TRACE FUNCTIONS OVER FINITE FIELDS AND THEIR APPLICATIONS

ÉTIENNE FOUVRY, EMMANUEL KOWALSKI, AND PHILIPPE MICHEL

Abstract. We survey our recent works concerning applications to analytic number theory of trace functions of ℓ-adic sheaves over finite fields.

1. Motivation

We begin by describing one of the motivating problems for our paper [7]. This concerns an equidistribution statement in the upper half-plane $\mathbf{H}$ of complex numbers with positive imaginary parts, or more precisely in the domain

$$F = \{z \in \mathbf{H} \mid |\operatorname{Re}(z)| \leqslant 1/2,\ |z| \geqslant 1\} \subset \mathbf{H}.$$

This closed subset of $\mathbf{H}$ is well-known to be a *fundamental domain* for the action of the modular group $\mathrm{SL}_2(\mathbf{Z})$ by homographies on $\mathbf{H}$, i.e., the restriction to $\mathrm{SL}_2(\mathbf{R})$ of the $\mathrm{SL}_2(\mathbf{R})$-action given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d} \tag{1.1}$$

for any $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$ (see Figure 1.)

This means that, for any $z \in \mathbf{H}$, there exists some element $\gamma \in \mathrm{SL}_2(\mathbf{Z})$ such that $\gamma \cdot z \in F$, and in fact γ is usually unique (it is unique if $\gamma \cdot z$ is in the interior of F). Consider in particular a prime number p, and the p points

$$z_0 = \frac{i}{p},\ z_1 = \frac{1+i}{p}, \dots, z_{p-1} = \frac{p-1+i}{p},$$

in $\mathbf{H}$. There are corresponding points $w_0, \dots, w_{p-1}$ in F, each equivalent to the respective point z_j under the action of $\mathrm{SL}_2(\mathbf{Z})$. Where are these points? Experiments quickly show that, as p increases, the points tend to range all over F. Indeed, one can prove that they become *equidistributed* as $p \to +\infty$, with respect to the probability measure

$$d\mu = \frac{3}{\pi}\frac{dxdy}{y^2}$$

on F (one checks indeed that $\mu(F) = 1$), which is a natural measure here because it is $\mathrm{SL}_2(\mathbf{R})$-invariant: $\mu(g^{-1}(A)) = \mu(A)$ for any $g \in \mathrm{SL}_2(\mathbf{R})$ and any measurable set $A \subset \mathbf{H}$.

Date: May 13, 2014, 12:42.

2010 *Mathematics Subject Classification.* 11B30,11F11,11F32,11F37,11G20,11T23,11L05,11N05, 11N13, 11N25,11N37,11N32, 11N35.

Key words and phrases. Trace functions, modular forms, exponential sums, arithmetic functions, Riemann Hypothesis over finite fields.

Ph. M. was partially supported by the SNF (grant 200021-137488) and the ERC (Advanced Research Grant 228304). É. F. thanks ETH Zürich, EPF Lausanne and the Institut Universitaire de France for financial support.

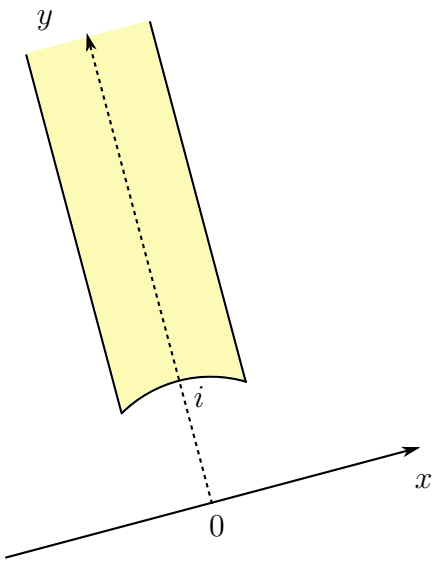

FIGURE 1. The leaning fundamental domain of Pisa

This equidistribution property means that

$$\lim_{p\to+\infty} \frac{1}{p}\sum_{j=0}^{p-1} \varphi\Big(\frac{j+w}{p}\Big) = \int_F \varphi(z)d\mu(z),$$

for any fixed $w \in \mathbf{H}$ (the case above being $w = i$) and for all continuous functions with compact support φ on $\mathbf{H}$ which are $\mathrm{SL}_2(\mathbf{Z})$-invariant: $\varphi(\gamma \cdot z) = \varphi(z)$ for all $z \in \mathbf{H}$.

In particular, in the spirit of Buffon's original needle problem, the following game is *fair*: for a large prime p, and a "random" integer $0 \leqslant j \leqslant p-1$, Players A and B take a bet as to whether the imaginary part of the corresponding w_j is $\geqslant 6/\pi$ or not. (Indeed, the measure of the subset

$$F_1 = \Big\{ z \in F \mid \mathrm{Im}(z) \geqslant \frac{6}{\pi} \Big\} \subset F$$

is easily computed to be $1/2$; thus, more precisely, the game is only fair in the limit when $p \to +\infty$.)

One picturesque question which our work can address in this context, is the following: can one player gain an edge in this game by selecting her bet according to some algebraic property of j modulo p, for instance by determining if j is of the form $f(n)$ modulo p for some fixed polynomial $f \in \mathbf{Z}[X]$ and $n \in \mathbf{Z}/p\mathbf{Z}$, and selecting her bet based on this value? As we will explain, this is not possible, at least asymptotically if p tends to infinity.

The key to solving this problem turns out to involve an intricate combination of methods of analytic number theory and concepts and results of algebraic geometry over finite fields. More precisely, the crucial notion is that of *trace functions* over a finite field k, which are certain complex-valued functions $k \longrightarrow \mathbf{C}$ which have very strong algebraic features. Most importantly, and as the critical ingredient in most of our applications, these functions satisfy a form of quasi-orthogonality, that follows from a very general form of the Riemann Hypothesis for algebraic varieties over finite fields, due to Deligne [4].

In the motivating problem, these functions can be seen to arise naturally. Suppose $f \in \mathbf{Z}[X]$ is a fixed non-constant polynomial, and we wish to show that there is no gain in the game that may be derived by betting that the point w_j is in F_1 if and only if j is a value of f, i.e., if and only if $j = f(n)$ for some $n \in \mathbf{Z}/p\mathbf{Z}$. If we denote by ξ_p the

function defined for $0 \leqslant j \leqslant p-1$ as the characteristic function of the set $f(\mathbf{Z}/p\mathbf{Z})$ of values of f, and by φ_1 the characteristic function of F_1, this fairness property amounts to showing that

$$\frac{1}{p}\Big(\sum_{j=0}^{p-1}\xi_p(j)\varphi_1(w_j) + \sum_{j=0}^{p-1}(1-\xi_p(j))(1-\varphi_1(w_j))\Big) \longrightarrow \frac{1}{2}$$

as $p \to +\infty$ (since the left-hand side is the proportion of success when betting that $w_j \in F_1$ using the strategy we described). What we actually show is that if

$$\delta_p = \frac{1}{p}\sum_{j=0}^{p-1}\xi_p(j)$$

is the average value of ξ_p, then

$$\frac{1}{p}\sum_{j=0}^{p-1}(\xi_p(j)-\delta_p)\varphi_1(w_j) \longrightarrow 0,$$

$$\frac{1}{p}\sum_{j=0}^{p-1}((1-\xi_p(j))-(1-\delta_p))(1-\varphi_1(w_j)) \longrightarrow 0,$$

from which the limit of the sum above is

$$\lim_{p\to+\infty}\frac{1}{p}\sum_{j=0}^{p-1}\varphi_1(w_j),$$

which converges to $1/2$ because the set F_1 was chosen so that the original game is fair.

These ideas, and the techniques we developed, have many other applications than the one we have described above, and we will survey, sometimes briefly, most of our main results.

More generally, we prove (considering only the first limit above, the second being just a variant) that

$$\frac{1}{p}\sum_{j=0}^{p-1}(\xi_p(j)-\delta_p)\varphi\Big(\frac{j+i}{p}\Big) \longrightarrow 0 \tag{1.2}$$

for all continuous and compactly supported functions φ on $\mathbf{H}$. A standard limiting procedure then extends the result to the characteristic function of F_1. Now, according to the well-known Weyl criterion, we may limit our attention to certain well-chosen functions. There is a standard choice of such functions in our case, which consists of the square-integrable eigenfunctions φ of the hyperbolic Laplace operator, together with certain other functions which we will not discuss further (which arise due to the non-compactness of F). The eigenfunctions φ are analogues of the eigenfunctions of the standard Laplace operator $-d^2/dx^2$ for periodic functions on $\mathbf{R}/\mathbf{Z}$, namely the exponentials $x \mapsto e(hx)$ for $h \in \mathbf{Z}$, where we denote $e(z) = e^{2i\pi z}$. They have an expansion

$$\varphi(z) = \sum_{m\in\mathbf{Z}-\{0\}}\lambda_\varphi(m)W_\varphi(2\pi|m|y)e(mx),$$

where W_φ is a certain Whittaker-Bessel function (depending only on the eigenvalue of φ for the hyperbolic Laplace operator) and the Fourier coefficients $\lambda_\varphi(m)$ are complex

numbers. Now observe that the left-hand side of (1.2) becomes

$$\sum_{m\neq 0}\lambda_\varphi(m)K_p(m)W_\varphi\Big(\frac{2\pi|m|}{p}\Big) \tag{1.3}$$

where

$$K_p(m)=\frac{1}{p}\sum_{j=0}^{p-1}(\xi_p(j)-\delta_p)e\Big(\frac{mj}{p}\Big). \tag{1.4}$$

This function K is naturally defined on $\mathbf{Z}/p\mathbf{Z}$, and its values on $\{0,\dots,p-1\}$ are obtained by reduction modulo p. We consider it to be of algebraic nature because it is a discrete Fourier transform modulo p of the function $\xi_p-\delta_p$, which has an algebraic definition in terms of the polynomial f. We have in fact proved in [7] that the limiting formula

$$\frac{1}{p}\sum_{m\neq 0}\lambda_\varphi(m)K_p(m)V(m)\longrightarrow 0 \tag{1.5}$$

holds as $p\longrightarrow+\infty$, for much more general cases of *trace functions* K_p and more general "nice" weight functions $V(m)$ than $V(m)=W_\varphi(2\pi|m|/p)$. This solves the fairness question we used as a motivation, and in fact it is a much more general and widely applicable result.

The outline of the remainder of this survey is the following. In the next section, we define rigorously trace functions; there appears then a crucial definition for analytic applications, that of the *conductor* of a trace function, which measures its complexity, in such a way that uniformity with respect to p may be considered. These definitions are illustrated in Section 3 with many examples. We next discuss the crucial, extremely deep and extremely powerful quasi-orthogonality property, which follows from the Riemann Hypothesis over finite fields, and how we use it in [7] and [8]. The last section discusses another, very concrete, application to the distribution of certain arithmetic functions in arithmetic progressions to large moduli, following [9].

We do not discuss some other papers, contenting ourselves with the following short indications: (1) in [10], we show that trace functions are "Gowers-uniform to all order", unless they have a very special shape, providing in particular the first explicit examples of functions on $\mathbf{Z}/p\mathbf{Z}$ with Gowers norms as small as those of "random" functions; (2) in [11], we use ideas of spherical codes to (roughly) bound from above the number of trace functions modulo p with bounded conductor; (3) in [12], we introduce a new method to estimate short exponential sums modulo primes, of length very close to $\sqrt{p}$, and obtain improvements for trace functions of the classical Polyá-Vinogradov bound.

Acknowledgments. We warmly thank U. Zannier for inviting one of us to present these results as a de Giorgi Colloquium at the Scuola Normale Superiore di Pisa and giving us the opportunity to present a written account of our work. We also thank F. Jouve and R. Cluckers for comments on this text.

1.1. **Notation.** We recall here some basic notation.
– The letters p will always refer to a prime number; for a prime p, we write $\mathbf{F}_p$ for the finite field $\mathbf{Z}/p\mathbf{Z}$. For a set X, $|X|$ is its cardinality, a non-negative integer or $+\infty$.
– The Landau and Vinogradov notation $f=O(g)$ and $f\ll g$ are synonymous, and $f(x)=O(g(x))$ for all $x\in D$ means that there exists an "implied" constant $C\geqslant 0$ (which may be a function of other parameters) such that $|f(x)|\leqslant Cg(x)$ for all $x\in D$. This definition *differs* from that of N. Bourbaki [2, Chap. V] since the latter is of

topological nature. We write $f \asymp g$ if $f \ll g$ and $g \ll f$. On the other hand, the notation $f(x) \sim g(x)$ and $f = o(g)$ are used with the asymptotic meaning of loc. cit.

2. Trace functions: definition

We present in this section the definition of trace functions over a finite field $\mathbf{F}_p$, and of the invariant which measures their complexity. This definition can, in fact, be presented from three different points of view: using automorphic forms (over $\mathbf{F}_p(T)$), using Galois representations (of the Galois group of $\mathbf{F}_p(T)$), or using étale sheaf theory. We use here the second because it is the most elementary, but the last is in fact the most convenient in many respects because it leads to the most flexible formalism, as we will describe. In order to be consistent with the terminology of our papers, we will use the language of sheaves after this section.

It is at least equally important to know the most common examples of trace functions, and for many applications to analytic number theory, one can in fact view trace functions as a kind of black box, building on the known very concrete examples and on the formalism these functions satisfy, especially the deep quasi-orthogonality property that encapsulates the Riemann Hypothesis over finite fields, as discussed in Section 4. In particular, readers who find the following definitions rather too abstract can just go through them very quickly, and concentrate their attention on the examples in the next section.

Let p be a fixed prime. We must first fix a different prime $\ell \neq p$, which plays an auxiliary role, and fix an isomorphism (of fields)

$$\iota : \bar{\mathbf{Q}}_\ell \longrightarrow \mathbf{C}$$

between a given algebraic closure of the field of ℓ-adic numbers and the field of complex numbers. In fact, we will mostly view this isomorphism as an algebraic identification, so that the reader may view $\bar{\mathbf{Q}}_\ell$ as just another name for $\mathbf{C}$; the main difference between the two, which is very important for the theory, is their different topological nature.

Let $K = \mathbf{F}_p(T)$ be the field of rational functions with coefficients in $\mathbf{F}_p$, and let $\bar{K}$ denote a separable closure of K (in which an algebraic closure $\bar{\mathbf{F}}_p$ of $\mathbf{F}_p$ is contained); elements of $\bar{K}$ can therefore be interpreted as "algebraic functions" on the projective line $\mathbf{P}^1(\bar{\mathbf{F}}_p)$, such as $\sqrt{f(X)}$ where $f \in \mathbf{F}_p[X]$ is a polynomial.

We let Π denote the Galois group of $\bar{K}$ over K. This groups contains a normal subgroup Π^g defined as the Galois group of $\bar{K}$ over the subfield $\tilde{K} = \bar{\mathbf{F}}_p(T)$, and the quotient Π/Π^g is naturally isomorphic to the Galois group of $\bar{\mathbf{F}}_p$ over $\mathbf{F}_p$, which is (topologically) generated by the arithmetic Frobenius automorphism $x \mapsto x^p$, or by its inverse, which is called the geometric Frobenius automorphism.

Definition 2.1 (ℓ-adic representation)**.** An *ℓ-adic Galois representation ϱ over* $\mathbf{F}_p$ is a continuous group homomorphism

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

for some finite-dimensional $\bar{\mathbf{Q}}_\ell$-vector space V, such that, for all but finitely many $x \in \mathbf{P}^1(\bar{\mathbf{F}}_p)$, the inertia group I_x at x is contained in the kernel of ϱ.

The dimension of V is called the rank of ϱ, and the set of $x \in \mathbf{P}^1(\bar{\mathbf{F}}_p)$ where I_x does not act trivially is called the set of singularities, or the set of ramification points, of ϱ, and is denoted $\mathrm{Sing}(\varrho)$. One also says that ϱ is *lisse at* $x \in \mathbf{P}^1(\bar{\mathbf{F}}_p)$ if I_x acts trivially on V.

We also say that ϱ is a *middle-extension ℓ-adic sheaf* modulo p, or sometimes just *ℓ-adic sheaf*. When using this language, we usually use curly letters, such as $\mathcal{F}$, instead of ϱ.

In this definition, as in classical algebraic number theory, the inertia group I_x is defined as the normal subgroup of the decomposition group

$$D_x = \{\gamma \in \Pi \mid \gamma(f(x)) = 0 \text{ for all } f \in \bar{K} \text{ such that } f(x) = 0\}$$

characterized by the condition

$$I_x = \{\gamma \in D_x \mid \gamma(f(x)) = f(x) \text{ for all } f \in \bar{K}\}.$$

These definitions make sense, because we can view an algebraic function $f \in \bar{K}$ as a "multi-valued function" on $\mathbf{P}^1(\bar{\mathbf{F}}_p)$: although $f(x)$ is not uniquely defined, all possible values are conjugates under the Galois group of $\bar{\mathbf{F}}_p/\mathbf{F}_p$, and the conditions defining D_x and I_x are invariant under this action (we also use the fact that γ also acts on $\bar{\mathbf{F}}_p \subset \bar{K}$).

It is immediately clear that this definition gives a relatively flexible formalism: we can form direct sums $\varrho_1 \oplus \varrho_2$, tensor products $\varrho_1 \otimes \varrho_2$, dual $\mathrm{D}(\varrho)$, of Galois representations, and we can define subrepresentations, quotient representations, and morphisms of representations (and therefore we can speak of isomorphic representations, or equivalently of isomorphic middle-extension sheaves).

For us, the point of this definition is that to each ℓ-adic representation ϱ is naturally attached a function $\mathbf{F}_p \longrightarrow \mathbf{C}$, which is called its trace function. From the representation theory point of view, it is just the restriction of the character $\mathrm{Tr}\,\varrho$ of the representation to special (conjugacy classes of) elements in Π.

Definition 2.2 (Trace function)**.** Let ϱ be an ℓ-adic Galois representation over $\mathbf{F}_p$. The *trace function* of ϱ is the function

$$t_\varrho : \mathbf{F}_p \longrightarrow \mathbf{C}$$

defined by

$$t_\varrho(x) = \iota\Big(\mathrm{Tr}(\varrho(\sigma_x \mid V^{I_x}))\Big),$$

for $x \in \mathbf{F}_p$, where σ_x denotes the conjugacy class of the geometric Frobenius automorphism at x, which generates topologically the quotient $D_x/I_x \simeq \mathrm{Gal}(\bar{\mathbf{F}}_p/\mathbf{F}_p)$ and V^{I_x} denotes the subspace of V invariant under the action of I_x, on which D_x/I_x acts naturally, while $\mathrm{Tr}(g \mid W)$ denotes the trace of an endomorphism g acting on a vector space W.

We have

$$t_{\varrho_1 \oplus \varrho_2} = t_{\varrho_1} + t_{\varrho_2}$$

and

$$t_{\varrho_1 \otimes \varrho_2}(x) = t_{\varrho_1}(x) t_{\varrho_2}(x)$$

for all x such that I_x acts trivially on ϱ_1 and ϱ_2 (at least).

In particular, the set of trace functions of ℓ-adic representations modulo p is closed under addition, and "almost" closed under products.

This set is infinite, and although not every function is of this form, it is dense for the uniform norm on functions modulo p. This implies that very few interesting analytic statements can be expected to hold for *all* trace functions. However, in applications, the trace functions that arise have two extra properties which rigidify the situation. Together, they explain the versatility and power of trace functions in analytic number theory.

The first condition is a restriction on the eigenvalues of the action of the Frobenius automorphisms on V.

Definition 2.3 (Weight 0 representation)**.** An ℓ-adic representation ϱ modulo p, acting on V, is *pointwise of weight* 0 if and only if the following condition holds:

For all finite extensions $k/\mathbf{F}_p$ and all $x \in \mathbf{P}^1(k)$ such that I_x acts trivially on V, the eigenvalues of $\varrho(\sigma_x)$ are algebraic numbers α such that all Galois conjugates of α have modulus 1.

In fact, it follows from non-trivial results of Deligne that one need only check the condition for x in the complement U of a finite set of points of $\mathbf{P}^1(\bar{\mathbf{F}}_p)$ such that ϱ is lisse at every point in U, and also that if ϱ is not lisse at $x \in \mathbf{P}^1(k)$ for some finite extension $k/\mathbf{F}_p$, it nevertheless satisfies the condition that there exists an integer $w \leqslant 0$ such that the eigenvalues of $\varrho(\sigma_x)$ are algebraic numbers α such that all Galois conjugates of α have modulus q^w (in particular, have modulus $\leqslant 1$).

Note that if ϱ, ϱ_1 and ϱ_2 are of weight 0, then so are $\varrho_1 \oplus \varrho_2$ and $\varrho_1 \otimes \varrho_2$, and the dual of ϱ, as well as any subrepresentation or quotient representation of ϱ. Furthermore, one can show that the trace function of the dual of a weight zero representation ϱ is the complex conjugate of the trace function of ϱ (which is easy at all unramified points, and the point is that it is also true for the possible ramified points, by a result of Gabber.)

A simple and immediate consequence of the definition is that the trace function of a representation of weight 0 satisfies

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

for all $x \in \mathbf{F}_p$. This is a first indication of how one can control the complexity of the trace function of an ℓ-adic representation.

Definition 2.4 (Trace function modulo p)**.** Let p be a prime number. A *trace function modulo* p is a function $t : \mathbf{F}_p \longrightarrow \mathbf{C}$ such that $t = t_{\mathcal{F}}$ for some ℓ-adic middle-extension sheaf $\mathcal{F}$ of weight 0.

Note that (as we will clearly see below) the sheaf $\mathcal{F}$ is not unique. However, it is unique if one restricts one's attention to representations with small complexity, in the sense that the conductor, which we now define, is small enough compared with p. This definition is absolutely essential for all analytic applications.

Definition 2.5 (Conductor)**.** Let $\varrho : \Pi \longrightarrow \mathrm{GL}(V)$ be an ℓ-adic representation modulo p of weight 0. The *conductor* of ϱ is

$$\mathbf{c}(\varrho) = \dim(V) + |\operatorname{Sing}(\varrho)| + \sum_{x \in \operatorname{Sing}(\varrho)} \operatorname{Swan}_x(\varrho),$$

where for a ramified point $x \in \mathbf{P}^1(\bar{\mathbf{F}}_p)$, we denote by $\operatorname{Swan}_x(\varrho)$ the Swan conductor of ϱ at x.

The precise definition of the Swan conductor, which measures fine properties of the representation ϱ at a ramified point, can be found for instance in [19, Ch. 1]. It is a rather subtle invariant, and we will mostly attempt to illustrate its meaning and properties with examples. For the moment, we will only mention that $\operatorname{Swan}_x(\varrho)$ is a non-negative integer. When $\operatorname{Swan}_x(\varrho) = 0$, one says that ϱ is *tamely ramified* at x. In a number of important cases, ϱ is tamely ramified at all $x \in \operatorname{Sing}(\varrho)$, in which case one says that ϱ itself is *tamely ramified*.

Finally, we remark that this definition of trace functions is not sufficient for certain constructions and applications, where slightly more general objects (constructible ℓ-adic sheaves) appear more naturally (see for instance [13]).

3. Trace functions: examples

The examples in this section are not only concrete examples of functions modulo p which are trace functions, but also examples of *operations* which may be performed on trace functions and lead to other trace functions. In all cases, it is very important to understand at least an upper-bound for the conductor of the associated ℓ-adic sheaves (or representations).

3.1. **Characters.** The simplest examples of trace functions are *character values*, involving either additive or multiplicative characters, or a product of them. Specifically, let p be a prime, and let

$$\psi : \mathbf{F}_p \longrightarrow \mathbf{C}^\times$$

be a non-trivial additive character (for instance, $\psi(x) = e(x/p)$ for $x \in \mathbf{F}_p$). Fix a rational function $f \in \mathbf{F}_p(X)$, which has no pole of order divisible by p, and define

$$t(x) = \begin{cases} e\left(\dfrac{f(x)}{p}\right) & \text{if } x \text{ is not a pole of } f \\ 0 & \text{otherwise.} \end{cases}$$

Then one can show that there exists an ℓ-adic middle-extension sheaf modulo p, denoted $\mathcal{L}_{\psi(f)}$, such that t is the trace function of $\mathcal{L}_{\psi(f)}$. Indeed, in contrast with most other examples we will discuss, this construction is elementary (see, e.g., [18, §11.11] for a discussion). A sheaf of the form $\mathcal{L}_{\psi(f)}$ is called an *Artin-Schreier sheaf*.

This sheaf is of rank 1 and ramified precisely at the poles of f (this applies also to the possible ramification at ∞, and uses the assumption that no pole is of order divisible by p). It has weight 0 (the image under ι of the only eigenvalue of Frobenius at any unramified $x \in \mathbf{P}^1(\mathbf{F}_p)$ is $\psi(f(x))$). As for the Swan conductors, if $x \in \mathbf{P}^1(\bar{\mathbf{F}}_p)$ is a pole of f, then one shows that $\mathrm{Swan}_x(\mathcal{L}_{\psi(f)})$ is *at most* equal to the order of the pole at x, and that there is equality at least if the pole is of order $< p$. Hence, if $f_2 \in \mathbf{F}_p[X]$ is the denominator of f, we have

$$\mathbf{c}(\mathcal{L}_{\psi(f)}) \leqslant 1 + 2\deg(f_2).$$

In particular, suppose we take now a rational function $f \in \mathbf{Q}(X)$, and we write $f = f_1/f_2$ with $f_i \in \mathbf{Z}[X]$. For all primes p large enough, we may reduce f modulo p to consider $f_1/f_2 \in \mathbf{F}_p(X)$. For each such p, we can form the corresponding sheaf $\mathcal{L}_{\psi(f)}$ modulo p, and what is essential is that the conductor of these sheaves is bounded by a constant depending only on f, and *not* on the prime p.

Similarly, let $\chi : \mathbf{F}_p^\times \longrightarrow \mathbf{C}^\times$ be a non-trivial multiplicative character, which can be seen as a Dirichlet character modulo p. Let $d \geqslant 2$ be the order of χ. Fix a rational function $f \in \mathbf{F}_p(X)$ such that f has no pole or zero of order divisible by d, and define

$$t(x) = \begin{cases} \chi(f(x)) & \text{if } x \text{ is not a zero or pole of } f \\ 0 & \text{otherwise.} \end{cases}$$

Then one can show (again, elementarily) that there exists an ℓ-adic middle-extension sheaf modulo p, denoted $\mathcal{L}_{\chi(f)}$, such that t is equal to the trace function of $\mathcal{L}_{\chi(f)}$ (through ι). Such sheaves are called *Kummer sheaves*.

This representation is of rank 1 and ramified precisely at the zeros and poles of f (in $\mathbf{P}^1(\bar{\mathbf{F}}_p)$). It has weight 0, the only (image under ι of an) eigenvalue of Frobenius at any unramified $x \in \mathbf{P}^1(\mathbf{F}_p)$ being equal to $\chi(f(x))$. Furthermore, it is *tamely ramified*. Thus

$$\mathbf{c}(\mathcal{L}_{\chi(f)}) \leqslant 1 + \deg(f_1) + \deg(f_2)$$

if $f = f_1/f_2$ with $f_i \in \mathbf{F}_p[X]$ coprime. As before, if we obtain f by reduction modulo p from a fixed rational function with rational coefficients, this family of sheaves indexed by p has conductor bounded independently of p.

One can combine these examples by tensor product, which amounts to multiplying the trace functions: if $\chi_1, \ldots, \chi_r$ are finitely many distinct non-trivial multiplicative characters, and $f_1, \ldots, f_{r+1}$ are rational functions in $\mathbf{F}_p(X)$, the middle-extension sheaf

$$\varrho = \mathcal{L}_{\chi_1(f_1)} \otimes \cdots \otimes \mathcal{L}_{\chi_r(f_r)} \otimes \mathcal{L}_{\psi(f_{r+1})}$$

has rank 1, weight 0, and it is unramified[1] at least outside of the union of the poles of f_{r+1} and the zeros and poles of $f_1, \ldots, f_r$, it has trace function

$$t_\varrho(x) = \chi_1(f_1)\cdots\chi_r(f_r)e\Big(\frac{f_{r+1}(x)}{p}\Big)$$

for all unramified $x \in \mathbf{F}_p$, and has conductor

$$\mathbf{c}(\varrho) \leqslant 1 + \sum_{i=1}^{r} (\deg(g_i) + \deg(h_i)) + \deg(h_{r+1}),$$

where $f_i = g_i/h_i$ with g_i, $h_i \in \mathbf{F}_p[X]$ coprime.

Example 3.1. If we take $f(X) = X^p - X \in \mathbf{F}_p[X]$, the trace function associated to $\mathcal{L}_{\chi(f)}$, for χ non-trivial, satisfies

$$t_{\mathcal{L}_{\chi(f)}}(x) = \chi(x^p - x) = 0$$

for all $x \in \mathbf{F}_p$. Thus, for any trace function t associated to a sheaf $\mathcal{F}$, we also have

$$t = t_{\mathcal{F}\oplus\mathcal{L}_{\chi(f)}},$$

illustrating the non-uniqueness of $\mathcal{F}$. Note however that this second sheaf has huge conductor in terms of p:

$$\mathbf{c}(\mathcal{F} \oplus \mathcal{L}_{\chi(f)}) \geqslant p.$$

3.2. **Point-counting functions.** The second class of examples is also relatively elementary. We consider a non-constant squarefree polynomial $f \in \mathbf{F}_p[X]$, and for $x \in \mathbf{F}_p$, we denote

$$n_f(x) = |\{y \in \mathbf{F}_p \mid f(y) = x\}|,$$

the number of pre-images of x in the field $\mathbf{F}_p$. In particular, $n_f(x) = 0$ if x is not of the form $x = f(y)$ for *some* $y \in \mathbf{F}_p$. One can then construct an ℓ-adic middle-extension sheaf $\mathcal{F}_f$ such that $t_{\mathcal{F}_f}(x) = n_f(x)$ for all $x \in \mathbf{F}_p$. (Indeed, if $K = \mathbf{F}_p(f(X))$, we have a Galois extension $\mathbf{F}_p(X)/\mathbf{F}_p(f(X))$, and since $\mathbf{F}_p(f(X))$ is isomorphic to $\mathbf{F}_p(X)$, this gives a homomorphism $\mathrm{Gal}(\bar{K}/K) \longrightarrow \mathrm{Gal}(\bar{K}/K)$ with image Π_f of finite index in $\Pi = \mathrm{Gal}(\bar{K}/K)$; then the representation ϱ_f which "is" $\mathcal{F}_f$ can be defined as the induced representation $\mathrm{Ind}_{\Pi_f}^{\Pi}(\bar{\mathbf{Q}}_\ell)$.)

Because the induced representation always contains a copy $\mathbf{1}$ of the trivial representation, it is often convenient to consider $\mathcal{F}'_f = \mathcal{F}_f/\mathbf{1}$, which has trace function

$$n_f^0(x) = t_{\mathcal{F}'_f}(x) = n_f(x) - 1.$$

The point (for applications) is that 1 is the average value of $n_f(x)$, so we have

$$\sum_{x\in\mathbf{F}_p} n_f^0(x) = \sum_{x\in\mathbf{F}_p} n_f(x) - p = \sum_{x\in\mathbf{F}_p} \sum_{\substack{y\in\mathbf{F}_p \\ f(y)=x}} 1 - p = \sum_{y\in\mathbf{F}_p} 1 - p = 0.$$

[1] Depending on the χ_i and f_i, it might be unramified at some extra points.

The representation $\mathcal{F}'_f$ has rank $\deg(f)-1$. It has weight 0, and it is unramified at all $x \in \mathbf{P}^1(\bar{\mathbf{F}}_p)$ such that the pre-image $f^{-1}(x) \subset \mathbf{P}^1(\bar{\mathbf{F}}_p)$ (with coefficients in $\bar{\mathbf{F}}_p$ now) consists of $\deg(f)$ different points, or in other words, at all regular values of f (where a regular value is one which is not a critical value, and a critical value is the image of a critical point $y \in \bar{\mathbf{F}}_p$, i.e., of a root of f').

If $p > \deg(f)$, it is known that $\mathcal{F}_f$ and $\mathcal{F}'_f$ are tamely ramified. In particular, for $\deg(f) \geqslant 2$ (so that $\mathcal{F}'_f$ is non-zero), we have

$$\mathbf{c}(\mathcal{F}'_f) \leqslant \deg(f) - 1 + \deg(f) - 1 = 2\deg(f) - 2.$$

The trace function of $\mathcal{F}_f$ counts the number of solutions in $\mathbf{F}_p$ of the equation $f(y) = x$, and therefore it is supported on the subset $f(\mathbf{F}_p) \subset \mathbf{F}_p$. In our motivating problem in Section 1, on the other hand, we considered the function ξ_p which is the characteristic function of this set. This is also related to trace functions of ℓ-adic sheaves, although it is not quite one. Rather, one shows (the construction is again elementary, see [8, Prop. 6.7]) that if $\deg(f) > p$, there exist middle-extension sheaves $\mathcal{F}_{f,i}$, $1 \leqslant i \leqslant m \leqslant \deg(f)$, all geometrically non-trivial, pointwise pure of weight 0 and tamely ramified, and algebraic numbers c_0, c_i for $1 \leqslant i \leqslant m$, such that

$$\xi_p(x) = c_0 + \sum_{i=1}^{m} c_i t_{\mathcal{F}_{f,i}}(x) \tag{3.1}$$

for all $x \in \mathbf{F}_p - S$, where $S \subset \mathbf{F}_p$ is a finite set of cardinality $\leqslant \deg(f)$. Moreover, the parameters m, $|c_i|$, $\mathbf{c}(\mathcal{F}_{f,i})$ which measure the complexity of this decomposition are all bounded in terms of $\deg(f)$, and

$$c_0 = \frac{|f(\mathbf{F}_p)|}{p} + O(p^{-1/2}),$$

where the implied constant depends only on $\deg(f)$.

Remark 3.2. If $f = X^2 + a$, then (for all odd primes p) the trace functions $n_f(x)$ and $n_f^0(x)$ are given by

$$n_f(x) = 1 + \Big(\frac{x-a}{p}\Big), \qquad n_f^0(x) = \Big(\frac{x-a}{p}\Big),$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol, while ξ_p is given by

$$\xi_p(x) = \frac{1}{2}\Big(1 + \Big(\frac{x-a}{p}\Big)\Big),$$

for $x \neq a$, and $\xi_p(a) = 1$. Thus we have

$$\mathcal{F}_f = \mathbf{1} \oplus \mathcal{L}_{\left(\frac{X-a}{p}\right)}, \qquad \mathcal{F}'_f = \mathcal{L}_{\left(\frac{X-a}{p}\right)}.$$

The general decomposition is therefore a (non-obvious) generalization of this fact.

3.3. **Exponential sums.** Our next example may seem very specialized, but it plays a critical role in many deep results in analytic number theory. Let ψ be the non-trivial additive character ψ modulo p given by $\psi(x) = e(x/p)$ (note that this depends on p). For an integer $n \geqslant 1$, we consider the (normalized) *hyper-Kloosterman sums* $\mathrm{Kl}_n(x;p)$ defined by

$$\mathrm{Kl}_n(x;p) = \frac{(-1)^{n-1}}{p^{(n-1)/2}} \sum_{\substack{y_1,\dots,y_n \in \mathbf{F}_p^\times \\ y_1\cdots y_n = x}} \cdots \sum \psi(y_1 + \cdots + y_n)$$

for $x \in \mathbf{F}_p^\times$. For instance we have $\mathrm{Kl}_1(x;p) = e(x/p)$

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

which is the classical Kloosterman sum with parameter x. It is now a highly non-trivial fact that there exists an ℓ-adic middle-extension sheaf $\mathcal{K}\ell_n$ modulo p, called a *Kloosterman sheaf* such that

$$t_{\mathcal{K}\ell_n}(x) = \mathrm{Kl}_n(x;p)$$

for $x \in \mathbf{F}_p^\times$. In fact, as far as we are aware, there is no proof of the existence of this representation directly in the framework of Galois representations: one must construct it as an ℓ-adic sheaf (a construction due to Deligne, and extensively studied by Katz [19], which was recently generalized by Heinloth, Ngô and Yun [17]).

Among the results of Deligne and Katz concerning Kloosterman sheaves are the following: $\mathcal{K}\ell_n$ is geometrically irreducible, it is of rank n, pointwise pure of weight 0, ramified only at 0 (if $n \geqslant 2$) and ∞ in $\mathbf{P}^1(\bar{\mathbf{F}}_p)$, and the ramification is tame at 0 (i.e., $\mathrm{Swan}_0(\mathcal{K}\ell_n) = 0$) and wild at ∞ with $\mathrm{Swan}_\infty(\mathcal{K}\ell_n) = 1$. Thus, for every prime p and $n \geqslant 2$, we have

$$\mathbf{c}(\mathcal{K}\ell_n) = n + 2 + 1 = n + 3,$$

and it is again crucial that the conductor is bounded independently of p.

To see how deep such results are, note that since $\mathcal{K}\ell_n$ is of weight 0 and unramified at $x \in \mathbf{F}_p^\times$, it follows that

$$|\mathrm{Kl}_n(x;p)| \leqslant n$$

for all primes p and all $x \in \mathbf{F}_p^\times$, for instance

$$\Big|\sum_{y \in \mathbf{F}_p^\times} e\Big(\frac{xy+\bar{y}}{p}\Big)\Big| \leqslant 2\sqrt{p}$$

for $x \in \mathbf{F}_p^\times$. This bound is the well-known *Weil bound for Kloosterman sums*, and it has countless applications in analytic number theory (due particularly to the presence of Kloosterman sums in the theory of automorphic forms, see e.g. [21] for a survey).

3.4. **Operating on trace functions.** It is a fundamental aspect of ℓ-adic sheaves and their trace functions that a flexible formalism is available in their study, and for applications. Besides the standard operations mentioned above (direct sums, tensor product, dual), we will illustrate this point here with one particular operation that is very relevant for our papers, in particular in [7, 8, 9]: the Fourier transform.

For a prime number p, a non-trivial additive character ψ and a function $\varphi : \mathbf{F}_p \longrightarrow \mathbf{C}$, we define here the Fourier transform $\mathrm{FT}_\psi(\varphi) : \mathbf{F}_p \longrightarrow \mathbf{C}$ by the formula

$$\mathrm{FT}_\psi(\varphi)(t) = -\frac{1}{\sqrt{p}} \sum_{x \in \mathbf{F}_p} \varphi(x)\psi(xt)$$

for $t \in \mathbf{F}_p$. If φ is a trace function as we defined them, the Fourier transform can not always be one, because the Fourier transform of a constant, for instance, is a delta function, which does not fit our framework well. However, exploiting the deep fact (due to Deligne [4, (3.4.1)]) that a middle-extension sheaf modulo p of weight 0 is, geometrically (i.e., over $\bar{\mathbf{F}}_p$) a direct sum of irreducible sheaves over $\bar{\mathbf{F}}_p$, one can define a *Fourier sheaf* modulo p to be one where no such geometrically irreducible component is isomorphic to

an Artin-Schreier sheaf $\mathcal{L}_{\psi(aX)}$ for some $a \in \bar{\mathbf{F}}_p$. Then Deligne showed that there exists an operation

$$\mathcal{F} \mapsto \mathrm{FT}_\psi(\mathcal{F})$$

at the level of Fourier ℓ-adic sheaves with the property that

$$t_{\mathrm{FT}_\psi(\mathcal{F})} = \mathrm{FT}(t_\mathcal{F}),$$

i.e., the trace function of the Fourier transform of a sheaf $\mathcal{F}$ is equal to the Fourier transform of the trace function of $\mathcal{F}$. This operation was studied in depth by Laumon [22], Brylinski and Katz [19, 20], and shown to satisfy the following properties (many of which are, intuitively, analogues of classical properties of the Fourier transform):

(1) If a Fourier sheaf $\mathcal{F}$ is pointwise of weight 0, then so is $\mathrm{FT}_\psi(\mathcal{F})$: this fact is extremely deep, as it relies on a refined application of the Riemann Hypothesis over finite fields.

(2) If $\mathcal{F}$ is geometrically irreducible, then so is $\mathrm{FT}_\psi(\mathcal{F})$ (as we will see in Section 4, this is to some extent an analogue of the unitarity property

$$\|\mathrm{FT}(\varphi))\|^2 = \|\varphi\|^2$$

of the Fourier transform of a function $\varphi : \mathbf{F}_p \longrightarrow \mathbf{C}$, where

$$\|\varphi\|^2 = \frac{1}{p}\sum_{x\in\mathbf{F}_p} |\varphi(x)|^2$$

is the standard L^2-norm.)

(3) Laumon [22] developed in particular a theory of "local Fourier transforms" which is an analogue of the stationary phase method in classical analysis, and which leads to very detailed information concerning the ramification properties of $\mathrm{FT}_\psi(\mathcal{F})$. Using this, we proved in [7] that the Fourier transform of sheaves has the important property that the conductor of $\mathrm{FT}_\psi(\mathcal{F})$ can be estimated solely in terms of the conductor of $\mathcal{F}$, and more precisely we showed:

$$\mathbf{c}(\mathrm{FT}_\psi(\mathcal{F})) \leqslant 10\,\mathbf{c}(\mathcal{F})^2.$$

This estimate is essential in analytic applications, since it implies that if p varies but $\mathcal{F}$ has bounded conductor, so do the Fourier transforms. In [13], we have extended such estimates to other linear transformations $\varphi \mapsto T\varphi$ of the type

$$(T\varphi)(x) = -\frac{1}{\sqrt{p}}\sum_{y\in\mathbf{F}_p} \varphi(y)\psi(f(x,y))$$

for arbitrary rational functions f.

Example 3.3. (1) As a first example, note that the function K_p defined by (1.4) for a fixed polynomial $f \in \mathbf{Z}[X]$ and ξ_p the characteristic function of $f(\mathbf{F}_p) \subset \mathbf{F}_p$ is (up to a factor $p^{1/2}$) the Fourier transform of $\xi_p - \delta_p$. By (3.1) and the remarks following, we see that we have (essentially)

$$K_p = \frac{1}{\sqrt{p}}\sum_{i=1}^{m} c_i t_{\mathrm{FT}(\mathcal{F}_{f,i})}, \tag{3.2}$$

i.e., $\sqrt{p}K_p$ is, if not a trace function, then at least a "short" linear combination of trace functions with bounded complexity in terms of $\deg(f)$. This is a crucial step in the proof of the results we mentioned in this first section.

(2) Consider the Artin-Schreier sheaf $\mathcal{L}_{\psi(X^{-1})}$ as in Section 3.1. Then we see that the Fourier transform $\mathcal{F} = \mathrm{FT}(\mathcal{L}_{\psi(X^{-1})})$ has trace function

$$t_{\mathcal{F}}(x) = -\frac{1}{\sqrt{p}} \sum_{y \in \mathbf{F}_p} \psi(y^{-1})\psi(xy) = \mathrm{Kl}_2(x; p).$$

In fact, this sheaf $\mathcal{F}$ is the same as the Kloosterman sheaf $\mathcal{K}\ell_2$ discussed in Section 3.3, and one can deduce all the basic properties of the latter from the general theory of the Fourier transform. The other Kloosterman sheaves $\mathcal{K}\ell_n$, for $n \geqslant 3$, can be constructed similarly using the operation of *multiplicative convolution* on trace functions.

4. Quasi-orthogonality of trace functions

The most important analytic property of trace functions lies in the quasi-orthogonality of trace functions of geometrically irreducible sheaves of weight 0, which is a very important and general form of the Riemann Hypothesis over finite fields as proved by Deligne [4].

Theorem 4.1 (Deligne)**.** *Let p be a prime number, $\ell \neq p$ a prime distinct from p and let $\mathcal{F}_1$, $\mathcal{F}_2$ be geometrically irreducible ℓ-adic sheaves modulo p which are pointwise of weight 0.*

(1) *If $\mathcal{F}_1$ is not geometrically isomorphic to $\mathcal{F}_2$, then we have*

$$\left|\sum_{x \in \mathbf{F}_p} t_{\mathcal{F}_1}(x)\overline{t_{\mathcal{F}_2}(x)}\right| \leqslant 3\,\mathbf{c}(\mathcal{F}_1)^2\,\mathbf{c}(\mathcal{F}_2)^2 p^{1/2}. \tag{4.1}$$

(2) *If $\mathcal{F}_1$ is geometrically isomorphic to $\mathcal{F}_2$, then there exists a complex number α with modulus 1 such that*

$$t_{\mathcal{F}_1}(x) = \alpha t_{\mathcal{F}_2}(x)$$

for all $x \in \mathbf{F}_p$, and

$$\left|\sum_{x \in \mathbf{F}_p} t_{\mathcal{F}_1}(x)\overline{t_{\mathcal{F}_2}(x)} - \alpha p\right| \leqslant 3\,\mathbf{c}(\mathcal{F}_1)^2\,\mathbf{c}(\mathcal{F}_2)^2 p^{1/2}. \tag{4.2}$$

Note that, in that case, we have $\mathbf{c}(\mathcal{F}_1) = \mathbf{c}(\mathcal{F}_2)$.

To be precise, it follows from the Grothendieck-Lefschetz trace formula (see, e.g., [3, Rapport, Th. 3.2]) and the Riemann Hypothesis [4, Th. 3.3.1] that both inequalities hold with right-hand side replaced by

$$(\dim H^1_c(\mathbf{A}^1 \times \bar{\mathbf{F}}_p, \mathcal{F}_1 \otimes \mathrm{D}(\mathcal{F}_2)))p^{1/2},$$

where we recall that $\mathrm{D}(\mathcal{F}_2)$ denotes the dual of $\mathcal{F}_2$, and from the Grothendieck-Ogg-Shafarevich formula for the Euler-Poincaré characteristic of a sheaf (see, e.g., [19, 2.3.1]), one obtains relatively easily a bound

$$\dim H^1_c(\mathbf{A}^1 \times \bar{\mathbf{F}}_p, \mathcal{F}_1 \otimes \mathrm{D}(\mathcal{F}_2)) \leqslant 3\,\mathbf{c}(\mathcal{F}_1)^2\,\mathbf{c}(\mathcal{F}_2)^2$$

(see, e.g., [11, Lemma 3.3]).

Remark 4.2. One useful interpretation of this result is as an approximate version of the orthogonality relations of characters of representations of finite (or compact) groups, which algebraically is related to Schur's Lemma. In particular, note that it implies that if $\mathbf{c}(\mathcal{F}_1)$ and $\mathbf{c}(\mathcal{F}_2)$ are small enough (roughty $\ll p^{1/8}$ in this version), the condition $t_{\mathcal{F}_1} = t_{\mathcal{F}_2}$ of equality of the trace functions suffices to imply that $\mathcal{F}_1$ and $\mathcal{F}_2$ are geometrically isomorphic. In [11], we use this fact, as well as bounds on the number of quasi-orthogonal

unit vectors in a finite-dimensional Hilbert space to bound from above the number of geometrically irreducible ℓ-adic sheaves with bounded complexity.

In order to illustrate how this theorem is used, we will state precisely a version of the main theorem of [7] and explain which sums of trace functions arise in the proof.

A holomorphic cusp form of integral weight[2] $k \geqslant 2$ and level $N \geqslant 1$ is a holomorphic function

$$f : \mathbf{H} \longrightarrow \mathbf{C}$$

such that

$$f\Big(\frac{az+b}{cz+d}\Big) = (cz+d)^k f(z)$$

for all elements in the subgroup $\Gamma_0(N)$ of elements $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$ such that N divides c, and furthermore

$$\int_{F_N} |f(z)|^2 y^k \frac{dxdy}{y^2} < +\infty, \tag{4.3}$$

where

$$F_N = \bigcup_{\gamma \in \Gamma_0(N)\backslash \mathrm{SL}_2(\mathbf{Z})} \gamma \cdot F$$

in terms of the fundamental domain F as in Section 1. It follows that $f(z+1) = f(z)$ and therefore f has a Fourier expansion, which holomorphy and the growth condition force to be

$$f(z) = \sum_{n \geqslant 1} n^{(k-1)/2} \varrho_f(n) e(nz)$$

for some coefficients $\varrho_f(n) \in \mathbf{C}$ (the normalizing factor $n^{(k-1)/2}$ has the effect of ensuring that $\varrho_f(n)$ is bounded in mean-square average).

Theorem 4.3 ([7])**.** *Let f be a fixed cusp form as above. For any prime p and for any function $K : \mathbf{Z} \to \mathbf{C}$ such that $K(n) = t_{\mathcal{F}}(n \,(\mathrm{mod}\, p))$ for some ℓ-adic representation $\mathcal{F}$ modulo p of weight 0, we have*

$$\sum_{n \leqslant X} \varrho_f(n) K(n) \ll \mathbf{c}(\mathcal{F})^9 X \Big(1 + \frac{p}{X}\Big)^{1/2} p^{-1/16+\varepsilon}$$

for $X \geqslant 1$ and any $\varepsilon > 0$, where the implied constant depends only on f and on $\varepsilon > 0$.

The trivial bound for the sum is

$$\Big|\sum_{n \leqslant X} \varrho_f(n) K(n)\Big| \leqslant \Big(\sum_{n \leqslant X} |\varrho_f(n)|^2\Big)^{1/2} \Big(\sum_{n \leqslant X} |K(n)|^2\Big)^{1/2} \ll \mathbf{c}(\mathcal{F}) X$$

by the well-known Rankin-Selberg estimate

$$\sum_{n \leqslant X} |\varrho_f(n)|^2 \sim c_f X$$

for some $c_f > 0$ as $X \to \infty$. Thus, assuming that the conductor is bounded by a fixed constant B, our theorem is non-trivial provided X is of size between p (or a bit smaller) and p^A for some fixed A. For the critical case $X = p$, we get a saving of size $p^{-1/16+\varepsilon}$ over the trivial bound.

[2] This notion of weight is not directly related in general to that of weight 0 sheaves.

In particular, if we have an integer $B \geqslant 1$, and for each prime p a trace function K_p in such a way that the associated representations satisfy $\mathbf{c}(\mathcal{F}_p) \leqslant B$, then it follows that for p large, there is no correlation between the phases of $\varrho_f(n)$ and those of $K_p(n)$.

Slightly more general versions of this theorem apply to the sums (1.3) which arose in Section 1, and combined with the decomposition (3.2) of the functions (1.4), this leads to the proof of the limit relations (1.5) discussed in the motivating problem.

A striking feature of Theorem 4.3 is the universality of the exponent 1/16 (which can be improved to 1/8 for "smooth" sums). This is a direct effect of Theorem 4.1 and the universality of the exponent 1/2 in the right-hand side of (4.1).

We outline the basic strategy of the proof, to indicate where the Riemann Hypothesis comes into play. First, by elementary decompositions, we may assume that the ℓ-adic sheaf $\mathcal{F}$ with trace function K is geometrically irreducible, and by dealing directly with Artin-Schreier sheaves $\mathcal{L}_{\psi(aX)}$, we may assume that $\mathcal{F}$ is a Fourier sheaf. Applying deep results from the theory of automorphic forms (especially the Kuznetsov formula, Hecke theory, and the amplification method) one reduces estimates for the sums in Theorem 4.3 to the study of certain sums of the type

$$\sum_{\boldsymbol{\alpha}\in X} c(\boldsymbol{\alpha})\mathcal{C}(K;\gamma(\boldsymbol{\alpha})) \tag{4.4}$$

where X is a certain finite set of parameters, $c(\boldsymbol{\alpha})$ are complex numbers, $\gamma(\boldsymbol{\alpha})$ is an element of the finite group $\mathrm{PGL}_2(\mathbf{F}_p)$, and the *correlation sums* $\mathcal{C}(K;\gamma(\boldsymbol{\alpha}))$ are defined by

$$\mathcal{C}(\varphi;\gamma) = \sum_{\substack{x\in\mathbf{F}_p\\ \gamma\cdot x\neq\infty}} \overline{\mathrm{FT}(\varphi)(x)}\,\mathrm{FT}(\varphi)(\gamma\cdot x)$$

for any function $\varphi : \mathbf{F}_p \longrightarrow \mathbf{C}$ and $\gamma \in \mathrm{PGL}_2(\mathbf{F}_p)$, which we view as acting on $\mathbf{P}^1(\mathbf{F}_p) = \mathbf{F}_p \cup \{\infty\}$ by the usual action (the same formula as in (1.1)).

From the theory of the Fourier transform, as explained in Section 3.4, we know that $\mathrm{FT}(K)$ is a trace function associated to a geometrically irreducible Fourier sheaf of weight 0 with conductor $\leqslant 10\,\mathbf{c}(\mathcal{F})^2$. Furthermore, for any ℓ-adic sheaf modulo p and $\gamma \in \mathrm{PGL}_2(\mathbf{F}_p)$, we have an elementary definition of an ℓ-adic sheaf $\gamma^*\mathcal{F}$ with trace function given by $x \mapsto t_{\mathcal{F}}(\gamma\cdot x)$, and with the same conductor as $\mathcal{F}$. Thus the factor $\mathrm{FT}(K)(\gamma\cdot x)$ is also the trace function of an ℓ-adic sheaf (geometrically irreducible, weight 0) with conductor $\leqslant 10\,\mathbf{c}(\mathcal{F})^2$.

It turns out that the reduction procedure implies that good estimates for the original sum follow if we can use in (4.4) a square-root cancellation estimate

$$|\mathcal{C}(K;\gamma(\boldsymbol{\alpha}))| \leqslant Cp^{1/2} \tag{4.5}$$

for *all* $\boldsymbol{\alpha} \in X$. On the other hand, Theorem 4.1 easily implies that if C is large enough in terms of the conductor of $\mathcal{F}$, we have an inclusion

$$\{\gamma \in \mathrm{PGL}_2(\mathbf{F}_p) \mid |\mathcal{C}(K;\gamma)| > Cp^{1/2}\} \subset \{\gamma \in \mathrm{PGL}_2(\mathbf{F}_p) \mid \gamma^*\,\mathrm{FT}_\psi(\mathcal{F}) \simeq \mathrm{FT}_\psi(\mathcal{F})\}$$

(where $\simeq$ denotes geometric isomorphism; we use here the reduction to geometrically irreducible $\mathcal{F}$).

The crucial point is that the right-hand side is a group, which we denote $\mathbf{G}_{\mathrm{FT}_\psi(\mathcal{F})} \subset \mathrm{PGL}_2(\mathbf{F}_p)$ and call the "Möbius group of $\mathrm{FT}_\psi(\mathcal{F})$". This group may well be non-trivial, or relatively large, so that (4.5) cannot in general be expected to hold in all cases.

Using the classification of subgroups of $\mathrm{PGL}_2(\mathbf{F}_p)$, we can nevertheless conclude using the following proposition:

Proposition 4.4. *Let p be a prime number and let $\mathcal{G}$ be a geometrically irreducible ℓ-adic Fourier sheaf of weight 0 modulo p. Then, if p is large enough compared to the conductor of $\mathcal{G}$, one of the following properties holds:*

(1) *The Möbius group $\mathbf{G}_{\mathcal{G}}$ contains an element of order p; in this case $\mathcal{G} \simeq \gamma^*\mathcal{L}_{\psi(aX)}$ for some $a \in \bar{\mathbf{F}}_p$ and some element $\gamma \in \mathrm{PGL}_2(\mathbf{F}_p)$, and $\mathbf{G}_{\mathcal{F}}$ is then conjugate in $\mathrm{PGL}_2(\mathbf{F}_p)$ to the subgroup*

$$U = \Big\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \mid t \in \mathbf{F}_p \Big\}.$$

(2) *The Möbius group $\mathbf{G}_{\mathcal{G}}$ has order coprime to p, and in this case it is contained in the union of at most 60 subgroups, each of which is either a conjugate of the normalizer in $\mathrm{PGL}_2(\mathbf{F}_p)$ of the diagonal torus*

$$T = \Big\{ \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \mid x, y \in \mathbf{F}_p^\times \Big\}$$

or a conjugate of the normalizer in $\mathrm{PGL}_2(\mathbf{F}_p)$ of a non-split torus

$$T_1 = \Big\{ \begin{pmatrix} a & b \\ \varepsilon b & a \end{pmatrix} \mid a^2 - \varepsilon b^2 \neq 0 \Big\},$$

where $\varepsilon \in \mathbf{F}_p^\times$ is a non-square.

We can then exploit the fact that the elements of the form $\gamma(\boldsymbol{\alpha})$ are explicit, and from their origin in the analytic steps, they have no particular algebraic structure. In particular, they are seen to *not* be conjugate to elements of the subgroup U in this proposition (for p large enough compared with the conductor), so that in the first case of the proposition (applied to $\mathcal{G} = \mathrm{FT}_\psi(\mathcal{F})$), we have the estimates (4.5) for all $\gamma(\boldsymbol{\alpha})$. If the second case of the proposition applies, on the other hand, we exploit a repulsion argument for each of the finitely many possible conjugates N of the normalizer of a torus to show, roughly, that if one $\gamma(\boldsymbol{\alpha})$ is in N (which may happen) then there can only be extremely few other $\boldsymbol{\alpha}'$ with $\gamma(\boldsymbol{\alpha}') \in \mathbf{G}_{\mathrm{FT}_\psi(\mathcal{F})}$. Such a small set of exceptions to the estimate (4.5) can then be handled.

Example 4.5. (1) Let

$$K(n) = e\Big(\frac{\bar{n}}{p}\Big),$$

the trace function of the Artin-Schreier sheaf $\mathcal{L}_{\psi(X^{-1})}$. Then $-\mathrm{FT}(K)$ is the trace function of the Kloosterman sheaf $\mathcal{K}\ell_2$ modulo p, and hence

$$\mathcal{C}(K;\gamma) = \sum_{\substack{x \in \mathbf{F}_p \\ cx+d \neq 0}} \mathrm{Kl}_2(x;p)\,\mathrm{Kl}_2\Big(\frac{ax+b}{cx+d};p\Big)$$

for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ (since Kloosterman sums $\mathrm{Kl}_2(x;p)$ are real numbers). One can show that $\mathbf{G}_{\mathcal{K}\ell_2} = 1$ is the trivial group, and hence there exists a constant $C \geqslant 1$ such that

$$\Big| \sum_{\substack{x \in \mathbf{F}_p \\ cx+d \neq 0}} \mathrm{Kl}_2(x;p)\,\mathrm{Kl}_2\Big(\frac{ax+b}{cx+d};p\Big) \Big| \leqslant Cp^{1/2}$$

for all p prime and $\gamma \neq 1$. We will see in the next section some other applications of special cases of this estimate.

(2) For p prime and $n \in \mathbf{F}_p$, define

$$K(n) = \mathrm{Kl}_2(n^2;p) - 1.$$

This is the trace function of an ℓ-adic sheaf modulo p, the symmetric square $\mathcal{F}$ of the pull-back of the Kloosterman sheaf $\mathcal{K}\ell_2$ under the map $x \mapsto x^2$, and this description also shows that the conductor of $\mathcal{F}$ is bounded independently of p. It is a non-trivial fact that $\mathbf{G}_{\mathrm{FT}_\psi(\mathcal{F})}$, in that case, is the subgroup of $\mathrm{PGL}_2(\mathbf{F}_p)$ stabilizing the subset $\{\infty, 0, 4, -4\}$ of $\mathbf{P}^1(\mathbf{F}_p)$, which is a dihedral group of order 8. (More precisely, the inclusion of $\mathbf{G}_{\mathrm{FT}_\psi(\mathcal{F})}$ in this group is elementary, because $\mathcal{F}$ is ramified exactly at these points, and the converse can be checked in different ways, none of which is elementary – maybe the most elegant is to use a result of Deligne and Flicker [5, Cor. 7.7].)

Prior to [7], the only instances of Theorem 4.3 that were considered in the literature (to our knowledge) where $K(n) = e(an/p)$, an additive character, or $K(n) = \chi(n)$, where χ is a non-trivial Dirichlet character modulo p. In the first case, there is an even stronger bound

$$\sum_{n \leqslant X} \varrho_f(n) e\Big(\frac{an}{p}\Big) \ll X^{1/2}(\log X)$$

due to Wilton, valid uniformly for all a modulo p (and in fact we use this estimate directly for $K(n) = e(an/p)$, or equivalently for Artin-Schreier sheaves $\mathcal{L}_{\psi(aX)}$.) The case of a multiplicative character is related to the subconvexity problem for the twisted special values of L-functions $L(f \otimes \chi, 1/2)$ (see for instance [23, Lecture 4] for a survey), and non-trivial estimates were first found by Duke, Friedlander and Iwaniec [6]. The bound in Theorem 4.3 recovers the best known result in terms of the modulus, due to Blomer and Harcos [1] (although the latter deals more generally with characters to all moduli $q \geqslant 1$, and not only primes). In this case, one sees that $\mathrm{FT}_\psi(\mathcal{L}_\chi) \simeq \mathcal{L}_{\bar{\chi}}$ and that $\mathbf{G}_{\mathcal{L}_{\bar{\chi}}}$ is either the diagonal torus T of Proposition 4.4 (2), or its normalizer in $\mathrm{PGL}_2(\mathbf{F}_p)$ (this last case occurring if and only if χ is a real character).

5. Distribution of arithmetic functions in arithmetic progressions

The result of Theorem 4.3 is not only interesting as a statement concerning modular forms. Generalizing the result to encompass Eisenstein series and not only cusp forms, and applying further methods from the analytic study of prime numbers, as well as more general properties of trace functions, we obtained in [8] a striking application to sums over primes (or against the Möbius function).

Theorem 5.1 ([8])**.** *Let p be a prime number and let $K = t_{\mathcal{F}}$ be the trace function of an ℓ-adic middle-extension $\mathcal{F}$ of weight 0 modulo p such that no geometrically irreducible component of $\mathcal{F}$ is geometrically isomorphic to a tensor product*

$$\mathcal{L}_\psi \otimes \mathcal{L}_\chi$$

where ψ is a possibly trivial additive character and χ a possibly trivial multiplicative character. There exists an absolute constant $B \geqslant 0$ such that

$$\sum_{n \leqslant X} \Lambda(n) K(n) \ll \mathbf{c}(\mathcal{F})^B X \Big(1 + \frac{p}{X}\Big)^{1/12} p^{-1/48+\varepsilon},$$

$$\sum_{n \leqslant X} \mu(n) K(n) \ll \mathbf{c}(\mathcal{F})^B X \Big(1 + \frac{p}{X}\Big)^{1/12} p^{-1/48+\varepsilon}$$

for any $\varepsilon > 0$, where the implied constant depends only on $\varepsilon > 0$. Here Λ denotes the von Mangoldt function and μ the Möbius function.

We remark that the restriction on $\mathcal{F}$ is, with current techniques, necessary: an estimate of this quality for $K(n) = \chi(n)$ would imply a non-trivial zero-free strip in the critical strip for the Dirichlet L-function $L(s, \chi)$. This assumption holds however in many cases, for instance whenever $\mathcal{F}$ is geometrically irreducible with rank at least 2, or if $\mathcal{F}$ is ramified at some point $x \in \mathbf{P}^1(\bar{\mathbf{F}}_p) - \{0, \infty\}$, or if $\mathcal{F} = \mathcal{L}_{\chi(f)}$ with χ non-trivial and $f \in \mathbf{F}_p(X)$ not a monomial, or if $\mathcal{F} = \mathcal{L}_{\psi(f)}$ with ψ non-trivial and $f \in \mathbf{F}_p(X)$ not a polynomial of degree $\leqslant 1$.

The interest of this theorem is when X is close to p (for X much larger, one can use periodicity instead). Prior to [8], only the following cases had been studied in this range:

(1) When $K(n) = \chi(f(n))$, where f is a polynomial of degree $\leqslant 2$ which is not a monomial (Vinogradov, Karatsuba);

(2) When $K(n) = e(f(n)/p)$ for certain rational functions $f \in \mathbf{Q}(X)$ (Vinogradov, Fouvry–Michel [14], ...)

The new ingredient concerning trace functions in the proof of this theorem is the following general estimate:

Theorem 5.2 ([8]). *Let p be a prime and let $K = t_{\mathcal{F}}$ be the trace function of an ℓ-adic middle-extension sheaf $\mathcal{F}$ of weight 0 modulo p such that no geometrically irreducible component of $\mathcal{F}$ is geometrically isomorphic to a tensor product*

$$\mathcal{L}_\psi \otimes \mathcal{L}_\chi$$

where ψ is a possibly trivial additive character and χ a possibly trivial multiplicative character. Let $\alpha = (\alpha(m))$ and $\beta = (\beta(n))$ be sequences of complex numbers supported on $M/2 \leqslant m \leqslant M$ and $N/2 \leqslant n \leqslant N$ respectively for some M, $N \geqslant 1$. There exists an absolute constant $B \geqslant 0$ such that we have

$$\sum_m \sum_n \alpha(m)\beta(n)K(mn) \ll \mathbf{c}(\mathcal{F})^B \|\alpha\|\|\beta\|(MN)^{1/2}\Big(\frac{1}{p^{1/4}} + \frac{1}{M^{1/2}} + \frac{p^{1/4}(\log p)^{1/2}}{N^{1/2}}\Big),$$

where the implied constant is absolute.

The basic idea of the proof is classical in analytic number theory: one reduces quickly using the Cauchy-Schwarz inequality to proving that

$$\Big|\sum_{x \in \mathbf{F}_p} K(x)\overline{K(ax)}e\Big(\frac{bx}{p}\Big)\Big| \ll \mathbf{c}(\mathcal{F})^B p^{1/2}$$

for all $(a, b) \in \mathbf{F}_p^\times \times \mathbf{F}_p$, with *at most* $\mathbf{c}(\mathcal{F})^B$ exceptions, where B and the implied constant are absolute.

But the Plancherel formula gives

$$\sum_{x \in \mathbf{F}_p} K(x)\overline{K(ax)}e\Big(\frac{bx}{p}\Big) = \sum_{t \in \mathbf{F}_p} \mathrm{FT}(K)(t)\overline{\mathrm{FT}(K)(-at+b)} = \mathcal{C}(K; \begin{pmatrix} -a & b \\ 0 & 1 \end{pmatrix}) \tag{5.1}$$

so that the sums in this theorem are special cases of the correlation sums $\mathcal{C}(K; \gamma)$ of the previous section, for γ restricted to the subgroup B_p of upper-triangular matrices in $\mathrm{PGL}_2(\mathbf{F}_p)$ (interestingly, one of the properties of the decomposition (4.4) used in the proof of Theorem 4.3 is that $\gamma(\boldsymbol{\alpha})$ is *not* upper-triangular in that case!) We then only need to check that, under the assumptions of Theorem 5.2, the intersection of the group $\mathbf{G}_{\mathrm{FT}_\psi(\mathcal{F})}$ with B_p has size bounded by $\mathbf{c}(\mathcal{F})^B$ for some absolute constant B. This is done using some analysis of the group $\mathbf{G}_{\mathcal{F}}$ (in particular, the non-obvious fact that it is the group of $\mathbf{F}_p$-rational points of an algebraic subgroup of $\mathbf{G}_{\mathcal{F}}$.)

The general estimate of Theorem 5.2 has further applications. One which is dear to our heart is found in [9]: combining it with versions of the Voronoi summation formula and with other tools from [8], we improve significantly the exponent of distribution for the ternary divisor function d_3 in arithmetic progression. We recall that

$$d_3(n) = \sum_{abc=n} 1,$$

so that the Dirichlet generating series of d_3 is

$$\sum_{n\geqslant 1} d_3(n)n^{-s} = \zeta(s)^3$$

for $\mathrm{Re}(s) > 1$.

Theorem 5.3 ([9]). *Let p be a prime number, and let a be an integer coprime to p. Let $\varepsilon > 0$ be a positive number. For all X such that $p \leqslant X^{1/2+1/46-\varepsilon}$, we have*

$$\Big| \sum_{\substack{n\leqslant X\\ n\equiv a \,(\mathrm{mod}\, p)}} d_3(n) - \frac{1}{p-1}\sum_{n\leqslant X} d_3(n)\Big| \ll \frac{X}{p}\frac{1}{(\log X)^A}$$

for any $A \geqslant 1$, where the implied constant depends on ε and A.

The essential qualitative point is that the exponent $1/2 + 1/46$ is beyond $1/2$, which is the limit where a result like this would almost trivially follow from the Generalized Riemann Hypothesis for Dirichlet characters. Going beyond $1/2$ in this problem was first achieved by Friedlander and Iwaniec [15], whose result was improved by Heath-Brown [16]. Our own result, although slightly less general (in that we only consider prime moduli p instead of all $q \geqslant 1$), is another significant improvement, and most importantly in our mind, the proof is rather straightforward in principle when using the results of [8]. In fact, the only specific trace functions modulo p that we use in the proof are given by

$$K(n) = \mathrm{Kl}_3(an; p)$$

for some $a \in \mathbf{F}_p^\times$. In particular, we make use of the following estimate, which was already used (implicitly) by Friedlander-Iwaniec and Heath-Brown:

Theorem 5.4 (Correlation of hyper-Kloosterman sums). *Let p be a prime and $(a, b) \in \mathbf{F}_p^\times \times \mathbf{F}_p$. There exists a constant $C \geqslant 1$, independent of p and a, such that*

$$\Big| \sum_{x\in\mathbf{F}_p^\times} \mathrm{Kl}_3(x; p)\overline{\mathrm{Kl}_3(ax; p)}e\Big(\frac{bx}{p}\Big)\Big| \leqslant C\sqrt{p}$$

for $(a, b) \neq (1, 0)$.

We see from (5.1) that this can be derived from the existence of (and conductor bound for) Kloosterman sheaves and the fact that the group $\mathbf{G}_{\mathrm{FT}_\psi(\mathcal{K}\ell_3)} \subset \mathrm{PGL}_2(\mathbf{F}_p)$ is trivial.

Previously, this exponential sum was handled by Friedlander and Iwaniec (and by Heath-Brown) by writing

$$\sum_{x\in\mathbf{F}_p^\times} \mathrm{Kl}_3(x; p)\overline{\mathrm{Kl}_3(ax; p)}e\Big(\frac{bx}{p}\Big) = \sum_{t\neq 0,-b} \mathrm{Kl}_2\Big(\frac{1}{t}; p\Big)\,\mathrm{Kl}_2\Big(\frac{a}{t+b}; p\Big) - \frac{1}{p^2}$$

(by a simple computation), and using a bound for the sum on the right-hand side, which was itself proved by Bombieri and Birch in the Appendix to [15].

One may observe that the sum on the right-hand side also arises as a correlation sum. Indeed, for $K(n) = e(\bar{n}/p)$ as in Example 4.5 (1), if we take

$$\gamma = \begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix}$$

for $(a, b, c) \in \mathbf{F}_p \times \mathbf{F}_p^\times \times \mathbf{F}_p^\times$, then a simple change of variable in the definition leads to the identity

$$\mathcal{C}(K; \gamma) = \sum_{\substack{t \in \mathbf{F}_p \\ x \neq -b}} \mathrm{Kl}_2\Big(\frac{1}{t}; p\Big) \mathrm{Kl}_2\Big(\frac{a}{t+b}; p\Big).$$

Thus Theorem 5.4 follows also from the fact that the group $\mathbf{G}_{\mathrm{FT}_\psi(\mathcal{K}\ell_2)}$ is trivial, which is not very difficult to prove. Interestingly, other correlations sums attached to the same sheaf appeared in other papers in the literature: we are aware of its occurrence in works of Pitt [25] and Munshi [24].

References

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

[2] N. Bourbaki: *Fonctions d'une variable réelle*, Paris, Hermann, 1976.

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

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

[5] P. Deligne and Y.Z. Flicker: *Counting local systems with principal unipotent local monodromy*, Ann. of Math. (2) 178 (2013), no 3, 921–982.

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

[7] É. Fouvry, E. Kowalski, Ph. Michel: *Algebraic twists of modular forms and Hecke orbits*, preprint (2012), `arXiv:1207.0617`.

[8] É. Fouvry, E. Kowalski, Ph. Michel: *Algebraic trace weights over the primes*, to appear in Duke Math. Journal.

[9] É. Fouvry, E. Kowalski, Ph. Michel: *On the exponent of distribution of the ternary divisor function,* to appear in Mathematika; `arXiv:1304.3199`.

[10] É. Fouvry, E. Kowalski, Ph. Michel: *An inverse theorem for Gowers norms of trace functions over* $\mathbf{F}_p$, Math. Proc. Cambridge Phil. Soc. 155 (2013), 277–295.

[11] É. Fouvry, E. Kowalski, Ph. Michel: *Counting sheaves with spherical codes*, Math. Res. Letters 20 (2013), 305–323.

[12] É. Fouvry, E. Kowalski, Ph. Michel: *The sliding sum method for short exponential sums*, preprint (2013), `arXiv:1307.0135`.

[13] É. Fouvry, E. Kowalski, Ph. Michel: *On the conductor of cohomological transforms*, preprint (2013), `arXiv:1310.3603`.

[14] É. Fouvry, Ph. Michel: *Sur certaines sommes d'exponentielles sur les nombres premiers*, Ann. Sci. École Norm. Sup. (4) 31 (1998), 93–130.

[15] 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).

[16] D. R. Heath-Brown: *The divisor function $d_3(n)$ in arithmetic progressions*, Acta Arith. 47 (1986), no. 1, 29–56.

[17] J. Heinloth, B-C. Ngô and Z. Yun: *Kloosterman sheaves for reductive groups*, Ann. of Math. (2) 177 (2013), no 1, 241–310.

[18] H. Iwaniec and E. Kowalski: *Analytic number theory*, A.M.S. Coll. Publ. 53 (2004).

[19] N.M. Katz: *Gauss sums, Kloosterman sums and monodromy groups*, Annals of Math. Studies 116, Princeton Univ. Press (1988).

[20] N.M. Katz: *Exponential sums and differential equations*, Annals of Math. Studies 124, Princeton Univ. Press (1990).

[21] E. Kowalski: *Poincaré and analytic number theory*, in "The scientific legacy of Poincaré", edited by É. Charpentier, É. Ghys and A. Lesne, Hist. Math., 36, Amer. Math. Soc., 2010.

[22] G. Laumon: *Transformation de Fourier, constantes d'équations fonctionnelles et conjecture de Weil*, Publ. Math. IHÉS, 65 (1987), 131–210.
[23] Ph. Michel, *Analytic number theory and families of automorphic L-functions*, in *Automorphic forms and applications*, 181–295, IAS/Park City Math. Ser., 12, Amer. Math. Soc. 2007.
[24] R. Munshi: *Shifted convolution sums for $GL(3) \times GL(2)$*, preprint (2012), `arXiv:1202.1157`.
[25] N. Pitt: *On shifted convolutions of $\zeta(s)^3$ with automorphic L-functions*, Duke Math. J. 77 (1995), no. 2, 383–406.

Université Paris Sud, Laboratoire de Mathématique, Campus d'Orsay, 91405 Orsay Cedex, France
E-mail address: `etienne.fouvry@math.u-psud.fr`

ETH Zürich – D-MATH, Rämistrasse 101, CH-8092 Zürich, Switzerland
E-mail address: `kowalski@math.ethz.ch`

EPFL/SB/IMB/TAN, Station 8, CH-1015 Lausanne, Switzerland
E-mail address: `philippe.michel@epfl.ch`