i = c_i^{-1} e_i$. Then $Y_i$ is an element of $J_{\chi_i}$. Of course, $Y_i^p = \gamma_i Y_{i+1}$ for some $\gamma_i \in R$. For $\mu = \prod \chi_i^{\mu(i)}$, let $Y^\mu = \prod Y_i^{\mu(i)}$, an element of $J_\mu$. We also put $e^\mu = \prod e_i^{\mu(i)}$, which is not equal to $e_\mu$.

Let $w_\mu = \langle X^\mu, Y^\mu \rangle$ and $w_i = \langle X_i^p, Y_i^p \rangle$. The main work in understanding $G$ is understanding the behavior of these numbers. Note that $w_\mu = \langle \epsilon_\mu, e^\mu \rangle$ and $w_i = \langle \chi_{i+1}, e_{\chi_i}^p \rangle$, so these numbers do not depend on $G$. We will understand them by computing with a specific choice of $G$. ¡h3¿Determining the $w$'s¡/h3¿ In this section, we take $G$ to be the constant group scheme $\mathbb{F}$ over $R$. Thus $A$ is the ring of functions $\mathbb{F} \to R$, and $B$ is the group algebra $R[\mathbb{F}]$. The elements $\epsilon_\mu$ of $A_{\overline{K}}$ belong to $A$, and the elements $e_\mu$ of $B_{\overline{K}}$ belong to $B$.

Obviously, $\epsilon_\mu \epsilon_\nu = \epsilon_{\mu\nu}$. Since $\Delta(\epsilon_\mu)$ is a $\mu$-eigenvector of $F^\times$, it must be a linear combination of the tensors $\epsilon_\mu \otimes 1$, $1 \otimes \epsilon_\mu$, and the $\epsilon_\nu \otimes \epsilon_\eta$ with $\nu\eta = \mu$. As $\langle \Delta(\epsilon_\mu), e_\mu \otimes 1 \rangle = \langle \epsilon_\mu, e_\mu \rangle = 1$, it follows that $\epsilon_\mu \otimes 1$ appears with coefficient 1, and similarly for $1 \otimes \epsilon_\mu$. We thus find

$$\Delta(\epsilon_\mu) = \epsilon_\mu \otimes 1 + 1 \otimes \epsilon_\mu + \sum_{\nu\eta=\mu} J_{\nu,\eta} \cdot \epsilon_\nu \otimes \epsilon_\eta$$

for some $J_{\nu,\eta} \in R$. Dualizing these expressions, we find $e_\mu e_\nu = J_{\mu,\nu} e_{\mu\nu}$ and

$$\Delta(e_\mu) = e_\mu \otimes 1 + 1 \otimes e_\mu + \sum_{\mu=\nu\eta} e_\nu \otimes e_\eta$$

We now consider $B_k = k[\mathbb{F}]$. The space $I_k$ is spanned by the elements $\langle x \rangle = \{x\} - \{0\}$, while the space $I^2$ is spanned by the elements

$$\langle x \rangle \langle y \rangle = \{x+y\} - \{x\} - \{y\} + \{0\} = \langle x+y \rangle - \langle x \rangle - \langle y \rangle$$

Thus $I_k/I_k^2$ is the quotient of the $k$-vector space with basis $\langle x \rangle$ (for $x \in \mathbb{F}$), by the relations $\langle x+y \rangle = \langle x \rangle + \langle y \rangle$. In other words, $I/I^2$ is exactly $k \otimes \mathbb{F}$. This decomposes as $\bigoplus_{i \in \mathcal{I}} k v_i$, where the product is over the embeddings of $\mathbb{F}$ into $k$ (as fields). By definition, $\mathbb{F}^\times$ acts on $v_i$ through $\chi_i$. It follows that $v_i = e_i$ (or rather, its image in $B_k$).

Since the $e_i$ are elements of $I_k$ whose image in $I_k/I_k^2$ form a basis, it follows from Nakayama's lemma that the $e_i$ generate $B_k$ as an algebra (and thus $B$ as well). Clearly, $I_k^p = 0$, and so the only non-zero monomials in the $e_i$ are the $e^\mu$, and so these form a $k$-basis for $I_k$. Thus the $e^\mu$ are an $R$-basis of $I$. In particular, $e_\mu$ and $e^\mu$ differ multiplicatively by a unit of $R$. (In fact, $e^\mu = w_\mu e_\mu$, by definition of $w_\mu$, so this shows that $w_\mu$ is a unit. We obtain a more precise statement below.)

If $x \in I$ then $x^p = 0$ modulo $p$. It follows that if $\mu(i) + \nu(i) \geq p$ for any $i$ (e.g., if $\mu\nu$ is a fundamental character) then $e^\mu e^\nu = 0 \bmod p$ ; thus $e_\mu e_\nu = 0 \bmod p$ as well, and so $J_{\mu,\nu} = 0$ mod $p$ as well. In particular, we see that $\Delta(\epsilon_i) = \epsilon_i \otimes 1 + 1 \otimes \epsilon_i$ modulo $p$. Suppose $n = \sum \mu(i)$. Then

$$\langle \epsilon^\mu, e^\mu \rangle = \langle \prod \epsilon_i^{\mu(i)}, \prod e_i^{\mu(i)} \rangle = \langle \prod \Delta_n(\epsilon_i)^{\mu(i)}, \bigotimes e_i^{\otimes \mu(i)} \rangle$$

where $\Delta_n \colon A \to A^{\otimes n}$ is repeated comultiplication (which is an algebra homomorphism). Considering this equation mod $p$, we can replace $\Delta_n(\epsilon_i)$ with $\epsilon_i \otimes 1 \otimes \cdots \otimes 1 + \cdots$, where the $\cdots$ are the symmetrical terms. The inner product is the coefficient of $\bigotimes \epsilon_i^{\otimes \mu(i)}$, which is easily seen to be $\prod \mu(i)!$. We have thus proven:

**Proposition.** We have $w_\mu = \prod \mu(i)!$ modulo $p$.

Similarly, we have

$$\langle \epsilon_i^p, e_i^p \rangle = \langle \Delta_p(\epsilon_i)^p, e_i^{\otimes p} \rangle$$

Now, we can write $\Delta_p(\epsilon_i)$ as $x+y$, where $x$ is a sum of things like $\epsilon_i \otimes 1 \cdots \otimes 1$, and $y$ is a multiple of $p$. We have $(x+y)^p = x^p + \cdots + y^p$, where the unwritten terms have both $y$'s and binomial coefficients, and are thus divisible by $p^2$; of course, $y^p$ is also divisible by $p^2$. We thus find $(x+y)^p = x^p \bmod p^2$. Now, the coefficient of $\epsilon_i^{\otimes p}$ in $x^p$ is $p!$, which is $-p$ modulo $p^2$. We have thus shown:

**Proposition.** $w_i = -p$ modulo $p^2$.

### Structure theorem

We now return to the general setting. The elements $X^\mu$ span an $R$-submodule of $I$, while the $Y^\mu$ span an $R$-submodule of the dual module $J$. The pairing $\langle X^\mu, Y^\mu \rangle$ is a unit, namely $w_\chi$ (and all other pairings vanish). It follows that the $X^\mu$ span $I$ and the $Y^\mu$ span $J$. Thus the $X_i$ generate $A$ as an algebra.

Recall $X_i^p = \delta_i X_{i+1}$ and $Y_i^p = \gamma_i Y_{i+1}$ and $\delta_i \gamma_i = w_i$. As $w_i = -p$ modulo $p^2$, we see that $v(\delta_i) \le e$, where $v$ is the valuation on $K$ with $v(p) = e$. We have thus shown:

**Theorem.** $A$ is isomorphic to the quotient of $R[X_i]$ by equations $X_i^p = \delta_i X_{i+1}$, where $\delta_i$ is an element of $R$ of valuation at most $e$.

### Existence theorem

**Theorem.** Let $(\delta_i)_{i \in \mathcal{I}}$ be elements of $R$ of valuation at most $e$, and let $A$ be the quotient of $R[X_i]$ by the equations $X_i^p = \delta_i X_{i+1}$. Then there is a unique structure of a Raynaud $F$-module scheme on $G = \operatorname{Spec}(A)$ such that $[t]X_i = \chi_i(t) X_i$ for all $t \in F^\times$.

*Proof.* Choose $c_i \in \overline{K}^\times$ such that $\delta_i = c_i^p / c_{i+1}$. Identify $A_{\overline{K}}$ with the ring of functions on $\mathbb{F}$ via $X_i = c_i \epsilon_i$. Then the monomials $X^\mu$, and the unit 1, form an $R$-basis for $A$.

Let $B$ be the $R$-dual of $A$, thought of as an $R$-submodule of $\overline{K}[\mathbb{F}]$. Since $X^\mu = (\prod c_i^{\mu(i)}) \epsilon_\mu$, the dual basis $Y_\mu$ is given by $Y_\mu = (\prod c_i^{-\mu(i)}) e_\mu$. Let $Y_i = c_i^{-1} e_i$. Then $Y^\mu = (\prod c_i^{-\mu(i)}) e^\mu$, which differs by a unit from $Y_\mu$. Thus the $Y^\mu$ (and 1) span $B$ as an $R$-module.

Of course, we still have $Y_i^p = \gamma_i Y_{i+1}$ with $\gamma_i \delta_i = p$. The point is that, due to the restriction on $v(\delta_i)$, this shows that $\gamma_i \in R$, and hence the span of the $Y^\mu$ is an algebra. Thus the dual to $A$ is closed under multiplication, which shows that $A$ is closed under comultiplication, from which one can show that $A$ is naturally a sub Hopf algebra of the ring of $\overline{K}$-valued functions on $\mathbb{F}$. This shows that $G$ can be endowed with an $\mathbb{F}$-module structure as stated.

By our previous work, any Raynaud $\mathbb{F}$-module structure on $G$ comes from one of the above form, for some choice of $c$'s. The choice of $c$'s does not change the resulting structure, as the $c$'s are uniquely determined up to a single $q-1$ root of unity, which can compensated for using $\mathbb{F}^\times$. Thus the Raynaud $\mathbb{F}$-module structure on $G$ is unique. □

We write $G_\delta$ for the Raynaud $\mathbb{F}$-module scheme corresponding to $\delta = (\delta_i)_{i \in \mathcal{I}}$. The structure theorem can be rephrased as: every Raynaud $\mathbb{F}$-module scheme is isomorphic to $G_\delta$, for some $\delta$. We leave the following proposition to the reader:

**Proposition.** The set of $\mathbb{F}$-module homomorphism maps $f \colon G_\delta \to G_{\delta'}$ correspond to sequence $(a_i)_{i \in \mathcal{I}}$ of elements of $R$ such that $a_{i+1} \delta_i = a_i^p \delta_i'$. The sequence $(a_i)$ corresponds to the map $f$ which is given on rings by $X_i' \mapsto a_i X_i$.

### UP for $F$-module schemes

**Proposition.** Suppose $e < p-1$ and $f\colon G \to G'$ is a map of Raynaud $\mathbb{F}$-module schemes over $R$ which induces an isomorphism over $K$. Then $f$ is an isomorphism.

*Proof.* Write $G = G_\delta$ and $G' = G_{\delta'}$, so that $f$ corresponds to a sequence $(a_i)$ with $a_{i+1}\delta_i = a_i^p \delta_i'$. Let $i$ be such that $v(a_i)$ is maximal. Then $v(a_{i+1} + \delta_i) \le v(a_i) + e$, while $v(a_i^p\delta_i') \ge pv(a_i)$. Thus $pv(a_i) \le v(a_i) + e$, i.e., $(p-1)v(a_i) \le e$. The hypothesis on $e$ forces $v(a_i) = 0$, and so all the $a_j$'s are units, and so $f$ is an isomorphism. □

**Proposition.** Suppose $e < p-1$. Then UP holds for Raynaud $\mathbb{F}$-module schemes over $K$.

*Proof.* Apply the previous proposition to the maximal and minimal prolongation, which are Raynaud $\mathbb{F}$-module schemes over $R$. This shows $G^+ = G^-$, which implies UP. □

This completes the proof of Raynaud's theorem.

## 1.3 Abelian varieties in mixed characteristic, including Néron models

### 1.3.1 Reduction theory of elliptic curves

This section is devoted to the behavior of elliptic curves over DVRs. The various types of reduction (good, multiplicative, additive) are defined, and their behavior under extension is studied. Then the behavior of torsion points under reduction is discussed. Finally, I prove the Néron–Ogg–Shafarevich theorem. A good reference for this section is Chapter VII of Silverman's "The arithmetic of elliptic curves" (MR0817210).

Let $R$ be a complete DVR, $\mathfrak{f}$ its maximal ideal, $K$ its field of fractions, $k$ its residue field, and $v$ the valuation with $v(\pi) = 1$, for $\pi$ a uniformizer. We are going to study elliptic curves over $K$, and their reduction modulo $\mathfrak{f}$. We assume throughout that $k$ does not have characteristic 2 or 3.

#### Minimal Weierstrass equations

Let $E/K$ be an elliptic curve given by a Weierstrass equation $y^2 = x^3 + ax + b$. Recall that the discriminant $\Delta = -16(4a^3 + 27b^2)$ is non-zero. A Weierstrass equation for $E$ is not unique: one can replaces $y$ with $u^3y$ and $x$ with $u^2x$, for $u \in K^\times$, which has the effect of changing $a$ to $u^{-4}a$ and $b$ to $u^{-6}b$. We say that a Weierstrass equation is minimal if $a$ and $b$ belong to $R$ and $v(a) < 4$ or $v(b) < 6$ (this is equivalent to asking that $v(\Delta)$ be minimal). A minimal Weierstrass equation is unique up to a change of variables as above with $u$ a unit. We let $\mathcal{E}$ be the projective scheme over $R$ defined by a minimal Weierstrass equation. We call this the minimal Weierstrass model for $E$. It is independent of the choice of minimal Weierstrass equation, up to isomorphism (since $u$ must be a unit in any change of variables).

We can now introduce most of the objects we will be interested in:

(i) We let $\overline{E}$ be $\mathcal{E}_k$, the special fiber of $\mathcal{E}$. We call this the reduction of $E$ modulo $\mathfrak{f}$. This is an irreducible projective curve over $k$, though possibly singular.

(ii) We let $\overline{E}_{\mathrm{sm}}$ be the smooth locus of $\overline{E}$. A basic fact is that $\overline{E}_{\mathrm{sm}}$ is a group variety: the group law can be defined using the secant line construction, as on an elliptic curve.

(iii) Since $\mathcal{E}$ is projective, $\mathcal{E}(R) = \mathcal{E}(K) = E(K)$. We therefore have a well-defined map $E(K) \to \overline{E}(k)$, which we call the reduction map.

(iv) We let $E_0(K)$ be the subset of $E(K)$ which reduces into $\overline{E}_{\mathrm{sm}}(k)$. Then $E_0(K)$ is a subgroup of $E(K)$, and the reduction map $E_0(K) \to \overline{E}_{\mathrm{sm}}(k)$ is a group homomorphism. In fact, it is surjective by Hensel's lemma.

(v) We let $E_1(K)$ be the kernel of the reduction map $E_0(K) \to \overline{E}_{\mathrm{sm}}(k)$.

### Types of reduction

The curve $\overline{E}$ is defined by the equation $y^2 = x^3 + \overline{a}x + \overline{b}$, where $\overline{a}$ and $\overline{b}$ are the images of $a$ and $b$ in $k$. This curve is an elliptic curve if and only if $\overline{\Delta} \neq 0$, which is equivalent to asking that $\Delta$ be a unit of $R$. If $\overline{E}$ is an elliptic curve, we say that $E$ has good reduction. In this case, $\mathcal{E}$ is a smooth scheme over $R$ and is naturally a group object in the category of schemes over $R$.

Now suppose $\overline{E}$ is singular, i.e., $\overline{\Delta} = 0$. We then say that $E$ has bad reduction. There are two possibilities. If $\overline{a} = \overline{b} = 0$ then $\overline{E}$ has a single singularity, at $(0, 0)$, and it is a cusp. The smooth locus $\overline{E}_{\mathrm{sm}}$ is isomorphic to $\mathbb{G}_a$, as a group variety. We therefore say that $E$ has additive reduction. If $\overline{a}$ or $\overline{b}$ is non-zero then both are non-zero (since $\overline{\Delta} = 0$), and $E$ has a single singularity, at $(-3b/2a, 0)$, and it is a node. The smooth locus $\overline{E}_{\mathrm{sm}}$ is isomorphic (over $\overline{k}$) to $\mathbb{G}_m$, as a group variety. We therefore say that $E$ has multiplicative reduction.

To summarize:

(i) $E$ has good reduction if and only if $\Delta$ is a unit of $R$.

(ii) $E$ has multiplicative reduction if and only if $\Delta \in \mathfrak{f}$ but $a$ and $b$ are units of $R$.

(iii) $E$ has additive reduction if and only if $a$ and $b$ are both in $\mathfrak{f}$.

We say that $E$ has semi-stable reduction if it has either good or multiplicative reduction. This is equivalent to one of $a$ or $b$ being a unit of $R$.

### Behavior of reduction type under extensions

**Proposition.** Let $K'/K$ be a finite extension. Suppose that either $K'/K$ is unramified or $E$ has semi-stable reduction over $K$. Then a minimal Weierstrass equation for $E$ over $K$ is still minimal over $K'$. It follows that the reduction type of $E$ over $K$ is the same as that over $K'$.

*Proof.* Let $v'$ be the valuation on $K'$. First suppose that $K'/K$ is unramified. Then for $x \in K$ we have $v(x) = v'(x)$. Thus if $v(a) < 4$ or $v(b) < 6$ then $v'(a) < 4$ or $v'(b) < 6$. Now suppose that $E$ has semi-stable reduction. Then either $v(a) = 0$ or $v(b) = 0$, and so $v'(a) = 0$ or $v'(b) = 0$, which shows that the equation is minimal over $K'$. □

**Theorem** (Semi-stable reduction theorem)**.** There exists a finite extension $K'/K$ such that $E$ has semi-stable reduction over $K'$.

*Proof.* Recall that we can make a change of variables to replace $(a, b)$ with $(a', b') = (u^{-4}a, u^{-6}b)$. First suppose that $3v(a) \leq 2v(b)$. Taking $u = a^{1/4}$, we find that $a' = 1$ is a unit and $b'$ is integral, so the new equation is minimal and has semi-stable reduction. Thus $E$ has semi-stable reduction over $K' = K(u)$. Now suppose that $3v(a) \geq 2v(b)$. Taking $u = b^{1/6}$, we find that $a'$ is integral and $b' = 1$ is a unit, so the new equation is minimal and has semi-stable reduction. Thus $E$ has semi-stable reduction over $K' = K(u)$. □

**Remark.** The proof shows that the extension $K'/K$ can always be taken to have degree at most 6.

Combining the above two results, we see that for all sufficiently large extensions $K'/K$, the curve $E_{K'}$ has either good or multiplicative reduction (independent of $K'$). We say that $E$ has potentially good or potentially multiplicative reduction accordingly. There is a simple test to determine which, in terms of the equation for $E$:

**Proposition.** $E$ has potentially good reduction if and only if $j(E) = -1728(4a)^3/\Delta$ is integral.

*Proof.* Since the $j$-invariant is independent of the model, we may as well assume that we have passed to an extension where $E$ is semi-stable and we are working with the minimal model. If $E$ has good reduction then $\Delta$ is a unit, and $j(E)$ is integral. If $E$ has multiplicative reduction then $\Delta$ is not a unit but $a$ is, and so $j(E)$ is not integral. □

**Example.** Suppose $E$ is the curve over $\mathbb{Q}_p$ given by $y^2 = x^3 + p$. Then $E$ has additive reduction. We have $a = 0$ and $b = p$, so $j = 0$ is integral, and so $E$ has potentially good reduction. Indeed, changing $y$ to $p^{1/2}y$ and $x$ to $p^{1/3}x$, we find that $E$ is isomorphic to $y^2 = x^3 + 1$ over $\mathbb{Q}_p(p^{1/6})$, which is still an elliptic curve mod $p$ (since $p \geq 5$).

**Reduction of torsion points**

We assume in this section that $E$ has good reduction. Since $\mathcal{E}$ is a proper smooth group over $R$, its $n$-torsion $\mathcal{E}[n]$ is a finite flat group scheme over $R$, for any $n$. We can therefore apply our knowledge of group schemes to its study.

**Proposition.** Let $G$ be a finite flat group scheme over $R$ whose order is prime to the residue characteristic. Then the reduction map $G(\overline{K}) \to G(\overline{k})$ is an isomorphism of Galois modules. In particular, $G(\overline{K})$ is an unramified Galois module.

*Proof.* The reduction map is obviously Galois equivariant, so it suffices to show it's a bijection. To do this, we can assume $k$ is algebraically closed. Since the order of $G$ is invertible on the base, $G$ is étale. Thus, if $G = \mathrm{Spec}(A)$, then $A$ is a product of copies of $R$. Clearly then, $G(\overline{K}) = G(K) = G(k)$. □

**Corollary.** Suppose $E$ has good reduction and $n$ is prime to the residue characteristic. Then the reduction map $E[n](\overline{K}) \to \overline{E}[n](\overline{k})$ is an isomorphism of Galois modules. In particular, $E[n](\overline{K})$ is an unramified Galois module.

Using Raynaud's theorem, we can say something about the $p$-torsion when the residue characteristic is $p$.

**Proposition.** Suppose $K$ is an extension of $\mathbb{Q}_p$ with $e < p - 1$. Let $G$ be a finite flat group scheme over $R$. Then the map $G(R) \to G(k)$ is injective.

*Proof.* Let $\Gamma$ be the group $G(R)$, regarded as a constant group scheme over $R$. There is a natural map $\Gamma \to G$ of group schemes over $R$, inducing the identity on $R$-points. Let $\overline{\Gamma}$ be the scheme-theoretic image of this map in $G$, which is a closed subgroup of $G$. (One can also describe $\overline{\Gamma}$ as the scheme-theoretic closure of $G(K)$ in $G$.) Since the map $\Gamma \to \overline{\Gamma}$ is an isomorphism on the generic fibers, Raynaud's theorem implies that it is an isomorphism. It follows that $\Gamma_k \to G_k$ is injective; since $\Gamma(R) \to \Gamma(k)$ is bijective (as $\Gamma$ is constant), the composite $G(R) = \Gamma(R) \to G(k)$ is injective. □

**Remark.** In the above situation, the reduction map need not be surjective. For example, let $G$ be the Kummer extension of $\mathbb{Z}/p\mathbb{Z}$ by $\mu_p$ corresponding to $a \in R$. If $A$ is a connected $R$-algebra, then $G(A)$ is the set of pairs $(i, z)$, where $i \in \mathbb{Z}/p\mathbb{Z}$ and $z \in A$ satisfies $z^p = a^i$. If $R$ does not contain a primitive $p$th root of unity or a $p$th root of $a$ then $G(R) = 0$. But if $k$ is perfect then $G_k$ is the trivial extension (since $a$ has a $p$th root), so $G(k) = \mathbb{Z}/p\mathbb{Z}$.

**Remark.** Without the assumption on $e$, the reduction map need not be injective. For example, take $G = \mu_p$ and suppose $K$ contains the $p$th roots of unity. Then $G(R) = \mu_p(K)$ has order $p$ but $G(k)$ is the trivial group.

**Corollary.** Suppose $E$ has good reduction and maintain the same assumptions on $K$. Then the reduction map $E[n](K) \to \overline{E}[n](k)$ is injective.

### The kernel of reduction

We now study the group $E_1(K)$, the kernel of the homomorphism $E_0(K) \to \overline{E}_{\mathrm{sm}}(k)$. Since points on $E_1(k)$ are $\mathfrak{f}$-adically close to the identity, the point at infinity, it makes sense to switch coordinates so that the identity is at $(0, 0)$. The projective equation for $E$ is

$$ZY^2 = X^3 + aZ^2X + bZ^3$$

We usually put $x = X/Z$ and $y = Y/Z$. We now put $u = X/Y$ and $v = Z/Y$ to obtain the equation

$$v = u^3 + auv^2 + bv^3$$

The point at infinity in projective coordinates is $[0 : 1 : 0]$, and thus corresponds to $(u, v) = (0, 0)$. The set $E_1(K)$ is given by the set of solutions to the above equation with $u$ and $v$ in $\mathfrak{f}$.

Let $F(u, v)$ be the right side of the above equation, so that the equation reads $v = F(u, v)$. We can then plug this expression for $v$ into the right side to find $v = F(u, F(u, v))$. Continuing in this way, we find $v = \phi(u)$, where $\phi(u)$ is the iterate $F(u, F(u, F(u, \ldots)))$. It is not difficult to see that $\phi(u)$ is a power series in $u$ with coefficients in $R$. Note that, because $R$ is complete, if $u$ is an element of $\mathfrak{f}$ then $\phi(u)$ is a well-defined element of $R$, and in fact $\mathfrak{f}$ since $\phi(0) = 0$. It is now an easy exercise to show:

**Proposition.** The map $\mathfrak{f} \to E_1(K)$ sending $u$ to $(u, \phi(u))$ is a bijection of sets taking 0 to the identity element of $E_1(K)$.

Using this bijection, we can transfer the group structure on $E_1(K)$ to a group structure on $\mathfrak{f}$, which we denote by $\oplus$. It is not hard to show that $\oplus$ is given by a power series over $R$, i.e., there exists a power series $G \in R[\![s, t]\!]$ such that $s \oplus t = G(s, t)$. Since 0 is the identity element, we have $G(s, 0) = G(0, s) = s$, and so $G(s, t) = s + t + \cdots$, where $\cdots$ are higher order terms. It follows that $\mathfrak{f}^n$ is a subgroup of $\mathfrak{f}$ under $\oplus$. Let $E_n(K)$ be the corresponding subgroup of $E_1(K)$. Clearly then, we have an isomorphism

$$E_n(K)/E_{n+1}(K) = \mathfrak{f}^n/\mathfrak{f}^{n+1} = k$$

We have thus proved:

**Proposition.** The group $E_1(K)$ has a decreasing filtration $\{E_n(K)\}_{n\geq 1}$ such that $E_n(K)/E_{n+1}(K)$ is isomorphic to $k$.

**Corollary.** Suppose $n$ is prime to the residue characteristic. Then the map $E_0(K)[n] \to E_{\mathrm{sm}}(k)[n]$ is injective.

*Proof.* The kernel is a subgroup of $E_1(K)$ killed by $n$, and therefore 0. □

**Corollary.** Suppose $k$ is finite of characteristic $p$. Then $E_1(K)$ is a pro-$p$ group.

### The quotient of $E(K)$ by $E_0(K)$

We have the following important result:

**Theorem.** The group $E(K)/E_0(K)$ is finite. In fact, if $E$ has split multiplicative reduction (i.e., $\overline{E}_{\mathrm{sm}}$ is isomorphic to $\mathbb{G}_m$ over $k$) then this group is cyclic of order $-v(j)$; otherwise, it has cardinality at most 4.

We will not prove this theorem. Some remarks:

(i) The finiteness statement follows immediately from the existence of Néron models. The more precise statement about the structure and cardinality of the group follows from the classification of Néron models. We will discuss these topics in the next section.

(ii) If $k$ is finite then the finiteness statement is easy, for $E(K)$ is then a compact group and $E_0(K)$ is an open subgroup; thus $E(K)/E_0(K)$ is both discrete and compact, and thus finite.

(iii) One can prove the theorem without Néron models through a case-by-case analysis. For instance, suppose $v(a) = 1$ and $v(b) \geq 2$. If $P = (x, y)$ is a point of $E(K)$ then

$$x(2P) = \frac{x^4 - 2ax^2 - 8bx + a^2}{4(x^3 + ax + b)}$$

Thus if $(x, y)$ reduces to the singular point $(0, 0)$, i.e., $v(x) \geq 1$, then the valuation of the numerator is equal to 2, while the valuation of the denominator is at least 2; thus $v(x(2P)) \leq 0$, and so $2p$ does not reduce to the singular point. This shows that $E(K)/E_0(K)$ is killed by 2. In fact, one can show that the sum of any two points reducing to the singular point does not reduce to the singular point, and so $E(K)/E_0(K) = \mathbb{Z}/2\mathbb{Z}$.

### The Néron–Ogg–Shafarevich criterion

Let $G_K$ be the absolute Galois group of $K$ and $I_K$ the inertia subgroup.

**Theorem.** Let $\ell$ be a prime different from the residue characteristic. Then:

(i) $E$ has good reduction if and only if $I_K$ acts trivially on $T_\ell(E)$.

(ii) $E$ has semi-stable reduction if and only if $I_K$ acts unipotently on $T_\ell(E)$.

*Proof.* First, note that $I_K$ acts trivially on $T_\ell(E)$ if and only if it does so on $E[\ell^n](\overline{K})$ for all $n$. Thus, if $E$ has good reduction then $I_K$ acts trivially on $T_\ell(E)$ by what we've already shown. Conversely, suppose $I_K$ acts trivially on $T_\ell(E)$. Thus all $\ell^n$ torsion points belong to $E(K^{\mathrm{un}})$. Let $d$ be the order of $E(K^{\mathrm{un}})/E_0(K^{\mathrm{un}})$, which is finite. Then $E_0(K^{\mathrm{un}})[\ell^n]$ is the kernel of the map $E(K^{\mathrm{un}})[\ell^n] \to E(K^{\mathrm{un}})/E_0(K^{\mathrm{un}})$, and thus has cardinality at least $\ell^{2n}/d$. Since the reduction map $E_0(K^{\mathrm{un}}) \to \overline{E}_{\mathrm{sm}}(\overline{k})$ is injective on $\ell$-power torsion, it follows that $\overline{E}_{\mathrm{sm}}(\overline{k})[\ell^n]$ has cardinality at least $\ell^{2n}/d$. But this is not true for $\mathbb{G}_m$ (where the cardinality is $\ell^n$) or $\mathbb{G}_a$ (where the cardinality is 1), and so $E$ cannot have multiplicative or additive reduction. Thus $E$ has good reduction.

Now suppose that $I_K$ acts unipotently on $T_\ell(E)$. It thus fixes some vector in $T_\ell(E)$, which implies that $E(K^{\mathrm{un}})[\ell^n]$ has cardinality at least $\ell^n$. Arguing as in the previous paragraph, we see that $\overline{E}_{\mathrm{sm}}$ cannot be $\mathbb{G}_a$, and so $E$ has semi-stable reduction.

Finally, suppose that $E$ has semi-stable reduction. The multiplication-by-$\ell^n$ map on the smooth locus $\mathcal{E}_{\mathrm{sm}}$ of $\mathcal{E}$ is flat, and so $\mathcal{E}_{\mathrm{sm}}[\ell^n]$ is a flat group scheme over $R$. Let $G$ be the scheme-theoretic closure in $\mathcal{E}_{\mathrm{sm}}[\ell^n]$ of the set of $\overline{K}$-points which extend to $\overline{R}$-points. Then $G$ is finite and flat, and $G_k = \overline{E}_{\mathrm{sm}}[\ell^n]$. Since $G$ has $\ell$-power order, it is étale, and so $G(K^{\mathrm{un}}) = \overline{E}_{\mathrm{sm}}[\ell^n](\overline{k})$, which contains $\mathbb{Z}/\ell^n\mathbb{Z}$ (since $E$ is semi-stable). Thus $E[\ell^n](K^{\mathrm{un}})$ contains $\mathbb{Z}/\ell^n\mathbb{Z}$ for all $n$, which shows that $I_K$ fixes a vector in $T_\ell(E)$. Since the determinant of $T_\ell(E)$ is the $\ell$-cyclotomic character, which is trivial on $I_K$, the result follows. □

**Corollary.** If $I_K$ acts trivially (or unipotently) on one $T_\ell(E)$ then it does so on all of them.

**Corollary.** $E$ has potentially good reduction if and only if $I_K$ acts through a finite quotient on $T_\ell(E)$.

**Corollary.** Isogenous curves have the same reduction type.

*Proof.* If $E$ and $E'$ are isogenous then $T_\ell(E)[1/\ell]$ and $T_\ell(E)[1/\ell]$ are isomorphic $\mathbb{Q}_\ell$ representations of $I_K$. □

The next section is an exposition on Néron models. I begin by discussing quasi-finite étale group schemes over DVRs: these are the sorts of things that occur as the prime-to-$p$ torsion of Néron models. Then, before going to the general theory, I discuss Néron models of elliptic curves, especially their relationship to Weierstrass and minimal regular models. I do two simple and explicit examples. Finally, I introduce Néron models of general abelian varieties. Two applications are discussed: the Néron–Ogg–Shafarevich criterion, and Grothendieck's generalization thereof; and the semi-stable reduction theorem.

### 1.3.2 Quasi-finite étale group schemes

At the end of the last section, we considered the group scheme obtained by taking the $n$ torsion in the smooth locus of a minimal Weierstrass model for an elliptic curve. This group scheme is typically not finite over the base. However, it is quasi-finite: all its fibers have finitely many points. We now study such group schemes in the étale case.

Let $R$ be a henselian DVR, and keep our usual notation ($K$, $k$, etc). Let $G$ be a quasi-finite étale group scheme over $R$ (assumed to be of finite presentation and commutative). Let $M = G(\overline{K})$ be the Galois module corresponding to $G_K$, and let $M_0 = G(\overline{k})$ be the one corresponding to $G_k$. Since $G$ is étale, the natural map $G(\overline{R}) \to G(\overline{k})$ is an isomorphism, and so we can regard $M_0$ as a submodule of $M$. It is obviously stable under the Galois action and fixed by inertia.

**Theorem.** The functor $G \mapsto (M, M_0)$ is an equivalence of categories.

Some comments:

(i) Let $G$ correspond to $(M, M_0)$ and let $H$ be a subgroup corresponding to $(N, N_0)$. Then $H$ is closed in $G$ if and only if $N_0 = M_0 \cap N$. In this case, $G/H$ is an étale quasi-finite group scheme, and it corresponds to $(M/N, M_0/N_0)$.

(ii) Let $\mathcal{G}$ be a finite group scheme over $K$, corresponding to the Galois module $M$. Then $\mathcal{G}$ admits a maximal extension to an étale quasi-finite group over $R$, by taking $M_0 = M^{I_K}$. It also admits a minimal such extension, by taking $M_0 = 0$; we call this the extension by zero.

(iii) Let $G$ correspond to $(M, M_0)$, and let $H$ be the closed subgroup corresponding to $(M_0, M_0)$. Then $H$ is the maximal closed subgroup of $G$ which is finite over $R$. Note that $H_k = G_k$.

(iv) Suppose $G$ is a quasi-finite flat group scheme over $R$ which is killed by $n$, and $n$ is invertible on $R$. Then $G_K$ and $G_k$ are both étale, and this implies that $G$ itself is étale. In particular, if $\mathcal{E}$ is some smooth commutative group variety over $R$ and $\mathcal{E}[n]$ is quasi-finite, then it is étale as well.

**Remark.** At the end of the previous section, we proved that if $E$ is an elliptic curve with semi-stable reduction, then $I_K$ fixes a vector in $T_\ell(E)$. Let us re-explain the argument with the above theory in hand. Let $G$ be the $\ell^n$-torsion in the smooth part of the minimal Weiestrass model. Then this is a quasi-finite étale group scheme over $R$. Let $H \subset G$ be the maximal finite subgroup. Then $H_k = G_k$, and so $H(\overline{k})$ contains a point of order $\ell^n$ (since it is the $\ell^n$-torsion in either an elliptic curve or $\mathbb{G}_m$). Since $H$ is étale over $R$, the map $H(K^{\mathrm{un}}) \to H(\overline{k})$ is an isomorphism. Of course, $H(K^{\mathrm{un}}) \subset E[\ell^n](K^{\mathrm{un}})$, so this shows that $E$ contains a point of order $\ell^n$ defined over $K^{\mathrm{un}}$.

### 1.3.3 Néron models of elliptic curves

A good reference for this section is Chapter IV of Silverman's book "Advanced topics in the arithmetic of elliptic curves" (MR1312368).

**Motivation**

Let $E/K$ be an elliptic curve and let $\mathcal{W}/R$ be its minimal Weierstrass model. Since $\mathcal{W}$ is proper over $R$, we have $\mathcal{W}(R) = \mathcal{W}(K) = E(K)$. However, $\mathcal{W}$ is typically singular. Its smooth locus $\mathcal{W}_{\mathrm{sm}}$ is a group scheme over $R$. Typically, it is not proper, and not all $K$-points of $E$ extend to $\mathcal{W}_{\mathrm{sm}}$. Those that do are the subgroup $E_0(K)$, which has finite index in $E(K)$.

The Néron model is an extension $\mathcal{E}$ of $E$ over $R$ which combines the desirable properties of $\mathcal{W}$ and $\mathcal{W}_{\mathrm{sm}}$: it is a smooth group scheme and all $K$-points extend to $R$-points. The identity component of $\mathcal{E}$ is $\mathcal{W}_{\mathrm{sm}}$, while the component group of $\mathcal{E}_k$ (at least for $k$ algebraically closed) is $E(K)/E_0(K)$. So all the points of $E(K)$ extend to points of $\mathcal{E}(R)$, and $E_0(K)$ is the subgroup of points which extend to the identity component of $\mathcal{E}$.

**Minimal regular models and Neron models**

Let $C/K$ be a curve. A regular model for $C$ is a proper flat scheme $\mathcal{C}$ over $R$ which is regular and whose generic fiber is $C$. A regular model $\mathcal{C}$ is minimal if for any other regular model $\mathcal{C}'$, there exists a map of schemes $\mathcal{C}' \to \mathcal{C}$ extending the identity on the generic fiber. The main theorem is that minimal regular models exist and are canonically unique. One can find a regular model for $C$ by starting with any model and repeatedly blowing-up and normalizing. From there, one can find a minimal regular model by blowing-down certain divisors in the special fiber.

Let $E/K$ be an elliptic curve and let $\mathcal{C}/R$ be its minimal regular model. The Néron model of $E$ is then the smooth locus in $\mathcal{C}$. (This can be taken as a definition, though a better definition is given below.)

**Example 1**

Consider the curve $y^2 = x^3 + p$ over $K = \mathbb{Q}_p$. The same equation defines the minimal Weierstrass model $\mathcal{W}$ over $R = \mathbb{Z}_p$. Clearly, $\mathcal{W}$ is smooth everywhere except for the point $P = (0, 0)$ in the special fiber.

We claim that $P$ is regular. To see this, let $A = R[x, y]/(y^2 = x^3 + p)$ be the ring of natural affine chart containing $P$, so that $P$ corresponds to the maximal ideal $\mathfrak{m} = (x, y, p)$. The ideal $\mathfrak{m}^2$ is generated by $x^2, xy, y^2, px, py, p^2$. But note that $y^2 = x^3 + p$, and $x^3 \in \mathfrak{m}$, so we may as well replace the generator $y^2$ with $p$, which means the generators $px$, $py$, and $p^2$ are unnecessary. Thus $\mathfrak{m}^2 = (x^2, xy, p)$. The quotient $\mathfrak{m}/\mathfrak{m}^2$ is has for a basis the images of $x$ and $y$, and is thus two dimensional over the residue field $A/\mathfrak{m}$. Since $A$ has Krull dimension 2, this establishes regularity.

It follows that $\mathcal{W}$ is a regular model for $E$, which is necessarily minimal since there are no divisors in the special fiber to blow-down. The Néron model $\mathcal{E}$ is the smooth locus of $\mathcal{W}$, i.e., $\mathcal{W} \setminus \{P\}$. In particular, the special fiber $\mathcal{E}_k$ is connected and isomorphic to $\mathbb{G}_a$. We have $E(K) = E_0(K)$ in this case.

**Example 2**

Now consider the curve $E$ defined by $y^2 = x^3 + p^2$. Again, this equation defines the minimal Weierstrass model $\mathcal{W}$ over $R$ and $P = (0, 0)$ in the special fiber is the unique singular point.

In this case, $P$ is not regular. Let $A = R[x, y]/(y^2 = x^3 + p^2)$ and $\mathfrak{m} = (x, y, p)$, similar to before. The generators of $\mathfrak{m}^2$ are similar to before. The difference is that one can no longer use the defining equation to find $p$ in $\mathfrak{m}^2$; in fact, the equation shows that $y^2$ is not needed as a generator of $\mathfrak{m}^2$. Thus $\mathfrak{m}^2 = (x^2, xy, px, py, p^2)$. The images of $x$, $y$, and $p$ in $\mathfrak{m}/\mathfrak{m}^2$ are linearly independent (and in fact a basis), so $\mathfrak{m}/\mathfrak{m}^2$ is 3 dimensional, and so $P$ is not regular.

To find the minimal regular model of $E$, we blow up at the point $P$. We'll only do the computations in the affine chart $\mathrm{Spec}(A)$. The blow-up algebra $B$ is the subring of $A[t]$ generated by $tx$, $ty$, and $tp$. This is considered as a graded ring by giving $t$ degree 1. The blow-up is $\mathrm{Proj}(B)$. Let $B_1$ be the degree 0 subring of $B[1/tx]$, and define $B_2$ and $B_2$ similarly but with $ty$ and $tp$. Let $U_i = \mathrm{Spec}(B_i)$. Then $\mathrm{Proj}(B)$ is covered by the $U_i$, so we first study them.

The ring $B_1$ can be presented as the quotient of $R[x, y/x, p/x]$ by the equations $x(p/x) = p$ and $(y/x)^2 = x + (p/x)^2$. (One should think of $y/x$ and $p/x$ as indeterminates.) The special fiber is therefore defined by the equations $x(p/x) = 0$ and $(y/x)^2 = x + (p/x)^2$. This is a union of three lines: when $x = 0$ we get $(y/x) = \pm(p/x)$ and when $(p/x) = 0$ we get $x = (y/x)^2$. The three lines intersect at the point $x = (p/x) = (y/x) = 0$.

The ring $B_2$ can be presented as the quotient of $R[y, x/y, p/y]$ by the equations $y(p/y) = p$ and $1 = y(x/y)^3 + (p/y)^2$. Its special fiber is defined by $y(p/y) = 0$ and $1 = y(x/y)^3 + (p/y)^2$. This is also a union of three lines: when $y = 0$ we get $(p/y) = \pm 1$ and when $(p/y) = 0$ we get $y = (x/y)^{-3}$. Note that these lines do not intersect, since $(p/y)$ is constant on each line of a different value. The two lines with $y = 0$ meet up with the two lines in $U_1$ with $x = 0$. Since $y/x$ can assume any non-zero value in $U_1$ and $x/y$ can assume any non-zero value in $U_2$, they glue to $\mathbb{P}^1$'s. The third line in $U_2$ is missing two points, and is contained in the third line in $U_1$.

Finally, the ring $B_3$ can be presented as the quotient of $R[x/p, y/p]$ by the equation $(y/p)^2 = p(x/p)^3 + 1$. Its special fiber consists of two lines, defines by $(x/p) = 0$ and $(y/p) = \pm 1$. Thus $U_3$ is contained in $U_1 \cup U_2$.

We thus see that the special fiber of $\mathrm{Proj}(B)$ has three components, two $\mathbb{P}^1$'s and one $\mathbb{A}^1$, and they are joined at a single point. However, $\mathrm{Proj}(B)$ is not the full blow-up of $\mathcal{W}$ at $P$, but only one chart. The other chart adds the missing point to the $\mathbb{A}^1$ in the special fiber.

This blow-up $\mathcal{C}$ is the minimal regular model for $E$. The Néron model $\mathcal{E}$ is obtained by deleting the intersection point in the special fiber. Thus $\mathcal{E}^\circ$ has three components, so its component group is necessarily $\mathbb{Z}/3\mathbb{Z}$.

#### Classification of minimal regular models

If $E$ has good reduction, then its minimal Weierstrass model is smooth, and coincides with its minimal regular model and Néron model. In this case, the special fiber of the minimal regular model is an elliptic curve.

In all other cases, the special fiber of the minimal regular model is made up of genus 0 curves, though they may have singularities and non-reduced behavior. This data is combinatorial, since one simply needs to record how many $\mathbb{P}^1$'s there are, how they intersect, and what their multiplicities are. It can be depicted as a sort of graph, with numbers on the edges to denote multiplicities. Néron and Kodiara classified all the possible special fibers; the graphs that occur turn out to be closely related to Dynkin diagrams. An important fact that follows from this classification is that, unless $E$ has split multiplicative reduction, the special fiber of its Neron model has at most 4 components.

### 1.3.4 Néron models for abelian varieties

#### Definition and basic properties

It is not at all clear how to extend our discussion of elliptic curves to higher dimensional abelian varieties: the theory of Weierstrass models relies on explicit equations, which are unavailable, while the more abstract theory of minimal regular models is specific to curves. The key observation is that the functor of points of the Néron model of an elliptic curve admits a nice description.

**Theorem.** Let $E/K$ be an elliptic curve, and let $\mathcal{E}/R$ be its Néron model. Let $\mathcal{X}/R$ be any smooth scheme, and let $X = \mathcal{X}_K$. Then the natural map

$$\mathrm{Hom}_R(\mathcal{X}, \mathcal{E}) \to \mathrm{Hom}_K(X, E)$$

is an isomorphism.

Given this description of $\mathcal{E}$, it is clear how the definition can be extended to any scheme:

**Definition.** Let $A/K$ be a smooth scheme. A Néron model for $A$ is a smooth scheme $\mathcal{A}/R$ which satisfies the Néron mapping property: the natural map

$$\mathrm{Hom}_R(\mathcal{X}, \mathcal{A}) \to \mathrm{Hom}_K(X, A)$$

is a bijection, for any smooth scheme $\mathcal{X}/R$ as above.

Some remarks:

(i) The definition of Néron model specifies its functor of points on smooth $R$-schemes. Since the Néron model itself is required to be a smooth $R$-scheme, Yoneda's lemma shows that Néron models are canonically unique, when they exist.

(ii) Although the definition applies to any smooth scheme $A/K$, we only consider the case where $A$ is an abelian variety.

(iii) The main existence result is that the Néron model of an abelian variety exists.

(iv) As a special case of the Néron mapping property, we see that the natural map $\mathcal{A}(R) \to A(K)$ is a bijection, i.e., all $K$-points of $A$ extend to $R$-points of $\mathcal{A}$. Thus, from the perspective of $K$-points, the Néron model behaves as if it were proper. This is not true for $K'$-points if $K'/K$ is a ramified extension!

(v) Formation of Néron models is compatible with passing to unramified extensions, but not to ramified extensions, in general. Precisely, suppose $K'/K$ is a finite extension, let $\mathcal{A}$ be the Néron model of $A$ and let $\mathcal{A}'$ be the Néron model of $A_{K'}$. Then there is a natural map $\mathcal{A}_{R'} \to \mathcal{A}'$. If $K'/K$ is unramified this map is an isomorphism, but when $K'/K$ is ramified it is typically not. In particular, the natural map $\mathcal{A}(R') \to A(K')$ need not be an isomorphism.

**Types of reduction** Let $A/K$ be an abelian variety with Néron model $\mathcal{A}$, and let $\mathcal{A}_0$ be the special fiber of $\mathcal{A}$. Let $\mathcal{A}_0^\circ$ be its identity component. A theorem of Chevalley states that every smooth connected group is an extension of an abelian variety by a smooth affine group. Thus there is an exact sequence

$$0 \to L \to \mathcal{A}_0^\circ \to B \to 0$$

where $B$ is an abelian variety and $L$ is a commutative smooth affine group. The group $L$ contains a maximal torus $T$ such that the quotient $U = L/T$ is unipotent (a product of $\mathbb{G}_a$'s). In other words, we can say that there is a canonical filtration

$$0 = F_0 \subset F_1 \subset F_2 \subset F_3 \subset F_4 = \mathcal{A}_0$$

where $T = F_1/F_0$ is a torus, $U = F_2/F_1$ is unipotent, $B = F_3/F_2$ is an abelian variety, and $F_4/F_3$ is finite étale (the component group). The dimensions of $T$, $U$, and $B$ are important invariants of $A$ refining the trichotomy of multiplicative/additive/good reduction in the case of elliptic curves.

We say that $A$ has good reduction if it extends to an abelian scheme over $R$. (An abelian scheme is a smooth proper group scheme with geometrically connected fibers.) This is equivalent to $\mathcal{A}_0$ (or just $\mathcal{A}_0^\circ$) being an abelian variety. If $A$ has good reduction then $\mathcal{A}$ is the unique abelian scheme extending it.

We say that $A$ has semi-stable reduction if $\mathcal{A}_0$ has no unipotent part, i.e., $\mathcal{A}_0^\circ$ is an extension of an abelian variety by a torus (what is called a semi-abelian variety).

**Neron–Ogg–Shafarevich**

Let $\ell$ be a prime different from the residue characteristic and $T_\ell(A)$ the $\ell$-adic Tate module of $A$, a representation of $G_K$.

**Theorem.** $A$ has good reduction if and only if $T_\ell(A)$ is unramified.

*Proof.* The proof is the same as the elliptic curve case. We briefly recall the details. First, if $A$ has good reduction then it extends to an abelian scheme $\mathcal{A}$, so $\mathcal{A}[\ell^n]$ is a finite étale group scheme over $R$, which implies that $T_\ell(A)$ is unramified. Conversely, suppose $T_\ell(A)$ is unramified. Then all the $\ell^n$-torsion of $A$ is defined over $K^{\mathrm{un}}$, and so $\mathcal{A}(K^{\mathrm{un}})[\ell^n]$ has cardinality $\ell^{2ng}$, where $g = \dim(A)$. Since the reduction map $A(K^{\mathrm{un}})[\ell^n] \to \mathcal{A}(\overline{k})$ is injective, we see that $\mathcal{A}(\overline{k})[\ell^n]$ has cardinality at least $\ell^{2ng}$. Using the fact that $\mathbb{G}_m$ and $\mathbb{G}_a$ have too little $\ell^n$-torsion, this implies that there can be no toric or unipotent part of $\mathcal{A}_0$, and so $\mathcal{A}_0^\circ$ is an abelian variety, which implies good reduction. □

There is an important generalization of this theorem, due to Grothendieck. We do not give the proof.

**Theorem.** $A$ has semi-stable reduction if and only if the action of $I_K$ on $T_\ell(A)$ is unipotent.

**Semi-stable reduction theorem**

**Theorem.** There exists a finite extension $K'/K$ such that $A_{K'}$ has semi-stable reduction.

*Proof.* We assume $K$ is a finite extension of $\mathbb{Q}_p$ for simplicity. By Grothendieck's extension of Néron–Ogg–Shafarevich, it is enough to show that $I_{K'}$ acts unipotently on $T_\ell(A)$ for some finite $K'/K$. In fact, we will show that this is true for any $\ell$-adic representation of $G_K$!

Thus let $V$ be a continuous $\ell$-adic representation of $G_K$. Since the wild inertia subgroup of $G_K$ is pro-$p$, its image in $\mathrm{GL}(V)$ must be finite. Thus, passing to a finite extension, we can assume wild inertia acts trivially. The quotient of $G_K$ by the wild inertia subgroup is topologically generated by two elements, $F$ (a lift of Frobenius) and $\tau$ (a generator of tame inertia), which satisfy the single relation $F\tau F^{-1} = \tau^q$. This equation shows that the transformations $\tau$ and $\tau^q$ of $V$ are conjugate. Thus if $\alpha_1, \ldots, \alpha_n$ are the eigenvalues of $\tau$ then $\alpha_i^q = \alpha_{\sigma(i)}$ for some permutation $\sigma \in S_n$. This implies that $\alpha_i^{q^n} = \alpha_i$ for all $i$, i.e., the $\alpha_i$ are roots of unity of order dividing $q^n - 1$. Thus, by passing to an extension of $K$ with ramification index $e = q^n - 1$ (which has the effect of replacing $\tau$ by $\tau^e$), the action of $\tau$ (and thus all of inertia) becomes unipotent. □

## 1.4 Jacobians

This section is an exposition on Jacobians. I start with the analytic theory, where the Jacobian is defined as the quotient of a vector space by a lattice. I then discuss the representability issues of the functor of points over arbitrary fields. Following this, I briefly sketch Weil's construction of the Jacobian and say something about the relative situation.

### 1.4.1 Analytic Theory

**Hodge theory of curves** Let $X$ be a (connected, smooth, projective) curve over the complex numbers of genus $g$. Let $V$ be the space of global holomorphic 1-forms on $X$, which is a complex vector space of dimension $g$. Let $\mathrm{H}^1_{\mathrm{dR}}(X)$ be the first de Rham cohomology group of $X$, i.e., the space of smooth closed real 1-forms modulo exact forms. Since every element of $V$ is closed, we have a natural map $V \to \mathrm{H}^1_{\mathrm{dR}}(X) \otimes_{\mathbb{R}} \mathbb{C}$.

**Lemma.** This map is injective.

*Proof.* Suppose $\omega \in V$ is closed, and write $\omega = df$. Then $f$ is holomorphic: indeed, in local coordinates, this expression implies that $df$ has no $d\overline{z}$, which is exactly the Cauchy–Riemann equations. Thus $f$ is a holomorphic function on all of $X$, and therefore constant, and so $\omega = 0$. □

**Theorem** (Hodge Decomposition)**.** The map $V \oplus \overline{V} \to \mathrm{H}^1_{\mathrm{dR}}(X) \otimes_{\mathbb{R}} \mathbb{C}$ is an isomorphism.

*Proof.* Let $J \colon T_x \to T_x$ be the multiplication by $i$ map on tangent spaces. For a complex 1-form $\omega$ on $X$, define $\omega^c$ by $-i\omega J$. Then $(-)^c$ induces an involution of $\mathrm{H}^1_{\mathrm{dR}}(X) \otimes_{\mathbb{R}} \mathbb{C}$. Clearly, $V$ lies in the 1 eigenspace of this operator, while $\overline{V}$ lies in the $-1$ eigenspace. Thus $V \cap \overline{V} = 0$ in $\mathrm{H}^1_{\mathrm{dR}}(X) \otimes_{\mathbb{R}} \mathbb{C}$, and so the map in question is injective. Since both spaces have complex dimension $2g$, it is also surjective. □

**Proposition.** Let $p \colon \mathrm{H}^1_{\mathrm{dR}}(X) \otimes_{\mathbb{R}} \mathbb{C} \to V$ be the projection map. Then $p$ induces an isomorphism of real vector spaces $\mathrm{H}^1_{\mathrm{dR}}(X) \to V$.

*Proof.* Suppose $\alpha$ is an element of $\mathrm{H}^1_{\mathrm{dR}}(X)$. Then in the decomposition $\alpha = \omega + \overline{\eta}$ with $\omega, \eta \in V$, we must have $\omega = \eta$, since $\alpha = \overline{\alpha}$. It follows that $\omega \mapsto \omega + \overline{\omega}$ is the inverse to $p$. □

**Proposition.** Let $\alpha, \beta \in \mathrm{H}^1_{\mathrm{dR}}(X)$. Let $\omega = p(\alpha)$ and $\eta = p(\beta)$. Then

$$\int_X \alpha \wedge \beta = 2\,\mathrm{Re} \int_X \omega \wedge \overline{\eta}$$

*Proof.* We have $\alpha = \omega + \overline{\omega}$ and $\beta = \eta + \overline{\eta}$, and so $\alpha \wedge \beta = \omega \wedge \overline{\eta} + \overline{\omega} \wedge \eta = 2\,\mathrm{Re}(\omega \wedge \overline{\eta})$. □

We define a Hermitian form $H$ on $V$ by

$$H(\omega, \eta) = 2i \int \omega \wedge \overline{\eta}$$

The factor of $i$ is required for the identity $H(\omega, \eta) = \overline{H(\eta, \omega)}$. The above proposition says that for $\alpha, \beta \in \mathrm{H}^1_{\mathrm{dR}}(X)$, we have $\int_X \alpha \wedge \beta = \mathrm{Im}(H(p(\alpha), p(\beta)))$.

#### Definition of the Jacobian

Let $L = \mathrm{H}_1(X, \mathbb{Z})$. Given $\gamma \in L$ and $\omega \in V$, we can integrate $\omega$ over $\gamma$ and get a number. For fixed $\gamma$, this defines a linear map $V \to \mathbb{C}$, and so we have a natural map $i \colon L \to V^*$.

**Proposition.** $i(L)$ is a lattice in $V^*$.

*Proof.* Let $i_{\mathbb{R}}$ denote the induced map $L \otimes \mathbb{R} \to V^*$. The real dual of $V^*$ is natural isomorphic to $V$, where an element $v \in V$ induces a linear map $V^* \to \mathbb{R}$ by taking the real value of the usual pairing. The dual of $i_{\mathbb{R}}$ is thus a real-linear map $V \to L^\vee \otimes \mathbb{R}$. Identifying $L^\vee \otimes \mathbb{R}$ with $\mathrm{H}^1_{\mathrm{dR}}(X)$, this map takes $\omega$ to $\omega + \overline{\omega}$. Indeed, the image of $\omega$ is supposed to be a real 1-form $\alpha$ such that $\mathrm{Re} \int_\gamma \omega = \int_\gamma \alpha$ for any $\gamma \in L$, and it is clear that $\alpha = \omega + \overline{\omega}$ does the job. Thus $i_{\mathbb{R}}$ is an isomorphism; indeed, it is the inverse to $p$. □

**Definition.** The Jacobian of $X$, denoted $\mathrm{Jac}(X)$, is the complex torus $V^*/L$.

We have a natural conjugate-linear isomorphism $j\colon V \to V^*$ given by $j(\omega) = H(-,\omega)$. Define a Hermitian form $H^*$ on $V^*$ by $H^*(\lambda,\mu) = H(j^{-1}(\mu), j^{-1}(\lambda))$; the order is switched so that $H^*$ is linear in its first slot.

**Proposition.** $\operatorname{Im} H^*(i(\gamma), i(\gamma')) = \langle \gamma, \gamma' \rangle$ for any $\gamma, \gamma' \in L$, where $\langle , \rangle$ is the intersection pairing on $L$.

*Proof.* The two pairings $\operatorname{Im} H^*$ and $\langle , \rangle$ both induce real skew-symmetric pairings on $L \otimes \mathbb{R}^*$. These can be transferred to the dual space, which is identified with $\mathrm{H}^1_{\mathrm{dR}}(X)$. The first is then given by $(\alpha,\beta) \mapsto \operatorname{Im} H(p(\alpha), p(\beta))$, while the second by $(\alpha, \beta) \mapsto \int_X \alpha \wedge \beta$. The equality of these two pairings has already been established. □

**Corollary.** $\mathrm{Jac}(X)$ is canonically a principally polarized abelian variety.

### Basic properties

Some elementary properties of $\mathrm{Jac}(X)$:

(i) The tangent space to $\mathrm{Jac}(X)$ at the identity is canonically isomorphic to $V^* = \mathrm{H}^1(X, \mathcal{O})$ (the identification here is Serre duality).

(ii) Dually, we see that the cotangent space to $\mathrm{Jac}(X)$ at 0 is $V$. It follows that $\mathrm{H}^0(\mathrm{Jac}(X), \Omega^1) = V = \mathrm{H}^0(X, \Omega^1)$.

(iii) We have a natural isomorphism $\mathrm{H}_1(\mathrm{Jac}(X), \mathbb{Z}) = \mathrm{H}_1(X, \mathbb{Z})$.

(iv) Fix a point $x \in X$. We then get a map $f_x\colon X \to \mathrm{Jac}(X)$ as follows. For $y \in X$, choose a path $\rho$ from $x$ to $y$ in $X$. We then get an element of $V^*$ by integrating over $\rho$. The choice of $\rho$ is not unique, but the difference of any two choices lies in $L$, so the resulting elements of $V^*$ is well-defined up to $i(L)$. The map $f_x$ takes $y$ to this element of $V^*/i(L)$.

(v) One can show that $f_x^*$ induces an isomorphism $\mathrm{H}^0(\mathrm{Jac}(X), \Omega^1) \to \mathrm{H}^0(X, \Omega^1)$.

We now give perhaps the most important property of $\mathrm{Jac}(X)$. Let $\mathrm{Pic}(X)$ denote the group of isomorphic classes of line bundles on $X$, and let $\mathrm{Pic}^0(X)$ be the subgroup consisting of those of degree 0.

**Proposition.** We have a natural isomorphism $\mathrm{Jac}(X) \to \mathrm{Pic}^0(X)$.

*Proof.* Consider the exponential sequence on $X$:

$$0 \to \mathbb{Z} \to \mathcal{O} \to \mathcal{O}^\times \to 0$$

Taking cohomology, we obtain an exact sequence

$$0 \to \mathrm{H}^1(X, \mathcal{O})/\mathrm{H}^1(X, \mathbb{Z}) \to \mathrm{H}^1(X, \mathcal{O}^\times) \to \mathrm{H}^2(X, \mathbb{Z}) \to 0$$

Identifying $\mathrm{H}^1(X, \mathcal{O})$ with $V^*$ by Serre duality, the group on the left is $\mathrm{Jac}(X)$. The group in the middle is $\mathrm{Pic}(X)$. We have a natural identification $\mathrm{H}^2(X, \mathbb{Z}) = \mathbb{Z}$, under which the right map is the degree map $\mathrm{Pic}(X) \to \mathbb{Z}$. Thus $\mathrm{Jac}(X)$ maps isomorphically to $\mathrm{Pic}^0(X)$. □

### 1.4.2 Algebraic Theory

A good reference for this section is Chapter III of Milne's notes on abelian varieties.

**Attempt at a definition**

The definition of the Jacobian in the complex case was inherently analytic, and does not carry over to the algebraic case. However, we can use one of the results we proved about the Jacobian, namely that its points parametrize degree 0 line bundles on $X$, to give a definition valid over any field. This is similar to our previous discussion of the dual abelian variety.

From now on, we fix a field $k$, and let $X/k$ be a curve. We want to give $\mathrm{Pic}^0(X)$ an algebraic structure. To do this, we must make sense of families of degree 0 line bundles on $X$ over a base scheme $T$. This is not hard to do: such a family is just a line bundle on $X_T = X \times T$, which restricts to degree 0 in each fiber. Let $F(T)$ be the set of isomorphism classes of such bundles. One might then hope that $F$ is representable, and define the Jacobian in this manner.

**First obstruction to representability**

Unfortunately, $F$ is not representable. There are two obstacles to representability. The first is that line bundles on $T$ cause problems. To see this, suppose $F$ were represented by some variety $J$. Let $L$ be a line bundle on $T$, and write $p\colon X_T \to T$ for the projection. The line bundle $p^*(L)$ on $X_T$ is trivial on each fiber, and therefore of degree 0 in fiber, and thus belongs to $F(T)$. There should therefore be a map $f\colon T \to J$ such that $p^*(L)$ is isomorphic to $f^*(\mathcal{L})$, where $\mathcal{L}$ is some universal bundle on $J$. However, since $p^*(L)$ is trivial in each fiber, $f$ must map all of $T$ to the same point (the one corresponding to the trivial bundle on $X$), and so $f^*(\mathcal{L})$ is trivial. But $p^*(L)$ need not be trivial.

This problem can be fixed by simply killing all the bundles that come from $T$. Precisely, define $G(T)$ to be the quotient of $F(T)$ by the subgroup $p^*(\mathrm{Pic}(T))$. Then the above paragraph shows that we should work with $G$ instead of $F$.

**Second obstruction to representability**

However, $F$ suffers from another problem which prevents it, and $G$, from being representable: it is not necessarily a sheaf. In fact, if $k'/k$ is a Galois extension with group $\Gamma$ then the natural map $F(k) \to F(k')^\Gamma$ need not be a bijection, which is a requirement for representability. (Note that $F$ and $G$ have the same field-value points, so this fails for $G$ as well.) Precisely, we have the following picture:

**Proposition.** Let $k'/k$ be a Galois extension of group $\Gamma$. Then there is a natural exact sequence

$$0 \to \mathrm{Pic}(X) \to \mathrm{Pic}(X_{k'})^\Gamma \to \mathrm{Br}(k)$$

In particular, given $L \in \mathrm{Pic}(X_{k'})$ there is an obstruction in $\mathrm{Br}(k)$ measuring the failure of $L$ to descend to $X$.

*Proof.* We first show that the left map is injective. In other words, if $L$ and $L'$ are two line bundles on $X$ which are isomorphic over $X_{k'}$, then $L$ and $L'$ are isomorphic. Let $i$ be an isomorphism over $X_{k'}$. For $\sigma \in \Gamma$, the map $i^\sigma$ is also an isomorphism $L \to L'$, and thus differs from $i$ by an element $c_\sigma$ of $\mathrm{Aut}(L) = (k')^\times$. One easily sees that $c$ satisfies the cocycle condition, and thus defines an element of $\mathrm{H}^1(\Gamma, (k')^\times)$, which vanishes by Hilbert's Theorem 90. Thus $c$ is a coboundary, i.e., of the form $c_\sigma = (\sigma\alpha)/\alpha$ for some $\alpha \in (k')^\times$. One easily sees that $\alpha^{-1}i$ is a Galois-invariant isomorphism $L \to L'$ over $X_{k'}$, and thus descends to $X$.

We now construct an element of $\mathrm{Br}(k)$ measuring the obstruction of an element of $\mathrm{Pic}(X_{k'})^\Gamma$ to come form $\mathrm{Pic}(X)$. The basic reason such an obstruction exists is because an element of $\mathrm{Pic}(X_{k'})^\Gamma$ is a line bundle $L$ on $X_{k'}$ such that $\sigma^*(L)$ is isomorphic to $L$ for each $\sigma \in \Gamma$, but these isomorphisms are not required to satisfy any sort of compatibilities, which is needed for descent. In fact, the failure of the compatibilities defines a 2-cocycle which gives the Brauer obstruction. Suppose $L \in \mathrm{Pic}(X_{k'})^{\Gamma}$, and for each $\sigma \in \Gamma$ choose an isomorphism $i_\sigma \colon L \to \sigma^*(L)$. Then $\sigma^*(i_\tau) \circ i_\sigma$ and $i_{\sigma\tau}$ are two isomorphisms $L \to (\sigma\tau)^* L$, and thus differ by an element $c_{\sigma,\tau}$ of $\mathrm{Aut}(L) = (k')^\times$. It is easy to see that $c$ satisfies the 2-cocycle condition, and thus defines an element of $\mathrm{H}^2(\mathrm{Gal}(k'/k), (k')^\times) \subset \mathrm{Br}(k)$. If this 2-cocycle is a coboundary, then the choice of $i$'s can be modified to give descent data on $L$, and $L$ belongs to $\mathrm{Pic}(X)$. This completes the proof. □

**Example.** Take $X$ to be a curve which is isomorphic to $\mathbb{P}^1$ over $k'$ but not over $k$, e.g., the curve over $k = \mathbb{R}$ given by $X^2 + Y^2 = -Z^2$. Then $\mathrm{Pic}(X_{k'})$ is isomorphic to $\mathbb{Z}$, and thus contains $\mathrm{Pic}(X)$ with finite index, and so $\Gamma$ acts trivially on $\mathrm{Pic}(X_{k'})$. But the bundle $\mathcal{O}(1)$ on $X_{k'}$ does not descend to $X$, as this would give an isomorphism $X \to \mathbb{P}^1$ over $k$.

**Remark.** Suppose $k$ is a finite extension of $\mathbb{Q}_p$ and $k'$ is an algebraic closure of $k$. Then $\mathrm{Br}(k) = \mathbb{Q}/\mathbb{Z}$, and Lichtenbaum showed that the image of the the map $\mathrm{Pic}(X_{k'})^\Gamma \to \mathrm{Br}(k)$ is $N^{-1}\mathbb{Z}/\mathbb{Z}$, where $N$ is the gcd of the degrees of divisors on $X$. Thus $\mathrm{Pic}(k) = \mathrm{Pic}(k')^\Gamma$ if and only if $X$ has a divisor of degree 1 defined over $k$.

**Remark.** We have not actually give an example where a line bundle of degree 0 fails to descend, which is the case of interest (as $F(k') = \mathrm{Pic}^0(X_{k'})$). I believe such an example exists if $X$ is a genus 1 curve over a finite extension of $\mathbb{Q}_p$ without a point.

### The case where a rational point exists

In fact, the failure of $G$ to satisfy descent only occurs when $X$ has not $k$-rational points. To see this, suppose $X$ has a $k$-rational point $x$. Define $\mathcal{G}_x(T)$ to be the category of pairs $(L, i)$ where $L$ is a fiberwise degree 0 line bundle on $X_T$, and $i$ is an isomorphism of $L|_{\{x\} \times T}$ with the trivial bundle $\mathcal{O}_T$. Define $G_x(T)$ to be the set of isomorphism classes in $\mathcal{G}_x(T)$. The key point is that objects of $\mathcal{G}_x(T)$ are rigid: they have no automorphisms. This means that if an isomorphism class is invariant, then it has canonical descent data. It follows that $G_x$ is a sheaf. On the other hand, we have the following lemma:

**Lemma.** The natural map $G_x \to G$, given by forgetting $i$, is an isomorphism.

*Proof.* If $L$ and $L'$ belong to $G_x$ and $L \cong L' \otimes p^*(L'')$ for some line bundle on $L''$ then, restricting to $\{x\} \times T$, one sees that $L''$ is trivial, and so $L \cong L'$; this proves injectivity. As for surjectivity, suppose $L$ is a line bundle on $X_T$ and let $L_0$ be its restriction to $\{x\} \times T$. Then $L \otimes p^*(L_0^{-1})$ is naturally an element of $G_x$ mapping to $L$ in $G(T)$. □

We thus see that, when $X$ has a $k$-point, $G$ is a sheaf.

**Theorem.** Suppose $X$ has a $k$-point. Then the sheaf $G$ is representable. The representing scheme is denoted $\mathrm{Jac}(X)$, and called the Jacobian of $X$.

If $X$ does not have a point then $G$ is not necessarily a sheaf, and thus not necessarily representable. However, one can replace $G$ with its sheafification, and this turns out to be representable. Thus one can define the Jacobian of $X$ even when $X$ does not have a point.

### Construction of the Jacobian

We now sketch the proof of the representability of $G$ in the case that $X$ has a $k$-rational point $x$. Let $X^{(r)}$ be the $r$th symmetric power of $X$, i.e., the quotient of $X^r$ by the action of the symmetric group $S_r$. Points on $X^{(r)}$ defined over $k'$ can be identified with effective divisors on $X_{k'}$ of degree $r$. We will consider $X^{(g)}$, where $g$ is the genus of $X$.

Let $D$ and $D'$ be effective divisors of degree $g$ on $X$. The Riemann–Roch theorem then implies that $\ell(D + D' - g[x]) \geq 1$. By semi-continuity, the locus $U \subset X^{(g)} \times X^{(g)}$ where equality holds is open, and it is not difficult to show that it is non-empty. (Taking $D' = g[x]$, one must find an effective divisor $D$ of degree $g$ with $\ell(D) = 1$, or, equivalently $\ell(K - D) = 0$. Simply pick $g$ points $x_1, \ldots, x_g$ of $X$ such that the restriction map $\mathrm{H}^0(X, \Omega^1) \to \prod_{i=1}^g T^*_{x_i}$ is an isomorphism.) Given $(D, D') \in U$, there is thus a non-zero meromorphic function $f$ on $X$, unique up to scaling, such that $D'' = \mathrm{div}(f) + D + D' - g[x]$ is effective. We define a map $U \to X^{(g)}$ by taking $(D, D')$ to $D''$. By working systematically with families of divisors, one shows that this is a map of schemes.

One can regard the above map as a rational map $X^{(g)} \times X^{(g)} \dashrightarrow X^{(g)}$. As such, it satisfies the axioms to be a group (it is a group object in the category of varieties with rational maps). Weil showed that any such rational group variety can be upgraded to an actual group variety. Precisely, there exists a group variety $J$ (unique up to isomorphism) and a unique isomorphism of rational group varieties $X^{(g)} \dashrightarrow J$

Finally, one must show that $J$ represents $G$. One first shows that $J$ is proper, and so the rational map $X^{(g)} \dashrightarrow J$ is an actual map. Then, one defines a map $f \colon \mathrm{Div}^0(X) \to J$ as follows. If $D$ is a degree 0 divisor such that $D + g[x]$ is effective, then one regards $D + g[x]$ as an element of $X^{(g)}$ and takes its image in $J$. If $D + g[x]$ is not effective, then one finds a degree 0 divisor $D'$ such that $D + D' + g[x]$ and $D' + g[x]$ are both effective, and defines $f(D) = f(D + D') - f(D')$. Working with families of divisors, $f$ gives a map of functors $G \to J$. One then verifies that it is a bijective on $T$-points.

### Basic properties

Many of the basic properties satisfied in the analytic case remain true in the algebraic case.

(i) One can show that $T_0(\mathrm{Jac}(X)) = \mathrm{H}^1(X, \mathcal{O})$ using the functor of points of $\mathrm{Jac}(X)$ and the interpretation of the tangent space in terms of dual numbers.

(ii) From this, one finds that $\mathrm{H}^0(\mathrm{Jac}(X), \Omega^1)$ is naturally isomorphic to $\mathrm{H}^0(X, \Omega^1)$.

(iii) One again has a map $f_x \colon X \to \mathrm{Jac}(X)$ given a base point $x \in X(k)$. On field points, this takes a point $y \in X(k)$ to the degree 0 divisor $[y] - [x]$. On $T$-points, it does the same thing, but one must use a relative notion of divisor.

(iv) By definition, $\mathrm{Jac}(X)(k)$ is isomorphic to $\mathrm{Pic}^0(X)$.

One again has a comparison between the first (co)homology groups of $X$ and $\mathrm{Jac}(X)$, though this now involves cohomology. This is most easily seen using Kummer theory. Suppose $n$ is prime to the characteristic of $k$, so that we have an exact sequence of sheaves on the étale site of $X$:

$$0 \to \mu_n \to \mathbb{G}_m \to \mathbb{G}_m \to 0$$

Taking cohomology over $\overline{k}$, and using the fact that every element of $\overline{k}^\times$ is an $n$th power, we see that $\mathrm{H}^1(X_{\overline{k}}, \mathbb{G}_m)[n] = \mathrm{H}^1(X_{\overline{k}}, \mu_n)$. Now, $\mathrm{H}^1(X_{\overline{k}}, \mathbb{G}_m) = \mathrm{Pic}(X_{\overline{k}})$; since all torsion in this group is of degree 0, we see that $\mathrm{H}^1(X_{\overline{k}}, \mathbb{G}_m)[n] = \mathrm{Jac}(X)[n](\overline{k})$. Replacing $n$ with $\ell^n$ and taking an inverse limit, we find $T_\ell(\mathrm{Jac}(X)) = \mathrm{H}^1(X_{\overline{k}}, \mathbb{Z}_\ell(1))$, where the (1) is a Tate twist.

### In relative situations

Suppose $C \to S$ is a family of smooth projective curves with geometrically connected fibers. One can then define a functor $G$ just as we did above. When $C$ has a section over $S$, this functor is representable by an abelian scheme $\mathrm{Jac}(C)$, which one would call the relative Jacobian.

One reason this is relevant for us is as follows. Suppose $R$ is a DVR with fraction field $K$, let $X/K$ be a curve, and let $J$ be its Jacobian. Suppose we can find a nice model of $X$ over $R$ (smooth, projective, geometrically connected fibers). Then the relative Jacboian of this model is an abelian scheme extending $J$. This shows that $J$ has good reduction.

## 1.5 Criterion for rank 0 (Theorem B)

In this section, we establish Theorem B from the Overview section, which is a criterion for an abelian variety to have rank 0. The idea of the proof is similar to that of the weak Mordell–Weil theorem, but here we control the ramification of the cohomology classes much more carefully. Most of the work goes into understanding a certain class of group schemes (the admissible ones) very precisely.

The purpose of this section is to establish the following criterion for an abelian variety to have rank 0,

**Theorem.** Let $A/\mathbb{Q}$ be an abelian variety, and let $N$ and $p$ be distinct prime numbers, with $N$ odd. Suppose the following conditions hold:

(i) $A$ has good reduction away from $N$.

(ii) $A$ has completely toric reduction at $N$.

(iii) The Jordan–Hölder constituents of $A[p](\overline{\mathbb{Q}})$ are 1-dimensional and either trivial or cyclotomic.

Then $A(\mathbb{Q})$ has rank 0.

**Remark.** The proof uses many special properties of $\mathbb{Q}$, but can be generalized slightly, as follows. Let $K$ be an imaginary quadratic number field, let $p$ be a rational prime, and let $\mathfrak{N}$ be a prime of $K$. Assume that $p$ does not divide the class number of $K$ and if $p \leq 3$ then $p$ is unramified in $K$. Then the obvious generalization holds: if $A/K$ be an abelian variety with good reduction away from $\mathfrak{N}$, completely toric reduction at $\mathfrak{N}$, and such that the Jordan–Hölder constituents of $A[p](\overline{K})$ are either trivial or cyclotomic then $A(K)$ has rank 0.

**Remark.** This theorem, and the proof presented here, comes from III.3 of Mazur's paper "Modular curves and the Eisenstein ideal" (MR488287). It is not stated there explicitly, however.

*Proof Idea.* Recall the proof of the weak Mordell–Weil theorem. Kummer theory gives an injection of $A(\mathbb{Q})/nA(\mathbb{Q})$ into $\mathrm{H}^1(G_{\mathbb{Q}}, A[n])$, so it suffices to prove the $\mathrm{H}^1$ is finite. However, it's not, because we have not restricted ramificiation. One can show that there is a finite set of places $S$ such that the image of $A(\mathbb{Q})/nA(\mathbb{Q})$ lands in $\mathrm{H}^1(G_{\mathbb{Q},S}, A[n])$. This $\mathrm{H}^1$ is finite, and this proves the weak Mordell–Weil theorem.

As one shrinks $S$, the $\mathrm{H}^1$ gets smaller and smaller, so it makes sense for us to take $S$ as small as possible. In general, one can take $S$ to be the set of places of bad reduction together with the divisors of $n$. So if work with $p$-power torsion, we can take $S = \{N, p\}$. However, this is still too big for us!

We can improve the situation using the following idea. Let $\mathcal{A}$ be the Néron model of $A$ over $\mathbb{Z}$, and let $G_n = \mathcal{A}[p^n]$. Then $\mathrm{H}^1(G_{\mathbb{Q},S}, A[p^n])$ is the étale cohomology group $\mathrm{H}^1_{\mathrm{et}}(\mathrm{Spec}(\mathbb{Z}[1/Np]), G_n)$ – restricting ramification to $S$ corresponds to taking étale cohomology over the ring of integers with the primes in $S$ removed. It is not true that $A(\mathbb{Q})/p^nA(\mathbb{Q})$ injects into $\mathrm{H}^1_{\mathrm{et}}(\mathrm{Spec}(\mathbb{Z}), G_n)$: this group is often zero. However, $A(\mathbb{Q})/p^nA(\mathbb{Q})$ does inject into $\mathrm{H}^1_{\mathrm{fppf}}(\mathrm{Spec}(\mathbb{Z}), G_n)$, and this is the group we will use. (This is a slight lie that we will correct below.)

The idea is to show that this $\mathrm{H}^1$ is bounded independent of $n$, which will establish that $A(\mathbb{Q})$ has rank 0. To do this, we need to understand the flat cohomology of $G_n$ very well, so we begin by studying groups like $G_n$ and their flat cohomology. □

### 1.5.1 Admissible groups

**Definition**

(i) A group scheme $G$ over $\mathbb{Z}[1/N]$ is pre-admissible if it is finite, flat, commutative, and killed by a power of $p$.

(ii) A group scheme $G$ over $\mathbb{Z}$ is pre-admissible if it is commutative, separated, of finite presentation, quasi-finite, flat, killed by a power of $p$, and its restriction to $\mathbb{Z}[1/N]$ is finite (and thus pre-admissible).

**Example.** Let $A$ be an abelian variety with good reduction away from $N$ and let $\mathcal{A}$ be its Néron model over $\mathbb{Z}$. Then $\mathcal{A}[p^n]$ is a pre-admissible group scheme over $\mathbb{Z}$.

Let $G$ be an pre-admissible group over $\mathbb{Z}[1/N]$. An admissible filtration on $G$ is a filtration

$$0 = F^0G \subset F^1G \subset \cdots \subset F^nG = G$$

by closed subgroups such that $F^{n+1}G/F^nG$ is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ or $\mu_p$ for each $n$. We say that $G$ is admissible if it has an admissible filtration. We say that a pre-admissible group over $\mathbb{Z}$ is admissible if its restriction to $\mathbb{Z}[1/N]$ is.

We make a similar definition for Galois representations. Precisely, let $V$ be a $\Gamma_{\mathbb{Q}}$-module. We say that $V$ is admissible if it possess a filtration $F^\bullet V$ by $\Gamma_{\mathbb{Q}}$-submodules such that $F^{n+1}V/F^nV$ is a one-dimensional $\mathbb{F}_p$ vector space on which $\Gamma_{\mathbb{Q}}$ acts either trivially or through the cyclotomic character.

**Detecting admissible filtrations**

**Proposition.** Let $G$ be a pre-admissible group over $\mathbb{Z}[1/N]$. Then $G$ is admissible if and only if $G(\overline{\mathbb{Q}})$ is admissible.

*Proof.* Let $V \subset G(\overline{\mathbb{Q}})$ be the first piece of an admissible filtration, let $H_0 \subset G_{\mathbb{Q}}$ be the subgroup it corresponds to, and let $H$ be the closure of $H_0$ in $G$. Over $\mathbb{Z}[1/Np]$, the group $H$ is finite étale, and therefore isomorphic to either $\mu_p$ or $\mathbb{Z}/p\mathbb{Z}$, depending on the Galois action on $H(\overline{\mathbb{Q}})$. Over $\mathbb{Z}_p$, the group $H$ is a finite flat commutative group which is generically isomorphic to $\mu_p$ or $\mathbb{Z}/p\mathbb{Z}$. If $p \neq 2$, Raynaud's theorem implies that the isomorphism extends over $\mathbb{Z}_p$; the same is true for $p = 2$ by a theorem of Fontaine. (Note: it is important here that we're over $\mathbb{Q}_p$, so that there is no ramification.) Thus $H$ is isomorphic, over $\mathbb{Z}[1/N]$, to $\mu_p$ or $\mathbb{Z}/p\mathbb{Z}$. Applying the same reasoning to $G/H$, the result follows by induction □

**Invariants**

Let $G$ be an admissible group over $\mathbb{Z}$. Following Mazur, we define several invariants:

(i) Let $\ell(G) = \log_p(\#G_{\mathbb{Q}})$. This coincides with the length of an admissible filtration on $G$.

(ii) Let $\delta(G) = \log_p(\#G_{\mathbb{Q}}) - \log_p(\#G_{\mathbb{F}_N})$.

(iii) Define $\alpha(G)$ to be the number of $\mathbb{Z}/p\mathbb{Z}$'s appearing in an admissible filtration of $G$ (over $\mathbb{Z}[1/N]$).

(iv) Let $h^i(G)$ be $\log_p(\#\mathrm{H}^i_{\mathrm{fppf}}(\mathrm{Spec}(\mathbb{Z}), G))$, for $i = 0, 1$.

Note that everything we're applying $\log_p$ to is a $p$th power.

**Remark.** Let $G$ be a group scheme over a base scheme $S$. The group $\mathrm{H}^1_{\mathrm{fppf}}(S, G)$ admits a fairly concrete description, as follows. A torsor for $G$ is a scheme $T/S$ equipped with an action of $G$ that is simply transitive, in the following sense: for any scheme $S'/S$ and any section $x \in T(S')$, the map $G(S') \to T(S')$ given by $g \mapsto gx$ is a bijection. An fppf (or étale) torsor is a torsor $T/S$ for which there exists an fppf (or étale) cover $S' \to S$ such that $T(S')$ is non-empty. Then $\mathrm{H}^1_{\mathrm{fppf}}(S, G)$ is naturally in bijection with the set of isomorphism classes of fppf torsors; similarly, $\mathrm{H}^1_{\mathrm{et}}(S, G)$ is in bijection with the set of isomorphism classes of étale torsors. Of course, $\mathrm{H}^0_{\mathrm{fppf}}(S, G) = \mathrm{H}^0_{\mathrm{et}}(S, G) = G(S)$ is even easier to describe.

### Elementary admissible groups

We say that an admissible group $G$ is elementary if $\ell(G) = 1$. Over $\mathbb{Z}[1/N]$, there are two elementary admissible groups: $\mathbb{Z}/p\mathbb{Z}$ and $\mu_p$. Recall the following result:

**Proposition.** Let $H$ be a pre-admissible group over $\mathbb{Q}_N$. Then extensions of $H$ to a pre-admissible group over $\mathbb{Z}_N$ correspond to unramified Galois submodules of $H(\overline{\mathbb{Q}}_N)$. In particular, if $H(\overline{\mathbb{Q}}_N)$ is unramified and one-dimensional, it admits exactly two such extensions: a finite one (corresponding to the full module) and the extension by zero $H^\flat$ (correspond to the zero submodule).

This applies to both $\mathbb{Z}/p\mathbb{Z}$ and $\mu_p$ over $\mathbb{Z}[1/N]$ (even though we only stated the above result locally). We thus see that there are four elementary admissible groups over $\mathbb{Z}$, namely: $\mathbb{Z}/p\mathbb{Z}$, $(\mathbb{Z}/p\mathbb{Z})^\flat$, $\mu_p$, and $\mu_p^\flat$.

**Proposition.** The invariants of the elementary admissible groups are given as follows:

| | $\mathbb{Z}/p\mathbb{Z}$ | $(\mathbb{Z}/p\mathbb{Z})^\flat$ | $\mu_p$ | $\mu_p^\flat$ |
|---|---|---|---|---|
| $\delta$ | 0 | 1 | 0 | 1 |
| $\alpha$ | 1 | 1 | 0 | 0 |
| $h^0$ | 1 | 0 | 0 if $p$ odd<br>1 if $p = 2$ | 0 |
| $h^1$ | 0 | 0 | 0 if $p$ odd<br>1 if $p = 2$ | $\epsilon$ |

Here $\epsilon$ is 0 if $p$ is odd and $N \not\equiv 1 \bmod p$, or if $p$ is even and $N = 3 \pmod 4$, and 1 otherwise.

*Proof.* The first three lines are obvious. We explain the fourth. For this proof, let $S = \mathrm{Spec}(Z)$. Since $\mathbb{Z}/p\mathbb{Z}$ is étale, $\mathrm{H}^1_{\mathrm{fppf}}(S, \mathbb{Z}/p\mathbb{Z}) = \mathrm{H}^1_{\mathrm{et}}(S, \mathbb{Z}/p\mathbb{Z})$: if $T$ is an fppf torsor for $\mathbb{Z}/p\mathbb{Z}$ over $S$ then there is an fppf cover $S' \to S$ such that $T_{S'}$ is isomorphic to $(\mathbb{Z}/p\mathbb{Z})_{S'}$, and thus étale; this implies that $T$ is étale since this is an fppf local property. Since $\mathbb{Z}/p\mathbb{Z}$ is a constant scheme, we have $\mathrm{H}^1_{\mathrm{et}}(S, \mathbb{Z}/p\mathbb{Z}) = \mathrm{Hom}(\pi_1^{\mathrm{et}}(S), \mathbb{Z}/p\mathbb{Z})$. However, $\pi_1^{\mathrm{et}}(S)$ is trivial: it is the Galois group of the maximal everywhere unramified extension of $\mathbb{Q}$, which is just $\mathbb{Q}$. Thus $h^1(\mathbb{Z}/p\mathbb{Z}) = 0$.

Let $G$ be the quotient of $\mathbb{Z}/p\mathbb{Z}$ by $(\mathbb{Z}/p\mathbb{Z})^\flat$; this is the push-forward of $\mathbb{Z}/p\mathbb{Z}$ along the inclusion $\mathrm{Spec}(\mathbb{F}_N) \to \mathrm{Spec}(\mathbb{Z})$. By what we already have shown, there is a short exact sequence

$$0 \to \mathrm{H}^0_{\mathrm{fppf}}(S, \mathbb{Z}/p\mathbb{Z}) \to \mathrm{H}^0_{\mathrm{fppf}}(S, G) \to \mathrm{H}^1_{\mathrm{fppf}}(S, (\mathbb{Z}/p\mathbb{Z})^\flat) \to 0$$

Both the $\mathrm{H}^0$'s are $\mathbb{Z}/p\mathbb{Z}$, which shows $h^1((\mathbb{Z}/p\mathbb{Z})^\flat) = 0$. We have the following short exact sequence on the fppf site of $S$:

$$0 \to \mu_p \to \mathbb{G}_m \xrightarrow{p} \mathbb{G}_m \to 0$$

Taking cohomology, we obtain an exact sequence

$$0 \to \mathbb{Z}^\times/(\mathbb{Z}^\times)^p \to \mathrm{H}^1_{\mathrm{fppf}}(S, \mu_p) \to \mathrm{H}^1_{\mathrm{fppf}}(S, \mathbb{G}_m)[p] \to 0$$

The group on the left is 0 if $p$ is odd, and $\mathbb{Z}/p\mathbb{Z}$ if $p = 2$. By the theory of fppf descent, $\mathrm{H}^1_{\mathrm{fppf}}(S, \mathbb{G}_m) = \mathrm{H}^1_{\mathrm{Zar}}(S, \mathbb{G}_m) = \mathrm{Pic}(S)$, which is just the ideal class group of $\mathbb{Q}$, which is trivial. Thus $h^1(\mu_p)$ is 1 if $p = 2$ and 0 otherwise.

Let $G$ be the quotient of $\mu_p$ by $\mu_p^\flat$: this is the push-forward of $\mu_p$ along the inclusion $\mathrm{Spec}(\mathbb{F}_N) \to \mathrm{Spec}(\mathbb{Z})$. We have an exact sequence

$$0 \to \mathrm{H}^0_{\mathrm{fppf}}(S, \mu_p) \to \mathrm{H}^0_{\mathrm{fppf}}(S, G) \to \mathrm{H}^1_{\mathrm{fppf}}(S, \mu_p^\flat) \to \mathrm{H}^1_{\mathrm{fppf}}(S, \mu_p) \to \mathrm{H}^1_{\mathrm{fppf}}(S, G)$$

If $p \neq 2$, then $\mathrm{H}^i_{\mathrm{fppf}}(S, \mu_p) = 0$ for $i = 0, 1$, and so the map $\mathrm{H}^0_{\mathrm{fppf}}(S, G) \to \mathrm{H}^1_{\mathrm{fppf}}(S, \mu_p^\flat)$ is an isomorphism. The source is just $\mu_p(\mathbb{F}_N)$, which has order $p$ if $p \mid N - 1$, and vanishes otherwise. Now suppose that $p$ is even. Then the map $\mathrm{H}^0_{\mathrm{fppf}}(S, \mu_p) \to \mathrm{H}^0_{\mathrm{fppf}}(S, G)$ is an isomorphism. Kummer theory shows that the unique non-trivial element of $\mathrm{H}^1_{\mathrm{fppf}}(S, \mu_2)$ is represented by the $\mu_2$ torsor $\mathrm{Spec}(\mathbb{Z}[\sqrt{-1}])$. This torsor splits over $\mathbb{F}_N$ if and only if $-1$ is a square mod $N$. Thus the kernel of $\mathrm{H}^1_{\mathrm{fppf}}(S, \mu_p) \to \mathrm{H}^1_{\mathrm{fppf}}(S, G)$ is $\mathbb{Z}/2\mathbb{Z}$ if $N$ is 1 mod 4, and 0 if $N$ is 3 mod 4. This completes the proof. □

**Proposition.** Let $G$ be an admissible group over $\mathbb{Z}$. Then $h^1(G) - h^0(G) \le \delta(G) - \alpha(G)$.

*Proof.* Let

$$0 \to G_1 \to G_2 \to G_3 \to 0$$

be a short exact sequence of admissible groups. From the first few terms of the long exact sequence in cohomology, we find

$$h^1(G_2) - h^0(G_2) \le (h^1(G_1) - h^0(G_1)) + (h^1(G_3) - h^0(G_3))$$

that is, $h^1 - h^0$ is sub-additive. It is clear that

$$\delta(G_2) - \alpha(G_2) = (\delta(G_1) - \alpha(G_1)) + (\delta(G_3) - \alpha(G_3))$$

i.e., $\delta - \alpha$ is additive (in fact, both $\delta$ and $\alpha$ are additive separately). Thus if the result is true for $G_1$ and $G_3$ then it is true for $G_2$. It thus suffices to prove the result for elementary admissible groups, which follows easily from the computation of the invariants. □

### Proof of criterion

We now prove the main theorem. Let $\mathcal{A}$ be the Néron model of $A$ over $\mathbb{Z}$ and let $\mathcal{A}^\circ$ be its identity component (i.e., throw out the non-identity components in each fiber). Let $G_n = \mathcal{A}^\circ[p^n]$. Since $A$ has good reduction away from $N$, it is clear that $G_n$ is pre-admissible. The condition on $A[p](\overline{\mathbb{Q}})$ exactly says that it is admissible, and so $A[p^n](\overline{\mathbb{Q}})$ is as well, since it is an iterated self-extension of $A[p](\overline{\mathbb{Q}})$. It follows that $G_n$ is admissible.

We now compute the invariants $\alpha$ and $\delta$ for $G_n$. We begin with $\delta$. We have $\ell(G_n) = 2gn$, where $g = \dim(A)$. Now, $(G_n)_{\mathbb{F}_N} = \mathcal{A}^\circ_{\mathbb{F}_N}[p^n]$. By hypothesis, $\mathcal{A}^\circ_{\mathbb{F}_N}$ is a torus of dimension $g$, and so its $p^n$ torsion has cardinality $p^{ng}$. Thus $(G_n)_{\mathbb{F}_N}$ has cardinality $p^{gn}$. We thus find $\delta(G_n) = gn$.

We now compute $\alpha$. Since $\alpha$ is additive and only depends on the group over $\mathbb{Z}[1/N]$, we see that $\alpha(G_n) = n\alpha(G_1)$, so it suffices to treat the $n = 1$ case. Note that $\alpha(G_1)$ is the number of $\mathbb{Z}/p\mathbb{Z}$'s appearing in $(G_1)_{\mathbb{F}_p}$; thus it is $\log_p$ of the order of the étale part of $\mathcal{A}_{\mathbb{F}_p}[p]$. Since $G_1$ is admissible, $\mathcal{A}_{\mathbb{F}_p}[p]$ has only $\mathbb{Z}/p\mathbb{Z}$'s and $\mu_p$'s in it, and so $\mathcal{A}_{\mathbb{F}_p}$ is ordinary. This implies that its étale part has order $p^g$, and so $\alpha(G_1) = g$. Thus $\alpha(G_n) = gn$.

We thus have $\delta(G_n) = gn$ and $\alpha(G_n) = gn$. It follows that $h^1(G_n) - h^0(G_n) \leq 0$. However, $h^0(G_n)$ is the $p^n$ torsion in $\mathcal{A}^\circ(\mathbb{Z}) \subset \mathcal{A}(\mathbb{Z}) = A(\mathbb{Q})$, which is bounded independent of $n$ by the Mordell–Weil theorem. It follows that $h^1(G_n)$ is bounded independent of $n$.

Consider now the short exact sequence of sheaves on the fppf site of Spec($\mathbb{Z}$):

$$0 \to G_n \to \mathcal{A}^\circ \xrightarrow{p^n} \mathcal{A}^\circ \to 0$$

A few remarks:

(i) The map $[p^n]\colon \mathcal{A} \to \mathcal{A}$ is not a surjection of fppf sheaves in general since the component group of $\mathcal{A}_{\mathbb{F}_N}$ might have $p$-torsion. This is why we use $\mathcal{A}^\circ$ instead of $\mathcal{A}$.

(ii) The map $[p^n]\colon \mathcal{A}^\circ \to \mathcal{A}^\circ$ is not a surjection of étale sheaves in general, since give a section $x \in \mathcal{A}^\circ(S)$, one cannot in general find an étale extension $S'/S$ a section $y \in \mathcal{A}^\circ(S')$ such that $p^n y = x$. This is why we must use fppf cohomology.

(iii) The map $[p^n]\colon \mathcal{A}^\circ \to \mathcal{A}^\circ$ is faithfully flat: the key point is that $\mathcal{A}^\circ$ is $p$-divisible, and so $[p^n]$ is surjective on points. This implies that $[p^n]\colon \mathcal{A}^\circ \to \mathcal{A}^\circ$ is a surjection of fppf sheaves.

Taking cohomology, we obtain an injection

$$\mathrm{H}^0_{\mathrm{fppf}}(\mathrm{Spec}(\mathbb{Z}), \mathcal{A}^\circ) \otimes \mathbb{Z}/p^n\mathbb{Z} \to \mathrm{H}^1_{\mathrm{fppf}}(\mathrm{Spec}(\mathbb{Z}), G_n)$$

It follows that the cardinality of $\mathcal{A}^\circ(\mathbb{Z}) \otimes \mathbb{Z}/p^n\mathbb{Z}$ is bounded as $n \to \infty$.

Let $C$ be the $\mathbb{F}_N$-points of the component group of $\mathcal{A}_{\mathbb{F}_N}$. Then there is an exact sequence

$$0 \to \mathcal{A}^\circ(\mathbb{Z}) \to \mathcal{A}(\mathbb{Z}) \to C$$

It follows that $\mathcal{A}^\circ(\mathbb{Z})$ is a finite index subgroup of $\mathcal{A}(\mathbb{Z}) = A(\mathbb{Q})$. In particular, $\mathcal{A}^\circ(\mathbb{Z})$ is finitely generated. Since the cardinality of $\mathcal{A}^\circ(\mathbb{Z}) \otimes \mathbb{Z}/p^n\mathbb{Z}$ is bounded, it follows that $\mathcal{A}^\circ(\mathbb{Z})$ has rank 0. Thus $A(\mathbb{Q})$ has rank zero, as it contains $\mathcal{A}^\circ(\mathbb{Z})$ with finite index.

# 2 Moduli of elliptic curves

## 2.1 Modular curves

In this section, we begin the study of modular curves. I introduce the basic modular curves $Y(N)$, $Y_0(N)$, and $Y_1(N)$ and describe them as quotients of the upper half-plane. I then discuss more general upper half-plane quotients, their complex structures, and their compactifications. Finally, I prove the genus formula and give some examples of it.

### 2.1.1 Sets of elliptic curves

**Cast of characters**

Let $Y(1)$ be the set of isomorphism classes of elliptic curves over the complex numbers. It turns out to be useful to consider some generalizations of this set as well; we mention here the three most common. Let $N$ be a positive integer. Define $Y_1(N)$ to be the set of isomorphism classes of pairs $(E, P)$ where $E$ is an elliptic curve and $P$ is a point of exact order $N$; an isomorphism $(E, P) \to (E', P')$ is an isomorphism $f\colon E \to E'$ such that $f(P) = P'$. Define $Y_0(N)$ to be the set of isomorphism classes of pairs $(E, G)$ where $E$ is an elliptic curve and $G \subset E$ is a cyclic subgroup of order $N$. And define $Y(N)$ to be the set of isomorphism classes of pairs $(E, i)$ where $E$ is an elliptic curve and $i$ is an isomorphism $(\mathbb{Z}/N\mathbb{Z})^2 \to E[N]$.

**Description of $Y(1)$**

The set $Y(1)$ is not hard to describe directly: the $j$-invariant gives a bijection $Y(1) \to \mathbb{C}$. However, this description does not easily generalize to the related sets, and so we seek an alternate way to get at $Y(1)$.

Every elliptic curve is of the form $\mathbb{C}/\Lambda$ for some lattice $\Lambda$. Scaling $\Lambda$ by an element of $\mathbb{C}^\times$ leads to an isomorphic curve, so we may as well assume that 1 is a generator of $\Lambda$. We can choose another generator $\tau$ that has positive imaginary part, i.e., $\tau$ belongs to $\mathfrak{h}$, the upper half-plane.

For $\tau \in \mathfrak{h}$, let $\Lambda_\tau$ be the lattice spanned by 1 and $\tau$, and let $E_\tau = \mathbb{C}/\Lambda_\tau$. We have thus shown that the map $\mathfrak{h} \to Y(1)$ given by $\tau \mapsto E_\tau$ is surjective. However, it is not injective. Then $az + b$ and $cz + d$ also form a basis for $\Lambda_\tau$. Define

$$\gamma(z) = \frac{az + b}{cz + d}.$$

Then $\Lambda_\tau$ is just $\Lambda_{\gamma(\tau)}$ scaled by $cz + d$, and so $E_\tau$ and $E_{\gamma(\tau)}$ are isomorphic. It is not difficult to see that $(\gamma, z) \mapsto \gamma z$ defines an action of $\Gamma$ on $\mathfrak{h}$. We have just shown that our map $\mathfrak{h} \to S$ descends to a map $\mathfrak{h}/\Gamma(1) \to Y(1)$. In fact:

**Theorem.** The map $\mathfrak{h}/\Gamma(1) \to Y(1)$ is a bijection.

*Proof.* Suppose that $\Lambda_\tau$ and $\Lambda_{\tau'}$ define isomorphic elliptic curves, so $\Lambda_\tau = \alpha\Lambda_{\tau'}$ for some scalar $\alpha$. Then $1 = \alpha(c\tau' + d)$ for some $c, d \in \mathbb{Z}$. In fact, $c$ and $d$ are coprime: if $e > 1$ divides both then we would have $1/e = \alpha((c/e)\tau + (d/e)) \in \alpha\Lambda_{\tau'} = \Lambda_\tau$, which is not the case. We can therefore find integers $a$ and $b$ such that $ad - bc = 1$. Let $\gamma$ be the corresponding element of $\Gamma(1)$. Then $\alpha\Lambda_{\tau'}$ is exactly $\Lambda_{\gamma(tau')}$, and so $\Lambda_\tau = \Lambda_{\gamma(\tau')}$. But this implies $\tau = \tau' + n$ for some integer $n$, and so $\tau = \gamma'(\gamma(\tau'))$ □

**Description of $Y_1(N)$**

This description does generalize well. For example, we can define a map $\mathfrak{h} \to Y_1(N)$ by $\tau \mapsto (E_\tau, 1/N)$, where by $1/N$ we really mean its image in $E_\tau = \mathbb{C}/\Lambda_\tau$. Then it is not difficult to see that this map descends to a bijection $\mathfrak{h}/\Gamma_1(N) \to Y_1(N)$, where $\Gamma_1(N)$ is the subgroup of $\Gamma(1)$ consisting matrices $\gamma$ Similarly, we have a bijection $\mathfrak{h}/\Gamma_0(N) \to Y_0(N)$ and $\mathfrak{h}/\Gamma(N) \to Y(N)$, where $\gamma$ belongs to $\Gamma_0(N)$ and $\gamma$ belongs to $\Gamma(N)$ if $\gamma = 1 \pmod{N}$.

### 2.1.2 Modular curves

Motivated by the above examples, we consider an arbitrary finite-index subgroup $\Gamma$ of $\Gamma(1)$ and define $Y_\Gamma$ to be the quotient $\mathfrak{h}/\Gamma$. We now study these spaces.

**Complex structure**

Write $\pi$ for the quotient map $\mathfrak{h} \to Y_\Gamma$. We give $Y_\Gamma$ a complex structure by declaring a function $f\colon U \to \mathbb{C}$ to be holomorphic if $\pi^*(f)\colon \pi^{-1}(U) \to \mathbb{C}$ is, where $U$ is an open subset of $Y_\Gamma$. One can show that this gives $Y_\Gamma$ the structure of a Riemann surface.

**Compactification**

The space $Y_\Gamma$ is never compact: for example, $Y(1)$ is missing the point at infinity. There is an elegant way to fill in the missing points, as follows.

Think of $\mathfrak{h}$ as the open half of the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ above the equator $\mathbb{P}^1(\mathbb{R})$. Define $\mathfrak{h}^*$ to be the union of $\mathfrak{h}$ and $\mathbb{P}^1(\mathbb{Q})$, the set of cusps. Then $\Gamma(1)$ naturally acts on $\mathfrak{h}^*$. We give $\mathfrak{h}^*$ a topology by declaring the circles tangent to $\mathbb{P}^1$ at cusps to be open neighborhoods of cusps. More precisely, a neighborhood basis of $\infty$ consists of the subsets of $\mathfrak{h}^*$ of the form $U_K = \{\infty\} \cup \{z \in \mathfrak{h} \mid \operatorname{Im}(z) > K\}$, as $K$ varies. A neighborhood basis of $\gamma(\infty)$ is given by $\gamma(U_K)$. As $\Gamma(1)$ acts transitively on the set of cusps, this defines the topology everywhere. We give $\mathfrak{h}^*$ a complex structure by defining a function to be holomorphic at $\infty$ if it is holomorphic and bounded on $U_K$, for $K$ sufficiently large, and then moving this condition to other cusps as before.

We now define $X_\Gamma$ to be $\mathfrak{h}^*/\Gamma$. This is again a reasonable topological space, and can be given a complex structure in a manner similar to $Y_\Gamma$. It is a compact Riemann surface. We call a point of $X_\Gamma$ a cusp if it is the image of a cusp under the quotient map $\pi$. We note that the set of cusps is $\mathbb{P}^1(\mathbb{Q})/\Gamma$, which is finite since $\Gamma(1)$ acts transitively on $\mathbb{P}^1(\mathbb{Q})$ and $\Gamma$ has finite index in $\Gamma(1)$.

**Stabilizers**

To understand $X_\Gamma$, it will be important to understand the stabilizers $\Gamma_z$ of points $z \in \mathfrak{h}^*$, as this is where $\pi$ ramifies. We being by considering the case $\Gamma = \Gamma(1)$. As $-1$ stabilizes everything, we let $\overline{\Gamma(1)}$ be the quotient $\Gamma(1)/\{-1\}$, which is sometimes easier to deal with.

**Proposition.** Suppose $z \in \mathfrak{h}$. There is a natural identification of $\Gamma(1)_z$ with $\operatorname{Aut}(E_z)$.

*Proof.* Suppose $\gamma \in \Gamma(1)$ stabilizes $z$. Via the basis $1, z$ of $\Lambda_z$, we can regard $\gamma$ as an automorphism $g$ of $\Lambda_z$. The condition $\gamma(z) = z$ exactly says $g(z) = (cz + d)z$, and so $g$ is $\mathbb{C}$-linear. Thus $g$ induces an automorphism of $E_z$. The converse reasoning is similar. □

We thus see that $\Gamma(1)_z = \operatorname{End}(E_z)^\times$. We know that $\operatorname{End}(E_z)$ is either $\mathbb{Z}$ or an order in a quadratic imaginary field; we thus get non-trivial units only if $\operatorname{End}(E_z)$ is $\mathbb{Z}[i]$ or $\mathbb{Z}[\rho]$, with $\rho = e^{2\pi i/3}$.

**Proposition.** Suppose $E$ is an elliptic curve with $\operatorname{End}(E) = \mathbb{Z}[i]$. Then $E$ is isomorphic to $E_i$.

*Proof.* Write $E = \mathbb{C}/\Lambda$ where $\Lambda$ is a lattice stable under $\mathbb{Z}[i]$. Since $\Lambda$ is a torsion-free $\mathbb{Z}[i]$-module of $\mathbb{Z}$-rank 2, it is a projective $\mathbb{Z}[i]$-module of rank 1, and therefore free since the class group of $\mathbb{Q}(i)$ is trivial. We thus see that $\Lambda = \mathbb{Z}[i]\alpha$ for some $\alpha \in \mathbb{C}^\times$, and so $E$ is isomorphic to $E_i$. □

**Proposition.** Suppose $E$ is an elliptic curve with $\mathrm{End}(E) = \mathbb{Z}[\rho]$. Then $E$ is isomorphic to $E_\rho$.

*Proof.* Exactly the same. □

**Proposition.** Let $z \in \mathfrak{h}$. We have one of three cases:

(i) $z$ belongs to $\Gamma(1)i$. In this case, $\overline{\Gamma(1)}_z \cong \mathbb{Z}/2\mathbb{Z}$.

(ii) $z$ belongs to $\Gamma(1)\rho$. In this case $\overline{\Gamma(1)}_z \cong \mathbb{Z}/3\mathbb{Z}$.

(iii) $z$ does not belong to $\Gamma(1)i$ or $\overline{\Gamma(1)}\rho$. In this case, $\Gamma(1)_z \cong 1$.

*Proof.* This follows from what we have shown above. □

Determining the stabilizers of cusps is easy:

**Proposition.** Let $z$ be a cusp. Then $\overline{\Gamma(1)}_z/\{\pm 1\} \cong \mathbb{Z}$.

*Proof.* □

### Ramification

**Proposition.** Let $\Gamma \subset \Gamma'$ be finite index subgroups of $\Gamma(1)$. Let $z \in \mathfrak{h}^*$ and let $p$ be its image in $X_\Gamma$. Then the ramification index of $f\colon X_\Gamma \to X_{\Gamma'}$ at $p$ is $[\overline{\Gamma}'_z : \overline{\Gamma}_z]$.

### The genus formula

Let $\Gamma \subset \Gamma(1)$ have finite index $d$. Let $\nu_2$ (resp. $\nu_3$) be the number of $\Gamma$-orbits of elliptic points of order 2 (resp. 3), and let $\nu_\infty$ be the number of cusps. Let $g$ be the genus of $X_\Gamma$. Then

**Proposition.** Let $g$ be the genus of $X_\Gamma$. Then

$$g = 1 + \frac{d}{12} - \frac{\nu_2}{4} - \frac{\nu_3}{3} - \frac{\nu_\infty}{2}$$

*Proof.* We apply the Riemann–Hurwitz formula to the map $f\colon X_\Gamma \to X(1)$. Since $X(1)$ has genus 0, this states

$$(2-2g) = 2d - \sum (e_p - 1)$$

where the sum is over all points $p$ of $X_\Gamma$, and $e_p$ denotes the ramification index of $p$ along the map $f$. Let $q_2$, $q_3$, and $q_\infty$ be the images of $i$, $\rho$, and $\infty$ in $X(1)$. Then $e_p = 1$ unless $p$ lies above $q_2$, $q_3$ or $q_\infty$.

The points of $f^{-1}(q_2)$ which are elliptic are unramified, while those which are not have ramification index 2. Since the total number of points, with multiplicity, is $d$, we see that the number of ramified points is $(d-\nu_2)/2$. We thus find

$$\sum_{f(p)=q_2} (e_p - 1) = \frac{d-\nu_2}{2}.$$

Similarly, we find

$$\sum_{f(p)=q_3} (e_p - 1) = \frac{2(d-\nu_3)}{3}.$$

It is clear that

$$\sum_{f(p)=q_\infty} (e_p - 1) = d - \# f^{-1}(q_\infty) = d - \nu_\infty.$$

We thus have

$$2 - 2g = 2d - \frac{d-\nu_2}{2} - \frac{2(d-\nu_3)}{3} - (d - \nu_\infty),$$

from which the result follows. □

**Example.** Suppose $N$ is prime, and consider the case $\Gamma = \Gamma_0(N)$. One can show that $d = N+1$, $\nu_2$ is 1 if $N = 2$, 2 if $N = 1 \pmod 4$, and 0 otherwise, $\nu_3$ is 1 if $N = 3$, 2 if $N = 1 \pmod 3$, and 0 otherwise, and $\nu_\infty = 2$. This amounts to

$$g = \left\lfloor \frac{N}{12} \right\rfloor + \begin{cases} -1 & N = 1 \pmod{12} \\ 1 & N = 11 \pmod{12} \end{cases}$$

For example, if $N \leq 13$ then $g = 0$, except if $N = 11$ when $g = 1$.

**Example.** The genus of $X_1(N)$ is 0 if and only if $N \leq 12$ and $N \neq 11$. These are exactly the values of $N$ for which an elliptic curve over $\mathbb{Q}$ can have an $N$-torsion point. This is not a coincidence!

This following section introduces modular forms and Hecke operators. I started by introducing modular forms of level 1, and gave several interpretations of them, e.g., as sections of line bundles on the modular curve, or as functions of lattices. I then talked about modular forms of higher level. Finally, I introduced Hecke operators and their action on modular forms, and proved that they commute.

### 2.1.3 Modular forms of level 1

**Definition**

In the previous section, we established bijections

$$\mathfrak{h}/\Gamma(1) \to \{\text{lattices in } \mathbb{C}\}/\text{homothety} \to \{\text{isom. cl. of elliptic curves}/\mathbb{C}\} = Y(1)$$

The first map takes a point $z \in \mathfrak{h}$ to the lattice $\Lambda_z = \langle 1, z \rangle$, while the second map takes a lattice $\Lambda$ to the elliptic curve $\mathbb{C}/\Lambda$. Recall that $\Gamma(1)$ is just notation for $\mathrm{SL}_2(\mathbb{Z})$. Furthermore, this space is identified with $\mathbb{A}^1$ via the $j$-invariant.

A modular function is a meromorphic function on this space which is meromorphic at infinity. The $j$-invariant is an example, as is any rational function of the $j$-invariant. Since $Y(1) = \mathbb{A}^1$, every modular function is a rational function of the $j$-invariant.

A modular form is a function on the space of lattices which is homogeneous under homothety, but not necessarily of degree 0. Precisely, a modular form of weight $k$ is a function $f$ on the set of lattices in $\mathbb{C}$ satisfying the following conditions:

(i) Homogeneity: $f(\alpha\Lambda) = \alpha^{-k} f(\Lambda)$ for $\alpha \in \mathbb{C}^\times$.

(ii) Holomorphicity: $f$ is a holomorphic function of $\Lambda$, in the sense that $z \mapsto f(\Lambda_z)$ is holomorphic on $\mathfrak{h}$.

(iii) Holomorphicity at $\infty$: $f$ is holomorphic at $\infty$, in the sense that $z \mapsto f(\Lambda_z)$ converges as $z \to \infty$. We denote this value by $f(\infty)$.

Clearly, a modular form of weight $k$ is the same thing as a holomorphic function $f$ on $\mathfrak{h}^*$ which satisfies $f(\gamma z) = (cz + d)^k f(z)$ for all $\gamma \in \Gamma(1)$, where we use our usual notation. A modular form is a cusp form if it vanishes at the cusps.

**Remark.** The above condition applied to $\gamma = -1$ shows that $f(z) = (-1)^k f(z)$. So if $k$ is odd then $f = 0$. Therefore, there are only interesting modular forms of even weight.

### Modular interpretation

A modular form does not have a well-defined value on an elliptic curve – in other words, it is not a function on $Y(1)$. Rather, it is a section of a line bundle. This line bundle has a nice modular interpretation.

Let $\mathcal{L}$ be the space of lattices in $\mathbb{C}$. There is a natural family of elliptic curves wt$\pi\colon E \to \mathcal{L}$: the fiber above $\Lambda \in \mathcal{L}$ is $E_\Lambda = \mathbb{C}/\Lambda$. Let $w$ be a parameter on $\mathbb{C}$, so that $dw$ spans the space of holomorphic 1-forms on $E_\Lambda$ for any $\Lambda$. Note that $dw$ is not invariant under homothety: indeed, $\alpha^*(dw) = \alpha dw$. However, if $f$ is a weight $k$ modular form then $f(dw)^k$ is invariant under homothety, as a section of the $k$-fold tensor power of wt$\pi_*(\Omega^1_{E/\mathcal{L}})$.

Let $\pi\colon E \to Y(1)$ be the universal elliptic curve over $Y(1)$. We must treat $Y(1)$ as a stack for this to work properly, something we'll briefly discuss in the future. Let $\omega = \pi_*(\Omega^1_{E/Y(1)})$. Then $\omega$ is a line bundle on $Y(1)$ whose fiber at an elliptic curve $E$ is $\Gamma(E, \Omega^1_{E/\mathbb{C}})$. The bundle $\omega$ is called the Hodge bundle. The above discussion shows that a weight $k$ modular form defines a section of $\omega^{\otimes k}$. In fact, every section that satisfies an appropriate condition at $\infty$ defines a weight $k$ modular form.

It is possible to state the above description more concretely. Let $f$ be a weight $k$ modular form, let $E$ be an elliptic curve, and let $\omega$ be a non-zero 1-form on $E$. Writing $E = E_\Lambda$, the above discussion shows that $f(\Lambda)(dw)^k$ is a well-defined element of $\Gamma(E, \Omega^1)^{\otimes k}$. However, $\omega^k$ is also such an element, and so we can divide to get a well-defined number. In other words, $f$ defines a function $F$ from the set of pairs $(E, \omega)$ to $\mathbb{C}$. The function $F$ has two important properties:

(i) Homogeneity: $F(E, \alpha\omega) = \alpha^{-k} F(E, \omega)$.

(ii) Invariance: if $(E, \omega)$ is isomorphic to $(E', \omega')$ (in the obvious sense), then $F(E, \omega) = F(E', \omega')$.

Any $F$ satisfying these two properties, and a holomorphicity condition that we do not state now, comes from a modular form $f$ of weight $k$.

### Geometric interpretation

There is another useful way to think about modular forms, in terms of the geometry of the modular curve $X(1)$. Suppose $f$ is a modular form of weight $2k$ on $\mathfrak{h}$. Then $f(\gamma z) = (cz+d)^k f(z)$, by definition. A simple computation shows that $\gamma^*(dz) = (cz+d)^{-2} dz$. Thus $f(z)(dz)^k$ is invariant under $\Gamma$. Let $\pi\colon \mathfrak{h}^* \to X(1)$ be the quotient map. Then $f(z)(dz)^k = \pi^*(\omega)$ for some meromorphic section $\omega$ of $(\Omega^1)^{\otimes k})$ over $X(1)$. The local behavior of $f$ and $\omega$ are related as follows.

**Proposition.** Let $x \in \mathfrak{h}^*$ and let $y = \pi(x) \in X(1)$. Then

$$\operatorname{ord}_y(\omega) = \begin{cases} \frac{1}{2}(\operatorname{ord}_x(f) - k) & \text{if } x = i \\ \frac{1}{3}(\operatorname{ord}_x(f) - 2k) & \text{if } x = \rho \\ \operatorname{ord}_x(f) - k & \text{if } x = \infty \\ \operatorname{ord}_x(f) & \text{otherwise} \end{cases}$$

*Proof.* We just explain the $x = i$ case. Let $z$ be a uniformizing parameter of $\mathfrak{h}$ at $x$ and let $w$ be one on $X(1)$ at $y$. Then $\pi^*(w) = z^2$ and $\pi^*(dw) = zdz$ (up to higher order terms and constants). So if $\omega = w^n(dw)^k$ then $\pi^*(\omega) = z^{2n+k}dz$. Thus $\operatorname{ord}_x(f) = 2n + k = 2\operatorname{ord}_y(\omega) + k$. □

**Corollary.** The space of modular forms of weight $2k$ is isomorphic to the space of sections $\omega$ of $(\Omega^1)^{\otimes k}$ over $X(1) = \mathbb{P}^1$ which are holomorphic away from $\pi(i)$, $\pi(\rho)$, and $\pi(\infty)$, and satisfy $\operatorname{ord}_{\pi(i)}(\omega) \geq -k/2$, $\operatorname{ord}_{\pi(\rho)}(\omega) \geq -2k/3$, $\operatorname{ord}_{\pi(\infty)}(\omega) \geq -k$. A similar statement is true for cusp forms, but where the last condition is changed to $\operatorname{ord}_{\pi(\infty)}(\omega) \geq 1 - k$.

**Corollary.** The space of modular forms of weight $2k$ (for $k > 0$) has dimension $\lfloor k/6 \rfloor + \epsilon$, where $\epsilon$ is 1 if $k \neq 1 \pmod 6$. The space of cusp forms has dimension one less.

*Proof.* Let $P = \pi(i)$, $Q = \pi(\rho)$, $\infty = \pi(\infty)$. Then the above corollary says the space of modular forms of weight $2k$ is identified with the space of sections of $(\Omega^1)^{\otimes k}(nP + mQ + k\infty)$, where $n = \lfloor k/2 \rfloor$ and $m = \lfloor 2k/3 \rfloor$. This bundle has degree $-2k + n + m + k = n + m - k$, and thus $n + m - k + 1$ sections. It is elementary to show that this agrees with the stated formula. A modified argument applies to the cuspidal case. □

**Example.** There are no non-zero modular forms of weight 2. There is exactly one non-zero form, up to scalars, of weight 4, 6, 8, and 10. Then there are two of weight 12, one of which is cuspidal.

### Fourier Expansion

Any modular form $f$ on $\mathfrak{h}$ is invariant under the translation $z \mapsto z + 1$, and can therefore be expanded in powers of $q = e^{2\pi i z}$. The expansion

$$f(z) = \sum_{n \in \mathbb{Z}} a_n q^n$$

is called the Fourier expansion or $q$-expansion of $f$. The condition that $f$ be holomorphic at $\infty$ amounts to $a_n = 0$ for $n < 0$; given this, cuspidality is equivalent to $a_0 = 0$.

### Examples

Given a lattice $\Lambda \subset \mathbb{C}$ and an even integer $k \geq 4$, put

$$G_k(\Lambda) = \sideset{}{'}\sum_{\lambda \in \Lambda} \frac{1}{\lambda^k},$$

where the prime means to omit $\lambda = 0$. Clearly, $G_k(\alpha\Lambda) = \alpha^{-k} G_k(\Lambda)$, and so $G_k$ has the right homogeneity property to be a modular form. Put $G_k(z) = G_k(\Lambda_z)$. Then

$$G_k(z) = \sideset{}{'}\sum_{n,m} \frac{1}{(nz + m)^k},$$

which shows that $G_k$ is a holomorphic function of $z$. Furthermore, as $z \to \infty$, only the terms with $n = 0$ survive, and so $G_k(\infty) = 2\zeta(k)$, where $\zeta$ is the Riemann zeta function; in particular, $G_k$ is holomorphic (but non-zero) at $\infty$. Thus $G_k$ is a modular form of weight $k$. It is called the Eisenstein series of weight $k$. The modular form $E_k = (2\zeta(k))^{-1} G_k$ is called the normalized Eisenstein series of weight $k$. Its $q$-expansion is given

**Proposition.** We have

$$E_k(z) = 1 - \frac{4k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n) q^n$$

where $B_k$ is the Bernoulli number and $\sigma_{k-1}(n)$ is the sum of the $(k-1)$st powers of the divisors of $n$.

Let $\Delta = E_4^3 - E_6^2$. Then $\Delta$ is a modular form of weight 12 whose constant term vanishes. Computing with the first two terms of the $q$-series of $E_2$ and $E_4$, one finds that $\Delta(z) = q + \cdots$, and so $\Delta$ is non-zero. It is therefore the unique (up to scaling) non-zero cusp form of weight 12. Its $q$-expansion is complicated, but it admits the following nice product formula (due to Jacobi):

**Proposition.** $\Delta(z) = q\prod_{n\geq 1}(1-q^n)$.

The modular forms $E_4$, $E_6$, and $\Delta$ admit nice modular interpretations. Recall that every elliptic curve over $\mathbb{C}$ is isomorphic to one of the form

$$y^2 = x^3 + ax + b.$$

Call this curve $E_{a,b}$. Then $E_{a,b}$ is isomorphic to $E_{u^4a,u^6b}$, but there are no other isomorphisms. Let $\omega_{a,b}$ be the holomorphic 1-form on $E_{a,b}$ given by $y^{-1}dx$. Then under the natural isomorphism $f\colon E_{a,b} \to E_{u^4a,u^6b}$, we have $f^*(\omega_{u^4a,u^6b}) = u^{-1}\omega_{a,b}$. Thus a given pair $(E,\omega)$ is isomorphic to $(E_{a,b},\omega_{a,b})$ for a unique value of $(a,b)$. Furthermore, if $(E,\omega)$ is isomorphic to $(E_{a,b},\omega_{a,b})$ then $(E,u\omega)$ is isomorphic to $(E_{u^{-4}a,u^{-6}b},\omega_{u^{-4}a,u^{-6}b})$.

Define $F_4(E,\omega)$ to be the unique value of $a$ such that $(E,\omega)$ is isomorphic to $(E_{a,b},\omega_{a,b})$ then we find $F_4(E,u\omega) = u^{-4}F_4(E,\omega)$. Thus $F_4$ defines a modular form of weight 4. Similarly, if we define $F_6$ using $b$ then $F_6$ is a modular form of weight 6. The function taking $(E,\omega)$ to the discriminant of the corresponding $E_{a,b}$ is a modular form of weight 12. These coincide with $E_4$, $E_6$, and $\Delta$, up to constants.

### 2.1.4 Modular forms of higher level

The above theory can be generalized by replacing $\Gamma(1)$ with an arbitrary finite-index subgroup $\Gamma$. We sketch the general picture.

A modular form of weight $k$ for $\Gamma$ is a function $f\colon \mathfrak{h} \to \mathbb{C}$ satisfying the following conditions:

(i) $f(\gamma z) = (cz+d)^k f(z)$ for all $\gamma \in \Gamma$.

(ii) $f$ is holomorphic on $\mathfrak{h}$.

(iii) $f$ is holomorphic at the cusps.

The last condition should be explained. At the cusp infinity, it means $f(z)$ converges as $z \to i\infty$. Suppose $x$ is some other cusp, and $\gamma(\infty) = x$ for some $\gamma \in \Gamma(1)$. Then $g(z) = (cz+d)^{-k}f(\gamma z)$ is a modular form for $\gamma^{-1}\Gamma\gamma$, and $f$ is holomorphic at $x$ if and only if $g$ is holomorphic at $\infty$.

The modular interpretation carries over: weight $k$ modular forms can be identified with sections of $\omega^{\otimes k}$ over $Y_\Gamma$ satisfying appropriate conditions at the cusps. Concretely, this means that a modular form assigns to every triple $(E,?,\omega)$ a number, and is homogenous in $\omega$ of the appropriate degree. Here, the ? is the extra data associated to the moduli problem: for instance, for $\Gamma = \Gamma_1(N)$ it would be a point of order $N$.

The geometric interpretation carries over as well: a weight $2k$ modular form gives a meromorphic section of $(\Omega^1)^{\otimes k}$ over $X_\Gamma$. As before, one can specify the local conditions on $X_\Gamma$ that correspond to holomorphicity and cuspidality. The most important case the following:

**Proposition.** The space of weight 2 cusp forms for $\Gamma$ is identified with $\mathrm{H}^0(X_\Gamma,\Omega^1)$. In particular, the dimension of the space of weight 2 cusp forms for $\Gamma$ is the genus of $X_\Gamma$.

### 2.1.5 Hecke operators

#### On lattices

Recall that $\mathcal{L}$ is the set of lattices in $\mathbb{C}$. Let $\mathbb{Z}[\Lambda]$ denote the free abelian group of $\Lambda$. Let $n$ be an integer. We define an endomorphism $T(n)$ of $\mathbb{Z}[\Lambda]$ by

$$T(n)[\Lambda] = \sum_{[\Lambda':\Lambda]=n} [\Lambda']$$

and extend linearly to all of $\mathbb{Z}[\Lambda]$. For a complex number $\alpha$, define an operator $H_\alpha$ by $H_\alpha[\Lambda] = [\alpha\Lambda]$.

**Proposition.** We claim the following:

(i) If $n$ and $m$ are coprime then $T(nm) = T(n)T(m)$.

(ii) We have $T(p^{n+1}) = T(p^n)T(p) - pT(p^{n-1})H_p$ for $p$ prime.

(iii) The operators $T(n)$ and $T(m)$ commute for all $n$ and $m$.

*Proof.* We will prove our claim:

(i) Let $\Lambda''$ be a lattice of index $nm$ in $\Lambda$. Since $n$ and $m$ are coprime, there is a unique intermediate lattice $\Lambda'$ in $\Lambda$ of index $m$. Thus

$$T(nm)\Lambda = \sum_{[\Lambda'':\Lambda]=nm} [\Lambda''] = \sum_{[\Lambda':\Lambda]=m} \sum_{[\Lambda'':\Lambda']=n} [\Lambda''] = T(n)T(m)\Lambda.$$

(ii) We prove the $n = 2$ case, for simplicity. We have

$$T(p)T(p)\Lambda = \sum_{\Lambda'' \subset \Lambda' \subset \Lambda} [\Lambda''],$$

where each inclusion has index $p$. We thus find

$$T(p)^2[\Lambda] = \sum_{[\Lambda'':\Lambda]=p^2} n_{\Lambda''}[\Lambda''],$$

where the coefficient is the number of subgroups of $\Lambda/\Lambda''$ of order $p$. If this quotient is cyclic of order $p^2$, the coefficient is 1. Otherwise, it is the cardinality of $\mathbb{P}^1(\mathbb{F}_p)$, which is $p+1$. In $T(p^2)\Lambda$, we get the same sum, but with all coefficients equal to 1. Thus the only difference is that in $T(p)^2[\Lambda]$ the coefficient of $[p\Lambda]$ is $p+1$, while in $T(p^2)[\Lambda]$ it has coefficient 1. Thus $T(p)^2[\Lambda] - T(p^2)[\Lambda] = pH_p[\Lambda]$.

(iii) By (ii), $T(p^n)$ is a polynomial in $T(p)$. Thus the $T(p^n)$ commute with each other. The result now follows from (a). ¡h3¿On modular forms of level 1¡/h3¿ A modular form $f$ of weight $k$ for $\Gamma(1)$ is a function $f\colon \mathcal{L} \to \mathbb{C}$ satisfying $H_\alpha f = \alpha^{-k} f$, together with some holomorphicity conditions. For a modular form $f$ of weight $2k$ and an integer $n$, we put

$$(T(n)f)(\Lambda) = n^{2k-1} \sum_{[\Lambda':\Lambda]=n} f(\Lambda').$$

Since $T(n)$ and $H_\alpha$ commute, this still has the appropriate homogeneity properties to be a modular form of weight $2k$. The following proposition shows that it has the appropriate holomorphicity properties.

Hence proved. □

**Proposition.** Suppose $f(z) = \sum_{n\geq 0} a_n q^n$. Let $p$ be a prime. Then $(T(p)f)(z) = \sum_{n\geq 0}(a_{pn} + p^{2k-1}a_{n/p})q^n$, where $a_{n/p} = 0$ if $p$ does not divide $n$.

*Proof.* We need to compute $(T(p)f)(\Lambda_z)$. The index $p$ sublattices of $\Lambda_z$ are $\langle p, z+i\rangle$ for $0 \leq i \leq p-1$ and $\langle 1, pz\rangle = \Lambda_{pz}$. We have

$$p^{2k-1}\sum_{i=0}^{p-1} f(\langle p, z+i\rangle) = p^{-1}\sum_{i=0}^{p-1} f(\Lambda_{(z+i)/p}) = p^{-1}\sum_{n\geq 0}\sum_{i=0}^{p-1} a_n e^{2\pi i n(z+i)/p}.$$

The sum over $i$ is equal to $p$ when $p \mid n$, and 0 otherwise. We thus obtain

$$\sum_{p|n} a_n e^{2\pi i n z/p} = \sum_{n\geq 0} a_{np} q^n.$$

On the other hand,

$$f(\Lambda_{pz}) = \sum_{n\geq 0} a_n q^{np} = \sum_{n\geq 0} a_{n/p} q^n.$$

Combining, we obtain the stated result. □

**Corollary.** $T(n)f$ is a modular form of weight $2k$ for any $n$. If $f$ is cuspidal, so is $T(n)f$.

**Proposition.** The above calculations establish holomorphicity/cuspidality for $T(p)f$. The general case follows from this, since the $T(p)$ generate the $T(n)$.

**Remark.** Suppose $f$ is a cusp form with $a_1 = 1$ (normalized) which is an eigenvector for $T(p)$. Then its eigenvalue is equal to $a_p$, as the linear coefficient of $T(p)f$ is $a_p$. In fact, this holds for composite $p$ as

**Moduli description**

If $\Lambda'$ is an index $n$ sublattice of $\Lambda$, then there is a degree $n$ isogeny $\varphi \colon E_{\Lambda'} \to E_\Lambda$ whose kernel has cardinality $n$. Furthermore, $\varphi^*(dw) = dw$ for this isogeny. It follows that we can expression the Hecke operators moduli-theoretically as follows:

$$f(E, \omega) = \sum_{\varphi\colon E' \to E} f(E', \varphi^*(\omega))$$

where the sum is over all isomorphism classes of isogenies $\varphi \colon E' \to E$ whose kernel has cardinality $n$.

**In higher level**

Suppose $\Gamma$ is a finite index subgroup of $\Gamma(1)$ of level $N$, meaning it contains $\Gamma(N)$. Then $Y_\Gamma$ can be described as elliptic curves together with some $N$-torsion data. The Hecke operators $T(n)$ act on modular forms for $\Gamma$ so long as $n$ is prime to $N$: indeed, an isogeny $\varphi \colon E' \to E$ induces an isomorphism on $N$-torsion, and so any $N$-torsion data can be transported along $\varphi$. These operators commute, as before.

This is the first section on the arithmetic moduli theory of elliptic curves. We begin by explaining why the natural moduli problem for elliptic curves is not representable by a scheme. I then proved (in a fair amount of detail) that the moduli problem of elliptic curves with full level 3 structure is representable by a scheme. Using this, I deduced representability for full level N structure, for N¿3. I then returned to the level 1 case and briefly explained the concept of Deligne–Mumford stack and how it applies here.

### 2.1.6 The main issue with representability

We have previously defined $Y(1)$ as a Riemann surface via an analytic construction, and saw that its points correspond bijectively to elliptic curves over $\mathbb{C}$. We now want to do this theory algebraically, over any field (or base scheme). As with any moduli problem, we first approach $Y(1)$ by defining its functor of points. If $S$ is some scheme, then a map $S \to Y(1)$ should correspond to a family of elliptic curves over $S$. We already have a good notion of this: an elliptic curve over $S$ is a smooth proper scheme $E \to S$ equipped with a section $0 \in E(S)$ such that each geometric fiber is a geometrically connected genus 1 curve.

We are therefore lead to the following functor: for a scheme $S$, let $F_{\Gamma(1)}(S)$ denote the set of isomorphism classes of elliptic curves over $S$. We would like to define $Y(1)$ to be the scheme (over $\mathbb{Z}$) that represents $F_{\Gamma(1)}$. Unfortunately, $F_{\Gamma(1)}$ is not representable!

The reason for this is the same reason that the $j$-invariant does detect isomorphism classes over a non-algebraically closed field. If $X$ is any scheme over $\mathbb{Q}$ then the map $X(\mathbb{Q}) \to X(\overline{\mathbb{Q}})$ is injective. Since the map $F_{\Gamma(1)}(\mathbb{Q}) \to F_{\Gamma(1)}(\overline{\mathbb{Q}})$ is not injective (because of this problem with the $j$-invariant), $F_{\Gamma(1)}$ cannot be representable. In fact, this shows that $F_{\Gamma(1)}$ is not even a sheaf.

It is worthwhile to examine the situation a little more closely. Suppose $E$ and $E'$ are elliptic curves over a field $k$ that become isomorphic over the separable closure $k^s$ — we say that $E$ and $E'$ are twisted forms. Choose an isomorphism $\varphi\colon E \to E'$ over $k^s$. If $\sigma$ is an element of the absolute Galois group of $k$, then $\varphi^\sigma$ defines a (possibly different) isomorphism $E \to E'$ over $k^s$. Thus $\psi_\sigma = \varphi^\sigma \varphi^{-1}$ is an automorphism of $E$ over $k^s$. It is not difficult to verify that $\sigma \mapsto \psi_\sigma$ satisfies the 1-cocycle condition, and that $\varphi$ can be chosen to be defined over $k$ if and only if $\psi$ is a 1-coboundary. This construction in fact defines a bijection

$$\{\text{twisted forms of } E \text{ (up to isom.)}\} \to \mathrm{H}^1(\Gamma_k, \mathrm{Aut}(E_{k^s})).$$

For example, if $E$ is a non-CM curve over a field $k$ of characteristic 0, then $\mathrm{Aut}(E_{k^s}) = \{\pm 1\}$ with trivial Galois action, and the $\mathrm{H}^1$ is just $k^\times/(k^\times)^2$, and so twisted forms of $E$ correspond to square classes in $k$. Explicitly, if $E$ is given by $y^2 = f(x)$ and $d \in k^\times$, then the twist of $E$ corresponding to $d$ is given by $dy^2 = f(x)$.

### 2.1.7 Full level 3 structure

The above discussion shows that the existence of automorpisms of elliptic curves prevents the functor $F_{\Gamma(1)}$ from being representable. This suggests that we might have better luck if we look at more rigid objects.

Let $N \geq 2$ be an integer. Let $E \to S$ be an elliptic curve, with $N$ invertible on $S$. A $\Gamma(N)$-structure on $E$ is a pair of sections $(P, Q) \in E(S)[N]$ such that the map $(P, Q) : (\mathbb{Z}/N\mathbb{Z})^2_S \to E[N]$ is an isomorphism of group schemes over $S$; this is the same as asking that $(P, Q)$ give a basis of the $N$-torsion of $E_s$ for each geometric point $s$ of $S$. One can prove the following result:

**Proposition.** If $N \geq 3$ then any automorphism of $E$ fixing a $\Gamma(N)$-structure is the identity.

In other words, elliptic cures equipped with $\Gamma(N)$-structures have no non-trivial automorphisms for $N \geq 3$. Let $F_{\Gamma(N)}(S)$ denote the set of isomorphism classes of data $(E, (P, Q))$, where $E$ is an elliptic curve over $S$ and $(P, Q)$ is a $\Gamma(N)$-structure on $E$. It follows from the above proposition, and general principles, that $F_{\Gamma(N)}$ is a sheaf for $N \geq 3$. In the rest of this section, we investigate the $N = 3$ case.

Suppose $E/S$ is an elliptic curve and $(P, Q)$ is a $\Gamma(3)$-structure on $E$. By Riemann–Roch, locally on $S$ there is a function $x$ having a double pole at 0; it is unique up to $x \mapsto ax + b$. There is also a function $y$ having a triple pole at 0, and no other poles: it is unique up to $y \mapsto ay + bx + c$. After possibly scaling $x$ and $y$, they satisfy an equation of the form $y^2 + a_1xy + a_3y = h(x)$, for some cubic $h$.

Since $P$ is 3-torsion, the divisor $3[P] - 3[0]$ is principal, and so there is a unique function (up to scaling) with this divisor. Since 1, $x$, and $y$ span the space of functions having a triple pole at 0, there is a unique such function of the form $y + ax + b$. In other words, we can choose our original function $y$ uniquely (up to scaling) so that $y$ has a triple zero at $P$. The function $y^2 + a_1xy + a_3y$ has valuation 3 at $P$, and so $h(x)$ does as well. Since $h$ is cubic and $x$ is a degree 2, this implies $h(x) = (x - x(P))^3$ and $x$ vanishes to order 1 at $P$. We now replace $x$ with $x - x(P)$.

We have thus shown that $x$ and $y$ can be chosen uniquely, up to scaling, such that the equation becomes $y^2 + a_1xy + a_3y = x^3$, and $P$ is the point $(0, 0)$.

Now, $3[Q] - 3[0]$ is also a principal divisor, and so is the divisor of a function of the form $y - Ax - B$. We claim $A$ is invertible on $S$. To prove this it suffices to treat the case where $S$ is a field, and show $A \neq 0$. Suppose $A = 0$, so that $y - B$ has a triple zero at $Q$. Since $y$ is a degree 3 function, this means $Q$ is the only point at which $y - B$ vanishes, and so plugging $y = B$ into the defining equation gives a polynomial in $x$ with a triple root. Thus $x^3 - (B^2 + a_1Bx + a_3B) = (x - x(Q))^3$. Comparing the coefficients of $x^2$, we see that $x(Q) = 0$. (Here we use that we are not in characteristic 3.) But then $y(Q)^2 + a_3y(Q) = 0$, and so $y(Q)$ is either equal to 0 or $-a_3$. Thus $Q$ is given in coordinates by $(0, 0)$ or $(0, -a_3)$. But these two points are $P$ and $-P$, and $Q$ is not equal to either of them. Thus $A \neq 0$.

Since $A$ is a unit, we can replace $y$ by $y/A^3$ and $x$ by $x/A^2$, and assume $A = 1$. As $y - x - B$ vanishes only at $Q$, when we substitute $y = x + B$ into the defining equation the resulting polynomial has a triple root. Thus $x^3 - ((x + B)^2 + a_1x(x + B) + a_3(x + B)) = (x - C)^3$, where $C = x(Q)$. Equating like powers, we find

$$\begin{cases} 3C = a_1 + 1 \\ -3C^2 = 2B + a_1B + a_3 \\ C^3 = B^2 + a_3B \end{cases}$$

The first two define $a_1$ and $a_3$ in terms of $B$ and $C$. Subtracting the third from $B$ times the second gives $C^3 + 3C^2B + (a_1 + 1)B^2 = 0$. But $a_1 + 1 = 3C$, and so this equation is equivalent to $(B + C)^3 = B^3$. We have thus proved the following result:

**Proposition.** Let $E/S$ be an elliptic curve and let $(P, Q)$ be a $\Gamma(3)$-structure on $E$. Then there exist unique functions $x$ and $y$ on $E$ such that the following conditions hold:

(i) $x$ and $y$ have poles of order 2 and 3 at 0, and no other poles, and $y^2/x^3 = 1$ at 0.

(ii) $y$ vanishes to order 3 at $P$, and $x$ vanishes at $P$ (to order 1).

(iii) $y - x - B$ vanishes to order 3 at $Q$, for some function $B$ on $S$.

Furthermore, if $C = x(Q)$ then $(B + C)^3 = B^3$, and $E$ is defined by the equation

$$y^2 + a_1xy + a_3y = x^3$$

with $a_1 = 3C - 1$ and $a_3 = -3C^2 - B - 3BC$.

**Remark.** In coordinates, $P = (0, 0)$ and $Q = (C, B + C)$.

The converse to this proposition holds as well. Namely, given functions $B$ and $C$ on $S$, let $E$ be the curve defined by the above equations, let $P$ be the point $(0, 0)$ and let $Q$ be the point $(C, B + C)$. Then as long as $E$ is an elliptic curve (i.e., the discriminant of the equation is a unit), the above statements hold. These statements imply $\mathrm{div}(y) = 3[P] - 3[0]$ and $\mathrm{div}(y - x - B) = 3[Q] - 3[0]$, and so $P$ and $Q$ are 3-torsion. Furthermore, since the discriminant is a unit, $C$ is a unit, which implies $Q \neq \pm P$. Thus $(P, Q)$ is a $\Gamma(3)$-structure on $E$. This proves the following theorem:

**Theorem.** Let $R = \mathbb{Z}[1/3, B, C][1/\Delta]/(B^3 = (B + C)^3)$. Then $Y(3) = \mathrm{Spec}(R)$ represents the functor $F_{\Gamma(3)}$.

### 2.1.8 Full level $N$ structure

We would now like to prove that $F_{\Gamma(N)}$ is representable for all $N \geq 3$. However, the argument we gave in the $N = 3$ is too complicated if $N$ is much larger than 3. We therefore need an alternate approach. Fortunately, we can use the work we did for $N = 3$ to (almost) prove the result in general. We begin with the following.

**Proposition.** Let $E/S$ be an elliptic curve. Let $F$ be the functor on schemes over $S$ which attaches to $S' \to S$ the set $F(S')$ of $\Gamma(N)$-structures on $E_{S'}$. Then $F$ is represented by a finite étale scheme $T \to S$.

*Proof.* Let $T_0 = E[N] \times E[N]$, a finite étale group scheme over $S$. Let $T$ be the kernel of the Weil pairing $T_0 \to (\mu_N)_S$. Then $T$ is a closed subscheme of $T_0$, and therefore finite étale over $S$. Furthermore, giving a map $S' \to T$ is the same as giving $P, Q \in E(S')[N]$ such that in each fiber the Weil pairing of $P$ and $Q$ is non-zero, i.e., $(P, Q)$ is a $\Gamma(N)$-structure. □

**Theorem.** Suppose $N$ is prime to 3. Then $F_{\Gamma(3N)}$ is represented by a smooth affine scheme $Y(3N)$ over $\mathbb{Z}[1/3N]$.

*Proof.* Let $Y(3N) \to Y(3)$ be the finite étale scheme constructed in the previous proposition using the universal family over $Y(3)$. Giving a map $S \to Y(3N)$ is the same as giving an elliptic curve $E/S$ with both $\Gamma(3)$ and $\Gamma(N)$ structures. But giving $\Gamma(3)$ and $\Gamma(N)$ structures is the same as giving a $\Gamma(3N)$ structure by the Chinese remainder theorem. Thus $Y(3N)$ represents $F_{\Gamma(3N)}$. Since $Y(3N)$ is finite étale over $Y(3)$, and $Y(3)$ is smooth and affine, the same is true for $Y(3N)$. □

**Proposition.** Suppose $N \geq 4$ is prime to 3. Then $\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$ acts freely on $F_{\Gamma(3N)}$ and the quotient sheaf is $F_{\Gamma(N)}$.

*Proof.* The action is by moving around the $\Gamma(3)$-structure. Suppose $(E, (P, Q), (P', Q'))$ were fixed by the action of $g \in \mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$, where $(P, Q)$ is a $\Gamma(3)$-structure and $(P', Q')$ is a $\Gamma(N)$-structure. Thus there is some automorphism $f \colon E \to E$ taking $(P, Q)$ to $g(P, Q)$ and $(P', Q')$ to $(P', Q')$ since $N \geq 3$, this implies $f = 1$, and so $g = 1$. This proves the action is free. Finally, suppose $E/S$ is an elliptic curve with $\Gamma(N)$-structure. Let $T \to S$ be the finite étale cover parametrizing $\Gamma(3)$-structures on $E$. Then $E_T$ canonically has a $\Gamma(3N)$ structure. This shows that the map $F_{\Gamma(3N)}(T) \to F_{\Gamma(N)}(S)$ is surjective, and so the map of sheaves $F_{\Gamma(3N)} \to F_{\Gamma(N)}$ is surjective. It is clear that $F_{\Gamma(3N)}/\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$ is the maximal quotient it factors through. □

**Theorem.** Suppose $N \geq 4$ is prime to 3. Then $F_{\Gamma(N)}$ is represented by a smooth affine scheme $Y(N)$ over $\mathbb{Z}[1/3N]$.

*Proof.* By the above proposition, the group $\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$ acts freely on $Y(3N)$. The quotient $Y(N)$ therefore exists as a smooth affine scheme over $\mathbb{Z}[1/3N]$, and represents the quotient sheaf $F_{\Gamma(N)}$. □

Obviously, this is still less than what we want. The full result is:

**Theorem.** Suppose $N \geq 3$. Then $F_{\Gamma(N)}$ is represented by a smooth affine scheme $Y(N)$ over $\mathbb{Z}[1/N]$.

*Proof Idea.* We just give the idea of the proof. One first proves that a suitable replacement of $F_{\Gamma(2)}$ is representable. This done explicitly, similar to what we did for $F_{\Gamma(3)}$. Then one reasons as we did above to show that if $N \geq 3$ is prime to 2 then $F_{\Gamma(N)}$ is representable by a smooth affine scheme over $\mathbb{Z}[1/2N]$. A patching argument now shows that for any $N$ prime to 6, $F_{\Gamma(N)}$ is representable by a smooth affine scheme over $\mathbb{Z}[1/N]$. A small amount of extra work is needed to remove the "$N$ prime to 6" condition. □

### 2.1.9 Stacks

Let us re-exaimine why $F_{\Gamma(1)}$ fails to be a sheaf. If we are given an elliptic curve $E$ over $k^s$ such that $\sigma^*(E)$ is isomorphic to $E$ for all Galois automorphisms $\sigma$, we cannot necessarily descend $E$ to $k$, because these isomorphisms may not be compatible. This suggests that we shouldn't simply take isomorphism classes of elliptic curves, but consider the whole category (or at least groupoid) of elliptic curves.

A stack (on some topological space or site) is a rule $\mathcal{F}$ that assigns to each open set $U$ a groupoid $\mathcal{F}(U)$ and to each inclusion $U' \subset U$ a restriction functor $\mathcal{F}(U) \to \mathcal{F}(U')$ such that appropriate analogs of the sheaf axioms hold. The glueing axiom states that if $U = \bigcup U_i$ is an open cover and $X_i$ are objects of $\mathcal{F}(U_i)$ equipped with isomorphisms $\varphi_{ij} \colon X_i|_{U_{ij}} \to X_j|_{U_{ij}}$ satisfying the cocycle codition, then the $X_i$ glue to an object $X$ over $U$.

Suppose $G$ is a finite group acting on a variety $X$. The usual quotient $X/G$ is not well-behaved when $G$ has fixed points: for instance, the fibers of $X \to X/G$ do not all have cardinality $G$. However, the is always a well-behaved stack quotient $[X/G]$. It has the following property: if $Y \to [X/G]$ is any map from a scheme $Y$, then the fiber product $X \times_{[X/G]} Y$ is a scheme, and the fibers of the projection map $X \times_{[X/G]} Y \to Y$ are permuted simply transitively by $G$; in other words, $X \times_{[X/G]} Y \to Y$ is a $G$-torsor. In fact, the functor of points of $[X/G]$ takes a scheme $Y$ to the groupoid of objects $(T, f, g)$ where $T$ is a scheme with a $G$-action, $f \colon T \to Y$ gives $T$ the structure of a $G$-torsor over $Y$, and $g \colon T \to X$ is a $G$-equivariant map.

The stack $[X/G]$ above is not a scheme. However, it is not very far off from a scheme. A Deligne–Mumford stack is stack having similar properties to $[X/G]$. More precisely, it is a stack $X$ for which there exists a map $\mathrm{wt}X \to X$ (with $\mathrm{wt}X$ a scheme) with the following property: if $T \to X$ is any map with $T$ a scheme, then $T \times_X \mathrm{wt}X$ is a scheme and the projection map to $T$ is surjective and étale. We say that $\mathrm{wt}X$ is an étale cover of $X$.

Geometric properties of Deligne–Mumford stacks are defined by appealing to analogies with schemes. For example, if $\mathrm{wt}X \to X$ is an étale cover of schemes then $X$ is smooth (over whatever base) if and only if $\mathrm{wt}X$ is. Thus one says that a Deligne–Mumford stack is smooth if it admits an étale cover by a smooth scheme.

### 2.1.10 Back to level 1

For a scheme $S$, let $\mathcal{F}_{\Gamma(1)}(S)$ denote the groupoid of elliptic curves over $S$. One can then show that $\mathcal{F}_{\Gamma(1)}$ is a stack for the fppf site of schemes. Furthermore, $Y_{\Gamma(N)} \to \mathcal{F}_{\Gamma(1)}$ is relatively representable and étale over $\mathbb{Z}[1/N]$: indeed, giving a map $S \to \mathcal{F}_{\Gamma(1)}$ is the same as giving an elliptic curve $E/S$, and the fiber product is then the scheme of $\Gamma(N)$-structures on $E$, which we know to be finite étale over $S$. From this, we see that $Y(1) = \mathcal{F}_{\Gamma(1)}$ is a Deligne–Mumford stack.

### 2.1.11 Other moduli problems

Let $N \geq 2$ be an integer. We only consider schemes over $\mathbb{Z}[1/N]$. Let $E/S$ be an elliptic curve. A $\Gamma_1(N)$-structure on $E$ is a section $P \in E(S)[N]$ of order $N$. Let $\mathcal{F}_{\Gamma_1(N)}$ be the stack associating to $S$ the groupoid of $(E, P)$ with $E$ an elliptic curve over $S$ and $P$ a $\Gamma_1(N)$-structure. For $N \geq 3$, this problem is rigid, and $\mathcal{F}_{\Gamma_1(N)}$ is equivalent to a sheaf $F_{\Gamma_1(N)}$.

**Theorem.** The stack $Y_1(N) = \mathcal{F}_{\Gamma_1(N)}$ is a smooth Deligne–Mumford stack over $\mathbb{Z}[1/N]$. For $N \geq 3$, it is a smooth affine scheme.

A $\Gamma_0(N)$-structure on $E$ is a closed étale subgroup $G \subset E$ which is cyclic of cardinality $N$ in each geometric fiber. Let $\mathcal{F}_{\Gamma_0(N)}$ be the associated stack.

**Theorem.** The stack $Y_0(N) = \mathcal{F}_{\Gamma_0(N)}$ is a smooth Deligne–Mumford stack over $\mathbb{Z}[1/N]$.

Unlike previous cases, this example is never a scheme, since multiplication by $-1$ preserves and $\Gamma_0(N)$-structure. In fact, the automorphism groups can even be larger than order 2 in certain cases. For example, suppose $N$ is a prime congruent to 1 modulo 4, and let $E$ be an elliptic curve with $\operatorname{End}(E) = \mathbb{Z}[i]$. Under a suitable basis $(P, Q)$ of $E[N]$, we have $iP = aP$ and $iQ = -aQ$, where $a$ is a square root of $-1$ in $\mathbb{Z}/N\mathbb{Z}$. Thus if $G$ is the subgroup of $E[N]$ generated by $P$ or $Q$ then $(E, G)$ has $i$ as an automorphism.

## 2.2 Modular forms and the Hecke algebra

This section covers three topics: the coarse moduli space, compactifying modular curves via generalized elliptic curves, and defining modular curves over all over $\mathbb{Z}$. I also briefly discuss the fibers in bad characteristic.

### 2.2.1 Coarse spaces

Let $\mathfrak{M}_0(N)$ be the stack over $\mathbb{Z}[1/N]$ assigning to a scheme $S$ the groupoid of pairs $(E, G)$ where $E/S$ is an elliptic curve and $G \subset S$ is a cyclic subgroup of order $N$. We saw last time that $\mathfrak{M}_0(N)$ is a Deligne–Mumford stack. (I called it $Y_0(N)$ last time, but it will be better to not use that notation from now on.) Recall that this means it has an étale cover by a scheme. Explicitly, suppose $p \nmid N$ is a prime, and let $Y(S)$ be the set of isomorphism classes of data $(E, G, (P, Q))$, where $(E, G)$ is as above and $(P, Q)$ is a $\Gamma(p)$-structure on $E$. Then $Y$ is a representable by a scheme over $\mathbb{Z}[1/pN]$, and this provides an étale cover of $\mathfrak{M}_0(N)$ over $\mathbb{Z}[1/pN]$. In fact, $\mathfrak{M}_0(N)[1/p]$ is the quotient stack $[Y/\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})]$.

One can also consider the quotient scheme $Y/\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$. It is not hard to show that this represents the sheafification of the presheaf $S \mapsto |\mathfrak{M}_0(N)(S)|$ on $\mathbb{Z}[1/pN]$-schemes, where $|\cdot|$ denotes the set of isomorphism classes. Using different values of $p$ and patching, we see that $S \mapsto |\mathfrak{M}_0(N)(S)$ is represented by a $\mathbb{Z}[1/N]$-scheme, which we denote by either $M_0(N)$ or $Y_0(N)$. This is called the ¡em¿coarse space¡/em¿ of $\mathfrak{M}_0(N)$, and is the universal scheme to which $\mathfrak{M}_0(N)$ maps. When $N = 1$ then $Y_0(N)$ is the familiar $j$-line $\cong \mathbb{A}^1$. Furthermore, the set of complex points of $Y_0(N)$ is identified with $\mathfrak{h}/\Gamma_0(N)$. It is affine and smooth over $\mathbb{Z}[1/N]$, as can be seen from its description as $Y/\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$ when some prime $p \nmid N$ is inverted.

### 2.2.2 Compactification

#### Level 1

Recall that, over the complex numbers, $Y(1) = \mathfrak{h}/\Gamma(1)$ is not compact, and that we compactified it by adding cusps; precisely, the compactification $X(1)$ was defined as $\mathfrak{h}^*/\Gamma(1)$, where $\mathfrak{h}^*$ is the union of $\mathfrak{h}$ and $\mathbb{P}^1(\mathbb{Q})$. The group $\Gamma(1)$ acts transitively on $\mathbb{P}^1(\mathbb{Q})$, and so $X(1)$ has a unique cusp.

We would like to give a moduli-theoretic meaning to the cusp. The valuative crtierion for properness suggests that this should be related to the problem of extending elliptic curves over DVR's, which is something we know a lot about. In particular, the semi-stable reduction theorem tells us that, after a possible base change, every elliptic curve over $K = \mathrm{Frac}(A)$ ($A$ a DVR) extends to either an elliptic curve over $A$ or a nodal cubic over $A$. This suggests that the cusp should correspond to a nodal cubic.

We now give some definitions. Let $n \geq 1$ be an integer and let $k$ be a field. The ¡em¿standard $n$-gon¡/em¿ over $k$, denoted $C_n$, is the quotient of $(\mathbb{P}^1)_k \times \mathbb{Z}/n\mathbb{Z}$ where $(\infty, i)$ is identified with $(0, i+1)$. When $n = 1$, this gives the nodal cubic; in general, it has $n$ irreducible components. Note that the smooth locus $C_n^{\mathrm{sm}}$ of $C_n$ is just $\mathbb{G}_m \times \mathbb{Z}/n\mathbb{Z}$, and is a group. Furthermore, the action of $C_n^{\mathrm{sm}}$ on itself extends to an action of $C_n^{\mathrm{sm}}$ on all of $C_n$: the $\mathbb{G}_m$ part fixes the singular points.

Note that $C_n^{\mathrm{sm}}[n]$ has order $n^2$. In fact, there is a natural short exact sequence

$$0 \to \mu_n \to C_n^{\mathrm{sm}}[n] \to \mathbb{Z}/n\mathbb{Z} \to 0$$

where the $\mu_n$ sits in the identity component of $C_n^{\mathrm{sm}}$.

A generalized elliptic curve over a base scheme $S$ is, roughly, a curve over $S$ whose fibers are either elliptic curves or $n$-gons. More precisely, it is a tuple $(E, +, e)$, where $E/S$ is a proper flat curve, $e \in E(S)$, and $+$ is a map $E^{\mathrm{sm}} \times E \to E$ such that: (1) $+$ (with $e$) gives $E^{\mathrm{sm}}$ the structure of a group and defines aan action on $E$; (2) the geometric fibers of $E$ are elliptic curves or $n$-gons.

Define $\overline{\mathfrak{M}}(1)(S)$ to be the groupoid of generalized elliptic curves $E/S$ such that the fibers are either elliptic curves or 1-gons. Then one has the following result:

**Theorem.** $\overline{\mathfrak{M}}(1)$ is a proper smooth Deligne–Mumford stack over $\mathbb{Z}$.

This can be proved similarly to how we handled the $\mathfrak{M}(1)$ case. Note that in the valuative criterion for properness for stacks, one is allowed to make an extension of the DVR. Thus the properness in this theorem corresponds exactly to the semi-stable reduction theorem.

### Higher level

Let $E/S$ be a generalized elliptic curve. A $\Gamma_0(N)$-structure on $E$ is a cyclic subgroup $G$ of order $N$ inside of $E^{\mathrm{sm}}$. The definition of $\overline{\mathfrak{M}}_0(N)$ should clearly be related to $\Gamma_0(N)$-structures on generalized elliptic curves, but there is some subtlety.

To explain this, we should first familiarize ourselves more with the cusps on $X_0(N)$. For simplicity, let's suppose that $N$ is prime. The set of cusps if $\mathbb{P}^1(\mathbb{Q})/\Gamma_0(N)$, which is easily seen to be isomorphic to $\mathbb{P}^1(\mathbb{F}_N)/G$, where $G \subset \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ is the group of upper triangular matrices. The group $G$ has two orbits on $\mathbb{P}^1(\mathbb{F}_n)$, namely, that of 0 (which is all of $\mathbb{A}^1(\mathbb{F}_N)$) and that of $\infty$ (which is fixed by $G$). Thus $X_0(N)$ has two cusps, which are denoted 0 and $\infty$, as they are the images of $0, \infty \in \mathbb{P}^1(\mathbb{Q})$.

Now, the 1-gon only admits one $\Gamma_0(N)$-structure, namely $\mu_N \subset \mathbb{G}_m$. Thus if we only considered 1-gons with $\Gamma(N)$-structure we would only by adding 1 point to $\mathfrak{M}_0(N)(\mathbb{C})$. This is insufficient since there are two cusps.

We can get more $\Gamma_0(N)$-structures by allowing $N$-gons. Up to isomorphism, it has two $\Gamma_0(N)$ structures, namely, the subgroups $\mu_N$ and $\mathbb{Z}/N\mathbb{Z}$. This therefore looks like the right thing to use. However, for technical reasons (that we'll soon see), we need our level structure to meet all the irreducible components. We therefore take $(C_1, \mu_N)$ and $(C_N, \mathbb{Z}/N\mathbb{Z})$ to be the cusps, which we call 0 and $\infty$. Note that 0 is distinguished from $\infty$ by the fact that its $\Gamma_0(N)$-structure lives in the identity component.

Here is the definition of the moduli problem, for any $N$. We let $\overline{\mathfrak{M}}_0(N)(S)$ be the groupoid of pairs $(E, G)$, where $E/S$ is a generalized elliptic curve and $G \subset E^{\mathrm{sm}}$ is a cyclic subgroup of order $N$ such that in each fiber, $G$ meets each irreducible component of $E$. We then have the following result:

**Theorem.** $\overline{\mathfrak{M}}_0(N)$ is a proper smooth Deligne–Mumford stack over $\mathbb{Z}[1/N]$.

### Maps between moduli spaces

Over the complex numbers, $X_0(N)$ is identified with $\mathfrak{h}^*/\Gamma_0(N)$. Thus if $N' \mid N$ then the inclusion $\Gamma_0(N) \subset \Gamma_0(N')$ induces a map $X_0(N) \to X_0(N')$. How can we see this in terms of the moduli problem?

For $\mathfrak{M}_0(N)$ it is not hard: given $(E, G)$ with $G$ a cyclic subgroup of order $N$, there is a unique subgroup $H \subset G$ of order $N'$. The map $\mathfrak{M}_0(N) \to \mathfrak{M}_0(N')$ takes $(E, G)$ to $(E, H)$.

The map is more subtle over the compactified space, since the objects are different: $\mathfrak{M}_0(N)$ includes $N$-gons while $\mathfrak{M}_0(N')$ does not. The construction is as follows. Let $(E, G)$ be a generalized elliptic curve with $\Gamma_0(N)$-structure. Again, there is a unique $H \subset G$ which is cyclic of order $N'$. We now contract the components of $E$ which do not meet $H$. This can be done canonically as follows. Let $f \colon E \to S$ be the structure map. Let $\mathcal{A}$ be the sheaf of graded rings on $S$ given by $\bigoplus_{n=0}^{\infty} f_*(\mathcal{O}_E(nH))$. Then the contraction is $\mathrm{Proj}(\mathcal{A})$.

### 2.2.3 Working over $\mathbb{Z}$

#### A problem

We would now like to study moduli problems over $\mathbb{Z}$, as opposed to $\mathbb{Z}[1/N]$. Of course, this means that we'll have to include the case of elliptic curves in characteristic $N$. This causes problems with the definitions of level structure due to the lack of points.

More precisely, suppose we want to study $\mathfrak{M}(N)$. Our definition of a $\Gamma(N)$-structure on an elliptic curve $E$ is a pair $(P, Q)$ of $N$-torsion points which form a basis for $E[N]$. However, if $E/\mathbb{F}_N$ is supersingular, then there are no non-zero $N$-torsion points, and so therefore no $\Gamma(N)$-structures (under this definition)! This will clearly prevent the moduli space from being proper.

This problem can be solved by using the notion of a Drinfeld level structure. Fortunately, we will not need to go down this road.

#### The case of $\Gamma_0(N)$

We are primarily interested in the case of $\Gamma_0(N)$-structures, with $N$ squarefree, where a simpler definition can be given over $\mathbb{Z}$. Namely, if $E/S$ is a generalized elliptic curve, a $\Gamma_0(N)$-structure on $E$ is a closed subgroup $G \subset E$ which is finite and flat over $S$ of order $N$. We let $\overline{\mathfrak{M}}_0(N)(S)$ be the groupoid of pairs $(E, G)$ as above. We have the following result:

**Proposition.** $\overline{\mathfrak{M}}_0(N)$ is a flat Deligne–Mumford stack over $\mathbb{Z}$.

**Remark.** We can define $\overline{\mathfrak{M}}_0(N)$ for any $N$. When $N$ is not squarefree, however, one needs an appropriate definition of cyclic. In general, the non-compactified space $\mathfrak{M}_0(N)$ is a Deligne–Mumford stack. However, if $p^2 \mid N$ then $\mu_p$ can appear in the automorphism group of a generalized elliptic curve with $\Gamma_0(N)$-structure, which implies that $\overline{\mathfrak{M}}_0(N)$ cannot be a Deligne–Mumford stack; it is an Artin stack, however.

#### The fiber in bad characteristic

Let us now consider the space $\overline{\mathfrak{M}}_0(N)_{\mathbb{F}_p}$, where $p \mid N$. Suppose $k$ is an algebraically closed field of characteristic $p$ and $E/k$ is an elliptic curve. If $E$ is supersingular, then $E[p] \cong \alpha_{p^2}$ has a unique subgroup of order $p$, namely $\alpha_p$. If $E$ is ordinary, then $E[p] \cong \mu_p \times (\mathbb{Z}/p\mathbb{Z})$. In this case, $E[p]$ has exactly two subgroups of order $p$, namely $\mathbb{Z}/p\mathbb{Z}$ and $\mu_p$. The cuspidal points admit exactly two structures as well, $\mu_p$ and $\mathbb{Z}/p\mathbb{Z}$.

Let $N' = N/p$. We can then define two maps $f, g \colon \overline{\mathfrak{M}}_0(N')_{\mathbb{F}_p} \to \overline{\mathfrak{M}}_0(N)_{\mathbb{F}_p}$, as follows. Let $E/S$ be a generalized elliptic curve, where $S$ is an $\mathbb{F}_p$-scheme, with a $\Gamma_0(N')$-structure $G$. Then we have the relative Frobenius map $F \colon E \to E^{(p)}$ and Verscheibung $V \colon E^{(p)} \to E$. We define $f(E, G) = (E, G, \ker(F))$ and $g(E, G) = (E^{(p)}, V^{-1}(G))$.

We can also define two maps $f', g' \colon \overline{M}_0(N)_{\mathbb{F}_p} \to \overline{M}_0(N')_{\mathbb{F}_p}$, as follows. Let $E/S$ be an elliptic curve, let $G$ be a $\Gamma_0(N')$-structure on $E$, and let $H$ be a $\Gamma_0(p)$-structure on $E$. We define $f'(E, G, H) = (E, G)$ and $g'(E, G, H) = (E/H, \text{image of} G \text{in} E/H)$.

It is clear that $f'f = \mathrm{id}$. It is also clear that $g'g = \mathrm{id}$, since $E^{(p)}/\ker(V)$ is canonically isomorphic to $E$. All these maps induce maps on the coarse spaces, and so we see:

**Theorem.** $M_0(N)_{\mathbb{F}_p}$ is obtained by taking two copies of $M_0(N')_{\mathbb{F}_p}$ and glueing the supsersingular loci (a finite set of points) by the Frobenius map.

The picture is especially nice when $N' = 1$, as then $M_0(N')$ is just $\mathbb{P}^1$.

In the following section, we establish basic results on the structure of the Hecke algebra and some of its natural modules. In particular, we show that the Hecke algebra $\mathbb{T}$ is a finite rank free $\mathbb{Z}$-algebra and that $\mathbb{T} \otimes \mathbb{Q}$ is semi-simple. We also show that the space of weight 2 cusp forms (at prime level) is a free module of rank 1; this is the multiplicity one theorem.

We recall that for a modular form $f$ for $\Gamma(1)$ and weight $k$, we defined the action of the Hecke operator $T_n$ on $f$ by

$$(T_n f)(\Lambda) = n^{k-1} \sum_{[\Lambda:\Lambda']=n} f(\Lambda')$$

In fact, the same formula can be used to define $T_n f$ when $f$ is a modular form on $\Gamma_0(N)$ of weight $k$ if $n$ is prime to $N$. From now on, we only care about the $k = 2$ case.

We proved the following results:

(i) The $T_n$ commute with each other.

(ii) If $n$ and $m$ are coprime then $T_{nm} = T_n T_m$.

(iii) We have the recurrence $T_{p^{n+1}} = T_p T_{p^n} - p T_{p^{n-1}}$.

(iv) If $f = \sum_{n \geq 1} a_n q^n$ and $p$ is prime then

$$T_p f = \sum_{n \geq 1} (a_{pn} + p^{k-1} a_{n/p}) q^n$$

In particular, $a_1(T_p f) = a_p(f)$. In fact, $a_1(T_n f) = a_n(f)$ holds for all $n$ prime to $N$.

We define wt$\mathbb{T}$ to be the infinite polynomial ring $\mathbb{Z}[T_p]$ with $p$ prime to $N$. We define $\mathbb{T}$ to be the image of wt$\mathbb{T}$ in $\mathrm{End}(S_2(N))$. Our goal today is to understand this ring and the structure of $S_2(N)$ as a module over it.

### 2.2.4 The Petersson inner product

Let $f, g \in S_2(N)$. Then $f(z)dz$ and $g(z)dz$ are 1-forms on the upper half-plane invariant under $\Gamma_0(N)$. It follows that $\overline{g(z)dz}$ is also invariant, and so $f(z)dz \wedge \overline{g(z)dz} = 2if(z)\overline{g(z)}dxdy$ is also invariant. We define

$$\langle f, g \rangle = \int_{\mathfrak{h}/\Gamma_0(N)} f(z)\overline{g(z)}dxdy.$$

This is called the Petersson inner product. The integral converges since $f$ and $g$ decay rapidly at the cusps (as they are cusp forms). It is clear that $\langle , \rangle$ is a positive definite Hermitian form on $S_2(N)$. Furthermore, we have the following result:

**Proposition.** The operators $T_p$ are self-adjoint. That is, $\langle T_p f, g \rangle = \langle f, T_p g \rangle$.

**Corollary.** The algebra $\mathbb{T} \otimes \mathbb{C}$ is semi-simple. The space $S_2(\Gamma)$ admits a basis consisting of Hecke eigenforms (i.e., forms $f$ which are eigenvectors of $T_p$ for all $p \nmid N$).

We call a Hecke eigenform $f$ normalized if $a_1(f) = 1$. We note that for such forms, $a_n(f)$ is the eigenvalue of $T_n$ acting on $f$, for $n$ prime to $N$. If $f$ is an eigenform then we get a ring homomorphism $\alpha\colon \mathbb{T} \to \mathbb{C}$ by mapping $T_p$ to the eigenvalue of $f$ under $T_p$. One calls $\alpha$ a system of eigenvalues. The space of cusp forms decomposes as

$$S_2(N) = \bigoplus S_2(N)_\alpha,$$

where $S_2(N)_\alpha$ is the space of forms $f$ with $T_p f = \alpha(T_p) f$ for all $p \nmid N$.

### 2.2.5 Multiplicity one

**Theorem.** Suppose $N$ is prime and $f, g \in S_2(N)$ are two normalized Hecke eigenforms with the same eigenvalues for all $p \neq N$. Then $f = g$.

**Corollary.** For any system of eigenvalues $\alpha$, the space $S_2(N)_\alpha$ is one-dimensional.

**Corollary.** There is a bijection between homomorphisms $\mathbb{T} \to \mathbb{C}$ and normalized cusp forms.

**Corollary.** $S_2(N)$ is free of rank 1 as a $\mathbb{T} \otimes \mathbb{C}$ module.

**Remark.** There is also a stronger version of this theorem: if $f$ and $g$ are two normalized Hecke eigenforms that have the same eigenvalues under $T_p$ for all but finitely many $p$ (or even all $p$ in a density 1 set of primes) then $f = g$.

**Remark.** The above theorem is false for $N$ composite, but for a somewhat silly reason. If $p \mid N$ and $f \in S_2(N/p)$ then both $f(z)$ and $f(pz)$ belong to $S_2(N)$, and they have the same $T_\ell$ eigenvalues for $\ell \nmid N$. This is the only problem. More precisely, we define the old subspace to be the subspace of $S_2(N)$ spanned by $f(z)$ and $f(pz)$ where $p|N$ and $f \in S_2(N/p)$. We define the new subspace to be the orthogonal complement of the old subspace. Then multiplicity one holds on the new subspace.

### 2.2.6 Hecke correspondences

In terms of lattices, $(T_p f)(\Lambda)$ is defined by summing the values of $f$ over the index $p$ sublattices of $\Lambda$. If $\Lambda$ corresponds to the elliptic curve $E$, then the sublattices correspond to isogenies $E \to E'$ of degree $p$. The set of such isogenies is exactly the fiber of $X_0(pN) \to X_0(N)$ above $E$. Summing over the $E'$ the corresponds to pushing forward along the map taking $E \to E'$ to $E'$.

Let us be more precise. We previously defined a $\Gamma_0(p)$ structure on $E$ to be a cyclic subgroup $G$ of order $p$. But it is equivalent to define a $\Gamma_0(p)$ structure to be an isogeny $E \to E'$ of degree $p$ with cyclic kernel. For the moment, it will be more convenient to think of $X_0(p)$ as the space of such isogenies. We think of $X_0(Np)$ as parametrizing data $(E \to E', G)$ where $E \to E'$ is an isogeny of degree $p$ and $G \subset E$ is a cyclic subgroup of size $N$.

Let $p$ be prime to $N$. There are two natural maps $X_0(Np) \to X_0(N)$, namely, take $(f : E \to E', G)$ to $(E, G)$ or $(E', f(G))$. Call these maps $p_1$ and $p_2$. The diagram

$$\begin{tikzcd} & X_0(pN) \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\ X_0(N) & & X_0(N) \end{tikzcd}$$

is called the Hecke correspondence.

### 2.2.7 Generalities on correspondences

Let $C$ be a smooth projective curve. A correspondence $C \dashrightarrow C$ is a pair of maps $p_1, p_2 \colon C' \to C$ with $p_1$ and $p_2$ finite maps; we assume $C'$ is a smooth projective curve as well. We can think of a non-constant function $f \colon C \to C$ as a correspondence by taking $C' = C$, $p_1 = \mathrm{id}$, and $p_2 = f$. In general, one thinks of correspondences as multi-valued functions, where $x \in C$ is mapped to the set $p_2(p_1^{-1}(x))$.

Correspondences act on the singular cohomology $\mathrm{H}^1(C, \mathbb{Z})$ by the formula $(p_2)_* p_1^*$. Here $p_1^*$ is the usual pull-back operation $\mathrm{H}^1(C, \mathbb{Z}) \to \mathrm{H}^1(C', \mathbb{Z})$, and $(p_2)_*$ is the adjoint to $p_2^*$ under the cup product pairing.

Correspondences also act on differential forms, via the same formula. Over the complex numbers, the isomorphism

$$\mathrm{H}^1(C,\mathbb{Z})\otimes\mathbb{C}=\mathrm{H}^0(C,\Omega^1)\oplus\overline{\mathrm{H}^0(C,\Omega^1)}$$

from Hodge theory is compatible with the action of correspondences.

Correspondences also act on divisors, by the same formula. This action preserves principal divisors, and thus induces a map on the Jacobian. If $f$ is the correspondence $(p_1, p_2)$ then the induced endomorphism of $\mathrm{Jac}(C)$ is $(p_2^*)^\vee p_1^*$, where $p_i^* \colon \mathrm{Jac}(C)\to\mathrm{Jac}(C')$ is the natural map, and $(-)^\vee$ is the dual isogeny.

### 2.2.8 Back to Hecke operators

By the above discussion, the big Hecke algebra wt$\mathbb{T}$ acts on $\mathrm{H}^1(X_0(N),\mathbb{Z})$, and the isomorphism

$$\mathrm{H}^1(X_0(N),\mathbb{Z})\otimes\mathbb{C}=S_2(N)\oplus\overline{S_2(N)}$$

is compatible with the actions on each side. (Here we have identified $S_2(N)$ with $\mathrm{H}^0(X_0(N),\Omega^1)$.) If an element of wt$\mathbb{T}$ acts by zero on $S_2(N)$, then it does so on $\overline{S_2(N)}$ as well, and therefore on $\mathrm{H}^1(X_0(N),\mathbb{Z})$. We therefore see that the image of wt$\mathbb{T}$ in $\mathrm{End}(\mathrm{H}^1(X_0(N),\mathbb{Z}))$ is just $\mathbb{T}$. Since this endomorphism ring is $M_{2g}(\mathbb{Z})$, we see that:

**Proposition.** The Hecke algebra $\mathbb{T}$ is a free $\mathbb{Z}$-module of finite rank.

**Corollary.** The Hecke eigenvalues of a normalized cusp form are algebraic integers.

**Corollary.** $\mathbb{T}\otimes\mathbb{Q}$ is a product of finitely many number fields.

We have shown that $S_2(N)$ is free of rank 1 as a module over $\mathbb{T}\otimes\mathbb{C}$. The same is clearly true for $\overline{S_2(N)}$. Thus $\mathrm{H}^1(X_0(N),\mathbb{Z})\otimes\mathbb{C}$ is free of rank 2 over $\mathbb{T}\otimes\mathbb{C}$. Now, $\mathrm{H}^1(X_0(N),\mathbb{Q})$ is a module over the semi-simple ring $\mathbb{T}\otimes\mathbb{Q}$ which becomes free of rank 2 when tensored up to $\mathbb{C}$; it is therefore necessarily free of rank 2 itself. We've proved:

**Proposition.** $\mathrm{H}^1(X_0(N),\mathbb{Q})$ is free of rank 2 as a $\mathbb{T}\otimes\mathbb{Q}$ module.

### 2.2.9 The Atkin–Lehner involution

The space $X_0(N)$ admits a natural involution $w$ defined as follows. If we think of the points of $X_0(N)$ as cyclic isogenies of degree $N$, then $w$ takes $f\colon E\to E'$ to the dual isogeny $f^\vee\colon E'\to E$. If we think of the points of $X_0(N)$ as elliptic curves with a cyclic subgroup of order $N$ then $w$ takes $(E,G)$ to $(E/G, E[N]/G)$.

Thinking of weight 2 cusp forms as 1-forms on $X_0(N)$, we get an action of $w$ by pull-back. In terms of functions on the upper half-plane, $(wf)(z)=f(-1/Nz)$. One verifies that this action of $w$ on $S_2(N)$ commutes with the Hecke operators $T_p$ for $p\nmid N$. Assuming now that $N$ is prime (or $N$ is arbitrary and we use the new subspace). Since $w$ commutes with $\mathbb{T}$, it preserves the $\mathbb{T}$ eigenspaces, and these are one dimensional. It follows that if $f$ is an eigenform then $wf=\pm f$.

## 2.3 The Eichler-Shimura theorem

In the first part of this section we prove the Eichler–Shimura theorem. In the second part, we study Shimura's construction, and use it to construct the Galois representation associated to a weight 2 cusp form.

### 2.3.1 The Eichler–Shimura theorem

#### Statement

In the previous section, we defined the Hecke correspondence $T_p \colon X_0(N) \dashrightarrow X_0(N)$ over the complex numbers (for $p \nmid N$). The definition (which we review below) makes sense over the rational numbers. Thus $T_p$ induces an endomorphism of the Jacobian $J_0(N)$ of $X_0(N)$, over $\mathbb{Q}$. As $J_0(N)$ has good reduction at $p$, it extends uniquely to an abelian scheme over $\mathbb{Z}_p$ (which we still denote by $J_0(N)$), and $T_p$ extends uniquely as well. The Eichler–Shimura theorem concerns the reduction of this endomorphism at $p$:

**Theorem.** We have $T_p = F + V$ on $J_0(N)_{\mathbb{F}_p}$, where $F$ is Frobenius and $V$ is Verscheibung.

#### A lemma on correspondences

**Lemma.** Let $\mathcal{O}$ be a complete DVR with field of fractions $K$ and residue field $k$. Let $X/\mathcal{O}$ be a proper smooth curve, and let $f, g \colon Y \rightrightarrows X$ be finite flat maps giving a correspondence from $X$ to itself. Let $J_K = \mathrm{Jac}(X_K)$ and let $h_K \colon J_K \to J_K$ be the map induces by $(f, g)$. Let $J/\mathcal{O}$ be the Néron model of $J$, let $h \colon J \to J$ be the map induced by $h_K$, and let $h_k \colon J_k \to J_k$ be its base change. Let $D_0$ be a degree 0 divisor on $X_k$ defining a point $x_0 \in J(k)$. Then $h_k(x_0)$ is represented by $g_*(f^*(D_0))$.

*Proof.* Lift $D_0$ to a relative divisor $D$ on $X$, which is possible since $X$ is smooth over $\mathcal{O}$. Let $D' = g_*(f^*(D))$, which is again a relative divisor on $X$. Then $D$ and $D'$ define $\mathcal{O}$-points $x$ and $y$ of $J$. Then $h(x)$ is the unique $\mathcal{O}$-point of $J$ extending the $K$-point $h_K(x_K)$. However, $h_K(x_K)$ is by definition the point represented by $g_*(f^*(D_K))$, and is therefore equal to $y_K$. Since $y$ extends $y_K$, we have $h(x) = y$. It follows that $h_k(x_0) = y_0$, the reduction of $y$. But $y_0$ is computed by $g_*(f^*(D_0))$, since these operations commute with base change. □

#### Hecke correspondences, integrally

Recall that $\overline{\mathfrak{M}}_0(N)$ is the proper Deligne–Mumford stack over $\mathbb{Z}$ parametrizing generalized elliptic curves with $\Gamma_0(N)$-structure and $\overline{M}_0(N)$ is its coarse space. We also use the notation $X_0(N)$ when $N$ is inverted. Let $\mathrm{wt}f \colon \overline{\mathfrak{M}}_0(Np) \to \overline{\mathfrak{M}}_0(N)$ be the map which forgets the level $p$ structure, and let $\mathrm{wt}g \colon \overline{\mathfrak{M}}_0(Np) \to \overline{\mathfrak{M}}_0(N)$ be the composition of $f$ with the Atkin–Lehner involution at $p$. Precisely, $\mathrm{wt}g(E, H) = E/H$, where $E$ is an elliptic curve with $\Gamma_0(N)$-structure and $H$ is a $\Gamma_0(p)$-structure. Then $\mathrm{wt}f$ and $\mathrm{wt}g$ induce maps $f, g \colon \overline{M}_0(Np) \to \overline{M}_0(N)$ on the coarse spaces, which over $\mathbb{C}$ is the usual Hecke correspondence $T_p$. Moreover, $f$ and $g$ are finite and flat. It follows (from the previous lemma) that we can compute $T_p$, as an endomorphism of $J_0(N)_{\mathbb{F}_p}$, by the formula $g_* f^*$ on divisors.

#### Proof of the theorem

We work over $\mathbb{F}_p$. Let $\mathrm{wt}i \colon \overline{\mathfrak{M}}_0(N) \to \overline{\mathfrak{M}}_0(Np)$ be the map defined by $\mathrm{wt}i(E) = (E, \ker F)$, where $F \colon E \to E^{(p)}$ is Frobenius. Let $\mathrm{wt}j \colon \overline{\mathfrak{M}}_0(N) \to \overline{\mathfrak{M}}_0(Np)$ be the composition of $\mathrm{wt}i$ and the Atkin–Lehner involution at $p$. Precisely, $\mathrm{wt}j(E) = (E^{(p)}, \ker V)$, where $V$ is Verscheibung. Let $i$ and $j$ be the maps induced on coarse spaces. As we explained in the previous section, we have

$$fi = \mathrm{id}, \quad gj = \mathrm{id}, \quad fj = F, \quad gi = F$$

where $F \colon \overline{M}_0(N) \to \overline{M}_0(N)$ is Frobenius.

Let $M_0(N)^{\mathrm{ord}}$ and $M_0(Np)^{\mathrm{ord}}$ be the ordinary locus, i.e., the coarse moduli space parametrizing ordinary elliptic curves with appropriate level structure. The above picture, combined with the analysis of $\Gamma_0(p)$-structures on ordinary curves, shows that $M_0(Np)^{\mathrm{ord}}$ is isomorphic, via $i$ and $j$, to $M_0(N)^{\mathrm{ord}} \amalg M_0(N)^{\mathrm{ord}}$. Furthermore, on the first copy $f$ restricts to the identity and $g$ to Frobenius, while on the second copy, $f$ restricts to Frobenius and $g$ to the identity.

The above discussion shows that, on $M_0(N)^{\mathrm{ord}}$, the correspondence $T_p$ is the sum (disjoint union) of the self-correspondences $(1, F)$ and $(F, 1)$ of $M_0(N)^{\mathrm{ord}}$. The effect of $(1, F)$ on divisors is simply Frobenius, while the effect of $(F, 1)$ is the dual of $(1, F)$, i.e., Verscheibung. We thus see that if $D$ is a divisor supported on the ordinary locus, then $T_p(D) = F(D) + V(D)$, as divisors. Since every divisor is linearly equivalent to one supported on the ordinary locus (this is true whenever a finite number of points are removed from a curve), it follows that $T_p = F + V$ holds on all of $J_0(N)$.

**The Tate module of $J_0(N)$**

Let $V_\ell$ be the rational Tate module of $J_0(N)_{\mathbb{F}_p}$. This carries an action of Frobenius $F$. Since $\mathbb{T}$ acts by endomorphisms of $J_0(N)$, it acts on $V_\ell$, and so we can regard $V_\ell$ as a module for $\mathbb{T} \otimes \mathbb{Q}_\ell$. As such, it is free of rank 2. We can therefore regard the action of $F$ as a matrix in $\mathrm{GL}_2(\mathbb{T} \otimes \mathbb{Q}_\ell)$.

**Proposition.** We have $\mathrm{tr}(F \mid V_\ell) = T_p$ and $\det(F \mid V_\ell) = p$

*Proof.* Recall that we have the Weil pairing $\langle, \rangle \colon V_\ell \times V_\ell \to \mathbb{Q}_\ell(1)$. For any $\ell \nmid N$, the endomorphism $T_\ell$ of $J_0(N)$ is induced by a correspondence $(f_\ell, g_\ell)$ as above. The adjoint of $T_\ell$ is induced by the transposed correspondence $(g_\ell, f_\ell)$. However, these are easily seen to be equal: $(f_\ell, g_\ell)$ sums over $p$-isogenies $E \to E'$, while $(g_\ell, f_\ell)$ sums over $p$-isogenies $E' \to E$, and these sets are in bijection via the dual isogeny. Thus each $T_\ell$ is its own adjoint under $\langle, \rangle$.

We thus see that the isomorphism $\phi \colon V_\ell \to V_\ell^*$ defined by $x \mapsto \langle -, x\rangle$ commutes with the natural action of $\mathbb{T}$ on each side. As $\phi(Fx) = V\phi(x)$, we see that $\mathrm{tr}(F \mid V_\ell) = \mathrm{tr}(V \mid V_\ell^*)$. But the matrix for $V$ on $V_\ell^*$ is just the transpose of the matrix of $V$ on $V_\ell$ (in appropriate bases), and so $\mathrm{tr}(F \mid V_\ell) = \mathrm{tr}(V \mid V_\ell)$. But now appealing to $T_p = F + V$ again, we see that $2T_p = 2\,\mathrm{tr}(F)$, and so $\mathrm{tr}(F) = T_p$. Multiplying the Eichler–Shimura relation by $F$, we find $F^2 - T_pF + p = 0$. Since $\mathrm{tr}(F|V_\ell) = T_p$, it follows that $\det(F|V_\ell) = p$. □

**Remark.** It is very important in the above theorem that the trace and determinant are taken in the sense of $\mathbb{T} \otimes \mathbb{Q}_\ell$ modules. If we just think of $F$ as an endomorphism of the $\mathbb{Q}_\ell$ vector space $V_\ell$, its determinant is $p^g$, where $g = \dim(J_0(N))$.

### 2.3.2 The Shimura Construction

**The construction**

Fix a prime number $N$ and a normalized weight 2 cuspidal eigenform $f \in S_2(N)$. Let $\alpha \colon \mathbb{T} \to \mathbb{C}$ be the homomorphism giving the eigenvalues, i.e., $\alpha(T_p)$ is the $T_p$-eigenvalue of $f$. Then $\alpha(\mathbb{T} \otimes \mathbb{Q})$ is a number field $K \subset \mathbb{C}$ (in fact, the field generated by the $a_p(f)$), while $\alpha(\mathbb{T})$ is an order $\mathcal{O}$ in $K$. Let $\mathfrak{a} \subset \mathbb{T}$ be the kernel of $\alpha$. Define

$$A_f = J_0(N)/\mathfrak{a}J_0(N)$$

By $\mathfrak{a}J_0(N)$ we mean $\sum_{T \in \mathfrak{a}} TJ_0(N)$, which is an abelian subvariety of $J_0(N)$. Thus $A_f$ is an abelian variety over $\mathbb{Q}$. This is the Shimura construction. In the rest of this section, we'll study $A_f$.

**Remark.** Everything we'll do only depends on $\mathfrak{a}$: the embedding of $K = (\mathbb{T}/\mathfrak{a}) \otimes \mathbb{Q}$ into $\mathbb{C}$ will not be used.

**Remark.** This can be done for composite $N$ as well as long as one works with newforms.

**Dimension**

The tangent space to $J_0(N)$ at the identity is $\mathrm{H}^0(X_0(N), \Omega^1)$. Over the complex numbers, this $\mathrm{H}^0$ is identified with $S_2(N)$, which have shown is a free module of rank 1 over $\mathbb{T}_{\mathbb{C}}$. It follows that $T_0(J_0(N))$ is a free module of rank 1 over $\mathbb{T}_{\mathbb{Q}}$. We thus see that

$$T_0(J_0(N))/\mathfrak{a}T_0(J_0(N)) = T_0(A_f)$$

is a one dimensional vector space of $K$. We have thus shown:

**Proposition.** $A_f$ is an abelian variety of rank $[K : \mathbb{Q}]$.

**Corollary.** If $K = \mathbb{Q}$ then $A_f$ is an elliptic curve.

### Good reduction

We have the following result on the good reduction of $A_f$:

**Proposition.** $A_f$ has good reduction away from $N$.

In fact, this is an immediate corollary of the following general result:

**Proposition.** Let $B$ be an abelian variety over a DVR with good reduction, and let $A$ be a subquotient of $B$. Then $A$ has good reduction.

*Proof.* Let $\ell$ be a prime different from the residue characteristic. Since $B$ has good reduction, $V_\ell(B)$ is an unramified representation, by Néron–Ogg–Shafarevich. As $V_\ell(A)$ is a subquotient of $V_\ell(B)$, it too is unramified. Another application of Néron–Ogg–Shafarevich shows that $A$ has good reduction. □

### Structure of the Tate module

We have a natural map $\mathcal{O} = \mathbb{T}/\mathfrak{a} \to \mathrm{End}(A_f)$. We can therefore regard $V_\ell(A_f)$ as $K \otimes \mathbb{Q}_\ell$-module. With this in mind, we have:

**Proposition.** Let $p \nmid \ell N$ be a prime, and let $F_p \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ be the Frobenius at $p$. Then $\mathrm{tr}(F_p \mid V_\ell) = a_p$ and $\det(F_p \mid V_\ell) = p$.

*Proof.* It suffices to prove this over $\mathbb{F}_p$. Here it follows immediately from our above results on the Eichler–Shimura theorem. (Just apply the idempotent of $\mathbb{T}$ that projects onto $K$.) □

### The Galois representation associated to $f$

Choose an embedding $K \to \overline{\mathbb{Q}}_\ell$. We have the following extremely important result:

**Theorem.** There exists a unique semi-simple representation $\rho\colon \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ satisfying the following conditions:

(i) $\rho$ is unramified away from $N\ell$.

(ii) $\mathrm{tr}(\rho(F_p)) = a_p$ for $p \nmid N\ell$.

(iii) $\det(\rho) = \chi_\ell$.

*Proof.* We first prove existence. The choice of embedding $K \to \overline{\mathbb{Q}}_\ell$ determines a place of $K$ above $\ell$, i.e., an idempotent $e$ in $K \otimes \mathbb{Q}_\ell$. Simply take $\rho$ to be the semi-simplification of the representation on the space $eV_\ell(A_f)$. The first two points follow from the previous section. That section also shows that $\det(\rho(F_p)) = p$ for all $p \nmid N\ell$, which implies that $\det(\rho) = \chi_\ell$.

Uniqueness follows from Chebotarev and an exercise in group theory. Precisely, suppose that $\rho'$ were a second representation satisfying the same conditions. Then $\mathrm{tr}(\rho(F_p)) = \mathrm{tr}(\rho'(F_p))$ holds for all $p \nmid N\ell$. Chebotarev then implies that $\mathrm{tr}(\rho(g)) = \mathrm{tr}(\rho'(g))$ for all $g$ in the Galois group, since the $F_p$ are dense. It now follows that $\rho$ and $\rho'$ are equivalent, as they are semi-simple representations with the same character. □

**Remark.** In fact, the representation $\rho$ constructed above is absolutely irreducible.

**Remark.** Instead of taking our data to be a form $f$ and an embedding of its coefficient field into $\overline{\mathbb{Q}}_\ell$, we could simply have considered a homomorphism $\mathbb{T} \to \overline{\mathbb{Q}}_\ell$. In other words, for any homomorphism $\alpha\colon \mathbb{T} \to \overline{\mathbb{Q}}_\ell$ we get a representation $\rho_\alpha$ as above. Furthermore, one has a decomposition

$$\mathrm{H}^1_{\mathrm{et}}(X_0(N)_{\overline{\mathbb{Q}}}, \overline{\mathbb{Q}}_\ell) = \bigoplus \rho_\alpha$$

where the sum is over all $\alpha\colon \mathbb{T} \to \overline{\mathbb{Q}}_\ell$.

**Application: strong multiplicity 1**

**Theorem.** Let $f, g \in S_2(N)$ (with $N$ prime) be normalized (i.e., $a_1 = 1$). Suppose there is a density 1 set of primes $S$ such that $f$ and $g$ are eigenvectors of $T_p$ for all $p \in S$ with the same eigenvalues. Then $f = g$.

*Proof.* Let $\alpha\colon S \to \mathbb{C}$ be the function taking $p$ to the $T_p$-eigenvalue of $f$ and $g$. Let $V \subset S_2(N)$ be the space of forms $h$ satisfying $T_p h = \alpha(p)h$. It suffices to show that $V$ is one-dimensional. Now, $V$ has a basis consisting of eigenforms for the full Hecke algebra $\mathbb{T}$. It thus suffices to show that if $h, h' \in V$ are normalized eigenforms for the full $\mathbb{T}$ then $h = h'$.

Let $K \subset \mathbb{C}$ contain the eigenvalues of $h$ and $h'$, and choose an embedding of $K$ into $\overline{\mathbb{Q}}_\ell$. We then get semi-simple representations $\rho, \rho'\colon \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ satisfying $\mathrm{tr}(\rho(F_p)) = a_p(h)$ and $\mathrm{tr}(\rho'(F_p)) = a_p(h')$ for all $p \nmid N\ell$. Thus $\mathrm{tr}(\rho(F_p)) = \mathrm{tr}(\rho'(F_p))$ holds for all $p \in S$. However, by Chebotarev then $F_p$ with $p \in S$ are dense in the Galois group, and so $\rho \cong \rho'$. We thus see that $a_p(h) = a_p(h')$ for all $p \nmid N\ell$. By using two different $\ell$'s, we get this for all $p \neq N$. But now $h = h'$ by the version of multiplicity one we previously proved. □

## 2.4 Criterion for non-existence of torsion (Theorem A)

We prove Theorem A (which we saw in the beginning), which states that if one can find an appropriate map from $X_0(N)$ to an abelian variety, then no elliptic curve over the rational numbers has a rational point of order $N$. Combined with Theorem B, this gives an axiomatic approach to proving Mazur's theorem.

### 2.4.1 Statement of criterion

Our purpose today is to prove the following result (Theorem A):

**Theorem** (Theorem 1)**.** Let $N > 7$ be a prime number. Suppose there exists an abelian variety $A/\mathbb{Q}$ and a map of varieties $f\colon X_0(N) \to A$ satisfying the following conditions:

(i) $A$ has good reduction away from $N$.

(ii) $A(\mathbb{Q})$ has rank 0.

(iii) $f(0) \neq f(\infty)$.

Then no elliptic curve defined over $\mathbb{Q}$ has a rational point of order $N$.

Combined with Theorem B, we have the following criterion:

**Theorem** (Theorem 2)**.** Let $N > 7$ be a prime number and let $p \neq N$ be a second prime number. Suppose there exists an abelian variety $A/\mathbb{Q}$ and a map $f\colon X_0(N) \to A$ satisfying the following:

(i) $A$ has good reduction away from $N$.

(ii) $A$ has completely toric reduction at $N$.

(iii) The Jordan–Holder constituents of $A[p](\overline{\mathbb{Q}})$ are 1-dimensional and either trivial or cyclotomic.

(iv) $f(0) \neq f(\infty)$.

Then no elliptic curve defined over $\mathbb{Q}$ has a rational point of order $N$.

**Remark.** The above theorem, and the proof presented here, comes from III.5 of Mazur's paper "Modular curves and the Eisenstein ideal" (MR488287). It is not stated there explicitly, however.

### 2.4.2 Initial Reduction

We will prove the following theorem in what follows. In this section, we explain how it implies Theorem 1.

**Theorem** (Theorem 3)**.** Suppose that $A$ and $f$ are as in the statement of the theorem. Suppose that $E/\mathbb{Q}$ has a point of order $N$. Then $E[N] \cong \mathbb{Z}/N\mathbb{Z} \oplus \mu_N$.

Note that $E$ having a point of order $N$ implies that $\mathbb{Z}/N\mathbb{Z}$ is a sub group scheme of $E[N]$; the Weil pairing implies that the quotient is $\mu_N$. The content of the above proposition is that this extension is split.

Before proving Theorem 1, we need two lemmas:

**Lemma.** Suppose we have $A$ and $f$ as in the statement of the theorem. Then $X_0(N)(\mathbb{Q})$ is finite.

*Proof.* By assumption, $A(\mathbb{Q})$ is finite. Since $f(0) \neq f(\infty)$, the map $f$ is non-constant, and so the fibers of the map $f\colon X_0(N)(\mathbb{Q}) \to A(\mathbb{Q})$ are finite. The result follows. □

**Lemma.** Let $E/\mathbb{Q}$ be an elliptic curve. Then $\mathrm{End}(E) = \mathbb{Z}$. (I.e., even if $E$ is CM, its extra endomorphisms are not defined over $\mathbb{Q}$.)

*Proof.* We have a ring homomorphism $\mathrm{End}(E) \otimes \mathbb{Q} \to \mathrm{End}(T_0 E)$, where $T_0 E$ is the tangent space at the identity. Of course, $\mathrm{End}(T_0 E) = \mathbb{Q}$. We know that over any field of characteristic 0, $\mathrm{End}(E)$ is either $\mathbb{Z}$ or an order in an imaginary quadratic field. Thus $\mathrm{End}(E) \otimes \mathbb{Q}$ is a number field. Since it admits a homomorphism to $\mathbb{Q}$, it must be $\mathbb{Q}$ itself, and so $\mathrm{End}(E) \cong \mathbb{Z}$. □

We now prove the Theorem 1, assuming Theorem 3. Let $E_1$ be an elliptic curve over $\mathbb{Q}$ having a point $P_1$ of order $N$. By Theorem 3, $E$ contains $\mu_N$ as a subgroup. Define $E_2 = E_1/mu_N$. The image $P_2$ of $P_1$ in $E_1$ is a point of order $N$. Thus, Theorem 3 yields a $\mu_N$ in $E_1$. Continuing, we obtain a sequence of isogenies

$$E_1 \to E_2 \to E_3 \to \cdots$$

each of degree $N$. Furthermore, if $P_i$ denotes the image of $P_1$ in $E_i$ then the isogeny $E_i \to E_{i+1}$ does not kill $P_i$.

By the first Lemma, two of these curves, say $E_i$ and $E_j$ (with $i < j$), must be isomorphic. Let $f\colon E_i \to E_j$ be an isomorphism, and let $g\colon E_i \to E_j$ be the degree $N^{j-i}$ isogeny defined above. Then $f^{-1}g$ is an endomorphism of $E_i$ of degree $N^{j-i}$ and does not kill $P_i$, a point of order $N$. It follows that $f^{-1}g$ cannot be multiplication by an integer (as it would have to be multiplication by $N^{(i-j)/2}$ by degree considerations, but this kills $P_i$), and so $\mathrm{End}(E) \neq \mathbb{Z}$. But this contradicts the second Lemma. We conclude that no such $E_1$ exists. This proves Theorem 1.

### 2.4.3 Proof of the Theorem 3

We fix an elliptic curve $E/\mathbb{Q}$ having a point $P$ of order $N$. Our goal is to show that the sequence

$$0 \to \mathbb{Z}/N\mathbb{Z} \to E[N] \to \mu_N \to 0$$

splits. Let $\mathcal{E}/\mathbb{Z}$ be the Néron model of $E$, and let $\mathcal{P}$ be the $\mathbb{Z}$-point of $\mathcal{E}$ extending $P$. We proceed in four steps.

**Step 1**

**Proposition.** $E$ has everywhere semi-stable reduction.

*Proof.* Suppose $E$ has additive reduction at a prime $p$. By the classification of the special fibers of Néron models, we know that $\mathcal{E}_{\mathbb{F}_p}$ has at most four components. Since $N > 7$ is prime, the image of $\mathcal{P}_{\mathbb{F}_p}$ in the component group must therefore vanish. It follows that $\mathcal{P}_{\mathbb{F}_p}$ is contained in the identity component. However, the identity component is $p$-torsion, since $E$ has additive reduction. We conclude that $p = N$.

By the semi-stable reduction theorem, there is an extension $K/\mathbb{Q}_p$ such that $E$ has semi-stable reduction over $K$. In fact, we can take $K$ to have degree at most 6 over $\mathbb{Q}_p$ (since $p = N > 3$). Let $\mathcal{O}$ be the ring of integers in $K$, let $k$ be the residue field of $\mathcal{O}$, and let $\mathcal{E}'$ be the Néron model of $E$ over $\mathcal{O}$. The Néron mapping property yields a map $f\colon \mathcal{E}_{\mathcal{O}} \to \mathcal{E}'$. We claim that $f$ maps $\mathcal{E}_k^\circ$ to the identity. Indeed, $\mathcal{E}_k^\circ$ is $\mathbb{G}_a$, while $(\mathcal{E}')_k^\circ$ is an elliptic curve or a torus. Since there are no non-constant maps from $\mathbb{G}_a$ to an elliptic curve or torus, the claim follows.

Let $\mathcal{P}' \in \mathcal{E}'(\mathcal{O})$ be the section extending $P$. We note $\mathcal{P}' = f(\mathcal{P})$ since the two sections agree over $K$. In particular, $\mathcal{P}'$ reduces to the identity element of $\mathcal{E}'(k)$. But this is a contradiction, as we have previously proved [RC46CS] the reduction map $\mathcal{E}'[N](\mathcal{O}) \to \mathcal{E}'[N](k)$ to be injective when $e < p - 1$ (and note here that $e \leq 6$ and $p = N > 7$). □

**Remark.** The key result above (injectivity of reduction) makes essential use of Raynaud's theorem.

**Step 2**

**Proposition.** Suppose $p \in \{2, 3\}$. Then $E$ has multiplicative reduction at $p$ and the reduction of $\mathcal{P}$ at $p$ is not contained in the identity component of $\mathcal{E}_{\mathbb{F}_p}$.

*Proof.* Suppose $E$ has good reduction at $p$. Since $N$ is prime to $p$, the reduction map on $N$-torsion is injective, and so the reduction of $\mathcal{P}$ at $p$ is a point of order $N$ on $\mathcal{E}_{\mathbb{F}_p}$. But by the Hasse bound, $\#\mathcal{E}(\mathbb{F}_p) \leq p + 1 + 2\sqrt{p} \leq 7.5$ and $N > 7$. This is a contradiction, and so $E$ must have bad reduction. We have previously ruled out additive reduction.

Since $E$ has multiplicative reduction, $\mathcal{E}_{\mathbb{F}}^\circ$ is a one dimensional torus over $\mathbb{F}_p$. Up to isomorphism, there are two such tori: $\mathbb{G}_m$, and the norm 1 piece of the restriction of scalars of $\mathbb{G}_m$ from $\mathbb{F}_{p^2}$. These have $p - 1$ and $p + 1$ points over $bF_p$. It follows that $\mathcal{P}$ cannot reduce into $\mathcal{E}_{\mathbb{F}}^\circ$, since $N > p + 1$. □

**Corollary.** We have $E[N] = \mathbb{Z}/N\mathbb{Z} \oplus \mu_N$ over $\mathbb{Q}_p$.

*Proof.* Let $G \subset E[N]$ be the subgroup scheme that reduces into the identity component of $\mathcal{E}_{\mathbb{F}_p}$. Then $G$ has order $N$ but does not intersect the $\mathbb{Z}/N\mathbb{Z}$ coming from $P$, and is thus a complementary space to this $\mathbb{Z}/N\mathbb{Z}$. □

**Step 3**

**Proposition.** Let $p \notin \{2, 3\}$ be a prime of bad reduction for $E$. Then the reduction of $\mathcal{P}$ at $p$ is not contained in the identity component of $\mathcal{E}_{\mathbb{F}_p}$.

*Proof.* First note: (1) the reduction of $\mathcal{P}$ is a point of order $N$ on $\mathcal{E}_{\mathbb{F}_p}$, since the reduction map is injective on torsion; and (2) $E$ has multiplicative reduction at $p$ (by Step 1). If $p = N$ then $\mathcal{E}^\circ(\mathbb{F}_N)$ has either $N-1$ or $N+1$ points and therefore does not contain a point of order $N$. We therefore assume $p \neq N$ in what follows.

We are interested in three $\mathbb{Z}[1/N]$ points of $X_0(N)$:

(i) The point $\infty$ corresponding to the generalized elliptic curve which is a 1-gon. The smooth locus is $\mathbb{G}_m$, and the $\Gamma_0(N)$-structure is $\mu_N$, which is contained in the identity component.

(ii) The point 0 corresponding to the generalized elliptic curve which is an $N$-gon. The smooth locus is $\mathbb{G}_m \times \mathbb{Z}/N\mathbb{Z}$ and the $\Gamma_0(N)$-structure is $\mathbb{Z}/N\mathbb{Z}$, which is not contained in the identity component.

(iii) The point $x$ corresponding to $(\mathcal{E}, \mathbb{Z}/N\mathbb{Z})$, where $\mathcal{E}$ is the curve we have at hand. Note that this defines a $\mathbb{Z}[1/N]$-point of $X_0(N)$ since $E$ has everywhere semi-stable reduction, and thus its minimal regular model is a generalized elliptic curve.

Let $\mathcal{A}/\mathbb{Z}[1/N]$ be the abelian scheme extending $A$. The reduction map $\mathcal{A}(\mathbb{Z}[1/N])_{\text{tors}} \to \mathcal{A}(\mathbb{F}_p)$ is injective for $p \neq 2$, since for such $p$ we have $e = 1 < p - 1$. But by hypothesis, every $\mathbb{Z}[1/N]$-point of $\mathcal{A}$ is torsion, and so we can say that the reduction map $\mathcal{A}(\mathbb{Z}[1/N]) \to \mathcal{A}(\mathbb{F}_p)$ is injective. Note that the map $f$ extends to a map $f \colon X_0(N) \to \mathcal{A}$ over $\mathbb{Z}[1/N]$ by the Néron mapping property.

By Step 2, $x$ and 0 reduced to the same point of $X_0(N)(\mathbb{F}_3)$. Thus $f(x)$ and $f(0)$ reduce to the same point of $\mathcal{A}(\mathbb{F}_3)$. It follows from the above reasoning that $f(x) = f(0)$ in $\mathcal{A}(\mathbb{Z}[1/N])$.

Suppose now that the reduction of $\mathcal{P}$ were contained in $\mathcal{E}^\circ_{\mathbb{F}_p}$. Then $x$ reduces to $\infty$ in $X_0(N)(\mathbb{F}_p)$. Thus $f(x)$ and $f(\infty)$ reduce to the same point of $\mathcal{A}(\mathbb{F}_p)$, and so $f(x) = f(\infty)$. But this contradicts $f(0) \neq f(\infty)$. The result follows. □

**Corollary.** Under the above hypotheses, $E[N] \cong \mathbb{Z}/N\mathbb{Z} \oplus \mu_N$ over $\mathbb{Q}_p$.

### Step 4

Let $\Gamma$ be the absolute Galois group of $\mathbb{Q}$, and let $\rho \colon \Gamma \to \mathrm{GL}_2(\mathbb{F}_N)$ be the representation given by $E[N]$. Let $K = \mathbb{Q}(\mu_N)$. We begin with the following observation:

**Proposition.** $\rho|_K$ is everywhere unramified.

*Proof.* We consider several cases:

(i) $p \neq N$ is a prime of good reduction. Then $\rho|_{\mathbb{Q}_p}$ is already unramified by the easy direction of Néron–Ogg–Shafarevich.

(ii) $p = N$ is a prime of good reduction. Then $\mathcal{E}[N]$ is a finite flat group scheme over $\mathbb{Z}_p$ containing $\mathbb{Z}/N\mathbb{Z}$ as a sub and $\mu_N$ as a quotient. The connected–étale sequence goes the opposite direction, and thus implies that $\mathcal{E}[N] = \mu_N \oplus \mathbb{Z}/N\mathbb{Z}$ over $\mathcal{O}$. Thus $\rho|_{\mathbb{Q}_p} = \mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/N\mathbb{Z}(1)$, and is thus unramified over $K$.

(iii) $p$ is a prime of bad reduction. By Steps 2 and 3 above, we have $E[N] = \mathbb{Z}/N\mathbb{Z} \oplus \mu_N$, and so the result follows.

Hence proved. □

**Theorem** (Herbrand)**.** Let $j > 1$ be odd. Then $H^j \neq 0$ only if $N$ divides $B_{N-j}$.

In particular, we see that $H^{-1} = H^{N-2}$ is non-zero only if $N$ divides $B_2 = 1/6$. Since this never happens, we see that $H^{-1} = 0$, and so $f_0 = 0$. But this means that $\rho|_K$ is the trivial representation, and so we can regard $\rho$ as a representation of $\mathrm{Gal}(K/\mathbb{Q})$. It follows that $\rho$ is semi-simple, since the order of this group is prime to $N$. This proves Theorem 3.

The purpose of the next section is to show that $J_0(N)$, and any quotient of it, has completely toric reduction at $N$. We do this in three steps: (1) We analyze the minimal regular model of $X_0(N)$ and show that its special fiber is a nodal curve whose irreducible components are $\mathbb{P}^1$'s; (2) We recall a theorem of Raynaud, relating the Néron model of the Jacobian of a curve to the Picard scheme of its minimal model; (3) Combining the first two results shows that the special fiber of the Néron model of $J_0(N)$ is the Picard scheme of a nodal curve whose irreducible components are $\mathbb{P}^1$'s. We explicitly compute this and find that it is a torus.

### 2.4.4 This sections's goal

Let $N > 7$ be a prime. Thanks to the previous section, our task is to find an abelian variety $A/\mathbb{Q}$ and a map $f\colon X_0(N) \to A$ satisfying the following four hypotheses:

(i) $A$ has good reduction away from $N$.

(ii) $A$ has completely toric reduction at $N$.

(iii) The Jordan–Holder constituents of $A[p](\overline{\mathbb{Q}})$ are trivial or cyclotomic.

(iv) $f(0) \neq f(\infty)$.

We know an example of an abelian variety to which $X_0(N)$ maps: the Jacobian $J_0(N)$ of $X_0(N)$. In fact, this is the universal such abelian variety. Thus to find $A$ we need only consider quotients of $J_0(N)$.

As we have seen, the first condition above ("$A$ has good reduction away from $N$") comes for free for quotients of $J_0(N)$, thanks to the following results:

(i) $X_0(N)$ admits a smooth model away from $N$.

(ii) The Jacobian of such a curve has good reduction away from $N$.

(iii) Any quotient of an abelian variety having good reduction has good reduction.

Here, we're going to show that the second condition ("$A$ has completely multiplicative reduction at $N$") also comes for free. The proof is similar to the above, but a bit more complicated. We proceed as follows:

(i) The special fiber of the minimal regular model of $X_0(N)$ at $N$ is a nodal curve whose irreducible components are $\mathbb{P}^1$.

(ii) The Jacobian of such a curve has completely toric reduction.

(iii) Any quotient of an abelian variety having completely toric reduction has completely toric reduction.

The first step will consume most of our time. For the second, we appeal to a theorem of Raynaud relating Néron and minimal regular models. The third step follows easily from the theory of Néron models.

### 2.4.5 Completely toric reduction for quotients

**Proposition.** Let $\mathcal{O}$ be a DVR with fraction field $K$ and residue field $k$. Let $A/K$ be an abelian variety and let $B$ be a quotient of $A$. Suppose $A$ has completely toric reduction. Then the same is true for $B$.

*Proof.* Let $f\colon A \to B$ be the quotient map. Since the isogeny category is semi-simple, there exists a map $g\colon B \to A$ such that $fg = [n]$ for some positive integer $n$. The maps $f$ and $g$ extend to maps of the Néron models $\mathcal{A}$ and $\mathcal{B}$ by the Néron mapping property. These extended maps still satisfy $fg = [n]$, since it holds generically. It follows that $f$ and $g$ induce maps between $\mathcal{A}_k$ and $\mathcal{B}_k$ satisfying $fg = [n]$. In particular, $f\colon \mathcal{A}_k \to \mathcal{B}_k$ is surjective, which shows that $\mathcal{B}_k$ is a torus. □

### 2.4.6 Raynaud's theorem on the relative Picard functor

Let $f\colon X \to S$ be a proper flat map. We define the relative Picard functor, denoted $\mathrm{Pic}_{X/S}$, to be the sheafification of the functor $S' \mapsto \mathrm{Pic}(X_{S'})$ on the big fppf site of $S$. A lot is known about this functor, but we'll only mention the few results we need. We refer to Raynaud's article (MR0282993) as a reference.

To begin with, we have the following result of Murre:

**Theorem.** If $S$ is a field then $\mathrm{Pic}_{X/S}$ is representable by a group scheme. (Note: nothing about the singularities of $X$ is assumed.)

Thus, when $S$ is a field, we have a component group $\mathrm{Pic}^0_{X/S}$, which we can think of as a subsheaf of $\mathrm{Pic}_{X/S}$. For a general base $S$, we define $\mathrm{Pic}^0_{X/S}$ to be the subsheaf of $\mathrm{Pic}_{X/S}$ consisting of those sections that restrict into $\mathrm{Pic}^0_{X_s/s}$ for every geometric point $s \to S$.

Suppose now that $S = \mathrm{Spec}(\mathcal{O})$ where $\mathcal{O}$ is a DVR. Let $K$ be the fraction field of $\mathcal{O}$ and $k$ the residue field. Suppose also that $X$ is a curve (i.e., its fibers are pure of dimension 1). Let $\{X_i\}$ be the irreducible components of the special fiber $X_k$. The local ring of $X_i$ at its generic point is artinnian; let $d_i$ be its length. This is the multiplicity of $X_i$ in $X$.

**Theorem.** Suppose that $X_K$ is smooth over $K$, $X$ is regular, and the gcd of the $d_i$ is 1. Let $\mathcal{J}$ be the Néron model of $\mathrm{Jac}(X_K)$ over $\mathcal{O}$ and let $\mathcal{J}^0$ be its identity component. Then $\mathrm{Pic}^0_{X/S}$ is representable by smooth group scheme over $\mathcal{O}$, and coincides with $\mathcal{J}^0$. In particular, $\mathcal{J}^0_k$ is isomorphic to $\mathrm{Pic}^0_{X_k/k}$.

**Remark.** The functor $\mathrm{Pic}_{X/S}$ is not necessarily representable by a scheme, but it is represented by an algebraic space. Let $E$ be the scheme-theoretic closure of the identity section of $\mathrm{Pic}_{X/S}$. If $\mathrm{Pic}_{X/S}$ were separated, this would simply be the identity section, but $\mathrm{Pic}_{X/S}$ can fail to be separated. The quotient sheaf $\mathrm{Pic}_{X/S}/E$ is representable by a separated and smooth group scheme over $\mathcal{O}$. It admits a degree function $\mathrm{Pic}_{X/S}/E \to \mathbb{Z}$, the kernel of which is the full Néron model $\mathcal{J}$ of $\mathrm{Jac}(X_K)$.

### 2.4.7 The minimal regular model of $X_0(N)$

Given Raynaud's theorem, to understand the Néron model of $J_0(N)$ we should first understand the minimal regular model of $X_0(N)$. One might guess that $\overline{M}_0(N)$, the coarse space of $\overline{\mathfrak{M}}_0(N)$, would be the minimal regular model. This is almost the case, but not quite: the automorphisms groups in $\overline{\mathfrak{M}}_0(N)$ cause its coarse space to be non-regular. However, the singularities are very mild and easy to resolve.

To study $\mathfrak{M}_0(N)$ and its coarse space, we first pass to a finite étale Galois cover which is a scheme, see what goes on there, and then take the quotient to obtain the coarse space.

**The covering space and its structure**

We now change notation and use $p$ in place of $N$. We assume that $p$ is a prime $> 3$. Let $\ell$ be a prime satisfying the following: (1) $\ell \neq p$; (2) $\ell > 2$; and (3) $\ell \neq \pm 1$ modulo $p$. Let $G = \mathrm{GL}_2(\mathbb{F}_\ell)$. The order of $G$ is $(\ell^2-1)(\ell^2-\ell) = \ell(\ell-1)^2(\ell+1)$, and is therefore prime to $p$. We work over $\mathbb{Z}[1/6\ell]$ in this section (we really only care about what goes on at $p$).
We will be concerned with the following moduli spaces:

(i) $\mathfrak{M}_0(p)$ and its coarse space $M_0(p)$.

(ii) The moduli space $\mathfrak{M}_0(p;\ell)$ of elliptic curves with $\Gamma_0(p)$- and $\Gamma(\ell)$-structure. This is a scheme since $\ell > 2$, so we denote it by $M_0(p;\ell)$.

(iii) The moduli space $M(\ell)$ of elliptic curves with $\Gamma(\ell)$-structure, which is smooth over $\mathbb{Z}[1/\ell]$.

We have a natural map $M_0(p;\ell) \to \mathfrak{M}_0(p)$ which is finite étale and Galois with group $G$. We have a natural identification $\mathfrak{M}_0(p) = [M_0(p;\ell)/G]$ and $M_0(p) = M_0(p;\ell)/G$. Note that $M_0(p;\ell)$ is affine. If we let $A$ be its coordinate ring then $M_0(p) = \mathrm{Spec}(A^G)$.

We need the following result, a proof of which can be found in Katz–Mazur:

**Theorem.** The scheme $M_0(p;\ell)$ is regular and flat over $\mathbb{Z}$.

**Proposition.** The scheme $M_0(p;\ell)_{\mathbb{F}_p}$ is Cohen–Macaulay and reduced. It is smooth away from the supersingular points. Each supersingular point is an ordinary node

**The structure of $M_0(p)$**

Recall that $M_0(p) = M_0(p;\ell)/G$, and that, if $M_0(p;\ell) = \mathrm{Spec}(A)$ then $M_0(p) = \mathrm{Spec}(A^G)$. Since $G$ is prime to $p$, formation of $G$ invariants commutes with reduction mod $p$. In particular, formation of the coarse space of $\mathfrak{M}_0(p)$ commutes with reduction mod $p$.

The element $-1 \in G$ acts trivially on $M_0(p;\ell)$. Let $\overline{G} = G/\{\pm 1\}$. Then $M_0(p) = M_0(p;\ell)/\overline{G}$. Let $x$ be a point in $M_0(p)$ in characteristic $p$ with automorphism group $H$. The group $\overline{G}$ transitively permutes the points of $M_0(p;\ell)$ above $x$, and the stabilizer of any point is a subgroup $\overline{G}$ isomorphic to $\overline{H} = H/\{\pm 1\}$. It follows that the strict completion $R$ of the local ring at $x$ is isomorphic to the $\overline{H}$-invariants of the strict completion of the local ring $S$ at any point $y$ over $x$. Now, since $p?3$, we have the following:

(i) If $j(x) \neq 0, 1728$ then $\overline{H}$ is trivial.

(ii) If $j(x) = 1728$ then $\overline{H} = \mathbb{Z}/2\mathbb{Z}$.

(iii) If $j(x) = 0$ then $\overline{H} = \mathbb{Z}/3\mathbb{Z}$.

Thus if $j(x) \neq 0, 1728$ then $R = S$ and $x$ is a regular point. Note that if $j(x)$ is 0 or 1728 then, since $\overline{H}$ does not fix a point in a neighborhood of $x$, it acts non-trivially on $S$. It follows that, for an appropriate choice of $u$, $v$, the generator of $\overline{H}$ acts by $u \mapsto \zeta u$, $v \mapsto \zeta^{-1} v$, where $\zeta$ is a primitive $k$th, where $k = \#\overline{H}$. It follows that $R = S^{\overline{H}}$ is generated by $U = u^k$, $V = v^k$, and $uv = p$. We have thus shown that following:

**Theorem.** Let $x$ be a characteristic $p$ point of $M_0(p)$ and let $R$ be the strict complete local ring at $x$.

(i) If $x$ is not supersingular, then $M_0(p)$ is smooth at $x$.

**Remark.** It is also true that the cuspidal points of $\overline{M}_0(p)$ are smooth, since they are smooth in the special fiber.

**The minimal regular model**

The scheme $\overline{M}_0(p)$ is a flat proper model of its generic fiber which is regular except at possibly two points. The singularities of these points can be resolved with one or two blow-ups. The result is that an additional $\mathbb{P}^1$ is added at $j = 1728$ if that point is supersingular, and two additional $\mathbb{P}^1$'s are added in a chain at $j = 0$ if that point is supersingular. The resulting model is minimal. We thus have:

**Proposition.** Let $C$ be the special fiber of the minimal regular model of $\overline{M}_0(N)$. Then $C$ is reduced curve, all of its components are $\mathbb{P}^1$'s, and all of its singularities are simple nodes.

### 2.4.8 The special fiber of the Néron model

**Proposition.** Let $C$ be a curve over an algebraically closed field $k$ with the following properties: $C$ is reduced, all of its components are $\mathbb{P}^1$'s, and all of its singularities are simple nodes. Then $\mathrm{Pic}^0_{C/k}$ is a torus.

*Proof.* Let $\Gamma$ be the graph corresponding to $C$: its vertices are the irreducible components of $C$, and there is one edge between two components at each point they touch. Given a line bundle on $C$, we get a line bundle on each component, and an identification of the fibers at the touching points. Every line bundle on $\mathbb{P}^1$ is of the form $\mathcal{O}(n)$. Thus if we assign to each vertex of $\Gamma$ an integer and to each edge an element of $\mathbb{G}_m$ then we can build a line bundle on $C$, and all line bundles are of this form. We have thus produced a surjection from a torus to $\mathrm{Pic}^0_{C/k}$, which proves the proposition. □

Actually, we can say a bit more. Suppose we have data as above defining some line bundle. For the bundle to be trivial, it must be trivial on each component, and so the integers at each node must be 0. The non-vanishing sections of the bundle on one of the components is given by $\mathbb{G}_m$. Given sections on each component (i.e., elements of $\mathbb{G}_m$ at each node), they glue if and only if at each edge the quotient of their values is equal to the value of the edge. In other words, the original data defines the trivial bundle if and only if the integers are 0 and the values on the edges are a 1-coboundary.

We thus see that the identity component of $\mathrm{Pic}_{C/k}$ is $\mathrm{H}^1(\Gamma, \mathbb{G}_m)$. This is a torus with character lattice $\mathrm{H}_1(\Gamma, \mathbb{Z})$.

**Theorem.** $J_0(N)$ has completely toric reduction at $N$.

*Proof.* This follows from the above computation, Raynaud's theorem, and the form established for the minimal regular model of $X_0(N)$ at $N$. □

### 2.4.9 Injectivity of the reduction map on torsion

To end this section, I want to give a proof of the following theorem. This result was crucial to previous section, and it was pointed out to me that we had not yet given a proof in all cases.

**Theorem.** Let $K/\mathbb{Q}_p$ be a finite extension with ramification index $< p - 1$. Let $\mathcal{O}$ be the ring of integers of $K$ and $k$ the residue field of $\mathcal{O}$. Let $A/K$ be an abelian variety, and let $\mathcal{A}/\mathcal{O}$ be its Néron model. Then the reduction map $\mathcal{A}(\mathcal{O})_{\mathrm{tors}} \to \mathcal{A}(k)_{\mathrm{tors}}$ is injective.

*Proof.* Let $G_0 = A(K)_{\text{tors}}$, regarded as a closed subscheme of $A(K)$. Note that as a group scheme, $G_0$ is constant. Let $G$ be the scheme-theoretic closure of $G_0$ in $\mathcal{A}$. Then $G$ is a flat group scheme over $\mathcal{O}$. And it is finite: every field-valued point of $G$ is defined over $K$, and every $K$-point of $G$ extends to an $\mathcal{O}$-point of $G$ by the Néron mapping property. Thus $G$ is proper, and therefore finite (since we know it to be quasi-finite). Since $G_0$ obviously extends to a constant group scheme over $\mathcal{O}$, Raynaud's theorem implies that $G$ itself is a constant group scheme. The theorem follows, since the reduction map $G(\mathcal{O}) \to G(k)$ is clearly injective for constant groups. □

# 3 Proof of Mazur's theorem

## 3.1 The Eisentein ideal and Eisenstein quotient of $J_0(N)$

In this section, we begin the proof of Mazur's theorem: if $N$ is a prime greater than 7 and not 13 then no elliptic curve over $\mathbb{Q}$ has a rational point of order $N$. We begin by analyzing $[0]-[\infty]$ as a point on $J_0(N)$. We show that it is a non-trivial torsion point of order dividing $N-1$ and compute the Hecke action on it. We then prove the theorem under the assumption that all eigenforms have rational coefficients. This hypothesis allows us to apply our criteria directly to quickly prove the theorem.

### 3.1.1 Our goal

Our purpose in the next two sections is to prove the following result, due to Mazur.

**Theorem.** Let $N$ be a prime greater than 7 and not 13. Then no elliptic curve over $\mathbb{Q}$ has a rational point of order $N$.

Recall that it is enough to construct a quotient $A$ of $J_0(N)$ such that $A(\mathbb{Q})$ has rank 0 and $0 \neq \infty$ in $A$. Furthermore, rank 0 is ensured if the Jordan–Holder constituents of $A[p](\overline{\mathbb{Q}})$ are trivial and cyclotomic.

Throughout this section, $N$ is as in the theorem. The prime 13 is excluded because $X_0(13)$ has genus 0, and therefore $J_0(N)$ is trivial. For the $N$ allowed by the theorem, $X_0(N)$ has positive genus.

### 3.1.2 The difference of the cusps

**Order on the Jacobian**

**Proposition.** The point $[0]-[\infty]$ of $J_0(N)$ is a non-trivial torsion point of order dividing $N-1$.

*Proof.* If $[0]-[\infty]=0$ in $J_0(N)$ then there would be a function $f$ on $X_0(N)$ with $\mathrm{div}(f)=[0]-[\infty]$. Such an $f$ would provide a degree 1 map to $\mathbb{P}^1$, and so $X_0(N)$ would have genus 0. But this is not the case for the $N$ under consideration.

Consider the modular form $\Delta(z)$ of weight 12 for $\Gamma(1)$ on the upper half-plane. A basic fact about this form is that it is nowhere vanishing; this can be seen either from the product formula for it, or in terms of its description as the discriminant. Its $q$-expansion begins $q+\cdots$. Now, $\Delta(z)$ and $\Delta(Nz)$ are both modular forms of weight 12 for $\Gamma_0(N)$, and both are non-vanishing on the upper half-plane. Thus $f(z)=\Delta(z)/\Delta(Nz)$ is a nowhere vanishing function on the upper half-plane which is invariant under $\Gamma_0(N)$. It therefore descends to a meromorphic function on $X_0(N)$ which is holomorphic and non-vanishing on $Y_0(N)$. The $q$-expansion of $f$ at $\infty$ begins $q^{-(N-1)}+\cdots$. It follows that $f$ has a pole of order $N-1$ at $\infty$, as a function on $X_0(N)$. (The function $q$ on $\mathfrak{h}$ descends to a local parameter at $\infty$ on $X_0(N)$.) Since the only other zero or pole of $f$ occurs at 0, and the divisor of $f$ has degree 0, we necessarily have $\mathrm{div}(f)=(N-1)[0]-(N-1)[\infty]$, which shows that $[0]-[\infty]$ is $(N-1)$-torsion. □

**Remark.** In fact, Ogg showed that the exact order of $[0]-[\infty]$ is $(N-1)/\gcd(N-1,12)$, but we will not need this statement.

**Remark.** Mazur proves that $[0]-[\infty]$ generates the entire torsion subgroup of the Mordell–Weil group of $J_0(N)$.

**Action of Hecke operators**

Recall that for primes $\ell \neq N$, we have the Hecke operator $T_\ell$, which we can regard as an endomorphism of $J_0(N)$.

**Proposition.** We have $T_\ell([0]-[\infty]) = (\ell+1)([0]-[\infty])$.

*Proof.* Consider the Hecke correspondence $f, g\colon X_0(N\ell) \rightrightarrows X_0(N)$. We have the following facts:

(i) The curve $X_0(N\ell)$ has 4 cusps. In fact, the set of cusps for $X_0(N\ell)$ is the product of the sets of cusps for $X_0(N)$ and $X_0(\ell)$. We can therefore represent its cusps as pairs $(x, y)$ with $x, y \in \{0, \infty\}$. The first coordinate is the $X_0(N)$ coordinate.

(ii) We have $f(x, y) = g(x, y) = x$. This can be seen as follows. The maps $f$ and $g$ lift to the identity map and multiplication by $\ell$ on $\mathfrak{h}^*$. The elements of $\mathbb{P}^1(\mathbb{Q})$ with $N$ in the denominator map to the cusp $\infty$ of $X_0(N)$, while all others map to 0. Since multiplication by $\ell$ cannot introduce a $N$ in the denominator, the statement follows.

(iii) The map $f$ has ramification index $\ell$ at $(*, 0)$ and index 1 at $(*, \infty)$. The map $g$ is the opposite. This is a simple calculation with stabilizer groups in $\Gamma_0(N)$ and $\Gamma_0(N\ell)$.

We thus see that $f^*([x]) = \ell[(x, 0)] + [(x, \infty)]$. Applying $g_*$, we find $g_*(f^*([x])) = (\ell+1)[x]$. □

### 3.1.3 The case of rational eigenforms

#### The quotient of the Jacobian

Say an abelian variety $A/\mathbb{Q}$ satisfies condition JH($p$) if the Jordan–Holder constituents of $A[p](\overline{\mathbb{Q}})$ are all trivial or cyclotomic. Note that this condition is isogeny invariant: it is equivalent to asking that the semi-simplified reduction of the rational $p$-adic Tate module of $A$ is a direct sum of trivial and cyclotomic characters.

We have shown that, up to isogeny, we have a decomposition $J_0(N) = \prod A_f$, where the product is over the (Galois orbits of) normalized weight 2 cuspidal eigenforms $f$. Let's suppose for simplicity that each $f$ has rational coefficient field, so the $A_f$'s are elliptic curves (and the Galois orbits are singletons). Given $p$, there is a maximal quotient (up to isogeny) of $J_0(N)$ that satisfies JH($p$), namely, the product of the $A_f$'s that do.

We want to more precisely define this quotient. For an eigenform $f$, let $\mathfrak{f}_f$ be the kernel of the homomorphism $\mathbb{T} \to \mathbb{Z}$ giving the eigenvalues of $f$. Recall that $A_f$ is by definition $J_0(N)/\mathfrak{f}_f J_0(N)$. Let $S$ be the set of those $f$ for which $A_f$ satisfies JH($p$), and let $I = \bigcap_{f \in S} \mathfrak{f}_f$. We define $A = J_0(N)/IJ_0(N)$. Up to isogeny, $A = \prod_{f \in S} A_f$, and so $A$ satisfies JH($p$). From our criterion for rank 0, we find:

**Proposition.** $A(\mathbb{Q})$ has rank 0.

How do we know that there will be a prime $p$ for which $A$ is actually non-trivial? In fact, the analysis of $[0]-[\infty]$ provides such a prime. Let $p$ be a prime dividing the order of $[0]-[\infty]$. Then $J_0(N)(\mathbb{Q})$ has a $p$-torsion point, and so $J_0(N)[p]$ has a copy of the trivial representation in it. This copy must come from one of the $A_f$'s, and this $A_f$ must satisfy JH($p$). From now on, we fix such a prime $p$.

#### Alternate characterization of I

To actually work with the quotient $A$, we will need to better understand the ideal $I$. We begin with the following observation:

**Lemma.** $f \in S$ if and only if $a_\ell(f) - (\ell+1)$ is divisible by $p$ for all $\ell$.

*Proof.* Suppose $f \in S$, so that $A_f$ satisfies JH($p$). Then the semi-simplification of $A_f[p]$ is isomorphic to trivial plus cyclotomic, as it is a 2-dimensional representation with cyclotomic determinant. It follows that the trace of the Frobenius at $\ell$ on $A_f[p]$, which is equal to $a_\ell(f)$ mod $p$, is $\ell+1$. Conversely, if all the $a_\ell(f)$ are equal to $\ell+1$ modulo $p$, then $A_f$ is (up to semi-simplification) the sum of trivial and cyclotomic, and thus satisfies JH($p$). □

Define the $p$-Eisenstein ideal $\mathfrak{a}$ to be the ideal of $\mathbb{T}$ generated by $p$ and the $T_\ell - (\ell+1)$.

**Lemma.** We have $\mathbb{T}/\mathfrak{a} = \mathbb{F}_p$. The ideal $\mathfrak{a}$ is maximal.

*Proof.* By assumption, there exists a form $f$ in $S$, and for such a form the image of $T_\ell - (\ell+1)$ in $\mathbb{T}/\mathfrak{f}_f \cong \mathbb{Z}$ is divisible by $p$ for all $\ell$. Thus the image of $\mathfrak{a}$ in $\mathbb{T}/\mathfrak{f}_f$ is not the unit ideal, and so $\mathfrak{a}$ is not the unit ideal. It is clear now that $\mathbb{T}/\mathfrak{a} = \mathbb{F}_p$, since the quotient is non-trivial, and every Hecke operator is identified with an integer. □

**Lemma.** The following are equivalent: (1) $f \in S$; (2) the image of $\mathfrak{a}$ in $\mathbb{T}/\mathfrak{f}_f$ is not the unit ideal; (3) $\mathfrak{f}_f \subset \mathfrak{a}$.

*Proof.* The image of $\mathfrak{a}$ in $bT/\mathfrak{f}_f \cong \mathbb{Z}$ is the ideal generated by $p$ and the $a_\ell(f) - (\ell+1)$. This is not the unit ideal if and only if $a_\ell(f) - (\ell+1)$ is divisible by $p$ for all $p$. Thus (1) and (2) are equivalent. Clearly, (3) implies (2). Conversely, if $\mathfrak{f}_f \not\subset \mathfrak{a}$ then $\mathfrak{f}_f + \mathfrak{a} = (1)$, since $\mathfrak{a}$ is maximal, and so the image of $\mathfrak{a}$ in $\mathbb{T}/\mathfrak{f}_f$ is the unit ideal. □

**Proposition.** *$I$ is the intersection of the minimal primes $\mathfrak{f}$ of $\mathbb{T}$ which are contained in $\mathfrak{a}$.*

*Proof.* The minimal primes of $\mathbb{T}$ are exactly the $\mathfrak{f}_f$. By definition, $I$ is the intersection of the $\mathfrak{f}_f$ with $f \in S$. By the previous lemma, this coincides with the $\mathfrak{f}_f$ which are contained in $\mathfrak{a}$. □

### The image of the cusps

**Lemma.** Let $R$ be a reduced noetherian ring, let $\mathfrak{a}$ be a maximal ideal, and let $I$ be the intersection of the minimal primes contained in $\mathfrak{a}$. Then $I_\mathfrak{a} = 0$.

*Proof.* The local ring $R_\mathfrak{a}$ is reduced, and its nilradical is $I_\mathfrak{a}$. □

**Corollary.** We have $I_\mathfrak{a} = 0$ for our ideals $\mathfrak{a}$ and $I$.

Suppose $X$ is a $\mathbb{T}$-module in which all elements are killed by a power of $p$. Then the action of $\mathbb{T}$ extends to one of the $p$-adic completion $\mathbb{T}_p = \varprojlim \mathbb{T}/p^n\mathbb{T}$. Since this is a complete semi-local ring, it is a product of local rings, the factors corresponding to the maximal ideals. In particular, the localization $\mathbb{T}_\mathfrak{a}$ is a direct factor of $\mathbb{T}_p$. It follows that $X$ decomposes as $X_\mathfrak{a} \oplus X'$, where $\mathbb{T}_\mathfrak{a}$ acts by zero on $X'$. We can identify $X_\mathfrak{a}$ with $X[\mathfrak{a}^\infty] = \bigcup_{n \ge 0} X[\mathfrak{a}^n]$, where $X[\mathfrak{a}^n]$ is the $\mathfrak{a}^n$-torsion in $X$.

**Lemma.** The map $J_0(N)[\mathfrak{a}^\infty] \to A[\mathfrak{a}^\infty]$ is an isomorphism.

*Proof.* Let $X = J_0(N)[p^\infty]$ and $Y = A[p^\infty]$. We then have a surjection of $\mathbb{T}_p$-modules $X \to Y$. The kernel of this map is $X \cap IJ_0(N)$, which is $IX$. (Let $t_1, \ldots, t_n$ generate $I$. Then $IJ_0(N)$ is the image of the map $J_0(N)^n \to J_0(N)$ given by the $t_i$. Any $p$-power torsion element in the image comes from one in the source.) We thus have an exact sequence

$$0 \to IX \to X \to Y \to 0$$

Now localize at $\mathfrak{a}$. On the one hand, this is an exact operation. On the other, $(IX)_\mathfrak{a} = I_\mathfrak{a} X_\mathfrak{a} = 0$, since $I_\mathfrak{a} = 0$. Thus the map $X_\mathfrak{a} \to Y_\mathfrak{a}$ is an isomorphism. □

**Proposition.** We have $[0] \neq [\infty]$ in $A$.

*Proof.* Let $P = [0] - [\infty]$, as a point of $J_0(N)$. Let $Q$ be a multiple of $P$ which is non-zero and $p$-torsion. Then $Q$ is killed by $\mathfrak{a}$. Indeed, we showed above that $T_\ell P = (\ell+1)P$, and so the same holds for $Q$. Thus $Q \in J_0(N)[\mathfrak{a}^\infty]$, and so its image in $A[\mathfrak{a}^\infty]$ is non-zero. □

### 3.1.4 The general case

**The problem**

In general, we have shown that we have a decomposition $V_pJ_0(N) = \prod_{f,\lambda} V_{f,\lambda}$, where the product is over pairs $(f, \lambda)$ consisting of a normalized weight 2 eigenform $f$ and a place $\lambda$ of its coefficient field $K_f$ above $p$, and $V_{f,\lambda}$ is 2-dimensional Galois representation over the field $K_{f,\lambda}$. We would like to take $A$ to be the quotient of $J_0(N)$ for which $T_pA$ is the product of those $V_{f,\lambda}$ which reduce mod $p$ to trivial plus cyclotomic. However, there might not be such a quotient. For instance, it could be that there is only one $f$, and that for some $\lambda$ the representation $V_{f,\lambda}$ has the right form, and for other $\lambda$ it does not.

What we'll do is take $A$ to be the quotient of $J_0(N)$ which is isogenous to the product of $A_f$'s over $f$'s for which $V_{f,\lambda}$ is of the right form for some $\lambda$. Since this might not hold for all $\lambda$, the abelian variety $A$ will not necessarily satisfy JH($p$), and so our criterion for rank 0 will not apply directly. However, one can make sense of the piece of the Mordell–Weil group of $A$ corresponding to the "good" $\lambda$, and the proof of Theorem B will show that this is finite. Then a simple commutative algebra result will allow us to deduce that $A(\mathbb{Q})$ is finite.

**The quotient A**

As before, choose a prime $p$ dividing the order of $[0] - [\infty]$ in $J_0(N)$, and define $\mathfrak{a}$ to be the ideal of $\mathbb{T}$ generated by $p$ and $T_\ell - (\ell + 1)$ for all $\ell \neq N$.

**Lemma.** We have $\mathbb{T}/\mathfrak{a} = \mathbb{F}_p$. The ideal $\mathfrak{a}$ is maximal.

*Proof.* The representation $J_0(N)[p]$ contains a copy of the trivial representation. Thus for some $(f, \lambda)$, the semi-simplified reduction of $V_{f,\lambda}$ contains the trivial representation, and is therefore the sum of trivial and cyclotomic. It follows that $a_\ell(f) = \ell + 1$ holds modulo $\lambda$, for all $\ell$. Thus the image of $\mathfrak{a}$ in $\mathbb{T}/\mathfrak{f}_f$ is contained in $\lambda$. It follows that $\mathfrak{a}$ is not the unit ideal, and so $\mathbb{T}/\mathfrak{a} = \mathbb{F}_p$, as before. □

Let $I$ be the intersection of the minimal primes of $\mathbb{T}$ contained in $\mathfrak{a}$, and let $A = J_0(N)/IJ_0(N)$. The following is proved exactly as before.

**Proposition.** We have $[0] \neq [\infty]$ in $A$.

## 3.2 The special fiber at $N$ of $J_0(N)$

In this section, we finish the proof Mazur's theorem. In the previous section, we constructed a quotient of $A$ of $J_0(N)$, and showed that to prove the theorem it was enough to establish that $A(\mathbb{Q})$ has rank 0. Unfortunately, our criterion for rank 0 does not apply directly to $A$. However, the proof of the criterion almost applies, and with a little extra work we show what we want.

### 3.2.1 Review

**From the previous section**

We are in the process of proving Mazur's theorem:

**Theorem.** Let $N$ be a prime $> 7$ and not 13. Then no elliptic curve over $\mathbb{Q}$ has a rational point of order $N$.

Let's recall where we left off. We picked a prime $p$ dividing the order of $[0] - [\infty]$ in $J_0(N)$, and defined $\mathfrak{a}$ to be the $p$-Eisenstein ideal, that is, the ideal of $\mathbb{T}$ generated by $p$ and $T_\ell - (\ell + 1)$ for all $\ell$. We showed that $\mathfrak{a}$ is not the unit ideal of $\mathbb{T}$, and thus a maximal ideal with $\mathbb{T}/\mathfrak{a} = \mathbb{F}_p$. We defined $I$ to be the intersection of the minimal primes of $\mathbb{T}$ contained in $\mathfrak{a}$, and $A$ to be the quotient $J_0(N)/IJ_0(N)$. To complete the proof, we need to show that $A(\mathbb{Q})$ has rank 0. That is what we'll do in this section.

**From the criterion for rank** 0

To show that $A(\mathbb{Q})$ has rank 0, we're going to adapt the proof of Theorem B from [PP69BM]. We now briefly recall how that went.

A $p$-power order group scheme over $\mathbb{Z}$ or $\mathbb{Z}[1/N]$ is called admissible if: (1) it's finite flat away from $N$; (2) it's quasi-finite étale away from $p$; and (3) over $\mathbb{Z}[1/N]$, admits a filtration with quotients $\mu_p$ or $\mathbb{Z}/p\mathbb{Z}$. We showed that a group scheme $G$ satisfying (1) and (2) is admissible if and only if the Galois module $G(\overline{\mathbb{Q}})$ satisfies JH($p$), i.e., it admits a filtration whose successive quotients are trivial or cyclotomic.

Let $G$ be an admissible group over $\mathbb{Z}$. We defined several numerical invariants:

(i) $\ell(G) = \log_p(\#G)$. (We'll use this notation for any group of $p$-power order.)

(ii) $\alpha(G)$ is the number of $\mathbb{Z}/p\mathbb{Z}$'s occurring in $G$.

(iii) $\delta(G) = \ell(G_{\mathbb{Q}}) - \ell(G_{\mathbb{F}_N})$.

(iv) $h^i(G)$ is the length of $\mathrm{H}^i_{\mathrm{fppf}}(\mathrm{Spec}(\mathbb{Z}), G)$.

We proved the following inequality:

$$h^1(G) - h^0(G) \leq \delta(G) - \alpha(G)$$

Suppose now that $A/\mathbb{Q}$ is an abelian variety satisfying the hypotheses of Theorem B, i.e., $A$ has good reduction away from $N$, completely toric reduction at $N$, and $A[p](\overline{\mathbb{Q}})$ satisfies JH($p$). Let $\mathcal{A}/\mathbb{Z}$ be the Néron model of $A$. Then $\mathcal{A}[p^n]$ is admissible. Using the hypotheses, we computed $\alpha(\mathcal{A}[p^n])$ and $\delta(\mathcal{A}[p^n])$, and found them to be approximately equal, from which we concluded that $h^1(\mathcal{A}[p^n]) - h^0(\mathcal{A}[p^n])$ is bounded as $n \to \infty$. Since $h^0(\mathcal{A}[p^n])$ is bounded (by the Mordell–Weil theorem), it follows that $h^1(\mathcal{A}[p^n])$ is bounded. By Kummer theory, $A(\mathbb{Q})$ injects into the inverse limit of $\mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}, \mathcal{A}[p^n])$, which shows that $A(\mathbb{Q})$ is finite. (Actually, one must use $\mathcal{A}^\circ$ here, but that does not cause a problem.)

### 3.2.2 Proof of rank 0

**Notation**

The quantities $N$, $p$, $\mathfrak{a}$, $I$, and $A$ are as above. We write $d$ for the $\mathbb{Z}_p$-rank of $\hat{\mathbb{T}}_{\mathfrak{a}}$ and $e$ for the idempotent of $\hat{\mathbb{T}}_p$ which projects onto $\hat{\mathbb{T}}_{\mathfrak{a}}$. We let $\mathcal{A}$ for the Néron model of $A$ over $\mathbb{Z}$.

**Admissibility**

**Proposition.** $\mathcal{A}[p^n]_{\mathfrak{a}}$ is admissible.

By our previous work, it is enough to show the following:

**Lemma.** $A[p^n]_{\mathfrak{a}}$ satisfies JH($p$).

*Proof.* We have a containment $A[p^n]_{\mathfrak{a}} \subset A[\mathfrak{a}^m]$ for some $m$. Furthermore, we have an exact sequence

$$0 \to A[\mathfrak{a}] \to A[\mathfrak{a}^m] \to A[\mathfrak{a}^{m-1}]$$

It thus suffices (by induction) to show that $A[\mathfrak{a}]$ satisfies JH($p$). Let $V = A[\mathfrak{a}]$, regarded as an $\mathbb{F}_p$-representation of the Galois group. Since $T_\ell$ acts by $\ell + 1$ on $V$, Eichler–Shimura implies that the Frobenius $F_\ell$ satisfies $F_\ell^2 - (\ell+1)F_\ell + \ell = 0$ on $V$, for all $\ell \neq p, N$. We have a factorization $T^2 - (\ell+1)T + \ell = (T - \ell)(T - 1)$, and so all generalized eigenvalues of $F_\ell$ on $V$ are 1 or $\ell$. The result now follows from the following general lemma. □

**Lemma.** Let $V$ be a finite dimensional representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ over $\mathbb{F}_p$ such that 1 and $\ell$ are the only generalized eigenvalues of $F_\ell$, for all but finitely many $\ell$. Then $V$ satisfies $\mathrm{JH}(p)$.

*Proof.* Let $W = V \oplus V^*(1)$, where (1) denotes tensoring with cyclotomic. If $m$ is the matrix for $F_\ell$ on $V$, then the matrix for $F_\ell$ on $V^*(1)$ is $\ell \cdot {}^t m^{-1}$. In particular, the dimension of the 1-generalized eigenspace of $F_\ell$ on $V$ is equal to the dimension of the $\ell$-generalized eigenspace of $F_\ell$ on $V^*(1)$. It follows that the dimension of the 1-generalized eigenspace of $F_\ell$ on $W$ is $n = \dim(V)$, and similarly for the $\ell$-generalized eigenspace. Thus the character of $W$ is the same as that of $n$ copies of the trivial plus $n$ copies of cyclotomic. It follows that every constituent of $W$ is trivial or cyclotomic, and so the same holds for $V$. □

**Computation of $\alpha$**

**Lemma.** Let $G$ be a $p$-divisible group with an action of $\hat{\mathbb{T}}_{\mathfrak{a}}$ such that the rational Tate module $V_p(G)$ is free of rank 2 as a $\hat{\mathbb{T}}_{\mathfrak{a}}[1/p]$-module. Then $\ell(G[p^n]) = 2nd + O(1)$.

*Proof.* Let $T$ be the integral Tate module of $G$. Since $T[1/p]$ is free of rank 2 over $\hat{\mathbb{T}}_{\mathfrak{a}}[1/p]$, it follows that $T$ contains a free $\hat{\mathbb{T}}_{\mathfrak{a}}$-module $T'$ of rank 2 with finite index. We have $\ell(G[p^n]) = \mathrm{len}(T/p^nT)$. Of course, $\mathrm{len}(T/p^nT) = \mathrm{len}(T'/p^nT') + O(1)$, and $\mathrm{len}(T'/p^nT') = 2\mathrm{len}(\hat{\mathbb{T}}_{\mathfrak{a}}/p^n\hat{\mathbb{T}}_{\mathfrak{a}}) = 2nd$. □

**Proposition.** $\alpha(\mathcal{A}[p^n]_{\mathfrak{a}}) = nd + O(1)$.

*Proof.* (For notational simplicity suppose $p \neq 2$, so that $\alpha$ counts the number of trivial constituents of $A[p^n]$ as a Galois representation.) Recall that the Hecke operators $T_\ell$ are self-adjoint under the Weil pairing on $J_0(N)[p^n]$. It follows that every element of $\hat{\mathbb{T}}_p$ is also self-adjoint. Let $e$ be the idempotent of this ring that gives the projection onto the direct factor $\hat{\mathbb{T}}_{\mathfrak{a}}$. Then $A[p^n]_{\mathfrak{a}} = J_0(N)[p^n]_{\mathfrak{a}} = eJ_0(N)[p^n]$. Now, $eJ_0(N)[p^n]$ and $(1-e)J_0(N)[p^n]$ are orthogonal under the Weil pairing, since $e$ is a self-adjoint idempotent. Since the Weil pairing on $J_0(N)[p^n]$ is perfect, it follows that it restricts to a perfect pairing on $A[p^n]_{\mathfrak{a}}$. Thus $A[p^n]_{\mathfrak{a}}$ is Cartier self-dual. Since it is made up of $\mathbb{Z}/p\mathbb{Z}$'s and $\mu_p$'s, the self-duality implies that there must be the same number of $\mathbb{Z}/p\mathbb{Z}$'s as $\mu_p$'s, and so the number of $\mathbb{Z}/p\mathbb{Z}$'s is half $\ell(A[p^n]_{\mathfrak{a}})$. This is $nd$, up to a bounded factor. (Note that $A[p^n]_{\mathfrak{a}} = G[p^n]$, where $G$ is the $p$-divisible group $A[p^\infty]_{\mathfrak{a}} = eA[p^\infty]$, and $V_p(G)$ is free of rank 2 over $\hat{\mathbb{T}}_{\mathfrak{a}}[1/p]$.) □

**Computation of $\delta$**

**Lemma.** Let $G/\mathbb{Q}_N$ be a $p$-divisible group, let $V = V_p(G)$ be its rational Tate module, and let $\mathcal{G}_n/\mathbb{Z}_p$ be the maximal extension of $G[p^n]$ to an étale quasi-finite group (see Néron models). Then $\delta(\mathcal{G}_n) = (\dim(V) - \dim(V^I))n + O(1)$, where $I \subset \mathrm{Gal}(\overline{\mathbb{Q}}_N/\mathbb{Q}_N)$ is the inertia subgroup.

*Proof.* Let $T = T_p(G)$ be the integral Tate module of $G$. Then $\mathcal{G}_n(\overline{\mathbb{Q}}_N) = T/p^nT$, and so

$$\ell((\mathcal{G}_n)_{\mathbb{Q}_N}) = \mathrm{len}(T/p^nT) = n\dim(V)$$

By definition, $\mathcal{G}_n(\overline{\mathbb{F}}_N) = \mathcal{G}_n(\overline{\mathbb{Q}}_N)^I = (T/p^nT)^I$, and so

$$\ell((\mathcal{G}_n)_{\mathbb{F}_N}) = \mathrm{len}((T/p^nT)^I).$$

We have an exact sequence

$$0 \to T^I/p^nT^I \to (T/p^nT)^I \to \mathrm{H}^1(I,T)[p^n] \to 0.$$

Since $\mathrm{H}^1(I,T)$ is a finitely generated $\mathbb{Z}_p$-module, its $p^n$-torsion is bounded independent of $n$. Thus

$$\mathrm{len}((T/p^nT)^I) = \dim(T^I/p^nT^I) + O(1) = n\dim(V^I) + O(1).$$

The result follows. □

**Lemma.** Let $B/\mathbb{Q}_N$ be an abelian variety and let $\mathcal{B}/\mathbb{Z}_N$ be its Néron model. Then $V_p(\mathcal{B}_{\mathbb{F}_N}) = V_p(B)^I$, where $I$ is the inertia group.

*Proof.* Since $\mathcal{B}[p^n]$ is étale over $\mathbb{Z}_N$, every $p^n$-torsion point over $\overline{\mathbb{F}}_N$ lifts to one over $\mathbb{Z}_N^{\mathrm{un}}$. Furthermore, the Néron mapping property shows that every $\mathbb{Q}_N^{\mathrm{un}}$ point of $B$ extends to one of $\mathcal{B}$. We therefore have an isomorphism

$$\mathcal{B}[p^n](\overline{\mathbb{F}}_N) = \mathcal{B}[p^n](\mathbb{Q}_N^{\mathrm{un}}) = \mathcal{B}[p^n](\overline{\mathbb{Q}}_N)^I.$$

Taking inverse limits gives the result. □

**Lemma.** Let $B/\mathbb{Q}_N$ be an abelian variety with completely toric reduction and let $U$ be a summand of $V = V_p(B)$ as a Galois representation. Then $\dim(U^I) = \frac{1}{2}\dim(U)$.

*Proof.* We have $\dim(V) = 2\dim(B)$. Let $\mathcal{B}$ be Néron model of $B$. Then $V^I = V_p(\mathcal{B}_{\mathbb{F}_N})$. Since the identity component of $\mathcal{B}_{\mathbb{F}_N}$ is a torus of the same dimension as $B$, $V_p(\mathcal{B}_{\mathbb{F}_N})$ has dimension $\dim(B)$. Thus $\dim(V) = 2\dim(V^I)$.

Since $B$ has semi-stable reduction, Grothendieck's extension of Néron–Ogg–Shafarevich shows that $I$ acts unipotently on $V$. In fact, a stronger result is true: for any $g \in I$ we have $(g-1)^2 = 0$ on $V$. It follows that if $U$ is any Galois stable subspace of $V$ then $(g-1)^2 = 0$ on $U$ for any $g \in I$, and so $\dim(U^g) \geq \frac{1}{2}\dim(U)$. Since $I$ acts unipotently, its image is pro-$p$, and so the wild subgroup $I^w$ acts trivially. The group $I/I^w$ is pro-cyclic. If $g$ is a topologcal generator, then $U^I = U^g$. We thus see that $\dim(U^I) \geq \frac{1}{2}\dim(U)$.

Suppose now that $U$ is a summand of $V$, and write $V = U \oplus U'$. Then $V^I = U^I \oplus (U')^I$. By the above paragraph, $\dim(U^I) \geq \frac{1}{2}\dim(U)$ and $\dim((U')^I) \geq \frac{1}{2}\dim(U')$. Since $\dim(V^I) = \frac{1}{2}\dim(V)$, it follows that both inequalitities are equalities, which proves the result. □

**Remark.** In the case $B = J_0(N)$, which is the only case of interest to us, we know that $V$ decomposes into a sum of $V_{f,\lambda}$, which are each two dimensional. Thus the fact that $(g-1)^2 = 0$ for $g \in I$ follows from $g$ being unipotent in this case.

**Remark.** The above lemma shows that the inertia group at $N$ acts unipotently and non-trivially on $V_{f,\lambda}$. This is an instance of local–global compatibility in the Langlands program.

**Proposition.** $\delta(\mathcal{A}[p^n]_{\mathfrak{a}}) = nd + O(1)$.

*Proof.* Let $e \in \hat{\mathbb{T}}_p$ be the idempotent projecting onto $\hat{\mathbb{T}}_{\mathfrak{a}}$. Let $\mathcal{G}$ be the $p$-divisible group $\mathcal{A}[p^\infty]$ and let $V$ be the rational Tate module of $G = \mathcal{G}_{\mathbb{Q}}$. Then $\mathcal{G}_{\mathfrak{a}} = e\mathcal{G}$ is a $p$-divisible group and the rational Tate module of its generic fiber is $V_{\mathfrak{a}} = eV$. We have $\mathcal{G}_{\mathfrak{a}}[p^n] = \mathcal{A}[p^n]_{\mathfrak{a}}$, as both are just $\mathcal{A}[\mathfrak{a}^\infty] \cap \mathcal{A}[p^n]$. The group $\mathcal{A}[p^n]$ is the maximal quasi-finite étale extension of its generic fiber, by the Néron mapping property, and the same is true for $\mathcal{A}[p^n]_{\mathfrak{a}}$, since it is a summand. Thus $\delta(\mathcal{A}[p^n]_{\mathfrak{a}}) = n(\dim(V_{\mathfrak{a}}) - \dim(V_{\mathfrak{a}}^I)) + O(1)$. But $V_{\mathfrak{a}}$ is a summand of $V$ and $A$ has completely toric reduction, so $\dim(V_{\mathfrak{a}}^I) = \frac{1}{2}\dim(V_{\mathfrak{a}})$, and so $\delta(\mathcal{A}[p^n]_{\mathfrak{a}}) = \frac{1}{2}\dim(V_{\mathfrak{a}})n + O(1)$. Since $V_{\mathfrak{a}}$ is a free $\hat{\mathbb{T}}_{\mathfrak{a}}[1/p]$-module of rank 2, we have $\frac{1}{2}\dim(V_{\mathfrak{a}}) = \mathrm{rk}(\hat{\mathbb{T}}_{\mathfrak{a}}) = d$. □

**Finiteness of the Eisenstein piece of the Mordell–Weil group**

**Proposition.** The group $\hat{\mathbb{T}}_{\mathfrak{a}} \otimes_{\mathbb{T}} A(\mathbb{Q})$ is finite.

*Proof.* Let $\mathcal{A}^\circ$ be the identity component of the Néron model of $\mathcal{A}$ and let $\mathcal{G}_n = \mathcal{A}^\circ[p^n]_{\mathfrak{a}}$. We have

$$\alpha(\mathcal{G}_n) = \alpha(\mathcal{A}[p^n]_{\mathfrak{a}}) = nd + O(1)$$

and

$$\delta(\mathcal{G}_n) = \delta(\mathcal{A}[p^n]_{\mathfrak{a}}) + O(1) = nd + O(1)$$

(the two $\delta$'s can only differ by the size of the component group of the special fiber of $\mathcal{A}$ at $N$). Since the group $\mathcal{G}_n$ is admissible, we have

$$h^1(\mathcal{G}_n) - h^0(\mathcal{G}_n) \leq \delta(\mathcal{G}_n) - \alpha(\mathcal{G}_n) = O(1)$$

Now, $\mathcal{G}_n(\mathbb{Z}) \subset \mathcal{A}(\mathbb{Z})[p^n] = A(\mathbb{Q})[p^n]$ has bounded cardinality as $n$ varies, by the Mordell–Weil theorem. Thus $h^0(\mathcal{G}_n)$ is bounded, and so $h^1(\mathcal{G}_n)$ is bounded as well.

From the short exact sequence

$$0 \to \mathcal{A}^\circ[p^n] \to \mathcal{A}^\circ \to \mathcal{A}^\circ \to 0$$

we obtain an injection

$$\mathcal{A}^\circ(\mathbb{Z}) \otimes \mathbb{Z}/p^n\mathbb{Z} \to \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}, \mathcal{A}^\circ[p^n])$$

Taking inverse limits over $n$, we have an injection

$$\mathcal{A}^\circ(\mathbb{Z}) \otimes \mathbb{Z}_p \to \varprojlim \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}, \mathcal{A}^\circ[p^n])$$

Applying the idempotent $e$, we obtain an injection

$$\mathcal{A}^\circ(\mathbb{Z}) \otimes_{\mathbb{T}} \hat{\mathbb{T}}_{\mathfrak{a}} \to \varprojlim \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}, \mathcal{G}_n)$$

Since the group on the right is finite, so is the group on the left. The inclusion $\mathcal{A}^\circ(\mathbb{Z}) \subset \mathcal{A}(\mathbb{Z}) = A(\mathbb{Q})$ has finite index, and so we conclude that $A(\mathbb{Q}) \otimes_{\mathbb{T}} \hat{\mathbb{T}}_{\mathfrak{a}}$ is finite as well □

**Completion of proof**

**Lemma.** Let $\mathcal{O}$ be an order in a number field and let $\mathfrak{a}$ be a maximal ideal in $\mathcal{O}$. Suppose that $M$ is a finitely generated $\mathcal{O}$ module and $M \otimes_{\mathcal{O}} \hat{\mathcal{O}}_{\mathfrak{a}}$ is finite. Then $M$ is finite.

**Lemma.** Let $M$ be a $\mathbb{T}/I$-module which is finitely generated as a $\mathbb{Z}$-module, and suppose that $M \otimes_{\mathbb{T}} \hat{\mathbb{T}}_{\mathfrak{a}}$ is finite. Then $M$ is finite.

*Proof.* The map $M \to \bigoplus M/\mathfrak{f}M$ has finite kernel and cokernel, where the sum is over the minimal primes $\mathfrak{f}$ of $\mathbb{T}$ contained in $\mathfrak{a}$. Thus the completion of $M/\mathfrak{f}M$ at $\mathfrak{a}$ is finite for each $\mathfrak{f}$. By the above lemma, each $M/\mathfrak{f}M$ is finite. Thus $M$ is finite as well. □

**Proposition.** $A(\mathbb{Q})$ has rank 0.

*Proof.* We have shown that $A(\mathbb{Q}) \otimes_{\mathbb{T}} \hat{\mathbb{T}}_{\mathfrak{a}}$ is finite. Since $A(\mathbb{Q})$ is a $\mathbb{T}/I$-module which is finitely generated as a $\mathbb{Z}$-module, the result follows from the above lemma. □

### 3.2.3 Set-up

We prove the theorem of Mazur and Tate: no elliptic curve over the rational numbers has a point of order 13.

**Statement of the main theorem**

**Theorem** (Main Theorem)**.** No elliptic curve over $\mathbb{Q}$ has a rational point of order 13.

To prove this theorem, we'll show that $X_1(13)$ has no rational points other than the cusps. The main work lies in showing that $J_(13)$ has Mordell–Weil rank 0.

**Notation**

We use the following notation in this section:

(i) $X = X_1(13)$ and $J = J_1(13)$. Both are considered over $\mathbb{Z}[1/13]$.

(ii) $G = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

(iii) $K = \mathbb{Q}(\sqrt[13]{1})$.

(iv) $K^+$ is the maximal real subfield of $K$.

**References**

We follow the article of Mazur and Tate (MR0347826) in our proof of the main theorem. We also need some results of Ogg, established in MR0337974.

### 3.2.4 Preliminaries on $X_1(13)$

**Definition**

We begin by giving the definition of $X = X_1(13)$. We work throughout over $\mathbb{Z}[1/13]$. Recall that $Y_1(13)$ parametrizes pairs $(E, P)$ where $E$ is an elliptic curve and $P$ is a point of order 13 on $E$. The space $X$ parametrizes pairs $(E, P)$ where $E$ is a generalized elliptic curves and $P$ is a point on $E$ of order 13 such that the group generated by $E$ meets all irreducible components of $E$ over the algebraic closure.

There is a natural map $X \to X_0(13)$ taking $(E, P)$ to $(E, G)$, where $G$ is the group generated by $P$. A simple calculation with the genus formula shows that $X_0(13)$ has genus 0 and $X_1(13)$ has genus 2.

**Cusps**

There are two cusps on $X_0(13)$: for one, $E$ is a 1-gon and $G$ is $\mu_{13}$, and for the other, $E$ is a 13-gon and $G = \mathbb{Z}/13\mathbb{Z}$. Both of these are defined over $\mathbb{Q}$. There are six points of $X$ lying over each of these two cusps: there are 12 choices of generator for the group $G$, but $(E, P)$ and $(E, -P)$ are isomorphic. When $E$ is the 1-gon, picking a generator involves picking a root of unity, and so these points are not rational: they are defined over $K^+$. When $E$ is a 13-gon, the generators of $G$ are rational, and so the corresponding points are rational as well. We thus see that $X$ has 12 cusps, 6 of which are defined over $\mathbb{Q}$.

**The twisted dihedral action**

The group $(\mathbb{Z}/13\mathbb{Z})^\times$ acts on $X$ by scaling the point $P$, that is, $a \in (\mathbb{Z}/13\mathbb{Z})^\times$ takes $(E, P)$ to $(E, aP)$. The element $-1 \in (\mathbb{Z}/13\mathbb{Z})^\times$ acts trivially on $X$, since $(E, P)$ and $(E, -P)$ are isomorphic. Let $\Gamma$ be the quotient of $(\mathbb{Z}/13\mathbb{Z})^\times$ by $\{\pm 1\}$, a cyclic group of order 6. Then $\Gamma$ acts faithfully on $X$. For $a \in \Gamma$, we let $\gamma_a$ be the corresponding automorphism of $X$.

There are also Atkin–Lehner type automorphisms of $X$. Let $\zeta$ be a primitive 13th root of unity, and let $(E, P)$ be a point of $Y_1(13)$. There is then a 13-torsion point $Q$ on $E$ which pairs with $P$ under the Weil pairing to $P$. This point is unique up to translation by $P$, i.e., if $Q_1$ is a second such point then $Q - Q_1$ is a multiple of $P$. Let $E'$ be the quotient of $E$ by the group generated by $P$, and let $P'$ be the image of $Q$ in $E'$, which is well-defined. Then $(E', P')$ is a new point on $Y_1(13)$. We let $\tau_\zeta$ be the map $Y_1(13) \to Y_1(13)$ taking $(E, P)$ to $(E', P')$. This can be extended to a map on $X$.

One readily verifies that $\tau_\zeta$ and $\tau_{\zeta^{-1}}$ induce the same map of $X$. We let $\Gamma'$ be the set of primitive 13th roots of unity modulo $\zeta = \zeta^{-1}$, a set of cardinality 6. Thus $\tau_\zeta$ is well-defined for $\zeta \in \Gamma'$.

One has the following relations between the $\gamma$'s and $\tau$'s:

$$\gamma_m \tau_\zeta = \tau_{\zeta^m}, \qquad \tau_\zeta \gamma_m \tau_\zeta^{-1} = \gamma_m^{-1}, \qquad \tau_\zeta^2 = 1$$

It follows that $\Delta = \Gamma \cup \Gamma'$ is a group of automorphisms of $X_1(13)$, and isomorphic to the dihedral group of order 12.

There is a natural action of $G$ on $\Delta$ which fixes $\Gamma$ and acts on $\Gamma'$ in the obvious manner. This action is compatible with the action on $X$. In other words, via the Galois action we can regard $\Delta$ as an étale group scheme over $\mathbb{Z}[1/13]$, and the action of this group scheme on $X$ is defined over $\mathbb{Z}[1/13]$.

The group $\Delta$ acts simply transitively on the cusps, while the subgroup $\Gamma$ acts simply transitively on the rational cusps.

### 3.2.5 Ogg's results

**The subgroup generated by the cusps**

**Proposition.** Let $P_i$ for $1 \leq i \leq 6$ be the 6 rational cusps on $X$. Then for $i \neq j$, the divisor $[P_i] - [P_j]$ defines a point of order 19 in $J(\mathbb{Q})$. Furthermore, these points all generate the same subgroup of $J(\mathbb{Q})$.

*Proof Idea.* The idea of the proof is similar to the computation of the order of $[0] - [\infty]$ in $J_0(N)$ given in "The Eisentein ideal and Eisenstein quotient of $J_0(N)$", but a bit more complicated. For $1 \leq a \leq 6$, let

$$E_{2,a}(z) = \sum_{n,m} \frac{1}{(mz+n)^2}$$

where the sum is over integers $n$ and $m$ such that $n = 0 \pmod{13}$ and $m = a \pmod{13}$. Then $\varphi_{i,j} = E_{2,i} - E_{2,j}$ is a weight 2 modular form for $\Gamma_1(13)$. As such, one can show that it must have 14 zeros in the fundamental domain for $\Gamma_1(13)$. On the other hand, one can compute the order of vanishing of $\varphi_{i,j}$ at the cusps (by computing the first few terms of $q$-series), and one finds that it has 14 zeros at the cusps. It follows that $\varphi_{i,j}$ only vanishes at the cusps, and so the full divisor of $\varphi_{i,j}$ is known; call it $D_{i,j}$. Since $\varphi_{i,j}/\varphi_{k,\ell}$ is a meromorphic function on $X$, its divisor represents the point 0 in $J$. It follows that $D_{i,j} = D_{k,\ell}$ as points of the $J$. This gives enough relations between the $[P_i] - [P_j]$ to establish the proposition. □

**The torsion subgroup of $J(Q)$**

**Proposition.** The torsion subgroup of $J(\mathbb{Q})$ is cyclic of order 19 and generated by the difference of any two cusps.

*Proof.* Observe the following:

(i) The only points on $X$ defined over $\mathbb{F}_4$ are the 6 rational cusps. Indeed, an elliptic curve over $\mathbb{F}_2$ or $\mathbb{F}_4$ cannot have a point of order 13 (by the Hasse bound), and the other 6 cusps are not rational over $\mathbb{F}_2$ or $\mathbb{F}_4$.

(ii) Similarly, the only points on $X$ defined over $\mathbb{F}_3$ are the 6 rational cusps.

(iii) There is a unique elliptic curve $E$ over $\mathbb{F}_9$ with a point of order 13, and $\mathrm{Aut}(E) = \mathbb{Z}/6\mathbb{Z}$. This yields 2 points on $Y_1(\mathbb{F}_9)$. The only other points on $X$ defined over $\mathbb{F}_9$ are the 6 rational cusps.

We thus see that $\#X(\mathbb{F}_2) = \#X(\mathbb{F}_4) = 6$, and so $\#J(\mathbb{F}_2) = 19$ by the following lemma. We also see that $\#X(\mathbb{F}_3) = 6$ and $\#X(\mathbb{F}_9) = 8$, and so $\#J(\mathbb{F}_3) = 19$ by the following lemma. Since the kernel of $J(\mathbb{Q})_{\text{tors}} \to J(\mathbb{F}_2)$ consists only of 2-torsion and the kernel of $J(\mathbb{Q})_{\text{tors}} \to J(\mathbb{F}_3)$ consists only of 3-torsion, the result follows. (In fact, $J(\mathbb{Q}) \to J(\mathbb{F}_3)$ is injective using the stronger results coming from Raynaud's theorem, so the analysis at 2 is unnecessary.) □

**Lemma.** Let $X/\mathbb{F}_q$ be a curve of genus 2 and let $J$ be its Jacobian. Then

$$\#J(\mathbb{F}_q) = -q + \tfrac{1}{2}\#X(\mathbb{F}_{q^2}) + \tfrac{1}{2}(\#X(\mathbb{F}_q))^2$$

*Proof.* Let $V = \mathrm{H}^1_{\mathrm{et}}(X_{\overline{\mathbb{F}}_q}, \mathbb{Q}_\ell)$, a four dimensional $\mathbb{Q}_\ell$ vector space, and let $F$ be Frobenius. Then

$$\#X(\mathbb{F}_q) = 1 + q - \mathrm{tr}(F \mid V),$$

and a similar formula holds for $\#X(\mathbb{F}_{q^2})$ using $F^2$ in place of $F$. Also,

$$\#J(\mathbb{F}_q) = \sum_{i=0}^{4}(-1)^i \,\mathrm{tr}(F \mid \bigwedge^i V).$$

Finally, note that $q^tF^{-1}$ is conjugate to $F$, by duality. The result is now a linear algebra exercise. □

**The image of $X$ in $J$**

**Proposition.** The image of $X(\mathbb{C})$ in $J(\mathbb{C})$ (via $P \mapsto [P] - [P_6]$) intersects $J(\mathbb{Q})_{\mathrm{tors}}$ only at the 6 points $[P_i] - [P_6]$.

*Proof Idea.* If $[P]-[P_6]$ belonged to $J(\mathbb{Q})_{\mathrm{tors}}$ for some $P \in X(\mathbb{C})$ then we would have an equality $[P]-[P_6] = n([P_1]-[P_6])$ in $J$, and so $[P]-n[P_1]+(n-1)[P_0]$ would be the divisor of a function on $X$. Ogg shows that this cannot happen unless $n([P_1]-[P_6]) = [P_i]-[P_6]$ for some $i$. □

**Corollary.** To prove the main theorem, it suffices to show that $J(\mathbb{Q})$ has rank 0.

Ogg notes that this is predicted by the BSD conjecture.

### 3.2.6 Structure of $J$

**Proposition.** $J$ is a simple abelian variety over $\mathbb{Q}$.

*Proof.* Suppose that

$$0 \to J_1 \to J \to J_2 \to 0$$

is an exact sequence of abelian varieties, with $J_1$ and $J_2$ elliptic curves. Then either $J_1$ or $J_2$ would have a point of order 19, which we have already shown to be impossible. (In this case, it is easy to see, as $J_1$ and $J_2$ have good reduction away from 13, and so reducing modulo 2 would give an elliptic curve over $\mathbb{F}_2$ with a point of order 19, which is impossible.) □

We observe that the element $\gamma_2$ generates $\Gamma$.

**Proposition.** The action of $\gamma_2$ on $J$ satisfies the polynomial $1 - x + x^2 = 0$.

*Proof.* This follows from the following observations:

(i) $\gamma_2$ satisfies the polynomial $x^6 - 1$ but not any polynomial of the form $x^d - 1$ with $d$ a proper divisor of 6. This is because $\gamma_2$ has exact order 6 on $X$, and hence $J$.

(ii) If $\gamma_2$ satisfies some polynomial then it satisfies some irreducible factor of that polynomial. This is because $J$ is simple.

(iii) The polynomial $x^6 - 1$ factors as $(1-x)(1+x)(1+x+x^2)(1-x+x^2)$, and $1-x$, $1+x$, and $1+x+x^2$ all divide polynomials of the form $x^d - 1$ with $d$ a proper divisor of 6.

It follows from the above proposition that the ring $D = \mathbb{Z}[\Delta]/(1-\gamma_2+\gamma_2^2)$ acts on $J$. This action is necessarily faithful since $D$ is an order in a simple algebra (in fact, $D \otimes \mathbb{Q} = M_2(\mathbb{Q})$). Note that $\mathbb{Z}[\gamma_2] \subset D$ is isomorphic to $\mathbb{Z}[\sqrt[3]{1}]$, and $\gamma_2 = -\sqrt[3]{1}$, a 6th root of unity. □

**Remark.** The action of $D$ on $J$ is defined over $K^+$, and so $\mathrm{End}(J_{K^+}) \otimes \mathbb{Q}$ contains $M_2(\mathbb{Q})$. It follows that $J$ is not simple over the field $K^+$.

### 3.2.7 The 19-torsion in $J$

Let $V = J[19](\overline{\mathbb{Q}})$ be the 19-torsion in $J$, which is a 4-dimensional vector space over the field $\mathbb{F}_{19}$. This space admits compatible actions of $G$ and $\Delta$. In particular, it is a module over the ring $\mathbb{Z}[\gamma_2] = \mathbb{Z}[\sqrt[3]{1}]$.

The prime 19 splits in $\mathbb{Z}[\sqrt[3]{1}]$; let $19 = \pi\overline{\pi}$ be its factorization. Let $V_\pi$ and $V_{\overline{\pi}}$ be the kernels of $\pi$ and $\overline{\pi}$ on $V$, so that $V = V_\pi \oplus V_{\overline{\pi}}$. The spaces $V_\pi$ and $V_{\overline{\pi}}$ are stable by $G$ and $\Gamma$, since $\gamma_2$ commutes with these groups, but interchanged under $\tau_\zeta$, since conjugation by $\tau_\zeta$ induces a non-trivial automorphism of $\mathbb{Z}[\gamma_2]$, and thus interchanges $\pi$ and $\overline{\pi}$.

**Proposition.** The Weil pairing on $V$ induces a Cartier duality between $V_\pi$ and $V_{\overline{\pi}}$.

*Proof.* It suffices to show that $V_\pi$ is self-orthogonal under the Weil pairing. Denote the pairing by $(,)$. We have $(\gamma_2 x, \gamma_2 y) = (x, y)$, since $\gamma_2$ induces the identity on $\mathrm{H}_2$ of $X$. On the other hand, $V_\pi$ and $V_{\overline{\pi}}$ are eigenspaces of $\gamma_2$ with eigenvalues $\zeta^{\pm 1}$, where $\zeta \in \mathbb{F}_{19}$ is a primitive 6th root of unity. Thus for $x, y \in V_\pi$ we have

$$(x, y) = (\gamma_2 x, \gamma_2 y) = (\zeta x, \zeta y) = \zeta^2 (x, y),$$

and so $(x, y) = 0$, since $\zeta^2 \neq 1$. □

We now define some more subspaces of $V$:

(i) Let $V(1) \subset V$ be the 1-dimensional space $J(\mathbb{Q})_{\mathrm{tors}}$ spanned by the cusps, on which $G$ acts trivially. Since $V(1)$ is stable by $\gamma_2$, it is contained in $V_\pi$ or $V_{\overline{\pi}}$. Relabeling if necessary, we can assume it's contained in $V_{\overline{\pi}}$.

(ii) Let $V(\gamma) \subset V$ be the space of vectors $v$ satisfying $av = \gamma_a v$ for $a \in \Gamma$, where here we identify $\Gamma$ with $\mathrm{Gal}(K^+/\mathbb{Q})$. One easily sees that the action of $\Gamma'$ interchanges $V(1)$ and $V(\gamma)$, and so $V(\gamma)$ is a one dimensional subspace of $V_\pi$.

(iii) Let $V(\chi)$ be the one dimensional $\mathbb{F}_{19}$ space on which $G$ acts through the mod 19 cyclotomic character. The inclusion $V(1) \to V_{\overline{\pi}}$ combined with the above proposition yields a surjection $V_\pi \to V(1)$.

**Proposition.** The sequence

$$0 \to V(\gamma) \to V_\pi \to V(\chi) \to 0$$

is exact.

*Proof.* The action of $G$ on $V(\gamma)$ factors through $\mathrm{Gal}(K^+/\mathbb{Q})$ and no smaller quotient, while the action of $G$ on $V(\chi)$ factors through $\mathrm{Gal}(\mathbb{Q}(\sqrt[19]{1})/\mathbb{Q})$ and no smaller quotient. Thus $V(\gamma)$ and $V(\chi)$ are non-isomorphic representations of $G$, and so the composite is 0. This proves the proposition. □

### 3.2.8 Rank 0

**First reduction**

To prove that $J(\mathbb{Q})$ has rank 0, it is enough to show that the map $\pi$ on $J(\mathbb{Q})$ is surjective, as a finitely generated $\mathbb{Z}[\sqrt[3]{1}]$-module on which $\pi$ acts surjectively is necessarily finite. We have the following diagram:

$$\begin{tikzcd}
J(\mathbb{Z}[1/13]) \arrow[r, "\pi"] \arrow[d] & J(\mathbb{Z}[1/13]) \arrow[r] \arrow[d] & \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13], J[\pi]) \arrow[d, "\rho"] \\
J(\mathbb{Q}_{13}) \arrow[r, "\pi"] & J(\mathbb{Q}_{13}) \arrow[r] & \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Q}_{13}, J[\pi])
\end{tikzcd}$$

To prove that $\pi$ is surjective on $J(\mathbb{Z}[1/13]) = J(\mathbb{Q})$, it is enough to show (i) that $\pi$ is surjective on $J(\mathbb{Q}_{13})$; and (ii) that $\rho$ is injective.

**Proof of (i)**

Let $\mathcal{J}$ be the Néron model of $J$ over $\mathbb{Z}_{13}$; note that $\mathcal{J}(\mathbb{Z}_{13}) = J(\mathbb{Q}_{13})$. Let $N$ be the kernel of the reduction map $\mathcal{J}(\mathbb{Z}_{13}) \to \mathcal{J}(\mathbb{F}_{13})$. We have the following diagram

$$\begin{tikzcd}
0 \arrow[r] & N \arrow[r] \arrow[d, "\pi"] & \mathcal{J}(\mathbb{Z}_{13}) \arrow[r] \arrow[d, "\pi"] & \mathcal{J}(\mathbb{F}_{13}) \arrow[r] \arrow[d, "\pi"] & 0 \\
0 \arrow[r] & N \arrow[r] & \mathcal{J}(\mathbb{Z}_{13}) \arrow[r] & \mathcal{J}(\mathbb{F}_{13}) \arrow[r] & 0
\end{tikzcd}$$

Now, $N$ is pro-13 and so 19, and thus $\pi$, acts bijectively on it. By the snake lemma, the cokernel of $\pi$ on $\mathcal{J}(\mathbb{Z}_{13})$ is isomorphic to the cokernel of $\pi$ on $\mathcal{J}(\mathbb{F}_{13})$, and so it suffices to show that this is 0. But $\mathcal{J}(\mathbb{F}_{13})$ is a finite set, and so the cokernel of $\pi$ on it vanishes if and only if the kernel of $\pi$ on it vanishes. Again, the snake lemma shows that the kernel of $\pi$ on $\mathcal{J}(\mathbb{F}_{13})$ is isomorphic to the kernel of $\pi$ on $\mathcal{J}(\mathbb{Z}_{13})$. We are thus reduced to showing that the kernel of $\pi$ on $J(\mathbb{Q}_{13})$ vanishes. Now, $J(\mathbb{Q}_{13})[\pi]$ is exactly $V_\pi^D$, where $D \subset G$ is the decomposition group at 13. As we have an exact sequence

$$0 \to V(\gamma) \to V_\pi \to V(\chi) \to 0,$$

it suffices to show that $V(\gamma)^D = 0$ and $V(\chi)^D = 0$. The former assertion follows since 13 ramifies in $K^+$, while the latter follows since 13 does not split in $\mathbb{Q}(\sqrt[19]{1})$ (as $13 \neq 1 \pmod{19}$).

**Proof of (ii)**

**Proposition.** We have a short exact sequence of group schemes over $\mathbb{Z}[1/13]$:

$$0 \to E \to J[\pi] \to \mu_{19} \to 0$$

where $E$ is finite étale and becomes constant over $\mathbb{Z}[1/13, \sqrt[13]{1}]$.

*Proof.* Let $E$ be the scheme theoretic closure of $V(\gamma)$ in $J[\pi]$. The restriction of $E$ to the finite étale cover $\mathbb{Z}[1/13, \sqrt[13]{1}]$ is trivial by Raynaud's theorem, as the restriction of $V(\gamma)$ to $\mathbb{Q}(\sqrt[13]{1})$ is trivial. The quotient $J[\pi]/E$ is generically isomorphic to $\mu_{19}$, as its Galois representation is $V(\chi)$, and so isomorphic to $\mu_{19}$ over $\mathbb{Z}[1/13]$ by Raynaud's theorem. □

Consider the diagram

$$\begin{tikzcd}
\mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13], E) \arrow[r] & \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13], J[\pi]) \arrow[r] \arrow[d, "\rho"] & \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13], \mu_{19}) \arrow[d, "\rho'"] \\
 & \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Q}_{13}, J[\pi]) \arrow[r] & \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Q}_{13}, \mu_{19})
\end{tikzcd}$$

To show that $\rho$ is injective, it suffices to show (iia) that $\rho'$ is injective; and (iib) that $\mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13], E) = 0$. We do this in the following two propositions.

**Proposition.** $\rho'$ is injective.

*Proof.* By Kummer theory, $\rho'$ is equivalent to the map

$$\mathbb{Z}[1/13]^\times/(\mathbb{Z}[1/13]^\times)^{19} \to \mathbb{Q}_{13}^\times/(\mathbb{Q}_{13}^\times)^{19}$$

This map is injective because $\mathbb{Z}[1/13]^\times = \{\pm 13^n\}$ and 13 is not a 19th power in $\mathbb{Q}_{13}$ (it has valuation 1). □

**Proposition.** $\mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13], E) = 0$.

*Proof.* We have $\mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13], E) = \mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13, \sqrt[13]{1}], \mathbb{Z}/19\mathbb{Z})^{\mathrm{Gal}(K/\mathbb{Q})}$, and so it suffices to show that $\mathrm{H}^1_{\mathrm{fppf}}(\mathbb{Z}[1/13, \sqrt[13]{1}], \mathbb{Z}/19\mathbb{Z}) = 0$. This amounts to showing that $K$ has no abelian extension of degree 19 unramified everywhere away from the prime $\lambda$ of $K$ above 13. Such abelian extensions correspond to quotients of the ray class group for the modulus $\lambda$. The ray class group is an extension of the class group by a quotient of the group of $\lambda$-adic units. Both of these groups are prime to 19, which proves the result. □

## 3.3 Ogg's theorem on the order of $[\ ] - [0]$ in $J_0(N)$

In this final section, we complete the classification of torsion groups of elliptic curves over the raionals. Actually, we only give an overview of the proof, as it is a rather lengthy case-by-case analysis.

### 3.3.1 Where we are!

Our goal in this course has been to prove the following theorem:

**Theorem** (The Main Theorem)**.** Let $G$ be a finite abelian group. Then $G = E(\mathbb{Q})_{\mathrm{tors}}$ for some elliptic curve $E/\mathbb{Q}$ if and only if $G$ is one of the following groups:

(i) $\mathbb{Z}/N\mathbb{Z}$ for $1 \leq N \leq 10$ or $N = 12$.

(ii) $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ with $N \in \{2, 4, 6, 8\}$.

We have proven the following:

**Theorem** (Mazur, Mazur – Tate)**.** Let $N$ be a prime $> 7$. Then no elliptic curve over $\mathbb{Q}$ has a rational point of order $N$.

To complete the proof of the theorem, then, it remains to show the following:

(i) $E(\mathbb{Q})$ does not have any $N$-torsion for $N \in \{14, 15, 16, 18, 20, 21, 24, 25, 27, 35, 49\}$.

(ii) $E(\mathbb{Q})$ does not have a subgroup of the form $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ for $N \in \{10, 12\}$.

(iii) The 15 groups that should occur as torsion subgroups actually do.

Most of the work was carried out by Kubert in MR0434947.

### 3.3.2 The groups that should occur do

Let $N$ be an integer with $1 \leq N \leq 10$ or $N = 12$. The curve $X_1(N)$ then has genus 0, and, in fact, these are the only values of $N$ for which this is true. Furthermore, $X_1(N)$ has rational points, namely, its rational cusps. It follows that $X_1(N)$ is isomorphic to $\mathbb{P}^1$ over $\mathbb{Q}$, and therefore has infinitely many rational points. In particular, $Y_1(N)$ has rational points. For $N \geq 4$, the scheme $Y_1(N)$ represents the moduli problem of elliptic curves with $N$-torsion, and so there exist elliptic curves over $\mathbb{Q}$ with a rational point of order $N$. For $N = 2, 3$, the same conclusion holds, and is easy to see directly. (Any curve of the form $y^2 + axy + by = x^3$ admits $(0, 0)$ as a 3-torsion point, and any curve of the form $y^2 = f(x)$ where $f$ has a rational root has a 2-torsion point.)

Now let $N \in \{4, 6, 8\}$. Let $Y$ be the moduli scheme parametrizing elliptic curves equipped with an injection from $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ (this does exist as a scheme), and let $X$ be the corresponding compactification. Then, again, $X$ has genus and a rational point, and is therefore isomorphic to

$\mathbb{P}^1$ over $\mathbb{Q}$. It follows that $Y(\mathbb{Q})$ is non-empty, and so $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ does occur as a subgroup of the Mordell–Weil group of an elliptic curve over $\mathbb{Q}$. It is easy to see that this conclusion continues to hold for $N = 2$: consider $y^2 = f(x)$ where $f$ has only rational roots.

We have thus shown that each of the 15 groups $G$ there exists an elliptic curve $E/\mathbb{Q}$ such that $E(\mathbb{Q})$ contains a subgroup isomorphic to $G$. In fact, one can find $E$ for which $E(\mathbb{Q})_{\text{tors}}$ is isomorphic to $G$. We just explain the idea in one case. Let's show that $\mathbb{Z}/5\mathbb{Z}$ occurs exactly as the torsion subgroup. If it didn't, that'd mean that every elliptic curve over $\mathbb{Q}$ with a rational 5-torsion point would also have an additional rational torsion point, which necessarily would be 2-torsion by Ogg's conjecture. In other words, the map $X_1(10) \to X_1(5)$ would induce a surjection on rational points. But this is a degree 3 map of rational curves, so that cannot happen: in fact, most points in the target will not be in the image.

**Remark.** For $N \geq 4$ it is possible to explicitly write down the universal elliptic curve with a point of order $N$ (or a subgroup of the form $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$). For example, the universal curve with a point of order 4 is given by

$$y^2 + xy - ty = x^3 - tx^2$$

The 4-torsion point is $(0, 0)$.

### 3.3.3 When $X_0(N)$ has genus 1

**Excluding certain torsion orders**

**Proposition.** Let $N$ belong to $\{11, 14, 15, 17, 19, 20, 21, 24, 27, 32, 36, 49\}$. Then no elliptic curve over $\mathbb{Q}$ has a point of order $N$.

*Proof.* The curve $X_0(N)$ has genus 1. Choosing one of the cusps as the origin, this gives $X_0(N)$ the structure of an elliptic curve. One can then use standard methods to study the rational points on $X_0(N)$, and one finds that it has rank 0. Determining the torsion subgroup is straightforward, and so one can explicitly write down all the rational points on $X_0(N)$. If all the rational points on $X_0(N)$ are cuspdial then we are done. If not there is more work to do.

Suppose there are non-cuspidal rational points on $X_0(N)$. Enumerate them as $x_1, \ldots, x_n$. Each can be represented by a pair $(E_i, G_i)$ where $E_i/\mathbb{Q}$ is an elliptic curve and $G_i$ is a cyclic subgroup of order $N$. If $y$ is a rational point of $X_1(N)$, corresponding to $(E, P)$, then $y$ maps to some $x_i$, and so $E$ is isomorphic to some twisted form of $E_i$. (Twisted form because $X_0(N)$ is only a coarse moduli space.) To prove the proposition, it therefore suffices to check that no twisted form of any $E_i$ has a point of order $N$. To do this, consider $E_i[N]$ as a Galois representation. Since $E_i$ admits a cyclic isogeny of degree $N$, we have a short exact sequence

$$0 \to (\mathbb{Z}/N\mathbb{Z})(\alpha_i) \to E_i[N] \to (\mathbb{Z}/N\mathbb{Z})(\beta_i) \to 0$$

where $\alpha_i$ and $\beta_i$ are homomorphisms $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to (\mathbb{Z}/N\mathbb{Z})^\times$. Suppose $E_i$ is non-CM, and let $E_i^{(d)}$ be the quadratic twist of $E_i$ by $d$. We then have an exact sequence

$$0 \to (\mathbb{Z}/N\mathbb{Z})(\alpha_i\chi_d) \to E_i^{(d)}[N] \to (\mathbb{Z}/N\mathbb{Z})(\beta_i\chi_d) \to 0$$

where $\chi_d$ is the quadratic character associated to $d$. Thus $E_i^{(d)}$ has a rational $N$-torsion point if and only if $\alpha_i$ is a quadratic character, or $E_i[N]$ is split and $\beta_i$ is a quadratic character. (When $E_i$ is CM a similar statement holds.) Thus, to prove the proposition, it suffices to check this is not the case for each $i$, which is a straightforward finite computation. □

**Example.** When $N = 21$ there are four non-cuspidal rational points on $X_0(N)$. The $E_i$ are given by:

$$\begin{aligned} y^2 &= x^3 + 45x - 18 \\ y^2 &= x^3 - 75x - 262 \\ y^2 &= x^3 - 1515x - 46106 \\ y^2 &= x^3 - 17235x - 870894 \end{aligned}$$

None of these curves are CM, and so it suffices to show that no quadratic twist of any of them has a 21-torsion point.

**Example.** When $N = 27$ there is one non-cuspidal rational point on $X_0(N)$. It is given by

$$y^2 + y = x^3 - 30x - 5$$

This curve has CM by $\sqrt{-27}$.

**Remark.** There are rational non-cuspdial points on $X_0(N)$ if and only if $N \in \{11, 14, 15, 17, 19, 21, 27\}$.

**Remark.** When $N \in \{11, 14, 15\}$, the curve $X_1(N)$ has genus 1. Since $X_1(N) \to X_0(N)$ is an isogeny, it follows that $X_1(N)$ has rank 0. One can therefore easily compute all of its rational points, and one finds that they are all cusps. This is easier than the argument given above.

**Remark.** Combined with the previous remark and what we already know (or will prove), the above proposition is really only needed for $N \in \{20, 21, 24, 27, 49\}$. There are non-cuspidal rational points only for $N \in \{21, 27\}$, so only in these cases does the proof require the additional computation.

### Excluding certain product groups

**Lemma.** Let $E/\mathbb{Q}$ be an elliptic curve such that $E(\mathbb{Q})$ contains a subgroup of the form $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ (resp. $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$). Then there exists a degree 2 isogeny $E \to E'$ defined over $\mathbb{Q}$ such that $E'$ admits a rational cyclic isogeny of degree 4 (resp. 8).

*Proof.* Let $P$ be the point of order 2 and $Q$ an independent point of order 2 (resp. 4). Let $E_1 = E/P$ and $E_2 = E/Q$ and let $f \colon E \to E_1$ and $g \colon E \to E_2$ be the natural isogenies. Consider the dual isogeny $f^\vee \colon E_1 \to E$. The image of $E_1[2]$ under $f^\vee$ is just the point $P$. This point is not in the kernel of $g$, and so $gf^\vee \colon E_1 \to E_2$ is a cyclic isogeny of degree 4 (resp. 8). □

**Proposition.** Let $E/\mathbb{Q}$ be an elliptic curve. Then $E(\mathbb{Q})$ does not contain a subgroup isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/10\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$.

*Proof.* Suppose $E$ did contain such a subgroup. Then there would be a 2-isogeny $E \to E'$ such that $E'$ admits a cyclic 20 or 24 isogeny, and thus defines a rational point on $X_0(20)$ or $X_0(24)$. By the above results, the only such points are cuspidal, which is a contradiction. □

### 3.3.4 The case of 16 torsion

This was dealt with by Lind in his 1940 thesis. The proof is supposed to be easy, but I haven't found it written down anywhere (and can't get the thesis). The curve $X_1(16)$ has genus 2, while $X_0(16)$ has genus 0.

### 3.3.5 The cases of 18 and 25 torsion

These cases are similar to the case of 13 torsion in that $X_0(N)$ has genus 0. This means that there's no map from $X_1(N)$ to an elliptic curve, and so one cannot reduce to the study of an elliptic curve. Kubert carries out a fairly direct generalization of the Mazur–Tate argument in these two cases.

The curve $X_1(18)$ has genus 2. The torsion in $J_1(18)$ is $\mathbb{Z}/21\mathbb{Z}$, and meets $X_1(18)$ only at the cusps. It therefore suffices to show that $J_1(18)$ has rank 0. The automorphism $\gamma_5$ of $X_1(18)$ satisfies $\gamma_5^2 + \gamma_5 + 1 = 0$ on $J_1(18)$, and so $J_1(18)$ contains the ring $\mathbb{Z}[x]/(x^2 + x + 1)$ in its endomorphism algebra. The prime 7 factors as $\pi\overline{\pi}$ in this ring, and Kubert shows that multiplication by $\pi$ is surjective on the Mordell–Weil group of $J_1(18)$.

The case of 25-torsion is more complicated.

### 3.3.6 The case of 35 torsion

At this point, it remains only to exclude 35 torsion. In fact, Kubert shows that $X_0(35)$ has no non-cuspidal rational points. The idea is to consider the quotient $E = X_0(35)/w_5$, where $w_5$ is the Atkin–Lehner involution at 5. This curve has genus 1. One shows that $E(\mathbb{Q}) = \mathbb{Z}/3\mathbb{Z}$, and computes the preimages of these three rational points in $X_0(35)$. They turn out to be cusps.