WEIL COHOMOLOGY IN PRACTICE

KIRAN S. KEDLAYA ET AL.

These are revised lecture notes from a course given by Kiran Kedlaya at UC San Diego in fall 2019 on the topic "Weil cohomology in practice". Thanks to the following students in the course for compiling the original draft of the notes: Samir Canning, Mingjie Chen, Patrick Girardet, Thomas Grubb, Jacob Keller, Bochao Kong, Woonam Lim, Zeyu Liu, Alex Mathers, Baiming Qiao, Nandagopal Ramachandran, Sankeerth Rao, Peter Wear, Wei Yin. (Special thanks to Peter Wear for coordinating this effort.) Thanks also to David Hansen for additional feedback.

The purpose of these notes is to present a somewhat idealized version of the course compared to what was actually delivered. To this end, I have filled in some details that were missing (or incorrect) in the original lectures; I have also shifted a couple of topics to improve the narrative flow.

I distributed five problem sets over the course of the term. I have included these as well as a few supplemental exercises that came up during the revision process.

There exist many excellent expositions about different parts of this material, some of which I surely do not yet know about. Suggestions for additional readings would be greatly appreciated.

Last revision: 22 Jun 2020.

Contents

1. Prehistory of the Weil conjectures (September 30)

In this lecture, we discuss the "prehistory" of the Weil conjectures from Gauss/Jacobi and Riemann/Dirichlet to Artin to Weil.

Readings 1.1. The primary source for this lecture is Weil's 1949 paper [115]. We will assume some familiarity with basic facts about algebraic number theory; there are many references for this, but we generally will follow Neukirch [92].

Since this topic is old and well-studied, many other expositions of it are available. A particularly detailed one has been given by Milne [87].

For context, let's start by formulating the Riemann hypothesis for Dedekind zeta functions.

Definition 1.2. Let K be a *number field*, i.e., a finite-degree field extension of the field of rational numbers $\mathbb{Q}$. Let $\mathcal{O}_K$ be the *ring of integers* of K, which is to say the integral closure of $\mathbb{Z}$ in K (more concretely, the elements of K which are roots of monic polynomials with integer coefficients). A basic fact about $\mathcal{O}_K$ is that it is a *Dedekind domain*, and so every nonzero ideal can be written uniquely as a product of powers of maximal ideals. (Note: we say "maximal ideals" rather than "prime ideals" only to exclude the zero ideal.)

The *Dedekind zeta function* of K is defined initially as the formal expression

$$\zeta_K(s) := \prod_{P \subset \mathcal{O}_K} \frac{1}{1 - \mathrm{Norm}_{K/\mathbb{Q}}(\mathfrak{p})^{-s}},$$

where the product is over all the maximal ideals of $\mathcal{O}_K$ and $\mathrm{Norm}_{K/\mathbb{Q}}(\mathfrak{p})$ is the cardinality of the quotient ring $\mathcal{O}_K/\mathfrak{p}$. For $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$, the product converges absolutely and so defines a holomorphic function without zeroes in that region. By unique factorization, we can rewrite the product as a sum

$$\zeta_K(s) = \sum_{I \subseteq \mathcal{O}_K} \mathrm{Norm}_{K/\mathbb{Q}}(I)^{-s}$$

where I now runs over all nonzero ideals of $\mathcal{O}_K$.

Theorem 1.3 (Hecke)**.** *The function $\zeta_K(s)$ extends meromorphically to $\mathbb{C}$, with a simple pole at $s = 1$ and no other poles.*

When $K = \mathbb{Q}$, $\zeta_K(s)$ is the usual Riemann zeta function. Like the latter, $\zeta_K(s)$ satisfies a functional equation relating its values at s and $1 - s$. The cleanest way to conceptualize this is to use the language of *places*, as follows.

Definition 1.4. Each maximal ideal $\mathfrak{p}$ of $\mathcal{O}_K$ corresponds to a dense embedding of K into a field complete with respect to a multiplicative absolute value, namely the fraction field of the $\mathfrak{p}$-adic completion of $\mathcal{O}_K$. These embeddings are called *finite places* of K. By Ostrowski's theorem, the only other dense embeddings of K into a field complete with respect to a (nontrivial) multiplicative absolute value are embeddings into $\mathbb{R}$ or $\mathbb{C}$, of which there are only finitely many; these are called *infinite places* of K. (For $K = \mathbb{Q}$, there is a unique infinite place, because $\mathbb{Q}$ maps in only one way into $\mathbb{R}$. Each prime number b corresponds to the embedding of $\mathbb{Q}$ into the p-adic numbers $\mathbb{Q}_p$.)

One may then define a *completed zeta function* $\Lambda_K(s)$ by adding to the product a suitable factor for each infinite place. This factor has the form

$$\pi^{-s/2}\Gamma(s/2) \qquad \text{or} \qquad 2(2\pi)^{-s}\Gamma(s)$$

(where Γ is Gauss's meromorphic interpolation of the factorial function) depending on whether the completion of $\mathbb{Q}$ is isomorphic to $\mathbb{R}$ (a *real place*) or $\mathbb{C}$ (a *complex place*). With these factors in place, the functional equation has the form

$$\Lambda_K(s) = \Lambda_K(1 - s).$$

Conjecture 1.5 (Riemann Hypothesis)**.** All nontrivial zeroes of $\zeta_K(s)$ (i.e., the ones not forced by the functional equation for $\Lambda_K(s)$) lie on the line $\mathrm{Re}(s) = 1/2$.

It was suggested by Artin that there should be a close analogy between number fields and *function fields*. This grows out of the observation that for any finite field $\mathbb{F}_q$, the ring of integers $\mathbb{Z}$ and the polynomial ring $\mathbb{F}_q[t]$ are both Euclidean domains and their maximal ideals have finite residue fields. To build out this perspective, let's make the following definition.

Definition 1.6. Fix a finite field $\mathbb{F}_q$. Let K be a *function field*, by which I mean a finite-degree extension of the field of rational functions $\mathbb{F}_q(t)$. We may then define the ring of integers $\mathcal{O}_K$ and the Dedekind zeta function $\zeta_K(s)$ using exactly the same formulas as in the number field case; the analogue of Dedekind's theorem also holds (with one minor quibble; see Remark 1.8). However, there is a key difference: in this case, the residue fields $\mathcal{O}_K/\mathfrak{p}$ all contain $\mathbb{F}_q$, so $\zeta_K(s)$ is a power series in q^{-s} rather than a more general Dirichlet series.

The discussion of places, and the definition of and functional equation for the completed zeta function $\Lambda_K(s)$, also extend to this setting, but again there is a key difference the "infinite places" in the function field setting look just like finite places after a change of coordinates, so there is no need to give a separate definition for the missing factors in the completed zeta function. We will come back to this point in Remark 1.8.

The analogue of the Riemann hypothesis for function fields was formulated by Artin. A proof was announced by Weil in 1940 [111], and a second proof in 1941 [112], but due to the precarious state of both Weil's life and world events in that period, the missing details from these announcements did not see print until 1948 [113, 114].

Theorem 1.7. *For K a function field, all nontrivial zeroes of $\zeta_K(s)$ (i.e., the ones not forced by the functional equation for $\Lambda_K(s)$) lie on the line* $\mathrm{Re}(s) = 1/2$.

Remark 1.8. The analogue of the Riemann zeta function here is the Dedekind zeta function for $K = \mathbb{F}_q(t)$, which one may easily calculate to be

$$\zeta_K(s) = \frac{1}{1 - q^{1-s}}$$

(see for example Definition 1.9 below). In this case, $\zeta_K(s)$ has no zeroes at all, so the Riemann hypothesis holds for particularly trivial reasons. Note however that $\zeta_K(s)$ has poles not only at $s = 1$, but also at $s = 1 + 2\pi i n/\log q$ for any $n \in \mathbb{Z}$. The completed zeta function is

$$\Lambda_K(s) = \frac{1}{(1 - q^{-s})(1 - q^{1-s})}.$$

For a general function field K, we will have

$$\zeta_K(s) = \frac{\text{polynomial in } q^{-s}}{1 - q^{1-s}}$$

and

$$\Lambda_K(s) = \frac{\text{polynomial in } q^{-s}}{(1 - q^{-s})(1 - q^{1-s})}.$$

From a modern point of view, we see that the properties of $\zeta_K(s)$ and $\lambda_K(s)$ amount to concrete statements about points on a certain algebraic curve over a finite field, whose proofs rely on now-standard techniques in algebraic geometry. For example, the proof of the functional equation for $\Lambda_K(s)$ uses the Riemann-Roch theorem for curves; Weil's first proof of the Riemann hypothesis uses the embedding of a curve in its Jacobian variety; and Weil's second proof uses the Hodge index theorem on the product of a curve with itself over the base field.

What one should keep in mind here is that none of this perspective was available to Weil. At the time he began his work, the subject of algebraic geometry only included varieties over the complex numbers; many of its best results did not comport with modern standards of rigor; commutative algebra had not yet developed to the point where it could be used to plug some of the gaps; and the key insights of Zariski, Serre, and Grothendieck needed to adapt sheaf theory into the modern foundations of algebraic geometry still lay years in the future (and would take Weil's work, and the Weil conjectures, as a primary impetus). As a result, the completion of Weil's announcements was delayed not just by geopolitical events, but also by Weil's need to build interim foundations on which to base his work. While these foundations are no longer in widespread use, and modern accounts of Weil's work typically reformulate his arguments using the theory of schemes, these reformulations are considered translations rather than completions.

In light of Remark 1.8, we now take the next step and reformulate the previous discussion in the language of algebraic geometry.

Definition 1.9. For K a function field, let X be the normalization of $\mathbb{A}^1_{\mathbb{F}_q}$ in K and let X° be the set of closed points in K. Then we have

$$\zeta_K(s) = \prod_{P \in X^\circ} \frac{1}{1 - \#\kappa(P)^{-s}} = \prod_{P \in X^\circ} \frac{1}{1 - q^{-d_P s}} \tag{1.10}$$

where $\kappa(P)$ denotes the residue field of P and $d_P = [\kappa(P) : \mathbb{F}_q]$. This can be rearranged to

$$\zeta_K(s) = \exp\Big(\sum_{n=1}^{\infty} \frac{q^{-ns}}{n} \#X(\mathbb{F}_{q^n}) \Big)$$

For example, for $K = \mathbb{F}_q(t)$, $X = \mathbb{A}^1_{\mathbb{F}_q}$ and so $X(\mathbb{F}_{q^n}) = q^n$ for all n; this recovers our earlier formula for $\zeta_K(s)$ in this case.

Remark 1.11. In the language of schemes, the previous discussion also applies in the case where K is a number field, taking X to be the normalization of $\mathrm{Spec}(\mathbb{Z})$ in $\mathrm{Spec}(\mathcal{O}_K)$. In particular, (1.10) carries over.

Definition 1.12. Following Weil, we now let X be an algebraic variety over $\mathbb{F}_q$ (or in modern language, a scheme of finite type over $\mathbb{F}_q$) and define $\zeta_X(s)$ as in (1.10):

$$\zeta_X(s) := \prod_{P \in X^\circ} \frac{1}{1 - \#K(P)^{-s}} = \prod_{P \in X^\circ} \frac{1}{1 - q^{-d_P s}} = \exp\Big(\sum_{n=1}^{\infty} \frac{q^{-ns}}{n} \#X(\mathbb{F}_{q^n}) \Big).$$

We then ask whether $\zeta_X(s)$ shares any of the previously observed properties when $\dim(X) > 1$. To get some clarity on this question, we consider some examples.

Example 1.13. Let $X = \mathbb{P}^n_{\mathbb{F}_q}$. Then

$$\zeta_X(s) = \frac{1}{(1-\tau)(1-q\tau)\cdots(1-q^n\tau)}, \quad \tau = q^{-s}.$$

We now consider a key example of Weil. At this point, our chain of inquiry, which so far has flowed naturally from the Riemann zeta function, links up with another thread from elementary number theory.

Example 1.14. Consider the *diagonal hypersurface* (or *Fermat hypersurface*)

$$a_0 x_0^{n_0} + a_1 x_1^{n_1} + \cdots + a_r x_r^{n_r} = b \qquad (n_i > 0, a_i, b \in \mathbb{F}_q).$$

Over $\mathbb{Q}$, rational points on varieties of this form were considered by Fermat, Euler, and others. The finite field case was considered first by Gauss in the setting where $q = p$ is prime, $r = 2$, and n_0, n_1, n_2 are small. For example, Gauss proved that for $p \neq 2$, the equation $x^2 - y^2 = 1$ has $p - 1$ solutions in $\mathbb{F}_p$.

Definition 1.15. Let p be the characteristic of the finite field $\mathbb{F}_q$. Let $\chi : \mathbb{F}_q^\times \to \mathbb{C}^\times$ be a multiplicative character and $\psi : \mathbb{F}_q \xrightarrow{\text{Trace}} \mathbb{F}_p \to \mathbb{C}^\times$ be an additive character (where $\mathbb{F}_p \to \mathbb{C}^\times$ is the map $x \mapsto e^{2\pi i x/p}$). The *Gauss sum* $g(\chi) = g(\chi, \psi)$ associated to χ is given by

$$g(\chi) := \sum_{x \in \mathbb{F}_q} \chi(x)\psi(x) \in \overline{\mathbb{Z}} \subset \mathbb{C}.$$

(We have $g(\chi) \in \overline{\mathbb{Z}}$ because $g(\chi)$ is a sum of roots of unity.)

Theorem 1.16. *The Gauss sum $g(\chi)$ has the following properties.*

1. *(Gauss) We have $|g(\chi)|^2 = g(\chi)g(\overline{\chi}) = q$.*
2. *(Davenport–Hasse) For an extension $\mathbb{F}_{q^v}$ of $\mathbb{F}_q$, put $\chi' := \chi \circ \mathrm{Norm}_{\mathbb{F}_{q^v}/\mathbb{F}_q}$ and $\psi' := \psi \circ \mathrm{Trace}_{\mathbb{F}_{q^v}/\mathbb{F}_q}$. Then*

$$-g(\chi') = \Big(-g(\chi) \Big)^v.$$

Proof. See Set 1 exercises. □

Note that the conjugates of $g(\chi)$ are all themselves Gauss sums for other characters for the same q; consequently, $g(\chi)$ is an algebraic integer all of whose conjugates in $\mathbb{C}$ have absolute value $\sqrt{q}$.

Theorem 1.17 (Weil)**.** *Consider the Fermat hypersurface*

$$a_0 x_0^{n_0} + a_1 x_1^{n_1} + \cdots + a_r x_r^{n_r} = 0 \qquad (n_i > 0, a_i \in \mathbb{F}_q).$$

Then the number of points over $\mathbb{F}_q$ is given by

$$q^r + \frac{q-1}{q} \sum_{(\chi_0, \cdots, \chi_r)} \chi_0(a_0)^{-1} \cdots \chi_r(a_r)^{-1} g(\chi_0) \cdots g(\chi_r)$$

where $(\chi_0, \ldots, \chi_r)$ runs over all tuples in which χ_i is a multiplicative character of $\mathbb{F}_q$ of order dividing $\gcd(n_i, q-1)$ and $\chi_0 \cdots \chi_r = 1$.

Proof. See Set 2 exercises. □

By combining this with the Davenport-Hasse relation, we see that if we fix the hypersurface and count points over $\mathbb{F}_q^v$ for varying v, the answer is of the form $\sum_i \pm \alpha_i^v$. This forms a prototype for the *Weil conjectures*, to be introduced in the next lecture.

2. The Weil conjectures and examples (October 2)

In this lecture, we give the full statement of Weil's conjecture together with some small examples.

Readings 2.1. We roughly follow [59, Appendix C].

Definition 2.2. Let $k = \mathbb{F}_q$ be a finite field of q elements, and X/k be a quasi-projective algebraic variety (or more generally, any k-scheme of finite type; we will add hypotheses later in the statement). For any integer $r \geq 1$, let $k_r = \mathbb{F}_{q^r}$ be one field extension of k with degree r (which is unique up to noncanonical isomorphism). Write the zeta function $\zeta_X(s)$ as $Z(X, q^{-s})$ for

$$Z(X, T) := \exp\left(\sum_{r=1}^{\infty} \frac{T^r}{r} \#X\left(k_r\right) \right) \in \mathbb{Z}[\![T]\!].$$

(The containment $\mathbb{Z}(X,T) \in \mathbb{Q}[\![T]\!]$ is more obvious here, but the prior description of $\zeta_X(s)$ as an infinite product shows that $Z(X,T) \in \mathbb{Z}[\![T]\!]$.)

Theorem 2.3 (Weil conjectures)**.** *The series $Z(X,T)$ has the following properties.*

(1) *(Rationality) The series $Z(X,T)$ represents a rational function of T. We will often make a minor misuse of language and say that $Z(X,T)$ is a rational function of T.*

(2) *(Functional equation) Suppose in addition that X is of pure dimension n. Then*

$$Z\left(X, \frac{1}{q^n T}\right) = \pm q^{nE/2} T^E Z(X,T)$$

for some integer E.

(3) *(Analogue of the Riemann hypothesis) Suppose in addition that X is smooth and projective*[1] *of (pure) dimension n. Then there is a unique factorization*

$$Z(X,T) = \frac{P_1(T) \cdots P_{2n-1}(T)}{P_0(T) \cdots P_{2n}(T)}$$

in which $P_i(T)$ factors over $\mathbb{C}$ as $\prod_j (1 - \alpha_{ij} T)$ where $|\alpha_{ij}| = q^{i/2}$ for all j. In particular, the integer E from (2) equals

$$E = \sum (-1)^i \deg(P_i).$$

Moreover, if X is geometrically irreducible, then $P_0(T) = 1 - T$ and $P_{2n}(T) = 1 - q^n T$.

(4) *(Betti numbers) Suppose in addition that there exist a number field K, a finite set S of prime ideals of $\mathcal{O}_K$, a maximal ideal $\mathfrak{p}$ of $\mathcal{O}_K$ not contained in S with residue field isomorphic to k, and a smooth projective scheme $\mathfrak{X}$ over $\mathcal{O}_{K,S}$ (the localization of $\mathcal{O}_K$ at the primes in S) such that X is isomorphic to the base extension $\mathfrak{X} \times_{\mathcal{O}_{K,S}} k$. (Informally, X is the "reduction of $\mathfrak{X}$ modulo $\mathfrak{p}$.") Then for any embedding $K \to \mathbb{C}$, the i-th Betti number of the topological space $(\mathfrak{X} \times_{\mathcal{O}_{K,S}} \mathbb{C})^{\mathrm{an}}$ equals $\deg(P_i)$.*

[1] Throughout these lectures, I will often say "projective" when I could say "proper" instead.

Example 2.4. If X is set-theoretically the disjoint union of an open subscheme Y and a closed subscheme S, then $X(k_r)$ is likewise the disjoint union of $Y(k_r)$ and $S(k_r)$, so formally

$$Z(X,T) = Z(Y,T) \cdot Z(S,T).$$

Let us apply this to the decomposition $\mathbb{P}^n = \mathbb{A}^n \sqcup \mathbb{P}^{n-1}$. We obtain:

$$Z(\mathbb{P}^n,T) = Z(\mathbb{A}^n,T) \cdot Z(\mathbb{P}^{n-1},T) = \frac{1}{1-q^nT} \cdot Z(\mathbb{P}^{n-1},T)$$

. In particular, as we have seen before,

$$Z(\mathbb{P}^1,T) = \frac{1}{(1-T)(1-qT)}$$

and similarly for $\mathbb{P}^n$ (see Set 2 exercises).

Example 2.5. For $X = C$ an elliptic curve, it can be shown by (relatively) elementary methods that

$$Z(C,T) = \frac{1-aT+qT^2}{(1-T)(1-qT)}$$

where a is an integer depending on C. It was shown by Hasse that moreover $|a| \leq 2q^{1/2}$; see [100, Chapter V] for an efficient proof.

Remark 2.6. Let's see in detail what the Weil conjectures say for $\mathbb{P}^1$ and C.

(1) Rationality is obviously true in both cases.
(2) The functional equation for $\mathbb{P}^1$:

$$Z(\mathbb{P}^1, \frac{1}{qT}) = \frac{qT^2}{(1-T)(1-qT)} \quad E = 2.$$

The functional equation for C:

$$Z(C, \frac{1}{qT}) = \frac{1-aT+qT^2}{(1-T)(1-qT)} \quad E = 0.$$

(3) The factorization for $\mathbb{P}^1$ is obvious, and the analogue of the Riemann hypothesis carries no new information. The factorization for C gives something nontrivial:

$$P_i(T) = \begin{cases} 1-T & i=0 \\ 1-aT+qT^2 & i=1 \\ 1-qT & i=2. \end{cases}$$

The analogue of the Riemann hypothesis asserts that the roots of $P_1(T)$ lie on the circle $|T| = q^{-1/2}$; given the shape of the factorization, this is equivalent to the Hasse bound.
(4) The Betti numbers of a topological $\mathbb{P}^1$ are $1,0,1$. The Betti numbers of a topological elliptic curve are $1,2,1$.

Remark 2.7. The factorization assertion was largely inspired by the example of Fermat hypersurfaces considered in the previous lecture. In that example, the numbers α_{ij} are the products of Gauss sums appearing in Weil's formula.

Remark 2.8. The Betti number statement is a proxy for a stronger statement that Weil was not in a position to formulate precisely: what we wanted is to have $P_i(T) = \det(1 - FT, V_i)$ where V_i is some "naturally occurring" vector space over a field and $F : V_i \to V_i$ is some endomorphism of the vector space. This perspective gives rise to the notion of *Weil cohomology* around which this course is centered.

But before we get there, note that the Betti number statement has a fair bit of power on its own. One important example computed by Weil in [115] is that of Grassmannian varieties, whose points correspond to subspaces of a fixed vector space. It is elementary to compute the number of points on a Grassmannian over a finite field (see Set 1 exercises); according to the Weil conjectures, this should then predict the Betti numbers of a Grassmannian over $\mathbb{C}$. These had been computed previously by Ehresmann using totally different methods.

Now that the Weil conjectures are a theorem, one can go further with this logic: in some cases, the first known computation of the Betti numbers of a topological space have used the Weil conjecture. A famous example is the Hilbert schemes of points on a smooth projective surface, by Göttsche [47].

We conclude this lecture with a very brief summary of how the Weil conjectures became a theorem. We will spend much of the course partially unpacking this summary.

(1) The rationality was first proved in 1958 by Dwork [37] using an interpretation of $Z(X,T)$ in terms of in terms of p-adic analysis (where p is the characteristic of the finite field).
(2) During the 1960s, Grothendieck [54, 55] led a heroic effort to develop modern foundations of algebraic geometry, including a theory of *étale cohomology* that was meant to simulate the role of topological (singular) cohomology for complex algebraic varieties. This led to a new proof of rationality (via a form of the Lefschetz trace formula as per Remark 2.8), together with the first proofs of the functional equation (arising from Poincaré duality) and the Betti number condition (arising from a comparison theorem with singular cohomology).
(3) Grothendieck proposed an approach to the analogue of the Riemann hypothesis via the so-called "Standard Conjectures" [56], but this approach never bore fruit.
(4) In the 1970s, Deligne [29] came up with a more *ad hoc* approach for part (3) and proved it. Shortly thereafter, he gave a more robust proof [30]; this paper (commonly known as "Weil II") is itself foundational in the study of zeta functions.
(5) An important simplification of "Weil II" was discovered by Laumon [76], inspired by the *stationary phase approximation* from classical analysis.
(6) Subsequently, Dwork's methods were adapted to give a parallel cohomology theory, again based on p-adic analysis, in which the entire étale-cohomological proof of the Weil conjectures can be emulated. For example, a p-adic adaption of Laumon's argument was given by Kedlaya [66].

3. Weil's cohomological metaconjecture (October 7)

In the previous lecture, we stated the Weil conjectures for an algebraic variety (or a scheme of finite type) X over a finite field $\mathbb{F}_q$, which imply that the zeta function

$$Z(X,T) = \exp\left(\sum_{n=1}^{\infty} \frac{T^n}{n} \#X(\mathbb{F}_{q^n})\right)$$

has properties that we identified as follows:

(1) (rationality)
(2) (functional equation)
(3) (Riemann Hypothesis)
(4) (Betti numbers)

As we pointed out in Remark 2.8, Weil went further and suggested an approach to these conjectures inspired by algebraic topology. In this lecture, we explain this approach.

Readings 3.1. We continue to follow [59, Appendix C].

Definition 3.2. Let R be a commutative algebra over $\mathbb{F}_q$; then the map $x \to x^q$ is an $\mathbb{F}_q$-homomorphism from R to itself. For any scheme X over $\mathbb{F}_q$, this construction induces a morphism $F : X \to X$ of $\mathbb{F}_q$-schemes, called the *absolute Frobenius* of X (more precisely, of X over $\mathbb{F}_q$). One easily sees that we have an action of F on the set $X(\overline{\mathbb{F}_q})$, whose set of fixed points is exactly $X(\mathbb{F}_q)$. Also if we consider the action of F^n on $X(\overline{\mathbb{F}_q})$, then the fixed points would be $X(\mathbb{F}_{q^n})$.

Remark 3.3. The inspiration for what follows is the general principle that the problem of counting fixed points of a self-map on a space should have something to do with computing traces of some associated linear map. A simple example of this principle is the following: if σ is a permutation of $\{1,\dots,n\}$, then the number of fixed points of σ is equal to the trace of the permutation matrix assocated to σ.

A vastly more sophisticated example is the *Lefschetz trace formula*. Let $T : S \to S$ be a continuous self-map of a topological space. Under suitable conditions, the quantity

$$\sum_i (-1)^i \operatorname{Trace}(T, H^i(S))$$

gives a weighted count of the fixed points of T; in particular, the nonvanishing of this quantity can be used to establish the existence of a fixed point of T (as in the Brouwer fixed point theorem).

With the above considerations Weil proposed the following.

Metaconjecture 3.4. (Weil) For some field K of characteristic 0, there is a series of contravariant "cohomology" functors

$$H^i : \{\text{algebraic varieties over } \mathbb{F}_q\} \to \{\text{finite dimensional vector spaces over } K\}$$

satsifying the following formula: for $i = 0, \ldots, 2d = 2\dim(X)$, satisfying the formula

$$\#X(\mathbb{F}_{q^n}) = \sum_{i=0}^{2d} (-1)^i \operatorname{Trace}(F^n | H^i(X))$$

for every positive integer n, where $F^n : H^i(X) \to H^i(X)$ denotes (by abuse of notation) the linear transformation induced by the morphism $F^n : X \to X$. (One can also formulate a similar metaconjecture in terms of a sequence of covariant "homology" functors H_i.)

Remark 3.5. Let us see what the metaconjecture says, or could say with some refinement, about the Weil conjectures.

Firstly, it immediately implies rationality because

$$Z(X,T) = \prod \det(1 - FT, H^i(X))^{(-1)^{i+1}}.$$

Note that here, we use crucially that K is of characteristic 0; otherwise, we would only get this relation modulo the characteristic of K.

Secondly, the functional equation would hold if the functors $H^i(X)$ satisfied "Poincaré Duality", in the sense of admitting a perfect, F-equivariant pairing

$$H^i(X) \times H^{2d-i}(X) \to K(-d)$$

where $K(-d)$ denotes the field K with the "twisted" F-action, sending 1 to q^d.

Thirdly, the Betti number statement would follow from an equality of dimensions between our $H^i(X)$ and the usual singular cohomology groups of the analytification.

It is not clear where the Riemann hypothesis would come from in this framework. We will discuss this later.

Let us note that we haven't talked much about the field of coefficients K which plays an important role in our cohomology theory here (except to note that it must be of characteristic 0). The following example shows that we cannot hope to take $K = \mathbb{Q}$.

Example 3.6. Suppose the metaconjecture holds for some K. Let $X/\mathbb{F}_q$ be a supersingular elliptic curve; we then have an action of $\operatorname{End}(X)$ on $H^1(X)$. As we have seen in the previous lecture, $H^1(X)$ is of dimension 2. However, if the endomorphisms of $X_{\overline{\mathbb{F}_q}}$ are all defined over $\mathbb{F}_q$, then $\operatorname{End}(X)$ whereas $\operatorname{End}(X)$ is a $\mathbb{Z}$-module of rank 4 contained in a (nonsplit) quaternion algebra over $\mathbb{Q}$. However, a quaternion algebra over $\mathbb{Q}$ cannot act on a 2-dimensional $\mathbb{Q}$-vector space unless it splits (i.e., is isomorphic to the matrix ring $\mathrm{M}_2(\mathbb{Q})$). Thus we cannot have $K = \mathbb{Q}$.

In this example, the quaterion algebra in question remains nonsplit after tensoring over $\mathbb{Q}$ with either $\mathbb{R}$ or $\mathbb{Q}_p$ (where p is the characteristic of $\mathbb{F}_q$). Consequently, the same argument rules out the possibility of satisfying the metaconjecture with $K = \mathbb{R}$ or $K = \mathbb{Q}_p$ (but it does not rule out extensions of these fields).

Remark 3.7. There are essentially two known approaches to constructing a Weil cohomology theory over a finite field of characteristic p.

- For $K = \mathbb{Q}_\ell$ where $\ell \neq p$ is prime (which is not precluded by Example 3.6), the construction of *étale cohomology* by Grothendieck et al. will satisfy the metaconjecture.

- For $K = \overline{\mathbb{Q}}_p$, the construction of *rigid cohomology* developed by Berthelot et al. will satisfy the metaconjecture. (Note that we cannot take $K = \mathbb{Q}_p$ because of Example 3.6.)

More on both of these later.

4. Curves and abelian varieties (October 9)

In this lecture, we study zeta functions for curves and abelian varieties.

Readings 4.1. We follow [81, Chapters VIII–IX]. For background on abelian varieties, see also [91].

Definition 4.2. Throughout this lecture, let X be a geometrically irreducible smooth projective curve of genus g over the finite field $k = \mathbb{F}_q$ of characteristic p. The field of rational functions $k(X)$ is finite over $k(t)$ for any element $t \in k(X)$ which is not in k (or equivalently, which is not integral over k; note that the geometrically irreducible condition implies that k is integrally closed in $k(X)$).

Let $\mathrm{Div}(X)$ be the free abelian group generated by the closed points X° of X; the elements of $\mathrm{Div}(X)$ are called *divisors* on X. We have a *degree* map

$$\begin{aligned} \deg : \mathrm{Div}(X) &\longrightarrow \mathbb{Z} \\ \sum a_i[P_i] &\longmapsto \sum a_i[\kappa(P_i) : k] \end{aligned}$$

where $\kappa(P)$ denotes the residue field of P. A divisor is called *effective* if it is a nonnegative linear combination of closed points; the degree of an effective divisor is also nonnegative.

Denote $\mathrm{Div}^0(X) := \deg^{-1}(0)$. Then for $f \in k(X)^\times$, the divisor

$$\mathrm{div}(f) = \sum_{P \in X^\circ} \mathrm{ord}_P(f)[P]$$

associated to f belongs to $\mathrm{Div}^0(X)$, hence

$$\mathrm{Pic}^0(X) := \mathrm{coker}(\mathrm{div} : k(X)^\times \to \mathrm{Div}^0(X))$$

is well-defined.

Remark 4.3. In what follows, it is helpful to bifurcate the discussion based on whether or not $X(k) = \emptyset$. For an example with $X(k) = \emptyset$, take the genus-2 curve

$$y^2 = 2x^6 - 2x^2 + 2$$

over $\mathbb{F}_3$. (Note: it is impossible to have $X(k) = \emptyset$ for a curve of genus 1 over a finite field; see the supplementary exercises.)

Suppose now that $X(k) \neq \emptyset$; then the degree map $\deg : \mathrm{Div}(X) \to \mathbb{Z}$ is evidently surjective. Specifically, if we fix a choice of $O \in X(k)$, we can define a map

$$\begin{aligned} \mathrm{cl} : \text{Effective divisors of degree } d &\longrightarrow \mathrm{Pic}^0(X) \\ D &\longmapsto [D - dO] \end{aligned}$$

which is surjective. For $d \geq 2g-1$, each fibre has order $\frac{q^{d-g+1}-1}{q-1}$ for $d \geq 2g-1$; this follows from the Riemann-Roch theorem, which implies that $h^0(X, \mathcal{L}) = \deg(\mathcal{L}) - g + 1$ for a line bundle $\mathcal{L}$ with $\deg(\mathcal{L}) \geq 2g - 1$. (We will use the full strength of Riemann-Roch a bit later.)

Now write

$$Z(X, T) = \prod_{x \in X^\circ} \frac{1}{1 - T^{\deg(x)}} = \sum_{D \geq 0} T^{\deg(D)}$$

where the last sum is over the effective divisors D on X (this is analogous to the equality between the sum and product representations of a Dedekind zeta function). Breaking this sum into two parts according to whether $\deg(D) \geq 2g - 1$ or $\deg(D) \leq 2g - 1$ leads to the following proposition.

Proposition 4.4. If $X(k) \neq \emptyset$, then $Z(X, T) = \frac{f(T)}{(1-T)(1-qT)}$ for some polynomial f with $\deg(f) \leq 2g$ and $f(1) = \#\,\mathrm{Pic}^0(X)$.

Remark 4.5. The equality $f(1) = \#\operatorname{Pic}^0(X)$, which crucially implies that f does not have a zero at $T = 1$, is analogous to a property of Dedekind zeta functions which we did not comment on earlier. For K a number field, the residue of $\zeta_K(s)$ at $s = 1$ (where the function has a simple pole) is given by the *class number formula*. It includes factors coming from the class number of $\mathcal{O}_K$ and the regulator of the unit lattice of K. In this context, there are no infinite places and so we see only a class number contribution.

Let us now see about getting rid of the condition that $X(k) \neq \emptyset$. Obviously X has points over *some* finite extension of k, so let us try passing from X to its base extension $X_{\mathbb{F}_{q^n}}$ for some positive integer n chosen so that $X(\mathbb{F}_{q^n}) \neq \emptyset$. We can then try to recover information about X using the identity

$$Z\left(X_{\mathbb{F}_{q^n}}, T^n\right) = \prod_{i=0}^{n-1} Z\left(X, \zeta_n^i T\right)$$

where ζ_n is a primitive n-th root of unity.

However, there is a strict loss of information between $Z(X,T)$ and $Z\left(X_{\mathbb{F}_{q^n}}, T\right)$, even for curves.

Example 4.6. If $Z(X_1,T) = \frac{1-aT+qT^2}{(1-T)(1-qT)}$, $Z(X_2,T) = \frac{1+aT+qT^2}{(1-T)(1-qT)}$ then $Z\left(X_{1,\mathbb{F}_{q^2}}, t\right) = Z\left(X_{2,\mathbb{F}_{q^2}}, t\right)$. This occurs when X_1 is an elliptic curve and X_2 is a quadratic twist; to make this explicit (assuming $p > 2$), let X_1 be a curve of the form

$$y^2 = x^3 + ax^2 + bx + c$$

and let X_2 be the curve

$$dy^2 = x^3 + ax^2 + bx + c$$

where d is a nonsquare in $\mathbb{F}_q^\times$.

A key observation is that the previous proof in the case $X(\mathbb{F}_q) \neq \emptyset$ only relies on the surjectivity of the degree map. Hence if could show such surjectivity always hold (without assuming $X(\mathbb{F}_q) \neq \emptyset$), then we do not have to worry about the existence of $O \in X(\mathbb{F}_q)$. Fortunately, this is the case.

Proposition 4.7. The degree map $\deg : \operatorname{Div}(X) \to \mathbb{Z}$ is always surjective, whether or not $X(k) \neq \emptyset$.

Proof. Since the degree map is clearly nonzero, we have $\deg(\operatorname{Pic}(X)) = e\mathbb{Z}$ for some positive integer e. Let us again compute $Z(X,T) = \sum_{D \geq 0} T^{\deg(D)}$ by breaking the sum in two as before; the second sum then runs over T^{de} with $d \geq d_0$, where d_0 is the smallest integer such that $d_0 e \geq 2g - 1$. As a result, we have

$$Z(X,T) = \frac{f\left(t^e\right)}{\left(1 - T^e\right)\left(1 - q^e T^e\right)}$$

and $f(1) = \#\operatorname{Pic}^e(X) \neq 0$. In particular, $Z(X,T)$ has a pole of order 1 at $T = 1$.

The same logic applies also to $X_{\mathbb{F}_{q^e}}$, so $Z\left(X_{\mathbb{F}_{q^e}}, T\right)$ has a pole of order 1 at $T = 1$. As a result, $Z\left(X_{\mathbb{F}_{q^e}}, T^e\right)$ has a pole of order 1 at $T = 1$. On the other hand,

$$Z\left(X_{\mathbb{F}_{q^e}}, T^e\right) = \prod_{i=0}^{e-1} Z\left(X, \zeta_e^i T\right) = Z\left(X,T\right)^e.$$

Comparing the pole orders at $T = 1$, we deduce that $e = 1$, which finishes the proof. □

Remark 4.8. Using the full strength of the Weil conjectures, one can prove more: for any fixed X, we have $X(\mathbb{F}_{q^n}) \neq \emptyset$ for *every* sufficiently large n. See the supplementary exercises.

Given Proposition 4.7, we can now reprise the proof of Proposition 4.4 to deduce the following.

Proposition 4.9. For any X, $Z(X,T) = \frac{f(T)}{(1-T)(1-qT)}$ for some polynomial f with $\deg(f) \leq 2g$ and $f(1) = \#\operatorname{Pic}^0(X)$.

Note that we currently only know that $\deg(f) \leq 2g$, whereas we expect equality. To resolve this, we must prove the functional equation using the Riemann-Roch theorem.

Proposition 4.10. We have $Z(X, 1/(qT)) = q^{-g}T^{2-2g}Z(X,T)$. Consequently, $f(q^{-1}T^{-1}) = q^{-g}T^{-2g}f(T)$ and $\deg(f) = 2g$.

Proof. Write $(q-1)Z(X,T)$ as a sum of two terms:

$$\alpha(T) := \sum_{0 \le \deg(\mathcal{L}) \le 2g-2} q^{h^0(\mathcal{L})} T^{\deg(\mathcal{L})}$$
$$\beta(T) := \sum_{\deg(L) \ge 2g-1} q^{h^0(\mathcal{L})} T^{\deg(\mathcal{L})} - \sum_{\deg(\mathcal{L}) \ge 0} T^{\deg(\mathcal{L})}.$$

We will prove that each of these satisfies the same functional equation that we desire for $Z(X,T)$. For $\beta(T)$, using the weak form of Riemann-Roch used earlier, we obtain

$$\beta(T) = \#\operatorname{Pic}^0(X) \left(\frac{q^g T^{2g-1}}{1-qT} - \frac{1}{1-T} \right)$$

and the functional equation is clear. To analyze $\alpha(T)$, we must use Riemann-Roch at full strength: for Ω the sheaf of Kähler differentials on X and $\mathcal{L}$ any line bundle on X,

$$h^0(X,\mathcal{L}) = \deg(\mathcal{L}) + 1 - g + h^0(\Omega \otimes \mathcal{L}^{-1}).$$

Since $\deg(\Omega) = 2g-2$, we may rewrite $\alpha(T)$ by substituting $\Omega \otimes \mathcal{L}^{-1}$ for $\mathcal{L}$. Using Riemann-Roch, we then obtain

$$\begin{aligned} \alpha(T) &= \sum_{0 \le \deg(\mathcal{L}) \le 2g-2} q^{h^0(\Omega \otimes \mathcal{L}^{-1})} T^{\deg(\Omega \otimes \mathcal{L}^{-1})} \\ &= \sum_{0 \le \deg(\mathcal{L}) \le 2g-2} q^{h^0(\mathcal{L}) - \deg(\mathcal{L}) - 1 + g} T^{2g-2-\deg(\mathcal{L})} \end{aligned}$$

and again read off the desired functional equation. □

We will show a bit later, using the Riemann-Roch theorem, that $Z(X,T)$ satisfies the functional equation; this will also show that $\deg(f) = 2g$. One can also establish the Riemann hypothesis in this framework, but we postpone this to a later lecture.

In the remainder of this lecture, we describe (without proofs) the relationship between curves and abelian varieties, and between the Weil conjectures in these two cases.

Definition 4.11. An *abelian variety* over a field k is a smooth, projective, geometrically connected k-scheme equipped with a commutative group structure. It turns out that the commutativity hypothesis is superfluous; see [91].

Example 4.12. Elliptic curves over k are abelian varieties of dimension 1. Products of elliptic curves give examples of higher-dimensional abelian varieties.

Definition 4.13. Given a curve of genus g, there are two different constructions giving rise to a g-dimensional abelian variety.

- The *Albanese construction*:

  $$\text{pointed curve } X/k \text{ of genus } g \rightsquigarrow \operatorname{Alb}(X)$$

  This is a covariant functor, and comes with a (functorial) map $X \to \operatorname{Alb}(X)$ sending the marked point to the identity. This map does not factor through any abelian subvariety of $\operatorname{Alb}(X)$, and induces a homomorphism

  $$\operatorname{Div}^0(X) \to \operatorname{Alb}(X)(k)$$

  which factors through $\operatorname{Pic}^0(X)$.
- The *Picard construction*:

  $$\text{curve } X/k \text{ of genus } g \rightsquigarrow \underline{\operatorname{Pic}}^0(X) := \text{Moduli space of degree-0 line bundles on } X.$$

  This is a contravariant functor. The following universal property holds: maps from an abelian variety S over k to $\underline{\operatorname{Pic}}^0(X)$ correspond to line bundles on $S \times_k X$ whose restriction to every fiber $s \times X$ has degree 0.

Remark 4.14. These two construction are related by the *Abel-Jacobi map*:

$$\mathrm{Alb}^\vee(X) \cong \underline{\mathrm{Pic}}^0(X)$$

where for an abelian variety A, the *dual variety* $A^\vee$ is defined as $\underline{\mathrm{Pic}}^0(A)$. Using the *Poincaré bundle*, we obtain a natural isomorphism $(A^\vee)^\vee \cong A$.

For a general abelian variety A over k, A and $A^\vee$ need not be isomorphic (although they are necessarily *isogenous*. However, one can construct a *principal polarization* giving rise to an isomorphism

$$\mathrm{Alb}(X) \cong \underline{\mathrm{Pic}}^0(X).$$

Example 4.15. Over $\mathbb{C}$, every abelian variety arises analytically as a complex torus $\mathbb{C}^g/\Lambda$. The dual variety is then $(\mathbb{C}^g/\Lambda)^\vee \cong \mathbb{C}^g/\Lambda^\vee$, where $\Lambda^\vee := \{\mu : \mathrm{Hom}_{\mathbb{R}}(\mathbb{C}^g, \mathbb{R}) |\ \mu(\Lambda) \subset \mathbb{Z}\}$.

The zeta functions of a curve A and its Jacobian $\mathrm{Jac}(X) := \underline{\mathrm{Pic}}^0(X)$ are related as follows.

Theorem 4.16. *Suppose A is an abelian variety over $k = \mathbb{F}_q$ of dimension g.*

(1) *The zeta function for A is*

$$Z(A,T) = \frac{P_1(T)\cdots P_{2g-1}(T)}{P_0(T)\cdots P_{2g}(T)}$$

where $P_0(T) = 1 - T$, $P_{2g}(T) = 1 - q^g T$, and $P_i(T) = \wedge^i P_1(T)$ for $i = 1, \ldots, 2g$ in the sense that if $P_1(T) = \prod_j (1 - \alpha_j T)$, then $P_i(T) = \prod_{j_1 < \cdots < j_i} (1 - \alpha_{j_1} \cdots \alpha_{j_i} T)$. Note that this implies $\#X(k) = P_1(1) = \prod_j (1 - \alpha_j)$.

(2) *If $A \cong \mathrm{Jac}(X)$, then $Z(X,T) = \frac{P_1(T)}{(1-T)(1-qT)}$ for the same P_1.*

5. Two approaches to RH for curves (October 14)

In this lecture we examine two of the three "elementary" approaches to the Riemann hypothesis for curves over finite fields (that is, the approaches that do not require Weil cohomology).

(1) Comparison of a curve with its Jacobian. This is the first proof announced by Weil.
(2) Intersection theory on the self-product of the curve. This is the second proof announced by Weil.
(3) Clever use of Riemann-Roch. This approach was introduced by Stepanov for hyperelliptic curves [101] and generalized to all curves by Bombieri [10].

This list is given in order of first appearance, but we will proceed in the opposite order, focusing in this lecture on the Bombieri–Stepanov method and then the second proof of Weil. We will turn to the first proof of Weil in a subsequent lecture.

Readings 5.1. For the Bombieri–Stepanov method, we continue to follow [81, Chapters VIII–IX]. For the second method of Weil, we follow [59, Exercise V.1.10].

Throughout this lecture, let X be a geometrically irreducible smooth projective curve of genus g over the finite field $k = \mathbb{F}_q$ of characteristic p, and write $q = p^a$. Let us first summarize what we established in the previous lecture.

Proposition 5.2. We have

$$Z(X,T) = \frac{P(T)}{(1-T)(1-qT)},$$

where $P(T) \in \mathbb{Z}[T]$ is a polynomial satisfying:

- $P(0) = 1$;
- $\deg(P(T)) = 2g$;
- $P(T) = 1 + a_1 T + \cdots + a_{2g-1} T^{2g-1} + q^g T^{2g}$, with $a_{g+i} = q^i a_{g-i}$.

Our goal is therefore to prove the Riemann hypothesis for X, which amounts to the assertion that the roots of $P(T)$ lie on the circle $|T| = q^{-1/2}$.

We give some initial preparation the Bombieri-Stepanov method.

Remark 5.3. Recall that if we can prove the Riemann Hypothesis for a base extension $X_{\mathbb{F}_{q^n}}$ of X, then this will imply the Riemann hypothesis for X because the zeroes and poles of $Z(X_{\mathbb{F}_{q^n}}, T)$ are the n-th powers of the zeroes and poles of $Z(X, T)$. In particular, we can arrange for q to be "sufficiently large" compared to g.

Definition 5.4. Let $\alpha_1^{-1}, \ldots, \alpha_{2g}^{-1}$ be the roots of $P(T)$, labeled so that $|\alpha_1| \leq \cdots \leq |\alpha_{2g}|$; the functional equation implies that $q/\alpha_i = \alpha_{2g-i}$. Utilizing the equality

$$\log\left(\frac{P(T)}{(1-T)(1-qT)}\right) = \sum_{N=0}^{\infty} \frac{\#X(\mathbb{F}_{q^N})}{N} T^N,$$

expanding power series, and matching coefficients, we obtain

$$\#X(\mathbb{F}_{q^N}) = q^N + 1 - \sum_{i=1}^{2g} \alpha_i^N$$

for all $N \geq 1$. In particular, the Riemann Hypothesis would imply

$$\left|\#X(\mathbb{F}_{q^N}) - q^N - 1\right| \leq Cq^{N/2}$$

for $N \geq 1$ and C a constant (we can take $C = 2g$). A key point here is that the reverse implication is also true!

Lemma 5.5. Assume there exists an integer $d \geq 1$ and a constant C_0 for which

$$|\#X(\mathbb{F}_{q^{dN}}) - q^{dN} - 1| \leq C_0 q^{dN/2}$$

for all N. Then the Riemann Hypothesis for X holds.

Proof. The hypothesis implies that

$$\sum_{N=0}^{\infty} (\alpha_1^{dN} + \cdots + \alpha_{2g}^{dN}) T^N$$

converges in the open disc $|T| < q^{-d/2}$ (say, by the root test). In particular, the power series

$$\sum_{i=1}^{2g} (1 - \alpha_i^d T)^{-1}$$

converges uniformly on $|T| < q^{-d/2}$ (i.e., there are no poles), so that $|\alpha_i| \geq q^{-1/2}$. The functional equation then tells us that $\alpha_{2g-i} = \alpha_i/q$, and hence we obtain $|\alpha_i| = q^{-1/2}$ for all i. □

We are thus reduced to proving an upper bound and a lower bound on $\#X(\mathbb{F}_q)$. We start with the former, again keeping in mind that we may apply this after performing a base change.

Theorem 5.6. *Let $q = p^s$, with s even and $q > (g+1)^4$. Then*

$$\#X(\mathbb{F}_q) \leq q + 1 + (2g+1)\sqrt{q}.$$

Proof. There is nothing to verify if $\#X(\mathbb{F}_q)$ is empty, so assume there is an $\mathbb{F}_q$-rational point on X and call it ∞. The goal is to write down a rational function on X with a controlled pole at ∞ and with zeroes at $X(\mathbb{F}_q) \setminus \{\infty\}$; this would then imply $\#X(\mathbb{F}_q) \leq 1 + P$, where P denotes the pole order of the function at ∞. To this end, let

$$H_m := \{f \in K(X) : \operatorname{div}(f) \geq -m\infty\}$$
$$H_m^{p^\mu} := \{f^{p^\mu} : f \in H_m\}.$$

Let us consider a function

$$f = \sum \nu_i s_i^q,$$

with $\nu_i \in H_l^{p^\mu}$ and $s_i \in H^m$. Suppose that f is not identically zero and that $\delta(f) = \sum \nu_i s_i = 0$. It follows that f vanishes on $X(\mathbb{F}_q) \setminus \{\infty\}$. If we assume moreover that $p^\mu < q$, then f is a perfect p^μ-th power and hence vanishes to order p^μ at each of its zeroes; in particular, we obtain

$$\#X(\mathbb{F}_q) \leq 1 + \deg(f)/p^\mu \leq 1 + l + mq/p^\mu.$$

Now we examine when such an f exists. By polar expansion around infinity, one may show that the map

$$\delta : H_l^{p^\mu} \cdot H_m^q \to H_{lp^\mu + m}$$

is in fact a well-defined linear morphism; moreover, if we assume additionally that $lp^u < q$, a straightforward calculation gives an isomorphism $H_l^{p^\mu} \cdot H_m^q \cong H_l^{p^\mu} \otimes H_m^q$, and hence

$$\dim_{\mathbb{F}_q} H_l^{p^\mu} \cdot H_m^q = \dim_{\mathbb{F}_q}(H_l^{p^\mu}) \dim_{\mathbb{F}_q}(H_m^q).$$

By Riemann-Roch,

$$\dim_{\mathbb{F}_q} H_l^{p^\mu} = \dim_{\mathbb{F}_q} H_l \geq \max\{1, l+1-g\}.$$

Hence δ will have a nontrivial kernel whenever

$$(l+g-1)(m+1-g) - (lp^\mu + m + 1 - g) > 0.$$

To optimize this, choose $\mu = s/2$ and $m = \sqrt{q} + 2g$. All of the requisite conditions will be satisfied if we can choose an *integer* l for which

$$q + \frac{g}{g+1}\sqrt{q} < l < \sqrt{q}.$$

This is possible so long as $q > (g+1)^4$; with these choices of l, m, μ the bound reduces to

$$\#X(\mathbb{F}_q) \leq 1 + \sqrt{q} + (\sqrt{q} + 2g)\sqrt{q}$$

as desired. □

The previous method does not directly give a lower bound. Instead, we use a trick to convert the lower bound problem into a collection of upper bound problems that can be treated as before.

Definition 5.7. For a Galois cover of curves $\pi : X \to S$ and an element $\sigma \in \mathrm{Gal}(X/S)$, let $N(X/S, \sigma)$ denote the number of points $P \in X(\overline{\mathbb{F}_q})$ which lie above a point of $S(\mathbb{F}_q)$ in an unramified way and for which σ acts as the Frobenius on P.

Lemma 5.8. Let $\pi : X \to S$ be a Galois cover of curves defined over $\mathbb{F}_q$, with $q > (g(X)+1)^4$. Then $N(X/S, \sigma) < q + 1 + (2g(X)+1)\sqrt{q}$.

Proof. Let ∞ be a point counted by $N(X/S, \sigma)$ (if there are no such points there is nothing to prove). Consider the endomorphism $\phi := \sigma^{-1} \circ \mathrm{Frob}$ on X; it suffices to bound the fixed points of $\overline{\phi}$ on $X_{\overline{\mathbb{F}}_q}$.

Maintaining the notation of Theorem 5.6, any nonconstant function in $\overline{\phi}^*(H_m)$ has a pole solely at ∞, since $\overline{\phi}^*(H_m) \subset H_{qm}$. Consider $f = \sum v_i \overline{\phi}^*(s_i)$ in $H_l^{p^\mu} \overline{\phi}^*(H_m)$, and set $\delta(f) = \sum v_i s_i$. As in the previous proof, if there exists a nonzero function f for which $\delta(f) = 0$, it follows that f vanishes at all points counted by $N(X/S, \sigma)$. One then proceeds as before to show that δ must have a nontrivial kernel once q is suitably chosen with respect to g. □

This becomes helpful when we combine all of the automorphisms σ.

Lemma 5.9. Let $\pi : X \to S$ be a Galois cover of curves defined over $\mathbb{F}_q$. Then

$$\left| \sum_{\sigma \in \mathrm{Gal}(X/S)} N(X/S, \sigma) - \#\,\mathrm{Gal}(X/S) \#S(\mathbb{F}_q) \right|$$

is bounded by a constant depending only on $g(X)$ and $\deg(\pi)$.

Proof. Both $\sum_{\sigma \in \mathrm{Gal}(X/S)} N(X/S, \sigma)$ and $\#\,\mathrm{Gal}(X/S) \#S(\mathbb{F}_q)$ can be written as a sum over points $P_0 \in S(\mathbb{F}_q)$. If P_0 is not a branch point of π, then P_0 makes identical contributions to both quantities. Thus the discrepancy comes only from fibers containing branch points, the number of which is controlled by the Riemann-Hurwitz formula. □

We now derive the desired lower bound, thus completing the Bombieri-Stepanov proof of the Riemann hypothesis for curves.

Lemma 5.10. There exist an integer $d \geq 1$ and a constants C_0 for which for all positive integers N,

$$\#X(\mathbb{F}_{q^{dN}}) \geq q^{dN} - C_0 q^{dN/2}.$$

Proof. If X itself can be written as a Galois cover of $\mathbb{P}^1$ via some map π, then Lemma 5.9 implies that an upper bound on $N(X/S, \sigma)$ for each nontrivial automorphism σ implies a lower bound on $N(X/S, \mathrm{id}_X) = \#X(\mathbb{F}_q)$. So in this case, we just apply Lemma 5.8 and we are done.

In general, X cannot always be written as a Galois cover of $\mathbb{P}^1$ (e.g., if it has trivial automorphism group and positive genus). However, we can always choose a finite separable morphism $X \to \mathbb{P}^1$ (perhaps after extending the base field, although this isn't really needed) and then take its Galois closure to obtain a Galois cover $Z \to X$ for which $Z \to X \to \mathbb{P}^1$ is also Galois. By applying Lemma 5.9 to both $Z \to X$ and $Z \to \mathbb{P}^1$, we may again reduce the desired lower bound to some instances of Lemma 5.8. □

We now shift our attention to Weil's second method, whose main tools are the intersection pairing on surfaces and the Hodge index theorem. We briefly recall these two objects.

Definition 5.11. Let S be a smooth projective surface over a field k. There is a unique bilinear pairing

$$\mathrm{Div}(S) \times \mathrm{Div}(S) \to \mathbb{Z},$$

called the *intersection pairing*, with the following properties.

- If D_1 and D_2 are effective divisors on S without common components, then
$$D_1 \cdot D_2 = \mathrm{length}_k(D_1 \times_k D_2).$$
In other words, the pairing measures usual intersections when possible.
- The pairing depends solely on linear equivalence; i.e. if $D_1 \sim_{\mathrm{lin}} D_1'$ and $D_2 \sim_{\mathrm{lin}} D_2'$, then $D_1 \cdot D_2 = D_1' \cdot D_2'$.

The intersection pairing furthermore can be shown to satisfy the *adjunction formula*: if $C \hookrightarrow S$ is a closed immersion and C is a smooth, projective, geometrically irreducible curve of genus g over k, then

$$C \cdot (C + K) = 2g - 2,$$

where K is the canonical divisor (or rather, "a" canonical divisor) on S. See [59, §V.1].

Having set up the intersection pairing on surfaces, we can state the Hodge Index Theorem [59, Theorem V.1.9].

Theorem 5.12 (Hodge Index Theorem)**.** *Let H be an ample divisor on the surface S. Then for any divisor D, $D \cdot H = 0$ implies $D \cdot D \leq 0$.* □

With this setup, we can proceed with Weil's proof. The idea is to apply the previous two theorems with $S = X \times_k X$. The surface S comes equipped with two natural divisors:

- Δ, the diagonal embedding $X \hookrightarrow X \times_k X$;
- and Γ, the graph of the Frobenius morphism.

One can verify (by working locally) that Δ and Γ have no common component and intersect transversally. Furthermore, the intersection $\Delta \times_S \Gamma$ is naturally identified with $X(\mathbb{F}_q)$. Thus utilizing the intersection pairing we can write

$$\#X(\mathbb{F}_q) = \Delta \cdot \Gamma.$$

We need the following preparatory lemma.

Lemma 5.13. Let H be an ample divisor, and D an arbitrary divisor, on $S = X \times_k X$. Let $\sigma(1,0)$ and $\sigma(0,1)$ be the divisors obtained by pulling back a hyperplane section on X from the first and second projections respectively.

(1) We have $D^2H^2 \leq (D \cdot H)^2$
(2) For $S = X \times X$, $D^2 \leq 2(D \cdot \sigma(1,0))(D \cdot \sigma(0,1))$, with equality if and only if $D = a\sigma(1,0) + b\sigma(0,1)$.

Proof. For the first statement, we may take an orthogonal decomposition of the space of divisors to write $D = aH + bE$, where $E \cdot H = 0$. Then $D^2H^2 = ((aH)^2 + (bE)^2)H^2$. By the Hodge index theorem, $(bE)^2 \leq 0$, so $D^2H^2 \leq (aH)^2H^2$. But now $(aH)^2H^2 = (H \cdot (aH + bE))^2 = (D \cdot H)^2$ as desired. For the second statement, apply the first statement to the ample divisor $\sigma(1,1) = \sigma(0,1) + \sigma(1,0)$. □

To finish Weil's proof we now need the following computations, which follow from adjunction:

$$\begin{aligned}\Delta^2 &= 2-2g\\ \Gamma^2 &= q(2-2g)\\ \Delta\cdot\sigma(1,0) &= 1\\ \Delta\cdot\sigma(0,1) &= 1\\ \{\Gamma\cdot\sigma(1,0),\Gamma\cdot\sigma(0,1)\} &= \{1,q\}.\end{aligned}$$

Now apply the previous lemma to $a\Gamma + b\Delta$ to obtain

$$a^2\Gamma^2 + 2ab\Gamma\cdot\Delta + b^2\Delta^2 \leq 2(a\Gamma+b\Delta)\cdot\sigma(0,1)(a\Gamma+b\Delta)\cdot\sigma(1,0).$$

Simplifying gives

$$0 \leq 2(a+b)(qa+b) - a^2q(2-2g) - b^2(2-2g) - 2ab\#X(\mathbb{F}_q).$$

In other words, we have a semipositive quadratic form in a and b represented by the matrix

$$\begin{pmatrix} q(2g-1) & 2(q+1)-2ab\#X(\mathbb{F}_q)\\ 2(q+1)-2ab\#X(\mathbb{F}_q) & (2g-1)\end{pmatrix};$$

by Sylvester's criterion, semipositivity implies

$$\frac{q(2g-1)^2}{4} - (q+1-\#X(\mathbb{F}_q))^2 \geq 0,$$

giving a bound as in Lemma 5.5 and thus completing the proof.

6. RH for abelian varieties (October 16)

In this lecture, we discuss Weil's first proof of the Riemann hypothesis for curves, and the Weil conjectures for abelian varieties.

Readings 6.1. We follow the presentation of Weil's proof given in [87]. For background on abelian varieties, see [91].

We begin by discussing Weil's construction of the Jacobian of a curve.

Definition 6.2. Let X be a smooth, projective, geometrically irreducible curve of genus g. Consider the symmetric product $\mathrm{Sym}^g X$. We want to obtain an abelian variety from the symmetric product. One begins by fixing a point on $\mathrm{Sym}^g X$ and constructing rational maps $m : \mathrm{Sym}^g X \times \mathrm{Sym}^g X \dashrightarrow \mathrm{Sym}^g X$ and $i : \mathrm{Sym}^g X \dashrightarrow \mathrm{Sym}^g X$, which serve as a birational multiplication law and inverse law, respectively. Weil proved that any such set up as above is birational to a genuine group variety, the *Jacobian* $\mathrm{Jac}(X)$. The construction of the multiplication law above, for example, comes from Riemann-Roch. A more modern way to view Jacobians is as the moduli space of degree 0 line bundles on a curve (as discussed in a previous lecture).

We want to study the action of Frobenius on the Jacobian. It makes sense to more generally discuss endomorphisms on abelian varieties. We will want to study the action of Frobenius, or more generally any endomorphism, on the Tate module of an abelian variety.

Definition 6.3. Let A be an abelian variety over a field k. Let ℓ be a prime nonzero in k. The *ℓ-adic Tate module* $T_\ell(A)$ is defined as

$$T_\ell(A) := \varprojlim A(\overline{k})[\ell^n].$$

The *rational ℓ-adic Tate module* $V_\ell(A)$ is the base extension $T_\ell(A)\otimes_{\mathbb{Z}_\ell}\mathbb{Q}_\ell$.

The Tate module is a free rank-$2g$ module over the ℓ-adic integers $\mathbb{Z}_\ell$ (this will follow from Example 6.8 below), and it comes equipped with an action of the absolute Galois group of k. It records information about endomorphisms faithfully, in the following sense.

Theorem 6.4. *For any abelian varieties A and B over a field k, and any prime ℓ nonzero in k, the map*

$$\mathbb{Z}_\ell\otimes_{\mathbb{Z}}\mathrm{Hom}(A,B) \to \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(A),T_\ell(B))$$

is injective. In particular, $\mathrm{Hom}(A,B)$ *is a finite $\mathbb{Z}$-module.*

Proof. See [91, §18, Theorem 3]. □

For $\alpha \in \mathrm{End}(A)$, the characteristic polynomial of α acting on the Tate module $T_\ell(A)$ is of degree $2g$ as a polynomial over $\mathbb{Z}_\ell$.

Corollary 6.5. The minimal and characteristic polynomials of α on $T_\ell(A)$ are defined over $\mathbb{Z}$.

Proof. By the Cayley-Hamilton theorem, the minimal polynomial of α kills α in $\mathbb{Z}_\ell \otimes_{\mathbb{Z}} \mathrm{End}(A)$. From the injection

$$\mathrm{End}(A) \hookrightarrow \mathbb{Z}_\ell \otimes \mathrm{End}(A),$$

one sees that the linear dependence relation between $1, \alpha, \ldots, \alpha^{2g-1}$ coming from the Cayley-Hamilton theorem provides a linear dependence relation in $\mathrm{End}(A)$. Because $\mathrm{End}(A)$ is a finite free $\mathbb{Z}$-module, the minimal polynomial and characteristic polynomial of α on the Tate module $T_\ell(A)$ are both actually polynomials defined over the integers. □

We will now use this result to obtain information about the number of rational points on an abelian variety over a finite field.

Definition 6.6. Let A be an abelian variety over a finite field $k = \mathbb{F}_q$, and let $F : A \to A$ be the Frobenius endomorphism over $\mathbb{F}_q$. Then one can see easily that $A(k) = A[1 - F]$, where the notation $A[\alpha]$ denotes the kernel of the endomorphism α (the key point is that $A[1 - F]$ is reduced, which is an easy local calculation).

Hence to count $A(k)$, we want to understand the kernel of $1 - F$. We do this using the degree function.

Definition 6.7. Let $\mathcal{L}$ be a symmetric ample line bundle on A. The *degree* map $\deg : \mathrm{End}(A) \to \mathbb{Q}$ is given by the formula

$$\alpha \mapsto c_1(\alpha^* \mathcal{L})^g / c_1(\mathcal{L})^g,$$

where c_1 denotes the first Chern class.

If α is an endomorphism with finite kernel (i.e., an *isogeny* from A to itself), then α defines a finite morphism $A \to A$, and the degree of this morphism is the same as the quantity $\deg(\alpha)$ defined above. It is also equal to the k-length of the scheme-theoretic kernel $A[\alpha]$, which agrees with the number of geometric points of the kernel if α is separable. (For example, $\alpha = 1 - F$ is always separable.)

On the other hand, if α does not have finite kernel, then from dimensional considerations one may see that $\deg(\alpha) = 0$.

Example 6.8. For any positive integer n, using the *theorem of the cube* we may see that $[n]^* \mathcal{L} \cong \mathcal{L}^{n^2}$. Therefore $\deg([n]) = n^{2g}$.

Proposition 6.9. The degree map on $\mathrm{End}(A)_{\mathbb{Q}}$ is a polynomial of degree $2g$.

Proof. As in Example 6.8, we see that $\deg(n\alpha) = n^{2g} \deg(\alpha)$ for any $\alpha \in \mathrm{End}(A)_{\mathbb{Q}}$. Hence to prove that deg is a polynomial of degree $2g$, it will suffice to show that for any fixed $\alpha, \beta \in \mathrm{End}(A)$, $\deg(\alpha + n\beta)$ is a polynomial in n of degree at most $2g$. This again can be deduced using the theorem of the cube; we omit the details. □

Proposition 6.10. For $\alpha \in \mathrm{End}(A)$, the determinant of α acting on $T_\ell(A)$ is equal to $\deg(A)$.

Proof. We now know that both quantities are polynomials of degree $2g$ on $\mathrm{End}(A)$. To compare them, we first examine what happens on ℓ-power torsion to see that

$$|\det(\alpha, T_\ell(A))|_\ell = |\deg(\alpha)|_\ell \qquad (\alpha \in \mathrm{End}(A)).$$

In particular, for any fixed $\alpha \in \mathrm{End}(A)$,

$$|\det(F(\alpha), T_\ell(A))|_\ell = |\deg(F(\alpha))|_\ell \qquad (F \in \mathbb{Z}[T]).$$

By an elementary argument (see the supplemental exercises), this is enough to deduce that $\det(\alpha, T_\ell A) = \deg(\alpha)$. □

In particular, $\#A(\mathbb{F}_{q^n}) = (1 - r_1^n) \dots (1 - r_{2g}^n)$ where $r_1, \dots, r_{2g}$ are the roots of the characteristic polynomial of Frobenius acting on $T_\ell(A)$ for any ℓ. It follows (see Set 3 exercises) that $Z(A, T)$ is of the form

$$\frac{P_1(T) \dots P_{2g-1}(T)}{P_0(T) \dots P_{2g}(T)},$$

where $P_1(T) = (1 - r_1 T) \dots (1 - r_{2g} T)$ and $P_i(T) = \wedge^i P_1(T)$.

Example 6.11. If A is the product of the abelian varieties A_1 and A_2, then $\#A(\mathbb{F}_{q^n}) = \#A_1(\mathbb{F}_{q^n}) \#A_2(\mathbb{F}_{q^n})$ for all n. From this, it follows that the polynomial P_1 for A is the product of the polynomials P_1 for A_1 and A_2.

Remark 6.12. From the interpretation of $\#A(\mathbb{F}_{q^n})$ as $\deg(1 - F^n)$, we see that it is invariant under isogeny, as then is the whole zeta function. Beware however that the *structure* of the abelian group $A(\mathbb{F}_{q^n})$ is not an isogeny invariant.

This leaves the matter of the Riemann hypothesis. As in Weil's second proof of the Riemann hypothesis for curves, the key is a positivity assertion, here given in terms of the Rosati involution.

Definition 6.13. Let $\lambda : A \to A^\vee$ be an isogeny induced by a polarization of A, and define $\mathrm{End}(A)_\mathbb{Q} := \mathrm{End}(A) \otimes_\mathbb{Z} \mathbb{Q}$. The *Rosati involution* is the map $\dagger : \mathrm{End}(A)_\mathbb{Q} \to \mathrm{End}(A)_\mathbb{Q}$ given by

$$\alpha \mapsto \lambda^{-1} \circ \alpha^\vee \circ \lambda := \alpha^\dagger.$$

Theorem 6.14. *The pairing* $(\alpha, \beta) = \mathrm{Trace}(\alpha \circ \beta^\dagger) : \mathrm{End}(A)_\mathbb{Q} \times \mathrm{End}(A)_\mathbb{Q} \to \mathbb{Q}$ *is positive definite.*

Proof. For $\mathcal{L}$ an ample line bundle defining the polarization, we obtain

$$\mathrm{Trace}(\alpha \circ \alpha^\dagger) = \frac{2g}{\mathcal{L}^g} (\mathcal{L}^{g-1} \cdot \alpha^* \mathcal{L}).$$

This number is positive for $\alpha \neq 0$ because $\mathcal{L}$ is ample. □

Theorem 6.15. *The analogue of the Riemann hypothesis holds for abelian varieties over k.*

Proof. Let $F : A \to A$ again be the Frobenius endomorphism of an abelian variety A over $k = \mathbb{F}_q$. Using the Weil pairing (or a more direct calculation), one may calculate that $F^\dagger \circ F = [q]$. In particular the ring $\mathbb{Q}[F] \subset \mathrm{End}(A)$ is stable under $\dagger$, and hence must be semisimple. Hence $\mathbb{Q}[F] \otimes_\mathbb{Q} \mathbb{R}$ must also be semisimple, meaning that it is a product of copies of $\mathbb{R}$ and $\mathbb{C}$; the action of $\dagger$ extends to $\mathbb{Q}[F] \otimes_\mathbb{Q} \mathbb{R}$ and (in order to obey positivity) must fix each copy of $\mathbb{R}$ and conjugate each copy of $\mathbb{C}$. Each eigenvalue r of $\mathbb{C}$ appears as the image of F in one of the factors, and the image of $F^\dagger$ in the same factor is $\overline{r}$; hence $\left| r^2 \right| = q$. □

As a corollary, we may now recover the Riemann hypothesis for a curve over k, by applying the previous theorem to its Jacobian.

Remark 6.16. Alternatively, one can go the other way and deduce the Riemann hypothesis for abelian varieties from the corresponding statement for curves. This seems not obvious at first, because an arbitrary abelian variety A over k is not isomorphic, or even isogenous, to a Jacobian. However, if we take X to be a transverse intersection of $\dim(A) - 1$ ample divisors (see the next remark), then $\mathrm{Jac}(X)$ maps surjectively onto A, and is in fact isogenous to the product of A with the kernel A' of the map. As per Example 6.11, the polynomial P_1 for $\mathrm{Jac}(X)$ factors as the product of the P_1 polynomials for A and A', so the Riemann hypothesis for X does imply the Riemann hypothesis for A.

Remark 6.17. In the previous remark, it is not immediately obvious that a transverse intersection exists because we are working over a finite field; the assertion of the Bertini smoothness theorem that a "generic" intersection is transverse is of no value. However, one may apply Bertini over $\overline{k}$ and then just make a suitable base extension; the latter does not affect the proof of the Riemann hypothesis. Alternatively, one can find such an intersection defined over k by using a probabilistic adaptation of the Bertini smoothness theorem to finite fields given by Poonen [93], or more directly a further adaptation by Bucur–Kedlaya [17] that directly address complete intersections rather than single hypersurface sections. The latter can be used to give a good bound on the genus of the curve X in terms of the dimension of A; see [13].

7. Inverse problems for zeta functions (October 21)

Note that for any fixed g and q, the Weil conjectures imply that there are only finitely many rational functions that can occur as the zeta function of a curve of genus g over $\mathbb{F}_q$, or of an abelian variety of dimension g over $\mathbb{F}_q$. (See supplemental exercises.)

In this lecture, we discuss the "inverse problems" of which zeta functions actually occur. One can give a relatively complete answer for abelian varieties; for curves this is much more subtle.

Readings 7.1. Our presentation of the Honda-Tate theorem follows [110]. For a discussion of numbers of points on curves over finite fields extending Remark 7.13, see [109].

Definition 7.2. To alleviate confusion, let us fix a bit of terminology here. Let A be an abelian variety over a finite field k, and let $F : A \to A$ be the Frobenius map. We refer to the characteristic polynomial $\det(T - F, T_\ell(A))$ of F (for any prime ℓ nonzero in k) as the *Weil polynomial* of A, and the reverse characteristic polynomial $\det(1 - FT, T_\ell(A))$ as the *L-polynomial* of A. The latter coincides with the factor $P_1(A, T)$ of the zeta function $Z(A, T)$.

Definition 7.3. An *isogeny* of abelian varieties A_1, A_2 over a field k is a finite k-linear morphism $f : A_1 \to A_2$ which is a homomorphism of group varieties. Any such morphism is surjective with finite (scheme-theoretic) kernel.

For any prime ℓ nonzero in k, the induced map $T_\ell(A_1) \to T_\ell(A_2)$ is itself an isogeny (it becomes an isomorphism after inverting ℓ). As we observed earlier, this implies that the characteristic polynomials of Frobenius of isogenous abelian varieties over a finite field coincide. Amazingly, this result has a converse; see Theorem 7.5 below.

Before stating the converse, we state (without further discussion) the key result that goes into its proof.

Theorem 7.4 (Tate)**.** *Let A_1, A_2 be abelian varieties over a finite field $k = \mathbb{F}_q$. Let ℓ be a prime nonzero in k. Then the map*

$$\mathbb{Z}_\ell \otimes_{\mathbb{Z}} \mathrm{Hom}(A_1, A_2) \to \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(A_1), T_\ell(A_2))$$

from Theorem 6.4 (which is injective without any condition on k) becomes a bijection if we restrict to G_k-equivariant maps on the right-hand side.

Proof. See [110, Theorem 6]. □

Theorem 7.5 (Tate)**.** *Let A_1, A_2 be abelian varieties over a finite field $k = \mathbb{F}_q$. Then the following conditions are equivalent.*

1. *A_1 and A_2 are isogenous.*
2. *The Weil polynomials of A_1, A_2 coincide.*
3. *The L-polynomials of A_1, A_2 coincide.*
4. *For each positive integer n, $\#A_1(\mathbb{F}_{q^n}) = \#A_2(\mathbb{F}_{q^n})$. (As noted earlier, this does not guarantee that the groups $A_1(\mathbb{F}_{q^n})$ and $A_2(\mathbb{F}_{q^n})$ are in fact isomorphic.*

Proof. The only issue is to prove that if the L-polynomials coincide, then A_1 and A_2 are isogenous. Recall that the action of Frobenius on $T_\ell(A_i) \otimes_{\mathbb{Z}} \mathbb{Q}$ is semisimple, and so is determined up to isomorphism by its characteristic polynomial. Consequently, $\mathrm{Hom}_{\mathbb{Z}_\ell[G_k]}(T_\ell(A_1), T_\ell(A_2))$ contains a map of full rank; by Tate's theorem, $\mathrm{Hom}(A_1, A_2)$ must do so also. (For example, if we start with an element of $\mathrm{Hom}_{\mathbb{Z}_\ell[G_k]}(T_\ell(A_1), T_\ell(A_2))$ of full rank, any sufficiently close ℓ-adic approximation contained in $\mathrm{Hom}(A_1, A_2)$ will have full rank over $\mathbb{Z}_\ell$, hence also over $\mathbb{Z}$.) □

Remark 7.6. When Theorem 7.5 implies the existence of an isogeny, it does not give much useful information about how to find the isogeny, or what its degree might be. Indeed, finding an explicit isogeny (or one of a specific form) is a sufficiently (apparently) hard computational problem that it has been proposed as the basis of cryptographic protocols; see for example [38].

Remark 7.7. It is not immediately obvious why Theorem 7.4 should depend on k being finite. Tate's proof ultimately comes down to the fact that there are only finitely many isomorphism classes of abelian varieties of a given dimension over a fixed finite field.

One can hope to prove a similar theorem for other base fields using more refined finiteness statements. Indeed, one such statement (which takes into account primes of bad reduction) was proved by Faltings, giving his *isogeny theorem* which extends Theorem 7.4 to the case where k is a number field (see Theorem 9.16). This in turn yields an analogue of Theorem 7.5 asserting that two abelian varieties over a number field are isogenous if and only if their L-functions coincide.

At this point, we know that abelian varieties over a finite field are characterized up to isogeny by their Weil polynomials or their L-polynomials. We next formulate the Honda-Tate theorem which pins down exactly which polynomials can occur. This theorem *almost* says that every polynomial consistent with the Weil conjectures occurs, but not quite: there are some multiplicity conditions that have to be enforced also.

Theorem 7.8 (Honda-Tate theorem, part 1)**.** *Fix a positive integer g and a prime power q. Then for every irreducible polynomial $P(T) \in \mathbb{Z}[T]$ such that:*

(i) $P(0) = 1$;
(ii) $\deg(P) = 2g$;
(iii) $P(\frac{1}{qT}) = q^{-g}T^{-2g}P(T)$;
(iv) *all roots of P in $\mathbb{C}$ have absolute value $q^{-1/2}$;*

there exist a positive integer e and a simple abelian variety A such that $P_1(A,T) = P(T)^e$.

The integer e is determined by the rational endomorphism algebra of A, which can itself be described explicitly in terms of P.

Theorem 7.9 (Honda-Tate theorem, part 2)**.** *With notation as in Theorem 7.8, $E := \mathrm{End}(A)_{\mathbb{Q}}$ is a central simple algebra of degree e over $\Phi = \mathbb{Q}(\pi_A)$, where π_A^{-1} is a root of P. The invariant of E at a place v of Φ is given by*

$$\begin{cases} \frac{1}{2} & \text{if } v \text{ is real} \\ [\Phi_v : \mathbb{Q}_p]\frac{\mathrm{ord}_v(\pi_A)}{\mathrm{ord}_v(q)} & \text{if } v \text{ lies over } p \\ 0 & \text{otherwise.} \end{cases}$$

Note that e is then the least common denominator of the local invariants.

Corollary 7.10. If the coefficient of T^g in P is nonzero modulo p, then $e = 1$. (This corresponds to the case of an *ordinary* abelian variety.)

Remark 7.11. We limit ourselves to a few words about the proof of Theorem 7.8. The construction starts by using analytic methods to construct a complex torus whose endomorphism ring contains $\mathbb{Z}[\pi_A]$. One then introduces a polarization to give this torus the structure of an abelian variety over $\mathbb{C}$. Since abelian varieties with complex multiplication occur as isolated points in moduli, they are forced to descend to some number field. Reducing modulo a suitable prime, we then get the desired abelian variety *except* that it might be defined not over $\mathbb{F}_q$ but over a finite extension. Finally, we use Tate's theorem to descend down to $\mathbb{F}_q$.

It is possible to supplant the use of complex analysis with more arithmetic methods; see [19]. However, the use of characteristic 0 methods to prove this statement, which is formulated exclusively in positive characteristic, remains unavoidable with current techniques.

Remark 7.12. Using the Honda-Tate theorem, one can tabulate isogeny classes of abelian varieties of dimension g over $\mathbb{F}_q$ for small values of g and q. This has been done in the LMFDB [80], using Sage to enumerate Weil polynomials. The theory behind this enumeration is described in [69] and [72].

Remark 7.13. The constraint imposed by the Honda-Tate theorem may be restricted to Jacobians, so it also implies a nontrivial restriction on the zeta function of a curve of genus g over $\mathbb{F}_q$. However, for dimensional reasons there are expected to be many fewer zeta functions of curves than zeta function of abelian varieties (once g is large enough), so it is natural to look for other constraints on zeta functions that are exclusive to curves.

One important set of constraints arises from positivity conditions. For any curve X over $\mathbb{F}_q$ (or indeed any algebraic variety at all), the following conditions obviously hold.

(1) We have $\#X(\mathbb{F}_q) \geq 0$.
(2) For all positive integers m and n, $\#X(\mathbb{F}_{q^{nm}}) \geq \#X(\mathbb{F}_{q^m})$.

These conditions impose certain "linear programming" constraints which have unexpectedly strong consequences. For example, suppose we want to know the maximum value of $\#X(\mathbb{F}_q)$ for X a curve of a given genus g. This question is more than theoretical in light of the *Goppa construction*: one can construct interesting error-correcting codes by fixing a rational point ∞ and a positive integer d, and considering the vectors

$$\{(f(x)_{x \in X(\mathbb{F}_q) \setminus \{\infty\}} : f \in K(X), \operatorname{div}(f) + d\infty \geq 0\}.$$

The corrective capacity of this code is limited by g (thanks to Riemann-Roch), so the quality of the code depends on $\#X(\mathbb{F}_q)$ being large relative to g.

Much is known about optimizing $\#X(\mathbb{F}_q)$ for fixed q, g; see `https://manypoints.org`. However, let us consider instead the asymptotic situation where q is fixed and $g \to \infty$. For fixed q, the Weil conjectures imply

$$\limsup_{g \to \infty} \frac{\#X(\mathbb{F}_q)}{g} \leq 2\sqrt{q}$$

but Ihara [60] discovered this is not best possible; Drinfeld-Vlăduț [36] improved the upper bound to $\sqrt{q} - 1$.

In the other direction, it is known that the bound $\sqrt{q} - 1$ is best possible when q is a square. For q not a square, it is known that

$$\limsup_{g \to \infty} \frac{\#X(\mathbb{F}_q)}{g} \geq c(q) > 0$$

but the optimal value is unknown.

We end with one concrete result with a curious history: the *class number 1 problem for function fields*.

Theorem 7.14. *There are exactly* 8 *isomorphism classes of curves of any genus g over any finite field $\mathbb{F}_q$ for which* $\operatorname{Jac}(X)(\mathbb{F}_q) = 1$. *The pairs (g, q) that occur are*

$$(1,2), (1,3), (1,4), (2,2), (2,2), (3,2), (3,2), (4,2).$$

Proof. It was originally "proved" by Leitzel–Madan–Queen [77] that there are only 7 such isomorphism classes, with the case $(g, q) = (4, 2)$ omitted; the list of these had previously been obtained by Madan–Queen [82]. Much later, it was discovered by Stirpe [102] that the case $(g, q) = (4, 2)$ actually does occur. The completeness of the list as given above is due independently to Mercuri–Stirpe [84] and Shen–Shi [99]. □

8. The Lang-Weil estimate (October 23)

In this lecture, we discuss the Lang-Weil theorem, which uses the Riemann hypothesis for curves to give a partial result towards the Weil conjectures for higher-dimensional varieties.

Readings 8.1. We follow the original presentation of Lang–Weil [75].

Our first statement of the Lang-Weil theorem is the following.

Theorem 8.2 (Lang-Weil 1)**.** *Let X be a scheme of finite type over $\mathbb{F}_q$ of dimension n. Let c be the number of irreducible components of X of dimension n which are also geometrically irreducible. Then we have the estimate*

$$\#X(\mathbb{F}_q) = c \cdot q^n + O(q^{n - \frac{1}{2}})$$

where the constant for the big-O notation depends only on the geometry of $X_{\overline{\mathbb{F}_q}}$.

Proof. We induction on the dimension of n, the case $n = 0$ being straightforward (and the source of the constant c). For $n > 0$, choose a projection map $X \to S$ from X onto a scheme S of dimension $n - 1$, such that X and S have the same value of c. We may then compute $\#X(\mathbb{F}_q)$ by summing over the fibers of the map over $\mathbb{F}_q$-points. By the induction hypothesis, the number of summands is $cq^{n-1} + Q(q^{n-3/2})$; by the Weil conjectures for curves, the number of points on each fiber is $q + O(q^{1/2})$ where the implied constant in the big-O notation can be bounded in terms of the geometry of the projection map. This yields the desired estimate. □

Here we illustrate by an example that c should count only geometrically irreducible components.

Example 8.3. Let $X := V(x^2+y^2) \subset \mathbb{A}^2_{\mathbb{F}_q}$, for a prime power q such that $q \equiv 3 \pmod 4$. By the quadratic residue criterion for -1,

$$\#X(\mathbb{F}_{q^k}) = \begin{cases} 1 & k \text{ odd} \\ 2q^k - 1 & k \text{ even.} \end{cases}$$

Note that $X_{\mathbb{F}_{q^k}}$ has only one irreducible component for odd k, but this component is not geometrically irreducible because it splits after a base change. Hence $c = 0$ for this case which coincides with the above calculation. If k is even, then $X_{\mathbb{F}_{q^k}}$ is a disjoint union of two geometrically irreducible points, hence $c = 2$ which also agrees with the calculation.

Here is a more precise version of the theorem, whose proof we omit.

Theorem 8.4 (Lang-Weil 2)**.** *Let X be a geometrically irreducible projective variety of dimension n admitting an embedding $X \hookrightarrow \mathbb{P}^m$ of degree d. Then we have the following inequality*

$$\left|\#X(\mathbb{F}_q) - q^n\right| \leq 2\binom{d-1}{2} q^{n-\frac{1}{2}} + A(n,m,d)q^{n-1}$$

where $A(n,m,d)$ is a constant dependent only on n, m, d.

Remark 8.5. The shape of the estimate in Lang-Weil is in some sense best possible: the exponent of q in the error term cannot be reduced. For example, if $X = C \times_{\mathbb{F}_q} \mathbb{A}^{n-1}$, then the error term for $\#X(\mathbb{F}_q)$ comes from the error term for $\#C(\mathbb{F}_q)$, and this can certainly be as large as $O(q^{1/2})$ (e.g., for elliptic curves).

On the other hand, under additional hypotheses one can hope for a better error estimate. For example, Let X be a smooth hypersurface in $\mathbb{P}^n$ over $\mathbb{F}_q$. Then using the Weil conjectures plus some additional knowledge (the *hard Lefschetz theorem* in Weil cohomology; see Theorem 9.14 for the étale version), one can show that

$$\#X(\mathbb{F}_q) = \#\,\mathbb{P}^{n-1}(\mathbb{F}_q) + O(q^{\frac{n-1}{2}})$$

where the first term is coming from the geometry of the ambient space and the second term is from the interesting middle cohomology $H^{n-1}(X)$.

As an intermediate case, consider what happens if we drop the smoothness hypothesis. Then the exponent in the error term is $O(q^{(n+d-2)/2})$ where d is the dimension of the singular locus X^{sing} (or -1 if $X^{\text{sing}} = \emptyset$).

All of these results can be made with completely explicit error terms. See [46].

9. Étale cohomology as a black box (October 28)

In this lecture, we give a "black box" description of étale cohomology for varieties over finite fields and number fields. This will be our first example of a Weil cohomology theory.

Readings 9.1. We make no attempt here to explain how étale cohomology is constructed. Introductory sources for that include Freitag–Kiehl [44], Milne [86], Tamme [104].

Definition 9.2. Let K be a number field. Let X be a smooth projective K-scheme. Then for any prime number ℓ, we get a collection of finite-dimensional $\mathbb{Q}_\ell$-vector spaces $H^i_{\text{et}}(X_{\overline{K}}, \mathbb{Q}_\ell)$, each with a continuous G_K-action. The G_K-action comes from the fact that we first base-extend X from K to $\overline{K}$ before taking cohomology. (Since K is of characteristic 0, there is no restriction on ℓ right now.)

Similarly, suppose that X is a smooth projective k-scheme where k is a finite field. Then for any prime number ℓ, we get a collection of finite-dimensional $\mathbb{Q}_\ell$-vector spaces $H^i_{\text{et}}(X_{\overline{k}}, \mathbb{Q}_\ell)$, each with a continuous G_k-action. However, the case where ℓ equals the characteristic of k is anomalous and is *not* considered a Weil cohomology theory. (We will describe a replacement later.)

Example 9.3. For $X = A$ an abelian variety over K, $H^1_{\text{et}}(X_{\overline{K}}, \mathbb{Q}_\ell) = V_\ell(A)^* = \text{Hom}_{\mathbb{Q}_\ell}(V_\ell(A), \mathbb{Q}_\ell(1))$ where $\mathbb{Q}_\ell(1) = V_\ell(\mathbb{G}_m)$. That is, $\mathbb{Q}_\ell(1)$ is a one-dimensional vector space over $\mathbb{Q}_\ell$ with the Galois action given by the ℓ-adic cyclotomic character. For $i > 1$, $H^i(X_{\overline{K}}, \mathbb{Q}_\ell) = \wedge^i_{\mathbb{Q}_\ell} H^1(X_{\overline{K}}, \mathbb{Q}_\ell)$.

Theorem 9.4 (Lefschetz trace formula)**.** *Let X be a smooth projective scheme over a finite field k. Then for any prime ℓ nonzero in k,*

$$\#X(\mathbb{F}_q) = \sum_{i=0}^{2\dim(X)} (-1)^i \operatorname{Trace}(F, H^i_{\text{et}}(X_{\overline{k}}, \mathbb{Q}_\ell)).$$

Consequently,

$$Z(X,T) = \prod_{i=0}^{2\dim(X)} \det(1 - FT, H^i_{\text{et}}(X_{\overline{k}}, \mathbb{Q}_\ell))^{(-1)^{1+i}}.$$

Let us inspect more closely the relationship between the constructions over K and over k.

Definition 9.5. Let $\mathfrak{p}$ be a maximal ideal of $\mathcal{O}_K$ with residue field $\kappa(\mathfrak{p})$. By choosing a place of $\overline{K}$ above $\mathfrak{p}$, we obtain an inclusion $\overline{K} \subset \overline{K}_\mathfrak{p}$ and a corresponding inclusion $G_{K_\mathfrak{p}} \subset G_K$.

Taking $G_{K_\mathfrak{p}}$ apart further, we have an exact sequence

$$1 \to I_\mathfrak{p} \to G_{K_\mathfrak{p}} \to G_{\kappa(\mathfrak{p})} \to 1,$$

where $I_\mathfrak{p}$ is the *inertia group*, and a tower of field extensions:

$$\begin{array}{c} \overline{K}_\mathfrak{p} \\ \Big| I_\mathfrak{p} \\ K_\mathfrak{p}^{\text{unr}} \\ \Big| G_{\kappa(\mathfrak{p})} \\ K_\mathfrak{p}. \end{array}$$

We say that a representation of $G_{K_\mathfrak{p}}$ is *unramified* if it restricts trivially to $I_\mathfrak{p}$, which is to say that it factors through $G_{\kappa(\mathfrak{p})}$. A representation of G_K is *unramified at* $\mathfrak{p}$ if its restriction to $G_{K_\mathfrak{p}}$ is unramified.

The following is a form of the *proper base change theorem*.

Proposition 9.6. Suppose that $\mathfrak{X}$ is a smooth projective scheme over the local ring $(\mathcal{O}_K)_\mathfrak{p}$. Then for any prime ℓ nonzero in $\kappa(\mathfrak{p})$, there is a natural isomorphism

$$H^i_{\text{et}}(\mathfrak{X}_{\overline{K}}, \mathbb{Q}_\ell)|_{G_{K_\mathfrak{p}}} \cong H^i_{\text{et}}(\mathfrak{X}_{\overline{\kappa(\mathfrak{p})}}, \mathbb{Q}_\ell)|_{G_{\kappa(\mathfrak{p})}}$$

which is equivariant for the actions of $G_{K_\mathfrak{p}}$ on both sides (the latter via $G_{K_\mathfrak{p}} \twoheadrightarrow G_{\kappa(\mathfrak{p})}$). In particular, the action of G_K on $H^i_{\text{et}}(\mathfrak{X}_{\overline{K}}, \mathbb{Q}_\ell)$ is unramified at $\mathfrak{p}$.

In other words, if X is a smooth projective K-scheme with good reduction at a prime ideal $\mathfrak{p}$, the action of G_K on its étale cohomology is unramified at $\mathfrak{p} \nmid \ell$.

Remark 9.7. In general, if X has *good reduction at* $\mathfrak{p}$, meaning that it extends in some way to a smooth proper scheme over $(\mathcal{O}_K)_\mathfrak{p}$, then this extension is not guaranteed to be unique up to isomorphism. It is unique if $\dim(X) = 1$, and in some isolated cases of higher dimension (e.g., when X is an abelian variety); but in general, when $\dim(X) \geq 2$ there can be multiple lifts which differ by a birational transformation. (Note that the lifts will have dimension ≥ 3, so constructions like *flips* become relevant.)

In particular, the reduction modulo $\mathfrak{p}$ is not uniquely determined by X. However, its zeta function is independent of choices by Proposition 9.9 below.

In the good reduction case, we may transfer the statement of the Lefschetz trace formula to an assertion about the G_K-action on étale cohomology.

Definition 9.8. We denote by $\text{Frob}_\mathfrak{p}$ any element of G_K which belongs to $G_{K_\mathfrak{p}}$ and projects to the inverse of the Frobenius element of $G_{\kappa(\mathfrak{p})}$. Such an element of G_K is called a *geometric Frobenius element* at $\mathfrak{p}$; the inverse of such an element is called an *arithmetic Frobenius element* at $\mathfrak{p}$.

Proposition 9.9. With notation as in Proposition 9.6,

$$Z(\mathfrak{X}_{\kappa(\mathfrak{p})}, T) = \prod_{i=0}^{2\dim(\mathfrak{X}_K)} \det(1 - \mathrm{Frob}_{\mathfrak{p}}\, T, H^i_{\mathrm{et}}(\mathfrak{X}_{\overline{K}}, \mathbb{Q}_\ell))^{(-1)^{1+i}}.$$

Remark 9.10. The case where $\mathfrak{p}$ divides ℓ behaves differently; in that case, even if X has good reduction at $\mathfrak{p}$, the action of G_K on ℓ-adic étale cohomology will not in general be unramified at $\mathfrak{p}$. For example, this is already true for $\mathbb{P}^1$ with $i = 2$.

Based on ideas of Tate, Fontaine managed to define a condition on p-adic Galois representations which is satisfied by étale cohomology in this context but "usually" fails for other representations; this is called the *crystalline* condition. Its study is part of the subject of *p-adic Hodge theory*, which we will not pursue here.

Remark 9.11. The case where the action of G_K is ramified at $\mathfrak{p}$, but $\mathfrak{p}$ does not divide ℓ, is also different. In this case, there is no well-defined action of $\mathrm{Frob}_{\mathfrak{p}}$ on all of $H^i_{\mathrm{et}}(\mathfrak{X}_{\overline{K}}, \mathbb{Q}_\ell)$, only on the subspace invariant under the action of $I_{\mathfrak{p}}$. The resulting action can be used to predict "missing" Euler factors needed to complete the L-function to achieve its functional equation (in addition to factors corresponding to infinite places, as in the case of Dedekind zeta functions).

However, these can be somewhat difficult to compute in practice. For elliptic curves, the relevant computation is Tate's algorithm, which is somewhat of a nuisance to implement from scratch (but is fortunately implemented in many existing software packages). The corresponding algorithm for curves of genus 2 has been worked out, but is even more complicated. Beyond that, few general results are known, except in a few special cases like cyclic covers of $\mathbb{P}^1$ [11] (but see [34] for some recent progress on general curves).

For empirical purposes (i.e., where a rigorous proof is not required), one can sometimes short-circuit this issue by guessing a value for the missing Euler factors and then verifying numerically (by contour integration) that the resulting L-function appears to satisfy the correct functional equation. A more extreme version of this is [41], in which L-functions are detected by guessing *all* of their Euler factors up to some bound.

We now ask whether Proposition 9.6 has a converse. That is, if X is a smooth projective K-scheme and the action of G_K on $H^i_{\mathrm{et}}(X_{\overline{K}}, \mathbb{Q}_\ell)$ is unramified at $\mathfrak{p} \nmid \ell$, does X have good reduction at $\mathfrak{p}$?

For abelian varieties, such a converse does hold.

Theorem 9.12 (Néron–Ogg–Shafarevich criterion)**.** *Let X be an abelian variety over K. For $\mathfrak{p}$ a prime ideal of $\mathcal{O}_K$ and ℓ a prime not equal to the characteristic of $\kappa(\mathfrak{p})$, X has good reduction at $\mathfrak{p}$ if and only if the action of G_K on $H^i_{\mathrm{et}}(X_{\overline{K}}, \mathbb{Q}_\ell)$ is unramified at $\mathfrak{p}$. (Note that on account of Example 9.3, it suffices to check the case $i = 1$.)*

Proof. See Serre-Tate [98]. □

Remark 9.13. For general X, there is no converse of Proposition 9.6. For example, if X is a curve of genus 2, then $H^i_{\mathrm{et}}(X_{\overline{K}}, \mathbb{Q}_\ell)$ is unramified at $\mathfrak{p}$ if and only if the same is true with X replaced by its Jacobian (as this does not change H^1_{et}). However, it is possible for the Jacobian of X to degenerate to a product of two elliptic curves (with the product polarization), which obstructs X itself from having good reduction (as otherwise taking the reduction would commute with taking the Jacobian, whereas the product of two elliptic curves cannot be a Jacobian).

One way to fix this is to consider "nonabelian étale cohomology". Roughly speaking, this means that étale cohomology arises from some sort of abelianization process on homotopy groups, which we replace with a less drastic quotienting operation.

We mention another key property of étale cohomology which was mentioned previously in our analysis of the Lang-Weil theorem (see Remark 8.5): the *hard Lefschetz theorem*.

Theorem 9.14 (hard Lefschetz theorem)**.** *Let X be a smooth projective scheme over K of dimension n. Let Y/K be a smooth ample hypersurface in X. Then the functoriality map*

$$H^i_{\mathrm{et}}(X_{\overline{K}}, \mathbb{Q}_\ell) \longrightarrow H^i_{\mathrm{et}}(Y_{\overline{K}}, \mathbb{Q}_\ell)$$

(1) *is G_K-equivariant,*
(2) *is an isomorphism for $i < n - 1$,*

(3) *and is injective for* $i = n - 1$.

Example 9.15. For $X = \mathbb{P}^n$,

$$H^i_{\text{et}}(X_{\overline{K}}, \mathbb{Q}_\ell) = \begin{cases} 0 & i \text{ odd} \\ \mathbb{Q}_\ell(i/2) & i \text{ even} \end{cases}$$

For Y a smooth ample hypersurface in X, using hard Lefschetz (and Poincaré duality) one sees that $H^i_{\text{et}}(Y_{\overline{K}}, \mathbb{Q}_\ell) \cong H^i_{\text{et}}(X_{\overline{K}}, \mathbb{Q}_\ell)$ in all degrees *except* $i = n - 1$ (the middle cohomology degree for Y).

As discussed earlier (Remark 7.7), Tate's theorem on isogenies of abelian varieties over finite fields (Theorem 7.4) extends to number fields. However, the proof of this theorem is outside the scope even of our suggested readings; see instead [22].

Theorem 9.16 (Faltings isogeny theorem)**.** *Let* A_1, A_2 *be abelian varieties over a number field* K*. For any prime* ℓ*,*

$$\text{Hom}(A_1, A_2) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \cong \text{Hom}_{\mathbb{Z}_\ell[G_K]}(T_\ell(A_1), T_\ell(A_2)).$$

As for Tate's theorem, this has the following corollary.

Corollary 9.17. Let A_1, A_2 be abelian varieties over a number field K. Then A_1 and A_2 are isogenous if and only if $V_\ell(A_1) \cong V_\ell(A_2)$ as $\mathbb{Q}_\ell[G_K]$-modules.

To make this look more like Tate's corollary, we must do a bit more group theory.

Proposition 9.18. Fix a finite set S of primes of the number field K, and let $G_{K,S}$ be the Galois group of the maximal algebraic extension of K unramified away from S. Let

$$\rho_1 : G_{K,S} \longrightarrow \text{GL}_n(\mathbb{Q}_\ell), \qquad \rho_2 : G_{K,S} \longrightarrow \text{GL}_n(\mathbb{Q}_\ell)$$

be two continuous representations with the property that

$$\text{Trace}(\text{Frob}_\mathfrak{p}, \mathbb{Q}_\ell^n) = \text{Trace}(\text{Frob}_\mathfrak{p}, \mathbb{Q}_\ell^n),$$

for every $\mathfrak{p} \notin S$. Then ρ_1 and ρ_2 are isomorphic up to semisimplification (that is, their irreducible constituents can be matched up).

Proof. By the Chebotarëv density theorem, the elements $\{\text{Frob}_\mathfrak{p}\}$ are dense in G_K (see Set 4 exercises); so the equality of traces holds identically on $G_{K,S}$. By the Brauer-Nesbitt theorem, two representations (on vector spaces over a field of characteristic 0) whose traces agree everywhere must have the same simplification. □

Corollary 9.19. Let A_1, A_2 be abelian varieties over a number field K. Then A_1 and A_2 are isogenous if and only if $\text{Trace}(\text{Frob}_\mathfrak{p}, T_\ell(A_1)) = \text{Trace}(\text{Frob}_\mathfrak{p}, T_\ell(A_2))$ for all but finitely many prime ideals $\mathfrak{p}$ of $\mathcal{O}_K$.

This sort of condition does not appears to be finitely verifiable, but in fact it is. We will discuss this point in the next lecture.

10. Comparing Galois representations: the Faltings–Serre method (October 30)

In the last lecture, we formulated Proposition 9.18, which asserts that two continuous ℓ-adic representations of $G_{K,S}$ (for K a number field and S a finite set of primes) are equal up to semisimplification if and only if their Frobenius traces coincide for all but finitely many primes. In this lecture, we discuss the *Faltings–Serre method* which makes it psosible to verify this sort of equality by a *finite* computation. This turns out to have numerous applications, which do not depend on computing with étale cohomology in any direct way.

Readings 10.1. The Faltings–Serre method has been codified for GL_2 by Livné [79]. See [15] for an extension to larger groups plus some practical improvements.

Before proceeding, let us discuss an example where such a comparison of representations comes up naturally: the modularity of elliptic curves, which had been conjectured in various forms by Taniyama, Shimura, and Weil.

Theorem 10.2 (Modularity of rational elliptic curves)**.** *Let E be an elliptic curve over $\mathbb{Q}$. Then there exists a modular form f (more precisely, a cuspidal weight 2 newform for the group $\Gamma_0(N)$ where N is the conductor of E) such that for every prime p at which E has good reduction (i.e., every prime not dividing N), the trace of Frobenius on $E/\mathbb{F}_p$ equals the Fourier coefficient $a_p(f)$.*

Proof. In the case where E is semistable (i.e., the conductor N is squarefree), this is part of the work of Wiles [116] and Taylor–Wiles [107] that completed the proof of Fermat's last theorem. The general case was resolved (based on the aforementioned papers and several intermediate results) by Breuil–Conrad–Diamond–Taylor [12]. □

Remark 10.3. Before Theorem 10.2 was proved in the 1990s, it had been verified in numerous examples, most conclusively by Cremona [26]. This relied on the older *Eichler-Shimura theorem*, which implies that for any f as in the theorem, there exists an elliptic curve E_f for which the desired conclusion holds.

Now suppose in that context, one had in mind a particular elliptic curve E for which one wanted to confirm the statement of the theorem. Since the conjecture includes a prediction for the level of the newform in terms of E, and the newforms for a given level form a computable (via Manin's method of modular symbols) finite-dimensional vector space, one could then do the finite computation to find all candidates for f, and quickly isolate a unique candidate with the first few Fourier coefficients correct.

However, this would not suffice to *prove* that E and E_f are isogenous: applying Corollary 9.19 would require an infinite number of equalities, which cannot *a priori* be established via a finite computation. Instead, one applies some form of the Faltings–Serre method as described below to conclude.

Remark 10.4. It should be emphasized that the Faltings-Serre method *cannot* be used to prove that an "abstract" infinite list of numbers matches the list of Frobenius traces of a given Galois representation; one must know that the first list itself comes from a Galois representation with some control on the ramified primes. Besides the control on ramification, we need no other information other than the ability to compute entries of both lists on demand.

For example, in the case of elliptic curves, it would not have been possible to rigorously establish modularity of a given elliptic curve E without knowledge of the existence of the other elliptic curve E_f produced from f via Eichler-Shimura. However, it is not necessary to know anything more about E_f beyond its existence, its ramified primes (or even just an upper bound on this set), and its Frobenius traces.

Remark 10.5. Let us now break down how to reduce a comparison of infinitely many Frobenius traces, as in Proposition 9.18, to a finite computation. This splits into two main steps.

(1) This is the hard step. We know that these representations can be factored through $\mathrm{GL}_n(\mathbb{Z}_\ell)$ (they are representations of compact groups into Hausdorff targets, so they have closed image), so we would like to see these two representations with images in $\mathrm{GL}_n(\mathbb{F}_\ell)$ have the same semisimplification (and in particular, the Frobenius traces are pairwise congruent modulo ℓ). To check that, we try to identify the two kernels and comparing them; this involves enumerating number fields with prescribed ramification.

(2) Use a group-theoretic argument to promote the mod-ℓ equality to an ℓ-adic equality. This turns out to be much easier since it doesn't depend on the set S; it is a matter of pure group theory.

The difficulty of both steps scales badly with ℓ, so generally one takes $\ell = 2$ in practice. (Fortunately, there is no requirement that the representations in question have good reduction at primes above 2.)

To settle the first step, we use the following basic theorem of algebraic number theory.

Theorem 10.6 (Hermite-Minkowski)**.** *Given a number field K, an integer $d \in \mathbb{Z}^+$, and a finite set of finite primes S of $\mathcal{O}_K$, there are only finitely many isomorphism classes of number fields L/K which are unramified away from S and satisfy $[L : K] = d$. Moreover, this list is effectively computable.*

Proof. One may reduce to the case $K = \mathbb{Q}$. In this case, one can bound the contribution of each prime in S to the discriminant to get a bound on the absolute discriminant of L, and then use a geometry of numbers argument (Minkowski's theorem) to limit the possibilities for L to a finite set. □

In particular, there are finitely many homomorphisms $G_{K,S} \to \mathrm{GL}_n(\mathbb{F}_\ell)$ and (in principle!) one can compute them all, and then group them according to semisimplification. By Chebotarëv density, it takes only finitely many Frobenius traces to rule out all but the right candidate.

To achieve the second step (for $\ell = 2$), we use the following theorem.

Definition 10.7. A subset T of a vector space V over a field (here $\mathbb{F}_2$) is *noncubic* if every homogeneous polynomial of degree 3 which vanishes on T also vanishes on all of V.

Theorem 10.8 (Livné, after Serre)**.** *With notation as in Proposition 9.18, if the Frobenius traces of ρ_1 and ρ_2 agree modulo 2 for all $h \in G_K$ (as evidenced by the first step) and agree "on the nose" for some finite set $T \subseteq G_K$ whose image in $G_{K,S}^{\mathrm{ab}}/2G_{K,S}^{\mathrm{ab}}$ is noncubic, then the traces agree for all $h \in G_K$.*

Remark 10.9. Note that if $G_{K,S}^{\mathrm{ab}}/2G_{K,S}^{\mathrm{ab}}$ has dimension ≤ 2 over $\mathbb{F}_2$, then it is impossible for T to be noncubic. This is easily fixed by adding extra primes to S.

Remark 10.10. The main practical difficulty in applying this method in practice is to make the enumeration of homomorphisms $G_{K,S} \to \mathrm{GL}_n(\mathbb{F}_\ell)$ efficient. There is a lot of work to be done in this direction; see for example [15].

Remark 10.11. To conclude this lecture, we justify our claim that the Faltings-Serre method is useful in practice by citing a few additional examples of its use in the literature.

- There is a conjectural analogue of the modularity theorem for elliptic curves over imaginary quadratic fields, in which the role of classical modular forms is played by Bianchi forms. However, one must consider not only elliptic curves, but also abelian surfaces with quaternionic multiplication (QM); instances of modularity in this context have been exhibited by Dieulefait–Guerberoff–Pacetti [33] for elliptic curves and Schembri [95] for QM abelian surfaces.
- Inspired by some examples arising in *mirror symmetry* in mathematical physics, numerous authors have identified examples of Calabi-Yau threefolds whose Frobenius traces can be computed in terms of modular forms. See [117] for a survey.
- There is a conjectural analogue of the modularity theorem for elliptic curves known as the *paramodularity conjecture* of Brumer–Kramer [14]. This has been verified in a small number of cases [15].

11. Dwork's proof of rationality (November 4)

In this lecture, we discuss Dwork's proof of rationality of the zeta function for varieties over finite fields. Dwork's proof does not itself use a finite-dimensional cohomology theory on X, but it did inspire the construction of p-adic Weil cohomology which we will see later.

Readings 11.1. The original paper of Dwork is [37]. An updated presentation of the proof was given by Koblitz [73].

Theorem 11.2 (Dwork, 1958)**.** *For any scheme X of finite type over $\mathbb{F}_q$, $Z(X,T)$ represents a rational function of T.*

The proof involves three key components.

- Some initial reduction steps to put the problem in a more convenient form.
- An extension of a theorem of Borel on power series.
- Use of p-adic analysis to check the hypothesis of Borel's theorem.

We start with the reduction steps. Recall that if X splits as a disjoint union of an open subscheme U and a closed subscheme S, then

$$Z(X,T) = Z(U,T)Z(S,T).$$

Similarly, if X is a union of two closed subschemes X_1 and X_2, then

$$Z(X,T) = \frac{Z(X_1,T)Z(X_2,T)}{Z(X_1 \cap X_2, T)}.$$

Using this logic (and induction on dimension), we may reduce to the case where X is affine and irreducible; in fact, we can assume that X is contained not just in an affine space $\mathbb{A}^n$, but in a torus $\mathbb{G}_m^n$.

We can even take this a bit further. Write X as the subscheme of $\mathbb{G}_m^n$ cut out by some Laurent polynomials $(P_1, \dots, P_n)$. Suppose that we know the rationality for the hypersurface cut out by any single (but not necessarily irreducible) Laurent polynomial; we may then deduce the same conclusion for an intersection of

k such hypersurfaces by induction on k. That is, we may assume from now on that X is the zero locus of a (not necessarily irreducible) Laurent polynomial $P \in k[x_1^{\pm}, \dots, x_n^{\pm}]$ in a torus $\mathbb{G}_m^n$.

We next turn to the Borel-Dwork theorem. It is motivated by a simple observation.

Lemma 11.3. Let $f(T) = \sum_{n=0}^{\infty} a_n T^n \in \mathbb{Z}[\![T]\!]$ be a power series which over $\mathbb{C}$ has radius of convergence strictly greater than 1. Then $f(T) \in \mathbb{Z}[T]$.

Proof. The root test implies $\limsup_{n\to\infty} a_n^{1/n} < 1$. The only integer with absolute value less than 1 is zero, giving the claim. □

In the setting of zeta functions we do not expect polynomials, and we don't have much control over any archimedean valuations, although we can at least prove that $Z(X,T)$ has *some* positive radius of convergence.

Lemma 11.4. As a power series over $\mathbb{C}$, $Z(X,T)$ has radius of convergence at least $q^{-\dim(X)}$.

Proof. Since we are assuming X is a toric hypersurface, it admits a finite morphism $f : X \to \mathbb{G}_m^d$ where $d = \dim(X)$. Then

$$\#X(\mathbb{F}_{q^n}) \leq \deg(f)\#(q^n - 1)^d,$$

so the claim follows by an elementary calculation using the expression

$$Z(X,T) = \exp\left(\sum \#X(\mathbb{F}_{q^n})\frac{T^n}{n}\right).$$

□

To finesse this issue, we bring in the other places of $\mathbb{Q}$. The statement

$$\{x \in \mathbb{Z} : |x| < 1\} = \{0\}$$

has an analogue over $\mathbb{Q}$ in the form of the *product formula* for the valuations of $\mathbb{Q}$:

$$\left\{x \in \mathbb{Q} : \left(\prod_{v \text{ a valuation of } \mathbb{Q}} |x|_v\right) \neq 1\right\} = \{0\}.$$

This forms the basis of Dwork's extension of Borel's theorem.

Theorem 11.5 (Borel, 1894, extended by Dwork in 58)**.** *Suppose $f(T) \in \mathbb{Z}[\![T]\!]$ has radius of convergence over $\mathbb{C}$ at least R and is meromorphic on $\mathbb{Q}_p$ for $|T| < r$ (that is, it is the ratio of two power series with radius of convergence at least r, as measured by the root test). If $R > r^{-1}$, then $f(T)$ represents a rational function of T. Additionally, if $f(T)$ itself has radius of convergence over $\mathbb{Q}_p$ at least r, $f(T)$ is a polynomial.*

Proof. We give only an outline of the proof here; the details are filled in one of the Set 4 exercises. suppose first that $f(T)$ has radius of convergence over $\mathbb{Q}_p$ at least r. By the root test, we have

$$|a_n|_\infty < C_\epsilon(R-\epsilon)^{-n}, |a_n|_p < C_\delta(r-\delta)^{-n}$$

and, since $f(T) \in \mathbb{Z}[\![T]\!]$, $|a_n|_l \leq 1$ for any other valuation. In particular, for $n \gg 0$,

$$\prod_{v \text{ a valuation of } \mathbb{Q}} |x|_v < \frac{C_\epsilon C_\delta}{(R-\epsilon)^n(r-\delta)^n} < 1.$$

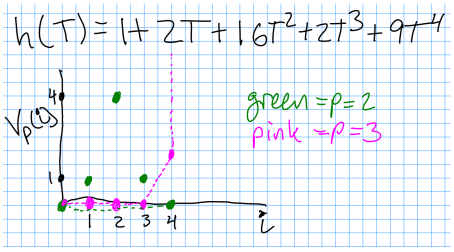
In particular, $a_n = 0$ for n sufficiently large, as desired.

In the general setting, write

$$f(T) = \frac{g(T)}{h(T)}$$

with $g(T), h(T) \in \mathbb{Q}_p[\![T]\!]$ having radius of convergence at least r. If $h(T)$ is a polynomial, we can clear denominators and argue that the result is a polynomial. Otherwise, $h(T)$ can have only finitely many zeroes in $\overline{\mathbb{Q}_p}$ with absolute value less than $r-\delta$ for any $\delta > 0$. Thus on $|T| < r - \delta$ we can write $h(T) = p_\delta(T)u_\delta(T)$, with $p_\delta(T) \in \mathbb{Q}_p[T]$ and $u_\delta(T)$ a unit in $\mathbb{Q}_p[\![T]\!]$. The idea then is to strip off $p_\delta(T)$ and apply the previous observation to $g(T)/u_\delta(T)$, but in a way that is uniform as δ varies. See the Set 4 exercises for the remaining details. □

Remark 11.6. As an aside, the equality $h(T) = p_\delta(T)u_\delta(T)$ in the above proof is governed by the *Newton polygon* of h. Suppose $h(T) = \sum a_n T^n$ is a polynomial with $a_0 = 1$; then the Newton polygon of h is the lower convex hull of the set $\{(i, v_p(a_i)\}$. See the following diagram for an example.



The main theorem is that if the Newton polygon has a segment of width w and slope s, then the original polynomial has exactly w roots of p-adic valuation $-s$.

At this point, Dwork proceeds by emulating Weil's analysis of Fermat hypersurfaces. Recall that we are assuming that $X = \operatorname{Spec} k[x_1^{\pm}, \dots, x_n^{\pm}]/(P)$ is a toric hypersurface. Write $q = p^a$ and let Θ be a nontrivial additive character of $\mathbb{F}_p$ (we do not yet specify where Θ is valued), so that for any positive integer s,

$$\Theta_{as}(x) = \Theta(x^{1+p+\cdots+p^{as-1}})$$

is a nontrivial additive character of $\mathbb{F}_{q^s}$. Then

$$\sum_{x_0 \in \mathbb{F}_{q^s}} \Theta_{as}(x_0 y) = \begin{cases} q^s & y = 0 \\ 0 & y \neq 0; \end{cases}$$

consequently,

$$q^s \# X(\mathbb{F}_{q^s}) = (q^s - 1)^n + \sum_{x_0, \dots, x_n \in \mathbb{F}_{q^s}^\times} \Theta_{as}(x_0 P(x_1, \dots, x_n)).$$

If we expand $x_0 P$ as a sum $\sum_j \alpha_j \mu_j(x)$ with $\alpha_j \in \mathbb{F}_q^\times$ and $\mu_j(x)$ a monomial in $x_0, \dots, x_n$, we also have

$$q^s \# X(\mathbb{F}_{q^s}) = (q^s - 1)^n + \sum_{x_0, \dots, x_n \in \mathbb{F}_{q^s}^\times} \prod_j \Theta_{as}(\alpha_j \mu_j(x)).$$

So far we have done nothing p-adic. We now make the key advance, expressing the character Θ in p-adic terms.

Definition 11.7. Let $\mathbb{Z}_q$ be the finite étale extension of $\mathbb{Z}_p$ with residue field $\mathbb{F}_q$. For $x \in \mathbb{F}_q^\times$, there is a unique $(q-1)$-st root of unity in $\mathbb{Z}_q$ congruent to x modulo p. We denote it by $[x]$ and call it the *Teichmüller lift* of x. We also write $[0] = 0$.

Form the product

$$(1+Y)^X (1+Y^p)^{(X^p - X)/p} (1+Y^{p^2})^{(X^{p^2} - X^p)/p^2} \cdots \in \mathbb{Q}[\![X, Y]\!]$$

using binomial expansions, and label its coefficients as

$$\sum_{n=0}^{\infty} \sum_{m=n}^{\infty} a_{m,n} X^n Y^m.$$

Then pick a primitive p-th root of unity ζ_p in $\overline{\mathbb{Q}}_p$, put $\lambda = \zeta_p - 1$, and define

$$\Theta(T) = \sum_{n=0}^{\infty} a_n T^n, \qquad = \sum_{m=n}^{\infty} a_{m,n} \lambda^m.$$

As in [73, Chapter 4], one verifies that $x \mapsto \Theta([x])$ is a nontrivial additive character of $\mathbb{F}_p$, so we may use it in place of $\Theta(x)$ in the previous calculations.

Remark 11.8. The power series $\Theta(T)$ is not uniquely determined by the fact that it gives rise to an additive character of $\mathbb{F}_p$ in this fashion; indeed, Dwork used a slightly different construction (see Set 4 exercises), but this discrepancy has no significant effect on the resulting argument.

Now define the power series

$$G(X_0, \dots, X_n) := \prod_j \Theta([\alpha_j]\mu_j(X))\Theta([\alpha_j]^p\mu_j(X)^p)\cdots\Theta([\alpha_j]^{p^{a-1}}\mu_j(X)^{p^{a-1}}),$$

so that

$$q^s \# X(\mathbb{F}_{q^s}) = (q^s-1)^n + \sum_{x_0,\dots,x_n \in \mathbb{F}_{q^s}^\times} \prod_j G([x])G([x^q])\cdots G([x^{q^{s-1}}]).$$

The sum can be written as the trace (for a suitable topology) of the operator $f \mapsto T(Gf)$ on $\mathbb{Q}_p[\![X_0, \dots, X_n]\!]$, where T is the "decimation" map

$$\sum a_{i_0,\dots,i_n} X_0^{i_0}\cdots X_0^{i_n} \mapsto \sum a_{pi_0,\dots,pi_n} X_0^{i_0}\cdots X_0^{i_n}$$

(see [73, §V.3, Lemma 3]). Since this operator does not act on a finite-dimensional space, it does not immediately give rationality of the zeta function; however, it does give p-adic meromorphicity, which is what we need to plug into Borel's theorem and complete the proof.

Remark 11.9. Notice that the key move here was indeed to use a trace formula, but on an infinite-dimensional vector space. This construction can be promoted to give what is sometimes called *Dwork cohomology*, which we do not treat as a Weil cohomology per se but is nonetheless extremely useful in the study of zeta functions.

Remark 11.10. If one were to specialize this argument back to Fermat hypersurfaces, it would yield an explicit p-adic analytic formula for Gauss sums. This was somehow missed by Dwork and his contemporaries, only to be appear later as the *Gross-Koblitz formula* [53].

12. Algebraic de Rham cohomology (November 18)

In this lecture, we lay the groundwork for the introduction of a p-adic Weil cohomology theory, by describing algebraic de Rham cohomology.

Readings 12.1. In preparation for the next lecture, we follow [68].

Definition 12.2. Let K be a field of characteristic 0. Let R be a K-algebra. Let us denote by $\Omega_{R/K}$ the module of Kähler differentials, i.e.

$$\Omega_{R/K} = \frac{\text{free module on } dr\,(r \in R)}{\langle dr\,(r \in K), d(r+s) - dr - ds, d(rs) - s\,dr - r\,ds\rangle}.$$

This admits a K-linear derivation $d : R \to \Omega_{R/K}$ given by $r \mapsto dr$ and is universal in the sense that if M is an R-module and $\delta : R \to M$ is a K-linear derivation, then δ factors uniquely as

$$R \to \Omega_{R/K} \to M$$

where the map on the right is R-linear and the map on the left is d.

Similarly, we define $\Omega_{X/K}$ when X is a K-scheme.

Example 12.3. If $R = K[t_1, ..., t_n]$, then $\Omega_{R/K} = R\,dt_1 \oplus \cdots \oplus R\,dt_n$ with the universal derivation d taking f to $\frac{\partial f}{\partial t_1}dt_1 + \cdots + \frac{\partial f}{\partial t_n}dt_n$.

Remark 12.4. By the previous example, If R is a finite type K-algebra, then $\Omega_{R/K}$ is a finite R-module. When R is a smooth K-algebra of dimension n, then $\Omega_{R/K}$ is finite and locally free of rank n. The converse is also true by the Jacobian criterion. Similarly, if X is a scheme of finite type over K, $\Omega_{X/K}$ is coherent; if X is smooth over K, then $\Omega_{X/K}$ is locally free.

Definition 12.5. For $i \geq 1$, we define $\Omega^i_{R/K} = \wedge^i_R \Omega_{R/K}$. That is, $\Omega^i_{R/K}$ is the free R-module on symbols $\omega_1 \wedge \cdots \wedge \omega_i$ with $\omega_j \in \Omega_{R/K}$, modulo relations of the form

$$(r\omega_1 + r'\omega_1') \wedge \cdots \wedge \omega_i - r(\omega_1 \wedge \cdots \wedge \omega_i) - r'(\omega_1' \wedge \cdots \wedge \omega_i)$$

(and similarly for each position) and $\omega_1 \wedge \cdots \wedge \omega_i$ whenever two of the ω_j are equal (alternating condition).

The map d induces a map $d^i : \Omega^i_{R/K} \to \Omega^{i+1}_{R/K}$ by

$$r\omega_1 \wedge \cdots \wedge \omega_i \mapsto dr \wedge \omega_1 \wedge \cdots \wedge \omega_i.$$

One checks that $d^{i+1} \circ d^i = 0$, so the $\Omega^i_{R/K}$ form a K-linear complex, called the *de Rham complex* of R/K.

Similarly, for X a scheme over K, we get a de Rham complex $\Omega^\bullet_{X/K}$.

Example 12.6. For $R = K[t_1, \ldots, t_n]$, $\Omega^\bullet_{R/K}$ has cohomology

$$h^0(\Omega^\bullet_{R/K}) = K \text{ and } h^i(\Omega^\bullet_{R/K}) = 0 \text{ for } i > 0.$$

Definition 12.7. How do we make sense of the "cohomology" of $\Omega^\bullet_{X/K}$? The correct notion is that of *hypercohomology*. In fancy terms, this means viewing the complex as an object in the (bounded) derived category of quasicoherent sheaves on X, then taking the derived global sections functor. In concrete terms, it is computed as follows.

For simlicity, let us assume that X is separated. Let $\{U_i\}$ be a cover of X by open affines; our condition that X is separated means that any intersection among the U_i is again affine. Define the double complex

$$D^{j,k} = \bigoplus \Gamma(U_{i_0} \cap \cdots \cap U_{i_j}, \Omega^k_{X/K})$$

with the j-differentials being the Čech differentials and the k-differentials being the de Rham differentials. Then form the associated *total complex*, whose i-th term is $\bigoplus_{j+k=i} D^{j,k}$ (with appropriate signs on the differentials to make this a complex), and take the cohomology to obtain $\mathbb{H}^i(X, \Omega^\bullet_{X/K})$.

Example 12.8. Take $X = \mathbb{P}^1$ and consider the covering by two copies of $\mathbb{A}^1$. The double complex in this case is

$$\begin{array}{ccccccc}
 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \\
0 \to & & K[x] \oplus K[x^{-1}] & \xrightarrow{f} & K[x, x^{-1}] & \to 0 & \\
 & & \downarrow{\scriptstyle f'} & & \downarrow{\scriptstyle g'} & & \\
0 \to & & K[x]dx \oplus K[x^{-1}]\,dx^{-1} & \xrightarrow{g} & K[x, x^{-1}]\,dx & \to 0 & \\
 & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & &
\end{array}$$

Here f is surjective with kernel K; f' is surjective with kernel $K \oplus K$; g is injective and the cokernel is generated by $x^{-1}\,dx$; and $\ker(g') = K$ and $\operatorname{coker}(g') = Kx^{-1}\,dx$. Keeping in mind that $dx^{-1} = -x^{-2}dx$, we find that

$$\mathbb{H}^0(X, \Omega^\bullet_{X/K}) = K; \;\; \mathbb{H}^1(X, \Omega^\bullet_{X/K}) = 0; \;\; \mathbb{H}^2(X, \Omega^\bullet_{X/K}) = K.$$

Theorem 12.9 (Grothendieck)**.** *Suppose X is a smooth projective (or proper) variety over $\mathbb{C}$. Then there is a natural isomorphism $\mathbb{H}^i(X, \Omega^\bullet_{X/\mathbb{C}}) \to H^i(X^{\mathrm{an}}, \mathbb{C})$ where X^{an} denotes the associated complex analytic variety (analytification of X).*

Proof. This follows by combining the following statements.

- By Serre's GAGA theorem [96], $\mathbb{H}^i(X, \Omega^\bullet_{X/\mathbb{C}}) \cong \mathbb{H}^i(X^{\mathrm{an}}, \Omega^\bullet_{X^{\mathrm{an}}/\mathbb{C}})$.
- By Dolbeaut's theorem [50, §0.3], $H^i(X^{\mathrm{an}}, \Omega^\bullet_{X^{\mathrm{an}}/\mathbb{C}}) \cong H^i(X^{C^\infty}, \Omega^\bullet_{X^{C^\infty}})$.
- By de Rham's theorem, $\mathbb{H}^i(X^{C^\infty}, \Omega^\bullet_{X^{C^\infty}}) \cong H^i(X^{\mathrm{an}}, \mathbb{C})$.

□

Remark 12.10. We can also define $\Omega^\bullet_{X/k}$ and $\mathbb{H}^\bullet(X, \Omega^\bullet_{X/k})$ when k is of characteristic p, but this can behave in unexpected ways. For example, if X/k is smooth and proper, $\mathbb{H}^\bullet(X, \Omega^\bullet_{X/k})$ is finite dimensional over k (because it can be computed in terms of the coherent cohomology groups of the individual terms of the complex), but can be of the "wrong dimension." Such phenomena can mostly be explained in terms of failure of degeneration of the Hodge-de Rham spectral sequence.

13. Monsky-Washnitzer cohomology (November 19)

In this lecture, we explain how to adapt algebraic de Rham cohomology to obtain the *Monsky-Washnitzer cohomology* of a smooth affine variety over a finite field. The construction globalizes naturally to smooth nonaffine varieties; the generalization to nonsmooth varieties is more difficult, and is part of Berthelot's theory of *rigid cohomology* which we do not discuss in detail here.

Readings 13.1. We continue to follow [68]. The original development of Monsky-Washnitzer cohomology is [90, 88, 89]. For Berthelot's rigid cohomology, start with [78].

Definition 13.2. Throughout this lecture, let k be a finite field. Let $W(k)$ be the the ring of Witt vectors of k; all you need to know about this ring is that it is a finite étale algebra over $\mathbb{Z}_p$ with $W(k)/pW(k) \cong k$. Let K be the fraction field of $W(k)$, which is obtained by inverting p.

Let $X = \mathrm{Spec}(\overline{A})$ be a smooth affine scheme over k. Since a Weil cohomology theory has to have coefficient field of characteristic 0 (otherwise it cannot completely control the zeta function), if we want to use differential forms it will have to be over some characteristic-0 lift of k.

Theorem 13.3 (Elkik, Arabia)**.** *There exists a smooth affine scheme $\mathfrak{X} = \mathrm{Spec}(A)$ over $W(k)$ with $\mathfrak{X} \times_{W(k)} k \cong X$.*

Proof. A crude summary of Elkik's proof is the following. Since X is smooth, there is no *local* obstruction to finding at least a formal lift (i.e., a lift to the formal scheme $\mathrm{Spf}\, W(k)$). Since X is affine, there is also no *global* obstruction. To complete the proof, one must show that the existence of a formal lift implies the existence of an algebraic lift. A streamlined presentation and generalization of Elkik's result has been given by Arabia [4]. □

Our first candidate for the cohomology of X is the de Rham cohomology of the generic fiber of a lift:

$$H^i(\mathfrak{X}_{W(k)[1/p]}, \Omega^\bullet).$$

Unfortunately, this is not independent of the choice of the lift $\mathfrak{X}$, so we need to try something else.

Our second attempt involves replacing $\mathfrak{X}$ with its p-adic completion.

Definition 13.4. Let $\hat{A}$ be the p-adic completion of A. We define the *module of continuous Kähler differentials* of $\hat{A}$ to be

$$\Omega_{\hat{A}[\frac{1}{p}]/K} := \varprojlim \Omega_{(A/p^n)/(W(k)/p^n)} \otimes_{W(k)} K.$$

Since A is smooth, this is is a finite projective $\hat{A}[\frac{1}{p}]$-module.

Our second candidate for the cohomology of X is the cohomology of the resulting de Rham complex:

$$H^\bullet(\Omega^\bullet_{\hat{A}[\frac{1}{p}]/K}).$$

Unfortunately, this turns out not to be finite-dimensional over K!

Example 13.5. Take $\overline{A} = k[X]$, $A = W(k)[X]$. Then $\hat{A}$ is the ring $W(k)\langle X\rangle$ of null power series (also called *strictly convergent* power series) over $W(k)$, and similarly

$$\hat{A}\left[\frac{1}{p}\right] = K\langle X\rangle = \left\{ \sum_{n=0}^{\infty} a_n X^n \in K[\![X]\!] | a_n \to 0 \text{ for the } p\text{-adic topology} \right\}.$$

It is easy to see that $\sum p^n X^{p^n - 1}$ belongs to $\hat{A}[\frac{1}{p}]$ but its antiderivative $\sum X^{p^n}$ does not. By similar considerations, one may show that $H^1(\Omega^\bullet_{\hat{A}[\frac{1}{p}]/K})$ is infinite-dimensional over K.

Definition 13.6. As prelude for the general case, let us see how to modify this example to eliminate the issue we have just seen. If we think of $K\langle X\rangle$ as the rigid analytic functions on the closed unit disc, the problem is that antidifferentiation preserves the radius of convergence but not the convergence at the boundary. (This is of course backwards from what happens in classical analysis, where it is differentiation of power series that has a similar problem.) To remedy this issue, we consider instead functions which are holomorphic on some *larger* disc: taking these together yields the ring

$$\begin{aligned} K\langle X\rangle^\dagger &= \left\{ \sum_{n=0}^\infty a_n X^n | a_n \in K, \lim_{n\to\infty} \sup |a_n|^{\frac{1}{n}} < 1 \right\} \\ &= \varinjlim_{\rho > 1} \left\{ \sum a_n X^n | a_n \in K, \lim_{n\to\infty} |a_n|\rho^n = 0 \right\}. \end{aligned}$$

One may verify easily that the sequence

$$0 \to K\langle X\rangle^\dagger \xrightarrow{d} K\langle X\rangle^\dagger \, dx \to 0$$

has cohomology K in degree 0 (the constants) and 0 in degree 1 (differentiation is surjective).

Returning to the general case, we introduce the following definition.

Definition 13.7. Let R be a ring and let $I = (x_1, \dots, x_n) \subset R$ be a finitely generated ideal. Let $\hat{R} := \varprojlim_{m\to\infty} R/I^m$ denote the I-adic completion of R. Note that for any $y_1, \dots, y_n \in R$ and any power series $c(X) = \sum_J c_J X_1^{j_1} \cdots X_n^{j_n} \in R[\![X_1, \dots, X_n]\!]$ for which there exists a function $f(x)$ with $\lim_{x\to\infty} f(x) = \infty$ and

$$c_J \in I^{f(j_1+\cdots+j_n)} \qquad (j_1, \dots, j_n \geq 0),$$

the evaluation $c(y_1, \dots, y_n)$ makes sense as an element of R.

We define the *weak completion* of R with respect to I to be the smallest subring $R^\dagger$ of $\hat{R}$ with the following property: if $y_1, \dots, y_n \in R^\dagger$ and $c(X) = \sum_J c_J X_1^{j_1} \cdots X_n^{j_n} \in R[\![X_1, \dots, X_n]\!]$ is a power series for which there exists a constant $C > 0$ with

$$c_J \in I^{\lfloor C(j_1+\cdots+j_n)\rfloor} \qquad (j_1, \dots, j_n \geq 0),$$

then $c(y_1, \dots, y_n) \in R^\dagger$.

Example 13.8. For $R = W(k)[X_1, \dots, X_n]$,

$$R^\dagger = W(k)\langle X_1, \dots, X_n\rangle^\dagger := \varinjlim_{\rho > 1} \left\{ \sum a_I X^I | a_I \in W(k), \lim_{n\to\infty} |a_I|\rho^{i_1+\cdots+i_n} = 0 \right\}$$

is the ring of rigid analytic functions on all possible polydiscs of radius > 1.

Theorem 13.9 (Fulton; see [45])**.** *If R is a noetherian ring, then any weak completion of R is again noetherian.*

Remark 13.10. Corresponding to the passage from adically complete rings to formal schemes, one may use weak completions to define *weak formal schemes*. This was done by Meredith [85].

Definition 13.11. With notation as before, let $A^\dagger$ be the weak completion of A with respect to the ideal (p). We may compute $A^\dagger$ by choosing a surjection $W(k)[X_1, \dots, X_n] \to A$ and then taking

$$A^\dagger = A \otimes_{W(k)[X_1,\dots,X_n]} W(k)\langle X_1, \dots, X_n\rangle^\dagger.$$

Using such a presentation, we may also define the module of continuous Kähler differentials $\Omega_{A^\dagger[\frac{1}{p}]/K}$ (for $A = W(k)[X_1, \dots, X_n]$ it will again be freely generated by $dX_1, \dots, dX_n$) and then define the *Monsky-Washnitzer cohomology* to be

$$H^i_{\mathrm{MW}}(X) = \mathrm{H}^i(\Omega^\bullet_{A^\dagger[\frac{1}{p}]/K}).$$

To see that this gives something well-defined, we need to verify that it is independent of the choice of lifting and the presentation.

Theorem 13.12 (Monsky–Washnitzer)**.** *The Monsky–Washnitzer cohomology groups of X are independent of the choice of lift A and of the presentation of A as a finitely generated $W(k)$-algebra. Moreover, they are (contravariantly) functorial in X.*

Proof. Let us sketch some ideas behind the proof. One starts with a form of the Poincaré lemma: the natural map

$$H^i_{\mathrm{MW}}(X) \to H^i_{\mathrm{MW}}(X \times_k \mathbb{A}^1_k)$$

is an isomorphism (using a particular presentation of X and the corresponding presentation of $X \times_k \mathbb{A}^1_k$ by adding one more variable). This can then be applied in the following ways.

- Adding generators to a presentation does not change cohomology. Given two presentations, we may then combine their generators to see that the cohomology groups given by the two presentations may be naturally identified.
- Given two different lifts of the same morphisms, we may "interpolate" between the two to see that they define the same morphism in cohomology.

This gives everything we need, except that it is not yet apparent that one can always lift a morphism at all (even if one does not specify in advance a lift of either the source or target). This again follows from the theorem of Arabia [4]. □

Remark 13.13. The proof of the Poincaré lemma gives more than just an isomorphism in cohomology, but also a chain homotopy witness for the fact that the composition $H^i_{\mathrm{MW}}(X) \to H^i_{\mathrm{MW}}(X \times_k \mathbb{A}^1_k) \to H^i_{\mathrm{MW}}(X)$ is the identity (mapping X to $X \times_k \mathbb{A}^1_k$ via the zero section). This makes it possible to globalize the definition of Monsky–Washnitzer cohomology to accommodate general smooth schemes over k.

The following is not *a priori* clear, and indeed was unknown to Monsky–Washnitzer.

Theorem 13.14 (Berthelot)**.** *The space* $\mathrm{H}^\bullet_{\mathrm{MW}}$ *is finite-dimensional over* K*.*

Proof. This was first proved by Berthelot in 2000 [8], using de Jong's theorem on alterations [27]. A stronger result, allowing coefficients in the p-adic analogue of a local system, was given by Kedlaya in 2006 [65]. □

What makes Monsky–Washnitzer computable in practice is the following comparison theorem with the de Rham cohomlogy of the generic fiber.

Theorem 13.15. *If* X *is nice enough, then*

$$H^i_{\mathrm{dR}}(\mathfrak{X}_K) \xrightarrow{\simeq} H^i_{\mathrm{MW}}(X).$$

Here "Nice" means that X *admits a smooth proper compactification* $\overline{X}$ *such that the complement* $\overline{X} - X$ *corresponds to a normal crossings divisor and* $\overline{X}$ *admits a smooth proper lifting over* $W(k)$*, in which the complement of a certain relative normal crossings divisor lifts* X*.*

When X is not nice, we could use de Jong's resolution of singularities and excision to reduce to the nice case.

Lemma 13.16 (Excision)**.** Suppose X is smooth and $Z \subset X$ is pure and smooth of codimension d, set $U = X - Z$, then we have a short exact sequence

$$\cdots \to H^{i-2d}_{\mathrm{dR}}(Z) \to H^i_{\mathrm{dR}}(X) \to H^i_{\mathrm{dR}}(U) \to H^{i-2d+1}_{\mathrm{dR}}(Z) \to \cdots$$

The final piece needed to make Monsky–Washnitzer cohomology into a Weil cohomology theory is the Lefschetz trace formula for Frobenius. We will state *and prove* this in the next lecture.

14. Frobenius actions and the Lefschetz–Monsky trace formula (November 20)

In this section, we will give an example of a Frobenius action on the cohomology of the affine piece of a hyperelliptic curve, and give a proof of the Lefschetz trace formula in the p-adic setting due to Monsky.

Readings 14.1. The Frobenius action on a hyperelliptic curve is presented as in [64]. The Lefschetz trace formula is presented as in Monsky [89]. See also [68].

Example 14.2. Let $k = \mathbb{F}_q$ be a finite field with characteristic p not equal to 2. Let $P(x) \in k[x]$ be a monic polynomial of degree $2g + 1$ with no repeated roots. Let $\overline{A}$ be the ring

$$\overline{A} := k[x, y, z]/(y^2 - P(x), yz - 1)$$

and let X the affine curve $X := \operatorname{Spec} \overline{A}$; X is then a hyperelliptic curve of genus g with one rational Weierstrass point at infinity, with all of its Weierstrass points removed. Choose $\tilde{P}(x) \in W(k)[x]$ a monic polynomial lifting $P(x)$; note that $\tilde{P}(x)$ has no repeated roots, since it is a lift of something with no repeated roots. Define the lift

$$A := W(k)[x, y, z]/(y^2 - \tilde{P}(x), yz - 1).$$

Remark 14.3. The space $H^1_{\mathrm{dR}}(\operatorname{Spec} A[\frac{1}{p}])$ has a basis given by

$$x^i \frac{dx}{y} \qquad (i = 0, \dots, 2g-1),$$
$$x^i \frac{dx}{y^2} \qquad (i = 0, \dots, 2g)$$

(see Set 5 exercises). Since X is hyperelliptic, it has an involution

$$x \mapsto x$$
$$y \mapsto -y$$
$$z \mapsto -z.$$

Lifting this involution, we find that the $x^i \frac{dx}{y}$ form the minus eigenspace and that the $x^i \frac{dx}{y^2}$ form the plus eigenspace.

Definition 14.4. With A, k as before, let $A^\dagger$ be the set of all sums of the form

$$\sum_{n=-\infty}^{\infty} \frac{Q_n(x)}{y^n},$$

with $Q_n(x) \in W(k)[x]$ such that $\deg Q_n(x) \leq 2g$. Moreover, we require as a convergence condition that there exist positive integers a, b such that the p-adic valuation of Q_n is at least $a|n| - b$.

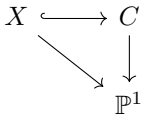
Lemma 14.5. Every element in A can be written as a finite sum

$$\sum_{n=n_0}^{n_1} \frac{Q_n(x)}{y^n}$$

for some integers n_0, n_1, with $Q_n(x) \in W(k)[x]$ of degree $\leq 2g$.

Proof. Note first that via the relation $yz = 1$ in A that every element in A is a polynomial in x, y, y^{-1}. We may then successively take remainders modulo $\tilde{P}(x)$ and use the relation $y^2 = \tilde{P}(x)$ to reduce the degrees of everything to at most $2g$. □

To compute the action of the q-power Frobenius on $H^1_{\mathrm{MW}}(X)$, it suffices to find a map $Q : A^\dagger \to A^\dagger$ lifting the action of Frobenius on $\overline{A}$. To do this, let C be the projective curve obtained from X by adding back in the Weierstrass points. We then get the following diagram:



where the vertical arrow is the two-to-one cover of $\mathbb{P}^1$ by the hyperelliptic curve C. Since we took out the Weierstrass points in defining X, the arrow $X \to \mathbb{P}^1$ is étale, so we may lift the standard Frobenius map $x \mapsto x^p$ on $\mathbb{P}^1$ to X. Using that $(A^\dagger, (p))$ is a henselian pair, we may then lift the induced Frobenius action

on X to $A^\dagger$. We may compute out this action explicitly, as

$$\begin{aligned} x &\mapsto x^q \\ y &\mapsto \sqrt{\tilde{P}(x^q)} \\ &= \sqrt{\tilde{P}(x)^q + p(*)} \\ &= \sqrt{y^{2g} + p(*)} \\ &= y^q(1 + p(*)z^{2q})^{1/2} \\ z &\mapsto y^{-1}, \end{aligned}$$

where the expression in the last line for y may be expanded out using the generalized binomial theorem and the instances of $(*)$ denote polynomial quantities which may be computed explicitly.

Remark 14.6. This example of the use of Monsky-Washnitzer cohomology for algorithmic computation is taken from [64]. Subsequently, Tuitman [105, 106] adapted this method to work for arbitrary curves. (As an aside, note that this also gives a method to compute Coleman's p-adic path integrals [6]; such computations play a pivotal role in the Chabauty–Coleman–Kim approach to finding rational points on curves, as in [5].)

In higher dimension, Abbott–Kedlaya–Roe [1] involves similar computations for smooth hypersurfaces in $\mathbb{P}^n$, using the explicit description of de Rham cohomology by Griffiths [48, 49]. This has been further adapted to smooth nondegenerate hypersurfaces in toric varieties by Costa–Harvey–Kedlaya (unpublished, but see [24] for a preview).

We now turn our attention to proving the Lefschetz trace formula.

Theorem 14.7 (Monsky)**.** *Let $X = \mathrm{Spec}(\overline{A})$ be a smooth affine scheme over a finite field $k = \mathbb{F}_q$ of characteristic p. Then*

$$\#X(\mathbb{F}_q) = \sum (-1)^i \ \mathrm{Trace}(q^n F^{-1}, H^i_{\mathrm{MW}}(X)).$$

Proof. We start with two key reductions.

- First, both sides of the desired equality are additive for a scissors decomposition (i.e. a union of an open set and its closed complement): this is obvious for the left side, and for the right side it follows from the excision exact sequence.
- Second, if $X = \mathrm{Spec}(\mathbb{F}_q)$, then $H^0_{\mathrm{MW}}(X) = K$, $H^i_{\mathrm{MW}}(X) = 0$ for $i > 0$, and the action of F on $H^0_{\mathrm{MW}}(X)$ is the identity map. Hence both sides of the desired equality equal 1.

From these observations, it follows that we may reduce the general case of the theorem to the case where $\#X(\mathbb{F}_q) = 0$. This equality has a key algebraic consequence: it implies that the ideal in $\overline{A}$ generated by all elements of the form $f^q - f$ for $f \in \overline{A}$ is trivial (see supplementary exercises). That is, we can find an equality of the form

$$1 = \sum \overline{a}_i(\overline{b}_i^q - \overline{b}_i)$$

for some $\overline{a}_i, \overline{b}_i \in \overline{A}$; we will use this equality in a crucial way to see that the right-hand side of the trace formula is also zero.

By the Elkik–Arabia theorem cited in the previous lectures, we may choose a lift $\phi : A^\dagger \to A^\dagger$ of Frobenius to $A^\dagger$. We may then find elements $a_i, b_i \in A^\dagger$ such that

$$1 \equiv \sum a_i(\varphi(b_i) - b_i) \mod p.$$

Since p belongs to the Jacobson radical of $A^\dagger$, we may multiply all of the a_i by a suitable unit to ensure that in fact

$$1 = \sum a_i(\varphi(b_i) - b_i).$$

Note that $\varphi : A^\dagger \to A^\dagger$ is finite flat of degree q^n (this is true mod p, and one can argue then that it is true in general). Define $\psi : A^\dagger \to A^\dagger$ to be the *reduced trace* of φ, i.e.

$$\psi = \frac{1}{q^n} \cdot \mathrm{Trace}(\varphi), \qquad \psi \circ \varphi = \mathrm{id}.$$

Since we know by Berthelot that $\dim H^i_{\mathrm{MW}}(X) < \infty$, ψ is not just a left inverse but a genuine inverse to φ by linear algebra. Thus, the actions of $q^n F^{-1}$ and $q^n\psi$ coincide, so we want to show that the alternating sum of traces of ψ is 0; in fact, we will show that *each* trace of ψ vanishes in this situation.

Note that $\psi(\varphi(a)b) = a\psi(b)$. Therefore, if we define $L_a : \Omega^+ \to \Omega^+$ on the de Rham complex to be multiplication by a, we then have

$$\psi \circ L_{\varphi(a)} = L_a \circ \psi,$$

implying that

$$\sum_i L_{a_i} \circ \psi \circ L_{\varphi(b_i)} = \sum_i L_{a_i} \circ L_{b_i} \circ \psi.$$

Take the traces of both sides to obtain

$$\begin{aligned}\mathrm{Trace}\left(\sum_i L_{a_i} \circ \psi \circ L_{\varphi(b_i)}\right) &= \mathrm{Trace}\left(\sum_i L_{a_i} \circ L_{b_i} \circ \psi\right) \\ &= \mathrm{Trace}\left(\sum_i L_{a_i b_i} \circ \psi\right)\end{aligned}$$

since composition of the multiplication operator becomes multiplication by the product. On the other hand, we may apply the fact that the trace of a product is invariant under cyclic permutations to get that

$$\mathrm{Trace}\left(\sum_i L_{a_i} \circ \psi \circ L_{\varphi(b_i)}\right) = \mathrm{Trace}\left(\sum_i L_{\varphi(b_i)} \circ L_{a_i} \circ \psi\right).$$

We thus get the equality

$$\mathrm{Trace}\left(\sum_i L_{\varphi(b_i)} \circ L_{a_i} \circ \psi\right) = \mathrm{Trace}\left(\sum_i L_{a_i b_i} \circ \psi\right),$$

so that

$$\begin{aligned}0 &= \mathrm{Trace}\left(\sum_i (L_{\varphi(b_i)} \circ L_{a_i} \circ \psi - L_{a_i b_i} \circ \psi)\right) \\ &= \mathrm{Trace}\left(\sum_i L_{a_i \varphi(b_i) - b_i} \circ \psi\right) \\ &= \mathrm{Trace}\left(L_{\sum_i a_i(\varphi(b_i) - b_i)} \circ \psi\right) \\ &= \mathrm{Trace}(\psi)\end{aligned}$$

since $\sum_i a_i(\varphi(b_i) - b_i) = 1$ by prior arrangement. Thus, $\mathrm{Trace}(\psi) = 0$ as desired. □

Remark 14.8. Recall that Monsky did not have the finite-dimensionality of $\dim H^i_{\mathrm{MW}}(X)$ at his disposal when he originally devised this argument. This required him to be more careful in two aspects. First, he had to introduce a suitable topology in order to argue that he could take traces (recall that there was a corresponding step in Dwork's proof of the rationality of zeta functions). Second, he could not assume that F is invertible, and so he had to set up the formula in a more cautious way.

15. Étale local systems (November 25)

In this lecture, we turn back to étale cohomology and introduce the notion of an *étale local system*.

Readings 15.1. The (profinite) étale fundamental group is introduced in SGA 1 [57].

First we cover the étale fundamental group. Its definition is motivated by the fact that in algebraic geometry, universal covers don't have a good equivalent, but finite covering space maps correspond to finite étale morphisms.

Definition 15.2. Let X be a connected scheme and $x \in X$ a geometric point, i.e., a map

$$x : \mathrm{Spec}(\overline{k}) \to X$$

where $\overline{k}$ is an algebraically closed field. (In fact we could even take $\overline{k}$ to be only separably closed, but never mind about that here.) To this data, one then attach a profinite fundamental group $\pi_1(X, x)$ the *profinite (étale) fundamental group of X with basepoint x*. This has the following properties.

(1) If $y \to X$ is another geometric point of X, then $\pi_1(X, x) \cong \pi_1(X, y)$ via an isomorphism which is well-defined up to conjugation. This parallels the topological case, in which such isomorphisms are paths joining the two basepoints, and two different paths give isomorphisms differing by conjugation by the loop formed by the two paths.

This also parallels Galois theory. Fixing an algebraic closure of a field gives one absolute Galois group, and choosing another closure gives another group that is isomorphic to the first one, but the isomorphism is only well-defined up to conjugation.

(2) If

$$f : X \to Y$$

is a morphism, then the composition

$$x \to X \xrightarrow{f} Y$$

defines a geometric point $f(x) \in Y$. There is then a homomorphism

$$f_* : \pi_1(X, x) \to \pi_1(Y, f(x)).$$

Note that this map is defined "on the nose", without any conjugation ambiguity. However, if we want to consider a map

$$\pi_1(X, \star) \to \pi_1(Y, \star)$$

with unrestricted base points, then we must first apply the previous point to line up the base points; consequently, the resulting map is only well-defined up to conjugation.

(3) If $X = \mathrm{Spec}(k)$ for some field k, and $x = \mathrm{Spec}(\overline{k})$, then $\pi_1(X, x) \cong \mathrm{Gal}(\overline{k}/k)$.

Remark 15.3. In the case where X is normal (and excellent, so that normalization of finite covers behaves well) and $\eta \in X$ is its generic point, we may give a concrete description of $\pi_1(X, x)$ where x is a geometric point mapping to η. Note that $\kappa(\eta) = k(X)$, the function field of X. We may identify $\pi_1(X, x)$ with the quotient of the absolute Galois group $G_{k(X)}$ corresponding to the compositum of every finite extension of $k(X)$ inside of which the normalization of X is finite and étale over X. This family of fields is closed under compositum because in the composite field, the normalization of X is a connected component of the fiber product of the covers.

Note that this definition does not give a good description of what happens when you change the basepoint to something not lying over η. It is thus hard to see how functoriality works, say, for the embedding of a closed subscheme into X.

Example 15.4. Suppose that $X = \mathrm{Spec}(\mathbb{Z}[1/N])$. Then $\pi_1(X, x) = G_{\mathbb{Q},S}$, where S is the set of primes dividing N and $G_{\mathbb{Q},S}$ is the Galois group of the compositum of all number fields unramified outside S. As an example of functoriality, for any p not dividing N, there is a diagram

$$\begin{array}{ccc} G_{\mathbb{Q}_p} & \longrightarrow & G_{\mathbb{Q},S} \\ \downarrow & \nearrow & \\ G_{\mathbb{F}_p} & & \end{array}$$

but we may obtain other such diagrams via conjugation in $G_{\mathbb{Q},S}$.

The general definition uses the following setup.

Definition 15.5. Let $\mathbf{FÉt}(X)$ denote the category of finite étale schemes over X. For a geometric point $x \in X$, we have a base change (fiber) functor

$$\omega_x : \mathbf{FÉt}(X) \to \mathbf{FÉt}(x)$$

where $\mathbf{F\acute{E}t}(x)$ is canonically equivalent to the category of finite sets (via the forgetful functor from schemes to sets: each object of $\mathbf{F\acute{E}t}(x)$ is a finite disjoint union of copies of x).

Definition 15.6. We define the *profinite fundamental group* of the scheme X with basepoint x to be

$$\pi_1(X, x) = \mathrm{Aut}(\omega_x)$$

That is, the group of natural isomorphisms of the fiber functor to itself.

Lemma 15.7. If $X = \mathrm{Spec}(k)$ for some field k, and $\overline{k}$ is an algebraic closure of k, then $\pi_1(X, \mathrm{Spec}(\overline{k}))$ is naturally isomorphic to $\mathrm{Gal}(\overline{k}/k)$ (the absolute Galois group of k).

Proof. See supplemental exercises. □

Remark 15.8. Definition 15.6 is an example of a *Tannakian* construction. The formalism of Tannakian categories typically involves categories of vector spaces (over a field of characteristic 0, for technical reasons). rather than finite sets; one can get into that setup by replacing finite sets with the free vector spaces that they generate. The standard treatments of Tannakian categories are [94] and [31].

Remark 15.9. The group $\pi_1(X, x)$ was originally called the *étale fundamental group*; for instance, this is the terminology used in [57]. The terminology *profinite fundamental group* is now preferred because, when X is not normal, there are infinite-degree étale covers of schemes that should be included in the construction of the étale fundamental group, but do not contribute to the profinite fundamental group as we have defined it. One way to get the "right" definition is to replace finite étale covers with *pro-étale covers* in the sense of Bhatt–Scholze [9].

Example 15.10. A typical example of the previous remark is the "banana". This is the connected, but not normal, scheme given by gluing two copies of $\mathbb{P}^1$ together at 0 and ∞. It has an infinite étale cover which looks like a helix, and is constructed as a set as follows

$$\frac{\mathbb{P}^1 \times \mathbb{Z}}{(0, 2n) \sim (0, 2n+1), (\infty, 2n+1) \sim (\infty, 2n+2)}$$

That is, it is countably many copies of $\mathbb{P}^1$ glued together alternately at 0 and ∞. There are deck transformations of this cover given by the action of $2\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$ on the second factor that are not included in the profinite fundamental group.

Remark 15.11. The issue of infinite étale covers becomes much more acute if one passes from algebraic geometry to analytic geometry. For example, in rigid analytic geometry, the Tate elliptic curve has important infinite covers analogous to those seen in complex geometry. However, there are even more exotic examples: the Gross-Hopkins period maps give rise to infinite covers of rigid analytic projective spaces [51, 52]. In perfectoid geometry, something similar happens with Hodge-Tate period maps [18].

Remark 15.12. Note that when applicable, functoriality gives Frobenius elements coming from points, although they are only defined up to conjugation in general. For example, you can take the image of Frobenius in the diagonal arrow in the diagram of Galois groups in Example 15.4.

Now we turn to lisse $\overline{\mathbb{Q}}_\ell$-sheaves.

Definition 15.13. Let ℓ be a prime. A *lisse $\overline{\mathbb{Q}}_\ell$-sheaf* "is" a finite-dimensional continuous representation

$$\pi_1(X) \to \mathrm{GL}(r, \overline{\mathbb{Q}}_\ell);$$

note that here we are being sloppy and dropping the basepoint in the fundamental group.

Remark 15.14. Note that

$$\mathrm{GL}(r, \overline{\mathbb{Q}}_\ell) = \bigcup_{E/\mathbb{Q}_\ell \mathrm{finite}} \mathrm{GL}(r, E)$$

and any continuous representation from a profinite group, such as $\pi_1(X)$, into $\mathrm{GL}(r, \overline{\mathbb{Q}}_\ell)$ factors through some $\mathrm{GL}(r, E)$ (see supplemental exercises). Consequently, the category of lisse $\overline{\mathbb{Q}}_\ell$-sheaves is the 2-colimit of the categories of lisse E-sheaves (meaning representations into $\mathrm{GL}(r, E)$) over all finite extensions $E/\mathbb{Q}_\ell$ within $\overline{\mathbb{Q}}_\ell$.

Remark 15.15. This is not the real definition, just a shortcut. It is not even a sheaf. We are on the wrong side of the Riemann-Hilbert correspondence.

There is an actual sheaf on some Grothendieck topology (a profinite version of the étale topology) that corresponds to a lisse $\overline{\mathbb{Q}}_\ell$-sheaf as we have defined it. If $\mathcal{F}$ is such a sheaf, we denote the corresponding representation by $\Lambda_{\mathcal{F}}$.

Definition 15.16. Suppose X is a scheme of finite type over a finite field k. Then there is an L-function associated to a lisse sheaf $\mathcal{F}$ on X, given by

$$L(X/k, \mathcal{F})(T) := \prod_{x \in X \text{ closed point}} \det\left(1 - T^{[\kappa(x):k]} \Lambda_{\mathcal{F}}(\mathrm{Frob}_x)\right)$$

Here Frob_x is the geometric Frobenius coming from the map

$$\pi_1(x) \to \pi_1(X).$$

The geometric Frobenius is the inverse of the map $t \mapsto t^{\#\kappa(x)}$, which is called the arithmetic Frobenius.

As with zeta functions, we may formally rewrite this L-function as

$$L(X/k, \mathcal{F})(T) = \exp\left(\sum_{n=1}^{\infty} s_n \frac{T^n}{n}\right), \qquad s_n := \sum_{[\kappa(x):k]=n} \mathrm{Trace}(\mathrm{Frob}_x).$$

Why do we care?

Example 15.17. Let

$$\pi : Y \to X$$

be a morphism between smooth proper schemes of finite type over a finite field k, and ℓ a prime number not dividing the characteristic of k. Then there exist lisse $\overline{\mathbb{Q}}_\ell$-sheaves $\mathcal{F}_i$ such that for each closed point $x \in X$,

$$\det\left(I - T\Lambda_{\mathcal{F}_i}(\mathrm{Frob}_x)\right)$$

is the i-th factor of the zeta function

$$Z(\pi^{-1}(x), T)$$

in the factorization predicted by the Weil conjectures. These $\mathcal{F}_i$ arise as *higher direct images* of the trivial sheaf $\overline{\mathbb{Q}}_\ell$ on Y.

Remark 15.18. Deligne conjectured that every lisse $\overline{\mathbb{Q}}_\ell$-sheaf on X "comes from geometry" in the sense that each of its irreducible subquotients is a twist (by a rank 1 sheaf) of a subquotient of something appearing in the previous example.

This conjecture was stated somewhat cautiously in [30]. It was motivated by the observation that when X is a curve, it followed from Drinfeld's geometric proof of the Langlands correspondence for the group GL_2 over the function field [35], which appeared at around the same time as [30]. For a general lisse sheaf on a curve, the conjecture follows from the extension of Drinfeld's work to GL_n given by L. Lafforgue [74].

For X of dimension greater than 1, we do not know of any plausible approach to proving Deligne's conjecture. However, one can extract a number of concrete predictions from it, concerning both ℓ-adic and p-adic Weil cohomologies, that can be verified using the case of curves as a black box. See for example the introduction to [71] and references cited therein.

16. Étale fundamental groups (November 26)

In this section, we'll continue the discussion of the étale fundamental group, giving the formulation of Deligne's "Weil II" theorem.

Readings 16.1. We follow the setup of [30]. While there is a whole parallel p-adic setup, we did not have time to say much it in these lectures (besides Remark 16.8 below); for more, see [70].

Remark 16.2. Let's recall our setup from last time. Let X be a smooth connected scheme over a finite field k of characteristic p. (We didn't require smoothness last time, but it will be convenient later to add this hypothesis.) Let ℓ be a prime nonzero in k. We "defined" the notion of a *lisse $\overline{\mathbb{Q}}_\ell$-sheaf* $\mathcal{F}$ in terms of an associated continuous representation $\Lambda_{\mathcal{F}} : \pi_1(X, x) \to \mathrm{GL}(r, \overline{\mathbb{Q}}_\ell)$, where x is a geometric point of X. For any closed point $y \in X$, we have a well-defined conjugacy class of elements $\mathrm{Frob}_y \in \pi_1(X, x)$.

Definition 16.3. Suppose now that X is geometrically irreducible. The *geometric profinite fundamental group* of X is defined as $\pi_1(X_{\overline{k}})$, where the basepoint is omitted.

The geometric fundamental group is related to the actual fundamental group by functoriality for the morphism $X_{\overline{k}} \to X \to k$. This sequence of maps behaves a bit like a homotopy fiber sequence, in that we obtain the following exact sequence.

Proposition 16.4. The sequence

$$1 \to \pi_1(X_{\overline{k}}) \to \pi_1(X) \to G_k \to 1$$

is exact.

Proof. See for example [57, Exposé IX, Théoréme 6.1]. □

Remark 16.5. In lieu of saying more about the proof of Proposition 16.4, we point out a philosophical observation: the group $\pi_1(X_{\overline{k}})$ has "many" representations, but $\pi_1(X)$ has "few" representations.

To wit, the exact sequence gives rise to a map $G_k \to \mathrm{Out}(\pi_1(X_{\overline{k}}))$ to the group of outer automorphisms of $\pi_1(X_{\overline{k}})$ (which makes sense without regard for the basepoint). The group $\mathrm{Out}(\pi_1(X_{\overline{k}}))$ acts on the set of isomorphism classes of continuous $\overline{\mathbb{Q}}_\ell$-representations of $\pi_1(X_{\overline{k}})$, and the class of any representation that extends to $\pi_1(X_{\overline{k}})$ must be a fixed point for this action. However, "most" classes are not fixed points.

Example 16.6. Let X be the scheme obtained from a smooth, projective, geometrically irreducible curve of genus g over k by removing a zero-dimensional closed subscheme of length m over k. While the group $\pi_1(X_{\overline{k}})$ is somewhat difficult to define, Grothendieck defined a quotient of it, the *tame profinite fundamental group* $\pi_1^{\mathrm{tame}}(X_{\overline{k}})$, which is much easier to compute: it is a certain profinite completion of the free group generated by $2g + m$ letters $a_1, \dots, a_g, b_1, \dots, b_g, c_1, \dots, c_m$ modulo the relation $[a_1, b_1] \cdots [a_g, b_g] c_1 \cdots c_m$, where the brackets denote commutators. (That is, we take the profinite completion of the ordinary fundamental group of a genus-g Riemann surface with m punctures.)

It is quite easy to write down continuous $\overline{\mathbb{Q}_\ell}$-representations of $\pi_1^{\mathrm{tame}}(X_{\overline{k}})$: this just comes down to writing down systems of matrices corresponding to the generators, with a bit of care to ensure that the resulting map extends to the profinite completion. However, most of these maps will not be preserved by the outer action of G_k.

Remark 16.7. Proposition 16.4 remains true for an arbitrary field k. In the case $k = \mathbb{Q}$, the resulting "mysterious" action of G_k on $\pi_1(X_{\overline{k}})$ gives rise to a "mysterious" action on finite covers of $\mathbb{P}^1$ branched over $\{0, 1, \infty\}$ (since these covers are rigid, they always give rise to curves defined over number fields) and in turn to a "mysterious" action on Grothendieck's *dessins d'enfants*.

Remark 16.8. Although we have not included in these lectures a detailed account, there is a parallel p-adic construction of "lisse sheaves" to which much of the following discussion carries over. Let us briefly indicate the analogue of Proposition 16.4 in this setup.

The analogue of a continuous ℓ-adic representation of $\pi_1(X_{\overline{k}})$ is an *overconvergent isocrystal*. For X affine, such an object can be described as a finite projective module ("vector bundle") over a dagger lift $A^\dagger$ equipped with an integrable K-linear connection, where K again denotes the fraction field of the ring of Witt vectors $W(k)$. (An additional condition must be imposed to ensure that this definition is independent of the choice of the dagger lift.)

The analogue of a continuous ℓ-adic representation of $\pi_1(X)$ is an *overconvergent F-isocrystal*. Such an object consists of an overconvergent isocrystal plus an isomorphism with its Frobenius pullback.

As in the étale case, it is relatively easy to manufacture overconvergent isocrystals from the definition, but most of these will not admit a compatible Frobenius action.

Definition 16.9. Fix now a lisse $\overline{\mathbb{Q}}_\ell$-sheaf $\mathcal{F}$. Let $H^i(X_{\overline{k}}, \mathcal{F})$ and $H^i_c(X_{\overline{k}}, \mathcal{F})$ denote *étale cohomology with coefficients in $\mathcal{F}$* and *étale cohomology with compact support with coefficients in $\mathcal{F}$* respectively. We will refer to these for short as H^i and H^i_c.

It turns out that H^i and H^i_c are finite-dimensional $\overline{\mathbb{Q}}_\ell$ vector spaces on which G_k acts continuously; that is, there are lisse $\overline{\mathbb{Q}}_\ell$-sheaves over the point $\mathrm{Spec}(k)$. These are special cases of higher direct images $R^i f_* \mathcal{F}, R^i f_! \mathcal{F}$ for $f : X \to S$ a smooth morphism, but in general these land in a large category than the lisse

$\overline{\mathbb{Q}}_\ell$-sheaves (namely, the category of *constructible* $\overline{\mathbb{Q}}_\ell$-sheaves). When f is smooth proper, they are indeed lisse lisse $\overline{\mathbb{Q}}_\ell$-sheaves.

To formulate Poincaré duality for étale cohomology, before stating it, we need three more constructions.

- The *forgetting supports* map $H^i_c \to H^i$, which is an isomorphism when X is proper.
- The *trace map*: when X is of dimension n, there is a G_k-equivariant map

$$H^{2n}_c(X_{\overline{k}}) \twoheadrightarrow \overline{\mathbb{Q}}_l(-n)$$

  where $\overline{\mathbb{Q}}_\ell(-n)$ is the $(-n)$-th Tate twist of $\overline{\mathbb{Q}}_\ell$. This map is an isomorphism if X is (smooth and) geometrically irreducible.
- The *cup product pairing*:

$$H^i_c(X_{\overline{k}}, \mathcal{F}) \times H^{2n-i}(X_{\overline{k}}, \mathcal{F}^\vee) \to H^{2n}_c(X_{\overline{k}}, \mathcal{F} \otimes \mathcal{F}^\vee) \to H^{2n}_c(X_{\overline{k}}, \overline{\mathbb{Q}}_l) \twoheadrightarrow \overline{\mathbb{Q}}_l(-n).$$

With this pairing defined, we can state Poincare duality in étale cohomology.

Proposition 16.10. (Poincare Duality) For X smooth over k, the cup product pairing is perfect. That is, it defines a G_k-equivariant isomorphism of either $H^i_c(X_{\overline{k}}, \mathcal{F})$ and $H^{2n-i}(X_{\overline{k}}, \mathcal{F}^\vee)$ with the space of maps of the other one into $\overline{\mathbb{Q}}_l(-n)$.

We also have a Lefschetz trace formula for étale cohomology. In the following formulation, it does not even require X to be smooth over k.

Proposition 16.11 (Lefschetz trace formula)**.** For X of finite type over k,

$$\sum_{x \in X(k)} \mathrm{Trace}(\mathrm{Frob}_x, \mathcal{F}) = \sum (-1)^i \,\mathrm{Trace}(\mathrm{Frob}_x, H^i_c(X_{\overline{k}}, \mathcal{F}))$$

This implies a factorization of the L-function of a now-familiar form.

Corollary 16.12.

$$L(X, \mathcal{F}) = \prod_{i=0}^{2\dim(X)} \det(1 - \mathrm{Frob}_k\, T, H^i_c(X_{\overline{k}}, \mathcal{F}))^{(-1)^{i+1}}$$

.

In order to state Weil II, we need to add a little more notation in order to measure archimedean absolute values.

Definition 16.13. Fix an algebraic (but in no way topological!) embedding $\iota : \overline{\mathbb{Q}}_\ell \to \mathbb{C}$, and let $|x|_\iota = |\iota(x)|$ be the induced absolute value on $\overline{\mathbb{Q}}_l$.

We say that a lisse $\overline{\mathbb{Q}}_\ell$-sheaf $\mathcal{F}$ is *ι-pure of weight $w \in \mathbb{R}$* (resp. *ι-mixed of weight $\leq w$*, *ι-mixed of weight $\geq w$*) if for all finite extensions k'/k and all $x \in X(k')$, all the eigenvalues of $\Lambda_{\mathcal{F}}(\mathrm{Frob}_x)$ have ι-absolute value equal to (resp. greater than or equal to, less than or equal to) $(\#k')^{w/2}$.

Remark 16.14. The construction of an embedding ι as above is not at all effective: it depends on the axiom of choice. However, while one cannot easily run the proof of Weil II without making such an artificial choice, in practice one only ever applies such an embedding to algebraic numbers, for which it is much less exotic: it amounts to choosing a place of $\overline{\mathbb{Q}}$ above ℓ.

In its simplest form, "Weil II" is the following statement.

Theorem 16.15 (Deligne's "Weil II" theorem)**.** *Let U be a smooth geometrically connected curve over a finite field k. Let $\mathcal{F}$ be a lisse $\overline{\mathbb{Q}}_\ell$-sheaf on U which is ι-pure of weight w. Then $H^1_c(U_{\overline{k}})$, viewed as a lisse $\overline{\mathbb{Q}}_\ell$-sheaf on* $\mathrm{Spec}(k)$*, is ι-mixed of weight $\leq w+1$.*

Remark 16.16. By Poincaré duality, if U is proper, then we get that $H^1_c(U_{\overline{k}})$ is ℓ-pure of weight $w+1$.

17. RH and Weil II (December 4)

In this lecture, we'll talk about the proof of the (Riemann hypothesis part of the) Weil conjectures and Deligne's Weil II theorem. As this is far too deep a topic to cover thoroughly in one lecture, we instead describe a few of the main tools used in the proof and sketch their application. emphasizing analogies and intuition.

Readings 17.1. The original source is Deligne's "Weil II" paper [30], but the use of the Fourier transform was introduced later by Laumon [76]. We follow most closely [66], which is written in terms of p-adic coefficients but can be translated fairly directly back to the ℓ-adic side. See also [62] for an approach to "Weil II" in the style of "Weil I" [29].

Remark 17.2. The major advance of Weil II over Weil I is to allow for cohomology with nonconstant coefficients. Roughly speaking, this will allow us to translate the Riemann hypothesis—a statement about the cohomology of a simple sheaf on a complicated space—into a statement about a complicated sheaf on a simple space. In particular, we will be able to induct on dimension and reduce consideration to one-dimensional spaces.

17.1. **Dévissage.**

Remark 17.3. Making this a little more precise, let's start with a smooth connected variety X of dimension n over a finite field k. Then we'll study a tower

$$X = X_n \to X_{n-1} \to \cdots X_1 \to k$$

where each map $f_i : X_i \to X_{i-1}$ presents X_i as a family of curves over a space with dimension one lower. We want to understand the cohomology of X by summing over fibers of $f = f_n : X \to X_{n-1}$ (the zeta function of X is literally a product over fibers of this map). The Leray spectral sequence gives a description of $H^i(X, \overline{\mathbb{Q}}_\ell)$ in terms of $H^i(X_{n-1}, R^j f_* \overline{\mathbb{Q}}_\ell)$. This in turn can be described in terms of cohomology of some sheaves on X_{n-2}, and so on; this dévissage[2] allows us to reduce the general problem to understanding families of curves. Note that the Weil conjectures for curves tell us what we need to know about $H^i(X_{n-1}, R^j f_* \overline{\mathbb{Q}}_\ell)$, but for the remaining steps we really need Weil II because we start already with nontrivial coefficients.

As reported earlier (Definition 16.9), the higher direct images $R^i f_* \mathcal{F}$ of a lisse $\overline{\mathbb{Q}}_\ell$-sheaf do not always exist in the category of lisse $\overline{\mathbb{Q}}_\ell$-sheaves, but only in some larger category of *constructible $\overline{\mathbb{Q}}_\ell$-sheaves*. However, any object in the constructible object defines a stratification on its space, and it "looks lisse" on each stratum (but the rank may vary between strata). In particular, there is always a dense open subset on which it restricts to something lisse. More precisely, given a morphism $f : X \to S$ and a lisse $\overline{\mathbb{Q}}_\ell$-sheaf $\mathcal{F}$ on X, there is always an open dense subset U of S on which the higher direct images $R^i f_* \mathcal{F}$ are lisse and their formation commutes with arbitrary base change, in particular to a point. That means that these objects really are computing "cohomology in fibers".

17.2. **Nearby cycles and monodromy.** Of course, in the previous discussion we cannot simply throw away the part of S where the higher direct images of $\mathcal{E}$ are not lisse. We need a *nearby cycles formalism* or *vanishing cycles formalism* to extract information about the fibers over $S - U$ from the fibers over U.

Remark 17.4. Here's the classical picture to keep in mind. Imagine we have a family of complex varieties over a base with a singularity over some degenerate point. Then we should be able to understand the cohomology of the singular fiber by removing it and looking at the cohomology of all the "nice" things surrounding it. Imagine looping around this bad fiber and looking at a homology class of the good fibers surrounding it. As you go around this bad fiber, you'll get a different class once you get to the end of a loop: this gives a monodromy action on the cohomology. (This picture is sometimes called a *Milnor fiber*.)

The intuition is that the cohomology of an "exceptional fiber" can be recovered from its neighbors by looking at monodromy invariants. So even if we don't understand $S - U$, we can try understand U near a point of S, and see what happens. The idea of vanishing cycles is that there are cycles that make sense in U, but as we move towards the exceptional point they get smaller and smaller until they vanish.

In the étale world, here's an example of this showing up.

[2]Translation: "unscrewing". The English word *vise* is related.

Example 17.5. Say X is a curve, and $\mathcal{E}$ is a lisse sheaf on $X - \{s\}$ for some point s, so that $\mathcal{E}$ corresponds to a representation of $\pi_1(X - \{s\})$. There is a surjection $\pi_1(X - \{s\}) \to \pi_1(X)$, but this won't be injective as we have covers coming from looping around s. Picking a geometric base point $\overline{x}$, we have an exact sequence

$$1 \to I_s \to \pi_1(X - s, \overline{x}) \to \pi_1(X, \overline{x}) \to 1$$

where I_s is defined by this sequence and called the *inertia group at* s. The number theory analogue is the inertia group in Galois theory. Roughly speaking, I_s is keeping track of the new cohomology coming from loops around s. (As an aside, this is closely related to the discussion of missing Euler factors in the L-function associated to a variety over a number field, from Remark 9.11.)

Remark 17.6. The p-adic analogue of inertia is much closer to differential geometry: it's related to actual monodromy of differential equations. Because a p-adic coefficient object is a vector bundle with connection, you can actually try to solve differential equations using power series (possibly after adjoining some extra ring elements, as in differential Galois theory) and study monodromy that way.

17.3. **Grothendieck-Ogg-Shafarevich.** If you're trying to make some kind of estimate about a zeta function, you need to control the dimensions of the spaces $H^i(X, \mathcal{F})$. It's therefore good to know things about Euler characteristics, as these are easier to control but still retain some of the needed information.

Definition 17.7. Let X be a curve over k with smooth compactification $\overline{X}$. Let $\mathcal{F}$ be a lisse sheaf (or a p-adic coefficient) on X. We define the *Euler characteristic* to be

$$\chi(\mathcal{F}) = \sum_{i=0}^{2} (-1)^i \dim H^i(X, \mathcal{F})$$

and as in normal cohomology, these are well behaved (eg, additive in short exact sequences, etc).

We'd expect this to be related to the Euler characteristic of X, which is $\chi(X) = \chi(\overline{\mathbb{Q}}_\ell)$, and should also be related to $\mathcal{F}$ somehow. A natural guess is

$$\chi(\mathcal{F}) = \chi(X) \operatorname{rank}(\mathcal{F})$$

but this formula is missing a correction factor as we will see in the following example.

Example 17.8. Let $f : Y \to X$ be a finite étale morphism of curves and put $\mathcal{F} = f_* \overline{\mathbb{Q}}_\ell$. Then $\chi(\mathcal{F}) = \chi(Y)$, which doesn't agree with our guess! There's a correction factor given by Riemann-Hurwitz, coming from the ramification at points of $\overline{X} - X$. These factors can be computed locally, so we can look at them one at a time. In the étale case, this is essentially the Artin conductor of the inertia representation, which detects only what is happening at a single bad point.

This suggests that the shape of the correct formula is

$$\chi(\mathcal{F}) = \chi(X) \operatorname{rank}(\mathcal{F}) + \sum_{x \in \overline{X} - X} (\text{correction at x})$$

where the correction term depends only on what is happening at x (say, on the formal completion at x). This is true, but we will not give a more precise formulation here.

17.4. **Deligne's study of weights on curves.** We now discuss, in very sketchy terms, how to use the above tools to get to Weil II.

Remark 17.9. Let us recall how weights were defined in the previous lecture. Let X be a curve over $k = \mathbb{F}_q$ and fix an embedding $\iota : \overline{\mathbb{Q}}_\ell \hookrightarrow \mathbb{C}$ or $\iota : \overline{\mathbb{Q}}_p \hookrightarrow \mathbb{C}$. Given a lisse $\overline{\mathbb{Q}}_\ell$-sheaf $\mathcal{F}$ on X, we say that $\mathcal{F}$ is *ι-pure of weight w* if for all $x \in X$, the eigenvalues of Frob on $\mathcal{F}_x$ have ι-absolute values $\#\kappa(x)^{w/2}$. We say that $\mathcal{E}$ is *ι-mixed of weight $\leq w$ or $\geq w$* if a corresponding condition holds.

It is not clear that one expects to be able to impose much effective control on these weights: we are manipulating ℓ-adic objects and attempting to keep track of archimedean information, and the two are not very compatible. However, Deligne proves a key theorem that imposes some control on the situation.

Theorem 17.10. *Suppose $\mathcal{F}$ is ι-*real *(that is, all of the Frobenius charpolys have coefficients in $\iota^{-1}(\mathbb{R})$). Then the irreducible subquotients (using Jordan-Holder filtrations because we're in an abelian category) of $\mathcal{F}$ are each ι-pure of some weight.*

Corollary 17.11. The same is true if $\mathcal{F}$ is a subquotient of something ι-real. (We say $\mathcal{F}$ is *ι-realizable* in this case.)

Somehow, if you are able to force the real numbers into the picture, really nice things happen. The idea comes from an argument of Rankin about modular forms, a real analysis argument which boils down to the fact that squares of real numbers are nonnegative. This is useful for us because the next lemma says that any pure coefficient can be written as a subquotient of something real. So studying real coefficients isn't as arbitrary as it may seem.

Lemma 17.12. If $\mathcal{F}$ is ι-pure of weight 0, then $\mathcal{F}^\vee \oplus \mathcal{F}$ is ι-real. More generally, if $\mathcal{F}$ is ι-pure of some other weight, then some twist of $\mathcal{F} \oplus \mathcal{F}^\vee$ is ι-real.

Proof. In the weight 0 case, all of the Frobenius eigenvalues have complex norm 1, so live on the unit circle. The coefficients of the characteristic polynomials are all symmetric functions in the Frobenius eigenvalues, so they will be real if the set of eigenvalues is stable under complex conjugation. Because we're living on the unit circle, this is the same as the eigenvalues being stable under inverses. The eigenvalues of $\mathcal{F}^\vee$ are precisely the inverses of the eigenvalues of $\mathcal{F}$, and the eigenvalues of $\mathcal{F}^\vee \oplus \mathcal{F}$ is the disjoint union of the eigenvalues of $\mathcal{F}^\vee$ and $\mathcal{F}$, so we are all set. □

To prove Theorem 17.10, we first guess what the weight should be, then use this guess to prove that things actually work. Let $\mathcal{E}$ be an irreducible subquotient of $\mathcal{F}$ of rank r. We want to show that $\mathcal{E}$ is ι-pure of some weight w. If this were true, then $\wedge^r \mathcal{E}$ would be an ι-pure, rank 1 object of weight rw. The following key lemma will let us understand rank 1 objects nicely.

Lemma 17.13. All rank 1 sheaves are ι-pure.

The proof uses geometric class field theory. Any rank 1 coefficient on a curve corresponds to a character, which can be explicitly written as a constant times a finite order character. So any eigenvalue will be a constant times a root of unity. As roots of unity always have weight 0, the weights of the eigenvalues are all just given by the weight of the constant, so the coefficient must be pure.

So we understand rank-1 objects well, and we can therefore guess what the weight of $\mathcal{F}$ must be:

Definition 17.14. The *determinantal weight* of a rank r sheaf $\mathcal{F}$ is $1/r$ times the weight of the determinant sheaf $\wedge^r \mathcal{F}$.

The hard part is then to show that the determinantal weights of $\mathcal{F}$ actually behave like weights with respect to operations like $\oplus$. This takes plenty of work, but eventually we can find some inequality between our guess and reality, then use positivity and duality to flip the inequality and get things on the nose.

17.5. **ℓ-adic Fourier transforms.** Then you combine this theorem with a Fourier transform construction. The key case is for $\mathbb{A}^1$, because we're just working with curves, and you can use the following trick in characteristic p. If you take $x \mapsto x^p + 1/x$, this gives a finite étale cover $\mathbb{G}_m \to \mathbb{A}^1$ of degree $p+1$. This type of thing lets us shove all of the missing points to a single point at ∞, so if we were thinking about $\mathbb{P}^1 - s_1, \ldots, s_k$, we can replace it with $\mathbb{A}^1$.

Remark 17.15. The previous argument shows for example that Belyi's theorem goes out the window in positive characteristic, unless one does something like restrict to tamely ramified maps.

Now the really rough idea of Fourier transforms is to take a function $f(x)$, multiply it by $e^{-2\pi i x\eta}$, and integrate with respect to x to get a new function in terms of η. In more geometric terms, you start with a function on $\mathbb{R}$, pull back to $\mathbb{R} \times \mathbb{R}$, twist by a biadditive character, then project on the second factor.

Translating this idea into our language, we'll start with a coefficient object on $\mathbb{A}^1$, pull it back to $\mathbb{A}^1 \times \mathbb{A}^1$ along the first projection, twist by the Artin-Schreier cover to get a family of coefficients that we mostly understand, then project onto the second factor. Because we rigged our cover so that we understand all of the coefficients in the family besides the original one, the fibers over this second projection are copies of $\mathbb{A}^1$

with coefficient objects that we understand away from a single point. This is the kind of thing that Deligne's theory of weights is good at dealing with, so we're now in a good situation; we then use nearby cycles to recover information about the original sheaf. (It is in this last step that we are forced to get something mixed rather than pure.)

Remark 17.16. In the p-adic world, a coefficient looks like a module over a Weyl algebra, a noncommutative ring containing both multiplication by a coordinate x and differentiation $\frac{d}{dx}$ in the same coordinate. There, the Fourier transform can be effected by interchanging these two variables (up to a sign).

Remark 17.17. A slogan for this argument is that one doesn't prove Weil II for a single sheaf in isolation. Instead, one proves something about a whole collection of sheaves at once.

18. Causal versus random: the Tate conjecture and equidistribution (December 9)

In this last lecture, we talk about two different-sounding things that we'll see are actually related: the Tate conjecture, and distribution problems. Roughly speaking, we'll see that the Tate conjecture is a question about the "causal" factors of zeta functions, whereas distribution problems are questions about "random" factors of zeta functions. The idea is that zeta functions should be made up of these two parts, the first coming from the geometry of the variety, and the second being something that we can (sometimes) show is really random in a suitable sense.

Readings 18.1. Since this lecture covers many disparate topics, suggestions for additional reading have been embedded in the text as appropriate.

18.1. **The Hodge conjecture.** The Tate conjecture is an analogue of the Hodge conjecture, so we'll start with that.

Definition 18.2. Let $X/\mathbb{C}$ be a smooth, proper variety of dimension n, and look at the singular cohomology $H^i(X^{\mathrm{an}}, \mathbb{C})$ of the associated complex analytic space. (From now on we'll sloppily write X for X^{an}.) We know that $H^i(X, \mathbb{Q})$ injects into $H^i(X, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} = H^i(X, \mathbb{C})$; that is, the $\mathbb{C}$-vector space contains a lattice which remembers which classes are rational.

We also have the Hodge decomposition

$$H^i(X, \mathbb{C}) \cong \bigoplus_{p+q=i} H^{p,q}, \qquad H^{p,q} := H^p(X, \Omega^q).$$

Let $Z \hookrightarrow X$ be a closed irreducible subvariety of pure codimension p. Via the cycle class map (i.e., by viewing Z as representing a homology class and then dualizing), Z gives rise to a class in $H^{2p}(X, \mathbb{Q}) \cap H^{p,p}(X)$.

Conjecture 18.3 (Hodge conjecture)**.** The intersection $H^{2p}(X, \mathbb{Q}) \cap H^{p,p}(X)$ is spanned by classes coming from subvarieties.

Not much is known about the Hodge conjecture.

Theorem 18.4 (Lefschetz (1,1) theorem)**.** *The Hodge conjecture holds for $p = 1$.*

Proof. See [50, Page 163] for a proof, and a lot of background. □

Example 18.5. Let $A/\mathbb{C}$ be an abelian variety. By the Lefschetz (1,1) theorem, $H^{1,1}(A)$ can be described using endomorphisms of A. In particular, if A has trivial endomorphism ring, then $H^{1,1}$ is 1-dimensional.

Remark 18.6. Besides the $p = 1$ case, not much is known about the Hodge conjecture. It is far from clear "where to look" for a subvariety corresponding to a particular (p, p)-class.

One case that can be handled is the case $p = 2$ when X is a *K3 surface*, which is to say a smooth projective surface such that $K_X \simeq \mathcal{O}_X$ and which is simply connected (say, in the sense that every geometrically connected finite etale cover of X splits).

For a K3 surface, the weight-2 Hodge structure in this case can be embedded into the square of a weight-1 Hodge structure coming from a certain abelian variety, via the *Kuga-Satake construction*. Using a similar construction, Deligne was able to prove the Weil conjectures for K3 surfaces before coming up with the general proof [28].

18.2. **The Tate conjecture.**

Definition 18.7. Let X be a smooth proper scheme over a finite field k. As in the complex case, for any codimension-i subvariety Z on X, the cycle class map gives us a class in $H^{2i}(X)$. By the Weil conjectures, this class is a Frobenius eigenvector with eigenvalue q^i.

The Tate conjecture is now the natural analogue of the Hodge conjecture.

Conjecture 18.8. The generalized q^i-eigenspace of $H^{2i}(X)$ is spanned by cycles coming from codimension-i subvarieties of X.

Remark 18.9. Note that we don't a priori know that Frobenius gives a diagonalizable matrix, so part of the conjecture is a semisimplicity statement. We can rephrase this to get a concrete prediction about zeta functions. In this language, the Tate conjecture says that

$$Z(X,T) = \frac{\cdots}{\cdots \det(1 - FT, H^{2i}(X)) \cdots}$$

has a pole at $T = q^{-i}$ of order equal to the dimension of the space spanned by the codimension i cycle classes. So given a variety, we can write down zeta functions explicitly, look at its poles, and get a prediction about cycle classes which would generally be hard to find by hand.

Remark 18.10. The Tate conjecture tends to be as hard as the Hodge conjecture. It's known for $i = 1$ for abelian varieties (this is equivalent to Tate's theorem) and for K3 surfaces by a recent result of Ito–Ito–Koshikawa [61].

The K3 case is already extremely hard and interesting, so let's look a little at it. For X a K3 surface,

$$Z(X,T) = \frac{1}{(1-T)P(T)(1-q^2T)}.$$

Here the outside terms of the denominator come from H^0 and H^4, and the inside two come from H^2, which is the interesting bit. Renormalizing to make the roots lie on the unit circle, we have $Q(T) := qP(q^{-1}T) = a_0T^{22} + a_1T^{21} + \cdots + a_{22}$ with $a_0 = a_{22} = q$, and we are interested in the multiplicity of the factor $(1 - T)$ in Q.

This renormalization introduces lots of powers of q in the denominators, so one might expect Q to no longer be integral. However, there is a crucial piece of information that we have not yet introduced in these lectures.

Proposition 18.11. The coefficients a_i of $Q(T)$ are all integers.

Sketch. This comes from Mazur's Newton above Hodge theorem [83]. In this case, the theorem says that the Newton polygon of $\det(1 - FT, H^2(X))$ lies above the Hodge polygon, which has integer slopes given by the second row of the Hodge diamond. This row is 1,20,1, so the Hodge polygon has slope 0 for one step, slope 1 for 20 steps, and slope 2 for the final step. Renormalizing and scaling to write this in terms of Q, the polygon starts at $(0, 1)$, goes down to $(1, 0)$, goes horizontally to $(21, 0)$, then goes up to $(22, 1)$. In particular, it never dips below the x-axis, so the coefficients of the Newton polygon must all have nnnegative q-adic valuation and therefore will be integers. □

Remark 18.12. Continuing with our K3 example, we also have a symmetry property $a_{22-j} = \pm a_j$ (where this sign is uniform over j). So we either have actual symmetry, or there's a sign flip after we cross the middle. It is possible but tricky to understand which of these actually happens using the geometry of the K3 surface.

As we are looking at H^2, we are in the case where $i = 1$, so our cycles are divisors. The well-studied Néron-Severi group $\mathrm{NS}(X)$ is the group of divisors modulo algebraic equivalence. Rephrasing the conjecture one more time, we are saying that the order of the zero of Q at 1 is equal to the Picard number of X, a/k/a the rank of $\mathrm{NS}(X)$. Call this order r; it must be an integer between 1 and 22, the 1 because there is automatically a Tate class corresponding to an ample divisor, and the 22 because this is the dimension of H^2 (see following remark). After renormalizing, we're looking for the order of vanishing of $Q(T)$ at $T = 1$. Call this order r; we have the *Artin-Tate formula* (a conjecture in general, but known for K3 surfaces)

$$\frac{Q(T)}{(1-T)^r}|_{T=1} = D\#\,\mathrm{Br}(X)$$

where D is the discriminant of the Néron-Severi lattice and $\mathrm{Br}(X)$ is the Brauer group: a finite group with order a perfect square.

This should remind us of the conjecture of Birch–Swinnerton-Dyer. It actually coincides with it in certain cases: when X is an *elliptic* K3 surface, the Brauer group coincides with the Tate-Shafarevich group. In general, just like the latter, the Brauer group carries an alternating pairing which forces its order to be a square.

Remark 18.13. In characteristic 0, the Picard number can only go up to 20, but in characteristic p the value 22 is actually possible! This is similar to the fact that the endomorphism ring of an elliptic curve in characteristic 0 has rank at most 2, but in characteristic p it can have rank as high as 4.

Remark 18.14. A nice thing here is that if we're handed a K3 surface over a small finite field, we can compute this polynomial, get our hands on r and make a guess about what the constants should be. For example, if $r = 1$ and X is a smooth quartic in $\mathbb{P}^3$, then $D = 4$ and so the right-hand side should be a square. In [72], Kedlaya–Sutherland checked that this is always true over $\mathbb{F}_2$.

Remark 18.15. This isn't quite the whole story; so far we've been focused on cycles that are defined over $\mathbb{F}_q$. If we pass to a finite extension, we might get more cycles, and the Tate conjecture would tell us that they should show up in the zeta function as well. For example, if $1+T$ divides $Q(T)$, then $-q^{-1}$ is an eigenvalue, and if we base change to $\mathbb{F}_{q^2}$, we square the eigenvalues and get an eigenvalue of q^{-2}. So the Tate conjecture is also saying that eigenvalues of $q^{-\zeta}$ for any root of unity ζ are "causal": once we base-change, they should also come from cycles. We should therefore be thinking about the factorization of $Q(T)$ into cyclotomic factors and noncyclotomic factors. The cyclotomic factors tell us about the geometric Néron-Severi rank.

There's a nice heuristic about point counting underlying all of this. Going back to the general Tate conjecture, any codimension-i subvariety Z of X will make some "geometric" contribution to the number of $\mathbb{F}_{q^r}$ points on X, coming from the $\mathbb{F}_{q^r}$ points on Z. The (unnormalized) factor of $(1-q^iT)$ in the denominator is just keeping track of this contribution. So if X has lots of codimension-i subspaces, we expect it to have more rational points than usual, giving rise to a larger pole in the zeta function. This heuristic seems very similar to the heuristic that led to the Birch–Swinnerton-Dyer conjecture: if an elliptic curve over a number field has high rank, then its reductions modulo primes are forced to have lots of points, which should again lead to a large pole of the L-function at $s = 1$.

Remark 18.16. For a final application in this section, we explain the key idea behind constructing a K3 surface over $\mathbb{Q}$ with geometric Picard number 1. This construction is due to van Luijk [108] and answers a question of Mumford.

If one starts with a K3 surface over $\mathbb{Q}$, its geometric Picard number can only *increase* under specialization, as the Néron-Severi lattice of a characteristic 0 K3 surface injects into the Néron-Severi lattice of a reduction. So in principle, we could try to prove that the geometric Picard number is 1 by reduction to a finite field. But there's a catch: the polynomial Q has integer coefficients and its degree is even (22), so its geometric Picard number is always even (the noncyclotomic part necessarily has even degree).

This seems to be the end of the story, until we realize (as van Luijk did) that we can apply the Artin-Tate formula at various different primes of good reduction and compare the answers. To wit, van Luijk constructs a family of K3 surfaces over $\mathbb{Z}$ whose reductions at 2 and 3 both have geometric Picard number 2 (the smallest possible value given the previous discussion). Using the Artin-Tate formula, he shows that the discriminants of the lattices in characteristics 2 and 3 are -12 and -9, which represent different elements of $\mathbb{Q}^*/\mathbb{Q}^{*2}$. But if the Néron-Severi lattice over $\mathbb{Z}$ were 2-dimensional, its discriminant would be the same modulo squares as each of these, as it would be a sublattice of full rank. As this cannot happen, the geometric Picard number must be 1.

For more results about the variation of Picard numbers under specialization, see [25, 20, 23].

18.3. **Distribution questions.** Individual zeta functions can be unpredictable, but we can make headway looking at distributions of lots of them. Here are three different flavors of questions that people study.

(1) Fix a finite field $\mathbb{F}_q$ and look at a class of varieties $\{X\}$ over $\mathbb{F}_q$. The zeta function of each variety is related to the number of points, so we can consider $\#X(\mathbb{F}_q)$ (or something related) as a random variable on the probability space of all such X.

(2) Look at a *geometric* family of varieties, i.e., look at the fibers of a map $X \to S$ over closed points of S where both X and S are varieties over $\mathbb{F}_q$. Now that we're looking over all closed points, we'll also be counting $\mathbb{F}_{q^r}$ points.
(3) Look at the same question for an *arithmetic* family $X \to \mathrm{Spec}(\mathcal{O}_K)$ for a number field K.

For an example of the first flavor, look at smooth plane curves over $\mathbb{F}_q$. We want to understand $\#X(\mathbb{F}_q)$ viewed as a random variable on the probability space of all such X. For a fixed degree d, this is a finite probability space as there are finitely many curves, so we should really average over all d in some fashion. We take all d up to some bound, compute the distribution, then take the limit as the bound goes to infinity.

Theorem 18.17. *The resulting distribution is a sum of $q^2 + q + 1$ individually independently distributed 0,1-random variable with total mean $q + 1$.*

Sketch. This is an application of Poonen's Bertini theorem [93] by Bucur-David-Feigon-Lalín [16]. Here's where this is coming from. The quantity $q^2 + q + 1$ is the number of points on $\mathbb{P}^2(\mathbb{F}_q)$. We can think of each point as a variable, and ask if that point is a rational point of a given plane curve X. At the point 0, we can locally expand out our curve as being cut out by the equation $a + bx + cy + \dots$. So 0 is on the curve exactly when $a = 0$, and it's a smooth point if $a = 0$ and $bc \neq 0$. If we exclude the case where $a = b = c = 0$ (by sieving), all other possibilities are equally likely. Now we have $q^3 - 1$ total possibilities, of which $q^2 - 1$ are good. So the probability that a given point is on a random curve is $(q + 1)/(q^2 + q + 1)$, and one uses Poonen's theorem to ensure that each point contributes independently to the count. □

Remark 18.18. Poonen's Bertini theorem asserts that given a smooth quasiprojective variety X, the probability that an ample hypersurface section of X is predicted by a product of local probabilities, each computing the probability that there is no failure of smoothness at a given point. It has spawned a sizable literature concerning questions of a similar flavor. Two notable examples are the papers of Bucur–Kedlaya [17], which extends Poonen's theorem by considering a complete intersection of multiple hypersurfaces (this came up previously in Remark 6.17), and of Erman–Wood [40], which allows the use of semiample hypersurfaces (at the cost of some degree of independence between points).

Remark 18.19. The standard reference for questions of type 2 is the book of Katz-Sarnak [63]. We won't talk much about these today, except to say that they can generally be settled by combining Weil II with a computation of a certain *monodromy group* attached to the family of varieties. See Theorem 18.27 below for an example.

It's generally harder to prove anything for questions of type 3 than type 2, but we expect similar answers. In each case, assuming X is smooth and proper, we have $Z(X, T) = \prod L_i(T)^{(-1)^{i+1}}$ where $L_i(T)$ is pure of weight i. We normalize to get $\overline{L}_i(T) = L_i(q^{-i/2}T)$, which has eigenvalues on the unit circle.

Philosophy 18.20. We expect the $\overline{L}_i(T)$ to behave like characteristic polynomials of *random* matrices in a certain compact Lie group G. Here randomness is measured with respect to the (unique) Haar measure on G.

Remark 18.21. This philosophy predates people thinking about finite fields. It was originally introduced to think about the Riemann zeta function, based on the idea (attributed to Pólya) that the zeroes of ζ should be (up to rotation) the eigenvalues of some self-adjoint operator on some Hilbert space. Since we have no idea what this operator should look like, we might hope that it behaves like a "random" operator, and indeed evidence (from Montgomery, Odlyzko, and others) suggests that the distribution of zeroes of ζ does have some features in common with the eigenvalues of suitable random matrices.

An important question of type 3 is the Sato-Tate conjecture (now a theorem) over $\mathbb{Q}$.

Definition 18.22. To formulate the Sato-Tate conjecture, let $E/\mathbb{Q}$ be an elliptic curve. For p a prime of good reduction, let a_p be the trace of Frobenius on $E_{\mathbb{F}_p}$; the Hasse bound says that $a_p \in [-2\sqrt{p}, 2\sqrt{p}]$, so we divide by $\sqrt{p}$ to renormalize a_p. This lets us compare the values of a_p as p varies over all primes of good reduction.

Theorem 18.23. *Given an elliptic curve $E/\mathbb{Q}$, as p varies over all primes of good reduction, the $a_p/\sqrt{p}$ are equidistributed (see below) in $[-2, 2]$ with respect to one of the following measures:*

- *If E has CM, the trace of a random matrix in $N(\mathrm{SO}(2), \mathrm{SU}(2))$ (the normalizer);*
- *else, the trace of a random matrix in* $\mathrm{SU}(2)$.

Remark 18.24. For a gif of this theorem in action, see `https://math.mit.edu/~drew/g1_r28_a1f.gif`. See also `https://math.mit.edu/~drew/g1SatoTateDistributions.html` for additional examples.

Definition 18.25. The notion of equidistribution comes from ergodic theory. Let X be a measure space with measure μ, and let $x_1, x_2, \ldots$ be a sequence in X. We say this sequence is *equidistributed* if for all continuous functions $f : X \to \mathbb{R}$, we have

$$\int_\mu f = \lim_{N\to\infty} \frac{f(x_1) + \cdots + f(x_N)}{N}.$$

The intuition here is that the left-hand side of this equation represents a "space average" of the function f, while the right-hand side represents a "time average" over a sequence of sample points.

This is now a (very very hard) theorem of Clozel–Harris–Taylor [21], Harris–Shephard-Barron–Taylor [58], and many more using automorphic forms. These show up because the relevant X comes from taking a Lie group G modulo conjugation, which gives it a natural measure coming from the Haar measure on G. When we look at a distribution problem on a space of conjugacy classes, it suffices to test the equidistribution property for $f = \chi$ an irreducible character by the Peter-Weyl theorem. For these, an argument of Serre (inspired by the prime number theorem; see [97, Chapter I, Appendix]) shows that the limiting property follows from analytic continuation of suitable L-function.

We could ask the same question over other number fields K. There's a similar conjecture, except if K contains an imaginary quadratic field M, there's a third option when E has CM in M. Then the random matrices are in $\mathrm{SO}(2)$. The CM cases can be settled using Hecke's theory of Grossencharacters, so the real issue is the non-CM cases.

Theorem 18.26. *The analogue of the Sato-Tate conjecture is known when K is either a totally real number field, or a CM field (a totally imaginary quadratic extension of a totally real field, such as an imaginary quadratic field).*

Proof. The first case was settled by Barnet-Lamb–Geraghty–Gee [7]. The second case was settled by the "paper of 10 authors" [3]. □

For comparison, let us also formulate the corresponding type 2 question; this amounts to looking at all elliptic curves over finite extensions of some $\mathbb{F}_q$.

Theorem 18.27. *Let $f : X \to S$ be a family of elliptic curves over a finite-type $\mathbb{F}_q$-scheme, and assume that the j-invariant is nonconstant in the family (that is, the family is* nonisotrivial*). Then as s varies over closed points of S, if we write a_s for the trace of Frobenius on $f^{-1}(s)$, the quantities $a_s/\sqrt{\#\kappa(s)}$ are equidistributed with respect to the distribution of traces of random matrices in* $\mathrm{SU}(2)$.

Proof. This is proved by Deligne in his Weil II paper [30, Section 3.5] □

Remark 18.28. In more general cases, you figure out what Lie group to use (either conjecturally or provably) by looking at the image of monodromy in the case of geometric families, or the image of Galois in the case of arithmetic families.

In the number field case, you look at the associated Galois representation; look at how big the image might possibly be; check that you aren't missing constraints that could make the image smaller; and then hope that this actually is the image. For example, for CM elliptic curves, the image of Galois must respect the endomorphisms and is thus much smaller than for non-CM elliptic curves; this is then reflected in the distribution of the Frobenius traces.

By looking carefully at the constraints involved, one can sometimes identify all possible candidates for the distribution and for the compact Lie group (the *Sato-Tate group*) giving rise to it. For example, if X/K is a genus 2 curve over a number field, or an abelian surface, there are 52 possible Sato-Tate groups [42], and one can easily distinguish the resulting distributions numerically. For an abelian threefold, there are 410 possible groups [43]. For further discussion of the Sato-Tate conjecture and its generalizations, see [103].

18.4. **Tying everything together.** We'll end with a remarkable application of random matrix theory to questions about point counting, which ties together our discussions of causality and randomness in zeta functions.

Remark 18.29. We start with a remarkable fact from probability theory due to Diaconis-Shahshahani [32]. Consider the k-th moment of the trace of a random matrix in the unitary group $\mathrm{U}(n)$. If we fix k and send n to ∞, we might expect the moment to grow; after all, we're taking the trace of a really big matrix! But this doesn't happen: the k-th moment *stabilizes* for $n \geq k$. In particular, a random matrix has bounded trace! Similar results hold for orthogonal and symplectic groups.

What does this mean for zeta functions? Let X be a smooth proper scheme over $\mathbb{F}_q$, and factor the i-th piece of the zeta function $L_i(T) = \det(1 - FT, H^i(X))$ into a causal part (the Tate classes if i is even, otherwise nothing) and a random part $P_r(T)$ (everything else). Then if the renormalized polynomial $P_r(q^{-i/2}T)$ really corresponds to a random matrix (say in the unitary group U or the unitary symplectic group $\mathrm{USp}(n)$, the trace should be fairly small even if the degree is large. It's therefore reasonable to expect that as the degree gets big, the zeta function will be dominated by the causal factors.

One application of this logic is a heuristic prediction about the distribution of the number of rational points of $\#X(\mathbb{F}_q)$ as X varies over all curves of a given genus [2]. This prediction involves point counts on $M_{g,n}$, which has a causal part (the stable cohomology) and a non-causal part (the unstable cohomology). If we predict that the unstable part should act randomly, then the causal part will dominate.

Exercises

Set 1.

(1) Let k be a finite field of order q and fix an additive character (homomorphism) $\psi : k \to \mathbb{C}^\times$. For $\chi : k^\times \to \mathbb{C}^\times$ a nontrivial multiplicative character, define the Gauss sum

$$G_\psi(\chi) = \sum_{x \in k^\times} \chi(x)\psi(x).$$

Prove that $G_\psi(\chi)G_\psi(\overline{\chi}) = q$, where $\overline{\chi}$ is the character for which $\overline{\chi}(x)$ is the complex conjugate of $\chi(x)$. (Hint: write the product as a sum over $x, y \in k^\times$, then regroup terms by the value of x/y.)

(2) Fix a choice of χ as above. For $P(T) = T^n + P_{n-1}T^{n-1} + \cdots + P_0 \in k[T]$ a monic polynomial, define

$$\lambda(P) = \chi(P_0)\psi(P_{n-1}).$$

(In particular, $\lambda(1) = 1$.) Show that

$$\lambda(P_1P_2) = \lambda(P_1)\lambda(P_2) \qquad (P_1, P_2 \in k[T])$$

and deduce that for each positive integer n, in $\mathbb{C}[\![U]\!]$ we have

$$\sum_{P \in k[T] \text{ monic}} \lambda(P)U^{\deg(P)} = \prod_{Q \in k[T] \text{ monic irreducible}} (1 - \lambda(Q)U^{\deg Q})^{-1}.$$

(3) Show that for n a nonnegative integer,

$$\sum_{P \in k[T] \text{ monic}, \deg(P) = n} \lambda(P)U^{\deg(P)} = \begin{cases} 1 & n = 0 \\ G_\psi(\chi)U & n = 1 \\ 0 & n > 1. \end{cases}$$

(4) With notation as in the previous problem, let k' be an extension of k of degree v. Let $\psi' : k' \to \mathbb{C}^\times$ be the additive character given by $\psi \circ \mathrm{Trace}_{k'/k}$. Given χ, let χ' be the multiplicative character given by $\chi \circ \mathrm{Norm}_{k'/k}$. For $P' \in k'[T]$ monic, define λ' by analogy with λ.

For $P \in k[T]$ monic irreducible, let P' run over the irreducible factors of P in $k'[T]$. Prove that

$$\prod_{P'} (1 - \lambda'(P')U^{v \deg(P')}) = \prod_{\rho=0}^{v-1} (1 - \lambda(P)(e^{2\pi i\rho/v}U)^{\deg(P)}).$$

(Hint: let $-\xi$ be a root of one of the factors P', and consider the field extensions $k(\xi)/k$ and $k'(\xi)/k'$.)

(5) Using all of the above, deduce the Davenport-Hasse relation

$$-G_{\psi'}(\chi') = (-G_\psi(\chi))^v.$$

Set 2. Throughout, let $\mathbb{F}_q$ denote a finite field of characteristic p.

(1) For X an algebraic variety over $\mathbb{F}_q$, we write the zeta function of X as $Z(X, q^{-s})$ for

$$Z(X,T) = \prod_{x \in X^\circ} (1 - T^{\deg(x)})^{-1},$$

where X° denotes the set of Galois orbits of $\overline{\mathbb{F}_q}$-points and $\deg(x)$ is the cardinality of such an orbit. Prove that in $\mathbb{Q}[\![T]\!]$, we have the equality

$$Z(X,T) = \exp\left(\sum_{n=1}^{\infty} \frac{T^n}{n} \#X(\mathbb{F}_{q^n})\right).$$

(2) For X equal to the n-dimensional projective space over $\mathbb{F}_q$, compute that

$$Z(X,T) = \frac{1}{(1-T)(1-qT)\cdots(1-q^nT)}.$$

(3) Prove that the following statements are equivalent.
   (i) The power series $Z(X,T)$ represents a rational function in T.
   (ii) There exist $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s \in \mathbb{C}$ such that

$$\#X(\mathbb{F}_{q^n}) = \alpha_1^n + \cdots + \alpha_r^n - \beta_1^n - \cdots - \beta_s^n \qquad (n = 1, 2, \dots).$$

(4) Let X be the *Grassmannian* of k-dimensional subspaces of m-space over $\mathbb{F}_q$.
   (i) Compute $\#X(\mathbb{F}_{q^n})$; your answer should be a polynomial in q^n depending on k and m. (Hint: count bases of subspaces, then divide by the number of bases of a given subspace.)
   (ii) Compute $Z(X,T)$.

(5) Choose $a_0, \dots, a_r \in \mathbb{F}_q^\times$. For d a positive integer dividing $q-1$, let X_d be the projective hypersurface $a_0 x_0^d + \cdots + a_r x_r^d = 0$.
   (i) Let G_d be the group of homomorphisms $\chi : \mathbb{F}_q^\times \to \mathbb{C}^\times$ of order d. For $\chi \in G_d$, extend the definition of χ to $\mathbb{F}_q$ by setting $\chi(0) = 1$ if $\chi = 1$ and $\chi(0) = 0$ otherwise. Show that

$$1 + (q-1)\#X_d(\mathbb{F}_q) = \sum_{(u_0,\dots,u_r) \in X_1} \sum_{\chi_0,\dots,\chi_r \in G_d} \prod_{i=0}^{r} \chi_i(u_i).$$

   (ii) Show that $\chi_0, \dots, \chi_r \in G_d$ are neither all equal to 1 or all distinct from 1, then

$$\sum_{(u_0,\dots,u_d) \in X_1} \prod_{i=0}^{r} \chi_i(u_i) = 0.$$

   (iii) Let T be the set of tuples $(\chi_0, \dots, \chi_r) \in G_d \setminus \{1\}$ with $\chi_0 \cdots \chi_r = 1$. For $(\chi_0, \dots, \chi_r) \in T$, define the *Jacobi sum*

$$j(\chi_0, \cdots \chi_r) = \frac{1}{q-1} \sum_{u_0,\dots,u_r \in \mathbb{F}_q : u_0 + \cdots + u_r = 0} \chi_0(u_0) \cdots \chi_r(u_r).$$

   Deduce from above that

$$\#X_d(\mathbb{F}_q) = 1 + q + \cdots + q^{r-1} + \sum_{(\chi_0,\cdots,\chi_r) \in T} \chi_0(a_0^{-1}) \cdots \chi_r(a_r^{-1}) j(\chi_0, \dots, \chi_r).$$

   (iv) Fix an additive character $\psi : \mathbb{F}_q \to \mathbb{C}^\times$. Show that

$$j(\chi_0, \dots, \chi_r) = \frac{1}{q} G(\chi_0, \psi) \cdots G(\chi_r, \psi)$$

   where $G(\chi, \psi)$ denotes the Gauss sum.

(6) Keep notation as in the previous exercise, but assume only that d is not divisible by p (not that it divides $q-1$).
   (i) Show that $\#X_d(\mathbb{F}_q) = \#X_e(\mathbb{F}_q)$ for $e = \gcd(d, q-1)$.

(ii) Using the Davenport-Hasse relation, show that the rationality, functional equation, and Riemann hypothesis hold for $Z(X_d, T)$.

Set 3. Throughout, let $\mathbb{F}_q$ denote a finite field of characteristic p. Assume the Weil conjectures for curves and abelian varieties unless otherwise specified.

(1) Let X be a nonzero abelian variety over $\mathbb{F}_q$. Prove that if $q \geq 5$, the group $X(\mathbb{F}_q)$ is nontrivial.

(2) Let X be a curve over $\mathbb{F}_q$ such that $\#X(\mathbb{F}_q) = 1$.
   (a) If $q = 3$ or $q = 4$, prove that
   $$Z(X, T) = \frac{1 - qT + qT^2}{(1 - T)(1 - qT)}.$$
   (b) If $q = 2$, prove that the genus of X is at most 4, and that there are at most 6 possibilities for $Z(X, T)$.
   (c) Optional: show that each of the 8 possibilities occurs for a unique X up to isomorphism.

(3) Let X be an abelian variety of dimension g over $\mathbb{F}_q$. Assuming only the existence of complex numbers $\alpha_1, \dots, \alpha_{2g}$ such that
$$X(\mathbb{F}_{q^n}) = (1 - \alpha_1^n) \cdots (1 - \alpha_{2g}^n) \qquad (n = 1, 2, \dots),$$
compute $Z(X, T)$.

(4) Using the Honda-Tate theorem, prove that if A_1, A_2 are abelian varieties over $\mathbb{F}_q$ and $P_1(A_1, T)$ divides $P_1(A_2, T)$, then A_1 is isogenous to the product of A_2 with some other abelian variety.

(5) Let X be a curve of genus g over $\mathbb{F}_q$. Prove the following refinement of the Weil bound due to Serre:
$$|\#X(\mathbb{F}_q) - q - 1| \leq g\lfloor 2\sqrt{q} \rfloor.$$
Hint: apply AM-GM to the numbers $\lfloor 2\sqrt{q} \rfloor + 1 + \alpha + \overline{\alpha}$ where α runs over the Frobenius eigenvalues.

(6) Let $P(T) = \sum_{i=0}^{2g} a_i T^i$ be a polynomial over $\mathbb{Z}$ such that $a_0 = 1$, $a_{g+i} = q^i a_{g-i}$ for all i, all roots of $P(T)$ in $\mathbb{C}$ lie on the circle $|T| = q^{-1/2}$, *and* a_g is not divisible by p (that is, P is an *ordinary Weil polynomial*). Use the Honda-Tate theorem to show that $P(T)$ occurs as $P_1(A, T)$ for some abelian variety A over $\mathbb{F}_q$ (without raising P to a power).

Set 4.

(1) Let K be a number field. Using the Chebotarëv density theorem, prove that the Frobenius elements corresponding to maximal ideals of $\mathfrak{o}_K$ are dense in the absolute Galois group G_K. (This is just an exercise in unwinding the definitions.)

(2) In this exercise, we prove the theorem of Borel stated in class on November 4.
   (a) Let $f(T) = \sum_{n=0}^{\infty} a_n T^n$ be a power series over an arbitrary field K. Prove that $f(T)$ represents a rational function over K if and only if for some positive integer m, the determinants of the $(m + 1) \times (m + 1)$ matrices $A_{n,m} = (a_{n+i+j})_{i,j=0}^{m}$ vanish for all sufficiently large n.
   (b) Let $f(T) = \sum_{n=0}^{\infty} a_n T^n$ be a power series over $\mathbb{Z}$. Let $r > 0$ be a real number such that over $\mathbb{Q}_p$, there exists a polynomial $P(T)$ of degree $d < m$ such that $P(T)f(T)$ converges for $|T| < r + \epsilon$ for some $\epsilon > 0$. (We do not assume that P has coefficients in $\mathbb{Z}$.) Prove that for some $C > 0$, $|\det(A_{n,m})|_p \leq C r^{-n(m-d)}$ for all n.
   (c) Let $f(T) = \sum_{n=0}^{\infty} a_n T^n$ be a power series over $\mathbb{Z}$. Let R and r be real numbers with $Rr > 1$ such that over $\mathbb{C}$, $f(T)$ converges for $|T| < R$; and over $\mathbb{Q}_p$, $f(T)$ is the ratio of two series that converge for $|T| < r$. Prove that f represents a rational function. (Hint: apply (b) with r replaced by $r - \epsilon$ for which $(R - \epsilon)(r - \epsilon) > 1$, then combine with a trivial bound on $|\det(A_{n,m})|_\infty$.)

(3) Let π be an element of an algebraic closure of $\mathbb{Q}_p$ satisfying $\pi^{p-1} = -p$. (You may use without proof the fact that $\mathbb{Z}_p[\pi]$ is a discrete valuation ring with maximal ideal (π).) Define the power series
$$E_\pi(T) = \exp(\pi(T - T^p)) \in \mathbb{Q}_p(\pi)[\![T]\!].$$
   (a) Prove that $E_\pi(T) \in 1 + \pi\mathbb{Z}_p[\pi][\![T]\!]$.
   (b) Prove that $E_\pi(T)$ has radius of convergence strictly greater than 1. In particular, it makes sense to evaluate it at any element of $\mathbb{Z}_p[\pi]$.

(c) Prove that if $t \in \mathbb{Z}_p$ satisfies $t^p = t$, then $E_\pi(t)^p = 1$. (Hint: check that in the identity

$$E_\pi(T)^p = \exp(\pi p T)\exp(-\pi p T^p)$$

it is valid to substitute t *separately* into the two factors on the right.)

(4) With notation as in the previous problem, let n be a positive integer and define

$$E_n(T) := \exp(\pi(T - T^{p^n})) = E_\pi(T)E_\pi(T^p)\cdots E_\pi(T^{p^{n-1}}) \in \mathbb{Q}_p(\pi)[\![T]\!].$$

Show that the formula $t \mapsto E_n([t])$ defines a nontrivial additive character on $\mathbb{F}_{p^n}$, where $[t]$ denotes the unique element of $\mathbb{Z}_{p^n}$ (the finite étale extension of $\mathbb{Z}_p$ with residue field $\mathbb{F}_{p^n}$) lifting t and satisfying $t^{p^n} = t$.

(5) Set $q = p^n$ and let

$$f = \sum_{I=(i_1,\dots,i_d)} a_I x_1^{i_1}\cdots x_d^{i_d} \in \mathbb{F}_q[x_1,\dots,x_d]$$

be a polynomial. Prove that for any positive integer m, the number of points $(x_1,\dots,x_d) \in (\mathbb{F}_{q^m}^\times)^d$ for which $f(x_1,\dots,x_d) = 0$ equals

$$\frac{(q^m-1)^d}{q^m}\left(1 + (q^m-1)\sum_{x_0,\dots,x_d \in \mathbb{F}_{q^m}^\times}\prod_{I:a_I\neq 0}\prod_{j=0}^{m-1} E_\pi(a_I([x_0][x_1]^{i_1}\cdots[x_d]^{i_d})^{q^j})\right).$$

Set 5.

(1) Define the rings

$$R = \mathbb{Z}[x_1,y_1,x_2,y_2,\dots], \quad R' = \mathbb{Q}[x_1,y_1,x_2,y_2,\dots], \quad F = \mathrm{Frac}(R) = \mathrm{Frac}(R').$$

Define the power series $x = 1 + x_1T + x_2T^2 + \cdots$, $y = 1 + y_1T + y_2T^2 + \cdots$, and

$$f = 1/\exp(\log(1/x) \star \log(1/y)) \in R'[\![T]\!]$$

where $\star$ denotes the *Hadamard product*:

$$(a_1T + a_2T^2 + \cdots) \star (b_1T + b_2T^2 + \cdots) = a_1b_1T + a_2b_2T^2 + \cdots$$

(a) Let V_1, V_2 be two finite-dimensional vector spaces over F equipped with endomorphisms φ_1, φ_2 satisfying, for some positive integer n,

$$\det(1 - \varphi_1 T, V_1)^{-1} \equiv 1 + x_1T + \cdots + x_nT^n \pmod{T^{n+1}F[\![T]\!]},$$
$$\det(1 - \varphi_2 T, V_2)^{-1} \equiv 1 + y_1T + \cdots + y_nT^n \pmod{T^{n+1}F[\![T]\!]},$$

Prove that

$$\det(1 - (\varphi_1\otimes\varphi_2)T, V_1 \otimes_F V_2)^{-1} \equiv f \pmod{T^{n+1}F[\![T]\!]}.$$

(Hint: pass to an algebraic closure of F and write everything in terms of eigenvalues. Remember that f is determined mod $T^{n+1}F[\![T]\!]$ by $x_1,\dots,x_n,y_1,\dots,y_n$.)

(b) Deduce that $f \in R[\![T]\!]$.

(2) Using the previous exercise, prove that there is a unique functor Λ from rings to rings with the following properties.

(a) The underlying functor from rings to additive groups takes R to $\Lambda(R) = 1 + TR[\![T]\!]$ with the usual series multiplication.

(b) For any ring R, the multiplication map $*$ on $\Lambda(R)$ satisfies

$$(1-aT)^{-1} * (1-bT)^{-1} = (1-abT)^{-1} \qquad (a,b \in R).$$

The ring $\Lambda(R)$ is (a form of) the ring of *big Witt vectors* with coefficients in R.

(3) Let X_1, X_2 be two varieties over $\mathbb{F}_q$. Prove that in $\Lambda(\mathbb{Z})$, we have

$$Z(X_1 \times_{\mathbb{F}_q} X_2, T) = Z(X_1,T) * Z(X_2,T).$$

(4) Let K be a field of characteristic 0. Let $P(x) \in K[x]$ be a monic polynomial of degree $2g+1$ with no repeated roots.

(a) Let X be the affine scheme $\operatorname{Spec} K[x,y]/(y^2 - P(x))$. Prove that $\Omega^1_{X/K}$ is freely generated by dx/y. (Hint: it suffices to check that dx/y is a nowhere vanishing section of $\Omega^1_{X/K}$. Treat the points where $y = 0$ and $y \neq 0$ separately.)

(b) Prove that $H^1_{\mathrm{dR}}(X)$ admits the basis

$$x^i \frac{dx}{y} \qquad (i = 0, \dots, 2g-1).$$

(Hint: for each integer $d \geq 2g$, write down a relation of the form $Q(x)dx/y$ with $\deg(Q) = d$.)

(c) Let Y be the affine scheme $\operatorname{Spec} K[x,y,z]/(y^2 - P(x), yz - 1)$. Prove that $H^1_{\mathrm{dR}}(Y)$ admits the basis

$$x^i \frac{dx}{y}, \qquad (i = 0, \dots, 2g-1); \qquad x^i \frac{dx}{y^2} \qquad (i = 0, \dots, 2g).$$

(5) Let $p > 2$ be a prime. Let $\overline{P} \in \mathbb{F}_p[x]$ be a monic polynomial of degree $2g+1$ with no repeated roots.

(a) Put $\overline{X} = \operatorname{Spec} \mathbb{F}_p[x,y]/(y^2 - \overline{P}(x))$. Prove that $H^1_{\mathrm{MW}}(\overline{X})$ admits the basis

$$x^i \frac{dx}{y} \qquad (i = 0, \dots, 2g-1).$$

(b) Put $\overline{Y} = \operatorname{Spec} \mathbb{F}_p[x,y,z]/(y^2 - \overline{P}(x), yz - 1)$. Prove that $H^1_{\mathrm{MW}}(\overline{Y})$ admits the basis

$$x^i \frac{dx}{y}, \qquad (i = 0, \dots, 2g-1); \qquad x^i \frac{dx}{y^2} \qquad (i = 0, \dots, 2g).$$

Supplementary exercises. These exercises were not assigned during the course, but were added subsequently.

(1) Using the Weil conjectures for curves, show that a curve X of genus 1 over a finite field k cannot satisfy $X(k) = \emptyset$.

(2) Using Tate's theorem, show that the zeta function of a curve C over a finite field $\mathbb{F}_q$ is uniquely determined by the sequence

$$\#\operatorname{Jac}(C)(\mathbb{F}_q), \#\operatorname{Jac}(C)(\mathbb{F}_{q^2}), \#\operatorname{Jac}(C)(\mathbb{F}_{q^3}), \dots.$$

Optional (and harder): use the Weil conjectures to show that $O(g)$ terms suffice, where g is the genus of C. See [67].

(3) Let X be a geometrically irreducible variety over a finite field $\mathbb{F}_q$. Using the Weil conjectures, show that there exists an integer N such that $X(\mathbb{F}_{q^n}) \neq \emptyset$ for all $n \geq N$.

(4) (Weil's lemma) Let ℓ be a prime. Let $P \in \mathbb{Q}_\ell[T]$ be a polynomial with roots $\alpha_1, \dots, \alpha_d \in \overline{\mathbb{Q}}_\ell$. Show that P is uniquely determined by the function $\mathbb{Z}[T] \to \mathbb{Q}$ given by

$$F \mapsto \left| \prod_{i=1}^{d} F(\alpha_i) \right|_\ell.$$

(5) Fix a positive integer g and a prime power q. Using the Weil conjectures, show that the number of polynomials that can occur as $P_1(T)$ for some abelian variety of dimension g over $\mathbb{F}_q$ is bounded. (Hint: each coefficient is both integral and bounded by some function of g and q.)

(6) Let $X = \operatorname{Spec}(\overline{A})$ be an affine scheme of finite type over $\mathbb{F}_q$. Prove that $\#X(\mathbb{F}_q) = 0$ if and only if the ideal in $\overline{A}$ generated by all elements of the form $f^q - f$ for $f \in \overline{A}$ is the unit ideal.

(7) Put $X = \operatorname{Spec}(k)$ for some field k and fix an algebraic closure $\overline{k}$ of k. Show that the profinite fundamental group $\pi_1(X, \operatorname{Spec}(\overline{k}))$, as defined in Definition 15.6, is canonically isomorphic to the absolute Galois group $G_k = \operatorname{Gal}(\overline{k}/k)$.

(8) Let G be a profinite topological group. Prove that any homomorphism $G \to \mathrm{GL}(r, \overline{\mathbb{Q}}_\ell)$ has image contained in $\mathrm{GL}(r, E)$ for some finite extension $E/\mathbb{Q}_\ell$. (Hint: one approach to this uses the Baire category theorem. See [63, Remark 9.0.7].)

References

[1] T.G. Abbott, K.S. Kedlaya, and D. Roe, Bounding Picard numbers of surfaces using p-adic cohomology, in *Arithmetic, Geometry and Coding Theory (AGCT 2005)*, Séminaires et Congrès 21, Soc. Math. France, 2009, 125–159.

[2] J. Achter, D. Erman, K.S. Kedlaya, M.M. Wood, and D. Zureick-Brown, A heuristic for the distribution of point counts for random curves over a finite field, *Phil. Trans. Royal Soc. A* **373** (2015).

[3] P. Allen, F. Calegari, A. Caraiani, T. Gee, D. Helm, B. Le Hung, J. Newton, P. Scholze, R. Taylor, and J. Thorne, Potential automorphy over CM fields, arXiv:1812.09999v1 (2018).

[4] A. Arabia, Relèvements des algèbres lisses et de leurs morphismes, *Comment. Math. Helvet.* **76** (2001), 607–639.

[5] J.S. Balakrishnan, N. Dogra, J.S. Müller, J. Tuitman, and J. Vonk, Explicit Chabauty–Kim for the split Cartan modular curve of level 13, *Annals of Math.* **189** (2019), 885–944.

[6] J.S. Balakrishnan and J. Tuitman, Explicit Coleman integration for curves, arXiv:1710.01673v3 (2020); to appear in *Math. Comp.*

[7] T. Barnet-Lamb, D. Geraghty, and T. Gee, The Sato-Tate conjecture for Hilbert modular forms, *J. Amer. Math. Soc.* **24** (2011), 411–469.

[8] P. Berthelot, Finitude et pureté cohomologique en cohomologie rigide (with an appendix by A.J. de Jong), *Invent. Math.* **128** (1997), 329–377.

[9] B. Bhatt and P. Scholze, The pro-étale topology for schemes, *Astérisque* **369** (2015), 99–201.

[10] E. Bombieri, Hilbert's 8th problem: an analogue, in *Mathematical Developments Arising from Hilbert Problems, Part 1*, Proc. Symp. Pure Math. 28.1, Amer. Math. Soc., Providence, 1976.

[11] I.I. Bouw and S. Wewers, Computing L-Functions and semistable reduction of superellipic curves, *Glasgow Math. J.* **59** (2017) 77–108.

[12] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises, *J. Amer. Math. Soc.* **14** (2001), 843–939.

[13] J. Bruce and W. Li, Effective bounds on the dimensions of Jacobians covering abelian varieties, arXiv:1804.11015v2 (2019); to appear in *Proc. Amer. Math. Soc.*

[14] A. Brumer and K. Kramer, Paramodular abelian varieties of odd conductor, *Trans. Amer. Math. Soc.* **366** (2014), 2463–2516 (corrigendum posted 2019).

[15] A. Brumer, A. Pacetti, C. Poor, G. Tornaría, J. Voight, and D. Yuen, On the paramodularity of typical abelian surfaces, *Algebra and Number Theory* **13** (2019), 1145–1195.

[16] A. Bucur, C. David, B. Feigon, and M. Lalín, Fluctuations in the number of points on smooth plane curves over finite fields, *J. Number Theory* **130** (2010), 2528–2541.

[17] A. Bucur and K.S. Kedlaya, The probability that a complete intersection is smooth, *J. Théorie Nombres Bordeaux* **24** (2012), 541–556.

[18] A. Caraiani, Perfectoid Shimura varieties, in *Perfectoid Spaces: Lectures from the 2017 Arizona Winter School*, Math. Surveys and Monographs 242, Amer. Math. Soc., Providence, 2019.

[19] C.-L. Chai and F. Oort, An algebraic construction of an abelian variety with a given Weil number, *Alg. Geom.* **2** (2015), 654–663.

[20] F. Charles, On the Picard number of K3 surfaces over number fields, *Algebra and Number Theory* **8** (2014), 1–17.

[21] L. Clozel, M. Harris, and R. Taylor, Automorphy for some l-adic lifts of automorphic mod l Galois representations, *Publ. Math. IHÉS* **108** (2008), 1–181.

[22] G. Cornell and J.H. Silverman, *Arithmetic Geometry*, Springer-Verlag, Berlin, 1986.

[23] E. Costa, A.-S. Elsenhans, and J. Jahnel, On the distribution of the Picard ranks of the reductions of a K3 surface, arXiv:1610.07823v2 (2017).

[24] E. Costa, D. Harvey, and K.S. Kedlaya, Zeta functions of nondegenerate hypersurfaces in toric varieties via controlled reduction in p-adic cohomology, in *ANTS XIII: Proceedings of the Thirteenth Algorithmic Number Theory Symposium*, Open Book Ser. 2, Math. Sci. Pub., 2019, 221–238.

[25] E. Costa and Y. Tschinkel, Variation of Néron-Severi ranks of reductions of K3 surfaces, *Exper. Math.* **23** (2014), 475–481.

[26] J. Cremona, *Algorithms for Modular Elliptic Curves*, second edition, Cambridge Univ. Press, Cambridge, 1997.

[27] A.J. de Jong, Smoothness, semi-stability and alterations, *Publ. Math. IHÉS* **83** (1996), 51–93.

[28] P. Deligne, La conjecture de Weil pour les surfaces K3, *Invent. Math.* **15** (1971/72), 206–226.

[29] P. Deligne, La conjecture de Weil, I, *Publ. Math. IHÉS* **43** (1974), 273–307.

[30] P. Deligne, La conjecture de Weil, II, *Publ. Math. IHÉS* **52** (1980), 137–252.

[31] P. Deligne, Catégories tannakiennes, in *The Grothendieck Festschrift, Volume 2*, Birkhäuser, 1990, 111–195.

[32] P. Diaconis and M. Shahshahani, On the eigenvalues of random matrices, *J. Appl. Prob.* **31** (1994), 49–62.

[33] L. Dieulefait, L. Guerberoff, and A. Pacetti, Proving modularity for a given elliptic curve over an imaginary quadratic field, *Math. Comp.* **79** (2010), 1145–1170.

[34] T. Dokchitser, V. Dokchitser, and A. Morgan, Tate module and bad reduction, arXiv:1809.10208v2 (2020).

[35] V.G. Drinfeld, Langlands' Conjecture for GL(2) over functional Fields, in *Proceedings of the International Congress of Mathematicians, Helsinki*, 1978, 565–574.

[36] V. Drinfeld and S. Vlăduț, The number of points of an algebraic curve [translation of the Russian original], *Functional Anal. Appl.* **17** (1983), 53–54.

[37] B. Dwork, On the rationality of the zeta function of an algebraic variety, *Amer. J. Math.* **82** (1960), 631–648.

[38] K. Eisenträger, S. Hallgren, K. Lauter, T. Morrison, and C. Petit, Supersingular isogeny graphs and endomorphism rings: reductions and solutions, in *Advances in Cryptology—EUROCRYPT 2018, Part III*, Lecture Notes in Comput. Sci. 10822, Springer, Cham, 2018, 329–368.

[39] R. Elkik, Solutions d'èquations à coefficients dans un anneau hensélien, *Ann. Scient. Éc. Norm. Sup.* **6** (1973), 553–604.

[40] D. Erman and M.M. Wood, Semiample Bertini theorems over finite fields, *Duke Math. J.* **164** (2015), 1–38.

[41] D. W. Farmer, S. Koutsoliotas, and S. Lemurell, Varieties via their L-functions, *J. Number Theory* **196** (2019), 364–380.

[42] F. Fité, K.S. Kedlaya, V. Rotger, and A.V. Sutherland, Sato-Tate distributions and Galois endomorphism modules in genus 2, *Compos. Math.* **148** (2012), 1390–1442.

[43] F. Fité, K.S. Kedlaya, and A.V. Sutherland, Sato-Tate groups of abelian threefolds: a preview of the classification, arXiv:1911.02071v2 (2020).

[44] E. Freitag and R. Kiehl, *Étale Cohomology and the Weil Conjecture* [translation of the German original], Ergeb. der Math. 13, Springer-Verlag, Berlin, 1988.

[45] W. Fulton, A note on weakly complete algebras, *Bull. Amer. Math. Soc.* **75** (1969), 591–593.

[46] S.R. Ghorpade and G. Lachaud, Étale cohomology, Lefschetz theorems and number of points of singular varieties over finite fields, *Mosc. Math. J.* **2** (2002), no. 3, 589–631.

[47] L. Göttsche, The Betti numbers of the Hilbert scheme of points on a smooth projective surface, *Math. Ann.* **286** (1990), 193–207.

[48] P.A. Griffiths, On the periods of certain rational integrals, I, *Annals of Math.* **90** (1969), 460–495.

[49] P.A. Griffiths, On the periods of certain rational integrals, II, *Annals of Math.* **90** (1969), 496–541.

[50] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, Wiley, New York, 1978.

[51] B.H. Gross and M.J. Hopkins, Equivariant vector bundles on the Lubin-Tate moduli space, in *Topology and Representation Theory* Contemp. Math. 158, Amer. Math. Soc., Providence, 1994, 23–88.

[52] B.H. Gross and M.J. Hopkins, The rigid analytic period mapping, Lubin-Tate space, and stable homotopy theory, *Bull. Amer. Math. Soc.* **30** (1994), 76–86.

[53] B. Gross and N. Koblitz, Gauss sums and the p-adic Γ-function, *Annals of Math.* **109** (1979), 569–581.

[54] A. Grothendieck, The cohomology theory of abstract algebraic varieties, in *Proceedings of the International Congress of Mathematicians, 1958*, 103–118.

[55] A. Grothendieck, Formule de Lefschetz et rationalité des fonctions L, *Séminaire Bourbaki* **9** (1964), 41–55.

[56] A. Grothendieck, Standard conjectures on algebraic cycles, in *Algebraic Geometry (Internat. Colloq., Tata Inst. Fund. Res., Bombay, 1968)*, Oxford Univ. Press, Oxford, 1969, 193–199.

[57] A. Grothendieck and M. Raynaud, *Revêtements Étales et Groupe Fondamental (SGA 1)*, Doc. Math. 3, Soc. Math. France, 2003.

[58] M. Harris, N.Shepherd–Barron, and R. Taylor, A family of Calabi-Yau varieties and potential automorphy, *Annals of Math.* **171** (2010), 779–813.

[59] R. Hartshorne, *Algebraic Geometry*, Graduate Texts in Math. 52, Springer, 2013.

[60] Y. Ihara, Some remarks on the number of rational points of algebraic curves over finite fields, *J. Fac. Sci. Univ. Tokyo* **28** (1982), 721–724.

[61] K. Ito, T. Ito, and T. Koshikawa, CM liftings of K3 surfaces over finite fields and their applications to the Tate conjecture, arXiv:1809.09604v2 (2018).

[62] N.M. Katz, L-functions and monodromy: four lectures on Weil II, *Adv. Math.* **160** (2001), 81–132.

[63] N.M. Katz and P. Sarnak, *Random Matrices, Frobenius Eigenvalues, and Monodromy*, Amer. Math. Soc. Colloq. Pub. 45, Amer. Math. Soc., Providence, 1999.

[64] K.S. Kedlaya, Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology, *J. Ramanujan Math. Soc.* **16** (2001), 323–338; errata, *ibid.* **18** (2003), 417–418.

[65] K.S. Kedlaya, Finiteness of rigid cohomology with coefficients, *Duke Math. J.* **134** (2006), 15–97.

[66] K.S. Kedlaya, Fourier transforms and p-adic "Weil II", *Compos. Math.* **142** (2006), 1426–1450.

[67] K.S. Kedlaya, Quantum computation of zeta functions of curves, *Comput. Complexity* **15** (2006), 1–19.

[68] K.S. Kedlaya, p-adic cohomology: from theory to practice, in *p-adic Geometry: Lectures from the 2007 Arizona Winter School*, Univ. Lecture Series 45, Amer. Math. Soc., 2008, 190–219.

[69] K.S. Kedlaya, Search techniques for root-unitary polynomials, in *Computational Arithmetic Geometry*, Contemp. Math. 463, Amer. Math. Soc., 2008, 71–82.

[70] K.S. Kedlaya, Notes on isocrystals, arXiv:1606.01321v5 (2018).

[71] K.S. Kedlaya, Étale and crystalline companions, I, arXiv:1811.00204v2 (2019).

[72] K.S. Kedlaya and A.V. Sutherland, A census of zeta functions of quartic K3 surfaces over $\mathbb{F}_2$, *LMS J. Comp. Math.* **19** special issue A (2016), 1–11.

[73] N. Koblitz, *p-adic Numbers, p-adic Analysis, and Zeta Functions*, Graduate Texts in Math. 58, Springer-Verlag, New York, 1984.

[74] L. Lafforgue, Chtoucas de Drinfeld et correspondance de Langlands, *Invent. Math.* **147** (2002), 1–241.

[75] S. Lang and A. Weil, Number of points of varieties in finite fields, *Amer. J. Math.* **76** (1954), 819–827.

[76] G. Laumon, Transformation de Fourier, constantes d'équations fonctionelles et conjecture de Weil, *Publ. Math. IHÉS* **65** (1987), 131–210.

[77] J. Leitzel, M. Madan, and C. Queen, Algebraic function fields with small class number, *J. Number Theory* **7** (1975) 11–27.

[78] B. Le Stum, *Rigid Cohomology*, Cambridge Tracts in Math. 172, Cambridge Univ. Press, Cambridge, 2007.

[79] R. Livné, Cubic exponential sums and Galois representations, in *Current Trends in Arithmetical Algebraic Geometry*, Contemp. Math. **67** (1987), 247–261.
[80] The LMFDB Collaboration, L-Functions and Modular Forms Database, `https://www.lmfdb.org`.
[81] D. Lorenzini, *An Invitation to Arithmetic Geometry*, Graduate Studies in Math. 9, Amer. Math. Soc., Providence, 1996.
[82] M. Madan and C. Queen, Algebraic function fields of class number one, *Acta Arith.* **20** (1972), 423–432.
[83] B. Mazur, Frobenius and the Hodge filtration, *Bull. Amer. Math. Soc.* **78** (1972), 653–667.
[84] P. Mercuri and C. Stirpe, Classification of algebraic function fields with class number one, *J. Number Theory* **154** (2015), 365–374.
[85] D. Meredith, Weak formal schemes, *Nagoya Math. J.* **45** (1971), 1–38.
[86] J.S. Milne, *Étale Cohomology*, Princeton Math. Ser. 33, Princeton Univ. Press, Princeton, N.J., 1980.
[87] J.S. Milne, The Riemann hypothesis over finite fields, from Weil to the present day, in *The Legacy of Bernhard Riemann after One Hundred and Fifty Years. Vol. II*, Int. Press, Somerville, MA, 2016, 487–565; reprinted in *ICCM Not.* **4** (2016), 14–52.
[88] P. Monsky, Formal cohomology, I: The cohomology sequence of a pair, *Annals of Math.* **88** (1968), 218–238.
[89] P. Monsky, Formal cohomology, I: Fixed point theorems, *Annals of Math.* **93** (1971), 315–343.
[90] P. Monksy and G. Washnitzer, Formal cohomology, I, *Annals of Math.* **88** (1968), 181–217.
[91] D. Mumford, *Abelian Varieties*, TIFR Pub. 13, Amer. Math. Soc., Providence, 2012.
[92] J. Neukirch, *Algebraic Number Theory*, Springer, Berlin, 1999.
[93] B. Poonen, Bertini theorems over finite fields, *Ann. Math.* **160** (2004), 1099–1127.
[94] N. Saavedra Rivano, *Catégories Tannakiennes*, Lecture Notes in Math. 265, Springer-Verlag, Berlin, 1972.
[95] C. Schembri, Examples of genuine QM abelian surfaces which are modular, *Res. Number Theory* **5** (2019), article 11.
[96] J.-P. Serre, Géométrie algébrique et géométrie analytique, *Ann. Inst. Fourier* **6** (1956), 1–42.
[97] J.-P. Serre, *Abelian ℓ-adic Representations and Elliptic Curves*, W.A. Benjamin, New York, 1968.
[98] J.-P. Serre and J. Tate, Good reduction of abelian varieties, *Ann. Math.* **88** (1968), 492–517.
[99] Q. Shen and S. Shi, Function fields of class number one, *J. Number Theory* **154** (2015), 375–379.
[100] J. Silverman, *The Arithmetic of Elliptic Curves*, second edition, Graduate Texts in Math. 106, Springer, Dordrecht, 2009.
[101] S. A. Stepanov, An elementary proof of the Hasse-Weil theorem for hyperelliptic curves, *J. Number Theory* **4** (1972), 118–143.
[102] C. Stirpe, A counterexample to 'Algebraic function fields with small class number', *J. Number Theory* **143** (2014), 402–404.
[103] A.V. Sutherland, Sato-Tate distributions, in *Analytic Methods in Arithmetic Geometry*, Contep. Math. 740, Amer. Math. Soc., 2019, 197–248.
[104] G. Tamme, *Introduction to Étale Cohomology* [translation of the German original], Universitext, Springer-Verlag, Berlin, 1994.
[105] J. Tuitman, Counting points on curves using a map to $\mathbb{P}^1$, *Math. Comp.* **85** (2016), 961–981.
[106] J. Tuitman, Counting points on curves using a map to $\mathbb{P}^1$, II, *Finite Fields Appl.* **45** (2017), 301–322.
[107] R. Taylor and A. Wiles, Ring-theoretic properties of certain Hecke algebras, *Ann. Math.* **141** (1995), 553–572.
[108] R.M. van Luijk, K3 surfaces with Picard number one and infinitely many rational points, *Algebra and Number Theory* **1** (2007), 1–15.
[109] J. Voight, Curves over finite fields with many points: an introduction, in *Computational aspects of algebraic curves*, Lecture Notes Series on Computing 13, World Scientific, Hackensack, NJ, 2005, 124–144.
[110] W.C. Waterhouse and J.S. Milne, Abelian varieties over finite fields, in *1969 Number Theory Institute*, Proc. Sympos. Pure Math. 20, Amer. Math. Soc., Providence, 1971, 53–64.
[111] A. Weil, Sur les fonctions algébriques à corps de constantes fini, *C. R. Acad. Sci. Paris* **210** (1940), 592–594.
[112] A. Weil, On the Riemann hypothesis in function-fields, *Proc. Nat. Acad. Sci.* **27** (1941), 345–347.
[113] A. Weil, Sur les courbes algébriques et les variétés qui s'en déduisent, Hermann et Cie., Paris, 1948.
[114] A. Weil, Variétés abéliennes et courbes algébriques, Hermann et Cie., Paris, 1948.
[115] A. Weil, Numbers of solutions of equations in finite fields, *Bull. Amer. Math. Soc.* **55** (1949), 497–508.
[116] A. Wiles, Modular elliptic curves and Fermat's last theorem, *Ann. Math.* **141** (1995), 443–551.
[117] N. Yui, Modularity of Calabi-Yau varieties: 2011 and beyond, in *Arithmetic and Geometry of K3 Surfaces and Calabi-Yau Threefolds*, Fields Inst. Commun. 67, Springer, New York, 2013, 101–139.