Disclaimer

These are the notes for the 2024–2025 graduate number theory course at Johns Hopkins (Math 617-618). This course did not cover the standard syllabus of the course, but had basic algebraic number theory, including class field theory, as a prerequisite, in order to cover more advanced topics. The notes have not been carefully checked, so use them at your own peril, and please let us know if you spot serious mistakes!

Number Theory I–II

Murilo Corato Zanarella (Part I), Yiannis Sakellaridis (Part II)

Last updated April 24, 2025

Contents

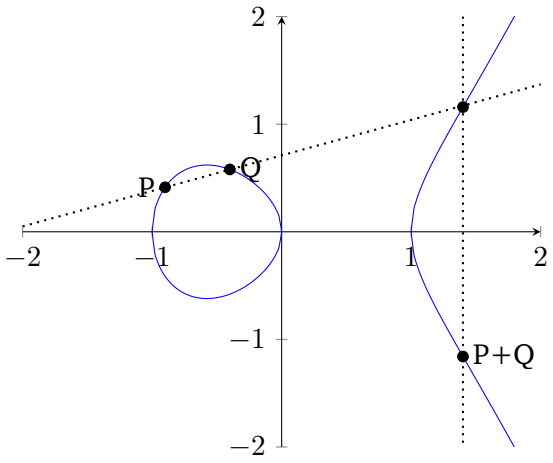
Part I
Fall 2024

0 Introduction

Definition 0.1. An elliptic curve (E, O) over a field k is a nonsingular projective algebraic curve E/k of genus 1 with a marked point $O \in E(k)$.

Example 0.1. Consider the projective curve with affine chart $E\colon y^2 = x^3 - n^2x$, taking $O = [0\colon 1\colon 0]$ the point at infinity. Then $E(\mathbb{Q}) \supseteq \{O, (0,0), (\pm n, 0)\}$, and this inclusion is strict if and only if n is a *congruent number*, i.e. the area of a right triangle with rational side lengths.

It turns out that the set $E(k)$ has naturally the structure of an abelian group. Pictorially, it is given as follows.



One of the goals of this class will be to understand the arithmetic of elliptic curves, i.e. the group $E(k)$ for k a number field.

Theorem 0.1 (Mordell–Weil)**.** *For E an elliptic curve over a number field k, the abelian group $E(k)$ is finitely generated: $E(k) \simeq \mathbb{Z}^r \times E(k)_{\mathrm{tors}}$, where $r = r_{\mathrm{alg}}(E/k)$ is called the* algebraic rank *of E/k.*

For $E/\mathbb{Q}$, one can attach an L-function roughly given by the Euler product

$$L(E/\mathbb{Q}, s) = \prod_p (1 - a_p p^{-s} + p^{1-2s})^{-1}, \quad \text{where } a_p = p + 1 - \#E(\mathbb{F}_p). \tag{0.1}$$

As we will see later, this converges absolutely for $\mathrm{Re}(s) > \frac{3}{2}$. We have the deep conjecture.

Conjecture 0.1 (Birch–Swinnerton-Dyer)**.** *$L(E/\mathbb{Q}, s)$ is an entire function, and if we denote $r_{\text{an}}(E/\mathbb{Q}) := \text{ord}_{s=1} L(E/\mathbb{Q}, s)$, then*

$$r_{\text{alg}}(E/\mathbb{Q}) = r_{\text{an}}(E/\mathbb{Q}). \tag{0.2}$$

All of the progress we have on this conjecture (which is only on cases of rank $\leqslant 1$) heavily rely on *modularity*: the relationship of elliptic curves and *modular forms* as follows.

Theorem 0.2 (Shimura–Taniyama conjecture, Breuil–Conrad–Diamond–Taylor–Wiles,...)**.** *There is a bijection*

$$\{E/\mathbb{Q} \text{ of conductor } N\}/\text{isogeny} \longleftrightarrow \{\text{rational Hecke eigenforms of weight } 2 \text{ and level } \Gamma_0(N)\} \tag{0.3}$$

For a modular form f, there is also a notion of an L-function $L(s, f)$ for which we can prove good analytic properties. In the above bijection, we have $L(s, E/\mathbb{Q}) = L(s, f_E)$, and this is how we can eventually prove that $L(s, E/\mathbb{Q})$ has good analytic properties.

Furthermore, to an elliptic curve $E/\mathbb{Q}$ we can attach *Galois representations*

$$\rho_{E,\ell} \colon \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \text{GL}_2(\mathbb{Q}_\ell) \tag{0.4}$$

and the association between modular forms and such Galois representations is an instance of global Langlands.

1 GL_1 case

Before delving into elliptic curves and modular forms, let's first discuss the theory of L-functions for GL_1.

1.1 L-functions of Dirichlet characters

Reference: [Bum97, Section 1.1].

Definition 1.1. A *Dirichlet character modulo N* is a group homomorphism $\chi \colon (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$. We say χ is *primitive* if it doesn't factor through $(\mathbb{Z}/M\mathbb{Z})^\times$ for $M \mid N$ proper divisors. In the case χ is primitive, we call N its *conductor*, and extend it to a function $\chi \colon \mathbb{Z} \to \mathbb{C}$ by declaring that $\chi(n) = 0$ whenever $(n, N) \neq 1$.

Attached to a Dirichlet character $\chi\colon \mathbb{Z} \to \mathbb{C}^\times$, we consider its L-function

$$L(s,\chi) = \sum_{n\geqslant 1} \frac{\chi(n)}{n^s} = \prod_p (1-\chi(p)p^{-s})^{-1}, \quad \mathrm{Re}(s) > 1. \tag{1.1}$$

For example $\zeta(s)$ is the L-function for the trivial character. Let's briefly recall how one proves that these L-functions extend meromorphically.

Theorem 1.1. *If $\chi(-1) = (-1)^\epsilon$ for $\epsilon \in \{0,1\}$, we denote*

$$\Lambda(s,\chi) = \pi^{-(s+\epsilon)/2}\Gamma\left(\frac{s+\epsilon}{2}\right) L(\chi,s). \tag{1.2}$$

Then $\Lambda(s,\chi)$ extends meromorphically to all s and satisfies the functional equation

$$\Lambda(1-s,\chi^{-1}) = \frac{i^\epsilon N^s}{\tau(\chi)}\Lambda(s,\chi) \tag{1.3}$$

where $\tau(\chi)$ is the Gauss sum

$$\tau(\chi) = \sum_{n \mod N} \chi(n) e^{2\pi i n/N}. \tag{1.4}$$

Proof. For simplicity, let's consider the case χ is not the trivial character. Consider the theta function

$$\theta_\chi(t) = \frac{1}{2}\sum_{n\in\mathbb{Z}} n^\epsilon \chi(n) e^{-\pi n^2 t} = \sum_{n\geqslant 1} n^\epsilon \chi(n) e^{-\pi n^2 t}. \tag{1.5}$$

This is defined such that so that it's Mellin transform is $\Lambda(s,\chi)$:

$$\int_0^\infty \theta_\chi(t) t^{(s+\epsilon)/2}\frac{\mathrm{d}t}{t} = \sum_{n\geqslant 1} n^\epsilon \chi(n) \frac{\pi^{-(s+\epsilon)/2}\Gamma((s+\epsilon)/2)}{(n^2)^{(s+\epsilon)/2}} = \Lambda(s,\chi). \tag{1.6}$$

A (twisted) Poisson summation formula tells us that

$$\theta_{\chi^{-1}}\left(\frac{1}{N^2 t}\right) = (iNt)^\epsilon \frac{Nt^{1/2}}{\tau(\chi)}\theta_\chi(t). \tag{1.7}$$

In particular we may write

$$\Lambda(s,\chi) = \int_{1/N}^\infty \theta_\chi(t) t^{(s+\epsilon)/2}\frac{\mathrm{d}t}{t} + \frac{\tau(\chi)}{i^\epsilon N^{1+\epsilon}}\int_0^{1/N} \theta_{\chi^{-1}}\left(\frac{1}{N^2 t}\right) t^{(s-1-\epsilon)/2}\frac{\mathrm{d}t}{t} \tag{1.8}$$

and changing variables $t \mapsto 1/(N^2 t)$ in the second term,

$$\Lambda(s,\chi) = \int_{1/N}^\infty \theta_\chi(t) t^{(s+\epsilon)/2}\frac{\mathrm{d}t}{t} + \frac{\tau(\chi)}{i^\epsilon N^s}\int_{1/N}^\infty \theta_{\chi^{-1}}(t) t^{(1-s+\epsilon)/2}\frac{\mathrm{d}t}{t}. \tag{1.9}$$

Now both θ_χ and $\theta_{\chi^{-1}}$ decay exponentially as $t\to\infty$ (here we are using χ, χ^{-1} are nontrivial) and thus the above right hand side extends to an entire function of s. The functional equation is also now clear. □

1.2 L-functions of Hecke characters

Reference: [Bum97, Section 1.7].

Now let's try to mimic the above discussion for an arbitrary number field F. The Riemann zeta function generalizes to Dedekind zeta functions

$$\zeta_F(s) = \sum_{\mathfrak{a} \subseteq \mathcal{O}_F} \frac{1}{\mathrm{Nm}(\mathfrak{a})^s}, \quad \mathrm{Re}(s) > 1. \tag{1.10}$$

One naïve guess of analogues of Dirichlet characters would be characters of the form $\chi_D \colon (\mathcal{O}_F/\mathfrak{n})^\times \to \mathbb{C}^\times$. But this doesn't quite give an L-function, since we would like to make sense of "$\chi_0(\mathfrak{a})$" for ideals $\mathfrak{a}$. Even "$\chi_0((\alpha))$" doesn't make sense.

A character χ_D as above only involves the finite places of F, and it turns out that the fix for the above problem is to also involve the archimedean places: we consider a continuous character

$$\chi_\infty \colon \prod_{v|\infty} F_v^\times = (\mathbb{R}^\times)^{r_1} \times (\mathbb{C}^\times)^{r_2} \to S^1 \subseteq \mathbb{C}^\times \tag{1.11}$$

and we try to make it so that $\chi_0((a)) := \chi_D(a)\chi_\infty^{-1}(a)$ is well-defined. Note that all the possibilities for χ_∞ as above are of the following form: for v real, we have

$$\chi_v(t) = \mathrm{sgn}(t)^{\epsilon_v} |t|_c^{\nu_v}, \quad \text{for } \epsilon_v \in \{0,1\},\ \nu_v \in i\mathbb{R} \tag{1.12}$$

and for v complex we have

$$\chi_v(t) = e^{in_v \arg(t)} |t|_v^{\nu_v}, \quad \text{for } n_v \in \mathbb{Z},\ \nu_v \in i\mathbb{R}. \tag{1.13}$$

Proposition 1.2. *Given χ_D, there is always a choice of χ_∞ such that $\chi_D(a) = \chi_\infty(a)$ for all $a \in \mathcal{O}_F^\times$.*

Proof. For notational simplicity, we will only consider the case F is totally real.

The first step is to choose ν_v so that $\chi_D(a) = \chi_\infty(a)$ for all $a \in \mathcal{O}_F^\times$ totally real. Let $\xi_1, \dots, \xi_{r_1-1}$ be a basis of the totally real units. Then we are looking at the system of equations

$$\sum_v \nu_v \log|\xi_j|_v \equiv \log \chi_D(\xi_j) \mod 2\pi i\mathbb{Z}. \tag{1.14}$$

This has r_1-1 equations and r_1 variables, and the determinants of the $(r_1-1)\times(r_1-1)$ minors are in absolute value, by definition, the regulator $\mathrm{Reg}(\xi_1, \dots, \xi_{r_1-1})$. This is nonzero, and thus we can find a solution ν_v. Note that we can choose them to be purely imaginary since all right hand sides are purely imaginary and since $\log|\xi_j|_v$ are real.

Now consider $\chi_D(a)/\prod_{v|\infty}|a|_v^{\nu_v}$ for $a \in \mathcal{O}_F^\times$. Since a^2 is totally real, this squares to 1, and thus is a function $\mathcal{O}_F^\times/\mathcal{O}_F^{\times,\text{tot real}} \to \pm 1$. Now we can choose ϵ_v such that this function is $\prod_{v|\infty}\operatorname{sgn}_v$, since $\mathcal{O}_F^\times/\mathcal{O}_F^{\times,+} \xrightarrow{\prod_v \operatorname{sgn}_v} \{\pm 1\}^{r_1}$. □

Definition 1.2. A unitary Hecke character of conductor $\mathfrak{n}$ and infinity type χ_∞ is a character $\chi_0 \colon I(\mathfrak{n}) \to S^1$ which in principal ideals is given by $\chi_0((a)) = \chi_D(a)\chi_\infty^{-1}(a)$ for χ_D of conductor $\mathfrak{n}$. It's L-function is given by

$$L(s,\chi_0) = \sum_{\mathfrak{a}\subseteq\mathcal{O}_F} \frac{\chi_0(\mathfrak{a})}{\operatorname{Nm}(\mathfrak{a})^s} = \prod_{\mathfrak{p}} \left(1 - \frac{\chi_0(\mathfrak{p})}{\operatorname{Nm}(\mathfrak{p})^s}\right)^{-1}, \quad \operatorname{Re}(s) > 1. \tag{1.15}$$

Remark 1.1. One can also consider non-unitary Hecke characters as above. The only change is that the L-function converges in some region $\operatorname{Re}(s) > \sigma$ where σ depends on the growth of χ_∞.

Theorem 1.3. *For a Hecke character χ_0 of conductor $\mathfrak{n}$, we define $L(s,\chi_\infty) = \prod_{v|\infty} L(s,\chi_v)$ as*

$$L(s,\chi_v) = \begin{cases} \Gamma_{\mathbb{R}}(s+\nu_v+\epsilon) & \text{if } v \text{ is real,} \\ \Gamma_{\mathbb{C}}(s+\nu_v+|n_v|/2) & \text{if } v \text{ is complex,} \end{cases} \tag{1.16}$$

where $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$ and $\Gamma_{\mathbb{C}}(s) = (2\pi)^{-s}\Gamma(s)$. Then $\Lambda(s,\chi_0) := L(s,\chi_0)L(s,\chi_\infty)$ extends meromorphically for all s, and we have the functional equation

$$\Lambda(1-s,\chi_0^{-1}) = \epsilon(\chi_0,s)\Lambda(s,\chi_0) \tag{1.17}$$

where $\epsilon(\chi_0,s) = (D_F \cdot \operatorname{Nm}(\mathfrak{n}))^{s-1/2}\epsilon(\chi_0)$ for some nonzero constant $\epsilon(\chi_0)$ of absolute value 1.

This was first proved by Hecke in a similar way to above. Instead, we will discuss parts of the proof following Iwasawa–Tate. This is often reffered to as *Tate's thesis*. We will work adelically.

Proposition 1.4. *There is a bijection between Hecke characters χ_0 and continuous characters $\chi \colon F^\times\backslash\mathbb{A}_F^\times \to \mathbb{C}^\times$ that matches conductors $\mathfrak{n}$ and infinity types χ_∞. For $\mathfrak{p} \nmid \mathfrak{n}$, we have that $\chi_{\mathfrak{p}}(\varpi_{\mathfrak{p}}) = \chi_0(\mathfrak{p})$.*

Remark 1.2. Here, the conductor of χ is defined as a product of local conductors, as follows. Consider $\chi_{\mathfrak{p}} \colon F_{\mathfrak{p}}^\times \to \mathbb{C}^\times$. If $\chi_{\mathfrak{p}}$ is unramified (i.e. $\chi_{\mathfrak{p}}|_{\mathcal{O}_{F_{\mathfrak{p}}}^\times} = 1$) then the local conductor is 1. Otherwise, let n be the smallest positive integer such that $\chi_{\mathfrak{p}}|_{1+\mathfrak{p}^n}$ is trivial, and then the local conductor is $\mathfrak{p}^n$.

Proof. We discuss how one constructs χ from χ_0.

We denote $\mathbb{A}_F^{\mathfrak{n},\times} \subseteq \mathbb{A}_F^\times$ the ideles away from $\mathfrak{n}$, embedded by

$$(\alpha_v)_{v\nmid\mathfrak{n}} \mapsto (\alpha_v)_v \tag{1.18}$$

where $\alpha_\mathfrak{p} = 1$ for $\mathfrak{p} \mid \mathfrak{n}$.

Given a Hecke character χ_0, we define the character

$$\chi^\mathfrak{n} \colon \mathbb{A}_F^{\mathfrak{n},\times} \to \mathbb{C}^\times \tag{1.19}$$

given by the unramified characters $\chi_\mathfrak{p}^\mathfrak{n}(\varpi_\mathfrak{p}) = \chi_0(\mathfrak{p})$ for $\mathfrak{p}$ non-archimedean, and by $\chi_\infty^\mathfrak{n} = \chi_\infty$ for the archimedean places. We note that $\chi^\mathfrak{n}$ is invariant by

$$F^{\mathfrak{n},1} = \{\alpha/\beta \in F^\times \colon \alpha, \beta \in \mathcal{O}_F \text{ coprime with } \mathfrak{n},\ \alpha \equiv \beta \mod \mathfrak{n}\}. \tag{1.20}$$

This is because if $\alpha \in \mathcal{O}_F$ is coprime to $\mathfrak{n}$, we have

$$\chi^\mathfrak{n}(\alpha) = \chi_\infty(\alpha) \prod_{\mathfrak{p}\nmid\mathfrak{n}} \chi_\mathfrak{p}(\varpi_\mathfrak{p})^{\nu_\mathfrak{p}(\alpha)} = \chi_\infty(\alpha) \prod_{\mathfrak{p}} \chi_0(\mathfrak{p})^{\nu_\mathfrak{p}(\alpha)} = \chi_\infty(\alpha)\chi_0((\alpha)) = \chi_D(\alpha). \tag{1.21}$$

Thus we have a character $\chi^\mathfrak{n} \colon F^{\mathfrak{n},1}\backslash\mathbb{A}_F^{\mathfrak{n},\times} \to \mathbb{C}^\times$.

Now we use that $F^\times \mathbb{A}_F^{\mathfrak{n},\times}$ is dense in $\mathbb{A}_F^\times$ by weak approximation. This implies that $\chi^\mathfrak{n}$ extends uniquely to a character $\chi \colon F^\times\backslash\mathbb{A}_F^\times$ with conductor dividing $\mathfrak{n}$. □

Let's briefly discuss how this interacts with the Langlands conjectures. Very roughly, we expect a relationship

$$\{\text{Hecke characters}\} \dashleftarrow\dashrightarrow \{1\text{-dim Galois representations}\} \tag{1.22}$$

and this is obtained by Class Field Theory and the above proposition: the Artin map $F^\times\backslash\mathbb{A}_F^\times \to \mathrm{Gal}(\bar{F}/F)^{ab}$ identifies $\mathrm{Gal}(\bar{F}/F)^{ab}$ with the profinite completion of $F^\times\backslash\mathbb{A}_F^\times$. In particular, since every homomorphism $\mathrm{Gal}(\bar{F}/F) \to \mathbb{C}^\times$ has finite image, we have an identification

$$\{\text{finite order Hecke characters over } F\} \longleftrightarrow \{\rho \colon \mathrm{Gal}(\bar{F}/F) \to \mathbb{C}^\times\}. \tag{1.23}$$

It is more interesting, however, to also consider characters $\rho_\ell \colon \mathrm{Gal}(\bar{F}/F) \to \bar{\mathbb{Q}}_\ell^\times$, since these can have infinite image. With a bit more work, one can also prove

$$\{\text{algebraic Hecke characters over } F\} \longleftrightarrow \left\{ \begin{array}{c} \text{compatible systems of characters} \\ \rho_\ell \colon \mathrm{Gal}(\bar{F}/F) \to \bar{\mathbb{Q}}_\ell^\times \text{ for all primes } \ell \end{array} \right\}. \tag{1.24}$$

Here, *algebraic* is the condition that χ_∞ restricted to the identity component is an algebraic map of $\mathbb{R}$-varieties. It is a general phenomena that an algebraicity condition at ∞ is what should capture the automorphic forms which correspond to compatible systems of Galois representations under Langlands correspondence.

1.3 Tate's thesis

Reference: [Bum97, Section 3.1].

For nontrivial Dirichlet characters $\chi_D\colon (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$, the proof of the functional equation relied on the integral representation

$$\Lambda(s,\chi) = \int_0^\infty \sum_{n\geqslant 1} n^\epsilon \chi_D(n) e^{-\pi n^2 t} t^{(s+\epsilon)/2} \frac{\mathrm{d}t}{t} = \int_0^\infty \sum_{n\geqslant 1} \chi_0((n))\chi_\infty(n)\Phi_\infty(nt^{1/2}) t^{s/2}\frac{\mathrm{d}t}{t} \tag{1.25}$$

for $\Phi_\infty(x) = x^\epsilon e^{-\pi x^2}$. Changing $t \mapsto t^2$ and then $t \mapsto t/n$ for each n, we have

$$\Lambda(s,\chi) = \int_{\mathbf{R}^\times} \sum_{n\geqslant 1} \chi_0((n))\chi_\infty(nt)\Phi_\infty(nt)|t|^s \,\mathrm{d}^\times t = \left(\int_{\mathbf{R}^\times} \Phi_\infty(t)\chi_\infty(t)|t|^s\,\mathrm{d}^\times t.\right)\cdot\left(\sum_{n\geqslant 1}\chi_0((n))|n|^{-s}\right). \tag{1.26}$$

Now we can write

$$\sum_{n\geqslant 1}\chi_0((n))|n|^{-s} = \prod_{p\nmid N}\left(\sum_{k\geqslant 0}\chi_p(p^k)|p^k|_p^s\right) = \prod_{p\nmid N}\int_{\substack{\mathbb{Q}_p^\times \\ |x|_p\leqslant 1}} \chi_p(x)|x|_p^s\,\mathrm{d}^\times x. \tag{1.27}$$

Hence, for $\Phi = \Phi_\infty \times \prod_{p\nmid N}\mathrm{char}(\mathbb{Z}_p)$ we have

$$\Lambda(s,\chi) = \int_{\mathbb{A}_{\mathbb{Q}}^{N,\times}} \Phi(x)\chi(x)|x|^s\,\mathrm{d}^\times x, \tag{1.28}$$

where $|x| = \prod_v |x_v|_v$.

Now let F be a number field. We consider the usual Haar measure $\mathrm{d}x = \prod_v \mathrm{d}x_v$ on $\mathbb{A}_F$. We consider the Haar measure $\mathrm{d}^\times x$ on $\mathbb{A}_F^\times$ given by

$$m_v = \begin{cases} 1 & \text{if } v \text{ is archimedean,} \\ (1-\frac{1}{q_v})^{-1} & \text{otherwise.} \end{cases} \tag{1.29}$$

These are such that $\mathrm{d}x_v(\mathcal{O}_{F_v}) = 1$ and $\mathrm{d}^\times x_v(\mathcal{O}_{F_v}^\times) = 1$ for non-archimedean places v.

Recall we have the absolute value $|\cdot|\colon \mathbb{A}_F^\times \to \mathbb{R}^\times$ given by $|x| = \prod_v |x_v|_v$, which is $F^\times$-invariant. This is such that $\mathrm{d}(\alpha x) = |\alpha|\,\mathrm{d}x$. Let denote $\mathbb{A}_F^{\times,1}$ the ideles of norm 1, i.e. the kernel of $|\cdot|$. Recall that $F^\times\backslash\mathbb{A}_F^{\times,1}$ is compact.

As in the classical proof, we will rely on a Poisson summation formula.

Definition 1.3. We consider the Schwartz space $\mathcal{S}(\mathbb{A}_F)$ to be finite linear combinations of functions of the form $\Phi(x) = \prod_v \Phi_v(x_v)$ where $\Phi_v \in \mathcal{S}(F_v)$ and where all but finitely many are $\mathrm{char}(\mathcal{O}_{F_v})$. For such v with $\Phi_v = \mathrm{char}(\mathcal{O}_{F_v})$, we say Φ is unramified at v.

Proposition 1.5. *For $\Phi \in \mathcal{S}(\mathbb{A}_F)$ and a choice of additive character $\psi\colon F\backslash\mathbb{A}_F \to \mathbb{C}$, we consider the Fourier transform*

$$\hat{\Phi}(x) = \int_{\mathbb{A}_F} \Phi(y)\psi(xy)\,\mathrm{d}y. \tag{1.30}$$

This is also in $\mathcal{S}(\mathbb{A}_F)$, and for $t \in \mathbb{A}_F^\times$ we have the Poisson summation formula

$$\sum_{\alpha\in F} \Phi(\alpha t) = \frac{1}{|t|}\sum_{\alpha\in F} \hat{\Phi}\left(\frac{\alpha}{t}\right). \tag{1.31}$$

Definition 1.4. For χ a Hecke chatacter and $\Phi \in \mathcal{S}(\mathbb{A}_F)$, their *zeta integral* is

$$\zeta(s,\chi,\Phi) = \int_{\mathbb{A}_F^\times} \Phi(x)\chi(x)|x|^s\,\mathrm{d}^\times x. \tag{1.32}$$

Note that we may write $\zeta(s,\chi,\Phi) = \zeta(0,\chi|\cdot|^s,\Phi)$, which we simply denote $\zeta(\chi|\cdot|^s,\Phi)$.

Formally, we have

$$\zeta(s,\chi,\Phi) = \prod_v \zeta_v(s,\chi_v,\Phi_v), \quad \text{where} \quad \zeta_v(s,\chi_v,\Phi_v) = \int_{F_v^\times} \Phi_v(x)\chi_v(x)|x|_v^s\,\mathrm{d}^\times x_v. \tag{1.33}$$

Roughly, what will happen is that the local L factors $L_v(s,\chi_v)$ will be the "gcd" of all local zeta integrals $\zeta_v(s,\chi_v,\Phi_v)$.

Remark 1.3. A non-unitary Hecke character can always be written as $\chi|\cdot|^\sigma$ for some $\sigma \in \mathbb{R}$ and some χ unitary. This means that it suffices to understand $\zeta(s,\chi,\Phi)$ for unitary χ, as $\zeta(s,\chi|\cdot|^\sigma,\Phi) = \zeta(s+\sigma,\chi,\Phi)$.

On what follows, χ will be a unitary Hecke character and $\Phi \in \mathcal{S}(\mathbb{A}_F)$, unless noted otherwise.

Proposition 1.6. *The local zeta integrals $\zeta_v(s,\chi_v,\Phi_v)$ define holomorphic functions for $\mathrm{Re}(s) > 0$. If $\chi_\mathfrak{p}, \Phi_\mathfrak{p}$ are unramified, then $\zeta_\mathfrak{p}(s,\chi_\mathfrak{p},\Phi_\mathfrak{p}) = (1-\chi(\mathfrak{p})q_\mathfrak{p}^{-s})^{-1}$.*

Proof. To prove convergence, it suffices to see that $\int_{F_v^\times} |\Phi_v(x)||x|_v^s\,\mathrm{d}^\times x$ converges. Over $|x|_v > 1$, the rapid decay of Φ_v makes it so that it converges for all s. Over the compact region $|x|_v \leqslant 1$, Φ_v is bounded, and thus it suffices to consider $\int_{|x|_v\leqslant 1}|x|_v^s\,\mathrm{d}^\times x$.

For $v = \mathfrak{p}$ non-archimedean, this is

$$\int_{|x|_\mathfrak{p} \leqslant 1} |x|_\mathfrak{p}^s \, \mathrm{d}^\times x = \sum_{k \geqslant 0} \int_{|x|_\mathfrak{p} = q_\mathfrak{p}^{-k}} |x|_\mathfrak{p}^s \, \mathrm{d}^\times x = \sum_{k \geqslant 0} q_\mathfrak{p}^{-ks} \tag{1.34}$$

which converges for $\mathrm{Re}(s) > 0$. A similar computation proves the last claim for $\chi_\mathfrak{p}, \Phi_\mathfrak{p}$ unramified.

For v real we are looking at $\int_{-1}^1 |t|^s \frac{\mathrm{d}t}{|t|}$ and for v complex we are looking at $\frac{1}{2\pi} \int_0^{2\pi} \int_0^1 r^{2s} \frac{\mathrm{d}r}{r} \, \mathrm{d}\theta$, both of which converge for $\mathrm{Re}(s) > 0$. □

Proposition 1.7. *$\zeta(s, \chi, \Phi)$ defines a holomorphic function for $\mathrm{Re}(s) > 1$, and the product decomposition holds for such s.*

Proof. This follows from the above, by comparison to the Dedekind zeta function. □

Proposition 1.8. *The local zeta integrals have meromorphic continuation to all $s \in \mathbb{C}$. There exists a meromorphic function $\gamma_v(s, \chi_v, \psi_v)$ (independent of Φ_v!) such that*

$$\zeta_v(1 - s, \chi_v^{-1}, \hat{\Phi}_v) = \gamma_v(s, \chi_v, \psi_v) \zeta_v(s, \chi_v, \Phi_v). \tag{1.35}$$

Proof. First we prove that such γ_v exists for $\mathrm{Re}(s) \in (0, 1)$. For this, we need to see that the ratios $\frac{\zeta_v(1-s, \chi_v^{-1}, \hat{\Phi}_v)}{\zeta_v(s, \chi_v, \Phi_v)}$ are independent of Φ. By changing χ, we assume $s = 0$ for notational simplicity. Expanding, we have

$$\zeta_v(\chi_v^{-1}|\cdot|, \hat{\Phi}_v) \zeta_v(\chi_v, \Phi_v') = \int_{F_v^\times} \left(\int_{F_v} \Phi_v(y) \psi(xy) \, \mathrm{d}y \right) \chi_v(x)^{-1} |x|_v \, \mathrm{d}^\times x \int_{F_v^\times} \Phi_v'(z) \chi_v(z) \, \mathrm{d}^\times z \tag{1.36}$$

and making the substitution $x \mapsto y^{-1}x$, this becomes

$$\frac{1}{m_v} \int_{F_v^\times} \int_{F_v^\times} \int_{F_v^\times} \Phi_v(y) \Phi_v'(z) \chi_v(yz) \psi(x) \chi_v(x)^{-1} |x|_v \, \mathrm{d}^\times x \, \mathrm{d}^\times y \, \mathrm{d}^\times z. \tag{1.37}$$

which is symmetric in Φ_v, Φ_v'. This proves that $\gamma_v(s, \chi_v, \psi_v)$ does not depend on Φ_v.

Now taking Φ_v resp. $\hat{\Phi}_v$ to vanish in a neighborhood of zero, the integrals $\zeta_v(s, \chi_v, \Phi_v)$ resp. $\zeta_v(1-s, \chi_v^{-1}, \hat{\Phi}_v)$ define holomorphic functions for all s. This implies that $\gamma_v(s, \chi_v, \psi_v)$ is meromorphic for $\mathrm{Re}(s) < 1$ resp. $\mathrm{Re}(s) > 0$. □

Proposition 1.9. *$\zeta(s, \chi, \Phi)$ has analytic continuation to all s, and is entire unless $\chi|_{\mathbb{A}_F^{\times,1}}$ is trivial. We have the functional equation*

$$\zeta(s, \chi, \Phi) = \zeta(1 - s, \chi^{-1}, \hat{\Phi}). \tag{1.38}$$

Proof. We write $\zeta(s,\chi,\Phi)=\zeta_1(s,\chi,\Phi)+\zeta_0(s,\chi,\Phi)$ where

$$\zeta_1(s,\chi,\Phi)=\int_{\substack{\mathbb{A}_F^\times\\|x|>1}}\Phi(x)\chi(x)|x|^s\,\mathrm{d}^\times x,\quad \zeta_0(s,\chi,\Phi)=\int_{\substack{\mathbb{A}_F^\times\\|x|<1}}\Phi(x)\chi(x)|x|^s\,\mathrm{d}^\times x. \tag{1.39}$$

By the rapid decay of Φ, the term $\zeta_1(s,\chi,\Phi)$ is entire. Now we write

$$\zeta_0(s,\chi,\Phi)=\sum_{\alpha\in F^\times}\int_{\substack{F^\times\backslash\mathbb{A}_F^\times\\|x|<1}}\Phi(\alpha x)\chi(x)|x|^s\,\mathrm{d}^\times x=\int_{\substack{F^\times\backslash\mathbb{A}_F^\times\\|x|<1}}\left(\sum_{\alpha\in F^\times}\Phi(\alpha x)\right)\chi(x)|x|^s\,\mathrm{d}^\times x. \tag{1.40}$$

We assume for simplicity that $\chi|_{\mathbb{A}_F^{\times,1}}$ is nontrivial (otherwise one needs to be a bit more careful with the poles). Then

$$\int_{\substack{F^\times\backslash\mathbb{A}_F^\times\\|x|<1}}\chi(x)|x|^s\,\mathrm{d}^\times x=\int_0^1\int_{\substack{F^\times\backslash\mathbb{A}_F^\times\\|x|=t}}\chi(x)|x|^s\,\mathrm{d}^\times x\frac{\mathrm{d}t}{t}=\int_0^1\chi(x_t)t^s\frac{\mathrm{d}t}{t}\int_{F^\times\backslash\mathbb{A}_F^{\times,1}}\chi(x)\,\mathrm{d}^\times x \tag{1.41}$$

for some section x_t of $|\cdot|\colon\mathbb{A}_F^\times\to\mathbb{R}_+$, and thus this is zero. So we may write

$$\zeta_0(s,\chi,\Phi)=\int_{\substack{F^\times\backslash\mathbb{A}_F^\times\\|x|<1}}\left(\sum_{\alpha\in F}\Phi(\alpha x)\right)\chi(x)|x|^s\,\mathrm{d}^\times x \tag{1.42}$$

which by Poisson summation is

$$\zeta_0(s,\chi,\Phi)=\int_{\substack{F^\times\backslash\mathbb{A}_F^\times\\|x|<1}}\left(\frac{1}{|x|}\sum_{\alpha\in F}\hat{\Phi}\left(\frac{\alpha}{x}\right)\right)\chi(x)|x|^s\,\mathrm{d}^\times x. \tag{1.43}$$

Making the change of variables $x\mapsto 1/x$, this becomes simply $\zeta_1(1-s,\chi^{-1},\Phi)$. Thus

$$\zeta(s,\chi,\Phi)=\zeta_1(s,\chi,\Phi)+\zeta_1(1-s,\chi^{-1},\hat{\Phi}), \tag{1.44}$$

which proves that $\zeta(s,\chi,\Phi)$ is entire and also proves the functional equation. □

Corollary 1.10. *If S denotes a set of places containing all archimedean ones and the ones that χ or ψ are ramified, then the partial L-function*

$$L^S(s,\chi):=\prod_{v\notin S}L_v(s,\chi_v) \tag{1.45}$$

extends to a meromorphic function to all s, and satisfy the functional equation

$$L^S(s,\chi)=\left(\prod_{v\in S}\gamma_v(s,\chi_v,\psi_v)\right)L^S(1-s,\chi^{-1}) \tag{1.46}$$

In particular, the completed $\Gamma(s,\chi)$ is meromorphic for all s, and if we further define

$$\epsilon_v(s,\chi_v,\psi_v) := \frac{\gamma_v(s,\chi_v,\psi_v)L_v(s,\chi_v)}{L_v(1-s,\chi_v^{-1})} = \frac{\zeta_v(1-s,\chi_v^{-1},\hat{\Phi}_v)}{L_v(1-s,\chi_v^{-1})}\cdot\left(\frac{\zeta_v(s,\chi_v,\Phi_v)}{L_v(s,\chi_v)}\right)^{-1}, \tag{1.47}$$

then we have the full functional equation

$$\Lambda(s,\chi) = \epsilon(s,\chi)\Lambda(1-s,\chi^{-1}) \tag{1.48}$$

for $\epsilon(s,\chi) = \prod_v \epsilon_v(s,\chi_v,\psi_v)$. As the next proposition shows, we can in fact prove that $\epsilon_v(s,\chi_v,\psi_v)$ are of *exponential type*, i.e. of the form AB^s for some $A, B \in \mathbb{C}^\times$. Ultimately, this is done by a computation with explicitly chosen Φ_v.

Proposition 1.11. *Consider $\Phi_v \in \mathcal{S}(F_v)$, and if v is archimedean we assume it is real analytic at 0. Then the function $\frac{\zeta_v(s,\chi_v,\Phi_v)}{L_v(s,\chi_v)}$ is entire. Moreover, there is a choice of Φ_v such that both this ratio and $\epsilon(s,\chi_v)$ are of exponential type.*

Proof. We prove the first part. The second part is proved by choosing particular functions Φ_v, and we leave it as an exercise.

If $v = \mathfrak{p}$ is non-archimedean, we have

$$\zeta_\mathfrak{p}(s,\chi_\mathfrak{p},\Phi_\mathfrak{p}) = \sum_{k\in\mathbb{Z}} q^{ks}\int_{|x|_\mathfrak{p}=q_\mathfrak{p}^k} \Phi_\mathfrak{p}(x)\chi_\mathfrak{p}(x)\,\mathrm{d}^\times x. \tag{1.49}$$

Since Φ_v is compactly supported, the contribution for k large is 0. Since Φ_v is locally constant, the contribution for k small enough is $\int_{|x|_\mathfrak{p}=q_\mathfrak{p}^k}\Phi_\mathfrak{p}(0)\chi_\mathfrak{p}(x)\,\mathrm{d}^\times x$. If $\chi_\mathfrak{p}$ is ramified, this is zero, and thus $\zeta_\mathfrak{p}(s,\chi_\mathfrak{p},\Phi_\mathfrak{p})$ is a finite sum. Similarly, if $\chi_\mathfrak{p}$ is unramified, then the difference $\zeta_\mathfrak{p}(s,\chi_\mathfrak{p},\Phi_\mathfrak{p}) - \Phi_\mathfrak{p}(0)\zeta_\mathfrak{p}(s,\chi_\mathfrak{p},\mathrm{char}(\mathcal{O}_{F_\mathfrak{p}}))$ is a finite sum, and the claim follows.

If v is archimedean, say Φ_v is real analytic for $|x|_v < \varepsilon$. Since $\int_{|x|_v>\varepsilon}\Phi(x)\chi_v(x)|x|_v^s\,\mathrm{d}^\times x$ extends holomorphically to all s, we need to understand the poles of

$$\int_{|x|_v<\varepsilon}\Phi(x)\chi_v(x)|x|_v^s\,\mathrm{d}^\times x. \tag{1.50}$$

For v real, consider the Taylor expansion $\Phi_v(x) = \sum_{n\geqslant 0} a_n x^n$. Then we are looking at

$$\int_0^\varepsilon \sum_{n\geqslant 0} a_n x^n |x|_v^s\,\mathrm{d}^\times x = 2\int_0^\varepsilon \sum_{\substack{n\geqslant 0\\ 2|n+\epsilon_v}} a_n x^{n+s+\nu_v}\frac{\mathrm{d}x}{x} \tag{1.51}$$

and the poles are such that $n+s+\nu_v = 0$ for some n with $2 \mid n+\epsilon_v$ and $a_n \neq 0$. Except for the condition $a_n \neq 0$, these are the same poles as $\Gamma_\mathbb{R}(s+\nu_v+\epsilon_v) = \pi^{-(s+\nu_v+\epsilon_v)/2}\Gamma\left(\frac{s+\nu_v+\epsilon_v}{2}\right)$.

For v complex, we are looking at

$$\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\varepsilon} r^{2s+2\nu_v} e^{in_v\theta}\Phi_v(re^{i\theta})\,\mathrm{d}^\times r\,\mathrm{d}\theta = \int_0^{\varepsilon} r^{2s+2\nu_v}\phi(r)\frac{\mathrm{d}r}{r} \tag{1.52}$$

where

$$\phi(r) = \frac{1}{2\pi}\int_0^{2\pi} e^{in_v\theta}\Phi_v(re^{i\theta})\,\mathrm{d}\theta. \tag{1.53}$$

If $\Phi_v(x) = \sum_{n,m\geqslant 0} a_{n,m}x^n\bar{x}^m$ is the Taylor expansion at 0, then

$$\phi(r) = \frac{1}{2\pi}\int_0^{2\pi}\sum_{n,m\geqslant 0} a_{n,m}e^{i\theta(n_v+n-m)}r^{n+m} = \sum_{\substack{n,m\geqslant 0\\ m-n=n_v}} a_{n,m}r^{n+m} \tag{1.54}$$

So the possible poles are at $n+m+2s+2\nu_v = 0$ for some $m,n\geqslant 0$ with $m-n=n_v$ and $a_{n,m}\neq 0$. Except for the condition $a_{n,m}\neq 0$, these are precisely the poles of $\Gamma_{\mathbb{C}}(s+\nu_v+\frac{|n_v|}{2}) = \pi^{-(s+\nu_v+\frac{|n_v|}{2})}\Gamma(s+\nu_v+\frac{|n_v|}{2})$. □

2 Modular forms

2.1 First definitions

Reference: Parts of [DS05, Sections 1.1, 1.2].

Recall that Riemann's proof of the functional equation for $\zeta(z)$ relies on the integral representation

$$\xi(s) = \int_0^\infty \frac{(\theta(t)-1)}{2}t^{s/2}\frac{\mathrm{d}t}{t}, \quad \text{where} \quad \theta(z) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 z} \text{ for } \mathrm{Re}(z)>0, \tag{2.1}$$

and on the following functional equation, which is a consequence of Poisson summation formula

$$\theta(z) = \frac{1}{\sqrt{z}}\theta(1/z). \tag{2.2}$$

If we write $\tau = iz$, then $\theta(\tau)$ has a Fourier expansion

$$\theta(\tau) = \sum_{n\in\mathbb{Z}} q^{n^2/2}, \quad \text{for } q = e^{2\pi i\tau} \tag{2.3}$$

and satisfies

$$\theta(-1/\tau) = \sqrt{\frac{\tau}{i}}\cdot\theta(\tau). \tag{2.4}$$

So one way to construct functions like $\zeta(s)$ is to look at functions $f\colon \mathcal{H}\to\mathbb{C}$ satisfying

1. $f(\tau+1) = f(\tau)$
2. $f(\tau) = \tau^k f(-1/\tau)$ for some $k \in \mathbb{Z}$.

If f is also meromorphic, we will see later that (1) implies that f has a Fourier expansion $f(\tau) = \sum_{n\in\mathbb{Z}} a_n q^n$. In the case that this is concentrated at $n > 0$, we formally "define"

$$L_f(s) := \int_0^\infty f(it) t^s \frac{\mathrm{d}t}{t} \tag{2.5}$$

then we have $L_f(s) = (2\pi)^{-s}\Gamma(s)\sum_{n>0}\frac{a_n}{n^s}$ and we can "prove" that

$$L_f(s) = (*)\cdot L_f(k-s). \tag{2.6}$$

Definition 2.1. We consider the action of $\mathrm{GL}_2(\mathbb{R})^+$ in $\mathcal{H} := \{z \in \mathbb{C}\colon \mathrm{Im}(z) > 0\}$ given by the projective transformations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d}. \tag{2.7}$$

Note that this is well defined since if $\tau = x + iy$ then

$$\mathrm{Im}(\gamma\tau) = \frac{ay(cx+d) - cy(ax+b)}{|c\tau+d|^2} = \frac{\mathrm{Im}(\tau)\det(\gamma)}{|c\tau+d|^2}. \tag{2.8}$$

Definition 2.2. A meromorphic function $f\colon \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight* k for some integer k if it satisfies

$$f\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\cdot\tau\right) = (c\tau+d)^k f(\tau) \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ \tau \in \mathcal{H}. \tag{2.9}$$

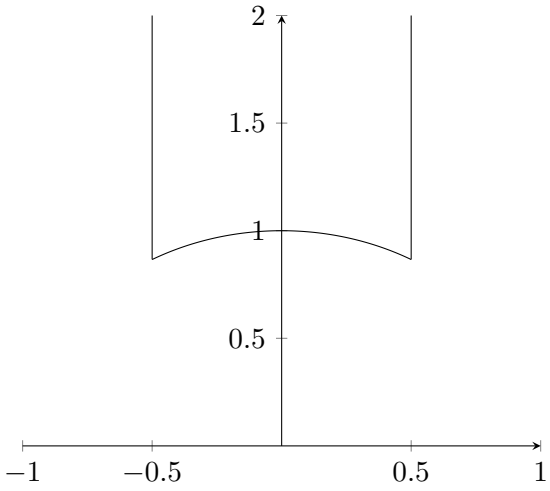
We denote $\nu_k\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tau\right) = (c\tau+d)^k$ the *automorphy factor of weight* k.

Remark 2.1. Note that $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ generate the subgroup $\mathrm{SL}_2(\mathbb{Z})$. We can also check that $\nu_k(\gamma,\tau)$ satisfy the cocycle condition

$$\nu_k(\gamma_1\gamma_2, z) = \nu_k(\gamma_1, \gamma_2 z)\nu_k(\gamma_2, z). \tag{2.10}$$

These mean that the condition $f(\gamma\tau) = \nu_k(\gamma,\tau)f(\tau)$ for all τ is equivalent to requiring it only for $\gamma \in \{T, S\}$.

The $\mathrm{SL}_2(\mathbb{Z})$-orbits of $\mathcal{H}$ have the fundamental domain $\{\mathrm{Re}(\tau) \in [-1/2, 1/2]\} \cap \{|\tau| > 1\}$:



Proposition 2.1. *Let $f\colon \mathcal{H} \to \mathbb{C}$ be a meromorphic function which satisfies $f(\tau+1) = f(\tau)$. Then f has a Fourier expansion*

$$f(\tau) = \sum_{n\in\mathbb{Z}} a_n q^n, \quad q = e^{2\pi i\tau}. \tag{2.11}$$

Proof. By Fourier, $f(\tau)$ has an expansion

$$f(\tau) = \sum_{n\in\mathbb{Z}} a_n(y) q^n, \quad q = e^{2\pi i\tau} \tag{2.12}$$

where we write $\tau = x+iy$. Then $\frac{\partial f}{\partial x} = \sum_{n\in\mathbb{Z}} 2\pi i n a_n(y) q^n$ and $\frac{\partial f}{\partial y} = \sum_{n\in\mathbb{Z}} (-2\pi a_n(y) q^n + a_n'(y) q^n)$ and hence Cauchy–Riemann equations are equivalent to $\sum_{n\in\mathbb{Z}} a_n'(y) q^n = 0$. Thus $a_n(y)$ are constant. □

Definition 2.3. We say that f as above is *holomorphic at ∞* if one of the equivalent conditions is satisfied:

1. the Fourier expansion $f(\tau) = \sum_{n\in\mathbb{Z}} a_n q^n$ is concentrated on $n \geqslant 0$,
2. f, as a function of q, extends holomorphically to the puncture point $q = 0$,
3. $\lim_{\mathrm{Im}(\tau)\to\infty} f(\tau)$ exists,
4. $f(\tau)$ is bounded as $\mathrm{Im}(\tau) \to \infty$.

Definition 2.4. A *modular form of weight k* is a holomorphic function $f\colon \mathcal{H} \to \mathbb{C}$ which is weakly modular of weight k and holomorphic at ∞. We denote by $M_k(\mathrm{SL}_2(\mathbb{Z}))$ the space of such functions. We say that f is a *cusp form of weight k* if moreover we have $a_0(f) = 0$, and denote by $S_k(\mathrm{SL}_2(\mathbb{Z}))$ the space of such functions.

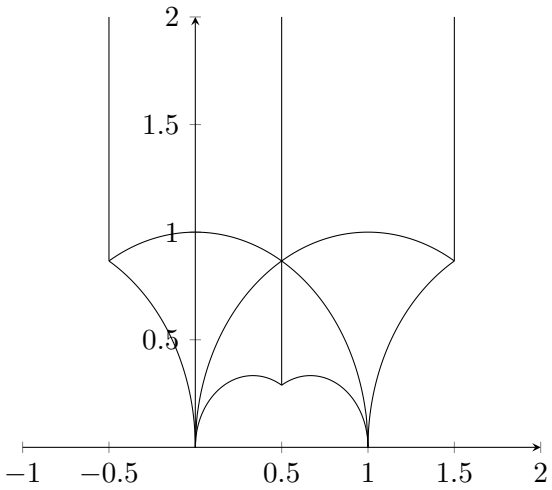
For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})^+$ and $f\colon \mathcal{H} \to \mathbb{C}$, we denote

$$(f[\gamma]_k)(\tau) = \det(\gamma)^{k-1}(c\tau + d)^{-k} f(\gamma\tau). \tag{2.13}$$

This is so that the weakly modular condition is equivalent to having $f[\gamma]_k = f$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. More generally, we define

Definition 2.5. For $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ a subgroup, a meromorphic function $f\colon \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight k and level Γ* if $f[\gamma]_k = f$ for all $\gamma \in \Gamma$.

Example 2.1. For $\Gamma = \Gamma(2) := \{A \equiv I_2 \mod 2\}$, a fundamental domain is



We note that there are three cusps: at ∞, at 0 and at 1. We want to make sense of holomorphicity at these cusps.

For $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ a finite index subgroup, there exists $N > 0$ such that $\begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \in \Gamma$. Then considering $q_N := e^{2\pi i\tau/N}$, any weakly modular form of level Γ has a Fourier expansion $f(\tau) = \sum_{n\in\mathbb{Z}} a_{n,N} q_N^n$ as above, and we can use this to make sense of holomorphicity at ∞: namely that this is concentrated on $n \geqslant 0$. For other cusps $c \in \mathbb{Q}$, we can find $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(c) = \infty$, and we consider $f[\gamma]_k$, which is weakly modular for $\gamma^{-1}\Gamma\gamma$. We then say that f is holomorphic at c if $f[\gamma]_k$ is holomorphic at ∞.

Definition 2.6. For $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ a finite index subgroup, a *modular form of weight k and level Γ* is holomorphic function $f\colon \mathcal{H} \to \mathbb{C}$ which is weakly modular of weight k and level Γ such that $f[\gamma]_k$ is holomorphic at ∞ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. We denote $M_k(\Gamma)$ the space of such functions. We say that f is a *cusp form of weight k and level Γ* if moreover we have $a_0(f[\gamma]_k) = 0$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, and we denote $S_k(\Gamma)$ the space of such functions.

Remark 2.2. We note that $[\gamma]_k \colon M_k(\Gamma) \to M_k(\gamma^{-1}\Gamma\gamma)$ and $[\gamma]_k \colon S_k(\Gamma) \to S_k(\gamma^{-1}\Gamma\gamma)$.

Example 2.2. For $\theta(\tau) = \sum_{n\in\mathbb{Z}} q_2^{n^2}$, we have that θ^2 satisfies

$$\theta^2(-1/\tau) = -i\tau\theta^2(\tau), \quad \theta^2(\tau+2) = \theta^2(\tau) \tag{2.14}$$

and is a modular form of weight 1 of level Γ where Γ is the subgroup generated by $\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 2 & 1\end{pmatrix}$. To see that, note that

$$\theta^2\left(\frac{1\cdot z+0}{2\cdot z+1}\right) = \frac{i(2z+1)}{z}\theta^2\left(-\frac{2z+1}{z}\right) = \frac{i(2z+1)}{z}\theta^2\left(-\frac{1}{z}\right) = (2z+1)\theta(z) \tag{2.15}$$

We have that $\Gamma \subseteq \Gamma(2)$ has index 2, as $\Gamma(2)$ is generated by Γ and $-I$. So θ^4 is modular of level $\Gamma(2)$. With some more care, one can make sense of $\theta(\tau)$ being a modular form of "weight $1/2$".

Now we tackle the question of convergence of $L_f(s) = \int_0^\infty f(it)t^s\frac{\mathrm{d}t}{t}$. Formally, if $f(\tau) = \sum_{n\in\mathbb{Z}} a_n q_N^n$, we have

$$L_f(s) = (2\pi)^{-s}\Gamma(s)\sum_{n\in\mathbb{Z}}\frac{a_n}{(n/N)^s} \tag{2.16}$$

Proposition 2.2. *Let $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ be a finite index subgroup and $f \in S_k(\Gamma)$. Choose $N > 0$ with $\begin{pmatrix}1 & N\\ 0 & 1\end{pmatrix} \in \Gamma$, so that f has Fourier expansion $f(\tau) = \sum_{n\geqslant 0} a_n(f)q_N^n$. Then there exists a constant $C > 0$ such that $|a_n(f)| \leqslant C(n/N)^{k/2}$ for all $n > 0$. In particular, $L_f(s)$ converges absolutely for $\mathrm{Re}(s) > \frac{k}{2} + 1$.*

Proof. First we see that $\mathrm{Im}(\tau)^{k/2}|f(\tau)|$ is Γ-invariant:

$$\mathrm{Im}(\gamma\tau)^{k/2}|f(\gamma\tau)| = \left(\frac{\mathrm{Im}(\tau)\det(\gamma)}{|c\tau+d|^2}\right)^{k/2}|(c\tau+d)^k f(\tau)| = \mathrm{Im}(\tau)|f(\tau)|. \tag{2.17}$$

Since f is holomorphic at all cusps, it is bounded near the cusps. Moreover, it decays exponentially at ∞ since f is a cusp form. This means that $\mathrm{Im}(\tau)^{k/2}f(\tau)$ is also bounded for $\mathrm{Im}(\tau)$ sufficiently large. Choosing a fundamental domain for the Γ action on $\mathcal{H}$, we can break it down as a union of a compact region and of regions near each cusp. The above discussion then implies that $\mathrm{Im}(\tau)^{k/2}|f(\tau)|$ is bounded for all τ, say

$$\mathrm{Im}(\tau)^{k/2}|f(\tau)| \leqslant C'. \tag{2.18}$$

Now we have

$$|a_n(f)| = \left| \frac{1}{N} \int_0^N f(x+iy) e^{-2\pi i n(x+iy)/N} \, \mathrm{d}x \right| \leqslant C' y^{-k/2} \int_0^N \frac{1}{N} |e^{-2\pi i n(x+iy)/N}| \, \mathrm{d}x = C' y^{-k/2} e^{2\pi n y/N}. \tag{2.19}$$

Picking out $y = N/n$ gives us $|a_n(f)| \leqslant C' e^{2\pi} (n/N)^{k/2}$. □

Corollary 2.3. *If $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$, then $L_f(s)$ extends to an entire function and satisfies $L_f(s) = (-1)^{k/2} \cdot L_f(k-s)$. (Note that k must be even since $-I \in \mathrm{SL}_2(\mathbb{Z})$ and $\nu_k(-I) = (-1)^k$)*

Proof. As usual, we write

$$L_f(s) = \int_1^\infty f(it) t^s \frac{\mathrm{d}t}{t} + \int_0^1 f(it) t^s \frac{\mathrm{d}t}{t} \tag{2.20}$$

and rewrite the second term as

$$\int_0^1 f(it) t^s \frac{\mathrm{d}t}{t} = \int_1^\infty f(-1/(it)) t^{-s} \frac{\mathrm{d}t}{t} = \int_1^\infty f(it) \cdot (it)^k t^{-s} \frac{\mathrm{d}t}{t}. \tag{2.21}$$

This means that

$$L_f(s) = \int_1^\infty f(it) t^s \frac{\mathrm{d}t}{t} + i^k \cdot \int_1^\infty f(it) t^{k-s} \frac{\mathrm{d}t}{t}. \tag{2.22}$$

Since $f(it)$ decays exponentially, this expression is entire for all s, and we see that $L_f(s) = i^k \cdot L_f(k-s)$. □

We will also later establish functional equations for certain modular forms of smaller levels, but we will have to work a bit harder. Eventually we will look at modular forms for the subgroups

$$\Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \colon c \equiv 0 \mod N \right\}. \tag{2.23}$$

2.2 Modular curves and dimension formulas

Reference: Parts of [DS05, Sections 2, 3].

For this subsection, we only work with modular forms of even weight. One can also do similar arguments for general weight, but we restrict to the even case for simplicity. In the case of even weight, we may assume without loss of generality that all $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ contain $\{\pm I\}$. **For this subsection, we will always assume that subgroups $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ are of finite index and contain $\{\pm I\}$.**

We will compute the dimension of the spaces $M_k(\Gamma)$ and $S_k(\Gamma)$ by relating them to algebraic geometry.

Definition 2.7. For $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$, we consider the *modular curve* $Y(\Gamma) = \Gamma\backslash\mathcal{H}$, and denote $q\colon \mathcal{H} \twoheadrightarrow Y(\Gamma)$ the quotient map. This can be given a structure of a Riemann surface with structure sheaf

$$\mathcal{O}_{Y(\Gamma)}(U) = \{f\colon \Gamma \cdot U \to \mathbb{C}\colon f \text{ is holomorphic and } \Gamma\text{-invariant}\}. \tag{2.24}$$

To write $Y(\Gamma)$ in terms of charts, one needs to be careful with the *elliptic points*. We refer to [DS05, Section 2] for a detailed description of $Y(\Gamma)$.

Definition 2.8. A point $z \in \mathcal{H}$ is *elliptic* for Γ if $\mathrm{Stab}_\Gamma(z) \subseteq \Gamma$ is larger than $\{\pm I\}$. We call $h_z := \#\mathrm{Stab}_\Gamma(z)/\{\pm I\}$ the *period* of z.

Example 2.3. For $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, the Γ-orbits of elliptic points are represented by i and ω, with periods 2 and 3 respectively. This means that for general Γ elliptic points can only have period 2 or 3, as $\mathrm{Stab}_\Gamma(z) \subseteq \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{Z})}(z)$. For $\Gamma = \Gamma(2)$ there are no elliptic points.

For $p \in \mathcal{H}$, write $\bar{p}$ for its image in $Y(\Gamma)$. Call $t_{\bar{p}}$ a uniformizer at the point $\bar{p}$. The structure of $Y(\Gamma)$ as a Riemann surface is such that the meromorphic function $\mathcal{H} \twoheadrightarrow Y(\Gamma) \xrightarrow{t_{\bar{p}}} \mathbb{P}^1(\mathbb{C})$ vanishes to order h_z on p.

Definition 2.9. We denote $\mathcal{H}^* := \mathcal{H} \sqcup \mathbb{P}^1(\mathbb{Q})$. This carries an a action of $\mathrm{GL}_2(\mathbb{Q})^+$, and for $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ we denote $X(\Gamma) := \Gamma\backslash\mathcal{H}^*$ the *compactified modular curve*. This carries the structure of a compact Riemann surface.

Again, we refer to [DS05, Section 2] for more details on this. If $\mathrm{Stab}_\Gamma(\infty) = \left\{\pm \begin{pmatrix} 1 & N* \\ 0 & 1 \end{pmatrix}\right\}$, then a uniformizer at $\bar{\infty}$ is $q_N := e^{2\pi i z/N}$. We call N the *width* of the cusp.

Remark 2.3. Really, the better way to think about $Y(\Gamma)$ and $X(\Gamma)$ is as a stacky quotient. In this viewpoint, what we are considering here corresponds to the coarse space attached to the stacky quotient.

Proposition 2.4. *$f\colon \mathcal{H} \to \mathbb{C}$ is weakly modular of weight $2k$ and level Γ if and only if $\omega_f(z) := f(z)(\mathrm{d}z)^k$ descends to a differential on $Y(\Gamma)$. Furthermore, f is meromorphic at all cusps if and only if $\omega_f(z)$ further extends to $X(\Gamma)$.*

Proof. This follows at once from the computation that

$$\frac{\mathrm{d}\frac{az+b}{cz+d}}{\mathrm{d}z} = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = (cz+d)^{-2}, \tag{2.25}$$

as then

$$f(\gamma z) = \nu_{2k}(\gamma, z) f(z) \iff f(\gamma z) = f(z)\left(\frac{\mathrm{d}z}{\mathrm{d}(\gamma z)}\right)^k \iff \omega_f(\gamma z) = \omega_f(z). \tag{2.26}$$

□

Given this, one maybe expect an identification of $M_k(\Gamma)$ and $H^0(X(\Gamma), \Omega^{\otimes k})$. However, this is not quite correct. One has to be careful with the elliptic points and with the cusps.

- For cusps: we look at ∞ for simplicity. If it has width N, then we write $\omega_f(z) = g(z)(\mathrm{d}q_N)^k$. Note that $\frac{\mathrm{d}q_N}{\mathrm{d}z} = \frac{2\pi i}{N} q_N$, so $\mathrm{d}z \sim \frac{\mathrm{d}q_N}{q_N}$. This means that
$$\mathrm{ord}_{\bar{\infty}}(\omega_f) = \mathrm{ord}_{\infty}(f) - k. \tag{2.27}$$
- For elliptic points: we write $\omega_f(z) = g(z)(\mathrm{d}t_{\bar{p}})^k$. If t_p is a uniformizer at p, then we have $t_{\bar{p}} \sim t_p^{h_p}$, and thus $\mathrm{d}t_{\bar{p}} \sim t_p^{h_p - 1}\,\mathrm{d}t_p$. Hence $g(z) \sim t_p^{\mathrm{ord}_p(f) - k(h_p-1)}(\mathrm{d}t_{\bar{p}})^k$, and thus
$$\mathrm{ord}_{\bar{p}}(\omega_f) = \frac{1}{h_p}(\mathrm{ord}_p(f) - k(h_p - 1)) = \frac{\mathrm{ord}_p(f)}{h_p} - k\left(1 - \frac{1}{h_p}\right). \tag{2.28}$$

Corollary 2.5. *Denote*
$$D = \sum_{p \in X(\Gamma)\ \textit{cusp}} p + \sum_{p \in X(\Gamma)\ \textit{elliptic}} (1 - 1/h_p) \cdot p \in \mathrm{Div}(X(\Gamma)) \otimes \mathbb{Q} \tag{2.29}$$
and $D_c = \sum_{p \in X(\Gamma)\ \textit{cusp}} p \in \mathrm{Div}(X(\Gamma))$. *Then we have identifications*
$$M_{2k}(\Gamma) \simeq H^0(X(\Gamma), \Omega^{\otimes k}(\lfloor kD \rfloor)), \quad S_{2k}(\Gamma) \simeq H^0(X(\Gamma), \Omega^{\otimes k}(\lfloor kD \rfloor - D_c)). \tag{2.30}$$
In particular, $M_{2k}(\Gamma)$ *is finite dimensional.*

The above computation also has the following interesting consequence. We denote by ϵ_h the number of elliptic points of $X(\Gamma)$ of period h, and by ϵ_∞ the number of cusps of $X(\Gamma)$.

Theorem 2.6 (Valence formula)**.** *Let* $f \in M_{2k}(\Gamma)$. *In fact, we may let* f *be weakly modular and meromorphic at all cusps. Then*
$$\sum_{\bar{p} \in X(\Gamma)\ \textit{cusp}} \mathrm{ord}_p(f) + \sum_{\bar{p} \in X(\Gamma)\ \textit{non cusp}} \frac{\mathrm{ord}_p(f)}{h_p} = k \cdot \left(2(g-1) + \frac{\epsilon_2}{2} + \frac{2\epsilon_3}{3} + \epsilon_\infty\right). \tag{2.31}$$

Proof. We have $\mathrm{div}(\omega_f) \sim k \cdot K_{X(\Gamma)}$, for a canonical divisor $K_{X(\Gamma)}$, which has degree $k \cdot 2(g-1)$. Thus
$$\sum_{\bar{p} \in X(\Gamma)} \mathrm{ord}_{\bar{p}}(\omega_f) = k \cdot 2(g-1), \tag{2.32}$$
and the claim follows from the above discussion. □

Remark 2.4. For $\bar{p}$ a cusp of width N, choose $\gamma \cdot p = \infty$. Then we are defining $\mathrm{ord}_p(f)$ to be such that, in the Fourier expansion $(f[\gamma]_k)(z) = \sum_{n\in\mathbb{Z}} a_n q_N^n$, we have $a_n = 0$ for $n < \mathrm{ord}_p(f)$ and $a_n \neq 0$ for $n = \mathrm{ord}_p(f)$. It is important to take the Fourier expansion with respect to q_N.

Given this, we can use Riemann–Roch to study the dimensions of the spaces $M_{2k}(\Gamma)$.

Proposition 2.7. *Denote $d = [\mathrm{SL}_2(\mathbb{Z})\colon \Gamma]$ the degree of the map $X(\Gamma) \to X(\mathrm{SL}_2(\mathbb{Z}))$. Then*

$$g(X(\Gamma)) = 1 + \frac{d}{12} - \frac{\epsilon_2}{4} - \frac{\epsilon_3}{3} - \frac{\epsilon_\infty}{2}. \tag{2.33}$$

Proof. First note that $X(\mathrm{SL}_2(\mathbb{Z}))$ has genus 0 (as can be easily seen topologically by the description of the fundamental domain). Riemann–Hurwitz for the finite map $X(\Gamma) \to X(\mathrm{SL}_2(\mathbb{Z}))$ tells us that

$$2g - 2 = -2d + \sum_{x\in X(\Gamma)} (e_x - 1). \tag{2.34}$$

The only possible ramified points are the cusps and the elliptic points, which are above one of $y_2 = i$, $y_3 = \omega$ and $y_\infty = \infty$. For $h \in \{2, 3\}$ we write

$$d = \sum_{x\in f^{-1}(y_h)} e_x = (\#f^{-1}(y_h) - \epsilon_h) \cdot h + \epsilon_h \cdot 1 \tag{2.35}$$

and thus

$$d - \#f^{-1}(y_h) = \sum_{x\in f^{-1}(y_h)} (e_x - 1) = (h-1)(\#f^{-1}(y_h) - \epsilon_h). \tag{2.36}$$

This allow us to write

$$\sum_{x\in f^{-1}(y_h)} (e_x - 1) = \left(1 - \frac{1}{h}\right)(d - \epsilon_h). \tag{2.37}$$

We similarly have

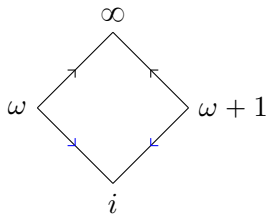
$$\sum_{x\in f^{-1}(y_\infty)} (e_x - 1) = d - \#f^{-1}(\infty) = d - \epsilon_\infty. \tag{2.38}$$

Now the claim follows. □

Corollary 2.8. *We may rewrite the valence formula as*

$$\sum_{\bar{p}\in X(\Gamma) \text{ cusp}} \mathrm{ord}_p(f) + \sum_{\bar{p}\in X(\Gamma) \text{ non cusp}} \frac{\mathrm{ord}_p(f)}{h_p} = \frac{kd}{6}. \tag{2.39}$$

Remark 2.5. Warning: I am not 100% sure if the following is correct, so take this remark with a grain of salt.[1] Thinking in terms of stacks, the left hand side is the "natural" way to count zeroes, and so it should be equal to the (stacky) degree of the line bundle $\Omega^{\otimes k}$ on $X(\Gamma)_{\text{stack}}$.[2] This degree is kd times the (stacky) degree of Ω on $X(\mathrm{SL}_2(\mathbb{Z}))_{\text{stack}}$. The (stacky) degree of the canonical bundle Ω is minus the (stacky) Euler characteristic. This can be computed with cell decompositions as $\chi = \sum_{\sigma \text{ cells}} \frac{(-1)^{\dim \sigma}}{\#\mathrm{Aut}(\sigma)}$, and applying this to the fundamental domain



we have: i) 3 vertices, with stabilizers of size $2, 3, \infty$, ii) 2 edges with stabilizers of size 1 and iii) 1 face with stabilizer of size 1. Thus

$$\chi = (\frac{1}{2} + \frac{1}{3} + 0) - (1 + 1) + (1) = -\frac{1}{6}. \tag{2.40}$$

For $k = 1$, we get

$$S_2(\Gamma) = H^0(X(\Gamma), \Omega), \tag{2.41}$$

which has dimension $g(X(\Gamma))$. In all other cases, it turns out that the computation is reduced to the simpler form of Riemann–Roch that $l(D) = \deg(D) - g + 1$ for $\deg(D) > 2g - 2$. Explicitly, this is

Theorem 2.9. *For $k \geqslant 1$ we have*

$$\dim M_{2k}(\Gamma) = (2k - 1)(g - 1) + \lfloor k/2 \rfloor \epsilon_2 + \lfloor 2k/3 \rfloor \epsilon_3 + k\epsilon_\infty, \tag{2.42}$$

and for $k \geqslant 2$ we have

$$\dim S_{2k}(\Gamma) = \dim M_{2k}(\Gamma) - \epsilon_\infty. \tag{2.43}$$

[1] More precisely, I am not sure how the cusps behave on the stacky framework, since they have infinite stabilizers. The reasoning I am outlining here definitely work if all stabilizers were finite (Deligne–Mumford stacks, or orbifolds), but I am assuming a similar reasoning also works in this more general case.

[2] Here, $X(\Gamma)_{\text{stack}}$ is the stacky quotient $\bar{\Gamma}\backslash\mathcal{H}^*$ for $\bar{\Gamma}$ the image of Γ on $\mathrm{PSL}_2(\mathbb{Z})$, so that the generic automorphism group has size 1. If one considers $\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$ then the generic automorphism group has size 2, and the left hand side is twice the "natural" count of zeroes.

Example 2.4. For $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, we have $g(X(\mathrm{SL}_2(\mathbb{Z}))) = 0$, and

$$\dim M_{2k}(\Gamma) = 1 - 2k + \lfloor k/2 \rfloor + \lfloor 2k/3 \rfloor + k = \begin{cases} \lfloor k/6 \rfloor + 1 & \text{if } k \not\equiv 1 \mod 6, \\ \lfloor k/6 \rfloor & \text{if } k \equiv 1 \mod 6. \end{cases} \tag{2.44}$$

Example 2.5. Looking at the fundamental domain of $X(\Gamma(2))$, we see that $d = 6$, $\epsilon_2 = \epsilon_3 = 0$ and $\epsilon_\infty = 3$. Thus $g(X(\Gamma(2))) = 0$, and

$$\dim M_{2k}(\Gamma(2)) = k + 1. \tag{2.45}$$

2.3 Modular forms of level $\mathrm{SL}_2(\mathbb{Z})$

Reference: [DS05, Sections 1.1, 1.2].

Later, we will see that the spaces $Y(\mathrm{SL}_2(\mathbb{Z}))$, and more generally $Y(\Gamma_0(N))$, are moduli spaces of elliptic curves. What follows is a first step towards the proof of that.

Proposition 2.10. *We have a bijection*

$$Y(\mathrm{SL}_2(\mathbb{Z})) \leftrightarrow \{\textit{lattices in } \mathbb{C}\}/\textit{homothety}. \tag{2.46}$$

given by $\tau \mapsto \mathbb{Z} \oplus \mathbb{Z}\tau$.

Furthermore, there is a bijection

$$\{f \colon \mathcal{H} \to \mathbb{C} \colon f[\gamma]_k = f \textit{ for all } \gamma \in \mathrm{SL}_2(\mathbb{Z})\} \leftrightarrow \{F \colon \{\textit{lattices in } \mathbb{C}\} \to \mathbb{C} \colon F(\lambda\Lambda) = \lambda^{-k}F(\Lambda)\}. \tag{2.47}$$

where F *corresponds to* $f(\tau) = F(\mathbb{Z} \oplus \mathbb{Z}\tau)$.

Proof. For the first bijection, it is easy to see that $\tau \mapsto \Lambda_\tau := \mathbb{Z} \oplus \mathbb{Z}\tau$ is surjective. Now if

$$\Lambda_\tau = \lambda\Lambda_{\tau'}, \tag{2.48}$$

this means that there is $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z})$ such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \tau \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda\tau' \\ \lambda \end{pmatrix}, \tag{2.49}$$

and in particular $\lambda = c\tau + d$ and $\tau' = \frac{a\tau+b}{c\tau+d}$. Note that $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z})^+ = \mathrm{SL}_2(\mathbb{Z})$, as both τ and τ' have positive imaginary part.

For the second claim, consider $\tau, \tau' \in \mathcal{H}$, $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and $\lambda \in \mathbb{C}^\times$ with $\tau' = \gamma\tau$ and $\Lambda_\tau = \lambda\Lambda_{\tau'}$ then as above we have $\lambda^k = \nu_k(\gamma, \tau)$, and thus

$$F(\Lambda_\tau) = \lambda^{-k}F(\Lambda_{\tau'}) \iff f(\tau) = \nu_k(\gamma, \tau)^{-1} f(\gamma\tau). \tag{2.50}$$

□

Similarly, we have

Proposition 2.11. *Denote* $\Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \colon c \equiv 0 \mod N \right\}$. *Then we have a bijection*

$$Y(\Gamma_0(N)) \leftrightarrow \{(\Lambda_1, \Lambda_2) \colon \Lambda_1 \subseteq \Lambda_2 \subseteq \mathbb{C} \text{ are lattices, } \Lambda_2/\Lambda_1 \simeq \mathbb{Z}/N\mathbb{Z}\}/\text{homothety} \tag{2.51}$$

which is given by

$$\tau \mapsto (\mathbb{Z} \oplus \mathbb{Z}\tau, \frac{1}{N}\mathbb{Z} \oplus \mathbb{Z}\tau). \tag{2.52}$$

We can use these descriptions to construct modular forms.

Definition 2.10. For $k \geqslant 2$, we consider the *Eisenstein series*

$$G_{2k}(\Lambda) := \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{\lambda^{2k}}. \tag{2.53}$$

Under the above, this corresponds to a function $G_{2k} \colon \mathcal{H} \to \mathbb{C}$ given by

$$G_{2k}(\tau) = \sum_{(n,m) \in \mathbb{Z}^2 \setminus \{(0,0)\}} \frac{1}{(n\tau + m)^{2k}} \tag{2.54}$$

which is of weight $2k$ for $\mathrm{SL}_2(\mathbb{Z})$.

Remark 2.6. For $k = 1$, this expression does not converge absolutely, and hence will fail to satisfy the transformation property for modular forms. We can still consider it as a conditionally convergent sum:

$$G_2(\tau) := 2\zeta(2) + \sum_{n \in \mathbb{Z} \setminus \{0\}} \sum_{m \in \mathbb{Z}} \frac{1}{(n\tau + m)^2}. \tag{2.55}$$

It is not hard to compute that

$$\lim_{\mathrm{Im}(\tau) \to \infty} G_{2k}(\tau) = 2\zeta(2k), \tag{2.56}$$

and hence that $G_{2k} \in M_{2k}(\mathrm{SL}_2(\mathbb{Z}))$ for $k \geqslant 2$. For $k = 1$, one can prove that G_2 satisfies the following instead:

Proposition 2.12 ([DS05, Exercise 1.2.8])**.** $G_2(\tau) - \frac{\pi}{\mathrm{Im}(\tau)}$ *is weight-*2 *invariant for* $\mathrm{SL}_2(\mathbb{Z})$. *In other words,*

$$(G_2[\gamma]_2)(\tau) = G_2(\tau) - \frac{2\pi i c}{c\tau + d}. \tag{2.57}$$

Proposition 2.13. *We have* $\sum_{k\geqslant 0} M_k(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}[G_4, G_6]$.

Proof. First we compare dimensions. We need to see that

$$\dim M_k(\mathrm{SL}_2(\mathbb{Z})) = \#\{a, b \in \mathbb{Z}_{\geqslant 0} \colon 4a + 6b = k\}. \tag{2.58}$$

This can be proved by induction on k, as $\dim M_{k+12}(\mathrm{SL}_2(\mathbb{Z})) = 1 + \dim M_k(\mathrm{SL}_2(\mathbb{Z}))$ and

$$\#\{a, b \in \mathbb{Z}_{\geqslant 0} \colon 4a + 6b = k + 12\} = 1 + \#\{a \in \mathbb{Z}_{\geqslant 0}, b \in \mathbb{Z}_{\geqslant 2} \colon 4a + 6b = k + 12\} \tag{2.59}$$

which is simply $1 + \#\{a, b \in \mathbb{Z}_{\geqslant 0} \colon 4a + 6b = k\}$. It remains to prove that G_4, G_6 are algebraically independent.

By the valence formula, we have

$$\frac{\mathrm{ord}_i(G_4)}{2} + \frac{\mathrm{ord}_\omega(G_4)}{3} + \sum_{\bar{p} \in X(\Gamma) \setminus \{\bar{i}, \bar{\omega}\}} \mathrm{ord}_p(G_4) = \frac{2}{6} = \frac{1}{3}, \tag{2.60}$$

hence the only zero of G_4 is at ω, of order 1. Similarly, we have

$$\frac{\mathrm{ord}_i(G_6)}{2} + \frac{\mathrm{ord}_\omega(G_6)}{3} + \sum_{\bar{p} \in X(\Gamma) \setminus \{\bar{i}, \bar{\omega}\}} \mathrm{ord}_p(G_6) = \frac{3}{6} = \frac{1}{2}, \tag{2.61}$$

so the only zero of G_6 is at i, of order 1. Now assume $F \in \mathbb{C}[x, y]$ was such that $F(G_4, G_6) = 0$. By declaring $\deg(x) = 4, \deg(y) = 6$, we may write $F(x, y) = \sum_k F_k(x, y)$ where $F_k(x, y)$ has degree k. Now $\sum_k F_k(G_4, G_6)$ is a sum of modular forms $F_k(G_4, G_6)$ of differing weights, and it we can see that this implies that each $F_k(G_4, G_6)$ is zero: we have

$$0 = F(G_4, G_6)(\gamma\tau) = \sum_k F_k(G_4, G_6)(\gamma\tau) = \sum_k (c\tau + d)^k F_k(G_4, G_6)(\tau). \tag{2.62}$$

Since this is true for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, we have that each $F_k(G_4, G_6)$ is identically zero. Hence we may assume without loss of generality that $F = F_k$ for some k.

We may also assume without loss of generality that $x \nmid F$. Then plugging in i, we get $0 = F(G_4(i), G_6(i)) = F(G_4(i), 0)$. But since $x \nmid F$ and $G_4(i) \neq 0$, this is a contradiction. □

The first nonzero cusp form is in $S_{12}(\mathrm{SL}_2(\mathbb{Z}))$, which is a multiple of

$$\left(\frac{G_4}{2\zeta(4)}\right)^3 - \left(\frac{G_6}{2\zeta(6)}\right)^2. \tag{2.63}$$

Definition 2.11. We define

$$\Delta = \frac{1}{(2\pi)^{12}}((60G_4)^3 - 27(140G_6)^2) \tag{2.64}$$

the *Ramanujan Delta function*, which is an element of $S_{12}(\mathrm{SL}_2(\mathbb{Z}))$. This is uniquely characterized by $\Delta \in S_{12}(\mathrm{SL}_2(\mathbb{Z}))$ and $a_1(\Delta) = 1$.

By the valence formula, we must have that Δ has a simple zero at ∞, and is nonvanishing at $\mathcal{H}$.

Proposition 2.14. *We have*

$$\Delta(\tau) = q\prod_{n\geqslant 1}(1-q^n)^{24}. \tag{2.65}$$

Proof. From the homework, we have

$$G_{2k}(\tau) = 2\zeta(2k) + 2\frac{(2\pi i)^{2k}}{(2k-1)!}\sum_{n\geqslant 1}\sigma_{2k-1}(n)q^n \tag{2.66}$$

for all $k \geqslant 1$. Using this, we see that $a_1(\Delta) = 1$. Since $\dim S_{12}(\mathrm{SL}_2(\mathbb{Z})) = 1$, it now suffices to prove that $\eta(z) := q^{1/24}\prod_{n\geqslant 1}(1-q^n)$ satisfies

$$\eta(-1/z) = \sqrt{-iz}\cdot\eta(z) \tag{2.67}$$

where the branch of $\sqrt{\cdot}$ is such that $\sqrt{-iz}$ is 1 for $z = i$. It turns out that this is equivalent to the transformation property for G_2. We have

$$\frac{\mathrm{d}\eta}{\mathrm{d}q} = \frac{1}{24}\frac{\eta}{q} - \sum_{n\geqslant 1} nq^{n-1}\frac{\eta}{(1-q^n)} = \frac{\eta}{q}\left(\frac{1}{24} - \sum_{n\geqslant 1} n\frac{q^n}{1-q^n}\right) = \frac{\eta}{q}\left(\frac{1}{24} - \sum_{n\geqslant 1}\left(\sum_{d|n} d\right)q^n\right) \tag{2.68}$$

and so

$$\frac{\mathrm{d}\eta}{\eta} = \frac{1}{8\pi^2}G_2\cdot\frac{\mathrm{d}q}{q} = \frac{i}{4\pi}G_2\cdot\mathrm{d}z. \tag{2.69}$$

Taking $\mathrm{d}\log$ of $\eta(-1/z) = \sqrt{-iz}\cdot\eta(z)$, it suffices to prove that

$$\frac{\mathrm{d}\eta}{\eta}(-1/z) = \frac{\mathrm{d}z}{2z} + \frac{\mathrm{d}\eta}{\eta}(z). \tag{2.70}$$

This is equivalent to

$$G_2(-1/z)\,\mathrm{d}(-1/z) = G_2(z)\,\mathrm{d}z - \frac{2\pi i}{z}\,\mathrm{d}z \iff G_2(-1/z) = z^2G_2(z) - 2\pi iz. \tag{2.71}$$

□

Definition 2.12. Define

$$j(\tau) := \frac{G_4(\tau)^3}{(2\zeta(4))^3 \cdot \Delta(\tau)} = \frac{1}{q} + 744 + 19884q + 21493760q^2 + \cdots . \tag{2.72}$$

Since $\Delta\colon \mathcal{H} \to \mathbb{C}$ is nonvanishing, j is a holomorphic function $j\colon Y(\mathrm{SL}_2(\mathbb{Z})) \to \mathbb{C}$.

Proposition 2.15. *$j\colon X(\mathrm{SL}_2(\mathbb{Z})) \to \mathbb{P}^1(\mathbb{C})$ is an isomorphism. In other words, we have $\mathbb{C}(X(\mathrm{SL}_2(\mathbb{Z}))) = \mathbb{C}(j)$.*

Proof. For $c \in \mathbb{C}$ we consider $j - c\colon \mathcal{H} \to \mathbb{C}$. As $j - c$ has a simple pole at ∞, the valence formula applied to $j - c$ says that there is exactly one point $\tau_c \in \mathcal{H}$ such that $j(\tau_c) = c$. Similarly, the only pole of j is at ∞, and thus $j\colon X(\mathrm{SL}_2(\mathbb{Z})) \to \mathbb{P}^1(\mathbb{C})$ is a bijection, hence an isomorphism. □

3 Elliptic curves

3.1 Geometry of elliptic curves

Reference: [Sil09, Sections III.2-III.4].

Let k be a field, and E/k be an elliptic curve.

Proposition 3.1. *E is isomorphic to the projective curve defined by a cubic equation of the form*

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

for some $a_1, a_2, a_3, a_4, a_6 \in k$, where O corresponds to the point $[0 : 1 : 0]$. This is called a Weierstraß form.

Proof. We apply Riemann–Roch for the divisors $D = nO$. First note that if D is a divisor with $\deg(D) > 0$, we have

$$l(D) - l(K - D) = \deg(D) - g + 1 = \deg(D), \tag{3.2}$$

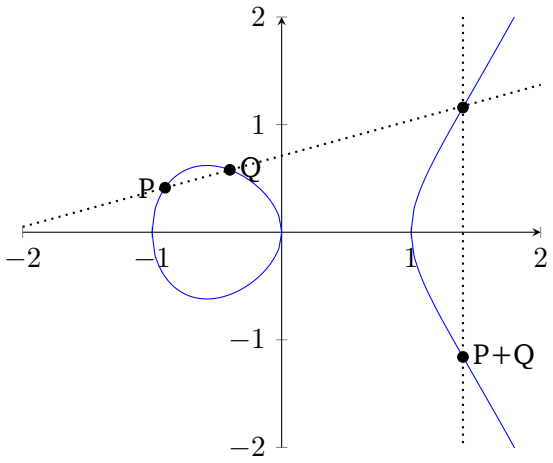
and since $\deg(K - D) = 2g - 2 - \deg(D) = -\deg(D) < 0$, we conclude that $l(D) = \deg(D)$.

We have $\mathcal{L}(O) = k$, and we choose functions x, y such that $\mathcal{L}(2O) = k \oplus k\cdot x$, $\mathcal{L}(3O) = k \oplus k\cdot x \oplus k\cdot y$. Then for $D = 6O$, we have $1, x, x^2, x^3, y, xy, y^2 \in \mathcal{L}(6O)$, and so they must be linearly dependent. Both x^3 and y^2 must be part of the relation since they have poles of order exactly 6. Scaling x, y we get the equation of the form above, say

$$f\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6. \tag{3.3}$$

This gives a map $E \to V(f) \subseteq \mathbb{P}^2(x,y)$ for f as above. Since both $V(f) \to \mathbb{P}^1(x)$ and $E \to \mathbb{P}^1(x)$ have degree 2, the map $E \to V(f)$ has degree 1. Since f is a cubic, we have either that $V(f)$ is smooth, in which case $E \to V(f)$ is an isomorphism, or that $V(f)$ is birational to a smooth curve of genus 0. In the later case, E would be birational (and hence isomorphic) to a smooth curve of genus 0, which is not the case. □

Upon choosing a Weierstraß form, we have the chord-and-tangent construction



The following is a more natural description of this group law, which in particular proves it is associative and independent of the choice of Weierstraß form.

Proposition 3.2. *We consider the map $E(k) \to \mathrm{Pic}^0(E)(k)$ given by $P \mapsto (P) - (O)$. This is a bijection of sets, which endow $E(k)$ with an abelian group structure. Moreover, this extends to an algebraic maps $m\colon E \times E \to E$ and $[-1]\colon E \to E$, making E into a group variety.*

Proof. If this was not injective, we would have $(P) \sim (Q)$ for some $P \neq Q$. this would give us a degree 1 map $E \to \mathbb{P}^1$, which would be an isomorphism.

Now if we have a Weierstrass form, we note that the chord-and-tangent construction satisfies that $(P) + (Q) \sim (P+Q) - (O)$: if l_1, l_2 are the linear equations defining the two dotted lines above, we have $\mathrm{div}(l_1/l_2) = (P) + (Q) - (P+Q) - (O)$. In other words, this is the same group structure coming from the map $E(k) \hookrightarrow \mathrm{Pic}^0(E)(k)$. This makes it clear that this group structure extends to algebraic maps $m\colon E \times E \to E$ and $[-1]\colon E \to E$.

Surjectivity of $E(k) \hookrightarrow \mathrm{Pic}^0(E)(k)$ also follows from the chord-and-tangent construction for a Weierstraß form: a general divisor

$$(P_1) + \cdots + (P_r) - (Q_1) - \cdots (Q_r), \tag{3.4}$$

by repeatedly using the chord and tangent construction, is linearly equivalent to $(P_1 + \cdots + P_r - Q_1 - \cdots - Q_r) - (O)$. □

Definition 3.1. An *isogeny* $f\colon E_1 \to E_2$ between elliptic curves is an algebraic map with $f(O_1) = O_2$.

Note that for $P \in E_2$ we may consider the morphism $\tau_P\colon E_2 \to E_2$ given by translation by P. With that, it is easy to see that any morphism $g\colon E_1 \to E_2$ factors uniquely as $g = \tau_P \circ f$ where $P = g(O_1)$ and $f = \tau_{-P} \circ g$ is an isogeny.

Proposition 3.3. *Let $f\colon E_1 \to E_2$ be an isogeny of elliptic curves. Then f respects the group structures.*

Proof. This follows at once from the functoriality of Pic^0: We have a commutative diagram

$$\begin{array}{ccc} E_1 & \xrightarrow{\quad f \quad} & E_2 \\ \downarrow & & \downarrow \\ \mathrm{Pic}^0(E_1) & \xrightarrow{f_*} & \mathrm{Pic}^0(E_2) \end{array} \tag{3.5}$$

□

Example 3.1. We have the following examples concerning isogenies.

1. $0\colon E \to E'$ the constant map to O'.
2. For $m \in \mathbb{Z}$, we have the multiplication-by-m maps $[m]\colon E \to E$: for $m \geqslant 0$, $[m]P = P + P + \cdots + P$ and for $m \leqslant 0$, $[m] = [-1] \circ [-m]$.
3. If $\mathrm{char}(k) = p > 0$, we have the Frobenius map $\phi\colon E \to E^{(p)}$. This has degree p. If k is perfect, this corresponds to $k(E^{(p)}) = k(E)^p \subseteq k(E)$ in terms of function fields.

Proposition 3.4. *If $\mathrm{char}(k) = p > 0$ and k is perfect, any nonzero isogeny $E_1 \to E_2$ factors uniquely up to isomorphism as $E_1 \to E_1^{(p^r)} \to E_2$ where $E_1^{(p^r)} \to E_2$ is separable.*

Proof. More generally, it is true that if $f\colon C_1 \to C_2$ is a morphism between smooth curves, then f factors uniquely as $C_1 \to C_1^{(p^r)} \to C_2$ where $C_1^{(p^r)} \to C_2$ is separable. To see this, consider the function fields $K_1 = k(C_1)$ and $K_2 =$

$k(C_2)$. Then f corresponds to an inclusion $K_2 \subseteq K_1$. We let $K_3 \subseteq K_1$ denote the separable closure of K_2 in K_1, and let C_3 be the corresponding smooth curve with $k(C_3) = K_3$. This gives the factorization $f\colon C_1 \to C_3 \to C_2$, where $C_1 \to C_3$ is purely inseparable and $C_3 \to C_2$ is separable.

It remains to see that a purely inseparable map $g\colon C_1 \to C_3$ is a power of the Frobenius. In terms of function fields, we have $[K_1\colon K_3] = \deg g$, and this must be a power of p, say $q = p^r$. Then $K_1^q \subseteq K_3$, as if $x \in K_1$, its minimal polynomial over K_3 is of the form $x^{p^e} - a$ with $e \leqslant r$ for some $a \in K_3$. So g factors as $C_1 \to C_1^{(q)} \to C_3$, and considering the degrees, $C_1^{(q)} \to C_3$ has degree 1, and hence is an isomorphism. □

Proposition 3.5. *Assume k is perfect and let E/k be an elliptic curve. There is an $\mathrm{Gal}(\bar{k}/k)$-equivariant equivalence of categories*

$$\{\textit{separable isogenies } f\colon E \to E' \textit{ over } \bar{k}\} \longleftrightarrow \{\textit{finite subgroups } \Phi \subseteq E(\bar{k})\} \tag{3.6}$$

such that if f_Φ and Φ correspond, then $\Phi = \ker f_\Phi := f_\Phi^{-1}(O')$. In particular, if $\Phi \subseteq E(\bar{k})$ is $\mathrm{Gal}(\bar{k}/k)$-stable, then f_Φ is defined over k.

Proof. For each $T \in \Phi$, we consider the translation morphism $\tau_T\colon E \to E$. This induces an isomorphism of function fields $\tau_T^* \in \mathrm{Aut}(\bar{k}(E))$. Denote $\bar{k}(E)^\Phi \subseteq \bar{k}(E)$ the subfield fixed by such τ_T^* for all $P \in \Phi$. By Galois theory, $\bar{k}(E)/\bar{k}(E)^\Phi$ is Galois with Galois group Φ. Let $f_\Phi\colon E \to E'$ be the map of smooth curves corresponding to this field inclusion.

We note that f_Φ is unramified. If $\phi \in \bar{k}(E')$ and $T \in \Phi$, we have $\phi(f_\Phi(P + T)) = \phi(f_\Phi(\tau_T(P))) = \tau_T^*(\phi \circ f_\Phi)(P)$, where $\phi \circ f_\Phi \in \bar{k}(E)$ is an element of $\bar{k}(E)^\Phi$, and thus $\phi(f_\Phi(P + T)) = \phi(f_\Phi(P))$. Thus $f_\Phi(P + T) = f_\Phi(P)$ for all $T \in \Phi$. If $Q \in E'$ and $f_\Phi(P) = Q$, then $f_\Phi^{-1}(Q) \supseteq \{P + T\colon T \in \Phi\}$, as so this is an equality as $\#\Phi = \deg f_\Phi$. In particular, f_Φ is unramified everywhere. By Riemann–Hurwitz, we have

$$2g(E) - 2 = (2g(E') - 2) \cdot \deg f_\Phi \tag{3.7}$$

and thus $g(E') = 1$. □

Let $\mathrm{Isog}(E_1, E_2)$ denote the set of isogenies between E_1 and E_2. It is an abelian group via $(\phi + \psi)(P) := \phi(P) + \psi(P)$.

Proposition 3.6. *There is a unique duality $\widehat{(\cdot)}\colon \mathrm{Isog}(E_1, E_2) \to \mathrm{Isog}(E_2, E_1)$ such that $\widehat{0} = 0$ and such that $\phi \circ \hat{\phi} = [\deg(\phi)]$. It satisfies $\deg(\hat{\phi}) = \deg(\phi)$ and $\hat{\hat{\phi}} = \phi$. We also have $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$.*

Remark 3.1. For abelian varieties, the group variety $\mathrm{Pic}^0(A)$ is not necessarily isomorphic to A. It is instead called the *dual abelian variety* $\hat{A}$. One still has a duality as above: $\mathrm{Isog}(A_1, A_2) \simeq \mathrm{Isog}(\hat{A}_2, \hat{A}_1)$.

Proof. Uniqueness is clear since if ϕ is nonzero and $\phi \circ f = \phi \circ g$, then we would have that $\phi \circ (f - g) = 0$, and thus that $f - g = 0$ since otherwise it would be surjective and then we would conclude $\phi = 0$.

We saw before that we have a canonical identification $E \simeq \mathrm{Pic}^0(E)$ as group varieties. Now an isogeny $\phi\colon E_1 \to E_2$, induces a pullback map $\phi^*\colon \mathrm{Pic}^0(E_2) \to \mathrm{Pic}^0(E_1)$, which we may think of as a map $\hat{\phi}\colon E_2 \to E_1$.

Now the composition $\phi \circ \hat{\phi}\colon E_2 \to E_1$ is identified with $\phi_* \circ \phi^*\colon \mathrm{Pic}^0(E_2) \to \mathrm{Pic}^0(E_1)$. We can easily compute $\phi_* \circ \phi^*\colon \mathrm{Div}(E_2) \to \mathrm{Div}(E_1)$:

$$(\phi_* \circ \phi^*)(P) = \phi_* \left(\sum_{Q \in \phi^{-1}(P)} e_Q \cdot Q \right) = \sum_{Q \in \phi^{-1}(P)} e_Q \cdot \phi(Q) = \left(\sum_{Q \in \phi^{-1}(P)} e_Q \right) \cdot P = \deg \phi \cdot P \tag{3.8}$$

and thus $\phi \circ \hat{\phi} = [\deg \phi]$.

We now prove that $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$. We can write $\phi + \psi\colon E_1 \to E_2$ as the composition

$$E_1 \xrightarrow{\Delta} E_1 \times E_1 \xrightarrow{(\phi, \psi)} E_2 \times E_2 \xrightarrow{m} E_2 \tag{3.9}$$

and thus if $\mathcal{L} \in \mathrm{Pic}^0(E_2)$, we have

$$(\phi+\psi)^*\mathcal{L} = \Delta^*(\phi, \psi)^* m^* \mathcal{L} = \Delta^*(\phi, \psi)^*(\mathcal{L} \boxtimes \mathcal{L}) = \Delta^*(\phi^*\mathcal{L} \boxtimes \psi^*\mathcal{L}) = \phi^*\mathcal{L} \otimes \psi^*\mathcal{L}. \tag{3.10}$$

The second equality requires an argument. Consider the line bundle $M(\mathcal{L}) := m^*\mathcal{L} \otimes \mathrm{pr}_1^*\mathcal{L}^{-1} \otimes \mathrm{pr}_2^*\mathcal{L}^{-1}$, which we wish to see is trivial. For $P \in E_2$, we consider $i_P\colon E \simeq P \times E \hookrightarrow E \times E$. Then

$$i_P^* M(\mathcal{L}) = \tau_P^*\mathcal{L} \otimes \mathcal{L}^{-1}. \tag{3.11}$$

Now if $\mathcal{L}$ is the line bundle $\mathcal{O}(Q)$ for some $Q \in E_2$, we have

$$i_P^* M(\mathcal{O}(Q)) = \mathcal{O}(Q - P) \otimes \mathcal{O}(Q)^{-1}$$

and since $(Q-P)-(Q) \sim (-P)-(O)$, we have $i_P^* M(\mathcal{O}(Q)) = \mathcal{O}((-P)-(O))$. So in general, if $\mathcal{L} \in \mathrm{Pic}(E_2)$ we have $i_P^* M(\mathcal{L}) = \mathcal{O}((-P) - (O))^{\otimes \deg \mathcal{L}}$. If $\mathcal{L} \in \mathrm{Pic}^0(E_2)$, we conclude that $i_P^* M(\mathcal{L})$ is trivial for all P. This means that $M(\mathcal{L}) \simeq \mathrm{pr}_2^*(\mathcal{L}')$ for some $\mathcal{L}'$. Applying the same argument but switching pr_1 and pr_2, we conclude that $M(\mathcal{L})$ is trivial.

This implies that $\widehat{[m]} = [m]$, and in particular $\deg[m] = m^2$. Now

$$(\deg \phi)^2 = \deg[\deg \phi] = \deg \phi \cdot \deg \hat{\phi} \tag{3.12}$$

and thus $\deg \hat{\phi} = \deg \phi$. Finally, we have

$$[\deg \phi] = \widehat{[\deg \phi]} = (\hat{\phi} \circ \phi)^\wedge = \hat{\phi} \circ \hat{\hat{\phi}}, \tag{3.13}$$

□

Remark 3.2. Restricting to separable morphisms, if ϕ corresponds to $\Phi \subseteq E_1(\bar{k})$, then $\hat{\phi}$ corresponds to $\phi(E_1[\deg \phi]) \subseteq E_2(\bar{k})$. This gives an alternate approach to prove this proposition: We construct $\hat{\phi}$ for separable ϕ as above. For the Frobenius $\mathrm{Frob}\colon E \to E^{(p)}$, one proves that $[p]$ is inseparable (as if ω is a holomorphic differential, we have $[p]^*\omega = p\omega = 0$ as per homework), and so $[p] = \mathrm{Ver} \circ \mathrm{Frob}$ for some isogeny $\mathrm{Ver}\colon E^{(p)} \to E$ called *Verschiebung*, which $\widehat{\mathrm{Frob}}$ is defined to be. It is a bit trickier to prove that $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$ with these definitions, but this can be done cleanly with Tate modules and the Weil pairing, as we will see later.

3.2 Weierstraß equations

Reference: [Sil09, Sections III.1].

Let k be a field. We consider a projective curve

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{3.14}$$

where $a_1, a_2, a_3, a_4, a_6 \in k$.

In the case that $\mathrm{char}(k) \neq 2, 3$ there exists a change of coordinates $y \mapsto y - \frac{(a_1x+a_3)}{2}$, $x \mapsto x - \frac{a_1^2+4a_4}{12}$ so that E is defined by $E\colon y^2 = x^3 + Ax + B$ for some polynomials $A, B \in \mathbb{Z}[\frac{1}{6}][a_1, a_2, a_3, a_4, a_6]$.

Definition 3.2. We denote

$$\Delta = -16(4A^3 + 27B^2), \quad j = -1728\frac{(4A)^3}{\Delta}. \tag{3.15}$$

We note that $\Delta, j\Delta \in \mathbb{Z}[a_1, a_2, a_3, a_4, a_6]$, so they still make sense for general k.

Proposition 3.7. *E is a nonsingular projective curve if and only if $\Delta \neq 0$.*

Proof. This is a straightforward computation, which is easy if $\mathrm{char}(k) \neq 2, 3$. □

Proposition 3.8. *j does not depend on the choice of Weierstraß model of E. Moreover, j is a complete invariant for isomorphism classes of elliptic curves over $\bar{k}$. We call $j(E)$ the* j-invariant of E.

Proof. This follows from analyzing possible isomorphisms between Weierstraß models. In the simple form $y^2 = x^3 + Ax + B$, they are all changes of variables $x \mapsto u^{-2}x$ and $y \mapsto u^{-3}y$ for some $u \in \bar{k}$, which correspond to $A \mapsto u^4 A$ and $B \mapsto u^6 B$. For $\mathrm{char}(k) \in \{2, 3\}$ one needs to be more careful with the computations.

Given a $j \in \bar{k}$, one can also construct an elliptic curve with this j-invariant, namely

$$y^2 + xy = x^3 - \frac{36}{j - 1728}x - \frac{1}{j - 1728} \tag{3.16}$$

if $j \neq 0, 1728$, and $y^2 + y = x^3$, $y^2 = x^3 + x$ for the remaining two values. □

3.3 Elliptic curves over $\mathbb{C}$

Reference: [Sil09, Chapter VI], or [DS05, Sections 1.3-1.5].

The main result we want to prove is the following.

Theorem 3.9. *The category of elliptic curves over $\mathbb{C}$ (with morphisms given by isogenies) is equivalent to the category of lattices $\{\Lambda \subseteq \mathbb{C}\}$ with morphisms given by*

$$\mathrm{Hom}(\Lambda_1, \Lambda_2) = \{\alpha \in \mathbb{C} \colon \alpha\Lambda_1 \subseteq \Lambda_2\}. \tag{3.17}$$

If E_Λ and Λ are matching objects, we have an isomorphism of Riemann surfaces $E_\Lambda(\mathbb{C}) \simeq \mathbb{C}/\Lambda$ which respect the group structures and is also compatible with morphisms.

Together with Propositions 2.10, 2.11 and 3.5, this also give us the following.

Corollary 3.10. *We have*

$$Y(\mathrm{SL}_2(\mathbb{Z})) \leftrightarrow \{E/\mathbb{C}\}/\simeq \tag{3.18}$$

and

$$\begin{aligned} Y(\Gamma_0(N)) &\leftrightarrow \{E_1 \to E_2 \text{ cyclic isogeny of degree } N\}/\simeq \\ &\leftrightarrow \{(E, C) \colon C \subseteq E \text{ is a subgroup isomorphic to } \mathbb{Z}/N\mathbb{Z}\}/\simeq . \end{aligned} \tag{3.19}$$

There are a few ways of approaching this. The fact that every $\mathbb{C}/\Lambda$ is isomorphic to some $E_\Lambda(\mathbb{C})$ is an instance of *Riemann's existence theorem*: compact Riemann surfaces are algebraic varieties. The converse is an example of *uniformization*. We will present proofs which use some of the work on modular forms for $\mathrm{SL}_2(\mathbb{Z})$ that we did above, although there are also approaches that do not involve modular forms.

Existence. In order to prove that $\mathbb{C}/\Lambda$ are algebraic varieties, we need to construct enough functions $\mathbb{C}/\Lambda \to \mathbb{C}$. This is the study of *elliptic functions*. Denote

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \tag{3.20}$$

the *Weierstraß $\wp$-function of Λ*.

Note that $\wp'_\Lambda(z) = -2\sum_{\lambda\in\Lambda} \frac{1}{(z-\lambda)^3}$ is Λ-periodic, with triple zeroes at Λ and simple zeroes at $\frac{1}{2}\Lambda\setminus\Lambda$. From this periodicity, we have that $\wp_\lambda(z+\lambda) = \wp_\Lambda(z) + c(\lambda)$ for $\lambda \in \Lambda$ and constants $c(\lambda)$. Since $\wp_\Lambda$ is even, we can plug in $z = -\lambda/2$ to conclude that $c(\lambda) = 0$. That is, $\wp_\Lambda$ is also Λ-periodic.

By integrating $\wp'_\Lambda/(\wp_\Lambda - c)$ on a fundamental domain of Λ, we have that $\wp_\Lambda - c$ must have exactly two zeroes modulo Λ. Since $\wp_\Lambda$ is even, these are either two simple zeroes $z, -z$ with $z \notin \frac{1}{2}\Lambda$, or a double zero at $z \in \frac{1}{2}\Lambda\setminus\Lambda$.

So analyzing zeroes and poles we have

$$(\wp'_\Lambda(z))^2 = C \prod_{\lambda \in (\frac{1}{2}\Lambda/\Lambda)\setminus\{0\}} (\wp_\Lambda(z) - \wp_\Lambda(\lambda)) = F(\wp_\Lambda) \tag{3.21}$$

for a cubic $F(x) \in \mathbb{C}[x]$. With a bit more work, we can see that the map

$$[\wp_\Lambda : \wp'_\Lambda : 1] \colon \mathbb{C}/\Lambda \to \mathbb{P}^2(\mathbb{C}) \tag{3.22}$$

identifies $\mathbb{C}/\Lambda$ with the closed subvariety of $\mathbb{P}^2(\mathbb{C})$ cut out by $y^2 = F(x)$. □

Remark 3.3. We can compute this cubic F explicitly from the Taylor expansion of $\wp_\Lambda$. We have

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\lambda\in\Lambda\setminus\{0\}} \frac{1}{\lambda^2}\left(\frac{1}{(1-z/\lambda)^2} - 1\right) = \frac{1}{z^2} + \sum_{n\geqslant 1}\sum_{\lambda\in\Lambda\setminus\{0\}} (n+1) z^n \lambda^{-(n+2)}, \tag{3.23}$$

and thus

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{n\geqslant 1} z^{2n}(2n+1)G_{2(n+1)}(\Lambda). \tag{3.24}$$

Matching the first terms, we can compute that $F(x) = 4x^3 - 60G_4(\Lambda)x - 140G_6(\Lambda)$. The discriminant of this Weierstraß equation is $(4\pi)^{12}\Delta(\Lambda)$, and it's j-invariant is $j(\Lambda)$. The fact that F does not have repeated roots is equivalent to Δ being non-vanishing, which we saw before.

Uniformization. Under the construction $\Lambda \mapsto E_\Lambda$ above, the j invariant of E_Λ is simply $j(\Lambda)$ because of the above remark. As we saw before, given any $c \in \mathbb{C}$ there is Λ with $j(\Lambda) = c$, and thus any elliptic curve over $\mathbb{C}$ is isomorphic to some E_Λ. □

Remark 3.4. We can also approach uniformization by thinking of $\mathrm{Pic}^0(E)$ as the Jacobian of E:

$$E(\mathbb{C}) \simeq \mathrm{Pic}^0(E)(\mathbb{C}) = \mathrm{Jac}(E) = \frac{H^0(E(\mathbb{C}),\Omega)^*}{H_1(E(\mathbb{C}),\mathbb{Z})} \simeq \mathbb{C}/\Lambda \tag{3.25}$$

as $\dim_{\mathbb{C}} H^0(E(\mathbb{C}),\Omega) = g(E) = 1$. and $H_1(E(\mathbb{C}),\mathbb{Z})$ is a free $\mathbb{Z}$-module of rank 2.

4 Automorphic representations of GL_2

4.1 Adelic quotients of GL_2 and modular curves

Theorem 4.1 (Strong approximation for SL_2)**.** *Let F be a number field. Then $\mathrm{SL}_2(F)$ is dense in $\mathrm{SL}_2(\mathbb{A}_{F,f})$.*

Proof. Let $N^+ = \left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \right\}$ and $N^- = \left\{ \begin{pmatrix} 1 & 0 \\ * & 1 \end{pmatrix} \right\}$ be unipotent subgroups of SL_2.

Then we have that $\mathrm{SL}_2(\mathbb{A}_{F,f})$ is generated by $N^{\pm}(\mathbb{A}_{F,f})$. This is because $\mathrm{SL}_2(F_{\mathfrak{p}})$ is generated by $N^{\pm}(F_{\mathfrak{p}})$ and $\mathrm{SL}_2(\mathcal{O}_{\mathfrak{p}})$ is generated by $N^{\pm}(\mathcal{O}_{\mathfrak{p}})$.

Let Z be the closure of $\mathrm{SL}_2(F)$ in $\mathrm{SL}_2(\mathbb{A}_{F,f})$. Then Z contains the closure of $N^{\pm}(F)$. By strong approximation for $\mathbb{G}_a$ (i.e. the Chinese remainder theorem), such closures are $N^{\pm}(\mathbb{A}_{F,f})$. Since Z is a subgroup, we conclude that $Z = \mathrm{SL}_2(\mathbb{A}_{F,f})$. □

Now if $K \subseteq \mathrm{SL}_2(\mathbb{A}_f)$ is an open compact, the above theorem tells us that $\mathrm{SL}_2(\mathbb{A}_f) = \mathrm{SL}_2(\mathbb{Q})K$ and thus that

$$\mathrm{SL}_2(\mathbb{Q})\backslash\mathrm{SL}_2(\mathbb{A})/K = \mathrm{SL}_2(\mathbb{Q})\backslash\left(\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{Q})K/K\right) = \Gamma\backslash\mathrm{SL}_2(\mathbb{R}) \tag{4.1}$$

where $\Gamma = \mathrm{SL}_2(\mathbb{Q}) \cap K$.

Since $\mathcal{H} = \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$, we have for $K_\infty = \mathrm{SO}_2(\mathbb{R})$ that

$$\mathrm{SL}_2(\mathbb{Q})\backslash\mathrm{SL}_2(\mathbb{A})/(K \times K_\infty) = \Gamma\backslash\mathcal{H} = Y(\Gamma). \tag{4.2}$$

Definition 4.1. For $N \in \mathbb{Z}_{\geqslant 1}$, consider the reduction-mod-N map $\mathrm{red}_N \colon \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. We consider the subgroups

$$\Gamma(N) := \ker(\mathrm{red}_N), \quad \Gamma_1(N) := \mathrm{red}_N^{-1}\left(\left\{\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}\right\}\right), \quad \Gamma_0(N) := \mathrm{red}_N^{-1}\left(\left\{\begin{pmatrix} * & * \\ 0 & * \end{pmatrix}\right\}\right). \tag{4.3}$$

We say that $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ is a *congruent subgroup* if it contains $\Gamma(N)$ for some N. In the homework you will check that Γ is congruent if and only if there exists K as above with $\Gamma = \mathrm{SL}_2(\mathbb{Q}) \cap K$.

Example 4.1. Denote $K(N), K_1(N), K_0(N) \subseteq \mathrm{SL}_2(\mathbb{A}_f)$ to be the subgroups of elements of $\prod_p \mathrm{SL}_2(\mathbb{Z}_p)$ congruent to $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$ modulo N. Then $\Gamma_?(N) = \mathrm{SL}_2(\mathbb{Q}) \cap K_?(N)$ for all three subscripts $? \in \{\varnothing, 1, 0\}$.

Remark 4.1. There exist finite index subgroups of $\mathrm{SL}_2(\mathbb{Z})$ which are *not* congruent subgroups. The arithmeticity of modular forms behaves wildly different for congruent and non-congruent subgroups. For example, if Γ is congruent then $M_k(\Gamma)$ has a basis of forms with algebraically integral Fourier expansion. Conversely, if a modular form f has algebraically integral Fourier expansion, then it is modular for some congruence subgroup.[3]

In some cases, it is preferable to consider GL_2 instead of SL_2.

Proposition 4.2. *For a number field F, let $K \subseteq \mathrm{GL}_2(\mathbb{A}_{F,f})$ be an open compact subgroup. Then the determinant map*

$$\det\colon \mathrm{GL}_2(F)\backslash \mathrm{GL}_2(\mathbb{A}_{F,f})/K \to F^\times \backslash \mathbb{A}_{F,f}^\times / \det(K) \tag{4.4}$$

is a bijection.

Proof. The map is clearly surjective.

Suppose $\mathrm{GL}_2(F) g_1 K$ and $\mathrm{GL}_2(F) g_2 K$ have the same image. This means that $\det(g_1) \in F^\times \det(g_2) \det(K)$. So we can find $x \in \mathrm{GL}_2(F)$ and $y \in K$ such that $\det(g_1) = \det(x g_2 y)$. Call $t = x g_2 y$.

Since $\mathrm{SL}_2(F)$ is dense in $\mathrm{SL}_2(\mathbb{A}_{F,f})$, it intersects nontrivially with $g_1 t^{-1}(tKt^{-1} \cap \mathrm{SL}_2(\mathbb{A}_{F,f}))$. In particular, there are $a \in \mathrm{SL}_2(F)$ and $b \in K$ with $g_1 t^{-1}(tbt^{-1}) = a$. Rearranging, this is $g_1 = a x g_2 y b^{-1} \in \mathrm{GL}_2(F) g_2 K$. □

Remark 4.2. Since $F^\times \backslash \mathbb{A}_{F,f}^\times / \hat{\mathcal{O}}_F \simeq \mathrm{Cl}(F)$, for open compacts K with $\det(K) = \hat{\mathcal{O}}_F$ we have

$$|\mathrm{GL}_2(F)\backslash \mathrm{GL}_2(\mathbb{A}_{F,f})/K| = \#\mathrm{Cl}(F). \tag{4.5}$$

Proposition 4.3. *Let $K \subseteq \mathrm{GL}_2(\mathbb{A})$ be an open compact subgroup with $\det(K) = \hat{\mathbb{Z}}$. We denote $Z \subseteq \mathrm{GL}_2$ the center, and by $Z(\mathbb{R})^+ \subseteq Z(\mathbb{R})$ the elements with positive entries. The natural inclusion $\mathrm{SL}_2(\mathbb{R}) \to \mathrm{GL}_2(\mathbb{R})$ induces a homeomorphism*

$$\Gamma\backslash \mathrm{SL}_2(\mathbb{R}) \simeq Z(\mathbb{R})^+ \mathrm{GL}_2(\mathbb{Q})\backslash \mathrm{GL}_2(\mathbb{A})/K. \tag{4.6}$$

where $\Gamma = K \cap \mathrm{SL}_2(\mathbb{Q})$.

[3] This is the "Unbounded denominators conjecture", recently proved by Calegari–Dimitrov–Tang.

Proof. By approximation, we have $\mathrm{GL}_2(\mathbb{A}) = \mathrm{GL}_2(\mathbb{Q})\mathrm{GL}_2(\mathbb{R})K$. Since $\mathrm{GL}_2(\mathbb{Q})$ contains elements with negative determinant, we may also write $\mathrm{GL}_2(\mathbb{A}) = \mathrm{GL}_2(\mathbb{Q})\mathrm{GL}_2(\mathbb{R})^+K$. So the right hand side is

$$Z(\mathbb{R})^+\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{Q})\mathrm{GL}_2(\mathbb{R})^+K/K = (Z(\mathbb{R})^+\Gamma)\backslash\mathrm{GL}_2(\mathbb{R})^+ \tag{4.7}$$

for $\Gamma = K \cap \mathrm{GL}_2(\mathbb{Q})^+$. Note that $K \subseteq \prod_p \mathrm{GL}_2(\mathbb{Z}_p)$, and thus $\Gamma = K \cap \mathrm{GL}_2(\mathbb{Z})^+ = K \cap \mathrm{SL}_2(\mathbb{Z}) = K \cap \mathrm{SL}_2(\mathbb{Q})$. Now the claim follows as $\mathrm{SL}_2(\mathbb{R}) = Z(\mathbb{R})^+\backslash\mathrm{GL}_2(\mathbb{R})^+$. □

Example 4.2. Denote $K_1(N), K_0(N) \subseteq \mathrm{GL}_2(\mathbb{A}_f)$ to be the subgroups of elements of $\prod_p \mathrm{GL}_2(\mathbb{Z}_p)$ congruent to $\begin{pmatrix} * & * \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$ modulo N. Then $\Gamma_?(N) = \mathrm{SL}_2(\mathbb{Q}) \cap K_?(N)$ for the two subscripts $? \in \{1, 0\}$.

4.2 Automorphic forms and representations

Reference: [Bum97, Sections 3.2, 3.3].

Let $f\colon \mathcal{H} \to \mathbb{C}$ be a modular form for a congruence subgroup $\Gamma = \mathrm{SL}_2(\mathbb{Q}) \cap K$ of weight k. We consider the function $F\colon \mathrm{GL}_2(\mathbb{R})^+ \to \mathbb{C}$ given by

$$F(g) = \tilde{\nu}_k(g, i)^{-1} f(g \cdot i). \tag{4.8}$$

where $\tilde{\nu}_k(g, z) := \det(g)^{-k/2}\nu_k(g, z)$. We note the following:

1. $\tilde{\nu}_k$ satisfies $\tilde{\nu}_k(g_1g_2, z) = \tilde{\nu}_k(g_1, g_2z)\tilde{\nu}_k(g_2, z)$, and so F is built so that the relation $f(\gamma z) = \tilde{\nu}_k(\gamma, z)f(z)$ translates to $F(\gamma g) = F(g)$. Note also that $\nu_k(g, i)$ is $Z(\mathbb{R})^+$ invariant. So we may think of F as a function

$$F\colon (Z(\mathbb{R})^+\Gamma)\backslash\mathrm{GL}_2(\mathbb{R})^+ = Z(\mathbb{R})^+\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A})/K \to \mathbb{C}. \tag{4.9}$$

2. F is not $SO(2)$-invariant on the right, and rather satisfies

$$F(gt) = \tilde{\nu}_k(gt, i)^{-1} f(gt \cdot i) = \frac{\tilde{\nu}_k(g, i)}{\tilde{\nu}_k(gt, i)} F(g) = \tilde{\nu}_k(t, i)^{-1} F(g), \quad t \in SO(2). \tag{4.10}$$

We write $SO(2) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \right\}$,[4] and if $t = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, we have $\tilde{\nu}_k(t, i)^{-1} = e^{ik\theta}$. Hence $F(gt) = e^{ik\theta}F(g)$.

[4] Note that this is parametrizing $SO(2)$ *clockwise*. This unusual convention is so that the character at the end is $e^{ik\theta}$.

3. Since f is holomorphic, it satisfies the Cauchy–Riemann equation: if we denote $\frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}$, we have $\frac{\partial f}{\partial \bar{z}} = 0$. This is translated to an expression for F involving the Lie algebra action of $\mathfrak{gl}_{2,\mathbb{C}}$ in $C^\infty(\mathrm{GL}_2(\mathbb{R}), \mathbb{C})$. We recall that this action is given as follows: for $X \in \mathfrak{gl}_2$ and $G \in C^\infty(\mathrm{GL}_2(\mathbb{R}), \mathbb{C})$, we consider the function $XG \in C^\infty(\mathrm{GL}_2(\mathbb{R}), \mathbb{C})$ given by

$$XG(g) = \frac{\mathrm{d}}{\mathrm{d}t} G(g \exp(tX))|_{t=0}. \tag{4.11}$$

We denote R, L, H, Z the following generators of $\mathfrak{gl}_{2,\mathbb{C}}$:

$$R = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.12}$$

Under the coordinates

$$g = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} y^{1/2} & xy^{-1/2} \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{4.13}$$

we have $\tilde{\nu}_k(g, i) = y^{-k/2} e^{-ik\theta}$ and we can compute that

$$R = e^{-2i\theta}\left(2iy\frac{\partial}{\partial x} + 2y\frac{\partial}{\partial y} - i\frac{\partial}{\partial \theta}\right) = e^{-2i\theta}\left(2iy\frac{\partial}{\partial z} - i\frac{\partial}{\partial \theta}\right), \tag{4.14}$$

$$L = e^{2i\theta}\left(-2iy\frac{\partial}{\partial x} + 2y\frac{\partial}{\partial y} + i\frac{\partial}{\partial \theta}\right) = e^{2i\theta}\left(-2iy\frac{\partial}{\partial \bar{z}} + i\frac{\partial}{\partial \theta}\right), \tag{4.15}$$

$$H = -i\frac{\partial}{\partial \theta}, \tag{4.16}$$

$$Z = a\frac{\partial}{\partial a}. \tag{4.17}$$

We have

$$0 = \frac{\partial f}{\partial \bar{z}}(g \cdot i) = \tilde{\nu}_k(g, i)\frac{\partial F(g)}{\partial \bar{z}} + F(g)\frac{\partial \tilde{\nu}_k(g, i)}{\partial \bar{z}} = \tilde{\nu}_k(g, i)\left(\frac{\partial F}{\partial \bar{z}} - \frac{ikF}{2y}\right)(g) \tag{4.18}$$

$$0 = \frac{\partial f}{\partial \theta}(g \cdot i) = \tilde{\nu}_k(g, i)\frac{\partial F(g)}{\partial \theta} + F(g)\frac{\partial \tilde{\nu}_k(g, i)}{\partial \theta} = \tilde{\nu}_k(g, i)\left(\frac{\partial F}{\partial \theta} - ikF\right)(g) \tag{4.19}$$

$$0 = \frac{\partial f}{\partial a}(g \cdot i) = \tilde{\nu}_k(g, i)\frac{\partial F(g)}{\partial a} + F(g)\frac{\partial \tilde{\nu}_k(g, i)}{\partial a} = \tilde{\nu}_k(g, i)\frac{\partial F}{\partial a}(g) \tag{4.20}$$

In other words, F is killed by

$$Z, \quad L, \quad \text{and} \quad H - k. \tag{4.21}$$

This discussion proves that F is an *automorphic form*, as follows. Note that the second point holds since f is holomorphic at the cusps.

Definition 4.2. The space of automorphic forms $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$ is the space of smooth functions $F\colon \mathrm{GL}_2(\mathbb{A}) \to \mathbb{C}$ satisfying

1. F is $\mathrm{GL}_2(\mathbb{Q})$-invariant on the left.
2. F is *slowly increasing*: for each $g_f \in \mathrm{GL}_2(\mathbb{A}_f)$, the quantity $F(g_f, g_\infty)$ is bounded by a polynomial in the entries of g_∞ and g_∞^{-1}.
3. F is *right $O(2)$-finite*: the space spanned by the functions $\{F(\cdot k)\colon k \in O(2)\}$ is finite dimensional.
4. There is an ideal $I \subseteq Z(U(\mathfrak{gl}_2))$ of finite codimension which annihilates F.

Remark 4.3. $Z(U(\mathfrak{gl}_2)) = \mathbb{C}[Z, \Delta]$ where $\Delta = -\frac{1}{4}\left(H^2 + \frac{RL}{2} + \frac{LR}{2}\right) = -\frac{1}{4}\left(H^2 - 2H + RL\right)$. In particular, if F is associated to a modular form of weight k as above, then $\Delta F = \frac{k}{2}\left(1 - \frac{k}{2}\right) F$.

The space $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$ has the following pieces of structure:

1. It is a $\mathrm{GL}_2(\mathbb{A}_f)$-module by right multiplication.
2. It is a $(\mathfrak{gl}_2, O(2))$-module: namely it carries actions of both $\mathfrak{gl}_2$ and $O(2)$ (by right multiplication), such that they induce the same action of $\mathfrak{so}_2$.

It is common to say that a space with two commuting actions as above is a $\mathrm{GL}_2(\mathbb{A})$-*representation*, although beware that in the archimedean place there is no $\mathrm{GL}_2(\mathbb{R})$-action, only actions of $\mathfrak{gl}_2$ and $O(2)$.

Definition 4.3. Consider D_k to be the $(\mathfrak{gl}_2, O(2))$-module defined to have basis $R^i v, L^i \bar{v}$ for $i \geqslant 0$ where $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} v = \bar{v}$ and such that $Zv = 0, Lv = 0, Hv = kv$ and $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} v = e^{ik\theta} v$.

Corollary 4.4. *We have*

$$\varinjlim_{\Gamma \text{ congruence subgroup}} M_k(\Gamma) \simeq \mathrm{Hom}_{(\mathfrak{gl}_2, O(2))}(D_k, \mathcal{A}(\mathrm{GL}_2(\mathbb{A})). \tag{4.22}$$

Consider a modular form $f\colon \mathcal{H} \to \mathbb{C}$. What does cuspidality mean in terms of F? For a modular form f of level Γ, let $\Gamma_\infty = \Gamma \cap N(\mathbb{Q})$, where N is the upper triangular unipotent subgroup. Then $\Gamma_\infty = \left\{ \begin{pmatrix} 1 & W* \\ 0 & 1 \end{pmatrix} \right\}$ for some $W \in \mathbb{Z}$. We have the Fourier expansion $f(z) = \sum_{n \geqslant 0} a_n(f) q_W^n$. Here, the constant term is

$$a_0(f) = \int_0^W f(x + iy)\,\mathrm{d}x \tag{4.23}$$

Recall that $x+iy$ is equal to $\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}y^{1/2} & 0\\ 0 & y^{-1/2}\end{pmatrix}\cdot i$ and so for $g=\begin{pmatrix}y^{1/2} & 0\\ 0 & y^{-1/2}\end{pmatrix}$, we may write the constant term as

$$\int_{\Gamma_\infty\backslash N(\mathbb{R})}\nu_k(ng,i)F(ng)\,\mathrm{d}n=\int_{\Gamma_\infty\backslash N(\mathbb{R})}\nu_k(n,gi)\nu_k(g,i)F(ng)\,\mathrm{d}n=\nu_k(g,i)\int_{\Gamma_\infty\backslash N(\mathbb{R})}F(ng)\,\mathrm{d}n \tag{4.24}$$

Now denoting $K_N=K\cap N(\mathbb{A}_f)$ and using that $N(\mathbb{A}_f)=N(\mathbb{Q})K_N$, we have

$$N(\mathbb{Q})\backslash N(\mathbb{A})/K_N=(K_N\cap N(\mathbb{Q}))\backslash N(\mathbb{R})=\Gamma_\infty\backslash N(\mathbb{R}), \tag{4.25}$$

and thus

$$a_0(f)=\nu_k(g,i)\cdot\int_{N(\mathbb{Q})\backslash N(\mathbb{A})}F(ng)\,\mathrm{d}n. \tag{4.26}$$

Definition 4.4. For $F\in\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$, we consider its *constant term (along N)* to be the function $c_{F,N}\colon\mathrm{GL}_2(\mathbb{A})\to\mathbb{C}$ given by

$$c_{F,N}(g)=\int_{N(\mathbb{Q})\backslash N(\mathbb{A})}F(ng)\,\mathrm{d}n. \tag{4.27}$$

We say F is a *cuspidal automorphic form* if $c_{F,N}$ is identically zero. We denote $\mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}))\subseteq\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$ the subspace of such cuspidal automorphic forms.

Remark 4.4. If F comes from a modular form f as before, note that the function $c_{F,N}$ accounts for constant terms at all cusps: if $\gamma\in\mathrm{SL}_2(\mathbb{Z})$, then the modular form $f[\gamma]_k$ corresponds to the function $\mathrm{GL}_2(\mathbb{R})^+\to\mathbb{C}$ given by

$$\big(g\mapsto\tilde{\nu}_k(g,i)^{-1}f[\gamma]_k(g\cdot i)=\tilde{\nu}_k(g,i)^{-1}\tilde{\nu}_k(\gamma,gi)^{-1}f(\gamma g\cdot i)=\tilde{\nu}_k(\gamma g,i)^{-1}f(\gamma g\cdot i)\big) \tag{4.28}$$

which is $F(\gamma(\cdot))\colon\mathrm{GL}_2(\mathbb{R})^+\to\mathbb{C}$. Adelically, this corresponds to $F((\cdot)\gamma_f^{-1})\colon\mathrm{GL}_2(\mathbb{A})\to\mathbb{C}$ where $\gamma_f\in\mathrm{GL}_2(\mathbb{A}_f)$ is the image of γ, and hence its constant term along N is $c_{F,N}((\cdot)\gamma_f^{-1})$.

Definition 4.5. An *irreducible automorphic representation* (π,V) is an irreducible $\mathrm{GL}_2(\mathbb{A})$-representation[5] which is isomorphic to a subspace of $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$. We say it is *cuspidal* if it is isomorphic to a subspace of $\mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}))$.

Remark 4.5. By a version of Schur's lemma, every irreducible automorphic representation (π,V) has a central character $\omega\colon\mathbb{Q}^\times\backslash\mathbb{A}^\times\to\mathbb{C}$. By twisting (π,V) by a power of $|\det|$, we may always assume that ω is unitary.

Some (hard) facts about automorphic representations:

[5]Recall that this means that it is a $\mathrm{GL}_2(\mathbb{A}_f)$-representation and a $(\mathfrak{gl}_2,O(2))$-module.

Theorem 4.5 (Harish–Chandra). *Given $I \subseteq Z(U(\mathfrak{gl}_2))$ of finite codimension, ρ an irreducible finite dimensional representation of $K_\infty = O(2)$ and $K \subseteq \mathrm{GL}_2(\mathbb{A}_f)$ an open compact subgroup, consider the subspace $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}), I, \rho)^K \subseteq \mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$ of automorphic forms F which i) is killed by I, ii) satisfy that the (finite dimensional) K_∞-representation $\{F(\cdot t) \colon t \in K_\infty\}$ is ρ-isotypical and iii) is K-invariant. Then $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}), I, \rho)^K$ is finite dimensional.*

Theorem 4.6 (Flath's tensor product theorem). *Let π be an irreducible automorphic representation.*[6] *Then there exist local representations π_v which are irreducible smooth representations of $\mathrm{GL}_2(\mathbb{Q}_p)$ when $v = p$ is non-archimedean, and π_∞ is an irreducible $(\mathfrak{gl}_2, O(2))$-module, such that*

$$\pi = \bigotimes_v{}' \, \pi_v. \tag{4.29}$$

Here, almost all π_p contain a $\mathrm{GL}_2(\mathbb{Z}_p)$-fixed vector ξ_p,[7] *and the restricted product is taken with respect to ξ_p.*

Theorem 4.7. *Let $\omega \colon \mathbb{Q}^\times \backslash \mathbb{A}^\times \to \mathbb{C}^\times$ be a unitary Hecke character. We consider $\mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}), \omega) \subseteq \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}))$ the subspace of automorphic forms with central character ω. We similarly consider $L^2(\mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A}), \omega)$ functions with central character ω which are in L^2 with respect to the inner product*

$$\langle f_1, f_2 \rangle = \int_{Z(\mathbb{A})\mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A})} f_1(g) \overline{f_2(g)} \, \mathrm{d}g. \tag{4.30}$$

Then $\mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}), \omega) \subseteq L^2(\mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A}), \omega)$. In particular, irreducible cuspidal automorphic representations are unitarizable.

For a cuspidal irreducible automorphic representation of $\mathrm{GL}_2(\mathbb{A})$, one can define an L-function $L(s, \pi)$ and prove its analytic properties and functional equation à la Tate's thesis. This was done by Godement–Jacquet. Without loss of generality, we may twist π by a power of $|\det|$ to assume that its central character ω is unitary (in this case the function equation will have center at $s = 1/2$). Godement–Jacquet considers zeta integrals

$$Z(s, \Phi, \beta) = \int_{\mathrm{GL}_2(\mathbb{A})} \Phi(g) \beta(g) |\det(g)|^{s + \frac{1}{2}} \, \mathrm{d}^\times g \tag{4.31}$$

[6] More generally, we can take π to be an *admissible* $\mathrm{GL}_2(\mathbb{A})$-representation. Here (π, V) being admissible means that for any every open compact $K \subseteq \mathrm{GL}_2(\mathbb{A}_f)$ and irreducible finite dimensional representation ρ of $K_\infty \times K$, the isotypic component $V(\rho)$ of ρ in V is finite dimensional. An irreducible automorphic representation is admissible by the above theorem of Harish–Chandra.

[7] In such case, $\pi_p^{K_p}$ is necessarily one-dimensional.

where $\Phi \in \mathcal{S}(\mathrm{Mat}_{2\times 2}(\mathbb{A}))$ and β are *matrix coefficients of* π: for $f_1, f_2 \in \pi$, we consider

$$\beta(g) = \beta(f_1, f_2, g) = \langle \pi(g) f_1, f_2 \rangle = \int_{Z(\mathbb{A})\mathrm{GL}_2(\mathbb{Q})\backslash \mathrm{GL}_2(\mathbb{A})} f_1(hg)\overline{f_2(h)}\,\mathrm{d}h. \tag{4.32}$$

5 Newform theory

Remark 5.1. We will focus on studying $M_k(\Gamma_0(N))$ (since these contain the modular forms corresponding to elliptic curves), although most of what we will do extend to $\Gamma_1(N)$. Furthermore, noting that

$$\begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}^{-1} \Gamma(N) \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} = \Gamma_0(N^2) \cap \Gamma_1(N) \supseteq \Gamma_1(N^2), \tag{5.1}$$

we have $\left[\begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}\right]_k : M_k(\Gamma(N)) \hookrightarrow M_k(\Gamma_1(N^2))$. So studying modular forms for congruence subgroups is reduced to studying modular forms for $\Gamma_1(N)$ for all N. In other words: $\varinjlim_N M_k(\Gamma(N)) = \varinjlim_N M_k(\Gamma_1(N))$.

One of the reasons we focus on the case $\Gamma = \Gamma_0(N)$ is because $Z(\hat{\mathbb{Z}}) \subseteq K_0(N)$. This is not true for $\Gamma_1(N)$, and to extend the discussion to $\Gamma_1(N)$ one needs to be a bit more careful with the action of the center. In other words, using that $K_0(N) = K_1(N)Z(\hat{\mathbb{Z}})$ and the above remark, we have

$$\varinjlim_N M_k(\Gamma_0(N)) = \mathrm{Hom}_{(\mathfrak{gl}_2, O(2))}\left(D_k, \mathcal{A}(\mathrm{GL}_2(\mathbb{A}))^{Z(\mathbb{A})}\right) = \mathrm{Hom}_{(\mathfrak{gl}_2, O(2))}\left(D_k, \mathcal{A}(\mathrm{PGL}_2(\mathbb{A}))\right). \tag{5.2}$$

5.1 Petersson inner product and Hecke operators

References: [DS05, Sections 5.1-5.5].

We note two pieces of extra structure that comes from thinking of modular forms adelically. We will denote $[\mathrm{GL}_2(\mathbb{A})] := Z(\mathbb{R})^+\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A})$, and recall that

$$[\mathrm{GL}_2(\mathbb{A})]/K = Z(\mathbb{R})^+\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A})/K = (K \cap \mathrm{SL}_2(\mathbb{Q}))\backslash\mathrm{SL}_2(\mathbb{R}). \tag{5.3}$$

First, we can try to consider an inner product on the space of functions $\mathcal{C}^\infty([\mathrm{GL}_2(\mathbb{A})])$. This is of course not quite true because of convergence issues, but let's ignore that and work formally for now. We want to consider

$$(F_1, F_2) = \int_{[\mathrm{GL}_2(\mathbb{A})]} F_1(g)\overline{F_2(g)}\,\mathrm{d}g \tag{5.4}$$

and if F_1, F_2 correspond to modular forms $f_1, f_2 \in M_k(\Gamma)$, this is

$$(F_1, F_2) = \int_{\Gamma\backslash \mathrm{SL}_2(\mathbb{R})} |\nu_k(g,i)|^{-2} f_1(gi)\overline{f_2(gi)}\,\mathrm{d}g = \int_{\Gamma\backslash\mathcal{H}} f_1(z)\overline{f_2(z)}y^{k-2}\,\mathrm{d}x\,\mathrm{d}y. \tag{5.5}$$

Here we are using that the Haar measure on $\mathrm{GL}_2(\mathbb{R})^+$ is $\frac{\mathrm{d}a\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\theta}{y^2}$.

Definition 5.1. For $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ a finite index subgroup, the *Petersson inner product* is the pairing $(\cdot,\cdot)\colon M_k(\Gamma) \times S_k(\Gamma) \to \mathbb{C}$ given by

$$(f_1, f_2) = \int_{\Gamma\backslash\mathcal{H}} f_1(z)\overline{f_2(z)}y^{k-2}\,\mathrm{d}x\,\mathrm{d}y. \tag{5.6}$$

We note that this is well-defined since f_2, being a cusp form, decays exponentially at all cusps.

Secondly, we have an action of $\mathrm{GL}_2(\mathbb{A}_f)$ by right multiplication on $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$. Since we are interested in the spaces $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))^K$, we can consider the following convolution operators.

Definition 5.2. For $K \subseteq \mathrm{GL}_2(\mathbb{A}_f)$ an open compact, the *Hecke algebra of level* K is the space $\mathcal{H}_K := \mathcal{C}_c^\infty(K\backslash\mathrm{GL}_2(\mathbb{A}_f)/K)$ with algebra structure given by convolution. The space $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))^K$ is a right $\mathcal{H}_K$-module via

$$(f * \phi)(g) = \int_{\mathrm{GL}_2(\mathbb{A}_f)} f(gh^{-1})\phi(h)\,\mathrm{d}h. \tag{5.7}$$

Here we normalize the local Haar measures $\mathrm{d}h_p$ so that K_p has volume 1. This makes it so that the identity element of $\mathcal{H}_K$ is simply $\mathrm{char}(K)$.

We also can consider local Hecke algebras $\mathcal{H}_{K_p} = \mathcal{C}_c^\infty(K_p\backslash\mathrm{GL}_2(\mathbb{Q}_p)/K_p)$, and we naturally have $\mathcal{H}_K = \otimes_p' \mathcal{H}_{K_p}$ where the restricted product is over the identity elements $e_p = \mathrm{char}(K_p)$.

Such Hecke algebras are very complicated in general, but when K_p is a maximal open compact subgroup, they are very simple to describe.

Theorem 5.1. *If $K_p = \mathrm{GL}_2(\mathbb{Z}_p)$, then $\mathcal{H}_{K_p} = \mathbb{C}[S_p, R_p^{\pm 1}]$ where $S_p = \mathrm{char}\left(K_p \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} K_p\right)$ and $R_p = \mathrm{char}\left(K_p \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} K_p\right)$.*

Definition 5.3. For any $K_p \subseteq \mathrm{GL}_2(\mathbb{Q}_p)$ open compact, we consider the elements $S_p, R_p \in \mathcal{H}_{K_p}$ to be $S_p := \mathrm{char}\left(K_p \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} K_p\right)$ and $R_p := \mathrm{char}\left(K_p \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} K_p\right)$.

Remark 5.2. Note that if $\pi \subseteq \mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$ is an irreducible automorphic representation with central character $\omega\colon \mathbb{Q}^\times\backslash\mathbb{A}^\times \to \mathbb{C}^\times$, then R_p acts on π simply by $\omega_p(p)^{-1}$. In particular, R_p acts trivially on $\mathcal{A}(\mathrm{GL}_2(\mathbb{A}))^{K_0(N)}$.

We now translate this action of the Hecke algebra classically.

Definition 5.4. Let $\Gamma_1, \Gamma_2 \subseteq \mathrm{SL}_2(\mathbb{Z})$ be two finite index subgroups and $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$. We define the double coset operator

$$[\Gamma_1\alpha\Gamma_2]_k\colon M_k(\Gamma_1) \to M_k(\Gamma_2). \tag{5.8}$$

If we write $\Gamma_1\alpha\Gamma_2 = \bigsqcup_i \Gamma_1\gamma_i$, this is defined as

$$f[\Gamma_1\alpha\Gamma_2]_k := \sum_i f[\gamma_i]_k. \tag{5.9}$$

When $\Gamma_1 = \Gamma_2 = \Gamma$, we denote $T_p\colon M_k(\Gamma) \to M_k(\Gamma)$ to be $T_p = \left[\Gamma \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma\right]_k$.

Remark 5.3. This definition of T_p is valid for all finite index subgroups $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$. However, this gives essentially trivial operators for non-congruence subgroups: If $\Gamma' \supseteq \Gamma$ is the smallest congruence subgroup containing Γ, we consider the projection map $[\Gamma\Gamma']_k\colon M_k(\Gamma) \to M_k(\Gamma')$ and let $M_k(\Gamma)^{\mathrm{prim}}$ denote its kernel. A conjecture of Atkin, proved by Berger, says that all T_p act trivially on $M_k(\Gamma)^{\mathrm{prim}}$.

Proposition 5.2. *Let $\Gamma = K \cap \mathrm{SL}_2(\mathbb{Q})$ be a congruence subgroup for some open compact $K \subseteq \mathrm{GL}_2(\mathbb{A}_f)$ with $\det(K) = \hat{\mathbb{Z}}$. Let $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$. Then if $f \in M_k(\Gamma)$ corresponds to $F \in \mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$, we have that $f[\Gamma\alpha\Gamma]_k$ corresponds to $F * \phi \in \mathcal{A}(\mathrm{GL}_2(\mathbb{A}))$ where*

$$\phi = \det(\alpha)^{k/2-1}\mathrm{char}(K\alpha K) \in \mathcal{H}_K. \tag{5.10}$$

Proof. We write $\Gamma\alpha\Gamma = \bigsqcup_j \Gamma\gamma_j$. Recall that we defined

$$(f[\gamma]_k)(z) = \det(\gamma)^{k-1}\nu_k(\gamma, z)^{-1}f(\gamma \cdot z), \tag{5.11}$$

and that we attached to f the function $F\colon \Gamma\backslash\mathrm{GL}_2(\mathbb{R})^+ \to \mathbb{C}$ given by

$$F(g) = \det(g)^{k/2}\nu_k(g, i)^{-1}f(g \cdot i). \tag{5.12}$$

So if we let $f' = f[\Gamma\alpha\Gamma]_k$, we have

$$\begin{aligned} F'(g) &= \det(g)^{k/2}\nu_k(g, i)^{-1}\sum_j (f[\gamma_j]_k)(g \cdot i) \\ &= \det(\alpha)^{k-1}\sum_j \det(g)^{k/2}\nu_k(g, i)^{-1}\nu_k(\gamma_j, g \cdot i)^{-1}f(\gamma_j g \cdot i). \end{aligned} \tag{5.13}$$

Noting that $\det(\gamma_j) = \det(\alpha)$, this is

$$F'(g) = \det(\alpha)^{k/2-1} \sum_j \det(\gamma_j g)^{k/2} \nu_k(\gamma_j g, i)^{-1} f(\gamma_j g \cdot i) = \det(\alpha)^{k/2-1} \sum_j F(\gamma_j g). \tag{5.14}$$

Now from $\Gamma \backslash \mathrm{GL}_2(\mathbb{R})^+ \simeq Z(\mathbb{R})^+ \mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A}) / K$, we have the automorphic forms $\tilde{F}, \tilde{F}' \colon Z(\mathbb{R})^+ \mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A}) \to \mathbb{C}$. If $(g_f, g_\infty) \in \mathrm{GL}_2(\mathbb{A})$, we take $\beta \in \mathrm{GL}_2(\mathbb{Q})$ such that $g_f K = \beta K$, and then

$$\tilde{F}'(g_f, g_\infty) = F'(\beta^{-1} g_\infty) = \det(\alpha)^{k/2-1} \sum_j F(\gamma_j \beta^{-1} g_\infty) = \det(\alpha)^{k/2-1} \sum_j \tilde{F}(\beta \gamma_j^{-1}, g_\infty). \tag{5.15}$$

Recall that the closure of Γ in $\mathrm{GL}_2(\mathbb{Q})$ is K. So taking closure of $\Gamma \alpha \Gamma = \bigsqcup_j \Gamma \gamma_j$, we obtain that $K \alpha K = \bigsqcup_j K \gamma_j$. Hence

$$\bigsqcup_j \beta \gamma_j^{-1} K = \beta \bigsqcup \gamma_j^{-1} K = \beta K \alpha^{-1} K = g_f K \alpha^{-1} K = g_f \bigsqcup_j \gamma_j^{-1} K = \bigsqcup_j g_f \gamma_j^{-1} K, \tag{5.16}$$

and thus

$$\tilde{F}'(g_f, g_\infty) = \det(\alpha)^{k/2-1} \sum_j \tilde{F}(g_f \gamma_j^{-1}, g_\infty) = \tilde{F} * \phi. \tag{5.17}$$

□

Corollary 5.3. *For $\Gamma = \Gamma_0(N)$ and any p, the action of T_p corresponds to the action of $p^{k/2-1} S_p$. In particular, the operators T_p on $M_k(\Gamma_0(N))$ commute for varying p.*

Proof. The first claim follows from the above as $\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$ is in $K_0(N)_\ell$ for all $\ell \neq p$. The second claim then follows as the S_p commute. □

Abstractly, it is easy to see how the action of $\mathcal{H}_K$ interacts with the Petersson inner product. For any $\phi \in \mathcal{H}_K$ and $F_1, F_2 \in \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}))$, we have

$$\begin{aligned} (F_1, \phi * F_2) &= \int_{[\mathrm{GL}_2(\mathbb{A})]} F_1(g) \overline{\int_{\mathrm{GL}_2(\mathbb{A})} F_2(gh^{-1}) \phi(h) \, \mathrm{d}h} \, \mathrm{d}g \\ &= \iint_{[\mathrm{GL}_2(\mathbb{A})] \times \mathrm{GL}_2(\mathbb{A})} F_1(gh) \overline{F_2(g) \phi(h)} \, \mathrm{d}(g, h) \\ &= \int_{[\mathrm{GL}_2(\mathbb{A})]} \left(\int_{\mathrm{GL}_2(\mathbb{A})} F_1(gh^{-1}) \overline{\phi(h^{-1})} \, \mathrm{d}h \right) \overline{F_2(g)} \, \mathrm{d}g \end{aligned} \tag{5.18}$$

which is simply $(\tilde{\phi} * F_1, F_2)$ for $\tilde{\phi} := (h \mapsto \overline{\phi(h^{-1})})$.

Corollary 5.4. *For $K_p = \mathrm{GL}_2(\mathbb{Z}_p)$ we have $\tilde{S}_p = S_p R_p^{-1}$. In particular, T_p for $p \nmid N$ is self-adjoint under the Petersson inner product on $S_k(\Gamma_0(N))$.*

Proof. The first claim follows from the above discussion as $\begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix}$ and $\begin{pmatrix} p^{-1} & 0 \\ 0 & 1 \end{pmatrix}$ are in the same double coset of $\mathrm{GL}_2(\mathbb{Z}_p)$.

The second claim then follows since T_p corresponds to $p^{k/2-1}S_p$ and R_p acts trivially on $S_k(\Gamma_0(N))$. □

Corollary 5.5. *The space $S_k(\Gamma_0(N))$ has a basis of eigenfunctions for the Hecke operators T_p with $p \nmid N$.*

Proof. By the results above, T_p are commuting self-adjoint operators, and so they are simultaneously diagonalizable by the spectral theorem in linear algebra. □

Remark 5.4. What does this correspond to in automorphic terms? Since cuspidal automorphic representations are unitarizable we have an orthogonal decomposition of $\mathcal{H}_{K_0(N)}$-modules

$$S_k(\Gamma_0(N)) = \bigoplus_{\substack{\pi \subseteq \mathcal{A}(\mathrm{GL}_2(\mathbb{A})) \\ \pi_\infty \simeq D_k}} \pi_f^{K_0(N)} \otimes V_\pi \tag{5.19}$$

for some multiplicity spaces $V_\pi = \mathrm{Hom}(\pi, \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A})))$. For $p \nmid N$, we have that $\pi_p^{K_0(N)_p}$ is one dimensional and in particular $\mathcal{H}_{K_0(N)_p}$ acts by scalars on $\pi_f^{K_0(N)}$. So a basis of eigenfunctions as above can be obtained simply by taking a basis of each of the terms on the right hand side.

We can also describe the effects of T_p on the Fourier expansion.

Proposition 5.6. *For $\Gamma = \Gamma_0(N)$, we have that*

$$a_n(T_p f) = \begin{cases} a_{np}(f) & \text{if } p \mid N, \\ a_{np}(f) + p^{k-1} a_{n/p}(f) & \text{if } p \nmid N. \end{cases} \tag{5.20}$$

In particular, $a_1(T_p f) = a_p(f)$.

Proof. We have that

$$T_p f = 1_N(p) \cdot f \left[\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \right]_k + \sum_{j=0}^{p-1} f \left[\begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix} \right]_k \tag{5.21}$$

and thus

$$(T_p f)(z) = p^{k-1} \sum_{n \geqslant 0} a_n(f) \left(1_N(p) e^{2i\pi npz} + \sum_{j=0}^{p-1} p^{-k} e^{2\pi i n(z+j)/p} \right) \tag{5.22}$$

which is

$$\sum_{n \geqslant 0} a_n(f) \left(p^{k-1} 1_N(p) q^{np} + \frac{e^{2\pi i n z/p}}{p} \sum_{j=0}^{p-1} e^{2\pi i n j/p} \right). \tag{5.23}$$

This implies that $a_n(T_p f) = a_{np}(f) + p^{k-1} 1_N(p) a_{n/p}(f)$. □

Definition 5.5. For any $p \nmid N$, we recursively define operators $T_{p^n} \colon M_k(\Gamma_0(N)) \to M_k(\Gamma_0(N))$ by

$$T_{p^{n+1}} = T_p T_{p^n} - p^{k-1} T_{p^{n-1}} \tag{5.24}$$

for $n \geqslant 1$. By the problem set, such T_{p^n} corresponds to $p^{n(k/2-1)} S_{p^n}$ where

$$S_{p^n} = \operatorname{char}\left(\mathrm{GL}_2(\mathbb{Z}_p) \begin{pmatrix} 1 & 0 \\ 0 & p^n \end{pmatrix} \mathrm{GL}_2(\mathbb{Z}_p) \right) \in \mathcal{H}_{\mathrm{GL}_2(\mathbb{Z}_p)}. \tag{5.25}$$

For $p \mid N$, we denote $T_{p^n} = T_p^n$. We also define T_n for $n \in \mathbb{Z}_{\geqslant 1}$ multiplicatively.

This is built such that $a_1(T_n f) = a_n(f)$. Soon, we will define *newforms* and prove that they are Hecke eigenfunctions for *all* Hecke operators.

5.2 Fricke involution and geometric description

References: [DS05, Sections 5.1-5.5].

We also give a description of Hecke operators in terms of the moduli problem. Recall that we have $M_{2k}(\Gamma_0(N)) = H^0(X_0(N), \Omega^{\otimes k}(\cdots))$. From the inclusion $\Gamma_0(Np) \subseteq \Gamma_0(N)$, we have an induced map $p_1 \colon X_0(Np) \to X_0(N)$. We also have that

$$\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}^{-1} \Gamma_0(Np) \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \Gamma_0(N) \cap \Gamma^0(p) \subseteq \Gamma_0(N), \tag{5.26}$$

which induces a map $p_2 \colon X_0(Np) \to X_0(N)$. This gives us a correspondence $p_{2,*} p_1^*$ which we also denote T_p because of the following proposition.

$$\begin{array}{ccccc} & & X_0(Np) & & \\ & \swarrow^{p_1} & & \searrow^{p_2} & \\ X_0(N) & & & & X_0(N) \end{array} \tag{5.27}$$

Proposition 5.7. *Consider $\Gamma_1, \Gamma_2 \subseteq \mathrm{SL}_2(\mathbb{Q})$ and $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$, and consider the diagram*

$$\begin{array}{ccc} X(\Gamma_1 \cap \alpha\Gamma_2\alpha^{-1}) & \xrightarrow{\sim} & X(\alpha^{-1}\Gamma_1\alpha \cap \Gamma_2) \\ \downarrow{\scriptstyle p_1} & & \downarrow{\scriptstyle p_2} \\ X(\Gamma_1) & & X(\Gamma_2). \end{array} \tag{5.28}$$

Then $\det(\alpha)^{k-1} p_{2,*} p_1^*$ corresponds to $[\Gamma_1\alpha\Gamma_2]_k$.

Proof. If $\Gamma_1\alpha\Gamma_2 = \bigsqcup_j \Gamma_1\gamma_j$, then $\Gamma_1(\alpha\Gamma_2\alpha^{-1}) = \bigsqcup_j \Gamma_1(\gamma_j\alpha^{-1})$, and thus we have a commutative diagram

$$\begin{array}{ccccc} M_k(\Gamma_1) & \xrightarrow{p_1^*} & M_k(\Gamma_1 \cap \alpha\Gamma_2\alpha^{-1}) & \xrightarrow{\alpha} & M_k(\alpha^{-1}\Gamma_1\alpha \cap \Gamma_2) \\ \downarrow{\scriptstyle \Gamma_1\alpha\Gamma_2} & & \downarrow & & \downarrow{\scriptstyle p_{2,*}} \\ M_k(\Gamma_2) & \xrightarrow{\alpha^{-1}} & M_k(\alpha\Gamma_2\alpha^{-1}) & \xrightarrow{\qquad\alpha\qquad} & M_k(\Gamma_2) \end{array} \tag{5.29}$$

where a label γ denote the action of $\det(\gamma)^{1-k}[\gamma]_k$. □

In terms of the moduli description of $Y_0(N) \leftrightarrow \{(E, C)\}/\simeq$, we have

$$p_1(E, C) = (E, pC), \quad p_2(E, C) = (E/NC, C). \tag{5.30}$$

Equivalently, in terms of $Y_0(N) \leftrightarrow \{(E_1 \to E_2)\}/\simeq$, we have

$$p_1(E_1 \to E_2) = (E_1 \to E_2'), \quad p_2(E_1 \to E_2) = (E_1' \to E_2) \tag{5.31}$$

where $E_1 \to E_1' \to E_2$ and $E_1 \to E_2' \to E_2$ are the unique ways of factoring the $\mathbb{Z}/Np\mathbb{Z}$-isogeny $E_1 \to E_2$ such that $E_1 \to E_1'$ and $E_2' \to E_2$ are of degree p.

Proposition 5.8. *Let $w_N \colon Y_0(N) \to Y_0(N)$ be the involution given by the dual isogeny: $w_N(E_1 \xrightarrow{\phi} E_2) = (E_2 \xrightarrow{\hat{\phi}} E_1)$. Then w_N is given $\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}$, that is, $\tau \mapsto -1/(N\tau)$. In particular, this extends to $w_N \colon X_0(N) \to X_0(N)$ and $p_2 = w_N p_1 w_{Np}$.*

Proof. In terms of lattices, $w_N(\Lambda_1 \subseteq \Lambda_2) = (\Lambda_2 \subseteq \frac{1}{N}\Lambda_1)$. Recalling that $\tau \in Y_0(N)$ corresponds to $(\mathbb{Z} \oplus \tau\mathbb{Z} \subseteq \frac{1}{N}\mathbb{Z} \oplus \tau\mathbb{Z})$, we have

$$w_N(\tau) \leftrightarrow w_N(\mathbb{Z}\oplus\tau\mathbb{Z} \subseteq \frac{1}{N}\mathbb{Z}\oplus\tau\mathbb{Z}) = (\frac{1}{N}\mathbb{Z}\oplus\tau\mathbb{Z} \subseteq \frac{1}{N}(\mathbb{Z}\oplus\tau\mathbb{Z})) = (\frac{1}{N\tau}\mathbb{Z}\oplus\mathbb{Z} \subseteq \frac{1}{N\tau}\mathbb{Z}\oplus\frac{1}{N}\mathbb{Z}) \tag{5.32}$$

which corresponds to $-1/(N\tau)$. □

Definition 5.6. We denote $w_N \colon M_k(\Gamma_0(N)) \to M_k(\Gamma_0(N))$ the *Fricke involution* to be the operator

$$w_N = i^k \sqrt{N}^{2-k} \left[\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} \right]_k, \tag{5.33}$$

that is,

$$(w_N f)(\tau) = i^k \sqrt{N}^k (N\tau)^{-k} f(-1/(N\tau)). \tag{5.34}$$

Adelically, we have $w_N = \prod_{p|N} w_{p^{\nu_p(N)}}$ where w_{p^e} is

$$\operatorname{char}\left(K_0(p^e) \begin{pmatrix} 0 & 1 \\ p^e & 0 \end{pmatrix} K_0(p^e) \right) = \operatorname{char}\left(K_0(p^e) \begin{pmatrix} 0 & 1 \\ p^e & 0 \end{pmatrix} \right) \in \mathcal{H}_{K_0(p^e)}. \tag{5.35}$$

These w_{p^e} are called *Atkin–Lehner involutions*.

Proposition 5.9. *For* all p, *we have* $T_p^* = w_N T_p w_N$ *on* $S_k(\Gamma_0(N))$.

Proof. In general, we have that

$$[\Gamma \alpha \Gamma]_k^* = [\Gamma \alpha' \Gamma]_k \tag{5.36}$$

for $\alpha' = \det(\alpha)\alpha^{-1}$. So

$$T_p^* = \left[\Gamma_0(N) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma_0(N) \right]_k \tag{5.37}$$

and the claim follows from computing that

$$\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}^{-1} \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}. \tag{5.38}$$

□

5.3 Newforms

References: [DS05, Sections 5.6-5.10].

We consider the following "degeneracy maps" $\iota_p^1, \iota_p^2 \colon S_k(\Gamma_0(N)) \to S_k(\Gamma_0(Np))$: ι_p^1 is induced from the inclusion $\Gamma_0(Np) \subseteq \Gamma_0(N)$ and ι_p^2 is induced from $\Gamma_0(Np) \simeq \Gamma_0(N) \cap \Gamma^0(p) \subseteq \Gamma_0(N)$. We write $\iota_p \colon S_k(\Gamma_0(N))^{\oplus 2} \to S_k(\Gamma_0(Np))$ to be $\iota_p(\alpha, \beta) = \iota_p^1(\alpha) + \iota_p^2(\beta)$.

Definition 5.7. The space of *old forms* is $S_k^{\text{old}}(\Gamma_0(N)) = \sum_{p|N} \iota_p(S_k(\Gamma_0(N/p))^{\oplus 2})$. The space of *new forms* is the orthogonal complement $S_k^{\text{new}}(\Gamma_0(N)) = (S_k^{\text{old}}(\Gamma_0(N)))^{\perp}$ under the Petersson inner product. A *newform* is an element of $S_k^{\text{new}}(\Gamma_0(N))$ which is an eigenfunction for the Hecke operators T_n with $(n, N) = 1$.

Note that since $\iota_p^2 = w_{Np}\iota_p^1 w_N = p^{1-k}\left[\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\right]_k$, we have $(\iota_p^2 f)(z) = f(pz)$.

Proposition 5.10. *The subspaces $S_k^{old}(\Gamma_0(N))$ and $S_k^{new}(\Gamma_0(N))$ are stable under the action of the Hecke operators T_n and under the action of the Fricke involution w_N.*

Proof. Since we have $\iota_p^1 = w_{Np}\iota_p^2 w_N$, it follows that $S_k^{\text{old}}(\Gamma_0(N))$ is stable under w_N. Since $w_N^* = w_N^{-1} = w_N$, so is $S_k^{\text{new}}(\Gamma_0(N))$. As $T_n^* = w_N T_n w_N$, it remains to check that T_n preserves $S_k^{\text{old}}(\Gamma_0(N))$: this implies that T_n^* preserves $S_k^{\text{old}}(\Gamma_0(N))$, and thus that T_n preserves $S_k^{\text{new}}(\Gamma_0(N))$.

Now consider $\ell \mid N$ and T_p acting on the image of ι_ℓ. If $\ell \neq p$, then T_p both on $S_k(\Gamma_0(N))$ and on $S_k(\Gamma_0(N/\ell))$ are convolution by the same function, so T_p commutes with $\iota_\ell^1, \iota_\ell^2$. For $\ell = p$, we have the commutative diagram

$$\begin{array}{ccc} S_k(\Gamma_0(N/p))^{\oplus 2} & \xrightarrow{\begin{pmatrix} T_p & 1 \\ -p^{k-1}1_{N/p}(p) & 0 \end{pmatrix}} & S_k(\Gamma_0(N/p))^{\oplus 2} \\ \downarrow{\scriptstyle \iota_p} & & \downarrow{\scriptstyle \iota_p} \\ S_k(\Gamma_0(N)) & \xrightarrow{T_p} & S_k(\Gamma_0(N)) \end{array} \tag{5.39}$$

We can check this on q-expansions, since $(\iota_p^2 f)(z) = f(pz)$: if $(\alpha, \beta) \in S_k(\Gamma_0(N/p))^{\oplus 2}$, then the image under the bottom is

$$(\alpha, \beta) \mapsto \alpha(z) + \beta(pz) \mapsto \sum_{n \geqslant 0} \left(a_{np}(\alpha) + a_n(\beta)\right) q^n \tag{5.40}$$

and under the top is

$$(\alpha, \beta) \mapsto \left(T_p\alpha + \beta, -p^{k-1}1_{N/p}(p)\alpha\right) \mapsto \sum_{n \geqslant 0} \left(a_{np}(\alpha) + a_n(\beta)\right) q^n \tag{5.41}$$

□

Since $(\iota_p^2 f)(z) = f(pz)$, in terms of q-expansions we have

$$(\iota_p^2 f)(z) = \sum_{n \geqslant 0} a_n(f) q^{pn}. \tag{5.42}$$

In particular, $g \in \sum_{p|N} \iota_2^p S_k(\Gamma_0(N/p))$ satisfies that $a_n(g) = 0$ whenever $(n, N) = 1$.

Lemma 5.11 (Main Lemma for $\Gamma_0(N)$, Atkin–Lehner)**.** *$f \in S_k(\Gamma_0(N))$ satisfies $a_n(f) = 0$ for all n with $(n, N) = 1$ if and only if $f \in \sum_{p|N} \iota_p^2 S_k(\Gamma_0(N/p))$.*

Proof. For $D \mid N$, a set of coset representatives for $\Gamma_0(D)\backslash\Gamma_0(N)$ is

$$\begin{pmatrix} 1 & 0 \\ aD & 1 \end{pmatrix} \quad \text{for } 0 \leqslant a < N/D. \tag{5.43}$$

So we may consider the projection

$$\mathrm{proj}_{S_k(\Gamma_0(D))} = \frac{1}{N/D} \sum_{0 \leqslant a < N/D} \left[\begin{pmatrix} 1 & 0 \\ aD & 1 \end{pmatrix} \right]_k \tag{5.44}$$

onto $\iota_p^1 S_k(\Gamma_0(D))$. Since $\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ aD & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}^{-1} = \begin{pmatrix} N & -aD \\ 0 & N \end{pmatrix}$, we have

$$\mathrm{proj}_{w_N S_k(\Gamma_0(D))} = \frac{1}{N/D} \sum_{0 \leqslant a < N/D} \left[\begin{pmatrix} 1 & a/(N/D) \\ 0 & 1 \end{pmatrix} \right]_k. \tag{5.45}$$

We denote this by $\pi_{N/D}$. Note that in terms of q-expansions we have

$$(\pi_d f)(z) = \sum_{n \geqslant 0} a_n(f) \frac{\sum_{a=0}^{d-1} e^{2\pi i n(z+a/d)}}{d} = \sum_{d|n} a_n(f) q^n. \tag{5.46}$$

Since $\iota_p^2 S_k(\Gamma_0(N/p)) = w_N \iota_p^1 S_k(\Gamma_0(N/p))$, we have

$$\sum_{p|N} \iota_p^2 S_k(\Gamma_0(N/p)) = \sum_{p|N} \mathrm{im}(\pi_p) = \sum_{p|N} \ker(1 - \pi_p) = \ker\left(\prod_{p|N} (1 - \pi_p) \right). \tag{5.47}$$

Here the second equality is since $\pi_p^2 = \pi_p$, and the third equality is since the π_p commute (as can be seen from their action on the q-expansions).

In terms of q-expansions, we have

$$\left(\prod_{p|N} (1 - \pi_p) \right) f(z) = \sum_{(n,N)=1} a_n(f) q^n \tag{5.48}$$

and the claim follows. □

Corollary 5.12. *If f is a newform, then $a_1(f) \neq 0$ and f is an eigenfunction for* all *Hecke operators T_n. If $a_1(f) = 1$, we call it a* normalized newform*, and it satisfies that $T_n f = a_n(f) f$ for all n. There is also a sign $\pm$ such that $w_N f = \pm f$.*

Proof. Suppose $f \in S_k(\Gamma_0(N))$ is a Hecke eigenfunction for all T_n with $(n, N) = 1$, say $T_n f = \lambda_n f$. Assume that $a_1(f) = 0$. Then we have $a_n(f) = a_1(T_n f) = a_1(\lambda_n f) = \lambda_n a_1(f) = 0$, for $(n, N) = 1$. By the Main Lemma, this implies that f is an old form. The contrapositive of this implies that if f is a newform, then $a_1(f) \neq 0$.

Now let $f \in S_k(\Gamma_0(N))$ be a normalized newform. Consider $g = T_n f - a_n(f)f$. Note that this is an eigenfunction for all T_m with $(m, N) = 1$. As $a_1(g) = 0$, g is an old form by the above. Since T_n preserves $S_k^{\text{new}}(\Gamma_0(N))$, g is also a newform, and thus $g = 0$. This implies that $T_n f = a_n(f)f$ for all n.

The last claim follows similarly: consider $g = w_N f - a_1(w_N f)f$. The same argument as above shows that $g = 0$ (here we are using that T_n and w_N commute if $(n, N) = 1$). Thus $w_N f = Cf$ for some $C \in \mathbb{C}$, and since $w_N^2 = 1$, we have $C^2 = 1$. □

Corollary 5.13. *If $f \in S_k(\Gamma_0(N))$ is a newform, we have*

$$\sum_{n>0} \frac{a_n(f)}{n^s} = \prod_p (1 - a_p(f)p^{-s} + 1_N(p)p^{k-1-2s})^{-1}. \tag{5.49}$$

Let $\pm$ be the sign such that $w_N f = \pm f$. Then

$$L_f(s) = (2\pi)^{-s}\Gamma(s) \prod_p (1 - a_p(f)p^{-s} + 1_N(p)p^{k-1-2s})^{-1} \tag{5.50}$$

is entire and satisfies the functional equation

$$L_f(k - s) = \pm N^{s-\frac{k}{2}} L_f(s). \tag{5.51}$$

Proof. As we have seen before, we have

$$L_f(s) = \int_0^\infty f(it)t^s \frac{\mathrm{d}t}{t} = (2\pi)^{-s}\Gamma(s) \sum_{n>0} \frac{a_n(f)}{n^s} \tag{5.52}$$

for all $\mathrm{Re}(s) > 1 + \frac{k}{2}$. Note that

$$\pm f(it) = (w_N f)(it) = i^k N^{k/2} (Nit)^{-k} f(-1/(Nit)) = N^{-k/2} t^{-k} f(i/(Nt)) \tag{5.53}$$

and so we may write, taking $u = 1/(Nt)$,

$$L_f(s) = \int_0^{1/\sqrt{N}} f(it)t^s \frac{\mathrm{d}t}{t} \pm \int_0^{1/\sqrt{N}} f(i/(Nu))N^{-s}u^{-s} \frac{\mathrm{d}u}{u}. \tag{5.54}$$

Thus

$$N^{s/2} L_f(s) = \int_0^{1/\sqrt{N}} (t^s N^{s/2} \pm t^{k-s} N^{k/2-s/2}) f(it) \frac{\mathrm{d}t}{t}. \tag{5.55}$$

Since $f(it)$ decays exponentially as $t \to 0$, this shows that $L_f(s)$ is entire. Moreover, the functional equation is now clear. □

Corollary 5.14. *We have a decomposition* $S_k(\Gamma_0(N)) = \bigoplus_{M|N} \bigoplus_{d|N/M} \iota_d^2 S_k^{new}(\Gamma_0(M))$.

5.4 Representation theoretically

If K is an open compact and Γ is the corresponding arithmetic subgroup, recall that

$$S_k(\Gamma) = \bigoplus_{\pi = \pi_f \otimes D_k} \pi_f^K \otimes V_\pi \tag{5.56}$$

where the sum is over irreducible cuspidal automorphic representations π, and V_π are multiplicity spaces.

Theorem 5.15. *Let* $\pi = \pi_f \otimes D_k$ *be a cuspidal irreducible automorphic representation with trivial central character. Then there is* N *with* $\pi_f^{K_0(N)} \neq 0$. *If* N *is chosen to be as smallest as possible, then* $\dim \pi_f^{K_0(N)} = 1$, *and also* $\dim V_\pi = 1$. *In particular, there is a bijection between*

$$\{\pi \subseteq \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}) \text{ with } \pi_\infty \simeq D_k \text{ and trivial central character}\} \tag{5.57}$$

and

$$\{\text{normalized newforms for some } \Gamma_0(N)\}. \tag{5.58}$$

Proof. We saw before that such N exists, and take it to be the smallest such N. Then $\pi_f^{K_0(N)} \otimes V_\pi \subseteq S_k^{\text{new}}(\Gamma_0(N))$. Since π is irreducible, $\pi_f^{K_0(N)}$ is an irreducible $\mathcal{H}_{K_0(N)}$-module. But it contains a newform, which is a subspace of dimension 1,[8] and thus $\dim \pi_f^{K_0(N)} = 1$. Furthermore, any element of $\pi_f^{K_0(N)} \otimes V_\pi$ has the same Hecke eigenvalues, and thus the same Fourier expansion up to a scalar. Hence $\dim(\pi_f^{K_0(N)} \otimes V_\pi) = 1$. □

Corollary 5.16. *Let* $\pi_1, \pi_2 \subseteq \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}))$ *be two irreducible automorphic representations with* $\pi_{1,\infty} \simeq \pi_{2,\infty} \simeq D_k$ *and trivial central character. Suppose* $\pi_1 \simeq \pi_2$ *as* $\mathrm{GL}_2(\mathbb{A}_f)$*-representations (that is, that* $\pi_{1,p} \simeq \pi_{2,p}$ *as* $\mathrm{GL}_2(\mathbb{Q}_p)$*-representations for all primes* p*). Then* $\pi_1 = \pi_2$ *as subspaces of* $\mathcal{A}_0(\mathrm{GL}_2(\mathbb{A}))$.

Proof. This follows at once from $V_{\pi_i} = \mathbb{C}$. □

More generally, we have the following harder theorems.

[8] Here we are using something slightly stronger than what we proved before: we are using that a newform is an eigenfunction for the entire algebra $\mathcal{H}_{K_0(N)}$, not just of the T_p. This follows from the same argument as before, once we verify that $\mathcal{H}_{K_0(N)}$ preserves $S_k^{\text{old}}(\Gamma_0(N))$.

Theorem 5.17 (Strong multiplicity one, Piatetski-Shapiro, Shalika)**.** *Let $\pi_1, \pi_2 \subseteq \mathcal{A}_0(\mathrm{GL}_n(\mathbb{A}_F))$ be cuspidal irreducible automorphic representations for a number field F, and let S be a finite set of places. If $\pi_{1,v} \simeq \pi_{2,v}$ for all $v \notin S$, then $\pi_1 = \pi_2$ as subspaces of $\mathcal{A}_0(\mathrm{GL}_n(\mathbb{A}_F))$.*

Theorem 5.18 (Newform theory for GL_n, Jacquet–Piatetski-Shapiro–Shalika)**.** *Let $\mathfrak{p}$ be a non-archimedean place of a number field F and $\pi_\mathfrak{p}$ an irreducible smooth representation of $\mathrm{GL}_n(F_\mathfrak{p})$. Consider the subgroups*

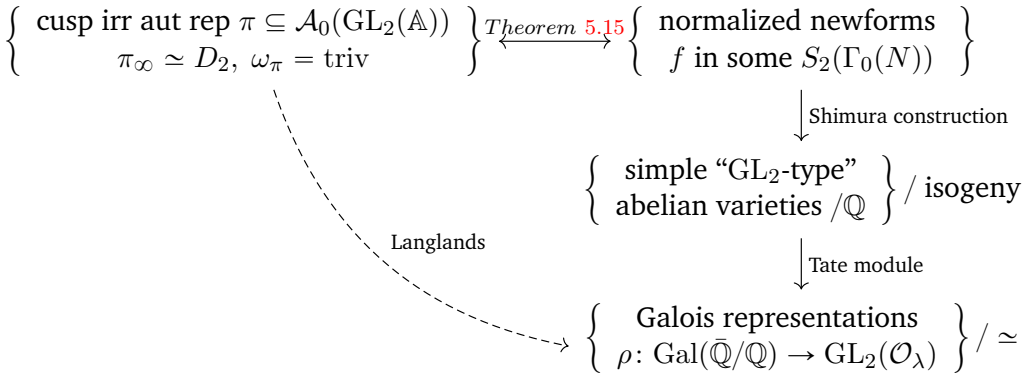
$$K(m) = \left\{ A \equiv \begin{pmatrix} * & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots \\ * & \cdots & * & * \\ 0 & \cdots & 0 & 1 \end{pmatrix} \mod \mathfrak{p}^m \right\} \subseteq \mathrm{GL}_n(F_\mathfrak{p}). \tag{5.59}$$

Then there exists $m \geqslant 0$ such that $\pi_\mathfrak{p}^{K(m)} \neq 0$, and we denote m_π the smallest such m. For such m_π, we have that $\pi_\mathfrak{p}^{K(m_\pi)}$ is one-dimensional.

Neither theorem is true for general G. Multiplicity one fails already for SL_n for $n > 2$. Newform theories have been proposed beyond GL_n only for some other classical groups.

6 Modular elliptic curves and the Eichler–Shimura relation

Our goal for this chapter is, for a normalized newform f of weight 2 and level $\Gamma_0(N)$, to construct an abelian variety $A_f/\mathbb{Q}$. This realizes some cases of global Langlands: if f correspond to an automorphic representation π, the Tate modules of A_f (to be defined later) are the Galois representations attached to π that are predicted by global Langlands.



(6.1)

In every level of this diagram, there is a compatible way to attach an L-function. We will mostly focus on the case that A_f is an elliptic curve, in which case we will prove the compatibility of such L-functions in the right column.

6.1 Shimura construction

References: [DS05, Sections 6.6, 7.5-7.9].

Since $X_0(N)$ are compact Riemann surfaces, they have the structure of an algebraic variety over $\mathbb{C}$. We first need to see that in fact they have models over $\mathbb{Q}$.

Proposition 6.1. *We have $\mathbb{C}(X_0(N)) = \mathbb{C}(j(z), j(Nz))$. The minimal polynomial $F_N(j, Y) \in \mathbb{C}(j)[Y]$ of $j(Nz)$ is a polynomial in j, and has coefficients in $\mathbb{Q}$, that is, $F_N(X, Y) \in \mathbb{Q}[X, Y]$. In particular, $X_0(N)$ has a model over $\mathbb{Q}$ with function field $\mathbb{Q}(j(z), j(Nz))$, which we denote $X_0(N)_\mathbb{Q}$*

Proof. Since $X_0(1)$ is the moduli space of elliptic curves, we know already that this has function field $\mathbb{C}(j)$.

First note that since $j(z) = j(\gamma z)$ for $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, we have that $F_N(j(z), j(N\gamma z)) = 0$ for all γ. Ranging through all $\gamma \in \Gamma_0(N)\backslash\Gamma_0(1)$, this gives $[\Gamma_0(1)\colon \Gamma_0(N)]$ roots of $F_N(j, Y)$. We claim they are all distinct. If $j(N\gamma_1 z) = j(N\gamma_2 z)$, then since $j\colon \Gamma_0(1)\backslash\mathcal{H} \simeq \mathbb{C}$, we must have that there exists $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(1)$ with

$$N\gamma_1(z) = \gamma(N\gamma_2(z)). \tag{6.2}$$

This means that

$$\begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} \gamma_1 = \gamma \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} \gamma_2 \tag{6.3}$$

and thus $\gamma_1\gamma_2^{-1} \in \Gamma_0(1) \cap \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}^{-1} \Gamma_0(1) \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} = \Gamma_0(N)$.

In particular, $\mathbb{C}(j(z), j(Nz)) \subseteq \mathbb{C}(X_0(N))$ have the same degree over $\mathbb{C}(j(z))$, and hence this is an equality. Hence

$$F_N(j, Y) = \prod_{\gamma \in \Gamma_0(N)\backslash\Gamma_0(1)} (Y - j(N\gamma z)). \tag{6.4}$$

Since the symmetric polynomials in the $j(N\gamma_i z)$ are holomorphic in $\mathcal{H}$, they are polynomials in $j(z)$, thus $F_N(X, Y) \in \mathbb{C}[X, Y]$. Say $F_N(X, Y) = \sum c_{n,m} X^n Y^m$. Recalling that $j(z) \in \mathbb{Z}[\![q]\!][q^{-1}]$, we can look at the q-expansion of $F(j(z), j(Nz))$ to see that $c_{n,m}$ are determined by a linear system with coefficients in $\mathbb{Z}$. Thus $F_N(X, Y) \in \mathbb{Q}[X, Y]$. □

Remark 6.1. Note that the Fricke involution $w_N \colon X_0(N) \to X_0(N)$, in terms of function fields, corresponds to

$$j(z) \mapsto j(-1/(Nz)) = j(Nz), \quad j(Nz) \mapsto j(-N/(Nz)) = j(z), \tag{6.5}$$

and in particular also descends to $X_0(N)_{\mathbb{Q}}$.

Remark 6.2. The natural maps $X_0(N) \to X_0(M)$ for $M \mid N$ also descend to maps $X_0(N)_{\mathbb{Q}} \to X_0(M)_{\mathbb{Q}}$. Given the above, it's not hard to see that the Galois closure of $\mathbb{C}(j(z), j(Nz))$ is simply the function field of $X(N)$: $\Gamma_0(1)$ acts transitively on the roots of $F_N(j, Y)$, and the stabilizer is $\bigcap_{\gamma \in \Gamma_0(1)} \gamma^{-1}\Gamma_0(N)\gamma = \pm I \cdot \Gamma(N)$. This gives a model $X(N)_{\mathbb{Q}}$ over $\mathbb{Q}$ such that $X(N)_{\mathbb{Q}} \to X(1)_{\mathbb{Q}}$ is Galois, and thus by Galois theory, the map $X_0(N) \to X_0(M)$ also descends to $X_0(N)_{\mathbb{Q}} \to X_0(M)_{\mathbb{Q}}$.

In particular, the above remarks imply that the Hecke correspondences T_p also descend to the models over $\mathbb{Q}$.

$$\begin{array}{ccc} & X_0(Np)_{\mathbb{Q}} & \\ {\scriptstyle p_1}\swarrow & & \searrow{\scriptstyle p_2} \\ X_0(N)_{\mathbb{Q}} & & X_0(N)_{\mathbb{Q}} \end{array} \tag{6.6}$$

Corollary 6.2. *Denote $M_k(\Gamma_0(N), \mathbb{Z}) \subseteq M_k(\Gamma_0(N))$ the subset of modular forms with Fourier expansion in $\mathbb{Z}[\![q]\!]$. Then $M_k(\Gamma_0(N)) = M_k(\Gamma_0(N), \mathbb{Z}) \otimes \mathbb{C}$. Denote $\mathbb{T}_{\mathbb{Z}}$ the $\mathbb{Z}$-subalgebra of $\mathrm{End}(M_k(\Gamma_0(N)))$ generated by the Hecke operators T_n. Then $\mathbb{T}_{\mathbb{Z}}$ is a free $\mathbb{Z}$-module of rank $\dim M_k(\Gamma_0(N))$.*

Proof. Assume k is even for simplicity. Then since $X_0(N)$ has a model over $\mathbb{Q}$, the space $M_k(\Gamma_0(N)) = H^0(X_0(N), \Omega^{\otimes k/2}(\cdots))$ also has a model over $\mathbb{Q}$. That

is, there is a submodule $M_k(\Gamma_0(N))_{\mathbb{Q}}$ with $M_k(\Gamma_0(N)) = M_k(\Gamma_0(N))_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{C}$, and is such that the Hecke operators act on $M_k(\Gamma_0(N))_{\mathbb{Q}}$.

For $R \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{C}\}$, we denote $\mathbb{T}_R$ the R-subalgebra of $\mathrm{End}(M_k(\Gamma_0(N)))$ generated by the image of the T_n. The existence of $M_k(\Gamma_0(N))_{\mathbb{Q}}$ above implies that $\mathbb{T}_{\mathbb{C}} = \mathbb{T}_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{C}$. We also have $\mathbb{T}_{\mathbb{Q}} = \mathbb{T}_{\mathbb{Z}} \otimes \mathbb{Q}$, and thus $\mathbb{T}_{\mathbb{C}} = \mathbb{T}_{\mathbb{Z}} \otimes \mathbb{C}$.

For $R \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{C}\}$, consider the pairing $\langle \cdot, \cdot \rangle \colon \mathbb{T}_R \times M_k(\Gamma_0(N), R) \to R$ given by $\langle T, f \rangle = a_1(Tf)$. Then this is a left and right nondegenerate: 1) if f is such that $\langle T, f \rangle = 0$ for all T, then $0 = a_1(T_n f) = a_n(f)$ for all n, so $f = 0$, 2) similarly, if T is such that $\langle T, f \rangle = 0$ for all f, then $0 = a_1(TT_n f) = a_1(T_n T f) = a_n(Tf)$ for all n, f, and so T acts trivially on $M_k(\Gamma_0(N), R)$.

This implies that $\dim_{\mathbb{C}} \mathbb{T}_{\mathbb{C}} = \dim_{\mathbb{C}} M_k(\Gamma_0(N))$. Take B such that elements of $M_k(\Gamma_0(N))$ are determined by its first B Fourier coefficients. Then $M_k(\Gamma_0(N), \mathbb{Z}) \subseteq \mathbb{Z}^B$ is a finite free $\mathbb{Z}$-module, and the above implies that $\mathbb{T}_{\mathbb{Z}}$ is also finite free, with $\mathrm{rank}_{\mathbb{Z}} \mathbb{T}_{\mathbb{Z}} = \mathrm{rank}_{\mathbb{Z}} M_k(\Gamma_0(N), \mathbb{Z})$. As, we have $\mathrm{rank}_{\mathbb{Z}} M_k(\Gamma_0(N), \mathbb{Z}) = \dim_{\mathbb{C}} M_k(\Gamma_0(N))$, and it remains to see that $M_k(\Gamma_0(N), \mathbb{Z}) \otimes \mathbb{C} \to M_k(\Gamma_0(N))$ is injective. This is because $f_1, \ldots, f_r \in M_k(\Gamma_0(N), \mathbb{Z})$ are linearly independent if and only if the matrix $(a_i(f_j))_{\substack{1 \leqslant i \leqslant B \\ 1 \leqslant j \leqslant r}} \in \mathrm{Mat}_{B \times r}(\mathbb{Z})$ has rank r, but this is the same as it having rank r as a matrix over $\mathbb{C}$. □

Remark 6.3. In fact, one can prove that $M_k(\Gamma_0(N))_{\mathbb{Q}} = M_k(\Gamma_0(N), \mathbb{Q})$, for instance by using Katz's definition of modular forms

Corollary 6.3. *If $f \in M_k(\Gamma_0(N))$ and $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q})$, then $f^\sigma(z) := \sum_{n \geqslant 0} \sigma(a_n(f)) q^n$ also belongs to $M_k(\Gamma_0(N))$.*

If f is a normalized newform, then $\mathbb{Q}(f) := \mathbb{Q}(a_n(f) \colon n \in \mathbb{Z}_{\geqslant 1})$ is a totally real number field, and $a_n(f)$ are algebraic integers.

Proof. We have that $\langle f, T_n f \rangle = \overline{a_n(f)} \langle f, f \rangle$. But we also have $\langle f, T_n f \rangle = \langle T_n^* f, f \rangle$. Remembering that $T_n^* = w_N T_n w_N$, we have $T_n^* f = a_n(f) f$ as well. Thus $\langle T_n^* f, f \rangle = a_n(f) \langle f, f \rangle$ and it follows that $a_n(f)$ is real. Since $\mathbb{T}_{\mathbb{Z}}$ is a free $\mathbb{Z}$-module, we also have that the eigenvalues of T_n are algebraic integers.

By the above corollary, f^σ is also a normalized newform with real Fourier coefficients for all $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q})$. Its stabilizer is a finite index subgroup of $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$, and thus $\mathbb{Q}(f)$ is a totally real number field. □

Definition 6.1. We denote $J_0(N)_{\mathbb{Q}}$ to be the Jacobian of $X_0(N)_{\mathbb{Q}}$.

Over $\mathbb{C}$, this is the complex torus given by

$$J_0(N)(\mathbb{C}) = \frac{H^0(X_0(N), \Omega)^*}{H_1(X_0(N), \mathbb{Z})} = \frac{S_2(\Gamma_0(N))^*}{H_1(X_0(N), \mathbb{Z})}. \tag{6.7}$$

As noted before, the Hecke operators T_p descend to correspondences over $X_0(N)_{\mathbb{Q}}$:

$$\begin{array}{ccccc} & & X_0(Np)_{\mathbb{Q}} & & \\ & \swarrow^{p_1} & & \searrow^{p_2} & \\ X_0(N)_{\mathbb{Q}} & & & & X_0(N)_{\mathbb{Q}} \end{array} \tag{6.8}$$

and so we may consider

$$T_p \colon \mathrm{Div}^0(X_0(N)) \xrightarrow{p_1^*} \mathrm{Div}^0(X_0(Np)) \xrightarrow{p_{2,*}} \mathrm{Div}^0(X_0(N)) \tag{6.9}$$

which induce an endomorphism $T_p \in \mathrm{End}_{\mathbb{Q}}(J_0(N)_{\mathbb{Q}})$. That is, we have an action $\mathbb{T}_{\mathbb{Z}} \to \mathrm{End}_{\mathbb{Q}}(J_0(N)_{\mathbb{Q}})$.

Theorem 6.4. *For $f \in S_2(\Gamma_0(M))$ a normalized newform, let $I_f \subseteq \mathbb{T}_{\mathbb{Z}}$ be the ideal attached to f, that is, $I_f = \ker(\lambda_f \colon \mathbb{T}_{\mathbb{Z}} \to \mathbb{C})$ where $\lambda_f(T_n) = a_n(f)$. We consider*

$$A_f := J_0(M)_{\mathbb{Q}}/I_f \cdot J_0(M)_{\mathbb{Q}}. \tag{6.10}$$

Note that I_f, and hence A_f, only depend on the Galois orbit of f. Then A_f is a simple abelian variety of dimension $\dim_{\mathbb{Q}} \mathbb{Q}(f)$ with $\mathrm{End}_{\mathbb{Q}}(A_f) \otimes \mathbb{Q} = \mathbb{Q}(f)$.

Moreover, for every N we have an isogeny decomposition[9]

$$J_0(N)_{\mathbb{Q}} \to \prod_f A_f^{m_f} \tag{6.11}$$

where the product runs through Galois orbits of newforms in $S_2(\Gamma_0(M))$ for $M \mid N$, and $m_f = \sigma_0(N/M) = \#\{\text{divisors of of } N/M\}$.

Proof. Let $V_f \subseteq S_2(\Gamma_0(M))$ be the subspace spanned by f^σ for all $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q})$, and let $\Lambda_f := H_1(X_0(M), \mathbb{Z})|_{V_f} \subseteq V_f^*$ be the image of $H_1(X_0(M), \mathbb{Z})$ under $S_2(\Gamma_0(M))^* \twoheadrightarrow V_f^*$.

We first note that $V_f = S_2(\Gamma_0(M))[I_f]$. We clearly have $V_f \subseteq S_2(\Gamma_0(M))[I_f]$, and so it suffices to see that both have dimension $\mathbb{Q}(f)$. This is clear for V_f, and we have

$$(S_2(\Gamma_0(M))[I_f])^* = S_2(\Gamma_0(M))^*/I_f = \mathbb{T}_{\mathbb{C}}/I_f = (\mathbb{T}_{\mathbb{Z}}/I_f) \otimes \mathbb{C} \tag{6.12}$$

and so $\dim_{\mathbb{C}} S_2(\Gamma_0(M))[I_f] = \mathrm{rank}_{\mathbb{Z}}(\mathbb{T}_{\mathbb{Z}}/I_f) = \mathrm{rank}_{\mathbb{Z}}(\mathbb{Z}[a_n(f),\ n \in \mathbb{Z}]) = \dim_{\mathbb{Q}} \mathbb{Q}(f)$.

[9] If we consider the *isogeny category* AbVar^0 of abelian varieties with $\mathrm{Hom}_{\mathrm{AbVar}^0}(A, B) = \mathrm{Hom}_{\mathbb{Q}}(A, B) \otimes \mathbb{Q}$, then this is a semisimple category. Note A_f are simple objects in this category since $\mathrm{End}_{\mathbb{Q}}(A_f) \otimes \mathbb{Q}$ is a field.

It follows that

$$A_f(\mathbb{C}) = S_2(\Gamma_0(M))^*/(I_f + H_1(X_0(M),\mathbb{Z})) = V_f^*/\Lambda_f. \tag{6.13}$$

We will skip the discussion of why A_f is an abelian variety, but let's at least see that $\Lambda_f \subseteq V_f^*$ is a lattice (in the case $\mathbb{Q}(f) = \mathbb{Q}$, this is enough to conclude $A_{f,\mathbb{C}}$ is an elliptic curve over $\mathbb{C}$). We clearly have by construction that $\Lambda_f \otimes \mathbb{R} = V_f^*$, and we also have

$$\dim_{\mathbb{R}}(V_f^*) = \dim_{\mathbb{R}}((H_1(X_0(M),\mathbb{Z})\otimes\mathbb{R})/I_f) = \operatorname{rank}_{\mathbb{Z}}(H_1(X_0(M),\mathbb{Z})/I_f) \geqslant \operatorname{rank}_{\mathbb{Z}}(\Lambda_f) \tag{6.14}$$

as we have a surjection $H_1(X_0(M),\mathbb{Z})/I_f \twoheadrightarrow \Lambda_f$. In particular, we have that $\dim A_f = \dim_{\mathbb{Q}} \mathbb{Q}(f)$.

By construction, A_f has an action of $\mathbb{T}_{\mathbb{Z}}/I_f \simeq \mathbb{Z}[a_n(f),\ n \in \mathbb{Z}]$, that is

$$\mathbb{Q}(f) \subseteq \operatorname{End}_{\mathbb{Q}}(A_f) \otimes \mathbb{Q}. \tag{6.15}$$

One can in fact prove that this is an equality[10], which implies that A_f is a simple abelian variety.

Now we define the map $J_0(N)_{\mathbb{Q}} \to \prod_f \prod_n A_f$ where f varies through Galois orbits of normalized newforms for some $S_2(\Gamma_0(M))$ with $M \mid N$ and n then varies through $n \mid N/M$, by

$$J_0(N)_{\mathbb{Q}} \xrightarrow{\iota_n^2} J_0(N/n)_{\mathbb{Q}} \xrightarrow{\iota^1_{N/Mn}} J_0(M)_{\mathbb{Q}} \twoheadrightarrow A_f. \tag{6.16}$$

We can check this is an isogeny by working over $\mathbb{C}$, and then this follows from Corollary 5.14, as we have the decomposition

$$S_2(\Gamma_0(N)) = \bigoplus_f \bigoplus_n \iota_n^2 V_f. \tag{6.17}$$

□

Definition 6.2. And elliptic curve $E/\mathbb{Q}$ is *modular* if there exists a normalized newform f such that E is isogenous to $E_f := A_f$.

In other words, modular elliptic curves are isogeny quotients of $J_0(N)_{\mathbb{Q}}$. Equivalently, an elliptic curve $E/\mathbb{Q}$ is modular iff there is a nontrivial algebraic map $X_0(N)_{\mathbb{Q}} \to E$ for some N. This is because $X_0(N)_{\mathbb{Q}} \to J_0(N)_{\mathbb{Q}}$ is initial among maps from $X_0(N)_{\mathbb{Q}}$ to abelian varieties.

[10]The easiest proof uses Tate modules: $\operatorname{End}_{\mathbb{Q}}(A_f)\otimes\mathbb{Q}_\ell \hookrightarrow \operatorname{End}_{\operatorname{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})}(V_\ell A_f)$, and one can prove that this right hand side is simply $\mathbb{Q}(f)\otimes\mathbb{Q}_\ell$.

6.2 L-functions of elliptic curves

Given a number field F and an elliptic curve E/F, we will attach an L-function $L(E/F, s)$ defined via an Euler product.

Elliptic curves over finite fields

References: [Sil09, Chapter V].

We let $k = \mathbb{F}_q$ be a finite field.

Definition 6.3. For E/k an elliptic curve, We denote $a(E/k) := q + 1 - \#E(k)$.

Proposition 6.5. *Let $\phi\colon E \to E$ be the q-Frobenius. Then $[a(E/k)] = \phi + \hat{\phi}$.*

Proof. First we note that $\phi - 1$ is a separable map. This is because if ω is an invariant differential, we have $\phi^*\omega = 0$, and thus $(\phi - 1)^*\omega = -\omega \neq 0$ This implies that $\#E(k) = \deg(\phi - 1)$. Now we have

$$[\#E(k)] = (\phi - 1)\widehat{(\phi - 1)} = \phi\hat{\phi} - \phi - \hat{\phi} + [1] = [q + 1] - (\phi + \hat{\phi}) \tag{6.18}$$

and the claim follows. □

Corollary 6.6 (Hasse's bound)**.** *Let E/k be an elliptic curve. Then $|a(E/k)| \leqslant 2\sqrt{q}$.*

Proof. This is a form of Cauchy–Schwartz. We have $0 \leqslant \deg(\phi - \hat{\phi})$, and this is

$$[\deg(\phi - \hat{\phi})] = (\phi - \hat{\phi})(\hat{\phi} - \phi) = -(\phi + \hat{\phi})^2 + 4\phi\hat{\phi} = [-a(E)^2 + 4q] \tag{6.19}$$

and thus $a(E/k) \leqslant 2\sqrt{q}$. □

Definition 6.4. If $X/\mathbb{F}_q$ is a variety over a finite field, we define it's zeta function to be the following power series

$$\zeta_{X/\mathbb{F}_q}(T) = \exp\left(\sum_{m \geqslant 1} \frac{\#X(\mathbb{F}_{q^m})}{m} T^m\right). \tag{6.20}$$

Corollary 6.7. *For $E/\mathbb{F}_q$ an elliptic curve, it's zeta function is*

$$\zeta_{E/\mathbb{F}_q}(T) = \frac{(1 - a(E/\mathbb{F}_q)T + qT^2)}{(1 - T)(1 - qT)}. \tag{6.21}$$

Proof. Let ϕ be the q-Frobenius. We have $[q^m + 1 - \#E(\mathbb{F}_{q^m})] = \phi^m + \hat{\phi}^m$. Hence if $\alpha, \beta \in \mathbb{C}$ are such that $\alpha\beta = q$ and $\alpha + \beta = a(E/\mathbb{F}_q)$, then $\#E(\mathbb{F}_{q^m}) = q^m + 1 - \alpha^m - \beta^m$. Then the claim follows from the computation that

$$\sum_{m \geqslant 1} \frac{\gamma^m}{m} T^m = -\log(1 - \gamma T), \tag{6.22}$$

and that $(1 - \alpha T)(1 - \beta T) = 1 - a(E/\mathbb{F}_q)T + qT^2$. □

Remark 6.4. If we denote $P_1(T) = 1 - a(E)T + qT^2$, then Hasse's bound is equivalent to $P_1(T)$ having non-positive discriminant, and in particular to both of its roots having absolute value $\sqrt{q}^{-1}$. This means that $P_1(q^{-s})$ only has roots with $\mathrm{Re}(s) = 1/2$. So in fact we have verified the entirety of the Weil conjectures for $E/\mathbb{F}_q$ (see [Sil09, Section V.2] for more details).

Proposition 6.8. *Let $E, E'/k$ be two elliptic curves which are isogenous over k. Then $\#E(k) = \#E'(k)$.*

Proof. Let $f\colon E \to E'$ be an isogeny. If ϕ is the $\#k$-Frobenius, then we have $f \circ \phi_E = \phi_{E'} \circ f$, and thus $f \circ (\phi_E - 1) = (\phi_{E'} - 1) \circ f$. Taking degrees, we conclude $\#E(k) = \deg(\phi_E - 1) = \deg(\phi_{E'} - 1) = \#E'(k)$. □

Theorem 6.9 (Tate's isogeny theorem)**.** *Let $E, E'/k$ be two elliptic curves. Then E, E' are isogenous if and only $a(E) = a(E')$.*

Definition 6.5. E/k is *supersingular* if $\mathrm{char}(k) \mid a(E)$, it's called *ordinary* otherwise.

Proposition 6.10. *Let $E/\mathbb{F}_q$ be an elliptic curve, and denote $p = \mathrm{char}(k)$. Then $\widehat{\mathrm{Frob}_p}$ is separable if and only if E is ordinary. Furthermore, we have*

$$E(\bar{\mathbb{F}}_q)[p^m] \simeq \begin{cases} 0 & \text{if } E \text{ is supersingular,} \\ \mathbb{Z}/p^m\mathbb{Z} & \text{if } E \text{ is ordinary.} \end{cases} \tag{6.23}$$

Proof. For the first claim, it suffices to see that $\widehat{\mathrm{Frob}_q}$ is separable if and only if E is ordinary. We have

$$a(E/\mathbb{F}_q)\omega = [a(E/\mathbb{F}_q)]^*\omega = \mathrm{Frob}_q^*\omega + \widehat{\mathrm{Frob}_q}^*\omega = \widehat{\mathrm{Frob}_q}^*\omega \tag{6.24}$$

and thus $p \mid a(E/\mathbb{F}_q)$ if and only if $\widehat{\mathrm{Frob}_q}$ is inseparable.

We have

$$\#E(\bar{\mathbb{F}}_q)[p^m] = \#\ker([p^m]) = \deg_s([p^m]) = \deg_s(\widehat{\mathrm{Frob}_p})^m, \tag{6.25}$$

and now the second claim follows from the first. □

We also note the following:

Proposition 6.11. *If $E/\mathbb{F}_q$ is supersingular, and $p = \mathrm{char}(\mathbb{F}_q)$, then in fact E is defined over $\mathbb{F}_{p^2}$.*

Proof. Since E is supersingular, $\widehat{\mathrm{Frob}_p}\colon E^{(p)} \to E$ is inseparable, and thus is isomorphic to $\mathrm{Frob}_p\colon E^{(p)} \to E^{(p^2)}$. Thus $E \simeq E^{(p^2)}$, and E is defined over $\mathbb{F}_{p^2}$. □

Reduction of elliptic curves

References: [Sil09, Sections VII.1-2].

Let $F_\mathfrak{p}$ be a local field with maximal ideal $\mathfrak{m}$ and residue field k.

Definition 6.6. Let $E/F_\mathfrak{p}$ be an elliptic curve. We say that E has *good reduction* if there exists an elliptic curve $\tilde{E}/\mathcal{O}_{F_\mathfrak{p}}$ extending E. For such a choice of $\tilde{E}$, we say its base change $\tilde{E}/k$ is a *reduction of E modulo $\mathfrak{m}$*.

It turns out that there is a "canonical" way to choose $\tilde{E}/\mathcal{O}_F$, as we now explain. Given a Weierstraß model of E over $F_\mathfrak{p}$, we can always clear denominators to get a Weierstraß model with coefficients in $\mathcal{O}_{F_\mathfrak{p}}$.

Definition 6.7. Given $E/F_\mathfrak{p}$, a *minimal Weierstraß model* is a Weierstraß model of E with coefficients in $\mathcal{O}_{F_\mathfrak{p}}$ such that $\nu_\mathfrak{m}(\Delta)$ is as smallest as possible among all such choices.

Proposition 6.12. *Given $E/F_\mathfrak{p}$ and a minimal Weierstraß model $E\colon F(x,y) = 0$ with $F(x,y) \in \mathcal{O}_{F_\mathfrak{p}}[x,y]$, we denote $\tilde{E}/k$ its* reduction mod $\mathfrak{m}$ *to be the Weierstraß equation $\tilde{E}\colon \tilde{F}(x,y) = 0$ where $\tilde{F} \in k[x,y]$ is the reduction of F. Then $\tilde{E}$, up to isomorphism, does not depend on the choice of minimal Weierstraß model.*

Proof Sketch. We illustrate this in the case $\mathrm{char}(k) \neq 2, 3$: we have a Weierstraß equation $y^2 = x^3 + Ax + B$ and the valid changes of coordinate are $x \mapsto u^{-2}x, y \mapsto u^{-3}y$, which sends $A \mapsto u^4 A$ and $B \mapsto u^6 B$. So $\Delta \mapsto u^{12}\Delta$. Hence a change of coordinates between minimal Weierstraß equations is such that $u \in \mathcal{O}_{F_\mathfrak{p}}^\times$.

Then $\tilde{E}/k$ is obtained by reducing the coefficients modulo $\mathfrak{m}$, and the characterization of the possible changes of coordinate above proves that such reduction is uniquely determined. □

When $E/F_\mathfrak{p}$ has bad reduction, the minimal Weierstraß models reduce to a singular Weistraß equation over k. Its nonsingular points $\tilde{E}^{ns}(k)$ still form a group by the chord and tangent construction, and we can fully characterize what they look like.

Proposition 6.13. *If $\tilde{E}/k$ is defined by a singular Weierstraß equation, then it has a unique singular k-point, and we have one of three cases:*

1. *(Additive reduction) $\tilde{E}^{ns}(k) \simeq (k, +)$,*
2. *(Split multiplicative reduction) $\tilde{E}^{ns}(k) \simeq k^\times$,*
3. *(Nonsplit multiplicative reduction) there is a quadratic extension k'/k such that*
$$\tilde{E}^{ns}(k) \simeq \ker\left((k')^\times \xrightarrow{\mathrm{Nm}_{k'/k}} k^\times\right).$$

Note that in the case of a singular equation, we have

$$a(\tilde{E}/\mathbb{F}_q) := q + 1 - \#\tilde{E}(\mathbb{F}_q) = \begin{cases} 0 & \text{in additive reduction,} \\ 1 & \text{in split multiplicative reduction,} \\ -1 & \text{in nonsplit multiplicative reduction.} \end{cases} \tag{6.26}$$

Remark 6.5. A more abstract way of thinking about $\tilde{E}$ is using *Néron models*. If R is a Dedekind domain with fraction field K and A/K is a smooth separated scheme, a Néron model of A is a smooth separated model $\mathcal{A}/R$ which satisfies the *Néron mapping property*

if X/R is smooth separated, then any map $X_K \to A$ can be extended uniquely to $X \to \mathcal{A}$.

A theorem of Néron says that such models exist if A are abelian varieties. In this case, $\mathcal{A}$ are also smooth group schemes. Now if $A = E$ is an elliptic curve, one can show that $\tilde{E}^{ns}(k) = \mathcal{E}_k^0(k)$ where $\mathcal{E}_k^0$ is the connected component of the special fiber of $\mathcal{E}_k$. Then $\mathcal{E}_{\bar{k}}^0$ is isomorphic to $\mathbb{G}_{a,\bar{k}}$ resp. $\mathbb{G}_{m,\bar{k}}$ in additive resp. multiplicative reduction.

L-function of elliptic curves

Let F be a number field, and for a prime $\mathfrak{p}$ we denote $k_\mathfrak{p} = \mathbb{F}_{q_\mathfrak{p}}$ the residue field of $F_\mathfrak{p}$. Given E/F an elliptic curve, we denote

$$a_\mathfrak{p}(E/F) := a(\tilde{E}/k_\mathfrak{p}). \tag{6.27}$$

Definition 6.8. The L-function of E/F is

$$L(E/F, s) := \prod_{\mathfrak{p}\text{ good}} (1 - a_\mathfrak{p}(E/F)q_\mathfrak{p}^{-s} + q_\mathfrak{p}^{1-2s})^{-1} \cdot \prod_{\mathfrak{p}\text{ bad}} (1 - a_\mathfrak{p}(E/F)q_\mathfrak{p}^{-s})^{-1} \tag{6.28}$$

This converges absolutely for $\mathrm{Re}(s) > \frac{3}{2}$ by Hasse's bound.

Remark 6.6. By the discussion above, if we write $L(E/F, s) = \prod_{\mathfrak{p}} L(\tilde{E}/k_{\mathfrak{p}}, s)$, then we can uniformly write

$$L(\tilde{E}/\mathbb{F}_{q_{\mathfrak{p}}}, s) = \sum_{m \geqslant 0} \frac{a(\tilde{E}/\mathbb{F}_{q_{\mathfrak{p}}^m})}{q_{\mathfrak{p}}^{ms}}, \tag{6.29}$$

where we take the $m = 0$ term to mean 1.

Remark 6.7. For $F = \mathbb{Q}$, there is always a choice of Weierstraß model over $\mathbb{Z}$ which is simultaneously a minimal model over all $\mathbb{Q}_p$. For a general number field, this is not always true: if $\Delta_{\text{min}} := \prod_{\mathfrak{p}} \mathfrak{p}^{\nu_{\mathfrak{p}}(\Delta_{\mathfrak{p},\text{min}})}$ is the supposed minimal discriminant, this may not be a principal ideal. One can prove that E/F admits a minimal Weierstraß equation over $\mathcal{O}_F$ if and only if Δ_{min} is principal.

6.3 Eichler–Shimura congruence relation

References: [DS05, Sections 8.6-8.7]. See also these notes.

We will prove that if $f \in S_2(\Gamma_0(N))$ is a normalized newform with $\mathbb{Q}(f) = \mathbb{Q}$, then $a_p(f) = a_p(E_f)$ for all $p \nmid N$. In other words, we will prove that $L(f, s)$ and $L(E_f, s)$ agree outside of finitely many factors.

We saw before that we have a complex uniformization

$$\begin{aligned} Y_0(N)_{\mathbb{C}} &= \{(E, C) \colon C \subseteq E \text{ is a } \mathbb{Z}/N\mathbb{Z} \text{ subgroup}\}/\sim \\ &= \{(E' \to E) \text{ cyclic } N\text{-isogeny}\}/\sim . \end{aligned} \tag{6.30}$$

Similarly, we can give a moduli description for $X_0(N)_{\mathbb{Q}}$.

Definition 6.9. For a base scheme S, an *elliptic curve over* S is a proper smooth morphism $p \colon E \to S$ with a section $e \colon S \to E$, whose geometric fibers are connected curves of genus 1.

Remark 6.8. It follows that an elliptic curve E/S is automatically a commutative group scheme over S.

Definition 6.10. And isogeny $f \colon E \to E'$ of elliptic curves over S is a morphism which is compatible with the identity sections. Equivalently, isogenies are morphisms $f \colon E \to E'$ which respect the group scheme structure.

If $\deg(f)$ is invertible in $\mathcal{O}_S$, similarly as in the case of fields we have that E' is determined by $\ker(f)$, which is a finite étale subgroup scheme of E. In such case, we say that f is *cyclic* if $\ker(f)$ has a generator étale locally.

Definition 6.11. We consider the moduli problem $\mathcal{E}_0(N)\colon \mathrm{Sch}_{\mathbb{Z}[1/N]} \to \mathrm{Set}$ to be

$$\begin{aligned}\mathcal{E}_0(N)(S) &= \{(E \to E') \text{ cyclic isogeny of degree } N\}/\sim \\ &= \{(E, C) \text{ where } C \subseteq E \text{ is cyclic of size } N\}/\sim\end{aligned} \tag{6.31}$$

where all the objects on the right hand side are over S.

Theorem 6.14. *$\mathcal{E}_0(N)$ has a coarse moduli scheme which is smooth over $\mathbb{Z}[1/N]$. We denote it by $Y_0(N)_{\mathbb{Z}[1/N]}$. It also has a compactification $X_0(N)_{\mathbb{Z}[1/N]}$, which is a projective smooth scheme. This notation is compatible with previous notations, as, over $\mathbb{Q}$, this compactification of the coarse moduli scheme is the previously defined $X_0(N)_{\mathbb{Q}}$.*

We denote $J_0(N)_{\mathbb{Z}[1/N]}$ the Jacobian of $X_0(N)_{\mathbb{Z}[1/N]}$. This is an abelian variety over $\mathbb{Z}[1/N]$ as $X_0(N)_{\mathbb{Z}[1/N]}$ is smooth.

Theorem 6.15 (Eichler–Shimura congruence relation)**.** *For $p \nmid N$, the image of $T_p \in \mathrm{End}(J_0(N)_{\mathbb{Q}})$ in $\mathrm{End}(J_0(N)_{\mathbb{F}_p})$ is equal to $\mathrm{Frob}_p + \mathrm{Ver}_p$.*

Proof Sketch. To analyze T_p in $X_0(N)_{\mathbb{F}_p}$, we need to make sense of $X_0(Np)_{\mathbb{F}_p}$. We make the following ad-hoc definition: for a $\mathbb{Z}[1/N]$-scheme S, we define

$$Y_0(Np)(S) = \{(E \to E', C) \text{ where } E \to E' \text{ has degree } p, \text{ and } C \subseteq E \text{ is cyclic of degree } N\}/\sim \tag{6.32}$$

and again this has a suitable compactification $X_0(Np)_{\mathbb{F}_p}$. However, this is not a smooth curve.

We consider the correspondence

$$\begin{array}{ccc} & X_0(Np)_{\mathbb{Z}[1/N]} & \\ {\scriptstyle\alpha}\swarrow & & \searrow{\scriptstyle\beta} \\ X_0(N)_{\mathbb{Z}[1,N]} & & X_0(N)_{\mathbb{Z}[1,N]} \end{array} \tag{6.33}$$

where $\alpha(E \to E', C) = (E, C)$ and $\beta(E \to E', C) = (E', C')$. Here C' is the image of C in E'. Over $\mathbb{Q}$, this is the same as the correspondence T_p, so we need to understand the correspondence

$$\begin{array}{ccc} & X_0(Np)_{\mathbb{F}_p} & \\ {\scriptstyle\alpha}\swarrow & & \searrow{\scriptstyle\beta} \\ X_0(N)_{\mathbb{F}_p} & & X_0(N)_{\mathbb{F}_p} \end{array} \tag{6.34}$$

Consider $\phi\colon E \to E'$ a degree p isogeny over $\mathbb{F}_p$. Then $\phi\hat{\phi} = [p]$ is inseparable, so either ϕ or $\hat{\phi}$ is inseparable. Since $\deg(\phi) = \deg(\hat{\phi}) = p$, this implies

that either ϕ or $\hat{\phi}$ is isomorphic to Frob_p. In other words, ϕ is, up to isomorphism, either Frob_p or Ver_p. This means that we have two loci $X_0(Np)^{\phi\simeq\mathrm{Frob}_p}_{\mathbb{F}_p}$ and $X_0(Np)^{\phi\simeq\mathrm{Ver}_p}_{\mathbb{F}_p}$ of $X_0(Np)_{\mathbb{F}_p}$. These loci intersect exactly at the points $(E \xrightarrow{\mathrm{Frob}_p} E^{(p)}, C)$ for which $\widehat{\mathrm{Frob}_p}$ is inseparable, which we saw is precisely the locus where E is supersingular.

Now both loci $X_0(Np)^{\phi\simeq\mathrm{Frob}_p}_{\mathbb{F}_p}$ and $X_0(Np)^{\phi\simeq\mathrm{Ver}_p}_{\mathbb{F}_p}$ are isomorphic to $X_0(N)_{\mathbb{F}_p}$, and thus $X_0(Np)_{\mathbb{F}_p}$ is the union of two copies of $X_0(N)_{\mathbb{F}_p}$ glued along the (finitely many) supersingular points. Moreover, the supersingular points are glued along the Frobenius map, which defines the following involution on the supersingular locus $X_0(N)^{ss}_{\mathbb{F}_p}$:

$$(E \xrightarrow{\mathrm{Frob}_p} E^{(p)}, C) \mapsto (E^{(p)} \xrightarrow{\mathrm{Frob}_p} E^{(p^2)}, C^{(p)}) \simeq (E^{(p)} \xrightarrow{\mathrm{Ver}_p} E, C^{(p)}). \quad (6.35)$$

We write this as the following diagram

$$\begin{array}{ccccc}
 & & X_0(N)^{ss}_{\mathbb{F}_p} & \xrightarrow[\sim]{\mathrm{Frob}_p} & X_0(N)^{ss}_{\mathbb{F}_p} \\
 & \swarrow & & & \searrow \\
X_0(N)_{\mathbb{F}_p} & & & & X_0(N)_{\mathbb{F}_p} \\
 & \searrow^{r} & & \swarrow_{s} & \\
\Big\downarrow{\scriptstyle \mathrm{id}} & & X_0(Np)_{\mathbb{F}_p} & & \Big\downarrow{\scriptstyle \mathrm{id}} \\
 & \swarrow_{\alpha} & & \searrow^{\beta} & \\
X_0(N)_{\mathbb{F}_p} & & & & X_0(N)_{\mathbb{F}_p}
\end{array} \quad (6.36)$$

Here $r(E, C) = (E \xrightarrow{\mathrm{Frob}_p} E^{(p)}, C^{(p)})$ and $s(E, C) = (E^{(p)} \xrightarrow{\mathrm{Ver}_p} E, C)$. So away from $X_0(N)^{ss}_{\mathbb{F}_p}$, we have that

$$T_p \equiv \begin{array}{ccc} & X_0(N)_{\mathbb{F}_p} & \\ {\scriptstyle \mathrm{id}}\swarrow & & \searrow{\scriptstyle \beta\circ r} \\ X_0(N)_{\mathbb{F}_p} & & X_0(N)_{\mathbb{F}_p} \end{array} + \begin{array}{ccc} & X_0(N)_{\mathbb{F}_p} & \\ {\scriptstyle \alpha\circ s}\swarrow & & \searrow{\scriptstyle \mathrm{id}} \\ X_0(N)_{\mathbb{F}_p} & & X_0(N)_{\mathbb{F}_p} \end{array} \quad (6.37)$$

But $\beta \circ r = \alpha \circ s = \mathrm{Frob}_p$, and thus $T_p \equiv \mathrm{Frob}_p + \mathrm{Frob}_p^\intercal$ in $X_0(N)_{\mathbb{F}_p} \backslash X_0(N)^{ss}_{\mathbb{F}_p}$.

This implies that the two endomorphisms T_p and $\mathrm{Frob}_p + \mathrm{Ver}_p \in \mathrm{End}(J_0(N)_{\mathbb{F}_p})$ agree in a Zariski dense subset, and hence must be the same. □

Corollary 6.16. *If $f \in S_2(\Gamma_0(N))$ is a normalized newform with $\mathbb{Q}(f) = \mathbb{Q}$, then $a_p(f) = a_p(E_f)$ for all $p \nmid N$.*

Proof. By definition, E_f is a quotient of $J_0(N)_{\mathbb{Q}}$ under which T_p acts as $[a_p(f)]$. By the Eichler–Shimura relation, we thus have that

$$[a_p(f)] = T_p = \mathrm{Frob}_p + \mathrm{Ver}_p \in \mathrm{End}(E_{f,\mathbb{F}_p}). \tag{6.38}$$

But we have seen before that $\mathrm{Frob}_p + \mathrm{Ver}_p = [a_p(E)]$ in $\mathrm{End}(E_{f,\mathbb{F}_p})$, and thus the claim follows. □

7 Galois representations of elliptic curves

7.1 Elliptic curves over complete DVRs

References: [Sil09, Sections IV.1-IV.3, VII.2-VII.3].

Let R be a complete DVR with maximal ideal $\mathfrak{m}$ and residue field k, let K be its fraction field. We consider E/K and a minimal Weierstraß model $\tilde{E}/R$.

We note that we have a reduction map

$$\widetilde{(\cdot)} \colon E(K) \to \tilde{E}(k) \tag{7.1}$$

sending $[x : y : z]$ to $[\lambda x : \lambda y : \lambda z]$ for any choice of $\lambda \in K$ with $\lambda x, \lambda y, \lambda z \in R$ and with at least one of them not in $\mathfrak{m}$. As an application of Hensel's lemma, we have

Proposition 7.1. *Let $E_0(K)$ be the preimage of $\tilde{E}^{ns}(k)$ under the reduction map. Then $E_0(K) \to \tilde{E}^{ns}(k)$ is a surjective homomorphism.*

It turns out that the quotient $E(K)/E_0(K)$ is finite:

Theorem 7.2 (Kodaira, Néron)**.** *If E has split multiplicative reduction, then $E(K)/E_0(K)$ is a cyclic group of order $\nu(\Delta) = -\nu(j)$. In all other cases, $E(K)/E_0(K)$ is a group of order at most 4.*

Denote $E_1(K)$ the kernel of the reduction map. It corresponds to points $[x : y : z]$ with $\nu(x), \nu(z) > \nu(y)$.

Remark 7.1. In terms of Néron models, recall that we have $\tilde{E}^{ns}(k) = \mathcal{E}_k^0(k)$. The surjective map $E(K) = \mathcal{E}(R) \to \mathcal{E}_k(k)$ identifies

$$E(K)/E_0(K) \simeq \Phi(k), \quad E_0(K)/E_1(K) \simeq \mathcal{E}_k^0(k) \tag{7.2}$$

where $\Phi = \mathcal{E}_k/\mathcal{E}_l^0$ is the group of components of $\mathcal{E}$, which is a finite étale group scheme.

To analyze the group $E_1(K)$, it is convenient to make a change of coordinates so that the identity is at $(0,0)$. Let $z = -x/y$ and $w = -z/y$. Then we are looking at

$$w = z^3 + a_1zw + a_2z^2w + a_3w^2 + a_4zw^2 + a_6w^3 = f(z,w). \tag{7.3}$$

and points of $E_1(K)$ correspond to solutions with $z, w \in \mathfrak{m}$. Note that if we are given z, then w is fully determined: we can write $w = f(z,w) = f(z,f(z,w)) = f(z,f(z,f(z,w))) = \cdots$ and this converges to a power series $w(z) \in \mathbb{Z}[a_1,a_2,a_3,a_4,a_6][\![z]\!]$. So the group structure of $E_1(K)$ is captured by a single function $z_3 = F(z_1,z_2)$ where $(z_3, w(z_3)) = (z_1,w(z_1)) + (z_2,w(z_2))$. An explicit computation shows

Proposition 7.3. *We have $F(X,Y) \in \mathbb{Z}[a_1,a_2,a_3,a_4,a_6][\![X,Y]\!]$ and $F(X,Y) = X + Y +$ (higher order terms).*

Definition 7.1. A *(1-dimensional commutative) formal group over R*[11] is a power series $F(X,Y) \in R[\![X,Y]\!]$ such that $F(X,Y) = X+Y+$(higher order terms), $F(X,F(Y,Z)) = F(F(X,Y),Z)$ and $F(X,Y) = F(Y,X)$. For a formal group F, we denote $F(\mathfrak{m})$ the set $\mathfrak{m}$ with group structure given by $x +_F y = F(x,y)$. A homomorphism of formal groups $f\colon F \to G$ is $f \in R[\![T]\!]$ such that $f(F(X,Y)) = G(f(X),f(Y))$.

Remark 7.2. It is automatic from this definition that $F(X,0) = X$, $F(0,Y) = Y$ and that there exists $i(T) \in R[\![T]\!]$ with $F(T,i(T)) = F(i(T),T) = 0$.

Definition 7.2. For a 1-dimensional formal group F, its *associated group* $F(\mathfrak{m})$, is the set $\mathfrak{m}$ equipped with the group operation $x +_F y := F(x,y)$.

Example 7.1.

1. For E/K an elliptic curve, the construction above gives a formal group $\hat{E}$ such that $E_1(K) = \hat{E}(\mathfrak{m})$.
2. $\widehat{\mathbb{G}_a}(x,y) = x + y$.
3. $\widehat{\mathbb{G}_m}(x,y) = x + y + xy = (1+x)(1+y) - 1$.
4. For any formal group F, we have the multiplication by m homomorphism $[m]\colon F \to F$ given by

$$[m](T) := \begin{cases} F(T,F(T,\cdots)) & \text{if } m > 0, \\ i([-m](T)) & \text{if } m < 0, \end{cases} \tag{7.4}$$

where F appears m times in the first line. Note that we have

$$[m](T) = mT + (\text{higher order terms}). \tag{7.5}$$

[11] This definition works for any ring R.

Lemma 7.4. *Let $f(T) = aT + (\text{higher order terms})$ be a homomorphism $f\colon F \to G$ of formal groups. Assume $a \neq 0$. Then f is an isomorphism if and only if $a \in R^\times$.*

Proof. If $g(T) = bT + (\text{higher order terms})$ is a homomorphism $g\colon G \to F$, it is clear that $fg(T) = abT + (\text{higher order terms})$, so having $a \in R^\times$ is necessary.

The inverse g of f can be constructed by induction: we first construct $g \in TR[\![T]\!]$ such that $f(g(T)) = T$. Let $g_1(T) = a^{-1}T$, and inductively define $g_{n+1}(T) = g_n(T) + \lambda T^{n+1}$ to satisfy $f(g_i(T)) \equiv T \mod T^{i+1}$. We are looking at

$$f(g_{n+1}(T)) = f(g_n(T) + \lambda T^{n+1}) \equiv f(g_n(T)) + a\lambda T^{n+1} \mod T^{n+2} \tag{7.6}$$

and thus we can choose λ to complete the induction. This constructs a right inverse of f.

Similarly, we can construct a left inverse $h \in TR[\![T]\!]$ with $h(f(T)) = T$, and then we must have $h(T) = h(f(g(T))) = g(T)$.

To check $g\colon G \to F$ is a homomorphism, we need to check

$$g(F(x, y)) = G(g(x), g(y)). \tag{7.7}$$

Equivalently, we may change $x \mapsto f(x)$ and $y \mapsto f(y)$ and apply f to both sides of the equation, so that it reduces to $F(f(x), f(y)) = f(G(x, y))$, which is given. □

Corollary 7.5. *Let m be an integer with $\operatorname{char}(k) \nmid m$. Then if F is a formal group as above, we have $F(\mathfrak{m})[m] = 0$. In particular, if E/K is an elliptic curve, then $E_0(K)[m] \simeq \tilde{E}^{ns}(k)[m]$.*

Proof. The first claim is immediate from the proposition since $[m]\colon F \to F$ is an isomorphism.

The second claim follows from the Snake lemma on the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \hat{E}(\mathfrak{m}) & \longrightarrow & E_0(K) & \longrightarrow & \tilde{E}^{ns}(k) & \longrightarrow & 0 \\ & & \sim\downarrow{\scriptstyle [m]} & & \downarrow{\scriptstyle [m]} & & \downarrow{\scriptstyle [m]} & & \\ 0 & \longrightarrow & \hat{E}(\mathfrak{m}) & \longrightarrow & E_0(K) & \longrightarrow & \tilde{E}^{ns}(k) & \longrightarrow & 0 \end{array} \tag{7.8}$$

□

7.2 Tate modules of elliptic curves

References: [Sil09, Sections III.7, VII.4, VII.7].

Let k be a perfect field.

Definition 7.3. Consider E/k an elliptic curve. For a rational prime ℓ, we consider the *ℓ-adic Tate module of E* to be

$$T_\ell E := \varprojlim_n E[\ell^n] \tag{7.9}$$

where the transition maps are $[\ell]\colon E[\ell^{n+1}] \to E[\ell^n]$.

Note that since each $E[\ell^n]$ carries an action of $\mathrm{Gal}(\bar{k}/k)$, $T_\ell E$ also carry such an action. Since the transition maps are surjective, $T_\ell E$ is also a continuous $\mathbb{Z}_\ell$-module.

Proposition 7.6. *As a $\mathbb{Z}_\ell$-module, we have that*

$$T_\ell E \simeq \begin{cases} \mathbb{Z}_\ell^2 & \text{if } \ell \neq \mathrm{char}(k), \\ \mathbb{Z}_\ell & \text{if } \ell = \mathrm{char}(k) \text{ and } E \text{ is ordinary}, \\ 0 & \text{if } \ell = \mathrm{char}(k) \text{ and } E \text{ is supersingular}. \end{cases} \tag{7.10}$$

Proof. This follows from what we have seen before that

$$E[\ell^n] \simeq \begin{cases} (\mathbb{Z}/\ell^n\mathbb{Z})^2 & \text{if } \ell \neq \mathrm{char}(k), \\ (\mathbb{Z}/\ell^n\mathbb{Z}) & \text{if } \ell = \mathrm{char}(k) \text{ and } E \text{ is ordinary}, \\ 0 & \text{if } \ell = \mathrm{char}(k) \text{ and } E \text{ is supersingular}, \end{cases} \tag{7.11}$$

and that the transition maps $E[\ell^{n+1}] \xrightarrow{[\ell]} E[\ell^n]$ are surjective. □

Definition 7.4. We define $V_\ell E := T_\ell E \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$ the ℓ-adic Galois representation attached to E, and we denote $\rho_{E,\ell}\colon \mathrm{Gal}(\bar{k}/k) \to \mathrm{GL}(V_\ell E)$.

Proposition 7.7. *Let E_1, E_2 be elliptic curves over k and $\ell \neq \mathrm{char}(k)$. Then*

$$\mathrm{Hom}(E_1, E_2) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \to \mathrm{Hom}_{\mathbb{Z}_\ell[\mathrm{Gal}(\bar{k}/k)]}(T_\ell E_1, T_\ell E_2) \tag{7.12}$$

is injective.

Proof. For any $\phi \in \mathrm{Hom}(E_1, E_2) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell$, we can choose a finitely generated submodule $M \subseteq \mathrm{Hom}(E_1, E_2)$ such that $\phi \in M \otimes_{\mathbb{Z}} \mathbb{Z}_\ell$. We claim its saturation M^{div} is still finitely generated. Consider the finite dimensional $\mathbb{R}$-vector space $M \otimes \mathbb{R}$. The degree map $\deg\colon M \to \mathbb{Z}$ extends to a continuous map $\deg\colon M \otimes$

$\mathbb{R} \to \mathbb{R}$. Then $M^{div} \subseteq M \otimes \mathbb{R}$ is a discrete submodule, since $\deg(M^{div} \backslash \{0\}) \geqslant 1$. Hence M^{div} is finitely generated, and thus free.

Let $\phi_1, \ldots, \phi_m$ be a basis of M^{div}, and we write $\phi = \sum_i \phi_i \otimes a_i$. Then ϕ is in the kernel of the above map iff for every n we have that $\sum_i \phi_i \otimes (a_i \mod \ell^n)|_{E_1[\ell^n]} = 0$. Since $\ell \neq \mathrm{char}(k)$, this is the same as having $\phi = \lambda \circ [\ell^n]$ for some $\lambda \in M \otimes_{\mathbb{Z}} \mathbb{Z}_\ell$, as M^{div} is saturated. This means that $a_i \mod \ell^n = 0$ for all i, n, and thus that $\phi = 0$.

Note that, a posteriori, this result implies that $\mathrm{Hom}(E_1, E_2)$ is finitely generated, and hence that it is a free $\mathbb{Z}$-module, so in fact we could have taken $M = \mathrm{Hom}(E_1, E_2)$. But we don't know a priori that this is free! □

Remark 7.3. The above map is an isomorphism in the following two cases:

1. (Tate) when k is a finite field.
2. (Faltings) when k is a number field.

Proposition 7.8. *Let F be a number field and E/F be an elliptic curve with good reduction at $\mathfrak{p}$, and $\mathfrak{p} \nmid \ell$. Denote $\tilde{E}/k_\mathfrak{p}$ the reduction of E. Then the $\mathrm{Gal}(\bar{F}/F)$-action on $T_\ell E$ is unramified at $\mathfrak{p}$, and the reduction map $T_\ell E \to T_\ell \tilde{E}$ is an isomorphism of $\mathbb{Z}_\ell[\mathrm{Gal}(\bar{k_\mathfrak{p}}/k_\mathfrak{p})]$-modules.*

Proof. Let $\mathfrak{p} \nmid m$. Let F'/F be a finite extension and $\mathfrak{p}'$ a prime above $\mathfrak{p}$. We consider $E(F') \subseteq E(F'_{\mathfrak{p}'}) \twoheadrightarrow \tilde{E}(k_{\mathfrak{p}'})$. Then $E(F')[m] \subseteq E(F'_{\mathfrak{p}'})[m] = \tilde{E}(k_{\mathfrak{p}'})[m]$ as $\hat{E}(\mathfrak{m}_{\mathfrak{p}'})[m] = 0$. In other words, the reduction map $E(F')[m] \to \tilde{E}(k_{\mathfrak{p}'})[m]$ is injective. This implies that $E(F')[m]$ is unramified at $\mathfrak{p}$, since the inertia (by definition) is the subgroup that acts trivially on $k_{\mathfrak{p}'}$.

Hence $E[m]$ is unramified at $\mathfrak{p}$, and $E[m] \to \tilde{E}[m]$ is injective. Since both groups are isomorphic to $(\mathbb{Z}/m\mathbb{Z})^2$, this is an isomorphism. □

Corollary 7.9. *In the situation of the proposition above, the characteristic polynomial of a Frobenius element $\mathrm{Frob}_\mathfrak{p} \in \mathrm{Gal}(\bar{F}/F)$ on $T_\ell E$ is $T^2 - a_\mathfrak{p}(E)T + q_\mathfrak{p}$.*

Proof. Since $[a_\mathfrak{p}(E)] = \mathrm{Frob}_\mathfrak{p} + \widehat{\mathrm{Frob}_\mathfrak{p}}$, we have $\mathrm{Frob}_\mathfrak{p}^2 + [q_\mathfrak{p}] = [a_\mathfrak{p}(E)]\mathrm{Frob}_\mathfrak{p}$ in $\mathrm{End}(\tilde{E})$, note that $\mathrm{Frob}_\mathfrak{p} \in \mathrm{End}(\tilde{E})$ induce the same action as $\mathrm{Frob}_\mathfrak{p} \in \mathrm{Gal}(\bar{F}/F)$ in $\mathrm{End}(T_\ell \tilde{E}) = \mathrm{End}(T_\ell E)$.

If this was not the characteristic polynomial, then $\mathrm{Frob}_\mathfrak{p}$ would act as a scalar in $T_\ell E$, say $\alpha \in \mathbb{Z}_\ell$. Then if $\alpha_n \in \mathbb{Z}$ is with $\alpha_n \equiv \alpha \mod \ell^n$ and $|\alpha_n| \leqslant \ell^n/2$, we would have $(\mathrm{Frob}_\mathfrak{p} - [\alpha_n])|_{E[\ell^n]} = 0$, and so

$$0 + \ell^{2n} \leqslant \deg(\mathrm{Frob}_\mathfrak{p} + [\alpha_n]) + \deg(\mathrm{Frob}_\mathfrak{p} - [\alpha_n]) = 2(q_\mathfrak{p} + \alpha_n^2) \leqslant 2q_\mathfrak{p} + \ell^{2n}/2, \tag{7.13}$$

which is a contradiction for n sufficiently large. □

In fact, the above proposition is an if and only if.

Theorem 7.10 (Néron–Ogg–Shafarevich)**.** *Let F be a number field, E/F be an elliptic curve and $\mathfrak{p}$ a prime of F. Then the following are equivalent.*

1. *E has good reduction at $\mathfrak{p}$.*
2. *For every $\mathfrak{p} \nmid \ell$, the Tate module $T_\ell E$ is unramified at $\mathfrak{p}$.*
3. *For a single $\mathfrak{p} \nmid \ell$, the Tate module $T_\ell E$ is unramified at $\mathfrak{p}$.*

Proof. (1) $\implies$ (2) follows from the above, and (2) $\implies$ (3) is obvious.

We consider the general case of an elliptic curve E and a prime $\mathfrak{p}$. We consider the reduction $\tilde{E}/k_\mathfrak{p}$ (which may be singular). Note that the inclusion $E[\ell^n] \subseteq E(\bar{F}_\mathfrak{p})[\ell^n]$ is an equality[12], as both are isomorphic to $(\mathbb{Z}/\ell^n\mathbb{Z})^2$. Thus

$$(T_\ell E)^{I_\mathfrak{p}} = \varprojlim_n E(F_\mathfrak{p}^{unr})[\ell^n]. \tag{7.14}$$

The group $E(F_\mathfrak{p}^{unr})/E_0(F_\mathfrak{p}^{unr})$ is finite since $\mathcal{O}_{F_\mathfrak{p}^{unr}}$ is a DVR, and we also have $E_0(F_\mathfrak{p}^{unr})[\ell^n] = \tilde{E}^{ns}[\ell^n]$ by the same proof as in the above proposition. We conclude that we have an injection

$$T_\ell \tilde{E}^{ns} \hookrightarrow (T_\ell E)^{I_\mathfrak{p}} \tag{7.15}$$

with finite cokernel (bounded by the Tamagawa number $\#E(F_\mathfrak{p}^{unr})/E_0(F_\mathfrak{p}^{unr})$). In particular,

$$(V_\ell E)^{I_\mathfrak{p}} \simeq V_\ell \tilde{E}^{ns}. \tag{7.16}$$

By our previous analysis of $\tilde{E}^{ns}$ in the case of bad reduction, we have

$$T_\ell \tilde{E}^{ns} \simeq \begin{cases} \mathbb{Z}_\ell^2 & \text{good reduction,} \\ \mathbb{Z}_\ell & \text{multiplicative reduction,} \\ 0 & \text{additive reduction,} \end{cases} \tag{7.17}$$

and thus $V_\ell E$ is necessarily ramified at $\mathfrak{p}$ if E has bad reduction at $\mathfrak{p}$. □

7.3 Artin L-functions

The discussion that follows can be done for Galois representations for $\mathrm{Gal}(\bar{F}/F)$ for a general number field F, but we stick with $F = \mathbb{Q}$ for simplicity.

[12]Here we are implicitly choosing an embedding $\bar{F} \hookrightarrow \bar{F}_\mathfrak{p}$, that is, a compatible choice of places above $\mathfrak{p}$ for all number fields above F.

Definition 7.5. A *compatible system of Galois representations* $(\rho_\ell, V_\ell)_\ell$ is a collection of continuous representations $\rho_\ell \colon \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(V_\ell)$ where V_ℓ are finite dimensional $\mathbb{Q}_\ell$-vector spaces, satisfying the following: for every prime p, the characteristic polynomials

$$\mathrm{char}\left(\rho_\ell(\mathrm{Frob}_p) \mid V_\ell^{I_p}\right) \in \mathbb{Q}_\ell[x], \quad \text{for } \ell \neq p \tag{7.18}$$

have coefficients in $\mathbb{Q}$ and are independent of $\ell \neq p$.

Definition 7.6. Given a compatible system $\rho = (\rho_\ell, V_\ell)_\ell$ of Galois representations, we define its *Artin L-function* $L(\rho, s) = \prod_p L_p(\rho, s)$ where

$$L_p(\rho, s) = \det\left(1 - p^{-s}\rho_\ell(\mathrm{Frob}_p^{-1}) \mid V_\ell^{I_p}\right)^{-1}. \tag{7.19}$$

Example 7.2.

1. The trivial compatible system $\rho = (\rho_\ell, \mathbb{Q}_\ell)$ has L-function $L(\rho, s) = (1 - p^{-s})^{-1} = \zeta(s)$.
2. We denote the *Tate twist* $\mathbb{Q}(1) = \{\mathbb{Q}_\ell(1)\}_\ell$ where $\mathbb{Q}_\ell(1) := V_\ell\mathbb{G}_m$. This is unramified at all $p \neq \ell$, and $\mathrm{Frob}_\mathfrak{p}$ acts as p. Thus for any compatible system $\rho = (\rho_\ell, V_\ell)_\ell$, we have

$$L(\rho \otimes \mathbb{Q}(1), s) = \det(1 - p^{-s}\rho_\ell(\mathrm{Frob}_p^{-1})p^{-1} \mid V_\ell^{I_p})^{-1} = L(\rho, s+1). \tag{7.20}$$

   It is also common to denote $\rho(k) := \rho \otimes \mathbb{Q}(1)^{\otimes k}$ for $k \in \mathbb{Z}$.

Proposition 7.11. *If $E/\mathbb{Q}$ is an elliptic curve and $\rho = (V_\ell E)_\ell$ is the compatible system of its Tate modules, then $L(E, s) = L(\rho, s - 1)$.*

Proof. If p is a prime of good reduction for E and $\ell \neq p$, we have that the characteristic polynomial of Frob_p in $V_\ell E$ is $T^2 - a_p(E)T + p$. In particular, the characteristic polynomial of $p^{-s}\mathrm{Frob}_p^{-1}$ is

$$T^2 - a_p(E)p^{-s-1}T + p^{-2s-1}, \tag{7.21}$$

and hence

$$L_p(\rho, s-1) = (1 - a_p(E)p^{-s} + p^{1-2s})^{-1} = L_p(E, s). \tag{7.22}$$

For primes p of bad reduction, we have seen before that $(V_\ell E)^{I_\mathfrak{p}} = V_\ell \tilde{E}^{ns}$, and then the claim follows from the previous classification of $\tilde{E}^{ns}$: i) if E has additive reduction, $L_p(E, s) = 1$ since $V_\ell \tilde{E}^{ns} = 0$, ii) if E has split multiplicative reduction, then $V_\ell \tilde{E}^{ns} = V_\ell \mathbb{G}_m \simeq \mathbb{Q}_\ell$, where Frob_p acts by multiplication by p, and thus

$$L_p(\rho, s-1) = (1 - p^{-(s-1)}p^{-1})^{-1} = (1 - p^{-s})^{-1} = L_p(E, s), \tag{7.23}$$

iii) similarly, in the case of nonsplit multiplicative reduction we have that Frob_p acts by $-p$, and thus $L_p(\rho, s-1) = (1 + p^{-s})^{-1} = L_p(E, s)$. □

7.4 Weil pairing and étale cohomology

Let k be a perfect field, and $\ell \neq \mathrm{char}(k)$. We consider E/k an elliptic curve.

Theorem 7.12. *There exists a perfect, alternating, Galois equivariant pairing, called the* Weil pairing

$$e\colon T_\ell E \times T_\ell E \to \mathbb{Z}_\ell(1). \tag{7.24}$$

Furthermore, if $\varphi\colon E_1 \to E_2$ is an isogeny, then $e(\varphi(P), Q) = e(P, \hat{\varphi}(Q))$. Equivalently, we have an isomorphism of Galois representations

$$\det T_\ell E := \bigwedge^2 T_\ell E \simeq \mathbb{Z}_\ell(1), \tag{7.25}$$

and this is such that an isogeny $\varphi\colon E_1 \to E_2$ induces the map $\mathbb{Z}_\ell(1) \simeq \det T_\ell E_1 \to \det T_\ell E_2 \simeq \mathbb{Z}_\ell(1)$ given by multiplication by $\deg(\varphi)$[13].

For an elementary proof, see [Sil09, Section III.8]. Instead, we will discuss this in terms of étale cohomology.

Proposition 7.13. *Let A/k be an abelian variety. We have $(T_\ell A)^* := \mathrm{Hom}(T_\ell A, \mathbb{Z}_\ell) \simeq H^1_{\text{ét}}(A_{\bar{k}}, \mathbb{Z}_\ell)$ as $\mathbb{Z}_\ell[\mathrm{Gal}(\bar{k}/k)]$-modules.*

Proof Sketch. We have that $H^1_{\text{ét}}(A_{\bar{k}}, \mathbb{Z}_\ell) = \mathrm{Hom}(\pi_1^{\text{ét}}(A_{\bar{k}}), \mathbb{Z}_\ell)$,[14] so it remains to show that $\pi_1^{\text{ét}}(A_{\bar{k}}) \otimes \mathbb{Z}_\ell = T_\ell A$.

By definition, we have

$$\pi_1^{\text{ét}}(A_{\bar{k}}) = \varprojlim_{\substack{\varphi\colon A' \to A_{\bar{k}} \\ \text{finite étale}}} \mathrm{Aut}(A' \to A_{\bar{k}}). \tag{7.26}$$

It turns out that every such A' is also an abelian variety: for the case of elliptic curves, Riemann–Hurwitz implies that $2 - 2g(A') = \deg(\varphi) \cdot (2 - 2g(A_{\bar{k}}))$, and thus A' is also a genus 1 curve. Since all such φ are separable, if follows that they are a factor of $[\deg(\varphi)]$:

$$[\deg(\varphi)]\colon A_{\bar{k}} \to A' \xrightarrow{f} A_{\bar{k}}. \tag{7.27}$$

Thus we may write

$$\pi_1^{\text{ét}}(A_{\bar{k}}) = \varprojlim_N \mathrm{Aut}(A_{\bar{k}} \xrightarrow{[N]} A_{\bar{k}}), \tag{7.28}$$

[13] Note that this is equivalent to $e(\varphi(P'), \varphi(Q')) = \deg(\varphi) \cdot e(P', Q')$. This is implied by $e(\varphi(P), Q) = e(P, \hat{\varphi}(Q))$ by taking $(P, Q) = (P', \varphi(Q'))$, and implies it by taking $(P', Q') = (P, \hat{\varphi}(Q))$ and cancelling $\deg(\varphi)$.

[14] This can be thought as an étale version of the Hurewicz map $\pi_1(X)^{ab} \simeq H_1(X)$.

and $A[N] \simeq \mathrm{Aut}(A_{\bar{k}} \xrightarrow{[N]} A_{\bar{k}})$ via $P \mapsto \tau_P$ the translation isomorphisms. Thus

$$\pi_1^{\text{ét}}(A_{\bar{k}}) \simeq \prod_{\ell} T_\ell A \tag{7.29}$$

and the claim follows. □

Proof of Theorem 7.12. Under the above identification, we have a cup product pairing

$$(T_\ell E)^* \times (T_\ell E)^* = H^1_{\text{ét}}(E_{\bar{k}}, \mathbb{Z}_\ell) \times H^1_{\text{ét}}(E_{\bar{k}}, \mathbb{Z}_\ell) \xrightarrow{\cup} H^2_{\text{ét}}(E_{\bar{k}}, \mathbb{Z}_\ell) \simeq \mathbb{Z}_\ell(-1) \tag{7.30}$$

where the last isomorphism is due to Poincaré duality. By properties of cup product, this is perfect, alternating and Galois equivariant.

This identifies $\det((T_\ell E)^*) \simeq \mathbb{Z}_\ell(-1)$, and taking duals this is the same as an identification $\det(T_\ell E) \simeq \mathbb{Z}_\ell(1)$. For the last claim, if $\varphi\colon E_1 \to E_2$ is an isogeny then we have the following commutative diagram by the functoriality of cup product and Poincaré duality.

$$\begin{array}{ccccc} H^1_{\text{ét}}(E_{1,\bar{k}}, \mathbb{Z}_\ell) \times H^1_{\text{ét}}(E_{1,\bar{k}}, \mathbb{Z}_\ell) & \xrightarrow{\cup} & H^2_{\text{ét}}(E_{1,\bar{k}}, \mathbb{Z}_\ell) & \xrightarrow{\sim} & \mathbb{Z}_\ell(-1) \\ \uparrow{\scriptstyle \varphi^* \times \varphi^*} & & \uparrow{\scriptstyle \varphi^*} & & \uparrow{\scriptstyle \deg(\varphi)} \\ H^1_{\text{ét}}(E_{2,\bar{k}}, \mathbb{Z}_\ell) \times H^1_{\text{ét}}(E_{2,\bar{k}}, \mathbb{Z}_\ell) & \xrightarrow{\cup} & H^2_{\text{ét}}(E_{2,\bar{k}}, \mathbb{Z}_\ell) & \xrightarrow{\sim} & \mathbb{Z}_\ell(-1) \end{array} \tag{7.31}$$

□

Remark 7.4. The usually defined Weil pairing (as in [Sil09, Section III.8]), agrees with the above identification $\det(T_\ell E) \simeq \mathbb{Z}_\ell$ up to a sign.

Remark 7.5. Note that if $E/\mathbb{F}_p$ and $\varphi\colon E \to E^{(p)} = E$ is the Frobenius isogeny, then $\varphi^* \in \mathrm{End}(H^i_{\text{ét}}(E_{\bar{k}}, \mathbb{Z}_\ell))$ agrees with the *geometric Frobenius* $\mathrm{Frob}_p^{-1} \in \mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_p)$.

7.5 Galois representations via étale cohomology

If C/k is a curve over a field perfect field k and $\ell \neq \mathrm{char}(k)$, we also have

$$H^1_{\text{ét}}(C_{\bar{k}}, \mathbb{Z}_\ell) = H^1_{\text{ét}}(\mathrm{Jac}(C)_{\bar{k}}, \mathbb{Z}_\ell). \tag{7.32}$$

In particular, we have the following decomposition as $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-modules from Theorem 6.4

$$H^1_{\text{ét}}(X_0(N)_{\bar{\mathbb{Q}}}, \mathbb{Q}_\ell(1)) = H^1_{\text{ét}}(J_0(N)_{\bar{\mathbb{Q}}}, \mathbb{Q}_\ell(1)) = \prod_f (V_\ell A_f)^*(1)^{\oplus m_f} = \prod_f (V_\ell A_f)^{\oplus m_f} \tag{7.33}$$

where the last equality $(V_\ell A)^*(1) \simeq V_\ell A$ is due to the Weil pairing[15].

As $\mathbb{T}_\mathbb{Z}$ acts on $X_0(N)$ by correspondences, this cohomology group also carries an action of $\mathbb{T}_\mathbb{Z}$[16]. For example, by Hodge theory we have

$$H^1_{\text{Betti}}(X_0(N), \mathbb{C}) = H^0(X_0(N), \Omega) \oplus \overline{H^0(X_0(N), \Omega)} = S_2(\Gamma_0(N)) \oplus \overline{S_2(\Gamma_0(N))} \tag{7.34}$$

as $\mathbb{T}_\mathbb{Z}$-modules, and thus by the discussion in Section 5.4

$$H^1_{\text{Betti}}(X_0(N), \mathbb{C}) = \bigoplus_{\substack{\pi \subseteq \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A})) \\ \pi_\infty \simeq D_2}} \pi_f^{K_0(N)} \otimes \mathbb{C}^2. \tag{7.35}$$

Recall that $\dim \pi_f^{K_0(N)} = \sigma_0(N/M)$ if M is the conductor of π_f. That is, if $f \leftrightarrow \pi$ as in Theorem 5.15, then $m_f = \dim \pi_f^{K_0(N)}$.

Since the action of $\mathbb{T}_\mathbb{Z}$ is by geometric correspondences, the Betti-étale comparison isomorphisms are $\mathbb{T}_\mathbb{Z}$ equivariant. So fixing an isomorphism of fields $\iota_\ell \colon \mathbb{C} \simeq \bar{\mathbb{Q}}_\ell$, we have

$$H^1_{\text{ét}}(X_0(N)_{\bar{\mathbb{Q}}}, \bar{\mathbb{Q}}_\ell) = \bigoplus_{\substack{\pi \subseteq \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A})) \\ \pi_\infty \simeq D_2}} \iota_\ell(\pi_f^{K_0(N)}) \boxtimes \rho_{\pi,\ell} \tag{7.36}$$

as $\mathcal{H}_{K_0(N)} \times \bar{\mathbb{Q}}_\ell[\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})]$-modules, where $\rho_{\pi,\ell}(1) = V_\ell A_f$ for $f \leftrightarrow \pi$. In fact, all this discussion generalizes to the congruence subgroups $\Gamma_1(N)$ as well, and thus in fact

$$\varinjlim_{\substack{K \subseteq \mathrm{GL}_2(\mathbb{A}_f) \\ \text{open compact}}} H^1_{\text{ét}}(X(K \cap \mathrm{GL}_2(\mathbb{Q}))_{\bar{\mathbb{Q}}}, \bar{\mathbb{Q}}_\ell) = \bigoplus_{\substack{\pi \subseteq \mathcal{A}_0(\mathrm{GL}_2(\mathbb{A})) \\ \pi_\infty \simeq D_2}} \iota_\ell(\pi_f) \boxtimes \rho_{\pi,\ell} \tag{7.37}$$

as $\mathrm{GL}_2(\mathbb{A}_f) \times \bar{\mathbb{Q}}_\ell[\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})]$-modules.

Remark 7.6. This is an example of the so-called *Kottwitz conjecture*. More generally, the cohomology of Shimura varieties are expected to realize certain cases of the global Langlands correspondence. More precisely, for certain connected reductive groups G over $\mathbb{Q}$, we can attach *Shimura varieties*, a collection $\{\mathrm{Sh}_G(K)\}_{\substack{K \subseteq G(\mathbb{A}_f) \\ \text{neat open compact}}}$ of quasi-projective algebraic varieties over a certain

[15] In general, for an abelian variety A, the Weil pairing is a perfect alternating pairing $T_\ell A \times T_\ell A^\vee \to \mathbb{Z}_\ell(1)$, but since A and $A^\vee$ are isogenous, we have $V_\ell A \simeq V_\ell A^\vee$. In fact we have a canonical such identification when A is a factor of a Jacobian of a curve, as above.

[16] In fact, the right hand side is a decomposition into $\mathbb{T}_\mathbb{Z}^{(N)}$-isotypic components, where $\mathbb{T}_\mathbb{Z}^{(N)} = \mathrm{im}(\mathbb{Z}[T_n \colon (n, N) = 1] \to \mathrm{End}(S_2(\Gamma_0(N))))$ is the Hecke algebra away from N.

number field E. These admit Hecke actions by finite étale correspondences

$$T_{KgK}: \begin{array}{ccc} \mathrm{Sh}_G(K \cap gKg^{-1}) & \xrightarrow{\simeq} & \mathrm{Sh}_G(g^{-1}Kg \cap K) \\ \downarrow & & \downarrow \\ \mathrm{Sh}_G(K) & & \mathrm{Sh}_G(K), \end{array} \tag{7.38}$$

and thus

$$H^i_{\text{ét}}(\mathrm{Sh}_G, \bar{\mathbb{Q}}_\ell) := \varinjlim_K H^i_{\text{ét}}(\mathrm{Sh}_G(K)_{\bar{E}}, \bar{\mathbb{Q}}_\ell) \tag{7.39}$$

is a $G(\mathbb{A}_f) \times \bar{\mathbb{Q}}_\ell[\mathrm{Gal}(\bar{E}/E)]$-module. For $\pi \subseteq \mathcal{A}(G(\mathbb{A}))$ an irreducible automorphic representation appearing in this cohomology, the Kottwitz conjecture gives a conjectural expression of

$$V_\pi := \sum_i (-1)^i (H^i_{\text{ét}}(\mathrm{Sh}_G, \bar{\mathbb{Q}}_\ell)[\pi]) \tag{7.40}$$

as virtual $\mathrm{Gal}(\bar{E}/E)$-representations in terms of the Langlands parameter of π. See for example Section 3.2.2 of these notes for a discussion on this. In the simplest cases, we expect

$$V_\pi(d/2) \stackrel{?}{=} a(\pi_f) \cdot [r_\mu \circ \sigma_\pi] \tag{7.41}$$

as virtual $\mathrm{Gal}(\bar{E}/E)$-representations, where

- $d = \dim \mathrm{Sh}_G(K)$,
- $a(\pi_f)$ is a certain integer,
- $\sigma_\pi \colon \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to {}^L G(\bar{\mathbb{Q}}_\ell)$ is the Langlands parameter of π,
- $r_\mu \colon {}^L G(\bar{\mathbb{Q}}_\ell) \to \mathrm{GL}(V_\mu)$ is a certain finite dimensional representation determined by the Shimura datum.

In general, this formula is more complicated in cases that the group G has nontrivial endoscopy.

8 Elliptic curves over global fields

8.1 Roadmap for the Mordell–Weil theorem

References: [Sil09, Section VIII.3], although our presentation is slightly different since we will use canonical heights (as in [Sil09, Section VIII.9]).

Theorem 8.1 (Mordell–Weil)**.** *Let E/F be an elliptic curve over a number field F. Then $E(F)$ is a finitely generated abelian group.*

The proof of this relies on two ingredients, which will be the focus of the later sections in this chapter.

Theorem 8.2 (Weak Mordell–Weil)**.** *Let E/F be an elliptic curve over a number field F, and let m be an integer. Then $E(F)/mE(F)$ is a finite group.*

Theorem 8.3 (Néron–Tate)**.** *There exists a height function $\hat{h}_{NT}\colon E(\bar{F}) \to \mathbb{R}_{\geqslant 0}$ satisfying the following properties*

1. *The pairing $\langle \cdot, \cdot \rangle_{NT}\colon E(\bar{F}) \times E(\bar{F}) \to \mathbb{R}$ given by*

$$\langle P, Q \rangle_{NT} = \frac{1}{2}(\hat{h}_{NT}(P+Q) - \hat{h}_{NT}(P) - \hat{h}_{NT}(Q)) \tag{8.1}$$

   is bilinear, and also satisfies that $\hat{h}_{NT}(P) = \langle P, P \rangle_{NT}$.

2. *For any $B > 0$, the set $\{P \in E(F)\colon \hat{h}_{NT}(P) < B\}$ is finite.*

Remark 8.1. Part (1) roughly means that we can think of $\hat{h}_{NT}$ as a quadratic function. In fact, it is easy to see that it implies that $\hat{h}_{NT}(mP) = m^2 \hat{h}_{NT}(P)$ for all $m \in \mathbb{Z}$, and that we have the parallelogram relation

$$\frac{\hat{h}_{NT}(P+Q) + \hat{h}_{NT}(P-Q)}{2} = \hat{h}_{NT}(P) + \hat{h}_{NT}(Q). \tag{8.2}$$

Proof of Mordell–Weil. Let $m \geqslant 2$ be an integer, and $P_1, \dots, P_n \in E(F)$ be a set of representatives for $E(F)/mE(F)$. We let $B_0 = \max_i(\hat{h}_{\mathrm{NT}}(P_i))$. From the parallelogram relation and the positivity of $\hat{h}_{NT}$, we note that we have

$$\hat{h}_{\mathrm{NT}}(P-Q) \leqslant \hat{h}_{\mathrm{NT}}(P-Q) + \hat{h}_{\mathrm{NT}}(P+Q) = 2\hat{h}_{\mathrm{NT}}(P) + 2\hat{h}_{\mathrm{NT}}(Q). \tag{8.3}$$

Let $Q_0 \in E(F)$, and we recursively write $Q_j = mQ_{j+1} + P_{i(j)}$ for $i(0), i(1), \dots \in \{1, \dots, n\}$. We have

$$\hat{h}_{\mathrm{NT}}(Q_{j+1}) = \frac{\hat{h}_{\mathrm{NT}}(Q_j - P_{i(j)})}{m^2} \leqslant \frac{2}{m^2}(\hat{h}_{\mathrm{NT}}(Q_j) + \hat{h}_{\mathrm{NT}}(P_{i(j)})) \leqslant \frac{2}{m^2}(\hat{h}_{\mathrm{NT}}(Q_j) + B_0). \tag{8.4}$$

Repeating this, we have

$$\hat{h}_{\mathrm{NT}}(Q_n) < \left(\frac{2}{m^2}\right)^n \hat{h}_{\mathrm{NT}}(Q_0) + \left(\frac{2}{m^2} + \frac{4}{m^4} + \frac{8}{m^6} + \cdots\right) B_0 \tag{8.5}$$

and thus

$$\hat{h}_{\mathrm{NT}}(Q_n) < \frac{1}{2^n}\hat{h}_{\mathrm{NT}}(Q) + B_0. \tag{8.6}$$

Now for any $\varepsilon > 0$, this implies that there is n with $\hat{h}_{NT}(Q_n) < B + \varepsilon$. This proves that $E(K)$ is generated by the finite sets $\{P \in E(F) \colon \hat{h}_{\mathrm{NT}}(P) < B + \varepsilon\}$. In fact, this implies that it is generated by the finite set $\{P \in E(F) \colon \hat{h}_{NT}(P) \leqslant B_0\}$. □

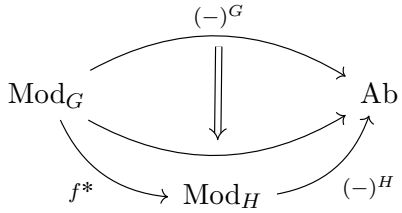
8.2 Group cohomology and Galois cohomology

Group cohomology

Let G be an abstract group. We consider the category Mod_G of $\mathbb{Z}[G]$-modules. We note that this is an abelian category with enough injectives[17] and projectives.

Definition 8.1. The invariants functor $(-)^G \colon \mathrm{Mod}_G \to \mathrm{Ab}$ is left exact and defines derived functors $H^*(G, -) \colon \mathrm{Mod}_G \to \mathrm{Ab}$, the *group cohomology of* G.

Remark 8.2. We note that this is also functorial in the first variable: if $f \colon H \to G$, then we have pullback maps $f^* \colon H^*(G, M) \to H^*(H, M)$. This will soon become clear in terms of cochains, but abstractly this comes from the natural transformation



(8.7)

since the functor $f^* \colon \mathrm{Mod}_G \to \mathrm{Mod}_H$ is exact.

Remark 8.3. Since $M^G = \mathrm{Hom}_{\mathrm{Mod}_G}(\mathbb{Z}, M)$, we also have that $H^*(G, M) = \mathrm{Ext}^*_{\mathrm{Mod}_G}(\mathbb{Z}, M)$.

Using the interpretation of Ext as a *left* derived functor on the first variable, we can compute $H^*(G, M)$ as follows. We consider the following free resolution of $\mathbb{Z}$ in Mod_G.

$$\cdots \to P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{d_0} \mathbb{Z} \to 0 \tag{8.8}$$

where $P_i = \mathbb{Z}[G^{i+1}]$ with diagonal G-action, and where

$$d_i(g_0, \ldots, g_i) = \sum_{j=0}^{i} (-1)^j (g_0, \ldots, \hat{g_j}, \ldots, g_i). \tag{8.9}$$

[17] As far as I understand, this requires the axiom of choice if G is infinite.

Then $H^*(G, M)$ is the cohomology of the complex

$$0 \to \mathrm{Hom}(P_0, M) \xrightarrow{d^0} \mathrm{Hom}(P_1, M) \xrightarrow{d^1} \cdots \tag{8.10}$$

This gives the presentation of $H^*(G, M)$ by *homogeneous cochains*. Rather, it is often easier to work with *inhomogeneous cochains* $C^i(G, M) = \mathrm{Maps}(G^i, M)$ by making the identification

$$\mathrm{Hom}(P_i, M) \simeq C^i(G, M), \quad f \mapsto ((g_1, \ldots, g_i) \mapsto f(1, g_1, g_1g_2, \ldots, g_1 \cdots g_i)). \tag{8.11}$$

In this way the differentials $d^i \colon C^i(G, M) \to C^{i+1}(G, M)$ are given by

$$(d^i c)(g_1, \ldots, g_{i+1}) = g_1 \cdot c(g_2, \ldots, g_{i+1}) + \sum_{j=1}^{i} (-1)^j c(g_1, \ldots, g_j g_{j+1}, \ldots, g_{i+1}) + (-1)^{i+1} c(g_1, \ldots, g_i). \tag{8.12}$$

Example 8.1. We of course have $H^0(G, M) = M^G$, and the above give us the description

$$H^1(G, M) = \frac{Z^1(G, M)}{B^1(G, M)} \tag{8.13}$$

where

$$Z^1(G, M) = \{c \colon G \to M \text{ s.t. } c(g_1g_2) = c(g_1) + g_1 \cdot c(g_2)\} \tag{8.14}$$

and

$$B^1(G, M) = \{(g \mapsto gm - m) \colon m \in M\}. \tag{8.15}$$

Example 8.2. If M is a trivial G-module, then $B^1(G, M) = 0$, and

$$H^1(G, M) = Z^1(B, M) = \mathrm{Hom}(G, M). \tag{8.16}$$

Remark 8.4. If we have an exact sequence $0 \to M_1 \to M_2 \to M_3 \to 0$, then the boundary map $H^0(G, M_3) \to H^1(G, M_1)$ in terms of inhomogeneous cochains is given as follows: if $m \in M_3^G$ and $\tilde{m} \in M_2$ is a lift, then we consider the cochain $c \in C^1(G, M_1)$ to be $c(g) = g\tilde{m} - \tilde{m}$. It is easy to see that $c \in Z^1(G, M_1)$ and that its image in $H^1(G, M_1)$ does not depend on the choice of $\tilde{m}$.

Galois cohomology

If G and M have a topology, it is not as easy to define group cohomology. The category TMod_G of continuous G-modules may not even be abelian![18] For our applications, we will only consider the case where M has *discrete* topology.

[18] Take $G = \mathbb{Z}$ with the discrete topology. Then $\mathrm{TMod}_{\mathbb{Z}} = \mathrm{TAb}$ is the category of topological abelian groups, which is additive but not abelian.

If G is a topological group, the category Mod_G of *discrete* G-modules is abelian and has enough injectives, so we similarly define $H^*(G, M)$.

Proposition 8.4. *If $G = \varprojlim_H G/H$ is profinite, then we also have the following:*

1. *$H^*(G, M) = Z^*(G, M)/B^*(G, M)$ where Z^*, B^* are similarly as above, but considering* continuous *cochains,*
2. *$H^*(G, M) = \varinjlim_H H^*(G/H, M^H)$ where for $H_1 \subseteq H_2$ the transition maps are the inflation maps*

$$H^*(G/H_2, M^{H_2}) \to H^*(G/H_1, M^{H_1}) \tag{8.17}$$

induced by $G/H_1 \twoheadrightarrow G/H_2$ and $M^{H_2} \subseteq M^{H_1}$.

Definition 8.2. If F'/F is a Galois extension of fields and M is a discrete $\mathrm{Gal}(F'/F)$-module, we denote

$$H^*(F'/F, M) := H^*(\mathrm{Gal}(F'/F), M). \tag{8.18}$$

If $F' = F^{sep}$, we also denote this by

$$H^*(F, M) := H^*(\mathrm{Gal}(F^{sep}/F), M). \tag{8.19}$$

Example 8.3. 1. If M is a finite $G_{\mathbb{F}_p}$-module, then

$$H^0(\mathbb{F}_p, M) = M^{G_{\mathbb{F}_p}}, \quad H^1(\mathbb{F}_p, M) = \frac{M}{(\mathrm{Frob}_p - 1)M}, \quad H^i(\mathbb{F}_p, M) = 0 \text{ for } i \geqslant 2. \tag{8.20}$$

2. If E/F is an elliptic curve over a number field F, we have the exact sequence

$$0 \to E[m] \to E(\bar{F}) \xrightarrow{[m]} E(\bar{F}) \to 0, \tag{8.21}$$

which induces

$$0 \to E(F)[m] \to E(F) \xrightarrow{[m]} E(F) \to H^1(F, E[m]) \to H^1(F, E) \to \cdots \tag{8.22}$$

and thus we have the *Kummer map*

$$\kappa \colon E(F)/mE(F) \hookrightarrow H^1(F, E[m]). \tag{8.23}$$

It turns out that $H^1(F, E[m])$ is infinite[19], but we will see in the next section how to bound the image of κ by a finite submodule, the *Selmer group* $\mathrm{Sel}_m(E/F) \subseteq H^1(F, E[m])$.

[19]For example it can happen that $E[m] = E(F)[m]$ and then $H^1(F, E[m]) = \mathrm{Hom}(G_F, E[m])$ which is infinite.

8.3 Selmer groups and weak Mordell–Weil

References: [Sil09, Sections VIII.1-2 and X.4].

Let F be a number field, E/F an elliptic curve and $m \in \mathbb{Z}$ an integer.

As mentioned in the previous section, we consider the exact sequence of $\mathrm{Gal}(\bar{F}/F)$ modules

$$0 \to E(\bar{F})[m] \to E(\bar{F}) \xrightarrow{[m]} E(\bar{F}) \to 0. \tag{8.24}$$

Taking Galois cohomology, we get the *Kummer map*

$$\kappa \colon E(F)/mE(F) \hookrightarrow H^1(F, E(\bar{F})[m]). \tag{8.25}$$

Concretely, $\kappa(P)$ is represented by the cocycle $\kappa(P)(\sigma) = \sigma(Q) - Q$ for a choice of $Q \in E(\bar{F})$ with $mQ = P$. The plan is to try to bound the image of κ.

Given an embedding $\bar{F} \hookrightarrow \bar{F}_v$, we can consider the analogous sequence $0 \to E(\bar{F}_v)[m] \to E(\bar{F}_v) \xrightarrow{[m]} E(\bar{F}_v) \to 0$ which fits into a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(\bar{F})[m] & \longrightarrow & E(\bar{F}) & \longrightarrow & E(\bar{F}) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \sim} & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E(\bar{F}_v)[m] & \longrightarrow & E(\bar{F}_v) & \longrightarrow & E(\bar{F}_v) & \longrightarrow & 0 \end{array} \tag{8.26}$$

They similarly induce local Kummer maps $\kappa_v \colon E(F_v)/mE(F_v) \hookrightarrow E[m]$

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(F)/mE(F) & \xrightarrow{\kappa} & H^1(F, E[m]) & \longrightarrow & H^1(F, E)[m] & \longrightarrow & 0 \\ & & \downarrow & & \downarrow{\scriptstyle \mathrm{loc}_v} & & \downarrow{\scriptstyle \mathrm{loc}_v} & & \\ 0 & \longrightarrow & E(F_v)/mE(F_v) & \xrightarrow{\kappa_v} & H^1(F_v, E[m]) & \longrightarrow & H^1(F_v, E)[m] & \longrightarrow & 0 \end{array} \tag{8.27}$$

Definition 8.3. The *m-Selmer group of E/F* is

$$\begin{aligned} \mathrm{Sel}_m(E/F) &:= \{c \in H^1(F, E[m]) \colon \mathrm{loc}_v(c) \in \mathrm{im}(\kappa_v) \text{ for all } v\} \\ &= \ker\left(H^1(F, E[m]) \to H^1(F, E) \xrightarrow{\prod_v \mathrm{loc}_v} \prod_v H^1(F_v, E) \right) \end{aligned} \tag{8.28}$$

The *Tate–Shafarevich group of E/F is*

$$Ш(E/F) := \ker\left(H^1(F, E) \to \prod_v H^1(F_v, E) \right). \tag{8.29}$$

By the above discussion, these fit into the so-called *fundamental exact sequence*

$$0 \to E(F)/mE(F) \to \mathrm{Sel}_m(E/F) \to Ш(E/F)[m] \to 0. \tag{8.30}$$

Theorem 8.5. *The m-Selmer group of E/F is finite.*

Proof. **Step 1.** For $\mathfrak{p} \nmid m$ of good reduction, we claim that

$$\mathrm{im}(\kappa_{\mathfrak{p}}) \subseteq \ker\left(H^1(F_{\mathfrak{p}}, E[m]) \to H^1(I_{\mathfrak{p}}, E[m])\right). \tag{8.31}$$

Indeed, we have seen that the reduction map $\widetilde{(\cdot)}\colon E(\bar{F}) \to \tilde{E}(\bar{k}_{\mathfrak{p}})$ induces an isomorphism $E(\bar{F}_v)[m] \xrightarrow{\sim} \tilde{E}(\bar{k}_{\mathfrak{p}})[m]$, and thus if $P \in E(F_{\mathfrak{p}})$ and $Q \in E(\bar{F}_{\mathfrak{p}})$ are such that $mQ = P$ and $\sigma \in I_{\mathfrak{p}}$, we have

$$\kappa_v\widetilde{(P)}(\sigma) = \sigma\widetilde{Q - Q} = \sigma\tilde{Q} - \tilde{Q} = 0, \tag{8.32}$$

which implies that $\kappa_v(P)(\sigma) \in E(\bar{F}_v)[m]$ is also trivial.

In other words, we have

$$\mathrm{Sel}_m(E/F) \subseteq \ker\left(H^1(F, E[m]) \to \prod_{\mathfrak{p}\nmid m \text{ good}} H^1(I_{\mathfrak{p}}, E[m])\right). \tag{8.33}$$

We will in fact prove that this right hand side it finite.

Step 2. We now reduce to the case where the action of $\mathrm{Gal}(\bar{F}/F)$ on $E(\bar{F})[m]$ is trivial. Let L/F be a finite Galois extension with $E[m] \subseteq E(L)$. Then we have an inflation-restriction exact sequence

$$0 \to H^1(L/F, E[m]) \to H^1(F, E[m]) \to H^1(L, E[m]). \tag{8.34}$$

Note that the restriction map also induces $\mathrm{Sel}_m(E/F) \to \mathrm{Sel}_m(E/L)$ since we have the commutative diagram for all $v' \mid v$

$$\begin{array}{ccc} E(F_v)/mE(F_v) & \xrightarrow{\kappa_v} & H^1(F_v, E[m]) \\ \downarrow & & \downarrow \\ E(L_{v'})/mE(L_{v'}) & \xrightarrow{\kappa_{v'}} & H^1(L_{v'}, E[m]) \end{array} \tag{8.35}$$

In particular, we have that

$$\ker\left(\mathrm{Sel}_m(E/F) \to \mathrm{Sel}_m(E/L)\right) \hookrightarrow H^1(L/F, E[m]). \tag{8.36}$$

This right hand side is a finite group, and thus we have the implication

$$\#\mathrm{Sel}_m(E/L) < \infty \implies \#\mathrm{Sel}_m(E/F) < \infty. \tag{8.37}$$

So we may assume without loss of generality that the $\mathrm{Gal}(\bar{F}/F)$-action on $E[m]$ is trivial.

Step 3. In the case $E(F)[m] = E[m]$, we have $H^1(F, E[m]) = \mathrm{Hom}(\mathrm{Gal}(\bar{F}/F), E[m])$. Combining everything, we have the diagram with exact rows

$$\begin{array}{ccccccc}
 & & \mathrm{Sel}_m(E/F) & & & & \\
 & & \downarrow & & & & \\
0 & \longrightarrow & \ker & \longrightarrow & H^1(F, E[m]) & \longrightarrow & \prod_{\substack{\mathfrak{p}\nmid m \\ \text{good}}} H^1(I_\mathfrak{p}, E[m]) \\
 & & \| & & \| & & \| \\
0 & \to & \mathrm{Hom}(\mathrm{Gal}(F^\Sigma/F), E[m]) & \to & \mathrm{Hom}(\mathrm{Gal}(\bar{F}/F), E[m]) & \to & \prod_{\substack{\mathfrak{p}\nmid m \\ \text{good}}} \mathrm{Hom}(I_\mathfrak{p}, E[m])
\end{array} \tag{8.38}$$

where $\Sigma = \{v \mid m\infty\} \cup \{\mathfrak{p} \text{ bad}\}$, and F^Σ is the maximal abelian extension of F unramified away from Σ. Since $E[m]$ has exponent m, we also have

$$\mathrm{Hom}(\mathrm{Gal}(F^\Sigma/F), E[m]) = \mathrm{Hom}(\mathrm{Gal}(F^{\Sigma,m}/F), E[m]) \tag{8.39}$$

where $F^{\Sigma,m}/F$ is the maximal subextension of F^Σ of exponent m. Now Class Field Theory[20] tells us that $F^{\Sigma,m}/F$ is a finite extension, and thus we conclude that $\mathrm{Sel}_m(E/F)$ is finite. □

Definition 8.4. The *ℓ-adic Selmer groups of E/F* are $\mathrm{Sel}_{\ell^\infty}(E/F) := \varinjlim_n \mathrm{Sel}_{\ell^n}(E/F)$ and $S_{\ell^\infty}(E/F) := \varprojlim_n \mathrm{Sel}_{\ell^n}(E/F)$.

These fit into the short exact sequences

$$0 \to E(F) \otimes \mathbb{Q}_\ell/\mathbb{Z}_\ell \to \mathrm{Sel}_{\ell^\infty}(E/F) \to Ш(E/F)[\ell^\infty] \to 0 \tag{8.40}$$

and

$$0 \to E(F) \otimes \mathbb{Z}_\ell \to S_{\ell^\infty}(E/F) \to \varprojlim_n Ш(E/F)[\ell^n] \to 0. \tag{8.41}$$

Conjecture 8.1 (Tate–Shafarevich Conjecture)**.** *The group $Ш(E/F)$ is finite.*

Assuming this conjecture, we for example have $E(F) \otimes \mathbb{Z}_\ell \xrightarrow{\sim} S_{\ell^\infty}(E/F)$, and so the Birch and Swinnerton-Dyer conjecture becomes equivalent to

[20] In fact this statement is much weaker than Class Field Theory, see [Sil09, Proposition VIII.1.6] for an elementary proof.

Conjecture 8.2 (Bloch–Kato conjecture for $T_\ell E$)**.** *We have*

$$\operatorname{rank}_{\mathbb{Z}_\ell} S_{\ell^\infty}(E/F) = \operatorname{ord}_{s=1} L(E/F, s). \tag{8.42}$$

It turns out that both terms can be defined purely in terms of $T_\ell E$: on the one hand, we have seen that[21]

$$L(E/F, s) = L((T_\ell E)^*, s) \tag{8.43}$$

and on the other hand we have that $S_{\ell^\infty}(E/F) = H^1_f(F, T_\ell E)$ is the *Bloch–Kato Selmer group of* $T_\ell E$. Analogous objects exist for *geometric*[22] Galois representations $\rho\colon \operatorname{Gal}(\bar{F}/F) \to \operatorname{GL}_n(\mathbb{Q}_\ell)$.

Conjecture 8.3 (Bloch–Kato conjecture)**.** *If ρ is a geometric Galois representation, we have*

$$\dim_{\mathbb{Q}_\ell} H^1_f(F, \rho) - \dim_{\mathbb{Q}_\ell} H^0(F, \rho) = \operatorname{ord}_{s=1} L(\rho^*, s) \tag{8.44}$$

where[23] $H^1(F, \rho) = \{c \in H^1(F, \rho)\colon \operatorname{loc}_v(c) \in H^1_f(F_v, \rho)$ *for all* $v\}$ *for*

$$H^1_f(F_v, \rho) = \begin{cases} \ker\left(H^1(F_v, \rho) \to H^1(I_v, \rho)\right) & \text{if } v \nmid \ell, \\ \ker\left(H^1(F_v, \rho) \to H^1(F_v, \rho \otimes \mathbb{B}_{crys})\right) & \text{if } v \mid \ell. \end{cases} \tag{8.45}$$

Remark 8.5. Similarly to the Birch–Swinnerton-Dyer conjecture, this L-function conjectured but not known to extend meromorphically. This means that the right hand side of this conjecture is not known to be well-defined in general.

8.4 Global heights

Given a projective variety $X/\bar{\mathbb{Q}}$ over a number field F, we want to construct a height function

$$h\colon X(\bar{F}) \to \mathbb{R} \tag{8.46}$$

that in some sense measure the arithmetic complexity of points of X.

Example 8.4. For $X = \mathbb{P}^1_{\mathbb{Q}}$, a point $X(\mathbb{Q}) = \mathbb{P}^1(\mathbb{Q})$ can be represented uniquely up to $\pm$ as $[a\colon b]$ with $a, b \in \mathbb{Z}$ relatively prime, and we define $h([a\colon b]) = \log \max(|a|, |b|)$.

[21] Technically speaking we haven't defined the Euler factors of the right hand side for primes above ℓ, but this can be done with ℓ-adic Hodge theory.

[22] This means that it is unramified at all but finitely many places, and that it is de Rham at places above ℓ.

[23] One needs to be more careful here about what one means by Galois cohomology here, since the coefficients have a nontrivial topology.

It will turn out that we will only define such functions up to a constant error, so we make the following definition.

Definition 8.5. For two functions $f, g\colon X(\bar{F}) \to \mathbb{R}$ we denote $f \sim g$ if the difference $f - g$ is bounded in absolute value.

Heights on projective spaces

For a number field F and a place v, we normalize the absolute value $|\cdot|_v$ as follows: if v is p-adic, then $|p|_v = p^{-1}$, if v is archimedean, then $|x|_v = |v(x)|$ for the usual complex absolute value $|\cdot|$. Note that this is so that if F'/F is an extension of number fields and $v' \mid v$, then $|x|_v = |x|_{v'}$ for $v \in F$.

Definition 8.6. We define a height function $h\colon \mathbb{P}^n(\bar{\mathbb{Q}}) \to \mathbb{R}$ as follows. For F a number field and $x_0, \dots, x_n \in F$, we define

$$h([x_0\colon \cdots\colon x_n]) = \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \cdot \log \max(|x_0|_v, \dots, |x_n|_v). \tag{8.47}$$

Remark 8.6. Note that this is well defined: if $\lambda \in F$, we have $h([x_0\colon \cdots\colon x_n]) = h([\lambda x_0\colon \cdots\colon \lambda x_n])$ because of the product formula

$$1 = \prod_v |\lambda|_v^{[F_v\colon \mathbb{Q}_v]}, \tag{8.48}$$

and we have that if F'/F is a finite extension, then $\frac{1}{[F'\colon \mathbb{Q}]} \sum_{v' \mid v} [F'_{v'}\colon \mathbb{Q}_v] = \frac{[F_v\colon \mathbb{Q}_v]}{[F\colon \mathbb{Q}]}$.

Remark 8.7. Note that for $P \in \mathbb{P}^1(\mathbb{Q})$, this recovers the definition from before: writing $P = [a\colon b]$ in lowest terms, we have $\max(|a|_p, |b|_p) = 1$ for all primes p.

Proposition 8.6. *The height function $h\colon \mathbb{P}^n(\bar{\mathbb{Q}}) \to \mathbb{R}$ lands in $\mathbb{R}_{\geqslant 0}$.*

Proof. We can always choose $[x_0\colon \cdots\colon x_n]$ such that one of the coordinates is 1. For that choice, each term in the summation above is non-negative. □

Proposition 8.7. *Let $f\colon \mathbb{P}^n \to \mathbb{P}^m$ be a morphism defined over $\bar{\mathbb{Q}}$. Then we have $h_{\mathbb{P}^m} \circ f \sim \deg(f) \cdot h_{\mathbb{P}^n}$.*

Proof. Denote $d = \deg f$, and write $f = (f_0\colon \cdots\colon f_m)$ for $f_0, \dots, f_m \in \bar{\mathbb{Q}}[X_0, \dots, X_m]$ homogeneous polynomials of degree d without common roots. For $P = [x_0\colon \cdots\colon x_n]$, denote $|P|_v = \max_i |x_i|_v$. Then we can easily bound

$$|f_i(P)|_v \leqslant C_{f,v} \cdot |P|_v^d \cdot \begin{cases} 1 & \text{if } v \text{ is nonarchimedean,} \\ \#\text{monomials} & \text{if } v \text{ is archimedean,} \end{cases} \tag{8.49}$$

where $C_{f,v}$ is the maximum of $|\cdot|_v$ over all coefficients of the f_i. Then for $C_f = \prod_v C_{f,v}^{[F_v\colon \mathbb{Q}_v]}$ and $C = C_f \cdot (\#\text{monomials})$, we have

$$\begin{aligned} h(f(P)) &= \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \log \max_i(|f_i(P)|_v) \\ &\leqslant \log(C) + \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \cdot d \cdot \log \max_i(|x_i|_v) \\ &= \log(C) + d \cdot h(P). \end{aligned} \tag{8.50}$$

For the opposite bound, we need to use Nullstelensatz. Since the f_i share no common root, we have $\operatorname{rad}(f_0, \dots, f_m) = (X_0, \dots, X_n)$. So there exists $e \in \mathbb{Z}$ and $g_{i,j} \in \bar{\mathbb{Q}}[X_0, \dots, X_n]$ such that $X_i^e = \sum_{j=1}^M g_{i,j} f_j$. Note $g_{i,j}$ are homogeneous of degree $e - d$. So

$$|P|_v^e \leqslant \max_{i,j}(|g_{i,j}(P)|_v) \cdot |f(P)|_v \cdot \begin{cases} 1 & \text{if } v \text{ is nonarchimedean,} \\ M & \text{if } v \text{ is archimedean,} \end{cases} \tag{8.51}$$

and since $g_{i,j}$ are homogeneous of degree $e - d$, we have

$$|g_{i,j}(P)|_v \leqslant C_{g,v} \cdot |P|_v^{e-d} \cdot \begin{cases} 1 & \text{if } v \text{ is nonarchimedean,} \\ \#\text{monomials} & \text{if } v \text{ is archimedean.} \end{cases} \tag{8.52}$$

Combining this and dividing by $|P|_v^{e-d}$, we conclude that $d \cdot h(P) \leqslant \log(C) + h(f(P))$ for some constant C similarly as above. □

Remark 8.8. Note that the to make the above implicit constant effective, one needs to effectively find the $g_{i,j}$ that are guaranteed to exist by Nullstelensatz.

Proposition 8.8. *For any $B, d > 0$, the set $\{P \in \mathbb{P}^n(\bar{\mathbb{Q}})\colon h(P) < B,\ [\mathbb{Q}(P)\colon \mathbb{Q}] \leqslant d\}$ is finite.*

Proof. For $P = [x_0\colon \cdots \colon x_n]$, we may assume without loss of generality that $x_0 \neq 0$, and then

$$h(P) = \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \cdot \log \max_{0 \leqslant i \leqslant n} |x_i|_v \geqslant \max_{1 \leqslant i \leqslant n} \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \cdot \log \max(|x_0|_v, |x_i|_v). \tag{8.53}$$

And so we are reduced to the case $n = 1$. That is, we need to see that the set

$$\{[1\colon x] \in \mathbb{P}^1(\bar{\mathbb{Q}})\colon h([1\colon x]) < B,\ [\mathbb{Q}(x)\colon \mathbb{Q}] = d\} \tag{8.54}$$

is finite. For such x, denote $F = \mathbb{Q}(x)$ and consider the minimal polynomial $T^d + a_1 T^{d-1} + \cdots + a_d \in \mathbb{Q}[T]$, with roots $x_1, x_2, \ldots, x_d \in \bar{\mathbb{Q}}$ with, say, $x_1 = x$. We have

$$|a_j|_v = \left| \pm \sum_{\substack{I \subseteq \{1,\ldots,d\} \\ \#I = j}} \prod_{i \in I} x_i \right|_v \leqslant \prod_{i=1}^{d} \max(1, |x_i|_v) \cdot \begin{cases} 1 & \text{if } v \text{ is nonarchimedean,} \\ \binom{d}{j} & \text{if } v \text{ is archimedean.} \end{cases} \tag{8.55}$$

and thus

$$\log \max_i(|a_i|_v) \leqslant \sum_{i=1}^{d} \log \max_i(1, |x_i|_v) + \begin{cases} 0 & \text{if } v \text{ is nonarchimedean,} \\ d \log(2) & \text{if } v \text{ is archimedean.} \end{cases} \tag{8.56}$$

Since x_i, x_j are all Galois conjugate, we have $h([1 \colon x_i]) = h([1 \colon x_j])$. Noting also that the right hand side is $\geqslant 0$, this implies that

$$h([1 \colon a_0 \colon \cdots \colon a_d]) = \frac{1}{[F \colon \mathbb{Q}]} \sum_v [F_v \colon \mathbb{Q}_v] \cdot \log \max(1, |a_0|_v, \ldots, |a_d|_v) \leqslant d \cdot \log(2) + d \cdot h([1 \colon x]). \tag{8.57}$$

Thus, the case $n = 1$ reduces to the case of $\mathbb{P}^d(\mathbb{Q})$, which is easy. □

Heights on projective varieties

Finally, we define heights for a general projective variety X.

Definition 8.7. Given a very ample divisor D on X, we may define a height function $h_D \colon X(\bar{F}) \to \mathbb{R}$, well-defined up to $\sim$, as follows: choosing sections generating $\mathcal{L}(D)$, we get an embedding $\phi_D \colon X \hookrightarrow \mathbb{P}^n$ and then we let $h_D = h_{\mathbb{P}^n} \circ \phi_D$. Note that different choices of bases of $\mathcal{L}(D)$ give embeddings, but they differ by an automorphism of $\mathbb{P}^n$, and so h_D is well defined up to $\sim$.

Theorem 8.9 (Weil height machine)**.** *There exists a unique collection of homomorphisms*

$$h_{(-)} \colon \mathrm{Pic}(X) \to \mathrm{Fun}(X(\bar{\mathbb{Q}}), \mathbb{R}) / \sim \tag{8.58}$$

for X projective varieties over $\bar{\mathbb{Q}}$, satisfying the following:

1. *If D is very ample, $h_D \sim h_{\mathbb{P}^n} \circ \phi_D$.*
2. *$h_{D+E} \sim h_D + h_E$.*

Moreover, this satisfies

3. *For $\phi \colon X \to Y$ a morphism, we have $h_{X, \phi^* D} \sim h_{Y,D} \circ \phi$.*

4. *(Northcott property) If D is ample, then for all $B, d > 0$, the set*

$$\{P \in X(\bar{F})\colon h_D(P) < B,\ [\mathbb{Q}(P)\colon \mathbb{Q}] \leqslant d\} \tag{8.59}$$

is finite.

5. *If D is effective, we have that h_D is bounded below outside of the base locus $b(|D|)$ of D.*

Proof sketch. The uniqueness is clear, since every divisor can be written as a difference of very ample divisors.

For existence, we need to check two things: i) if $D = \operatorname{div}(f)$ is principal, then $h_D \sim 0$, and ii) if D, E are very ample, then $h_{D+E} \sim h_D + h_E$.

For i), writing $D = D_1 - D_2$ as a difference of very ample divisors, we have $H^0(\mathcal{L}(D_2)) = f \cdot H^0(\mathcal{L}(D_1))$, and so if $\phi_{D_1}\colon X \to \mathbb{P}^n$ is $[g_0\colon \cdots\colon g_n]$, then we may choose $\phi_{D_2}\colon X \to \mathbb{P}^n$ to be $[fg_0\colon \cdots\colon fg_n]$ and

$$\max_i(|(fg_i)(P)|_v) = \max_i(|f(P)|_v|g_i(P)|_v) = |f(P)|_v \cdot \max_i(|g_i(P)|_v), \tag{8.60}$$

and thus $h_{\mathbb{P}^n}(\phi_{D_1}(P)) = h_{\mathbb{P}^n}(\phi_{D_2}(P))$ by the product formula.

For ii), one considers the Segre embedding $S\colon \mathbb{P}^N \times \mathbb{P}^M \to \mathbb{P}^{NM+N+M}$. This is such that we have the commutative diagram

$$\begin{array}{ccc} X & = & X \\ \downarrow{\scriptstyle \Delta} & & \downarrow{\scriptstyle \phi_{D_1+D_2}} \\ X \times X & & \mathbb{P}^n \\ \downarrow{\scriptstyle \phi_{D_1} \times \phi_{D_2}} & & \downarrow{\scriptstyle f} \\ \mathbb{P}^N \times \mathbb{P}^M & \xrightarrow{S} & \mathbb{P}^{NM+N+M} \end{array} \tag{8.61}$$

for f is a hyperplane section. This is because $S^*(\mathcal{O}(1)) = \mathcal{O}(1) \boxtimes \mathcal{O}(1)$, and thus

$$(S \circ (\phi_{D_1} \times \phi_{D_2}) \circ \Delta)^* \mathcal{O}(1) = \Delta^*(\phi_{D_1} \times \phi_{D_2})^*(\mathcal{O}(1) \boxtimes \mathcal{O}(1)) = \Delta^*(\mathcal{L}(D_1) \boxtimes \mathcal{L}(D_2)) = \mathcal{L}(D_1 + D_2). \tag{8.62}$$

Since f is a hyperplane section, we have $h_{\mathbb{P}^n} \circ \phi_{D_1+D_2} = h_{\mathbb{P}^{NM+N+M}} \circ f \circ \phi_{D_1+D_2}$
It remains to see that $h(S(x, y)) = h(x) + h(y)$, which follows from

$$\log \max_{\substack{0 \leqslant i \leqslant N \\ 0 \leqslant j \leqslant M}} (|x_i y_j|_v) = \log \left(\max_{0 \leqslant i \leqslant N} (|x_i|_v) \max_{0 \leqslant j \leqslant M} (|y_j|_v) \right) = \log \max_{0 \leqslant i \leqslant N} (|x_i|_v) + \log \max_{0 \leqslant j \leqslant M} (|y_j|_v) \tag{8.63}$$

By (2), both (3) and (4) reduce immediately to the case that D is very ample. In such case, they easily follow from (1).

For (5), let D be effective and write $D = D_1 - D_2$ as a difference of very ample divisors. Let $\phi_{D_2} = [f_0 \colon \cdots f_n] \colon X \to \mathbb{P}^n$. Since D is effective, we have $D_1 + (f_i) = D + D_2 + (f_i) \geqslant D_2 + (f_i) \geqslant 0$, and thus $f_i \in \mathcal{L}(D_1)$. This let us write $\phi_{D_1} = [f_0 \colon \cdots \colon f_m] \colon X \to \mathbb{P}^m$ for some $m \geqslant n$. All $f_0, \dots, f_m$ are regular outside the support of D_1, so for $x \notin \operatorname{supp}(D_1)$, this implies that

$$h_{\mathbb{P}^m}(\phi_{D_1}(x)) - h_{\mathbb{P}^n}(\phi_{D_2}(x)) \geqslant 0, \tag{8.64}$$

where the left hand side is $\sim h_D(x)$. By varying the decomposition $D = D_1 - D_2$, we can extend this to all $x \notin b(|D|)$. More precisely, we take $D = (D + E_i) - E_i$ where E_i are finitely many very ample divisors with $\bigcap_i E_i = \varnothing$. $\square$

Heights on abelian varieties

We will define the so-called *canonical heights* for abelian varieties, which are also called the *Néron–Tate heights*.

We will use a geometric input which we will only prove in the case of elliptic curves:

Proposition 8.10 (Theorem of the cube)**.** *Let k be a field and A/k an abelian variety. Consider D a divisor of A. If $\pi_1, \pi_2, \pi_3 \colon A^3 \to A$ are the three projections, then*

$$(\pi_1+\pi_2+\pi_3)^*D-(\pi_1+\pi_2)^*D-(\pi_1+\pi_3)^*D-(\pi_2+\pi_3)^*D+\pi_1^*D+\pi_2^*D+\pi_3^*D \tag{8.65}$$

is linearly equivalent to 0.

Proof in the case of elliptic curves. Denote $E = A$ for our elliptic curve, and by $C(D)$ the above expression.

If $P, Q \in E$, consider the map $i_{P,Q} \colon E \to E^3$ given by $R \mapsto (P, Q, R)$. Then

$$i_{P,Q}^* C(D) = \tau_{P+Q}^* D - \tau_P^* D - \tau_Q^* D + D. \tag{8.66}$$

We claim that this is principal. Indeed, if $D = R$ for some $R \in E$, we have

$$i_{P,Q}^* C(R) = (R - P - Q) - (R - P) - (R - Q) + R \sim 0. \tag{8.67}$$

Since $C(D)$ is linear in D, this implies the general case.

Now varying P, Q, this implies that we must have $C(D) \sim \pi_{12}^* D'$ for some $D' \in \operatorname{Div}(E^2)$, where $\pi_{12} \colon E^3 \to E^2$ is the projection the first two factors.

The same is true for π_{13} and π_{23}, and this implies that $C(D)$ is trivial. More precisely: for $i_{P,R} \colon E \to E^3$ given by $Q \mapsto (P, Q, R)$, we have

$$0 \sim i_{P,R}^* C(D) \sim (\pi_{12} \circ i_{P,R})^* D' = i_P^* D' \tag{8.68}$$

where $i_P\colon E \to E^2$ is $Q \mapsto (P, Q)$. Varying P, this implies that $C(D) \sim \pi_1^* D''$ for some $D'' \in \mathrm{Div}(E)$. Then pulling back by $i_{Q,R}\colon E \to E^3$ given by $P \mapsto (P, Q, R)$ we have

$$0 \sim i_{Q,R}^* C(D) \sim (\pi_1 \circ i_{Q,R})^* D'' = D''. \tag{8.69}$$

Thus $C(D) \sim 0$. □

Corollary 8.11. *In the same setting as above, we have that*

$$[m]^* D \sim \frac{m^2+m}{2} D + \frac{m^2-m}{2} [-1]^* D \tag{8.70}$$

If $\pi_1\colon \pi_2\colon E^2 \to E$ are the two projections, we also have that

$$(\pi_1 + \pi_2)^* D + (\pi_1 - \pi_2)^* D \sim 2 \cdot \pi_1^* D + \pi_2^* D + \pi_2^* [-1]^* D \tag{8.71}$$

Proof. For the first claim, pull back the proposition by $A \to A^3$ given by $(P) \mapsto (mP, P, -P)$ to get that

$$[m]^* D - [m+1]^* D - [m-1]^* D + [m]^* D + D + [-1]^* D \sim 0, \tag{8.72}$$

and use this to inductively prove the claim: taking $m = 1$ we get $[2]^* D \sim 3D + [-1]^* D$, and for the induction hypothesis we get

$$\begin{aligned} [m+1]^* D &\sim 2[m]^* D + D + [-1]^* D - [m-1]^* D \\ &\sim (m^2 + m + 1 - \frac{m^2-m}{2}) D + (m^2 - m + 1 - \frac{m^2 - 3m + 2}{2}) [-1]^* D \\ &= \frac{m^2+3m+2}{2} D + \frac{m^2+m}{2} [-1]^* D. \end{aligned} \tag{8.73}$$

For the second claim, simply pull back the above proposition by $A^2 \to A^3$ given by $(P, Q) \mapsto (P, Q, -Q)$. □

Theorem 8.12 (Canonical heights on abelian varieties)**.** *There exists a unique homomorphism $\hat{h}_{(\cdot)}\colon \mathrm{Pic}(A) \to \mathrm{Fun}(A(\bar{F}), \mathbb{R})$ satisfying the following:*

1. *We have $\hat{h}_D \sim h_D$.*
2. *$\hat{h}_{D+E} = \hat{h}_D + \hat{h}_E$.*

Moreover it satisfies

3. *If $f\colon A_1 \to A_2$ is a morphism of abelian varieties, then $\hat{h}_{f^* D} = \hat{h}_D \circ f$.*

4. *If D is ample, then for all $B, d > 0$ the set*

$$\{P \in A(\bar{F})\colon \hat{h}_D(P) < B,\ [\mathbb{Q}(P)\colon \mathbb{Q}] \leqslant d\} \tag{8.74}$$

is finite.

5. *If $[-1]^*D \sim D$ is ample, then $\hat{h}_D$ is non-negative, and $\hat{h}_D(P) = 0$ if and only if P is a torsion point.*
6. *If $[-1]^*D \sim D$, then $\langle P, Q\rangle_{NT,D} = \frac{1}{2}(\hat{h}_D(P+Q) - \hat{h}_D(P) - \hat{h}_D(Q))$ is bilinear, and we also have $\hat{h}_{NT,D}(P) = \langle P, P\rangle_{NT,D}$.*
7. *If $[-1]^*D \sim -D$, then $\hat{h}_D(P+Q) = \hat{h}_D(P) + \hat{h}_D(Q)$.*

Proof. We define $\hat{h}_D$ separately for $[-1]^*D \sim D$ and $[-1]^*D \sim -D$. In general we may write $2D = (D + [-1]^*D) + (D - [-1]^*D)$ and thus define

$$\hat{h}_{NT,D} = \frac{1}{2}\left(\hat{h}_{NT,D+[-1]^*D} + \hat{h}_{NT,D-[-1]^*D}\right). \tag{8.75}$$

Denote $e = 1$ resp. $e = 2$ if $[-1]^*D \sim -D$ resp. $[-1]^*D \sim D$. Then $[m]^*D \sim m^e D$. We define

$$\hat{h}_D(P) = \lim_{n\to\infty} \frac{h_D([2^n]P)}{2^{ne}}. \tag{8.76}$$

Note that this is forced from (1) and (2), and thus the uniqueness statement is clear. We need to see that this is well-defined: let C be a constant such that $|h_D([2]Q) - 2^e h_D(Q)| < C$ for all Q. Then for $n \geqslant m$,

$$\left|\frac{h_D([2^n]P)}{2^{ne}} - \frac{h_D([2^m]P)}{2^{me}}\right| \leqslant \sum_{i=m}^{n-1}\left|\frac{h_D([2^{i+1}]P)}{2^{(i+1)e}} - \frac{h_D([2^i]P)}{2^{ie}}\right| < \sum_{i=m}^{n-1}\frac{C}{2^{(i+1)e}} < \frac{C}{2^{me}} \tag{8.77}$$

so the sequence is Cauchy.

The above computation also implies that $|\hat{h}_D(P) - h_D(P)| < C$, and thus (1) follows. (2), (3) and (4) also easily follows from the corresponding claim for h_D.

For (5), if $D \sim [-1]^*D$ is ample, we may take some m such that mD is very ample. Then the positivity of the Weil height on projective spaces give us that $\hat{h}_{mD}$ is non-negative. Since

$$\hat{h}_D = \frac{1}{m}\hat{h}_{mD}, \tag{8.78}$$

the same is true for $\hat{h}_D$. Now if $\hat{h}_D(P) = 0$, we have

$$\hat{h}_D([m]P) = m^2\hat{h}_D(P) = 0, \tag{8.79}$$

and thus by the Northcott property the set $\{[m]P\colon m \in \mathbb{Z}\}$ is finite. This implies P is a torsion point.

For (6) and (7), the corollary above for A together with (3) for A^2 tell us that

$$\hat{h}_D(P+Q) + \hat{h}_D(P-Q) = 2 \cdot \hat{h}_D(P) + \hat{h}_D(Q) + \hat{h}_D(-Q). \tag{8.80}$$

But if $[-1]^*D \sim \pm D$, then $\hat{h}_D(-Q) = \pm \hat{h}_D(Q)$ by (3), and so both claims follow. Alternatively, one can avoid using (3) for A^2: we still have

$$h_D(P+Q) + h_D(P-Q) \sim 2 \cdot h_D(P) + h_D(Q) + h_D(-Q) \tag{8.81}$$

and $h_D(-Q) \sim \pm h_D(Q)$, and applying this for (mP, mQ) as we vary m also implies the claim. □

Remark 8.9. If $f\colon A_1 \to A_2$ is an arbitrary morphism, then we have

$$\hat{h}_{f^*D} = \hat{h}_D \circ f - \hat{h}_D(f(O_{A_1})). \tag{8.82}$$

By (3), it suffices to consider $f = t_x$ a translation morphism. By (1) we have that

$$\hat{h}_{t_x^*D} - \hat{h}_D \circ t_x \tag{8.83}$$

is bounded. The idea is that since this is a function of degree $\leqslant 2$, it must be constant. Concretely, consider separately the cases $[-1]^*D \sim (-1)^e D$ with $e \in \{1,2\}$, and then taking $P = mP'$ we have

$$(\hat{h}_{t_x^*D} - \hat{h}_D \circ t_x)(P) = m^2 \hat{h}_{t_x^*D} - m^2 \hat{h}_D(P') - \hat{h}_D(x) - 2m \cdot \langle P', x \rangle_D \tag{8.84}$$

and since this is bounded, it must be constant equal to $-\hat{h}_D$.

Lastly, we see that we have a completely canonical height pairing between A and its dual $A^\vee$.

Definition 8.8. Recall that the dual abelian variety $A^\vee$ satisfies $A^\vee(\bar{\mathbb{Q}}) = \mathrm{Pic}^0(A/\bar{\mathbb{Q}})$. We define

$$\langle \cdot, \cdot \rangle\colon A(\bar{\mathbb{Q}}) \times A^\vee(\bar{\mathbb{Q}}) \to \mathbb{R} \tag{8.85}$$

to be the pairing

$$\langle P, \mathcal{L} \rangle := \hat{h}_{\mathcal{L}}(P). \tag{8.86}$$

Proposition 8.13. *Let $\mathcal{P} \subseteq A \times A^\vee$ be the Poincaré bundle. Then $\langle P, \mathcal{L} \rangle = \hat{h}_{\mathcal{P}}(P, \mathcal{L})$. In particular, by taking duals, we also have $\langle P, \mathcal{L} \rangle = \hat{h}_P(\mathcal{L})$, and thus $\langle \cdot, \cdot \rangle$ is bilinear.*

Proof. For $\mathcal{L} \in A^\vee(\bar{\mathbb{Q}})$, consider $i\colon A \to A \times A^\vee$ given by $i(P) = (P, \mathcal{L})$. Then $i^*\mathcal{P} = \mathcal{L}$ by definition, and thus

$$\hat{h}_{\mathcal{P}}(P, \mathcal{L}) = \hat{h}_{\mathcal{L}}(P) - \hat{h}_{\mathcal{P}}(0, \mathcal{L}). \tag{8.87}$$

Similarly,

$$\hat{h}_{\mathcal{P}}(P, \mathcal{L}) = \hat{h}_P(\mathcal{L}) - \hat{h}_{\mathcal{P}}(P, \mathcal{O}_A), \tag{8.88}$$

which specializes to $\hat{h}_{\mathcal{P}}(0, \mathcal{L}) = 0$ when $P = 0$. □

Corollary 8.14. *If $\mathcal{L}$ is a line bundle on A and $\phi_{\mathcal{L}}\colon A \to A^\vee$ is the corresponding morphism $\phi_{\mathcal{L}}(P) = t_P^*\mathcal{L} \otimes \mathcal{L}^{-1}$, then*

$$\langle P, \phi_{\mathcal{L}}(Q)\rangle = \hat{h}_{\mathcal{L}}(P + Q) - \hat{h}_{\mathcal{L}}(P) - \hat{h}_{\mathcal{L}}(Q). \tag{8.89}$$

In particular, if $[-1]^*\mathcal{L} \simeq \mathcal{L}$, then $\langle P, \phi_{\mathcal{L}}(Q)\rangle = 2\langle P, Q\rangle_{\mathcal{L}}$.

Proof sketch. This follows from $(\mathrm{id} \times \phi_{\mathcal{L}})^*\mathcal{P} \simeq m^*\mathcal{L} \otimes \mathrm{pr}_1^*\mathcal{L}^{-1} \otimes \mathrm{pr}_2^*\mathcal{L}^{-1}$ as line bundles on $A \times A$. We will not prove this, although in the case of elliptic curves it can be easily proven in the same way we proved the theorem of the cube. □

Example 8.5. If $A = E$ is an elliptic curve with Weierstraß form $E \subseteq \mathbb{P}^2$, then we get the principal polarization $\phi_{\mathcal{O}(1)}\colon E \simeq E^\vee$ given by $P \mapsto (-P) - (O)$. Note that this is the negative of the "usual" one!

Example 8.6. If C is a smooth connected curve of genus $g \geqslant 1$, consider $A = \mathrm{Alb}(C)$ its Albanese and $J = \mathrm{Jac}(C)$ its Jacobian. They are canonically duals $A^\vee = J$, and also canonically isomorphic, where $J \xrightarrow{\sim} A = J^\vee$ corresponds to ϕ_Θ where $\Theta = i(C) + \cdots i(C)$ is the theta divisor. Here $i\colon C \to J$ is the Abel–Jacobi map, and $i(C)$ is being summed $g - 1$ times. This gives a canonical height pairing

$$\langle \cdot, \cdot \rangle_C\colon C(\bar{\mathbb{Q}}) \times C(\bar{\mathbb{Q}}) \xrightarrow{i \times i} J(\bar{\mathbb{Q}}) \times J(\bar{\mathbb{Q}}) \xrightarrow{\mathrm{id} \times \phi_\Theta} J(\bar{\mathbb{Q}}) \times J^\vee(\bar{\mathbb{Q}}) \xrightarrow{\langle \cdot, \cdot \rangle} \mathbb{R} \tag{8.90}$$

$$\langle \cdot, \cdot \rangle_{2\Theta}$$

8.5 Local heights

The discussion in this section will be a bit informal, as it is very technical. The goal is to discuss the relationship between the Néron–Tate canonical heights and the *Beilinson–Bloch height pairings*. Roughly, the canonical height pairing

$$\langle \cdot, \cdot \rangle\colon A(F) \times A^\vee(F) \to \mathbb{R} \tag{8.91}$$

can be decomposed into local height pairings

$$\langle \cdot, \cdot \rangle = \sum_v \langle \cdot, \cdot \rangle_v \tag{8.92}$$

where $\langle \cdot, \cdot \rangle_v$ are given by *arithmetic intersection numbers* on a suitable model $\mathcal{A}/\mathcal{O}_{F_v}$ of A.

Throughout, we will denote F a number field, X, Y will denote smooth projective varieties over F, and A will denote abelian varieties over F.

We continue the discussion of the previous section. Recall that we defined the height function on $\mathbb{P}^n$ by

$$h_{\mathbb{P}^n}(x) = \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \log \max_i (|x_i|_v) \tag{8.93}$$

where $x = [x_0\colon \cdots \colon x_n]$ with $x_i \in F$ where F is a number field. We would like to decompose $h_{\mathbb{P}^n}$ into a sum of local heights. Note that each term $\log \max_i(|x_i|_v)$ is not well-defined, the sum is only well-defined because of the product formula. In order to fix this, we can do the following: For a divisor D of degree d, say given by the equation $P \in \bar{\mathbb{Q}}[X_0, \ldots, X_n]$, we may define the local heights $\lambda_{\mathbb{P}^n, D, v}\colon \mathbb{P}^n(F_v) \backslash \mathrm{supp}(D) \to \mathbb{R}$ by

$$\lambda_{\mathbb{P}^n, D, v}(x) = \log \max \left(\frac{|x_0|_v^d}{|P(x)|_v}, \ldots, \frac{|x_n|_v^d}{|P(x)|_v} \right). \tag{8.94}$$

Note that this is only well-defined up to a constant, since we can have different choices of F for the same D. This depends on the choice of divisor D, not just on its rational equivalence class: $D + \mathrm{div}(f)$ is given by the equation Pf, and so

$$\lambda_{\mathbb{P}^n, D+\mathrm{div}(f), v} = \lambda_{\mathbb{P}^n, D, v} - \log |f|_v. \tag{8.95}$$

Nevertheless, by the product formula we have

$$\frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \cdot \lambda_{\mathbb{P}^n, D, v}(x) = d \cdot h_{\mathbb{P}^n}(x) = h_{\mathbb{P}^n, D}(x), \quad \text{for } x \in \mathbb{P}^n(F) \backslash \mathrm{supp}(D). \tag{8.96}$$

Similarly as in the discussion in the previous section, one can prove the following.

Definition 8.9. For X defined over a number field F, we denote

$$X_D := \bigsqcup_v X(F_v) \backslash \mathrm{supp}(D). \tag{8.97}$$

For two functions $\lambda_1, \lambda_2\colon X_D \to \mathbb{R}$, we denote $\lambda_1 \sim \lambda_2$ if their difference is zero for all but finitely many v, and bounded otherwise. We denote by $\lambda_1 \equiv \lambda_2$ if their difference is zero for all but finitely many v, and constant otherwise.

Theorem 8.15 (Local Weil height machine)**.** *There exists a unique collection of functions*

$$\lambda_{(-)}\colon \mathrm{Div}(X/F) \to \mathrm{Fun}(X_{(-)}, \mathbb{R})/\sim \tag{8.98}$$

as X/F runs through projective varieties such that

1. *For $X = \mathbb{P}^n$, this is as defined above.*
2. *$\lambda_D + \lambda_E \sim \lambda_{D+E}$.*
3. *If $f\colon X \to Y$ is a suitable*[24] *morphism, $\lambda_{f^*D} \sim \lambda_D \circ f$.*
4. *If $D = \mathrm{div}(f)$, then $\lambda_D \sim \{-\log|f|_v\}_v$.*

Moreover, we have that

$$h_D \sim \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \cdot \lambda_{D,v} \tag{8.99}$$

in $\mathrm{Fun}(X(F)\backslash \mathrm{supp}(D), \mathbb{R})/\sim$.

Theorem 8.16 (Canonical local heights)**.** *There exists a unique collection of functions*

$$\hat{\lambda}_{(-)}\colon \mathrm{Div}(A/F) \to \mathrm{Fun}(A_{(-)}, \mathbb{R})/\equiv \tag{8.100}$$

as A/F varies through abelian varieties such that

1. *$\hat{\lambda}_D \sim \lambda_D$.*
2. *$\hat{\lambda}_D + \hat{\lambda}_E \equiv \hat{\lambda}_{D+E}$.*
3. *If $f\colon A_1 \to A_2$ is a suitable morphism of abelian varieties, $\hat{\lambda}_{f^*D} \equiv \hat{\lambda}_D \circ f$.*
4. *If $D = \mathrm{div}(f)$, then $\hat{\lambda}_D \equiv \{-\log|f|_v\}_v$.*

Moreover, we have that

$$\hat{h}_D \equiv \frac{1}{[F\colon \mathbb{Q}]} \sum_v [F_v\colon \mathbb{Q}_v] \cdot \hat{\lambda}_{D,v} \tag{8.101}$$

in $\mathrm{Fun}(A(F)\backslash \mathrm{supp}(D), \mathbb{R})/\equiv$.

Remark 8.10. One can also remove the $\equiv$ ambiguity on the definition of the local heights by requiring that

$$\lim_{N\to\infty} N^{-2g} \sum_{\substack{P\in A[N] \\ P\notin \mathrm{supp}(D)}} \hat{\lambda}_{D,v}(P) = 0. \tag{8.102}$$

This removes the ambiguity in (2), but not in (3) and in the comparison with $\hat{h}_D$.

[24] You cannot always pull back divisors, for example if $f(X) \subseteq D$. Here we cannot pull back the line bundle associated to D since λ_D depends on D, not just on its linear equivalence class.

Denote $Z(A/F)$ the free group on $A(F)$ and by $Z^0(A/F)$ the subgroup of degree 0 elements. Another way to remove the ambiguity above is to consider the pairing

$$\langle \cdot, \cdot \rangle_v \colon Z^0(A/F_v) \times \mathrm{Div}(A/F_v) \dashrightarrow \mathbb{R} \tag{8.103}$$

where $\langle \sum_i n_i \cdot a_i, D \rangle := n_i \cdot \hat{\lambda}_{D,v}(a_i)$. This is such that

1. $\langle a, \mathrm{div}(f) \rangle_v = -\log |f(a)|_v$ (for $a = \sum n_i a_i$ we denote $f(a) = \prod f(a_i)^{n_i}$),
2. If $f \colon A_1 \to A_2$ is a suitable morphism, $\langle f_* a, b \rangle_v = \langle a, f^* b \rangle_v$.

In fact, this pairing can be reinterpreted in terms of *Beilinson–Bloch heights*. Let X/F be smooth projective. We consider $v \nmid \infty$, although there is a parallel discussion for the case $v \mid \infty$. We may choose a regular proper model $\mathcal{X}/\mathcal{O}_{F_v}$. If $a \in Z^0(X/F_v)$, $b \in \mathrm{Div}(X/F_v)$, we can consider their Zariski closures $\bar{a} \in Z^0(\mathcal{X}/\mathcal{O}_{F_v})$, $\bar{b} \in \mathrm{Div}(\mathcal{X}/\mathcal{O}_{F_v})$. Note that $\bar{a}$ has dimension 1, as $\dim \mathrm{Spec}(\mathcal{O}_{F_v}) = 1$, and $\bar{b}$ has codimension 1. If A, B meet transversally, then we consider

$$[a, b]_v := -(\bar{a} \cdot \bar{b}) \cdot \log q_v \tag{8.104}$$

where $(\bar{a} \cdot \bar{b})$ denotes the (arithmetic) intersection number of $\bar{a}, \bar{b}$.

Example 8.7. These arithmetic intersection numbers are quite concrete. For example if X/F is a curve with good reduction at v, and $a, b \in X(F)$ are distinct points, then $(A \cdot B)$ is the largest $n \in \mathbb{Z}_{\geqslant 0}$ such that "$a \equiv b \mod \mathfrak{m}_v^n$." Let's make this precise: Say $X = \mathrm{Proj}(\mathcal{O}_{F_v}[x_0, \dots, x_n]/I)$, and assume $\mathcal{X} := \mathrm{Proj}(\mathcal{O}_{F_v}[x_0, \dots, x_n]/\mathcal{I})$ is smooth, where $\mathcal{I} = I \cap \mathcal{O}_{F_v}[x_0, \dots, x_n]$. Then we can write the special fiber as $\mathcal{X}_{k_v} = \mathrm{Proj}(k_v[x_0, \dots, x_n]/\tilde{I})$ where $\tilde{I} = \mathcal{I} \mod \mathfrak{m}_v$. For a point $a \in X$ we can take coordinates $a = (a_0 \colon \cdots \colon a_n)$ where $a_i \in \mathcal{O}_{F_v}$ are not all in $\mathfrak{m}_v$, and similarly for $b = (b_0 \colon \cdots \colon b_n)$. Then $\tilde{a} := \bar{a}_{k_v} = (\tilde{a_0} \colon \cdots \tilde{a_n}) \in \mathcal{X}_{k_v}$. Then $[a, b]_v = 0$ unless $\tilde{a} = \tilde{b}$. In such case, without loss of generality we assume $a_0, b_0 = 1$. Considering the affine chart $Y \subset X$ with $x_0 \neq 0$, we reduce to the affine case, say $Y = \mathrm{Spec}(F[x_1, \dots, x_n]/J)$ with $\mathcal{Y} = \mathrm{Spec}(\mathcal{O}_{F_v}[x_1, \dots, x_n]/\mathcal{J})$ smooth. Then $\bar{a}$ is given by the ideal $(x_1 - a_1, \dots, x_n - a_n)$, and similarly for $\bar{b}$. Thus the intersection number is

$$\begin{aligned}
&\mathrm{length}\left(\mathcal{O}_{F_v}[x_1, \dots, x_n]_{(\bar{a})}/(\mathcal{J}, x_1 - a_1, x_1 - b_1, \dots, x_n - a_n, x_n - b_n)\right) \\
=&\mathrm{length}\left(\mathcal{O}_{F_v}/(a_1 - b_1, \dots, a_n - b_n)\right) \\
=&\min_i \nu_v(a_i - b_i) \\
=&\max_{n \geqslant 0}(a_i \equiv b_i \mod \mathfrak{m}_v^n \text{ for all } i).
\end{aligned} \tag{8.105}$$

Theorem 8.17 (Beilinson–Bloch height pairing)**.** *Under certain (conjectural*[25]*) assumptions on X, the above construction can be extended to pairings*

$$[\cdot,\cdot]_v\colon Z^0(X/F_v)\times \operatorname{Div}(X/F_v) \dashrightarrow \begin{cases} \mathbb{Q}\cdot\log q_v & \text{if } v\nmid\infty,\\ \mathbb{R} & \text{if } v\mid\infty. \end{cases} \tag{8.106}$$

These satisfy

1. $[a,\operatorname{div}(f)]_v = [F_v\colon\mathbb{Q}_v]\cdot\log|f(a)|_v$,
2. *If $f\colon X\to Y$ is a suitable morphism, then $[a,f^*b]_v=[f_*a,b]_v$,*
3. *If X has good reduction at v, then $[\cdot,\cdot]_v$ takes values in $\mathbb{Z}\cdot\log q_v$.*

They allow us to define

$$[\cdot,\cdot]\colon Z^0(X/F)\times\operatorname{Pic}(X/F)\to\mathbb{R} \tag{8.107}$$

by $[\cdot,\cdot]=\frac{1}{[F\colon\mathbb{Q}]}\sum_v[\cdot,\cdot]_v$, where to make sense of the right hand side we choose representatives with disjoint support.[26]

Theorem 8.18. *If A is an abelian variety, then $[F_v\colon\mathbb{Q}_v]\cdot\langle\cdot,\cdot\rangle_v=-[\cdot,\cdot]_v$ in $Z^0(A/F_v)\times\operatorname{Pic}^0(A/F_v)$. In particular, $\langle\cdot,\cdot\rangle=-[\cdot,\cdot]$ in $Z^0(A/F)\times\operatorname{Pic}^0(A/F)$.*

Sketch of proof. Since $[\cdot,\operatorname{div}(f)]_v$ and $\langle\cdot,\operatorname{div}(f)\rangle_v$ are similar, the pairing $\{\cdot,\cdot\}_v:=[\cdot,\cdot]_v+[F_v\colon\mathbb{Q}_v]\cdot\langle\cdot,\cdot\rangle_v$ descends to

$$\{\cdot,\cdot\}_v\colon Z^0(A/F_v)\times\operatorname{Pic}(A/F_v)\to\mathbb{R}. \tag{8.108}$$

Moreover, both $[\cdot,\cdot]_v$ and $\langle\cdot,\cdot\rangle_v$ are v-adically continuous, and thus so is $\{\cdot,\cdot\}_v$. So, for example, if $a\in Z^0(A/F_v)$ and we consider

$$\{a,\cdot\}_v\colon\operatorname{Pic}^0(A/F_v)=A^\vee(F_v)\to\mathbb{R}, \tag{8.109}$$

then this is a continuous group homomorphism. The source is compact and $\mathbb{R}$ has no nontrivial compact subgroups, hence this is trivial. □

Remark 8.11. In fact, if A has good reduction at v and the local canonical heights $\hat\lambda_b$ are normalized as in the remark above, then we even have

$$[F_v\colon\mathbb{Q}_v]\cdot\langle a,b\rangle_v=(\bar a,\bar b)\cdot\log q_v \tag{8.110}$$

for $a\in A(F)$ and $b\in\operatorname{Div}(A/F)$. More generally, $\hat\lambda_b$ can be computed in terms of intersection numbers in the Néron model of A.

[25] These are theorems if X is a curve or an abelian variety, for example.

[26] In fact, there are also such conjectural pairings for cycles in other dimensions $A^p(X)\times A^q(X)\to\mathbb{R}$ where $p+q+1=\dim X$ and $A^p(X)\subseteq\operatorname{CH}^p(X)$ is the subset of ℓ-adic étale cohomologically trivial classes. Conjecturally, $A^p(X)$ are independent of ℓ.

Example 8.8. Let C/F be a smooth connected curve, and denote $C \xrightarrow{i} A = \mathrm{Alb}(C)$ its Albanese variety and $J = \mathrm{Jac}(C)$ its Jacobian. They are canonically isomorphic via $\phi_\Theta \colon J \xrightarrow{\sim} J^\vee = A$ where Θ is the theta divisor as in Example 8.6. Consider $S^\vee \colon \mathrm{Pic}^0(C/F) \xrightarrow{\sim} J(F)$ the inverse of $J(F) = \mathrm{Pic}^0(A/F) \xrightarrow{j^*} \mathrm{Pic}^0(C/F)$. Consider also $S \colon Z^0(X/F) \xrightarrow{j_*} Z^0(A/F) \xrightarrow{\Sigma} A(F)$ where Σ is summing with the group structure on $A(F)$. Then we have a commutative diagram

$$\begin{array}{ccccc}
 & & Z^0(C/F) \times \mathrm{Pic}^0(C/F) & \xrightarrow{-[\cdot,\cdot]} & \mathbb{R} \\
 & & \downarrow{\scriptstyle j_* \times S^\vee} & & \| \\
 & & Z^0(A/F) \times \mathrm{Pic}^0(A/F) & \xrightarrow{-[\cdot,\cdot]} & \mathbb{R} \\
 & & \| & & \| \\
S^\vee \times S & & Z^0(A/F) \times \mathrm{Pic}^0(A/F) & \xrightarrow{\langle\cdot,\cdot\rangle} & \mathbb{R} \\
 & & \downarrow{\scriptstyle \Sigma \times \mathrm{id}} & & \| \\
 & \searrow & A(F) \times J(F) & \xrightarrow{\langle\cdot,\cdot\rangle} & \mathbb{R} \\
 & & \downarrow{\scriptstyle \phi_\Theta^{-1} \times \mathrm{id}} & & \| \\
 & & J(F) \times J(F) & \xrightarrow{\langle\cdot,\cdot\rangle_{2\Theta}} & \mathbb{R}
\end{array} \tag{8.111}$$

where the first square is because $[j_* a, b] = [a, j^* b]$, the third is because any $D \in \mathrm{Pic}^0(A/F)$ is antisymmetric and so $\hat{h}_D \colon A(F) \to \mathbb{R}$ is linear, the fourth because of Example 8.6. Noting that

$$\begin{array}{ccccc}
Z^0(C/F) & \hookrightarrow & \mathrm{Div}^0(C/F) & \twoheadrightarrow & \mathrm{Pic}^0(C/F) \\
\downarrow{\scriptstyle S} & & & & \downarrow{\scriptstyle S^\vee} \\
A(F) & & \xrightarrow{\qquad\qquad \phi_\Theta \qquad\qquad} & & J(F)
\end{array} \tag{8.112}$$

is commutative, we conclude that for all $x, y \in Z^0(C/F)$ we have

$$-[x, y] = \langle S(x), S^\vee(y) \rangle = \langle S(x), S(y) \rangle_{2\Theta}. \tag{8.113}$$

Local heights of curves

Let's fully describe $[\cdot,\cdot]_v$ in the case of a smooth connected curve C/F_v. We choose $\mathcal{C}/\mathcal{O}_{F_v}$ a regular proper model. Its special fiber $\mathcal{C}_{k_v}$ may be singular, say with irreducible components $F_1, \ldots, F_n$. Then as a divisor on $\mathcal{C}$, we have

$$\mathcal{C}_{k_v} = \sum_{i=1}^{n} c_i \cdot F_i \tag{8.114}$$

for some $c_i \in \mathbb{Z}_{\geqslant 0}$.

Proposition 8.19. *Let $V = \mathbb{Q}^n = \operatorname{span}(e_1, \dots, e_n)$ and equip it with a bilinear pairing given by $\langle e_i, e_j \rangle := (c_i F_i \cdot c_j F_j)$. Then this is negative semidefinite with 1-dimensional kernel spanned by $c := e_1 + \cdots + e_n$. In particular, the image of*

$$\begin{aligned} V &\longrightarrow V^* \\ v &\longmapsto \langle \cdot, v \rangle \end{aligned} \tag{8.115}$$

is precisely $\{f \in V^* \colon f(c) = 0\}$.

Proof. Since $\mathcal{C}_{k_v} = \operatorname{div}(\varpi_v)$ is principal, we have $\langle c, e_i \rangle = 0$ for all i. Since F_i, F_j intersect transversely for $i \neq j$, we have $\langle e_i, e_j \rangle \geqslant 0$ for $i \neq j$.

Thus the matrix $A = (a_{i,j})_{i,j}$ of the pairing $\langle \cdot, \cdot \rangle$ symmetric, nonnegative off-diagonal and has zero sum of rows and columns. Then for any $v = \sum_i v_i \cdot e_i$ we have

$$\langle v, v \rangle = \sum_{i,j} v_i v_j a_{i,j} = \sum_{i,j} \left(v_i v_j a_{i,j} - \frac{v_i^2 a_{i,j} + v_j^2 a_{i,j}}{2} \right) = \sum_{i,j} \frac{(v_i - v_j)^2}{2} a_{i,j} \leqslant 0, \tag{8.116}$$

with equality if and only if

$$a_{i,j} \neq 0 \implies v_i = v_j. \tag{8.117}$$

It's a fact that for $\mathcal{C}$ as above, $\mathcal{C}_{k_v}$ is connected. Thus the claim follows. □

Corollary 8.20. *If $a \in Z^0(C/F_v)$, there is $A \in Z^0(\mathcal{C}/\mathcal{O}_{F_v}) \otimes \mathbb{Q}$ extending a with $(A, F_i) = 0$ for all i. Such A is unique up to multiples of $\mathcal{C}_{k_v}$.*

Proof. If $\bar{a} \in Z^0(\mathcal{C}/\mathcal{O}_{F_v})$ is the Zariski closure of a, then

$$(\bar{a} \cdot \mathcal{C}_{k_v}) = \deg(\mathcal{O}_{\mathcal{C}_{k_v}}(\bar{a}_{k_v})) = \deg(\mathcal{O}_{\mathcal{C}_{F_v}}(a)) = \deg(a) = 0 \tag{8.118}$$

where the second equality is relying on the fact that $\mathcal{C}$ is flat. By the above proposition, this means that there is $v \in V$ with $\langle v, e_i \rangle = (\bar{a} \cdot F_i)$ for all i. Thus $A = \bar{a} - \sum_{i=1}^n v_i c_i \cdot F_i$ is an extension of a with $(A \cdot F_i) = 0$ for all i. □

Definition 8.10. If $a \in Z^0(C/F_v)$ and $b \in \operatorname{Div}^0(C/F_v)$ have disjoint supports, we define

$$[a, b]_v = -(A \cdot \bar{b}) \cdot \log q_v \tag{8.119}$$

where $A \in Z^0(\mathcal{C}/\mathcal{O}_{F_v})$ is an extension as in the above corollary.

Part II
Spring 2025

9 Complex multiplication

References: [Ser67a] is the nicest, but leaving a lot to the reader. Also [Mil05, § 11], [Sil94, § II], [Cox22, § 11].

9.1 Quadratic orders and their class groups

Recall that an elliptic curve E (over $\mathbb{C}$ or $\overline{\mathbb{Q}}$) has as endomorphism ring either $\mathbb{Z}$ or an order $R' \subset R$ in an imaginary quadratic field K (where we denote by R the ring of integers of K). Assume the latter.

Proposition 9.1. *There is a (unique) positive integer f (the* conductor *of R') such that $R' = R_f := \mathbb{Z} + f\mathbb{Z}$. Moreover, every finite R'-module $\Lambda \subset K$ with $\mathrm{End}(\Lambda) = R_f$ is invertible (= locally free of rank one), hence projective.*

Proof. The first statement follows simply from the facts that $R/\mathbb{Z} \simeq \mathbb{Z}$ and $\mathbb{Z} \subset R_f$.

Given Λ as in the proposition, let $\Lambda^\vee$ be the dual lattice under the trace pairing; it is straightforward to verify that its endomorphism ring is also R_f. Moreover, $R_f^\vee = g'(f\tau)^{-1}R_f$, where τ is a generator of R over $\mathbb{Z}$ and g is its monic irreducible polynomial, and it is straightforward to verify that

$$\Lambda \cdot g'(f\tau)\Lambda^\vee = R_f.$$

This means that, as abstract modules,

$$\Lambda \otimes \left(g'(f\tau)\Lambda^\vee\right) = R_f,$$

and the existence of such an inverse is equivalent to Λ being locally free of rank one. (See [Eis95, Theorem 11.6], [Sta25, 0B8M].)

□

This identifies the set of isomorphism classes of elliptic curves with CM by $R' = R_f$ with the Picard group of invertible R'-modules, $\mathrm{Pic}(R')$, also called the *ring class group* of R'.

Let $K \supset \mathbb{Q}$ be a quadratic imaginary extension, $R' = R_f$ an order in K as above, and M an invertible R'-module. Let $T = \mathrm{GL}_{R'}(M) \subset \mathrm{GL}_{\mathbb{Z}}(M)$; it is a linear group over $\mathbb{Z}$ with canonical identifications $T(\mathbb{Q}) = K$ and $T(\mathbb{Z}) = R'$. Moreover, it is easily seen to be independent of the choice of M.

Proposition 9.2. *There is a canonical identification of* $\mathrm{Pic}(R_f)$ *with the quotient* $\{$*Fractional Ideals of* R *prime to* $f\}/\{$*Principal ideals of* R *generated by elements of* $F^\times$ *which are congruent to an element of* $(\mathbb{Z}/f)^\times$ *modulo* $f.\}$

The identification is characterized by sending an actual ideal $I \subset R$ *to the class in* $\mathrm{Pic}(R_f)$ *generated by* $I \cap R_f$.

The same quotient can be written as (here the index f *means "finite adeles")*

$$K^\times \backslash \mathbb{A}_{K,f}^\times / \prod_p {R'_p}^\times = T(\mathbb{Q}) \backslash T(\mathbb{A}_{\mathbb{Q},f}) / T(\widehat{\mathbb{Z}}),$$

and ${R'_p}^\times = \{r \in R_p^\times | r \mod f \in (\mathbb{Z}_p/f)^\times\}$.

By "congruent to an element of $(\mathbb{Z}/f)^\times$ modulo f" we mean the subgroup of $F^\times$ generated by elements of R that are prime to f, and have image in $(\mathbb{Z}/f)^\times \subset (R/fR)^\times$.

Proof. Let us start from the last, adelic statement, which is quite general.[27] Given an invertible R'-module M, choose a trivialization over an open subset U of $\mathrm{Spec}\mathbb{Z}$, and trivializations over the formal neighborhood of the complement (a finite set S of primes). The quotients of the trivializations over the punctured formal neighborhoods give rise to an element of $T(\mathbb{Q}_S)$. Changing the trivializations, this element is modified by the action of $T(U) = ({R'}^S)^\times$ (where the exponent S here denotes the S-units), respectively by $T(\mathbb{Z}_S) = (R'_S)^\times$ (where the index S means product of completions at places in S). Taking the colimit over S gives a map $\mathrm{Pic}(R') \to T(\mathbb{Q}) \backslash T(\mathbb{A}_{\mathbb{Q},f}) / T(\widehat{\mathbb{Z}})$, which is clearly a homomorphism. To prove that it is an isomorphism, we will use the following fact:

For every prime p *of* $\mathbb{Z}$*, the category of free* $R' \otimes \mathbb{Z}_{(p)}$*-modules is equivalent (via the obvious functor) to the category consisting of triples* (M_K, M_p, τ)*, where* M_K*,* M_p *are free modules for* K *and* $R' \otimes \mathbb{Z}_p$*, and* τ *is an isomorphism of the corresponding* $R' \otimes \mathbb{Q}_p$*-modules.*

This is obvious for the underlying $\mathbb{Z}_{(p)}$-modules, and automatically extends to the $R' \otimes \mathbb{Z}_{(p)}$-structure.

The bijectivity of the map above now follows: To prove injectivity, suppose that two invertible R'-modules M, M' give rise to the same class in the double quotient, then there is an open U as above and trivializations over U and the formal neighborhoods of the complements so that the transition maps coincide. To show that the modules are isomorphic, it suffices to that the trivializations over U, which give rise to an isomorphism $M_U \xrightarrow{\sim} M'_U$, extend to

[27] It generalizes to a theorem of Beauville–Laszlo and its variants, see https://mathoverflow.net/q/112593.

isomorphisms of M and M' over Zariski open neighborhoods of the points over the complement of U. But this is equivalent to showing that they extend to isomorphisms between $M \otimes \mathbb{Z}_{(p)}$ and $M' \otimes \mathbb{Z}_{(p)}$, for all p in the complement of U, and the aforementioned equivalence of categories ensures that. Similarly for surjectivity.

The calculation of ${R'_p}^\times$ is straightforward from the presentation $R' = \mathbb{Z} + fR$.

The translation of this isomorphism in terms of ideals is achieved first by noticing that an element of $T(\mathbb{A}_{\mathbb{Q},f})$ can be acted up on by an element of $K^\times = T(\mathbb{Q})$ so that its components at places in S belong to $T(\mathbb{Z}_S) = (R'_S)^\times$, and that element of $T(\mathbb{Q})$ is determined modulo elements whose S-localization belongs to $(R'_S)^\times$. □

Hence, $\mathrm{Pic}(R')$ is identified with the quotient of the finite idele class group of K by the open compact subgroup $\widehat{R'}^\times$. In particular, it is finite. The corresponding class number is usually denoted by $h(R')$.

9.2 CM elliptic curves

By "an elliptic curve E with CM by R'" we will mean an elliptic curve E over some (implicit) algebraically closed field in characteristic zero, together with an isomorphism $\mathrm{End}(E) \simeq R'$. However, we will consider the set $Y_{R'}$ of isomorphism classes of elliptic curves with CM by R' *without taking into account this isomorphism*, i.e., $Y_{R'}$ denotes isomorphism classes of elliptic curves whose endomorphism ring has an isomorphism with R' (which is not fixed). Fixing the isomorphism $\mathrm{End}(E) \simeq R'$ is equivalent to fixing the embedding $K \hookrightarrow k$ (the field), via $R' \simeq \mathrm{End}(E) \to \mathrm{End}(\mathrm{Lie}(E)) = k$.

For an elliptic curve E with CM by R', and an invertible R'-module M, define

$$M \star E := \mathrm{Hom}_{R'}(M, E),$$

i.e., the functor that associates to a scheme S over k the group $\mathrm{Hom}_{R'}(M, E(S))$. Explicitly, if ${R'}^m \to (R'^n) \to M$ is a presentation of M, $M \star E$ is the kernel of $E^n \to E^m$. This is naturally an elliptic curve with CM by R'. Hence, we get an action of $\mathrm{Pic}(R')$ on the set $Y_{R'}$ of isomorphism classes of elliptic curves with CM by R'.

Exercise: Prove that, over $\mathbb{C}$, if $E(\mathbb{C}) \simeq \mathbb{C}/\Lambda$ then $(M \star E)(\mathbb{C}) \simeq \mathbb{C}/M^* \cdot \Lambda$, where M^* is a representative for the inverse to M in the Picard group, and $\cdot$ denotes multiplication in the Picard group.

Remark 9.1. Note that the nontrivial conjugation of K over $\mathbb{Q}$ acts on R', and composing with it we get a different action of R', which modifies the action of $\mathrm{Pic}(R')$ compatibly with complex conjugation on $\mathrm{Pic}(R')$.

Proposition 9.3. *The set $Y_{R'}$ is a torsor for $\mathrm{Pic}(R')$ under the above action.*

(We could also consider both sets as groupoids, and the statement remains true.)

Proof. We will first prove it with $\mathbb{C}$-coefficients. The corollary that follows will then say that all such curves are defined over $\overline{\mathbb{Q}}$, i.e., the corresponding points on the coarse moduli space $Y(1)$ are algebraic, and then it follows for every algebraically closed field in characteristic zero.

Such an elliptic curve over $\mathbb{C}$ has $\mathbb{C}$-points $E(\mathbb{C}) = \mathbb{C}/\Lambda$, where Λ is up to homothety equal to an invertible fractional ideal M of R' with respect to an embedding $K \hookrightarrow \mathbb{C}$. In particular, fixing this embedding, we have $E \simeq M^* \star E_0$, where $E_0 = \mathbb{C}/R'$ and M^* is a dual ideal, hence all such elliptic curves are in the orbit of E_0 under the action of $\mathrm{Pic}(R')$; moreover, it is clear that different elements of $\mathrm{Pic}(R')$ map E_0 to nonisomorphic elliptic curves over $\mathbb{C}$. □

Corollary 9.4. *The set $Y_{R'}$ of isomorphism classes of elliptic curves over $\mathbb{C}$ with CM by R' is finite; hence, the automorphism group of $\mathbb{C}$ over $\mathbb{Q}$ acts on them through a finite quotient, and all such curves are defined over an algebraic extension of $\mathbb{Q}$.*

Remark 9.2 (**Important remark!**). Here and later, when we say "defined over a number field L," we will mean that there is a model for the elliptic curve over L. However, remember that elliptic curves have nontrivial automorphism groups, and that, given such a model, there are infinitely many models ("forms") of the elliptic curve over L, parametrized by $H^1(L, \mathrm{Aut}(E))$. [Standard fact of Galois descent.] We will sometimes implicitly fix such a model in arguments, and leave it to the reader to check that the conclusions do not depend on the choice.

9.3 The theorem(s) of complex multiplication

For what follows, we imagine the set $Y_{R'}$ to be defined in terms of isomorphism classes over a fixed algebraically closed field in characteristic zero, and let $\overline{\mathbb{Q}}$ be the algebraic closure of $\mathbb{Q}$ in that field.

We now know that we have an actions of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and $\mathrm{Pic}(R')$ on the set $Y_{R'}$, although the action of the latter depends on the choice of an embedding

$K \hookrightarrow \overline{\mathbb{Q}}$. It is clear from the definitions that these two actions combine to an action of the semidirect product

$$\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \ltimes \mathrm{Pic}(R').$$

Moreover, since $Y_{R'}$ is a torsor for $\mathrm{Pic}(R')$, it follows that the restriction of the action to $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ is given by a character

$$\eta : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{Pic}(R').$$

Theorem 9.5. *The actions of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ and $\mathrm{Pic}(R')$ on $Y_{R'}$ are compatible under the reciprocity map normalized to send a prime to its arithmetic Frobenius element, i.e., η is induced by the inverse of the reciprocity map.*

Corollary 9.6. *For $\Lambda \subset \mathbb{C}$ a lattice with CM by R', the j-invariant $j(\Lambda)$ is an algebraic number, defined over the ring class field H_f of R'.*

For any ideal I of R prime to f, with Artin symbol $\sigma \in \mathrm{Gal}(H_f/K)$, and any fractional ideal Λ of R_f as above (i.e., $\mathrm{End}(\Lambda) = R_f$), we have

$$j(I\Lambda)^\sigma = j(\Lambda). \tag{9.1}$$

Moreover, $K(j(\Lambda)) = H_f$ and $[\mathbb{Q}(j(\Lambda)) : \mathbb{Q}] = [K(j(\Lambda)) : K] = h(R_f)$.

Proof. Theorem 9.5 implies that the Galois action factors through the ring class field of R', hence $j(\Lambda) \in H_f$, and that it is not fixed by any element of the Galois group of H_f over K, hence $K(j(\Lambda)) = H_f$. For the degree of $\mathbb{Q}(j(\Lambda))$ over $\mathbb{Q}$ it suffices to observe that the $\mathbb{Q}$-Galois orbit of the point $[\Lambda]$ on the modular curve is the same set $Y_{R'}$. □

The j-invariants of CM elliptic curves are called *singular moduli*.

Remark 9.3. It is actually the case that $j(\Lambda)$ is an *algebraic integer* in H_f. We may cover this later in the course.

Level structures

There is a strengthening of Theorem 9.5 to full level structures. First, note the following.

Lemma 9.7. *For any elliptic curve E with CM by R', and any prime p, the p-adic Tate module T_pE is a free rank one R'_p-module.*

Proof. This is very easy to see if we base change from the field of definition of E (an abelian extension of K) to $\mathbb{C}$. □

Fix an embedding $K \hookrightarrow k$ (our algebraically closed field in characteristic zero), and let $Y_{R'}(\infty)$ denote the set of isomorphism classes of elliptic curves with CM by R' equipped with full level structure, i.e., R'_p-equivariant isomorphisms $T_pE \simeq R'_p$ for all p. (This would not make sense without first fixing the embedding of K, since the level structure requires fixing the R'-action.) We can repeat the steps above to define an action of the idele class group on $Y_{R'}(\infty)$, as follows: As is clear from the proof of Proposition 9.2, the quotient $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ can be identified with the set of isomorphism classes of invertible R'-modules M equipped with trivializations of their completions at all primes p, $M_p := M \otimes_{\mathbb{Z}} \mathbb{Z}_p \simeq R'_p$.

We now calculate the Tate modules for the action of $\mathrm{Pic}(R')$ on $Y_{R'}$:

$$T_p(M\star E) = \lim_n \mathrm{Hom}_{R'}(M, E)[p^n] = \lim_n \mathrm{Hom}_{R'}(M/p^n, E[p^n]) = \mathrm{Hom}_{R'_p}(M_p, T_pE).$$

In particular, a trivialization of M_p and a trivialization of T_pE give rise to a trivialization of $T_p(M\star E)$, hence we get an action of $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ on $\mathrm{Pic}_{R'}(\infty)$, which is easily seen to be simply transitive.

We also have an action of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ on this set, commuting with the action of $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$, hence given by a character, that we will denote by the same letter as before,

$$\eta : \mathrm{Gal}(\overline{\mathbb{Q}}/K)^{\mathrm{ab}} \to T(\mathbb{Q})\backslash T(\mathbb{A}_f) = K^\times\backslash \mathbb{A}_{K,f}^\times.$$

The strengthening of Theorem 9.5 is the following.

Theorem 9.8. *The actions of* $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ *and* $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ *on* $Y_{R'}(\infty)$ *are compatible under the reciprocity map normalized as in Theorem 9.5, i.e.,* η *is the inverse of the reciprocity homomorphism.*

Remark 9.4. The set $Y_{R'}(\infty)$ is the image of a map of Shimura varieties. Namely, let $\mathcal{H}^\pm = \mathbb{C} \smallsetminus \mathbb{R}$, identified with the hermitian symmetric domain associated to the group $G = \mathrm{GL}_2$, by identifying the point i with the homomorphism $h_i : U(1) \to \mathrm{GL}_2$ sending $\exp(2\pi i\theta)$ to $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. Let $\tau \in \mathcal{H}^\pm$ so that the field $K = \mathbb{Q}(\tau)$ is quadratic. Note that by this presentation we've fixed an embedding $K \hookrightarrow \mathbb{C}$, as well as a lattice $\Lambda = \langle 1, \tau\rangle$, so that the elliptic curve $\mathbb{C}/\Lambda$ has CM by an order R' in K. Let T = the stabilizer of h_i, which is a $\mathbb{Q}$-subtorus of G. It is isomorphic to the restriction of scalars of $\mathbb{G}_m$ from K to $\mathbb{Q}$, with an integral model, inherited from G, that can be checked to coincide with the group of R'-automorphisms of the lattice Λ.

The sign is a bit arbitrary. Does it match any other choice?

The point τ is called a *special/CM point*, and the pair (T, τ) is called a *special/CM pair* in (G, X) (where $X = \mathcal{H}^\pm$ the symmetric space of G). It

Check sign of $\pm\tau$.

defines an embedding from the Shimura variety of T to that of G,

$$S_T = T(\mathbb{Q})\backslash T(\mathbb{A}_f) \to S_G = G(\mathbb{Q})\backslash (G(\mathbb{A}_f) \times X),$$

sending a class $[a]$ to the class $[\tau, a]$. Otherwise said, if we choose τ as a base point to present X as a quotient of $G(\mathbb{R})$, and hence S_G as a quotient of $[G] = G(\mathbb{Q})\backslash G(\mathbb{A})$, then the map $S_T \to S_G$ is induced from the map $[T] \to [G]$; however, this is not the best way to view this, since it really depends on the choice of special point.

See [Mil05, § 12].

Remark 9.5. If we fix a model for a representative of $Y_{R'}$ over an abelian extension L/K, from the action on the Tate modules we get a homomorphism

$$\mathrm{Gal}(\overline{\mathbb{Q}}/L) \to \prod_p {R'_p}^\times,$$

which, in general, *depends on the model*. (Why? What happens when we twist by a Galois 1-cocycle into the automorphism group?)

This is slightly finer information than the restriction of the map $\mathrm{Gal}(\overline{\mathbb{Q}}/K) \to T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ to $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$. Of course, the (finite) unit group ${R'}^\times$ induces an automorphism of the elliptic curve and the trivialization of its Tate module, and hence acts trivially on the set $Y_{R'}(\infty)$ of isomorphism classes of such data. But for a fixed elliptic curve defined over E, the Galois action gives a lift of this homomorphism, restricted to $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$, to $\prod_p {R'_p}^\times$ (which depends on the choice of E). In particular, the torsion points of the elliptic curve are not necessarily defined over the maximal abelian extension of K, but over that of L. We can, however, obtain points defined over the maximal abelian extension of K by considering the image of those torsion points in the quotient of the elliptic curve by its automorphism group, the group of units ${R'}^\times$; the quotient $E/{R'}^\times$ is always isomorphic to $\mathbb{P}^1$.

9.4 Proof of the theorems of complex multiplication

Here is the crux of the whole story, where we establish the compatibility of a "Hecke" action (that of idele class groups) with a Galois action. As is often the case, and as we saw in the Hecke–Galois correspondence for elliptic curves (Corollary 6.16, which relied on the Eichler–Shimura relation), the argument will boil down to reductions modulo primes.

It is enough to prove the analog of Theorem 9.8 for $Y_{R'}(N)$ for every $N \geqslant 1$, where $Y_{R'}(N)$ is defined in an analogous way, in terms of level structures modulo N. (This includes Theorem 9.5, when $N = 1$.) We have simultaneous,

commuting actions of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$ and $K^\times\backslash\mathbb{A}_{K,f}^\times$ on this finite set, and the action of the latter factors through some finite quotient $K^\times\backslash\mathbb{A}_{K,f}^\times/K_N$ (which can be made explicit, but we won't bother to). We can choose a finite abelian extension L of K such that

1. the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$ factors through $\mathrm{Gal}(L/K)$, i.e., one (equivalently all) representative E and its N-torsion points are defined over L;
2. the inverse reciprocity map r_K^{-1}, composed with projection to the finite quotient $K^\times\backslash\mathbb{A}_{K,f}^\times/K_N$, factors through $\mathrm{Gal}(L/K)$.

Hence, we have two homomorphisms $\eta, r_K^{-1} : \mathrm{Gal}(L/K) \to K^\times\backslash\mathbb{A}_{K,f}^\times/K_N$, and we want to show that they coincide.

Consider the set of primes $\mathfrak{p}$ of K such that

1. $\mathfrak{p}$ is prime to N and the conductor f, and unramified in L;
2. $\mathfrak{p}$ is of degree 1 over $\mathbb{Q}$, i.e., $N(\mathfrak{p}) = p$ for some rational prime p;
3. E has good reduction at one (hence all) the primes $\mathfrak{B}$ of L over $\mathfrak{p}$.

For such primes, the embedding $\mathfrak{p} \subset R'$ induces an isogeny $\alpha : E \to \mathfrak{p} \star E$, which gives rise to an isomorphism of N-torsion points. (Here, by abuse of notation, we denote $\mathfrak{p} \cap R'$ simply by $\mathfrak{p}$, i.e., we think of a prime ideal of R that is prime to f as a prime ideal of R'.) We will prove the following.

> There is an isomorphism
>
> $$\mathfrak{p} \star \bar{E} \simeq \bar{E}^{(p)}, \tag{9.2}$$
>
> where $\bar{E}$ denotes the reduction of E modulo $\mathfrak{B}$ and the exponent denotes the p-power Frobenius twist of E, which, composed with the reduction of the isogeny α, is equal to the p-Frobenius map $\bar{E} \to \bar{E}^{(p)}$.

Indeed, since $\mathfrak{p}$ is of degree 1, the degree of the isogeny is $N(\mathfrak{p}) = p$. Moreover, it is purely inseparable, as can be seen from calculating the map on tangent spaces: one can easily show that $\mathrm{Lie}(M \star E) = \mathrm{Hom}_{R'}(M, \mathrm{Lie}(E))$ for every invertible module M, and from this one sees that $\mathrm{Lie}(E) \to \mathrm{Lie}(\mathfrak{p} \star E)$ reduces to the zero map modulo $\mathfrak{B}$, since $\mathfrak{B}$ divides $\mathfrak{p}$. Such an isogeny has to coincide with the composition of an isomorphism $\bar{E}^{(p)} \simeq \mathfrak{p} \star \bar{E}$ with the Frobenius map.

This is enough to prove the theorem at level 1: Frobenius elements associated to primes $\mathfrak{p}$ as above are dense in the Galois group $\mathrm{Gal}(\bar{K}/K)$ (by Chebotarev's density theorem, regarding the degree-1 condition) and therefore each $\sigma \in \mathrm{Gal}(L/K)$ can be represented as $[\mathrm{Fr}_\mathfrak{p}]$ for infinitely many such

$\mathfrak{p}$'s. Choose an invertible module M representing the class of $r_K^{-1}(\sigma)$ in $\mathrm{Pic}(R')$. The result above shows that the reductions of ${}^\sigma E$ and $M \star E$ at infinitely many primes $\mathfrak{B}$ are isomorphic, hence ${}^\sigma E$ and $M \star E$ have to be isomorphic over $\bar{\mathbb{Q}}$. (At the level of j-invariants, this says that if the difference of the j-invariants of ${}^\sigma E$ and $M \star E$ cannot be divisible by infinitely many primes, unless it is zero – note that this makes sense even if we don't know integrality of the j-invariants, by only reducing modulo primes where they are integral.) This proves Theorem 9.5.

To extend this result to N-level structures, let us again take $\sigma \in \mathrm{Gal}(L/K)$ and M as above, and fix an isomorphism $\beta : {}^\sigma E \xrightarrow{\sim} M \star E$, as above. By enlarging L, if necessary, we may assume that the isomorphism is defined over L. Choose a prime $\mathfrak{p}$ as above (after enlarging L), again with $r_K^{-1}(\pi_\mathfrak{p}) = \mathrm{Fr}_\mathfrak{p} = \sigma$ in $\mathrm{Gal}(L/K)$, and identify M with $\mathfrak{p}$ (that is, with $\mathfrak{p} \cap R'$) as an R'-module, to consider β as an isomorphism $\beta : {}^\sigma E \to \mathfrak{p} \star E$. Those choices are unique up to $\mathrm{Aut}(E) = {R'}^\times$.

The action of σ on N-torsion points gives rise to an isomorphism

$$E[N] \xrightarrow{\sim} {}^\sigma E[N], \tag{9.3}$$

On the other hand, we will see that the action of $K_\mathfrak{p}^\times \subset \mathbb{A}_{K,f}^\times$ on $Y_{R'}(N)$ is unramified, with the action of a uniformizer

$$[\pi_\mathfrak{p}] : E[N] \to \mathfrak{p} \star E[N] \tag{9.4}$$

being equivalent to the isomorphism between $E[N]$ and $\mathfrak{p} \star E[N]$ induced by the isogeny α as above.

Assuming this for now, we need to prove that the two isomorphisms (9.3) and (9.4) are compatible with β *at the level of isomorphism classes*, i.e., up to an automorphism of E.

Reducing mod $\mathfrak{B}$, and remembering, Proposition 7.8, that reduction does nothing to N-torsion, since N is prime to p, (9.3) reduces to the isomorphism induced by the Frobenius map $\bar{E} \to \bar{E}^{(p)}$. We have already seen that (9.4) also reduces to the isomorphism induced by the Frobenius map $\bar{E} \to \bar{E}^{(p)}$, for *some* isomorphism $\bar{\beta}' : \bar{E}^{(p)} \xrightarrow{\sim} \mathfrak{p} \star \bar{E}$. Hence, we need to show that $\bar{\beta}'$ is equal to the reduction $\bar{\beta}$ of β, up to an automorphism of E. Since $\bar{\beta}$ and $\bar{\beta}'$ commute with the action of R', they differ by an element of $\mathrm{Aut}_{R'}(\bar{E})$. We claim that this is equal to ${R'}^\times$ and therefore such an automorphism lifts to characteristic zero. Indeed, $\mathrm{End}_{R'}(\bar{E}) \simeq K$, therefore, $\mathrm{End}_{R'}(\bar{E})$ is an order between R' and R. We claim that it is equal to R' Since p is prime to f, this order is already maximal when localized at p. At other primes, since for $(m, p) = 1$ the reduction map is an isomorphism on m-torsion, we must have

$$\mathrm{End}_{R'}(\bar{E})/m \simeq \mathrm{End}_{R'/m}(\bar{E}[m]) = \mathrm{End}_{R'/m}(E[m]) = R'/m.$$

Therefore, $\mathrm{End}_{R'}(\bar{E}) = R'$, and we are done.

We have not explained yet why the action of $[\pi_{\mathfrak{p}}]$ on $Y(N)$ is the one induced from the isogeny α above – we need to unwrap the definition of this action. This requires going back to the proof of Proposition 9.2, where to identify the class of the idele $(1, 1, \dots, \pi_{\mathfrak{p}}, 1, \dots)$ with the ideal $\mathfrak{p}$ we can choose the trivialization away from $\mathfrak{p}$ induced by the embedding $\mathfrak{p} \subset \mathfrak{o}$, and the trivialization in a formal neighborhood of $\mathfrak{p}$ induced by multiplication by $\pi_{\mathfrak{p}}$: $\mathfrak{o}_{\mathfrak{p}} \xrightarrow{\sim} \mathfrak{p}_{\mathfrak{p}}$. Since N is prime to $\mathfrak{p}$, action of this idele on N-level structures will be the one induced by the embedding $\mathfrak{p} \hookrightarrow \mathfrak{o}$.

9.5 Applications and examples

[Stub, but see Chao Li's minor thesis for a many nice examples.]

9.6 Extension: Abelian varieties and Shimura varieties

[Stub; see [Mil05, § 12].]

The theory of complex multiplication extends to abelian varieties and their moduli spaces.

The general theory of Shimura varieties is founded on the replication of CM theory: While for moduli of abelian varieties CM theory is a theorem (reconciling the rational structure on the moduli space with class field theory), for general Shimura varieties the restriction of Galois action on special pairs is a *postulate* which characterizes the "canonical models" of those.

9.7 The Galois representation

Let E be an elliptic curve defined over a number field L. For any prime L, the l-adic Tate module $T_l E$ gives rise to a representation of $G_L := \mathrm{Gal}(\bar{\mathbb{Q}}/L)$ on a 2-dimensional $\mathbb{Z}_l$-space, whose determinant is the l-adic cyclotomic character (by Theorem 7.12).

If E has CM by an order $\mathfrak{o}$ in a quadratic field K, then $G_{KL} := \mathrm{Gal}(\bar{\mathbb{Q}}/KL)$ acts through a character $G_{KL} \to \mathfrak{o}_l^\times \subset \mathrm{GL}_{\mathbb{Z}_l}(T_l E)$. It follows from Theorem 9.8 that the collection of those characters, up to a global unit in $\mathfrak{o}$ (i.e., as homomorphisms into $(\prod_l \mathfrak{o}_l^\times)/\mathfrak{o}^\times$), are given by the norm character of class field theory (see Remark 9.5).

If L does not contain K (but we may and will assume that KL is abelian over K), the Galois group $\mathrm{Gal}(K/\mathbb{Q}) = \mathrm{Gal}(LK/L)$ acts nontrivially on G_L^{ab}, and the representation on the l-adic Tate module is *dihedral*, i.e., has non-abelian image in the normalizer of a torus in $\mathrm{GL}_2(\mathbb{Z}_l)$.

10 Lubin–Tate theory

References: [Ser67b, § 3], [Neu99, § V.4–5].

10.1 Introduction: Complex multiplication, locally

Consider first the identification of the maximal abelian extension of $\mathbb{Q}$ with the cyclotomic field $\mathbb{Q}(\zeta_\infty)$. If we are only interested in the maximal abelian extension of the local field $\mathbb{Q}_p$, we can obtain it by attaching cyclotomic units – that is, torsion points in the multiplicative group $\mathbb{G}_m$ – to $\mathbb{Q}_p$. However, with prime-to-p torsion we only get unramified extensions. If we'd like to obtain ramified extensions, it is enough to look at p-power torsion. Now, one thing to observe is that, if we consider the (standard, smooth) *integral* model of the group $\mathbb{G}_m$ over $\mathbb{Z}$ or $\mathbb{Z}_p$, the p-power torsion elements are all congruent to 1 mod p, i.e., they live in the kernel of the reduction map $\mathbb{G}_m(\mathbb{Z}_p) \to \mathbb{G}_m(\mathbb{F}_p)$.

The local story of complex multiplication is the same: If K_v is a p-adic completion of a global field K, we can get all abelian extensions of K_v by localizing abelian extensions of K. (Why?) To fix ideas, let us assume that v is a place that splits completely in the Hilbert class field $\tilde{K}$ of K, so that K_v is also a localization of $\tilde{K}$ (in a way that we fix, $K_v = \tilde{K}_{\mathfrak{B}}$), and let us take a CM elliptic curve E with CM by the maximal order of K_v, hence defined over $\tilde{K}$. (If we are starting with a local field K_v, we can always globalize it this way to K – why?) Then all torsion points of E are algebraic over K_v, and generate its maximal abelian extension. Again, the prime-to-p torsion points don't go very far in terms of ramification – they only generate a field of finite ramification index over K_v, and if E has good reduction at $\mathfrak{B}$, they are unramified. Let us also assume that E has good reduction at $\mathfrak{B}$. Then, all ramification will be obtained from the p-power torsion of E, i.e., from its p-adic Tate module. Moreover, the following proposition shows that these points are in the residual neighborhood of the identity.

Proposition 10.1. *If E is an elliptic curve with CM by an order in an imaginary quadratic field K, and has good reduction at a prime $\mathfrak{B}$ over p, then the reduction $\bar{E}$ is supersingular if and only if there is a unique prime in K over p (i.e., p is inert or ramified in K).*

Proof. The assumptions and conclusion are isogeny invariant (e.g., good reduction is equivalent to the l-adic Tate module being unramified by the Néron–Ogg–Shafarevich Theorem 7.10, and that is an isogeny invariant), so we may without loss of generality assume that E has CM by the maximal order $R \subset K$. We fix this CM structure $\theta : R \xrightarrow{\sim} \mathrm{End}(E)$, which, remember, is equivalent to

fixing an embedding of K into the algebraically closed field (say k) of definition of E. Let $\mathfrak{B} \cap R = \mathfrak{p}$.

If p splits in K, i.e., $p = \mathfrak{p}\mathfrak{p}'$, we will show that there is a separable endomorphism of $\bar{E}$ of degree a (nonzero) power of p; it will then follow that $\bar{E}[p] \neq 0$, i.e., $\bar{E}$ is ordinary. Choose a positive integer m such that $\mathfrak{p}^m$ is principal, $\mathfrak{p}^m = \mu R$, and let μ' be the Galois conjugate of μ, so that $\mathfrak{p}'^m = \mu' R$. Then $\mu\mu' = N_{\mathbb{Q}}^K \mu = p^m$, hence the endomorphism $\theta(\mu')$ (and its reduction mod $\mathfrak{B}$) has order p^m. On the other hand, it acts on the tangent space (Lie algebra) by μ', which is $\neq 0$ modulo $\mathfrak{B}$; therefore, it is separable modulo $\mathfrak{B}$. (Note that $\mathfrak{B}$, here, can be taken to be a prime of the ring of a large enough finite extension L of K, so that $\theta(\mu')$ is defined over L.)

If, on the other hand, p is inert or ramified in K, so that there is a unique prime $\mathfrak{p}$ of R over it, then the Frobenius endomorphism π of $\bar{E}$ (that is, the q-Frobenius, if $R_L/\mathfrak{B} = \mathbb{F}_q$, with q a power of p) is contained in $\bar{\theta}(R) \subset \mathrm{End}(\bar{E})$. (If it were not, the endomorphism ring would be larger and the curve would be supersingular, so we would be done, but actually the claim is true in the supersingular case, as well, because π commutes with $\bar{\theta}(R)$, hence has to lie in it.) Assume $\pi = \bar{\theta}(\mu)$, then the dual isogeny π' is given by $\theta(\mu')$, where μ' is the Galois conjugate of μ. But the degree of π is q^2 (a power of p), hence $\mu\mu' = q^2$, which means that they both have to lie in the prime $\mathfrak{p}$. But then they only differ by a unit, $\mu' = \mu \cdot u$, and $\pi' = \pi \cdot \bar{\theta}(u)$, with $\bar{\theta}(u)$ an automorphism. Hence, π' is purely inseparable, which is equivalent to the curve being supersingular.

[See [Lan87, Theorem 13.4.12], which also includes a calculation of the endomorphism ring in the split case.] □

10.2 Formal $\mathfrak{o}$-modules

Lubin–Tate theory provides a complete generalization of these two examples, which produces a maximal abelian totally ramified extension for *any* p-adic field. Recall that 1-dimensional formal groups (understood to be abelian) over any ring A were defined in Definition 7.1.

We are now interested in the case $A =$ an $\mathfrak{o}$-algebra, where $\mathfrak{o}$ is the valuation ring of a p-adic field K. A *formal $\mathfrak{o}$-module* over A (understood again to be 1-dimensional) is a formal group law $\mathfrak{G}$ together with a homomorphism of algebras, $\mathfrak{o} \to \mathrm{End}\,\mathfrak{G}$ which induces the given homomorphism $\iota : \mathfrak{o} \to A$ at the level of Lie algebras. "Endomorphisms," here, are taken in the category of "formal Lie varieties," which is the analog of Lie algebras for formal neighborhoods (rather than just tangent spaces); see [Wei]. Explicitly, this means that the action of each element $a \in \mathfrak{o}$ is represented by a power series $[a] \in A[[X]]$, and the fact that at the level of Lie algebras the action is the one given by ι

means that this power series is of the form $[a] = \iota(a)X + O(X^2)$.

We now consider the case $A = \mathfrak{o}$. Fix a uniformizer $\pi \in \mathfrak{p} \subset \mathfrak{o}$. A *Lubin–Tate module* over $\mathfrak{o}$ for the prime element π is a formal $\mathfrak{o}$-module F over $\mathfrak{o}$ such that $[\pi]_F$ reduces to the Frobenius modulo $\mathfrak{p}$, that is, if $\mathfrak{o}/\mathfrak{p} = \mathbb{F}_q$, then $[\pi]_F \equiv X^q \mod \mathfrak{p}$.

Theorem 10.2 (Lubin–Tate)**.** *There is a unique, up to isomorphism, Lubin–Tate module for each prime element π. Moreover, a coordinate can be chosen so that $[\pi]_F$ is any given power series which is $\equiv \pi X$ modulo higher-order terms and $\equiv X^q$ modulo $\mathfrak{p}$.*

Proof. A lot of calculation, but quite straightforward; see [LT65] or [Neu99, Theorem 4.6]. □

10.3 Explicit local class field theory

Let F be a formal $\mathfrak{o}$-module over $\mathfrak{o}$. We can consider the $\mathfrak{p}^n$-torsion points $F[\mathfrak{p}^n]$ (in the prime ideal of an algebraic closure of K), and, if we choose a uniformizer π, organize them in an inverse system, the π-adic Tate module

$$T_\pi F = \lim F[\mathfrak{p}^n]$$

(inverse limit under multiplication by π).

Proposition 10.3. *If F_π is a Lubin–Tate module (for some uniformizer π), then, for every n, $F_\pi[\mathfrak{p}^n]$ is a free $\mathfrak{o}/\mathfrak{p}^n$-module of rank 1. Hence, $T_\pi F_\pi$ is a free $\mathfrak{o}$-module of rank 1.*

Remark 10.1. In particular, $F_\pi[p^n]$ is a free $\mathbb{Z}/p^n$-module of rank $h = [K : \mathbb{Q}_p]$. This is the *height* of the formal group, and its associated p-divisible group, forgetting the $\mathfrak{o}$-module structure, see § 10.6 below. Note that the assumption that F is a Lubin–Tate module is necessary; the additive formal group $F(x, y) = x + y$ with $\mathfrak{o}$-module structure $[z] = z$ has no torsion points.

Proof. Since there is a unique LT module up to isomorphism, it suffices to calculate explicitly with the unique formal $\mathfrak{o}$-module F with $e(X) := [\pi]_F(X) = \pi X + X^q$. By induction on n, one shows that the iterated polynomial $e^n(X) = e \circ e \circ \cdots \circ e(X) = [\pi^n]_F(X)$ is separable of degree q^n. The torsion group $F[\mathfrak{p}^n]$ consists of the zeroes of this polynomial, and therefore has q^n elements. Any element which does not belong to $F[\mathfrak{p}^{n-1}]$ must generate it freely over $\mathfrak{o}/\mathfrak{p}^n$ (which also has size q^n). □

Fixing π, now, as in the proposition (i.e., both for the definition of F and for the Tate module), we have commuting actions of $\mathrm{Gal}(\bar{K}/K)$ and $\mathfrak{o}$ on $T_\pi F_\pi$, hence we get a homomorphism $\eta_\pi : \mathrm{Gal}(\bar{K}/K)^{\mathrm{ab}} \to \mathrm{Aut}_{\mathfrak{o}} T_\pi F_\pi = \mathfrak{o}^\times$.

Theorem 10.4 (Lubin–Tate). *The composition of the Artin reciprocity map* $r_K : K^\times \to \mathrm{Gal}(\bar{K}/K)^{\mathrm{ab}}$ *with* η_π *is the unique homomorphism which is* $u \mapsto u^{-1}$ *on* $\mathfrak{o}^\times$ *and trivial (i.e.,* 1*) on* π*. In particular, the field* K_π *generated by the torsion points of* F_π *is a maximal totally ramified abelian extension of* K*, and* $K^{\mathrm{ab}} = K^{ur}K_\pi$.

The proof of this theorem will be given in § 10.5. Let us first recall what the reciprocity map of local class field theory is.

10.4 Recollection of local class field theory

There are many incarnations of the local reciprocity map, but maybe the most standard and laziest one (in that it packages everything of essence into the formalism) is the following:

> Assume that L/K is a finite extension of local (nonarchimedean) fields. Then the inverse reciprocity map
>
> $$r_{L/K}^{-1} : G := \mathrm{Gal}(L/K) \to K^\times/N_K^L L^\times$$
>
> (factoring through $\mathrm{Gal}(L/K)^{\mathrm{ab}}$, where it is an isomorphism) is given by a cup product in Tate cohomology,
>
> $$\hat{H}^{-2}(G, \mathbb{Z}) \to \hat{H}^0(G, L^\times)$$
>
> by $u_{L/K}$ = the fundamental class of $\mathrm{Br}_{L/K} = H^2(G, L^\times)$.

Let us unpack the statement: First, the Brauer group $\mathrm{Br}_{L/K}$ has many important incarnations, e.g., as the group of Morita equivalence classes[28] of central simple algebras over K that split in L, but for our purposes it is enough to know that it is canonically isomorphic to $\frac{1}{d}\mathbb{Z}/\mathbb{Q}$, where $d = [L : K]$, through a series of isomorphisms

$$H^2(G, L^\times) \hookrightarrow H^2(\mathrm{Gal}(\bar{K}/K), \bar{K}^\times) \xrightarrow{\sim} H^2(\mathrm{Gal}(\breve{K}/K) = \hat{\mathbb{Z}}, \breve{K}^\times) \xrightarrow{\sim} H^2(\mathrm{Gal}(\breve{K}/K) = \hat{\mathbb{Z}}, \mathbb{Z})$$

(where all cohomology groups are by definition taken with continuous cocycles, $\breve{K}$ denotes the maximal unramified extension of K, and the last map is induced from the valuation map), together with the connecting homomorphism

$$\mathbb{Q}/\mathbb{Z} = \mathrm{Hom}(\hat{\mathbb{Z}}, \mathbb{Z}) = H^1(\hat{\mathbb{Z}}, \mathbb{Z}) \xrightarrow{\sim} H^2(\hat{\mathbb{Z}}, \mathbb{Z}).$$

[28] Explicitly: $A \sim B$ if both A and B are matrix algebras over isomorphic division algebras.

The Tate cohomology groups connect Galois homology and cohomology, and in particular by definition we have

$$\hat{H}^{-2}(G,\mathbb{Z}) = H_1(G,\mathbb{Z}) = G^{\mathrm{ab}},$$

and $\hat{H}^0(G, L^\times) = K^\times / N_K^L L^\times$.

So, we have all the ingredients and, in principle, using standard operations in (Tate's) Galois cohomology can compute the map.

However, there is a much easier calculation, based on the following proposition.

Proposition 10.5. *Let L/K be an abelian extension that contains the maximal unramified extension $\breve{K}$. The reciprocity map $r_{L/K} : K^\times \to \mathrm{Gal}(L/K)$ is characterized among all such homomorphisms by the properties (1) that it reduces to the map uniformizer $\mapsto$ Frobenius under the quotient $\mathrm{Gal}(L/K) \to \mathrm{Gal}(\breve{K}/K)$, and (2) for every finite subextension $F \subset L$, the composite with $\mathrm{Gal}(L/K) \to \mathrm{Gal}(F/K)$ is trivial on the norm subgroup $N_K^F F^\times \subset K^\times$ (or, equivalently, on any uniformizer belonging to such a norm subgroup).*

Proof. That the reciprocity map (homomorphism) has these properties is clear from the existence of this map and its compatibility with the "inverse reciprocity maps" for finite extensions, defined above.

Since $K^\times$ is generated as a group by its uniformizing elements, it is enough to verify the characterization on those. If π is a uniformizer, and $L_\pi \subset L$ is the fixed field of $r(\pi)$, then by property (1) the extension $\mathrm{Gal}(L/L_\pi)$ is maximal unramified, while L_π is totally ramified, and $L = L_\pi \cdot \breve{K}$. If f is any homomorphism satisfying properties (1) and (2), the restrictions of f and r to $\breve{K}$ coincide by (1), therefore it suffices to check that $f(\pi)$ is trivial on L_π. By (2), it suffices to check that π is a norm from any finite subextension $F \subset L_\pi$, but this again follows from the construction of r as the inverse to a limit of isomorphisms $\mathrm{Gal}(F/K) \xrightarrow{\sim} K^\times / N_K^F F^\times$. □

This proposition can be used to calculate $r_{L/K}(\pi)$ for various abelian extensions L, first by passing to the maximal unramified extension $\breve{L} = L\breve{K}$, and then restricting to L.

10.5 Proof of the main theorem

The Lubin–Tate Theorem 10.4 will be proved through a series of steps. As we saw above, the extension L generated by the torsion points of F_π is abelian, given by a homomorphism $\eta_\pi : \mathrm{Gal}(\bar{K}/K) \to \mathfrak{o}^\times$. By Proposition 10.5, it suffices to show that

1. L is totally ramified;
2. η_π is surjective (hence a bijection $\mathrm{Gal}(L/K) \to \mathfrak{o}^\times$);
3. for any uniformizer $\bar{\pi} = u\pi$ of K, the element of $\mathrm{Gal}(\breve{L}/K)$ which is equal to $\eta_\pi^{-1}(u)$ on L and equal to the Frobenius on $\breve{K}$ satisfies property (2) of Proposition 10.5 (with respect to the extension $\breve{L}/K$).

Let $F = F_\pi$ be "the" Lubin–Tate module corresponding to π. We prove the first and second statements by showing that the points of $F[\mathfrak{p}^n] \smallsetminus F[\mathfrak{p}^{n-1}]$ are roots of an Eisenstein polynomial of degree $q^{n-1}(q-1)$, and therefore generate a totally ramified Galois extension L_n. Moreover, since the Galois action on the roots of such a polynomial is transitive, and on the other hand has to factor through the reduction of the character η_π to $(\mathfrak{o}/\mathfrak{p}^n)^\times$ (which also has the same order), L_n is abelian of order $q^{n-1}(q-1)$ over K.

To show that torsion points satisfy an Eisenstein polynomial, it is enough to work with any model of F_π, e.g., with our favorite one where $[\pi]$ is given by the polynomial $e(X) = \pi X + X^q$. As in the proof of Proposition 10.3, we will denote by e^n the n-fold composition of e with itself, and then we can compute that (the analog of the cyclotomic polynomial)

$$\phi_n(X) = \frac{e^n(X)}{e^{n-1}(X)}$$

is the Eisenstein polynomial

$$e^{n-1}(X)^{q-1} + \pi.$$

This computation actually tells us more, that we can use to prove the third fact: That for any $\lambda_n \in F[\mathfrak{p}^n] \smallsetminus F[\mathfrak{p}^{n-1}]$ the norm of $-\lambda_n$ is $N_K^{L_n}(-\lambda_n) = \pi$. That shows that π is a norm for *every* finite subextension of L, and therefore L is fixed by $r_{L/K}(\pi)$ – hence, the claim of the theorem holds, at least, for π, i.e., $\eta_\pi \circ r_{L/K}(\pi) = 1$.

To prove it for an arbitrary $\bar{\pi} = u\pi$, we will also invoke the Lubin–Tate group $F_{\bar{\pi}}$. The idea is that, after base changing to the ring of integers of the completion $\hat{K}$ of the maximal unramified extension $\breve{K}$, the two formal $\mathfrak{o}$-modules become isomorphic. More precisely, consider the Galois twist of F_π which is defined by the homomorphism

$$(\sigma := \mathrm{Fr}_q \mapsto [u^{-1}]) \in \mathrm{Hom}(W(\breve{K}/K), [\mathfrak{o}^\times]),$$

considered as a 1-cocycle on the Weil group $W(\breve{K}/K)$ with values in $\mathrm{Aut}(F_\pi \otimes \breve{\mathfrak{o}})$. When we reduce modulo any power of the prime ideal, this cocycle extends to the Galois group, and therefore defines a new rational structure on $\mathrm{Aut}(F_\pi \otimes$

$\breve{\mathfrak{o}}/\breve{\mathfrak{p}}^n$), compatibly for every n. Passing to the limit over all n, we get a new formal $\mathfrak{o}$-module $F_{\pi,\bar{\pi}}$ *which is isomorphic to the original one after tensoring with* $\hat{\mathfrak{o}}$, the ring of integers of the completion $\hat{K}$. (Passing to the completion is necessary here, because the Weil cocycle only becomes a Galois cocycle in the completion.)

Proposition 10.6. *The formal $\mathfrak{o}$-module $F_{\pi,\bar{\pi}}$ constructed above is isomorphic to the Lubin–Tate module $F_{\bar{\pi}}$.*

Explicitly, for any chosen models of F_π, $F_{\bar{\pi}}$, there exists an invertible power series $\phi = \epsilon X + O(X^2) \in \hat{\mathfrak{o}}[[X]]$ ($\epsilon \in \hat{\mathfrak{o}}^\times$) with the following properties:

1. ${}^\sigma\phi = \phi \circ [u]_\pi$;
2. $\phi \circ F_\pi = F_{\bar{\pi}} \circ (\phi \times \phi)$;
3. $\phi \circ [a]_\pi = [a]_{\bar{\pi}} \circ \phi$ *for all* $a \in \mathfrak{o}$.

(Note that the first property implies that $\epsilon^{\sigma-1} = u$; such an element has to be genuinely in the completion $\hat{\mathfrak{o}}$ and not in $\breve{\mathfrak{o}}$ when u is not a root of unity!)

Proof. The explication is just a restatement of the proposition, and it is also the way that this proposition appears in the original article of Lubin and Tate, and virtually all of its accounts in the literature.

It is "obvious" that the Galois twist is another formal $\mathfrak{o}$-module over $\mathfrak{o}$. (Start from: Why is the Galois twist of a free $\mathfrak{o}$-module a free $\mathfrak{o}$-module?) That means that there is an invertible power series ϕ satisfying Property (1) of the proposition, so that with respect to the parameter $Y = \phi(X)$ the module becomes isomorphic to $\mathscr{G}_{\bar{\pi}}$, i.e., the map $f \mapsto f \circ \phi^{-1}$ (where ϕ^{-1} is the inverse under composition) defines an isomorphism

$$\hat{\mathfrak{o}}[[X]] \text{ with the new Galois action} = \mathfrak{o}[[X]] \otimes_{\mathfrak{o}} \hat{\mathfrak{o}},$$

with the "usual" Galois action on the right hand side. Explicitly, if we denote the twisted Frobenius action by $\sigma\star$ and the usual one (on coefficients) by an exponent σ, then

$${}^\sigma f(\phi(X)) = \sigma \star (f \circ \phi)(X).$$

Moreover, if we take Properties (2) and (3) as the definition of $F_{\bar{\pi}}$ and $[a]_{\bar{\pi}}$, they will automatically satisfy the axioms of formal $\mathfrak{o}$-modules, since we twisted by an automorphism of formal $\mathfrak{o}$-modules.

To identify this with $F_{\bar{\pi}}$, we need to show that $\bar{\pi}$ acts as the q-Frobenius modulo $\mathfrak{p}$. This follows from the construction: Suppose that ϕ is as above, intertwining the two Galois actions. Then, modulo $\mathfrak{p}$ we have

$$\begin{aligned} f([\bar{\pi}]_{\bar{\pi}}(\phi(X))) = f(\phi([\bar{\pi}]_\pi(X)) &= f \circ \phi([u]_\pi \circ [\pi]_\pi(X)) \\ &= f \circ \phi([u]_\pi(X^q)) = f \circ {}^\sigma\phi(X^q) = f((\phi(X))^q). \end{aligned}$$

□

Using Proposition 10.6, we can think of F_π and $F_{\bar\pi}$, as two different rational structures on a formal $\mathfrak{o}$-module over $\hat{\mathfrak{o}}$. The set of torsion points does not depend on the rational structure, and hence the field extension $\hat L$ of $\hat K$ that they generate does not depend on it.

Since the Galois action, to obtain the second rational structure, was twisted by the character $\mathrm{Fr}_{\mathfrak{p}} \mapsto [u^{-1}]$, it immediately follows that the corresponding characters η_π, $\eta_{\bar\pi}$ are related[29] by $\eta_{\bar\pi}(\tau) = \eta_\pi(\tau)u^{\mathrm{val}(\tau)}$, where by $\mathrm{val}(\tau)$ we denote the power of Frobenius that τ restricts to in $\mathrm{Gal}(\check K/K)$. But we already know that $\eta_{\bar\pi} \circ r_{L/K}(\bar\pi) = 1$, and therefore $\eta_\pi \circ r_{L/K}(\bar\pi) = u^{-1}$.

10.6 Finite and p-divisible groups

[No proofs here. Just a tourist guide.]

10.6.1 Definition of p-divisible groups

Given a formal group F over a ring A (not necessarily 1-dimensional, but still assumed abelian throughout), we say that F is *p-divisible* if the p-power morphism $[p] : F \to F$ is an isogeny, i.e., presents $A[[X]]$ as a finite *locally free*[30] $A[[X]]$-module. The kernel of this morphism will then be locally free of rank p^h over itself (this will follow from our discussion of finite, connected p-torsion groups below), for some integer h called the *height* of F, and, more generally, the kernel of $[p^n]$ will be finite free of rank p^{nh}.

A different (but related) notion of p-divisible groups (without the word "formal," and also called *Barsotti–Tate groups*) is that of an inductive system

$$G : \cdot \to G_n \to G_{n+1} \to \cdot$$

of finite, locally free abelian group schemes over A, with the transition maps being monomorphisms identifying G_n with the p^n-torsion in G_{n+1}, and the multiplication-by-p maps $G_{n+1} \to G_n$ being epimorphisms. It then follows that the rank of G_n is of the form p^{nh} for some locally constant function h on A, called the *height* of G.

[29] The action on torsion points has the *inverse* twist! Explicitly, if x is a torsion point for $F_{\bar\pi}$ and $\bar\sigma$ a Frobenius element, then ${}^{\bar\sigma}(\phi(x)) = {}^{\sigma}\phi({}^{\bar\sigma}x) = \phi \circ [u] \circ [\eta_\pi(\bar\sigma)](x) = [u\eta_\pi(\bar\sigma)](\phi(x))$.

[30] Some of the literature, e.g., Tate, says "free" here, but this seems unnatural; is it automatic? Of course, there's no difference for the local rings that we will consider.

10.6.2 Cartier duality

Most of the statements about finite groups, with bibliographical references for their proofs, can be found in [Mil17, Chapter 11]. Shoutout to Patrick Walls for excellent slides summarizing the theory.

In this section, we will say "group" for a *finite (abelian) group* over a base ring A, by which we mean an abelian group scheme, locally free of finite order over A. Note that the entire discussion here will be about abelian group schemes, hence we will omit the word "abelian" throughout.

Cartier duality is the adaptation of Pontryagin duality of classical Fourier analysis to this setting. Namely, the Cartier dual of a finite group G is the "group of characters," i.e., homomorphisms $G \to \mathbb{G}_m$. We can apply this definition to points over any A-algebra, thus obtaining a functor that is representable by another finite group scheme, the Cartier dual G' of G. At the level of algebras, the coordinate ring $A[G]$ is a Hopf algebra – skipping the definitions here, but in the commutative case these are precisely the axioms required to make $\operatorname{spec} A[G]$ into a group scheme, such as a comultiplication map $A[G] \to A[G] \otimes_A A[G]$ – and the coordinate ring of $A[G']$ is the dual A-module, with multiplication becoming comultiplication, and vice versa. As is clear from this description, $(G')' = G$. The functor $G \to G'$ is an anti-equivalence of the category of finite groups over A.

From now on, we work over a complete noetherian local ring A. There is a canonical short exact sequence

$$0 \to G^0 \to G \to G^{\text{et}} \to 0, \tag{10.1}$$

with G^0 connected and G^{et} étale.

Now, let's assume that $A = k$, a perfect field of characteristic $p > 0$. We then have the (relative)[31] p-Frobenius morphism, for every group G over k,

$$F_G : G \to G^{(p)},$$

and its Cartier dual (the morphism obtained by applying Frobenius to the Cartier duals), the *Verschiebung* morphism

$$V_G : G^{(p)} \to G.$$

In this setting, the exact sequence (10.1) splits (non-canonically), and we have equivalences

[31] Recall: The relative Frobenius morphism in coordinates is induced from $\underline{X} \mapsto \underline{X^p} : k[\underline{X}]/(\underline{f^{(p)}}(\underline{X})) \to k[\underline{X}]/(\underline{f}(\underline{X}))$, where $f^{(p)}$ is the polynomial obtained from f by raising its coefficients to the p-th power.

$$G \text{ is étale} \iff F_G \text{ is an isomorphism;}$$
$$G \text{ is connected} \iff F_G \text{ is nilpotent.}$$

The Cartier dual conditions are called *multiplicative* and *unipotent*, i.e.,

$$G \text{ is multiplicative} \iff V_G \text{ is an isomorphism;}$$
$$G \text{ is unipotent} \iff V_G \text{ is nilpotent.}$$

They admit the following equivalent characterizations, hence the name.

Proposition 10.7. *G is multiplicative iff it embeds into copies of $\mathbb{G}_m$ over the algebraic closure, and unipotent if it has a normal filtration with quotients equal to subgroups of $\mathbb{G}_a$.*

Proof. See [Dem86, § II.8–9]. □

Here is a table with the basic examples, where we denote by α_p the kernel of Frobenius: $\mathbb{G}_m \to \mathbb{G}_m$.

G	G'	F_G	V_G
α_p	α_p	0	0
$\mathbb{Z}/p\mathbb{Z}$	μ_p	$\simeq$	0
$\mathbb{Z}/l\mathbb{Z}$	μ_l	$\simeq$	$\simeq$.

In particular, p-power torsion groups over k are built out of constituents G such that G or its Cartier dual is connected. In the next section, we will start by classifying finite unipotent groups in terms of semilinear data, and will then use Cartier duality to extend this classification to all p-power torsion groups.

10.6.3 Witt vectors and Dieudonné modules

A *unipotent* (affine, finite type for everything that follows) group scheme over k is one that embeds into the subgroup of upper triangular matrices in some GL_n, or, more abstractly, one that admits a filtration by normal subgroups whose successive quotients are subgroups of $\mathbb{G}_a$. This is a general definition, but now we will continue to assume that our groups are abelian.

First, we would like to extend the definition of Verschiebung to not necessarily finite commutative affine group schemes G. We do this by considering the diagram

$$k[G] \to (k[G]^{\otimes p})^{S_p} \xrightarrow{\lambda} k[G] \otimes_\sigma k,$$

where the first arrow is the p-fold comultiplication map (landing in symmetric tensors), and λ is the unique k-linear map sending $a \otimes a \otimes \cdots \otimes a$ to a. (Note that it is k-linear with respect to the Frobenius-twisted action of k, i.e., $ca \otimes$

$ca \otimes \cdots \otimes ca$ goes to $c^p a$!) When $k[G]$ is finite over k, and $A = k[G']$ is the dual Hopf algebra, the dual of this sequence of maps,

$$A \leftarrow (A^{\otimes^p})_{S_p} \leftarrow A \otimes_\sigma k$$

where the right arrow is sending $a \otimes 1$ to $a \otimes \cdots \otimes a$ and the left arrow is multiplication, is precisely the (relative) Frobenius morphism, $G' \to G'^{(p)}$; hence the definition of Verschiebung reduces to the previous one.

Remark 10.2. At least over fields, Cartier duality extends beyond the finite setting, to a duality between affine commutative group schemes and formal groups, see [Dem86, § 4]. This is implicit behind the extension of the Verschiebung morphism given above.

Now, we introduce the ring of (p-typical) Witt vectors $W = \lim_n W_n$. We will take its construction as a black box, and only mention the following facts:

1. It is a ring-scheme of infinite type over $\mathbb{F}_p$, with a (fixed) isomorphism of schemes $W = \mathbb{A}^\infty$, $W_n = \mathbb{A}^n$, i.e., every element is represented by a sequence $(a_0, a_1, a_2, \dots)$, with the truncated Witt ring W_n corresponding to $(a_0, \dots, a_{n-1})$.
2. At the level of points, $W(\mathbb{F}_p) \simeq \mathbb{Z}_p$ and, more generally, $W(\mathbb{F}_q) \simeq \mathbb{Z}_q$, the ring of integers in the unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$ (here, q is a power of p). **In other words, the Witt construction gives a way to understand p-adic rings as $\mathbb{F}_q$-points of some ring scheme.** *Compare: the equal-characteristic construction of arc spaces L^+X of varieties X over $\mathbb{F}_p$, with $L^+\mathbb{G}_a(\mathbb{F}_q) \simeq \mathbb{F}_q[[t]]$.*

   Explicitly, if $[a]$ denotes the Teichmüller lift to $\mathbb{Z}_q$ of an element in $\mathbb{F}_q$ (i.e., the unique multiplicative section of the reduction map), the isomorphism is given by

   $$(a_0, a_1, a_2, \dots) \mapsto [a_0] + [a_1^{1/p}]p + [a_2^{1/p^2}]p^2 + \dots.$$

   (Why the inverse Frobenius twists in the coefficients? It is necessary in order to make the addition law polynomial!)
3. Multiplication by p is given by $p \cdot (a_0, a_1, a_2, \dots) = (0, a_0^p, a_1^p, a_2^p, \dots)$.
4. It carries a p-Frobenius morphism (induced by this isomorphism of schemes),

   $$F \text{ or } \sigma : (a_0, a_1, a_2, \dots) \mapsto (a_0^p, a_1^p, a_2^p, \dots),$$

   and the Verschiebung is given by

   $$V : (a_0, a_1, a_2, \dots) \mapsto (0, a_0, a_1, \dots).$$

In particular, the truncated Witt groups W_n are V-torsion. They are also pro-unipotent: The quotients $p^n W/p^{n+1}W$ are isomorphic to $\mathbb{G}_a$. The following theorem will present them as co-generators of the corresponding category.

Be careful from now on to distinguish between W as a ring scheme and the ring $W(k)$ of its points over a perfect field k. For example, on $W(k)$ we can invert the p-power Frobenius (because k is perfect), but on W we can't (because the $1/p$-th power is not polynomial).

Define the Dieudonné ring D_k to be the ring generated by $W(k)$ and elements F, V with relations

$$F\lambda = {}^{\sigma}\lambda F, \quad V \cdot {}^{\sigma}\lambda = \lambda V, \quad FV = VF = p$$

for all $\lambda \in W(k)$, where σ denotes the Frobenius on W. A *Dieudonné module* is a module for D_k.

Define the ind-group $W_\infty := \operatorname{colim} W_n$ over $\mathbb{F}_p$, where W_n is the n-truncated Witt ring, embedded in W_{n+1} via $(a_1, a_2, \ldots, a_n) \mapsto (0, a_1, a_2, \ldots, a_n)$. It is a module for $W(k)$ under the following twisted action on W_n: $\lambda \star w_n = ({}^{\sigma^{1-n}}\lambda_n)w_n$, where λ_n denotes the n-th truncation of λ.

At the level of points, $W_\infty(\mathbb{F}_p) \simeq \mathbb{Q}_p/\mathbb{Z}_p$.

We have $\operatorname{End} W_{n,k} = D_k/D_k V^n$ [DG70, Proposition V.1.3.4].

Theorem 10.8. *The functor*

$$G \mapsto M(G) = \operatorname{Hom}(G, W_\infty) = \operatorname{colim} \operatorname{Hom}(G, W_n)$$

is an anti-equivalence of categories

$$\{\textit{Unipotent (abelian) group schemes over } k\} \leftrightarrow \{\textit{finite-type } V\textit{-nilpotent Dieudonné modules}\},$$

restricting to an equivalence

$$\{\textit{Finite unipotent groups over } k\} \leftrightarrow \{V\textit{-nilpotent Dieudonné modules of finite } W(k)\textit{-length.}\}$$

Remarks 10.1.

1. Recall that the length of a module is the longest l such that there is a chain of submodules $0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_l$; in particular, the free $W(k)$-module has infinite length.
2. We have
$$M(G^{(p)}) = \sigma^* M(G), \tag{10.2}$$
which is where the semilinear structure comes from.

Theorem 10.9. *There is an anti-equivalence of categories*

$$\{\textit{Finite } p\textit{-power torsion groups over } k\} \leftrightarrow \{\textit{Dieudonné modules of finite } W(k)\textit{-length.}\}$$

This equivalence $G \mapsto M(G)$ is characterized by the following properties:

1. *On finite unipotent groups, it is the one given by Theorem 10.8.*
2. *Under Cartier duality, $M(G') = M(G)^* := \mathrm{Hom}_{W(k)}(M(G), W_\infty(k))$, with Dieudonné module structure given by $(Ff)(m) = {}^{\sigma}f(Vm)$ and $(Vf)(m) = {}^{\sigma^{-1}}f(Fm)$. (Here, $f \in M(G)^*$, $m \in M(G)$ and σ is Frobenius on $W_\infty(k)$.)*

See [Dem86, III.5–6] for the statements and discussion, and [DG70, Theorem V.1.4.3] for the proof of the critical Theorem 10.8.

10.6.4 Dieudonné modules of p-divisible groups

Assume that A is a complete noetherian local ring whose residue field is of characteristic p. Then we have the following.

Proposition 10.10. *The functor of p-power torsion gives an equivalence of categories between p-divisible formal Lie groups over A, and* connected *p-divisible (Barsotti–Tate) groups.*

See [Tat67, Proposition 1].

This equivalence gives rise to a notion of *dimension* for a p-divisible group (take the identity components of the G_n's, which form a connected p-divisible group, and take the dimension of the associated formal group), which is an invariant that it complementary to the height. It can be computed in terms of the discriminant ideal of the group [Tat67, Proposition 2], and there is a Cartier dual p-divisible group G' (by taking the Cartier duals of the G_n's, and the duals of the p-power maps), satisfying

$$\dim G + \dim G' = h,$$

see [Tat67, Proposition 3].

Remarks 10.2.

1. A Lubin–Tate group for the ring of integers $\mathfrak{o}$ of a p-adic field K is p-divisible of height $h = [K : \mathbb{Q}_p]$.
2. The equivalence of categories between connected and formal p-divisible groups explains why, for an elliptic curve E with CM by an order R and a good reduction at a prime over p, the corresponding formal group $\hat{E}$ automatically admits endomorphisms by the completion R_p,

Say over what, and check/correct the completion

Theorem 10.9 implies the following.

Theorem 10.11. *The functor $M(\operatorname{colim} G_n) = \lim M(G_n)$ gives rise to an anti-equivalence of categories*

$$p\text{-divisible groups over } k \leftrightarrow \text{Dieudonné modules over } k,$$

where the right hand side is the category of free *finite-rank modules M over the Witt ring $W(k)$, equipped with an endomorphism F such that*

- *F is semilinear with respect to the p-power Frobenius σ, i.e., $F(am) = {}^{\sigma}am$, for all $a \in W(k)$;*
- *$FM \supset pM$.*

Remarks 10.3.

1. The Verschiebung morphism is uniquely determined by the conditions above and $FV = VF = p$.
2. The rank of the Dieudonné module corresponds to the height of the p-divisible group.
3. The dimension of G is the dimension of the k-vector space $M(G)/FM(G)$.
4. For connected p-divisible groups, the Dieudonné module admits another description in terms of invariant differentials and de Rham cohomology for the corresponding (by Proposition 10.10) formal p-divisible group; see [Wei].

10.7 Moduli of p-divisible groups and the local Langlands correspondence

[Stub. I point to [Wei, RV14] for a nicer and more detailed discussion. The main effort here is to highlight the reasons behind some technical choices. There may be errors due to my misunderstanding – please check the original sources.]

Generalizing Lubin–Tate theory, appropriate moduli spaces of 1-dimensional formal p-divisible groups of height h can be used to prove the local Langlands conjecture for GL_h, e.g., in the following form:

> There is a canonical bijection (characterized, e.g., by properties provided by Henniart [Hen93]) between irreducible h-dimensional representations of the Weil group of $\mathbb{Q}_p$ and irreducible supercuspidal representations of $\mathrm{GL}_h(\mathbb{Q}_p)$.

10.7.1 Deformations of p-divisible groups

In the theory of complex multiplication, the maximal abelian extension of an imaginary quadratic field is realized by the Galois action on a moduli space, the moduli space of CM elliptic curves (by a fixed order) with full level structure.

There was roughly a mention of moduli spaces for Lubin–Tate theory, but this is because, in reality, it is based on a moduli space with a single point – the moduli of Lubin–Tate modules for a fixed uniformizer π. We will generalize this to construct nonabelian extensions (in the form of higher-dimensional representations of the Galois group), but for simplicity we will restrict to the case of the field $K = \mathbb{Q}_p$.

This is harmless, because the theory of general Galois extensions of $\mathbb{Q}_p$ contains the Galois theory of every finite extension of $\mathbb{Q}_p$, unlike the abelian case. But, if one desires, one can replace $\mathbb{Q}_p$ by an arbitrary p-adic field, and this is the approach of Harris–Taylor in [HT01].

We can view a formal group over $\mathbb{Z}_p$ as a deformation of its special fiber. Fix again a perfect field k of characteristic p. The classification of deformations proceeds as follows.

1. When k is algebraically closed, there is a *unique, up to isomorphism 1-dimensional* divisible formal group *of height* h over k. This can be deduced from Theorem 10.11, see [Wei, p.10].
2. When k is not algebraically closed, all "forms" of such a formal group over k are again described by Theorem 10.11. Fixing such a form G_0 corresponds to fixing the Lubin–Tate module F_π modulo $\mathfrak{p}$. (For the purposes of the local Langlands conjecture, we can actually avoid this choice, as we will see, working directly over $k = \overline{\mathbb{F}_p}$.)
3. Fixing G_0 of height h over k, there is a classification of pairs (G, ι), where G is a 1-dimensional formal group G over a local Artinian algebras A with residue field k, and ι is an isomorphism of special fibers, $\iota : G_0 \xrightarrow{\sim} G \otimes_A k$. This deformation problem is pro-represented by the ring $R_h := W(k)[[X_1, \dots, X_{h-1}]]$, i.e., there is a universal formal group over this ring such as G is obtained by base change via a morphism $R_h \to A$.

Remark 10.3. Note that part of the assertion that this problem is pro-representable by a formal scheme is the fact that the problem is rigid, i.e., the pair (G, ι) has no automorphisms. More precisely, in the setting above, for p-divisible groups G, G' over A, the specialization map $\mathrm{Hom}(G, G') \to \mathrm{Hom}(G_0, G'_0)$, where by an index 0 we denote special fibers, is injective. Here, the Homs are taken in the isogeny category.

10.7.2 A host of problems (and how to solve them)

Fix a height h and let $M_0 = \mathrm{Spf} R_h$ (the deformation ring of 1-dimensional p-divisible formal groups of height h).

Naively speaking, we could hope to add p^n-level structures, obtaining a tower of moduli spaces M_n, and recover the local Langlands correspondence for $\mathrm{GL}_h(\mathbb{Q}_p)$ from simultaneous actions of $\mathrm{GL}_h(\mathbb{Q}_p)$ and the Galois group on the cohomology of $M_\infty = \lim M_n$.

There are several problems to address here.

1. First of all, for such a formal group G over $\mathbb{Z}_p$ (or, more generally, over $W(k)$, where k is a perfect field), a level structure

$$(\frac{1}{p^n}\mathbb{Z}/\mathbb{Z})^h \xrightarrow{\sim} G[p^n]$$

   cannot exist as an isomorphism of group schemes (since G is connected, and its torsion over the special fiber is trivial), but only at the level of sections over infinitesimal deformations of $\mathrm{Spec}k$. However, those don't behave nicely under specialization (again, because G is not étale), and therefore we need a modification of the notion of level structure, introduced by Drinfeld, where the maps above are not always isomorphisms at the level of sections. See [HT01] for references.

   Alternatively, the approach of Rapoport–Zink [RZ96] (see also [RV14] for a nice exposition) is to remove the points of the deformation space over the special fiber. But this is a formal scheme, and all of its points (consider as a ringed space) lie over the special fiber. But the generic fiber of such a formal scheme makes sense in the category of *rigid analytic spaces*, and one needs to work in this framework. Thus, we have the rigid analytic space $\mathcal{M}_0$ corresponding to the generic fiber of M_0, and we hope to define a tower of rigid analytic spaces $\mathcal{M}_n$ over it, adding level structures.

2. We will completely gloss over this, but we need a manageable and sufficiently "finite" cohomology theory for these formal schemes or rigid analytic spaces – this is not automatic, as they "spread out" over $W(k)$, which is not "proper." One applies Berkovich's version of "vanishing cycles" over the special fiber of $\mathrm{Spf}W(k)$ to get some reasonable cohomology theory.

3. If we did construct those moduli spaces as stated above, we would get an action of $\mathrm{GL}_h(\mathbb{Z}_p)$ on them (from its action on $(\mathbb{Q}_p/\mathbb{Z}_p)^h$), *not* of $\mathrm{GL}_h(\mathbb{Q}_p)$. This is similar to the fact that, in Lubin–Tate theory, we obtained a homomorphism $\mathrm{Gal}(\bar{K}/K) \to \mathfrak{o}^\times$, and not the inverse reciprocity map on the nose. As in the Lubin–Tate case, one can reconstruct the reciprocity map from the $\mathrm{GL}_h(\mathbb{Z}_p)$-action (as in Scholze [Sch13]), but it would be nicer to have a tower of spaces with a $\mathrm{GL}_h(\mathbb{Q}_p)$-action.

The rigid-analytic Rapoport–Zink spaces have a slightly modified moduli problem built into them, in order to get $\mathrm{GL}_h(\mathbb{Q}_p)$-actions on the tower $\mathcal{M} = \lim \mathcal{M}_n$, which at the same time removes the need to choose a form over a non-algebraically closed residue field like $\mathbb{F}_p$ (i.e., the analog of choosing the uniformizer π in Lubin–Tate theory).[32]

10.7.3 Revisiting the Lubin–Tate construction

For this, let us revisit Lubin–Tate theory for a p-adic field K, and try to construct the reciprocity map by working directly over $k = \overline{\mathbb{F}_p}$, and over the completion $\hat{\mathfrak{o}}$ of the Witt ring $W(k) = \breve{\mathfrak{o}}$. As we saw in Proposition 10.6, all Lubin–Tate formal $\mathfrak{o}$-modules are isomorphic over $\hat{\mathfrak{o}}$, but of course in this setting we only get an action of the inertia group I_K on the torsion $G[\mathfrak{p}^\infty]$ of such a group. To recover an action of the entire Galois (or rather Weil) group, we can set up a moduli problem as above: Fix G_0 over k, and consider triples (G, ι, α), where G is a formal $\mathfrak{o}$-module over $\hat{\mathfrak{o}}$, $\iota : G_0 \xrightarrow{\sim} G \otimes_{\hat{\mathfrak{o}}} k$, and α is an isomorphism of $\mathfrak{o}$-modules, $K/\mathfrak{o} \xrightarrow{\sim} G_{\hat{K}}[\mathfrak{p}^\infty]$. Now, any element τ in the Galois group acts on such data, but τ carries the isomorphism ι to an isomorphism between the τ-twists,

$$ {}^\tau G_0 \xrightarrow{\sim} {}^\tau G \otimes_{\hat{\mathfrak{o}}} k. $$

Of course, the τ-twists are isomorphic to the original objects (I stress again that there is a unique isomorphism class!), but not in any canonical way, so this does nothing to give us an action of τ on level structures. However, *if τ is in the Weil group, ${}^\tau G_0$ is just a power of the Frobenius twist of G_0. Let us for now assume that this power is a positive number m; we can compose with the corresponding power of the relative Frobenius map*,

$$ F^m_{G_0} : G_0 \to {}^\tau G_0 $$

to get a map

$$ G_0 \to {}^\tau G \otimes_{\hat{\mathfrak{o}}} k. \tag{10.3} $$

This is almost what we want, except that the last map is not an isomorphism, because Frobenius isn't. Thus, we need to pass to the *quasi-isogeny* category of formal groups, i.e., to a new category with the same objects, where we invert isogenies, defining new Hom-groups

$$ \mathrm{Hom}'(G_1, G_2) = \mathrm{Hom}(G_1, G_2)[\frac{1}{p}]. $$

[32] Harris–Taylor [HT01] also define a $\mathrm{GL}_h(\mathbb{Q}_p)$-action, without explicitly modifying the moduli problem or mentioning quasi-isogenies, but this reinterpretation is much cleaner.

Now the isogenies become isomorphisms, but they don't induce isomorphisms of Tate modules any more, except rationally.

Therefore, to get a set with a Weil group action, we need to replace the triples above by triples (G, ι, α) where G is as above, ι is a *quasi-isogeny* of its special fiber with G_0, and α is a trivialization of its *rational* Tate module $V_pG = T_pG \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$, $\alpha : K \xrightarrow{\sim} V_pG$. To finish our reworking of Lubin–Tate theory, we know that G and ${}^\tau G$ must be isomorphic (because there is only one isomorphism class of Lubin–Tate modules over $\hat{\mathfrak{o}}$), ι and the isogeny (10.3) fix a choice of (quasi-)isogeny ${}^\tau G \to G$, identifying their rational Tate modules. The τ-twist of α, now, gives rise to a homomorphism $\mathcal{W}_{\mathbb{Q}_p} \to \mathrm{GL}_1(\mathbb{Q}_p)$, which one can check is the inverse homomorphism to the inverse reciprocity map.

Exercise(s) 10.1. Check the last assertion against what we have proven!

10.7.4 The local Langlands correspondence for GL_h

Returning to general height h, there is an infinite tower of rigid analytic spaces $\mathcal{M}$ whose points parametrize triples (G, ι, α) where G is a p-divisible group over $\hat{\mathfrak{o}}$, ι is a quasi-isogeny of its special fiber with "the" p-divisible group G_0 of height h over $k = \overline{\mathbb{F}_p}$, and $\alpha : \mathbb{Q}_p^h \xrightarrow{\sim} V_pG$ is a trivialization of its rational Tate module. It admits commuting actions $\mathrm{GL}_h(\mathbb{Q}_p)$ and the Weil group $\mathcal{W}_{\mathbb{Q}_p}$, but there is a *third* commuting action to consider, that of the *automorphism group of* G_0 (in the quasi-isogeny category!).

This automorphism group can be computed using Dieudonné theory: In the quasi-isomorphism category, Dieudonné theory remains the same, except that we need to invert p, which means that instead of free $W(k)$-modules we consider $W(k)[\frac{1}{p}]$-vector spaces (i.e., $\breve{\mathbb{Q}}_p$-vector spaces, when $k = \overline{\mathbb{F}_p}$), equipped with a semilinear F-action by automorphisms. These are called *isocrystals*, and the h-dimensional ones are parametrized by the set $\mathrm{GL}_h(\breve{\mathbb{Q}}_p)/\sim$, where $\sim$ refers to Frobenius-twisted conjugacy: $b \simeq g^{-1} b \sigma(g)$. Indeed, if we fix a basis for such a crystal, $M = \mathbb{Q}_p^h$, the element b measures the ration between F and the standard Frobenius on $\mathbb{Q}_p^h$.

The unique 1-dimensional p-divisible group of height h over k corresponds to $M = \breve{\mathbb{Q}}_p^h$ with the usual Frobenius action twisted by the element

$$b = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 1 & \\ & & & \ddots & \\ p & & & & 0 \end{pmatrix}$$

(up to Frobenius-twisted conjugacy), and its endomorphism ring in the isogeny category is the subalgebra of $\mathrm{Mat}_h(\breve{\mathbb{Q}}_p)$ fixed by the twisted Frobenius action. It is the central h^2-dimensional division algebra D_h over $\breve{\mathbb{Q}}_p$ with invariant ("slope") $\frac{1}{h}$.

We set $J = D_h^\times$, which is an inner form of GL_h. The Jacquet–Langlands correspondence furnishes a bijection between the irreducible smooth representations of the (compact modulo center) group $J(\mathbb{Q}_p)$ and the discrete series representations of $\mathrm{GL}_h(\mathbb{Q}_p)$. Assume that ρ is an irreducible representation of $J(\mathbb{Q}_p)$, with $\rho' = JL(\rho)$ supercuspidal.

The proof of the local Langlands conjecture by [HT01] states, approximately, that there is a correspondence LL between irreducible supercuspidal representations of $\mathrm{GL}_h(\mathbb{Q}_p)$ and irreducible h-dimensional representations of $\mathcal{W}_{\mathbb{Q}_p}$, such that the ρ-isotypic part of the appropriate compactly supported cohomology $H_c(\mathcal{M})$ of the tower is isomorphic to

$$\rho \boxtimes \rho' \boxtimes LL(\rho'),$$

as a representation of $J(\mathbb{Q}_p) \times \mathrm{GL}_h(\mathbb{Q}_p) \times \mathcal{W}_{\mathbb{Q}_p}$. See [RV14, § 7] for more precise statements.

11 Geometric class field theory and sthukas for function fields

References:

- Yun, "Introduction to shtukas and their moduli." ArXiv:2411.10248.
- Bhatt, "Geometric class field theory", in Oberwolfach Arbeitsgemeinschaft "The Geometric Langlands Conjecture" report.
- V. Lafforgue, "Shtukas for reductive groups and Langlands correspondence for function fields."
- Milne, "Abelian Varieties".
- Serre, "Algebraic groups and class fields" [Ser88].
- Peter Toth, "Geometric abelian class field theory", Utrecht Masters Thesis.

[Caution: some formulas may be correct only up to inversion — haven't been checked carefully! Please let me know if you find such errors.]

11.1 Drinfeld's shtukas

We will now discuss the case of $K = \mathbb{F}_q(C)$, the function field of a curve C over a finite field $\mathbb{F}_q$, i.e., a finite extension of $\mathbb{F}_q(t)$. As function fields are invariant under compactifications and normalization, we may and will assume that C is smooth and projective.

We will proceed in inverse historical order. First, we will give a definition of Drinfeld's shtukas for GL_h (in order to see the analogy with Dieudonné theory for p-divisible groups), and then we will specialize to the case of GL_1, to see how they provide a geometric enhancement (and proof) of global class field theory. We will restrict ourselves to the unramified setting (i.e., we will construct the Hilbert class field of K), and we will also assume, for simplicity, that C contains an $\mathbb{F}_q$-point ∞ (which we fix).

Fix a smooth projective curve C over $\mathbb{F} = \mathbb{F}_q$, $q = p^r$, and let $k = \mathbb{F}(C)$, its function field. In this section, we will use Fr_S to denote the relative q-Frobenius morphism, on any scheme S over $\mathbb{F}$, i.e., $\mathrm{Fr}_S : S \to S$.

Definition 11.1. Let S be a scheme over $\mathbb{F}$. An *S-shtuka* of rank h over C is the data $(\mathscr{E}, \underline{x}, \tau)$, where

- $\mathscr{E}$ is a vector bundle of rank h on $C \times S$ (all products here are fiber products over $\mathrm{spec}\,\mathbb{F}$);

- $\underline{x} = (x_i)_I$ is a finite collection of S-points (the "legs" of the shtuka), $x_i \in C(S)$, indexed by a finite set I;
- α is an isomorphism of vector bundles outside of those points, between $\mathscr{E}$ and its S-Frobenius twist, i.e., if $U \subset C \times S$ denotes the complement of the graphs of these points, α is an isomorphism

$$\alpha : \mathscr{E}|_U \xrightarrow{\sim} (I_C \times \mathrm{Fr}_S)^* \mathscr{E}.$$

Consider an $\overline{\mathbb{F}}$-valued shtuka, and fix a leg x_i. We can trivialize the bundle in a formal neighborhood of x_i, $\mathscr{E} \otimes \mathcal{O} \simeq \mathcal{O}^h$, where $\mathcal{O} = \mathcal{O}_{x_i}$ (non-canonically isomorphic to $\overline{\mathbb{F}}((t))$) and also its Frobenius twist. (We can choose the induced trivialization, but we don't need to here – we are losing some information.) The Frobenius map, then, becomes an element of $\mathrm{GL}_h(\mathcal{F})$, where $\mathcal{F} = \mathcal{F}_{x_i}$ is the fraction field of the completed local ring $\mathcal{O}$. Modding out by all the possible trivializations, *we get a well-defined double coset* in $G(\mathcal{O})\backslash G(\mathcal{F})/G(\mathcal{O})$, where $G = \mathrm{GL}_h$. The *Cartan decomposition* states that these double cosets are in bijection with dominant coweights $\lambda : \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_h$, $\lambda_i \in \mathbb{Z}$, through the map that assigns a local uniformizer t to the double coset of the diagonal matrix $(t^{\lambda_1}, \ldots, t^{\lambda_h})$. Such coweights have a natural partial ordering ($\lambda \geqslant \mu \iff \lambda - \mu$ is a sum of positive coroots), and we can impose bounds on an $\overline{\mathbb{F}}$-valued shtuka, by imposing a bound on the coweight at each leg. This construction can also be made geometrically (with the coweights corresponding to strata on the affine Grassmannian, and the ordering corresponding to their closure relations), hence such bounds make sense for every test scheme S.

Remark 11.1. Let us compare this with Dieudonné modules, freely using the notation of the previous section: Let k be a perfect field in characteristic p. A free module M of rank h over $W(k)$ can be thought of as a vector bundle of rank h on the formal neighborhood of a k-valued point on a "characteristic zero" curve (e.g., on $\operatorname{spec}\mathbb{Z}$, if $k = \mathbb{F}_p$). A Dieudonné module structure consists of a semilinear "Frobenius" morphism F, such that $FM \supset pM$, see Theorem 10.11. The semilinear condition can be thought of as a linear map $M \to \sigma^* M$ (10.2). Moreover, F is an isomorphism between the isocrystal $M[\frac{1}{p}]$ and its Frobenius twist. Thus, Dieudonné modules correspond to shtukas on $\mathrm{Spf} W(k)$.

The condition $FM \supset pM$ is more specific than that: it means that the coweight associated to the Dieudonné module at its leg is of the form $p \mapsto (p, p, \cdots, p, 1, 1, \cdot, 1)$. This is a *minuscule coweight*: Its values on all root spaces are 0 or ± 1. It is the coweight corresponding to one of the fundamental representations (i.e., exterior powers of the standard representation) of the Langlands dual group (=GL_h again). If, in addition, we impose that the Dieudonné

module corresponds to a 1-dimensional formal group, i.e., $\dim_k M/FM = 1$, then the coweight is forced to be $(1, 0, 0, \ldots, 0)$, the weight of the standard representation of the dual group.

These qualifications are very useful when trying to translate between function fields and number fields or their localizations, e.g., global Shimura varieties and local Shimura "varieties" (rigid analytic spaces) require minuscule cocharacters, but locally there are now general moduli of shtukas using the technology of diamonds [SW20].

11.2 The case of GL_1

11.2.1 Picard groups and étale fundamental groups

We will be using $\mathbb{A}$ to denote the adeles of our function field $K = \mathbb{F}_q(C)$, and $[G]$ for the quotient $G(K)\backslash G(\mathbb{A})$, for any linear algebraic group G over K. By a straightforward adaptation of Proposition 9.2, the quotient $[\mathbb{G}_m]/\prod_v \mathfrak{o}_v^\times$ is the divisor class group, identified with the $\mathbb{F}$-points of the Picard group (i.e., the Jacobian) of C.

We will take advantage of the fact that the Picard group is an algebraic group. It is convenient to adopt the following moduli description of Pic_C: it represents the functor that sends an $\mathbb{F}$-scheme S to isomorphism classes of line bundles over $C \times S$ *together with an identification of their restriction to* $\{\infty\} \times S$ *with the trivial line bundle*. (This trivialization rigidifies the functor, i.e., kills automorphisms. Having done so, the bijection

$$[\mathbb{G}_m]/\prod_v \mathfrak{o}_v^\times = \mathrm{Pic}_C(\mathbb{F}).$$

is a bijection of sets, not stacks in sets, but this doesn't matter at this point.)

Now, let us recall/learn some facts about étale fundamental groups. If $\bar{s} \to X$ is a scheme together with a geometric point (i.e., $\bar{s} = \mathrm{Spec} F$, for some algebraically closed field F), then the *étale fundamental group* $\pi_1(X, \bar{s})$ is defined, as the group of *automorphisms of the fiber functor*

$$\{\text{Finite étale covers of } X\} \ni (Y \to X) \mapsto |Y_{\bar{s}} = Y \times_X \bar{s}| \in \{\text{Sets}\}$$

Ah! We should discuss Galois categories; see [Stacks Project, Tag 0BL6]. The quick way to say this is that this functor gives rise to an equivalence of categories

$$\{\text{Finite étale covers of } X\} \to \{\text{Finite sets with a continuous action of } \pi_1(X, \bar{s})\}.$$

A choice of different geometric point (even with different residue field!) gives rise to an isomorphic fundamental group, canonically up to inner automorphisms.

Now we return to the setting of the curve C over $\mathbb{F}$, with function field k. We think of it as the residue field for the generic point η of C, and of its absolute Galois group Γ_k as the étale fundamental group $\pi_1(\eta, \bar{\eta})$, where $\bar{\eta}$ corresponds to the chosen separable closure of k. There is a generalization map

$$\{\text{Finite étale covers of } C\} \to \{\text{Finite étale covers of } \eta\},$$

giving rise to a homomorphism

$$\Gamma_k = \pi_1(\eta, \bar{\eta}) \to \pi_1(C, \bar{\eta}).$$

Exercise(s) 11.1. 1. Show that this map identifies $\pi_1(C, \bar{\eta})$ as the Galois group of the *maximal everywhere unramified extension of* k. (You can consult the literature.)

If $\Gamma_k \to \mathbb{C}^\times$ is a character of the Galois group that is unramified at every place, it factors through the Galois group of the maximal unramified extension of k. Hence, at the unramified level class field theory reduces to an isomorphism

$$\widehat{\mathrm{Pic}_C(\mathbb{F})} \simeq \pi_1(C, \bar{\eta})^{ab}. \tag{11.1}$$

(Note that abelianization eliminates the dependence on base point.)

For geometric class field theory, we will actually construct isomorphisms of the character groups

$$\mathrm{Hom}(\widehat{\mathrm{Pic}_C(\mathbb{F})}, \Lambda^\times) \simeq \mathrm{Hom}(\pi_1(C), \Lambda^\times),$$

where $\Lambda = \mathbb{Z}_l$, for some auxilliary prime $l \neq p$.

11.3 The Lang isogeny; shtukas

This approach to class field theory was developed by Rosenlicht and Lang, see Serre [Ser88]. We will use it to construct the correspondence in one direction only, from characters of the Picard group to characters of the étale fundamental group. One can prove the full bijection by this method, but we will instead use Deligne's method to construct the map in the opposite direction.

To motivate the construction that follows, consider the short exact sequence

$$1 \to \pi_1(C_{\overline{\mathbb{F}_q}}, \bar{\eta}) \to \pi_1(C, \bar{\eta}) \to \mathrm{Gal}(\overline{\mathbb{F}_q}/\mathbb{F}_q) \to 0.$$

The subgroup $\pi_1(C_{\overline{\mathbb{F}_q}}, \bar{\eta})$ is called the *geometric* étale fundamental group. Having chosen an $\mathbb{F}_q$-point ∞, we get a splitting of the short exact sequence, which we can use to project the isomorphism (11.1) of class field theory to an homomorphism

$$\pi_1(C, \bar{\eta})^{ab} \twoheadrightarrow \mathrm{Pic}^0_C(\mathbb{F}).$$

This corresponds to a connected étale cover of the curve with Galois group $\mathrm{Pic}^0_C(\mathbb{F})$, and this is the cover that we are about to construct. First, some generalities.

Let G be a connected commutative group scheme over $\mathbb{F}_q$. Here we will denote by σ the q-Frobenius morphism

$$\sigma : G \to G,$$

as well as its base change to $G_{\overline{\mathbb{F}_q}}$ (the relative q-Frobenius morphism).

The *Lang isogeny* is the map

$$L : G \ni g \mapsto {}^{\sigma}g \cdot g^{-1} \in G.$$

Theorem 11.1 (Lang's theorem)**.** *Let G be a connected (not necessarily commutative) group scheme of finite type over a finite field $\mathbb{F} = \mathbb{F}_q$. (You can restrict to connected, Zariski-closed subgroups of GL_n, if you are not comfortable with general group schemes.) Let σ be the absolute q-Frobenius morphism on G, and consider the* Lang map

$$L : g \mapsto {}^{\sigma}g \cdot g^{-1}.$$

Then, the induced map on $\bar{\mathbb{F}}$-points $G(\bar{\mathbb{F}}) \to G(\bar{\mathbb{F}})$ is surjective.

Remark 11.2. For $G = \mathbb{G}_a$, the Lang map is the *Artin–Schreier map* $t \mapsto t^q - t$.

Another way to state this theorem is that the Weil group H^1 of $G(\bar{\mathbb{F}})$ is trivial; indeed, any Weil group cocycle c is determined by the image $c(\sigma)$ of Frobenius, which can be any element $x \in G(\bar{\mathbb{F}})$, without restrictions. The statement is that the coboundary action $g * x := {}^{\sigma}gxg^{-1}$ has a unique orbit.

Proof. Since this doesn't affect $\bar{\mathbb{F}}$-points, we can pass to the reduced scheme, and assume that G is reduced, hence smooth. (Every reduced group scheme over a perfect field is smooth, [Sta25, Lemma 047P].) Since G is connected, it is enough to prove that every orbit of the coboundary action is open — then there is only one orbit. This, in turn, follows from the following.

> For every $x \in G(\bar{\mathbb{F}})$ the coboundary map $g \mapsto {}^{\sigma}gxg^{-1}$ is surjective on tangent spaces: $T_1 G_{\bar{\mathbb{F}}} \to T_x G_{\bar{\mathbb{F}}}$.

Indeed, surjectivity on tangent spaces at closed points, for a morphism of nonsingular varieties, is equivalent to smoothness of the morphism [Hartshorne, Proposition 10.4]. Smooth morphisms are open, and therefore every x belongs to an open orbit.

To prove the claim, we can consider the map $G \times G \to G$ sending $(g_1, g_2) \mapsto {}^{\sigma}g_1 x g_2^{-1}$. In the g_1-variable, the map is zero on tangent spaces, while in the g_2-variable, the map is clearly an isomorphism of tangent spaces. (Note that this proves more – that the Lang map is a smooth morphism at every point.)

□

For a connected commutative G, this implies the following.

Corollary 11.2. *The map L is an isogeny with kernel $G(\mathbb{F})$, considered as a constant group scheme over $\mathbb{F}$.*

Proof. Indeed, the proof of Theorem 11.1 shows that the L is a smooth morphism. This implies, since G is connected, that it is an epimorphism. Moreover, it implies the fiber over every point is smooth; in particular, the kernel of the morphism is the smooth scheme given by the equation ${}^{\sigma}g = g$, which is the constant group scheme $G(\mathbb{F}_q)$. □

Now, we apply this to the Picard group. Note that it is not connected, and that the image of the Lang isogeny will lie in the identity (degree-zero) component Pic^0_C, while the kernel is the infinite group $\mathrm{Pic}_C(\mathbb{F})$. (But actually we will only use the restriction of the Lang map to the neutral component.)

Exercise(s) 11.2. 1. Show that the Lang map $\mathrm{Pic}_C \to \mathrm{Pic}^0_C$ admits the following moduli description: For any $\mathbb{F}$-scheme S, it sends the class of a line bundle $\mathcal{E}$ on $C \times S$ (with a trivialization over $\{\infty\} \times S$) to the class of $(Id_C \times \sigma_S)^*\mathcal{E} \otimes \mathcal{E}^{-1}$, where σ_S denotes the Frobenius morphism on S (with the induced trivialization).

Now consider the degree-zero Abel–Jacobi map $AJ_\infty : C \to \mathrm{Pic}^0_C$, which on geometric points is given by $c \mapsto (c) - (\infty)$. We form the fiber product

$$\begin{array}{ccc} \mathrm{Sht}^0 & \longrightarrow & \mathrm{Pic}^0_C \\ \downarrow & & \downarrow L \\ C & \xrightarrow{AJ_\infty} & \mathrm{Pic}^0_C\,. \end{array}$$

The left vertical arrow is a finite étale cover of C with Galois group $\mathrm{Pic}^0_C(\mathbb{F}_q)$, hence gives rise to a homomorphism

$$\varphi^0 : \pi_1(C) \to \mathrm{Pic}^0_C(\mathbb{F}_q).$$

This is almost, but not quite, the inverse reciprocity map of class field theory. Let

$$\varphi : \pi_1(C) \times_{\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})} \mathbb{Z} \to \mathrm{Pic}_C(\mathbb{F}_q) \tag{11.2}$$

be the product of φ^0 with the map that factors through $\pi_1(C) \to \mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$ and sends the Frobenius element σ to the divisor (∞).

Exercise(s) 11.3. 2. Give a moduli description for Sht^0.

3. (Important exercise!) Prove that (11.2) is compatible with local Artin maps: For any $c \in |C|$, it sends Fr_c to the class of $(c) \in \mathrm{Pic}(\mathbb{F}_q)$. In particular, it is surjective.

   Another way to formulate this compatibility is by the action of *Hecke operators, which in the abelian case come from actual maps (not more general correspondences)*: For any $c \in |C|$ and $n \in \mathbb{Z}$, let $T_{c,n} : \mathrm{Pic}_C \to \mathrm{Pic}_C$ be the *Hecke modification* sending a line bundle $\mathcal{E}$ to the line bundle $\mathcal{E}(nc)$. Then, this is compatible with the action of Fr_c^n on $\pi_1(C)$.

4. Prove that (11.2) is independent of the choice of ∞. (Here, you will need to think of $\mathrm{Pic}(\mathbb{F}_q)$ as "isomorphism classes of line bundles on C defined over $\mathbb{F}_q$", rather than "line bundles trivialized at ∞", in order for this statement to make sense.) Notice that φ^0 is then recovered from φ as $\varphi^0(\tau) = \varphi(\tau)\varphi(\mathrm{Fr}_\infty)^{-\deg(\tau)}$, where $\deg(\tau)$ is the image of τ in $\mathbb{Z}\sigma \simeq \mathbb{Z}$.

11.4 Deligne's construction

Let $\Lambda = \mathbb{Z}_\ell$, for some auxilliary prime $\ell \neq p$. (It can be replaced by the ring of integers in any finite extension of $\mathbb{Q}_\ell$.) The inverse reciprocity map constructed in the previous section gives rise to a map

$$\mathrm{Hom}(\widehat{\mathrm{Pic}_C(\mathbb{F})}, \Lambda^\times) \to \mathrm{Hom}(\pi_1(C, \bar{\eta}), \Lambda^\times). \tag{11.3}$$

Because (11.2) is surjective, this map is injective. Here we will construct the inverse map, showing that it is bijective. For that purpose, we interpret

$$\mathrm{Hom}(\pi_1(C, \bar{\eta}), \Lambda^\times)$$

as $\mathrm{Loc}_{1,C}$:= *isomorphism classes of rank-one étale local systems $\mathcal{L}$ on C* (=locally constant sheaves for the étale topology), together with a trivialization $\bar{\eta}^*\mathcal{L} \simeq \Lambda$.

[Some technicalities here: the Λ-adic local system is really defined as a limit over all powers of ℓ of Λ/ℓ^j-étale local systems.]

On the other hand, the other side

$$\mathrm{Hom}(\widehat{\mathrm{Pic}_C(\mathbb{F})}, \Lambda^\times)$$

will be interpreted as *character local systems* on Pic_C.

Definition 11.2. Let $G/\mathbb{F}$ be a commutative (smooth, finite type) algebraic group. A *character local system* on G is given by pair $(\mathcal{L}, \psi)$ where $\mathcal{L} \in \mathrm{Loc}_{1,G}$, and

$$\psi : m^*\mathcal{L} \simeq p_1^*\mathcal{L} \otimes p_2^*\mathcal{L}$$

is an isomorphism on $G \times G$ that satisfies the cocycle condition. Here, $m : G \times G \to G$ is the multiplication map, while the p_i's are the projection maps.

The idea behind character local systems is that the trace of Frobenius at the stalks of such a local system gives rise to an actual character $G(\mathbb{F}) \to \Lambda^\times$, and the same is true simultaneously for all finite extensions of $\mathbb{F}$, by taking powers of the Frobenius morphism. Character local systems form a group under tensor product. We let $\mathrm{CharLoc}(G)$ denote the group of character local systems (defined over $\mathbb{F}$) on G. [Obviously, there is more structure here than just sets, but we will stick to sets.]

Exercise(s) 11.4. 1. For any $\chi \in \mathrm{Hom}(G(\mathbb{F}), \Lambda^\times)$, construct a local system $\mathcal{L}_\chi$ on G by reduction of the Lang isogeny $G \to G$ via χ, i.e., the associated $\Lambda^\times$-torsor over G is[33] $G \times^{(G(\mathbb{F}_q),\chi)} \Lambda^\times$. Explain why this is naturally a character local system, and prove that the "Frobenius trace" map gives an inverse to this construction, giving rise to an isomorphism of groups

$$\mathrm{CharLoc}(G) \to \mathrm{Hom}(\widehat{G(\mathbb{F})}, \Lambda^\times).$$

Remark 11.3. This exercise points to something amazing, which is a basic theme of all geometric representation theory: there are "sheaves" (local systems) whose Frobenius traces produce characters on $G(\mathbb{F}')$ for *all* finite extensions $\mathbb{F}'/\mathbb{F}$, and one obtains all characters of those finite groups in this manner.

Thus, the isomorphism of class field theory can now be reformulated as an isomorphism

$$\mathrm{CharLoc}(\mathrm{Pic}_C) \xrightarrow{\sim} \mathrm{Loc}_C^1 . \tag{11.4}$$

[33] This requires sheafification modulo finite quotients of $\mathbb{Z}_l$; the notation $G \times^{(G(\mathbb{F}_q),\chi)} \Lambda^\times$ stands for the limit over all powers of ℓ of the sheaf associated to the presheaf sending an étale cover $U \to G$ to $\Gamma(U, G) \times^{(G(\mathbb{F}),\chi)} (\Lambda/\ell^j)^\times$, where $\Gamma(U, G)$ is the set of sections $U \to (G \times_G U)$, where the fiber product is taken with respect to the Lang map.

Theorem 11.3. *The* unbased *Abel–Jacobi map*

$$AJ : C \ni c \mapsto (c) \in \mathrm{Pic}_C$$

induces, by pullback, an isomorphism (11.4).

Exercise(s) 11.5. 2. Unfold the definitions to make sure that this is the same map as (11.3). Thus, we already know that it is injective.

The following should look familiar, if you have ever seen the construction of the Picard scheme (see Milne, §III.1–5): For any $d \geqslant 0$, let $C^{(d)}$ be d-th symmetric power of the curve, classifying degree-d divisors on C. (It is a smooth algebraic curve, identified with the invariant-theoretic quotient C^d/S_d; it can be covered by Zariski open sets $U^{(d)} = \mathrm{Spec}(\mathbb{F}[U]^{\otimes d})^{S_d}$, where U ranges over open affine subsets of C.)

The Abel–Jacobi map extends naturally to a map $AJ : C^{(d)} \to \mathrm{Pic}^d_C$, and its fiber over a line bundle $\mathcal{E}$ can be identified with the projectivization of $H^0(C, \mathcal{E})$. When $d > 2g-2$, the Riemann–Roch theorem implies that this fiber is isomorphic to $\mathbb{P}^{d-g+1}$, for any $\mathcal{E}$.

Let $\bar{s} \to C^{(d)}_{\bar{\eta}}$ be a geometric point lifting the given geometric point of $C^{(d)}$ (induced by our choice of geometric point on C); there is a homotopy exact sequence, see [Stacks Project, Tag 0BUM]

$$\pi_1(C^{(d)}_{\bar{\eta}}, \bar{s}) \to \pi_1(C^{(d)}, \bar{s}) \to \pi_1(\mathrm{Pic}^d_C, \bar{\eta}) \to 0,$$

which, since projective spaces are simply connected, shows that

$$\pi_1(C^{(d)}, \bar{s}) \simeq \pi_1(\mathrm{Pic}^d_C, \bar{\eta})$$

(when $d > 2g-2$). Thus, local systems on Pic^d_C and on $C^{(d)}$ are in bijection (under pullback).

Given, now, a 1-dimensional étale local system $\mathcal{L}$ on C, we construct the restriction of a character local system to Pic_C, as follows:

Exercise(s) 11.6. 3. Consider the local system $\mathcal{L} \boxtimes \cdots \boxtimes \mathcal{L}$ on C^d; show that it is the pullback of a local system $\mathcal{L}^{(d)}$ on $C^{(d)}$. (Hint: Use its S_d-equivariance — but that won't be quite enough.)

4. Now, for $d > 2g-2$, let $r(\mathcal{L})$ be the corresponding local system on Pic^d_C. For $c \in |C|$, consider the map $T_c : \mathrm{Pic}_C \to \mathrm{Pic}_C$ sending $\mathcal{E} \mapsto \mathcal{E}(c)$. Show that there is a canonical isomorphism of local systems

$$T_c^* r(\mathcal{L})|_{\mathrm{Pic}^d_C} = r(\mathcal{L})|_{\mathrm{Pic}^d_C} \otimes \mathcal{L}_c, \tag{11.5}$$

where $\mathcal{L}_c$ is the stalk of $\mathcal{L}$ at c.

5. Using the previous exercise, prove that there is a unique extension and "upgrade" of $r(\mathcal{L})$ to a character sheaf on Pic_C, satisfying the "Hecke eigensheaf" property (11.5). Show that its pullback via the Abel–Jacobi map recovers $\mathcal{L}$, giving rise to the bijection (11.4).

12 The Waldspurger and Gross–Zagier theorems – an overview

The main reference for this section is [YZZ13]; although we will not follow their proof in later sections, they include clean and precise formulations of the theorems. We will be referring as the "Gross–Zagier formula" for the more general version that is due to Yuan–Zhang–Zhang, proven in this book.

12.1 The pairings

The second part of the course will cover the formulas of Waldspurger and Gross–Zagier. Both have to do with the diagonal embedding $T \hookrightarrow T \times G$, where $T = \mathrm{Res}_{K/\mathbb{Q}}(\mathbb{G}_m)$ (K an imaginary quadratic field), and $G = \mathrm{GL}_2$, or, more canonically, the algebraic group corresponding to $\mathrm{GL}_{\mathbb{Q}}(K)$. More precisely, there is a $\mathbb{G}_m \subset T$ (corresponding to the embedding $\mathbb{Q}^\times \subset K^\times$) which maps into the center of $T \times G$, and we should really be thinking about the quotient by that,

$$U_1 = T/\mathbb{G}_m \hookrightarrow \tilde{G} := (T \times G)/\mathbb{G}_m. \tag{12.1}$$

(The algebraic group $T/\mathbb{G}_m$ is isomorphic to the unitary group U_1 associated to the extension $K/\mathbb{Q}$.)

Let $[G]$ denote the adele class quotient $G(\mathbb{Q})\backslash G(\mathbb{A})$, and simiilarly for T. Let S_G, S_T denote the Shimura varieties of G and T, so, at the level of points,

$$S_T = T(\mathbb{Q})\backslash T(\mathbb{A}_f) \text{ (a finite set)}, \ S_G = G(\mathbb{Q})\backslash (G(\mathbb{A}_f) \times \mathcal{H}^{\pm}).$$

12.1.1 The Waldspurger pairing

(Automorphic representations jargon alert. We won't cover these notions, but we'll later specialize to the representations generated by holomorphic modular forms of weight 2, at least for the Gross–Zagier formula.)

The embedding (12.1) induces an embedding $[T/\mathbb{G}_m] \hookrightarrow [\tilde{G}]$, through which we can take the Haar measure on $[T/\mathbb{G}_m]$ (the choice of normalization – Tamagawa measure – to be discussed later) to a distribution $\delta_{[T]}$ on $\tilde{G}$. The Waldspurger formula states, roughly, that, for "any" automorphic representation $\chi \otimes \pi$ of $\tilde{G}$, that is, for any idele class character χ of T and an automorphic representation π of G satisfying $\chi \cdot \omega_\pi = 1$ (where ω_π is the central character of π), the $\chi \boxtimes \pi$-isotypic component of the inner product

$$\langle \delta_{[T]}, \delta_{[T]} \rangle_{[T \times G]} \tag{12.2}$$

is "proportional" to a certain central L-value $L(\chi \times \pi, \frac{1}{2})$ (see § 12.2 for the definition of the L-function). Here, the inner product is not really defined, because these are distributions, but we can formally think of its $\chi \boxtimes \pi$-isotypic component as the sum, over an orthonormal basis consisting of automorphic forms ϕ, of

$$\langle \delta_{[T]}, \phi \rangle \cdot \langle \phi, \delta_{[T]} \rangle = \left| \int_{[T/\mathbb{G}_m]} \chi(t)\phi(t)dt \right|^2 . \tag{12.3}$$

This sum is still divergent, but the actual formula will be a relation between the hermitian forms (12.3) and the L-value.

We really don't mean "any" automorphic representation here, but those that belong (weakly) in $L^2([\tilde{G}])$, such as cuspidal representations with unitary central character, so that we can talk about inner products; similarly, χ should be unitary. (Moreover, the formula doesn't quite apply to 1-dimensional representations, i.e., automorphic characters of G.) For such representations, the complex conjugate $\bar{\pi}$ can be identified with the smooth dual $\tilde{\pi}$ through the L^2-pairing, and the Hermitian form (12.3) can be expanded to a bilinear pairing

$$B_W : (\chi \boxtimes \pi) \otimes (\chi^{-1} \boxtimes \tilde{\pi}) \to \mathbb{C}. \tag{12.4}$$

The main property of the Waldspurger pairing B_W is that *it is invariant, in both arguments, under the adelic points of $T/\mathbb{G}_m$ embedded as in* (12.1).

12.1.2 The Gross–Zagier pairing

Similarly, we have an embedding $S_T \hookrightarrow S_G$, which, as we saw in § 9.3, depends on the choice of special point. This gives rise to an embedding $S_T \hookrightarrow S_T \times S_G$. Recall that S_T is zero-dimensional. Replacing S_G by its compactification $\overline{S_G}$ (=the inverse limit of compactified modular curves $X(N)$), we can consider the Néron–Tate height pairings on points (extended bilinearly to divisors). The Gross–Zagier formula states, roughly, that the $\chi \boxtimes \pi$-isotypic component of the pairing

$$\langle S_T, S_T \rangle_{NT} \tag{12.5}$$

is "proportional" to the central derivative $L'(\chi \times \pi, \frac{1}{2})$ of the same L-function as before. The representation π is restricted, here, to the ones corresponding to holomorphic modular forms of weight 2, and χ is restricted here to a character of the Galois group[34] of K^{ab}/K, that is, $\chi_\infty = 1$. By "isotypic component" we mean isotypic under the actions of the groups of finite adeles, $T(\mathbb{A}_f) \times G(\mathbb{A}_f)$.

[34]Here, we will use the normalization of class field theory that sends uniformizers to *geometric* Frobenii, so that the Hecke and Galois action on S_T are compatible by CM theory.

Explicitly, this means the following: Recall, from Theorem 6.4, that the Jacobian at a finite level has an isotypic decomposition (in the isogeny category) in terms of abelian varieties indexed[35] by Galois orbits of holomorphic weight-2 normalized newforms – which is the same as Galois orbits of the $G(\mathbb{A}_f)$-representations generated by those forms. (Galois orbits, here, refers to the Galois action on the Fourier coefficients of the eigenforms.) We can then average the points of S_T against the character χ to get a "divisor" on $\overline{S_G}$,

$$S_T(\chi) := \int_{S_T} \chi(\tau) \cdot [\tau] d\tau,$$

that is, a compatible sequence of divisors in the tower of compactified modular curves. Here, the integral over S_T means averaging, i.e., at every finite level $X(N)$ (through which the character χ factors) we are averaging over the finite image of S_T. This divisor is defined over the abelian closure of K (since, by CM theory, the points of S_T are), and has coefficients in whichever number field L the character χ is valued. We then apply "the" Abel–Jacobi map (*normalized by sending a specific divisor, that we will not describe yet, to zero*), to get a "*point*" $P(\chi) := AJ(S_T)(\chi) \in J(K^{\mathrm{ab}}) \otimes_{\mathbb{Z}} L$ (i.e., a family of points, compatible under pushforwards, in the Jacobians at finite level), and project to the π-isotypic component, to get a "point"

$$P(\chi \boxtimes \pi) \in J(K^{\mathrm{ab}}) \otimes_{\mathbb{Z}} L,$$

where we may need to enlarge L, so that it contains the coefficient field of π.

The Gross–Zagier theorem is about the Néron–Tate height pairing

$$\langle P(\chi \boxtimes \pi), P(\chi^{-1} \boxtimes \tilde{\pi}) \rangle_{NT(J)},$$

which again is a formal thing, which has to be projected to finite level in order to make sense of it. More precisely, we can again define a bilinear form between π and $\tilde{\pi}$ (or rather, their "finite" components), as follows: If we fix a representative A_π for the isomorphism class (in the isogeny category of abelian varieties over $\mathbb{Q}$) of the abelian variety corresponding to π, we set

$$\pi_f = \mathrm{Hom}^0_{\mathbb{Q}}(J, A_\pi)$$

[35] Theorem 6.4 was stated for the congruence subgroups $\Gamma_0(N)$, but similar statements hold for other congruence subgroups like $\Gamma(N)$ – in fact, we can dispense with the word "newform," and just consider eigenforms up to the equivalence relation of having the same eigenvalues for almost all Hecke operators T_p. Passing to congruence subgroups like $\Gamma(N)$ – or at least $\Gamma_1(N)$; see Remark 5.1 – is important because their intersection over all N's is the identity, leading to a presentation of the Shimura variety S_G as $\lim_N Y(N)$.

(homomorphisms defined over $\mathbb{Q}$, but in the isogeny category, i.e., $\mathrm{Hom}^0 = \mathrm{Hom} \otimes_{\mathbb{Z}} \mathbb{Q}$). This is a representation of $G(\mathbb{A}_f)$ with coefficients the field $M = \mathrm{End}^0_{\mathbb{Q}}(A)$, which is of dimension equal to $\dim A_f$ over $\mathbb{Q}$, see [YZZ13, 3.2.1].[36] Choosing an embedding $\iota : M \hookrightarrow \mathbb{C}$ and extending scalars to $\mathbb{C}$, we have an isomorphism

$$\pi_f \otimes_M \pi_\infty \simeq \pi^\iota,$$

where π^ι is the representation generated by a Galois conjugate of f. To show this, one needs to appeal to the multiplicity-one statement of Theorem 5.15; see [YZZ13, 3.2.3]. (Note that A_f is determined by the Galois conjugacy class of f, not f itself.) We fix the embedding ι from now on, and without loss of generality assume that $\pi^\iota = \pi$; the Gross–Zagier theorem holds for any such choice of ι. The (smooth) dual of π_f is defined similarly, using the dual abelian variety $A_f^\vee$. The pairing $\pi_f \otimes \tilde{\pi}_f \to \mathbb{C}$ is given by

$$\langle F, \tilde{F} \rangle = \mathrm{Vol}(X(N))^{-1} (F_N \circ F_N^\vee), \tag{12.6}$$

where N is large enough so that the maps F, $F^\vee$ factor through $X(N)$. The pairing is valued in M, and then we compose with an embedding $M \hookrightarrow \mathbb{C}$ to make it $\mathbb{C}$-valued.

We then define a pairing

$$B_{GZ} : (\chi \boxtimes \pi_f) \otimes (\chi^{-1} \boxtimes \tilde{\pi}_f) \to \mathbb{C}, \tag{12.7}$$

by

$$F \otimes \tilde{F} \mapsto \langle F(P(\chi)), \tilde{F}(P(\chi^{-1})) \rangle_{NT(A_\pi)}, \tag{12.8}$$

where the height pairing now is on $A(K^{\mathrm{ab}}) \otimes_{\mathbb{Z}} L$. As with the Waldspurger pairing, the Gross–Zagier pairing B_{GZ} is *invariant, in both arguments, under the finite adelic points of $T/\mathbb{G}_m$ embedded as in* (12.1).

There will be a lot of things to be straightened out about the rough statements above, besides the need to translate "proportional" into a precise equality. But let us first discuss the L-functions that appeared in the statements.

12.2 L-functions and ϵ-factors (root numbers)

Both formulas concern an L-function, whereby, in this course, we will mean an assignment

$$\text{Automorphic representation } \pi \mapsto \text{Meromorphic function } L(\pi, r, s),$$

[36] Notice that, even in the case of CM elliptic curves, since we are taking automorphisms defined over $\mathbb{Q}$, we have $M = \mathbb{Q}$.

where $L(\pi, r, s)$ is an *automorphic L-function* as defined by Langlands. That is, if π is an automorphic representation for a group G, r is an algebraic (complex) representation of the L-group, ${}^L G \to \mathrm{GL}(V)$, and $L(\pi, r, s)$ is defined by an Euler product, whose factors at places v where the local component π_v of π is unramified are $L(\pi_v, r, s) = \det(I - q_v^{-s} c(\pi_v))^{-1}$, where $c(\pi_v)$ is the *Satake parameter* of π_v (a conjugacy class in ${}^L G$). For the full definition, we need to assume the local Langlands correspondence, which assigns to every π_v a homomorphism $\phi_v : \mathcal{W}_v \to {}^L G$ (where $\mathcal{W}_v$ is the local Weil group at v), to define, at nonarchimedean places,

$$L(\pi_v, r, s) = \det(I - q_v^{-s} \phi_v(\mathrm{Fr}_v)|_{V^{\phi_v(I_v)}})^{-1},$$

i.e., as a local Artin L-function associated to the representation $r \circ \phi_v$. We skip the definition at archimedean places, for now, but emphasize that in these notes L will stand for the *completed* L-function.

Example 12.1. The basic example, at least for the Gross–Zagier formula is that of weight-2 modular forms (with trivial nebentypus). Their local unramified L-factor is given by the inverse of $L_p(f, s)^{-1} = (1 - p^{-s} a_p(f) + p^{1-2s})$, with the functional equation is centered at $s = 1$. This can be factored $(1 - p^{\frac{1}{2}-s}\alpha_p)(1 - p^{\frac{1}{2}-s}\alpha_p^{-1})$. The Satake parameter here is the diagonal element $(\alpha_p, \alpha_p^{-1})$, and the "automorphic" parametrization of the L-function is $L(\pi, s) = L(f, s + \frac{1}{2})$, centering the functional equation at $\frac{1}{2}$.

For Gross–Prasad and Waldspurger, the automorphic representation is on the group $\tilde{G} = (T \times G)/\mathbb{G}_m$, whose L-group can be identified with ${}^L\tilde{G} = \mathrm{GO}_2 \times_{\mathbb{G}_m} \mathrm{GL}_2$, where the fiber product is taken over the determinant in one copy and the *inverse* of the determinant in the other, and r is the tensor product of the standard representations of GO_2 and GL_2, hence a degree-4 L-function. If $\chi \boxtimes \pi$ is the automorphic representation, we can denote this L-function by $L(\chi \times \pi, s)$, although a more customary notation is $L(\chi \otimes \Pi, s)$, where Π stands for the *base change* of Π from $\mathbb{Q}$ to K. For the purposes of this definition, the base change is an entirely formal thing, and we don't need to know the actual existence of an automorphic base change to an automorphic representation of $\mathrm{GL}_2(\mathbb{A}_K)$ (even though it is known by the work of Langlands). Simply, we restrict the local Langlands parameters ϕ_v of the automorphic representation to the decomposition groups of the Galois group of $\overline{\mathbb{Q}}/K$, and form the corresponding L-functions there, regarding now χ as an automorphic representation of the group $\mathbb{G}_m$ over K (i.e., as an idele class character of K).

To be clear, no local Langlands correspondence will explicitly appear anywhere in the story. We will have "working definitions" of these L-functions, which, independently and in ways that will not concern us in this course, have been checked to coincide with local Langlands.

When the character χ of T is trivial, this L-function is simply

$$L(\Pi, s) = L(\pi, s) \otimes L(\pi \otimes \eta, s), \tag{12.9}$$

where η is the quadratic character associated to $K/\mathbb{Q}$, and, when no r appears, we mean the standard L-function (i.e., the L-function associated to the standard, 2-dimensional representation of G).

It is important to observe that the tensor product representation of

$${}^L\tilde{G} = \mathrm{GO}_2 \times_{\mathbb{G}_m} \mathrm{GL}_2$$

is *symplectic*, and the functional equation relates the aforementioned (complete) L-function to itself,

$$L(\chi \times \pi, s) = \epsilon(\chi \times \pi, s) L(\chi \times \pi, 1 - s),$$

where the (global) *epsilon factor* is the product of some constant $\epsilon = \epsilon(\chi \times \pi, \frac{1}{2})$ by an exponential in $\frac{1}{2} - s$. The representation being symplectic, the number ϵ is known to be ± 1, and it is called the (global) *root number* of $\chi \times \pi$. Root numbers will play an important role in the precise formulation of the theorems.

12.3 Local obstructions; the theorem of Tunnel and Saito

The theorems of Waldspurger and Gross–Zagier actually apply to complementary situations, or rather are "trivial" (equalities of two sides which are known a priori to be zero) in complementary situations: The theorem of Waldspurger is trivial when $\epsilon = -1$, and Gross–Zagier is trivial when $\epsilon = 1$. The fact that the corresponding L-value, respectively derivative, is zero in this setting is obvious from the functional equation; the fact that the other side is zero is deeper, but due to local, representation-theoretic, not arithmetic reasons, as we will explain now.

The global root number is a product of local root numbers, $\epsilon = \prod_v \epsilon_v$, also equal to ± 1, and almost all equal to 1. (We will be referring, from now on, to general number fields F, for statements that apply in this generality; the field K can be any quadratic extension of F. The theorem of Waldspurger is as general as this, and the Gross–Zagier theorem generalizes to the setting where F is totally real and K is a CM extension.) The local root numbers determine whether

$$\mathrm{Hom}_{T(F_v)}(\chi_v \boxtimes \pi_v, \mathbb{C})$$

is nonzero or not: By a theorem of Tunnel and Saito, it is nonzero iff

$$\epsilon_v = \chi_v(-1)\eta_v(-1). \tag{12.10}$$

This explains why the equalities of Waldspurger and Gross–Zagier are trivial when the global root number is -1, respectively $+1$, but it actually shows that the left hand side is "trivially" zero in many more cases – whenever (12.10) does not hold at one place. Surely, the nonvanishing of the L-value or its derivative, as long as the *global* root number is right, cannot be dictated by the *local* root numbers. Therefore, the theorems as sketched cannot be true – some modification is needed.

The modification needed has to do with the fact that, when $\mathrm{Hom}_{T(F_v)}(\chi_v \boxtimes \pi_v, \mathbb{C}) = 0$, the corresponding Hom-space for a "pure inner form" of the problem is nonzero. More precisely, we can replace G_v by $G'_v =$ the multiplicative group of the quaternion division algebra over F_v, in which T_v still embeds. We also need to replace π_v by $\pi'_v =$ the Jacquet–Langlands lift of π_v to G'_v; this is a representation of $G'_v(F_v)$ with the same L- and ϵ-factors. The conditions for nonvanishing of the Hom, in the case of G'_v, are complementary to those of G_v.

Thus, *starting* from an automorphic representation $\chi \boxtimes \pi$, we have its local root numbers ϵ_v, and those indicate an inner form G'_v of G_v at every place for which the local Hom-spaces are $\neq 0$. We can hope to have valid versions of the theorems of Waldspurger and Gross–Zagier if we formulate their analogs for an inner form G' of G, whose localizations are G'_v.

When $\epsilon = 1$, this works: Since $\prod_v \chi_v(-1)\eta_v(-1) = 1$, this implies that $\epsilon_v = -\chi_v(-1)\eta_v(-1)$ at an *even* set Σ of places, and there is a unique quaternion algebra D_Σ over F that is ramified at precisely those places. The theorem of Waldspurger holds in this setting (i.e., for $G' = D_\Sigma^\times$), while when $\epsilon = -1$ it holds trivially for *any* quaternion algebra, as an equality $0 = 0$.

For the Gross–Zagier theorem, we observe that we don't need a $T(\mathbb{R})$ invariant functional on π_∞ – this representation does not appear in the Gross–Zagier pairing (12.7). On the other hand, $\chi_\infty = 1$ and $\eta_\infty(-1) = -1$, since K is imaginary. It is known that $\epsilon_\infty = \epsilon(\pi_\infty, \frac{1}{2})\epsilon(\pi_\infty \otimes \eta_\infty, \frac{1}{2})$ is -1 in this case. Therefore, when the global root number ϵ is 1, there is an odd, hence nontrivial, set of finite primes p where condition (12.10) fails, and the two sides of the Gross–Zagier theorem are "trivially" zero. When $\epsilon = -1$, there is an *even* set Σ of finite places where (12.10) fails, and, as above, we can replace G by $G' = D_\Sigma^\times$, as above, in order to get the correct theorem. In this process, the tower of modular curves/Shimura variety S_G will be replaced by the Shimura variety $S_{G'}$ of G', which is a tower of *Shimura curves* – their complex points are of the form $\Gamma\backslash\mathcal{H}$ for Γ a congruence subgroup *for* G', and they are compact.

They admit a modular interpretation in terms of abelian surfaces with "quaternionic multiplication" by an order in D (e.g., see Chao Li's notes), which can be understood in terms of an embedding $G' \hookrightarrow \mathrm{Sp}_4$ and the moduli interpretation of quotients of the Siegel upper half space by congruence subgroups (generalizing the case of elliptic curves to higher-rank abelian varieties).

12.4 Formulation of the Waldspurger and Gross–Zagier theorems

Fix an automorphic representation $\chi \boxtimes \pi$ of $\tilde{G}$, which in the Gross–Zagier case will be required, at infinity, to be $\chi_\infty = 1$ and $\pi_\infty =$ weight-2 discrete series.

Form the set Σ as above – hence, for Waldspurger's formula when $\epsilon = \epsilon(\chi \boxtimes \pi, \frac{1}{2}) = 1$, Σ is the set of all places where (12.10) fails, while for the Gross–Zagier formula when $\epsilon = -1$ it is the set of *finite* places where it fails. In both cases, the set Σ contains an even number of places, and we let G' be the multiplicative group of the quaternion algebra that is ramified precisely at Σ. We let $\pi' =$ the Jacquet–Langlands lift of π to G' (= the unique automorphic representation of G' with the same Hecke eigenvalues as π, for all finite places $p \notin \Sigma$). And similarly define, in the Gross–Zagier case,

$$\pi_f = \mathrm{Hom}^0_{\mathbb{Q}}(J', A_\pi),$$

where J'_A is the Jacobian of the Shimura variety for G' (compactified, if $G = G'$).

Fix a factorization $\pi' = \bigotimes'_v \pi'_v$, where π_v is a unitary irreducible representation of $G'(\mathbb{Q}_v)$. The factorization is required to be compatible with inner products, where we fix the inner product on $L^2([G'/\mathbb{G}_m])$ with respect to *Tamagawa measure*, which assigns total volume 2 to $[G'/\mathbb{G}_m]$. We also endow $[U_1] = [T/\mathbb{G}_m]$ with Tamagawa measure, giving it total mass 1, and we factorize the corresponding Haar measure dt on $U_1(\mathbb{A})$ as a product $\prod_v dt_v$ of local Haar measures on $U_1(\mathbb{Q}_v)$ so that $dt_p(U_1(\mathbb{Z}_p)) = 1$ for almost every p.

We choose a global embedding $T \hookrightarrow G'$, and define $T(\mathbb{Q}_v) \times T(\mathbb{Q}_v)$-invariant bilinear forms

$$\alpha'_v : (\chi_v \boxtimes \pi'_v) \otimes (\chi_v^{-1} \boxtimes \tilde{\pi}_v) \to \mathbb{C}$$

by

$$\alpha'(u \otimes \tilde{u}) = \int_{U_1(\mathbb{Q}_v)} \langle \pi_v(t)u, \tilde{u} \rangle \chi^{-1}(t) dt.$$

We have the following *fact*: At split nonarchimedean places (i.e., $G'_v \simeq G_v \simeq \mathrm{GL}_2$), if K, χ_v and π_v are unramified at v and we fix an integral model

$G'_v \simeq \mathrm{GL}_2$ over $\mathbb{Z}_v$ so that the embedding $T_v \hookrightarrow G'_v$ corresponds to a basis of $K \otimes \mathbb{Z}_v$ over $\mathbb{Z}_v$, and if $dt_v(U_1(\mathbb{Z}_v)) = 1$, then

$$\alpha'_v(u \otimes \tilde{u}) = \frac{\zeta_v(2) L(\chi_v \times \pi_v, \frac{1}{2})}{L(\eta_v, 1) L(\pi_v, \mathrm{Ad}, 1)} =: c_v^{-1},$$

where "Ad" denotes the adjoint representation of the dual group on $\mathfrak{pgl}_2$. We then define a normalized functional (at *every* place v) by

$$\alpha_v = c_v \alpha'_v.$$

Theorem 12.1 (Waldspurger's theorem)**.** *In the setting above, with $\epsilon = 1$, the Waldspurger pairing (12.4), defined (completely analogously) for G', admits an Euler product*

$$B_W = \frac{\zeta(2) L(\chi \times \pi, \frac{1}{2})}{8 L(\eta, 1)^2 L(\pi, \mathrm{Ad}, 1)} \prod_v \alpha_v.$$

All L-functions (including ζ), here, are completed L-functions (i.e., include the archimedean factor).

I think there's something wrong with the way that "Tamagawa measure" is used in [YZZ13]; to double-check the precise factors.

Remarks 12.1.

1. It may be hard at first to digest all the L-values that appear above, but the most important ones are the ones that vary with π (and χ), namely, the central L-value $L(\chi \times \pi, \frac{1}{2})$ in the numerator, and the adjoint L-value $L(\pi, \mathrm{Ad}, 1)$ in the denominator. The factor $L(\eta, 1)$ can also be seen as an adjoint L-value (for the action of the dual group of T on $\mathfrak{u}_1$), and the rest of the factors can be seen as being related to our choices of measures.
2. Note that we took the trouble to normalize the factors α_v, so that the Euler product is actually finite, i.e., almost all factors (when evaluating on a pair of vectors $u \otimes \tilde{u} \in \pi \otimes \tilde{\pi}$) are equal to 1. But a cleaner way to understand the formula is as a regularized Euler product

$$B_W = \frac{1}{8 L(1, \eta)} \prod_v{}^{*} \alpha'_v.$$

The remaining factor of $1/8L(1, \eta)$ can also be understood (and eliminated, by an appropriate reformulation), but we won't get into that.

Similarly, to formulate the theorem of Gross–Zagier (as extended by Yuan–Zhang–Zhang) when $\epsilon = 1$, first we fix factorizations

$$\pi'_f = \bigotimes_p \pi'_p \quad \tilde{\pi}'_f = \bigotimes_p \tilde{\pi}'_p$$

(with the product over finite places only), compatible with the duality pairing (12.6).

Theorem 12.2 (Gross–Zagier–Yuan–Zhang–Zhang theorem)**.** *In the setting above, with $\epsilon = -1$, the Gross–Zagier pairing* (12.4)*, defined (completely analogously) for G', admits an Euler product*

$$B_{GZ} = \frac{\zeta^{\infty}(2) L^{\infty\prime}(\chi \times \pi, \frac{1}{2})}{4 L^{\infty}(\eta, 1)^2 L^{\infty}(\pi, \mathrm{Ad}, 1)} \prod_{p<\infty} \alpha_p,$$

with the same local functionals α_v as above. Here, L^{∞} denotes the L-functions without their archimedean factors.

13 L-factors and ϵ-factors

This section introduces the theory of integral representations of L-functions, through the examples that are relevant to the Waldspurger and Gross–Zagier formulas. It is written in the form of exercises, that will be given as an assignment. The section assumes familiarity with Iwasawa–Tate theory. We fix a number field F, and denote by $\mathbb{A}$ its adeles.

A good, straightforward reference is [Cog03].

Exercise(s) 13.1. Consider the space $M_k(\Gamma(N), \chi)$ of holomorphic modular forms of weight k with Nebentypus χ. Show that the map

$$f \mapsto F(g) := (f|_k g)(i) = (ad - bc)^{k/2}(ci + d)^{-k} f(gi)$$

defines an embedding into $C^\infty((\Gamma(N), \chi)\backslash \mathrm{GL}_2(\mathbb{R})^+)$, the space of smooth functions on $\mathrm{GL}_2(\mathbb{R})^+$ that vary by the character χ of $\Gamma(N)$. Explain how one can further embed the latter into $C^\infty(\mathrm{GL}_2(\mathbb{Q})\backslash \mathrm{GL}_2(\mathbb{A}))$, obtaining the automorphic form associated to f.

13.1 The Hecke integral and its unfolding

For an automorphic representation π on GL_2, the *Hecke integral* (as reinterpreted adelically by Jacquet–Langlands) is the functional

$$\pi \to \mathbb{C}$$

given by

$$I(\phi, s) := \int_{[\mathbb{G}_m]} \phi \begin{pmatrix} x & \\ & 1 \end{pmatrix} |x|^s d^\times x. \tag{13.1}$$

We'll omit discussing convergence of integrals – "whenever they make sense." Typically, convergence issues are easy for cusp forms, but some regularization is required for Eisenstein series.

Exercise(s) 13.2. Show that, appropriate choices of measures and up to a shift in the s-parameter, for holomorphic modular forms the integral above translates to the integral denoted by L_f in (2.5).

We will relate the Hecke integral (13.1) to the (standard) L-function of π, and discuss its functional equation, generalizing the discussion of § 2.1.

As with holomorphic modular forms, we can use the Fourier expansion of ϕ in order to write the Hecke integral in terms of the Fourier coefficients of ϕ; here, the precise generalization of the q-expansion is the Fourier transform

with respect to the upper unipotent subgroup $N = \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix}$. We need to recall from Tate's thesis that, after fixing a single nontrivial character of the compact group of adele classes,

$$\psi : [\mathbb{G}_a] = F\backslash\mathbb{A} \to \mathbb{C}^\times,$$

the Pontryagin dual of $[\mathbb{G}_a]$ is F via the map $f \ni \gamma \mapsto \psi_\gamma(x) := \psi(\gamma x)$. Thus, we have a Fourier expansion

$$\phi(g) = \sum_{\gamma \in F} W_{\phi,\gamma}(g),$$

where $W_{\phi,\gamma}(g)$ is the Fourier coefficient

$$W_{\phi,\gamma}(g) = \int_{[\mathbb{G}_a]} \phi\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g\right) \psi^{-1}(\gamma x) dx.$$

(Probability Haar measure on $[\mathbb{G}_a]$.)

Assume, now, that ϕ is a cusp form. Then there is no zeroth Fourier coefficient. The others can be expressed in terms of the first Fourier coefficient, by exploiting automorphicity, as follows:

$$\begin{aligned} W_{\phi,\gamma}(g) &= \int_{[\mathbb{G}_a]} \phi\left(\begin{pmatrix} \gamma^{-1} & \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g\right) \psi^{-1}(x) dx \\ &= \int_{[\mathbb{G}_a]} \phi\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g\right) \psi^{-1}(x) dx = W_{\phi,1}\left(\begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g\right). \end{aligned}$$

The function $W_\phi(g) := W_{\phi,1}(g)$ is also called the *Whittaker function* of ϕ. It satisfies the transformation property $W_\phi(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g) = \psi(x) W_\phi(g)$, and therefore the map $\phi \to W_\phi$ defines a map

$$\pi \to C^\infty((N,\psi)\backslash G(\mathbb{A})).$$

Important fact (proof omitted):

Proposition 13.1. *For every place v, every irreducible (admissible) representation π_v of $G(F_v)$ admits at most* one, up to multiple *embedding into $C^\infty((N,\psi)\backslash G(F_v))$. Its image $\mathcal{W}(\pi_v)$ is called the* Whittaker model *of π_v. If a representation admits a (nonzero) Whittaker model, it is called* generic*. For $G = \mathrm{GL}_2$, all but the finite-dimensional (= algebraic, up to character twist) representations of $G(F_v)$ are generic.*

This implies that there is a factorization

$$W_\phi = \prod_v W_{\phi,v}, \tag{13.2}$$

where $W_{\phi,v}$ belongs to the Whittaker model of π_v. In fact, we can choose an isomorphism $\pi = \bigotimes_v' \pi_v$, $\phi \mapsto \otimes_v \phi_v$, so that $W_{\phi,v}$ depends only[37] on ϕ_v.

Combining this with (13.1), the integral over $[\mathbb{G}_m] = F^\times \backslash \mathbb{A}^\times$ and the sum over $F^\times$ can be combined to an integral over $\mathbb{A}^\times$ at least when it is absolutely convergent (which – fact – happens when $\Re(s) \gg 0$),

$$(13.1) = \int_{\mathbb{A}^\times} W_\phi \begin{pmatrix} x & \\ & 1 \end{pmatrix} |x|^s d^\times x.$$

(The same argument works for Eisenstein series, with appropriate regularization, because the regularized contribution of the zero-th Fourier coefficient will be zero.)

Together with the Whittaker factorization (13.2), we now have an Eulerian integral (for $\Re(s) \gg 0$),

$$(13.1) = \prod_v I(W_{\phi,v}, s),$$

where $I(W_{\phi,v}, s) = \int_{F_v^\times} W_{\phi,v} \begin{pmatrix} x & \\ & 1 \end{pmatrix} |x|^s d^\times x$.

13.2 Functional equation

Let π be a generic automorphic representation of $G = \mathrm{GL}_2$. For $\phi \in \pi$, define $\tilde{\phi}(g) = \phi(g^{-t})$ (inverse transpose). This is another automorphic form, belonging to the space of the dual automorphic representation $\tilde{\pi}$.

Theorem 13.2 (Global functional equation)**.** *We have*

$$I(\phi, s) = I(\tilde{\phi}, -s).$$

Proof. This is quite a trivial calculation:

$$I(\tilde{\phi}, -s) = \int_{[\mathbb{G}_m]} \phi \begin{pmatrix} x^{-1} & \\ & 1 \end{pmatrix} |x|^{-s} d^\times x = \int_{[\mathbb{G}_m]} \phi \begin{pmatrix} x & \\ & 1 \end{pmatrix} |x|^s d^\times x = I(\phi, s).$$

□

[37] To make this factorization more precise, we would need a formulation in terms of global and local functionals similar to the one in Waldspurger's theorem; see [LM15, § 4].

Now we turn our attention to the local factors of (13.2).

Proposition 13.3. *Let v be a nonarchimedean place with residue field of order q, and let π_v be an irreducible generic representation of $G(F_v)$. The set of functions $s \mapsto I(W_v, s)$, as W_v varies in the Whittaker model of π_v, is a fractional $\mathbb{C}[q^s, q^{-s}]$-ideal of $\mathbb{C}(q^{-s})$ containing the constants.*

Before we sketch the proof in an exercise, let us give our *working definition* of local L-factors – a definition which does not mention the Langlands correspondence, as promised in § 12.2.

Definition 13.1 (Local L-factor)**.** In the setting of Proposition 13.3, we let

$$L(\pi_v, \frac{1}{2} + s) = P(q^{-s})^{-1},$$

where P^{-1} is the generator of the fractional ideal which satisfies $P(0) = 1$.

Exercise(s) 13.3. In this exercise, we sketch of proof of Proposition 13.3].

1. Since the field is nonarchimedean, we are working, by definition, in the context of *smooth* representations, i.e., every vector has an open stabilizer. Use (or even prove!) the Iwasawa decomposition, $G(F_v) = N(F_v)\varpi^{\Lambda}G(\mathfrak{o}_v)$, where Λ denotes the group of cocharacters into the diagonal torus, to show that for any $W_v \in \mathcal{W}(\pi_v)$ there is an $n \geqslant 0$ such that $W_v(\pi_v^{\lambda}) = 0$ when λ is of the form $x^{\lambda} = \begin{pmatrix} x^a & \\ & x^b \end{pmatrix}$ with $\alpha(\lambda) := a - b \leqslant n$ (i.e., when λ is "sufficiently antidominant"). *(Hint: Just use the fact that W_v is left-(N, ψ)-equivariant and right-J-invariant, for some open subgroup J.)*
2. Similarly, show that for any compact subset Ω of $G(F_v)$, when λ is *sufficiently dominant*, the restriction of any W_v to $\varpi^{\lambda} \cdot \Omega$ can be identified with the restriction of a function on $N\backslash G(F_v)$ (i.e., with trivial character on N, rather than ψ), equivariantly with respect to all $g \in \Omega$. That is, given Ω, there is an n, and for every Whittaker function W_v there is a function $W_v^0 \in C^{\infty}(N\backslash G(F_v))$ such that

$$W_v^0|_{\varpi^{\Lambda_n}\Omega} = W_v|_{\varpi^{\Lambda_n}\Omega},$$

where $\Lambda_n = \{\lambda | \alpha(\lambda) \geqslant n\}$. *(This is easier than it sounds!)*

The last point implies that the elements in $\mathcal{W}(\pi_v)$, the Whittaker model of π_v, are equal to elements in the image of a morphism $\pi_v \to C^{\infty}(N\backslash G(F_v))$ on "sufficiently dominant elements." You can take this conclusion as given,

or you can try to prove it by considering the last statement as saying that "$C^\infty((N,\psi)\backslash G(F_v))$ and $C^\infty(N\backslash G(F_v))$ are asymptotically Hecke-equivariant."

For the last step of this exercise, assume this conclusion, as well as the fact that

> Any embedding $\pi_v \hookrightarrow C^\infty(N\backslash G(F_v))$ is $A(F_v)$-finite, that is, it lies in a finite sum of generalized spaces for $A(F_v)$.

This is still not enough to prove Proposition 13.3; we need more facts about the "Kirillov model" of a representation (i.e., the restriction of its Whittaker model to the diagonal torus). Nonetheless, you can use the description above to calculate the local L-factor.

3. Calculating the local L-factor $L(\pi_v, s)$ in terms of the exponents (=generalized eigencharacters for A appearing in the asymptotics) of the Whittaker model of π_v.

Theorem 13.4 (Local functional equation)**.** *Let π_v be a generic irreducible representation of $G(F_v)$. The local Hecke integrals $I(W_{\phi,v}, s)$, originally defined for $\Re(s) \gg 0$, admit a meromorphic continuation to all $\mathbb{C}$, and there is a meromorphic function $\gamma(\pi_v, s, \psi)$ with the property that*

$$I(\tilde{W}_n, -s) = \omega_{\pi_v}(-1)\gamma(\pi, s+\frac{1}{2}, \psi)I(W_v, s),$$

where $\tilde{W}_n(g) = W_n(wg^{-t})$, $w = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}$, and ω_{π_v} is the central character of π_v.

Note that $\tilde{W}_n$ is a Whittaker function for the *inverse* character of ψ. The map $W_n \to \tilde{W}_n$ is a linear map from the ψ-Whittaker model of π to the ψ^{-1}-Whittaker model of its dual[38] representation $\tilde{\pi}$.

There is no clear reason why we should separate $\omega_{\pi_v}(-1)$ from the gamma factor, or why we should choose this representative w for the nontrivial element of the Weyl group and not another; the compatibilities with other theories of ϵ-factors simply work this way!

Sketch of proof. In the nonarchimedean case, the meromorphic continuation follows from Proposition 13.3. (The archimedean case is similar, but the analog of Exercise 13.3 is a little more complicated, and we will skip it.)

[38]The fact that the automorphism $g \mapsto g^{-t}$ of GL_n sends an irreducible representation to its dual is a theorem of Gelfand and Kazhdan. Its proof is based on character theory; eventually, the theorem follows simply from the fact that this automorphism takes the conjugacy class of an element g to the conjugacy class of its inverse.

The functionals $W_v \mapsto I(W_v, s)$ and $W_v \mapsto I(\tilde{W}_v, -s)$ are equivariant with respect to the character $|\bullet|^{-s}$ of the subgroup $F_v^\times = \begin{pmatrix} * & \\ & 1 \end{pmatrix}$. We will then appeal to another multiplicity-one result, similar to the one for Whittaker models from Proposition 13.1, again without proof:

Proposition 13.5. *We have* $\dim \mathrm{Hom}_{F^\times}(\pi_v, \mathbb{C}_\chi) \leqslant 1$ *for any irreducible representation* π_v *of* $G(F_v)$*, and any character* χ *of* $F^\times$*.*

This implies that, for any s where $I(-, s)$ and $I(-, -s)$ do not have a pole, there is a scalar of proportionality between the two functionals, independent of W_v. As, for a fixed W_v, those vary meromorphically in s, the scalar is meromorphic in s. □

Finally, the local ϵ-factor is defined by the equation

$$\gamma(\pi, s, \psi) = \frac{\epsilon(\pi, s, \psi) L(\tilde{\pi}, 1 - s)}{L(\pi, s)}. \tag{13.3}$$

Since, by Proposition 13.3, $I(W_v, s)$ is any $\mathbb{C}[q^s, q^{-s}]$-multiple of $L(\pi, \frac{1}{2} + s)$, and $I(\tilde{W}_v, -s)$ is any $\mathbb{C}[q^s, q^{-s}]$-multiple of $L(\tilde{\pi}, \frac{1}{2} - s)$, it follows that $\epsilon(\pi, s)$ is entire, nonvanishing element of $\mathbb{C}[q^s, q^{-s}]$, hence a constant times an integral power of q^{-s}:

$$\epsilon(\pi, s) = c q^{-f(\pi)s}.$$

The integer exponent $f(\pi)$ is called the *conductor* of π, and it is known to be nonnegative.

Note, that, by the definitions, it follows immediately that

$$\gamma(\pi, \frac{1}{2} + s, \psi)\gamma(\tilde{\pi}, \frac{1}{2} - s, \psi^{-1}) = 1.$$

14 Waldspurger's theorem via the relative trace formula

14.1 Relative trace formulas

We will discuss a proof of Waldspurger's theorem due to Jacquet [Jac86], for the case when the character χ of the quadratic extension K/F is trivial. In this case, the base change L-function is given simply by (12.9), where $L(\pi, s)$ and $L(\pi \otimes \eta, s)$ are simply the standard L-functions associated to π and its twist by the character η (composed with the determinant). Jacquet has another proof (also based on the relative trace formula) for general χ, but it is a little more complicated.

Since the character χ is trivial, the central character of π should also be trivial, which is equivalent to saying that π is an automorphic representation of PGL_2. We will change notation, and let G, T be PGL_2 and the $\mathrm{Res}_{K/F}(\mathbb{G}_m)/\mathbb{G}_m$, respectively. Let $A \simeq \mathbb{G}_m$ be a split torus in G. There are two main ideas in the proof:

1. The first idea is to exploit (12.9) and compare the squared absolute value of the T-period (12.3) with the product of the A-period on π and the A-period on $\pi \otimes \eta$.
2. The second idea is to group all automorphic representations together. In other words, to return to the pairing (12.2), and decompose it spectrally. The benefit of this is that the pairing (12.2) has a completely geometric description (the distribution $\delta_{[T]}$ is simply the integral over the subspace $[T] \subset [G]$), and there is a geometric way to calculate it, and perform the comparison. We will then use the action of the Hecke algebra to separate individual representations.

As mentioned, however, the pairing (12.2) does not make sense as a number. We will interpret it as a distribution. More precisely, let $\mathcal{S}(G(\mathbb{A}))$ denote the convolution algebra of Schwartz measures on $G(\mathbb{A})$ (Schwartz functions times a Haar measure). We will calculate

$$\mathrm{RTF}^T(f) := \langle \delta_{[T]} | f | \delta_{[T]} \rangle := \langle f \cdot \delta_{[T]}, \delta_{[T]} \rangle, \tag{14.1}$$

where

$$f \cdot \delta_{[T]}(g) = \int_{G(\mathbb{A})} f(g) \delta_{[T]}(g) = \sum_{\gamma \in T(F) \backslash G(F)} \int_{T(\mathbb{A})} f(t\gamma g) dt.$$

(Can you confirm this formula?)

Now, 14.1 rigorously makes sense, and will be our first version of the "RTF functional" (for: relative trace formula) associated to the pair (G,T). It is a functional on $\mathcal{S}(G(\mathbb{A}))$, but it is left and right invariant by $T(\mathbb{A})$ – for example, as the last expression makes clear, it depends only on the measure[39] $\varphi_f : g \mapsto \int_{T(\mathbb{A})} f(t\gamma g)$ on $T(\mathbb{A})\backslash G(\mathbb{A})$, which is the pushforward of f through the natural quotient map. Note, also, that in this case we have $T(\mathbb{A})\backslash G(\mathbb{A}) = X(\mathbb{A})$, where $X = T\backslash G$ (as follows from Hilbert 90!), so we can regard (14.1) as a $T(\mathbb{A})$-invariant "distribution"[40] on $X(\mathbb{A})$. The automorphic function (dividing by a Haar measure here)

$$g \mapsto \sum_{\gamma \in X(F)} \varphi_f(\gamma g)$$

which appears in (14.1) will be denoted by Θ_f or Θ_{φ_f}, and called a "theta series" for the space X. Note, also, that when f is a convolution, $f(g) = \int_{G(\mathbb{A})} f_1(x) f_2(x^{-1}g)dx$, then (14.1) can be rewritten as the inner product

$$\int_{[G]} \Theta_{f_1} \cdot \Theta_{f_2^\vee}, \tag{14.2}$$

where $f_2^\vee(g) = f_2(g^{-1})$. This point of view will be important for the spectral decomposition. Through this pairing, the RTF functional can be seen as a distribution on $X(\mathbb{A}) \times X(\mathbb{A})$, invariant under the diagonal action of $G(\mathbb{A})$. Our more sophisticated version of the "RTF functional" will replace $T(\mathbb{A})$-invariant distributions on $X(\mathbb{A})$, or $G^{\mathrm{diag}}(\mathbb{A})$-invariant distributions on $(X \times X)(\mathbb{A})$, by "distributions on the adelic points of the stack $T\backslash G/T = (X \times X)/G$." But we will discuss this after we realize the need for it.

To compare with the product of two A-periods, we similarly define distributions $\delta_{[A]}$ and $\delta_{[A],\eta}$ (the latter being $\phi \mapsto \int_{[A]} \phi(a)\eta(a)da$), and define a similar functional

$$\mathrm{RTF}^A(f) := \langle \delta_{[A]} | f | \delta_{[A],\eta} \rangle$$

on $\mathcal{S}(G(\mathbb{A}))$. *(This is not absolutely convergent, and requires regularization!)* The first version of the main geometric theorem is, then, the following.

Theorem 14.1. *For every $f \in \mathcal{S}(G(\mathbb{A}))$ there is a "transfer" $f' \in \mathcal{S}(G(\mathbb{A}))$ with*

$$\langle \delta_{[T]} | f | \delta_{[T]} \rangle = \langle \delta_{[A]} | f' | \delta_{[A],\eta} \rangle.$$

[39] The notation is a bit sloppy in distinguishing between functions and measures. You can fix invariant measures and perform the translations everywhere, until we need to specify our choices of such measures.

[40] A more appropriate word is "generalized function," which is by definition an element in the dual space to smooth, compactly supported distributions.

Moreover, this transfer map is compatible with the action of spherical Hecke algebras, in the following sense: Let v be a place of F where K/F is unramified, and assume that f is bi-$G(\mathfrak{o}_v)$-invariant, with transfer f'. Then, for every $h \in \mathcal{H}(G(F_v), G(\mathfrak{o}_v))$, $h \cdot f'$ is a transfer of $h \cdot f$; where $\cdot$ denotes the action of the spherical Hecke algebra by left or, equivalently, right multiplication.

Remark 14.1. The transfer map $f \mapsto f'$ cannot be unique, since, as we saw, the RTF functional only depends on the image of f modulo the left- or right-$T(\mathbb{A})$ actions. But it can be shown that it is unique as a linear map between the *coinvariant quotients*

$$\mathcal{T} : \mathcal{S}(G(\mathbb{A}))_{(T\times T)(\mathbb{A}))} \to \mathcal{S}(G(\mathbb{A}))_{(A\times(A,\eta))(\mathbb{A}))}. \tag{14.3}$$

(The coinvariant space V_G of a G-representation V is defined as the quotient of V by the subspace generated by all elements $v - \pi(g)v$, $v \in V$, $g \in G$; when a character η of G appears in the notation, we simply mean the twisted representation $V \otimes \eta$. If V is a topological representation, we may want to define this by modding out by the closure of that subspace.)

Before we discuss what goes into the proof of Theorem 14.1, let us see how it implies a weak version of Waldspurger's theorem.

Theorem 14.2. *Let π be a generic automorphic representation of G such that $\delta_{[T]}|_\pi \neq 0$. Then, $L(\pi, \frac{1}{2})L(\pi \otimes \eta, \frac{1}{2}) \neq 0$.*

Proof. For simplicity, we will sketch the proof when π is cuspidal (although the general case is not much harder, in this case). Taking f to be a convolution of two functions, so that (14.1) can be written as the inner product (14.2) of two theta series for $T\backslash G$, one first proves that these theta series are in $L^2([G])$, and decomposes them according to the Plancherel decomposition of $L^2([G])$. This gives rise to an absolutely convergent expression,

$$\langle \Theta_{\varphi_1}, \Theta_{\varphi_2} \rangle = \int_{\hat{G}^{\mathrm{aut}}} \langle \Theta_{\varphi_1}, \Theta_{\varphi_2} \rangle_\pi d\pi,$$

where $\langle\ ,\ \rangle_\pi$ is the "projection" of the pairing to the space of the automorphic representation π. This expression is called the *spectral side* of the relative trace formula.[41] If π is cuspidal, then it occurs discretely in that decomposition, and we can take the measure $d\pi$ to be 1 at π. We have a similar decomposition for theta series with respect to $[A]$ and $([A], \eta)$ *(except that those are not L^2, and*

[41] The relative trace formula is really the equality of this spectral expression with the geometric expression that will be discussed after this proof.

the decomposition of the regularized pairing requires some care!), which we will denote by the appropriate exponents,

$$\langle \Theta^A_{\varphi_1}, \Theta^{A,\eta}_{\varphi_2} \rangle = \int_{\hat{G}^{\mathrm{aut}}} \langle \Theta^A_{\varphi_1}, \Theta^{A,\eta}_{\varphi_2} \rangle_\pi d\pi.$$

We can consider the forms $J^T_\pi : \varphi_1 \otimes \varphi_2 \mapsto \langle \Theta_{\varphi_1}, \Theta_{\varphi_2} \rangle_\pi$ as linear functionals on the coinvariant space $\mathcal{S}(G(\mathbb{A}))_{(T\times T)(\mathbb{A}))}$, called the *(global) relative characters* associated to π. Explicitly,

$$J^T_\pi(\varphi_1 \otimes \varphi_2) = \sum_\phi \int_{[G]} \Theta_{\varphi_1}(g)\phi(g)dg \cdot \int_{[G]} \Theta_{\varphi_2}(g)\bar{\phi}(g)dg,$$

where ϕ runs over an orthonormal basis of π. Let us similarly denote by J^A_π the corresponding forms given by $\langle \Theta^A_{\varphi_1}, \Theta^{A,\eta}_{\varphi_2} \rangle_\pi$.

Theorem 14.1 tells us that there is a linear map $\mathcal{T}$ such that

$$\mathrm{RTF}^A \circ \mathcal{T} = \mathrm{RTF}^T.$$

We would like to upgrade this to a statement about the spectral decompositions, i.e., to say that

$$J^A_\pi \circ \mathcal{T} = J^T_\pi. \tag{14.4}$$

If we could deduce that, it would follow that the nonvanishing of J^T_π implies the nonvanishing of J^A_π, which, for generic representations, is equivalent to $L(\pi, \frac{1}{2})L(\pi \otimes \eta, \frac{1}{2}) \neq 0$, by the Hecke integral (§ 13.1).

To deduce (14.4), fix a finite set S of places containing all archimedean and ramified places for K/F, as well as the places where the representation π of interest is ramified. For any matching functions $f_S \leftrightarrow f'_S$ in $\mathcal{S}(G(F_S))$ (i.e., over the product of $v \in S$), we consider the following functional *on the spherical Hecke algebras away from* S:

$\mathrm{RTF}^T_S(h) := \mathrm{RTF}^T(h \otimes f_S)$

and

$\mathrm{RTF}^A_S(h) := \mathrm{RTF}^A(h \otimes f'_S)$.

(Note that we are suppressing the dependence on f_S, f'_S from the notation.)

The resulting test functions $h \otimes f_S$ and $h \otimes f'_S$ are matching, by the fundamental lemma for the full Hecke algebra of Theorem 14.1. We therefore have an equality of functionals

$$\mathrm{RTF}^T_S = \mathrm{RTF}^A_S$$

on $\mathcal{H}^S := \bigotimes_{v\notin S} \mathcal{H}(G(F_v), G(\mathfrak{o}_v))$. We similarly denote by $J^T_{\pi,S}$ and $J^A_{\pi,S}$ the corresponding functionals of the spectral decomposition, so that we have an equality

$$\int_{\hat{G}^{\mathrm{aut}}} J^T_{\pi,S}(h)d\pi = \int_{\hat{G}^{\mathrm{aut}}} J^A_{\pi,S}(h)d\pi \tag{14.5}$$

for every $h \in \mathcal{H}^S$.

But, because h acts by a scalar $\hat{h}(\pi)$ on the space of unramified (i.e., $G(\mathfrak{o}_v)$-invariant) elements of π, we have $J^T_{\pi,S}(h) = \hat{h}(\pi) \cdot J^T_{\pi,S}(1_{G(\mathfrak{o}^S)})$, and similarly for $J^A_{\pi,S}$; we will denote the scalars $J^T_{\pi,S}(1_{G(\mathfrak{o}^S)})$, $J^A_{\pi,S}(1_{G(\mathfrak{o}^S)})$ simply by $J^T_{\pi,S}$, $J^A_{\pi,S}$.

Recall that the Hecke algebra $\mathcal{H}(G(F_v), G(\mathfrak{o}_v))$ is freely generated by an operator T_v (supported on the double coset of the matrix $\begin{pmatrix} \varpi_v & \\ & 1 \end{pmatrix}$), and therefore an irreducible unramified representation π_v corresponds to a point $[\pi_v] = \mathrm{Spec}_{\max}\mathcal{H}(G(F_v), G(\mathfrak{o}_v)) = \mathrm{Spec}_{\max}\mathbb{C}[T_v] \simeq \mathbb{C}$, so that $\hat{h}_v(\pi_v) =$ the evaluation of the corresponding polynomial at the point $[\pi_v]$. (This point is, up to some normalization, the *Satake parameter* of π_v.) The scalar $\hat{h}(\pi)$ is then the product of $\hat{h}_v(\pi_v)$'s, when $h = \otimes_{v \notin S} h_v$; the representation π gives rise to a set of Satake parameters $([\pi_v])_{v \notin S} \in \prod_v \mathbb{C}_v$, where we add an index v to denote the v-th copy of $\mathbb{C}$.

The expressions on the two sides of (14.5) are absolutely convergent, and therefore can be thought of as *measures* on the space $\prod_v \mathbb{C}_v$. The proof of (14.4) would be complete[42] if we could show that the two measures are complete. For that purpose, we can use the Stone–Weierstrass theorem, for an appropriate compact subset of $\prod_v \mathbb{C}_v$. Namely, the *unitary* unramified representations form a compact subset[43] $K_v \subset \prod_v \mathbb{C}_v$, and for $[\pi_v]$ in this subset we have $\overline{\hat{h}_v}(\pi_v) = \widehat{h_v^\vee}(\pi_v)$, where $h_v^\vee(g) = h_v(g^{-1})$. Hence, the restriction of polynomials on $\mathbb{C}_v$ to K_v is an algebra that is invariant under complex conjugation; it also separates points, and by the Stone–Weierstrass theorem it follows that $\mathbb{C}[\mathbb{C}_v]$ is dense in $C(K_v)$. Thus, from the equality of the expressions (14.5) for all $h \in \prod_v \mathbb{C}_v$ we can deduce the equality of the measures (which are supported on $\prod_v K_v$), and we are done.

□

14.2 Geometric comparison

Before we come to the proof of Theorem 14.1, we will develop the geometric expression of the two RTF functionals. (The spectral expansion appeared in the proof of Theorem 14.2 above.) In general, a relative trace formula involves the pairing of two theta series coming from G-spaces X_1 and X_2 (or, when

[42] For almost all π with respect to the automorphic measure $d\pi$, but then it is easy to deduce the equality everywhere on the support of the continuous/Eisenstein part of this measure by a continuity argument.

[43] There is a very well-known, classical description of this in terms of *principal series*, *complementary series*, and *1-dimensional representations*, but we won't get into that here.

the X_i are homogeneous, $X_i = H_i\backslash G$, the pairing of two period distributions associated to subgroups H_1 and H_2), and its geometric side is a decomposition of the pairing indexed by G^{diag}-orbits on $X_1 \times X_2$ (or by $H_1 \times H_2$-orbits on G).

When everything converges, as is the case with RTF^T, this geometric expression arises simply from unfolding the definitions. Indeed, we can rewrite (14.1) as

$$\sum_{\xi \in T(F)\backslash G(F)/T(F)} \mathrm{Vol}(T_\xi(F)\backslash T_\xi(\mathbb{A})) \int_{T_\xi(\mathbb{A})\backslash (T\times T)(\mathbb{A})} f(t_1^{-1}\xi t_2) d(t_1, t_2), \quad (14.6)$$

with the integral on the right hand side, to be denoted by $O_\xi^T(f)$, called the *orbital integral*[44] for f at ξ.

A similar expression holds for the relative trace formula RTF^A, except that some of the integrals are divergent and need to be regularized. We'll discuss the nature of this regularization later, so let us formally write

$$\mathrm{RTF}^A(f') = \sum_{\xi \in A(F)\backslash G(F)/A(F)} \mathrm{Vol}(A_\xi(F)\backslash A_\xi(\mathbb{A})) \int_{A_\xi(\mathbb{A})\backslash (A\times A)(\mathbb{A})} f'(a_1^{-1}\xi a_2)\eta(a_2) d(a_1, a_2). \quad (14.7)$$

The idea, now, is to compare the geometric expressions above (almost) orbit-by-orbit. Namely, we have the invariant-theoretic quotients $T\backslash G /\!\!/ T = \mathrm{Spec} F[T\backslash G]^T$ and $A\backslash G /\!\!/ A = \mathrm{Spec} F[A\backslash G]^A$, and the following lemma.

Lemma 14.3. *Both quotients are isomorphic to $\mathbb{A}^1$ (i.e., the ring of invariants are polynomial in one generator); moreover, they are* canonically *equal, if we consider the pair (G, T) as a Galois twist of the pair (G, A).*

Proof. For PGL_2 and the diagonal torus A, it is very easy to see that the ring of $A \times A$-invariants is freely generated by $\xi \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{ad}{\det}$.

"Galois twist" means that we can identify the two pairs over the algebraic closure (in fact, over the quadratic extension K/F), in such a way that the Galois action on (G'_K, T_K) (we use G' for another copy of G here – an F-group that is F-isomorphic to, but not identified with, G) is the twist of the Galois action on (G_K, A_K) by a 1-cocycle $\mathrm{Gal}(K/F) \to \mathrm{Aut}(G_K, A_K)$; we can take this cocycle to be sending the nontrivial element of the Galois group to conjugation by the permutation matrix $w = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}$.

[44] In the equivalent formulation where the test functions $\varphi_1 \otimes \varphi_2$ live on $X \times X$ and ξ corresponds to a G^{diag}-orbit there, the orbital integral O_ξ is defined as an integral over the adelic points of this G-orbit. One important observation here is that stabilizers don't depend on whether we represent the quotient as $T\backslash G/T$ or as $(X \times X)/G$; in particular, the volume factor before the orbital integral does not change.

It is then clear that this automorphisms acts trivially on the variable ξ, and therefore the corresponding twist of $T\backslash G' /\!\!/ T$ is trivial, i.e., we have a canonical isomorphism $T\backslash G' /\!\!/ T \simeq A\backslash G /\!\!/ A$. □

By repeatedly using the Galois twist of the previous lemma, we will be proving statements in the easier setting of (G, A), and transferring them to (G, T). We will keep using the coordinate $\xi = \frac{ad}{\det}$ of the lemma.

Proposition 14.4. *Let $U = \mathbb{A}^1_\xi \smallsetminus \{0, 1\}$. The space $X = A\backslash G$ restricts to an A-torsor over U, and the space $Y = T\backslash G$ restricts to a T-torsor. Moreover, there is a birational T-equivariant map $\mathrm{Res}_{K/F}\mathbb{G}_a \dashrightarrow Y$ (where we think of T as the kernel of the norm map $K^\times \to F^\times$), which is an isomorphism over a Zariski open neighborhood of $\xi = 1$, and similarly for X, if we replace K by $F \oplus F$. Finally, there is a G-automorphism of X (resp. of Y) which descends to the automorphism $\xi \mapsto 1 - \xi$ on $\mathbb{A}^1$.*

Proof. Let us start with the linearization, in the case of the split torus: The space X can be identified with $(\mathbb{P}^1 \times \mathbb{P}^1) \smallsetminus \Delta\mathbb{P}^1$, in such a way that $\xi(x, y) = \frac{x}{x-y}$. The map $\mathbb{A}^2 \ni (x, y) \mapsto (x^{-1}, y)$ restricts to the birational map asserted in the proposition, is defined away from $\{xy = 1\}$, and has image the complement of $\xi = 0$. Moreover, the automorphism $Ag \mapsto Awg$ of $A\backslash G$ acts as $\xi \mapsto 1 - \xi$ on the quotient, and therefore everything we prove for the complement of $\xi = 0$ holds for the complement of $\xi = 1$, as well.

Therefore, for the split torus, we are reduced with proving the statement about torsors, which we can prove for the A-action on $\mathbb{A}^2$, removing the set $\zeta := xy = 0$ (which corresponds to $\xi = 1$). But it is clear that the equation $xy = \zeta$ represents an A-torsor (trivial, of course).

Now we twist everything by the "conjugation by w" involution. It is easy to see that the birational map that we defined above is compatible with this involution, if we also act on the (x, y)-plane by switching x and y. The resulting form of $\mathbb{A}^2$ can be identified with $\mathrm{Res}_{K/F}\mathbb{G}_a$, under the action of the Galois twist T of A. Therefore, all the statements remain true; of course, the T-torsors can now be nontrivial, as one can see by noticing that the coordinate ζ is equal to the norm map (hence, $\zeta \neq 0$ corresponds to a trivial T-torsor iff ζ is a norm from $K^\times$). □

We will use this proposition to compare the geometric sides of the relative trace formulas. First of all, the proposition implies that all the terms in the geometric expansions (14.6), (14.7) *except* for those living over $\xi = 0, 1 \in \mathbb{A}^1$ are simply orbital integrals, without volume terms. These are Euler products of the corresponding local orbital integrals. Therefore, up to examining the finitely many terms over $\xi = 0, 1$ (which we will skip in this course, for lack of

time!), the geometric comparison of Theorem 14.1 follows from the following local statement, which not only proves that theorem, but more precisely proves equality of the individual geometric terms with $\xi \neq 0, 1$.

Fix a place v, and, for $f \in \mathcal{S}(T\backslash G(F_v))$, denote by $\pi_{T,!}f$ the orbital integral function on $U(F_v) = F_v \smallsetminus \{0, 1\}$,

$$\pi_{T,!}f(\xi) = \int_{T(F_v)} f(\tilde{\xi}t)dt,$$

where $\tilde{\xi}$ is a lift of ξ to $G(F_v)$. Similarly, choose a *global* section $\xi \mapsto \tilde{\xi}$ of the map $A\backslash G \to A\backslash G /\!\!/ A$, and use it to define the local pushforward

$$\pi_{A,!}f(\xi) = \int_{A(F_v)} f(\tilde{\xi}a)\eta(a)da.$$

(Note that if we choice a different section, here, the presence of η might introduce a negative sign; but globally these signs would cancel out, since η is an automorphic character and we chose the sections globally, hence any two choices $\tilde{\xi}, \tilde{\xi}'$ are translates of each other by $A(F)$.) These definitions depend on choices of Haar measures on $T(F_v)$, $A(F_v)$, which need to be made carefully for some of the following theorems to hold, but we won't get into that. (Better: define Schwartz spaces as spaces of *smooth measures*; then these pushforwards don't depend on choices.)

We use the same notation for $f \in \mathcal{S}(G(F_v))$, identifying it with its pushforward[45] to $T\backslash G$, resp. $A\backslash G$.

Theorem 14.5. *1. (Matching:) The pushforwards map Schwartz functions to smooth functions on $U(F_v)$, and for every $f \in \mathcal{S}(G(F_v))$ there is an $f' \in \mathcal{S}(G(F_v))$ such that $\pi_{T,!}(f) = \pi_{A,!}(f')$.*

2. (Fundamental lemma for the Hecke algebra:) If v is a place where K/F is unramified,[46] and $f \in \mathcal{H}(G(F_v), G(\mathfrak{o}_v))$, then f' can be taken to be equal to f.

Proof. We will only prove the matching. The fundamental lemma needs some tedious calculations, which in this case can be done "by hand;" see [Jac86]. Note that if K splits at v, the two statements are tautologies, so now we assume that it doesn't split.

First of all, smoothness follows from the smoothness of the map $T\backslash G \to \mathbb{A}^1$ (and similarly for A), restricted to U. Indeed, locally around a point of $U(F_v)$ (in the Hausdorff topology), i.e. over a neighborhood V of that point, we have

[45] The pushforward is again dependent on choices of Haar measures, which

[46] ...and some integrality assumption for the choice of section $\tilde{\xi}$, which we omit...

$T\backslash G(F_v)|_V$ is either empty or a product $T(F_v) \times V$ ($T(F_v)$-equivariantly), and smoothness of the pushforward follows easily from this.

The matching statement now means that $\pi_{T,!}(\mathcal{S}(G(F_v)))$, as a subspace of $C^\infty(U(F_v))$, is contained in $\pi_{A,!}(\mathcal{S}(G(F_v)))$. The product statement that we used above also implies that all smooth, compactly supported functions on $U(F_v)$ are contained in the image of the pushforward map, provided that their support lies in the image of $G(F_v)$ (which is all of $U(F_v)$ in the split case, but not in the nonsplit case). Therefore, there remains to compare limiting behaviors of the pushforwards as $\xi \to 0, 1$. By the automorphism mentioned in Proposition 14.4, it is enough to check in a neighborhood of $\xi = 1$. By the linearization map of this proposition, it is enough to compare orbital integrals for the $A(F_v)$-action on F_v^2 and the $T(F_v)$-action on K_v, in a neighborhood of $\zeta = 0$ (where ζ is the invariant coordinate as in the proof of that proposition, i.e., the origin of both vector spaces corresponds to $\zeta = 0$).

We keep using the same symbols of pushforwards, except that our test functions now live on F_v^2, resp. K_v. In the split case, we analyze the functions of the form $\pi_{A,!}(f)$ by taking their Mellin transforms in the ζ, variable:

$$\widehat{\pi_{A,!}(f)}(\chi) = \int_{F_v^\times} \pi_{A,!}(f)(\zeta)\chi^{-1}(\zeta)d^\times\zeta.$$

Assuming that the section $\zeta \mapsto \tilde{\zeta}$ used to define $\pi_{A,!}$ is $\tilde{(}\zeta) = (1, \zeta)$ (this depends on the section used for $\tilde{\xi}$, but its choice does not matter for the matching statement), we use the definition to write the Mellin transform above as

$$\int_{F_v^\times}\int_{F_v^\times} f(a^{-1}, a\zeta)\eta(a)\chi^{-1}(\zeta)\, d^\times a\, d^\times\zeta = \int_{F_v^\times}\int_{F_v^\times} f(x,y)\eta\chi^{-1}(x)\chi^{-1}(y)\, d^\times x\, d^\times y,$$

which is a double Tate integral (one in the x-variable, and one in the y-variable). We know that such an integral is meromorphic in χ with (at most) simple poles at $\chi^{-1}\eta = 1$ and at $\chi^{-1} = 1$. A more careful study of it will prove that $\pi_{A,!}(f)$ has the following behavior close to $\zeta = 0$.

$$\pi_{A,!}(f) = c_1(\zeta) + \eta(\zeta)c_2(\zeta),$$

where c_1, c_2 are smooth functions, and that all functions of this form (in a neighborhood of $\zeta = 0$) can be obtained.

On the other hand, the pushforwards $\pi_{T,!}$ can be analyzed as follows. Given that $T(F_v)$ is compact, we can first make our test function $T(F_v)$-invariant, by averaging it. This produces a projection $\mathcal{S}(K_v) \to \mathcal{S}(K_v)^{T(F_v)}$, and for any function f in the image we have that $\pi_{T,!}f(\zeta) = f(\tilde{\zeta})$ for any lift $\tilde{\zeta}$ of ζ. This implies that

$$\pi_{T,!}(\mathcal{S}(K_v)) = \mathcal{S}(F_v) \cdot 1_{N(K_v^\times)},$$

i.e., the restrictions of smooth functions on the line to the image of the norm map. Since $1_{N(K_v^\times)} = \frac{1+\eta}{2}$, this means that

$$\pi_{T,!}(\mathcal{S}(K_v)) \subset \pi_{A,!}(\mathcal{S}(F_v^2)),$$

and the analogous statement for $T\backslash G/T$ and $A\backslash G/(A,\eta)$.

□

14.3 The missing orbits: Pure inner forms

Theorem 14.5 on the comparison of the two RTFs, and the consequent Theorems 14.1, 14.2, are asymmetric: They do not postulate the existence of a matching function for RTF^T for any test function for RTF^A. The proof of the theorem, relying on Proposition 14.4, reveals the reason: While all points of the GIT quotient $\mathbb{A}^1_\xi = A\backslash G /\!\!/ A$ are accounted for by $A(F)$-orbits on $A\backslash G$ (for F global or local), this is not the case for the quotient $T\backslash G /\!\!/ T$, where many points of $\mathbb{A}^1_\xi$ correspond to nontrivial T-torsors contained in $T\backslash G$. To account for those, and obtain a full matching of test functions, we need to "twist" the pair (G,T) by nontrivial T-torsors.

Let us collect a few facts about torsors for a smooth algebraic group G over a field F.

- By definition, a(n étale) torsor is a G-scheme R which is isomorphic to G (as a G-scheme) over an étale (i.e., separable) extension of F.
- Choosing such an isomorphism over the algebraic closure, and comparing Galois actions, we get a cocycle, whose cohomology class in $H^1(F,G) := H^1_{\mathrm{cont}}(\mathrm{Gal}(\bar{F}/F, G(\bar{F}))$ (we use $\bar{F}$ for separable closure here) determines the isomorphism class of the torsor; conversely, such a cocycle allows us to "twist" the usual Galois action on G to obtain a torsor, so the set of isomorphism classes of G-torsors is identified with $H^1(F,G)$.
- A G-torsor is trivializable (i.e., isomorphic to G as a G-space) iff it has a point over F.
- The G-automorphism group of a G-torsor is a form of G. More precisely, over the algebraic closure, the set of G-automorphisms of G (as a G-space under, left or right multiplication) is G (acting, respectively, by right or left multiplication), so the G-automorphism group of a G-torsor over F has to be a form of G. This form is determined by applying 1st Galois cohomology to the sequence of maps $G \to \mathrm{Inn(G)} \to \mathrm{Aut(G)}$, where $\mathrm{Inn}(G)$ is the group of inner automorphisms of G (identified, as an algebraic group, with the quotient of G by its center).

- Traditionally, a class in $H^1(F, \mathrm{Inn}(\mathrm{G}))$ is called an "inner form" of G, and a class in $H^1(F, G)$ is called a "pure inner form." This terminology is motivated by thinking about the corresponding form of the group, but one should keep in mind, when using those terms, that they contain more information than just an isomorphism class of forms of G. For example, if G is abelian, then all inner forms correspond to isomorphic groups, but it is not true that all G-torsors are trivial.

Going back to our setting, for any T-torsor R, we can consider the twisted space $Y^R = Y \times^T R$, which still has an action of $T = \mathrm{Aut}_T(R)$. (Recall that Y denotes the space $T\backslash G$.) If we denote by $R^\vee$ the dual T-torsor (which is R as a variety, but with the action of T inverted), then we have $R \times^T R^\vee \simeq T$, reflecting the group structure (inversion) in $H^1(F, T)$. Actually, in our setting, since T splits over a quadratic extension, the group $H^1(F, T)$ is a 2-group, and for simplicity of notation we will identify R with $R^\vee$.

We have a canonical isomorphism $Y^R /\!\!/ T = Y \times^T R /\!\!/ T = (Y \times *) /\!\!/ T = Y /\!\!/ T$, so it makes sense to talk about the fibers of X^R over the same quotient space $\mathbb{A}^1_\xi$. And, those fibers that used to be isomorphic to the T-torsor R before, are now isomorphic to $R \times^T R = T$, i.e., are trivial(izable). That means that we can recover the "missing" points of $U(F) \subset \mathbb{A}^1_\xi(F)$ from F-points of Y^R. With this observation, Theorem 14.5 upgrades to the following.

Theorem 14.6. *1. (Matching:) The pushforwards map Schwartz functions to smooth functions on $U(F_v)$, and we have an identity of subspaces of $C^\infty(U(F_v))$,*

$$\bigoplus_R \pi_{T,!}(\mathcal{S}(Y^R(F_v)) = \pi_{A,!}(\mathcal{S}(X(F_v))),$$

where the sum ranges over isomorphism classes of T-torsors over F_v. (There are only two isomorphism classes, since F_v is local.)

2. (Fundamental lemma for the Hecke algebra:) If v is a place where K/F is unramified, and $f \in \mathcal{H}(G(F_v), G(\mathfrak{o}_v))$, then f' can be taken to be equal to f.

Note that there is no change in the stamement of the fundamental lemma; the "basic function" still comes from the original form of the space Y.

To extract spectral content from the twisted spaces Y^R, we should upgrade them to varieties with an action of a group like G. For that purpose, we note that, given a T-torsor R, we get a (left) G-torsor R_G by $R_G = G \times^T R$. In terms of the classification of torsors by Galois cohomology, this operation is compatible with the map $H^1(F, T) \to H^1(F, G)$ induced by the embedding of T in G. The torsor R_G carries a (right) action of its G-automorphism group G^R, which as we recalled above is a form of G. Here, explicitly, G^R will be

$PD^\times$, for some quaternion algebra D over F. More precisely, we have the following description of these groups.

Proposition 14.7. *1. The group $H^1(F,T)$ classifies isomorphism classes of binary quadratic forms with the same discriminant as the norm form on K.*

2. *The group $H^1(F,G)$ classifies isomorphism classes of quaternion (i.e., 4-dimensional) central simple algebras over F.*

3. *The boundary map $H^1(F,\mathrm{PGL}_2) \to \mathrm{Br}(F) = H^2(F,\mathbb{G}_m)$ for the long exact sequence corresponding to $1 \to \mathbb{G}_m \to \mathrm{GL}_2 \to \mathrm{PGL}_2 \to 1$ identifies $H^1(F,\mathrm{PGL}_2)$ with the 2-torsion in $\mathrm{Br}(F)$. If F is local, this 2-torsion group is $\mathbb{Z}/2$ (which we will identify by the "invariant" map with the 2-torsion in $\mathbb{Q}/\mathbb{Z}$), i.e., there is a unique ramified (i.e, different from Mat_2) quaternion algebra over F. If F is global, restriction maps in cohomology give rise to a short exact sequence*

$$0 \to \mathrm{Br}(F) \to \bigoplus_v \mathrm{Br}(F_v) \xrightarrow{\sum_v \mathrm{inv}_v} \frac{1}{2}\mathbb{Z}/\mathbb{Z} \to 0;$$

therefore, a quaternion algebra over F is determined by the even set Σ of places where it is ramified.

4. *The map $H^1(F,T) \to H^1(F,G)$ is injective. Locally, at places v where K/F is nonsplit it is an isomorphism. Globally, its image corresponds to those quaternion algebras whose set Σ of ramified places consists of places that do not split in K/F.*

Proof. The first and second statement follow by identifying T and G with the automorphism groups of appropriate algebraic data. For T, it is the automorphism group of the triple (V,q,ω), where V is the vector space $\mathrm{Res}_{K/F}\mathbb{G}_a$ over F, q is the norm form, and ω is the volume form corresponding to some choice of basis. (This choice of volume form allows us to talk about the "discriminant" of the norm form as a number, but any other choice changes that number by a square, and the sets of isomorphism classes of binary quadratic forms with discriminants differing by a square are of course in canonical bijection, through scaling.) For G, it can be identified with the automorphism group of Mat_2, considered as an F-algebra; forms of Mat_2 are precisely the quaternion algebras.

The third statement is a standard result in class field theory, and can be found in most treatments of class field theory in the literature.

Injectivity in the fourth statement follows from the compatibility of the short exact sequences $1 \to \mathbb{G}_m \to \mathrm{Res}_{K/F}\mathbb{G}_m \to T \to 1$ and $1 \to \mathbb{G}_m \to \mathrm{GL}_2 \to G \to 1$, and the long exact sequences in cohomology (including

Hilbert '90: $H^1(F, \mathrm{Res}_{K/F}\mathbb{G}_m) = H^1(K, \mathbb{G}_m) = H^1(K, \mathrm{GL}_n) = 1$). Locally, the unique quaternion division algebra contains every quadratic field extension – this is also a standard result in class field theory, which can be used to identify the 2-torsion of the Brauer group with $\mathbb{Z}/2$, but can also be recovered from it by noticing that the existence of nontrivial T-torsors R (e.g., the fibers of the norm map $K \to F$ over non-norms), together with injectivity, imply that the map $H^1(F, T) \to H^1(F, G)$ has to be an isomorphism when K/F is a quadratic field extension of local fields, hence $T = \mathrm{Aut}_T(R)$ embeds into $PD^\times = \mathrm{Aut}_G(R_G)$. Globally, the characterization of the image of the map $H^1(F, T) \to H^1(F, G)$ can be obtained by combining the local maps with the reciprocity exact sequence of the third statement.

□

The twisted space Y^R can be rewritten as $Y \times^G R_G$, and therefore carries an action of G^R. The functional RTF^T can now be redifined as a functional on the direct sum

$$\bigoplus_R \mathcal{S}(Y^R(\mathbb{A})),$$

where R ranges over isomorphism classes of T-torsors over F. On each summand, it is again given by the $[T]$-period of the corresponding theta series on Y^R, which can also be rewritten as the inner product of two theta series on Y^R, and spectrally decomposed like the pairing (14.2). Theorem 14.2 upgrades to the following.

Theorem 14.8. *Let π be a generic automorphic representation of G. The following are equivalent.*

1. $L(\pi, \frac{1}{2})L(\pi \otimes \eta, \frac{1}{2}) \neq 0$.
2. *There is a quaternion algebra D, and an automorphic representation π_D of $PD^\times$, with the same Hecke eigenvalues as π at all nonarchimedean places v where D splits and π_v is unramified, such that $\delta_{[T]}|_{\pi_D} \neq 0$.*

15 The Gross–Zagier theorem via the relative trace formula

Finally, for the final (1!) lecture of this class, we sketch the adaptation of the relative trace formula, due to Wei Zhang, in order to prove the Gross–Zagier theorem. References include [Zha12b], [Zha12a], although they are written mostly with generalizations to higher unitary groups in mind; there is also an incomplete draft by L. Cai–Y. Tian–X. Yuan–W. Zhang, hopefully to be released soon, that will reprove the full Gross–Zagier theorem, as generalized in [YZZ13], by the relative trace formula approach.

Again, for simplicity, we will only consider the case of a trivial character on the torus, and automorphic representations (corresponding to holomorphic modular forms of weight 2) with trivial central character. We will seek an "arithmetic" analog of the comparison of RTFs that we used for the proof of (a weak version of) Waldspurger's theorem, but I should stress that the cited papers of Zhang use a *different* RTF comparison, suited for general idele class characters of T. Therefore, we will just formulate expectations and the general pattern of such a comparison, without presenting the actual comparison that appears in the literature.

Also for simplicitly, instead of treating the case of a general CM field K/F, we will take $F = \mathbb{Q}$. We start by writing down the completely analogous pairing to the relative trace formula, adapting the divergent (12.5) to

$$H(f) := \langle S_T | f | S_T \rangle_{NT},$$

where, we remind, S_T is the (0-dimensional) Shimura variety of the torus T, embedded in S_G (the inverse limit of modular curves), but also, here, identified with its image in the Jacobian J of S_G through the Abel–Jacobi map. The test function f lives in $\prod_{v<\infty} \mathcal{S}(G(F_v))$, i.e., we omit the archimedean place(s), and again the notation $\langle S_T | f | S_T \rangle_{NT}$ simply means the Néron–Tate pairing between $f \cdot S_T$ and S_T. The insertion of f reduces this pairing to a finite-level modular curve, where it makes sense as a finite number. Of course, eventually one also needs to include twists by T-torsors in the definition of H, to get the full theorem of [YZZ13].

The idea is to compare the distribution H to the *derivative* of the relative trace formula RTF^A for the split torus, where by derivative we mean

$$I' := \frac{d}{ds}\Big|_{s=0} \mathrm{RTF}^A_s,$$

where

$$\mathrm{RTF}^A_s(f) := \langle \delta_{[A],|\bullet|^s} | f | \delta_{[A],\eta|\bullet|^s} \rangle.$$

Here, as before, for an idele class character χ of $A = \begin{pmatrix} * & \\ & 1 \end{pmatrix} \simeq \mathbb{G}_m$, $\delta_{[A],\chi}$ denotes the distribution $\phi \mapsto \int_{[A]} \phi(a)\chi(a)da$.

In order to perform this comparison geometrically, we will need to take the derivatives on the geometric side of RTF_s^A (i.e., the derivatives of orbital integrals in the s-parameter), and compare them with a similar geometric expansion of the functional $H(f)$.

To simplify things, let us assume that f is supported away from the preimage of $0, 1 \in \mathbb{A}^1_\xi$, so that the geometric expression of RTF_s^A, by a straightforward generalization of (14.7), is just given by a sum

$$\sum_{\xi \in F \smallsetminus \{0,1\}} O^A_{\xi,s}(f),$$

where $O^A_{\xi,s}$ is the orbital integral

$$O^A_{\xi,s}(f) = \int_{(A\times A)(\mathbb{A})} f'(a_1^{-1}\xi a_2)\eta(a_2)|a_1^{-1}a_2|^s d^\times a_1\, d^\times a_2.$$

This is an Euler product of orbital integrals, and therefore its derivative at $s = 0$, which we will denote by $O'_\xi(f)$, is a sum over primes (when $f = \prod_v f_v$),

$$O'_\xi(f) = \sum_v \prod_{w\neq v} O_{\xi,w}(f_w)O'_{\xi,v}(f_v),$$

giving rise to a similar decomposition of $I'(f)$ of the form $\sum_v I'_v(f)$.

On the arithmetic side, one decomposes the height pairing in terms of local heights, as explained in § 8.5,[47]

$$H(f) = \sum_v H_v(f),$$

and of course we would like to match the term $H_v(f)$ with the term $I'_v(f')$, for suitable matching functions $f \leftrightarrow f'$.

The next task is to break up $H_v(f)$ into a similar sum indexed by $\xi \in F$ (or in $F \smallsetminus \{0, 1\}$), for suitable choices of test functions, and write the contribution of each ξ as an Euler product, as on the split side. Note that, although H_v is defined in terms of a local-at-v height pairing, it is the local height pairing on a global arithmetic surface – not really "local."

To do that, at least at nonarchimedean places, we need to use the *p-adic uniformization* of formal neighborhoods of points on the special fiber of the

[47] This requires arithmetic models satisfying various assumptions, which we gloss over here.

modular and Shimura curves at v, which in the case of modular curves is related to Serre–Tate theory [KM85, § 13] (identifying formal deformations of an abelian curve mod p with formal deformations of its p-divisible group), and generalizes to a theorem of Rapoport–Zink [RZ96, Theorem 6.30]. For example, the formal neighborhood of the modular curve (of given prime-to-p level) at a supersingular point modulo p can be identified with a quotient of the form

Check me!

$$G_p(\mathbb{Q})\backslash(\mathcal{N}_G \times G(\mathbb{A}_f^p)/K^p),$$

where

- $\mathcal{N}_G$ is the Lubin–Tate moduli space of formal p-divisible groups of height 2;
- $G = \mathrm{GL}_2$, and K^p is the subgroup in the adeles outside of p determining the level;
- G_p is the inner form corresponding to the quaternion algebra which ramifies at p and ∞.

Using this, the v-height $H_v(f)$ decomposes into a sum of the form

$$\sum_{\xi \in T(F)\backslash G(F)/T(F)} \left(\prod_{w \neq v} O_{\xi,w}(f_w) \right) \cdot \langle \mathcal{N}_T | \xi_v | \mathcal{N}_T \rangle_{\mathcal{N}_G},$$

where the last factor is a local height pairing on the "local Shimura variety" $\mathcal{N}_G$. (This is very sketchy, and for a "good" choice of local factor f_v, which has disappeared from the notation.)

Finally, the *Arithmetic Fundamental Lemma* states that, for regular $\xi_v \neq 0, 1$, the derivative $O'_{\xi,v}(f_v)$ of the local orbital integrals (for the split torus) when f_v is the "basic function" $1_{G(\mathfrak{o}_v)}$ is *equal* to the local height pairing $\langle \mathcal{N}_T | \xi_v | \mathcal{N}_T \rangle_{\mathcal{N}_G}$.

This allows for equating the functionals H and I' for appropriate test functions, from which the Gross–Zagier formula can be deduced.

References

[Bum97] Daniel Bump. *Automorphic forms and representations*, volume 55 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[Cog03] J. W. Cogdell. Analytic theory of L-functions for GL_n. In *An introduction to the Langlands program (Jerusalem, 2001)*, pages 197–228. Birkhäuser Boston, Boston, MA, 2003.

[Cox22] David A. Cox. *Primes of the form $x^2 + ny^2$—Fermat, class field theory, and complex multiplication*. AMS Chelsea Publishing, Providence, RI, third edition, [2022] ©2022. With contributions by Roger Lipsett.

[Dem86] Michel Demazure. *Lectures on p-divisible groups*, volume 302 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. Reprint of the 1972 original.

[DG70] Michel Demazure and Pierre Gabriel. *Groupes algébriques. Tome I: Géométrie algébrique, généralités, groupes commutatifs*. Masson & Cie, Éditeurs, Paris; North-Holland Publishing Co., Amsterdam, 1970. Avec un appendice *Corps de classes local* par Michiel Hazewinkel.

[DS05] Fred Diamond and Jerry Shurman. *A first course in modular forms*, volume 228 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2005.

[Eis95] David Eisenbud. *Commutative algebra*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. With a view toward algebraic geometry.

[Hen93] Guy Henniart. Caractérisation de la correspondance de Langlands locale par les facteurs ϵ de paires. *Invent. Math.*, 113(2):339–350, 1993.

[HT01] Michael Harris and Richard Taylor. *The geometry and cohomology of some simple Shimura varieties*, volume 151 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2001. With an appendix by Vladimir G. Berkovich.

[Jac86] Hervé Jacquet. Sur un résultat de Waldspurger. *Ann. Sci. École Norm. Sup. (4)*, 19(2):185–229, 1986.

[KM85] Nicholas M. Katz and Barry Mazur. *Arithmetic moduli of elliptic curves*, volume 108 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1985.

[Lan87] Serge Lang. *Elliptic functions*, volume 112 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1987. With an appendix by J. Tate.

[LM15] Erez Lapid and Zhengyu Mao. A conjecture on Whittaker-Fourier coefficients of cusp forms. *J. Number Theory*, 146:448–505, 2015.

[LT65] Jonathan Lubin and John Tate. Formal complex multiplication in local fields. *Ann. of Math. (2)*, 81:380–387, 1965.

[Mil05] J. S. Milne. Introduction to Shimura varieties. In *Harmonic analysis, the trace formula, and Shimura varieties*, volume 4 of *Clay Math. Proc.*, pages 265–378. Amer. Math. Soc., Providence, RI, 2005.

[Mil17] J. S. Milne. *Algebraic groups*, volume 170 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2017. The theory of group schemes of finite type over a field.

[Neu99] Jürgen Neukirch. *Algebraic number theory*, volume 322 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999. Translated from the 1992 German original and with a note by Norbert Schappacher, With a foreword by G. Harder.

[RV14] Michael Rapoport and Eva Viehmann. Towards a theory of local Shimura varieties. *Münster J. Math.*, 7(1):273–326, 2014.

[RZ96] M. Rapoport and Th. Zink. *Period spaces for p-divisible groups*, volume 141 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1996.

[Sch13] Peter Scholze. The local Langlands correspondence for GL_n over p-adic fields. *Invent. Math.*, 192(3):663–715, 2013.

[Ser67a] J.-P. Serre. Complex multiplication. In *Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965)*, pages 292–296. Academic Press, London, 1967.

[Ser67b] J.-P. Serre. Local class field theory. In *Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965)*, pages 129–162. Academic Press, London, 1967.

[Ser88] Jean-Pierre Serre. *Algebraic groups and class fields*, volume 117 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1988. Translated from the French.

[Sil94] Joseph H. Silverman. *Advanced topics in the arithmetic of elliptic curves*, volume 151 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994.

[Sil09] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009.

[Sta25] The Stacks project authors. The Stacks project. https://stacks.math.columbia.edu, 2025.

[SW20] Peter Scholze and Jared Weinstein. *Berkeley lectures on p-adic geometry*, volume 207 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2020.

[Tat67] J. T. Tate. p-divisible groups. In *Proc. Conf. Local Fields (Driebergen, 1966)*, pages 158–183. Springer, Berlin-New York, 1967.

[Wei] J. Weinstein. The geometry of Lubin–Tate spaces. http://math.bu.edu/people/jsweinst/FRGLecture.pdf.

[YZZ13] Xinyi Yuan, Shou-Wu Zhang, and Wei Zhang. *The Gross-Zagier formula on Shimura curves*, volume 184 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2013.

[Zha12a] Wei Zhang. Gross-Zagier formula and arithmetic fundamental lemma. In *Fifth International Congress of Chinese Mathematicians. Part 1, 2*, volume 51, pt. 1, 2 of *AMS/IP Stud. Adv. Math.*, pages 447–459. Amer. Math. Soc., Providence, RI, 2012.

[Zha12b] Wei Zhang. On arithmetic fundamental lemmas. *Invent. Math.*, 188(1):197–252, 2012.