COUNTING POINTS ON SHIMURA VARIETIES

YIHANG ZHU

Abstract. These notes are based on an online mini-course taught at BICMR, Peking University, in August 2021. The goal is to provide an informal introduction to the Langlands–Kottwitz program of applying trace formula methods to the understanding of cohomology of Shimura varieties.

Contents

Main references for the course: [Kot92], [Kot90], [Kis10], [Kis17], [KSZ]. Also cf. the expository article [Zhu20].

1. Lecture 1

1.1. **Hasse–Weil zeta functions.** Let X be a smooth projective variety over $\mathbb{Q}$. For almost every (i.e. avoiding finitely many) prime p, there exists a "good integral model" $\mathcal{X}_p$ over $\mathbb{Z}_{(p)}$, i.e., a smooth projective scheme over $\mathbb{Z}_{(p)}$ whose generic fiber is X. In fact, one can simply take any finite-type model $\mathcal{X}$ of X over $\operatorname{Spec}\mathbb{Z}$ and then define $\mathcal{X}_p$ to be $\mathcal{X} \times_{\operatorname{Spec}\mathbb{Z}} \operatorname{Spec}\mathbb{Z}_{(p)}$ for almost every p.

We can then define the local zeta function:

$$\zeta_p(X,s) := \exp\Big(\sum_{n=1}^{\infty} \#\mathcal{X}_p(\mathbb{F}_{p^n})\frac{p^{-ns}}{n}\Big)$$

By Lefschetz trace formula and proper smooth base change, we can rewrite this as

$$\zeta_p(X,s) := \prod_{i=0}^{2\dim X} \det\Big(1 - \mathrm{Frob}_p \cdot T \;\Big|\; \mathrm{H}^i_{\text{ét}}\big(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell\big)\Big)^{(-1)^{i+1}}\Big|_{T=p^{-s}}.$$

Date: September 27, 2021.

I would like to thank Liang Xiao for live-texing the notes during the lectures. I have made some edits afterwards, but have chosen to keep the roughness of the original notes. All the mistakes and inaccuracies are the responsibility of mine (Y.Z.), and I apologize to the reader for the unsystematic formatting and incomplete bibliography. Please send corrections or comments to yhzhu@umd.edu.

Here ℓ is a prime different from p, and Frob_p is the geometric Frobenius element at p. Since the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-action on $\mathrm{H}^i_{\text{ét}}\left(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell\right)$ is unramified at p, the action of Frob_p makes sense.

By the second expression, the zeta function does *not* depend on the choice of the integral model $\mathcal{X}_p$. Also, by the first expression, it does *not* depend on the choice of the prime ℓ.

Finall, we define

$$\zeta(X,s) = \prod_{\text{almost all } p} \zeta_p(X,s).$$

This of course depends on the finite set of primes that we have avoided, but we suppress that from the notation. The infinite product converges absolutely when $\Re s \gg 0$.

Conjecture 1.2. *The function $\zeta(X,s)$ admits a meromorphic continuation to $\mathbb{C}$.*

For example, if $X = \operatorname{Spec}\mathbb{Q}$, then $\zeta(X,s)$ (with the product taken over all primes) is Riemann's zeta function $\zeta(s)$. It has meromorphic continuation by Riemann.

Theorem 1.3 (Eichler–Shimura)**.** *Take $X = X_0(N)$ to be the (compactified) modular curve. Then*

$$\zeta(X,s) = \underbrace{\zeta(s)}_{\textit{comes from } \mathrm{H}^0} \cdot \underbrace{\zeta(s-1)}_{\textit{comes from } \mathrm{H}^2} \cdot \underbrace{\prod_{i=1}^{g} L(f_i,s)^{-1}}_{\textit{comes from } \mathrm{H}^1},$$

where $f_1, \dots, f_g$ form a Hecke eigen-basis of $S_2(\Gamma_0(N))$, and $L(f_i,s)$ is the L-function of f_i built from the Hecke eigenvalues of f_i.

Each of $L(f_i,s)$ admits a meromorphic continuation to $\mathbb{C}$ (by Hecke), so the same is true for $\zeta(X_0(N)), s)$.

Remark 1.4. If we replace $\mathrm{H}^i_{\text{ét}}\left(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell\right)$ by $\mathrm{H}^i_{\text{ét}}\left(X_{\overline{\mathbb{Q}}}, \mathcal{L}\right)$ for $\mathcal{L}$ a suitable local system on X (built from representations of $G = \mathrm{GL}_2$), then we see higher weight modular forms in the analogue of $\zeta(X,s)$.

1.5. **Towards Hasse–Weil zeta function for more general Shimura varieties.** Let (G,X) be a Shimura datum, i.e.

- G is a reductive group over $\mathbb{Q}$, e.g. GL_2,
- X is a $G(\mathbb{R})$-conjugacy class of $\mathbb{R}$-homomorphism $\mathbb{S} := \mathrm{Res}_{\mathbb{C}/\mathbb{R}}\,\mathbb{G}_m \to G_{\mathbb{R}}$

satisfying Deligne's axioms. Let $K \subset G(\mathbb{A}_f)$ be a compact open subgroup. Then we define (the complex points of) the Shimura variety

$$\mathrm{Sh}_K := \mathrm{Sh}_K(G,X)(\mathbb{C}) = G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K = \coprod_{i=1}^{m} X_i/\Gamma_i,$$

where each X_i is a connected component of X, and Γ_i is an arithmetic subgroup of $G(\mathbb{Q})$ which acts on X_i.

Assume that K is small enough. ("Neat" is the technical term.) As defined, one can show that $\mathrm{Sh}_K(G,X)$ is a complex manifold. By a theorem of Bailey and Borel, $\mathrm{Sh}_K(G,X)$ is a quasi-projective variety over $\mathbb{C}$. By later theorems of Shimura, Deligne, Borovoi, and Milne, $\mathrm{Sh}_K(G,X)$ admits a canonical model over a number field $E \subseteq \mathbb{C}$; this field E is called the *reflex field* of (G,X).

Remark 1.6. In a lot of cases namely the PEL type case (P = polarization, E = endomorphism, L = level structure), the canonical model of Sh_K over E can be directly defined as a moduli space of abelian varieties equipped with polarizations, endomorphism structures, and level structures. This will also lead to integral models. For example, modular curves, Siegel modular varieties, and some unitary Shimura varieties, belong to the PEL type case.

In fact, Shimura originally studied various types of moduli spaces of abelian varieties with additional structures; later Deligne generalized the idea of Shimura to give a more group theoretic approach to Shimura varieties, and introduced the concept of canonical models over the reflex field.

Remark 1.7. More recently, Kisin (hyperspecial level at $p > 2$), Madapusi Pera–Kim (hyperspecial level at $p = 2$), Kisin–Pappas (some parahoric level at p) have constructed integral models for Shimura varieties of abelian type, which are more general than PEL type but still do not cover all Shimura varieties as defined by Deligne.

Expectation 1.8. The reduction modulo p, or rather the set of $\overline{\mathbb{F}}_p$-points, of a suitable integral model also has a group theoretic description similar to $\mathrm{Sh}_K(\mathbb{C}) = G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$. This is the Langlands–Rapoport Conjecture.

Assumption 1.9. *For simplicity, in this course we will often assume $E = \mathbb{Q}$.*

Conjecture 1.10. *The Hasse–Weil ζ-function of a Shimura vairiety can be expressed in terms of automorphic L-functions.*

1.11. **Langlands' idea to study the Hesse–Weil ζ-function of Shimura varieties.** The local zeta function $\zeta_p(\mathrm{Sh}_K, s)$ encodes $\{\#\mathscr{S}_K(\mathbb{F}_{p^n}) \mid n\}$, where $\mathscr{S}_K$ is a suitable integral model of Sh_K over $\mathbb{Z}_{(p)}$. If one wants to relate $\zeta_p(\mathrm{Sh}_K, s)$ with automorphic representations of G (roughly: subrepresentations of the right regular $G(\mathbb{A})$-representation on $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))$), one typically uses the trace formula of Selberg and Arthur relating spectral information on $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))$ with orbital integrals, i.e., integrals of some functions on $G(\mathbb{A})$ over a conjugacy class of $G(\mathbb{A})$ (roughly speaking).

Langlands' idea is to relate the set $\mathscr{S}_K(\mathbb{F}_{p^n})$ with the orbital integrals. At least in the PEL case, this amounts to counting abelian varieties with additional structures over a finite field in terms of orbital integrals.

Remark 1.12. When G/Z_G contains a $\mathbb{Q}$-split torus (e.g. $G = \mathrm{GL}_2$, $G/Z_G = \mathrm{PGL}_2 \supset \mathbb{G}_m$), Sh_K is not projective over E. In this case, we need to compactify the Shimura varieties in order to have the "correct" definition of the Hasse–Weil zeta function.[1]

Similarly, on the automorphic side, $G(\mathbb{Q})\backslash G(\mathbb{A})$ is not compact for such G. In this case, for a function $f \in \mathcal{C}_c^\infty(G(\mathbb{A}))$, the trace of f on $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))$ does not make sense. (One needs a certain truncation process.) The trace formula becomes an identity between two quantities whose definitions are really complicated.

Remark 1.13. For applications, we are not just satisfied with understanding how $\mathrm{Gal}(\overline{E}/E)$ acts on $\mathrm{H}^i_{\text{ét}}\left(\mathrm{Sh}_{K,\overline{E}}, \mathbb{Q}_\ell\right)$. Actually, we want to also understand the commuting actions of

[1] From the point of view of étale cohomology, there are at least three possible choices for defining the Hasse–Weil zeta function. The usual cohomology of Sh_K, the compact support cohomology of Sh_K, and the intersection cohomology of the canonical Baily–Borel compactification of Sh_K. It is the third one that best fits Langlands' idea of using the Arthur–Selberg trace formula.

$\mathrm{Gal}(\overline{E}/E)$ and the Hecke algebra $\mathcal{H}(G(\mathbb{A}_f)//K)$ on $\mathrm{H}^i_{\text{ét}}\Big(\mathrm{Sh}_{K,\overline{E}},\mathbb{Q}_\ell\Big)$. Here $\mathcal{H}(G(\mathbb{A}_f)//K)$ is the convolution algebra consisting of K-bi-invariant compactly supported smooth functions on $G(\mathbb{A}_f)$, and its action comes from the $G(\mathbb{A}_f)$-action on the tower $\varprojlim_K \mathrm{Sh}_K$. For this, we need to understand: for a fixed $f \in \mathcal{H}(G(\mathbb{A}_f)//K)$, the trace

$$\mathrm{Tr}\Big(f \times \mathrm{Frob}_p^a \,\Big|\mathrm{H}^i_{\text{ét}}\Big)$$

for all but finitely many p (depending on f).

For the fixed f and for almost all p, we have $K = K^pK_p$ with $K^p \subset G(\mathbb{A}_f^p)$ and $K_p \subset G(\mathbb{Q}_p)$, and we have $f = f^pf_p$ with $f^p \in \mathcal{H}(G(\mathbb{A}_f^p)//K^p)$ and $f_p = \mathbf{1}_{K_p} : G(\mathbb{Q}_p) \to \{0,1\}$. By linearity, it is enough to consider the case when $f^p = \mathbf{1}_{K^pgK^p}$ for some $g \in G(\mathbb{A}_f^p)$.

Then we have

$$\sum_i (-1)^i \mathrm{Tr}\Big(f \times \mathrm{Frob}_p^a \;\Big|\; \mathrm{H}^i_{\text{ét}}\Big) = \#\text{ fixed points of the correspondence}$$

$$\begin{tikzcd}
\mathscr{S}_{(K^p\cap g^{-1}K^pg)\cdot K_p,\overline{\mathbb{F}}_p} \arrow[r, "\mathrm{Frob}_p^a"] \arrow[d, "g"] & \mathscr{S}_{(K^p\cap g^{-1}K^pg)\cdot K_p,\overline{\mathbb{F}}_p} \arrow[d] \\
\mathscr{S}_{K^pK_p,\overline{\mathbb{F}}_p} & \mathscr{S}_{K^pK_p,\overline{\mathbb{F}}_p}
\end{tikzcd}$$

This is quite similar to computing the cardinality of $\mathscr{S}_K(\mathbb{F}_{p^n})$.

Remark 1.14. Instead of looking at $\mathrm{H}^i_{\text{ét}}(\mathrm{Sh}_{K,\overline{\mathbb{Q}}},\mathbb{Q}_\ell)$, we can also look at $\mathrm{H}^i_{\text{ét}}(\mathrm{Sh}_{K,\overline{\mathbb{Q}}},\mathcal{L})$ for a local system $\mathcal{L}$ associated with a representation of G. This generalization should be straightforward, and we will not spend too much time on it.

1.15. **Integral models.** Let (G,X) be a Shimura datum, with reflex field E. Let $K \subset G(\mathbb{A}_f)$ be an open compact subgroup. Fix a prime p such that

- $K = K^pK_p$ with $K^p \subset G(\mathbb{A}_f^p)$ and $K_p \subset G(\mathbb{Q}_p)$
- K_p is hyperspecial, i.e., there exists a connected reductive group scheme $\mathcal{G}$ over $\mathbb{Z}_p$ such that $\mathcal{G}_{\mathbb{Q}_p} \cong G_{\mathbb{Q}_p}$ and such that $K_p = \mathcal{G}(\mathbb{Z}_p) \subset G(\mathbb{Q}_p)$. (Equivalently, K_p is the stabilizer of a hyperspecial point in the Bruhat–Tits building.)

For fixed K, our assumptions on p are satisfied for almost all p.

For example, $G = \mathrm{GL}_2$, $K = \Big\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\widehat{\mathbb{Z}}) \;\Big|\; \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \mathbf{1} \pmod{N}\Big\}$. The assumptions on p are satisfied if $p \nmid N$ (by taking $\mathcal{G} = \mathrm{GL}_2/\mathbb{Z}_p$).

In the sequel, we will also tacitly assume that K^p is sufficiently small. (The technical condition is called "neat".) In practice the following condition is good enough: There exists a faithful representation $G \to \mathrm{GL}_n$ defined over $\mathbb{Q}$ such that the image of K^p inside $\mathrm{GL}_n(\mathbb{A}_f^p)$ is contained in

$$\Big\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\widehat{\mathbb{Z}}^{(p)}) \;\Big|\; \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \mathbf{1} \pmod{N}\Big\}$$

for some $N \geq 3$ coprime to p.

Let v be a place of E above p. We denote by $\mathcal{O}_{E,(v)}$ the localization of $\mathcal{O}_E$ at the prime v.

Expectation 1.16. There exists a "canonical" smooth integral model $\mathscr{S}_K$ over $\mathcal{O}_{E,(v)}$ of the E-scheme Sh_K.

Theorem 1.17 (Kisin [Kis10], Madapusi Pera–Kim [KMP16])**.** *This is true if (G,X) is of abelian type (which is more general than Hodge type).*

Also, it is expected that if Sh_K is proper, so should be $\mathscr{S}_K$. If Sh_K is not proper, we expect the Bailey–Borel compactification of Sh_K extends to a similar compactification of $\mathscr{S}_K$. These statements have been proved by Madapusi Pera [MP19] in the Hodge type case.

2. Lecture 2

2.1. **Canonicity of integral models.** We briefly explain what a "canonical" integral model means. In particular, if Sh_K is not projective, the idea is that we want to "forbid" arbitrarily deleting points from the special fiber.

Suppose given an element $g \in G(\mathbb{A}_f^p)$, and open compact subgroups $U^p, K^p \subset G(\mathbb{A}_f^p)$ such that $g^{-1}U^pg \subseteq K^p$. Then we obtain a morphism (defined over E and is finite étale)

$$[g] : \mathrm{Sh}_{U^pK_p} \longrightarrow \mathrm{Sh}_{K^pK_p}$$

which on $\mathbb{C}$-points is given by

$$\begin{aligned} G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/U^pK_p &\longrightarrow G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K^pK_p \\ (x,y) &\longmapsto (x,yg). \end{aligned}$$

We obtain a limit

$$\varprojlim_{K^p} \mathrm{Sh}_{K^pK_p} =: \mathrm{Sh}_{K_p}$$

with transition maps [1]. (This limit exists in the category of schemes.)

Implicitly, the integral models for fixed $K_p = \mathcal{G}(\mathbb{Z}_p)$ and different choices of K^p should satisfy: every morphism $[g] : \mathrm{Sh}_{U^pK_p} \to \mathrm{Sh}_{K^pK_p}$ as above extends uniquely to a finite étale morphism $\mathscr{S}_{U^pK_p} \to \mathscr{S}_{K^pK_p}$. Then we can form the inverse limit of $\mathcal{O}_{E,(v)}$-schemes

$$\mathscr{S}_{K_p} := \varprojlim_{K^p} \mathscr{S}_{K^pK_p}$$

whose generic fiber is identified with Sh_{K_p}. Moreover, the $G(\mathbb{A}_f^p)$-action on Sh_{K_p} also extends to $\mathscr{S}_{K_p}$.

Note: In order to characterize $\mathscr{S}_{K^pK_p}$ we just need to characterize $\mathscr{S}_{K_p}$ together with a $G(\mathbb{A}_f^p)$-action. This is because $\mathscr{S}_{K^pK_p} \cong \mathscr{S}_{K^pK_p}/K^p$. Moreover, the $G(\mathbb{A}_f^p)$-action on $\mathscr{S}_{K_p}$ is determined by the action on Sh_{K_p}, since for each K^p the generic fiber $\mathrm{Sh}_{K^pK_p}$ is Zariski dense in $\mathscr{S}_{K^pK_p}$.[2]

Now we need to characterize the $\mathcal{O}_{E,(v)}$-scheme $\mathscr{S}_{K_p}$.

Characterizing condition: For any $\mathcal{O}_{E,(v)}$-scheme T which is regular and formally smooth over $\mathcal{O}_{E,(v)}$, every E-morphism $T_E \to \mathrm{Sh}_{K_p}$ extends uniquely to an $\mathcal{O}_{E,(v)}$-morphism $T \to \mathscr{S}_{K_p}$.

(This looks a lot like the valuative criterion of properness.)

Example 2.2. To get a feel of what the inverse limit $\varprojlim_{K^p} \mathrm{Sh}_{K^pK_p}$ looks like, maybe one can try to understand the "set of connected components" of modular curves $X(N)$; it is $\operatorname{Spec} \mathbb{Q}(\zeta_N)$. Taking inverse limit over N relatively prime to p, we obtain $\operatorname{Spec} \mathbb{Q}(\zeta_N; p \nmid N)$.

[2] This denseness follows from the separatedness of $\mathscr{S}_{K^pK_p}$ over $\mathcal{O}_{E,(v)}$, which is known in all cases where $\mathscr{S}_{K^pK_p}$ has been constructed. In general one should impose the denseness as one of the axioms.

Remark 2.3. By the work of Lan–Stroh [LS18], in the abelian type case with no assumption on Sh_K being projective, we have

$$\mathrm{H}^i_{\text{ét},c}\Big(\mathrm{Sh}_{K,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell\Big) \cong \mathrm{H}^i_{\text{ét},c}\Big(\mathscr{S}_{K,\overline{\mathbb{F}}_p}, \mathbb{Q}_\ell\Big).$$

Note that we may apply the Lefschetz trace formula to the RHS to get

$$\sum_i (-1)^i \mathrm{Tr}\Big(\mathrm{Fr}_{q^a} \;\Big|\; \mathrm{H}^i_{\text{ét},c}\Big(\mathscr{S}_{K,\overline{\mathbb{F}}_p}, \mathbb{Q}_\ell\Big)\Big) = \#\mathscr{S}_K(\mathbb{F}_{q^a})$$

when $\mathbb{F}_q$ is the residue field of v.

2.4. **Conjectural formula for the number of points.** Here and later, we always assume that $K = K^pK_p$ with K_p hyperspecial and K^p sufficiently small (see above). We will state the general formula for now, and later will focus on the special case when $G = \mathrm{GL}_2$.

We assume the following:

- G_{der} is simply connected.
- The maximal $\mathbb{R}$-split torus in Z_G is $\mathbb{Q}$-split. (Sometimes, we say "Z_G is cuspidal"; this condition is automatic for Shimura varieties of Hodge type.)

(The above assumptions can be removed [KSZ], but it makes it a lot harder to state the conjectures.) For example, $G = \mathrm{GL}_2$ or GSp_{2g} satisfy these assumptions.

Conjecture 2.5 (Kottwitz, [Kot90])**.** *Let $\mathbb{F}_q$ is the residue field of v. Let m be a positive integer, and write $q^m = p^n$. We have*

$$\#\mathscr{S}_K(\mathbb{F}_{q^m}) = \sum_{(\gamma_0,\gamma,\delta)} c_1(\gamma_0,\gamma,\delta)\cdot c_2(\gamma_0,\gamma,\delta)\cdot \mathrm{O}_\gamma(\mathbf{1}_{K^p})\cdot \mathrm{TO}_\delta(f_n)$$

Here $(\gamma_0, \gamma, \delta)$ runs through a *certain subset* of $G(\mathbb{Q}) \times G(\mathbb{A}_f^p) \times G(\mathbb{Q}_{p^n})$ modulo certain equivalence relation $\sim$:

$$(\gamma_0,\gamma,\delta) \sim (\gamma_0',\gamma',\delta')$$

if the following three conditions are satisfied:

- γ_0 and γ_0' are conjugate in $G(\overline{\mathbb{Q}})$,
- γ and γ' are conjugate in $G(\mathbb{A}_f^p)$,
- δ and δ' are σ-conjugate in $G(\mathbb{Q}_{p^n})$, i.e., $\delta = c\delta'\sigma(c)^{-1}$ for some $c \in G(\mathbb{Q}_{p^n})$. Here, σ is the arithmetic p-Frobenius on $\mathbb{Q}_{p^n}$.

Now we explain the other terms in the formula:

- $c_1(\gamma_0,\gamma,\delta)$ is a volume term,
- $c_2(\gamma_0,\gamma,\delta)$ comes from the size of a certain Galois cohomology group,
- $\mathrm{O}_\gamma(\mathbf{1}_{K^p})$ is the orbital integral of $\mathbf{1}_{K^p} : G(\mathbb{A}_f^p) \to \{0,1\}$ on the conjugacy class of γ in $G(\mathbb{A}_f^p)$,
- $\mathrm{TO}_\delta(f_n)$ is the integral of f_n on the σ-conjugacy class of δ inside $G(\mathbb{Q}_{p^n})$; here $f_n : G(\mathbb{Q}_{p^n}) \to \{0,1\}$ is the characteristic function of a certain $\mathcal{G}(\mathbb{Z}_{p^n})$-double coset in $G(\mathbb{Q}_{p^n})$ determined by the Shimura datum (G, X).

2.6. **Conjecture in the special case of** GL_2**.** Take $(G,X)=(\mathrm{GL}_2,\mathcal{H}^{\pm})$. Take the open compact subgroup $K=K^pK_p$ where $K_p=\mathrm{GL}_2(\mathbb{Z}_p)$ and K^p is small enough. Here we in particular require that there exists $N\geq 3$ such that $p\nmid N$ and

$$K^p\subset\left\{\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\mathrm{GL}_2(\widehat{\mathbb{Z}}^{(p)})\middle|\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\equiv\mathbf{1}\pmod N\right\}.$$

In this case, the reflex field is $\mathbb{Q}$, and the integral model $\mathscr{S}_K=\mathscr{S}_{K^pK_p}$ over $\mathbb{Z}_{(p)}$ is characterized by: for each $\mathbb{Z}_{(p)}$-scheme R,

$$\mathscr{S}_K(R)=\left\{(\mathcal{E},\eta)\middle|\begin{array}{l}\mathcal{E}\text{ is an elliptic curve over }R,\\ \eta\text{ is a }K^p\text{-level structure}\end{array}\right\}/\text{isomorphisms}$$

Here a K^p*-level structure* means the following: For each connected component R_i of R and each geometric point $\bar{x}$ of R_i, we have a $\pi_1^{\text{ét}}(R_i,\bar{x})$-stable K^p-orbit of isomorphisms

$$\left(\widehat{\mathbb{Z}}^{(p)}\right)^{\oplus 2}\xrightarrow{\simeq}T^{(p)}(\mathcal{E}_{\bar{x}}),$$

and these should satisfy the natural compatibilities when we vary $\bar{x}$.

Recall: For F a field, two semisimple elements of $\mathrm{GL}_n(F)$ are conjugate in $\mathrm{GL}_n(F)$ if and only if they are conjugate in $\mathrm{GL}_n(\overline{F})$.

Theorem 2.7. *We have*

$$\#\mathscr{S}_K(\mathbb{F}_{p^n})=\sum_{(\gamma_0,\delta)}c_1(\gamma_0,\delta)\cdot\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p})\cdot\mathrm{TO}_\delta(f_n).$$

Here

- γ_0 is an element of $G(\mathbb{Q})$, up to conjugacy (which is the same as up to conjugacy by $\mathrm{GL}_2(\overline{\mathbb{Q}})$), and is $\mathbb{R}$-elliptic, i.e. $\gamma_0\in T(\mathbb{R})$ for T a maximal torus in $G_\mathbb{R}$ such that $T(\mathbb{R})$ is compact modulo $Z_G(\mathbb{R})$. In other words, either γ_0 is central, i.e. $\gamma_0=\left(\begin{smallmatrix}\lambda&\\&\lambda\end{smallmatrix}\right)$ with $\lambda\in\mathbb{Q}^\times$, or the characteristic polynomial of γ_0 is irreducible over $\mathbb{R}$.
- (The element $\gamma\in G(\mathbb{A}_f^p)$ in the general conjecture is determined by γ_0 up to conjugacy, so it disappears here.)
- $\delta\in G(\mathbb{Q}_{p^n})$ such that the "naive norm" $\delta\cdot\sigma(\delta)\cdot\sigma^2(\delta)\cdots\sigma^{n-1}(\delta)\in G(\mathbb{Q}_{p^n})$ is conjugate to γ_0. This δ is taken up to σ-conjugacy in $G(\mathbb{Q}_{p^n})$.
- Writing G_{γ_0} for the centralizer of γ_0 in $G=\mathrm{GL}_2$, we define the orbital integral

$$\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p})=\int_{G_{\gamma_0}(\mathbb{A}_f^p)\backslash G(\mathbb{A}_f^p)}\mathbf{1}_{K^p}(x^{-1}\gamma_0x)dx.$$

  Here the measure dx is the quotient Haar measure on $G_{\gamma_0}(\mathbb{A}_f^p)\backslash G(\mathbb{A}_f^p)$ given by the Haar measure on $G(\mathbb{A}_f^p)$ normalized such that $\mathrm{vol}(K^p)=1$, and an arbitrary Haar measure on $G_{\gamma_0}(\mathbb{A}_f^p)$.
- Similarly, writing $G(\mathbb{Q}_{p^n})_{\delta\sigma}$ for the σ-centralizer of δ, namely $\left\{g\in G(\mathbb{Q}_{p^n})\,\middle|g\delta\sigma(g)^{-1}=\delta\right\}$, we define the twisted orbital integral

$$\mathrm{TO}_\delta(f_n)=\int_{G(\mathbb{Q}_{p^n})_{\delta\sigma}\backslash G(\mathbb{Q}_{p^n})}f_n(x^{-1}\delta\sigma(x))dx$$

  Here the quotient measure is given by the Haar measure on $G(\mathbb{Q}_{p^n})$ normalized such that $\mathrm{vol}(\mathrm{GL}_2(\mathbb{Z}_{p^n}))=1$, and an arbitrary Haar measure on $G(\mathbb{Q}_{p^n})_{\delta\sigma}$.

Here $f_n : \mathrm{GL}_2(\mathbb{Q}_{p^n}) \to \{0,1\}$ is the characteristic function of

$$\mathrm{GL}_2(\mathbb{Z}_{p^n})\begin{pmatrix} p & \\ & 1 \end{pmatrix}\mathrm{GL}_2(\mathbb{Z}_{p^n}) \subseteq \mathrm{GL}_2(\mathbb{Q}_{p^n}).$$

(Recall the Cartan decomposition

$$\mathrm{GL}_2(\mathbb{Q}_{p^n}) = \coprod_{a,b\in\mathbb{Z}, a\geq b} \mathrm{GL}_2(\mathbb{Z}_{p^n})\begin{pmatrix} p^a & \\ & p^b \end{pmatrix}\mathrm{GL}_2(\mathbb{Z}_{p^n}),$$

and $\begin{pmatrix} p & \\ & 1 \end{pmatrix} = \mu(p)$ where $\mu : z \mapsto \begin{pmatrix} z & \\ & 1 \end{pmatrix}$ is a Hodge cocharacter for the Shimura datum.)

In fact, $G(\mathbb{Q}_{p^n})_{\delta\sigma}$ is the $\mathbb{Q}_p$-points of a reductive group $J_{n,\delta}$ over $\mathbb{Q}_p$ defined as follows. For R an $\mathbb{Q}_p$-algebra,

$$J_{n,\delta}(R) = \Big\{ g \in G(\mathbb{Q}_{p^n} \otimes_{\mathbb{Q}_p} R) \;\Big|\; g\delta\sigma(g)^{-1} = \delta \Big\}.$$

Alternatively, we can define a new group $G_n := \mathrm{Res}_{\mathbb{Q}_{p^n}/\mathbb{Q}_p} G$ which is a reductive group over $\mathbb{Q}_p$, and $\theta \in \mathrm{Aut}_{\mathbb{Q}_p}(G_n)$ corresponding to $\sigma \in \mathrm{Gal}(\mathbb{Q}_{p^n}/\mathbb{Q}_p)$. Then GL_2 is the centralizer of θ on G_n, yet $J_{n,\delta}$ is the centralizer of $\mathrm{Ad}_\delta \circ \theta$.

- We now define $c_1(\gamma_0, \delta)$. Given (γ_0, δ) we can define a unique inner form I of G_{γ_0} such that
  - $I_\mathbb{R}$ is compact modulo Z_G,
  - $I_{\mathbb{Q}_\ell} \cong G_{\gamma_0}$ for $\ell \neq p$,
  - $I_{\mathbb{Q}_p} \cong J_{n,\delta}$.

  Note: in the above the isomorphisms should mean isomorphisms as inner forms.

  We define $c_1(\gamma_0, \delta)$ to be the volume of $I(\mathbb{Q})\backslash I(\mathbb{A}_f)$, with respect to the counting measure on $I(\mathbb{Q})$ and the Haar measure on $I(\mathbb{A}_f) \cong G_{\gamma_0}(\mathbb{A}_f^p) \times J_{n,\delta}(\mathbb{Q}_p)$ fixed before (in the definition of O_γ and TO_δ). Here $I(\mathbb{Q})\backslash I(\mathbb{A}_f)$ is a nice space: From the fact that $I_\mathbb{R}$ is compact modulo Z_G and Z_G is cuspidal, we know that $I(\mathbb{Q}) \subset I(\mathbb{A}_f)$ is discrete and the quotient has finite volume.

2.8. **Constraints on γ_0.** If γ_0 shows up, then $\det\gamma_0 = p^n$. This is because there exists $\delta \in G(\mathbb{Q}_{p^n})$ such that $\gamma_0 \sim \delta\sigma(\delta)\cdots\sigma^{n-1}(\delta)$. So

$$\det\gamma_0 = \mathrm{Nm}_{\mathbb{Q}_{p^n}/\mathbb{Q}_p}(\det\delta).$$

If $\mathrm{TO}_\delta(f_n) \neq 0$, we must have $c \in G(\mathbb{Q}_{p^n})$ such that $c\delta\sigma(c)^{-1} \in \mathrm{GL}_2(\mathbb{Z}_{p^n})\begin{pmatrix} p & \\ & 1 \end{pmatrix}\mathrm{GL}_2(\mathbb{Z}_{p^n})$. Thus $\det c \cdot \det\delta \cdot \sigma(\det c)^{-1}$ has p-adic valuation 1. This further implies that $\det\delta$ has p-adic valuation 1, and thus $\det\gamma_0 \in \mathbb{Q}^\times$ has p-adic valuation n.

Similarly, if $\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \neq 0$, then γ_0 is conjugate to some elements in $K^p \subseteq \mathrm{GL}_2(\widehat{\mathbb{Z}}^{(p)})$. So $\det\gamma_0$ has ℓ-adic valuation 0 for every $\ell \neq p$. Yet $\det\gamma_0 > 0$ because the characteristic polynomial is irreducible over $\mathbb{R}$.

Putting everything together, we obtain $\det\gamma_0 = p^n$.

Exercise 2.9. Show that only finitely many γ_0's (up to conjugacy) show up in the formula.

Moreover, we point out that the σ-conjugacy class of δ is determined by γ_0 and the relation $\delta\sigma(\delta)\cdots\sigma^{n-1}(\delta) \sim \gamma_0$. (This is a statement special for GL_2. We will prove this next time.) Therefore the inner form I of G_{γ_0}, whose definition depends on γ_0 and δ, is determined by γ_0.

2.10. **Classification of γ_0's.**

(1) Central case: (only appears if n is even) $\gamma_0 = \begin{pmatrix} p^{n/2} & \\ & p^{n/2} \end{pmatrix}$ or $\begin{pmatrix} -p^{n/2} & \\ & -p^{n/2} \end{pmatrix}$.

In this case, $G_{\gamma_0} = G = \mathrm{GL}_2$, $I = D^\times$, where D is the quaternion algebra ramified at p and ∞.

(2) Non-central case: In this case, set $F := \mathbb{Q}(\text{eigenvalues of } \gamma_0) = \mathbb{Q}(\gamma_0)$, which is an imaginary quadratic field (because γ_0 is $\mathbb{R}$-elliptic). Then,

$$G_{\gamma_0} \cong \mathrm{Res}_{F/\mathbb{Q}} \mathbb{G}_m = \text{“}F^\times\text{''} \hookrightarrow \mathrm{GL}_2$$

where one defines the embedding by viewing F as $\mathbb{Q} \oplus \mathbb{Q}$ and considering the left multiplication action of $F^\times$ on F. Since G_{γ_0} is a torus and I is an inner form of it, we have $I = G_{\gamma_0}$.

Exercise 2.11. Compute $\#\mathscr{S}_K(\mathbb{F}_5)$ with $K = \widehat{\Gamma(4)} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\widehat{\mathbb{Z}}) \;\middle|\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \mathbf{1} \pmod 4 \right\}$.

3. Lecture 3

Lemma 3.1. *Let $\gamma_0 \in \mathrm{GL}_2(\mathbb{Q}_p)$ be a semisimple element, and $\delta \in \mathrm{GL}_2(\mathbb{Q}_{p^n})$ be such that $\delta \cdot \sigma(\delta) \cdots \sigma^{n-1}(\delta) \sim \gamma_0$. Then the σ-conjugacy class of δ is uniquely determined by γ_0.*

Proof. We use the following two facts.

Fact 1: Suppose G is a reductive group over $\mathbb{Q}_p$ with G_{der} simply connected. Let $\gamma_0 \in G(\mathbb{Q}_p)$ be a semisimple element, and suppose $\delta \in G(\mathbb{Q}_{p^n})$ is such that $\delta\sigma(\delta)\cdots\sigma^{n-1}(\delta)$ is conjugate to γ_0 in $G(\overline{\mathbb{Q}}_p)$. Then $J_{n,\delta}$ is an inner form of G_{γ_0}.

Fact 2: Keep the setting in Fact 1. The set

$$\left\{ \delta' \in G(\mathbb{Q}_{p^n}) \;\middle|\; \delta'\sigma(\delta')\cdots\sigma^{n-1}(\delta') \sim \gamma_0 \right\} \big/ \sigma\text{-conj.}$$

(where $\sim$ means $G(\overline{\mathbb{Q}}_p)$-conjugate) is in bijection with

$$\mathrm{Ker}\left(\mathrm{H}^1(\mathbb{Q}_p, J_{n,\delta}) \to \mathrm{H}^1(\mathbb{Q}_p, \mathrm{Res}_{\mathbb{Q}_{p^n}/\mathbb{Q}_p} G) \right).$$

(Recall that $J_{n,\delta} \subset \mathrm{Res}_{\mathbb{Q}_{p^n}/\mathbb{Q}_p} G$.) Indeed, if δ and δ' both satisfy the similar condition, then there exists $g \in \mathrm{Res}_{\mathbb{Q}_{p^n}/\mathbb{Q}_p} G(\overline{\mathbb{Q}}_p)$ such that $g\delta\theta(g)^{-1} = \delta'$, where θ is the $\mathbb{Q}_p$-automorphism of $\mathrm{Res}_{\mathbb{Q}_{p^n}/\mathbb{Q}_p} G$ corresponding to $\sigma \in \mathrm{Gal}(\mathbb{Q}_{p^n}/\mathbb{Q}_p)$. Then we obtain a cocycle $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to J_{n,\delta}(\overline{\mathbb{Q}}_p)$ sending $\tau \mapsto g^{-1}\tau(g)$. This determines an element in the above kernel.

In our special case $G = \mathrm{GL}_2$, we claim that $\mathrm{H}^1(\mathbb{Q}_p, J_{n,\delta}) = 0$. This uses the following

Fact 3: Suppose F is a non-archimedean field of characteristic zero, and J, J' are reductive groups over F that are inner forms of each other. Then there is a canonical isomorphism $\mathrm{H}^1(F, J) \cong \mathrm{H}^1(F, J')$. (This is a deep result, based on Kneser's theorem that $\mathrm{H}^1(F, J) = 0$ for any semisimple simply connected group J over F.)

Now using Fact 3, we see that

$$\mathrm{H}^1(\mathbb{Q}_p, J_{n,\delta}) \cong \mathrm{H}^1(\mathbb{Q}_p, G_{\gamma_0})$$

But for $\gamma_0 \in G(\mathbb{Q}_p) = \mathrm{GL}_2(\mathbb{Q}_p)$ semisimple, there are only three possibilities of G_{γ_0}:

- if γ_0 is central, $G_{\gamma_0} = G$,

- if γ_0 is non-central and the characteristic polynomial is irreducible over $\mathbb{Q}_p$, $G_{\gamma_0} = \mathrm{Res}_{F/\mathbb{Q}_p} \mathbb{G}_m$, where F is the quadratic extension of $\mathbb{Q}_p$ generated by the eigenvalues of γ_0,
- if the characteristic polynomial of γ_0 has two distinct roots over $\mathbb{Q}_p$, then $G_{\gamma_0} = \mathbb{G}_m \times \mathbb{G}_m$.

Then $\mathrm{H}^1(\mathbb{Q}_p, J_{n,\delta}) = \mathrm{H}^1(\mathbb{Q}_p, G_{\gamma_0}) = 0$ by Hilbert 90 and Shapiro's lemma. □

3.2. **General way of computing (twisted) orbital integrals.** (We will only discuss the twisted case, since one can recover the untwisted case by setting $n = 1$.) Let G be a reductive group over $\mathbb{Q}_p$ and $n \in \mathbb{N}$. Let $\delta \in G(\mathbb{Q}_{p^n})$. We assume the following:

- $J_{n,\delta}$ is a reductive group,
- the σ-conjugacy class of δ in $G(\mathbb{Q}_{p^n})$ is a closed subset,

These conditions are always satisfied in the point counting formula, but let us indicate some sufficient conditions which imply these assumptions. For $n = 1$, it suffices to require that δ is semi-simple. For $n \geq 2$, it suffices to require that the automorphism $\mathrm{Ad}_\delta \circ \theta$ of $\mathrm{Res}_{\mathbb{Q}_{p^n}/\mathbb{Q}_p} G$ is semi-simple.

Now fix Haar measures dg on $G(\mathbb{Q}_{p^n})$ and dj on $J_{n,\delta}(\mathbb{Q}_p)$, and fix a function $f \in \mathcal{C}_c^\infty(G(\mathbb{Q}_{p^n}))$. Then we define the twisted orbital integral

$$\mathrm{TO}_\delta(f) := \int_{J_{n,\delta}(\mathbb{Q}_p)\backslash G(\mathbb{Q}_{p^n})} f(x^{-1}\delta\sigma(x))dx,$$

where dx is the quotient measure dg/dj. Fix a sufficiently small compact open *σ-invariant* subgroup $K \subset G(\mathbb{Q}_{p^n})$ such that f is K-bi-invariant. (This always exists.) Then

$$\mathrm{TO}_\delta(f) = \sum_{x \in J_{n,\delta}(\mathbb{Q}_p)\backslash G(\mathbb{Q}_{p^n})/K} f(x^{-1}\delta\sigma(x)) \cdot \frac{\mathrm{vol}_{dg}(K)}{\mathrm{vol}_{dj}(xKx^{-1} \cap J_{n,\delta}(\mathbb{Q}_p))}$$

Here, one can prove that the sum is taken over a finite set[3], and that the evaluation $f(x^{-1}\delta\sigma(x))$ makes sense (as f is K-bi-invariant and left multiplying by elements in $J_{n,\delta}(\mathbb{Q}_p)$ onto x does not change the value of $x^{-1}\delta\sigma(x)$). The volumes are computed with respect to the indicated Haar measures.

The hard part when applying this formula is usually to make explicit the indexing set $J_{n,\delta}(\mathbb{Q}_p)\backslash G(\mathbb{Q}_{p^n})/K$ of the summation.

3.3. **Proof of Theorem 2.7.** Set $q = p^n$. First, recall that $\mathscr{S}_K(\mathbb{F}_q)$ is the set of isomorphism classes of pairs (E, η), where E is an elliptic curve over $\mathbb{F}_q$, and η is a $\mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$-stable K^p-orbit of isomorphisms

$$\widehat{\mathbb{Z}}^{(p)} \oplus \widehat{\mathbb{Z}}^{(p)} \xrightarrow{\simeq} T^{(p)}(E_{\overline{\mathbb{F}}_q}).$$

(Here K^p acts on $\widehat{\mathbb{Z}}^{(p)} \oplus \widehat{\mathbb{Z}}^{(p)}$ via $K^p \subset \mathrm{GL}_2(\mathbb{Z}^{(p)})$, and $\mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ acts on $T^{(p)}(E_{\overline{\mathbb{F}}_q})$.)

The basic idea is that starting with E over $\mathbb{F}_q$, we can already cook up the pair (γ_0, δ) in the following way. Given E, for any prime $\ell \neq p$, the ℓ-adic Tate module $T_\ell(E) := T_\ell(E_{\overline{\mathbb{F}}_q})$

[3] Essentially, we are looking at the intersection

$$\sigma\text{-conjugacy class of } x \,\cap\, \mathrm{supp}(f)$$

modulo the action of K. Now the σ-conjugacy class of x in $G(\mathbb{Q}_{p^n})$ is closed (by our assumption) and $\mathrm{supp}(f)$ is compact, so their intersection is compact. Yet K is open so the quotient is finite.

is acted on by the q-Frobenius endomorphism $\pi \in \mathrm{End}(E)$. If we choose a basis of $T_\ell(E)$, then π is given by some matrix $\gamma_\ell \in \mathrm{GL}_2(\mathbb{Q}_\ell)$.

Fact: the characteristic polynomial of γ_ℓ has coefficients in $\mathbb{Z}$ and is independent of ℓ. In more fancy form, consider $\mathbb{Q}[\pi] \subseteq \mathrm{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$. Then $\mathbb{Q}[\pi]$ is a field, and we denote by $\min(\pi; \mathbb{Q})$ the minimal polynomial of π over $\mathbb{Q}$. Then the characteristic polynomial of γ_ℓ is $\min(\pi; \mathbb{Q})$ or $\min(\pi; \mathbb{Q})^2$ (for any $\ell \neq p$).

We define an element $\gamma_0 \in G(\mathbb{Q})$ whose minimal polynomial is $\min(\pi; \mathbb{Q})$; such γ_0 is unique up to conjugacy, and moreover $\gamma_0 \sim \gamma_\ell$ for every $\ell \neq p$.

Observation: This γ_0 is $\mathbb{R}$-elliptic, i.e., it is either central or has characteristic polynomial irreducible over $\mathbb{R}$. This is because the characteristic polynomial of γ_0 is $T^2 - \mathrm{Tr}(\gamma_0)T + \det(\gamma_0)$, and the $\mathbb{R}$-elliptic condition is equivalent to this polynomial having non-positive discriminant, i.e., $\mathrm{Tr}(\gamma_0)^2 \leq 4\det(\gamma_0)$. By general facts about elliptic curves over $\mathbb{F}_q$, we know that

$$\mathrm{Tr}(\gamma_0) = q + 1 - \#F(\mathbb{F}_q), \quad \det(\gamma_0) = q.$$

Thus the last inequality is precisely Hasse's bound: $|q + 1 - \#E(\mathbb{F}_q)| \leq 2\sqrt{q}$.

To construct δ, we need to consider the Dieudonné module $M_0 = M_0(E)$ of $E[p^\infty]$. Recall that this is a free $\mathbb{Z}_q$-module of rank 2 together with a σ-linear map $F : M_0 \to M_0$ (i.e. $F(a \cdot x) = \sigma(a)F(x)$ for $a \in \mathbb{Z}_q$ and $x \in M_0$), and a σ^{-1}-linear map $V : M_0 \to M_0$ such that $FV = VF = p$. If we choose a basis of M_0, then F becomes the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \longmapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \sigma(x) \\ \sigma(y) \end{pmatrix}$$

for some fixed $\delta = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q}_q)$. The element δ is independent of the choice of basis up to σ-conjugacy. Note that V can be reconstructed from F as long as F satisfies the condition $M_0 \supset F(M_0) \supset pM_0$. In our case, since M_0 comes from an elliptic curve, we have $M_0 \supset F(M_0) \supset pM_0$ *and* $\dim_{\mathbb{F}_q} F(M_0)/pM_0 = 1$. We say that M_0 is a *Dieudonné module of height* 2 *and dimension* 1. Correspondingly, $\delta \in \mathrm{GL}_2(\mathbb{Z}_q)\left(\begin{smallmatrix} p & \\ & 1 \end{smallmatrix}\right)\mathrm{GL}_2(\mathbb{Z}_q)$.[4]

Moreover, $\delta \cdot \sigma(\delta) \cdots \sigma^{n-1}(\delta) \sim \gamma_0$ holds (by some general theory of elliptic curves).

Summary: We have constructed a map from the set of elliptic curves E over $\mathbb{F}_q$ to the set of pairs (γ_0, δ) up to conjugacy and σ-conjugacy, respectively. Thus, we may then write

$$\#\mathscr{S}_K(\mathbb{F}_q) = \sum_{(\gamma_0, \delta)} N(\gamma_0, \delta), \quad \text{with} \quad N(\gamma_0, \delta) = \#\Big\{(E, \eta) \;\Big|\; E \text{ gives rise to } (\gamma_0, \delta)\Big\}.$$

Now we have two things to prove:

(1) Suppose $N(\gamma_0, \delta) \neq 0$. We need to prove $N(\gamma_0, \delta) = c_1(\gamma_0, \delta) \cdot \mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n)$.

(2) If (γ_0, δ) is such that $\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n) \neq 0$, then γ_0 comes form some elliptic curve E over $\mathbb{F}_q$.

[4] In general, a Dieudonné module M over $\mathbb{Z}_q$ is a finite free $\mathbb{Z}_q$-module equipped with a σ-linear operator F and a σ^{-1}-linear operator V such that $FV = VF = p$ on M. The existence of V is equivalent to $pM \subset FM \subset M$. We define the *height* of M to be $\mathrm{rank}_{\mathbb{Z}_q} M$, and define the *dimension* of M to be $\dim_{\mathbb{F}_q}(FM/pM)$. If we take a $\mathbb{Z}_q$-basis of M, then F is given by $\delta \circ \sigma$ for some $\delta \in \mathrm{GL}_n(\mathbb{Q}_q)$. We have $\delta \in \mathrm{GL}_n(\mathbb{Z}_q)\mathrm{diag}(p, \cdots, p, 1, \cdots, 1)\,\mathrm{GL}_n(\mathbb{Z}_q)$, where the number of 1's is equal to the dimension of M in the above sense.

The rest of this lecture is devoted to proving (1). We take some elliptic curve E_0 over $\mathbb{F}_q$, which gives rise to invariants (γ_0, δ). We need the following:

Theorem 3.4 (Honda–Tate). *If $E/\mathbb{F}_q$ gives rise to the same invariant (γ_0, δ) (up to conjugacy and σ-conjugacy) then E is quasi-isogenous to E_0. The converse is also true. In particular,*

$$N(\gamma_0, \delta) = \#\Big\{(E, \eta) \;\Big|\; E \textit{ is quasi-isogenous to } E_0\Big\}.$$

To use this theorem, we define

$$Y := \Big\{(E, \eta, \iota) \;\Big|\; \iota \text{ is a quasi isogeny } E \to E_0\Big\}.$$

Define the algebraic group I_{E_0} over $\mathbb{Q}$: for every $\mathbb{Q}$-algebra R,

$$I_{E_0}(R) = \Big(\operatorname{End}_{\mathbb{F}_q}(E_0) \otimes_{\mathbb{Z}} R\Big)^{\times}.$$

Then I_{E_0} is a reductive group over $\mathbb{Q}$. For example, $I_{E_0}(\mathbb{Q})$ is the set of self-quasi-isogenies of E_0.

Then we deduce

$$N(\gamma_0, \delta) = \Big|I_{E_0}(\mathbb{Q})\backslash Y\Big|.$$

We now define some "local variants" of Y. We define Y^p to be the set of $\mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$-stable K^p-orbits of embeddings $(\widehat{\mathbb{Z}}^{(p)})^{\oplus 2} \hookrightarrow T^{(p)}(E_0) \otimes_{\mathbb{Z}} \mathbb{Q}$. Note that the right hand side of the embedding is non-canonically isomorphic to $(\mathbb{A}_f^p)^{\oplus 2}$ and is equipped with an action of $\mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$. The group K^p only acts on the left hand side. We do not require the images of the embeddings to be $T^{(p)}(E_0)$. Observe that Y^p is the same as the set of π-stable K^p-orbits of such embeddings, which is further the same as, after choosing a basis of $T^{(p)}(E_0)$, the set of $(\gamma_\ell)_{\ell \neq p}$-stable K^p-orbits of embeddings $(\widehat{\mathbb{Z}}^{(p)})^{\oplus 2} \hookrightarrow (\mathbb{A}_f^p)^{\oplus 2}$. (Here $(\gamma_\ell)_{\ell \neq p} \in \mathrm{GL}_2(\mathbb{A}_f^p)$ is the matrix of π.) Since a K^p-orbit of embeddings $(\widehat{\mathbb{Z}}^{(p)})^{\oplus 2} \hookrightarrow (\mathbb{A}_f^p)^{\oplus 2}$ is given by $g \in \mathrm{GL}_2(\mathbb{A}_f^p)$ up to right multiplication by K^p, we see that Y^p is further identified with the set

$$\Big\{g \in \mathrm{GL}_2(\mathbb{A}_f^p)/K^p \;\Big|\; g^{-1}(\gamma_\ell)_{\ell \neq p} g \in K^p\Big\}$$

We also define

$$Y_p = \left\{\mathbb{Z}_q\text{-lattices } \Lambda \subset M_0(E_0)[\tfrac{1}{p}] \;\middle|\; p\Lambda \subset F\Lambda \subset \Lambda \text{ and } \dim_{\mathbb{F}_q} F\Lambda/p\Lambda = 1.\right\}$$

Here $F : M_0(E_0)[\frac{1}{p}] \to M_0(E_0)[\frac{1}{p}]$ is induced by the F on $M_0(E_0)$. In other words, we require that (Λ, F) is a Dieudonné module of type height 2 and dimension 1 in its own right. After choosing a basis of $M_0(E_0)$ (as how we get the element δ), this set Y_p is the same as

$$\Big\{\mathbb{Z}_q\text{-lattices } \Lambda \subset \mathbb{Q}_q^{\oplus 2} \;\Big|\; p\Lambda \subset \delta \cdot \sigma(\Lambda) \subset \Lambda \text{ and } \dim_{\mathbb{F}_q}(\delta \cdot \sigma(\Lambda)/p\Lambda) = 1\Big\}$$

which is the then the same as (by setting $\Lambda = g \cdot \mathbb{Z}_q^{\oplus 2}$)

$$\Big\{g \in \mathrm{GL}_2(\mathbb{Q}_q)/\mathrm{GL}_2(\mathbb{Z}_q) \;\Big|\; g^{-1}\delta\sigma(g) \in G(\mathbb{Z}_q)\Big(\begin{smallmatrix} p & \\ & 1 \end{smallmatrix}\Big)G(\mathbb{Z}_q)\Big\}.$$

There is a natural map

$$Y \longrightarrow Y^p \times Y_p$$

sending a tuple (E,η,ι) to the following element in $Y^p \times Y_p$: the composition

$$(\widehat{\mathbb{Z}}^{(p)})^{\oplus 2} \xrightarrow{\eta} T^{(p)}(E)_{\mathbb{Q}} \xrightarrow{\iota} T^{(p)}(E_0)_{\mathbb{Q}}$$

defines an element in Y^p. Similarly, we have an F-equivariant map

$$\iota : M_0(E)[\tfrac{1}{p}] \longrightarrow M_0(E_0)[\tfrac{1}{p}],$$

and $\iota(M_0(E)) \subset M_0(E_0)[\frac{1}{p}]$ is a lattice Λ belonging to Y_p (which follows from the fact that $M_0(E)$ is itself a Dieudonné module of height 2 and dimension 1).

Theorem 3.5 (Tate's isogeny theorem)**.** *The natural map $Y \to Y^p \times Y_p$ constructed above is a bijection. Moreover the natural action of $I_{E_0}(\mathbb{Q})$ on Y corresponds to the action of $I_{E_0}(\mathbb{Q})$ on $Y^p \times Y_p$ given by the composite map*

$$I_{E_0}(\mathbb{Q}) \hookrightarrow I_{E_0}(\mathbb{A}_f) \cong I(\mathbb{A}_f) = G_{\gamma_0}(\mathbb{A}_f^p) \times J_{n,\delta}(\mathbb{Q}_p)$$

followed by the natural $G_{\gamma_0}(\mathbb{A}_f^p) \times J_{n,\delta}(\mathbb{Q}_p)$-action on $Y^p \times Y_p$. (Here I is the $\mathbb{Q}$-group associated with (γ_0,δ) as usual, and actually we have $I_{\mathbb{A}_f} \cong (I_{E_0})_{\mathbb{A}_f}$.)

Then $N(\gamma_0,\delta) = \# I_{E_0}(\mathbb{Q}) \backslash Y^p \times Y_p$, where $I_{E_0}(\mathbb{Q})$ acts as indicated above. Using the group-theoretic descriptions of Y^p and Y_p, one can prove as an exercise that the last quantity is equal to

$$\mathrm{vol}(I_{E_0}(\mathbb{Q}) \backslash I(\mathbb{A}_f)) \cdot \mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \cdot \mathrm{TO}(f_n).$$

In addition, one can prove that

$$\mathrm{vol}(I_{E_0}(\mathbb{Q}) \backslash I(\mathbb{A}_f)) = \mathrm{vol}(I(\mathbb{Q}) \backslash I(\mathbb{A}_f)) = c_1(\gamma_0,\delta).$$

Note that $I(\mathbb{Q}) \hookrightarrow I(\mathbb{A}_f)$ and $I_{E_0}(\mathbb{Q}) \hookrightarrow I(\mathbb{A}_f)$ only agree up to $I(\mathbb{A}_f)$-conjugacy, but this does not matter for computing volumes.

4. Lecture 4

4.1. **Continuation of the proof of Theorem 2.7.**

Our target formula is

$$\# \mathscr{S}_K(\mathbb{F}_{p^n}) = \sum_{(\gamma_0,\delta)} c_1(\gamma_0,\delta) \cdot \mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n).$$

Last time: If γ_0 comes from some elliptic curve E_0 over $\mathbb{F}_q = \mathbb{F}_{p^n}$, then

$$\#\Big\{(E,\eta) \in \mathscr{S}_K(\mathbb{F}_q) \;\Big|\; E \text{ is associated with } \gamma_0\Big\} = c_1(\gamma_0,\delta) \cdot \mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n).$$

Today, we prove statement (2) from the last lecture. Namely, if a pair (γ_0,δ) (with γ_0 $\mathbb{R}$-elliptic, and $\gamma_0 \sim \delta\sigma(\delta)\cdots\sigma^{n-1}(\delta)$) is such that

$$\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n) \neq 0$$

then γ_0 indeed comes from some elliptic curve E over $\mathbb{F}_q = \mathbb{F}_{p^n}$. In the following, we fix an eigenvalue $\pi \in \overline{\mathbb{Q}}$ of γ_0.

Step I: We show that π is a Weil q-number (i.e., an algebraic integer π such that for every complex embedding $\mathbb{Q}(\pi) \hookrightarrow \mathbb{C}$, the absolute value of π is $\sqrt{q}$.)

If γ_0 is central, then $\pi = \sqrt{q} \in \mathbb{Q}$, and clearly it is a Weil q-number. In the non-central case, write $\bar{\pi}$ for the unique Galois conjugate of π. Recall that $\det\gamma_0 = q$ (if $\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p})\mathrm{TO}_\delta(f_n) \neq 0$). This implies that $\pi\bar{\pi} = q$. Thus we only need to show that π is an algebraic integer. It then suffices to show that the trace of γ_0 lies in $\mathbb{Z}$. (We already know that it is in $\mathbb{Q}$).

Over $\mathbb{Q}_{p^n}$, γ_0 is conjugate to $\delta\sigma(\delta)\cdots\sigma^{n-1}(\delta)$ and δ is σ-conjugate to some element in $\mathrm{GL}_2(\mathbb{Z}_q)\begin{pmatrix} p & \\ & 1\end{pmatrix}\mathrm{GL}_2(\mathbb{Z}_q)$. WLOG, $\delta \in \mathrm{GL}_2(\mathbb{Z}_q)\begin{pmatrix} p & \\ & 1\end{pmatrix}\mathrm{GL}_2(\mathbb{Z}_q)$ and thus $\delta\sigma(\delta)\cdots\sigma^{n-1}(\delta) \in \mathrm{M}_2(\mathbb{Z}_q)$. In particular, $\mathrm{Tr}(\gamma_0) \in \mathbb{Z}_q \cap \mathbb{Q}$, i.e. the p-adic valuation of $\mathrm{Tr}(\gamma_0)$ is non-negative.

Similarly, γ_0 is conjugate to some element in K^p as $\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \neq 0$, so $\mathrm{Tr}(\gamma_0)$ is equal to the trace of some element in $K^p \subseteq \mathrm{GL}_2(\widehat{\mathbb{Z}}^{(p)})$, i.e. the ℓ-adic valuation of $\mathrm{Tr}(\gamma_0)$ is non-negative for any $\ell \neq p$.

Combining the two paragraphs above, we deduce that $\mathrm{Tr}(\gamma_0) \in \mathbb{Z}$. So π is an algebraic integer.

Step II: We need to employ the existence part of Honda–Tate theory as follows. (The uniqueness part was stated in Theorem 3.4)

Theorem 4.2 (Honda–Tate)**.** *If π is a Weil q-number, then π comes from some simple abelian variety A over $\mathbb{F}_q$ in the following sense: $\mathbb{Q}(\pi)$ can be embedded into $\mathrm{End}_{\mathbb{F}_q}(A) \otimes \mathbb{Q}$ in such a way that π corresponds to the q-Frobenius endomorphism. (Moreover, A is uniquely determined by π up to isogeny over $\mathbb{F}_q$.)*

Now our π comes from some simple abelian variety A over $\mathbb{F}_q$, and we need to show that $\dim A = 1$. (Then it will follow that γ_0 comes from the elliptic curve A in the way we described.) In general, given a Weil q-number π, the dimension of the corresponding simple abelian variety A can be computed from the properties of π. More precisely, if $\mathbb{Q}(\pi) = \mathbb{Q}$, then $\dim A = 1$ and A is a supersingular elliptic curve. If $\mathbb{Q}(\pi) = \mathbb{Q}(\sqrt{p})$, then $\dim A = 2$. (For us this case does not appear since γ_0 is $\mathbb{R}$-elliptic.) In all the other cases, $\dim A$ is determined by the valuation of π at places of $\mathbb{Q}(\pi)$ above p together with the residue degrees of these places.

For us, $\mathbb{Q}(\pi)$ is either $\mathbb{Q}$ (when γ_0 is central), or an imaginary quadratic field (when γ_0 is non-central). Thus we have three possibilities:

(1) $\mathbb{Q}(\pi) = \mathbb{Q}$. We win, as A is a supersingular elliptic curve.
(2) $\mathbb{Q}(\pi)$ is imaginary quadratic, and p is non-split in $\mathbb{Q}(\pi)$. For such π it is known that A is a supersingular elliptic curve, so we also win.
(3) $\mathbb{Q}(\pi)$ is imaginary quadratic, and p splits in $\mathbb{Q}(\pi)$. In this case, let v_1, v_2 be the two distinct places of $\mathbb{Q}(\pi)$ above p. Then $v_1(\pi)$ and $v_2(\pi)$ are two non-negative integers whose sum is n (since $N_{\mathbb{Q}(\pi)/\mathbb{Q}}(\pi) = p^n$). The dimension of A is the order of $\frac{v_1(\pi)}{n}$ in $\mathbb{Q}/\mathbb{Z}$. In particular, $\dim A = 1$ if and only if $\{v_1(\pi), v_2(\pi)\} = \{n, 0\}$. (When this is the case, A is an ordinary elliptic curve.)

Thus we only need to do some work in case (3). In the following, we assume that $\mathbb{Q}(\pi)$ is imaginary quadratic (i.e., γ_0 is non-central), and assume that p splits in $\mathbb{Q}(\pi)$. Our goal is to show that the two places v_1, v_2 of $\mathbb{Q}(\pi)$ above p satisfy $\{v_1(\pi), v_2(\pi)\} = \{n, 0\}$. We write $\bar{\pi}$ for the complex conjugate of π, viewed as in $\mathbb{Q}(\pi)$. Thus π and $\bar{\pi}$ are the two distinct eigenvalues of γ_0.

Definition 4.3. Let F be a complete discretely valued field with a fixed uniformizer π, and let γ be a semisimple element in $\mathrm{GL}_N(F)$. We say that γ has a *polar decomposition* if $\gamma = \nu(\pi) \cdot k$, where

- ν is a cocharacter of GL_N over F commuting with γ,
- $k \in \mathrm{GL}_N(F)$ such that all eigenvalues of k in $\overline{F}$ have valuation 0.

Exercise 4.4. Every semisimple $\gamma \in \mathrm{GL}_N(F)$ admits a unique polar decomposition $\gamma = \nu(\pi)k$. We call $\nu(\pi)$ the *radial part* of γ (with respect to the choice of uniformizer π). **Hint:** first loosen the definition of a polar decomposition by allowing ν and k to be defined over $\overline{F}$. Prove existence and uniqueness in that setting. Then deduce from uniqueness that ν and k must be defined over F. (Use that $\nu(\pi) \in \mathrm{GL}_N(F)$ if and only if ν is defined over F.)

Fact 4.5 (nontrivial)**.** Suppose $\gamma \in \mathrm{GL}_N(\mathbb{Q}_p)$ is semisimple and there exists some $\delta \in \mathrm{GL}_N(\mathbb{Q}_{p^n})$ such that γ is conjugate to $\delta \cdot \sigma(\delta) \cdots \sigma^{n-1}(\delta)$. Then there exists an integer $t \geq 1$ such that the radial part of γ^t with respect to the uniformizer p is $\mathrm{GL}_N(\widehat{\mathbb{Q}_p^{\mathrm{ur}}})$-conjugate to $\nu_\delta^{nt}(p)$, where ν_δ is the Newton cocharacter of δ.

In our case, $\mathrm{TO}_\delta(f_n) \neq 0$ implies that δ is σ-conjugated to element in $\mathrm{GL}_2(\mathbb{Z}_q)\begin{pmatrix} p & \\ & 1 \end{pmatrix}\mathrm{GL}_2(\mathbb{Z}_q)$. Fact: in this case, ν_δ has only two choices up to conjugacy:

$$\text{either } \nu_\delta : z \mapsto \begin{pmatrix} z & \\ & 1 \end{pmatrix} \quad \text{or } \nu_\delta : z \mapsto \begin{pmatrix} z^{1/2} & \\ & z^{1/2} \end{pmatrix}$$

(In the second case, ν_δ is a *fractional cocharacter*, i.e., a formal fractional power of an actual cocharacter.) This fact is a special case of Mazur's inequality. Concretely, the condition on δ implies that the isocrystal over $\widehat{\mathbb{Q}_p^{\mathrm{ur}}}$ corresponding to δ comes from a p-divisible group of height 2 and dimension 1. By the Dieudonné–Manin classification, the slopes must be either $(1, 0)$ or $(\frac{1}{2}, \frac{1}{2})$.

By Fact 4.5, there exists t such that the radial part of γ_0^t is conjugate to either $\begin{pmatrix} p^{nt} & \\ & 1 \end{pmatrix}$ or $\begin{pmatrix} p^{nt/2} & \\ & p^{nt/2} \end{pmatrix}$. The two places v_1, v_2 determine two embeddings $\iota_1, \iota_2 : \mathbb{Q}(\pi) \hookrightarrow \mathbb{Q}_p$. We set $\lambda_i = \iota_i(\pi)$ for $i = 1, 2$. Thus over $\mathbb{Q}_p$, the characteristic polynomial of γ_0 factorizes as $(X - \lambda_1)(X - \lambda_2)$. Since γ_0 is not central, $\lambda_1 \neq \lambda_2$.

Now γ_0 is $\mathbb{Q}_p$-conjugate to $\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$, and the latter has a polar decomposition

$$\begin{pmatrix} p^{\nu_p(\lambda_1)} & \\ & p^{\nu_p(\lambda_2)} \end{pmatrix}\begin{pmatrix} k_1 & \\ & k_2 \end{pmatrix}.$$

This implies that the radial part of γ_0 is conjugate to $\begin{pmatrix} p^{v_p(\lambda_1)t} & \\ & p^{v_p(\lambda_2)t} \end{pmatrix}$; but it is also conjugate to $\begin{pmatrix} p^{nt} & \\ & 1 \end{pmatrix}$ or $\begin{pmatrix} p^{nt/2} & \\ & p^{nt/2} \end{pmatrix}$. In the first case, $\{v_p(\lambda_1), v_p(\lambda_2)\} = \{n, 0\}$, and we win. In the second case, we deduce that $v_p(\lambda_1) = v_p(\lambda_2) = \frac{n}{2}$. We show a contradiction. First, we show that some power of γ_0 is central. We claim that $\pi/\bar{\pi} \in \mathbb{Q}(\pi)$ has all non-archimedean valuations zero: at a place v of $\mathbb{Q}(\pi)$ coprime to p, say $v|\ell$, γ_0 is conjugate to some element in $\mathrm{GL}_2(\mathbb{Z}_\ell)$, and so $v(\pi) = v(\bar{\pi}) = 0$; for the places v_1, v_2 above p, we have $v_1(\pi/\bar{\pi}) = v_p(\lambda_1) - v_p(\lambda_2) = 0$, and $v_2(\pi/\bar{\pi}) = v_p(\lambda_2) - v_p(\lambda_1) = 0$. Also, for all complex embeddings $\mathbb{Q}(\pi) \to \mathbb{C}$, the absolute value of $\pi/\bar{\pi}$ is 1 (since $\det\gamma_0 = \pi\bar{\pi} = q$, as always). This implies that $\pi/\bar{\pi}$ is a root of unity. Since π and $\bar{\pi}$ are the two distinct eigenvalues of γ_0, we conclude that some power of γ_0 is central. Recall that γ_0 itself is non-central. This will contradict with the following lemma, and our proof of the point counting formula is complete once the lemma is proved.

Lemma 4.6. *Suppose a pair (γ_0, δ) with γ_0 $\mathbb{R}$-elliptic and $\gamma_0 \sim \delta\sigma(\delta)\cdots\sigma^{n-1}(\delta)$ is such that $\mathrm{O}_{\gamma_0}(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n) \neq 0$, and suppose some power of γ_0 is central. Then γ_0 is central.*

Proof. Suppose for the sake of contradiction that γ_0^k is central, and γ_0 is non-central. Since $\det(\gamma_0) = q$, we have $\gamma_0^k = \begin{pmatrix} q^{k/2} & \\ & q^{k/2} \end{pmatrix}$. The two eigenvalues of γ_0 are each a k-th root of $q^{k/2}$

and multiply to be q, so they must be of the form $\zeta\sqrt{q}$ and $\zeta^{-1}\sqrt{q}$, for some k-th root of unity ζ. The two eigenvalues must be distinct, so $\zeta^2 \neq 1$.

Yet γ_0 is conjugate to some element in K^p. Recall that $K^p \subset \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \in \mathrm{GL}_2(\widehat{\mathbb{Z}}^{(p)}) \;\middle|\; \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \equiv \mathbf{1} \pmod{N} \right\}$ for some fixed $N \geq 3$ and $p \nmid N$. So we can find some prime power $\ell^i | N$ with $\ell \neq p$. This implies that $\gamma_0 \in G(\mathbb{Q}_\ell)$ is conjugate to $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \in \mathrm{GL}_2(\mathbb{Z}_\ell)$ which is congruent to $\mathbf{1}$ modulo ℓ^i. Hence $\zeta\sqrt{q} \equiv \zeta^{-1}\sqrt{q} \equiv 1 \mod \ell^i$ inside $\overline{\mathbb{Z}}_\ell$. Since both ζ and $\sqrt{q}$ are units in $\overline{\mathbb{Z}}_\ell$, we have $\zeta^2 \equiv 1 \mod \ell^i$, i.e. $v_\ell(\zeta^2 - 1) \geq i$. This contradicts with the following exercise applied to $\mu = \zeta^2$. □

Exercise 4.7. Let ℓ^i be a prime power which is ≥ 3. Suppose μ is a root of unity in $\overline{\mathbb{Q}}$ and $v_\ell(\mu - 1) \geq i$. Then $\mu = 1$.

Remark 4.8. Abstractly, in the proof of Lemma 4.6 we used the following property called "neat": K^p being "neat" implies that for any $\gamma_0 \in G(\mathbb{Q}) \cap K^p$, the eigenvalues of γ_0 generate a torsion-free subgroup of $\overline{\mathbb{Q}}^\times$. In particular, the eigenvalues cannot differ from each other by a non-trivial root of unity.

Remark 4.9. Suppose an elliptic curve E over $\mathbb{F}_q$ gives rise to some $\gamma_0 \in \mathrm{GL}_2(\mathbb{Q})$. Then E is supersingular if and only if some power of γ_0 is central. More precisely, γ_0^k is central if and only if $\mathrm{End}_{\mathbb{F}_{q^k}}(E_{\mathbb{F}_{q^k}}) \otimes \mathbb{Q}$ is a quaternion algebra over $\mathbb{Q}$. Thus Lemma 4.6 implies that case (2) beneath Theorem 4.2, i.e., the case where E is supersingular and γ_0 is non-central, never shows up in our point counting! In other words, the classification of γ_0 appearing in the point counting formula into the non-central case and the central case is the same as the classification into the ordinary case (i.e., γ_0 comes from some ordinary elliptic curve) and the supersingular case (i.e., γ_0 comes from some supersingular elliptic curve).

Note that abstractly case (2) does exist. For example, take $q = p = 3$, and $\pi = \sqrt{-3}$, $\gamma_0 = \left(\begin{smallmatrix} \sqrt{-3} & \\ & -\sqrt{-3} \end{smallmatrix} \right)$. Then π corresponds to a supersingular elliptic curve E over $\mathbb{F}_3$ with $\mathrm{End}_{\mathbb{F}_3}(E) \otimes \mathbb{Q} = \mathbb{Q}(\sqrt{-3})$ and $\mathrm{End}_{\mathbb{F}_9}(E_{\mathbb{F}_9}) \otimes \mathbb{Q} = D_{3,\infty}$. Our proof shows that such an elliptic curve over $\mathbb{F}_3$ does not show up in $\mathscr{S}_K(\mathbb{F}_3)$, which means it does not admit a K^p-level structure over $\mathbb{F}_3$ for K^p neat. (This particular elliptic curve will contribute to $\mathscr{S}_K(\mathbb{F}_9)$.)

5. Lecture 5

5.1. Back to the general formula.

Let (G, X) be a Shimura datum. Assume

- G_{der} is simply connected, and
- Z_G is cuspidal, i.e, a maximal $\mathbb{R}$-split subtorus is $\mathbb{Q}$-split.

Take the level structure to be $K = K^p K_p$, with K^p small enough and K_p hyperspecial. Thus there exists a connected reductive group scheme $\mathcal{G}$ over $\mathbb{Z}_p$ whose generic fiber is $G_{\mathbb{Q}_p}$ such that $K_p = \mathcal{G}(\mathbb{Z}_p)$. (In fact K_p and $\mathcal{G}$ determine each other.) Fix a prime v of the reflex field E over p.

We assume that the conjectural canonical integral model $\mathscr{S}_K$ over $\mathcal{O}_{E,(v)}$ exists. Take some p^n such that the residue field of v is contained in $\mathbb{F}_{p^n}$. Then the conjectural counting point formula (see [Kot90]) is

$$\#\mathscr{S}_K(\mathbb{F}_{p^n}) = \sum_{(\gamma_0, \gamma, \delta)} c_1(\gamma_0, \gamma, \delta) c_2(\gamma_0) \cdot \mathrm{O}_\gamma(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n).$$

Here

- $(\gamma_0, \gamma, \delta)$ runs through $G(\mathbb{Q}) \times G(\mathbb{A}_f^p) \times G(\mathbb{Q}_{p^n})$, satisfying
  - γ_0 is $\mathbb{R}$-elliptic (which in particular implies that γ_0 is semisimple), i.e., there exists a maximal torus $T \subset G_{\mathbb{R}}$ over $\mathbb{R}$ such that $\gamma_0 \in T(\mathbb{R})$ and $(T/Z_G)(\mathbb{R})$ is compact,
  - γ is stably conjugate to γ_0, i.e, γ is conjugate to γ_0 over $G(\overline{\mathbb{A}}_f^p)$, where $\overline{\mathbb{A}}_f^p := \mathbb{A}_f^p \otimes_{\mathbb{Q}} \overline{\mathbb{Q}}$. (In particular, if we write $\gamma = (\gamma_\ell)_{\ell \neq p}$ with $\gamma_\ell \in G(\mathbb{Q}_\ell)$, then γ_ℓ is conjugate to γ_0 over $G(\overline{\mathbb{Q}}_\ell)$).[5]
  - $\delta \in G(\mathbb{Q}_{p^n})$ and $\delta\sigma(\delta)\cdots\sigma^{n-1}(\delta)$ is $G(\overline{\mathbb{Q}}_p)$-conjugate to γ_0, i.e., the stable conjugacy class of γ_0 is the degree n norm of δ.
- Given $(\gamma_0, \gamma, \delta)$ as above (plus some later hypothesis), one defines a certain Galois cohomological invariant $\alpha(\gamma_0, \gamma, \delta)$ (Kottwitz invariant) lying in some finite abelian group, where the abelian group depends only on γ_0. In the formula, the summation is over those $(\gamma_0, \gamma, \delta)$ with $\alpha(\gamma_0, \gamma, \delta) = 0$ up to an equivalence relation: $(\gamma_0, \gamma, \delta) \sim (\gamma_0', \gamma', \delta')$ if
  - γ_0 is conjugate to γ_0' by $G(\overline{\mathbb{Q}})$,
  - γ is conjugate to γ by $G(\mathbb{A}_f^p)$, and
  - δ is σ-conjugate to δ' in $G(\mathbb{Q}_{p^n})$.

  Here one key point is that if $(\gamma_0, \gamma, \delta) \sim (\gamma_0', \gamma', \delta')$, then the abelian group containing $\alpha(\gamma_0, \gamma, \delta)$ and that containing $\alpha(\gamma_0', \gamma', \delta')$ are canonically identified, and we have $\alpha(\gamma_0, \gamma, \delta) = \alpha(\gamma_0', \gamma', \delta')$.

Now, we continue to discuss the summand $c_1(\gamma_0, \gamma, \delta)c_2(\gamma_0) \cdot \mathrm{O}_\gamma(\mathbf{1}_{K^p}) \cdot \mathrm{TO}_\delta(f_n)$.

- $c_1(\gamma_0, \gamma, \delta)$: given $(\gamma_0, \gamma, \delta)$, write $\gamma = (\gamma_\ell)_{\ell \neq p}$ with $\gamma_\ell \in G(\mathbb{Q}_\ell)$. Then G_{γ_ℓ} is an inner form of $G_{\gamma_0, \mathbb{Q}_\ell}$, and similarly $J_{n,\delta}$ is an inner form of $G_{\gamma_0, \mathbb{Q}_p}$. We want a global inner form I of G_{γ_0} over $\mathbb{Q}$ such that $I_{\mathbb{R}}/Z_{G,\mathbb{R}}$ is compact as a real group, $I_{\mathbb{Q}_\ell} \cong G_{\gamma_\ell}$ as inner forms of $G_{\gamma_0, \mathbb{Q}_\ell}$, and $I_{\mathbb{Q}_p} \cong J_{n,\delta}$ as inner forms of $G_{\gamma_0, \mathbb{Q}_p}$. Then we define
$$c_1(\gamma_0, \gamma, \delta) = \mathrm{vol}\Big(I(\mathbb{Q}) \backslash I(\mathbb{A}_f)\Big).$$

  Note: for a general $(\gamma_0, \gamma, \delta)$, there is no reason why such a global inner form I should exist, but if $\alpha(\gamma_0, \gamma, \delta) = 0$, then I exists! Actually, one can think of the condition that $\alpha(\gamma_0, \gamma, \delta) = 0$ as some subtle strengthening of the existence of I.
- $c_2(\gamma_0) := \#\,\mathrm{Ker}\,\Big(\text{Ш}(G_{\gamma_0}) \to \mathrm{H}^1(\mathbb{Q}, G)\Big)$, where
$$\text{Ш}(G_{\gamma_0}) := \bigcap_v \mathrm{Ker}(\mathrm{H}^1(\mathbb{Q}, G_{\gamma_0}) \to \mathrm{H}^1(\mathbb{Q}_v, G_{\gamma_0})).$$
- $\mathrm{O}_\gamma(\mathbf{1}_{K^p})$ is the orbital integral
$$\int_{G(\mathbb{A}_f^p)_\gamma \backslash G(\mathbb{A}_f^p)} \mathbf{1}_{K^p}(x^{-1}\gamma x)dx.$$
- $\mathrm{TO}_\delta(f_n)$ is the twisted orbital integral
$$\int_{J_{n,\delta}(\mathbb{Q}_p) \backslash G(\mathbb{Q}_{p^n})} f_n(x^{-1}\delta\sigma(x))dx.$$

[5] In the GL_2-case, we did not have γ because γ_0 determines the conjugacy class of γ and we may just take $\gamma = \gamma_0$.

Here $f_n : G(\mathbb{Q}_{p^n}) \to \{0, 1\}$ is defined as follows: for any $h \in X$, h is a $G(\mathbb{R})$-conjugacy class of homomorphisms $\operatorname{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m \to G_{\mathbb{R}}$. It gives rise to

$$h_{\mathbb{C}} : (\operatorname{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m)_{\mathbb{C}} \cong \mathbb{G}_m \times \mathbb{G}_m \longrightarrow G_{\mathbb{C}},$$

where the first $\mathbb{G}_m$ corresponds to $\mathrm{id} : \mathbb{C} \to \mathbb{C}$ and the second $\mathbb{G}_m$ corresponds to the complex conjugation $\mathbb{C} \to \mathbb{C}$. Then we define

$$\begin{aligned} \mu_h : \mathbb{G}_{m,\mathbb{C}} &\longrightarrow G_{\mathbb{C}} \\ z &\longmapsto h_{\mathbb{C}}(z, 1). \end{aligned}$$

This is the *Hodge cocharacter associated to h*. The $G(\mathbb{C})$-conjugacy class of μ_h is defined over E (by the definition of E). In particular, if F is a field extension of E, we get a conjugacy class of cocharacters of G such that the whole conjugacy class is defined over F.

If G is quasi-split over F, this determines a $G(F)$-conjugacy class of F-rational cocharacters $\mathbb{G}_{m,F} \to G_F$. Now G is quasi-split over $\mathbb{Q}_{p^n}$ (since it is already quasi-split over $\mathbb{Q}_p$ by the existence of $\mathcal{G}$), and we have $\mathbb{Q}_{p^n} \supset E_v \supset E$. (In fact, one can show that E_v must be unramified over $\mathbb{Q}_p$ if $\mathcal{G}$ exists. Hence our assumption that $\mathbb{F}_{p^n}$ contains the residue field of E_v implies that $\mathbb{Q}_{p^n}$ contains E_v.) This then gives rise to a $G(\mathbb{Q}_{p^n})$-conjugacy class of cocharacters of $G_{\mathbb{Q}_{p^n}}$. This conjugacy class contains a canonical $\mathcal{G}(\mathbb{Z}_{p^n})$-conjugacy class, whose members are characterized by the condition that they extend to cocharacters of $\mathcal{G}_{\mathbb{Z}_{p^n}}$ (i.e., $\mathbb{Z}_{p^n}$-group scheme homomorphisms $\mathbb{G}_m \to \mathcal{G}_{\mathbb{Z}_{p^n}}$).

Then f_n is the characteristic function of $\mathcal{G}(\mathbb{Z}_{p^n})\mu(p)\mathcal{G}(\mathbb{Z}_{p^n})$, where μ is a member of the above-mentioned $\mathcal{G}(\mathbb{Z}_{p^n})$-conjugacy class. It turns out that f_n is independent of the choice of μ.

For example, when $G = \mathrm{GL}_2$, we can take μ to be $z \mapsto \left(\begin{smallmatrix} 1 & \\ & 1 \end{smallmatrix}\right)$, and then f_n is the characteristic function of $\mathrm{GL}_2(\mathbb{Z}_q)\left(\begin{smallmatrix} p & \\ & 1 \end{smallmatrix}\right)\mathrm{GL}_2(\mathbb{Z}_q)$.

Remark 5.2. If $\mathrm{TO}_\delta(f_n) \neq 0$, then the image of δ under the Kottwitz homomorphism

$$\kappa : G(\widehat{\mathbb{Q}_p^{\mathrm{ur}}}) \longrightarrow \pi_1(G)_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)},$$

has to be equal to a fixed element in $\pi_1(G)_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$, namely the natural image of μ for any Hodge cocharacter μ for X. We need this condition on δ in order to define the Kottwitz invariant $\alpha(\gamma_0, \gamma, \delta)$.

Remark 5.3. Why should we expect that for $(\gamma_0, \gamma, \delta)$ contributing to the formula, the global I exists?

In the PEL case, we start with an abelian variety A over $\mathbb{F}_{p^n}$ equipped with a PEL structure, then we can construct the triple $(\gamma_0, \gamma, \delta)$ as follows:

- γ corresponds to the p^n-Frobenius endomorphism $\pi \in \mathrm{End}(A)$ acting on $T^{(p)}(A)$ (which is equipped with PE structure),
- similarly, δ corresponds to the Frobenius action on the Dieudonné module $M_0(A)$,
- The global I is isomorphic to the $\mathbb{Q}$-group given by sending any $\mathbb{Q}$-algebra R to

$$\Big(\Big(\mathrm{End}_{\mathbb{F}_{p^n}}(A, \text{PE str.})\Big) \otimes_{\mathbb{Z}} R\Big)^{\times}.$$

The point is that: if we start with some random $(\gamma_0, \gamma, \delta)$, there is no reason for such I to exists. But if we start with some $(\gamma_0, \gamma, \delta)$ that comes from abelian varieties with PE structure (which is related to the condition that $\alpha(\gamma_0, \gamma, \delta) = 0$), then such I naturally exists.

Remark 5.4. Recall that $\#\mathscr{S}_K(\mathbb{F}_q) = \sum_i (-1)^i \mathrm{Tr}\Big(\mathrm{Frob}_q \Bigm| \mathrm{H}^i_{\text{ét}}\Big(\mathscr{S}_{K,\overline{\mathbb{F}}_q}, \mathbb{Q}_\ell\Big)\Big)$. More generally, we can also consider:

$$\sum_i (-1)^i \mathrm{Tr}\Big(\mathrm{Frob}_q \times f^p \Bigm| \mathrm{H}^i_{\text{ét}}\Big(\mathscr{S}_{K,\overline{\mathbb{F}}_q}, \mathbb{Q}_\ell\Big)\Big)$$

for some $f^p \in \mathcal{H}(G(\mathbb{A}_f^p)//K^p)$. For this, there is a similar conjectural formula, with $\mathrm{O}_\gamma(\mathbf{1}_{K^p})$ replaced by $\mathrm{O}_\gamma(f^p)$.

5.5. **Known cases of the counting point conjecture.**

- Kottwitz (1992, [Kot92]) proved the conjecture for PEL type Shimura varieties of type A and C.
- Between 1990's and now, there are some sporadic cases beyond PEL type, but closely related to PEL type.
- (Kisin–Shin–Z [KSZ]) All abelian type cases, and removed the assumptions that G_{der} is simply connected and Z_G is cuspidal.

In the next lecture, we will discuss briefly the proof of this conjecture.

5.6. **Informal introduction to trace formulas.** Recall that the idea of Langlands is that one can "compare" the formula for $\#\mathscr{S}_K(\mathbb{F}_{p^n})$ with trace formulas from representation theory. We give a brief introduction to the latter, following [Art05, §1].

Basic idea of trace formulas. Assume that we have some "nice" topological group H (Hausdorff, locally compact, unimodular), and we have a discrete subgroup $\Gamma \subset H$. A basic mathematical object to study is $L^2(\Gamma\backslash H)$ as an H-representation. Here H acts by right translation R, i.e., for $h \in H$, we define

$$\begin{aligned} R(h) : L^2(\Gamma\backslash H) &\longrightarrow L^2(\Gamma\backslash H) \\ \varphi &\longmapsto \big(x \mapsto \varphi(xh)\big). \end{aligned}$$

Fundamental question: How does R "decompose" into irreducible unitary representations of H?

For example, for $H = \mathbb{R}$ and $\Gamma = \mathbb{Z}$, unitary irreducible representations of H are parametrized by $y \in i\mathbb{R}$:

$$\begin{aligned} \pi_y : H &\longrightarrow \mathrm{GL}_1(\mathbb{C}) \\ x &\longmapsto e^{yx}. \end{aligned}$$

Then we have an isometry:

$$\begin{aligned} L^2(\mathbb{Z}\backslash\mathbb{R}) &\xrightarrow{\sim} L^2(\mathbb{Z}) \\ \varphi &\longmapsto \hat{\varphi}, \end{aligned}$$

where $\hat{\varphi} : \mathbb{Z} \to \mathbb{C}$ records the Fourier coefficients of φ:

$$\hat{\varphi}(n) = \int_{\mathbb{Z}\backslash\mathbb{R}} \varphi(x) e^{-2\pi i n x} dx.$$

The right-translation representation R of H on $L^2(\mathbb{Z}\backslash\mathbb{R})$ corresponds to the following representation of H on $L^2(\mathbb{Z})$. For $x \in H = \mathbb{R}$ and $\psi \in L^2(\mathbb{Z})$, we have

$$(x \cdot \psi)(n) = e^{2\pi i n x}\psi(x)$$

This means that we have

$$\begin{aligned} L^2(\mathbb{Z}) &\xrightarrow{\sim} \widehat{\bigoplus}_{n\in\mathbb{Z}}\mathbb{C} \\ \psi &\longmapsto (\psi(n))_n. \end{aligned}$$

Here the n-th copy of $\mathbb{C}$ is $\pi_{2\pi i n}$ as an H-representation. From this, we deduce

$$L^2(\mathbb{Z}\backslash\mathbb{R}) \cong \widehat{\bigoplus}_{y\in 2\pi i\mathbb{Z}}\pi_y$$

as H-representations. Thus we say $L^2(\mathbb{Z}\backslash\mathbb{R})$ *decomposes discretely.*

In contrast, for $H = \mathbb{R}$ and $\Gamma = \{0\}$, the theory of Fourier transform gives a "direct integral decomposition":

$$L^2(\Gamma\backslash H) \cong \int_{y\in i\mathbb{R}} \pi_y dy$$

In this case, we say that $L^2(\Gamma\backslash H)$ is a *continuous spectrum.*

The key distinction between the two cases is whether $\Gamma\backslash H$ is compact!

From now on, assume that $\Gamma\backslash H$ is compact. In this case, just like for $\mathbb{Z}\backslash\mathbb{R}$, we have a discrete decomposition

$$L^2(\Gamma\backslash H) \cong \widehat{\bigoplus}_{\pi} m_\pi \cdot \pi,$$

where the direct sum runs over all irreducible unitary representations π of H, and m_π is a finite multiplicity.

Fix a Haar measure dh on H (which is bi-invariant since H is unimodular). The trace formula studies the associated action of $\mathcal{C}_c(H)$ (a ring under convolution) on $L^2(\Gamma\backslash H)$. For $f \in \mathcal{C}_c(H)$, we define

$$\begin{aligned} R(f) : L^2(\Gamma\backslash H) &\longrightarrow L^2(\Gamma\backslash H) \\ \varphi &\longmapsto \Big[R(f)(\varphi)\Big](x) := \int_H \varphi(xh)f(h)dh. \end{aligned}$$

In fancy language,

$$R(f) := \int_H R(h)f(h)dh.$$

The trace formula computes the trace of a function in $\mathcal{C}_c(H)$ acting on $L^2(\Gamma\backslash H)$. More precisely, for $f \in \mathcal{C}_c(H)$, the operator $R(f)$ on $L^2(\Gamma\backslash H)$ is of trace class, and we have

$$\mathrm{Tr}\Big(R(f) \mid L^2(\Gamma\backslash H)\Big) = \sum_\pi m_\pi \mathrm{Tr}(f \mid \pi).$$

This is called the *spectral expansion* of the trace formula.

There is also the *geometric expansion*:

$$\mathrm{Tr}\Big(R(f) \mid L^2(\Gamma\backslash H)\Big) = \sum_{\gamma\in\Gamma} \mathrm{vol}(\Gamma_\gamma\backslash H_\gamma) \cdot \mathrm{O}_\gamma(f)$$

where $\mathrm{O}_\gamma(f)$ is the orbital integral

$$\int_{H_\gamma\backslash H} f(x^{-1}\gamma x)dx$$

as usual.

Of course, the geometric expansion is equal to the spectral expansion. This equality is called *Selberg's trace formula for a compact quotient.*

Exercise 5.7. In the case $\mathbb{Z}\backslash\mathbb{R}$, the equality between the geometric expansion and the spectral expansion amounts to the Poisson summation formula.

Remark 5.8. In the case when $\Gamma\backslash H$ is non-compact, we have the following problems.

- $\mathrm{Tr}(R(f) \mid L^2(\Gamma\backslash H))$ does not make sense.
- The geometric expansion and the spectral expansion as above do not make sense.

Now we want to apply this idea to $\Gamma\backslash H = G(\mathbb{Q})\backslash G(\mathbb{A})$ for some reductive group G over $\mathbb{Q}$. The bad news is that $G(\mathbb{Q})\backslash G(\mathbb{A})$ is often non-compact, even if we replace G by G^{ad}. For example, $G = \mathrm{GL}_2$ or SL_2 or PGL_2, the quotient $G(\mathbb{Q})\backslash G(\mathbb{A})$ is not compact.

5.9. **Arthur's invariant trace formula.** Arthur, in a long series of papers, developed the *invariant trace formula.* For details we refer the reader to [Art05]. As a result of this theory we have a conjugation-invariant distribution

$$\begin{aligned} \mathrm{I} : \mathcal{C}_c^\infty(G(\mathbb{A})) &\longrightarrow \mathbb{C} \\ f &\longmapsto \mathrm{I}(f) \end{aligned}$$

which is equal to $f \mapsto \mathrm{Tr}\Big(R(f) \mid L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))\Big)$ when $\Gamma\backslash H$ is compact but more complicated in general. The point is that even when $\Gamma\backslash H$ is non-compact, I is well defined, but of course it cannot be $f \mapsto \mathrm{Tr}\Big(R(f) \mid L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))\Big)$ since the latter does not make sense.

For simplicity, assume that G_{der} is simply connected. Then I has a geometric expansion:

$$\mathrm{I}(\cdot) = \sum_{\substack{\gamma\in G(\mathbb{Q})/\mathrm{conj} \\ \text{elliptic}}} \mathrm{vol}(G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A})) \cdot \mathrm{O}_\gamma(\cdot) + \text{some much more complicated terms}$$

The volume $\mathrm{vol}(G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A}))$ is often called the Tamagawa number $\tau(G_\gamma)$ of G_γ, when using the Tamagawa measure on $G_\gamma(\mathbb{A})$. We denote the sum

$$\sum_{\substack{\gamma\in G(\mathbb{Q})/\mathrm{conj} \\ \text{elliptic}}} \mathrm{vol}(G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A})) \cdot \mathrm{O}_\gamma(\cdot)$$

by $\mathrm{I}_{\mathrm{ell}}(\cdot)$, called the *elliptic part of the invariant trace formula.*

There is also a spectral expansion of I. First we have a direct sum decomposition

$$L^2(G(\mathbb{Q})\backslash G(\mathbb{A})) = L^2_{\mathrm{disc}} \oplus L^2_{\mathrm{cont}},$$

where $L^2_{\mathrm{disc}} = \widehat{\bigoplus}_\pi m_\pi^{\mathrm{disc}} \cdot \pi$, summing over unitary irreducible representations of $G(\mathbb{A})$, and L^2_{cont} is a continuous spectrum. Then

$$\mathrm{I}(\cdot) = \sum_\pi m_\pi^{\mathrm{disc}} \cdot \mathrm{Tr}(\cdot \mid \pi) + \text{some much more complicated terms.}$$

6. Lecture 6

6.1. **Stabilization.** The invariant trace formula has a problem: The distribution $\mathrm{I}(\cdot)$, or the simpler contributions $\mathrm{O}_\gamma(\cdot)$ and $\mathrm{Tr}(\cdot \mid \pi)$ in the geometric and spectral expansions, are not *stable*. There is a precise definition of a *stable distribution* on $G(\mathbb{A})$ which reflects the rough idea of invariance under conjugation by $G(\mathbb{A} \otimes_{\mathbb{Q}} \overline{\mathbb{Q}})$ (however note that we do not actually have a well-defined conjugation action of $G(\mathbb{A} \otimes_{\mathbb{Q}} \overline{\mathbb{Q}})$ on $G(\mathbb{A})$). The so-called *theory of endoscopy* was envisioned by Langlands in order to resolve this instability.

What we want: a stable distribution $\mathrm{S} : \mathcal{C}_c^\infty(G(\mathbb{A})) \to \mathbb{C}$ together with a geometric expansion and a spectral expansion into simpler stable distributions, i.e.,

$$\begin{aligned} \mathrm{S} &= \sum \text{stable distributions of a geometric nature} \\ &= \sum \text{stable distributions of a spectral nature.} \end{aligned}$$

As an illustration of some of the basic ideas, we sketch how to "stabilize" the distribution

$$\mathrm{I}_{\mathrm{ell}}(\cdot) = \sum_{\substack{\gamma_0 \in G(\mathbb{Q})/\mathrm{conj} \\ \gamma_0 \text{ elliptic}}} \tau(G_{\gamma_0}) \cdot \mathrm{O}_{\gamma_0}(\cdot).$$

For a more detailed introduction see [Har11]. As before, G is a reductive group over $\mathbb{Q}$ with simply connected derived subgroup.

Step 1: rewrite the sum as

$$\mathrm{I}_{\mathrm{ell}}(\cdot) = \sum_{\substack{\gamma_0 \in G(\mathbb{Q})/\text{stable conj} \\ \gamma_0 \text{ elliptic}}} \tau(G_{\gamma_0}) \cdot \sum_{\substack{\gamma \in G(\mathbb{A})/\mathrm{conj} \\ \gamma \text{ stable conj to } \gamma_0 \\ \mathrm{inv}(\gamma_0,\gamma)=0}} \mathrm{O}_\gamma(\cdot).$$

Here $\mathrm{inv}(\gamma_0, \gamma) \in \mathfrak{K}(\gamma_0)$, where $\mathfrak{K}(\gamma_0)$ is a finite abelian group coming from Galois cohomology. The point is that $\mathrm{inv}(\gamma_0, \gamma) = 0$ if and only if γ is $G(\mathbb{A})$-conjugate to some $\gamma_0' \in G(\mathbb{Q})$ which is stably conjugate to γ_0. Also the Tamagawa number $\tau(G_{\gamma_0})$ depends only on the stable conjugacy class of γ_0, namely, $\tau(I) = \tau(I')$ if I is an inner form of I', which is proved by Kottwitz.

Remark: The above formula is very similar to the point counting formula recalled below:

$$\sum_{\substack{(\gamma_0,\gamma,\delta) \\ \gamma_0 \in G(\mathbb{Q})_{\mathbb{R}\text{-ell}}/\text{stable conj} \\ \delta,\gamma \text{ adelic, up to conj or } \sigma\text{-conj} \\ \alpha(\gamma_0,\gamma,\delta)=0}} c_1 c_2 \mathrm{O}_\gamma(\cdot) \mathrm{TO}_\delta(\cdot).$$

Note the similar roles played by $\mathrm{inv}(\gamma_0, \gamma)$ and $\alpha(\gamma_0, \gamma, \delta)$.

Step 2: Apply Fourier inversion to the finite abelian group $\mathfrak{K}(\gamma_0)$. Write $\mathfrak{K}(\gamma_0)^D$ for the Pontryagin dual of $\mathfrak{K}(\gamma_0)$. We have

$$\begin{aligned}
\mathrm{I}_{\mathrm{ell}}(\cdot) &= \sum_{\substack{\gamma_0 \in G(\mathbb{Q})/\text{stable conj} \\ \gamma \in G(\mathbb{A})/\text{conj} \\ \gamma_0 \overset{\text{st}}{\sim} \gamma \\ \mathrm{inv}(\gamma,\gamma_0)=0}} \tau(G_{\gamma_0}) \mathrm{O}_\gamma(\cdot) \\
&= \sum_{\substack{\gamma_0 \in G(\mathbb{Q})/\text{stable conj} \\ \gamma \in G(\mathbb{A})/\text{conj} \\ \gamma_0 \overset{\text{st}}{\sim} \gamma \\ \text{no condition on } \mathrm{inv}(\gamma,\gamma_0)}} \frac{\tau(G_{\gamma_0})}{|\mathfrak{K}(\gamma_0)|} \sum_{\kappa \in \mathfrak{K}(\gamma_0)^D} \langle \kappa, \mathrm{inv}(\gamma_0, \gamma) \rangle \mathrm{O}_\gamma(\cdot) \\
&= \sum_{\gamma_0 \in G(\mathbb{Q})/\text{stable conj}} \frac{\tau(G_{\gamma_0})}{|\mathfrak{K}(\gamma_0)|} \sum_{\kappa \in \mathfrak{K}(\gamma_0)^D} \mathrm{O}^\kappa_{\gamma_0}(\cdot).
\end{aligned}$$

where

$$\mathrm{O}^\kappa_{\gamma_0}(\cdot) := \sum_{\substack{\gamma \in G(\mathbb{A})/\text{conj} \\ \gamma \overset{\text{st}}{\sim} \gamma_0}} \langle \kappa, \mathrm{inv}(\gamma_0, \gamma) \rangle \mathrm{O}_\gamma(\cdot)$$

is called the *κ-orbital integral* along γ_0.

For example, if $\kappa = 1$, this is just

$$\mathrm{O}^{\kappa=1}_{\gamma_0} = \sum_{\substack{\gamma \in G(\mathbb{A})/\text{conj} \\ \gamma \overset{\text{st}}{\sim} \gamma_0}} \mathrm{O}_\gamma(\cdot) =: \mathrm{SO}_{\gamma_0}(\cdot),$$

called the *stable orbital integral*. This is a stable distribution on $G(\mathbb{A})$!

Idea: For a nontrivial κ, $\mathrm{O}^\kappa_{\gamma_0}(\cdot)$ should “come from” a stable distribution on a different group. More precisely, from (γ_0, κ), we can construct a new reductive group H_κ over $\mathbb{Q}$ called an *endoscopic group* (which is also equipped with additional data relating the Langlands dual groups of H_κ and G) and construct an element $\gamma_\kappa \in H_\kappa(\mathbb{Q})$ up to stable conjugacy. We then want to relate $\mathrm{O}^\kappa_{\gamma_0}(\cdot)$ with $\mathrm{SO}_{\gamma_\kappa}(\cdot)$, where the latter is a stable distribution on $H_\kappa(\mathbb{A})$. This is done in the next step.

Step 3: (Hard!) For any $f \in \mathcal{C}^\infty_c(G(\mathbb{A}))$, we want to find $f^{H_\kappa} \in \mathcal{C}^\infty_c(H_\kappa(\mathbb{A}))$ called the *Langlands–Shelstad transfer* of f such that

$$\mathrm{O}^\kappa_{\gamma_0}(f) = \mathrm{SO}_{\gamma_\kappa}(f^{H_\kappa})$$

up to a *transfer factor* whose definition is very delicate. Here f^{H_κ} should depend only on H_κ (as an endoscopic group) and f but NOT on γ_κ and γ_0.

This involves extremely hard work by Langlands–Shelstad, Waldspurger, Hales, Laumon, and Ngo among others (including the Fundamental Lemma proved by Ngo).

Step 4: Put everything together.

Theorem 6.2 (Kottwitz 1980s, assuming Step 3)**.** *We have*

$$\mathrm{I}_{\mathrm{ell}}(f) = \sum_{\substack{H \text{ endoscopy} \\ \text{groups of } G}} i(G,H) \mathrm{ST}^H_{\mathrm{ell},*}(f^H)$$

where

- f^H *is the Langlands–Shelstad transfer of* f*,*
- $\mathrm{ST}^H_{\mathrm{ell},*}(\cdot)$ *is the "elliptic and* (G,H)*-regular part of the stable trace formula for* H*", given by*

$$\sum_{\substack{\gamma_H \in H(\mathbb{Q})/\textit{stable conj} \\ \gamma_H \textit{ is ellptic and } (G,H)\textit{-regular}}} \tau(H)\mathrm{SO}_{\gamma_H}(\cdot).$$

This is a stable distribution on $H(\mathbb{A})$*.*

Remark 6.3. Arthur later stabilized all terms in the geometric and spectral expansions of $\mathrm{I}(\cdot)$. This implies that

$$\mathrm{I}(f) = \sum_H i(G,H) \cdot \mathrm{S}^H(f^H)$$

Here S^H is a stable distribution on $H(\mathbb{A})$, which has a geometric expansion and a spectral expansion into simpler stable distributions.

Let us come back to the point counting formula

$$\sum (-1)^i \mathrm{Tr}\Big(f^p \times \mathrm{Frob}_{p^n} \Big| \mathrm{H}^i_{\text{ét},c}(\mathrm{Sh}_K)\Big) = \sum_{\substack{(\gamma_0,\gamma,\delta) \\ \alpha(\gamma_0,\gamma,\delta)=0}} c_1 c_2 \mathrm{O}_\gamma(f^p)\mathrm{TO}_\delta(f_n).$$

Kottwitz stabilized this formula in a similar way as the stabilization of $\mathrm{I}_{\mathrm{ell}}$ discussed above.

Theorem 6.4 (Kottwitz)**.** *The right hand side of the point counting formula is equal to*

$$\sum_{\substack{H \textit{ endscopy} \\ \textit{groups of } G}} i(G,H)\mathrm{ST}^H_{\mathrm{ell},*}(f^H_{\mathrm{Sh}}).$$

Here $f^H_{\mathrm{Sh}} = f^H_\infty f^H_p f^{H,p,\infty}$*, where* $f^{H,p,\infty} \in \mathcal{C}^\infty_c(H(\mathbb{A}^p_f))$ *is the Langlands–Shelstad transfer of* $f^p \in \mathcal{C}^\infty_c(G(\mathbb{A}^p_f))$*, while* $f^H_\infty \in \mathcal{C}^\infty(H(\mathbb{R}))$ *(a linear combination of pseudo-coefficients of discrete series) and* $f^H_p \in \mathcal{C}^\infty_c(H(\mathbb{Q}_p))$ *(a twisted transfer of* f_n*) are explicitly constructed. (The functions* f^H_∞ *and* f^H_p *are* not *given by Langlands–Shelstad transfer; the dependence of the formula on* n *is in the function* f^H_p*.)*

Expectation 6.5. If $\mathrm{Sh}_K(G,X)$ is projective, then for any H we should have

$$\mathrm{ST}^H_{\mathrm{ell},*}(f^H_{\mathrm{Sh}})) = \mathrm{S}^H(f^H_{\mathrm{Sh}}).$$

In particular, in the projective case

$$\sum (-1)^i \mathrm{Tr}\Big(f^p \times \mathrm{Frob}_{p^n} \Big| \mathrm{H}^i_{\text{ét},c}(\mathrm{Sh}_K)\Big) = \sum_H i(G,H)\mathrm{S}^H(f^H_{\mathrm{Sh}}).$$

In general, it is expected that

$$\sum (-1)^i \mathrm{Tr}\Big(f^p \times \mathrm{Frob}_{p^n} \Big| \mathrm{IH}^i_{\text{ét}}(\overline{\mathrm{Sh}}_K)\Big) = \sum_H i(G,H)\mathrm{S}^H(f^H_{\mathrm{Sh}}), \tag{6.5.1}$$

where $\mathrm{IH}^i_{\text{ét}}(\overline{\mathrm{Sh}}_K)$ is the intersection cohomology of the Bailey–Borel compactification $\overline{\mathrm{Sh}}_K$. In other words, the "difference" between $\mathrm{IH}^i_{\text{ét}}(\overline{\mathrm{Sh}}_K)$ and $\mathrm{H}^i_{\text{ét},c}(\mathrm{Sh}_K)$ should be accounted for precisely by the difference between $\mathrm{S}^H(f^H_{\mathrm{Sh}})$ and $\mathrm{ST}^H_{\mathrm{ell},*}(f^H_{\mathrm{Sh}})$.

Remark 6.6. In general, from (6.5.1), we expect to be able to relate the LHS of (6.5.1) to automorphic L-functions. To achieve this we need more ingredients, including Arthur's multiplicity conjectures. The point is that we need to "destabilize" and relate S^H back to automorphic representations of G. In general these ingredients are still highly conjectural. For a detailed explanation of the whole process see [Kot90].

For some classical groups, everything can be made to work, see for instance [Mor10] or [Zhu18].

6.7. **Proof of point counting formula.** Let (G, X) be a Shimura datum of Hodge type, i.e., there exists an embedding of Shimura data $\iota : (G, X) \hookrightarrow (\mathrm{GSp}(V), \mathfrak{H}^{\pm})$. We only explain the proof of the point counting formula in this case, without touching upon the general abelian type case. The references (including the general abelian type case) are the series of the papers [Kis10, Kis17, KSZ].

Stage 1: Constructing the anonical integral models.

Fix $K = K^p K_p \subset G(\mathbb{A}_f)$ with K_p hyperspecial and K^p small. Up to shrinking K^p and replacing ι by a different choice, we may assume that we have a closed embedding

$$\mathrm{Sh}_K(G, X) \hookrightarrow \mathrm{Sh}_U(\mathrm{GSp}(V)) \times_{\mathbb{Q}} E$$

where $U = U^p U_p \subset \mathrm{GSp}(V)(\mathbb{A}_f)$ with U_p hyperspecial in $\mathrm{GSp}(V)(\mathbb{Q}_p)$ and $U_p = \mathrm{GSp}(V_{\mathbb{Z}_p})$ for a self-dual $\mathbb{Z}_p$-lattice $V_{\mathbb{Z}_p}$ in $V_{\mathbb{Q}_p}$. Here, $\mathrm{Sh}_U(\mathrm{GSp})$ admits an integral model $\mathscr{S}_U(\mathrm{GSp})$ over $\mathbb{Z}_{(p)}$, which is a moduli space of polarized abelian schemes. Thus we have

$$\mathrm{Sh}_K(G) \hookrightarrow \mathrm{Sh}_U(\mathrm{GSp}) \times E \hookrightarrow \mathscr{S}_U(\mathrm{GSp}) \times_{\mathbb{Z}_{(p)}} \mathcal{O}_{E,(v)}.$$

Definition 6.8. The canonical integral model $\mathscr{S}_K(G)$ is the normalization of the Zariski closure of $\mathrm{Sh}_K(G)$ in $\mathscr{S}_U(\mathrm{GSp}) \times_{\mathbb{Z}_{(p)}} \mathcal{O}_{E,(v)}$.

The hard part is to prove $\mathscr{S}_K = \mathscr{S}_K(G)$ is smooth over $\mathcal{O}_{E,(v)}$. There are several steps here:

Step 1: First realize the embedding $G \to \mathrm{GSp}(V)$ as the stabilizer of certain tensors $(s_\alpha)_\alpha$ on V. We can arrange that each s_α extends to a $\mathbb{Z}_p$-linear tensor on $V_{\mathbb{Z}_p}$. Moreover, letting $\mathcal{G}$ be the reductive model over $\mathbb{Z}_p$ of $G_{\mathbb{Q}_p}$ such that $K_p = \mathcal{G}(\mathbb{Z}_p)$, we may arrange that $\mathcal{G} \hookrightarrow \mathrm{GSp}(V_{\mathbb{Z}_p})$ is the scheme-theoretic stabilizer of $(s_\alpha)_\alpha$.

Step 2: Given a finite extension F of $\mathbb{Q}_p$ with residue field κ, if $x \in \mathscr{S}_K(\mathcal{O}_F)$, then via the map $\mathscr{S}_K \to \mathscr{S}_U \times \mathcal{O}_{E,(\mathfrak{p})}$, we obtain an abelian scheme $\mathcal{A}_x$ on $\mathcal{O}_F$, giving rise to a p-adic representation of $\mathrm{Gal}(\overline{F}/F)$ on $T_p(\mathcal{A}_{x,\overline{F}})$. The $\mathbb{Z}_p$-module $T_p(\mathcal{A}_{x,\overline{F}})$ can be identified with $V_{\mathbb{Z}_p}$, and the tensors s_α's give rise to tensors on $T_p(\mathcal{A}_{x,\overline{F}})$ which are $\mathrm{Gal}(\overline{F}/F)$-invariant. By p-adic comparison, s_α "transports" to a tensor $s_{\alpha,0}$ on the rationalized Dieudonné module $M_0(\mathcal{A}_{x,\kappa})[\frac{1}{p}]$ of the special fiber $\mathcal{A}_{x,\kappa}$.

By some integral p-adic Hodge theory (in particular Breuil–Kisin modules of crystalline lattices and their relationship with p-divisible groups), one can prove that $s_{\alpha,0}$ is in fact a tensor on $M_0(\mathcal{A}_{x,\kappa})$, i.e., without inverting p.

Step 3: Write down a deformation space of the p-divisible group $A_{x,\kappa}[p^\infty]$ equipped with the integral tensors $s_{\alpha,0}$. This space has been defined amd studied by Faltings and is formally smooth over $W(\kappa)$.

Step 4: Relate the space in Step 3 with the local structure of $\mathscr{S}_K$, proving that $\mathscr{S}_K$ is smooth.

Stage 2: Classify "isogeny classes". We work with

$$\mathscr{S}_{K_p}(\overline{\mathbb{F}}_p) := \varprojlim_{K^p} \mathscr{S}_{K^pK_p}(\overline{\mathbb{F}}_p).$$

If we understand this set together with the $G(\mathbb{A}_f^p)$-action and the Frobenius action, then we can infer the point counting formula.

Definition 6.9. Two points $x, x' \in \mathscr{S}_{K^p}(\overline{\mathbb{F}}_p)$ are called *isogenous* if there exists a quasi-isogeny $f : \mathcal{A}_x \to \mathcal{A}_{x'}$ such that

- f takes each $s_{\alpha,0}$ on $M_0(\mathcal{A}_x)[\frac{1}{p}]$ to $s_{\alpha,0}$ on $M_0(\mathcal{A}_{x'})[\frac{1}{p}]$, and
- f preserves similar tensors on ℓ-adic Tate modules ($\ell \neq p$).

Kisin [Kis17] classified isogeny classes in a group theoretic manner, which is similar to Honda–Tate theory. This uses the language of Galois gerbs, due to Langlands–Rapoport [LR87].

One key ingredient of the classification is a generalization of Tate's special lifting theorem that every abelian variety over $\overline{\mathbb{F}}_p$ is isogenous to the reduction of a CM abelian variety. The generalization is stated as follows.

Theorem 6.10 (Kisin, [Kis17]). *Every isogeny class in $\mathscr{S}_K(\overline{\mathbb{F}}_p)$ contains a point which is the reduction of a* special point*, i.e. a point on Sh_K coming from $\mathrm{Sh}(T,h)$, where $(T,h) \hookrightarrow (G,X)$ with T a torus.*

Stage 3: Parametrize points in a fixed isogeny class.

Fix $x_0 \in \mathscr{S}_{K_p}(\overline{\mathbb{F}}_p)$. We want to parametrize the isogeny class of x_0. This x_0 gives rise to $\mathcal{A}_{x_0}$ equipped with tensors on $T^{(p)}(\mathcal{A}_{x_0}) \otimes_{\mathbb{Z}} \mathbb{Q}$ and tensors on $M_0(\mathcal{A}_{x_0})[\frac{1}{p}]$ (with Frobenius action).

Consider

$$X^p = \left\{\text{isoms } V_{\mathbb{A}_f^p} \xrightarrow{\sim} T^p(\mathcal{A}_{x_0}) \otimes_{\mathbb{Z}} \mathbb{Q} \text{ preserving tensors}\right\}$$

and

$$X_p = \left\{ \text{lattices } \Lambda \subset M_0(\mathcal{A}_{x_0})[\tfrac{1}{p}] \;\middle|\; \begin{array}{l} (\Lambda, F) \text{ is a Dieudonné module of dimension } = \dim \mathcal{A}_{x_0} \\ +\text{compatibility with } s_{\alpha,0}\text{'s} \end{array} \right\}.$$

Remark: Here X_p is an affine Deligne–Lusztig set, and it has a purely group-theoretic description.

We have a bijection

$$\text{the isogeny class of } x_0 \longleftrightarrow I_x(\mathbb{Q})\backslash(X^p \times X_p).$$

where $I_x(\mathbb{Q})$ is the group of self-quasi-isogenies of $\mathcal{A}_x$ preserving ℓ-adic and crystalline tensors. This I_x is a reductive group over $\mathbb{Q}$, just like I_{E_0} in the GL_2-case.

Stage 4: After rewriting X^p and X_p in a more group theoretic way, we obtain a reductive group I over $\mathbb{Q}$, such that $I(\mathbb{A}_f)$ naturally acts on $X^p \times X_p$. (To make the analogy with the GL_2 case, I is like the $\mathbb{Q}$-group attached to (γ_0, δ), or more precisely the stable limit of the $\mathbb{Q}$-group attached to (γ_0^n, δ) as n becomes more and more divisible.)

Also, $I_x(\mathbb{Q})$ naturally acts on $X^p \times X_p$, and this action factors through an embedding

$$\iota : I_x(\mathbb{Q}) \hookrightarrow I(\mathbb{A}_f).$$

Problem: $\iota(I_x(\mathbb{Q})) \neq I(\mathbb{Q})$ in general. Rather, they are only conjugate by $I^{\mathrm{ad}}(\mathbb{A}_f)$. If $I^{\mathrm{ad}}(\mathbb{A}_f)$-conjugacy is the same as $I(\mathbb{A}_f)$-conjugacy (for example in the GL_2 case), we are happy.

Upshot: In reality, we encounter $I_x(\mathbb{Q})\backslash X^p \times X_p$. Ideally, we would like to work with $I(\mathbb{Q})\backslash X^p \times X_p$, because this is described purely group-theoretically. But these two sets are not known to be the same!

The discrepancy is measured by $\tau_x \in I^{\mathrm{ad}}(\mathbb{A}_f)$ such that

$$\iota(I_x(\mathbb{Q})) = \mathrm{Int}(\tau_x)(I(\mathbb{Q})).$$

We need extra new ideas to "control" the τ_x's for different isogeny classes, and to show that with the suitable control, they do not really affect the desired point counting formula. This is done in [KSZ].

6.11. **What about more general level structure?** We only give an incomplete list of mathematicians, without giving precise references. The reader is encouraged to explore their work.

Drinfeld level structure Harris–Taylor, Shin , Scholze.

Parahoric level structure Pappas–Zhu, Kisin–Pappas, Rong Zhou, van Hoften.

6.12. **Appendix. The idea of comparing trace formulas.** We briefly discuss how the point counting formula for the Shimura variety can be compared with Arthur–Selberg trace formulas in an ideal situation (a situation without any technical problems). [6]

We assume that $G(\mathbb{Q})\backslash G(\mathbb{A})$ is compact (or equivalently, that G is anisotropic mod center over $\mathbb{Q}$; in this case the Shimura variety is projective), that stable conjugacy is the same as conjugacy for the group G, and that the automorphic multiplicity of any automorphic representation of G is one. An example is when G is the multiplicative group of an indefinite quaternion algebra over $\mathbb{Q}$.

In a vague way, our ultimate goal is to describe the cohomology $\mathrm{H}^*_{\text{ét}}(\mathrm{Sh}_K(G)_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$ (viewed as a virtual representation) or the limit $\varprojlim_K \mathrm{H}^*_{\text{ét}}(\mathrm{Sh}(G)_{K,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$. We would like to have a decomposition

$$\mathrm{H}^i_{\text{ét}}(\mathrm{Sh}_K(G)_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) = \bigoplus_\pi \pi_f^K \otimes \rho_\pi$$

that is equivariant with respect to the Hecke action and the Galois action. Here the Hecke algebra $\mathcal{H}(G(\mathbb{A}_f)//K)$ only acts on the first factor π_f^K and the Galois group only acts on the second factor ρ_π, and conjecturally, ρ_π is the Galois representation associated to π (and "modified according to the representation of ${}^LG$ associated with X"). Our goal is to understand ρ_π in terms of Satake parameters of the automorphic representation π. Langlands' idea is explained in the following way: for $f^p \in \mathcal{H}(G(\mathbb{A}_f^p)//K^p)$ and some appropriate Hecke operator $T_p^{(n)} \in \mathcal{H}(G(\mathbb{Q}_p)//\mathcal{G}(\mathbb{Z}_p))$,

[6] We thank Liang Xiao for providing a draft for the following material. All errors or inaccuracies are the responsibility of Y.Z.

$$\begin{tikzcd}
\sum_{\pi} \mathrm{Tr}(f^p \mid \pi_f^K)\mathrm{Tr}(\mathrm{Frob}_{p^n} \mid \rho_\pi) \arrow[d, equal, "\text{spectral. exp.}"] & \sum_{\pi} \mathrm{Tr}(f^p \mid \pi_f^K)\mathrm{Tr}(T_p^{(n)} \mid \pi_p^{\mathcal{G}(\mathbb{Z}_p)}) \arrow[d, equal, "\text{spectral. exp.}"] \\
\mathrm{Tr}\Big(f^p \times \mathrm{Frob}_{p^n} \mid \mathrm{H}^*_{\text{ét}}(\mathrm{Sh}_K(G)_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)\Big) \arrow[d, equal, "\text{Point counting}"] & \mathrm{Tr}\Big(f^p \times T_p^{(n)} \mid L^2(G(\mathbb{Q})\backslash G(\mathbb{A})\Big) \arrow[d, equal, "\text{geom. exp.}"] \\
\sum_{(\gamma_0,\gamma,\delta)} c_1c_2\mathrm{O}_\gamma(f^p)\mathrm{TO}_\delta(f_n) \arrow[r, leftrightarrow, dashed] & \sum_{\gamma} \tau(G_\gamma)\mathrm{O}_\gamma(f^pT_p^{(n)})
\end{tikzcd}$$

We hope to relate the trace of Frob_{p^n}-action on ρ_π with the trace of the Hecke operator action, namely the two terms on the top row. The point counting formula translates the left column into some sum on the bottom row, and the Arthur–Selberg trace formula translates the right column into some similar sum on the bottom row. As both sides on the bottom row are indexed by similar objects, it makes sense to expect that one can compare the corresponding terms. We then would be able to deduce the equality on the top row.

In addition, we may choose the away-from-p Hecke operators f^p so that they kill all but only one π in the process (assuming there is no issues of L-packets). Then this process truly provides a description of ρ_π, in terms of Satake parameters.

(A general principle of trace formula is: one trace formula by itself relating the spectral side and geometric side does not seem to provide much insight. It becomes extremely powerful when we *compare two trace formulas*. Often times, there is a way to prove that their geometric expansions are equal by matching the orbits. This way, we prove an equality between the two spectral expansions, which often proves surprisingly strong statements.)

References

[Art05] James Arthur. An introduction to the trace formula. In *Harmonic analysis, the trace formula, and Shimura varieties*, volume 4 of *Clay Math. Proc.*, pages 1–263. Amer. Math. Soc., Providence, RI, 2005.

[Har11] Michael Harris. An introduction to the stable trace formula. In *On the stabilization of the trace formula*, volume 1 of *Stab. Trace Formula Shimura Var. Arith. Appl.*, pages 3–47. Int. Press, Somerville, MA, 2011.

[Kis10] Mark Kisin. Integral models for Shimura varieties of abelian type. *J. Amer. Math. Soc.*, 23(4):967–1012, 2010.

[Kis17] Mark Kisin. Mod p points on Shimura varieties of abelian type. *J. Amer. Math. Soc.*, 30(3):819–914, 2017.

[KMP16] Wansu Kim and Keerthi Madapusi Pera. 2-adic integral canonical models. *Forum Math. Sigma*, 4:e28, 34, 2016.

[Kot90] Robert E. Kottwitz. Shimura varieties and λ-adic representations. In *Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988)*, volume 10 of *Perspect. Math.*, pages 161–209. Academic Press, Boston, MA, 1990.

[Kot92] Robert E. Kottwitz. Points on some Shimura varieties over finite fields. *J. Amer. Math. Soc.*, 5(2):373–444, 1992.

[KSZ] Mark Kisin, Sug Woo Shin, and Yihang Zhu. The stable trace formula for Shimura varieties of abelian type. *preprint, available at* `http://math.umd.edu/~yhzhu/KSZ.pdf`.

[LR87] R. P. Langlands and M. Rapoport. Shimuravarietäten und Gerben. *J. Reine Angew. Math.*, 378:113–220, 1987.

[LS18] Kai-Wen Lan and Benoît Stroh. Nearby cycles of automorphic étale sheaves, II. In *Cohomology of arithmetic groups*, volume 245 of *Springer Proc. Math. Stat.*, pages 83–106. Springer, Cham, 2018.

[Mor10] Sophie Morel. *On the cohomology of certain noncompact Shimura varieties*, volume 173 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2010. With an appendix by Robert Kottwitz.

[MP19] Keerthi Madapusi Pera. Toroidal compactifications of integral models of Shimura varieties of Hodge type. *Ann. Sci. Éc. Norm. Supér. (4)*, 52(2):393–514, 2019.

[Zhu18] Yihang Zhu. The stabilization of the Frobenius–Hecke traces on the intersection cohomology of orthogonal Shimura varieties. *arXiv e-prints*, page arXiv:1801.09404, January 2018.

[Zhu20] Yihang Zhu. Introduction to the LanglandsKottwitz method. In Thomas Haines and Michael Harris, editors, *Shimura Varieties*, London Mathematical Society Lecture Note Series, pages 115–150. Cambridge University Press, 2020.