A STUDY IN SUMS OF PRODUCTS

ÉTIENNE FOUVRY, EMMANUEL KOWALSKI, AND PHILIPPE MICHEL

ABSTRACT. We give a general version of cancellation in exponential sums that arise as sums of products of trace functions satisfying a suitable independence condition related to the Goursat-Kolchin-Ribet criterion, in a form that is easily applicable in analytic number theory.

1. INTRODUCTION

In many (perhaps surprisingly many) applications to number theory, exponential sums over finite fields of the type

$$(1.1) \qquad \sum_{x \in \mathbf{F}_p} K(\gamma_1 \cdot x) \cdots K(\gamma_k \cdot x) e\left(\frac{hx}{p}\right)$$

arise naturally, for some positive integer $k \geqslant 1$, where

- The function K is a "trace function" over $\mathbf{F}_p$, of weight 0, for instance

$$K(x) = e\left(\frac{f(x)}{p}\right)$$

for some fixed polynomial $f \in \mathbf{Z}[X]$, a Kloosterman sum

$$K(x) = -\frac{1}{\sqrt{p}} \sum_{y \in \mathbf{F}_p^\times} e\left(\frac{y^{-1} + xy}{p}\right),$$

or its generalization to hyper-Kloosterman sums

$$K(x) = \mathrm{Kl}_r(x;p) = \frac{(-1)^{r-1}}{p^{(r-1)/2}} \sum_{t_1 \cdots t_r = x} e\left(\frac{t_1 + \cdots + t_r}{p}\right)$$

for some $r \geqslant 2$;

- For $1 \leqslant i \leqslant k$, $\gamma_i \in \mathrm{PGL}_2(\mathbf{F}_p)$ acts on $\mathbf{F}_p$ by fractional linear transformation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot x = \frac{ax + b}{cx + d},$$

for instance $\gamma_i \cdot x = a_i x + b_i$ for some $a_i \in \mathbf{F}_p^\times$ and $b_i \in \mathbf{F}_p$ (and the sum (1.1) is restricted to those $x \in \mathbf{F}_p$ which are not poles of any of the γ_i);

- Finally, $h \in \mathbf{F}_p$.

The goal is usually to prove, except in special "diagonal" cases, an estimate of the type

$$\sum_{x \in \mathbf{F}_p} K(\gamma_1 \cdot x) \cdots K(\gamma_k \cdot x) e\left(\frac{hx}{p}\right) \ll \sqrt{p},$$

where the implied constant is independent of p and h, when K has suitably bounded "complexity".

Note that if $K(x)$ is a Kloosterman sum, or another similar normalized exponential sum in one variable, then opening the sums expresses (1.1) as a $(k+1)$-variable character sum, and (because of the normalization) the goal becomes to have square-root cancellation with respect to all variables.

We emphasize that we do not assume that the γ_i are distinct. Furthermore, such sums also arise with some factors $K(\gamma_i \cdot x)$ replaced with their conjugate $\overline{K(\gamma_i \cdot x)}$, or indeed with factors $K_i(x)$ which are not directly related. Such cases will be also handled in this paper.

As a sample of situations where such sums have arisen, we note:

- In many proofs of the Burgess estimate for short character sums, one has to deal with cases where $h = 0$ and $K_i(x) = \chi(x + a_i)$ or $\overline{\chi(x + a_i)}$ for some multiplicative character χ (see, e.g., [14, Cor. 11.24, Lem. 12.8]);
- Cases where $k = 2$ and γ_1, γ_2 are diagonal are found in the thesis of Ph. Michel and his subsequent papers, e.g. [21];
- For $k = 2$, $\gamma_1 = 1$, $h = 0$, we obtain the general "correlation sums" (for the Fourier transform of K) defined in [5]; these are crucial to our works [5, 6, 7];
- Special cases of this situation of correlation sums can be found (sometimes implicitly) in earlier works of Iwaniec [13], of Pitt [23] and of Munshi [22];
- The case $k = 2$, γ_1 and γ_2 diagonal, h arbitrary and K a Kloosterman sum in two variables (or a variant with K a Kloosterman sum in one variable and γ_1, γ_2 not upper-triangular) occurs in the work of Friedlander and Iwaniec [11], and it is also used in the work of Zhang [24] on gaps between primes;
- Cases where k is arbitrary, the γ_i are upper-triangular and distinct, and h may be non-zero appear in the work of Fouvry, Michel, Rivat and Sárközy [10, Lemma 2.1], indeed in a form involving different trace functions $K_i(\gamma_i \cdot x)$ related to symmetric powers of Kloosterman sums;
- The sums for k arbitrary and $h = 0$, with K a hyper-Kloosterman sum appear in the works of Fouvry, Ganguly, Kowalski and Michel [9] and Kowalski and Ricotta [19] (with γ_i diagonal);
- This last case, but with arbitrary h and the γ_i being translations also appears in the work of Irving [12], and (for very different reasons) in work of Kowalski and Sawin [20];
- Another instance, with $k = 4$, h arbitrary and γ_i upper-triangular, occurs in the work of Blomer and Milićević [1, §11].

The principles arising from algebraic geometry and algebraic group theory (in particular the so-called Goursat-Kolchin-Ribet criterion, as developed by Katz), together with the general form of the Riemann Hypothesis over finite fields of Deligne lead to square root cancellation in such sums in (also possibly surprisingly) many circumstances. However, this principle is not fully stated in a self-contained manner in any reference. Thus, this paper

is devoted to a review of these principles. We have aimed to give statements that can be quoted easily in applications similar to the ones above, possibly with some additional algebraic leg-work.

All estimates will be derived ultimately from the following form of the Riemann Hypothesis over finite fields (see Section 4 for a sketch). Given two constructible ℓ-adic sheaves $\mathcal{F}$ and $\mathcal{G}$ on $\mathbf{A}^1_{\mathbf{F}_p}$, lisse on some dense open set U, we will denote

$$\langle \mathcal{F}, \mathcal{G} \rangle = \dim H^2_c(\mathbf{A}^1 \times \bar{\mathbf{F}}_p, \mathcal{F} \otimes \mathrm{D}(\mathcal{G})) = \dim H^2_c(U \times \bar{\mathbf{F}}_p, \mathcal{F} \otimes \mathrm{D}(\mathcal{G})),$$

where we denote by $\mathrm{D}(\mathcal{G})$ the middle-extension dual of $\mathcal{G}$ (see the notation section for details). We have:

Proposition 1.1. *Let $k \geqslant 1$ and let $\mathbf{F} = (\mathcal{F}_i)_{1 \leqslant i \leqslant k}$ be any k-tuple of ℓ-adic middle-extension sheaves on $\mathbf{A}^1_{\mathbf{F}_p}$ such that the $\mathcal{F}_i$ are of weight 0, and let $\mathcal{G}$ be an ℓ-adic middle-extension sheaf of weight 0. Let K_i be the trace function of $\mathcal{F}_i$ and M that of $\mathcal{G}$. If*

(1.2)
$$\left\langle \bigotimes_i \mathcal{F}_i, \mathcal{G} \right\rangle = 0$$

then we have

$$\left| \sum_{x \in \mathbf{F}_p} K_1(x) \cdots K_k(x) \overline{M(x)} \right| \leqslant C \sqrt{p},$$

where $C \geqslant 0$ depends only on k and on the conductors[1] of $\mathcal{F}_i$ and of $\mathcal{G}$.

Thus, we will concentrate below on finding and explaining criteria that ensure that the vanishing property (1.2) holds, deriving bounds for the corresponding sums from this proposition. However, for convenience, we will state formally a number of special cases of the resulting estimates. Moreover, it can happen in some applications that the precise main term in diagonal situations is important[2], and we give some statements that contain such estimates.

We begin by defining a special class of trace functions for which we can give a general estimate for (1.1).

Definition 1.2 (Bountiful sheaves). We say that an ℓ-adic sheaf $\mathcal{F}$ on $\mathbf{A}^1_{\mathbf{F}_p}$ is *bountiful* provided the following conditions hold:
- The sheaf $\mathcal{F}$ is a middle extension, pointwise pure of weight 0, of rank $r \geqslant 2$;
- The geometric monodromy group of $\mathcal{F}$ is equal to either SL_r or Sp_r (we will say that $\mathcal{F}$ is of SL_r-type, or Sp_r-type, respectively);
- The projective automorphism group

(1.3)
$$\mathrm{Aut}_0(\mathcal{F}) = \{\gamma \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p) \mid \gamma^* \mathcal{F} \simeq \mathcal{F} \otimes \mathcal{L} \text{ for some rank } 1 \text{ sheaf } \mathcal{L}\}$$

    of $\mathcal{F}$ is trivial.

If $\mathcal{F}$ is of SL_r-type, we will also need to understand the set

$$\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F}) = \{\gamma \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p) \mid \gamma^* \mathcal{F} \simeq \mathrm{D}(\mathcal{F}) \otimes \mathcal{L} \text{ for some rank } 1 \text{ sheaf } \mathcal{L}\},$$

which we define for any middle-extension ℓ-adic sheaf $\mathcal{F}$.

---

[1] See the notation section for a remainder of the definition.
[2] See for instance [20].

This definition implies that $\mathrm{Aut}_0(\mathcal{F})$ acts on $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$ by left-multiplication: for elements $\gamma \in \mathrm{Aut}_0(\mathcal{F})$ and $\gamma_1 \in \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F}$, we have $\gamma_1 \gamma \in \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$. This action is simply transitive (if $\gamma_1,\, \gamma_2 \in \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$, we get $\gamma = \gamma_2 \gamma_1^{-1} \in \mathrm{Aut}_0(\mathcal{F})$ with $\gamma_2 = \gamma \gamma_1$). This means that $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$ is either empty or is a right coset $\xi \,\mathrm{Aut}_0(\mathcal{F})$ of $\mathrm{Aut}_0(\mathcal{F})$.

There is another extra property: if $\gamma \in \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$, the fact that $\mathrm{D}(\mathrm{D}(\mathcal{F})) \simeq \mathcal{F}$ implies that $\gamma^2 \in \mathrm{Aut}_0(\mathcal{F})$.

In particular,[3] for a sheaf with $\mathrm{Aut}_0(\mathcal{F}) = 1$ (e.g., a bountiful sheaf), there are only two possibilities: either $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$ is empty, or it contains a single element $\xi_\mathcal{F}$, and the latter is an involution: $\xi_\mathcal{F}^2 = 1$. If this second case holds, we say that $\xi_\mathcal{F}$ is the *special involution* of $\mathcal{F}$. (For instance, we will see that for hyper-Kloosterman sums $\mathcal{K}\ell_r$ with r odd, there is a special involution which is $x \mapsto -x$).

The diagonal cases, where there is no cancellation in (1.1), will be classified by means of the following combinatorial definitions:

Definition 1.3 (Normal tuples). Let p be a prime, $k \geqslant 1$ an integer, $\boldsymbol{\gamma}$ a k-tuple of $\mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ and $\boldsymbol{\sigma}$ a k-tuple of $\mathrm{Gal}(\mathbf{C}/\mathbf{R}) = \{1, c\}$, where c is complex conjugation.

(1) We say that $\boldsymbol{\gamma}$ is *normal* if there exists some $\gamma \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ such that

$$|\{1 \leqslant i \leqslant k \mid \gamma_i = \gamma\}|$$

is odd.

(2) If $r \geqslant 3$ is an integer, we say that $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is r-*normal* if there exists some $\gamma \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ such that

$$|\{1 \leqslant i \leqslant k \mid \gamma_i = \gamma\}| \geqslant 1$$

and

$$|\{1 \leqslant i \leqslant k \mid \gamma_i = \gamma \text{ and } \sigma_i = 1\}| - |\{1 \leqslant i \leqslant k \mid \gamma_i = \gamma \text{ and } \sigma_i \neq 1\}| \not\equiv 0 \,(\mathrm{mod}\, r).$$

(3) If $r \geqslant 3$ is an integer, and $\xi \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ is a given involution, we say that $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is r-*normal with respect to* ξ if there exists some $\gamma \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ such that

$$|\{1 \leqslant i \leqslant k \mid \gamma_i = \gamma\}| \geqslant 1$$

and

$$(1.4) \qquad \left(\sum_{\substack{1 \leqslant i \leqslant k \\ (\gamma_i, \sigma_i) = (\gamma, 1)}} 1 + \sum_{\substack{1 \leqslant i \leqslant k \\ (\gamma_i, \sigma_i) = (\xi\gamma, c)}} 1 \right) - \left(\sum_{\substack{1 \leqslant i \leqslant k \\ (\gamma_i, \sigma_i) = (\gamma, c)}} 1 + \sum_{\substack{1 \leqslant i \leqslant k \\ (\gamma_i, \sigma_i) = (\xi\gamma, 1)}} 1 \right) \not\equiv 0 \,(\mathrm{mod}\, r).$$

Example 1.4. (1) The basic example of a pair $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ which is not r-normal arises when k is even and it is of the form

$$((\gamma_1, \gamma_1, \ldots, \gamma_{k/2}, \gamma_{k/2}), (1, c, \ldots, 1, c))$$

since we then have

$$|\{1 \leqslant i \leqslant k \mid \gamma_i = \gamma \text{ and } \sigma_i = 1\}| = |\{1 \leqslant i \leqslant k \mid \gamma_i = \gamma \text{ and } \sigma_i = c\}|$$

for any $\gamma \in \{\gamma_1, \ldots, \gamma_{k/2}\}$.

(2) Let $\xi \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ be an involution. Some basic examples of pairs $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ which are not r-normal with respect to ξ are the following:

---

[3] See Lemma 3.1 for a more general statement, based on these properties, that limits the possible structure of $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$.

- If k is even, pairs

$$((\gamma_1, \xi\gamma_1, \ldots, \gamma_{k/2}, \xi\gamma_{k/2}), (1, 1, \ldots, 1, 1))$$

(for instance, if the γ_i are distinct, the left-hand side of (1.4) is then

$$(1 + 0) - (0 + 1) = 0$$

for each $\gamma \in \{\gamma_1, \ldots, \gamma_{k/2}\}$),
- For $r = 3$, $k = 7$, pairs

$$((\gamma, \xi\gamma, \xi\gamma, \gamma, \gamma, \xi\gamma, \gamma), (1, c, c, c, 1, 1, 1))$$

where the left-hand side of (1.4) for γ (resp. $\xi\gamma$) is

$$(3 + 2) - (1 + 1) = 3 \equiv 0 \,(\mathrm{mod}\, 3) \qquad (\text{resp. } (1 + 1) - (2 + 3) = -3).$$

After these definitions, we have first an abstract statement, from which estimates follow immediately from Proposition 1.1. In this statement, for a sheaf $\mathcal{F}$ and $\sigma \in \mathrm{Aut}(\mathbf{C}/\mathbf{R})$, we denote $\mathcal{F}^\sigma = \mathcal{F}$ if σ is the identity, and $\mathcal{F}^\sigma = \mathrm{D}(\mathcal{F})$ if $\sigma = c$ is complex conjugation.

Theorem 1.5 (Abstract sums of products). *Let p be a prime and let $\mathcal{F}$ be a bountiful ℓ-adic sheaf on $\mathbf{A}^1_{\mathbf{F}_p}$.*

(1) Assume that $\mathcal{F}$ is of Sp_r-type. For every $k \geqslant 1$, every k-tuple $\boldsymbol{\gamma}$ of elements in $\mathrm{PGL}_2(\bar{\mathbf{F}}_p)$, and every $h \in \mathbf{F}_p$, we have

$$\Big\langle \bigotimes_{1 \leqslant i \leqslant k} \gamma_i^* \mathcal{F}, \mathcal{L}_{\psi(hX)} \Big\rangle = 0$$

provided that either $\boldsymbol{\gamma}$ is normal or that $h \neq 0$.

(2) Assume that $\mathcal{F}$ is of SL_r-type. For every $k \geqslant 1$, for all k-tuples $\boldsymbol{\gamma}$ of elements of $\mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ and $\boldsymbol{\sigma}$ of elements of $\mathrm{Aut}(\mathbf{C}/\mathbf{R})$, and for all $h \in \mathbf{F}_p$, we have

$$\Big\langle \bigotimes_{1 \leqslant i \leqslant k} \gamma_i^* (\mathcal{F}^\sigma), \mathcal{L}_{\psi(hX)} \Big\rangle = 0$$

provided that either $h \neq 0$, or that $h = 0$ and either

- *$\mathcal{F}$ has no special involution, and $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is r-normal;*
- *$\mathcal{F}$ has a special involution ξ, $p > r$, and $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is r-normal with respect to ξ.*

To be concrete, we get:

Corollary 1.6 (Bountiful sums of products). *Let p be a prime and let K be the trace function modulo p of a bountiful sheaf $\mathcal{F}$ with conductor c. Then, for any $k \geqslant 1$, there exists a constant $C = C(k, c)$ depending only on c and k such that:*

(1) If $\mathcal{F}$ is self-dual, so that K is real-valued, then for any k-tuple $\boldsymbol{\gamma}$ of elements of $\mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ and for any $h \in \mathbf{F}_p$, provided that either $\boldsymbol{\gamma}$ is normal, or $h \neq 0$, we have

$$\Big| \sum_{x \in \mathbf{F}_p} K(\gamma_1 \cdot x) \cdots K(\gamma_k \cdot x) e\Big(\frac{hx}{p}\Big) \Big| \leqslant C\sqrt{p}.$$

(2) If $\mathcal{F}$ is of SL_r-type with $r \geqslant 3$, and $p > r$, then for k-tuples $\boldsymbol{\gamma}$ of elements of $\mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ and $\boldsymbol{\sigma}$ of $\mathrm{Aut}(\mathbf{C}/\mathbf{R})$, and for any $h \in \mathbf{F}_p$, provided either that $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is r-normal, or r-normal with respect to the special involution of $\mathcal{F}$, if it exists, or that $h \neq 0$, we have

$$\Big| \sum_{x \in \mathbf{F}_p} K(\gamma_1 \cdot x)^{\sigma_1} \cdots K(\gamma_k \cdot x)^{\sigma_k} e\Big(\frac{hx}{p}\Big) \Big| \leqslant C\sqrt{p}$$

5

where we put $K(\gamma_i \cdot x)^{\sigma_i} = \sigma_i(K(\gamma_i \cdot x))$.

This is intuitively best possible, because if $\mathcal{F}$ is self-dual and γ is not normal, so that the distinct elements γ_j in γ appear each with even multiplicity $2n_j$, we get for $h = 0$ the sum

$$\sum_{x \in \mathbf{F}_p} \prod_j K(\gamma_j \cdot x)^{2n_j}$$

in which there is no cancellation to be expected. The corresponding optimality holds for sheaves of SL_r-type, but this is less obvious.

It is sometimes important to determine even in this case what is the main term that may arise. Examples are given by the following statement:

Corollary 1.7. *Let p be a prime and let K be the trace function modulo p of a bountiful sheaf $\mathcal{F}$ with conductor c. Assume furthermore:*

- *That the arithmetic monodromy group of $\mathcal{F}$ is equal to the geometric monodromy group,*
- *If $\mathcal{F}$ is of SL_r-type and has a special involution ξ, that $\xi^* \mathcal{F} \simeq \mathrm{D}(\mathcal{F})$.*

Then, for any $k \geqslant 1$, there exists a constant C depending only on c and k such that:

(1) If $\mathcal{F}$ is of Sp_{2g}-type, then for any k-tuple $\boldsymbol{\gamma}$ of elements of $\mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ which is not normal and for any $h \in \mathbf{F}_p$, there exists an integer $m(\boldsymbol{\gamma}) \geqslant 1$ such that

$$\left| \sum_{x \in \mathbf{F}_p} K(\gamma_1 \cdot x) \cdots K(\gamma_k \cdot x) - m(\boldsymbol{\gamma})p \right| \leqslant C\sqrt{p}.$$

If k is even and $\boldsymbol{\gamma}$ consists of pairs of $k/2$ distinct elements, then $m(\boldsymbol{\gamma}) = 1$. In general,

$$m(\boldsymbol{\gamma}) = \prod_{\gamma \in \boldsymbol{\gamma}} A(n_\gamma)$$

where γ runs over all elements occuring in the tuple $\boldsymbol{\gamma}$, n_γ is the multiplicity of γ in the tuple and $A(n)$ is the multiplicity of the trivial representation of Sp_{2g} in the n-th tensor power of the standard representation of Sp_{2g}.

(2) If $\mathcal{F}$ is of SL_r-type with $r \geqslant 3$, then for k-tuples $\boldsymbol{\gamma}$ of elements of $\mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ and $\boldsymbol{\sigma}$) of $\mathrm{Aut}(\mathbf{C}/\mathbf{R})$, such that $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is not r-normal, or not r-normal with respect to the special involution of $\mathcal{F}$ if it exists, there exists an integer $m(\boldsymbol{\gamma}, \boldsymbol{\sigma}) \geqslant 1$ such that

$$\left| \sum_{x \in \mathbf{F}_p} K(\gamma_1 \cdot x)^{\sigma_1} \cdots K(\gamma_k \cdot x)^{\sigma_k} - m(\boldsymbol{\gamma}, \boldsymbol{\sigma})p \right| \leqslant C\sqrt{p}.$$

If k is even, $\boldsymbol{\gamma}$ consists of $k/2$ pairs of elements which are distinct or distinct modulo the special involution if it exists, and for each such pair (γ_i, γ_j), one of σ_i is the identity and the other is c, then $m(\boldsymbol{\gamma}, \boldsymbol{\sigma}) = 1$. Otherwise, $m(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is bounded in terms of k and r only.

The proofs of Theorem 1.5, Corollaries 1.6 and 1.7 will be found in Section 4, after we develop a more general framework in Section 2. Many examples of (trace functions of) bountiful sheaves, and also of the more general situation of the next section, together with more statements of the resulting estimates, are found in Section 3. Readers may wish to first read through this last section in order to see more examples of the estimates we obtain.

There is a certain inevitable tension in this paper between the fact that, on the one hand, we deal with rather general phenomena, and on the other hand most applications involve

6

extremely concrete special cases. We hope to write a fuller book-length account of trace functions over finite fields that will resolve this conflict by providing much more detailed explanations and examples, but in the meantime, the current text should provide precise references for many applications. Any reader in a state of doubt concerning sums of the type considered here is welcome to contact the authors. We also mention that the arXiv version of this text contains some more details and examples.

Notation and conventions. (1) An ℓ-adic sheaf over an algebraic variety X defined over $\mathbf{F}_p$ will always mean a constructible $\bar{\mathbf{Q}}_\ell$-sheaf for some $\ell \neq p$; whenever the trace function of such sheaves are mentioned, it is assumed that an isomorphism $\iota : \bar{\mathbf{Q}}_\ell \longrightarrow \mathbf{C}$ has been chosen once and for all, and that the trace function is seen as complex-valued through this isomorphism.

(2) For a lisse sheaf $\mathcal{F}$ (resp. a middle-extension sheaf $\mathcal{F}$ on $\mathbf{A}^1$) we denote by $\mathrm{D}(\mathcal{F})$ the dual lisse sheaf (resp. the middle-extension dual $j_*(\mathrm{D}(j^*\mathcal{F}))$ where $j : U \hookrightarrow \mathbf{A}^1$ is the open immersion of a dense open set where $\mathcal{F}$ is lisse). If ϱ is a finite-dimensional representation of a group G, we denote by $\mathrm{D}(\varrho)$ the contragredient representation.

(3) The conductor of an ℓ-adic middle-extension sheaf $\mathcal{F}$ on $\mathbf{A}^1_{\mathbf{F}_p}$ is defined as

$$\mathbf{c}(\mathcal{F}) = \mathrm{rank}(\mathcal{F}) + |\mathrm{Sing}(\mathcal{F})| + \sum_{x \in \mathrm{Sing}(\mathcal{F})} \mathrm{Swan}_x(\mathcal{F}),$$

where $\mathrm{Sing}(\mathcal{F})$ is the set of singularities of $\mathcal{F}$ in $\mathbf{P}^1(\bar{\mathbf{F}}_p)$ and $\mathrm{Swan}_x(\mathcal{F})$ the Swan conductor at x.

(4) We denote by $\mathrm{Z}(G)$ the center of a group G, and by G^0 the connected component of the identity in a topological or algebraic group G.

2. A GENERAL FRAMEWORK

We provide in this section, and the next, a very general statement concerning sheaves with trace functions of the type appearing in (1.1). This will be presented in a purely algebraic manner, and later sections will provide the diophantine interpretation that leads to the results of the first section.

We first make a definition that encapsulates some of the content of the Goursat-Kolchin-Ribet criterion of Katz (see[16, §1.8]):

Definition 2.1 (Generous tuple)**.** Let $k \geqslant 1$ be an integer and p a prime. Let $U \subset \mathbf{A}^1_{\mathbf{F}_p}$ be a dense open set. Let $\mathcal{F} = (\mathcal{F}_i)$ be a tuple of ℓ-adic middle-extension sheaves on $\mathbf{A}^1_{\mathbf{F}_p}$, all lisse on U. Denote by

$$\varrho_i : \pi_1(U \times \bar{\mathbf{F}}_p, \bar{\eta}) \longrightarrow \mathrm{GL}(V_i)$$

the ℓ-adic representations corresponding to $\mathcal{F}_i$, and

$$\varrho = \bigoplus_{1 \leqslant i \leqslant k} \varrho_i$$

We say that $\boldsymbol{\mathcal{F}}$ is U-*generous* if:

(1) The sheaves $\mathcal{F}_i$ are geometrically irreducible and pointwise pure of weight 0 on U;

7

(2) For all i, the normalizer of the connected component of the identity G_i^0 of the geometric monodromy group G_i of $\mathcal{F}_i$ is contained in $\mathbf{G}_m G_i^0 \subset \mathrm{GL}(V_i)$ and its Lie algebra is simple (in particular, G_i^0 acts irreducibly on V_i);

(3) For all $i \neq j$, the pairs (G_i^0, Std_i) and (G_j^0, Std_j) are Goursat-adapted in the sense of [16, p. 24], where Std_i denotes the tautological representations $G_i \subset \mathrm{GL}(V_i)$;

(4) Let G be the Zariski closure of the image of ϱ and let $\tilde{\varrho}_i : G \longrightarrow \mathrm{GL}(V_i)$ be the representation such that ϱ_i is the composition

$$\pi_1(U \times \bar{\mathbf{F}}_p, \bar{\eta}) \xrightarrow{\varrho} G \xrightarrow{\tilde{\varrho}_i} \mathrm{GL}(V_i) \ ;$$

then for all $i \neq j$, and all 1-dimensional characters χ of G, there is no isomorphism

(2.1) $$\tilde{\varrho}_i \simeq \tilde{\varrho}_j \otimes \chi, \ \text{or} \ \mathrm{D}(\tilde{\varrho}_i) \simeq \tilde{\varrho}_j \otimes \chi$$

as representations of G.

We say that $\boldsymbol{\mathcal{F}}$ is *strictly U-generous* if it is generous and the monodromy groups G_i are connected.

Remark 2.2. The last condition holds in particular if, for $i \neq j$, there is no rank 1 sheaf $\mathcal{L}$ such that

$$\mathcal{F}_i \simeq \mathcal{F}_j \otimes \mathcal{L}, \ \text{or} \ \mathrm{D}(\mathcal{F}_i) \simeq \mathcal{F}_j \otimes \mathcal{L},$$

and we will usually check it in this form.

Example 2.3. We just give quick examples here, leaving more detailed discussions to Section 3.

(1) Let $U = \mathbf{G}_m$. Given $n \geqslant 1$ even (resp. odd) and a k-tuple (a_i) of distinct elements of $\mathbf{F}_p^\times$ (resp. elements distinct modulo ± 1), we take $\mathcal{F}_i = [\times a_i]^* \mathcal{K}\ell_n$, where $\mathcal{K}\ell_n$ is the n-variable Kloosterman sheaf with trace function $\mathrm{Kl}_n(x; p)$ (see Section 3).

Then $(\mathcal{F}_i)$ is strictly U-generous. This follows from the theory of Kloosterman sheaves, in particular the computation of the geometric monodromy groups by Katz [15], and the fact that there does not exist a rank 1 sheaf $\mathcal{L}$ and a geometric isomorphism

$$[\times a]^* \mathcal{K}\ell_n \simeq \mathcal{K}\ell_n \otimes \mathcal{L} \ \text{or} \ [\times a]^* \mathcal{K}\ell_n \simeq \mathrm{D}(\mathcal{K}\ell_n) \otimes \mathcal{L},$$

for $a \neq 1$ if n is even, and for $a \notin \{\pm 1\}$ if n is odd. (In other words, we have $\mathrm{Aut}_0(\mathcal{K}\ell_r) = 1$, and for $r \geqslant 3$ odd, $\mathrm{Aut}_0^d(\mathcal{K}\ell_r)$ contains the unique special involution $x \mapsto -x$; see Section 3 for details).

(2) Given $\mathcal{F}_0$ self-dual and lisse on $\mathbf{G}_m$, with geometric monodromy group equal to Sp_r, such that the projective automorphism group of $\mathcal{F}_0$ is trivial, and a k-tuple (a_i) of distinct elements of $\mathbf{F}_p^\times$, we may take $\mathcal{F}_i = [\times a_i]^* \mathcal{F}_0$ on $U = \mathbf{G}_m$, and $(\mathcal{F}_i)$ is then strictly $\mathbf{G}_m$-generous.

(3) Given $\mathcal{F}_0$ lisse on $\mathbf{G}_m$ with geometric monodromy *containing* SL_r for some $r \geqslant 3$, such that

$$\mathrm{Aut}_0(\mathcal{F}_0) \cap \mathbf{T} = 1,$$

where $\mathbf{T} \subset \mathrm{PGL}_2$ is the diagonal torus, and a k-tuple $\boldsymbol{a} = (a_i)$ of elements of $\mathbf{F}_p^\times$ distinct modulo ± 1, then the tuple $([\times a_i]^* \mathcal{F}_0)$ is $\mathbf{G}_m$-generous.

Indeed, all conditions of the definition are clearly met, except maybe for the non-existence of isomorphisms

$$\mathrm{D}(\varrho_i) \simeq \varrho_j \otimes \chi$$

for $i \neq j$. But such an isomorphism would imply that there exists a rank 1 lisse sheaf $\mathcal{L}$ on $\mathbf{G}_m$ and a geometric isomorphism

$$[\times a_i]^* \mathrm{D}(\mathcal{F}_0) \simeq [\times a_j]^* \mathcal{F}_0 \otimes \mathcal{L},$$

and this implies that

$$\begin{pmatrix} (a_i a_j^{-1})^2 & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{Aut}_0(\mathcal{F}_0) = 1,$$

so that $a_i = \pm a_j$, contradicting our assumption on the tuple $\boldsymbol{a}$.

(4) Given a U-generous tuple (resp strictly U-generous tuple), any subtuple is still U-generous (resp. strictly U-generous). Similarly, if $V \subset U$ is another dense open set, the restrictions to V of a U-generous tuple is V-generous (and similarly for strictly generous tuples).

We now come back to the development of the general theory. The crucial point is the following lemma:

Lemma 2.4 (Katz). *Let $\mathcal{F}$ be U-generous. Then the connected component of the identity of the geometric monodromy group G of the sheaf*

$$\bigoplus_i \mathcal{F}_i$$

on U is equal to the product

$$G^0 = \prod_{1 \leqslant i \leqslant k} G_i^0$$

of the connected components of the geometric monodromy groups G_i of $\mathcal{F}_i$. If $\mathcal{F}$ is strictly generous, then $G = G^0$.

Let $\pi : V \times \bar{\mathbf{F}}_p \to U \times \bar{\mathbf{F}}_p$ be the finite abelian étale covering corresponding to the surjective homomorphism

$$\pi_1(U \times \bar{\mathbf{F}}_p, \bar{\eta}) \longrightarrow G/G^0,$$

so that $V = U$ and π is the identity on $U \times \bar{\mathbf{F}}_p$ if $\mathcal{F}$ is strictly U-generous. Then the geometric monodromy group of

$$\pi^* \left(\bigoplus_i \mathcal{F}_i \right)$$

is equal to G^0. Furthermore, the restriction to G^0 of any irreducible representation of G is irreducible.

Proof. In view of the definition, the computation of the monodromy groups is a special case of the Goursat-Kolchin-Ribet Proposition of Katz [16, Prop. 1.8.2] (noting that, with the notation there, if the normalizer of G_i^0 is contained in $\mathbf{G}_m G_i^0$, then G_i^0 acts irreducibly on V_i, because any subrepresentation is stable under the action of $\mathbf{G}_m G_i^0 \supset N_{\mathrm{GL}(V_i)} G_i^0 \supset G_i$).

For the last part, let τ be an irreducible representation of G. Note that

$$G \subset \prod_i (\mathbf{G}_m G_i^0) \subset \mathrm{Z}(G) G^0$$

by the second condition in the definition of a generous tuple, and the fact that any $g \in G$ is of the form

$$g = (\xi_i g_i)$$

9

for some $\xi_i \in \mathbf{G}_m \cap G_i \subset Z(G_i)$ and $g_i \in G_i^0$, so that $g = zh$ with $z = (\xi_i) \in Z(G)$ and $h = (g_i) \in G^0$. It follows that for any $g = zh \in G$, we have

$$\tau(g) = \tau(zh) = \tau(z)\tau(h).$$

Since $\tau(z)$ is a scalar (because τ is G-irreducible and z is central), we see that any G^0-invariant subspace is also G-invariant. $\qquad\square$

Remark 2.5. This result would not extend if we allow G_i not contained in $\mathbf{G}_m G_i^0$: for instance, if $G = O_{2r}$, so that $G^0 = SO_{2r}$, there exist irreducible representations of G which split in two irreducible subrepresentations when restricted to G^0.

We then state a preliminary result, which for convenience we express in the language of Tannakian categories. For a U-generous tuple $\boldsymbol{\mathcal{F}}$, we denote by $\mathcal{T}(\boldsymbol{\mathcal{F}})$ the Tannakian category of sheaves on $U \times \bar{\mathbf{F}}_p$ generated by the sheaves $\mathcal{F}_i$.

Proposition 2.6. *Let $\boldsymbol{\mathcal{F}}$ be U-generous, and let $\pi : V \times \bar{\mathbf{F}}_p \to U \times \bar{\mathbf{F}}_p$ be the finite abelian étale covering corresponding to the surjective homomorphism*

$$\pi_1(U \times \bar{\mathbf{F}}_p, \bar{\eta}) \longrightarrow G/G^0.$$

(1) The the category $\mathcal{T}(\boldsymbol{\mathcal{F}})$ is equivalent as a Tannakian category to the category of representations of the linear algebraic group G, a functor from the latter to $\mathcal{T}(\boldsymbol{\mathcal{F}})$ giving this equivalence is

$$\Lambda \mapsto \Lambda \circ \varrho_{\boldsymbol{\mathcal{F}}}$$

where $\varrho_{\boldsymbol{\mathcal{F}}}$ is the representation of $\pi_1(U \times \bar{\mathbf{F}}_p, \bar{\eta})$ corresponding to the lisse sheaf

$$\bigoplus_i \mathcal{F}_i.$$

Furthermore the restriction to G^0 of a representation of G corresponds to the functor π^*.
(2) If $\mathcal{G}$ is an irreducible object of $\mathcal{T}(\boldsymbol{\mathcal{F}})$, then we have a geometric isomorphism

$$\pi^*\mathcal{G} \simeq \bigotimes_i \Lambda_i(\pi^*\mathcal{F}_i)$$

where Λ_i is an irreducible representation of G_i^0 for each i. Two such sheaves have isomorphic restriction to $V \times \bar{\mathbf{F}}_p$ if and only if the respective Λ_i are the same.

Proof. The first part is a standard fact. To deduce (2), we simply note that from the last part of Lemma 2.4, the pullback $\pi^*\mathcal{G}$ is geometrically irreducible if $\mathcal{G}$ is geometrically irreducible. We then obtain the stated formula from the classification of irreducible representations of a direct product. $\qquad\square$

We now present a first classification theorem that is well-suited to cases where all sheaves involved are self-dual.

Theorem 2.7 (Diagonal classification)**.** *Let $\boldsymbol{\mathcal{F}}$ be U-generous and let $\pi : V \times \bar{\mathbf{F}}_p \to U \times \bar{\mathbf{F}}_p$ be the finite abelian étale covering corresponding to the surjective homomorphism*

$$\pi_1(U \times \bar{\mathbf{F}}_p, \bar{\eta}) \longrightarrow G/G^0.$$

Let $\mathcal{G}$ be an ℓ-adic sheaf which is geometrically irreducible and lisse on U. Let

$$\boldsymbol{n} = (n_1, \ldots, n_k)$$

10

be a k-tuple of positive *integers. Denote*

$$\mathcal{F}_{\boldsymbol{n}} = \bigotimes_{1 \leqslant i \leqslant k} \mathcal{F}_i^{\otimes n_i}.$$

We have

$$\langle \mathcal{F}_{\boldsymbol{n}}, \mathcal{G} \rangle \neq 0$$

only if there exists a geometric isomorphism

(2.2)
$$\pi^* \mathcal{G} \simeq \bigotimes_i \Lambda_i(\pi^* \mathcal{F}_i)$$

on $V \times \bar{\mathbf{F}}_p$, where, for all i, Λ_i is an irreducible representation of the group G_i^0 which is also a subrepresentation of the representation $\mathrm{Std}_i^{\otimes n_i}$ of G_i^0, with Std_i denoting the natural faithful representation of G_i^0 corresponding to $\pi^* \mathcal{F}_i$.

In fact, for $\mathcal{G}$ given as above, we have

$$\langle \mathcal{F}_{\boldsymbol{n}}, \mathcal{G} \rangle \leqslant \prod_{1 \leqslant i \leqslant k} \mathrm{mult}_{\Lambda_i}(\mathrm{Std}_i^{\otimes n_i}),$$

where $\mathrm{mult}_{\Lambda_i}(\mathrm{Std}_i^{\otimes n_i})$ denotes the multiplicity of Λ_i in $\mathrm{Std}_i^{\otimes n_i}$.

If $\boldsymbol{\mathcal{F}}$ is strictly U-generous, then equality holds in this formula, and in particular the left-hand side is non-zero if and only if $\mathcal{G}$ is of the form $\bigotimes_i \Lambda_i(\mathcal{F}_i)$ with Λ_i as above.

In general, if $\mathcal{G}$ is of the form (2.2), then there exists a character χ of G/G^0 such that

$$\langle \mathcal{F}_{\boldsymbol{n}}, \mathcal{G} \otimes \chi \rangle \neq 0.$$

If all n_i are equal to 1, we denote $\mathcal{F}_{(1,\ldots,1)} = \mathcal{F}$. Then

$$\langle \mathcal{F}, \mathcal{G} \rangle = 0$$

unless $\mathcal{G} \simeq \mathcal{F}$, and

$$\langle \mathcal{F}, \mathcal{G} \rangle = 1$$

in that case.

The crucial point in the proof is the following very simple fact:

Lemma 2.8. *With the notation of the theorem, assume that*

$$\langle \mathcal{F}_{\boldsymbol{n}}, \mathcal{G} \rangle \neq 0.$$

Then $\mathcal{G}$ is geometrically isomorphic to an object of $\mathcal{T}(\boldsymbol{\mathcal{F}})$.

Proof. The co-invariant formula for lisse sheaves states that

$$\langle \mathcal{F}_{\boldsymbol{n}}, \mathcal{G} \rangle = \dim(\mathcal{F}_{\boldsymbol{n}, \bar{\eta}} \otimes \mathrm{D}(\mathcal{G}_{\bar{\eta}}))_G.$$

The irreducibility of $\mathcal{G}$, and the semi-simplicity of the representations involved, shows that if this dimension is non-zero, then $\mathcal{G}$ is geometrically isomorphic to a subsheaf of $\mathcal{F}_{\boldsymbol{n}}$. But clearly this sheaf is itself an object of $\mathcal{T}(\boldsymbol{\mathcal{F}})$, hence the result by transitivity. $\qquad \square$

Proof of the theorem. By the lemma, $\mathcal{G}$ is geometrically isomorphic to an object of $\mathcal{T}(\boldsymbol{\mathcal{F}})$. Since it is also geometrically irreducible, Lemma 2.4 shows that $\pi^* \mathcal{G}$ is also geometrically irreducible. Thus, by the proposition, it follows that

$$\pi^* \mathcal{G} \simeq \bigotimes_{1 \leqslant i \leqslant k} \Lambda_i(\pi^* \mathcal{F}_i),$$

where the Λ_i are some irreducible representations of the group G_i^0. We have then

$$\langle \mathcal{F}_{\boldsymbol{n}}, \mathcal{G} \rangle \leqslant \langle \pi^* \mathcal{F}_{\boldsymbol{n}}, \pi^* \mathcal{G} \rangle = \dim(\mathcal{F}_{\boldsymbol{n}, \bar\eta} \otimes \mathrm{D}(\mathcal{G}_{\bar\eta}))^{G^0},$$

where we can use invariants instead of coinvariants because the representations are semisimple. But the G^0-invariants of the generic fibre of

$$\pi^* \mathcal{F}_{\boldsymbol{n}} \otimes \mathrm{D}(\pi^* \mathcal{G}) = \bigotimes_{1 \leqslant i \leqslant k} \left(\pi^* \mathcal{F}_i^{\otimes n_i} \otimes \mathrm{D}(\Lambda_i(\pi^* \mathcal{F}_i)) \right)$$

are isomorphic (under the equivalence of the proposition) to the invariants of G^0 on

$$\boxtimes_{1 \leqslant i \leqslant k} \left(\mathrm{Std}_i^{\otimes n_i} \otimes \mathrm{D}(\Lambda_i) \right)$$

hence to the tensor product over i of the G^0-invariants of

$$\mathrm{Std}_i^{\otimes n_i} \otimes \mathrm{D}(\Lambda_i).$$

Thus we get the inequality for the dimension, and in particular the G^0-invariant space is non-zero if and only if Λ_i is a subrepresentation of $\mathrm{Std}_i^{\otimes n_i}$ for all $1 \leqslant i \leqslant k$, and this gives a necessary condition for the G-invariant space to be non-zero.

In the opposite direction, if $\mathcal{G}$ is given by (2.2) with Λ_i an irreducible subrepresentation of $\mathrm{Std}_i^{\otimes n_i}$, then we have

$$(\mathcal{F}_{\boldsymbol{n}, \bar\eta} \otimes \mathrm{D}(\mathcal{G}_{\bar\eta}))^{G^0} \neq 0.$$

This invariant space is naturally a representation of G/G^0; since it is non-zero, it contains at least one character χ; one then checks easily that

$$(\mathcal{F}_{\boldsymbol{n}, \bar\eta} \otimes \mathrm{D}(\mathcal{G}_{\bar\eta} \otimes \chi))^{G} \neq 0.$$

Finally, if $n_i = 1$ and $\langle \mathcal{F}, \mathcal{G} \rangle \neq 0$, then since $\mathcal{F}$ is irreducible in this case (e.g. because its restriction to G^0 is irreducible as $\boxtimes_i \mathrm{Std}_i$), Schur's Lemma gives the result. $\qquad\square$

We state separately a more general version of Theorem 2.7 which is useful when some sheaves are not self-dual.

Theorem 2.9 (Diagonal classification, 2). *Let $\boldsymbol{\mathcal{F}}$ be U-generous and let $\pi : V \times \bar{\mathbf{F}}_p \to U \times \bar{\mathbf{F}}_p$ be the finite abelian étale covering corresponding to the surjective homomorphism*

$$\pi_1(U \times \bar{\mathbf{F}}_p, \bar\eta) \longrightarrow G/G^0.$$

Let $\mathcal{G}$ be an ℓ-adic sheaf which is geometrically irreducible and lisse on U. Let

$$\boldsymbol{m} = (m_1, \ldots, m_k), \qquad \boldsymbol{n} = (n_1, \ldots, n_k)$$

be k-tuples of integers such that $n_i + m_i \geqslant 1$ for all i. Denote

$$\mathcal{F}_{\boldsymbol{m}, \boldsymbol{n}} = \bigotimes_{1 \leqslant i \leqslant k} \left(\mathcal{F}_i^{\otimes m_i} \otimes \mathrm{D}(\mathcal{F}_i)^{\otimes n_i} \right).$$

We have

$$\langle \mathcal{F}_{\boldsymbol{m}, \boldsymbol{n}}, \mathcal{G} \rangle \neq 0$$

only if there exists a geometric isomorphism

$$(2.3) \qquad\qquad \pi^* \mathcal{G} \simeq \bigotimes_i \Lambda_i(\pi^* \mathcal{F}_i)$$

12

on $V \times \bar{\mathbf{F}}_p$, where, for all i, Λ_i is an irreducible representation of the group G_i^0 which is also a subrepresentation of the representation $\mathrm{Std}_i^{\otimes m_i} \otimes \mathrm{D}(\mathrm{Std}_i)^{\otimes n_i}$ of G_i^0, with Std_i denoting the natural faithful representation of G_i^0 corresponding to $\pi^* \mathcal{F}_i$.

In fact, for $\mathcal{G}$ given as above, we have

$$\langle \mathcal{F}_{\boldsymbol{m},\boldsymbol{n}}, \mathcal{G} \rangle \leqslant \prod_{1 \leqslant i \leqslant k} \mathrm{mult}_{\Lambda_i}(\mathrm{Std}_i^{\otimes m_i} \otimes \mathrm{D}(\mathrm{Std}_i)^{\otimes n_i}),$$

where $\mathrm{mult}_{\Lambda_i}(\mathrm{Std}_i^{\otimes m_i} \otimes \mathrm{D}(\mathrm{Std}_i)^{\otimes n_i})$ denotes the multiplicity of Λ_i in $\mathrm{Std}_i^{\otimes m_i} \otimes \mathrm{D}(\mathrm{Std}_i)^{\otimes n_i}$. If $\boldsymbol{\mathcal{F}}$ is strictly U-generous, then there is equality, and the converse also holds.

In general, if $\mathcal{G}$ is given by (2.3), then there exists a character χ of G/G^0 such that

$$\langle \mathcal{F}_{\boldsymbol{m},\boldsymbol{n}}, \mathcal{G} \otimes \chi \rangle \neq 0.$$

Clearly, the case $\boldsymbol{n} = (0, \ldots, 0)$ recovers Theorem 2.7.

Proof. This is the same as that of Theorem 2.7, mutatis mutandis. $\qquad\square$

Here is a simple corollary that can be very helpful:

Corollary 2.10. *Let $\boldsymbol{\mathcal{F}} = (\mathcal{F}_i)_{1 \leqslant i \leqslant k}$ be U-generous. Let $\mathcal{G}$ be an ℓ-adic sheaf. Let $\boldsymbol{\sigma}$ be a k-tuple of elements of $\mathrm{Aut}(\mathbf{C}/\mathbf{R})$. If*

$$\mathrm{rank}\,\mathcal{G} < \prod_i \mathrm{rank}\,\mathcal{F}_i,$$

then we have

$$\Big\langle \bigotimes_{1 \leqslant i \leqslant k} \mathcal{F}_i^{\sigma_i}, \mathcal{G} \Big\rangle = 0.$$

Proof. Note that this corresponds to the previous situation, with $\boldsymbol{m}$ and $\boldsymbol{n}$ such that $m_i + n_i = 1$ for all i.

By considering a geometrically irreducible subsheaf of $\mathcal{G}$, we may assume that it is geometrically irreducible (since a subsheaf still satisfies the dimension bound and $\langle \cdot, \cdot \rangle$ is "bilinear" with respect to direct sums). By the previous arguments, if

$$\Big\langle \bigotimes_{1 \leqslant i \leqslant k} \mathcal{F}_i^{\sigma_i}, \mathcal{G} \Big\rangle \neq 0$$

then we would have

$$\pi^* \mathcal{G} \simeq \bigotimes_i \Lambda_i(\pi^* \mathcal{F}_i),$$

where Λ_i is irreducible and occurs in Std_i. But this implies that $\Lambda_i \simeq \mathrm{Std}_i$, and in particular that $\mathrm{rank}\,\mathcal{G} = \prod_i \mathrm{rank}\,\mathcal{F}_i$. $\qquad\square$

3. Examples

We collect here examples of trace functions for which the results stated in the introduction or in the previous section apply, and state some of the resulting bounds for convenience. These examples are taken for the most part from the many results of Katz, who has computed the monodromy groups of many classes of sheaves over $\mathbf{A}^1$ using a variety of techniques.

3.1. **General construction.** Quite generally, let $(\mathcal{F}_i)_{i\in I}$ be any finite tuple of middle-extension sheaves of weight 0 on $\mathbf{A}^1_{\mathbf{F}_p}$ such that the geometric monodromy groups G_i of the restriction of $\mathcal{F}_i$ to a dense open set U_i where it is lisse, is such that G_i^0 is any of the groups

$$\mathrm{SL}_r, \text{ for } r \geqslant 3, \qquad \mathrm{SO}_{2r+1}, \text{ for } r \geqslant 1,$$
$$\mathrm{Sp}_r, \text{ for } r \text{ even} \geqslant 2,$$
$$\mathbf{F}_4, \quad \mathbf{E}_7, \quad \mathbf{E}_8, \quad \mathbf{G}_2.$$

Then we can always extract a convenient generous subtuple as follows: let U be the intersection of the U_i, and let $J \subset I$ be any set of representatives of I for the equivalence relation defined by $i \sim j$ if and only if

$$\mathcal{F}_i \simeq \mathcal{F}_j \otimes \mathcal{L}, \text{ or } \mathrm{D}(\mathcal{F}_i) \simeq \mathcal{F}_j \otimes \mathcal{L}$$

on U for some rank 1 sheaf $\mathcal{L}$ lisse on U. Then $\boldsymbol{\mathcal{F}} = (\mathcal{F}_i)_{i\in J}$ is U-generous.

Indeed, condition (1) is clear, and (2) holds by the restrictions on G_i^0 (see also [18, 9.3.6] for the normalizer condition, and note that in the exceptional cases indicated, all automorphisms of the groups are inner, which implies the normalizer condition). Also, by [16, Examples 1.8.1], the representations corresponding to $i \neq j$ in J are always Goursat-adapted, and finally the restriction to the representatives of the equivalence relation ensures the last condition.

Note that for any multiplicities n_i, $m_i \geqslant 0$ for $i \in I$, we have then geometric isomorphisms

$$\bigotimes_{i\in I} \mathcal{F}_i^{\otimes n_i} \otimes \bigotimes_{i\in I} \mathrm{D}(\mathcal{F}_i)^{\otimes m_i} \simeq \mathcal{L} \bigotimes_{i\in J} \mathcal{F}_i^{\otimes n_i'} \otimes \bigotimes_{i\in J} \mathrm{D}(\mathcal{F}_i)^{\otimes m_i'}$$

for some rank 1 sheaf $\mathcal{L}$ (depending on (n_i, m_i)) and

$$n_i' = \sum_{j\sim i} n_j, \qquad m_i' = \sum_{j\sim i} m_j,$$

and it is therefore possible to use many of the results for the generous tuple $\boldsymbol{\mathcal{F}}$ to derive corresponding statements that apply to the original one. For an example of applying this principle, see the discussion of the Bombieri–Bourgain sums in Section 5.

In applications of this strategy, especially in the SL_r case, the following lemma will be useful:

Lemma 3.1. *Let $\mathcal{F}$ be an ℓ-adic sheaf modulo p. Then $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$ is either empty or is of the form $\xi\,\mathrm{Aut}_0(\mathcal{F})$ for some $\xi \in N(\mathrm{Aut}_0(\mathcal{F}))$ such that $\xi^2 \in \mathrm{Aut}_0(\mathcal{F})$.*

For $\mathrm{Aut}_0(\mathcal{F}) = 1$, we recover the fact that $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$ is either empty or contains only an involution; if $\mathrm{Aut}_0(\mathcal{F})$ is equal to its normalizer, e.g., if it is a maximal and non-normal subgroup, then it shows that $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$ is either empty or equal to $\mathrm{Aut}_0(\mathcal{F})$, which means that $1 \in \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F})$, or in other words that

$$\mathcal{F} \simeq \mathrm{D}(\mathcal{F}) \otimes \mathcal{L}$$

for some rank 1 sheaf $\mathcal{L}$. This means that, in some sense, $\mathcal{F}$ is "almost" self-dual.

Proof. More generally, consider a subgroup H of a group G, and a coset $T \subset G$ of the form $T = \xi H$ that satisfies $g^2 \in G$ for all $g \in T$ (as is the case of $T = \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F}) \subset G = \mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ for the subgroup $H = \mathrm{Aut}_0(\mathcal{F})$).

We claim first that this situation occurs if and only if $T = \xi H$ for some $\xi \in G$ such that $\xi H \xi = H$.

Indeed, $(\xi g)(\xi g) \in H$ for all $g \in H$ is equivalent to $\xi g \xi \in H$ for all $g \in H$, i.e., to $\xi H \xi \subset H$. But then the converse inclusion $\xi H \xi \supset H$ also holds by taking the inverse:

$$\xi^{-1} H \xi^{-1} = (\xi H \xi)^{-1} \subset H^{-1} = H.$$

Now from $\xi H \xi = H$, we get first in particular $\xi^2 \in H$, and then

$$H = \xi H \xi = \xi (H \xi^2) \xi^{-1} = \xi H \xi^{-1}$$

implies that $\xi \in N(H)$. This gives the result in our case, and we may also note that the converse holds, namely if $\xi \in N(H)$ satisfies $\xi^2 \in H$, then

$$\xi H \xi = \xi H \xi^2 \xi^{-1} = \xi H \xi^{-1} = H.$$

$\square$

Looking at the list of simple groups at the beginning of this section, it is clear that the only significant omission is that of $G_i^0 = \mathrm{SO}_{2r}$ for $r \geqslant 2$; in that case, it is indeed not true that the normalizer O_{2r} is contained in $\mathbf{G}_m G_i^0$ (see also Remark 2.5 (2) below). This complication may be problematic in some applications, since geometric monodromy groups O_{2r} do occur naturally (e.g., for certain hypergeometric sheaves and for elliptic curves over function fields, see Section 3). However, we have not (yet) encountered such cases in analytic number theory, and one can expect that some analogues of our statements could be proved using the classification of representations of O_{2r} and their restrictions to SO_{2r}.

3.2. **Even rank Kloosterman sums.** For $r \geqslant 2$ even, the normalized Kloosterman sums

$$\mathrm{Kl}_r(x; p) = -\frac{1}{p^{(r-1)/2}} \sum_{t_1 \cdots t_r = x} e\left(\frac{t_1 + \cdots + t_r}{p}\right)$$

are the trace functions of a self-dual bountiful sheaf $\mathcal{K}\ell_r$ on $\mathbf{A}^1_{\mathbf{F}_p}$. Indeed, the geometric monodromy group is then Sp_r by [15, Th. 11.1], and the projective automorphism group is trivial by Proposition 3.6 below. In addition, one knows that the arithmetic monodromy group of $\mathcal{K}\ell_r$ is equal to its geometric monodromy group, so that Corollary 1.7 applies to this sheaf. The conductor is given by $\mathbf{c}(\mathcal{K}\ell_r) = r + 3$ (the rank is r, there are two singularities, one with zero Swan conductor, the other with Swan conductor 1).

Hence, from Corollary 1.6, we get:

Corollary 3.2. *Let $r \geqslant 2$ be an even integer. Let $k \geqslant 1$ be an integer. There exists a constant $C \geqslant 1$, depending only on k and r such that for any prime p, any $h \in \mathbf{F}_p$ and any $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k) \in \mathrm{PGL}_2(\mathbf{F}_p)$ and $h \in \mathbf{F}_p$, such that either*

- *we have $h \neq 0$, or;*
- *some component of $\boldsymbol{\gamma}$ occurs with odd multiplicity, i.e., $\boldsymbol{\gamma}$ is normal, as in Definition 1.3.*

Then we have

$$\left| \sum_{x \in \mathbf{F}_p} \mathrm{Kl}_r(\gamma_1 \cdot x; p) \cdots \mathrm{Kl}_r(\gamma_k \cdot x; p) e\left(\frac{hx}{p}\right) \right| \leqslant C p^{1/2}$$

where the sum runs over x such that all $\gamma_i \cdot x$ are defined.

3.3. Odd rank Kloosterman sums.

For $r \geqslant 2$ odd, the normalized Kloosterman sums

$$\mathrm{Kl}_r(x;p) = \frac{1}{p^{(r-1)/2}} \sum_{t_1 \cdots t_r = x} e\Big(\frac{t_1 + \cdots + t_r}{p}\Big)$$

are the trace functions of a non-self-dual bountiful sheaf $\mathcal{K}\ell_r$ on $\mathbf{A}^1_{\mathbf{F}_p}$ of SL_r type, with conductor uniformly bounded over p, with special involution $x \mapsto -x$. Indeed, the geometric monodromy group is SL_r by [15, Th. 11.1], and the projective automorphism group is trivial by Proposition 3.6 below, and we also have a geometric isomorphism

$$\mathrm{D}(\mathcal{K}\ell_r) \simeq [\times(-1)]^* \mathcal{K}\ell_r.$$

In addition, one knows that the arithmetic monodromy group of $\mathcal{K}\ell_r$ is equal to its geometric monodromy group, and hence Corollary 1.7 also applies to this sheaf of SL_r-type. The conductor is $r + 3$ as for even-rank Kloosterman sums.

Hence, from Corollary 1.6, we get:

Corollary 3.3. *Let $r \geqslant 2$ be an odd integer. Let $k \geqslant 1$ be an integer. There exists a constant $C \geqslant 1$, depending only on k and r such that for any prime p, any $h \in \mathbf{F}_p$ and any $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k) \in \mathrm{PGL}_2(\mathbf{F}_p)^k$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k) \in \mathrm{Aut}(\mathbf{C}/\mathbf{R})^k$, such that either*

- *we have $h \neq 0$, or;*
- *the pair $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is r-normal with respect to $x \mapsto -x$.*

Then we have

$$\Big| \sum_{x \in \mathbf{F}_p} \mathrm{Kl}_r(\gamma_1 \cdot x; p)^{\sigma_1} \cdots \mathrm{Kl}_r(\gamma_k \cdot x; p)^{\sigma_k} e\Big(\frac{hx}{p}\Big) \Big| \leqslant C p^{1/2}$$

where the sum runs over x such that all $\gamma_i \cdot x$ are defined.

Concretely, recall (see (1.4) and the examples following) that to say that the pair $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is r-normal with respect to $x \mapsto -x$ means that for *some* component γ of $\boldsymbol{\gamma}$, we have

$$r \nmid (a_1 + a_2) - (b_1 + b_2),$$

where:

- a_1 is the number of i with $\gamma = \gamma_i$ and $\sigma_i = 1$
- a_2 is the number of i with $\gamma = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \gamma_i$ and $\sigma_i = c$
- b_1 is the number of i with $\gamma = \gamma_i$ and $\sigma_i = c$
- b_2 is the number of i with $\gamma = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \gamma_i$ and $\sigma_i = 1$.

3.4. Hypergeometric sums.

Hyper-Kloosterman sums have been generalized by Katz [16, Ch. 8] to hypergeometric sums, which are analogues of general hypergeometric functions. Some give rise to bountiful sheaves, and many to generous tuples. We recall the definition: given a prime number p, integers m, $n \geqslant 1$, with $m + n \geqslant 1$, and tuples $\boldsymbol{\chi} = (\chi_i)_{1 \leqslant i \leqslant n}$ and $\boldsymbol{\varrho} = (\varrho_j)_{1 \leqslant j \leqslant m}$ of multiplicative characters of $\mathbf{F}_p^\times$, the hypergeometric sum $\mathrm{Hyp}(\boldsymbol{\chi}, \boldsymbol{\varrho}, t; p)$ is defined (see [16, 8.2.7]) for $t \in \mathbf{F}_p$ by

$$\mathrm{Hyp}(\boldsymbol{\chi}, \boldsymbol{\varrho}, t; p) = \frac{(-1)^{n+m-1}}{p^{(n+m-1)/2}} \sum_{N(\boldsymbol{x}) = tN(\boldsymbol{y})} \prod_i \chi_i(x_i) \overline{\prod_j \varrho_j(y_j)} e\Big(\frac{T(\boldsymbol{x}) - T(\boldsymbol{y})}{p}\Big)$$

where

$$N(\boldsymbol{x}) = x_1 \cdots x_n, \qquad N(\boldsymbol{y}) = y_1 \cdots y_m,$$
$$T(\boldsymbol{x}) = x_1 + \cdots + x_n, \qquad T(\boldsymbol{y}) = y_1 + \cdots + y_m$$

so that the sum is over all $(n+m)$-tuples $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbf{F}_p{}^{n+m}$ such that

$$x_1 \cdots x_n = t y_1 \cdots y_m.$$

If $n = r$, $m = 0$, and $\chi_i = 1$ for all i, then we recover the Kloosterman sums $\mathrm{Kl}_r(t; p)$. If $n = 2$, $m = 0$, and $\chi_2 = 1$ but χ_1 is non-trivial, we obtain Salié-type sums. This indicates that such sums should arise naturally in formulas like the Voronoi summation formula for automorphic forms with non-trivial nebentypus.

Katz shows (see [16, Th. 8.4.2]) that if no character χ_i coincides with a character ϱ_j (in which case one says that $\boldsymbol{\chi}$ and $\boldsymbol{\varrho}$ are *disjoint*), then for any $\ell \neq p$, there exists an irreducible ℓ-adic middle-extension sheaf $\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})$ on $\mathbf{A}^1_{\mathbf{F}_p}$, of weight 0, with trace function given by $\mathrm{Hyp}(\boldsymbol{\chi}, \boldsymbol{\varrho}, t; p)$. This sheaf is lisse on $\mathbf{G}_m$, except if $m = n$, in which case it is lisse on $\mathbf{G}_m - \{1\}$. It has rank $\max(m, n)$. Moreover, these results of Katz show that the conductor of $\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is bounded in terms of m and n only.

The basic results of Katz concerning the geometric monodromy group G of the hypergeometric sheaf $\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})$ depend on the following definitions of exceptional tuples of characters (see [16, Cor. 8.9.2, 8.10.1]):

Definition 3.4. Let k be a finite field and let $\boldsymbol{\chi}$ and $\boldsymbol{\varrho}$ be an n-tuple and an m-tuple of characters of $k^\times$.

(1) For $d \geqslant 1$, the pair $(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is d-Kummer-induced if $d \mid (n, m)$ and if there exist n/d and m/d-tuples $\boldsymbol{\chi}^*$ and $\boldsymbol{\varrho}^*$ such that $\boldsymbol{\chi}$ consists of all characters χ such that χ^d is a component of $\boldsymbol{\chi}^*$, and $\boldsymbol{\varrho}$ consists of all characters ϱ such that ϱ^d is a component of $\boldsymbol{\varrho}^*$.

(2) Assume $n = m$. For integers a, $b \geqslant 1$ such that $a + b = n$, the pair $(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is (a, b)-Belyi-induced if there exist characters α and β with $\beta \neq 1$ such that $\boldsymbol{\chi}$ consists of all characters χ such that either $\chi^a = \alpha$ or $\chi^b = \beta$, and if $\boldsymbol{\varrho}$ consists of all characters ϱ such that $\varrho^n = \alpha\beta$.

(3) Assume $n = m$. For integers a, $b \geqslant 1$ such that $a + b = n$, the pair $(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is (a, b)-inverse-Belyi-induced if and only if $(\overline{\boldsymbol{\varrho}}, \overline{\boldsymbol{\chi}})$ is (a, b)-Belyi-induced.

We say that $(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is Kummer-induced (resp. Belyi-induced, inverse-Belyi-induced) if there exists some $d \geqslant 2$ (resp. some a, $b \geqslant 1$) such that the pair is d-Kummer-induced (resp. (a, b)-Belyi-induced, (a, b)-inverse-Belyi-induced).

We then have the following:

- If $n = m$, let Λ denote the multiplicative character

$$\Lambda = \prod_i \chi_i \overline{\varrho_i}.$$

Assume that $(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is neither Kummer-induced, Belyi-induced, nor inverse-Belyi-induced. Then G^0 is either trivial, SL_n, SO_n or Sp_n; if $\Lambda = 1$, it is either SL_n or Sp_n; if $\Lambda \neq 1$ but $\Lambda^2 = 1$, then G^0 is either 1 or SO_n or SL_n, and if $\Lambda^2 \neq 1$, then G^0 is either 1 or SL_n (see [16, Th. 8.11.2]). The problem of determining which case occurs is discussed by Katz; most intricate is the criterion for G^0 to be trivial (see [16, §8.14–8.17]), which is however applicable in practice.

17

- If $n \neq m$, let $r = \max(n, m)$ be the rank of the sheaf. Assume that $(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is not Kummer induced. Then, provided $p > 2\max(n, m) + 1$, and p does not divide an explicit positive integer, we have: $G^0 = \mathrm{SL}_r$ if $n - m$ is odd (and $G \neq G^0$ if $|n - m| = 1$); $G^0 = \mathrm{SL}_r$, SO_r or Sp_r if $n - m$ is even and either $r \notin \{7, 8, 9\}$ or $|n-m| \neq 6$ (see [16, Th. 8.11.3]). Here also, more precise criteria for which G^0 arises exist, as well as a classification of the few exceptional possibilities when $|n - m| = 6$ and $r \in \{6, 7, 8\}$.

Example 3.5. If $\boldsymbol{\varrho}$ is the empty tuple, $n \geqslant 2$ and $\boldsymbol{\chi}$ is an n-tuple where all components are trivial, then it follows immediately from the definition that $(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is not Kummer-induced. Thus the last result recovers, for p large enough in terms of n, the fact that the geometric monodromy group of $\mathcal{K}\ell_n$ contains SL_n if n is odd, and contains either SO_n or Sp_n if n is even.

In order to apply the results of the previous section, it is of course very useful to have some information concerning the projective automorphism groups of hypergeometric sheaves. Many cases are contained in the following result:

Proposition 3.6. (1) *Let* $\boldsymbol{\chi}_1$, $\boldsymbol{\varrho}_1$ *and* $\boldsymbol{\chi}_2$, $\boldsymbol{\varrho}_2$ *be any* n_1-*tuple (resp.* m_1-*tuple,* n_2-*tuple,* m_2-*tuple) with* $\boldsymbol{\chi}_1$ *disjoint from* $\boldsymbol{\varrho}_1$ *and* $\boldsymbol{\chi}_2$ *disjoint from* $\boldsymbol{\varrho}_2$, *and with* $m_1 + n_1 \geqslant 1$, $m_2 + n_2 \geqslant 1$. *Let* $a \in \mathbf{F}_p^\times$. *Then we have a geometric isomorphism*

$$(3.1) \qquad\qquad [\times a]^* \mathcal{H}yp(\boldsymbol{\chi}_1, \boldsymbol{\varrho}_1) \simeq \mathcal{H}yp(\boldsymbol{\chi}_2, \boldsymbol{\varrho}_2),$$

if and only if $a = 1$ *and* $\boldsymbol{\chi}_1 \sim \boldsymbol{\chi}_2$ *and* $\boldsymbol{\varrho}_1 \sim \boldsymbol{\varrho}_2$.
(2) *Let* $m \neq n$ *with* $m + n \geqslant 1$ *be integers with* $\max(m, n) \geqslant 2$ *and* $(m, n) \neq (1, 2)$, $(m, n) \neq (2, 1)$. *Let* $\boldsymbol{\chi}$ *and* $\boldsymbol{\varrho}$ *be disjoint tuples of characters of* $\mathbf{F}_p^\times$. *The projective automorphism group* $\mathrm{Aut}_0(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}))$ *is then trivial.*
(3) *With notation as in* (2), *the set* $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}))$ *is non-empty if and only if the integer* $n - m$ *is odd, and the tuples* $\boldsymbol{\chi}$ *and* $\boldsymbol{\varrho}$ *are both invariant under inversion. In this case, the special involution is* $x \mapsto -x$, *i.e., we have*

$$[\times(-1)]^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathrm{D}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})).$$

(4) *If* $n = m \geqslant 2$, *then for any disjoint* n-*tuples* $(\boldsymbol{\chi}, \boldsymbol{\varrho})$, *the group* $\mathrm{Aut}_0(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}))$ *is a subgroup of the finite group*

$$\Gamma = \left\{ 1, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \right\} \subset \mathrm{PGL}_2(\bar{\mathbf{F}}_p).$$

(5) *With notation as in* (4), *the set* $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}))$ *is either empty or is a subset of* Γ, *which is of the form* $T = \xi H$ *for some subgroup* $H \subset \Gamma$ *and some* $\xi \in N_\Gamma(H)$ *such that* $\xi^2 \in H$.

Proof. (1) For all $a \in \bar{\mathbf{F}}_p^\times$, the components of $\boldsymbol{\chi}$ (resp. $\boldsymbol{\varrho}$) can be recovered from the sheaf $[\times a]^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})$ as the tame characters occuring in the representation of the inertia group at 0 (resp. at ∞) corresponding to this sheaf, and the multiplicity appears as the size of the associated Jordan block (see [16, Th. 8.4.2 (6), (7), (8)]). Thus (3.1) is only possible if $\boldsymbol{\chi}_1 \sim \boldsymbol{\chi}_2$ and $\boldsymbol{\varrho}_1 \sim \boldsymbol{\varrho}_2$.
We assume this is the case now, i.e., that

$$[\times a]^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}).$$

We then obtain $a = 1$ from [16, Lemma 8.5.4] and the fact that the Euler-Poincaré characteristic of a hypergeometric sheaf is -1.

(2) We may assume that $n > m$, using inversion otherwise. Assume that $\gamma \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p)$ is such that
$$\gamma^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \otimes \mathcal{L}$$
for a rank 1 sheaf $\mathcal{L}$. By comparing ramification behavior we see that γ must be diagonal (if $\gamma^{-1}(0) \neq 0$, then $\mathcal{L}$ must be tamely ramified at 0 to have the tensor product tamely ramified at $\gamma^{-1}(0)$, as $\gamma^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})$ is; but then the inertia invariants at $\gamma^{-1}(0)$ are zero for the tensor product, a contradiction to [16, Th. 8.4.2 (6)], and the case of $\gamma^{-1}(\infty) \neq \infty$ gives a similar contradiction).

Thus $\gamma \in \mathrm{Aut}_0(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}))$ implies a geometric isomorphism
$$[\times a]^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \otimes \mathcal{L}$$
on some dense open set $j : U \hookrightarrow \mathbf{G}_m$. By [16, Lemma 8.11.7.1], under the current assumption $(n, m) \neq (2, 1)$, this implies that $\mathcal{L} \simeq \mathcal{L}_\Lambda$ for some multiplicative character Λ.

But then we have
$$\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \otimes \mathcal{L}_\Lambda \simeq \mathcal{H}yp(\Lambda\boldsymbol{\chi}, \Lambda\boldsymbol{\varrho})$$
by [16, 8.3.3] where $\Lambda\boldsymbol{\chi} = (\Lambda\chi_i)_i$ and $\Lambda\boldsymbol{\varrho} = (\Lambda\varrho_j)_j$. We are therefore reduced to a geometric isomorphism
$$[\times a]^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathcal{H}yp(\Lambda\boldsymbol{\chi}, \Lambda\boldsymbol{\varrho}),$$
and by (1), it follows that $a = 1$, i.e., $\gamma = 1$.

(3) As in the previous case, we see that any element $\gamma \in \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}))$ must be diagonal, so that $\gamma \cdot x = ax$ for some $a \in \mathbf{F}_p^\times$. Since γ, if it exists, is an involution, we obtain $a^2 = 1$, and therefore the only possibility for the special involution is $x \mapsto -x$.

We now assume that
$$[\times(-1)]^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathrm{D}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})) \otimes \mathcal{L}$$
for some rank 1 sheaf $\mathcal{L}$.

Again by [16, Lemma 8.11.7.1], the sheaf $\mathcal{L}$ is a Kummer sheaf $\mathcal{L}_\Lambda$. We have
$$\mathrm{D}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})) \otimes \mathcal{L} \simeq \mathcal{H}yp_{\bar{\psi}}(\overline{\boldsymbol{\chi}}, \overline{\boldsymbol{\varrho}}) \otimes \mathcal{L}_\Lambda \otimes [\times(-1)^{n-m}]^* \mathcal{H}yp(\Lambda\overline{\boldsymbol{\chi}}, \Lambda\overline{\boldsymbol{\varrho}})$$
by combining [16, 8.3.3] and [16, Lemma 8.7.2] (using also the fact that Kummer sheaves are geometrically multiplication invariant). Thus the assumption means that
$$[\times(-1)]^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq [\times(-1)^{n-m}]^* \mathcal{H}yp(\Lambda\overline{\boldsymbol{\chi}}, \Lambda\overline{\boldsymbol{\varrho}}).$$

If $n - m$ is even, this can not happen by (1); if $n - m$ is odd, on the other hand, this happens if and only if $\Lambda\overline{\boldsymbol{\chi}} \sim \boldsymbol{\chi}$ and $\Lambda\overline{\boldsymbol{\varrho}} \sim \boldsymbol{\varrho}$, as claimed.

(4) Let $n = m \geqslant 2$ and $\gamma \in \mathrm{Aut}_0(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}))$ so that
$$\gamma^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \otimes \mathcal{L}$$
for some rank 1 sheaf $\mathcal{L}$. The right-hand side is ramified at $\{0, 1, \infty\}$ (because $n \geqslant 2$ and the description of local monodromy from [16, Th. 8.4.2 (8)] shows that the ramification of the hypergeometric sheaf cannot be eliminated by tensoring with a character), and hence γ must permute the points 0, 1, ∞. This shows that $\gamma \in \Gamma$.

(5) Arguing as in (4) with an isomorphism
$$\gamma^* \mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho}) \simeq \mathrm{D}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})) \otimes \mathcal{L}$$

19

we see that $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{H}yp(\boldsymbol{\chi}, \boldsymbol{\varrho})) \subset \Gamma$. Then the statement is just the conclusion of Lemma 3.1 in this special case. $\qquad\square$

In view of these results, one can feel confident that sums of products of hypergeometric sums can be handled using the results of this paper, at least in many cases. The trickiest case would be when $G^0 = O_r$ with r even (in view of Remark 2.5), which does occur (e.g., if $n - m \geqslant 2$ is even, n is even, the tuples $\boldsymbol{\chi}$ and $\boldsymbol{\varrho}$ are stable under inversion, and $\prod \chi_i$ is non-trivial of order 2, see [16, Th. 8.8.1, Lemma 8.11.6]).

3.5. Fourier transforms of multiplicative characters.
Many examples of sheaves with suitable monodromy groups are discussed in [16, 7.6–7.14], arising from Fourier transforms of other (rather simple) sheaves. We discuss one illustrative case, encouraging the reader to look at Katz's results if he or she encounters similar-looking constructions.

We consider a polynomial $g \in \mathbf{F}_p[X]$ and a non-trivial multiplicative character χ modulo p. We assume that no root of g is of order divisible by the order of χ. We then form the sheaf

$$\mathcal{F}_{\chi,g} = \mathrm{FT}_\psi(\mathcal{L}_{\chi(g)})$$

i.e., the Fourier transform of the Kummer sheaf with trace function $\chi(g(x))$, where ψ is the additive character $e(\cdot/p)$. The trace function of $\mathcal{F}_{\chi,g}$ is

$$K_{\chi,g}(x) = -\frac{1}{\sqrt{p}} \sum_{y \in \mathbf{F}_p} \chi(g(y))\psi(xy).$$

Proposition 3.7. *With notation as above, let r be the number of distinct roots of g in $\bar{\mathbf{F}}_p$. Assume that $r \geqslant 2$ and $p > 2r + 1$. Assume furthermore that the only solutions of the equations*

$$(3.2) \qquad\qquad\qquad\qquad x_1 - x_2 = x_3 - x_4$$

where $(x_1, \ldots, x_4)$ range over the roots of g in $\bar{\mathbf{F}}_p$ are given by $x_3 = x_1$, $x_4 = x_2$ or $x_2 = x_1$ and $x_3 = x_4$. Then $\mathcal{F}_{\chi,g}$ is a middle-extension sheaf of weight 0, of rank r, lisse on $\mathbf{G}_m$, and with geometric monodromy group containing SL_r. Furthermore, we have

$$\mathrm{Aut}_0(\mathcal{F}_{g,\chi}) \simeq \{a \in \bar{\mathbf{F}}_p \mid g(aX) = cg(X - \alpha) \text{ for some } c \in \bar{\mathbf{F}}_p^\times, \ \alpha \in \mathbf{F}_p\},$$

and $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F}_{\chi,g}) = \varnothing$ if $r \geqslant 3$. The conductor of $\mathcal{F}_{g,\chi}$ is bounded in terms of $\deg(g)$ only.

We leave this proposition without proof (see the `arXiv` version of this paper for details); all the results follow from the work of Katz [16, 7.9.1–7.9.3] on such trace functions. Concerning the last statement, note that if $r = 2$, the sheaf is of Sp_2-type (since $\mathrm{Sp}_2 = \mathrm{SL}_2$) to that $\mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F}_{\chi,g})$ is not relevant in that case.

4. SUMS OF PRODUCTS WITH FRACTIONAL LINEAR TRANSFORMATIONS

We can now quickly prove the results stated in Section 1 using the framework established previously.

Proof of Theorem 1.5. We begin with the easier Sp_r-type case. Let $\boldsymbol{\gamma}^*$ be the tuple of distinct elements of $\boldsymbol{\gamma}$, and n_γ the multiplicity of any such element in $\boldsymbol{\gamma}$. Let U be the common open set in $\mathbf{A}^1$ where all $\gamma \in \boldsymbol{\gamma}^*$ are defined. Arguing as in Example 2.3 (2), we see that the tuple $\boldsymbol{\mathcal{F}} = (\gamma^*\mathcal{F})_{\gamma \in \boldsymbol{\gamma}^*}$ is strictly U-generous, simply because $\mathcal{F}$ is bountiful of Sp_r-type.

By Theorems 2.7 and 2.9, we see that if

$$\Big\langle \bigotimes_{1 \leqslant i \leqslant k} \gamma_i^* \mathcal{F}, \mathcal{L}_{\psi(-hX)} \Big\rangle \neq 0,$$

there must exist some geometric isomorphism

$$\mathcal{L}_{\psi(hX)} \simeq \bigotimes_{\gamma \in \gamma^*} \Lambda_\gamma (\gamma^* \mathcal{F})$$

where Λ_γ are irreducible representations of the geometric monodromy group $G = \mathrm{Sp}_r$ of $\mathcal{F}$ such that Λ_γ is a subrepresentation of $\mathrm{Std}^{\otimes n_\gamma}$. Just for dimension reasons, each Λ_γ must be a one-dimensional character. But Definition 1.2 implies in particular that G has no non-trivial character, so that $\Lambda_\gamma = 1$, which implies that $\mathcal{L}_{\psi(hX)}$ must be geometrically trivial, i.e., that $h = 0$.

This already proves the first part of Theorem 1.5 when $h \neq 0$. Now assume $h = 0$. Then the condition that the trivial representation be a subrepresentation of $\mathrm{Std}^{\otimes n_\gamma}$ holds if and only if n_γ is even, and thus

$$\Big\langle \bigotimes_{1 \leqslant i \leqslant k} \gamma_i^* \mathcal{F}, \mathcal{L}_{\psi(-hX)} \Big\rangle \neq 0$$

if and only if all multiplicities n_γ are even, which means if and only if $\boldsymbol{\gamma}$ is *not* normal.

We now come to the SL_r-type case. If $\mathcal{F}$ has a special involution ξ, let $\mathcal{L}$ be a rank 1 sheaf such that

(4.1) $$\xi^* \mathcal{F} \simeq \mathrm{D}(\mathcal{F}) \otimes \mathcal{L},$$

and we note that (as a character of the fundamental group of $U \times \bar{\mathbf{F}}_p$) the sheaf $\mathcal{L}$ has order dividing r (by taking the determinant on both sides). For convenience, we let $\xi = 1$ and $\mathcal{L} = \bar{\mathbf{Q}}_\ell$, if there is no special involution.

Let $\boldsymbol{\gamma}^*$ be a tuple of representatives of the elements of $\boldsymbol{\gamma}$ for the equivalence relation

$$\gamma_i \sim \gamma_j \text{ if and only if } (\gamma_i = \gamma_j \text{ or } \gamma_i = \xi \gamma_j)$$

(which is indeed an equivalence relation because $\xi^2 = 1$).

Then, arguing as in Example 2.3 (3), we see that the tuple $\boldsymbol{\mathcal{F}} = (\gamma^* \mathcal{F})_{\gamma \in \gamma^*}$ is strictly U-generous, because $\mathcal{F}$ is bountiful of SL_r-type and because

$$\gamma_i^* \mathcal{F} \simeq \mathrm{D}(\gamma_j^* \mathcal{F}) \otimes \mathcal{L}',$$

for some rank 1 sheaf $\mathcal{L}'$, implies that

$$\gamma_i \gamma_j^{-1} \in \mathrm{Aut}_0^{\mathrm{d}}(\mathcal{F}),$$

and thus either does not occur (if $\mathcal{F}$ has no special involution) or happens only if $\gamma_i = \xi \gamma_j$, so that $\gamma_i \sim \gamma_j$, which is excluded for distinct components of $\boldsymbol{\gamma}^*$.

For $\gamma \in \boldsymbol{\gamma}^*$, we denote

$$n_\gamma^1 = |\{i \mid \gamma_i = \gamma \text{ and } \sigma_i = 1\}| + |\{i \mid \gamma_i = \xi \gamma \text{ and } \sigma_i = c\}|,$$
$$n_\gamma^c = |\{i \mid \gamma_i = \gamma \text{ and } \sigma_i = c\}| + |\{i \mid \gamma_i = \xi \gamma \text{ and } \sigma_i = 1\}|,$$

so that, by bringing together equivalent γ_i's, we obtain a geometric isomorphism

(4.2) $$\bigotimes_{1 \leqslant i \leqslant k} \gamma_i^* (\mathcal{F}^{\sigma_i}) \simeq \bigotimes_{\gamma \in \gamma^*} (\gamma^* \mathcal{F})^{\otimes n_\gamma^1} \otimes \mathrm{D}(\gamma^* \mathcal{F})^{\otimes n_\gamma^c} \otimes \mathcal{L}_0$$

21

for some rank 1 sheaf $\mathcal{L}_0$, which is a tensor product of sheaves of the form $\gamma^*\mathcal{L}$ or $\gamma^*(\mathrm{D}\,\mathcal{L})$. In particular, $\mathcal{L}_0$ has order dividing r since $\mathcal{L}$ does.

We now get from Theorem 2.9 that if

$$\Big\langle \bigotimes_{1\leqslant i\leqslant k} \gamma_i^*(\mathcal{F}^\sigma), \mathcal{L}_{\psi(hX)}\Big\rangle = \Big\langle \bigotimes_{\gamma\in\boldsymbol{\gamma}^*} (\gamma^*\mathcal{F})^{\otimes n_\gamma^1} \otimes \mathrm{D}(\gamma^*\mathcal{F})^{\otimes n_\gamma^c}, (\mathcal{L}_0\otimes\mathcal{L}_{\psi(hX)}\Big\rangle \neq 0,$$

then

$$\mathcal{L}_0\otimes\mathcal{L}_{\psi(hX)} \simeq \bigotimes_{\gamma\in\boldsymbol{\gamma}^*} \Lambda_\gamma(\gamma^*\mathcal{F})$$

where Λ_γ is an irreducible representation of SL_r which is a subrepresentation of the tensor product $\mathrm{Std}^{\otimes n_\gamma^1}\otimes\mathrm{D}(\mathrm{Std})^{\otimes n_\gamma^c}$. Since SL_r has no non-trivial 1-dimensional characters, this shows that this condition cannot occur unless Λ_γ is trivial for all γ, which implies then that

(4.3) $$\mathcal{L}_0\otimes\mathcal{L}_{\psi(hX)} \simeq \bar{\mathbf{Q}}_\ell$$

is trivial.

If $\mathcal{F}$ has no special involution, this immediately implies that $h = 0$. If $\mathcal{F}$ has a special involution, on the other hand, we recall that $\mathcal{L}_0$ has order r, while $\mathcal{L}_{\psi(hX)}$ has order p if $h \neq 0$. Hence (4.3) is impossible if $p > r$ and $h \neq 0$, and moreover, in that case we also get from (4.3) that $\mathcal{L}_0$ must be trivial.

Thus, in all cases of Theorem 1.5, we reduce to understanding the case $h = 0$. Since Λ_γ is trivial, we have also the condition that the trivial representation is a subrepresentation of the tensor product

$$(\gamma^*\mathcal{F})^{\otimes n_\gamma^1} \otimes \mathrm{D}(\gamma^*\mathcal{F})^{\otimes n_\gamma^c}$$

for all γ in $\boldsymbol{\gamma}^*$.

But the trivial representation of SL_r is a subrepresentation of $\mathrm{Std}^{\otimes n}\otimes\mathrm{D}(\mathrm{Std})^{\otimes m}$ if and only if $r \mid n - m$ (see, e.g., [19, Proof of Prop. 4.4]), and this means that if there is a main term, then $r \mid n_\gamma^1 - n_\gamma^c$ for all $\gamma \in \boldsymbol{\gamma}^*$, which means precisely that $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is not r-normal (if there is no special involution) or not r-normal with respect to ξ (if there is one). $\qquad\square$

Remark 4.1. We see from the proof that the condition $p > r$ in Theorem 1.5 (when $\mathcal{F}$ has a special involution) can be relaxed: especially, it is not needed if we have

$$\xi^*\mathcal{F} \simeq \mathrm{D}(\mathcal{F})$$

(i.e. if $\mathcal{L}$ in (4.1) can be taken to be the trivial sheaf, since we only used $p > r$ to deduce that $\mathcal{L}_0$ in (4.2) is trivial, which is automatically true in this case).

For completeness, we explain the proof of Proposition 1.1 (see, e.g., [5, §8,§9] for similar arguments).

Proof of Proposition 1.1. Let $U \subset \mathbf{A}^1$ be the maximal open set where all sheaves $\mathcal{F}_i$ and $\mathcal{G}$ are lisse. We have

$$|(\mathbf{A}^1 - U)(\mathbf{F}_p)| \leqslant \sum_i \mathbf{c}(\mathcal{F}_i) + \mathbf{c}(\mathcal{G}).$$

Since the sheaves are all mixed of weights $\leqslant 0$, we have

$$\Big| \sum_{x\in U(\mathbf{F}_p)} K_1(x)\cdots K_k(x)\overline{M(x)} - \sum_{x\in\mathbf{F}_p} K_1(x)\cdots K_k(x)\overline{M(x)}\Big| \leqslant C_1|(\mathbf{A}^1 - U)(\mathbf{F}_p)|$$

where C_1 is the product of the ranks of the sheaves. This means that it is enough to deal with the sum over $x \in U(\mathbf{F}_p)$.

By the Grothendieck–Lefschetz trace formula we have

$$\sum_{x \in U(\mathbf{F}_p)} K_1(x) \cdots K_k(x) \overline{M(x)} = -\operatorname{tr}(\operatorname{Fr} \mid H_c^1(U \times \bar{\mathbf{F}}_p, \bigotimes_i \mathcal{F}_i \otimes \mathrm{D}(\mathcal{G})))$$

since the H_c^0 and H_c^2 terms vanish, by assumption for H_c^2 and because we have a tensor product of middle-extension sheaves for H_c^0.

By Deligne's proof of the Riemann Hypothesis [3], since the tensor product is of weight 0, all eigenvalues of Frobenius acting on the cohomology space have modulus $\leqslant \sqrt{p}$, and hence

$$\left| \sum_{x \in U(\mathbf{F}_p)} K_1(x) \cdots K_k(x) \overline{M(x)} \right| \leqslant \dim H_c^1(U \times \bar{\mathbf{F}}_p, \bigotimes_i \mathcal{F}_i \otimes \mathrm{D}(\mathcal{G})) \times \sqrt{p}.$$

Finally, using the Euler-Poincaré formula, one sees that the dimension of this space is bounded in terms of the conductors of $\mathcal{F}_i$ and of $\mathcal{G}$, and in terms of k (see, e.g., [8, Lemma 3.3] for some details). $\qquad\square$

As already mentioned, Corollary 1.6 is an immediate consequence of Theorem 1.5 and Proposition 1.1. Corollary 1.7 is similar, except that in the argument of Proposition 1.1, there is a main term in the trace formula which is (for the Sp-type case) given by

$$\operatorname{tr}(\operatorname{Fr} \mid H_c^2(U \times \bar{\mathbf{F}}_p, \bigotimes \gamma_i^* \mathcal{F} \otimes \mathrm{D}(\mathcal{G}))).$$

However, the extra assumption that the geometric monodromy group coincides with the arithmetic monodromy group means that all eigenvalues of the Frobenius acting on H_c^2 are equal to p. Hence this contribution is equal to

$$p \dim H_c^2(\bigotimes \gamma_i^* \mathcal{F} \otimes \mathrm{D}(\mathcal{G})) = p \langle \bigotimes \gamma_i^* \mathcal{F}, \mathcal{G} \rangle$$

and for $\mathcal{G}$ given (as in the proof of Theorem 1.5) by

$$\mathcal{G} = \bigotimes_{\gamma \in \boldsymbol{\gamma}^*} \Lambda_\gamma(\gamma^* \mathcal{F})$$

with Λ_γ an irreducible representation of G which is a subrepresentation of $\operatorname{Std}^{\otimes n_\gamma}$, we have

$$\langle \bigotimes \gamma_i^* \mathcal{F}, \mathcal{G} \rangle = \prod_{\gamma \in \boldsymbol{\gamma}^*} \operatorname{mult}_{\Lambda_\gamma}(\operatorname{Std}^{\otimes n_\gamma})$$

where each multiplicity is at most k, and is equal to 1 if $n_\gamma = 1$. The result follows immediately.

The case of SL_r-type is similar and left to the reader; the extra condition that $\xi^* \mathcal{F} \simeq \mathrm{D}(\mathcal{F})$ (without a twist by a non-trivial rank 1 sheaf) allows us to deduce (4.2) with $\mathcal{L}_0$ trivial, from which the non-vanishing of H_c^2 follows when $(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ is not r-normal with respect to the special involution. (We already observed that under this condition we do not need to assume $p > r$ in Theorem 1.5).

5. Applications

We present here some applications of the general case developed in Section 2, going beyond the results of the introduction and of the previous section. The first recovers an estimate of Katz used by Fouvry and Iwaniec in their study of the divisor function in arithmetic progressions [4], the second discusses briefly the sums of Bombieri and Bourgain [2].

5.1. The Fouvry-Iwaniec sum.

In [4], for primes p and $(\alpha, \beta) \in \mathbf{F}_p^{\times 2}$, the exponential sum

$$S(\alpha, \beta; p) = \sum_t \mathrm{Kl}_2(\alpha(t-1)^2) \, \mathrm{Kl}_2((t-1)(\alpha t - \beta)) \, \mathrm{Kl}_2(\beta(t^{-1}-1)^2) \, \mathrm{Kl}_2((t^{-1}-1)(\beta t^{-1} - \alpha))$$

arises, where the sum is over $t \in \mathbf{F}_p^\times - \{1, \beta/\alpha\}$, and we abbreviate $\mathrm{Kl}_2(x) = \mathrm{Kl}_2(x; p)$. This is not of the type of Section 1, since the arguments of the Kloosterman sums are not simply of the form $\gamma_i \cdot t$. However, it fits the general framework of Section 2 with the 4-tuple

$$\boldsymbol{\mathcal{F}} = (f_i^* \mathcal{K}\ell_2)_{1 \leqslant i \leqslant 4},$$

where

$$f_1 = \alpha(X-1)^2, \qquad f_2 = (X-1)(\alpha X - \beta)$$
$$f_3 = \beta(X^{-1}-1)^2, \qquad f_4 = (X^{-1}-1)(\beta X^{-1} - 1).$$

Let $U = \mathbf{G}_m - \{1, \beta/\alpha\}$. We claim that this 4-tuple is U-generous if $\alpha \neq \beta$ (which is certainly a necessary condition, since otherwise $f_1 = f_2$). Indeed, since the geometric monodromy group of each $f_i^* \mathcal{F}$ is $\mathrm{SL}_2 = \mathrm{Sp}_2$ (because the geometric monodromy group of $\mathcal{K}\ell_2$ is SL_2, and SL_2 has no finite index algebraic subgroup), we need to check that there is no geometric isomorphism

$$f_i^* \mathcal{K}\ell_2 \simeq f_j^* \mathcal{K}\ell_2 \otimes \mathcal{L}$$

for $i \neq j$ and a rank 1 sheaf $\mathcal{L}$. But taking the dual and then tensoring, such an isomorphism implies

$$f_i^* \mathrm{End}(\mathcal{K}\ell_2) \simeq f_j^* \mathrm{End}(\mathcal{K}\ell_2),$$

on the open set $V = f_i^{-1}(\mathbf{G}_m)$ where the left-hand side of the original isomorphism (hence also the right-hand side) is lisse. Since $\mathrm{End}(\mathcal{K}\ell_2) \simeq \bar{\mathbf{Q}}_\ell \oplus \mathrm{Sym}^2(\mathcal{K}\ell_2)$, this implies that

$$f_i^* \mathrm{Sym}^2(\mathcal{K}\ell_2) \simeq f_j^* \mathrm{Sym}^2(\mathcal{K}\ell_2),$$

on V.

But since $\mathrm{Sym}^2(\mathcal{K}\ell_2)$ is ramified at 0 and ∞, the ramification loci S_i of the sheaves $f_i^* \mathrm{Sym}^2(\mathcal{K}\ell_2)$ are, respectively

$$S_1 = \{1, \infty\}, \qquad S_2 = \{1, \beta/\alpha, \infty\},$$
$$S_3 = \{0, 1\}, \qquad S_4 = \{0, 1, \beta\},$$

and are therefore distinct, proving the desired property of U-generosity.

Since the sum $S(\alpha, \beta; p)$ concerns the tensor product of

$$f_1^* \mathcal{K}\ell_2 \otimes f_2^* \mathcal{K}\ell_2 \otimes f_3^* \mathcal{K}\ell_2 \otimes f_4^* \mathcal{K}\ell_2$$

with the trivial sheaf $\bar{\mathbf{Q}}_\ell$, which is a tensor product of the trivial representations, which is not a subrepresentation of Std, it follows therefore that

$$\langle f_1^* \mathcal{K}\ell_2 \otimes f_2^* \mathcal{K}\ell_2 \otimes f_3^* \mathcal{K}\ell_2 \otimes f_4^* \mathcal{K}\ell_2, \bar{\mathbf{Q}}_\ell \rangle = 0$$

and hence by Proposition 1.1 that

$$S(\alpha, \beta; p) \ll p^{1/2}$$

for all primes p and $\alpha \neq \beta$ in $\mathbf{F}_p^\times$, where the implied constant is absolute. In the Appendix to [4], Katz gives a precise estimate of the implied constant.

5.2. **The Bombieri-Bourgain sums.** The Bombieri-Bourgain sums are defined by

$$S = \sum_{x \in \mathbf{F}_p} \prod_{1 \leqslant i \leqslant k} K_i(x + a_i) M(x)$$

(see [17, p. 513]) where

$$M(x) = e\left(\frac{bx + G(x)}{p}\right) \chi(g(x)),$$

$$K_i(x) = -\frac{1}{\sqrt{p}} \sum_{y \in \mathbf{F}_p} \chi_i(f_i(y)) e\left(\frac{g_i(y)}{p}\right) e\left(\frac{xy}{p}\right)$$

for some $b \in \mathbf{F}_p$ and $(a_1, \ldots, a_k) \in \mathbf{F}_p{}^k$, where

- $(\chi, \chi_1, \ldots, \chi_k)$ are non-trivial multiplicative characters modulo p,
- $f_i \in \mathbf{F}_p[X]$, $g \in \mathbf{F}_p[X]$ are non-zero polynomials,
- $g_i \in \mathbf{F}_p[X]$ and $G \in \mathbf{F}_p[X]$ may be zero.

This sum is of the type considered in Section 2, with

$$\mathcal{F}_i = [+a_i]^* \operatorname{FT}_\psi(\mathcal{L}_{\psi(g_i)} \otimes \mathcal{L}_{\chi(f_i)}),$$
$$\mathcal{G} = \mathcal{L}_{\psi(G+bX)} \otimes \mathcal{L}_{\chi(g)}$$

(or rather those $\mathcal{F}_i$ corresponding to the distinct parameters since this is not assumed to be the case).

Under (different) suitable conditions on these parameters, Bombieri and Bourgain [2, Lemma 33] and Katz [17, Th. 1.1] give estimates for S of the type

$$S \ll p^{1/2}$$

where the implied constant depends only on k and the degrees of the polynomials involved. Both proofs actually avoid involving monodromy groups which illustrates that sometimes an estimate for a sum of products might be easier to obtain than those involved in the previous sections. We show how to recover quickly the desired square-root cancellation in the case that occurs for the application considered by Bombieri and Bourgain, by a hybrid of Katz's argument and those of the previous sections.

In [2], the conditions are: p is odd, $g_i = G = 0$, $1 \leqslant \deg(f_i) \leqslant 2$, $\deg(g) \geqslant 2$, the f_i and g have only simple roots, and all χ_i and χ are equal and are of order 2. We then first note that if some f_i has degree 1, the resulting Fourier transform

$$\operatorname{FT}_\psi(\mathcal{L}_{\chi(f_i)})$$

is geometrically isomorphic to a tensor product

$$\mathcal{L}_{\psi(\alpha X)} \otimes \mathcal{L}_{\chi(X)}$$

25

(we use here that $\chi = \bar{\chi}$), so that by combining these with $\mathcal{G}$ we may assume that all f_i are of degree 2. Note that g is replaced by $X^k g$, where k is the number of i with $\deg(f_i) = 1$. Since χ has order 2, we have either k even and

$$\mathcal{L}_{\chi(X^k g)} \simeq \mathcal{L}_{\chi(g)},$$

so that the previous assumptions on g remain valid, or k odd and

$$\mathcal{L}_{\chi(X^k g)} \simeq \mathcal{L}_{X\chi(g)} \simeq \mathcal{L}_{\chi(\tilde{g})},$$

where $\tilde{g} = g/X$ if $g(0) = 0$, or $\tilde{g} = Xg$ otherwise; in the first case it may be that $\deg(\tilde{g}) = 1$, but in that case the unique zero of $\tilde{g}$ is in $\mathbf{G}_m$ since g has simple roots. In particular, in all cases, we see that g is replaced by a polynomial with at least one (simple) root in $\mathbf{G}_m$.

If all f_i were of degree 1, we are left with

$$\sum_x \chi(g(x))\psi(hx),$$

with g non-constant, which satisfies the desired conditions. We therefore assume that some f_i are of degree 2.

For a polynomial f_i of degree 2, by completing squares, we see that the Fourier transform

$$\mathrm{FT}_\psi(\mathcal{L}_{\chi(f_i)})$$

is geometrically isomorphic to a tensor product of $\mathcal{L}_{\psi(hX)}$ for some h and of the Fourier transform corresponding to a polynomial of the form $X^2 + c_i$. We may therefore assume that all f_i are of this form.

Finally, it is easy to see that

$$\mathrm{FT}_\psi(\mathcal{L}_{\chi(X^2 + c_i)}) \simeq [x \mapsto c_i x^2/4]^* \mathcal{K}\ell_2.$$

In particular, such sheaves are of rank 2, lisse on $\mathbf{G}_m$ and have geometric monodromy group $G_i = G_i^0 = \mathrm{SL}_2$. We therefore obtain a strictly $\mathbf{G}_m$-generous tuple by taking for $\mathcal{F}_i$ the Fourier transforms corresponding to the c_i's, modulo the equivalence relation $c_i \sim c_j$ if and only if

$$c_i c_j^{-1} \in \mathrm{Aut}_0([x \mapsto x^2/4]^* \mathcal{K}\ell_2).$$

We can now conclude: since g has a simple zero in $\mathbf{G}_m$, the sheaf $\mathcal{G}$ is ramified at at least one point inside $\mathbf{G}_m$, and therefore the irreducible sheaf $\mathcal{G}$ can not be a subsheaf of the tensor product

$$\bigotimes_i \mathcal{F}_i^{\otimes n_i}$$

which is lisse on $\mathbf{G}_m$.

Remark 5.1. Even if $\deg(g) = 1$, $g = \alpha X$ and $\alpha \neq 0$, we can obtain the square-root bounds provided we have at least one sheaf $\mathcal{F}_i$: by the results of Section 2, the condition

$$\left\langle \bigotimes_i \mathcal{F}_i^{\otimes n_i}, \mathrm{D}(\mathcal{G}) \right\rangle \neq 0$$

would imply that $\mathrm{D}(\mathcal{G})$ is geometrically isomorphic to

$$\bigotimes_i \mathrm{Sym}^{m_i}(\mathcal{F}_i)$$

for some $m_i \geqslant 0$. By rank considerations, we have $m_i = 0$, and this implies that $\mathcal{G}$ is geometrically trivial, which is impossible since g is non-constant.

References

[1] V. Blomer and D. Milićević: *The second moment of twisted modular L-functions*, preprint `arXiv: 1404.7845`.

[2] E. Bombieri and J. Bourgain: *On Kahane's ultraflat polynomials*, J. Europ. Math. Soc. 11 (2009), 627–703.

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

[4] É. Fouvry and H. Iwaniec: *The divisor function over arithmetic progressions*, with an Appendix by N. Katz, Acta Arith. 61 (1992), 271–287.

[5] É. Fouvry, E. Kowalski, Ph. Michel: *Algebraic twists of modular forms and Hecke orbits*, GAFA, to appear.

[6] É. Fouvry, E. Kowalski, Ph. Michel: *Algebraic trace weights over the primes*, Duke Math. Journal 163 (2014), 1683–1736.

[7] É. Fouvry, E. Kowalski, Ph. Michel: *On the exponent of distribution of the ternary divisor function*, Mathematika (2014), `doi:10.1112/S0025579314000096`.

[8] É. Fouvry, E. Kowalski, Ph. Michel: *Counting sheaves using spherical codes*, Math. Proc. Cambridge Phil. Soc. 20 (2013), 305–323.

[9] É. Fouvry, S. Ganguly, E. Kowalski, Ph. Michel: *Gaussian distribution for the divisor function and Hecke eigenvalues in arithmetic progressions*, Commentarii Math. Helv. 89 (2014), 979–1014

[10] É. Fouvry, Ph. Michel, J. Rivat and A. Sárközy: *On the pseudorandomness of the signs of Kloosterman sums*, J. Aust. Math. Soc. 77 (2004), 425–436.

[11] 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).

[12] A. Irving: *The divisor function in arithmetic progressions to smooth moduli*, preprint `arXiv:1403.8031`

[13] H. Iwaniec: *Small eigenvalues of Laplacian for* $\Gamma_0(N)$, Acta Arith. 56 (1990), no. 1, 65–82.

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

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

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

[17] N.M. Katz: *On a question of Bombieri and Bourgain*, Forum Math. 21 (2009), 511–517.

[18] N.M. Katz and P. Sarnak: *Random matrices, Frobenius eigenvalues and monodromy*, Colloq. Publ. 45, A.M.S 1999.

[19] E. Kowalski and G. Ricotta: *Fourier coefficients of* $GL(N)$ *automorphic forms in arithmetic progressions*, GAFA 24 (2014), 1229–1297.

[20] E. Kowalski and W. Sawin: *Kloosterman paths and the shape of exponential sums*, preprint (2014).

[21] Ph. Michel: *Minorations de sommes exponentielles*, Duke Math. J. 95 (1998), 227–240.

[22] R. Munshi: *Shifted convolution sums for* $GL(3) \times GL(2)$, Duke Math. J. 162 (2013), 2345–2362.

[23] N. Pitt: *On shifted convolutions of* $\zeta(3)$ *with automorphic L-functions*, Duke Math. J. 77 (1995), no. 2, 383–406.

[24] Y. Zhang: *Bounded gaps between primes*, Annals of Math. 179 (2014), 1121–1174.

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`