Trace functions over finite fields and applications

É. Fouvry, E. Kowalski, Ph. Michel

DRAFT

Version of December 3, 2014

etienne.fouvry@math.u-psud.fr
kowalski@math.ethz.ch
philippe.michel@epfl.ch

Contents

Chapter 1. Introduction and motivation 1
1.1. Objectives 1
1.2. A statement of the Riemann Hypothesis 1
1.3. Illustrations 2
1.4. What comes next? 3
Notation 4
Conventions 4

Chapter 2. Trace functions: definition and invariants 5
2.1. Galois representations 5
2.2. Representations of weight 0 8
2.3. The Kummer representations 11
2.4. The Artin-Schreier representations 13
2.5. A summary 15
2.6. Analogies and alternate approaches 15

Chapter 3. Formalism of trace functions 16
3.1. Swan conductor 16
3.2. Morphisms, isomorphisms 18
3.3. Irreducible representations 20
3.4. Complex conjugation 21
3.5. Addition 22
3.6. Multiplication 24
3.7. Change of variable 26
3.8. Summing over solutions of a polynomial equation 28
3.9. Summary 29

Chapter 4. The Riemann Hypothesis 30
4.1. A version of Deligne's theorem 30
4.2. More precise computations 34
4.3. The black-box version 35
4.4. First consequences of Deligne's Theorem 35
4.5. Extensions 39

Chapter 5. The Fourier transform and other cohomological transforms 41
5.1. The discrete Fourier transform 41
5.2. The sheaf-theoretic Fourier transform 42
5.3. Applications of the Fourier transform 43
5.4. Precise constructions 44

Appendix A. Appendix A: trace functions 45

Bibliography 46

CHAPTER 1

Introduction and motivation

1.1. Objectives

The first goal of this book is to make accessible to analytic number theorists the power of Deligne's general version of the Riemann Hypothesis over finite fields, as it applies to bounds for exponential sums in two or more variables, or to sums of more general summands. Readers of this book should be able to use it to prove, say, that

$$\Bigl|\sum_{x\in(\mathbf{Z}/p\mathbf{Z})^\times}\ldots\Bigr|\leqslant Cp^{3/2}$$

for all primes p and some absolute constant $C\geqslant 1$, just as readily as the Weil bounds for exponential sums in one variable allows them to check estimates like

$$\Bigl|\sum_{x\in\mathbf{Z}/p\mathbf{Z}}e\Bigl(\frac{x^3+3x+1}{p}\Bigr)\Bigr|\leqslant C\sqrt{p}$$

(for all primes p and some absolute constant $C\geqslant 1$) using straightforward statements that enscapsulate Weil's proof of the Riemann Hypothesis for *curves* over finite fields.

1.2. A statement of the Riemann Hypothesis

To motivate and illustrate the type of statements we will be able to derive from Deligne's general form of the Riemann Hypothesis over finite fields, we begin by stating a concrete version of the Riemann Hypothesis. This has the flavor of a "quasi-orthogonality" property of special functions defined over finite fields.

The statement contains keywords which are probably unfamiliar to most readers (they have not yet been defined!), and one of the goal of the book will be to explain these, and to understand how to apply successfully this result.

Theorem 1.2.1 (Deligne). *Let p be a prime number. Let $c\geqslant 1$ be a parameter. Let K_1, K_2 be* trace functions *modulo p*

$$K_i\,:\,\mathbf{F}_p\longrightarrow\mathbf{C}$$

which are ***geometrically irreducible*** *modulo p, with* **conductors** *respectively c_1 and c_2.*

(1) *If there does not exist a complex number α with $|\alpha|=1$ such that $K_2=\alpha K_1$, then we have*

$$\Bigl|\sum_{x\in\mathbf{F}_p}K_1(x)\overline{K_2(x)}\Bigr|\leqslant 3c_1^2c_2^2\sqrt{p}. \tag{1.1}$$

(2) *Otherwise, we have*

$$\Bigl|\sum_{x\in\mathbf{F}_p}K_1(x)\overline{K_2(x)}-\bar{\alpha}p\Bigr|\leqslant 3c_1^2c_2^2\sqrt{p}.$$

This theorem can be used, in many instances, purely as a "black box". Indeed, it is a fact that in most applications to analytic number theory, we will want to apply it to obtain a bound like (1.1) in cases where *the functions* K_1 *and* K_2 *are concretely given functions*. For instance, to prove (a weak form of) the Hasse bound for the number of points on an elliptic curve over $\mathbf{F}_p$, one would take

$$K_1(x) = \Bigl(\frac{x^3 + \alpha x + \beta}{p}\Bigr), \qquad K_2(x) = 1,$$

where $(\frac{\cdot}{p})$ is the Legendre symbol, and the discriminant of the cubic polynomial $X^3 + \alpha X + \beta \in \mathbf{F}_p[X]$ is non-zero. To prove (a weak form of) the Weil bound for Kloosterman sums, we could take

$$K_1(x) = e\Bigl(\frac{ax}{p}\Bigr), \qquad K_2(x) = e\Bigl(-\frac{\bar{x}}{p}\Bigr) \quad \text{if } x \neq 0, \quad K_2(0) = 0$$

It is then conceivable that one will find in a reference text (for instance, in this book), a statement to the effect that the specific functions K_1 and K_2 of interest *are* trace functions, and hopefully further information such as their conductor. Then Theorem 1.2.1 becomes applicable and the question, to prove a square-root cancellation bound, becomes simply: is K_2 proportional to K_1, with a proportionality constant of modulus 1? We will illustrate this with the two examples above and more in Section 1.3 below.

Remark 1.2.2. We spoke of "weak forms" of the respective bounds because Theorem 1.2.1 will usually give an estimate of size $C\sqrt{p}$, where C is not the best possible constant, whereas Hasse and Weil proved their respective estimates with $C = 2$ which is, in a certain sense, best possible (see below for more details).

1.3. Illustrations

We illustrate the remarks above by establishing using Theorem 1.2.1 a number of instances of the Riemann Hypothesis over finite fields, all of which are statements that have appeared in practice in important results of analytic number theory. To implement Theorem 1.2.1, we will make references to later parts of this book, and especially to Appendix A that lists known trace functions and their invariants. We expect that many basic applications of trace functions could arise in the future in similar manner.

1.3.1. The Hasse bound. Historically, one of the first example of the Riemann Hypothesis over finite fields was provided by Hasse's theorem concerning the number of solutions (x, y) in a finite field k of a cubic equation

$$y^2 = x^3 + \alpha x + \beta$$

where α and β are such that the polynomial $x^3 + \alpha x + \beta$ has no multiple roots. Precisely, Hasse proved that this number, say $N(\alpha, \beta)$, satisfies $N(\alpha, \beta) = p - a$ for some integer a such that

$$|a| \leqslant 2\sqrt{|k|}. \tag{1.2}$$

It is the exponent 1/2 hidden in $\sqrt{|k|} = |k|^{1/2}$ which is the sign of the Riemann Hypothesis.

There are today a few different proofs of (1.2), some of which are quite elementary from the technical point of view. (For very specific choices of α and β, such as $\alpha = -1$, $\beta = 0$, some arguments go back to Gauss). We now show how to bring this bound in the framework of Theorem 1.2.1.

1.3.2. The Weil bound. The Weil bound for Kloosterman sums is probably the most important instance of the Riemann Hypothesis over finite fields, as far as analytic applications are concerned. For a prime p and $a \in \mathbf{F}_p^\times$, we denote

$$\mathrm{Kl}_2(a;p) = \frac{1}{\sqrt{p}} \sum_{x \in \mathbf{F}_p^\times} e\Big(\frac{ax+\bar{x}}{p}\Big),$$

where $\bar{x}$ denotes the inverse of x modulo p. The precise statement of Weil's result is

$$|\,\mathrm{Kl}_2(a;p)| \leqslant 2 \tag{1.3}$$

for all $a \in \mathbf{F}_p^\times$. Again, we now show how to prove a weaker bound using Theorem 1.2.1.

1.3.3. The Friedlander-Iwaniec sums. Among the more recent applications of the Riemann Hypothesis over finite fields, one of the most important concerns an exponential sum that first arose in the work of Friedlander and Iwaniec [**37**] on the exponent of distribution of the ternary divisor function. Indeed, the estimate for this sum that was given by Birch and Bombieri is one of the key ingredients in the proof of the bounded gap property for primes by Y. Zhang [**82**].

We consider the sum as it appeared in [**37**], namely

$$\mathrm{FI}(a,b;p) = \frac{1}{p^{3/2}} \sum_{x,y,z \in \mathbf{F}_p^\times} e\Big(\frac{1}{p}(\ldots)\Big)$$

where p is a prime number and $(a,b) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times$. (In fact, this sum arises most naturally in a different manner, that we will describe at a later stage.)

Birch and Bombieri proved that there exists an absolute constant $C \geqslant 0$ such that

$$\mathrm{FI}(a,b;p) \leqslant C$$

for all primes p and all a, b invertible modulo p. We now explain how to prove this.

1.3.4. The Pierce sums. In her work proving the first non-trivial bound for the 3-part of the class group of an imaginary quadratic field, L. Pierce [**69**] had to estimate sums of the type

$$\Pi(a,b,c,d;p) = \frac{1}{p^{3/2}} \sum_{x,y,z} \Big(\frac{4x^3-a(z+b)^2}{p}\Big)\Big(\frac{4y^3-az^2}{p}\Big) e\Big(\frac{cx-cy+dz}{p}\Big),$$

where p is a prime and $(a,b,c,d) \in \mathbf{F}_p^\times \times {\mathbf{F}_p}^3$ are parameters.

Again, the desired outcome is that these should be uniformly bounded, provided $b \neq 0$ or $(c,d) \neq (0,0)$. This estimate was indeed given by N. Katz [**57**]: there exists an absolute constant $C \geqslant 0$ such that

$$\Pi(a,b,c,d;p) \leqslant C$$

for all primes $p \geqslant 3$, if $d \in \mathbf{F}_p^\times$.

We will see here that this sum can also be approached according to Theorem 1.2.1.

1.4. What comes next?

In the next two chapters, we will establish a rigorous definition of the terms used in Theorem 1.2.1. We will then be able to show how it follows from another version of the Riemann Hypothesis involving the algebraic objects that properly underly the trace functions. Then we will show how a rich formalism leads to the construction of many trace functions (and of estimates for their conductors). Throughout, we attempt to illustrate the evolving formalism with concrete interpretations of the meaning of various ideas and notions from algebraic geometry. In some cases, this can be done entirely rigorously, while in others one should see these interpretations as giving intuition and heuristic understanding only. As a typical example, we will see that the property that a trace function K is geometrically irreducible can be interpreted as the statement that

$$\sum_{x\in\mathbf{F}_p} |K(x)|^2$$

is close to p, and that it can often be deduced *rigorously* from this fact (in practice, it follows when the conductor of K is sufficiently small compared with p).

Notation

- If k is a finite field, we denote by $\bar{k}$ a fixed algebraic closure of k; for any integer $\nu \geqslant 1$, we further write k_ν for the unique extension of degree ν of k contained in $\bar{k}$.
- For any field K, we denote by $\mathbf{P}^1(K)$ the set of points on the projective line over K, i.e., $\mathbf{P}^1(K) = K \cup \{\infty\}$.
- We denote by $|X|$ the cardinality of any set, with $|X| = \infty$ if X is infinite.
- For any finite set X, we denote by $C(X)$ the space of complex-valued functions on X, and we usually view it as a Hilbert space with the inner product

$$\langle \varphi_1, \varphi_2 \rangle = \frac{1}{|X|} \sum_{x\in X} \varphi_1(x)\overline{\varphi_2(x)}$$

for φ_1, $\varphi_2 \in C(X)$.

Conventions

We attempt to keep the following notational conventions:

- ...

CHAPTER 2

Trace functions: definition and invariants

We will begin in this chapter by giving a rigorous definition of trace functions over finite fields, thus allowing us to also define their main invariants unambiguously. We then construct two of the most important classes of examples from scratch, to illustrate this definition. Most importantly, at the end of this chapter, the quasi-orthogonality form of the Riemann Hypothesis over finite fields that was presented in the previous chapter will have become a precise mathematical statement, where all terms are defined.

The definition also allows us to develop straightforwardly a basic formalism of trace functions (addition, multiplication, etc). This will be extended later to much deeper facts.

Readers who are unfamiliar with infinite Galois theory can omit the first two sections and continue with Section 2.5, where we summarize the qualitative features of trace functions that arise from the definitions. On the other hand, this summary should probably be skipped by those readers who have read until then.

2.1. Galois representations

Let k be a finite field of characteristic p, and $\bar{k}$ an algebraic closure of k. We have then field extensions

$$k \hookrightarrow \bar{k} \hookrightarrow \bar{k}(T) \hookrightarrow \overline{k(T)}$$

as well as

$$k \hookrightarrow k(T),$$

where $\overline{k(T)}$ denotes a separable closure of $k(T)$.

We denote then

$$\Pi_k = \mathrm{Gal}(\overline{k(T)}/k(T)) \tag{2.1}$$

$$\Pi_k^g = \mathrm{Gal}(\overline{k(T)}/\bar{k}(T)). \tag{2.2}$$

These Galois groups are complicated groups, and they carry a compact topology that will allow us to speak of continuity of maps from either of them to any topological space. By Galois theory, Π_k^g is a normal subgroup of Π_k and the quotient is isomorphic to $\mathrm{Gal}(\bar{k}/k)$.

For any $x \in \bar{k}$, two subgroups of Π_k are defined which play a crucial role in defining trace functions: the decomposition group at x and the inertia group at x. They are in fact only well-defined up to conjugacy, and they are best considered from the point of view of abstract algebra. However, we wish to give a description that is more "concrete" and will allow us to get basic understanding of their nature (see in particular the examples in the next sections).

To do this, we first recall that $\overline{k(T)}$ is the union of finite Galois extensions $L/k(T)$, and that an element $\sigma \in \Pi_k$ is uniquely defined by its restrictions to each such extension L. These are arbitrary collections $(\sigma_L)_L$ of elements in $\mathrm{Gal}(L/k(T))$, subject to an obvious compatibility condition when we consider $L \subset L'$: the restriction to L of $\sigma_{L'}$ must be σ_L.

We will first define “small” decomposition and inertia groups as subgroups of the Galois groups $\mathrm{Gal}(L/k(T))$, and then combine them in a similar manner.

We first view elements of $k(T)$ as functions on $\mathbf{P}^1(\bar{k}) = \bar{k} \cup \{\infty\}$ with values in $\mathbf{P}^1(\bar{k})$, using the usual interpretation of $1/\infty = 0$ and $1/0 = \infty$.

Similarly, we can think of elements of $\overline{k(T)}$ as algebraic “multi-valued” functions on $\mathbf{P}^1(\bar{k})$. In fact, any element $Z \in \overline{k(T)}$ satisfies a (separable) polynomial equation

$$Z^d + a_{d-1}(T)Z^{d-1} + \cdots + a_1(T)Z + a_0(T) = 0$$

for some $d \geqslant 1$, with $a_i(T) \in k(T)$. If the degree d is chosen to be minimal, and $x \in \mathbf{P}^1(\bar{k})$ is not a pole of any coefficient a_i, then one says that Z is *defined at* x. One then defines the set of “values” of $Z(x)$ as the set of roots of the polynomial equation

$$z^d + a_{d-1}(x)z^{d-1} + \cdots + a_1(x)z + a_0(x) = 0$$

in $\bar{k}$.

Let $L/k(T)$ be a finite Galois extension. By the primitive element theorem, there exists a single element $Y \in L$ such that $L = k(T)(Y)$. Given a fixed $x \in \bar{k}$, we can always arrange to find Y which is defined at x. Then L, as a $k(T)$-vector space, has basis $(1, Y, \ldots, Y^{d-1})$ where $d = [L : k(T)]$.

If we fix one of the values α of Y at x (in the sense above), then we can define in a consistent manner a value $Z_\alpha(x)$ of any $Z \in L$ at x by

$$Z_\alpha(x) = \sum_{i=0}^{d-1} f_i(x)\alpha^i$$

for any

$$Z = \sum_{i=0}^{d-1} f_i Y^i \in L$$

which is defined at x.

We then put

$$\bar{D}_{L,\alpha} = \{\sigma \in \mathrm{Gal}(L/k(T)) \mid \text{for all } Z \in L, \text{ we have } Z_\alpha(x) = 0 \text{ if and only if } (\sigma(Z))_\alpha(x) = 0\}.$$

It is a fact that $\bar{D}_{L,\alpha} \subset G$ only depends on x up to conjugation. It contains a normal subgroup

$$\bar{I}_{L,\alpha} = \{\sigma \in \mathrm{Gal}(L/k(T)) \mid \text{for all } Z \in L, \text{ we have } Z_\alpha(x) = (\sigma(Z))_\alpha(x)\},$$

with

$$\bar{D}_{L,\alpha}/\bar{I}_{L,\alpha} \simeq \mathrm{Gal}(k(\alpha)/k)$$

by the map of restriction of $\sigma \in \bar{D}_{L,\alpha}$ to constant functions.

It is a non-obvious fact (which depends on the axiom of choice) that we can make consistent choices of α for all finite Galois extensions L in such a way that the subgroup

$$D_x = \{\sigma \in \Pi_k \mid \sigma \mid L \in \bar{D}_{L,\alpha} \text{ for all } L/k(T)\}$$

makes sense, with a normal subgroup

$$I_x = \{\sigma \in \Pi_k \mid \sigma \mid L \in \bar{I}_{L,\alpha} \text{ for all } L/k(T)\}$$

such that

$$D_x/I_x \simeq \mathrm{Gal}(\bar{k}/k).$$

Lemma 2.1.1. *For any x, the inertia group at x is a subgroup of Π_k^g.*

Note that, since I_x is only well-defined up to conjugacy, this statement makes sense because Π_k^g is normal in Π_k.

Proof. Taking constant functions $Z \in \bar{k} \subset \overline{k(T)}$ and applying the definition, we see that any $\sigma \in I_x$ fixes $\bar{k}$, hence fixes the field generated by $k(T)$ and $\bar{k}$, which is $\bar{k}(T)$. □

Recall that, for any $\nu \geqslant 1$, we denote by k_ν the unique extension of k of degree ν contained in $\bar{k}$. Let $\nu \geqslant 1$ be an integer such that $x \in \mathbf{P}^1(k_\nu)$. Because of the above, there exist elements in D_x/I_x mapping to the Frobenius automorphism $x \mapsto x^{|k|^\nu}$ in $\mathrm{Gal}(\bar{k}/k_\nu)$. Such an element is only well-defined up to conjugacy, however. We will denote by $\mathrm{Fr}_{x,\nu}$ or $\mathrm{Fr}_{x,|k|^\nu}$ any element in D_x with class in D_x/I_x conjugate to the *inverse* of this Frobenius automorphism. These will be called *geometric Frobenius elements* at x, relative to ν (or to $|k|^\nu$).

Remark 2.1.2. It is important to take the degree ν into account, since viewing x as an element of $\mathbf{P}^1(k_{d\nu})$, for some $d \geqslant 1$, instead of $\mathbf{P}^1(k_\nu)$ has the effect of changing the Frobenius conjugacy classes, raising it to the d-th power.

We can now define trace functions in a very general context; these will be specialized later to the trace functions of interest to analytic number theory.

Definition 2.1.3. Let k be a finite field, and E an arbitrary field.

(1) A *Galois representation of* k over E is a homomorphism

$$\varrho \,:\, \Pi_k \longrightarrow \mathrm{GL}(V),$$

where V is a finite-dimensional E-vector space. The dimension of V is called the *rank* of ϱ.

(2) Let $\nu \geqslant 1$ be an integer. The ν-th trace function of ϱ, or trace function if $\nu = 1$, is the function

$$\begin{cases} k_\nu \longrightarrow E \\ x \mapsto t_\varrho(x;\nu) \end{cases}$$

defined by

$$t_\varrho(x;\nu) = \mathrm{Tr}(\varrho(\mathrm{Fr}_{x,\nu}) \mid V^{I_x}),$$

where V^{I_x} denotes the subspace in V of vectors invariant under the action of I_x.

Remark 2.1.4. This function is well-defined, because the ambiguity in the definition of $\mathrm{Fr}_{x,\nu}$ does not change the value of the trace: first, the trace is invariant under conjugation, and second, although the Frobenius conjugacy class is only well-defined in D_x/I_x, the action on V^{I_x} does not depend on the representative in D_x of such elements. In other words, we have the following elementary lemma:

Lemma 2.1.5. *Let G be a group, D a subgroup of G and I a normal subgroup of D. Let*

$$\varrho \,:\, G \longrightarrow \mathrm{GL}(V)$$

be any representation of G.

(1) *The subspace $w = V^I$ of I-invariants in V is stable under the action of D.*

(2) *Let $\sigma_0 \in D/I$ be given. Then for any $\sigma \in D$ mapping to σ_0 modulo I, the trace of σ on W is the same.*

(3) *Moreover, for any $g \in G$, the trace in* (2) *is the same when replacing D by gDg^{-1} and I by gIg^{-1}.*

Proof. (1) Let $\sigma \in D$ be given, and $w \in W$. Then for any $\tau \in I$, we get

$$\tau(\sigma w) = \sigma(\sigma^{-1}\tau\sigma)w = \sigma w$$

(since $\sigma^{-1}\tau\sigma \in I$ by assumption) so that $\sigma w \in W$.

(2) Given $\sigma \in D$ mapping to $\sigma_0 \in I$, any other element σ' with the same property satisfies $\sigma' = \sigma i$ for some $i \in I$; then for any $w \in W$, we have $\sigma'(w) = \sigma(i(w)) = \sigma(w)$ since $w \in V^I$. Hence σ' induces the same linear map on W as σ does, which certainly implies that they have the same traces.

(3) If we replace D and I by their respective conjugates gDd^{-1} and gIg^{-1}, we replace V^I by gV^I (as a simple computation shows) and σ by $g\sigma g^{-1}$, and we leave to the reader to check that the trace of $g\sigma g^{-1}$ on gV^I is the same as the trace of σ on V^I. □

In practice, the Galois representations of use for us have the property that I_x acts trivially on V for all but finitely many x. Any such x is called *unramified.* Precisely:

Definition 2.1.6. Let k be a finite field and E an arbitrary field. Let

$$\varrho : \Pi_k \longrightarrow \mathrm{GL}(V),$$

be a Galois representation of k over E. For $x \in \mathbf{P}^1(\bar{k})$, we say that ϱ is *unramified at* x, or *lisse at* x, if I_x acts trivially on ϱ, i.e.,

$$\varrho(\sigma)v = v$$

for $\sigma \in I_x$ and $v \in V$.

The set of points which are *ramified* is called the set of singularities of ϱ, and is denoted $\mathrm{Sing}(\varrho)$.

2.2. Representations of weight 0

The trace functions that occur in this book are of a special type. They are complex-valued, and the eigenvalues of Frobenius are highly restricted. However, we do not simply take Galois representations corresponding to $E = \mathbf{C}$, because the continuity restriction would be too stringent, and exclude too many examples. Instead, we assume that we can "transfer" a representation corresponding to a different field to $\mathbf{C}$:

Definition 2.2.1 (Analytic representation). Let k be a finite field, and let V be a finite-dimensional complex vector space. A homomorphism

$$\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$$

is an *analytic Galois representation* on V if there exists a topological field E, an embedding $\iota : E \hookrightarrow \mathbf{C}$ and a *continuous* Galois representation

$$\tilde{\varrho} : \Pi_k \longrightarrow \mathrm{GL}_{\dim(V)}(E)$$

of k over E, such that

$$\varrho(\sigma) = \iota(\tilde{\varrho}(\sigma))$$

for all $\sigma \in \Pi_k$. We then say that $\tilde{\varrho}$ is *associated to* ϱ.

If ϱ is an analytic Galois representation as above, we get immediately the relation

$$t_\varrho(x;\nu) = \iota(t_{\tilde{\varrho}}(x;\nu))$$

for any $\nu \geqslant 1$ and $x \in k_\nu$. Moreover, because unramified points are defined algebraically, ϱ is unramified at x if and only if $\tilde{\varrho}$ is.

Now we recall an important definition:

Definition 2.2.2 (Weil numbers). Let $w \in \mathbf{R}$ be a real number and q a power of a prime. Let E be a field. An element $\alpha \in E$ is a q-Weil number of weight w if

$$|\sigma(\alpha)| = q^{w/2}$$

for any embedding $\alpha : E \hookrightarrow \mathbf{C}$. If $w = 0$, we say that α is a Weil number of weight 0.

Example 2.2.3. (1) Roots of unity (in any field of characteristic zero) are Weil numbers of weight 0.

(2) Gauss sums of non-trivial characters of $\mathbf{F}_q^\times$ are q-Weil numbers of weight 1 in $\mathbf{C}$.

(3) If α is a q-Weil number of weight w, then $\bar{\alpha}$ is also one, and $1/\alpha$ is a q-Weil number of weight $-w$. In particular, if α is of weight 0 then $\bar{\alpha} = 1/\alpha$ is also of weight 0.

(4) If α (resp. β) is a q-Weil number of weight w (resp. w'), then $\alpha\beta$ is a q-Weil number of weight $w + w'$.

Exercise 2.2.4. (1) Show that any $\alpha \in E$ is a q-Weil number of any weight if the characteristic of E is positive.

(2) If E has characteristic zero, prove that any q-Weil number of any weight is algebraic over $\mathbf{Q}$.

(3) Prove that if $\alpha \in \mathbf{C}$ is both an algebraic integer and a Weil number of weight 0, then α is a root of unity.

(4) Let $\alpha \in E$. Show that if α^ν is a Weil number of weight 0 for some integer $\nu \geqslant 1$, then α is a Weil number of weight 0.

We then have the second definition:

Definition 2.2.5 (Weight w representation). Let k be a finite field and E an arbitrary field. Let $w \in \mathbf{R}$ be given. A Galois representation

$$\varrho : \Pi_k \longrightarrow \mathrm{GL}(W)$$

of k over E is of weight w if and only if, for all $\nu \geqslant 1$ and $x \in \mathbf{P}^1(k_\nu)$, if ϱ is unramified at x, the eigenvalues of $\varrho(\mathrm{Fr}_{x,|k|^\nu})$ are $|k|^\nu$-Weil numbers of weight w.

In particular, the definition implies that for any weight 0 representation over k and any $x \in k_\nu$ which is unramified, we have

$$|t_\varrho(x)| \leqslant \dim(V),$$

since it is the sum of $\dim(V)$ eigenvalues, each of which is of modulus 1.

We finally combine these two definitions to obtain the set of representations of most interest to us:

Definition 2.2.6 (ℓ-adic representations; trace functions). Let k be a finite field of characteristic p. Let $\ell \neq p$ be a prime number.

(1) An *ℓ-adic representation* ϱ of k on a finite-dimensional complex vector space V is an analytic Galois representation

$$\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$$

of weight 0 associated to a representation

$$\tilde{\varrho} : \Pi_k \longrightarrow \mathrm{GL}(W) \tag{2.3}$$

where W is a finite-dimensional $\bar{\mathbf{Q}}_\ell$-vector space, such that the set $\mathrm{Sing}(\varrho)$ is finite.

(2) A *trace function* of k is any function

$$K : k \longrightarrow \mathbf{C}$$

such that $K = t_\varrho$ for some ℓ-adic representation ϱ of k. We say that ϱ is *associated* to K, and similarly that any representation (2.3) is associated to K.

The representation ϱ associated to a trace function is not unique, as we will soon see. However, there is usually a "best" possible representation, defined as the "least complicated" one. In order to define this, we need a basic invariant measuring the complexity of an ℓ-adic representation. We already have at hand two basic numerical invariants that, intuitively, are related to the complexity of ϱ: its rank, and the number of singular points. However, their combination does *not* suffice to get a good theory of trace functions (see ....). The missing ingredient is however quite delicate.

Definition 2.2.7 (Swan conductor). Let k be a finite field, ℓ a prime different from the characteristic of k and

$$\varrho : \Pi_k \longrightarrow \mathrm{GL}(W)$$

a continuous representation, where W is a finite-dimensional $\bar{\mathbf{Q}}_\ell$-vector space.

The *Swan conductor* of ϱ at $x \in \mathbf{P}^1(\bar{k})$ is denoted $\mathrm{Swan}_x(\varrho)$. Similarly, if π is an analytic Galois representation associated to ϱ, we denote $\mathrm{Swan}_x(\pi) = \mathrm{Swan}_x(\varrho)$, which does not depend on the choice of ϱ. This is a non-negative integer, which is zero if ϱ or π is unramified at x.

We will give a precise definition in Appendix A for completeness, and we will see how some basic formalism allows one to handle these invariants in Section 3.1. For the moment it is merely a name.

It turns out then that the following invariant encapsulates the complexity of the trace functions, in all applications that we know.

Definition 2.2.8 (Conductor). Let k be a finite field.

(1) The conductor of an ℓ-adic representation ϱ is

$$\mathrm{c}(\varrho) = \mathrm{rank}(\varrho) + |\,\mathrm{Sing}(\varrho)| + \sum_{x \in \mathrm{Sing}(\varrho)} \mathrm{Swan}_x(\varrho).$$

(2) The conductor of a trace function $K : k \longrightarrow \mathbf{C}$ is

$$\mathrm{c}(K) = \min\{\mathrm{c}(\varrho) \mid K = t_\varrho\},$$

where ϱ runs over ℓ-adic representations with trace function K.

If ϱ satisfies $t_\varrho = K$ and $\mathrm{c}(\varrho) = \mathrm{c}(K)$, then we say that ϱ is a *minimal* representation associated to K.

Note that $\mathrm{c}(\varrho) \geqslant 0$ is an integer. It satisfies $\mathrm{c}(\varrho) \geqslant 1$ unless ϱ is the zero representation.

This completes the basic definitions of the objects we want to handle. All this looks very formal, and the precise combination of conditions seems (maybe) arbitrary. There are some hidden, rather deep, properties which depend on the precise combination and which we will use very frequently.

Theorem 2.2.9 (Deligne). (1) *Let k be a finite field and let $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ be a continuous representation of k where V is a finite $\bar{\mathbf{Q}}_\ell$-vector space. Suppose there exists a finite set $X \subset \mathbf{P}^1(\bar{k})$ such that ϱ is of weight 0 outside X, i.e., for all $\nu \geqslant 1$ and $x \in \mathbf{P}^1(k_\nu)$ such that $x \notin X$ and ϱ is unramified at x, the eigenvalues of $\varrho(\mathrm{Fr}_{x,|k|^\nu})$ are Weil numbers of weight 0. Then ϱ is of weight 0.*

(2) *Let k be a finite field and let $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ be an ℓ-adic representation. For any $\nu \geqslant 1$ and any $x \in \mathbf{P}^1(k_\nu)$, we have*

$$|t_\varrho(x;\nu)| \leqslant \mathrm{rank}(\varrho).$$

The meaning of the first part is that one need only check the weight 0 condition, in our definition of an ℓ-adic representation, for all but finitely many x. The bound in the second part is obvious for x unramified since $t_{\varrho}(x;\nu)$ is the trace of a matrix with all eigenvalues 1 by definition of weight 0, and the deep part is that the estimate extends to ramified x.

Proof. The first part is [**15**,] and the second is [**15**,] (see also [**51**,]). □

2.3. The Kummer representations

We provide in this section the construction of the ℓ-adic representations corresponding to trace functions of the type

$$K(x) = \chi(f(x))$$

where χ is a multiplicative character of $k^{\times}$ and $f \in k(T)$ is a rational function.

More precisely, we prove:

Theorem 2.3.1. *Let k be a finite field of characteristic p and let χ be a non-trivial multiplicative character of $k^{\times}$ and $f \in k(T)$ a non-zero rational function.*

There exists a representation

$$\mathcal{L}_{\chi(f)} : \Pi_k \longrightarrow \mathbf{C}^{\times}$$

such that:

(1) *It is unramified at all x which is not a zero or a pole of f;*

(2) *For each x which is a zero or a pole of f of order not divisible by d, the representation $\mathcal{L}_{\chi(f)}$ is ramified at x; its Swan conductor at x is* 0*, In particular, $\mathcal{L}_{\chi(f)}$ is everywhere tame.*

(3) *For all $x \in k$ not a zero or pole of f, we have*

$$t_{\mathcal{L}_{\chi(f)}}(x) = \chi(f(x))$$

while for a zero or a pole of f, we have

$$t_{\mathcal{L}_{\chi(f)}}(x) = 0.$$

Definition 2.3.2 (Kummer representations). Let k be a finite field of characteristic p and let χ be a non-trivial multiplicative character of $k^{\times}$ and $f \in k(T)$ a rational function. The representation $\mathcal{L}_{\chi(f)}$ is called the *Kummer representation* associated to f and χ.

Proof. Let $d \geqslant 2$ be the order of χ. Thus $d \mid |k| - 1$, and in particular k contains all d-th roots of unity of $\bar{k}$. We consider the equation

$$X^d = f(T)$$

with unknown $X \in \overline{k(T)}$. Let Y be a fixed solution in $\overline{k(T)}$, and let $L = k(T)(Y)$. Under the assumption on f, this is an extension of $k(T)$ of degree d.

We first claim that L is a Galois extension of $k(T)$, and that if G denotes the Galois group of L over $k(T)$, the map

$$\kappa \begin{cases} G \longrightarrow \mu_d \\ \sigma \mapsto \frac{\sigma(Y)}{Y} \end{cases}$$

is an isomorphism, where μ_d is the group of d-roots of unity in k (we observed above that these roots of unity all lie in k).

Indeed, that L is a Galois extension follows from the fact that $d \mid |k| - 1$ is coprime to p, so that $L/k(T)$ is separable, and that any two solutions Y_1 and Y_2 are non-zero and are related by $(Y_1/Y_2)^d = 1$, so that $Y_1 = \xi Y_2$ for some d-th root of unity $\xi \in k$. Thus, all solutions belong to L since one of them does.

This observation also shows that $\kappa(\sigma) = \sigma(Y)/Y$ is an element of μ_d for any $\sigma \in G$, and therefore κ is set-theoretically well-defined. It is furthermore a group-homoamorphism since

$$\begin{aligned}\kappa(\sigma\tau) &= \frac{\sigma\tau(Y)}{Y}\\ &= \sigma\Big(\frac{\tau(Y)}{Y}\Big)\frac{\sigma(Y)}{Y}\\ &= \frac{\tau(Y)}{Y}\frac{\sigma(Y)}{Y} = \kappa(\sigma)\kappa(\tau)\end{aligned}$$

for σ and τ in G, since $\tau(Y)/Y$, being in $\mu_d \subset k$, satisfies $\sigma(\tau(Y)/Y) = \tau(Y)/Y$.

Since Y generates L over $k(T)$, and since $\kappa(\sigma) = 1$ if and only if $\sigma(Y) = Y$, we see that κ is injective. Since $|G| = d = |\mu_d|$, it follows that κ is an isomorphism.

Because χ is of order d, hence also χ^{-1}, there exists a homomorphism

$$\tilde{\chi} : \mu_d \longrightarrow \mathbf{C}^\times$$

such that for all $x \in k^\times$ we have

$$\chi(x^{-1}) = \tilde{\chi}(x^{(|k|-1)/d}). \tag{2.4}$$

Now we can consider the composition

$$\varrho : \Pi_k \longrightarrow G \xrightarrow{\kappa} \mu_d \xrightarrow{\tilde{\chi}} \mathbf{C}^\times, \tag{2.5}$$

and we claim that it has the desired properties.

The first map is defined, surjective and continuous by Galois theory; with G and μ_d given the discrete topology, and $\mathbf{C}^\times$ the usual complex topology, it follows that ϱ is continuous. We denote it ϱ.

We next compute the decomposition groups and inertia groups at $x \in \bar{k}$, or rather their images $\bar{D}_x$ and $\bar{I}_x$ in G. For simplicity, we first assume that $x \in k$ and is not a pole of f.

As a $k(T)$-vector space, the field L admits $(1, Y, \ldots, Y^{d-1})$ as a basis. Therefore any element $Z \in L$ can be written

$$Z = f_0 + f_1 Y + \cdots + f_{d-1} Y^{d-1}$$

for some unique $f_i \in k(T)$. We fix an element α in $\bar{k}$ such that $\alpha^d = f(x)$. We view α as being $Y(x)$, and this allows us to speak consistently of $Z(x)$ for any $Z \in L$, namely

$$Z(x) = f_0(x) + \alpha f_1(x) + \cdots + \alpha^{d-1} f_{d-1}(x),$$

whenever x is not a pole of any f_i. Let $\sigma \in G$ and $\gamma = \kappa(\sigma)$. Then

$$\sigma Z = f_0 + \gamma f_1 Y + \cdots + \gamma^{d-1} f_{d-1} Y^{d-1},$$

takes value

$$(\sigma Z)(x) = f_0 + \gamma\alpha f_1(x) + \cdots + \gamma^{d-1}\alpha^{d-1} f_{d-1}(x),$$

at x.

We then distinguish a few cases:

(1) If $f(x) = 0$, so that $\alpha = 0$, we have then $Z(x) = (\sigma Z)(x) = f_0(x)$ for all σ and all Z, and therefore $\bar{D}_x = \bar{I}_x = G$. In particular, ϱ is ramified at zeros of f.

(2) If the extension $k(\alpha)/k$ has degree d, then $(1, \alpha, \ldots, \alpha^{d-1})$ are k-linearly independent, and hence $Z(x) = 0$ if and only if $f_i(x) = 0$ for all x, in which case $(\sigma Z)(x) = 0$

independently of σ. Thus, in that case, $\bar{D}_x = G$. But since $\alpha \neq 0$, simply taking $Z = Y$ we get $(\sigma Z)(x) = \gamma Z(x)$ and therefore $\sigma \in \bar{I}_x$ if and only if $\gamma = 1$, i.e., if and only if $\sigma = 1$. Thus $\bar{I}_x$ is trivial. In particular, ϱ is unramified at x.

(3) In general...[TODO]

Now we compute the Frobenius elements corresponding to $x \in k$. If x is a zero or pole of f, then any element is a Frobenius automorphism (since $\bar{D}_x = \bar{I}_x$), while the invariants in $\mathbf{C}$ of the image of ϱ (namely of the image of χ) is 0 because χ is non-trivial so that

$$t_\varrho(x) = 0$$

in that case.

If $[k(\alpha) : k] = d$, then $\bar{D}_x = G$ and $\sigma \in G$ is a Frobenius element if and only if σ^{-1} acts on values of $Z \in L$ by $y \mapsto y^{|k|}$. Since Y generates L as a field, this implies that

$$(\sigma^{-1}Y)(x) = Y(x)^{|k|} = \alpha^{|k|},$$

which by comparison with $\sigma^{-1}(Y) = \kappa(\sigma)^{-1}Y$, means that

$$\kappa(\sigma)^{-1} = \alpha^{|k|-1}.$$

Since $\bar{I}_x = 1$, we then have by definition

$$t_\varrho(x) = \tilde{\chi}(\kappa(\sigma)) = \tilde{\chi}(\alpha^{1-|k|}) = \tilde{\chi}((\alpha^d)^{(1-|k|)/d}) = \tilde{\chi}(f(x)^{(1-|k|)/d}) = \chi(f(x))$$

(see (2.4) and (2.5)). □

Remark 2.3.3. Note that it is quite possible for $\mathcal{L}_{\chi(f)}$ to be unramified at a zero or pole of f. Indeed, this is the case when the order of the pole (or zero) is divisible by d. As an example, if χ is non-trivial and of order 2, the singularities of $\mathcal{L}_{\chi(f)}$ are those $x \in \mathbf{P}^1(\bar{k})$ such that x is a simple zero or pole of f.

2.4. The Artin-Schreier representations

We provide in this section, similarly to the previous one, the construction of the ℓ-adic representations corresponding to trace functions of the type

$$K(x) = \psi(f(x))$$

where $f \in k(T)$ is a rational function and ψ is a non-trivial additive character of k.

More precisely, we prove:

Theorem 2.4.1. *Let k be a finite field of characteristic p, ψ a non-trivial additive character of k and let $f \in k(T)$ be a rational function. There exists a representation*

$$\mathcal{L}_{\psi(f)} : \Pi_k \longrightarrow \mathbf{C}^\times$$

such that:

(1) *It is unramified at all x which is not a pole of f;*

(2) *If x is a pole of order $d \geqslant 1$ and $d < p$, then the Swan conductor of $\mathcal{L}_{\psi(f)}$ at x is equal to d;*

(3) *For all $x \in k$ not a pole of f, we have*

$$t_{\mathcal{L}_{\psi(f)}}(x) = \psi(f(x))$$

while for a pole of f of order $< p$, we have

$$t_{\mathcal{L}_{\psi(f)}}(x) = 0.$$

Definition 2.4.2 (Artin-Schreier representation). Let k be a finite field of characteristic p and let $f \in k(T)$ be rational function. The representation $\mathcal{L}_{\psi(f)}$ is called the *Artin-Schreier representation* associated to f (with respect to ψ).

PROOF. The argument is similar to the previous section. For simplicity we only consider the case of the prime field, so that $k = \mathbf{F}_p$. We consider this time the equation

$$X^p - X = -f(T)$$

with unknown $X \in \overline{k(T)}$. Let Y be a fixed solution in $\overline{k(T)}$, and let $L = k(T)(Y)$. Under the assumption on f, this is an extension of $k(T)$ of degree p.

We first claim that L is a Galois extension of $k(T)$, and that if G denotes the Galois group of L over $k(T)$, the map

$$A \begin{cases} G \longrightarrow k \\ \sigma \mapsto \sigma(Y) - Y \end{cases}$$

is an isomorphism (where $k = \mathbf{F}_p$ is viewed as an abelian group under addition).

Indeed, that L is a Galois extension follows from the fact that L is separable (the derivative of the polynomial $X^p - X \in k(T)[X]$ is -1 which is invertible) and moreover from the fact that any two solutions Y_1 and Y_2 satisfy $(Y_1 - Y_2)^p = Y_1 - Y_2$ and therefore $Y_1 - Y_2$ is a constant in k. Thus, all solutions belong to L since one of them does.

For the same reason, $A(\sigma) = \sigma(Y) - Y$ is an element of k for any $\sigma \in G$, and therefore A is set-theoretically well-defined. It is furthermore a group-homomorphism since

$$\begin{aligned} A(\sigma\tau) &= \sigma\tau(Y) - Y \\ &= \sigma(\tau(Y) - Y) + \sigma(Y) - Y \\ &= \tau(Y) - Y + \sigma(Y) - Y = A(\sigma) + A(\tau) \end{aligned}$$

for σ and τ in G, since $\tau(Y) - Y$ being in k implies that $\sigma(\tau(Y) - Y) = \tau(Y) - Y$.

Since Y generates L over $k(T)$, and since $A(\sigma) = 0$ if and only if $\sigma(Y) = Y$, we see that A is injective. Since $|G| = p = |k|$, it follows that A is an isomorphism.

Now we can define the composition

$$\varrho : \Pi_k \longrightarrow G \xrightarrow{A} k \xrightarrow{\psi} \mathbf{C}^\times.$$

The first map is defined, surjective and continuous by Galois theory; with G and k given the discrete topology, the composition is continuous. We will see that ϱ satisfies the properties claimed for $\mathcal{L}_{\psi(f)}$.

We begin by computing images $\bar{D}_x$ and $\bar{I}_x$ in G of the decomposition groups and inertia groups, for simplicity only at $x \in k$. We also assume that x is not a pole of f.

As a $k(T)$-vector space, the field L admits $(1, Y, \ldots, Y^{p-1})$ as a basis. Therefore any element $Z \in L$ can be written

$$Z = f_0 + f_1 Y + \cdots + f_{p-1} Y^{p-1}$$

for some unique $f_i \in k(T)$. We fix an element α in $\bar{k}$ such that $\alpha^p - \alpha = -f(x)$. We view α as being $Y(x)$, and this allows us to speak consistently of $Z(x)$ for any $Z \in L$, namely

$$Z(x) = f_0(x) + \alpha f_1(x) + \cdots + \alpha^{p-1} f_{p-1}(x),$$

whenever x is not a pole of any f_i. Let $\sigma \in G$ and $\gamma = A(\sigma) \in k$. Then

$$\sigma Z = f_0 + f_1(Y + \gamma) + \cdots + f_{p-1}(Y + \gamma)^{d-1},$$

takes value

$$(\sigma Z)(x) = f_0 + (\gamma + \alpha) f_1(x) + \cdots + (\gamma + \alpha)^{p-1} f_{p-1}(x),$$

at x.

We distinguish two cases:

(1) If $f(x) = 0$, then we can take $\alpha = 0$; taking $Z = Y$, we find that $Y(x) = \alpha = 0$, but $(\sigma Z)(x) = Y + \gamma$, and therefore $\sigma \in \bar{D}_x$ if and only if $\gamma = 0$. In other words, we have $\bar{D}_x = 1$, and also $\bar{I}_x = 1$ as a consequence. Thus x is unramified.

(2) If the extension $k(\alpha)/k$ has degree p, then $(1, \alpha, \ldots, \alpha^{p-1})$ are k-linearly independent, and hence $Z(x) = 0$ if and only if $f_i(x) = 0$ for all x, in which case $(\sigma Z)(x) = 0$ independently of σ. Thus, in that case, $\bar{D}_x = G$. But since $\alpha \neq 0$, simply taking $Z = Y$ we get $(\sigma Z)(x) = Z(x) + \gamma$ and therefore $\sigma \in \bar{I}_x$ if and only if $\gamma = 0$, i.e., if and only if $\sigma = 1$. Thus $\bar{I}_x$ is trivial, ϱ is unramified at x.

(3) In general...[TODO]

Finally, we compute the Frobenius elements corresponding to $x \in k$. If $f(x) = 0$ then it is trivial, and hence

$$t_\varrho(x) = \psi(A(1)) = 1 = \psi(f(x))$$

in that case.

If $[k(\alpha) : k] = p$, then $\bar{D}_x = G$ and $\sigma \in G$ is a Frobenius element if and only if σ^{-1} acts on values of $Z \in L$ by $y \mapsto y^{|k|}$. Since Y generates L as a field, this means that

$$(\sigma^{-1}Y)(x) = Y(x)^{|k|},$$

which by comparison with $\sigma^{-1}(Y) - Y = -A(\sigma)$, implies that

$$Y(x)^{|k|} - Y(x) = -A(\sigma)$$

and hence that $A(\sigma) = -\alpha^{|k|} + \alpha = f(x)$. Thus, by definition

$$t_\varrho(x) = \psi(A(\sigma)) = \psi(f(x)).$$

□

Remark 2.4.3. As in the case of Kummer representations, it may be that $\mathcal{L}_{\psi(f)}$ is unramified at some pole of f: the simplest case is when $k = \mathbf{F}_p$ and $f = X^p - X$.

2.5. A summary

We recall that the goal of this section is simply to summarize, in the simplest possible terms, some basic qualitative features of trace functions that arise from the previous discussion.

2.6. Analogies and alternate approaches

The setting of Galois representaitons that we have used is not the only way to define trace functions. In fact, a more geometric point of view, that of ℓ-adic sheaves, is often more convenient, because it generalizes to higher-dimensional situations in a way that is much more flexible. We will see the usefulness of this in discussing the Fourier transform and other linear transformations on spaces of trace functions. For the moment, we simply make the following definition:

Definition 2.6.1 (Middle-extension sheaves). Let k be a finite field, ℓ a prime distinct from the characteristic of k. An ℓ-adic *middle-extension sheaf* $\mathcal{F}$ on k is an ℓ-adic representation of k.

The need for this definition, and the name, may seem obscure, but it is useful, if only to make our statements compatible with those in the literature (e.g., in the works of Katz).

In particular, we can speak of the trace function of an ℓ-adic middle-extension sheaf. We will often just say "let $\mathcal{F}$ be a middle-extension sheaf"; then any occurence of ℓ refers, if not specificed otherwise, to the corresponding ℓ-adic field.

CHAPTER 3

Formalism of trace functions

The power of the theory of trace functions relies on the fact that they satisfy a very flexible and general formalism. In this chapter, we present the basic cases of this formalism: we show that many basic operations, when applied to trace functions, lead to other trace functions. Furthermore (and this is essential), the conductor remains under control when performing these operations. In fact, the formalism operates primarily at the level of the Galois representations, and this is important in many contexts. Then, the fact that representations with a given conductor c are transformed into representations with conductor bounded by a function of c only is viewed as a form of "continuity" of these operations.

The first section "explains" in some more detail how to compute the Swan conductor at x of a representation, following [**51**,]. Although this is not a full definition, we will see here and in Chapter 4 that the given description is sufficient to (at least) estimate the Swan conductor in many settings. The next few sections present basic formalism from representation theory, using [**59**] as a standard reference for some facts that we do not prove here.

Some readers may wish to skip these first sections, until Section 3.5, or even might prefer in a first reading to first look at Chapter 4 to see the Riemann Hypothesis in action, before coming back to learn the formalism.

In Chapter 5, we will consider much deeper constructions, including the very important existence of the Fourier transform for Galois representations.

This chapter uses mostly the language of representation theory. After this, however, we will switch for the most part to speaking in sheaf-theoretic terms.

3.1. Swan conductor

Let x be any point of $\mathbf{P}^1(\bar{k})$. The Swan conductor at x of an ℓ-adic representation of k is computed using a certain filtration $(I(t))_{t\geqslant 0}$ of the inertia group $I = I_x$. This filtration has the following properties:

$$I(0) = I, \qquad I(s) \subset I(t) \text{ if } t \geqslant s,$$
$$\bigcap_{t>0} I(t) = 1, \qquad \bigcap_{s>t} I(s) = I(t) \text{ if } t > 0.$$

Associated to this filtration, one can define (see [**51**, Chapter 1]), for any continuous representation $\varrho : I \longrightarrow \mathrm{GL}(V)$ of I_x, where V is a finite-dimensional ℓ-adic vector space, a canonical decomposition

$$V = \bigoplus_{t\geqslant 0} V(t)$$

which satisfies

$$V(s)^{I(s)} = 0 \text{ for } s > 0 \text{ and hence } V(s)^{I(t)} = 0 \text{ if } 0 < t \leqslant s \tag{3.1}$$

$$V(s)^{I(t)} = V(s) \text{ if } t > s \geqslant 0 \tag{3.2}$$

(note how $t = 0$ plays a special role). Even if ϱ does not extend to a representation of the whole group Π_k, we will speak of ϱ being unramified (meaning $V^I = V$) or tamely ramified (meaning that the Swan conductor vanishes).

Now we have (see [**51**, Def. 1.6]):

Fact. *In these terms, the Swan conductor is given by*

$$\mathrm{Swan}_x(\varrho) = \sum_{t \geqslant 0} t \dim V(t). \tag{3.3}$$

The sum over t is finite, as there are only finitely many $t \geqslant 0$ such that $V(t) \neq 0$. These are called the breaks *of ϱ at x.*

Example 3.1.1. From the formula, it follows that ϱ is unramified or tamely ramified at x, i.e, $\mathrm{Swan}_x(\varrho) = 0$, if and only if $V(t) = 0$ for $t > 0$. Note that, in general, $V(0)$ is not the same as $V^{I(0)} = V^I$.

In analytic applications, we are most interested in estimating from above the Swan conductor. The following lemma often reduces the problem to a simpler invariant.

Lemma 3.1.2. *Let ϱ be a continuous representation of I acting on V as above. Assume $V \neq 0$, and let*

$$\lambda_x(\varrho) = \max\{t \geqslant 0 \mid V(t) \neq 0\}.$$

(1) *We have*

$$\lambda_x(\varrho) \leqslant \mathrm{Swan}_x(\varrho) \leqslant \mathrm{rank}(\varrho)\lambda_x(\varrho).$$

(2) *For any $t \geqslant 0$, we have $t > \lambda_x(\varrho)$ if and only if $t > 0$ and $I(t)$ acts trivially on V.*

Proof. (1) Let $\lambda = \lambda_x(\varrho)$. Since $\dim V(\lambda) \geqslant 1$, and all terms in (3.3) are non-negative, we get

$$\lambda \leqslant \lambda \dim V(\lambda) \leqslant \mathrm{Swan}_x(\varrho),$$

and the second estimate is also clear.

(2) First, $t > \lambda_x(\varrho)$ implies that $t > 0$. Moreover, by definition, $t > \lambda_x(\varrho)$ if and only if all $V(s)$, for $s \geqslant t$, are zero. Hence $t > \lambda_x(\varrho)$ if and only if

$$V = \bigoplus_{s<t} V(s).$$

We now show that this last condition, for $t > 0$, is equivalent with $V^{I(t)} = V$. Indeed, assuming it first, we get

$$V^{I(t)} = \bigoplus_{s<t} V(s)^{I(t)} = \bigoplus_{s<t} V(s) = V$$

by (3.2). Conversely, if $V^{I(t)} = V$ and $t > 0$, then we get

$$V = \Bigl(\bigoplus_{s\geqslant 0} V(s)\Bigr)^{I(t)} = \bigoplus_{s\geqslant 0} V(s)^{I(t)} = \bigoplus_{s<t} V(s)^{I(t)} = \bigoplus_{s<t} V(s)$$

by (3.1) and (3.2). □

Remark 3.1.3. Note that the condition $t > 0$ is important in (2): if $t = 0$, the condition $t > \lambda_x(\varrho)$ always fails, but $I(0) = I$ may nevertheless act trivially on V (and does in fact if and only if ϱ is unramified at x).

3.2. Morphisms, isomorphisms

An important notion that is rather hidden at the level of trace functions is that of *morphism* between representations.

Definition 3.2.1 (Morphism of representations). Let G be a group, E a field and

$$\varrho_1 : G \longrightarrow \mathrm{GL}(V_1), \qquad \varrho_2 : G \longrightarrow \mathrm{GL}(V_2)$$

two representations of G on E-vector spaces. A morphism between ϱ_1 and ϱ_2, or intertwiner from ϱ_1 to ϱ_2, is an E-linear map

$$\Phi : V_1 \longrightarrow V_2$$

such that

$$\Phi(\varrho_1(g)v) = \varrho_2(g)\Phi(v)$$

for all $g \in G$ and $v \in V_1$.

The space of interwiners from ϱ_1 to ϱ_2 is denoted

$$\mathrm{Hom}_G(V_1, V_2) \text{ or } \mathrm{Hom}_G(\varrho_1, \varrho_2).$$

If an intertwiner Φ is invertible, it is easily checked that its inverse Φ^{-1} is also an intertwiner. One then says that Φ is an isomorphism from ϱ_1 to ϱ_2.

It is clear that the identity map on the vector space underlying a representation is an intertwiner from the representation to itself, and the the composition of two intertwiners is an interwiner.

From the point of view of ℓ-adic representations over a finite field, we can therefore speak of morphisms or of isomorphism. However, a crucial variant is often essential:

Definition 3.2.2 (Geometric isomorphism). Let k be a finite field, E a field and

$$\varrho_1 : \Pi_k \longrightarrow \mathrm{GL}(V_1), \qquad \varrho_2 : \Pi_k \longrightarrow \mathrm{GL}(V_2)$$

two representations of Π_k on E-vector spaces. An E-linear map $\Phi : V_1 \longrightarrow V_2$ is a *geometric isomorphism* if Φ is an isomorphism of the restrictions of ϱ_i to the geometric Galois group Π_k^g.

Two representations of Π_k are said to be *geometrically isomorphic* is such a Φ exists.

By the invariance of the trace of linear maps under conjugation, it follows that whenever two finite-dimensional representations of Π_k are isomorphic, their trace functions are equal. It is natural to ask what happens if the representations are only geometrically isomorphic. A partial answer is the following:

Proposition 3.2.3. *Let k be a finite field, E a field and*

$$\varrho_1 : \Pi_k \longrightarrow \mathrm{GL}(V_1), \qquad \varrho_2 : \Pi_k \longrightarrow \mathrm{GL}(V_2)$$

two irreducible representations of Π_k on E-vector spaces. Then ϱ_1 and ϱ_2 are geometrically isomorphic if and only if there exists a morphism

$$\chi : \mathrm{Gal}(\bar{k}/k) \longrightarrow E^\times$$

such that $\varrho_2 \simeq \varrho_1 \otimes \chi$. In this case there exists $\alpha \in E^\times$ such that

$$t_{\varrho_2}(x; \nu) = \alpha^\nu t_{\varrho_1}(x; \nu) \tag{3.4}$$

for all $\nu \geqslant 1$ and $x \in k_\nu$. Furthermore, if ϱ_1 and ϱ_2 are ℓ-adic representations of weight 0, *then α is a Weil number of weight* 0.

In this statement, $\varrho_1 \otimes \chi$ denotes the representation of Π_k on V_1 given by

$$g \mapsto \chi(g)\varrho_1(g).$$

Proof. The first part is [**59**, Prop. 2.8.2]. Then since χ factors through the geometric Galois group, the value of $\chi(\sigma)$ at any Frobenius element $\sigma = \mathrm{Fr}_{x,\nu}$ depends only on $\nu \geqslant 1$ and not on $x \in k_\nu$. Moreover, for any $x \in k_\nu$, we have $(\varrho_1 \otimes \chi)^{I_x} = \varrho_1^{I_x}$ since $I_x \subset \Pi_k^g$ is in the kernel of χ.

Let $\sigma = \mathrm{Fr}_{x,1}$ for some $x \in k$ and $\alpha = \chi(\sigma) \in E^\times$. The element σ^ν is a Frobenius element for $x \in k_\nu$. Hence the value of χ on Frobenius elements for k_ν is α^ν. Thus we get (3.4) with this value of α.

Finally if the representations ϱ_i are ℓ-adic of weight 0, then comparing eigenvalues at any $x \in k_\nu$ for some $\nu \geqslant 1$ which is unramified for ϱ_1, we see that α^ν is a ratio of Weil numbers of weight 0, in particular that it is of weight 0, and hence so is also α (see Exercise 2.2.4, (4)). □

Character theory of finite-dimensional representations gives a stronger property of the trace of any finite-dimensional representation ϱ of any group G: seen as a function on G (not just on a subset such as the Frobenius elements), the trace of ϱ *determines* ϱ up to isomorphism of representations (under some conditions of semisimplicity, see e.g. [**59**, Prop. 2.7.38] for a precise statement).

As we will see in Chapter 4, one interpretation of the Riemann Hypothesis over finite fields is that this property extends to show that the trace function of an ℓ-adic representation ϱ determines it up to isomorphism, *provided* ϱ has small enough conductor with respect to the size of the base finite field. This leads to a concrete parallel for arithmetic and geometric isomorphism of ℓ-adic representations that helps in reasoning about these objects: (1) "two representations are arithmetically isomorphic" means roughly that they have the same trace functions; (2) "two representations are geometrically isomorphic" means roughly that their trace functions are proportional, with a proportionality factor of modulus 1.

We next illustrate the notions of geometric isomorphism by classifying the Kummer and Artin-Schreier representations up to isomorphism.

Theorem 3.2.4 (Classification). *Let k be a finite field and ℓ a prime different from the characteristic of k. Let f_1, $f_2 \in k(T)$ be rational functions.*

(1) *For a multiplicative character χ of $k^\times$ of order $d \geqslant 1$, the Kummer representations $\mathcal{L}_{\chi(f_1)}$ and $\mathcal{L}_{\chi(f_2)}$ are geometrically isomorphic if and only if*

$$\frac{f_1}{f_2} = cg^d$$

for some $c \in k^\times$ and some $g \in k(T)$.

(2) *For a non-trivial additive character ψ of k, the Artin-Schreier representations $\mathcal{L}_{\psi(f_1)}$ and $\mathcal{L}_{\psi(f_2)}$ are geometrically isomorphic if and only if*

$$f_1 - f_2 = g^{|k|} - g + c$$

for some $c \in k$ and some $g \in k(T)$.

Proof. In both case, the "if" part is easy to see. For instance if $f_1/f_2 = cg^d$ then we get

$$\mathcal{L}_{\chi(f_1)} = \mathcal{L}_{\chi(cg^d f_2)} \simeq \mathcal{L}_{\chi(cg^d)} \otimes \mathcal{L}_{\chi(f_2)}$$

geometrically, and the first factor is geometrically trivial. The additive case is similar. Thus, the point is to derive the converse. □

3.3. Irreducible representations

In representation theory, a crucial notion is that of subrepresentations and of *irreducible representations.*

Definition 3.3.1. Let G be a group, E a field and

$$\varrho \,:\, G \longrightarrow \mathrm{GL}(V)$$

a representation of G on an E-vector space.

(1) A subrepresentation of ϱ is a subspace $W \subset V$ such that $\varrho(g)W = W$ for all $g \in G$. In this case, the restriction of $\varrho(g)$ to W defines a representation $G \longrightarrow \mathrm{GL}(W)$.

(2) The representation ϱ is *irreducible* if and only if $V \neq 0$ and there is no subrepresentation $W \subset V$ except for $W = 0$ and $W = V$.

In our context, it turns out to often be useful to look at representations of Π_k which satisfy a slightly stronger property than irreducibility:

Definition 3.3.2. Let k be a finite field, E a field and

$$\varrho \,:\, \Pi_k \longrightarrow \mathrm{GL}(V)$$

a representation of Π_k on an E-vector space. Then ϱ is said to be geometrically irreducible if the restriction of ϱ to the geometric Galois group Π_k^g is irreducible, i.e., if $V \neq 0$, and there is no non-zero proper subspace $W \subset V$ such that $\varrho(\sigma)W = W$ for all $\sigma \in \Pi_k^g$.

For emphasis, a representation $\Pi_k \longrightarrow \mathrm{GL}(V)$ which is irreducible will often be called *arithmetically irreducible*, to avoid any ambiguity with geometric irreducibility.

It may not be easy to determine whether a representation is irreducible or not. However, one case is very simple, and quite important in practice:

Proposition 3.3.3. *Let G be a group, E a field and*

$$\chi \,:\, G \longrightarrow E^\times = \mathrm{GL}_1(E)$$

a one-dimensional E-representation of G. Then G is irreducible.

Indeed, there is simply no subspace of E which is both non-zero and proper!

In particular, we get:

Corollary 3.3.4. *Let k be a finite field, $f \in k(T)$ a rational function. Then for any additive character ψ of k, the Artin-Schreier representation $\mathcal{L}_{\psi(f)}$ is geometrically irreducible. Similarly if $f \neq 0$ and χ is a multiplicative character of $k^\times$, the Kummer representation $\mathcal{L}_{\chi(f)}$ is geometrically irreducible.*

It is natural to ask how much stronger geometric irreducibility is in comparison with arithmetic irreducibility. This is answered by the following proposition:

Proposition 3.3.5. *Let G be a group, $H \lhd G$ a normal subgroup with $G/H = A$ an abelian group. Let $\varrho : G \longrightarrow \mathrm{GL}(V)$ be a finite-dimensional irreducible representation of G on an E-vector space. Then either the restriction of ϱ to H is a direct sum of finitely many irreducible subrepresentations, all isomorphic to each other, and we say that the restriction of ϱ to H is* isotypic, *or there exists a proper finite index subgroup $\tilde{H} > H$ of G and an irreducible representation π of $\tilde{H}$ such that*

$$\varrho \simeq \mathrm{Ind}_{\tilde{H}}^{G}(\pi),$$

in which case we say that ϱ is induced.

We refer to [**59**, Prop. 2.8.1] for the proof, and the terminology if it is unfamiliar. The important corollary in our case is the following fact:

Corollary 3.3.6. *Let k be a finite field, E a field and*

$$\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$$

an irreducible finite-dimensional representation of Π_k on an E-vector space. Then either ϱ is geometrically isotypic, i.e., there exists an irreducible representation

$$\pi : \Pi_k^g \longrightarrow \mathrm{GL}(W)$$

and an integer $n \geqslant 1$ such that we have an isomorphism

$$V \simeq W^n$$

as representations of Π_k^g, or the trace function of ϱ on k is identically zero. In the first case, n is unique and π is unique up to isomorphism.

This corollary implies that in most applications of trace functions, the distinction between geometric and arithmetic irreducibility is not very important, since an isotypic representation behaves often in much the same way as an irreducible one, and obviously those representations with zero trace function are not too complicated...

Proof. We have $\Pi_k^g \lhd \Pi_k$ with quotient $\mathrm{Gal}(\bar{k}/k)$ which is abelian, so we can use the previous proposition. If ϱ is geometrically isotypical, we are done, so we may assume that

$$\varrho \simeq \mathrm{Ind}_{\tilde{H}}^{\Pi_k} \pi$$

for some proper subgroup $\tilde{H}$ of Π_k containing Π_k^g, and some irreducible representation π of $\tilde{H}$.

By group theory and Galois theory, the subgroups $\tilde{H}$ of Π_k containing Π_k^g are of the form

$$\tilde{H} = \mathrm{Gal}(\overline{k(T)}/k_\nu(T))$$

for some $\nu \geqslant 1$, where k_ν is the extension of k of degree ν in $\bar{k}$. The essential point is that no element in a Frobenius conjugacy class corresponding to $x \in k$ belongs to $\tilde{H}$ if $\nu \neq 1$.

On the other hand, because the inertia group at x is a subgroup of Π_k^g (Lemma 2.1.1), one sees that

$$V^{I_x} \simeq \mathrm{Ind}_{\tilde{H}}^{\Pi_k} \pi^{I_x},$$

from which a standard property of induction (see, e.g., [**59**, Example 2.7.44 (3)]) shows that the trace of any element not in I_x is zero. In particular, the trace function of ϱ at x is zero. □

3.4. Complex conjugation

A most basic operation is complex conjugation. The following result is, for the most part, elementary, but its precision is actually by no means immediate.

Proposition 3.4.1. *Let k be a finite field and $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ an ℓ-adic representation over k. Let $\mathrm{D}(\varrho)$ be the contragredient representation*

$$\mathrm{D}(\varrho) : \Pi_k \, \mathrm{GL}(V')$$

where V' is the dual of V, acting by

$$(g \cdot \lambda)(v) = \lambda(\varrho(g^{-1})v)$$

for any $\lambda \in V'$. Then $\mathrm{D}(\varrho)$ is an ℓ-adic representation, with

$$\mathrm{rank}(\mathrm{D}(\varrho)) = \mathrm{rank}(\varrho), \qquad \mathrm{Sing}(\mathrm{D}(\varrho)) = \mathrm{Sing}(\varrho), \tag{3.5}$$

$$\mathrm{Swan}_x(\mathrm{D}(\varrho)) = \mathrm{Swan}_x(\varrho) \tag{3.6}$$

for all x, and in particular

$$\mathrm{c}(\mathrm{D}(\varrho)) = \mathrm{c}(\varrho).$$

Furthermore, we have

$$t_{\mathrm{D}(\varrho)}(x) = \overline{t_\varrho(x)}$$

for all $x \in k$.

Proof. By its definition, $\mathrm{D}(\varrho)$ is an analytic Galois representation of k associated to the dual of the ℓ-adic representation underlying ϱ.

The first property in (3.5) is then immediate, and the second as well as (3.6) follow from [**51**, Lemma 1.5]. Of course, these imply the equality of the conductor.

By representation theory, the eigenvalues of $\mathrm{D}(\varrho)(g)$ are the inverses of the eigenvalues of $\varrho(g)$. In particular, for any unramified $x \in k_\nu$, since the eigenvalues of $\varrho(\mathrm{Fr}_{x,\nu})$ are Weil numbers of weight 0, the eigenvalues of $\mathrm{D}(\varrho(\mathrm{Fr}_{x,\nu}))$ are their conjugates, and hence

$$t_{\mathrm{D}(\varrho)}(x;\nu) = \overline{t_\varrho(x;\nu)}$$

for x unramified.

If x is ramified, the equality

$$t_{\mathrm{D}(\varrho)}(x;\nu) = \overline{t_\varrho(x;\nu)}$$

still holds, but is much deeper: it is a result of Gabber (see [**38**, Th. 3], and [**54**, p. 31]). □

3.5. Addition

Addition is the most basic algebraic operation. In the context of trace functions, its properties are very easy. What is less obvious but very important is how to decompose any trace function in a sum of irreducible ones. We consider these two properties in turn.

Proposition 3.5.1. *Let k be a finite field, E a field and*

$$\varrho_1 : \Pi_k \longrightarrow \mathrm{GL}(V_1), \qquad \varrho_2 : \Pi_k \longrightarrow \mathrm{GL}(V_2)$$

two representations of Π_k on E-vector spaces. Then the representation $\varrho_1 \oplus \varrho_2$ has trace function

$$t_{\varrho_1} + t_{\varrho_2}.$$

If ϱ_1 and ϱ_2 are ℓ-adic representations, then we have

$$\mathrm{rank}(\varrho_1 \oplus \varrho_2) = \mathrm{rank}(\varrho_1) + \mathrm{rank}(\varrho_2), \qquad \mathrm{Sing}(\varrho_1 \oplus \varrho_2) \subset \mathrm{Sing}(\varrho_1) \cup \mathrm{Sing}(\varrho_2),$$

$$\mathrm{Swan}_x(\varrho_1 \oplus \varrho_2) = \mathrm{Swan}_x(\varrho_1) + \mathrm{Swan}_x(\varrho_2),$$

hence

$$\mathrm{c}(\varrho_1 \oplus \varrho_2) \leqslant \mathrm{c}(\varrho_1) + \mathrm{c}(\varrho_2).$$

Proof. Since $(\varrho_1 \oplus \varrho_2)^{I_x} = \varrho_1^{I_x} \oplus \varrho_2^{I_x}$, the formula for the trace function is clear, and so are is that for the rank and the inclusion of $\mathrm{Sing}(\varrho_1 \oplus \varrho_2)$ in the union of the singularities of ϱ_1 and ϱ_2. The formula for the Swan conductor is a consequence of (3.3) and the additivity

$$(V_1 \oplus V_2)(t) = V_1(t) \oplus V_2(t)$$

of the filtrations of Section 3.1 (see [**51**, Prop. 1.1 (3)]). □

Exercise 3.5.2. We will say that two representations have *disjoint singularities* when $\mathrm{Sing}(\varrho_1) \cap \mathrm{Sing}(\varrho_2) = \varnothing$.

Now we decompose trace functions:

Proposition 3.5.3. *Let k be a finite field, and*

$$\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$$

an ℓ-adic representation of Π_k.

(1) *There exists an integer $0 \leqslant m \leqslant \mathrm{rank}(\varrho)$ and a family $(\varrho_i)_{1 \leqslant i \leqslant m}$ of arithmetically irreducible ℓ-adic representations such that*

$$t_\varrho = \sum_{i=1}^{m} t_{\varrho_i}. \tag{3.7}$$

(2) *There exists an integer $0 \leqslant m \leqslant \mathrm{rank}(\varrho)$ and a family $(\varrho_i)_{1 \leqslant i \leqslant m}$ of arithmetically irreducible ℓ-adic representations such that*

$$t_\varrho = \sum_{i=1}^{m} t_{\varrho_i}$$

and each ϱ_i is geometrically isotypic.

Proof. We may assume $V \neq 0$. Then the Jordan-Hölder-Noether Theorem (see, e.g., [**59**, Th. 2.7.1]) shows that there exist $m \geqslant 1$ and subrepresentations

$$0 = V_0 \subset V_1 \subset \cdots \subset V_{m-1} \subset V_m = V$$

such that each quotient representation $\varrho_i = V_i/V_{i-1}$ for $1 \leqslant i \leqslant m$ is (arithmetically) irreducible. This clearly implies that the trace functions of ϱ and that of $\pi = \bigoplus \varrho_i$ coincide on the unramified x, which gives (3.7) for x unramified.

It is again a non-trivial property that the equality holds if x is ramified; this depends on having ℓ-adic representations (and not arbitrary coefficient fields). Precisely, in the setting of Section 3.1, the fact that $I(0) = I_x$ and that sending a representation V to $V(0)$ is *exact* (see [**51**, Prop. 1.1 (3)]) imply that

$$\Bigl(\bigoplus \varrho_i\Bigr)^{I_x} = \bigoplus \varrho_i^{I_x},$$

so that the trace function of $\bigoplus \varrho_i$ at x is

$$\sum_{i=1}^{m} t_{\varrho_i}(x),$$

from which (3.7) follows at x.

In order to prove (2), it is enough to start with (1) and to observe that by Proposition 3.3.5, each arithmetically irreducible representation ϱ_i is either geometrically isotypic (in which case we keep it), or *induced* (in the sense of the statement of the proposition). In that case, Corollary 3.3.6 shows that the trace function of ϱ_i is identically zero. Thus we may drop it without changing the trace function, and therefore obtain a decomposition like (3.7) where each ϱ_i is geometrically isotypic. □

3.6. Multiplication

Multiplication is, in principle, also an easy construction, since the product of the traces of two matrices can always be written as the trace of the tensor product. However, just like the product of two primitive Dirichlet characters might not be primitive, we have to be a bit careful. It is convenient to introduce the following terminology:

Definition 3.6.1. Let k be a finite field, $c \geqslant 1$ an integer and

$$\varphi : k \longrightarrow \mathbf{C}$$

an arbitrary function.

(1) If $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ is an ℓ-adic representation, then we say that φ is c-close to ϱ, or to the trace function of ϱ, if the conductor of ϱ is at most c, and if there exists a set $X \subset k$ with $|X| \leqslant c$ such that

$$\varphi(x) = t_{\varrho}(x)$$

for $x \notin X$, and if moreover $|\varphi(x)| \leqslant c$ for all $x \in X$.

(2) We say that φ is an approximate trace function with conductor $\leqslant c$ if there exists an ℓ-adic representation ϱ such that φ is c-close to ϱ.

Then we have:

Proposition 3.6.2. *Let k be a finite field and*

$$\varrho_1 : \Pi_k \longrightarrow \mathrm{GL}(V_1), \qquad \varrho_2 : \Pi_k \longrightarrow \mathrm{GL}(V_2)$$

two ℓ-adic representations of Π_k with respective trace functions K_1 and K_2. Let $\varrho = \varrho_1 \otimes \varrho_2$.

(1) *The representation ϱ is an ℓ-adic representation. It satisfies*

$$\mathrm{rank}(\varrho) = \mathrm{rank}(\varrho_1)\,\mathrm{rank}(\varrho_2), \qquad \mathrm{Sing}(\varrho) \subset \mathrm{Sing}(\varrho_1) \cup \mathrm{Sing}(\varrho_2),$$

and for any x, we have

$$\mathrm{Swan}_x(\varrho) \leqslant \mathrm{rank}(\varrho)(\mathrm{Swan}_x(\varrho_1) + \mathrm{Swan}_x(\varrho_2)).$$

Moreover

$$\mathrm{c}(\varrho) \leqslant 5\,\mathrm{c}(\varrho_1)^2\,\mathrm{c}(\varrho_2)^2.$$

(2) *The function $\varphi = K_1K_2$ on k is c-close to ϱ, where*

$$c = 5c_1^2c_2^2.$$

(3) *If ϱ_1 and ϱ_2 have disjoint singularities, then the trace function of ϱ is exactly K_1K_2.*

We begin with the lemma that establishes the bound for the Swan conductor.

Lemma 3.6.3. *Let k be a finite field and*

$$\varrho_1 : \Pi_k \longrightarrow \mathrm{GL}(V_1), \qquad \varrho_2 : \Pi_k \longrightarrow \mathrm{GL}(V_2)$$

two ℓ-adic representations of Π_k. For any $x \in \mathbf{P}^1(\bar{k})$, we have

$$\mathrm{Swan}_x(\varrho_1 \otimes \varrho_2) \leqslant \mathrm{rank}(\varrho_1)\,\mathrm{rank}(\varrho_2)(\mathrm{Swan}_x(\varrho_1) + \mathrm{Swan}_x(\varrho_2)).$$

Note that the formula also holds for unramified x, since the left-hand side is equal to zero, and so is the right-hand side.

Proof. Let $r_i = \mathrm{rank}(\varrho_i)$. We have

$$\mathrm{Swan}_x(\varrho_1 \otimes \varrho_2) \leqslant r_1 r_2 \lambda_x(\varrho_1 \otimes \varrho_2)$$

by the first part of Lemma 3.1.2. Then, for any $t > 0$, note that if $t > \max(\lambda_x(\varrho_1), \lambda_x(\varrho_2))$, the group $I(t)$ acts trivially on both the underlying vector spaces V_1 and V_2 (by the second part of Lemma 3.1.2), hence also on $V_1 \otimes V_2$, so that (similarly) we have $t > \lambda_x(\varrho_1 \otimes \varrho_2)$. This means that

$$\lambda_x(\varrho_1 \otimes \varrho_2) \leqslant \max(\lambda_x(\varrho_1), \lambda_x(\varrho_2)),$$

and we deduce

$$\mathrm{Swan}_x(\varrho_1 \otimes \varrho_2) \leqslant r_1 r_2 \max(\lambda_x(\varrho_1, \lambda_x(\varrho_2))) \leqslant r_1 r_2 \max(\mathrm{Swan}_x(\varrho_1), \mathrm{Swan}_x(\varrho_2))),$$

which is (slightly more precise than) Lemma 3.6.3. □

Proof of Proposition 3.6.2. First of all, the representation ϱ is unramified outside $S = \mathrm{Sing}(\varrho_1) \cup \mathrm{Sing}(\varrho_2)$, since

$$\varrho^{I_x} = (\varrho_1 \otimes \varrho_2)^{I_x} = \varrho_1 \otimes \varrho_2$$

if I_x acts trivially on both ϱ_1 and ϱ_2. For $x \notin S$, linear algebra says that the eigenvalues of $\mathrm{Fr}_{x,\nu}$ on $\varrho_1 \otimes \varrho_2$ are all the products $\alpha\beta$ of the eigenvalues of $\mathrm{Fr}_{x,\nu}$ on ϱ_1 and ϱ_2 respectively. Hence they are Weil numbers of weight 0. Now by Theorem 2.2.9 (1), it follows that ϱ is of weight 0.

For all x, the Swan conductor at x of ϱ is estimated by Lemma 3.6.3. Denoting $r_i = \mathrm{rank}(\varrho_i)$, $n_i = |\mathrm{Sing}(\varrho_i)|$ and $c_i = \mathrm{c}(\varrho_i)$, we then compute

$$\begin{aligned}\mathrm{c}(\varrho) &\leqslant r_1 r_2 + n_1 + n_2 + r_1 r_2 \sum_{x \in S} (\mathrm{Swan}_x(\varrho_1) + \mathrm{Swan}_x(\varrho_2)) \\ &\leqslant c_1 c_2 + c_1 + c_2 + c_1 c_2 (c_1 + c_2) \leqslant 5 c_1^2 c_2^2.\end{aligned}$$

This finishes the proof of (1).

(2) By linear algebra again, we have

$$\varphi(x) = K_1(x) K_2(x)$$

for all $x \notin S$. We have $|S| \leqslant c_1 + c_2 \leqslant c$ and

$$|\varphi(x)| \leqslant r_1 r_2 \leqslant c$$

for $x \in S$ by Theorem 2.2.9, (2), so that the result follows from (1) and the definition of a function c-close to ϱ.

(3) If the singularities of ϱ_1 and ϱ_2 are disjoint, then we can always use one of the formulas

$$\varrho^{I_x} = \varrho_1 \otimes \varrho_2^{I_x}$$

for $x \notin \mathrm{Sing}(\varrho_1)$ or

$$\varrho^{I_x} = \varrho_1^{I_x} \otimes \varrho_2$$

for $x \notin \mathrm{Sing}(\varrho_2)$, to conclude that

$$\varphi(x) = K_1(x) K_2(x)$$

for all x without exceptions. □

EXERCISE 3.6.4. Let $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ be an ℓ-adic representation. Let

$$\Lambda : \mathrm{GL}(V) \longrightarrow \mathrm{GL}(W)$$

be a continuous homomorphism. Show that $\Lambda \circ \varrho$ is an ℓ-adic representation, and prove that for any x, we have

$$\mathrm{Swan}_x(\Lambda \circ \varrho) \leqslant \dim(W) \,\mathrm{Swan}_x(\varrho).$$

EXERCISE 3.6.5. Let k be a finite field and

$$\varrho_1 : \Pi_k \longrightarrow \mathrm{GL}(V_1), \qquad \varrho_2 : \Pi_k \longrightarrow \mathrm{GL}(V_2)$$

two ℓ-adic representations of Π_k. Show that $\mathrm{Sing}(\varrho_1 \otimes \varrho_2) = \mathrm{Sing}(\varrho_1) \cup \mathrm{Sing}(\varrho_2)$ if ϱ_1 and ϱ_2 have disjoint singularities.

3.7. Change of variable

Changing the variable in a trace function is one of the easiest operations. We distinguish first a change of variable by a fractional linear transformation, since these arise quite often in applications.

If K is any field, we recall that $\mathrm{PGL}_2(K)$ acts on the projective line $\mathbf{P}^1(\bar{K})$ by

$$\gamma \cdot x = \frac{ax+b}{cx+d}$$

where γ is the class of the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(K).$$

For a fixed γ, the operation $f \mapsto f \circ \gamma$ is an automorphism of the field $K(T)$ of rational functions over K. It induces a group homomorphism

$$\gamma_* : \mathrm{Gal}(\overline{K(T)}/K(T)) \longrightarrow \mathrm{Gal}(\overline{K(T)}/K(T))$$

where $\overline{K(T)}$ is an algebraic closure of $K(T)$, in the following manner: for any finite Galois extension $E/k(T)$, generated by a root Z of an equation

$$a_0(T) + a_1(T)X + \cdots + a_{d-1}(T)X^{d-1} + X^d = 0$$

with $a_i \in K(T)$, we get an associated Galois extension $E_\gamma/k(T)$ generated by a root Z_γ of

$$a_0(\gamma(T)) + a_1(\gamma(T))X + \cdots + a_{d-1}(\gamma(T))X^{d-1} + X^d = 0,$$

together with an isomorphism

$$E \longrightarrow E_\gamma$$

mapping Z to Z_γ. Then for any $\sigma \in \mathrm{Gal}(E/K(T))$, we obtain a corresponding element $\gamma_*(\sigma) \in \mathrm{Gal}(E_\gamma/K(T))$ by composition

$$E_\gamma \longrightarrow E \xrightarrow{\sigma} E \longrightarrow E_\gamma.$$

This is an homomorphism

$$\mathrm{Gal}(E/K(T)) \longrightarrow \mathrm{Gal}(E_\gamma/K(T))$$

and putting together all extensions E, this gives the desired homomorphism.

Proposition 3.7.1 (Fractional linear transformation). *Let k be a finite field and let $\gamma \in \mathrm{PGL}_2(k)$.*

(1) *Let E be any field. If $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ is any Galois representation of k on an E-vector space V, the representation $\gamma^*\varrho$ defined by*

$$\gamma^*\varrho = \varrho \circ \gamma_* : \Pi_k \longrightarrow \mathrm{GL}(V)$$

satisfies

$$t_{\gamma*\varrho}(x;\nu) = t_\varrho(\gamma(x);\nu)$$

for any $\nu \geqslant 1$ and any $x \in \mathbf{P}^1(k_\nu)$.

(2) *Let $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ be an ℓ-adic representation of k. Then we have*

$$\mathrm{Sing}(\gamma^*\varrho) = \gamma^{-1}(\mathrm{Sing}(\varrho)),$$

and for all $x \in \mathbf{P}^1(\bar{k})$, we have

$$\mathrm{Swan}_x(\gamma^*\varrho) = \mathrm{Swan}_{\gamma^{-1}(x)}(\varrho).$$

In particular, we have

$$\mathrm{c}(\gamma^*\varrho) = \mathrm{c}(\varrho).$$

(3) *If ϱ is irreducible, resp. geometrically irreducible, resp. isotypic, resp. geometrically isotypic, then so is $\varphi_*\varrho$.*

Proof. TODO □

Example 3.7.2. Among the most important examples are the elements

$$u_h = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad c_a = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$$

in $\mathrm{PGL}_2(k)$, which act by $x \mapsto x + a$ and $x \mapsto ax$, respectively. We will write

$$[+h]^*\varrho = u_h^*\varrho, \qquad [-h]^*\varrho = u_{-h}^*\varrho, \qquad [\times a]^*\varrho = c_a^*\varrho. \tag{3.8}$$

More generally, consider now an arbitrary non-constant rational function $\varphi \in k(T)$. We wish to obtain in a similar manner a change of variable from $t_\varrho(x)$ to $t_\varrho(\varphi(x))$. However, in general, the function $x \mapsto t_\varrho(\varphi(x))$ is only an approximate trace function (as in Section 3.6). Moreover, if P is not a fractional linear transformation, this operation is not bijective anymore, so the conductor may become a bit larger.

Proposition 3.7.3 (Polynomial change of variable). *Let k be a finite field of characteristic p and let $\varphi \in k(T)$ be a non-constant rational function. Let $d \geqslant 1$ be the degree of φ as a map $\mathbf{P}^1(\bar{k}) \longrightarrow \mathbf{P}^1(\bar{k})$.*

(1) *Let E be any field. If $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ is any Galois representation of k on an E-vector space V, the representation $\varphi^*\varrho$ defined by*

$$\varphi^*\varrho = \varrho \circ \varphi_* : \Pi_k \longrightarrow \mathrm{GL}(V)$$

satisfies

$$\mathrm{Sing}(\varphi^*\varrho) \subset \varphi^{-1}(\mathrm{Sing}(\varrho))$$

and

$$t_{\varphi*\varrho}(x;\nu) = t_\varrho(\varphi(x);\nu)$$

for any $\nu \geqslant 1$ and any $x \in \mathbf{P}^1(k_\nu)$ such that $\varphi(x) \notin \mathrm{Sing}(\varrho)$.

(2) *Let $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ be an ℓ-adic representation of k. Assume that no zero or pole of φ is of order $\geqslant p$. Then we have*

$$\mathrm{c}(\varphi^*\varrho) \leqslant 3d^2\,\mathrm{c}(\varrho)^2.$$

Moreover, the function $t_\varrho \circ \varphi$ is c-close to $t_{\varphi*\varrho}$, where $c = 3d^2\,\mathrm{c}(\varrho)^2$.

TODO. □

Example 3.7.4. We can easily illustrate the fact that the trace function of $\varphi^*\varrho$ is not always equal to $t_\varrho \circ \varphi$. Let k have odd characteristic, and let χ_2 denote the non-trivial real character of $k^\times$. Take $\varphi = T^2$ and consider the ℓ-adic representation $\varrho = \mathcal{L}_{\chi_2(X)}$ associated to this character (see Section 2.3). Then we have

$$t_\varrho(\varphi(x)) = t_\varrho(x^2) = \begin{cases} 1 & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

But on the other hand, the construction shows that $\varphi^*\varrho$ is the trivial one-dimensional representation, and therefore that it satisfies

$$t_{\varphi^*\varrho}(x) = 1$$

for *all* $x \in k$, including $x = 0$.

3.8. Summing over solutions of a polynomial equation

The next construction is also quite frequently useful in application. From the representation-theoretic point of view, it is a version of *induction*. Analytically, given a fixed non-constant rational function $\varphi \in k(T)$, it will replace a trace function K by the function $\varphi_* K$ defined by

$$(\varphi_* K)(x) = \sum_{\substack{y \in \mathbf{P}^1(k) \\ \varphi(y) = x}} K(y).$$

In particular, for $K = 1$, we obtain the function that counts the number of solutions of the equation $\varphi(y) = x$.

To define the corresponding representations, we need some preliminaries. Let $\varphi \in k(T)$ be non-constant. The field $k(\varphi(T))$ generated by φ is a subfield of $k(T)$, and the corresponding extension $k(T)/k(\varphi(T))$ is of finite degree $d \geqslant 1$, which is also the degree of φ as a map $\mathbf{P}^1(\bar{k}) \longrightarrow \mathbf{P}^1(\bar{k})$. Its Galois group can thus be naturally identified with a finite-index subgroup of Π_k, which we denote $\Pi_{k,\varphi}$.

Proposition 3.8.1 (Summing over solutions). *Let k be a finite field of characteristic p and let $\varphi \in k(T)$ be a non-constant rational function. Let $d \geqslant 1$ be its degree. Let $S_\varphi \subset \mathbf{P}^1(\bar{k})$ be the set of singular values of φ, i.e., the set of those $x \in \mathbf{P}^1(\bar{k})$ such that $\varphi^{-1}(x)$ contains less than d points.*

(1) *Let E be any field. If $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ is any Galois representation of k on an E-vector space V, the representation $\varphi_*\varrho$ defined by*

$$\varphi_*\varrho = \mathrm{Ind}_{\Pi_{k,\varphi}}^{\Pi_k} \varrho \,:\, \Pi_k \longrightarrow \mathrm{GL}(V)$$

has rank $d\,\mathrm{rank}(\varrho)$. It satisfies

$$t_{\varphi_*\varrho}(x;\nu) = \sum_{\substack{y \in \mathbf{P}^1(k_\nu) \\ \varphi(y) = x}} t_\varrho(y;\nu)$$

for any $\nu \geqslant 1$ and any $x \in \mathbf{P}^1(k_\nu)$.

(2) *Let $\varrho : \Pi_k \longrightarrow \mathrm{GL}(V)$ be an ℓ-adic representation of k. Assume that $d < p$. Then we have*

$$\mathrm{Sing}(\varphi_*(\varrho)) \subset \varphi(\mathrm{Sing}(\varrho)) \cup S_\varphi$$

and

$$\mathrm{Swan}_x(\varphi_*\varrho) = \sum_{\substack{y \in \mathbf{P}^1(\bar{k}) \\ \varphi(y)=x}} \mathrm{Swan}_x(\varrho)$$

for all x. Furthermore we have

$$\mathrm{c}(\varphi_*\varrho) \leqslant 5d^2\,\mathrm{c}(\varrho)^2.$$

TODO. □

3.9. Summary

CHAPTER 4

The Riemann Hypothesis

4.1. A version of Deligne's theorem

With the precise formalism of ℓ-adic representations in place, we may state a version of Deligne's Riemann Hypothesis in this context. We will then show in the next section how to deduce the "black-box" version of Theorem 1.2.1. It will then be shown how to exploit the more intrinsic version in this chapter.

The following encapsulates Deligne's (almost) most general version of the Riemann Hypothesis, as well as some important features from algebraic geometry, in order to give a convenient statement for applications.

Theorem 4.1.1 (Deligne). *Let k be a finite field and ℓ a prime different from the characteristic p of k. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ denote two middle-extension ℓ-adic sheaves on k. Denote $c_i = \mathrm{c}(\mathcal{F}_i)$ and denote K_i the trace function of $\mathcal{F}_i$. Assume that $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically irreducible.*

(1) *If $\mathcal{F}_1$ and $\mathcal{F}_2$ are* not *geometrically isomorphic then*

$$\Big|\sum_{x\in k} K_1(x)\overline{K_2(x)}\Big| \leqslant 4c_1^2c_2^2\sqrt{|k|}. \tag{4.1}$$

(2) *If $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic then there exists a Weil-number α of weight* 0 *such that $K_2 = \alpha K_1$, and we have*

$$\Big|\sum_{x\in k} K_1(x)\overline{K_2(x)} - \bar{\alpha}|k|\Big| \leqslant 4c_1^2c_2^2\sqrt{|k|}. \tag{4.2}$$

Remark 4.1.2. The constant 4 has no particular meaning. In fact, in many concrete cases, we will see that it is possible to replace the factor $4c_1^2c_2^2$ by some other constant that may be much sharper.

We explain the main steps of the argument. This does not by any means provide even a sketch of the full proof, but it will reduce this statement – which is not found in this exact form in the works of Deligne, as far as we know – to precise results of Deligne, Grothendieck and others that are found in the literature. Moreover, this will introduce some important invariants of Galois representations or sheaves, which will play some role in some of the later applications.

Let

$$S = \sum_{x\in k} K_1(x)\overline{K_2(x)}$$

denote the sum that we wish to understand. We also denote by r_i the rank of $\mathcal{F}_i$, and by n_i the number of singularities of $\mathcal{F}_i$. We also denote by ι the isomorphism $\iota : \bar{\mathbf{Q}}_\ell \longrightarrow \mathbf{C}$ that is used to transfer the ℓ-adic trace function to complex-valued functions.

Step 1. Let U be the complement in the projective line of the union of the set of singularities of $\mathcal{F}_1$ and $\mathcal{F}_2$. The sum

$$T = \sum_{x \in U(k)} K_1(x)\overline{K_2(x)}$$

satisfies

$$|S - T| \leqslant (n_1 + n_2)\|K_1\overline{K_2}\|_\infty \leqslant (n_1 + n_2)r_1 r_2 \tag{4.3}$$

by Theorem 2.2.9 (2).

Step 2. Let

$$\mathcal{F} = \mathcal{F}_1 \otimes \mathrm{D}(\mathcal{F}_2). \tag{4.4}$$

Because the sheaves $\mathcal{F}_i$ are of weight 0 and unramified on U, this is a sheaf of weight 0 unramified on U. Its trace function K satisfies

$$K(x) = K_1(x)\overline{K_2(x)}$$

for $x \in U(k)$ (see Proposition 3.4.1, noting that here do not need the deeper property of coincidence of the trace functions at ramified x) and hence

$$T = \sum_{x \in U(k)} K(x).$$

Remark 4.1.3. In fact, it follows from a deep result of Gabber [] that

$$t_{\mathrm{D}(\varrho_2)}(x) = \overline{t_{\varrho_2}(x)}$$

for all $x \in k$, including possible ramified points.

Step 3. We now appeal to the trace formula of Grothendieck:

Theorem 4.1.4 (Trace formula; Grothendieck, Verdier). *Let k be a finite field, U the complement of finitely many points in the projective line over k. Let $\mathcal{F}$ be an ℓ-adic middle-extension sheaf over k, lisse on U. We have*

$$\sum_{x \in U(k)} t_{\mathcal{F}}(x) = M_0 - M_1 + M_2,$$

where

$$M_i = \mathrm{Tr}(\mathrm{Fr}_k \mid H^i_c(U \times \bar{k}, \mathcal{F})).$$

This result introduces the étale cohomology groups $H^i_c(U \times \bar{k}, \mathcal{F})$ of the sheaf $\mathcal{F}$. These were defined and studied by Grothendieck and his collaborators. They are finite-dimensional $\bar{\mathbf{Q}}_\ell$-vector spaces, with an action of the Galois group $\mathrm{Gal}(\bar{k}/k)$. In particular, the geometric Frobenius automorphism $\mathrm{Fr}_k \in \mathrm{Gal}(\bar{k}/k)$ acts on $H^i_c(U \times \bar{k}, \mathcal{F})$, and M_i is the corresponding trace. Thus M_i is an element of $\bar{\mathbf{Q}}_\ell$.

Applying this result to $\mathcal{F}$ given by (4.4) and applying ι on both sides, we derive

$$T = N_0 - N_1 + N_2,$$

where

$$N_i = \iota\Big(\mathrm{Tr}(\mathrm{Fr}_k \mid H^i_c(U \times \bar{k}, \mathcal{F}_1 \otimes \mathrm{D}(\mathcal{F}_2)))\Big).$$

Step 4. We apply the following non-trivial property of the étale cohomology groups:

Proposition 4.1.5. *Let k be a finite field, $U \neq \varnothing$ the complement of finitely many points in the projective line over k. Let $\mathcal{F}$ be an ℓ-adic middle-extension sheaf over k, lisse on U. We have*

$$H^0_c(U \times \bar{k}, \mathcal{F}) = 0.$$

TODO. □

This reduces the formula above to

$$T = -N_1 + N_2.$$

Step 5. We now consider the H^2_c cohomology group. Another property of these groups is:

Proposition 4.1.6. *Let k be a finite field, U the complement of finitely many points in the projective line over k. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be geometrically irreducible ℓ-adic middle-extension sheaves over k, both lisse on U. We have*

$$H^2_c(U \times \bar{k}, \mathcal{F}_1 \otimes \mathrm{D}(\mathcal{F}_2)) = 0$$

unless $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic. If they are geometrically isomorphic, say

$$\mathcal{F}_2 \simeq \gamma^{\deg(\cdot)} \otimes \mathcal{F}_1,$$

then $H^2_c(U \times \bar{k}, \mathcal{F}_1 \otimes \mathrm{D}(\mathcal{F}_2))$ has dimension 1 and Fr_k acts on this vector space by multiplication by $\gamma^{-1}|k|$.

TODO. □

This leads to $N_2 = 0$ unless $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic, and $N_2 = \bar{\alpha}|k|$ if they are, where $\alpha = \iota(\gamma)$ in terms of the Weil number α with $\mathcal{F}_2 \simeq \gamma^{\deg(\cdot)} \otimes \mathcal{F}_1$. (Since α has weight 0, we have $\iota(\gamma^{-1}) = \bar{\alpha}$.)

Step 6. To study N_1, we rely on the Riemann Hypothesis, as proved by Deligne. Here is a version of his result:

Theorem 4.1.7 (Deligne). *Let k be a finite field, U the complement of finitely many points in the projective line over k. Let $\mathcal{F}$ be an ℓ-adic middle-extension sheaf over k, lisse on U. i*

Then all eigenvalues of Fr_k acting on $H^1_c(U \times \bar{k}, \mathcal{F})$ are $|k|$-Weil numbers of weight at most 1.

This result applied to N_1 gives immediately

$$|N_1| \leqslant \dim H^1_c(U \times \bar{k}, \mathcal{F})\sqrt{|k|}.$$

Remark 4.1.8. In many cases, the eigenvalues of the Frobenius are in fact of weight exactly 1, and therefore their modulus is $|k|^{1/2}$. This exponent $1/2$ is the mark of the Riemann Hypothesis, as the statement can be interpreted as saying that certain L-functions have zeros on the critical line $\mathrm{Re}(s) = 1/2$ (see ... below).

Step 7. The combination of the previous steps leads to

$$|S| \leqslant \dim H^1_c(U \times \bar{k}, \mathcal{F})\sqrt{|k|} + (n_1 + n_2)r_1r_2 \tag{4.5}$$

if $\mathcal{F}_1$ and $\mathcal{F}_2$ are not geometrically isomorphic, and

$$|S - \bar{\alpha}|k|| \leqslant \dim H^1_c(U \times \bar{k}, \mathcal{F})\sqrt{|k|} + (n_1 + n_2)r_1r_2 \tag{4.6}$$

otherwise. In order to deduce a useful estimate, we are therefore reduced to finding a bound for the dimension of the first cohomology group. For this purpose, we use the following other deep result:

Theorem 4.1.9 (Euler-Poincaré formula; Grothendieck–Ogg–Shafarevich). *Let k be a finite field, U the complement of $n \geqslant 0$ points in the projective line over k. Let $\mathcal{F}$ be an ℓ-adic middle-extension sheaf over k, lisse on U. We have*

$$h^0 - h^1 + h^2 = \operatorname{rank}(\mathcal{F})(2-n) - \sum_{x \in \operatorname{Sing}(\mathcal{F})} \operatorname{Swan}_x(\mathcal{F}), \tag{4.7}$$

where

$$h^i = \dim H_c^i(U \times \bar{k}, \mathcal{F}).$$

The point of this formula is that the alternating sum of the dimensions of the cohomology groups, among which only that of H_c^1 is really mysterious, has an expression in terms of elementary invariants of $\mathcal{F} = \mathcal{F}_1 \otimes \mathrm{D}(\mathcal{F}_2)$.

In order to exploit this formula, we need finally to estimate the Swan conductors of $\mathcal{F}$ in terms of invariants of the factors $\mathcal{F}_1$ and $\mathcal{F}_2$.

Step 8. We use for this purpose the bound of Lemma 3.6.3, which gives

$$\operatorname{Swan}_x(\mathcal{F}_1 \otimes \mathcal{F}_2) \leqslant r_1 r_2(\operatorname{Sing}(\mathcal{F}_1) + \operatorname{Sing}(\mathcal{F}_2))$$

for all $x \notin U$. In particular, summing over x, we get

$$\begin{aligned} \sum_{x \in \operatorname{Sing}(\mathcal{F}_1 \otimes \mathcal{F}_2)} \operatorname{Swan}_x(\mathcal{F}_1 \otimes \mathcal{F}_2) &\leqslant r_1 r_2 \sum_{x \in \operatorname{Sing}(\mathcal{F}_1) \cup \operatorname{Sing}(\mathcal{F}_2)} \max(\operatorname{Swan}_x(\mathcal{F}_1), \operatorname{Swan}_x(\mathcal{F}_2))) \\ &\leqslant r_1 r_2 \Big(\sum_{x \in \operatorname{Sing}(\mathcal{F}_1)} \operatorname{Swan}_x(\mathcal{F}_1) + \sum_{x \in \operatorname{Sing}(\mathcal{F}_2)} \operatorname{Swan}_x(\mathcal{F}_2)) \Big) \\ &\leqslant r_1 r_2 (s_1 + s_2) \end{aligned} \tag{4.8}$$

where s_i is the sum of the Swan conductors of $\mathcal{F}_i$.

Step 9. We can now bring everything together. From the consequence (4.8) to Lemma 3.6.3 and Theorem 4.1.9 we obtain a bound for the dimension of the first cohomology group, namely

$$\begin{aligned} \dim H_c^1(U \times \bar{k}, \mathcal{F}) &= \dim H_c^2(U \times \bar{k}, \mathcal{F}) + (n-2)\operatorname{rank}(\mathcal{F}) + \sum_{x \in \operatorname{Sing}(\mathcal{F})} \operatorname{Swan}_x(\mathcal{F}) \\ &\leqslant 1 + r_1 r_2 (n_1 + n_2 - 2) + r_1 r_2 (s_1 + s_2) \\ &\leqslant r_1 r_2 (n_1 + s_1 + n_2 + s_2) \leqslant r_1 r_2 (c_1 + c_2) \end{aligned}$$

(since the rank of a geometrically irreducible representation is at least 1).

If $\mathcal{F}_1$ and $\mathcal{F}_2$ are not geometrically isomorphic, we then get from (4.5) the estimate

$$\begin{aligned} |S| &\leqslant r_1 r_2 \Big(n_1 + n_2 + (c_1 + c_2)\sqrt{|k|} \Big) \\ &\leqslant r_1 r_2 \Big((c_1 + c_2)(1 + \sqrt{|k|}) \Big) \\ &\leqslant 2 c_1^2 c_2^2 (1 + \sqrt{|k|}) \leqslant 4 c_1^2 c_2^2 \sqrt{|k|} \end{aligned}$$

since $|k| \geqslant 2$. And if $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic, we get instead from (4.6) the bound

$$|S - \bar{\alpha}|k|| \leqslant 4 c_1^2 c_2^2 \sqrt{|k|}$$

in the same manner. These estimates are exactly what was claimed in Theorem 4.1.1.

4.2. More precise computations

The steps of the previous section can be applied also in particular cases, and lead to stronger estimates than the generic version of Theorem 4.1.1, when one uses precise information on ramification to compute *exactly* the dimension of the first cohomology space using the Euler-Poincaré formula. We illustrate this point here in a few cases, to obtain in particular the Hasse bound (1.2) and the Weil bound (1.3). Other applications will come later.

Example 4.2.1 (Hasse bound). Let p be an odd prime and let

$$K(x) = \Bigl(\frac{x^3 + \alpha x + \beta}{p}\Bigr)$$

where α and β are elements of $\mathbf{F}_p$ such that the cubic polynomial $P = X^3 + \alpha X + \beta$ has three distinct roots in $\bar{\mathbf{F}}_p$. For any fixed $\ell \neq p$, there is then a corresponding ℓ-adic Kummer sheaf $\mathcal{F} = \mathcal{L}_{(P(X)/p)}$ as in Section 2.3 with trace function

$$t_{\mathcal{F}}(x) = K(x)$$

for all $x \in \mathbf{F}_p$. Let $S \subset \bar{\mathbf{F}}_p$ be the set of zeros of P. Then $\mathcal{F}$ is lisse on $U = \mathbf{A}^1 - S$. We have

$$\sum_{x \in \mathbf{F}_p} \Bigl(\frac{x^3 + \alpha x + \beta}{p}\Bigr) = \sum_{x \in U(\mathbf{F}_p)} t_{\mathcal{F}}(x),$$

since $t_{\mathcal{F}}(x) = 0$ for any root $x \in S$ of P. The trace formula applied to $\mathcal{F}$ (which is geometrically irreducible because it has rank 1) on U gives

$$\sum_{x \in U(\mathbf{F}_p)} t_{\mathcal{F}}(x) = -\operatorname{Tr}(\mathrm{Fr}_p \mid H^1_c(U \times \bar{\mathbf{F}}_p, \mathcal{F}))$$

since the other cohomology groups of $\mathcal{F}$ vanish (by Proposition 4.1.5 and Proposition 4.1.6, the second applied to $\mathcal{F}_1 = \mathcal{F}$ and $\mathcal{F}_2$ the trivial rank 1 sheaf; then $\mathcal{F}_1$ is not geometrically isomorphic to $\mathcal{F}_1$ because $\mathcal{F}$ is not geometrically trivial).

Since $\mathcal{F}$ has rank 1 and is *tamely* ramified at the $n = 4$ distinct points $x \in S \cup \{\infty\}$, the Euler-Poincaré formula (4.12) leads to

$$\dim H^1_c(U \times \bar{\mathbf{F}}_p, \mathcal{F}) = n - 2 = 4 - 2 = 2.$$

Deligne's Theorem 4.1.7 then applies to derive the bound

$$\Bigl|\sum_{x \in \mathbf{F}_p} \Bigl(\frac{x^3 + \alpha x + \beta}{p}\Bigr)\Bigr| \leqslant 2\sqrt{p},$$

which is the precise Hasse bound (1.2). An additional geometric ingredient (a form of Poincaré duality) shows that, in this case, the two eigenvalues of the Frobenius acting on $H^1_c(U \times \bar{\mathbf{F}}_p, \mathcal{F})$ are of weight 1 exactly (and not just of weight $\leqslant 1$, as given by Theorem 4.1.7).

Example 4.2.2 (The Weil bound). Let p be a prime number and $a \in \mathbf{F}_p^\times$. We can proceed along similar lines with

$$K(x) = e\Bigl(\frac{ax + \bar{x}}{p}\Bigr)$$

to estimate Kloosterman sums. Let $\varphi(X) = aX + X^{-1} \in \mathbf{F}_p(X)$ and $\mathcal{F} = \mathcal{L}_{\psi(\varphi(X))}$ over $\mathbf{F}_p$, where ψ is the additive character $\psi(x) = e(x/p)$ (and we use any prime $\ell \neq p$ to define $\mathcal{F}$). Then $\mathcal{F}$ is of rank 1, ramified only at 0 and ∞, and the trace formula gives

$$\mathrm{Kl}_2(a;p) = -\frac{1}{\sqrt{p}} \sum_{x \in \mathbf{F}_p} t_{\mathcal{F}}(x) = \frac{1}{\sqrt{p}} \operatorname{Tr}(\mathrm{Fr}_p \mid H^1_c(U \times \bar{\mathbf{F}}_p, \mathcal{F})),$$

where $U = \mathbf{A}^1 - \{0\}$.

By (4.12), we obtain

$$\dim H^1_c(U \times \bar{\mathbf{F}}_p, \mathcal{F}) = 0 + \mathrm{Swan}_0(\mathcal{F}) + \mathrm{Swan}_\infty(\mathcal{F}) = 2$$

by the computation of Swan conductors of Artin-Schreier sheaves (where we use the fact that $a \neq 0$ to see that the Swan conductor at ∞ is really 1).

Therefore Theorem 4.1.7 gives the bound

$$|\mathrm{Kl}_2(a;p)| \leqslant 2$$

as in (1.3).

Exercise 4.2.3. Let p be a prime and $\psi(x) = e(x/p)$ for $x \in \mathbf{F}_p$. Prove that

$$\Big| \sum_{x \in \mathbf{F}_p} \psi\Big(\frac{a}{x+d} + \frac{b}{x} + cx\Big)\Big| \leqslant 3\sqrt{p}$$

for all $(a, b, c, d) \in \mathbf{F}_p$ such that either (1) $d = 0$ and $a + b \neq 0$; or (2) $d \neq 0$ and $(a, b, c) \neq (0, 0, 0)$.

4.3. The black-box version

4.4. First consequences of Deligne's Theorem

Interpreted as a quasi-orthogonality statement, Theorem 4.1.1 has a number of very important consequences that go in the direction of showing that, for small enough conductors, the trace function of a geometrically irreducible middle-extension sheaf determines it uniquely. These statements, besides their intrinsic interest, suggest some fascinating problems and questions mixing algebraic geometry and analytic number theory, which we will present here, and discuss later in some cases.

Proposition 4.4.1. *Let k be a finite field and ℓ a prime different from the characteristic p of k. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ denote two middle-extension ℓ-adic sheaves on k. Denote $c_i = \mathrm{c}(\mathcal{F}_i)$ and denote K_i the trace function of $\mathcal{F}_i$. Assume that $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically irreducible.*

(1) *If $K_1 = K_2$ and $c_i \leqslant \frac{1}{2}|k|^{1/8}$, then $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic.*

(2) *Let*

$$Z = \{x \in k \mid K_1(x) = 0\}.$$

Then we have

$$|Z| \leqslant |k|\Big(1 - \frac{1}{2c_1^2}\Big),$$

provided $c_1 \leqslant \frac{1}{2}|k|^{1/8}$.

(3) *Let*

$$Z' = \{x \in k \mid K_1(x) = K_2(x)\}.$$

Then if $\mathcal{F}_1$ and $\mathcal{F}_2$ are not geometrically isomorphic, we have

$$|Z'| \leqslant |k|\Big(1 - \frac{1}{4c^2}\Big),$$

provided c_1, $c_2 \leqslant \frac{1}{2}|k|^{1/8}$.

(4) *If* K_1 *is identically zero, then*

$$c \geqslant \frac{1}{2}|k|^{1/8}.$$

Proof. All of these facts follow by playing the two parts of Theorem 4.1.1 against each other. We denote by c the maximum of c_1 and c_2.

(1) We proceed by contraposition, assuming that $\mathcal{F}_1$ is not geometrically isomorphic to $\mathcal{F}_2$. Denoting

$$S = \sum_{x \in k} K_1(x)\overline{K_2(x)} = \sum_{x \in k} |K_1(x)|^2,$$

we obtain then from (4.10) the bound

$$|S| \leqslant 4c^4\sqrt{|k|}.$$

On the other hand, since $K_2 = K_1$, the second bound (4.11), applied to $(\mathcal{F}_1, \mathcal{F}_1)$ and $\alpha = 1$, gives

$$|S - |k|| \leqslant 4c^4\sqrt{|k|},$$

and therefore

$$|k| \leqslant 8c^4\sqrt{|k|}$$

by the triangle inequality, which gives the result since $8^{1/4} \leqslant 2$.

(2) If $K_1(x) = 0$ for all $x \in Z$, we get

$$|k| - 4c_1^4\sqrt{|k|} \leqslant \sum_{x \in k-Z} |K_1(x)|^2 \leqslant (|k| - |Z|)c_1^2,$$

hence

$$|Z| \leqslant |k|\Big(1 - \frac{1}{c^2}\Big) + 4c^2\sqrt{|k|} \leqslant |k|\Big(1 - \frac{1}{2c^2}\Big)$$

if $c_1 \leqslant \frac{1}{2}|k|^{1/8}$.

(3) We write

$$\begin{aligned}\sum_{x \in k} |K_1(x)|^2 &= \sum_{x \in Z'} K_1(x)\overline{K_2(x)} + \sum_{x \in k-Z'} |K_1(x)|^2 \\ &= \sum_{x \in k} K_1(x)\overline{K_2(x)} + \sum_{x \in k-Z'} (|K_1(x)|^2 - K_1(x)\overline{K_2(x)}).\end{aligned}$$

By the two parts of Theorem 4.1.1, we get again

$$|k| - 4c^4\sqrt{|k|} \leqslant \sum_{x \in k} |K_1(x)|^2,$$

and

$$\Big|\sum_{x \in k} K_1(x)\overline{K_2(x)}\Big| \leqslant 4c^4\sqrt{|k|},$$

so that

$$|k| - 4c^4\sqrt{|k|} \leqslant 4c^4\sqrt{|k|} + 2(|k| - |Z'|)c^2,$$

which gives

$$|Z'| \leqslant |k|\Big(1 - \frac{1}{2c^2}\Big) + 8c^2\sqrt{|k|} \leqslant |k|\Big(1 - \frac{1}{4c^2}\Big)$$

if $c \leqslant \frac{1}{2}|k|^{1/8}$.

(4) If K_1 is identically zero, then we get

$$0 = \sum_{x \in k} |K_1(x)|^2 \geqslant |k| - 4c_1^4\sqrt{|k|},$$

hence the result. □

This leads to the following questions:

QUESTION 4.4.2. Let k be a finite field. What is the *injectivity threshold* $IT(k)$ defined as the smallest c for which there exist two geometrically irreducible middle-extension sheaves over k, not geometrically isomorphic, with the same trace function?

QUESTION 4.4.3. Let k be a finite field. What is the *zero threshold* $ZT(k)$ defined as the smallest c for which there exist a geometrically irreducible middle-extension sheaf over k with trace function identically zero?

Proposition 4.4.1 shows that $IT(k) \geqslant \frac{1}{2}|k|^{1/8}$ and $ZT(k) \geqslant \frac{1}{2}|k|^{1/8}$. This can be improved by simply being more careful in the reduction of the Riemann Hypothesis to simple estimates (see [**34**]). However, in the absence of tools to study analytic properties of trace functions of large conductor, it is very unclear how to proceed when the Riemann Hypothesis ceases to give non-trivial estimates.

On the other hand, direct constructions lead only to upper bounds of size about $|k|$. For instance, if χ is a non-trivial multiplicative character of $k^{\times}$, then the Kummer sheaf $\mathcal{F}_2 = \mathcal{L}_{\chi(F)}$ with

$$F = X^{|k|} - X + 1$$

has trace function identically 1, i.e., equal to that of the trivial sheaf $\mathcal{F}_1$. In that case, $\mathrm{c}(\mathcal{F}_2) \leqslant |k| + 2$. Similarly, the trace function of $\mathcal{L}_{\chi(G)}$, for

$$G = X^{|k|} - X,$$

is identically zero.

The next corollary of the Riemann Hypothesis is also very interesting, as it reveals an important feature of the conductor: its height-like properties.

PROPOSITION 4.4.4. *Let k be a finite field and ℓ a prime different from the characteristic p of k. Let $c \geqslant 1$ be a real number. There are only finitely many geometric isomorphism classes of geometrically irreducible middle-extension ℓ-adic sheaves on k with conductor at most c. In particular, there are only finitely many geometrically irreducible trace functions on K with conductor at most c, up to multiplication by complex numbers of modulus* 1.

The key to this finiteness property is the following fact about unit vectors in a finite-dimensional real inner-product space:

LEMMA 4.4.5. *Let E be a finite-dimensional real inner-product space. Let $\xi > 0$ be given. If Y is any subset of unit vectors in E such that*

$$\langle x, y\rangle \leqslant 1 - \xi$$

for all $x \neq y$ in Y, then Y is finite.

PROOF. For any unit vector v in E, let

$$C_v = \{w \in E \mid \|w\| = 1, \quad |\langle v, w\rangle| > 1 - \frac{\xi}{4}\}.$$

Then, by homogeneity, the measure μ of C_v (with respect to the surface measure on the unit sphere of E, normalized for instance so that the whole sphere has measure 1) is independent of v, and $\mu > 0$ because it contains an open neighborhood of v on the unit sphere.

We claim that $C_x \cap C_y = \varnothing$ if x and y are on Y and distinct. Indeed, if $w \in C_x \cap C_y$, then since

$$\langle v_1, v_2 \rangle = 1 - \frac{1}{2}\|v_1 - v_2\|^2$$

for any vectors v_1, v_2 of length 1, we get

$$\|w - x\| < \sqrt{\frac{\xi}{2}}, \qquad \|w - y\| < \sqrt{\frac{\xi}{2}},$$

and therefore

$$\|x - y\| < \sqrt{2\xi},$$

and

$$\langle x, y \rangle = 1 - \frac{1}{2}\|x - y\|^2 > 1 - \xi,$$

which contradicts the assumption on Y. Thus the sets C_x are disjoint for $x \in Y$, and hence $\mu|Y| \leqslant 1$, so that Y is finite. □

This lemma means that the angle between two distinct unit vectors in X is always bounded away from zero (by $\arccos(1 - \xi) \in]0, \pi/2]$, if $\xi \leqslant 1$) In the combinatorics literature, sets such as Y in this statement are called *spherical codes*, and the property above is the beginning of a very delicate theory dealing with the problem of finding good estimates (especially upper bounds) for the maximal size of such a set Y, given k and ξ. We refer to [**63**,] for details and references.

Proof of Proposition 4.4.4. We denote by $M_k(c)$ a set of representatives, up to geometric isomorphism, of geometrically irreducible middle-extension ℓ-adic sheaves with conductor $\leqslant c$. We also write $X_k(c)$ for the set of the trace functions of these representatives, so that

$$X_k(c) \subset C(k),$$

where we recall that $C(k)$ denotes the space of complex-valued functions on k.

We begin by observing that by restriction of the underlying representation to the subgroup

$$\Pi_{k_\nu} \lhd \Pi_k$$

of Π_k, we can find an injection of $M_k(c)$ in $M_{k_\nu}(c)$, where k_ν is the extension of k of degree $\nu \geqslant 1$. Indeed, since Π_{k_ν} has the same geometric Galois subgroup Π_k^g as Π_k, a geometrically irreducible representation remains geometrically irreducible after restriction to Π_{k_ν}, and this restriction has the same conductor. Moreover if the restrictions of $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic, then tautologically so are $\mathcal{F}_1$ and $\mathcal{F}_2$, which shows the injectivity.

In particular, it is enough to prove that $M_{k_\nu}(c)$ is finite for some $\nu \geqslant 1$ to finish the proof. For this purpose, we pick ν such that

$$c < \frac{1}{2}|k|^{\nu/8}. \tag{4.9}$$

Then by Proposition 4.4.1, the map

$$M_{k_\nu}(c) \longrightarrow X_{k_\nu}(c)$$

is also injective, and its image is a set of non-zero vectors in the space $C(k_\nu)$. We have an $\mathbf{R}$-linear isomorphism of $C(k_\nu)$ with $E = \mathbf{R}^{2|k|^\nu}$ by mapping a function to the vector with coordinates given by the real and imaginary parts of its values $\varphi(x)$. We equip the real vector space E with the inner-product

$$\langle v_1, v_2\rangle = \frac{1}{|k|^{\nu}}\sum_i x_i y_i,$$

so that

$$\langle v_1, v_2\rangle = \frac{1}{|k|^{\nu}}\operatorname{Re}\Bigl(\sum_{x\in k_\nu} K_1(x)\overline{K_2(x)}\Bigr)$$

for v_i the vectors corresponding to $K_i \in X_{k_\nu}(c)$. Let Y be the set of unit vectors $v/\|v\|$ associated to $K \in X_{k_\nu}(c)$; this is well-defined since $K \neq 0$ by the above condition on ν.

For v_1, v_2 distinct in Y, corresponding to trace functions K_1, K_2, we have then

$$\langle v_1, v_2\rangle = \frac{1}{\|K_1\|\|K_2\|}\operatorname{Re}\Bigl(\frac{1}{|k|^{\nu}}\sum_{x\in k_\nu} K_1(x)\overline{K_2(x)}\Bigr).$$

From (4.9) and Theorem 4.1.1, we deduce

$$\|K_i\| \geqslant \frac{3}{4}, \qquad \frac{1}{|k|^{\nu}}\Bigl|\sum_{x\in k_\nu} K_1(x)\overline{K_2(x)}\Bigr| \leqslant \frac{1}{4},$$

and hence distinct vectors v_1, v_2 in Y satisfy

$$\langle v_1, v_2\rangle \leqslant \frac{4}{9}.$$

By Lemma 4.4.5, the set Y is therefore finite. □

4.5. Extensions

We have stated Deligne's Theorem for geometrically irreducible sheaves only. Although this is justified in many applications, the decomposition results of trace functions in terms of isotypic sheaves from Proposition 3.5.3 allows one to state a more general version which is also very useful.

THEOREM 4.5.1. *Let k be a finite field and ℓ a prime different from the characteristic p of k. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ denote two middle-extension ℓ-adic sheaves on k. Denote $c_i = \mathrm{c}(\mathcal{F}_i)$ and denote K_i the trace function of $\mathcal{F}_i$. Assume that $\mathcal{F}_1$ and $\mathcal{F}_2$ are arithmetically irreducible and geometrically isotypic.*

(1) *If $\mathcal{F}_1$ and $\mathcal{F}_2$ are* not *geometrically isomorphic then*

$$\Bigl|\sum_{x\in k} K_1(x)\overline{K_2(x)}\Bigr| \leqslant 4c_1^2c_2^2\sqrt{|k|}. \tag{4.10}$$

(2) *If $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic then there exists a Weil-number α of weight 0 such that $K_2 = \alpha K_1$, and an integer $r \geqslant 1$ such that we have*

$$\Bigl|\sum_{x\in k} K_1(x)\overline{K_2(x)} - r\bar{\alpha}|k|\Bigr| \leqslant 4c_1^2c_2^2\sqrt{|k|}. \tag{4.11}$$

As a corollary, we will show a very convenient analytic criterion for geometric irreducibility, due to Katz.

THEOREM 4.5.2. *Let k be a finite field and ℓ a prime different from the characteristic p of k. Let $\mathcal{F}$ be a middle-extension ℓ-adic sheaf on k. Then $\mathcal{F}$ is geometrically irreducible if and only if*

$$\lim_{\nu\to+\infty}\frac{1}{|k|^{\nu}}\sum_{x\in k_\nu}|t_\varrho(x;\nu)|^2 = 1$$

or if and only if

$$\frac{1}{|k|^{\nu}}\sum_{x\in k_{\nu}}|t_{\varrho}(x;\nu)|^2=1+O(|k|^{-\nu/2})$$

where the implied constant depends only on $\mathrm{c}(\mathcal{F})$*.*

Another result that can be very useful are the extensions of the trace formula and the Euler-Poincaré formula beyond the open set U where the representations we apply these results to are unramified. We state these results over the affine line for simplicity.

Theorem 4.5.3. *Let k be a finite field. Let $\mathcal{F}$ be an ℓ-adic middle-extension sheaf over k.*

(1) *We have*

$$\sum_{x\in k}t_{\mathcal{F}}(x)=M_0-M_1+M_2,$$

where

$$M_i=\mathrm{Tr}(\mathrm{Fr}_k\mid H^i_c(\mathbf{A}^1\times\bar{k},\mathcal{F})).$$

(2) *We have*

$$\chi(\mathbf{A}^1,\mathcal{F})=\mathrm{rank}(\mathcal{F})-\sum_{x\in\mathrm{Sing}(\mathcal{F})-\{\infty\}}\mathrm{drop}_x(\mathcal{F})-\sum_{x\in\mathrm{Sing}(\mathcal{F})}\mathrm{Swan}_x(\mathcal{F})\tag{4.12}$$

where

$$\chi(\mathbf{A}^1,\mathcal{F})=\dim H^0_c(\mathbf{A}^1\times\bar{k},\mathcal{F})-\dim H^1_c(\mathbf{A}^1\times\bar{k},\mathcal{F})+\dim H^2_c(\mathbf{A}^1\times\bar{k},\mathcal{F}),$$

and for any $x\in\mathbf{P}^1(\bar{k})$, we denote

$$\mathrm{drop}_x(\mathcal{F})=\mathrm{rank}(\mathcal{F})-\dim\mathcal{F}^{I_x}.$$

CHAPTER 5

The Fourier transform and other cohomological transforms

This chapter describes one of the most important tools in applications of trace functions to analytic number theory. In a nutshell, it explains that the *discrete Fourier transform* of functions defined on a finite field k can be interpreted at the level of sheaves, as was discovered by Deligne: the Fourier transform of a trace function is again a trace function. As we will see through numerous illustrations, this has very far-reaching consequences, arising from the preminent position of Fourier transforms in harmonic analysis.

5.1. The discrete Fourier transform

We begin with a short definition and description of the discrete Fourier transform for functions on a finite field. This is also the occasion to fix the notation and normalization.

Definition 5.1.1. Let k be a finite field. Fix a non-trivial additive character

$$\psi : k \longrightarrow \mathbf{C}^{\times}.$$

The *Fourier transform with respect to* ψ is the linear operator

$$\mathrm{FT}_{\psi} : C(k) \longrightarrow C(k)$$

defined by

$$\mathrm{FT}_{\psi}(\varphi)(y) = -\frac{1}{\sqrt{|k|}}\sum_{x\in k}\varphi(x)\psi(xy)$$

for any $\varphi \in C(k)$ and any $y \in k$.

Remark 5.1.2. For any $\nu \geqslant 1$, the additive character ψ defines naturally a character

$$\psi_{\nu} : k_{\nu} \longrightarrow \mathbf{C}^{\times}$$

which is non-trivial if ψ is. We will simply denote by $\mathrm{FT}_{\psi} : C(k_{\nu}) \longrightarrow C(k_{\nu})$ the Fourier transform on k_{ν} with respect to this character ψ_{ν}.

The choice of normalization is justified by the unitarity that is one of the basic properties of this operator, and which will correspond to the fact that we wish weight 0 representations to be sent to weight 0 representations in the sheaf-theoretic version of the Fourier transform.

We recall that the standard inner product defined on $C(k)$ is given by

$$\langle\varphi_1,\varphi_2\rangle = \frac{1}{|k|}\sum_{x\in k}\varphi_1(x)\overline{\varphi_2(x)}$$

for any φ_1 and $\varphi_2 \in C(k)$.

Proposition 5.1.3. *Let k be a finite field and let ψ a non-trivial additive character of k.*

(1) *The Fourier transform is invertible and in fact it satisfies*

$$\mathrm{FT}_{\psi}(\mathrm{FT}_{\psi}(\varphi)) = [\times(-1)]^{*}\varphi$$

for any $\varphi \in C(k)$, where

$$([\times(-1)]^*\varphi)(x) = \varphi(-x).$$

(2) *The Fourier transform is unitary on $C(k)$, i.e., for any φ_1 and φ_2 in $C(k)$, we have*

$$\langle \varphi_1, \varphi_2 \rangle = \langle \mathrm{FT}_\psi(\varphi_1), \mathrm{FT}_\psi(\varphi_2) \rangle.$$

In particular, we have

$$\sum_{x \in k} |\varphi(x)|^2 = \sum_{y \in k} |\, \mathrm{FT}_\psi(\varphi)(y)|^2$$

for any $\varphi \in C(k)$.

Proof. These are standard facts that follow from the orthogonality of characters. We prove the unitarity to illustrate the normalization we use: by definition, we have

$$\begin{aligned}\langle \mathrm{FT}_\psi(\varphi_1), \mathrm{FT}_\psi(\varphi_2) \rangle &= \frac{1}{|k|^2} \sum_y \Bigl(\sum_{x_1 \in k} \varphi_1(x)\psi(x_1 y) \Bigr) \Bigl(\sum_{x_2 \in k} \overline{\varphi_2(x)} \psi(-x_2 y) \Bigr) \\ &= \frac{1}{|k|} \sum_{x_1 \in k} \sum_{x_2 \in k} \sum_{y \in k} \psi((x_1 - x_2)y) \\ &= \frac{1}{|k|} \sum_{x \in k} \varphi_1(x)\overline{\varphi_2(x)}.\end{aligned}$$

□

Exercise 5.1.4. Fix a finite field k and a non-trivial additive character ψ.

(1) Let $a \in k$ be given. Show that the Fourier transform with respect to ψ of $\varphi(x) = \psi(ax)$ is given by

$$\mathrm{FT}_\psi(\varphi)(y) = \begin{cases} \sqrt{|k|} & \text{if } a + y = 0 \\ 0 & \text{otherwise.} \end{cases}$$

(2) Let $\chi : k^\times \longrightarrow \mathbf{C}^\times$ be a non-trivial multiplicative character of k. Extend χ to k by putting $\chi(0) = 0$. Show that

$$\mathrm{FT}_\psi(\chi)(y) = w_\psi(\chi)\bar{\chi}(y)$$

for all $y \in k$, where

$$w_\psi(\chi) = -\frac{1}{\sqrt{|k|}} \sum_{x \in k} \chi(x)\psi(x)$$

is the Gauss sums of χ with respect to ψ. Show also that $w_\psi(\chi)$ is a Weil number of weight 0.

The first part of this exercise, which is of course classical, illustrates a very special case of the Fourier transform: additive characters, although they are uniformly bounded independently of k, are mapped to functions with L^∞-norm that grows with the size of the finite field.

5.2. The sheaf-theoretic Fourier transform

The main result of this section is a very deep theorem of Deligne. To state it, we first introduce a subclass of ℓ-adic representations, which are those for which the Fourier transform is defined.

Definition 5.2.1 (Fourier sheaf). Let k be a finite field, $\ell \nmid p$ a prime number. A middle-extension sheaf $\mathcal{F}$ over k is *of Fourier type* or *a Fourier sheaf* if none of its geometric Jordan-Hölder component is geometrically isomorphic to an Artin-Schreier sheaf $\mathcal{L}_\psi$ attached to some (possibly trivial) additive character ψ.

This definition may look tricky, but it should be thought of as being generic. For instance, a sheaf $\mathcal{F}$ is of Fourier type if any of the following conditions hold:

- If $\mathcal{F}$ is geometrically irreducible, and not geometrically isomorphic to some Artin-Schreier sheaf $\mathcal{L}_\psi$; indeed, in that case $\mathcal{F}$ is its own unique geometric Jordan-Hölder component (in particular, if $\mathcal{F}$ is geometrically irreducible and of rank 2, or if $\mathcal{F}$ is geometrically irreducible and is ramified at least at some point $x \in \bar{k}$).
- If $\mathcal{F}$ is everywhere tamely ramified, and does not have a trivial Jordan-Hölder component.

Remark 5.2.2. In Appendix A, we will indicate for each of the trace functions that we describe, whether they are of Fourier type or not.

Theorem 5.2.3 (Deligne). *Let k be a finite field and let $\psi : k \longrightarrow \mathbf{C}^\times$ be a non-trivial additive character of k.*

For any Fourier sheaf $\mathcal{F}$, there exists a Fourier sheaf $\mathcal{G} = \mathrm{FT}_\psi(\mathcal{F})$, called the Fourier transform of $\mathcal{F}$ with respect to ψ, *such that*

(1) *For any $\nu \geqslant 1$ we have*

$$t_{\mathcal{G}}(\cdot;\nu) = \mathrm{FT}_\psi(t_{\mathcal{F}}(\cdot;\nu)), \tag{5.1}$$

i.e, we have

$$t_{\mathcal{G}}(y;\nu) = -\frac{1}{|k|^{\nu/2}} \sum_{x\in k_\nu} t_{\mathcal{F}}(x;\nu)\psi(\mathrm{Tr}_{k_\nu/k}(xy))$$

for any $y \in k_\nu$.

(2) *We have*

$$\mathrm{c}(\mathcal{G}) \leqslant 10\,\mathrm{c}(\mathcal{F})^2.$$

Remark 5.2.4. In fact, Deligne's theorem gives a specific geometric construction that shows that the map sending $\mathcal{F}$ to its Fourier transform is natural (it is a functor). We did not describe it in this mannre and hence there is some ambiguity in our statement. However, according to Proposition 4.4.1 (1), the upper bound on the conductor ensures that a sheaf $\mathcal{G}$ satisfying (5.1) is unique, if it exists and is geometrically irreducible, and if $\mathrm{c}(\mathcal{F})$ is small enough compared with $|k|$.

We will now motivate the result, but first we present some basic applications that will show that Theorem 5.2.3 is extremely deep, and in fact contains already quite general instances of the Riemann Hypothesis.

Example 5.2.5. Let p be a prime number and let $f \in \mathbf{F}_p(X)$ be a rational function that is not of the form $f = g^p - g + aX + b$ for some $g \in \mathbf{F}_p(X)$ and some a, $b \in \mathbf{F}_p$. This is the case, for instance, if f is a polynomial of degree d with $2 \leqslant d < p$. Let ψ be the additive character $\psi(x) = e(x/p)$. The Artin-Schreier sheaf $\mathcal{L}_{\psi(f)}$ (see Section 2.4) is then of Fourier type, by the classification of Artin-Schreier sheaves (Theorem 3.2.4 (2)).

5.3. Applications of the Fourier transform

We present here some of the basic examples of use of the Fourier transform. This will also motivate the statements of some of its additional deeper properties, some of which have little (if any) obvious counterpart at the level of the discrete Fourier transform of functions.

Example 5.3.1 (Deligne's bound for hyper-Kloosterman sums).

5.4. Precise constructions

We explain the actual construction of the Fourier transform and some of its properties.

APPENDIX A

Appendix A: trace functions

This Appendix is a reference list of the most important examples of trace functions, together with a description of their invariants. In each case, we begin by stating the results for the functions themselves (in a manner suitable for Theorem 1.2.1), and then we discuss the geometric and algebraic invariants of the underlying representations or sheaves.

Example 1 (Kummer functions).

Example 2 (Artin-Schreier functions).

Example 3 (Kloosterman and hyper-Kloosterman sums).

Example 4 (Hypergeometric functions).

Bibliography

[1] A. Beauville: *Les familles stables de courbes elliptiques sur* $\mathbf{P}^1$ *admettant quatre fibres singulières*, C. R. Acad. Sc. Paris 294 (1982), 657–660.

[2] A. Beauville: *Finite subgroups of* $\mathrm{PGL}_2(K)$, Contemporary Math. 522, 23–29, A.M.S (2010).

[3] V. Blomer, G. Harcos and Ph. Michel: *Bounds for modular L-functions in the level aspect*, Ann. Sci. Ecole Norm. Sup. (4) 40 (2007), no. 5, 697–740.

[4] V. Blomer, G. Harcos and Ph. Michel: *A Burgess-like subconvex bound for twisted L-functions*, with an Appendix by P. Zhao, Forum Math. 19 (2007), 61–105.

[5] V. Blomer and G. Harcos: *Hybrid bounds for twisted L-functions*, J. reine und angew. Mathematik 621 (2008), 53–79.

[6] E. Bombieri, J. Friedlander and H. Iwaniec: *Primes in arithmetic progressions to large moduli,* Acta Math. 156 (1985), 203–251.

[7] E. Bombieri, J. Friedlander and H. Iwaniec: *Primes in arithmetic progressions to large moduli. II,* Math. Ann. 277 (1987), no. 3, 361–393.

[8] A. Borel: *Linear algebraic groups*, Grad. Texts in Math. 126 (2nd ed.) Springer Verlag 1991.

[9] W. Bosma, J. Cannon and C. Playoust: *The Magma algebra system, I. The user language*, J. Symbolic Comput., 24 (1997), 235–265; also `http://magma.maths.usyd.edu.au/magma/`

[10] V.A. Bykovski: *A trace formula for the scalar product of Hecke series and its applications*, J. Math. Sciences 89 (1998), 915–932.

[11] Z. Chatzidakis, L. van den Dries and A. Macintyre: *Definable sets over finite fields*, J. reine angew. Math. 427 (1992), 107–135

[12] J.B. Conrey and H. Iwaniec: *The cubic moment of central values of automorphic L-functions*, Ann. of Math. (2) 151 (2000), no. 3, 1175–1216.

[13] J.H. Conway and N.J.A. Sloane: *Sphere packings, lattices and groups*, Grund. der Math. Wiss. 290, Springer–Verlag, 1988.

[14] H. Davenport: *On character sums in finite fields*, Acta Math. 71 (1939), 99–121.

[15] P. Deligne: *Cohomologie étale*, S.G.A $4\frac{1}{2}$, L.N.M 569, Springer Verlag (1977).

[16] P. Deligne: *La conjecture de Weil, II*, Publ. Math. IHÉS 52 (1980), 137–252.

[17] P. Deligne: letter to V. Drinfeld, dated June 18, 2011, 9 pages.

[18] P. Deligne: *Counting ℓ-adic representations, in the function field case*, lecture at the Newton Institute, July 2009, `http://www.newton.ac.uk/programmes/NAG/seminars/072710001.html`

[19] P. Deligne and Y. Flicker: *Counting local systems with principal unipotent local monodromy*, Annals of Math. (to appear), `http://www.math.osu.edu/~flicker.1/df.pdf`

[20] P. Delsarte, J.M. Goethals and J.J. Seidel: *Spherical codes and designs*, Geometriae Dedicata 6 (1977), 363–388.

[21] V. Drinfeld: *The number of two-dimensional irreducible representations of the fundamental group of a curve over a finite field*, Functional Anal. Appl. 15 (1981), 294–295 (1982).

[22] J-M. Deshouillers and H. Iwaniec: *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. math. 70 (1982/83), no. 2, 219–288.

[23] W.D. Duke, J. Friedlander and H. Iwaniec: *Bounds for automorphic L-functions*, Invent. math. 112 (1993), 1–8.

[24] W.D. Duke, J. Friedlander and H. Iwaniec: *The subconvexity problem for Artin L-functions*, Invent. math. 149 (2002), no. 3, 489–577.

[25] A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi: *Higher transcendental functions*, Vol. II, McGraw Hill (1955).

[26] H. Esnault and M. Kerz: *A finiteness theorem for Galois representations of function fields over finite fields (after Deligne)*, Acta Mathematica Vietnamica 37 (2012), 351–362; `arXiv: 1208.0128v3`.

[27] É. Fouvry: *Autour du théorème de Bombieri-Vinogradov*, Acta Math. 152 (1984), no. 3-4, 219–244.

[28] É. Fouvry: *Sur le problème des diviseurs de Titchmarsh*, J. reine angew. Math. 357 (1985), 51–76.

[29] É. Fouvry and H. Iwaniec: *Primes in arithmetic progressions*, Acta Arith. 42 (1983), no. 2, 197–218.

[30] É. Fouvry, Ph. Michel and E. Kowalski: *Algebraic twists of modular forms and Hecke orbits*, preprint available at `arXiv:1207.0617`

[31] É. Fouvry, E. Kowalski, Ph. Michel: *Algebraic trace weights over the primes*, preprint `arXiv: 1211.6043v1`.

[32] É. Fouvry, S. Ganguly, E. Kowalski, Ph. Michel: *Algebraic geometry and Fourier coefficients of cusp forms in arithmetic progressions*, Preprint (2012).

[33] É. Fouvry, E. Kowalski, Ph. Michel: *An inverse theorem for Gowers norms of trace functions over* $\mathbf{F}_p$, Math. Proc. Cambridge Phil. Soc. (to appear).

[34] É. Fouvry, E. Kowalski, Ph. Michel: *Counting sheaves using spherical codes*, Math. Res. Letters (to appear).

[35] É Fouvry and S. Ganguly: *Orthogonality between the Möbius function, additive characters, and Fourier coefficients of cusp forms*, preprint (2012).

[36] É. Fouvry, P. Michel, J. Rivat and A. Sárközy, *On the pseudorandomness of the signs of Kloosterman sums*, Journal of the Australian Mathematical Society, Volume 77, December 2004, 425–436.

[37] J. B. Friedlander and H. Iwaniec:*Incomplete Kloosterman sums and a divisor problem*, (With an appendix by Bryan J. Birch and Enrico Bombieri), Ann. of Math. (2) 121, 319–350 (1985).

[38] K. Fujiwara: *Independence of ℓ for intersection cohomology (after Gabber)*, in Algebraic Geometry 2000, Azumino, Adv. Stud. Pure Math. 36 (2002), 145–151.

[39] I.S. Gradshteyn and I.M. Ryzhkik: *Tables of integrals, series and products*, 5th ed. (edited by A. Jeffrey), Academic Press (1994).

[40] B.J. Green and T. Tao: *The distribution of polynomials over finite fields, with applications to the Gowers norms*, Contrib. Discr. Math. 4 (2009), 1–36; `arXiv:0711.3191`.

[41] B.J. Green, T. Tao and T. Ziegler: *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Annals of Math. 176 (2012), 1231–1372; `arXiv:1009.3998`.

[42] D. R. Heath-Brown: *The divisor function $d_3(n)$ in arithmetic progressions*, Acta Arith. 47 (1986), no. 1, 29–56.

[43] D. R. Heath-Brown, *The density of rational points on cubic surfaces*, Acta Arith. 79 (1997), 17–30.

[44] H. Helfgott and A. Venkatesh: *Integral points on elliptic curves and 3-torsion in class groups*, Journal of the A.M.S 19 (2006), 527–550; `arXiv:math.NT/0405180v2`

[45] H. Iwaniec: *Fourier coefficients of modular forms of half-integral weight*, Invent. math. 87 (1987), no. 2, 385–401.

[46] H. Iwaniec: *Small eigenvalues of Laplacian for* $\Gamma_0(N)$, Acta Arith. 56 (1990), no. 1, 65–82.

[47] H. Iwaniec: *Topics in classical automorphic forms*, Grad. Studies in Math. 17, A.M.S (1997).

[48] H. Iwaniec: *Introduction to the spectral theory of automorphic forms*, Biblioteca de la Revista Matematica Iberoamericana, Revista Matematica Iberoamericana, Madrid, 1995.

[49] H. Iwaniec and E. Kowalski: *Analytic number theory*, A.M.S. Coll. Publ. 53 (2004).

[50] G.A. Kabatjanskii and V.I. Levenshtein: *Bounds for packings on the sphere and in space*, Problemy Peredači Informacii 14 (1978), 3–25.

[51] N.M. Katz: *Gauss sums, Kloosterman sums and monodromy groups*, Annals of Math. Studies 116, Princeton Univ. Press (1988).

[52] N.M. Katz: *Exponential sums and differential equations*, Annals of Math. Studies 124, Princeton Univ. Press (1990).

[53] N.M. Katz: *Rigid local systems*, Annals of Math. Studies 139, Princeton Univ. Press (1993).

[54] N.M. Katz: *Moments, monodromy and perversity*, Annals of Math. Studies 159, Princeton Univ. Press (2005).

[55] N.M. Katz: *Convolution and equidistribution: Sato-Tate theorems for finite-field Mellin transforms*, Annals of Math. Studies 180, Princeton Univ. Press (2011).

[56] N.M. Katz: *Affine cohomological transforms, perversity, and monodromy*, Journal of the A.M.S. 6 (1993), 149–222.

[57] N.M. Katz: *On a question of Lillian Pierce*, Forum Math. ??, ???.

[58] H. Kim and P. Sarnak: *Refined estimates towards the Ramanujan and Selberg conjectures*, J. American Math. Soc. 16 (2003), 175–181.

[59] E. Kowalski: *An introduction to the representation theory of groups*, Grad. Studies Math. 155, A.M.S, 2014.

[60] E. Kowalski, O. Robert and J. Wu: *Small gaps in coefficients of L-functions and $\mathfrak{B}$-free numbers in small intervals*, Rev. Mat. Iberoamericana 23 (2007), 281–326.

[61] L. Lafforgue: *Chtoucas de Drinfeld et correspondance de Langlands*, Invent. math. 147 (2002), 1–241.

[62] G. Laumon: *Transformation de Fourier, constantes d'équations fonctionnelles et conjecture de Weil*, Publ. Math. IHÉS, 65 (1987), 131–210.

[63] V.I. Levenshtein: *Universal bounds for codes and designs*, in "Handbook of coding theory", 499–648, North-Holland, Amsterdam, 1998.

[64] H. Liu: *Gowers uniformity norm and pseudorandom measures of the pseudorandom binary sequences*, International J. Number Th. 7 (2005), 1279–1302.

[65] Ph. Michel and A. Venkatesh: *The subconvexity problem for* GL_2, Publ. Math. I.H.É.S 111, 171–271 (2010).

[66] Y. Motohashi: *On sums of Hecke-Maass eigenvalues squared over primes in short intervals*, preprint `arXiv:1209.4140v1`.

[67] R. Munshi: *Shifted convolution sums for* $GL(3) \times GL(2)$, preprint (2012), `arXiv:1202.1157`.

[68] H. Niederreiter and J. Rivat: *On the Gowers norms of pseudorandom binary sequences*, Bull. Aust. Math. Soc. 79 (2009), 259–271.

[69] L. Pierce: ..., J. London Math. Soc. ???.

[70] N. Pitt: *On shifted convolutions of* $\zeta(s)^3$ *with automorphic L-functions*, Duke Math. J. 77 (1995), no. 2, 383–406.

[71] D. Ramakrishnan and J. Rogawski, *Average values of modular L-series via the relative trace formula*, Pure Appl. Math. Q. 1 (2005), no. 4, Special Issue: In memory of Armand Borel. Part 3, 701–735.

[72] P. Sarnak: *Diophantine problems and linear groups*, in Proceedings of the I.C.M., 1990, Kyoto, Springer (1991), 459–471.

[73] J-P. Serre: *Représentations linéaires des groupes finis*, 2ème Édition, Hermann, 1971.

[74] A. Strömbergsson: *On the uniform equidistribution of long closed horocycles*, Duke Math. J. 123 (2004), no. 3, 507–547.

[75] P. Sarnak and A. Ubis: *The horocycle flow at prime times*, preprint `http://arxiv.org/abs/1110.0777v1`

[76] G. Szegö: *Orthogonal polynomials*, A.M.S. Coll. Publ. 23 (1939).

[77] T. Tao and V. Vu: *Additive combinatorics*, Cambridge studies adv. math. 105, Cambridge Univ. Press 2006.

[78] T. Tao and T. Ziegler: *The inverse conjecture for the Gowers norm over finite fields via the correspondence principle*, Analysis and PDE 3 (2010), 1–20; `arXiv:0810.5527`.

[79] T. Tao and T. Ziegler: *The inverse conjecture for the Gowers norm over finite fields in low characteristics*, Annals of Combinatorics 16 (2012), 121–188; `arXiv:1101.1469`.

[80] A. Venkatesh: *Sparse equidistribution problems, period bounds and subconvexity*, Ann. of Math. (2) 172 (2010), no. 2, 989–1094

[81] G.N. Watson: *A treatise on the theory of Bessel functions*, 2nd ed., Cambridge Math. Library, Cambridge Univ. Press (1996).

[82] Y. Zhang: *Bounded gaps between primes*, Annals of Math. ?? (2014), ???–???.