Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)
Vol. XXI (2020), 1453-1530

Stratification and averaging for exponential sums: bilinear forms with generalized Kloosterman sums

EMMANUEL KOWALSKI, PHILIPPE MICHEL AND WILL SAWIN

Abstract. We introduce a new comparison principle for exponential sums over finite fields in order to study "sum-product" sheaves that arise in the study of general bilinear forms with coefficients given by trace functions modulo a prime q. When these functions are hyper-Kloosterman sums with characters, we succeed in establishing cases of this principle that lead to non-trivial bounds below the Pólya-Vinogradov range. This property is proved by a subtle interplay between étale cohomology in its algebraic and diophantine incarnations. We give a first application of our bilinear estimates concerning the first moment of a family of L-functions of degree 3.

Mathematics Subject Classification (2010): 11T23 (primary); 11L05, 11N37, 11N75, 11F66, 14F20, 14D05 (secondary).

Contents

Ph. M. and E. K. were partially supported by a DFG-SNF lead agency program grant (grants 200021L_153647 and 200020L_175755). W. S. was partially supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zurich Foundation. Ph. M. and W. S. were partially supported by NSF Grant No. DMS-1440140.

Received April 30, 2018; accepted in revised form July 02, 2019.
Published online December 2020.

1. Introduction

1.1. Presentation of the results

Let $q \geqslant 1$ be an integer and let $K(\cdot; q)$ be a complex-valued q-periodic arithmetic function. A recurrent problem in analytic number theory is to evaluate how such functions correlate with other natural arithmetic functions $f(n)$, where f could be the characteristic function of an interval, or that of the primes, or the Fourier coefficients of some automorphic form. When facing such problems, one is often led to the problem of bounding non-trivially some bilinear forms

$$B(K, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum\sum_{m \leqslant M, n \leqslant N} \alpha_m \beta_n K(mn; q),$$

where the ranges of the variables $M, N \geqslant 1$ usually depend on q, and $\boldsymbol{\alpha} = (\alpha_m)_{m \leqslant M}$, $\boldsymbol{\beta} = (\beta_n)_{n \leqslant N}$ are complex numbers which, depending on the initial problem, are quite arbitrary. One of the main objectives is to improve on the trivial bound

$$\|K\|_\infty \|\boldsymbol{\alpha}\|_2 \|\boldsymbol{\beta}\|_2 (MN)^{1/2}$$

for ranges of M and N that are as small as possible compared to q; indeed, this uniformity is often more important than the strength of the saving compared to the trivial bound.

A natural benchmark is the *Pólya-Vinogradov* method, which often provides non-trivial bounds as long as $M, N \geqslant q^{1/2}$. Indeed, obtaining a result below that range is usually extremely challenging. When the modulus q is composite, a number of techniques exploiting the possibility of factoring q (starting with the Chinese Remainder Theorem) become available, and results exist in fair generality.

In this paper, we will only consider the case where q is a prime, and when K is a trace function (see [7] for a background survey).

The landmark result in this setting is the work of Burgess [3], which provides a non-trivial bound for the sum

$$\sum_{n \leqslant N} \chi(n)$$

when χ is a non-trivial Dirichlet character modulo q and $N \geqslant q^{3/8+\eta}$, for any $\eta > 0$. This is therefore well below Pólya-Vinogradov range. The ideas of Burgess (especially the "$+ab$ shifting trick") combine successfully the multiplicativity of χ and the (almost) invariance of intervals by additive translations.

Another twist of Burgess's method was given by the works of Karatsuba and Vinogradov, Friedlander-Iwaniec [12] and subsequently Fouvry-Michel [11] to bound non-trivially the bilinear sums $B(K, \boldsymbol{\alpha}, \boldsymbol{\beta})$ for various choices of functions K and ranges M, N shorter than $q^{1/2}$. In particular, using some version of the Sato-Tate equidistribution laws due to Katz [17], Fouvry and Michel considered

$$K(x; q) = e\left(\frac{x^k + ax}{q}\right), \ k \in \mathbf{Z} - \{0, 1, 2\}, \ a \in \mathbf{F}_q^\times, \quad (x, q) = 1,$$

and proved that for any $\delta > 0$, there exists $\eta > 0$ such that

$$\sum_{m \leqslant M, n \leqslant N}\sum \alpha_m \beta_n K(mn; q) \ll \|\boldsymbol{\alpha}\|_2 \|\boldsymbol{\beta}\|_2 (MN)^{1/2-\eta} \tag{1.1}$$

as long as

$$M, N \geqslant q^{\delta} \text{ and } MN \geqslant q^{3/4+\delta}. \tag{1.2}$$

The condition $MN \geqslant q^{3/4+\delta}$ is believed to be a barrier in this setting analogous to the condition $N > q^{1/4+\delta}$ in the Burgess bound for short character sums.

In our previous paper [21], motivated by the study of moments of L-functions (especially in our papers with Blomer, Milićević and Fouvry [1,2]), we obtained bounds of type (1.1) when $K(\cdot; q)$ is a hyper-Kloosterman sum, namely

$$\mathrm{Kl}_k(x; q) = \frac{1}{q^{\frac{k-1}{2}}} \sum_{\substack{y_1, \cdots, y_k \in \mathbf{F}_q^\times \\ y_1 \cdots y_k = x}} e\left(\frac{y_1 + \cdots + y_k}{q}\right),$$

where $k \geqslant 2$ is some fixed integer. More precisely, we proved that (1.1) holds as long as

$$M, N \geqslant q^{\delta} \text{ and } MN \geqslant q^{7/8+\delta}$$

for some $\delta > 0$. The argument was delicate and quite difficult.

In this second paper, we introduce a new approach that is both more robust and more powerful. The main complete exponential sum that needs to be bounded in this general setting is a difference of two exponential sums, which in previous work was bounded by estimating separately the main terms on both sides. Here, we show that the two underlying cohomology groups are equal, hence the main terms cancel, without explicitly calculating them. To establish the desired cohomological comparison, we define a stratification of the parameter space, and show using vanishing cycles that if the result fails at any point of one of the strata, it fails on the generic point. Using a variant of Katz's diophantine criterion of irreducibility, this implies that the original exponential sum estimate fails on average over the stratum. We check that the strata are defined by equations of a specific type, which makes the averaged exponential sum estimate amenable to classical analytic techniques, specifically separation of variables.

Remark 1.1. As the referee pointed out to us, a similar stratification strategy is present in the paper [26] of J. Xu on multiplicative character sums, where the key applications are related to multi-variable Burgess estimates. The main differences are that in Xu's method the stratification is more abstract, whereas for us it is explicit, and Xu's method relies on the higher moments of the exponential sums, while we use only the first moment.

Our main application in this paper is the proof of the estimate (1.1) in the full range (1.2) for generalized hyper-Kloosterman sums with character twists, whose definition we now recall. Let $k \geqslant 1$ be an integer, and let $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_k)$ be a tuple of k Dirichlet characters modulo q, each of which might be trivial. The $(k-1)$-dimensional generalized Kloosterman sums associated to $\boldsymbol{\chi}$ are the exponential sums defined for $x \in \mathbf{F}_q^\times$ by

$$\mathrm{Kl}_k(x; \boldsymbol{\chi}, q) = \frac{1}{q^{\frac{k-1}{2}}} \sum_{\substack{y_1,\cdots,y_k \in \mathbf{F}_q^\times \\ y_1 \cdots y_k = x}} \chi_1(y_1) \cdots \chi_k(y_k) e\left(\frac{y_1 + \cdots + y_k}{q}\right).$$

The hyper-Kloosterman sums (which correspond to the case $\chi_i = 1$) were introduced by Deligne [25], and these generalisations were introduced and studied by Katz in [16, Chapter 4]. As an application of the Riemann Hypothesis over finite fields, Deligne and Katz established the highly non-trivial pointwise bounds

$$|\mathrm{Kl}_k(x; \boldsymbol{\chi}, q)| \leqslant k.$$

The finer properties of these sums were studied in great depth by Katz in [16] and [17]. Among other things, Katz proved equidistribution statements that describe precisely the distribution of generalized Kloosterman sums inside $\mathbf{C}$, at least for most possible choices of $\boldsymbol{\chi}$.

A special case of our main result, Theorem 4.1, is the following:

Theorem 1.2. *Assume that $\boldsymbol{\chi}$ has Property* NIO *of Definition* 2.1*, for instance all χ_i are trivial. For any $\delta > 0$ there exists $\eta > 0$ such that for any integer $k \geqslant 2$, any prime number q, and any integers $M, N \geqslant 1$ such that*

$$M,\ N \geqslant q^{\delta},\ MN \geqslant q^{3/4+\delta}$$

we have

$$\sum\sum_{m \leqslant M, n \leqslant N} \alpha_m \beta_n \mathrm{Kl}_k(amn; \boldsymbol{\chi}, q) \ll \|\boldsymbol{\alpha}\|_2 \|\boldsymbol{\beta}\|_2 (MN)^{1/2-\eta}$$

for any $a \in \mathbf{F}_q^\times$ and for arbitrary families of complex numbers $\boldsymbol{\alpha} = (\alpha_m)_{m \leqslant M}$ and $\boldsymbol{\beta} = (\beta_n)_{n \leqslant N}$. The implied constant depends only on δ and k.

Property NIO (short for "Not Induced or Orthogonal") is an elementary combinatorial property that we define below in Section 2; it is easy to check, and it is "generically" satisfied in some sense. For instance, the case $\boldsymbol{\chi} = (1, \ldots, 1)$ corresponding to hyper-Kloosterman sums themselves has NIO, and so does $(1, \ldots, 1, \chi)$ if k is odd.

The exponent $3/4 = 2 \times 3/8$ seem to be a recurring barrier: it occurs in classical subconvexity estimates for L-functions, and more recently (see [6, 8]) when dealing with sums of the shape

$$\sum_{\substack{p \leqslant N \\ p \text{ prime}}} K(p; q),$$

where p ranges over prime numbers, or

$$\sum_{n \leqslant N} \lambda_f(n) K(n; q),$$

where K is a general trace function modulo q and $(\lambda_f(n))_{n \leqslant N}$ are the Hecke eigenvalues of a fixed Hecke eigenform f (cuspidal or Eisenstein).

For special bilinear forms, where one of the variables is smooth, *i.e.*, for

$$B(K, \boldsymbol{\alpha}, 1_N) = \sum\sum_{m \leqslant M, n \leqslant N} \alpha_m K(mn; q),$$

the barrier occurs at a shorter range, and we again are able to prove an estimate that reaches this barrier.

A special case of Theorem 4.3 is:

Theorem 1.3. *Assume that* $\boldsymbol{\chi}$ *has* NIO. *For any* $\delta > 0$ *there exists* $\eta > 0$ *such that for* $k \geqslant 2$ *an integer,* q *a prime and* $M, N \geqslant 1$ *some integers satisfying*

$$M, N \geqslant q^{\delta}, \; MN^2 \geqslant q^{1+\delta}$$

we have

$$\sum\sum_{m \leqslant M, n \leqslant N} \alpha_m \mathrm{Kl}_k(amn; \boldsymbol{\chi}, q) \ll \|\boldsymbol{\alpha}\|_2 (MN^2)^{1/2-\eta}$$

for any $a \in \mathbf{F}_q^{\times}$ *and for any tuple of complex numbers* $\boldsymbol{\alpha} = (\alpha_m)_{m \leqslant M}$*, where the implicit constant depends on* δ *and* k.

In particular, for $M = N$, we obtain a non-trivial bound as long as

$$M = N \geqslant q^{1/3+\delta}$$

for some $\delta > 0$. If we denote by $d_2(n)$ the classical divisor function, we deduce the following result:

Corollary 1.4. *Assume that* $\boldsymbol{\chi}$ *has* NIO. *For any* $\delta > 0$*, there exists* $\eta > 0$ *such that for any integer* $k \geqslant 2$*, any prime number* q*, and any* $N \geqslant q^{2/3+\delta}$*, we have*

$$\sum_{n \leqslant N} d_2(n) \mathrm{Kl}_k(an; \boldsymbol{\chi}, q) \ll N q^{-\eta},$$

for any $a \in \mathbf{F}_q^{\times}$ *where the implicit constant depends on* δ *and* k.

It is of considerable interest to generalize results like Theorem 1.2 to other trace functions K modulo q. We believe that the methods in this paper could be applicable when K satisfies suitable big monodromy assumptions, and has the following property: K belongs to a family K_a parameterized by non-trivial additive characters $x \mapsto e(ax/p)$ of $\mathbf{F}_q$, and this family satisfies a relation of the type $K_{a^{\mu}}(x) = K(a^{\nu}x)$ for some fixed non-zero integers μ and ν. For instance, this holds for the generalized Kloosterman sums with $\mu = 1$, $\nu = k$ when defining

$$K_a(x) = \frac{1}{q^{\frac{k-1}{2}}} \sum_{\substack{y_1, \cdots, y_k \in \mathbf{F}_q^{\times} \\ y_1 \cdots y_k = x}} \chi_1(y_1) \cdots \chi_k(y_k) e\left(\frac{a(y_1 + \cdots + y_k)}{q}\right).$$

1.2. Applications to moments of L-functions

As with our previous paper [21], Theorems 1.2 and 1.3 have applications to the evaluation of moments of L-functions indexed by Dirichlet characters modulo q. As a simple illustration, we will prove in Section 3 the following result, which generalizes some recent work of Zacharias [28]:

Theorem 1.5. *Let f be a primitive holomorphic cusp form of level 1. For q prime, let ξ be a non-trivial Dirichlet character modulo q. There exist an absolute constant $\delta > 0$ such that*

$$\frac{1}{q-1}\sum_{\chi \,(\mathrm{mod}\, q)} L(f\otimes\chi, 1/2)L(\xi\chi, 1/2) = 1 + O_f\big(q^{-\delta}\big).$$

Remark 1.6. Zacharias established this asymptotic for $\xi = 1$ using amongst other ingredients the bounds from [21] for $K(x) = \mathrm{Kl}_3(x; (1,1,1), q)$; he evaluated more generally a mollified version of this average, enabling him to establish that, for q large, there is a positive proportion of $\chi \,(\mathrm{mod}\, q)$ such that $L(f\otimes\chi, 1/2)$ and $L(\chi, 1/2)$ are both non-vanishing. Most likely a similar result may be established in our case.

As in [1,2,21], we also expect that our results will prove useful to estimate other averages of certain L-functions of degree 3 and 4 indexed by Dirichlet characters. For instance, we may consider:

- The twisted first moment

$$\frac{1}{q-1}\sum_{\chi \,(\mathrm{mod}\, q)} L(f\otimes\chi, 1/2)L(\xi\chi, 1/2)\prod_i \varepsilon_{\xi_i\chi}^{k_i},$$

  where $\boldsymbol{\xi} = (\xi_i)_i$ a tuple of characters of modulus q (possibly trivial) and $\boldsymbol{k} = (k_i)_i$ is a family of integers;
- The shifted second moment

$$\frac{1}{q-1}\sum_{\chi \,(\mathrm{mod}\, q)} L(f\otimes\chi, 1/2)L(f\otimes\xi\overline{\chi}, 1/2).$$

1.3. Principle of the stratification and averaging method

We denote $K(x) = \mathrm{Kl}_k(ax; \boldsymbol{\chi}, q)$ for a fixed k-tuple $\boldsymbol{\chi}$ with Property NIO and a fixed $a \in \mathbf{F}_q^\times$.

As in our previous work (and [11,12]), the proof starts with an application of the $+ab$-shifting trick of Karatsuba and Vinogradov. Let us recall that the shifting trick builds on the almost invariance of an interval under sufficiently small translations. The interval to be shifted here is that of the n variable (either directly for Theorem 1.3 or after an application of Cauchy's inequality for Theorem 1.2) and the shift is by product $+ab$ with $(a, b) \in [A, 2A[\times[B, 2B[$ for A, B suitable parameters (such that $AB = N$). As $K(mn; q)$ depends only on the congruence class of $mn \pmod q$ the replacement of $n \leftrightarrow n + ab$ leads to the following transformations

$$\begin{aligned} mn \pmod q &\leftrightarrow m(n+ab) = am(\overline{a}n + b) = s(r+b) \pmod q, \\ (m_1 n, m_2 n) \pmod q &\leftrightarrow (am_1(\overline{a}n+b), am_2(\overline{a}n+b)) \\ &= (s_1(r+b), s_2(r+b)) \pmod q, \end{aligned}$$

with (r, s), (r, s_1, s_2) taking values in $\mathbf{F}_q \times \mathbf{F}_q^\times$ or $\mathbf{F}_q \times (\mathbf{F}_q^{\times 2} - \Delta(\mathbf{F}_q^{\times 2}))$. Under suitable assumptions on A, M, N one can then show that the above maps are essentially injective (*i.e.*, have fibers bounded in size by $q^{o(1)}$). However, these maps are far from being surjective, so performing such a change of variable will result in a loss. This can be tamed by an application of the Hölder inequality with a sufficiently large exponent, which we denote by $2l$ in the sequel. This process leads then to the problem of bounding sums of the shape

$$\sum_{\boldsymbol{b} \in \mathcal{B}} \big|\Sigma_I(K, \boldsymbol{b})\big|, \quad \sum_{\boldsymbol{b} \in \mathcal{B}} \big|\Sigma_{II}(K, \boldsymbol{b})\big|,$$

where $\mathcal{B}$ denotes the set of $2l$-uples of integers $\boldsymbol{b} = (b_1, \cdots, b_{2l}) \in [B, 2B[^{2l}$ and

$$\begin{aligned} \Sigma_I(K, \boldsymbol{b}) &= \sum_{r \in \mathbf{F}_q} \sum_{s \in \mathbf{F}_q^\times} \mathbf{K}(sr, s\boldsymbol{b}), \\ \Sigma_{II}(K, \boldsymbol{b}) &= \sum_{r \in \mathbf{F}_q} \sum\sum_{\substack{s_1, s_2 \in \mathbf{F}_q^\times \\ s_1 \neq s_2}} \mathbf{K}(s_1 r, s_1 \boldsymbol{b}) \overline{\mathbf{K}(s_2 r, s_2 \boldsymbol{b})}, \end{aligned}$$

where

$$\mathbf{K}(r, \boldsymbol{b}) = \prod_{i=1}^{l} K(r + b_i) \overline{K(r + b_{i+l})}. \tag{1.3}$$

The goal is to give individual bounds for sums $\Sigma_I(K, \boldsymbol{b})$ and $\Sigma_{II}(K, \boldsymbol{b})$ with square-root cancellation, namely we wish to prove that

$$\Sigma_I(K, \boldsymbol{b}) \ll q, \qquad \Sigma_{II}(K, \boldsymbol{b}) \ll q^{3/2}.$$

A key fact is that these bounds do not always hold, but it will be enough to prove them outside a sufficiently small subset $\mathcal{B}^{\mathrm{diag}}$ of "diagonal" tuples $\boldsymbol{b}$. This subset will be the set of $\mathbf{F}_q$-points of a proper algebraic subvariety $\mathcal{V}^{\mathrm{diag}} \subset \mathbf{A}^{2l}_{\mathbf{F}_q}$. In fact, it is crucial (to avoid the loss involved in Hölder's inequality) to prove the required estimates outside of a variety $\mathcal{V}^{\mathrm{diag}}$ with large codimension, and we will do this with

$$\operatorname{codim}\left(\mathcal{V}^{\mathrm{diag}}\right) \geqslant \frac{l-1}{2}. \tag{1.4}$$

The outcome is that by taking l very large, we obtain non-trivial estimates of $B(K, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and $B(K, \boldsymbol{\alpha}, 1_N)$ in the ranges defined by

$$MN \geqslant q^{3/4+\delta} \text{ and } MN^2 \geqslant q^{2/3+\delta}$$

for any $\delta > 0$.

We now sketch the proof in the case of general bilinear forms (the special bilinear forms are easier). Setting

$$\mathbf{R}(r, \boldsymbol{b}) = \sum_{s \in \mathbf{F}_q^\times} \mathbf{K}(sr, s\boldsymbol{b}),$$

we observe that

$$\Sigma_{II}(K, \boldsymbol{b}) = \sum_{r \in \mathbf{F}_q} |\mathbf{R}(r, \boldsymbol{b})|^2 - \sum_{s \in \mathbf{F}_q^\times} \sum_{r \in \mathbf{F}_q} |\mathbf{K}(sr, s\boldsymbol{b})|^2.$$

This is the difference of two sums of positive terms, which therefore individually will have main terms, and we need these main terms to compensate exactly for $\boldsymbol{b} \notin \mathcal{V}^{\text{diag}}$. Our argument for this in [21] relies on separate evaluations of both sums to witness the coincidence of the main terms. But one can check that this evaluation only holds outside of a codimension 1 subvariety, which is far from (1.4) except in the case $l = 2$.

In this paper, we compare directly the two terms in the difference. This comparison is not a combinatorial or analytic rearrangement of terms, but is a cohomological comparison using the ideas of ℓ-adic cohomology to interpret exponential sums. Using this formalism, we interpret the functions

$$(r, \boldsymbol{b}) \to \mathbf{K}(r, \boldsymbol{b}),\ \ \mathbf{R}(r, \boldsymbol{b})$$

as trace functions of ℓ-adic sheaves $\mathcal{K}$ and $\mathcal{R}$ on $\mathbf{A} \times \mathbf{A}^{2l}$, which are pointwise pure of weight 0 and mixed of weight $\leqslant 1$ respectively. The functions

$$(r, \boldsymbol{b}) \to |\mathbf{K}(r, \boldsymbol{b})|^2,\ \ |\mathbf{R}(r, \boldsymbol{b})|^2$$

are the trace functions of the endomorphisms sheaves $\mathrm{End}(\mathcal{K})$ and $\mathrm{End}(\mathcal{R})$. By means of the Grothendieck-Lefschetz trace formula and of Deligne's most general form of the Riemann Hypothesis over finite fields [4], the desired bound

$$\sum_{r \in \mathbf{F}_q} |\mathbf{R}(r, \boldsymbol{b})|^2 - \sum_{s \in \mathbf{F}_q^\times} \sum_{r \in \mathbf{F}_q} |\mathbf{K}(sr, s\boldsymbol{b})|^2 \ll q^{3/2}$$

for a given $\boldsymbol{b}$ can be interpreted as stating that the specialized sheaves $\mathcal{K}_{\boldsymbol{b}}$ and $\mathcal{R}_{\boldsymbol{b}}$ have decompositions into geometrically irreducible components whose multiplicities precisely match.

This interpretation relies on the relationship between $\mathcal{K}_{\boldsymbol{b}}$ and $\mathcal{R}_{\boldsymbol{b}}$. As $\mathcal{R}_{\boldsymbol{b}}$ is obtained from applying a cohomology functor to $\mathcal{K}_{\boldsymbol{b}}$, each irreducible component ϱ of $\mathcal{K}_{\boldsymbol{b}}$ defines a summand $\widetilde{\varrho}$ of $\mathcal{R}_{\boldsymbol{b}}$. We check explicitly that these summands $\widetilde{\varrho}$ are nontrivial, which implies that the exponential sums match if and only if all the summands $\widetilde{\varrho}$ are themselves irreducible, and are pairwise non-isomorphic as ϱ varies.

The sheaf $\mathcal{R}_{\boldsymbol{b}}$ is a sheaf on the affine line, lisse away from a finite set of singular points that vary depending on $\boldsymbol{b}$. Using Deligne's semicontinuity theorem, and assuming that the local monodromy of $\mathcal{R}_{\boldsymbol{b}}$ is tame, we can show that the decomposition into irreducible components of $\mathcal{R}_{\boldsymbol{b}}$ is constant on any set of parameters $\boldsymbol{b}$ over which this varying finite set $S_{\boldsymbol{b}}$ of singular points does not itself develop singularities (*i.e.*, over which the size of $S_{\boldsymbol{b}}$ is constant). The tameness condition can be verified for large primes (which is sufficient for us) by expressing $\mathcal{R}_{\boldsymbol{b}}$ as the characteristic p fiber of a sheaf defined in characteristic zero. This reduces the problem to the generic points of the strata of the stratification of the parameter space by the number of singular points in $S_{\boldsymbol{b}}$.

To get a handle on this stratification, we first calculate the set of singular points. By an explicit inductive argument, we show how the strata can be expressed by equations in the coefficients b_i and auxiliary variables; these equations split into sums of different terms involving different subsets of the b_i's. We can then estimate the average of the complete sums $\Sigma_{II}(K, \boldsymbol{b})$ over a single stratum using only estimates for one-variable exponential sums, as long as the number of equations and auxiliary variables is not too large (which means that we must keep control of these numbers in the inductive argument). From this average estimate and the geometric interpretation, we deduce that the sheaves $\mathcal{K}_{\boldsymbol{b}}$ and $\mathcal{R}_{\boldsymbol{b}}$ have the same decomposition into irreducibles when $\boldsymbol{b}$ belongs to a stratum of sufficiently large dimension. This proves the desired result for all $\boldsymbol{b}$ except those in low-dimensional strata, which we simply consider as part of "diagonal" subset. It is therefore crucial that our induction is efficient enough to get a good bound on the codimension of this subset.

Stratifications where the validity of a desired estimate on a stratum only depends on its validity at the generic point exist for arbitrary families of complete exponential sums, arising from the stratification of a constructible ℓ-adic sheaf into lisse sheaves. They can often be computed by vanishing cycles methods, such as Deligne's semicontinuity theorem. We expect that proving estimates for individual strata by passing to the average and applying elementary analytic methods (which are known to perform very well when the number of variables to average over is large enough) will be a useful strategy for many families of exponential sums.

Remark 1.7. (1) It would be reasonable to expect that the correct codimension is

$$\operatorname{codim}\left(\mathcal{V}^{\mathrm{diag}}\right) \geqslant l + o(l)$$

as $l \to +\infty$, which would indeed be best possible (it is easy to see that the codimension is $\leqslant l$). A lower bound of this quality was established in [11] in the case $K(x) = e((x^k + a)/q)$ already mentioned. Although the bound (1.4) only goes half of the way to this expectation, it is nevertheless sufficient for our purpose, and it seems that even the full lower bound would not help in improving the exponents $3/4$ and $2/3$ in Theorems 1.2 and 1.3.

(2) Readers who have some familiarity with either [11] or [21] will have noticed that we will make a compromise in our argument: the new variables s and (s_1, s_2) belong to the subsets $[A, 2AM]$ and $[A, 2AM]^2 - \Delta([A, 2AM]^2)$ of the larger sets $\mathbf{F}_q^\times$ or $\mathbf{F}_q^{\times 2} - \Delta(\mathbf{F}_q^{\times 2})$, so that we lose something by "forgetting" this fact by positivity. It is certainly possible to compensate for this loss using the completion method, introducing additional twists by additive characters in the s-variable, and handling them by arguments similar to those of [21, Section 4.5]. However, when l is very large, the improvement in the final bounds is very small (because of (1.4)), and more importantly the final limiting exponents $3/4$ and $2/3$ are not improved. So we have chosen to avoid the completion step, in order to simplify an already complex argument. It should be noted however that, for small values of l, the completion step is worth pursuing, and that is was crucial in [21] to obtain non-trivial bounds for $l = 2$ (which was the only case that could be handled in [21], because, as noted earlier, the diagonal variety in that paper was of codimension 1).

Notation

For any prime number ℓ, we fix an isomorphism $\iota : \overline{\mathbf{Q}}_\ell \to \mathbf{C}$. Let q be a prime number. Given an algebraic variety $X_{\mathbf{F}_q}$, a prime $\ell \neq q$ and a constructible $\overline{\mathbf{Q}}_\ell$-sheaf $\mathcal{F}$ on X, we denote by $t_{\mathcal{F}} : X(\mathbf{F}_q) \longrightarrow \mathbf{C}$ its trace function, defined by

$$t_{\mathcal{F}}(x) = \iota\Big(\operatorname{Tr}\big(\operatorname{Fr}_{x,\mathbf{F}_q} \mid \mathcal{F}_x\big)\Big),$$

where $\mathcal{F}_x$ denotes the stalk of $\mathcal{F}$ at x. More generally, for any finite extension $\mathbf{F}_{q^d}/\mathbf{F}_q$, we denote by $t_{\mathcal{F}}(\cdot; \mathbf{F}_{q^d})$ the trace function of $\mathcal{F}$ over $\mathbf{F}_{q^d}$, namely

$$t_{\mathcal{F}}\Big(x; \mathbf{F}_{q^d}\Big) = \iota\Big(\operatorname{Tr}\big(\operatorname{Fr}_{x,\mathbf{F}_{q^d}} \mid \mathcal{F}_x\big)\Big).$$

An ℓ-adic sheaf will aways means a $\overline{\mathbf{Q}}_\ell$-sheaf. For standard facts in ℓ-adic cohomology (such as proper base change, cohomological dimension, etc), we refer to the books of Fu [13] and Milne [23], and to the notes of Deligne [25].

We will usually omit writing down ι. In any expression where some element z of $\overline{\mathbf{Q}}_\ell$ has to be interpreted as a complex number, we mean to consider $\iota(z)$.

We denote by $\mathcal{F}^\vee$ the dual of a constructible sheaf $\mathcal{F}$; if $\mathcal{F}$ is a middle-extension sheaf, we will use the same notation for the middle-extension dual.

Let ψ (respectively χ) be a non-trivial additive (respectively multiplicative) character of $\mathbf{F}_q$. We denote by $\mathcal{L}_\psi$ (respectively $\mathcal{L}_\chi$) the associated Artin-Schreier (respectively Kummer) sheaf on $\mathbf{A}^1_{\mathbf{F}_q}$ (respectively on $(\mathbf{G}_m)_{\mathbf{F}_q}$), as well (by abuse of notation) as their middle extension to $\mathbf{P}^1_{\mathbf{F}_q}$. The trace functions of the latter are given by

$$t_\psi\Big(x;\mathbf{F}_{q^d}\Big)=\psi\Big(\mathrm{Tr}_{\mathbf{F}_{q^d}/\mathbf{F}_q}(x)\Big)\quad\text{if }\ x\in\mathbf{F}_{q^d},\qquad t_\psi\Big(\infty;\mathbf{F}_{q^d}\Big)=0,$$
$$t_\chi\Big(x;\mathbf{F}_{q^d}\Big)=\chi\Big(\mathrm{Nr}_{\mathbf{F}_{q^d}/\mathbf{F}_q}(x)\Big)\quad\text{if }\ x\in\mathbf{F}_{q^d}^\times,\qquad t_\chi(0;\mathbf{F}_{q^d})=t_\chi\Big(\infty;\mathbf{F}_{q^d}\Big)=0.$$

For the trivial additive or multiplicative character, the trace function of the middle-extension is the constant function 1.

Given $\lambda\in\mathbf{F}_{q^d}$, we denote by $\mathcal{L}_{\psi_\lambda}$ the Artin-Schreier sheaf of the character of $\mathbf{F}_{q^d}$ defined by $x\mapsto\psi(\mathrm{Tr}_{\mathbf{F}_{q^d}/\mathbf{F}_q}(\lambda x))$.

If $X_{\mathbf{F}_q}$ is an algebraic variety, ψ (respectively χ) is an ℓ-adic additive character of $\mathbf{F}_q$ (respectively ℓ-adic multiplicative character) and $f:X\longrightarrow\mathbf{A}^1$ (respectively $g:X\longrightarrow\mathbf{G}_m$) is a morphism, we denote by either $\mathcal{L}_{\psi(f)}$ or $\mathcal{L}_\psi(f)$ (respectively by $\mathcal{L}_{\chi(g)}$ or $\mathcal{L}_\chi(g)$) the pullback $f^*\mathcal{L}_\psi$ of the Artin-Schreier sheaf associated to ψ (respectively the pullback $g^*\mathcal{L}_\chi$ of the Kummer sheaf). These are lisse sheaves on X with trace functions $x\mapsto\psi(f(x))$ and $x\mapsto\chi(g(x))$, respectively. The meaning of the notation $\mathcal{L}_\psi(f)$, which we use when putting f as a subscript would be typographically unwieldy, will always be unambiguous, and no confusion with Tate twists will arise.

Given a variety $X/\mathbf{F}_q$, an integer $k\geqslant 1$ and a function c on X, we denote by $\mathcal{L}_\psi(cs^{1/k})$ the sheaf on $X\times\mathbf{A}^1$ (with coordinates (x,s)) given by $\alpha_*\mathcal{L}_{\psi(c(x)t)}$, where α is the covering map $(x,s,t)\mapsto(x,s)$ on the k-fold cover

$$\Big\{(x,s,t)\in X\times\mathbf{A}^1\times\mathbf{A}^1\ \mid\ t^k=s\Big\}.$$

Given a field extension $L/\mathbf{F}_p$, and elements $\alpha\in L^\times$ and $\beta\in L$, we denote by $[\times\alpha]$ the scaling map $x\mapsto\alpha x$ on $\mathbf{A}^1_L$, and by $[+\beta]$ the additive translation $x\mapsto x+\beta$. For a sheaf $\mathcal{F}$, we denote by $[\times\alpha]^*\mathcal{F}$ (respectively $[+\alpha]^*\mathcal{F}$) the respective pull-back operation.

We will usually not indicate base points in étale fundamental groups; whenever this occurs, it will be clear that the properties under consideration are independent of the choice of a base point.

ACKNOWLEDGEMENTS. Ph. M. and W. S. thank the students of the Arizona Winter School 2016 who studied the paper [21] and worked on generalizing some of its steps. They also thank MSRI where parts of this work were completed during the "Analytic Number Theory" programme during the first semester 2017.

We thank É. Fouvry and I. Shparlinski for comments, and the referee for useful remarks, especially for the reference to the paper [26] of J. Xu.

2. Preliminaries

We begin by defining Property NIO, and a useful variant called CGM (for "Connected Geometric Monodromy"). These are motivated by results of Katz (see [17, Corollary 8.9.2, Theorem 8.8.1-8.8.2]).

Definition 2.1. Let A be a finite cyclic group and $\boldsymbol{\chi} = (\chi_1, \dots, \chi_k)$ a tuple of characters of A. Let $\Lambda = \chi_1 \cdots \chi_k$.

(1) The tuple $\boldsymbol{\chi}$ is *Kummer-induced* if there exists a divisor d of k, $d \neq 1$, and a tuple $(\xi_1, \dots, \xi_{k/d})$ of characters of A such that the χ's are all the characters with $\chi^d = \xi_j$ for some j, with multiplicity;
(2) The tuple $\boldsymbol{\chi}$ is *self-dual* if there is a character ξ such that the set of characters $\chi \in \boldsymbol{\chi}$, with multiplicity, is stable under $\chi \mapsto \xi\chi^{-1}$. The character ξ is called a "dualizing character";
(3) A self-dual tuple $\boldsymbol{\chi}$ is *alternating* if k is even and $\Lambda = \xi^{k/2}$, and otherwise, it is *symmetric*;
(4) A tuple $\boldsymbol{\chi}$ has Property NIO if it is not Kummer-induced and, if k is even, if it is not self-dual symmetric;
(5) A tuple $\boldsymbol{\chi}$ has Property CGM if it is not Kummer-induced, and $\chi_1 \cdots \chi_k = 1$, and one of the following conditions holds:
   - k is odd;
   - $\boldsymbol{\chi}$ is not self-dual;
   - k is even, $\boldsymbol{\chi}$ is self-dual and alternating, and the dualizing character ξ is trivial.

Example 2.2. We consider Dirichlet characters modulo q in these examples.

(1) Consider the case $k = 2$ and q odd, $\boldsymbol{\chi} = (\chi_1, \chi_2)$. Denote by $\chi_{(2)}$ the non-trivial real character of $\mathbf{F}_q^\times$. Then $\boldsymbol{\chi}$ is:

- Kummer-induced if and only if $\chi_2 = \chi_1\chi_{(2)}$;
- If not Kummer-induced, always self-dual alternating, taking $\xi = \chi_1\chi_2$ as dualizing character.

In particular, for $\boldsymbol{\chi} = (1, \chi_2)$, the alternating case is $\chi_2 = 1$, corresponding to the "classical" Kloosterman sum, and the non self-dual case is $\chi_2^2 \neq 1$. The Kummer-induced tuple $\boldsymbol{\chi} = (1, \chi_{(2)})$ corresponds to Salié sums.

(2) If k is odd, then $\boldsymbol{\chi}$ has NIO if and only if it is not Kummer-induced. In particular, this is the case if $\chi_1 = \dots = \chi_{k-1} = 1$.

(3) If $\chi_1 = \cdots = \chi_k = 1$, then $\boldsymbol{\chi}$ has NIO.

In the next section, we will need the following useful lemma which bounds the number of integral points in a box that satisfy a system of polynomial equations modulo q. We thank the referee for giving us a convenient reference.

Lemma 2.3. *Let $k \geqslant 1$ be an integer and let $A > 0$. Let $X_{\mathbf{Z}} \subset \mathbf{A}^k_{\mathbf{Z}}$ be an algebraic variety of dimension $d \geqslant 0$ given by the vanishing of $\leqslant A$ polynomials of degree $\leqslant A$. Let p be a prime number and $0 \leqslant B < p/2$ an integer. Then*

$$\left|\left\{x = (x_1, \dots, x_k) \in \mathbf{F}_p^k \mid x \in X\big(\mathbf{F}_p\big) \text{ and } B \leqslant x_i \leqslant 2B \text{ for } 1 \leqslant i \leqslant k\right\}\right| \ll B^d$$

where the implied constant depends only on k and A, and the notation $B \leqslant x_i \leqslant 2B$ means that the unique integer between 1 and $p-1$ congruent to x_i modulo p belongs to the interval $[B, 2B]$.

See [26, Lemma 1.7] for a proof.

3. An application to moments of L-functions

In this section, we will prove Theorem 1.5, which we recall is a variation of a recent result of Zacharias [28].

Let f be a primitive cusp form of level 1, trivial nebentypus and weight k_f, with Hecke eigenvalues $\lambda_f(n)$. For Dirichlet characters χ and ξ modulo q, we consider the L-function

$$L((f \oplus \xi) \otimes \chi, s) = L(f \otimes \chi, s) L(\chi\xi, s)$$

of degree 3. Note that for $\mathrm{Re}(s) > 1$, we have the Dirichlet series expansion

$$L((f \oplus \xi) \otimes \chi, s) = \sum_{n \geqslant 1} \chi(n)\big(\lambda_f \star \xi\big)(n) n^{-s}.$$

We wish to evaluate the average

$$\mathcal{M} = \frac{1}{q-1} \sum_{\chi \ (\mathrm{mod}\, q)} L((f \oplus \xi) \otimes \chi, 1/2),$$

proving that $\mathcal{M} = 1 + O(q^{-\alpha})$ for some $\alpha > 0$.

The proof is very similar to [28, Section 6.2], which corresponds to the case $\xi = 1$, so we will only sketch certain steps.

We assume for simplicity that ξ is even (*i.e.*, $\xi(-1) = 1$), and we will only evaluate the *even* moment

$$\mathcal{M}^+ = \frac{2}{q-1} \sum_{\chi \ (\mathrm{mod}\, q)}^{+} L((f \oplus \xi) \otimes \chi, 1/2),$$

where $\sum^+$ restricts the sum to even primitive characters modulo q. We will prove that $\mathcal{M}^+ = \frac{1}{2} + O(q^{-\alpha})$ for some $\alpha > 0$. The sum over odd characters satisfies the same asymptotics, hence this implies Theorem 1.5.

Define $\Gamma_{\mathbf{R}}(s)=\pi^{-s/2}\Gamma(s/2)$ and let $L_\infty(s)=L_\infty(f,s)L_\infty(\chi\xi,s)$, where

$$L_\infty(f,s)=\Gamma_{\mathbf{R}}\left(s+\frac{k-1}{2}\right)\Gamma_{\mathbf{R}}\left(s+\frac{k+1}{2}\right),\quad L_\infty(\chi\xi,s)=\Gamma_{\mathbf{R}}(s),$$

are the archimedean L-factors of $L(f,s)$ and $L(\chi\xi,s)$ respectively. Further, let

$$\varepsilon((f\oplus\xi)\otimes\chi)=\varepsilon(f)\varepsilon_\chi^2\varepsilon_{\chi\xi},$$

where ε_η denotes the normalized Gauss sum of a Dirichlet character. Define then the completed L-function

$$\Lambda((f\oplus\xi)\otimes\chi,s)=q^{3s/2}L_\infty(s)L((f\oplus\xi)\otimes\chi,s).$$

For ξ and χ even, we then have the functional equation

$$\Lambda((f\oplus\xi)\otimes\chi,s)=\varepsilon((f\oplus\xi)\otimes\chi)\Lambda(f\otimes\overline{\chi}\oplus\overline{\chi\xi},1-s).$$

Let $0<\alpha<1/4$ be a parameter to be fixed later. For χ even, non-trivial and not equal to ξ^{-1}, we apply the approximate functional equation to the L-function $L((f\oplus\xi)\otimes\chi,s)$, in an unbalanced form ([15, Theorem 5.3] with q replaced by the conductor q^3 and $X=q^{1/2-2\alpha}$). After adding the contribution of the character ξ^{-1}, which is $\ll q^{-1/5+\varepsilon}$ for any $\varepsilon>0$, this gives $\mathcal{M}^+=\mathcal{M}_1+\mathcal{M}_2$, where

$$\mathcal{M}_1=\frac{2}{q-1}\sideset{}{^+}\sum_{\chi\ (\mathrm{mod}\, q)}\sum_{n\geqslant 1}\frac{\chi(n)(\lambda_f\star\xi)(n)}{n^{1/2}}\mathcal{V}\left(\frac{n}{q^{2-2\alpha}}\right),$$

$$\mathcal{M}_2=\frac{2}{q-1}\sideset{}{^+}\sum_{\chi\ (\mathrm{mod}\, q)}\varepsilon((f\oplus\xi)\otimes\chi)\sum_{n\geqslant 1}\overline{\frac{\chi(n)(\lambda_f\star\xi)(n)}{n^{1/2}}}\mathcal{V}\left(\frac{n}{q^{1+2\alpha}}\right),$$

where the function $\mathcal{V}$ is defined by

$$\mathcal{V}(y)=\frac{1}{2i\pi}\int_{(1)}\frac{L_\infty\left(\frac{1}{2}+s\right)}{L_\infty\left(\frac{1}{2}\right)}G(s)y^{-s}\frac{ds}{s},\qquad G(s)=\exp\left(s^2\right),$$

for $y>0$. Shifting the s-contour to the right if $y\geqslant 1$ or to $\mathrm{Re}(s)=-1/2$ if $y\leqslant 1$, we deduce that

$$y^i\mathcal{V}^{(i)}(y)\ll_{A,i,f}(1+y)^{-A}$$

for any $A>0$ and $i\geqslant 0$, and

$$\mathcal{V}(y)=1+O(y^{1/2})\text{ for }y\leqslant 1.$$

It follows from the first of these bounds that, for any $\kappa>0$, the contribution to both sums of the integers $n\geqslant q^{3/2+\kappa}$ is $\ll_{A,f,\kappa} q^{-A}$ for any $A\geqslant 0$.

We first bound $\mathcal{M}_1$. We add to $\mathcal{M}_1$ the contribution of the trivial character, up to an error term bounded by $O(q^{-1/5})$, and perform the summation over the even characters χ. We obtain

$$\begin{aligned}\mathcal{M}_1 &= \sum_{n\equiv\pm1\,(\mathrm{mod}\,q)} \frac{(\lambda_f\star\xi)(n)}{n^{1/2}}\mathcal{V}\left(\frac{n}{q^{2-2\alpha}}\right)+O\big(q^{-1/5}\big)\\ &= \mathcal{V}\left(\frac{1}{q^{2-2\alpha}}\right)+O\big(q^{-\alpha+\varepsilon}\big)=1+O\big(q^{-\alpha+\varepsilon}\big),\end{aligned}$$

for any $\varepsilon>0$, where the first term $\mathcal{V}(q^{-2+2\alpha})$ is the contribution of the trivial solution $n=1$ of the congruence $n\equiv\pm1\,(\mathrm{mod}\,q)$.

Now we consider $\mathcal{M}_2$. We add to $\mathcal{M}_2$ the contribution of the trivial character, up to an error of size $\ll q^{\varepsilon+\frac{1}{2}+\alpha-1}\ll q^{\varepsilon+\alpha-1/2}$, for any $\varepsilon>0$. We then perform the summation over χ even. We have

$$\begin{aligned}\frac{1}{q-1}\sideset{}{^+}\sum_{\chi\,(\mathrm{mod}\,q)}\varepsilon((f\oplus\xi)\otimes\chi)\overline{\chi}(n) &= \frac{\varepsilon(f)}{q-1}\sideset{}{^+}\sum_{\chi\,(\mathrm{mod}\,q)}\varepsilon_\chi^2\varepsilon_{\chi\xi}\overline{\chi}(n)\\ &= \frac{\varepsilon(f)}{q^{1/2}}\left(\mathrm{Kl}_3(n;\xi,q)+\mathrm{Kl}_3(-n;\xi,q)\right),\end{aligned}$$

where we abbreviate

$$\mathrm{Kl}_3(\pm n;\xi,q)=\mathrm{Kl}_3(\pm n;(1,1,\xi),q).$$

Hence we have

$$\mathcal{M}_2=\frac{\varepsilon(f)}{q^{1/2}}\sum_n\frac{\overline{(\lambda_f\star\xi)(n)}}{n^{1/2}}(\mathrm{Kl}_3(n;\xi,q)+\mathrm{Kl}_3(-n;\xi,q))\mathcal{V}\left(\frac{n}{q^{1+2\alpha}}\right)+O\big(q^{-1/5}\big).$$

We open the Dirichlet convolution

$$\overline{(\lambda_f\star\xi)(n)}=\sum_{ab=n}\lambda_f(a)\overline{\xi(b)}.$$

By standard techniques (dyadic subdivisions, inverse Mellin transform to separate the variables), we establish that $\mathcal{M}_2$ is, up to a factor $\ll q^\varepsilon$ for any $\varepsilon>0$, bounded by the sum of $\ll(\log q)^2$ bilinear sums of the type

$$\mathcal{M}_2(M,N)=\frac{1}{(qMN)^{1/2}}\sum_{m,n}\lambda_f(m)\overline{\xi(n)}\mathrm{Kl}_3(amn;\xi,q)V\left(\frac{m}{M}\right)W\left(\frac{n}{N}\right),$$

where

$$1\leqslant MN\leqslant q^{1+2\alpha},$$

$a = 1$ or -1, and V and W are smooth functions, compactly supported in $[1, 2]$, such that

$$x^i V^i(x),\ x^i W^i(x) \ll_{f,\varepsilon} q^{i\varepsilon}$$

for any $\varepsilon > 0$ and $i \geqslant 0$.

We set $M = q^{\mu}$ and $N = q^{\nu}$. The trivial bound is

$$\mathcal{M}_2(M, N) \ll q^{\varepsilon} \left(\frac{MN}{q}\right)^{1/2} = q^{(\mu+\nu)/2-1/2+\varepsilon}$$

for any $\varepsilon > 0$, which is $\ll q^{-\alpha+\varepsilon}$ if $\mu + \nu \leqslant 1 - 2\alpha$. Now assume that

$$1 - 2\alpha \leqslant \mu + \nu \leqslant 1 + 2\alpha.$$

Estimating the sum over n by the Pólya-Vinogradov technique (completion), summing trivially over the m variable, we obtain

$$\mathcal{M}_2(M, N) \ll q^{\varepsilon} \left(\frac{M}{N}\right)^{1/2} \ll q^{1/2-\nu+\alpha+\varepsilon}$$

for any $\varepsilon > 0$. This bound is $\ll q^{-\alpha+\varepsilon}$ if $\nu \geqslant \frac{1}{2} + 2\alpha$. We then assume that

$$\nu \leqslant \tfrac{1}{2} + 2\alpha.$$

If ν is small, so that μ is large, we apply [8, Theorem 1.2] to the sum over m, summing trivially over n. We get

$$\mathcal{M}_2(M, N) \ll N q^{-1/8+\alpha+\varepsilon} = q^{-1/8+\nu+\alpha+\varepsilon}$$

for any $\varepsilon > 0$. Again, this is $\ll q^{-\alpha+\varepsilon}$ provided $\nu \leqslant \frac{1}{8} - 2\alpha$. Now assume that

$$\tfrac{1}{8} - 2\alpha \leqslant \nu \leqslant \tfrac{1}{2} + 2\alpha.$$

Then $\frac{1}{2} - 4\alpha \leqslant \mu \leqslant \frac{7}{8} + 4\alpha$. The general bilinear form estimate in [6, Theorem 1.17] gives

$$\mathcal{M}_2(M, N) \ll q^{\varepsilon+\alpha} \min\left(N^{-1} + M^{-1} q^{1/2}, M^{-1} + N^{-1} q^{1/2}\right)^{1/2},$$

which is $\ll q^{-\alpha+\varepsilon}$ provided $\alpha \leqslant 1/32$ and

$$\max(\mu, \nu) \geqslant \tfrac{1}{2} + 2\alpha.$$

We finally consider the case when $\alpha \leqslant 1/32$ and

$$\tfrac{1}{2} - 4\alpha \leqslant \mu,\ \nu \leqslant \tfrac{1}{2} + 2\alpha.$$

In this situation, we can then apply Theorem 1.2 for the triple $\boldsymbol{\chi} = (1, 1, \xi)$, which has Property NIO for any ξ by Example 2.2 (2). We obtain the bound

$$\mathcal{M}_2(M, N) \ll q^{\varepsilon} \left(\frac{MN}{q}\right)^{1/2} (MN)^{-\eta} \ll q^{2\alpha+\varepsilon}(MN)^{-\eta} \ll q^{2\alpha-3\eta/4+\varepsilon}$$

for any $\varepsilon > 0$, where $\eta > 0$ is the saving exponent in Theorem 1.2 when the parameter δ there is $\delta = \frac{1}{4} - 8\alpha$. Hence, for $\alpha > 0$ fixed and small enough, we obtain

$$\mathcal{M}_2(M, N) \ll q^{-\eta'+\varepsilon}$$

for some fixed $\eta' > 0$ and any $\varepsilon > 0$, where the implied constant depends on ε and f.

4. Reduction to complete exponential sums

In this section, we will state the general forms of Theorems 1.2 and 1.3, and reduce their proofs to certain bounds for families of exponential sums over finite fields. In fact, we begin with slightly more general bilinear sums.

Let q be a prime number, and let $K : \mathbf{F}_q \to \mathbf{C}$ be any function. Let M, N be integers such that $1 \leqslant M, N \leqslant q - 1$. Let $\mathcal{M}$ be a subset of the positive integers $m \leqslant q - 1$ of cardinality M. We set $M^+ = \max_{m\in\mathcal{M}} m$. Let finally

$$\mathcal{N} = \{n \mid 1 \leqslant n < N\}.$$

Given tuples of complex numbers $\boldsymbol{\alpha} = (\alpha_m)_{m\in\mathcal{M}}$ and $\boldsymbol{\beta} = (\beta_n)_{n\in\mathcal{N}}$, we set

$$B(K, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{m\in\mathcal{M},}\sum_{n\in\mathcal{N}} \alpha_m \beta_n K(mn).$$

We will prove the following:

Theorem 4.1. *Fix an integer $k \geqslant 2$. Let q be a prime and let $a \in \mathbf{F}_q^{\times}$. Let $\boldsymbol{\chi}$ be a k-tuple of Dirichlet characters modulo q. Suppose that $\boldsymbol{\chi}$ has Property* NIO*, and define $K(x) = \mathrm{Kl}_k(ax; \boldsymbol{\chi}, q)$. With notations as above, for any integer $l \geqslant 2$ and any $\varepsilon > 0$, we have*

$$B(K, \boldsymbol{\alpha}, \boldsymbol{\beta}) \ll q^{\varepsilon} \|\alpha\|_2 \|\beta\|_2 (MN)^{1/2} \left(\frac{1}{M} + \left(\frac{q^{\frac{3}{4}+\frac{3}{4l}}}{MN}\right)^{\frac{1}{l}}\right)^{1/2},$$

where the implied constant depends only on (k, l, ε), provided one of the following two conditions holds:

$$q^{\frac{3}{2l}} \leqslant N < \frac{1}{2} q^{\frac{1}{2}-\frac{3}{4l}},$$

$$q^{\frac{3}{2l}} \leqslant N,\ NM^+ < \frac{1}{2} q^{1-\frac{3}{2l}}.$$

Remark 4.2. This bound is non-trivial only for l large enough, precisely for $l \geqslant 9$. As we will explain, this limitation results from our simplifying choice of not applying the completion method to detect that an auxiliary variable belongs to some interval in $\mathbf{F}_q$.

In the special case of "type I" sums, we obtain:

Theorem 4.3. *With the same notation and assumption as in Theorem* 4.1*, especially assuming that* $\boldsymbol{\chi}$ *has* NIO*, and with the additional condition that* $\beta_n = 1$ *for* $n \in \mathcal{N}$*, for any integer* $l \geqslant 1$ *and any* $\varepsilon > 0$*, we have*

$$B(K, \boldsymbol{\alpha}, \mathbf{1}) \ll q^{\varepsilon} \|\boldsymbol{\alpha}\|_1^{1-\frac{1}{l}} \|\boldsymbol{\alpha}\|_2^{\frac{1}{l}} M^{\frac{1}{2l}} N \left(\frac{q^{1+\frac{1}{l}}}{MN^2}\right)^{1/2l},$$

where the implied constant depends on (k, l, ε)*, provided one of the following two conditions holds:*

$$q^{\frac{1}{l}} \leqslant N \leqslant \frac{1}{2} q^{1/2+1/2l},$$

$$q^{\frac{1}{l}} \leqslant N,\ NM^+ \leqslant \frac{1}{2} q^{1+1/2l}.$$

Remark 4.4. As l gets large, this bound is non-trivial if

$$M^+ N \leqslant q,\ MN^2 \geqslant q^{1+\delta}$$

for some $\delta > 0$. In particular for $M = M^+ = N$, this is non trivial if

$$N \geqslant q^{1/3+\delta}.$$

4.1. The type II bilinear sum

We now start the proof of the reduction step for Theorem 4.1.

Applying Cauchy's inequality, we obtain

$$|B(K, \boldsymbol{\alpha}, \boldsymbol{\beta})| \leqslant \|\beta\|_2 \left(\sum_n |\sum_m \alpha_m K(mn)|^2\right)^{1/2} \ll \|\beta\|_2 \Big(\|\alpha\|_2^2 N + S^{\neq}\Big)^{1/2},$$

where

$$S^{\neq} = \sum_{m_1 \neq m_2} \alpha_{m_1} \overline{\alpha_{m_2}} \sum_n K(m_1 n) \overline{K(m_2 n)}.$$

We now use the $+ab$-shift trick of Karatsuba-Vinogradov as in [11,21]. For this we introduce two integer parameters $A, B \geqslant 1$ such that $AB \leqslant N$. Using the notation $a \sim A$ for $A \leqslant a < 2A$, we then have

$$S^{\neq} = \frac{1}{AB} \sum_{a \sim A, b \sim B} \sum\sum_{m_1 \neq m_2} \alpha_{m_1} \overline{\alpha_{m_2}} \sum_{n+ab \in \mathcal{N}} K(m_1(n+ab)) \overline{K(m_2(n+ab))}.$$

Using the fact that $\mathcal{N}$ is an interval, we deduce as in [11, page 126, (7.2)] (see also [21, (2.11)] that

$$S^{\neq} \ll \frac{\log q}{AB} \sum\sum_{\substack{a,m_1,m_2,n\\ a\sim A, m_1\neq m_2}} |\alpha_{m_1}\alpha_{m_2}| \left|\sum_{b\sim B} K(m_1(n+ab))\overline{K(m_2(n+ab))}e(bt)\right|$$

for some $t \in \mathbf{R}$ and n varying over an interval of length $\ll N+AB$. For $(r,s_1,s_2)\in(\mathbf{F}_q^{\times})^3$ set

$$\nu(r,s_1,s_2) = \sum\sum_{\substack{a,m_1\neq m_2,n\\ a\sim A, \overline{a}n\equiv r, am_i\equiv s_i}} \left|\alpha_{m_1}\alpha_{m_2}\right|,$$

so that

$$S^{\neq} \ll \frac{\log q}{AB}\sum\sum_{r,s_1,s_2} \nu(r,s_1,s_2) \left|\sum_{b\sim B} K(s_1(r+b))\overline{K(s_2(r+b))}e(bt)\right|$$

(by the change of variable $r=\overline{a}\cdot n,\ s_i = a\cdot m_i,\ i=1,2$). We have

$$\sum\sum_{r,s_1,s_2}\nu(r,s_1,s_2) = \sum\sum_{a,n,m_1\neq m_2}|\alpha_{m_1}\alpha_{m_2}| \leqslant AN\|\alpha\|_1^2 \leqslant AMN\|\alpha\|_2^2$$

and

$$\sum\sum_{r,s_1,s_2}\nu(r,s_1,s_2)^2 = \sum\sum_{a,n,m_1\neq m_2}|\alpha_{m_1}||\alpha_{m_2}| \sum\sum_{\substack{a',n',m_1'\neq m_2'\\ \overline{a}'n'\equiv\overline{a}n,\ a'm_i'\equiv am_i \pmod q}} |\alpha_{m_1'}||\alpha_{m_2'}|.$$

Now assume that

$$2AN < q. \tag{4.1}$$

Then the equation $\overline{a}'n' \equiv \overline{a}n \pmod q$ is equivalent to $an' \equiv a'n \pmod q$, which is equivalent to $an' = a'n$. Therefore if we fix a and n', the integers a' and n are determined up to $q^{o(1)}$ values.

Suppose that a, a', n, n' are so chosen. For $i=1,2$, we then have

$$\sum\sum_{\substack{m_i,m_i'\\ am_i\equiv a'm_i' \pmod q}} |\alpha_{m_i}||\alpha_{m_i'}| \leqslant \sum\sum_{\substack{m_i,m_i'\\ am_i\equiv a'm_i' \pmod q}} |\alpha_{m_i}|^2 + \sum\sum_{\substack{m_i,m_i'\\ am_i\equiv a'm_i' \pmod q}} |\alpha_{m_i'}|^2 \ll \|\alpha\|_2^2.$$

Indeed, since $\mathcal{M}$ is a subset of $[1,q-1]$, once m_i (respectively m_i') is given, the congruence $am_i \equiv a'm_i' \pmod q$ uniquely determines m_i' (respectively m_i). Therefore

$$\sum\sum_{r,s_1,s_2}\nu(r,s_1,s_2)^2 \ll q^{o(1)}AN\|\alpha\|_2^4. \tag{4.2}$$

Alternatively, if we assume instead of (4.1) that

$$2AM^{+} < q, \tag{4.3}$$

then the same reasoning with the equation $am_1 \equiv a'm'_1 \pmod q$ also leads to (4.2).

Fix an integer $l \geqslant 2$. We apply Hölder's inequality in the following form:

$$\begin{aligned}
&\sum_{r,s_1,s_2}\sum \nu^{1-\frac{1}{l}+\frac{1}{l}} \Bigl|\sum_{b\sim B} \cdots\Bigr| \\
&\leqslant \Bigl(\sum_{r,s_1,s_2}\sum \nu\Bigr)^{1-\frac{1}{l}} \Bigl(\sum_{r,s_1,s_2}\sum \nu \Bigl|\sum_{b\sim B}\cdots\Bigr|^{l}\Bigr)^{1/l} \\
&\leqslant \Bigl(\sum_{r,s_1,s_2}\sum \nu\Bigr)^{1-\frac{1}{l}} \Bigl(\sum_{r,s_1,s_2}\sum \nu^2\Bigr)^{1/2l} \Bigl(\sum_{r,s_1,s_2}\sum \Bigl|\sum_{b\sim B}\cdots\Bigr|^{2l}\Bigr)^{1/2l} \\
&\leqslant q^{\varepsilon}\|\alpha\|_2^2 (AN)^{1-\frac{1}{2l}} M^{1-\frac{1}{l}} \Bigl(\sum_{\boldsymbol{b}\in\mathcal{B}} |\Sigma_{II}(K,\boldsymbol{b})|\Bigr)^{1/2l},
\end{aligned}$$

where $\mathcal{B} = [B, 2B[^{2l}$, and

$$\Sigma_{II}(K,\boldsymbol{b}) = \sum_{r\in\mathbf{F}_q}\sum_{\substack{s_1,s_2\in\mathbf{F}_q^{\times}\\ s_1\neq s_2}}\sum \mathbf{K}(s_1r, s_1\boldsymbol{b})\overline{\mathbf{K}(s_2r, s_2\boldsymbol{b})}$$

is the exponential sum defined in (1.3), where

$$\mathbf{K}(r,\boldsymbol{b}) = \prod_{i=1}^{l} K(r+b_i)\overline{K(r+b_{i+l})}.$$

We observe at this point that the sum $\Sigma_{II}(K,\boldsymbol{b})$ is independent of the parameter a such that $K(x) = \mathrm{Kl}_k(ax;\boldsymbol{\chi},q)$, by changing the variables s_1 and s_2 to as_1 and as_2 respectively.

We will estimate these sums in different ways depending on the position of $\boldsymbol{b}$. Precisely:

Theorem 4.5. *There exist affine varieties*

$$\mathcal{V}^{\Delta} \subset \mathcal{W} \subset \mathbf{A}^{2l}_{\mathbf{Z}}$$

defined over $\mathbf{Z}$ *such that*

$$\operatorname{codim}\bigl(\mathcal{V}^{\Delta}\bigr) = l, \qquad \operatorname{codim}(\mathcal{W}) \geqslant \frac{l-1}{2}$$

which have the following property: for any prime q large enough, depending only on k, for any tuple $\boldsymbol{\chi}$ of characters of $\mathbf{F}_q^\times$ with Property NIO*, for any $a \in \mathbf{F}_q^\times$, and for all $\boldsymbol{b} \in \mathbf{F}_q^{2l}$, with*

$$K(x) = \mathrm{Kl}_k(ax; \boldsymbol{\chi}, q),$$

we have

$$\Sigma_{II}(K, \boldsymbol{b}) \ll q^3 \text{ if } \boldsymbol{b} \in \mathcal{V}^\Delta\big(\mathbf{F}_q\big), \tag{4.4}$$

$$\Sigma_{II}(K, \boldsymbol{b}) \ll q^2 \text{ if } \boldsymbol{b} \in (\mathcal{W} - \mathcal{V}^\Delta)\big(\mathbf{F}_q\big), \tag{4.5}$$

$$\Sigma_{II}(K, \boldsymbol{b}) \ll q^{3/2} \text{ if } \boldsymbol{b} \notin \mathcal{W}\big(\mathbf{F}_q\big). \tag{4.6}$$

In all cases, the implied constant depends only on k.

We emphasize that the varieties $\mathcal{V}^\Delta$ and $\mathcal{W}$ are independent of the tuple of characters. After a number of preliminaries, the final proof of this theorem will be found in Section 14 (see page 1527).

We will apply these estimates for the parameters $\boldsymbol{b}$ belonging to the box $[B, 2B)^{2l}$, and for this we use Lemma 2.3.

Let $\mathcal{B}^{\mathcal{V}}$ (respectively $\mathcal{B}^{\mathcal{W}}$) be the set of $\boldsymbol{b} \in \mathcal{B}$ such that $\boldsymbol{b} \in \mathcal{V}^\Delta(\mathbf{F}_q)$ (respectively $\boldsymbol{b} \in \mathcal{W}(\mathbf{F}_q)$). Since the subvarieties $\mathcal{V}^\Delta$ and $\mathcal{W}$ are defined over $\mathbf{Z}$, it follows from Lemma 2.3 that

$$\begin{aligned}\sum_{\boldsymbol{b}} |\Sigma_{II}(K, \boldsymbol{b})| &\ll q^3\big|\mathcal{B}^{\mathcal{V}}\big| + q^2\big|\mathcal{B}^{\mathcal{W}}\big| + q^{3/2}B^{2l} \\ &\ll q^3 B^{2l-\mathrm{codim}(\mathcal{V}^\Delta)} + q^2 B^{2l-\mathrm{codim}(\mathcal{W})} + q^{3/2}B^{2l}.\end{aligned} \tag{4.7}$$

We have $\mathrm{codim}(\mathcal{V}^\Delta) = l$ and $\mathrm{codim}(\mathcal{W}) \geqslant (l-1)/2$ by Theorem 4.5. We choose B so that the first and third terms in (4.7) are equal, namely

$$B = q^{3/2l}.$$

We also choose A so that $AB = N$, *i.e.*

$$A = N/B = Nq^{-\frac{3}{2l}}.$$

Writing $\mathrm{codim}(\mathcal{W}) = \gamma l$, we deduce that

$$|B(K, \boldsymbol{\alpha}, \boldsymbol{\beta})| \leqslant \|\beta\|_2\big(\|\alpha\|_2^2 N + S^{\neq}\big)^{1/2},$$

where

$$S^{\neq} \ll \frac{q^\varepsilon}{AB}\|\alpha\|_2^2 (AN)^{1-\frac{1}{2l}} M^{1-\frac{1}{l}} \Big(q^2 B^{(2-\gamma)l)} + q^{3/2}B^{2l}\Big)^{1/2l}.$$

Hence

$$\begin{aligned}|B(K,\boldsymbol{\alpha},\boldsymbol{\beta})| &\ll q^{\varepsilon}\|\alpha\|_2\|\beta\|_2(MN)^{1/2}\left(\frac{1}{M}+\left(\frac{q^2B^{-\gamma l}}{AM^2N}+\frac{q^{\frac{3}{2}}}{AM^2N}\right)^{\frac{1}{2l}}\right)^{1/2}\\ &\ll q^{\varepsilon}\|\alpha\|_2\|\beta\|_2(MN)^{1/2}\left(\frac{1}{M}+\left(\frac{q^{2-\frac{3}{2}\gamma+\frac{3}{2l}}}{(MN)^2}+\frac{q^{\frac{3}{2}+\frac{3}{2l}}}{(MN)^2}\right)^{\frac{1}{2l}}\right)^{1/2}.\end{aligned} \tag{4.8}$$

This holds under the condition that

$$A = Nq^{-\frac{3}{2l}} \geqslant 1$$

and that either of (4.1) or (4.3) hold.

In particular, since $\gamma \geqslant 1/3$, the second term on the right-hand side of (4.8) is smaller than the third. This implies Theorem 4.1. Theorem 1.2 follows by choosing l large enough depending on δ.

4.2. Bounding type I sums

We turn now to Theorem 4.3, and consider the special bilinear form

$$B(K,\boldsymbol{\alpha},1_{\mathcal{N}}) = \sum_{m\in\mathcal{M},}\sum_{n\in\mathcal{N}} \alpha_m K(mn).$$

Given $l \geqslant 2$, a trivial bound is

$$B(K,\boldsymbol{\alpha},1_{\mathcal{N}}) \leqslant \|\alpha\|_1^{1-\frac{1}{l}}\|\alpha\|_2^{\frac{1}{2l}}M^{\frac{1}{2l}}N.$$

Proceeding as before, we get

$$\begin{aligned}B(K,\boldsymbol{\alpha},1_{\mathcal{N}}) &= \frac{1}{AB}\sum_{a\sim A,}\sum_{B\sim B}\sum_{m\in\mathcal{M}}\alpha_m\sum_{n+ab\in\mathcal{N}}K(m(n+ab))\\ &\ll_{\varepsilon}\frac{q^{\varepsilon}}{AB}\sum_{r\in\mathbf{F}_q,}\sum_{s\in\mathbf{F}_q^{\times}}\nu(r,s)\left|\sum_{b\sim B}\eta_b K(s(r+b))\right|,\end{aligned}$$

with

$$\nu(r,s) = \sum_{\substack{a\sim A,\\ am=s,}}\sum_{\substack{m\in\mathcal{M},\\ \overline{a}n\equiv r \,(\mathrm{mod}\, q)}}\sum_{n\in\mathcal{N}} |\alpha_m|$$

and $|\eta_b| \leqslant 1$. We have

$$\sum_{r,s}\nu(r,s) \ll AN\sum_{m\in\mathcal{M}}|\alpha_m|.$$

We also have

$$\sum_{r,s} \nu(r,s)^2 = \sum_{\substack{a,m,n,a',m',n' \\ am\equiv a'm', a'n\equiv an' \,(\mathrm{mod}\, q)}} \cdots \sum |\alpha_m||\alpha_{m'}|.$$

Assuming that

$$2AN < q \ \text{ or } \ 2AM^+ < q \tag{4.9}$$

we show by the same reasoning as above that

$$\sum_{r,s} \nu(r,s)^2 \ll \sum_{a,m} |\alpha_m|^2 \sum_{\substack{n,a',m',n' \\ am=a'm' \\ a'n=an' \,(\mathrm{mod}\, q)}} \cdots \sum 1 \ll_\varepsilon q^\varepsilon AN \sum_m |\alpha_m|^2.$$

We next apply Hölder's inequality in the form

$$\begin{aligned}
&\sum_{r\in\mathbf{F}_q, s\in\mathbf{F}_q^\times} \sum \nu(r,s) \Bigl| \sum_{B<b\leqslant 2B} \eta_b K(s(r+b)) \Bigr| \\
&\leqslant \Bigl(\sum_{r,s} \nu(r,s) \Bigr)^{1-\frac{1}{l}} \Bigl(\sum_{r,s} \nu(r,s)^2 \Bigr)^{\frac{1}{2l}} \Bigl(\sum_{r,s} \Bigl| \sum_{B<b\leqslant 2B} \eta_b K(s(r+b)) \Bigr|^{2l} \Bigr)^{\frac{1}{2l}} \\
&\ll_\varepsilon q^\varepsilon (AN)^{1-\frac{1}{2l}} \|\boldsymbol{\alpha}\|_1^{1-\frac{1}{l}} \|\boldsymbol{\alpha}\|_2^{\frac{1}{l}} \Bigl(\sum_{r,s} \Bigl| \sum_{B<b\leqslant 2B} \eta_b K(s(r+b)) \Bigr|^{2l} \Bigr)^{\frac{1}{2l}}.
\end{aligned}$$

Expanding the $2l$-th power, we have

$$\sum_{r\in\mathbf{F}_q, s\in\mathbf{F}_q^\times} \sum \Bigl| \sum_{B<b\leqslant 2B} \eta_b K(s(r+b)) \Bigr|^{2l} \leqslant \sum_{\boldsymbol{b}\in\mathcal{B}} \bigl| \Sigma_I(K,\boldsymbol{b}) \bigr|,$$

with

$$\Sigma_I(K,\boldsymbol{b}) = \sum_{r\in\mathbf{F}_q} \sum_{s\in\mathbf{F}_q^\times} \mathbf{K}(sr, s\boldsymbol{b}) = \sum_{r\in\mathbf{F}_q} \mathbf{R}(r,\boldsymbol{b}). \tag{4.10}$$

Note that $\Sigma_I(K,\boldsymbol{b})$ is independent of the choice of $a \in \mathbf{F}_q^\times$ such that $K(x) = \mathrm{Kl}_k(ax; \boldsymbol{\chi}, q)$. We have reached the bound

$$B(K,\boldsymbol{\alpha},1_{\mathcal{N}}) \ll q^\varepsilon \|\boldsymbol{\alpha}\|_1^{1-\frac{1}{l}} \|\boldsymbol{\alpha}\|_2^{\frac{1}{l}} M^{\frac{1}{2l}} N \Bigl(\frac{(MN)^{-1}}{AB^{2l}} \sum_{\boldsymbol{b}\in\mathcal{B}} \bigl| \Sigma_I(K,\boldsymbol{b}) \bigr| \Bigr)^{\frac{1}{2l}}. \tag{4.11}$$

As before, we can prove different bounds on $\Sigma_I(K,\boldsymbol{b})$ depending on the position of $\boldsymbol{b}$.

Theorem 4.6. *Let $\mathcal{V}^{\Delta}$ and $\mathcal{W}$ be the affine varieties on Theorem* 4.5*. For any prime q large enough, depending only on k, for any tuple $\boldsymbol{\chi}$ with Property* NIO*, for any $a \in \mathbf{F}_q^{\times}$ and for all $\boldsymbol{b} \in \mathbf{F}_q^{2l}$, with*

$$K(x) = \mathrm{Kl}_k(ax; \boldsymbol{\chi}, q),$$

we have

$$\Sigma_I(K, \boldsymbol{b}) \ll q^2 \text{ if } \boldsymbol{b} \in \mathcal{V}^{\Delta}(\mathbf{F}_q), \tag{4.12}$$

$$\Sigma_I(K, \boldsymbol{b}) \ll q^{3/2} \text{ if } \boldsymbol{b} \in (\mathcal{W} - \mathcal{V}^{\Delta})(\mathbf{F}_q), \tag{4.13}$$

$$\Sigma_I(K, \boldsymbol{b}) \ll q \text{ if } \boldsymbol{b} \notin \mathcal{W}(\mathbf{F}_q). \tag{4.14}$$

In all cases, the implied constant depends only on k.

This is also proved ultimately in Section 14 (page 1528).

Taking this for granted, and using the same notation $\mathrm{codim}(\mathcal{W}) = \gamma l$ as before, we have therefore

$$\begin{aligned}\sum_{\boldsymbol{b} \in \mathcal{B}} \big|\Sigma_I(K, \boldsymbol{b})\big| &\ll \big|\mathcal{B}^{\mathcal{V}}\big| q^2 + \big|\mathcal{B}^{\mathcal{W}}\big| q^{3/2} + |\mathcal{B}| q \\ &\ll B^l q^2 + B^{(2-\gamma)l} q^{3/2} + B^{2l} q,\end{aligned}$$

by Lemma 2.3. Choosing

$$B = q^{1/l}$$

to equate the first and third terms above and

$$A = N/B = Nq^{-1/l},$$

we obtain from (4.11) the estimate

$$\begin{aligned}B(K, \boldsymbol{\alpha}, 1_{\mathcal{N}}) \ll_{k,\varepsilon} \ll q^{\varepsilon} \|\boldsymbol{\alpha}\|_1^{1-\frac{1}{l}} \|\boldsymbol{\alpha}\|_2^{\frac{1}{l}} M^{\frac{1}{2l}} N \left(\frac{(MN)^{-1}}{AB^{2l}} \Big(qB^{2l} + q^{1/2} B^{(3-\gamma)l} \Big) \right)^{\frac{1}{2l}} \\ \ll_{k,\varepsilon} q^{\varepsilon} \|\boldsymbol{\alpha}\|_1^{1-\frac{1}{l}} \|\boldsymbol{\alpha}\|_2^{\frac{1}{l}} M^{\frac{1}{2l}} N \left(\frac{q^{1+\frac{1}{l}}}{MN^2} + \frac{q^{\frac{3}{2}-\gamma+\frac{1}{l}}}{MN^2} \right)^{1/2l},\end{aligned}$$

assuming that (4.9) holds and that $A \geqslant 1$. Since $\gamma \geqslant 1/2$ (by Theorem 4.5), the second term on the right-hand side of the last inequality is smaller than the first. Together with (4.9), this leads to Theorem 4.3, and Theorem 1.3 follows by letting l get large.

5. Algebraic preliminaries

We collect in this section some definitions and statements of algebraic geometry that we will use later. Most are standard, but we include some proofs for completeness and by lack of a convenient reference.

Let $C_{\mathbf{F}_q}$ be a smooth and geometrically connected curve with smooth projective model S. The conductor of a constructible ℓ-adic sheaf $\mathcal{F}$ on C is defined by

$$\mathbf{c}(\mathcal{F}) = g(S) + \operatorname{rank}(\mathcal{F}) + |\operatorname{Sing}(\mathcal{F})| + \sum_{x \in \operatorname{Sing}(\mathcal{F})} \operatorname{Swan}_x(\mathcal{F}) + \dim H_c^0(C_{\overline{\mathbf{F}}_q}, \mathcal{F}),$$

where $g(S)$ is the genus of S, $\operatorname{Sing}(\mathcal{F})$ is the set of points of S where the middle-extension of $\mathcal{F}$ is not lisse and $\operatorname{Swan}_x(\mathcal{F})$ is the Swan conductor at x.

Let $C_{\mathbf{F}_q}$ be a curve (not necessarily smooth or irreducible). Let $(C_i)_{i\in I}$ be the geometrically irreducible components of $C_{\overline{\mathbf{F}}_q}$ and $\pi_i : \widetilde{C}_i \to C_i$ their canonical desingularization. We define the conductor of a constructible ℓ-adic sheaf $\mathcal{F}$ on $C_{\mathbf{F}_q}$ by

$$\mathbf{c}(\mathcal{F}) = \sum_{i\in I} \mathbf{c}\big(\pi_i^*(\mathcal{F}|C_i)\big) + \sum_{x\in C_{sing}} m_x(C),$$

where C_{sing} is the singular set of C and $m_x(C)$ the multiplicity of x as a singularity of C.

If $C_{\mathbf{F}_q}$ is a curve, f is a function on C and $\mathcal{F}$ an ℓ-adic sheaf on C, then

$$\mathbf{c}\left(\mathcal{F} \otimes \mathcal{L}_{f(x)}\right) \ll \mathbf{c}\left(\mathcal{L}_{f(x)}\right)^2 \mathbf{c}(\mathcal{F})^2, \tag{5.1}$$

where the implied constant is absolute.

We will use the following version of Deligne's Riemann Hypothesis over finite fields [4].

Proposition 5.1. *Let $\mathbf{F}_q$ be a finite field with q elements and let C be a curve over $\mathbf{F}_q$. Let $\mathcal{F}$ and $\mathcal{G}$ be constructible ℓ-adic sheaves on C which are mixed of weights $\leqslant 0$ and pointwise pure of weight 0 on a dense open subset. Suppose that the restriction of $\mathcal{F} \otimes \mathcal{G}^\vee$ to any geometrically irreducible component of C has no trivial summand. We then have*

$$\sum_{x\in C(\mathbf{F}_q)} t_{\mathcal{F}}(x;\mathbf{F}_q)\overline{t_{\mathcal{G}}(x;\mathbf{F}_q)} \ll \sqrt{q},$$

where the implied constant depend only on the conductors of $\mathcal{F}$ and of $\mathcal{G}$.

Proof. If C is smooth and geometrically connected, and $\mathcal{F}$ and $\mathcal{G}$ are geometrically irreducible middle-extensions, this is deduced from Deligne's results in [5, Lemma 3.5]; the extension to general $\mathcal{F}$ and $\mathcal{G}$ satisfying our assumptions is immediate. For a general smooth curve, one need only apply the bound to each component separately.

For a general curve, observe that the difference between the sum over C and the sum over a desingularization of C is the sum over the singular points of $t_{\mathcal{F}}(x; \mathbf{F}_q)\overline{t_{\mathcal{G}}(x; \mathbf{F}_q)}$ minus the sum over points of the desingularization lying over singular points of $t_{\mathcal{F}}(x; \mathbf{F}_q)\overline{t_{\mathcal{G}}(x; \mathbf{F}_q)}$. Since the size of both those sets of points may be bounded in terms of the sum of the multiplicities of singular points, and the value of $t_{\mathcal{F}}(x; \mathbf{F}_q)\overline{t_{\mathcal{G}}(x; \mathbf{F}_q)}$ at those points may be bounded in terms of the conductors, this contribution is also bounded in terms of the conductors. ◻

We will also use a criterion for a sheaf to be lisse that might be well-known but for which we do not know of a suitable reference.

Lemma 5.2. *Let* $\mathrm{Spec}(\mathcal{O})$ *be an open dense subset of the spectrum of the ring of integers in a number field and* $U \to \mathrm{Spec}(\mathcal{O})$ *a reduced scheme of finite type. Let* ℓ *be a prime number invertible in* $\mathcal{O}$*. Let* $r \geqslant 1$ *be an integer and let* $\mathcal{F}$ *be a constructible* ℓ*-adic sheaf on* U*.*

Assume that:

(1) *For any finite-field valued point* $\mathrm{Spec}(k) \to \mathrm{Spec}(\mathcal{O})$*, the sheaf* $\mathcal{F}_k$ *on* U_k *is lisse of rank* r*;*
(2) *For any finite-field valued point* $\mathrm{Spec}(k) \to \mathrm{Spec}(\mathcal{O})$*, any generic point* η *of* U_k*, and any* $s \in \Gamma(\mathrm{Spec}(\mathcal{O}^{et}_\eta), \mathcal{F})$*, if* s *is non-zero at the special point of the étale local ring* $\mathcal{O}^{et}_\eta$*, then it is non-zero at the generic point.*

Then $\mathcal{F}$ *is lisse on* U*.*

Proof. Let $x \in U_k \subset U$ and let s be a non-zero section of $\mathcal{F}$ over the étale local ring $\mathcal{O}^{et}_x$ at x. Since (the pullback of) $\mathcal{F}$ is lisse on $\mathcal{O}^{et}_{x,k}$ by Assumption (1), the generic point of $\mathcal{O}^{et}_{x,k}$ belongs to the support of s. Hence (the pullback of) s is non-zero at the special point of $\mathcal{O}^{et}_\eta$, which maps to the generic point $\mathcal{O}^{et}_{x,k}$ (for some generic point η of U_k). By Assumption (2), we deduce that the generic point of $\mathcal{O}^{et}_\eta$ belongs to the support of (the pullback of) s. Since this generic point maps to the generic point of $\mathcal{O}^{et}_x$, this means that the support of s contains the generic point of $\mathcal{O}^{et}_x$, hence because the support of s is closed, it is the whole $\mathrm{Spec}(\mathcal{O}^{et}_x)$.

Now let $(s_1, \dots, s_r)$ be a basis of the stalk $\mathcal{F}_x = \Gamma(\mathcal{O}^{et}_x, \mathcal{F})$. These sections define a morphism

$$\overline{\mathbf{Q}}_\ell^r \to \mathcal{F}_{\mathcal{O}^{et}_x}$$

whose induced map on stalks is, by the above, injective. By Assumption (1) and the fact that the rank of the stalk of a constructible ℓ-adic sheaf is a constructible function, the rank of the stalk of $\mathcal{F}$ at every point is r. Hence both stalks have the same dimension, thus the induced map on stalks is an isomorphism. This means that $\mathcal{F}$ is locally constant at x, and we conclude that $\mathcal{F}$ is lisse. ◻

6. Generalized Kloosterman sheaves

In this section, we summarize the basic properties of the generalized Kloosterman sheaves whose trace functions are the sums $\mathrm{Kl}_k(x;\boldsymbol{\chi},q)$. These were defined by Katz in [16, Theorem 4.1.1], building on Deligne's work [25, Sommes trig., Theorem 7.8]. They are special cases of the hypergeometric sheaves defined by Katz in [17, 8.2.1].

Throughout this section, we fix a prime number p, a prime number $\ell\neq p$, and we consider a finite field $\mathbf{F}_q$ of characteristic p with q elements and a non-trivial ℓ-adic additive character ψ of $\mathbf{F}_q$. We fix an integer $k\geqslant 2$ coprime to q, and a tuple $\boldsymbol{\chi}=(\chi_1,\ldots,\chi_k)$ of ℓ-adic characters of $\mathbf{F}_q^\times$. We denote by $\Lambda(\boldsymbol{\chi})$ (or Λ if $\boldsymbol{\chi}$ is understood) the product $\chi_1\cdots\chi_k$.

Proposition 6.1 (Generalized Kloosterman sheaves). *There exists a constructible $\overline{\mathbf{Q}}_\ell$-sheaf $\mathcal{K}\ell=\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ on $\mathbf{P}^1_{\mathbf{F}_q}$, called a* generalized Kloosterman sheaf*, with the following properties:*

(1) *For any $d\geqslant 1$ and any $x\in\mathbf{G}_m(\mathbf{F}_{q^d})$, we have*

$$\begin{aligned}
t_{\mathcal{K}\ell}\big(x;\mathbf{F}_{q^d}\big)&=\mathrm{Kl}_k\big(x;\boldsymbol{\chi},\mathbf{F}_{q^d}\big)\\
&=\frac{(-1)^{k-1}}{q^{d(k-1)/2}}\sum_{x_1\cdots x_k=x}\chi_1\big(N_{\mathbf{F}_{q^d}/\mathbf{F}_q}x_1\big)\cdots\chi_k(N_{\mathbf{F}_{q^d}/\mathbf{F}_q}x_k)\psi\Big(\mathrm{Tr}_{\mathbf{F}_{q^d}/\mathbf{F}_q}(x_1+\cdots+x_k)\Big);
\end{aligned}$$

(2) *The sheaf $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ is lisse of rank k on $\mathbf{G}_m$;*
(3) *On $\mathbf{G}_m$, the sheaf $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ is geometrically irreducible and pure of weight 0;*
(4) *The sheaf $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ is tamely ramified at 0, and its $I(0)$-decomposition is*

$$\bigoplus_{\chi\in\boldsymbol{\chi}}\mathcal{L}_\chi\otimes J\big(n_\chi\big),$$

where $J(n)$ is a unipotent Jordan block of size n, and n_χ is the multiplicity of χ in $\boldsymbol{\chi}$;
(5) *The sheaf $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ is wildly ramified at ∞, with a single break equal to $1/k$, and with Swan conductor equal to 1;*
(6) *The stalks of $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ at 0 and ∞ both vanish;*
(7) *If $\gamma\in\mathrm{PGL}_2(\overline{\mathbf{F}}_q)$ is non-trivial, there does not exist a rank 1 sheaf $\mathcal{L}$ such that we have a geometric isomorphism*

$$\gamma^*\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})\simeq\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})\otimes\mathcal{L}$$

over a dense open set;
(8) *The conductor of $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ is $k+3$.*

Proof. Let $j:\mathbf{G}_m\to\mathbf{P}^1$ be the open inclusion. We define

$$\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})=j_!\mathrm{Kl}(\psi;\boldsymbol{\chi};1,\ldots,1)\left(\frac{n-1}{2}\right),$$

where the sheaf on the right-hand side is the lisse sheaf on $\mathbf{G}_m$ defined by Katz in [16, 4.1.1]. We also have a formula in terms of hypergeometric sheaves, namely

$$\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}) = j_!\mathcal{H}_1(!, \psi; \boldsymbol{\chi}, \emptyset)\left(\frac{n-1}{2}\right)$$

(see [17, 8.4.3]). Assertions (1) and (2) are, respectively, assertions (2) and (1) of [16, 4.1.1]. Assertion (3) results from the identification with hypergeometric sheaves and [17, Theorem 8.4.2 (1), (4)].

Assertions (4) and (5)) are given in [17, Theorem 8.4.2 (6)]. Assertion (6) is clear from the definition as an extension by zero of a sheaf on $\mathbf{G}_m$.

Finally, (7) is a special case of [10, Proposition 3.6 (2)], and (8) follows from the definition of the conductor and the previous statements. □

All parts of Definition 2.1, including the definition of Property CGM and Property NIO, make sense for tuples of ℓ-adic characters of $\mathbf{F}_q^\times$. When we wish to emphasize the base finite field, we will speak of Property CGM or NIO over $\mathbf{F}_q$. The names CGM and NIO are justified by the following theorem of Katz.

Theorem 6.2 (Katz). *Assume that $k \geqslant 2$, that $p > 2k+1$ and that $\boldsymbol{\chi}$ is not Kummer induced. Let G be the geometric monodromy group of $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$. We then have $G^0 = G^{0,der}$, the derived group. Moreover*

(1) *If k is odd, then $G^0 = G^{0,der} = \mathrm{SL}_k$;*
(2) *If k is even, then $G^0 = G^{0,der}$ is either*
   - SO_k *if $\boldsymbol{\chi}$ is self-dual and symmetric;*
   - Sp_k *if $\boldsymbol{\chi}$ is self-dual and alternating;*
   - SL_k *if $\boldsymbol{\chi}$ is not self-dual.*

Finally, if $\boldsymbol{\chi}$ has CGM*, then $G = G^0$ is either SL_k or Sp_k.*

Proof. The claims about G^0 are proved by Katz in [17, Theorem 8.11.3 and Corollary 8.11.2.1].

To evaluate G, note that when $G^0 = \mathrm{SL}_k$, G is contained in GL_k. To show $G = G^0$, it suffices to show the determinant is trivial. But the determinant character is $\mathcal{L}_\Lambda$ by [17, Lemma 8.11.6], and we have assumed Λ trivial.

If $G^0 \neq \mathrm{SL}_k$ then k is even and $\boldsymbol{\chi}$ is self-dual. Let ξ be the dualizing character (Definition 2.1). Under the assumptions $\Lambda = 1$ and $\xi = 1$, we always have $\Lambda = \xi^{k/2}$, so the self-duality is alternating. Thus $G^0 = \mathrm{Sp}_k$, hence G is contained in GSp_k, and it suffices to show that the similitude character is trivial, *i.e.*, that $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ is actually self-dual and not just self-dual up to a twist. This follows from [17, Theorem 8.8.1]. Reviewing Definition 2.1, we obtain the desired statements. □

The need to sometimes increase the base field is justified by the following lemma that will allow us to work with tuples satisfying the weaker CGM Property.

Lemma 6.3. *Assume that $\boldsymbol{\chi}$ has* NIO. *Then there exists an ℓ-adic character χ_0, possibly over a finite extension $\mathbf{F}_{q^\nu}$ of $\mathbf{F}_q$, such that the tuple $\chi_0\boldsymbol{\chi}$ has* CGM *over $\mathbf{F}_{q^\nu}$.*

Proof. If k is even and $\boldsymbol{\chi}$ is self-dual alternating, take χ_0 to be the inverse of a square root of the duality character. Otherwise, take χ_0 to be the inverse of a k-th root of Λ. ☐

For convenience, we will most often simply denote $\mathcal{K}\ell_k = \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ since we assume that ψ and $\boldsymbol{\chi}$ are fixed.

The next lemma computes precisely the local monodromy of $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ at ∞.

Lemma 6.4. *Assume $p > k \geqslant 2$. Denote by $\widetilde{\psi}$ the additive character $x \mapsto \psi(kx)$ of $\mathbf{F}_q$. Then, as representations of the inertia group $I(\infty)$ at ∞, there exists an isomorphism*

$$\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}) \simeq [x \mapsto x^k]_* \left(\mathcal{L}_{\chi_{(2)}^{k+1}} \otimes \mathcal{L}_\Lambda \otimes \mathcal{L}_{\widetilde{\psi}}\right),$$

where $\chi_{(2)}$ is the unique non-trivial character of order 2 of $\mathbf{F}_q^\times$.

Proof. This follows from a more precise result of L. Fu [13, Proposition 0.8] (who describes the local representations of the decomposition group). ☐

7. Sheaves and statement of the target theorem

As in the previous section, we fix a prime number p, a prime number $\ell \neq p$, and we consider a finite field $\mathbf{F}_q$ of characteristic p with q elements and a non-trivial ℓ-adic additive character ψ of $\mathbf{F}_q$. We assume that $p > 2k + 1$.

Let $\boldsymbol{\chi}$ be a k-tuple of ℓ-adic characters of $\mathbf{F}_q^\times$. We define

$$\mathcal{F} = \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}),$$

a constructible ℓ-adic sheaf on $\mathbf{A}^1_{\mathbf{F}_q}$. *In this section we impose no further conditions on $\boldsymbol{\chi}$.*

Fix $l \geqslant 2$. For $1 \leqslant i \leqslant 2l$, let $f_i = s(r + b_i)$ on $\mathbf{A}^{2+2l}$ with coordinates $(r, s, \boldsymbol{b})$.

We now define the "sum-product" sheaf

$$\mathcal{K}(\boldsymbol{\chi}) = \bigotimes_{1\leqslant i\leqslant l} f_i^*\mathcal{F} \otimes f_{i+l}^*\mathcal{F}^\vee$$

on $\mathbf{A}^{2+2l}_{\mathbf{F}_q}$.

Let $V/\mathbf{Z}$ be the open subset of $\mathbf{A}_{\mathbf{Z}}^{2+2l}$ where $s(r+b_i) \neq 0$ for all i, so that $\mathcal{K}(\chi)$ is lisse on $V_{\mathbf{F}_q}$ for all q. Let $\pi : \mathbf{A}^{2+2l} \to \mathbf{A}^{1+2l}$ be the projection $(r, s, \boldsymbol{b}) \mapsto (r, \boldsymbol{b})$ (defined over $\mathbf{Z}$). We define

$$\mathcal{R}(\boldsymbol{\chi}) = R^1\pi_!\mathcal{K}(\boldsymbol{\chi}),$$

a constructible ℓ-adic sheaf on $\mathbf{A}_{\mathbf{F}_q}^{1+2l}$.

We will most often drop the dependency on $\boldsymbol{\chi}$ in these notation and write $\mathcal{K} = \mathcal{K}(\boldsymbol{\chi})$ and $\mathcal{R} = \mathcal{R}(\boldsymbol{\chi})$.

We define the diagonal variety $\mathcal{V}^{\Delta}$ by the condition

$$\mathcal{V}^{\Delta} = \left\{ \boldsymbol{b} \in \mathbf{A}^{2l} \;\mid\; \text{for all } i, \text{there exists } j \neq i \text{ such that } b_i = b_j \right\}.$$

Note that $\mathcal{V}^{\Delta}$ does not depend on the tuple of characters considered.

Lemma 7.1. *Outside $\mathcal{V}^{\Delta}$, we have $R^0\pi_!\mathcal{K} = R^2\pi_!\mathcal{K} = 0$.*

Proof. This is very similar to [21, Lemma 4.1 (2)]. By the proper base change theorem, the stalk of $R^i\pi_!\mathcal{K}$ at $x = (r, \boldsymbol{b}) \in \mathbf{A}^{1+2l}$ is

$$H_c^i\left(\mathbf{A}^1_{\overline{\mathbf{F}}_q}, \bigotimes_{i=1}^{l} [\times(r+b_i)]^*\mathcal{F} \otimes [\times(r+b_{i+2})]^*\mathcal{F}^{\vee}\right),$$

where s is the coordinate on $\mathbf{A}^1$. This cohomology group vanishes for $i = 0$ and any x, and it vanishes for $i = 2$ and $x \notin \mathcal{V}^{\Delta}$ by [10, Theorem 1.5]. ☐

We now compute the local monodromy at infinity of the sheaf $\mathcal{K}$. For any additive character ψ, we denote by $\widetilde{\psi}$ the character $x \mapsto \psi(kx)$.

Lemma 7.2.

(1) *Let $r \in \mathbf{F}_q$ and $\boldsymbol{b} \in \mathbf{F}_q^{2l}$ be such that $r + b_i \neq 0$ for all i. Let $(r+b_i)^{1/k}$ be a fixed k-th root of $r + b_i$ in $\overline{\mathbf{F}}_q$. Define signs $\varepsilon_i = 1$ for $1 \leqslant i \leqslant l$ and $\varepsilon_i = -1$ for $l+1 \leqslant i \leqslant 2l$.*
The local monodromy at $s = \infty$ of $\mathcal{K}_{r,\boldsymbol{b}}$ is isomorphic to the local monodromy at $s = \infty$ of the sheaf

$$\bigoplus_{(\zeta_2,\ldots,\zeta_{2l}) \in \boldsymbol{\mu}_k^{2l-1}} \mathcal{L}_{\widetilde{\psi}}\left(\left((r+b_1)^{1/k} + \sum_{i=2}^{2l} \varepsilon_i\zeta_i(r+b_i)^{1/k}\right)s^{1/k}\right),$$

where $\boldsymbol{\mu}_k$ is the group of k-th roots of unity in $\overline{\mathbf{F}}_q$.

(2) *Let K be a field of characteristic $p \nmid k$, and let $r \in K$ and $\boldsymbol{b} \in K^{2l}$ be such that $r + b_i \neq 0$ for all i. Assume that K contains all k-th roots $(1 + b_i/r)^{1/k}$ of $1 + b_i/r$ for all i. Let ψ be a non-trivial ℓ-adic additive character and let $\boldsymbol{\chi}$ be a k-tuple of multiplicative characters of a finite subfield of K. The local monodromy at $t = \infty$ of the lisse sheaf*

$$\widetilde{\mathcal{K}} = \bigotimes_{1 \leqslant i \leqslant l} \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(t(1 + b_i/r)) \otimes \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(t(1 + b_{i+l}/r))^{\vee}$$

on $\mathbf{G}_{m,K}$ is isomorphic to the local monodromy at $t = \infty$ of the sheaf

$$\bigoplus_{(\zeta_2,\ldots,\zeta_{2l}) \in \boldsymbol{\mu}_k^{2l-1}} \mathcal{L}_{\widetilde{\psi}} \left(\left((t(1 + b_1/r))^{1/k} + \sum_{i=2}^{2l} \varepsilon_i \zeta_i (t(1 + b_i/r))^{1/k} \right) \right).$$

Proof. Since Lemma 6.4 has the same form as [21, Lemma 4.9], up to the additional factor $\mathcal{L}_{\Lambda}$, the first assertion may be proved exactly like [21, Lemma 4.16 (1)] (with $\lambda = 0$ there), replacing throughout the tensor product

$$\bigotimes_{i=1}^{2} [\times(r + b_i)]^* \mathcal{K}\ell_k \otimes [\times(r + b_{i+2})]^* \mathcal{K}\ell_k^{\vee}$$

by

$$\bigotimes_{i=1}^{l} [\times(r + b_i)]^* \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}) \otimes [\times(r + b_{i+2})]^* \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})^{\vee}$$

(note that the factors involving Λ cancel-out at the end). The second statement is proved in the same manner. ☐

Let $\widetilde{Z} \subset \mathbf{A}_{\mathbf{Z}}^{1+2l}$ be the image of

$$\widetilde{\mathcal{Z}} = \left\{ (r, \boldsymbol{b}, \boldsymbol{x}) \in \mathbf{A}^{1+4l} \mid x_i^k = r + b_i \text{ for } 1 \leqslant i \leqslant 2k,\ \sum_{i=1}^{l} x_i = \sum_{i=l+1}^{2l} x_i \right\} \subset \mathbf{A}_{\mathbf{Z}}^{1+4l} \quad (7.1)$$

under the projection onto $(r, \boldsymbol{b})$. Let

$$Z = \widetilde{Z} \cup \bigcup_{1 \leqslant i \leqslant 2l} \{r = -b_i\}.$$

Let U be the complement of Z. We emphasize that $\widetilde{Z}$, Z and U are defined over $\mathbf{Z}$, and independent of $\boldsymbol{\chi}$.

Lemma 7.3. *The subscheme $\widetilde{Z}$ of $\mathbf{A}^{2l+1}$ is closed and irreducible, and $\mathcal{R}$ is lisse on $U_{\mathbf{F}_q}$.*

Proof. This is analogue to [21, Lemma 4.26, (1) and (2)], so we will be brief.[1] The projection $(r, \boldsymbol{b}, \boldsymbol{x}) \mapsto (r, \boldsymbol{b})$ from the subscheme

$$\mathcal{Z}' = \left\{ (r, \boldsymbol{b}, \boldsymbol{x}) \in \mathbf{A}^{1+4l} \mid x_i^k = r + b_i \text{ for } 1 \leqslant i \leqslant 2k \right\}$$

to $\mathbf{A}^{1+2l}$ is finite, since the domain is defined by adjoining the coordinates $(x_1, \ldots, x_{2l})$ to $\mathbf{A}^{1+2l}$, and each satisfies a monic polynomial equation. Thus the closed subscheme $\widetilde{\mathcal{Z}}$ defined by (7.1) is also finite over $\mathbf{A}^{1+2l}$, and its image $\widetilde{Z}$ is closed. Moreover, the subscheme (7.1) is the divisor in $\mathcal{Z}'$ given by the equation

$$\sum_{i=1}^{l} x_i = \sum_{i=l+1}^{2l} x_i .$$

In particular, this subscheme, and consequently its projection $\widetilde{Z}$, is irreducible.

To prove that $\mathcal{R}$ is lisse on $U_{\mathbf{F}_q}$, we use Deligne's semicontinuity theorem [22]. The sheaf $\mathcal{K}$ is lisse on the complement of the divisors given by the equations $r = -b_i$ and $s = 0$ in $\mathbf{A}^{2+2l}$. We compactify the s-coordinate by $\mathbf{P}^1$ and work on

$$X = \left(\mathbf{A}^1 \times \mathbf{P}^1 \times \mathbf{A}^{2l}\right) \cap \{(r, s, \boldsymbol{b}) \mid (r, \boldsymbol{b}) \in U\}.$$

By extending by 0, we view $\mathcal{K}$ as a sheaf on X which is lisse on the complement in X of the divisors $s = 0$ and $s = \infty$ (because U is contained in the complement of the divisors $r = -b_i$ and thus X is as well). Let

$$\pi^{(2)} : X \longrightarrow U$$

denote the projection $(r, s, \boldsymbol{b}) \mapsto (r, \boldsymbol{b})$. Then $\pi^{(2)}$ is proper and smooth of relative dimension 1 and $\mathcal{R}|U = R^1\pi_*^{(2)}\mathcal{K}$.

Since the restrictions of $\mathcal{K}$ to the divisors $s = \infty$ and $s = 0$ are zero, this sheaf is the extension by zero from the complement of those divisors to the whole space of a lisse sheaf. Deligne's semicontinuity theorem [22, Corollary 2.1.2] implies that the sheaf $\mathcal{R}$ is lisse on U if the Swan conductor is constant on each of these two divisors. By Proposition 6.1, the generalized Kloosterman sheaf has tame ramification on $s = 0$, hence any tensor product of generalized Kloosterman sheaves (such as $\mathcal{K}$) has tame ramification, hence Swan conductor 0, on $s = 0$. On the other hand, Lemma 7.2 gives a formula for the local monodromy representation of $\mathcal{K}$ at $s = \infty$ as a sum of pushforward of representations from the tame covering $x \mapsto x^k$. Since the Swan conductor is additive and since the Swan conductor is invariant under pushforward by a tame covering (see, *e.g.*, [16, 1.13.2]), it follows that

$$\begin{aligned}&\mathrm{Swan}_\infty(\mathcal{K}_{r,\boldsymbol{b}})\\ &= \sum_{\zeta_2,\ldots,\zeta_{2l} \in \boldsymbol{\mu}_k} \mathrm{Swan}_\infty\left(\mathcal{L}_\psi\left(\left(\left((r+b_1)^{1/k} + \sum_{i=2}^{2l} \varepsilon_i \zeta_i (r+b_i)^{1/k}\right) s^{1/k}\right)\right)\right) = k^{2l-1}\end{aligned}$$

by definition of U, since the Swan conductor of $\mathcal{L}_{\psi(at)}$ is 1 for $a \neq 0$. ☐

[1] To avoid confusion, note that what is called Z in [21] is not the analogue of what is called Z here.

Lemma 7.4. *The subscheme Z is a hypersurface in $\mathbf{A}^{1+2l}_{\mathbf{Z}}$. It is defined by the vanishing of a polynomial P in $\mathbf{Z}[r, b_1, \dots, b_{2l}]$ such that, for any fixed $\boldsymbol{b} \notin \mathcal{V}^{\Delta}$, the polynomial $P_{\boldsymbol{b}} = P(\cdot, \boldsymbol{b})$ of the variable r is not zero.*

Proof. First we check that $\widetilde{Z}$ is a hypersurface in $\mathbf{A}^{1+2l}_{\mathbf{Z}}$. It is the projection of the closed subscheme

$$\widetilde{\mathcal{Z}} = \left\{ (r, \boldsymbol{b}, \boldsymbol{x}) \in \mathbf{A}^{1+4l} \mid x_i^k = r + b_i \text{ for } 1 \leqslant i \leqslant 2l,\ \sum_{i=1}^{l} x_i = \sum_{i=l+1}^{2l} x_i \right\} \subset \mathbf{A}^{1+4l}.$$

This closed subscheme is pure of dimension $2l$, since the first $2l$ equations let us eliminate the variables b_i and the last equation is nontrivial. The projection $\widetilde{\mathcal{Z}} \to \widetilde{Z}$ is finite (as already observed in the proof of the previous lemma) and hence $\widetilde{Z}$ is a closed subscheme of $\mathbf{A}^{2l+1}$ that is pure of dimension $2l$, *i.e.*, a hypersurface. Since Z is the union of $\widetilde{Z}$ and the hyperplanes with equation $r + b_i = 0$, it is also a hypersurface.

Let $P \in \mathbf{Z}[r, \boldsymbol{b}]$ be a polynomial whose vanishing set is $\widetilde{Z}$. Suppose $\boldsymbol{b}$ is such that $P_{\boldsymbol{b}}$ is the zero polynomial in the variable r, *i.e.*, such that the projection $\widetilde{\mathcal{Z}}_{\boldsymbol{b}} \to \mathbf{A}^1$ given by $(r, \boldsymbol{x}) \mapsto r$ is surjective.

The scheme $C \subset \mathbf{A}^{1+2l}$ given by the equations

$$x_i^k = r + b_i \qquad 1 \leqslant i \leqslant 2k$$

is a curve and the projection $C \to \mathbf{A}^1$ given by $(r, \boldsymbol{x}) \mapsto r$ is finite. The fiber $\widetilde{\mathcal{Z}}_{\boldsymbol{b}}$ is the intersection of C and the hyperplane

$$\sum_{i=1}^{l} x_i = \sum_{i=l+1}^{2l} x_i,$$

so that $P_{\boldsymbol{b}} = 0$ if and only if the function

$$F = \sum_{i=1}^{l} x_i - \sum_{i=l+1}^{2l} x_i$$

vanishes on an irreducible component of C.

If we assume that $\boldsymbol{b} \notin \mathcal{V}^{\Delta}$ then by definition there exists some i such that $b_i \neq b_j$ for all $j \neq i$. Locally on $\mathbf{A}^1$ with coordinate r near the point $r = -b_i$, the covering maps $x_k^j = r + b_j$ for $j \neq i$ are étale, so the functions x_j (on the curve C) "belong" to the étale local ring R of $\mathbf{A}^1$ at $-b_i$. The function x_i, however, does not belong to R, hence the function F is non-zero in an algebraic closure of the fraction field of R, which is also an algebraic closure of the function field of any irreducible component of C. This concludes the proof. □

Definition 7.5. The sheaf $\mathcal{R}^*$ on $U_{\mathbf{F}_q}$ is the maximal quotient of the sheaf $\mathcal{R}|U_{\mathbf{F}_q}$ that is pure of weight 1 (see [4]).

Define $f\colon U \to \mathbf{A}^{2l}$ over $\mathbf{Z}$ by $(r, \boldsymbol{b}) \mapsto \boldsymbol{b}$.

Below, by $\mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{G})$, where $\mathcal{G}$ is a lisse sheaf on $V_{\mathbf{F}_q,\boldsymbol{b}}$, we mean the $\pi_1(V_{\mathbf{F}_q,\boldsymbol{b}}\times \overline{\mathbf{F}}_q)$-homomorphisms, etc.

Let $\boldsymbol{b} \in \mathbf{A}^{2l}_{\mathbf{F}_q}$ and let $\kappa(\boldsymbol{b})$ be the residue field of $\boldsymbol{b}$. Since $\mathcal{R}_{\boldsymbol{b}} = R^1\pi_!\mathcal{K}_{\boldsymbol{b}}$ by the proper base change theorem, there exists a natural $\mathrm{Gal}(\overline{\kappa(\boldsymbol{b})}/\kappa(\boldsymbol{b}))$-equivariant morphism

$$\mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}}) \longrightarrow \mathrm{End}_{U_{\boldsymbol{b}}}(\mathcal{R}_{\boldsymbol{b}}).$$

Since every $V_{\boldsymbol{b}}$-endomorphism of $\mathcal{K}_{\boldsymbol{b}}$ preserves the weight filtration, the image of this morphism is contained in the subring of endomorphisms of $\mathcal{R}_{\boldsymbol{b}}$ that preserve the weight filtration, and hence we have an induced morphism

$$\theta_{\boldsymbol{b}}\colon\ \mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}}) \longrightarrow \mathrm{End}_{U_{\boldsymbol{b}}}(\mathcal{R}^*_{\boldsymbol{b}}),$$

which by construction is still Frobenius-equivariant.

In the next definition, we already describe the subvariety $\mathcal{W}$ of Theorem 4.5; in particular, we see that it is independent of the tuple of characters $\boldsymbol{\chi}$, since this is the case for X_∞ and Z. The difficulty will be to prove that it satisfies the required properties.

Definition 7.6. We denote $X_\infty = \mathbf{A}^{2l} - \mathcal{V}^\Delta$, and for any integer $j \geqslant 0$, we let

$$X_j = \{\boldsymbol{b} \in X_\infty \mid |Z_{\boldsymbol{b}}| \leqslant j\}.$$

We define $\mathcal{W}$ to be the union of $\mathcal{V}^\Delta$ and of all irreducible components of all X_j of dimension strictly less than $(3l+1)/2$.

By definition, we therefore have the codimension bound

$$\mathrm{codim}(\mathcal{W}) \geqslant \frac{l-1}{2}. \tag{7.2}$$

Our main geometric goal will be to prove the following result:

Theorem 7.7. *Assume that $\boldsymbol{\chi}$ has* NIO. *If p is large enough, depending only on k and l, then the natural morphism $\theta_{\boldsymbol{b}}$ is an isomorphism for all $\boldsymbol{b} \in \mathbf{A}^{2l}(\mathbf{F}_q) - \mathcal{W}(\mathbf{F}_q)$. Furthermore, each geometrically irreducible component of $\mathcal{R}^*_{\boldsymbol{b}}$ has rank greater than one.*

The basic strategy to be used is as follows:

(1) We show that for q large enough and for $\boldsymbol{b} \in \mathbf{A}^{2l}(\mathbf{F}_q)$ outside an explicit subscheme $\mathcal{W}_1$ of codimension $l-1$, the natural morphism $\theta_{\boldsymbol{b}}$ is injective. This reduces the target statement to a proof that the dimensions $\mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}})$ and $\mathrm{End}_{U_{\boldsymbol{b}}}(\mathcal{R}^*_{\boldsymbol{b}})$ are equal;

(2) We show that, when these dimensions agree for the generic point of an irreducible component of a stratum, this implies the corresponding statement on the whole irreducible component;
(3) Finally, we prove the target theorem at the generic point of an irreducible component of a stratum with dimension $> (3l+1)/2$.

The most difficult part is the last one. This we prove by showing the strata can be covered by the vanishing sets of equations of a certain type in products of curves. Using this description, and a variant of Katz's Diophantine criterion for irreducibility, we show that the dimension of the space of endomorphisms of $\mathcal{K}$ is equal to that of the space of endomorphisms of $\mathcal{R}$ that are invariant under the Galois group of the function field of this cover. Finally, by a vanishing cycles argument, we show that the Galois group in fact acts trivially.

Remark 7.8. We have defined U, the stratification X_j, and $\mathcal{W}$ as objects over the integers rather than over a finite field $\mathbf{F}_q$. This is used in a few different places: first, when comparing the generic point and the special point of a stratum, we use a tameness property of the sheaf $\mathcal{R}$, which we verify by showing that the sheaf is defined over the integers. Second, when describing the defining equations of the strata, at one point we make a large characteristic assumption. Third, we need the set $\mathcal{W}$ to be uniform in q to allow us to apply Lemma 2.3.

8. Integrality

We fix an integer $n \geqslant 1$ and an integer $k \geqslant 2$. Let ℓ be a prime number. We denote in this section $S = \operatorname{Spec}(\mathbf{Z}[\boldsymbol{\mu}_n, 1/n\ell])$. For any ℓ-adic character $\widetilde{\chi}$ of $\boldsymbol{\mu}_n$, we have an associated lisse ℓ-adic sheaf $\mathcal{L}_{\widetilde{\chi}}$ over S defined by Kummer theory. If $\mathbf{F}_q$ is a residue field of S of characteristic $p \nmid n\ell$, so that $q \equiv 1 \bmod n$, then there is a natural isomorphism between the group of ℓ-adic characters $\widetilde{\chi}$ of $\boldsymbol{\mu}_n$ and the group of ℓ-adic characters χ of order dividing n of $\mathbf{F}_q^\times$, such that $\chi(x) = \widetilde{\chi}(\xi)$, where ξ is the n-th root of unity in $\mathbf{Z}[\boldsymbol{\mu}_n, 1/n\ell]$ mapping to $x^{(q-1)/n}$. We then have a natural isomorphism $\mathcal{L}_{\widetilde{\chi},\mathbf{F}_q} = \mathcal{L}_\chi$ of ℓ-adic sheaves.

Proposition 8.1. *Let $\widetilde{\boldsymbol{\chi}}$ be a k-tuple of characters of $\boldsymbol{\mu}_n$. There exists an ℓ-adic sheaf $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ on $\mathbf{A}_S^{1+2l}$, lisse on U_S, with the following property: for any prime $p \nmid \ell n$, for any finite field $\mathbf{F}_q$ of characteristic p which is a residue field of a prime ideal in $\mathbf{Z}[\boldsymbol{\mu}_n, 1/n\ell]$, for any non-trivial additive character ψ of $\mathbf{F}_q$, we have*

$$\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})|\mathbf{A}_{\mathbf{F}_q}^{1+2l} = \mathcal{R}(\boldsymbol{\chi}),$$

where $\boldsymbol{\chi}$ is the k-tuple of ℓ-adic characters of $\mathbf{F}_q^\times$ corresponding to $\widetilde{\boldsymbol{\chi}}$.

Proof. We will first construct a sheaf $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ over S with the desired specialization property, and we will then check that the sheaf thus defined is lisse on U_S. The existence statement is a fairly straightforward generalization of [21, Lemma

4.27], but we give full details since the precise construction is needed to check the lisseness assertion.

Let $X_1 \subset \mathbf{G}_m^{k+1}$ be the subscheme over S with equation

$$x_1 \cdots x_k = t$$

and let

$$f_1 \,:\, X_1 \longrightarrow \mathbf{A}^1$$

be the projection $(x_1, \ldots, x_k, t) \mapsto t$. Let X_2 be the subscheme of $\mathbf{G}_m^{2lk} \times \mathbf{A}^{2+2l}$ over S defined by the equations

$$\prod_{j=1}^{k} x_{i,j} = s(r + b_i), \qquad 1 \leqslant i \leqslant 2l,$$

and let $f_2 \,:\, X_2 \longrightarrow \mathbf{A}^{1+2l}$ be the projection

$$f_2(x_{1,1}, \ldots, x_{2l,k}, r, s, \boldsymbol{b}) = (r, \boldsymbol{b}).$$

Let further $X \subset X_2$ be the closed subscheme over S defined by the equation $x_{1,1} = 1$. The morphism

$$\mathbf{G}_m \times X \to X_2$$

defined by

$$\big(t, x_{1,1}, \ldots, x_{2l,k}, r, s, \boldsymbol{b}\big) \mapsto \big(tx_{1,1}, \ldots, tx_{2l,k}, r, t^k s, \boldsymbol{b}\big)$$

is an isomorphism, with inverse given by

$$\big(x_{1,1}, \ldots, x_{2l,k}, r, s, \boldsymbol{b}\big) \mapsto \left(x_{1,1}, 1, \frac{x_{1,2}}{x_{1,1}}, \ldots, \frac{x_{2l,k}}{x_{1,1}}, r, \frac{s}{t^k}, \boldsymbol{b}\right).$$

Let now $p \nmid n\ell$ be a prime and $\mathbf{F}_q$ a finite field of characteristic p that is a residue field of a prime ideal in S. Let ψ be a non-trivial additive character of $\mathbf{F}_q$. We have an isomorphism

$$\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})\left(\frac{1-k}{2}\right)[1-k] \simeq Rf_{1,!}\mathcal{L}_\psi(x_1 + \cdots + x_k) \otimes \bigotimes_{i=1}^{k} \mathcal{L}_{\chi_i}(x_i)$$

of sheaves on $\mathbf{A}^1_{\mathbf{F}_q}$. By definition and Lemma 7.1, it follows that

$$\mathcal{R}(\boldsymbol{\chi}) = R^{2l(k-1)+1} f_{2,!}\left(\mathcal{L}_\psi\left(\sum_{j=1}^{k}\left(\sum_{i=1}^{l} x_{i,j} - \sum_{i=1}^{l} x_{l+i,j}\right)\right)\right.$$
$$\left.\otimes \bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l} \mathcal{L}_{\chi_j}\big(x_{i,j}/x_{l+i,j}\big)\right).$$

We now translate this by "transport of structure" to $\mathbf{G}_m \times X \simeq X_2$. First, we have $f_2 = f \circ p_2$ where p_2 is the projection $\mathbf{G}_m \times X \to X$. Next, let $f : X \to \mathbf{A}^{1+2l}$ be the projection onto $(r, \boldsymbol{b})$, and let $g : X \to \mathbf{A}^1$ be defined by

$$g(x_{1,1}, \ldots, x_{2l,k}, r, s, \boldsymbol{b}) = \sum_{j=1}^{k} \left(\sum_{i=1}^{l} x_{i,j} - \sum_{i=1}^{l} x_{l+i,j} \right).$$

Let g' be the function

$$g' = \sum_{j=1}^{k} \left(\sum_{i=1}^{l} x_{i,j} - \sum_{i=1}^{l} x_{l+i,j} \right)$$

on X_2. Then g' corresponds to tg under the isomorphism $X_2 \simeq \mathbf{G}_m \times X$. Moreover, the sheaves $\mathcal{L}_{\chi_j}(x_{i,j}/x_{l+i,j})$ are transported to $\mathcal{L}_{\chi_j}(x_{i,j}/x_{l+i,j})$ under this isomorphism (since both variables involved are multiplied by t). We conclude that

$$\mathcal{R}(\boldsymbol{\chi})[-2l(k-1)-1] \simeq R(f \circ p_2)_! \left(\mathcal{L}_\psi(tg) \otimes \bigotimes_{j=1}^{k} \bigotimes_{i=1}^{l} \mathcal{L}_{\chi_j}\big(x_{i,j}/x_{l+i,j}\big) \right)$$

on $\mathbf{A}^{1+2l}_{\mathbf{F}_q}$.

We can now apply the strategy of [21, Lemma 4.23]. By the projection formula, we have

$$\begin{aligned} &Rp_2!\left(\mathcal{L}_\psi(tg) \otimes \bigotimes_{j=1}^{k} \bigotimes_{i=1}^{l} \mathcal{L}_{\chi_j}\big(x_{i,j}/x_{l+i,j}\big) \right) \\ &= \left(\bigotimes_{j=1}^{k} \bigotimes_{i=1}^{l} \mathcal{L}_{\chi_j}\big(x_{i,j}/x_{l+i,j}\big) \right) \otimes Rp_{2!}\mathcal{L}_\psi(tg) \end{aligned}$$

and $Rp_{2!}\mathcal{L}_\psi(tg)$ is the pullback along g of the Fourier transform of the extension by zero of the constant sheaf on $\mathbf{G}_{m,\mathbf{F}_q}$, which is $(Ru_*\overline{\mathbf{Q}}_\ell[-1])_{\mathbf{F}_q}$ for $u : \mathbf{G}_m \to \mathbf{A}^1$ the inclusion.

We then define the sheaf

$$\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}}) = R^{2l(k-1)} f_! \left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big) \otimes \bigotimes_{j=1}^{k} \bigotimes_{i=1}^{l} \mathcal{L}_{\widetilde{\chi}_j}\big(x_{i,j}/x_{l+i,j}\big) \right)$$

over S. The preceeding computation gives an isomorphism $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})_{\mathbf{F}_q} \simeq \mathcal{R}(\boldsymbol{\chi})$ over $\mathbf{F}_q$.

Furthermore, since the complex

$$Rf_! \left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big) \otimes \bigotimes_{j=1}^{k} \bigotimes_{i=1}^{l} \mathcal{L}_{\chi_j}\big(x_{i,j}/x_{l+i,j}\big) \right),$$

is supported in degree $2l(k-1)$ over $U_{\mathbf{F}_q}$ for all $\mathbf{F}_q$, the corresponding complex

$$Rf_!\left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j}\big(x_{i,j}/x_{l+i,j}\big)\right)$$

is supported in a single degree on S.

We will now check that $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ is lisse on U_S. By the specialization property and Lemma 7.3, we know that $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ is lisse on $U_{\mathbf{F}_q}$ for any residue field $\mathbf{F}_q$ of characteristic $p\nmid \ell n$, and that it has constant rank. Because it is a constructible sheaf, its rank is a constructible function, and hence it has the same rank everywhere on U_S.

Write $\mathcal{R}^{\mathrm{univ}}=\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ for simplicity. We show that $\mathcal{R}^{\mathrm{univ}}$ is lisse on U_S by contradiction. By the criterion in Lemma 5.2, if $\mathcal{R}^{\mathrm{univ}}$ is *not* lisse on U_S, then there exists a finite-field-valued point (say over $\mathbf{F}_q$) and a section of $\mathcal{R}^{\mathrm{univ}}$ over the étale local ring $\mathcal{O}_\eta^{et}$ for some generic point η of $U_{\mathbf{F}_q}$ which is non-zero at the special point, but zero at the generic point. If we denote by i the inclusion of η in $\mathrm{Spec}(\mathcal{O}_\eta^{et})$, then such a section corresponds to a morphism $i_*\overline{\mathbf{Q}}_\ell\to\mathcal{R}^{\mathrm{univ}}$ over this local ring that is non-trivial at the generic point. Because

$$\mathcal{R}^{\mathrm{univ}}=R^{2l(k-1)}f_!\left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j}\big(x_{i,j}/x_{l+i,j}\big)\right)$$

and the complex

$$Rf_!\left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j}\big(x_{i,j}/x_{l+i,j}\big)\right)$$

is supported in a single degree, we obtain a nontrivial map.

$$Ri_*\overline{\mathbf{Q}}_\ell[-2l(k-1)]\to Rf_!\left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j}\big(x_{i,j}/x_{l+i,j}\big)\right). \tag{8.1}$$

We then apply the Verdier duality functor, taking our base scheme $S=\mathrm{Spec}(\mathcal{O}_\eta^{et})$. In this case our dualizing complex is $\overline{\mathbf{Q}}_\ell$ and we set $\mathrm{D}(\mathcal{F})=\mathrm{Hom}(\mathcal{F},\overline{\mathbf{Q}}_\ell)$. Later, we will apply also apply Verdier duality on schemes of finite type over S (see, *e.g.*, [13, Chapter 8, Chapter 10.1] for the ℓ-adic formalism of Verdier duality in this setting). As usual, for a scheme of finite type over S with structural morphism ϖ, we set $\mathrm{D}(\mathcal{F})=\mathrm{Hom}(\mathcal{F},\varpi^!\overline{\mathbf{Q}}_\ell)$. Dualizing the morphism (8.1), we obtain a morphism

$$\mathrm{D}\,Rf_!\left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j}\big(x_{i,j}/x_{l+i,j}\big)\right)\to\mathrm{D}\,Ri_*\overline{\mathbf{Q}}_\ell[2l(k-1)], \tag{8.2}$$

that is also nontrivial, since by double-duality its dual is (8.1).

We have

$$\mathrm{D}\, Ri_*\overline{\mathbf{Q}}_\ell = Ri_!\,\mathrm{D}\,\overline{\mathbf{Q}}_\ell = Ri_!i^!\overline{\mathbf{Q}}_\ell = Ri_!\overline{\mathbf{Q}}_\ell[-2] = Ri_*\overline{\mathbf{Q}}_\ell[-2],$$

where the last two equalities follow respectively from the fact that i is the inclusion of a smooth divisor of codimension one and the fact that i is proper. The left-hand side of (8.2) is

$$\begin{aligned}
&Rf_*\,\mathrm{D}\left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j}\big(x_{i,j}/x_{l+i,j}\big)\right)\\
=&Rf_*\,\mathrm{D}\left(g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\right)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j^{-1}}\big(x_{i,j}/x_{l+i,j}\big),
\end{aligned}$$

since duality is local, and therefore commutes with twisting with a locally constant sheaf. Hence the existence of a non-trivial morphism (8.2) would lead to a morphism

$$i^*Rf_*\,\mathrm{D}\,g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j^{-1}}\big(x_{i,j}/x_{l+i,j}\big)\to\overline{\mathbf{Q}}_\ell\big[2l(k-1)+2\big]$$

that is nontrivial at η. Finally, this would force the stalk of the sheaf

$$i^*Rf_*\,\mathrm{D}\,g^*\big(Ru_*\overline{\mathbf{Q}}_\ell\big)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j^{-1}}\big(x_{i,j}/x_{l+i,j}\big)$$

in degree $-2l(k-1)-2$ to be nontrivial at the generic point of $\mathbf{A}^{2l+1}$. We will now prove that this last property fails.

Away from the vanishing set of g, the sheaf $g^*(Ru_*\overline{\mathbf{Q}}_\ell)$ is the constant sheaf $\overline{\mathbf{Q}}_\ell$, so its dual is $\overline{\mathbf{Q}}_\ell[2(2l(k-1))]$, where $2l(k-1)$ is the relative dimension of X.

On the other hand, we claim that the morphism g is smooth in a Zariski-open neighborhood of the vanishing set of g. To check this, because $g' = gt$, it suffices to check that g' is smooth in a neighborhood of its vanishing set. Examining just the contribution $\sum_{j=1}^{k} x_{i,j}$ to g', observe that the only equation defining X_2 involving $(x_{i,1},\dots,x_{i,k})$ is of the form $\prod_{j=1}^{k} x_{i,j} = \alpha$, so the derivative of this contribution in a transverse direction is nonzero, and g' is smooth, unless $x_{i,1} = x_{i,2} = \cdots = x_{i,k}$. In this case, all the x_i are equal to some k-th root of $s(r+b_i)$, and thus

$$g' = \sum_{i=1}^{l}(s(r+b_i))^{1/k} - \sum_{i=l+1}^{2l}(s(r+b_i))^{1/k},$$

which is non-zero when $(r,\boldsymbol{b})\in U$.

Since g is smooth in a neighborhood of the vanishing locus of g, the sheaf $\mathrm{D}\, g^*(Ru_*\overline{\mathbf{Q}}_\ell) = g^! \,\mathrm{D}(Ru_*\overline{\mathbf{Q}}_\ell)$ is there a shift (and Tate twist) of $g^*\,\mathrm{D}(Ru_*\overline{\mathbf{Q}}_\ell)$, which is a shift (and Tate twist) of $g^* Ru_!\overline{\mathbf{Q}}_\ell$, and thus vanishes on the zero-set of g. We conclude that $\mathrm{D}\, g^*(Ru_*\overline{\mathbf{Q}}_\ell)$ is everywhere supported in degree

$$-4l(k-1).$$

Finally, we observe that f is an affine morphism from a scheme of dimension $2l(k-1)$. By results of Gabber (see [14, XV, Theorem 1.1.2]), the support of the sheaf

$$R^d f_*\,\mathrm{D}\left(g^*(Ru_*\overline{\mathbf{Q}}_\ell)\otimes\bigotimes_{j=1}^{k}\bigotimes_{i=1}^{l}\mathcal{L}_{\widetilde{\chi}_j^{-1}}\left(x_{i,j}/x_{l+i,j}\right)\right)$$

has dimension $2l(k-1)-d-4l(k-1)$ relative to S. Hence, its stalk in degree $2-2l(k-1)$ has support of dimension

$$2l(k-1)+2l(k-1)-2-4l(k-1)=-2$$

and therefore vanishes at the generic point of the special fiber, which has dimension -1 (relative to $\mathrm{Spec}(\mathcal{O}_\eta^{et})$). This is the desired contradiction. ☐

9. Injectivity

Let

$$\mathcal{W}_1=\mathcal{V}^\Delta\cup\{\boldsymbol{b}\in\mathbf{A}^{2l}\ \mid\ \text{at most two coordinates of } \boldsymbol{b} \text{ have multiplicity } 1\}.$$

This is a closed subvariety of codimension $l-1$ of $\mathbf{A}^{2l}_{\mathbf{Z}}$. The goal of this section is to prove the following injectivity statement for $\theta_{\boldsymbol{b}}$:

Theorem 9.1. *Let $p>2k+1$ be a prime and let $\mathbf{F}_q$ be a finite field of characteristic p with q elements. Let $\boldsymbol{\chi}$ be a k-tuple of ℓ-adic characters of $\mathbf{F}_q^\times$ with Property* CGM.

For p large enough, depending only on (k,l) and for $\boldsymbol{b}\in\mathbf{A}^{2l}(\mathbf{F}_q)$ outside $\mathcal{W}_1(\mathbf{F}_q)$, the natural morphism

$$\theta_{\boldsymbol{b}}\colon\ \mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}})\longrightarrow\mathrm{End}_{U_{\boldsymbol{b}}}\left(\mathcal{R}_{\boldsymbol{b}}^*\right)$$

is injective.

We begin with a lemma. First, we observe that for any $\boldsymbol{b}$, and any geometrically irreducible component $\mathcal{H}$ of $\mathcal{K}_{\boldsymbol{b}}$, we can meaningfully speak of the weight one part of $R^1\pi_!\mathcal{H}$, since $\mathcal{H}$ is defined over a finite field extension of $\mathbf{F}_q$.

Lemma 9.2. *For any $\boldsymbol{b} \in \mathbf{A}^{2l}(\mathbf{F}_q)$, the morphism $\theta_{\boldsymbol{b}}$ is injective if, and only if, for any geometrically irreducible component $\mathcal{H}$ of $\mathcal{K}_{\boldsymbol{b}}$, the weight one part of $R^1\pi_!\mathcal{H}$ is non-zero.*

Proof. Since $\mathcal{K}_{\boldsymbol{b}}$ is pointwise pure, hence geometrically semisimple, it is geometrically isomorphic to a direct sum

$$\bigoplus_{i\in I} \mathcal{F}_i^{\oplus n_i}$$

for some geometrically irreducible sheaves $\mathcal{F}_i$ and some integers $n_i \geqslant 1$. Then

$$R^1\pi_!\mathcal{K}_{\boldsymbol{b}} \simeq \bigoplus_{i\in I} \big(R^1\pi_!\mathcal{F}_i\big)^{\oplus n_i},$$

and the maximal weight one quotient of $R^1\pi_!\mathcal{K}_{\boldsymbol{b}}$ is also the corresponding direct sum of the maximal weight one quotients $(R^1\pi_!\mathcal{F}_i)^{w=1}$ of $R^1\pi_!\mathcal{F}_i$, with multiplicity n_i. If one of these quotients vanishes, then any $u \in \mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}})$ that is non-zero only on the corresponding summand $\mathcal{F}_i$ satisfies $\theta_{\boldsymbol{b}}(u) = 0$.

Conversely, suppose that all the quotients $(R^1\pi_!\mathcal{F}_i)^{w=1}$ are non-zero. By Schur's Lemma, the endomorphism algebra $\mathrm{End}_{U_{\boldsymbol{b}}}(\mathcal{R}^*_{\boldsymbol{b}})$ is isomorphic to a product of matrix algebras $M_{n_i}(\overline{\mathbf{Q}}_\ell)$. For each i, $\theta_{\boldsymbol{b}}$ maps an endomorphism u to the endomorphism of $(R^1\pi_!\mathcal{F}_i)^{w=1,\oplus n_i}$ represented by a block matrix with diagonal scalar matrices in each block, whose entries are the coefficients of the matrix in $M_{n_i}(\overline{\mathbf{Q}}_\ell)$ corresponding to u. Since the blocks have non-zero size, such a matrix is zero if and only if u is zero. □

Let G be the geometric monodromy group of $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$. Let $\boldsymbol{b} \in \mathbf{A}^{2l}(\mathbf{F}_q)$. We denote by $B \subset \mathbf{A}^1$ the set of values $\{b_i\}$. For any family $\boldsymbol{\varrho} = (\varrho_x)_{x\in B}$ of irreducible representations of G, we denote by $\mathcal{H}_{\boldsymbol{\varrho}}$ the sheaf

$$\mathcal{H}_{\boldsymbol{\varrho}} = \bigotimes_{x\in B} \varrho_x(\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}))(s(r+x)),$$

on $\mathbf{A}^2$ with coordinates (r, s).

Lemma 9.3. *Assume that $\boldsymbol{\chi}$ has* CGM. *Any geometrically irreducible component $\mathcal{H}$ of $\mathcal{K}_{\boldsymbol{b}}$ is isomorphic to $\mathcal{H}_{\boldsymbol{\varrho}}$ for some family $\boldsymbol{\varrho} = (\varrho_x)_{x\in B}$ such that, for all $x \in B$, the representation ϱ_x is an irreducible summand of the representation* $\mathrm{Std}^{\otimes n_1} \otimes (\mathrm{Std}^\vee)^{\otimes n_2}$*, where*

$$n_1 = \sum_{\substack{1\leqslant i\leqslant l\\ b_i=x}} 1, \qquad n_2 = \sum_{\substack{l+1\leqslant i\leqslant 2l\\ b_i=x}} 1. \tag{9.1}$$

Proof. Write

$$\mathcal{K}_{\boldsymbol{b}} = \bigotimes_{x\in B} \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(s(r+x))^{\otimes n_1} \otimes \big(\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(s(r+x))^\vee\big)^{\otimes n_2}.$$

By the Goursat-Kolchin-Ribet criterion (see [17] or [10]), which may be applied since the sheaf $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$ has geometric monodromy group SL_k or Sp_k by Theorem 6.2, the sheaf

$$\bigoplus_{x\in B}\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(s(r+x))$$

has geometric monodromy group $G^{|B|}$, so that its irreducible components correspond exactly to the tuples $\boldsymbol{\varrho}$. □

Lemma 9.4. *Let $\boldsymbol{b}$ be a point in $\mathbf{A}^{2l}-\mathcal{V}^{\Delta}$. Let $\mathcal{H}_{\boldsymbol{\varrho}}$ be an irreducible component of $\mathcal{K}_{\boldsymbol{b}}$. Then the rank of $R^1\pi_!\mathcal{H}_{\boldsymbol{\varrho}}$ on the dense open set where $P_{\boldsymbol{b}}(r)\neq 0$ is equal to the rank of $\mathcal{H}_{\boldsymbol{\varrho}}$ divided by k.*

Proof. Note that the set where $P_{\boldsymbol{b}}$ does not vanish is indeed a dense open subset by Lemma 7.4.

Let r be such that $P_{\boldsymbol{b}}(r)\neq 0$. Then by proper base change, the stalk of $R^1\pi_!\mathcal{H}_{\boldsymbol{\varrho}}$ at r is equal to $H^1_c(\mathbf{G}_{m,\overline{\mathbf{F}}_q},\mathcal{H}_{\boldsymbol{\varrho},r})$.

Because $P_{\boldsymbol{b}}(r)\neq 0$, Lemma 7.2 shows that the local monodromy representation at ∞ of $\mathcal{K}_{\boldsymbol{b},r}$ is isomorphic to a sum of sheaves of the form $\mathcal{L}_{\psi}(\alpha\cdot s^{1/k})$ for nonzero α. Each sheaf $\mathcal{L}_{\psi}(\alpha\cdot s^{1/k})$ has all breaks $1/k$ at ∞, so the same is true for $\mathcal{K}_{\boldsymbol{b},r}$.

The sheaf $\mathcal{H}_{\boldsymbol{\varrho},r}$ is a summand of $\mathcal{K}_{\boldsymbol{b},r}$, hence it also lisse on $\mathbf{G}_m$, tamely ramified at 0, and has all breaks $1/k$ at ∞. Moreover, it also satisfies

$$H^0_c\Big(\mathbf{G}_{m,\overline{\mathbf{F}}_q},\mathcal{H}_{\boldsymbol{\varrho}}\Big)=H^2_c\Big(\mathbf{G}_{m,\overline{\mathbf{F}}_q},\mathcal{H}_{\boldsymbol{\varrho}}\Big)=0,$$

and therefore the Euler-Poincaré characteristic formula for a lisse sheaf on $\mathbf{G}_m$ implies that

$$\begin{aligned}\dim H^1_c\Big(\mathbf{G}_{m,\overline{\mathbf{F}}_q},\mathcal{H}_{\boldsymbol{\varrho},r}\Big)&=-\chi\Big(\mathbf{G}_{m,\overline{\mathbf{F}}_q},\mathcal{H}_{\boldsymbol{\varrho},r}\Big)=\mathrm{Swan}_0\Big(\mathcal{H}_{\boldsymbol{\varrho},r}\Big)+\mathrm{Swan}_\infty\Big(\mathcal{H}_{\boldsymbol{\varrho},r}\Big)\\&=\frac{1}{k}\,\mathrm{rk}\,\big(\mathcal{H}_{\boldsymbol{\varrho}}\big).\end{aligned}$$

□

In the next lemmas, we fix a point $\boldsymbol{b}$ in $\mathbf{A}^{2l}-\mathcal{V}^{\Delta}$, and an index i such that $b_i\neq b_j$ for $j\neq i$.

We denote $\epsilon=-1$ if $1\leqslant i\leqslant l$, and $\epsilon=1$ if $l+1\leqslant i\leqslant 2l$. For any character χ, we denote n_χ the multiplicity of χ in $\boldsymbol{\chi}$, which is 0 if $\chi\notin\boldsymbol{\chi}$.

For an irreducible component

$$\mathcal{H}_{\boldsymbol{\varrho}}=\bigotimes_{x\in B}\varrho_x(\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}))(s(r+x))$$

of $\mathcal{K}_{\boldsymbol{b}}$ (all are of this type by Lemma 9.3), we denote

$$\mathcal{M}_{\boldsymbol{\varrho}}=\bigotimes_{\substack{x\in B\\ x\neq b_i}}\varrho_x(\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}))(s(r+x)).\tag{9.2}$$

Since $\mathcal{M}_{\boldsymbol{\varrho}}$ is tamely ramified at 0, its local monodromy representation at $s = 0$ can be expressed as a sum of Jordan blocks, which we write

$$\bigoplus_{\eta} \mathcal{L}_{\eta} \otimes J(m_{\eta}),$$

where η runs over a finite set of characters.

Lemma 9.5. *With notation as above, the rank of the weight one part of $R^1\pi_!\mathcal{H}_{\boldsymbol{\varrho}}$ on the nonempty open set where $P_{\boldsymbol{b}}(r) \neq 0$ is equal to*

$$\sum_{\eta} \max(m_{\eta} - n_{\eta^{\epsilon}}, 0).$$

Proof. Because b_i occurs with multiplicity one in B, the representation ϱ_{b_i} is necessarily the standard representation if $i \leqslant l$ or its dual if $i > l$ (see (9.1)), and in any case has rank k. This implies that

$$\operatorname{rk}\left(\mathcal{H}_{\boldsymbol{\varrho}}\right) = k \operatorname{rk}\left(\mathcal{M}_{\boldsymbol{\varrho}}\right)$$

and hence by Lemma 9.4, we have

$$\operatorname{rk}\left(R^1\pi_!\mathcal{H}_{\boldsymbol{\varrho}}\right) = \operatorname{rk}\left(\mathcal{M}_{\boldsymbol{\varrho}}\right) = \sum_{\eta} m_{\eta},$$

so that it suffices to show that the weight < 1 part of $R^1\pi_!\mathcal{H}_{\boldsymbol{\varrho}}$ has the rank

$$\sum_{\eta} \min\left(m_{\eta}, n_{\eta^{\epsilon}}\right).$$

To prove this, observe that the weight < 1 part is the sum over the singularities of the sheaf of the local monodromy invariants (see, *e.g.*, [21, Lemma 4.22(2)]). Because $\mathcal{H}_{\boldsymbol{\varrho},r}$ is a summand of $\mathcal{K}_{\boldsymbol{b},r}$ which by Lemma 7.2 has no nontrivial local monodromy invariants at ∞, $\mathcal{H}_{\boldsymbol{\varrho},r}$ has no nontrivial local monodromy invariants at ∞.

If $i \leqslant l$, then the local monodromy representation at 0 is given by

$$\begin{aligned}\mathcal{H}_{\boldsymbol{\varrho},r} &= \mathcal{M}_{\boldsymbol{\varrho}} \otimes \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(s(r+b_i)) = \left(\bigoplus_{\eta} \mathcal{L}_{\eta} \otimes J(m_{\eta})\right) \otimes \left(\bigoplus_{\chi \in \boldsymbol{\chi}} \mathcal{L}_{\chi} \otimes J(n_{\chi})\right)\\ &= \bigoplus_{\eta} \bigoplus_{\chi \in \boldsymbol{\chi}} \mathcal{L}_{\eta\chi} \otimes J(m_{\eta}) \otimes J(n_{\chi}).\end{aligned}$$

The dimension of the invariant subspace of $\mathcal{L}_{\eta\chi} \otimes J(m_{\eta}) \otimes J(n_{\chi})$ is zero unless $\eta\chi = 1$, in which case it is $\min(m_{\eta}, n_{\chi})$, hence the result follows in that case. If $l + 1 \leqslant i \leqslant 2l$, the same calculation applies, except that $\mathcal{L}_{\chi^{-1}}$ appears instead of $\mathcal{L}_{\chi}$. $\square$

The next lemma continues with the same notation.

Lemma 9.6. *Assume that* $\boldsymbol{\chi}$ *has* CGM. *Then the rank of the weight one part of* $R^1\pi_!\mathcal{H}_{\boldsymbol{\varrho}}$ *is at least two.*

Proof. By the previous lemma, it is enough to prove that

$$\sum_{\eta} \max(m_\eta - n_{\eta^\epsilon}, 0) \geqslant 2. \tag{9.3}$$

Since $\boldsymbol{b} \notin \mathcal{W}_1$, there are at least three elements of B that occur with multiplicity one, say b_i, b_j and $b_{j'}$.

Let $\delta = 1$ if $j \leqslant l$ and $\delta = -1$ if $j > l$, so that ϱ_{b_j} is the standard representation if $\delta = 1$ and the dual representation if $\delta = -1$.

Let

$$\mathcal{M}'_{\boldsymbol{\varrho}} = \bigotimes_{\substack{x\in B\\ x\neq b_i, b_j}} \varrho_x\big(\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})\big)(s(r+x))$$

so that

$$\mathcal{M}_{\boldsymbol{\varrho}} = \mathcal{M}'_{\boldsymbol{\varrho}} \otimes \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})\big(s\big(r+b_j\big)\big)$$

if $\delta = 1$ and

$$\mathcal{M}_{\boldsymbol{\varrho}} = \mathcal{M}'_{\boldsymbol{\varrho}} \otimes \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})\big(s\big(r+b_j\big)\big)^{\vee}$$

if $\delta = -1$.

Let $\mathcal{L}_\theta \otimes J(r)$ be a Jordan block in the local monodromy representation of $\mathcal{M}'_{\boldsymbol{\varrho}}$ at $s = 0$. We estimate the contribution from this factor in the local monodromy representation (9.2) of $\mathcal{M}_{\boldsymbol{\varrho}}$.

This contribution contains a direct sum

$$\bigoplus_{\chi\in\boldsymbol{\chi}} \mathcal{L}_{\chi^\delta\theta} \otimes J\big(n_\chi + r - 1\big). \tag{9.4}$$

If the character θ is nontrivial, then the tuple of characters $\theta^\epsilon \boldsymbol{\chi}^{\epsilon\delta}$ cannot be equal to $\boldsymbol{\chi}$, up to permutation because this would contradict the CGM assumption. Hence, there exists a character χ such that $n_\chi > n_{\chi^{\delta\epsilon}\theta^\epsilon}$, and therefore the Jordan blocks (9.4) include a character $\eta = \chi^\delta\theta$ with $m_\eta > n_{\eta^\epsilon}$. Hence these blocks have a contribution

$$\geqslant \min\big(n_\chi + r - 1 - n_{\chi^{\delta\epsilon}\theta^\epsilon}, 0\big) \geqslant r$$

to the sum on the left-hand side of (9.3).

On the other hand, if θ is trivial, then the character χ with n_χ maximal contributes

$$\geqslant \min\big(n_\chi + r - 1 - n_\chi, 0\big) = r - 1.$$

In particular, we obtain (9.3) except if the local monodromy of $\mathcal{M}'_{\boldsymbol{\varrho}}$ at zero consists of at most one unipotent Jordan block of rank two, or of at most one nontrivial character of rank one, plus a sum of any number of trivial representations. This conditions means that local monodromy representation of $\mathcal{M}'_{\boldsymbol{\varrho}}$ at zero is either trivial or is a pseudoreflection (unipotent or not).

In the first case, we have a sheaf with trivial local monodromy at 0 that is expressed as a tensor product. Then all the tensor factors must have scalar local monodromy at 0. This is impossible here, since one of the tensor factors is $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(s(r+b_{j'}))$ or its dual, and the local monodromy of this sheaf is not scalar (because $k \geqslant 2$).

If the local monodromy representation is a pseudoreflection, then when it is expressed as a tensor product, all but one of the tensor factors must be one-dimensional, and the remaining factor must have local monodromy that is given by a pseudoreflection times a scalar. Again, because one of the tensor factors is $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(s(r+b_{j'}))$ or its dual, this must be the special factor, and this can only happen when $k = 2$ by Proposition 6.1. All the remaining tensor factors are one-dimensional. But since the geometric monodromy group is SL_2 in that case (because $\boldsymbol{\chi}$ has CGM), and the only one-dimensional representation of SL_2 is the trivial representation, and this only appears in even tensor powers of the standard representation, we conclude that all remaining factors must have even multiplicity. This is a contradiction, since we have three factors with multiplicity one, and the sum of the multiplicities is $2l$, which is even. □

Now Theorem 9.1 follows immediately from Lemma 9.2 and Lemma 9.6.

10. Specialization statement

We continue with the previous notation. Recall that $X_\infty = \mathbf{A}^{2l} - \mathcal{V}^\Delta$ and that X_j is defined in Definition 7.6. We recall that we have the projection $f\colon U \to \mathbf{A}^{2l}$.

Lemma 10.1. *For each j, the subvariety X_j is closed in X_∞.*

For each irreducible component X of X_j that intersects the characteristic zero part, the morphism

$$f\colon Z \cap f^{-1}(X - X \cap X_{j-1}) \to X - X \cap X_{j-1}$$

is finite étale.

Proof. These claims follow from Lemma 7.4. Indeed, Z is the solution set of a family of nonzero polynomials in one variable indexed by points of $X_\infty = \mathbf{A}^{2l} - \mathcal{V}^\Delta$. The set X_j is constructible, so to show it is closed it suffices to show that it is closed under specialization. The polynomial factorizes completely over any geometric generic point into one distinct factor for each root, raised to some power, and each factor has at most one root over the special point, so the number of roots over the special point is at most the number of roots over the generic point, as desired.

To check that $Z \cap f^{-1}(X - X \cap X_{j-1})$ is finite étale over $X - X \cap X_{j-1}$, we consider the polynomial $P(r)$ over the étale local ring of a point of $X - X \cap X_{j-1}$, which is an integral strict Henselian local ring, and use the fact that the polynomial has the same number of roots over the special point and over the generic point.

By the previous discussion each linear factor over the geometric generic point must admit a root over the residue field, which means the polynomial is monic. Because it is monic, and the ring is strict henselian, we can factor it into a product of irreducible factors, each with exactly one root in the residue field. Over the generic point each such factor will have only one root in the residue field, hence have only one root in the fraction field. Therefore, because the generic point has characteristic zero, so all polynomials are separable, each such factor is a power of $(x-\alpha)$ where α is its unique root, so the polynomial is a product of linear factors, with at most one distinct linear factor with each possible root in the residue field, hence its vanishing set is the disjoint union of the vanishing sets of these linear factors and thus is finite étale. ☐

Fix $j \geqslant 0$. Let $X \subset X_j \subset \mathbf{A}^{2l}$ be an irreducible component of X_j over $\mathbf{Z}$ which intersects the characteristic zero part. We consider a finite field $\mathbf{F}_q$ of characteristic $p > 2k+1$ such that $X_{\overline{\mathbf{F}}_q}$ is irreducible and nonempty.

Lemma 10.2. *Let $\boldsymbol{\chi}$ be a k-tuple of characters of $\mathbf{F}_q^\times$. The sheaf $\mathcal{R}^*|(U \cap f^{-1}(X_{\overline{\mathbf{F}}_q} - X_{\overline{\mathbf{F}}_q} \cap X_{j-1}))$ is tamely ramified around the divisor $Z \cup \{\infty\}$.*

Proof. Let n be the lcm of the orders of the characters χ_i. By the remarks before Proposition 8.1, there exists a tuple $\widetilde{\boldsymbol{\chi}}$ of characters of $\boldsymbol{\mu}_n$ such that $\boldsymbol{\chi}$ is associated to this tuple. Let $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ be the sheaf over $\mathbf{Z}[\boldsymbol{\mu}_n, 1/(n\ell)]$ given by Proposition 8.1. This sheaf $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ is lisse on the open set $U \cap f^{-1}(X_j - X \cap X_{j-1})$, whose complement is the étale divisor $Z \cup \{\infty\}$. Hence, by Abyankhar's Lemma [24, Exposé XIII, Section 5], the sheaf $\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})$ is tamely ramified, and hence so is

$$\mathcal{R}^{\mathrm{univ}}(\widetilde{\boldsymbol{\chi}})|\mathbf{A}^{1+2l}_{\mathbf{F}_q} = \mathcal{R}(\boldsymbol{\chi}),$$

and also $\mathcal{R}^*(\boldsymbol{\chi})$. ☐

Proposition 10.3. *Let η be the generic point of $X_{\overline{\mathbf{F}}_q}$, and let $\bar{\eta}$ be a geometric generic point over η. Let $\boldsymbol{\chi}$ be a k-tuple of characters of $\mathbf{F}_q^\times$ with Property* CGM. *Suppose that*

$$\dim \mathrm{End}_{U_{\bar{\eta}}}\left(\mathcal{R}^*_{\bar{\eta}}\right) = \dim \mathrm{End}_{V_{\bar{\eta}}}\left(\mathcal{K}_{\bar{\eta}}\right).$$

Let $\boldsymbol{b} \in X(\mathbf{F}_q)$ such that $\boldsymbol{b} \notin X_{j-1}$ and $\boldsymbol{b} \notin \mathcal{W}_1$. Then we have

$$\dim \mathrm{End}_{U_b}\left(\mathcal{R}^*_{\boldsymbol{b}}\right) = \dim \mathrm{End}_{V_b}\left(\mathcal{K}_{\boldsymbol{b}}\right).$$

Proof. Consider the sheaf

$$\mathcal{E} = R^2 f_!\big(\mathcal{R}^* \otimes \mathcal{R}^{*,\vee}\big)$$

on $\mathbf{A}^{2l}_{\mathbf{F}_q}$. We claim that:

(a) The restriction of $\mathcal{E}$ to $X_j - X_{j-1}$ is lisse;
(b) We have an isomorphism

$$\mathcal{E}_{\bar{\eta}} \simeq \mathrm{End}_{U_{\bar{\eta}}}\left(\mathcal{R}^*_{\bar{\eta}}\right)(-1);$$

(c) We have an isomorphism

$$\mathcal{E}_{\boldsymbol{b}} \simeq \mathrm{End}_{U_{\boldsymbol{b}}}(\mathcal{R}^*_{\boldsymbol{b}})(-1).$$

Moreover, let $g\colon V \to \mathbf{A}^{2l}$ be the map $(r, s, \boldsymbol{b}) \mapsto \boldsymbol{b}$ over $\mathbf{Z}$ and

$$\widetilde{\mathcal{E}} = R^4 g_!\big(\mathcal{K} \otimes \mathcal{K}^\vee\big)$$

on $\mathbf{A}^{2l}_{\mathbf{F}_q}$. We claim that:

(a') The restriction of $\widetilde{\mathcal{E}}$ to $X_j - X_{j-1}$ is lisse;
(b') We have an isomorphism

$$\widetilde{\mathcal{E}}_{\bar\eta} \simeq \mathrm{End}_{V_{\bar\eta}}\big(\mathcal{K}_{\bar\eta}\big)(-1);$$

(c') We have an isomorphism

$$\widetilde{\mathcal{E}}_{\boldsymbol{b}} \simeq \mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}})(-1).$$

Assuming these facts, we have

$$\begin{aligned}\dim \mathrm{End}_{U_{\boldsymbol{b}}}\big(\mathcal{R}^*_{\boldsymbol{b}}\big) &= \dim \mathcal{E}_{\boldsymbol{b}} = \dim \mathcal{E}_{\bar\eta} = \dim \mathrm{End}_{U_{\bar\eta}}\big(\mathcal{R}^*_{\bar\eta}\big)\\ &= \dim \mathrm{End}_{V_{\bar\eta}}\big(\mathcal{K}_{\bar\eta}\big) = \dim \widetilde{\mathcal{E}}_{\bar\eta} = \dim \widetilde{\mathcal{E}}_{\boldsymbol{b}} = \dim \mathrm{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}}),\end{aligned}$$

with the identities following from respectively (c), (a), (b), the assumption, (b'), (a'), and (c'). (In particular, when we apply assumption (a) and (a'), we use the fact that $\boldsymbol{b}$ is a specialization of $\bar\eta$, hence they lie on the same connected component of $X_j - X_{j-1}$, and so any lisse sheaf on $X_j - X_{j-1}$ has equal ranks at these two points.)

We now prove the claims. The assertions (b)/(b') and (c)/(c') follow from the proper base change theorem, Poincaré duality, and semisimplicity.

Assertion (a) is a consequence of Deligne's semicontinuity theorem and the tameness of $\mathcal{R}^*$. Specifically, by Lemma 10.1, we know that U, over $X_{\mathbf{F}_q} - (X_{\mathbf{F}_q} \cap X_{j-1})$, is the complement of a finite étale divisor inside a morphism smooth and proper of relative dimension one, and $\mathcal{R}^* \otimes \mathcal{R}^{*,\vee}$ is a lisse sheaf on it. By Lemma 10.2, the Swan conductor of $\mathcal{R}^* \otimes \mathcal{R}^{*,\vee}$ at this divisor vanishes, and so by Deligne's semicontinuity theorem [22, Corollary 2.1.2] the cohomology sheaf is lisse.

Assertion (a'): Let $Y = X_{\mathbf{F}_q} - (X_{\mathbf{F}_q} \cap X_{j-1})$. Then $\mathcal{K} \otimes \mathcal{K}^\vee$ is lisse on $V \times_{\mathbf{A}^{2l}} Y$. Let $\big(\mathcal{K} \otimes \mathcal{K}^\vee\big)^{\pi_1(V \times_{\mathbf{A}^{2l}} Y)}$ be its (geometric) monodromy invariants. Then there is a natural map

$$\big(\mathcal{K} \otimes \mathcal{K}^\vee\big)^{\pi_1(V \times_{\mathbf{A}^{2l}} Y)} \to \mathcal{K} \otimes \mathcal{K}^\vee$$

over $V \times_{\mathbf{A}^{2l}} Y$, where we interpret $\big(\mathcal{K} \otimes \mathcal{K}^\vee\big)^{\pi_1(V \times_{\mathbf{A}^{2l}} Y)}$ as a constant sheaf. This induces by functoriality a map

$$R^4 g_!\big(\mathcal{K} \otimes \mathcal{K}^\vee\big)^{\pi_1(V \times_{\mathbf{A}^{2l}} Y)} \to R^4 g_! \mathcal{K} \otimes \mathcal{K}^\vee$$

over Y. Because V is an open subset of $\mathbf{A}^{2l+2}$ whose fibers under g are all nonempty, the top cohomology of a constant sheaf along g is a constant sheaf, so this gives a map

$$\left(\mathcal{K} \otimes \mathcal{K}^\vee\right)^{\pi_1(V \times_{\mathbf{A}^{2l}} Y)} \to R^4 g_! \mathcal{K} \otimes \mathcal{K}^\vee.$$

We claim that this last map is an isomorphism. It is sufficient to check this on the stalk at each point $\boldsymbol{b}$. To do this, first check that the monodromy group of $\mathcal{K} \otimes \mathcal{K}^\vee$ over $V \times_{\mathbf{A}^{2l}} Y$ is equal to the monodromy of the same sheaf on $V_{\boldsymbol{b}}$. This can be done using Goursat-Kolchin-Ribet, since $\boldsymbol{\chi}$ has CGM and $p > 2k + 1$. We also use the fact that, because Z is finite etale over Y, and Z includes $\{-b_1, \ldots, -b_{2l}\}$, no b_i, b_j that are distinct generically on the Y stratum can become equal at any point of Y.

Next observe that this map is simply the natural map from the monodromy invariants of $\mathcal{K} \otimes \mathcal{K}^\vee$ to the monodromy coinvariants of $\mathcal{K} \otimes \mathcal{K}^\vee$. Because the monodromy is semisimple, it is an isomorphism. □

11. Diophantine preliminaries for the proof of the generic statement

This section uses independent notation from the rest of the paper. In particular, we will use the letter k to denote finite fields.

We will use the following variant of the Diophantine Criterion for irreducibility of Katz (compare [20, page 25] and [21, Lemma 4.14]).

Lemma 11.1. *Let w be an integer. Let X be a geometrically irreducible separated scheme of finite type over a finite field k, and let U be a normal open dense subset of X. Let ℓ be a prime different from the characteristic of k. Let $\mathcal{F}$ be an ℓ-adic sheaf on X, mixed of weights $\leqslant w$ on X, and lisse and pure of weight w on U. We have then*

$$\dim \operatorname{End}_{\pi_1(U \times \overline{\mathbf{F}}_q)}(\mathcal{F}|U) = \limsup_{\nu \to +\infty} \frac{1}{|k|^{\nu(\dim(X_{\overline{\mathbf{F}}_q}) + w)}} \sum_{x \in X(k_\nu)} \left|t_{\mathcal{F}}(x; k_\nu)\right|^2, \quad (11.1)$$

where k_ν is the extension of k of degree ν in a fixed algebraic closure.

In particular, if the right-hand side of the formula above is equal to 1*, then $\mathcal{F}|U$ is geometrically irreducible.*

Proof. Let $n = \dim(X_{\overline{\mathbf{F}}_q})$. Up to performing a Tate twist on $\mathcal{F}$, we may assume that $w = 0$. For any $x \in X(k_\nu)$ we have then

$$\left|t_{\mathcal{F}}(x; k_\nu)\right|^2 \leqslant \operatorname{rk}(\mathcal{F})^2,$$

hence by trivial counting we get

$$\frac{1}{|k|^{n\nu}}\sum_{x\in X(k_\nu)}\big|t_{\mathcal{F}}(x;k_\nu)\big|^2=\frac{1}{|k|^{n\nu}}\sum_{x\in U(k_\nu)}\big|t_{\mathcal{F}}(x;k_\nu)\big|^2+\frac{1}{|k|^{n\nu}}\sum_{x\in(X-U)(k_\nu)}\big|t_{\mathcal{F}}(x;k_\nu)\big|^2$$
$$=\frac{1}{|k|^{n\nu}}\sum_{x\in U(k_\nu)}\big|t_{\mathcal{F}}(x;k_\nu)\big|^2+O_{\mathcal{F}}\big(|k|^{-\nu}\big).$$

This shows that we may restrict the sum on the right-hand side of (11.1) to $U(k_\nu)$.

Since $\mathcal{F}$ and its dual $\mathcal{F}^\vee$ are lisse and pointwise pure of weight 0 on U, the sheaf $\mathrm{End}(\mathcal{F}) = \mathcal{F}\otimes\mathcal{F}^\vee$ is also lisse and pointwise pure of weight 0 on U. Moreover, for all $x\in U(k_\nu)$, we have

$$t_{\mathrm{End}(\mathcal{F})}(x;k_\nu)=|t_{\mathcal{F}}(x;k_\nu)|^2.$$

By the Grothendieck-Lefschetz trace formula, we have

$$\frac{1}{|k|^{n\nu}}\sum_{x\in U(k_\nu)}\big|t_{\mathcal{F}}(x;k_\nu)\big|^2=\frac{1}{|k|^{n\nu}}\mathrm{Tr}\big(\mathrm{Fr}_{k_\nu}|H^{2n}_c(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F}))\big)$$
$$+\frac{1}{|k|^{n\nu}}\sum_{i=0}^{2n-1}(-1)^i\mathrm{Tr}\big(\mathrm{Fr}_{k_\nu}|H^i_c(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F}))\big).$$

By Deligne's Riemann Hypothesis [4], all eigenvalues of the Frobenius of k_ν acting on the cohomology group $H^i_c(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F}_{\bar{k}}))$ have modulus $\leqslant |k|^{i/2}$, and therefore

$$\big|\mathrm{Tr}\big(\mathrm{Fr}_{k^\nu}\big|H^i_c\big(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F})\big)\big)\big|\leqslant\dim\big(H^i_c\big(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F})\big)\big)|k|^{i\nu/2},$$

so that we derive

$$\frac{1}{|k|^{n\nu}}\sum_{x\in U(k_\nu)}\big|t_{\mathcal{F}}(x;k_\nu)\big|^2=\frac{1}{|k|^{n\nu}}\mathrm{Tr}\big(\mathrm{Fr}_{k_\nu}|H^{2n}_c\big(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F})\big)\big)+O\big(|k|^{-\nu/2}\big).$$

On the other hand, we have a Frobenius-equivariant isomorphism

$$H^{2n}_c\big(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F})\big)\simeq\mathrm{End}(\mathcal{F})_{\pi_1(U\times\overline{\mathbf{F}}_q)}(-n).$$

The eigenvalues of Frobenius on $\mathrm{End}(\mathcal{F})_{\pi_1(U\times\overline{\mathbf{F}}_q)}(-n)$ have modulus q^n. Therefore

$$|k|^{-n\nu}\mathrm{Tr}\big(\mathrm{Fr}_{k^\nu}|H^{2n}_c\big(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F})\big)\big)$$

is the sum of the ν-th power of $\dim H^{2n}_c(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F}))$ complex numbers, each of of modulus 1, and by a standard lemma, we have therefore

$$\limsup_{\nu\to+\infty}\frac{1}{|k|^{n\nu}}\mathrm{Tr}\big(\mathrm{Fr}_{k^\nu}|H^{2n}_c\big(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F})\big)\big)=\dim H^{2n}_c\big(U\times\overline{\mathbf{F}}_q,\mathrm{End}(\mathcal{F})\big)$$
$$=\dim\mathrm{End}_{\pi_1(U\times\overline{\mathbf{F}}_q)}(\mathcal{F}),$$

by the geometric semi-simplicity of $\mathcal{F}|U$. □

This result, combined with the injectivity statement, reduces the desired isomorphism to a bound on exponential sums, where $\boldsymbol{b}$ are summed over a stratum of the stratification. The technique we will use to obtain cancellation is a form of separation of variables, where we essentially obtain cancellation in the sum over each individual coordinate b_i.

We now describe a general geometric form of the type of separation of variables that we will use.

- Let m and N be natural numbers. Let S be a finite set;
- Let $\mathcal{O}_K$ be the ring of integers of a number field, and B a separated scheme of finite type over $\mathcal{O}_K[1/N]$;
- Let C_i for $i \in S$ be curves over B. Let A be a smooth geometrically irreducible curve over $\mathbf{Z}[1/N]$. We will use s as a variable for points of A and x_i for points of C_i;
- We denote $\mathcal{C} = C_1 \times_B \cdots \times_B C_n$. We view functions on C_i as functions on $\mathcal{C}$ by composing with the i-th projection;
- For $1 \leqslant j \leqslant m$, let $\boldsymbol{f}_j = (f_{i,j})_{1 \leqslant i \leqslant n} \in \Gamma$ be a tuple of functions on the curves C_i, and let g_j be a function on B;
- Let $Y \subseteq \mathcal{C}$ be the common zero locus of the m functions
$$\Sigma_j := g_j + \sum_{i \in S} f_{i,j} \in \Gamma(\mathcal{C}, \mathcal{O}_{\mathcal{C}}), \ \ j = 1, \cdots, m;$$
- Let $\pi : Y \times A \to Y$ be the obvious projection, and $g_i : Y \times A \to C_i \times A$ the obvious morphisms;
- Let ℓ be a prime number dividing N. For $i \in S$, and q some prime ideal of $\mathcal{O}_K$ coprime to N, we assume given a lisse ℓ-adic sheaf $\mathcal{F}_i$, pointwise pure of weight 0, on $C_i \times A_{\mathbf{F}_q}$. We denote by $(\varrho, x_i, s) \mapsto t_i(\varrho, x_i, s; k)$ the trace function of $\mathcal{F}_i$ over some finite extension $k/\mathbf{F}_q$;
- For $s \in A(k)$ and $\varrho \in B(k)$ we set
$$\mathcal{F}_{i,\varrho,s} := \mathcal{F}_i | C_i \times_B \{\varrho\} \times \{s\}$$
the sheaf on $C_i \times k$ obtained by restricting to the fiber of ϱ and "freezing" the s-variable. We assume that for any q, any $k/\mathbf{F}_q$ and any point $s \in A(k)$ the conductor of $\mathcal{F}_{i,\varrho,s}$ is bounded by some constant $C \geqslant 1$;
- For q some prime of $\mathcal{O}_K$ coprime with n, we are given a lisse ℓ-adic sheaf $\mathcal{G}$, pointwise pure of weight 0, on $B \times A_{\mathbf{F}_q}$. We denote by $(\varrho, s) \mapsto t_*(\varrho, s; k)$ its trace function.

We make the following "twist-independence" assumption:

(TI). For all i, for all $\varrho \in B$ and for all $s_1 \neq s_2$ in A, the lisse sheaf $\mathcal{F}_{i,\varrho,s_1} \otimes \mathcal{F}^{\vee}_{i,\varrho,s_2}$ on each geometrically irreducible component of $C_{i,\varrho}$ has no geometrically irreducible component that is of rank 1.

The implicit constants associated with the symbols $O(\cdots)$ or $\ll$ are assumed to depend on $\mathcal{C}$, A, the maps $(\boldsymbol{f}_j)_{j=1,\cdots,m}$, and the conductors of the sheaves involved.

The main estimate on exponential sums we will need is the following

Proposition 11.2. *Assume that Assumption* (TI) *holds. We have*

$$\sum_{(\varrho,\boldsymbol{x})\in Y(k)} \left| \sum_{s\in A(k)} t_*(\varrho,s;k) \prod_{i=1}^{n} t_i(\varrho,x_i,s;k) \right|^2 = \sum_{(\varrho,\boldsymbol{x})\in Y(k)} \sum_{s\in A(k)} |t_*(\varrho,s,k)|^2 \prod_{i=1}^{n} |t_i(\varrho,x_i,s;k)|^2 + O\left(|k|^{\dim B+|S|/2+2}\right). \tag{11.2}$$

Remark 11.3. One can often show (by fibering by curves) that as $|k|\to\infty$ the first term on the righthand side of (11.2) satisfies

$$\sum_{(\varrho,\boldsymbol{x})\in Y(k)} \sum_{s\in A(k)} |t_*(\varrho,s,k)|^2 \prod_{i=1}^{n} |t_i(\varrho,b_i,s;k)|^2 \gg |k|^{\dim(Y\times A)_{\overline{\mathbf{F}}_q}},$$

while the error term is

$$\ll |k|^{(n-m+1)-1/2} \ll |k|^{\dim(Y\times A)_{\overline{\mathbf{F}}_q}-1/2}$$

as soon as

$$m \leqslant \frac{|S|-3}{2}.$$

Example 11.4. Take B a point, $C_i = A = \mathbf{G}_m$, $\mathcal{F}_i = [(b_i,s)\mapsto b_i s]^*\mathcal{K}\ell_2$ on $\mathbf{G}_m^2$, and $\mathcal{G} = \overline{\mathbf{Q}}_\ell$. Define $f_{i,j}(b_i) = b_i^j$ and Y be the subvariety of $\mathbf{G}_m^n$ defined by the equations

$$\sum b_i = \cdots = \sum b_i^m = 0.$$

One has $\dim V_{\overline{\mathbf{F}}_q} = n-m$ for q large enough. Then (TI) is satisfied and Proposition 11.2 states that

$$\sum_{\substack{b_1,\dots,b_n\in\mathbf{F}_q^\times \\ \sum b_i=\cdots=\sum b_i^m=0}} \!\!\cdots\! \sum \left| \sum_{s\in\mathbf{F}_q^\times} \prod_{i=1}^{n} \mathrm{Kl}_2(b_i s;q) \right|^2 = \sum_{\substack{b_1,\dots,b_n\in\mathbf{F}_q^\times \\ \sum b_i=\cdots=\sum b_i^m=0}} \!\!\cdots\! \sum \sum_{s\in\mathbf{F}_q^\times} \prod_{i=1}^{n} |\mathrm{Kl}_2(b_i s;q)|^2 + O\big(q^{(n-m+1)-1/2}\big),$$

provided $m \leqslant (n-3)/2$.

Proof. We will omit the indication of the finite field, which is always k, in the notation for trace functions. Opening the square, we have

$$\sum_{(\varrho,\boldsymbol{x})\in Y(k)}\Bigl|\sum_{s\in A(k)} t_*(\varrho,s)\prod_{i=1}^{n} t_i(\varrho,x_i,s)\Bigr|^2$$

$$= \sum_{(\varrho,\boldsymbol{x})\in Y(k)}\sum_{s\in A(k)} |t_*(\varrho,s)|^2\prod_{i=1}^{n}|t_i(\varrho,x_i,s)|^2 \qquad (11.3)$$

$$+ \sum\sum_{\substack{s_1,s_2\in A(k)\\ s_1\neq s_2}}\sum_{(\varrho,\boldsymbol{x})\in Y(k)} t_*(\varrho,s_1)\overline{t_*(\varrho,s_2)}\prod_{i=1}^{n} t_i(\varrho,x_i,s_1)\overline{t_i(\varrho,x_i,s_2)}.$$

We detect the condition $(\varrho,\boldsymbol{x})\in Y(k)$ through additive characters. Thus, let ψ a non-trivial character of k. For $\boldsymbol{x}=(x_i)_{i\in S}\in\mathcal{C}(k)$, we have

$$\delta_{(\varrho,\boldsymbol{x})\in Y(k)} = \prod_{j=1}^{m}\frac{1}{|k|}\sum_{\lambda_j\in k}\psi\bigl(\lambda_j\Sigma_j(\varrho,\boldsymbol{x})\bigr) = \frac{1}{|k|^m}\sum_{\boldsymbol{\lambda}\in k^m}\psi\Bigl(g_j(\varrho)+\sum_{j=1}^{m}\sum_{i\in S}\lambda_j f_{i,j}(x_i)\Bigr)$$

$$= \frac{1}{|k|^m}\sum_{\boldsymbol{\lambda}\in k^m}\psi(g_{\boldsymbol{\lambda}}(\varrho))\prod_{i=1}^{n}\psi(f_{i,\boldsymbol{\lambda}}(x_i)),$$

where $\boldsymbol{\lambda}=(\lambda_j)_{j\leqslant m}$, and

$$g_{\boldsymbol{\lambda}}(\varrho)=\sum_{j=1}^{m}\lambda_j g_j(\varrho), \qquad f_{i,\boldsymbol{\lambda}}(x_i)=\sum_{j=1}^{m}\lambda_j f_{i,j}(x_i).$$

Thus the second sum on the right-hand side of (11.3) is equal to

$$\frac{1}{|k|^m}\sum\sum_{\substack{s_1,s_2\in A(k)\\ s_1\neq s_2}}\sum_{\boldsymbol{\lambda}\in k^m}\sum_{(\varrho,\boldsymbol{x})\in\mathcal{C}(k)}\psi(g_{\boldsymbol{\lambda}}(\varrho))t_*(\varrho,s_1)\overline{t_*(\varrho,s_2)}$$

$$\times\prod_{i\in S} t_i(\varrho,x_i,s_1)\overline{t_i(\varrho,x_i,s_2)}\psi(f_{i,\boldsymbol{\lambda}}(x_i))$$

$$=\frac{1}{|k|^m}\sum\sum_{\substack{s_1,s_2\in A(k)\\ s_1\neq s_2}}\sum_{\boldsymbol{\lambda}\in k^m}\sum_{\varrho\in B(k)}\psi(g_{\boldsymbol{\lambda}}(\varrho))t_*(\varrho,s_1)\overline{t_*(\varrho,s_2)}$$

$$\times\prod_{i\in S}\Biggl(\sum_{x_i\in C_{i,\varrho}(k)} t_i(\varrho,x_i,s_1)\overline{t_i(\varrho,x_i,s_2)}\psi(f_{i,\boldsymbol{\lambda}}(x_i))\Biggr).$$

For $s_1\neq s_2$, it follows from the twist-independence assumption and the Riemann Hypothesis (Proposition 5.1 and (5.1)) that for each $i\in S$, we have

$$\sum_{x_i\in C_{i,\varrho}(k)} t_i(\varrho,x_i,s_1)\overline{t_i(\varrho,x_i,s_2)}\psi(f_{i,\boldsymbol{\lambda}}(x_i)) \ll |k|^{1/2}$$

and $t_*(\varrho, s_1)\overline{t_*(\varrho, s_2)} \ll 1$ for all $\varrho \in B(k)$. Hence the sum above is $\ll |k|^{\dim B + |S|/2 + 2}$, which concludes the proof. □

12. Parameterization of strata

The goal of this section is to give a convenient parameterization of the irreducible components of the strata of the stratification (X_j) (Definition 7.6).

Let j be an integer with X_j non-empty. Let $X \subset X_j \subset \mathbf{A}^{2l}$ be an irreducible component of X_j over $\mathbf{Z}$ which intersects the characteristic zero part. Let $\overline{\eta}$ be a geometric generic point of X.

We will show that X is the projection of a space defined by equations of a certain explicit type; more precisely, these will be exactly of the type that can be handled using Lemma 11.2, allowing us to evaluate the sums that appear in Lemma 11.1. To describe these equations and to perform an inductive process, where we express better and better approximations of X as the image of such space, we need to package certain data, which we do using the following definitions.

Definition 12.1. A *perspective datum* Π on X is a tuple

$$\Pi = \big(m, S, B, (C_i), (b_i), (f_{i,j}), (g_j)\big)$$

where

- $m \geqslant 0$ is an integer;
- $S \subseteq \{1, \dots 2l\}$;
- B is a separated scheme of finite type over $\bar{\mathbf{Q}}$;
- $(C_i)_{i \in S}$ is a family of relative curves over B;
- $(b_i)_{i \in S}$ is a family of functions $b_i : B \to \mathbf{A}^1$ if $i \notin S$ and $b_i : C_i \to \mathbf{A}^1$ if $i \in S$, such that if $i \in S$, the function b_i is not constant on any irreducible component of any geometric fiber of $C_i \to B$;
- $(f_{i,j})_{\substack{i \in S \\ 1 \leqslant j \leqslant m}}$ is a family of functions $f_{i,j} : C_i \to \mathbf{A}^1$;
- $(g_j)_{1 \leqslant j \leqslant m}$ is a family of functions $g_j : B \to \mathbf{A}^1$.

To simplify the notation, we will sometimes write $\Pi \cdot m, \dots, \Pi \cdot (g_j)$ for the corresponding data.

Let Π be a perspective datum over X. We denote $\mathcal{C}_\Pi$ the fiber product over B of the curves C_i for $i \in S$, and $\mathcal{Y}_\Pi$ the subvariety of $\mathcal{C}_\Pi$ defined as the zero locus of the functions

$$g_j + \sum_{i \in S} f_{i,j}$$

for $1 \leqslant j \leqslant m$, where we extend the functions $f_{i,j}$ and the functions g_j by pullback to $\mathcal{C}_\Pi$.

A *perspective* over X is a triple $(\Pi, Y, \bar{\gamma})$ where

- Π is a perspective datum on X;
- Y is an irreducible component of $\mathcal{Y}_\Pi$;
- $\bar{\gamma}$ is a geometric point of Y;

such that the morphism $g\colon \mathcal{Y}_\Pi \to \mathbf{A}^{2l}$ defined by $(b_1, \dots, b_{2l})$ induces a quasi-finite morphism

$$Y - g^{-1}(\mathcal{V}^\Delta) \to \mathbf{A}^{2l} - \mathcal{V}^\Delta$$

which maps $\bar{\gamma}$ to $\bar{\eta}$.

The goal of this section will be to construct a perspective on X where Y is irreducible and the image of the map $Y \to \mathbf{A}^{2l}$ is X. More precisely, the main result is the following:

Theorem 12.2. *There exists a perspective $(\Pi, Y, \bar{\gamma})$ on X such that Y is irreducible, $\bar{\gamma}$ is a geometric generic point of Y, and*

$$2l - |\Pi \cdot S| + 2\Pi \cdot m \leqslant 4(2l - \dim(X)).$$

The reader is encouraged to first finish reading the proof of the main theorems of this paper, assuming that this statement holds, since this will illustrate how the perspective data is exploited in the final steps.

The basic strategy is the following:

(1) We start with a perspective with S as large as possible, m as small as possible, but $\bar{\gamma}$ potentially a quite special point of Y (Lemma 12.3). We plan to reduce $\dim Y$ while keeping the growth of m and the loss of $|S|$ controlled by a step-by-step induction;
(2) At each step, we find some equations that are satisfied at $\bar{\gamma}$ but not at the generic point of Y (Lemmas 12.4 and 12.5);
(3) We construct a new perspective by adding these new equations (which may require also adjoining some new variables to B and C_i), lowering $\dim Y$ (Lemma 12.6). However, the solution set in Y of these new equations might not contain any irreducible components of the solution set in $\mathcal{Y}_\Pi$ of the new equations, since they may instead be absorbed into other irreducible components of $\mathcal{Y}_\Pi$. To deal with this, we must assume $\mathcal{Y}_\Pi = Y$;
(4) We can ensure that this condition holds by a Diophantine argument, which requires increasing $|S|$ (Lemma 12.9). This requires certain irreducibility assumptions on B and on the curves C_i, which we ensure in Lemma 12.10 by a direct construction;
(5) Finally, we prove Theorem 12.2 by showing that an induction involving all these steps terminates in a suitable perspective.

We begin by the exhibiting trivial examples of perspectives that will be used to start the induction process (or to terminate it in a trivial case).

Lemma 12.3.

(1) *The tuple*

$$\Pi_0 = \left(0, \{1, \ldots, 2l\}, \mathrm{Spec}(\bar{\mathbf{Q}}), \left(\mathbf{A}^1\right)_{1 \leqslant i \leqslant 2l}, (\mathrm{Id}_{\mathbf{A}^1})_{1 \leqslant i \leqslant 2l}, \emptyset, \emptyset\right)$$

is a perspective datum, and $(\Pi_0, X, \bar{\eta})$ *is a perspective;*

(2) *The tuple*

$$\Pi_1 = (0, \emptyset, X, \emptyset, (b_i | X), \emptyset, \emptyset)$$

is a perspective datum and $(\Pi_1, X, \bar{\eta})$ *is a perspective.*

Proof. This is an elementary check. In (1), we have $\mathcal{Y}_{\Pi_0} = \mathbf{A}^{2l}$, and the morphism $X \to \mathbf{A}^{2l}$ is quasi-finite, while in (2) we have $\mathcal{C}_{\Pi_1} = X$, with the same conclusion. □

In the next three lemmas, we begin the proof of the second step by studying how the roots of the polynomial $P_{\boldsymbol{b}}$, which are the r-coordinates of the points in the fiber $Z_{\boldsymbol{b}}$, can change under specialization.

Let F be an algebraically closed field. Let $r_0, b_1, \ldots, b_{2l}$ be elements of F. Formally, the polynomial $P_{\boldsymbol{b}} \in F[r]$ is the product

$$P_{\boldsymbol{b}} = \prod_{i=1}^{2l} (r + b_i) \prod_{(\zeta_i) \in \boldsymbol{\mu}_k^{2l}} \left(\sum_{i=1}^{2l} \zeta_i (r + b_i)^{1/k} \right).$$

This expansion makes sense unambiguously in an algebraic closure K of the complete local field $F((r - r_0))$, provided we fix a choice of k-th roots of $r + b_i$ in K. In particular, the order of vanishing of $P_{\boldsymbol{b}}$ at r_0 is the sum of the valuation of the factors, where the valuation on $F((r - r_0))$ is extended uniquely to K.

For $1 \leqslant i \leqslant 2l$, fix k-th roots $(r_0 + b_i)^{1/k}$ of $r_0 + b_i$ in F consistent with the choice of $(r + b_i)^{1/k}$ in K. Then the multiplicity of the factor

$$\sum_{i=1}^{2l} \zeta_i (r + b_i)^{1/k}$$

at r_0 is

$$\begin{cases} 0 & \text{if} \quad \displaystyle\sum_{i=1}^{2l} \zeta_i (r_0 + b_i)^{1/k} \neq 0, \\ 1/k & \text{if} \quad \displaystyle\sum_{i=1}^{2l} \zeta_i (r_0 + b_i)^{1/k} = 0 \text{ but } \sum_{\substack{1 \leqslant i \leqslant 2l \\ r_0 + b_i = 0}} \zeta_i \neq 0, \end{cases}$$

and otherwise it is equal to the multiplicity of the formal power series

$$\sum_{\substack{1\leqslant i\leqslant 2l\\ r_0+b_i\neq 0}} \zeta_i(r+b_i)^{1/k} \in F[[r]] \subset K$$

at r_0, when one choses the branch of $(r+b_i)^{1/k}$ with constant coefficient $(r_0+b_i)^{1/k}$.

Lemma 12.4. *Let R be a local integral domain with algebraically closed residue field F, and let K be an algebraic closure of the fraction field of R. Let $b_1,\dots,b_{2l}$ be elements of R, and $\bar{\boldsymbol{b}} \in F^{2l}$ their reductions modulo the maximal ideal. Let r_0 be some root of $P_{\bar{\boldsymbol{b}}} \in F[r]$. Assume that there exist at least two roots of $P_{\boldsymbol{b}}$ in K that reduce to r_0. For $1 \leqslant i \leqslant 2k$, fix a k−th root of r_0+b_i in F.*

Consider an algebraic closure $\widetilde{K}$ of $K((u))$. For $\boldsymbol{\zeta} \in \boldsymbol{\mu}_k^{2l}$, let $n(\boldsymbol{\zeta}) \geqslant 0$ be the multiplicity of

$$\sum_{i=1}^{2l} \zeta_i(r+b_i)^{1/k}$$

at r_0, as defined above.

There is no solution $(u_0, v_1, \dots, v_{2l}) \in R^{1+2l}$ of the system of equations

$$v_i^k = u_0 + b_i \tag{12.1}$$

$$u_0 + b_i = 0, \text{ for all } i \text{ such that } r_0 + b_i = 0 \tag{12.2}$$

$$\sum_{i=1}^{2l} \zeta_i v_i = 0, \text{ for all } \boldsymbol{\zeta} \text{ such that } \sum_{i=1}^{2l} \zeta_i (r_0+b_i)^{1/k} = 0 \in F \tag{12.3}$$

$$\sum_{\substack{1\leqslant i\leqslant 2l\\ r_0+b_i\neq 0}} \zeta_i v_i^{1-kt} = 0, \text{ if } n(\boldsymbol{\zeta}) \geqslant 2 \text{ and } 0 \leqslant t \leqslant n(\boldsymbol{\zeta})-1. \tag{12.4}$$

Proof. Suppose that there exists a solution $u_0 \in R$. We estimate from below the multiplicity of u_0 as a root of $P_{\boldsymbol{b}}$. For each factor of $P_{\boldsymbol{b}}$, the valuation at u_0 is at least the valuation of the corresponding factor of $P_{\bar{\boldsymbol{b}}}$ at r_0, hence by summing, the order of vanishing of $P_{\boldsymbol{b}}$ at u_0 is at least the order of vanishing of $P_{\bar{\boldsymbol{b}}}$ at r_0. But this contradicts the assumption that there exist two roots of $P_{\boldsymbol{b}}$ reducing to r_0. □

Lemma 12.5. *Let R be a local integral domain with algebraically closed residue field F containing a primitive k-th root of unity. Let $b_1,\dots,b_{2l}$ be elements of R and $\bar{\boldsymbol{b}}$ the reduction of $\boldsymbol{b}$ modulo the maximal ideal. Assume that* $\deg(P_{\bar{\boldsymbol{b}}}) < \deg(P_{\boldsymbol{b}})$.

(1) *If $\boldsymbol{b} \notin \mathcal{V}^{\Delta}$, then for any $\boldsymbol{\zeta} = (\zeta_i) \in \boldsymbol{\mu}_k^{2l}$ there exists an integer $n_{\boldsymbol{\zeta}} \geqslant 0$ such that*

$$\sum_{i=1}^{2l} \zeta_i \bar{b}_i^{n_{\boldsymbol{\zeta}}} \neq 0 \in F;$$

(2) *There exist some* $\boldsymbol{\zeta} = (\zeta_i) \in \boldsymbol{\mu}_k^{2l}$ *and some integer* ν *with* $0 \leqslant \nu \leqslant n_{\boldsymbol{\zeta}} - 1$ *such that*

$$\sum_{i=1}^{2l} \zeta_i b_i^n \neq 0 \in R.$$

Proof. Writing

$$\sum_{i=1}^{2l} \zeta_i (r+b_i)^{1/k} = r^{1/k} \sum_{i=1}^{2l} \zeta_i (1+b_i/r)^{1/k} = r^{1/k} \sum_{t=0}^{\infty} \left(\prod_{j=0}^{t-1} \frac{1/k - j}{1+j} \right) \left(\sum_{i=1}^{2l} \zeta_i b_i^t \right) \frac{1}{r^t}$$

for $(\zeta_i) \in \boldsymbol{\mu}_k^{2l}$, we first see that if (1) fails, then the left-hand side is identically 0, which implies that $\boldsymbol{b} \in \mathcal{V}^{\Delta}$. Then we obtain

$$\deg(P_{\boldsymbol{b}}) = 2l + k^{2l-1} - \sum_{(\zeta_i) \in \boldsymbol{\mu}_k^{2l}} m_{\boldsymbol{\zeta}},$$

where $m_{\boldsymbol{\zeta}} \geqslant 0$ is the largest integer such that

$$\sum_{i=1}^{2l} \zeta_i b_i^t = 0$$

for $0 \leqslant t \leqslant m_{\boldsymbol{\zeta}}$. If condition (2) does not hold, we therefore deduce that $\deg(P_{\boldsymbol{b}}) \leqslant \deg(P_{\bar{\boldsymbol{b}}})$, which contradicts the assumption. ☐

The next lemma is one of the key ingredients of the proof of Theorem 12.2.

Lemma 12.6. *Let* Π *be a perspective datum on* X *and* $(\Pi, Y, \bar{y})$ *a perspective. If* $\mathcal{Y}_{\Pi}$ *is irreducible, so that* $Y = \mathcal{Y}_{\Pi}$, *and* $\bar{y}$ *is not a geometric generic point of* $\mathcal{Y}_{\Pi}$, *then there exists a perspective* $(\Pi', Y', \bar{y}')$ *with*

$$\Pi' \cdot S = \Pi \cdot S, \quad \dim(\Pi' \cdot B) \leqslant \dim(\Pi \cdot B) + 1 \quad \dim(Y') < \dim(Y).$$

Proof. Let $\bar{\alpha}$ be a geometric generic point of Y, and $\bar{\beta}$ its image in $\mathbf{A}^{2l}$. By definition of a perspective, the fiber of $Y \to \mathbf{A}^{2l}$ over $\bar{\eta}$ is finite, and since it contains $\bar{y}$, it cannot contain the point $\bar{\alpha}$ that specializes to $\bar{y}$. Hence $\bar{\beta} \neq \bar{\eta}$, and since $\bar{\alpha}$ specializes to $\bar{y}$, it follows that $\bar{\beta}$ specializes to $\bar{\eta}$. In particular, we deduce that $\bar{\beta} \notin \mathcal{V}^{\Delta}$.

By definition, $\bar{y}$ is a geometric generic point of $X \subset X_j$. If $\bar{\beta}$ was a point of X_j, it would follow that $\beta = \eta$, which is not the case. Hence the fiber of $f : Z \to \mathbf{A}^{2l} - \mathcal{V}^{\Delta}$ over $\bar{\beta}$ has $\geqslant j+1$ points, whereas the fiber over $\bar{\eta}$ has j points.

Consider now the local ring R of the closure of $\bar{\beta}$ at the point $\bar{\eta}$. It has algebraically closed residue field. The polynomial $P_{\bar{\beta}} \in R[r]$ has $\geqslant j+1$ roots, and the specialization $P_{\bar{\eta}}$ has j roots. So either there exist two roots of $P_{\bar{\eta}}$ that have the same image in the residue field, or $\deg(P_{\bar{\beta}}) > \deg(P_{\bar{\eta}})$.

Case 1 (two roots coincide).

Let r_0 be the common reduction of at least two roots of $P_{\bar{\beta}}$. We will apply Lemma 12.4 to R and to this r_0. We define the multiplicity $n(\boldsymbol{\zeta})$ for $\boldsymbol{\zeta} \in \boldsymbol{\mu}_k^{2l}$ as in that lemma.

We consider the covering $\widetilde{B} \to B \times \mathbf{A}^1$, with coordinate u on $\mathbf{A}^1$, obtained by adjoining k-th roots v_i of $u+b_i$ for all $i \notin S$. We then define B' as the complement in $\widetilde{B}$ of the zero locus of $u+b_i$ for all $i \notin S$ such that $r_0+b_i \neq 0$. For $i \notin S$, the functions b_i define functions $B' \to \mathbf{A}^1$ by composing with the projection $B' \to B$.

For $i \in S$, we consider the curve $\widetilde{C}_i \to \widetilde{B}$ obtained from the base change of $C_i \times \mathbf{A}^1 \to B \times \mathbf{A}^1$ to B' by adjoining a k-th root v_i of $u+b_i$, so we have a diagram

$$\begin{array}{ccccc} C_i & \longleftarrow & C_i \times_B \widetilde{B} & \longleftarrow & \widetilde{C}_i \\ \downarrow & & \downarrow & & \\ B & \longleftarrow & \widetilde{B}. & & \end{array}$$

If $r_0+b_i \neq 0$, we define C_i' as the complement in C_i of the zero locus of $u+b_i$, and otherwise we define $C_i' = \widetilde{C}_i$. In all cases, the morphism $C_i' \to C_i$ allows us to define a function $b_i\colon C_i' \to \mathbf{A}^1$. The fibers of this function over a geometric point of B' project to geometric fibers of $C_i \to B$, hence irreducible components project to irreducible components, and so b_i is not constant on any irreducible component of any geometric fiber, since Π is a perspective datum.

We next define the scheme $\mathcal{C}' \to B'$ as the fiber product for $i \in S$ of the curves C_i' over B'.

There exists a lift $\bar{\gamma}'$ of $\bar{\gamma}$ in $\mathcal{C}'$ such that $u(\bar{\gamma}') = r_0$ (indeed, we can lift $\bar{\gamma}$ to the fiber product of the $\widetilde{C}_i$ over $\widetilde{B}$, and the resulting point lies in $\mathcal{C}'$ since $r_0+b_i = 0$ if $u+b_i = 0$). We fix such a lift. This choice defines canonical k-th roots of $u(\bar{\gamma}')+b_i(\bar{\gamma}') = r_0+b_i$, and we will use these later.

The functions g_j, $1 \leqslant j \leqslant m$ and $f_{i,j}$ of the perspective datum Π extend to B' and C_i', respectively, by composing with the projections $B' \to B$ and $C_i' \to C_i$. We will now add additional functions (corresponding to a change of the value of the parameter m).

Precisely, let $m' = m+m_1+m_2+m_3$, where m_1 (respectively m_2, m_3) is the number of equations (12.2) in Lemma 12.4 (respectively number of equations (12.3) or (12.4)). We define the additional functions g_j and $f_{i,j}$ for $m+1 \leqslant j \leqslant m'$, making a one-to-one correspondance between the values of j and the equations of those three types.

If j corresponds to an equation (12.2), *i.e.*, to an integer i with $1 \leqslant i \leqslant 2l$ such that $r_0 + b_i = 0$, then we define

$$\begin{cases} f_{i',j} = u + b_i & \text{for } i' \in S \text{ if } i' = i \\ f_{i',j} = 0 & \text{for } i' \in S \text{ if } i' \neq i \\ g_j = 0, \end{cases}$$

if $i \in S$, and otherwise we define

$$\begin{cases} f_{i',j} = 0 & \text{for } i' \in S \\ g_j = u + b_i. \end{cases}$$

If j corresponds to an equation (12.3), *i.e.*, to some $\boldsymbol{\zeta} \in \boldsymbol{\mu}_k^{2l}$ such that

$$\sum_{i=1}^{2l} \zeta_i (r_0 + b_i)^{1/k} = 0,$$

we define

$$\begin{cases} f_{i,j} = \zeta_i v_i & \text{for } i \in S \\ g_j = \sum_{i \notin S} \zeta_i v_i. \end{cases}$$

Finally, if j corresponds to an equation (12.4), *i.e.*, to $\boldsymbol{\zeta} \in \boldsymbol{\mu}_k^{2l}$ and t such that $n(\boldsymbol{\zeta}) \geqslant 2$ and $0 \leqslant t \leqslant n(\boldsymbol{\zeta}) - 1$, then we define

$$\begin{cases} f_{i,j} = \zeta_i v_i^{1-kt} & \text{if } i \in S \text{ and } r_0 + b_i \neq 0 \\ g_j = \sum\limits_{\substack{i \notin S \\ r_0 + b_i \neq 0}} \zeta_i v_i^{1-kt}, \end{cases}$$

(note that by the definition of C_i', the function v_i is non-vanishing). We now have defined the perspective datum

$$\Pi' = (m', S, B', (C_i')_{i \in S}, (b_i), (f_{i,j})_{\substack{i \in S \\ 1 \leqslant j \leqslant m'}}, (g_j)_{1 \leqslant j \leqslant m'}).$$

The associated variety, *i.e.*, the vanishing locus $\mathcal{Y}'$ of

$$g_j + \sum_{i \in S} f_{i,j}$$

for $1 \leqslant j \leqslant m'$, contains $\bar{\gamma}'$ by construction (see Lemma 12.4 again). Let Y' be an irreducible component of $\mathcal{Y}'$ containing $\bar{\gamma}'$. We claim that $(\Pi', Y', \bar{\gamma}')$ is the required perspective.

First, for $y \in \mathcal{Y}$, the points of the fiber of $\mathcal{Y}' \to \mathcal{Y}$ over y are determined by the value of the function u on $\mathcal{Y}'$, whose values lie in the set of roots of the polynomial

$P_{\boldsymbol{b}(y)}$. In particular, the fiber is finite, and hence $\mathcal{Y}'$ is quasi-finite over $\mathcal{Y}$. It follows on the one hand that Y' has dimension $\leqslant \dim(Y)$, and on the other hand that Y' is quasi-finite over $\mathbf{A}^{2l} - \mathcal{V}^{\Delta}$. So $(\Pi', Y', \bar{\gamma}')$ is a perspective.

We have $\dim(B') \leqslant \dim(B) + 1$. It remains therefore to check that $\dim(Y') < \dim(Y)$. We have already observed that $\dim(Y') \leqslant \dim(Y)$. Suppose the dimensions were equal. Then, since $Y' \to Y$ is quasi-finite, the geometric generic point $\bar{\gamma}'$ would map to $\bar{\alpha}$ in Y, and therefore to $\bar{\beta}$ in $\mathbf{A}^{2l}$. By applying finally Lemma 12.4, we obtain a contradiction: since two roots of $P_{\bar{\beta}}$ reduce to the same root of $P_{\bar{\eta}}$, there cannot be solutions in R of the system of equations (12.2), (12.3), (12.4), whereas this is exactly what we obtain from the fact that $\bar{\beta}$ is the image of $\bar{\gamma}'$.

Case 2 (the degree drops).

We now consider instead Lemma 12.5, and define integers $n_{\boldsymbol{\zeta}}$ for $\zeta \in \boldsymbol{\mu}_k^{2l}$ as the least integer $\geqslant 0$ such that

$$\sum_{i=1}^{2l} \zeta_i \bar{b}_i^{n_\zeta} \neq 0$$

at $\bar{\eta}$ (this exists by statement (1) in the lemma). We define $m' = m + m_1$, where m_1 is the number of pairs $(\boldsymbol{\zeta}, \nu)$ with $\boldsymbol{\zeta} \in \boldsymbol{\mu}_k^{2l}$ and $0 \leqslant \nu \leqslant n_{\boldsymbol{\zeta}}$. For $m+1 \leqslant j \leqslant m'$, corresponding in one-to-one fashion to $(\boldsymbol{\zeta}, \nu)$, we define

$$\begin{cases} f_{i,j} = \zeta_i b_i^{\nu} & \text{for } i \in S \\ g_j = \sum_{i \notin S} \zeta_i b_i^{\nu}. \end{cases}$$

Then $\Pi' = (m', S, B, (C_i), (b_i), (f_{i,j})_{\substack{i \in S \\ 1 \leqslant j \leqslant m'}}, (g_j)_{1 \leqslant j \leqslant m'})$ is a perspective datum (since the b_i have not changed, the non-constancy condition is also unchanged). The point $\bar{\gamma}$ belongs to the associated variety $\mathcal{Y}' \subset \mathcal{Y}_\Pi \subset \mathcal{C}_\Pi$ (by definition of $n_{\boldsymbol{\zeta}}$), so $(\Pi', Y', \bar{\gamma})$ is a perspective, where Y' is the irreducible component of $\mathcal{Y}'$ containing $\bar{\gamma}$. By Lemma 12.5, on the other hand, $\bar{\alpha}$ does not lie in $\mathcal{Y}'$, so all its irreducible components, including Y', have dimension $< \dim(\mathcal{Y}_\Pi) = \dim(Y)$. □

In the next lemma, we produce from a perspective another one with a specific value of the parameter m.

Lemma 12.7. *Let $(\Pi, Y, \bar{\gamma})$ be a perspective on X. There exists a perspective $(\Pi', Y', \bar{\gamma})$ such that*

$$\Pi' \cdot S = \Pi \cdot S, \quad \Pi' \cdot B = \Pi \cdot B, \quad \Pi' \cdot (C_i) = \Pi \cdot (C_i) \quad \Pi' \cdot (b_i) = \Pi \cdot (b_i)$$
$$\Pi' \cdot m = \dim(\Pi \cdot B) + |\Pi \cdot S| - \dim(Y)$$
$$Y' \text{ is isomorphic to } Y, \qquad \mathcal{Y}_\Pi \subset \mathcal{Y}'_\Pi \text{ as } B\text{-schemes}.$$

Proof. Let $m' = \dim(\Pi \cdot B) + |\Pi \cdot S| - \dim(Y)$. It is the codimension of Y in $\mathcal{C}_\Pi$. Let X be the subspace of $\Gamma(\mathcal{C}_\Pi, \mathcal{O})$ generated by the functions

$$h_j = g_j + \sum_{1 \leqslant j \leqslant m} f_{i,j}$$

for $1 \leqslant j \leqslant m$. We claim that for any integer ν with $0 \leqslant \nu \leqslant m'$, there exist $(\varphi_1, \dots, \varphi_\nu)$ in X such that all irreducible components of the zero locus $V(\varphi_1, \dots, \varphi_\nu)$ in $\mathcal{C}_\Pi$ that contain Y have codimension ν in $\mathcal{C}_\Pi$.

We prove this by induction on ν. The statement is true for $\nu = 0$. Assume that $\nu \leqslant m'$ and that the property holds for $\nu - 1$ and the functions $(\varphi_1, \dots, \varphi_{\nu-1})$. Let W be an irreducible component of the zero locus $V(\varphi_1, \dots, \varphi_{\nu-1})$. It has codimension $\nu - 1 < m' = \operatorname{codim}(Y)$ in $\mathcal{C}_\Pi$ so Y is a proper closed irreducible subset of W. Hence there exists j such that h_j does not vanish identically on W, and in particular the set of $\varphi \in X$ such that φ does not vanish on W is a non-empty Zariski-open subset of X. Taking intersection of these open sets, there exists $\varphi_\nu \in X$ such that φ_ν is non-vanishing on all irreducible components W containing Y. It follows that $(\varphi_1, \dots, \varphi_\nu)$ satisfy the induction assumption.

For $\nu = m'$, this means that all irreducible components of $V(\varphi_1, \dots, \varphi_{m'})$ containing Y have codimension $m' = \operatorname{codim}(Y)$ in $\mathcal{C}_\Pi$. Hence Y is one of the irreducible components of $V(\varphi_1, \dots, \varphi_{m'})$.

For $1 \leqslant \nu \leqslant m'$, write

$$\varphi_\nu = \sum_{1 \leqslant j \leqslant m} \alpha_{\nu,j} h_j.$$

We define

$$g'_\nu = \sum_{1 \leqslant j \leqslant m} \alpha_{\nu,j} g_j, \qquad f'_{i,\nu} = \sum_{1 \leqslant j \leqslant m} \alpha_{\nu,j} \sum_{i \in S} f_{i,j},$$

for $i \in S$ and $1 \leqslant \nu \leqslant m'$ so that

$$g'_\nu + \sum_{i \in S} f'_{i,\nu} = \varphi_\nu.$$

Then

$$\Pi' = (m', S, B, (C_i)_{i \in S}, (b_i), (f'_{i,j})_{\substack{i \in S \\ 1 \leqslant j \leqslant m'}}, (g'_j)_{1 \leqslant j \leqslant m'})$$

is a perspective datum on X; by construction Y is an irreducible component of $\mathcal{Y}_{\Pi'}$ and $\mathcal{Y}_\Pi \subset \mathcal{Y}'_\Pi$ as B-schemes, so $(\Pi', Y, \bar{\gamma})$ is a perspective with the desired properties. □

In the next lemma, we have a single perspective, so we do not use the selector notation.

Lemma 12.8. *Let $(\Pi, Y, \bar{\gamma})$ be a perspective on X. For any $T \subset S$ and $b \in B$, we put*

$$\widetilde{\Gamma}_{T,b} = \prod_{i \in T} \Gamma\big(C_{i,b}, \mathcal{O}_{C_{i,b}}\big), \qquad \Gamma_{T,b} = \prod_{i \in T} \big(\Gamma(C_{i,b}, \mathcal{O}_{C_{i,b}})/\kappa_b\big),$$

where the κ_b is the residue field at b. The spaces $\widetilde{\Gamma}_{T,b}$ and $\Gamma_{T,b}$ are κ_b-vector spaces. For $1 \leqslant j \leqslant m$, we denote $\boldsymbol{f}_{T,j,b} = (f_{i,j})_{i \in T} \in \widetilde{\Gamma}_{T,b}$.

Assume that S is not empty, that B is irreducible, and that the generic fiber of $C_i \to B$ is geometrically irreducible for all $i \in S$.

One of the following properties holds:

(a) *The scheme $\mathcal{Y}_\Pi$ has a unique geometrically irreducible component whose projection to B is dominant;*
(b) *There exists a proper subset $T \subset S$ such that the images of $(\boldsymbol{f}_{T,1,\eta}, \dots, \boldsymbol{f}_{T,m,\eta})$ span a subspace of $\Gamma_{T,\eta}$ of dimension $\leqslant m - (|S| - |T|)/2$, for η the generic point of B.*

Proof. There exists a number field and an open dense subset $\mathcal{O}$ of its ring of integers in a number field such that the persective datum is defined over $\mathcal{O}$. We fix one model of Π over $\mathcal{O}$, and we will use the same notation for its components as for the original objects over $\bar{\mathbf{Q}}$. We assume that property (b) does *not* hold and we will show that (a) holds. We will do this by studying fibers of $\mathcal{Y}_\Pi \to B$ over finite-valued field points of a suitable dense open subset of B, using the point-counting criterion for irreducibility over finite fields.

For $b \in B$, the condition that the all curves $C_{i,b}$ are geometrically irreducible is a constructible condition. So is the condition $(\boldsymbol{f}_{T,1,b}, \dots, \boldsymbol{f}_{T,m,b})$ generate a subspace of $\Gamma_{T,b}$ of dimension $> m - (|S| - |T|)/2$ for all proper subsets T of S.

By assumption, including the negation of (b), these properties both hold at the generic point, hence we can find a dense open subset B° where both properties hold.

Let $\mathrm{Spec}(\kappa) \to \mathrm{Spec}(\mathcal{O})$ be a finite-field valued point of $\mathrm{Spec}(\mathcal{O})$. Fix $b \in B^\circ(\kappa)$. Let ψ be a fixed non-trivial additive character of κ. We denote $V = \mathcal{Y}_{\Pi,b,\kappa}$. We compute $|V(\kappa)|$ using additive characters (as in the proof of Proposition 11.2). For $\boldsymbol{\lambda} \in \kappa^m$ and $x \in \mathcal{C}_\Pi(\kappa)$, we denote

$$f_{\boldsymbol{\lambda}}(x) = \sum_{j=1}^{m} \lambda_j \sum_{i \in S} f_{i,j}(x)$$

and

$$\xi(\boldsymbol{\lambda}) = \psi\left(\sum_{j=1}^{m} \lambda_j g_j(b)\right).$$

We have

$$|V(\kappa)| = \frac{1}{|\kappa|^m} \sum_{x\in\mathcal{C}_{\Pi,b}(\kappa)} \prod_{j=1}^{m} \sum_{\lambda\in\kappa} \psi\left(\lambda\left(g_j(x) + \sum_{i\in S} f_{i,j}(x)\right)\right)$$
$$= \frac{1}{|\kappa|^m} \sum_{x\in\mathcal{C}_{\Pi,b}(\kappa)} \prod_{j=1}^{m} \sum_{\lambda\in\kappa} \psi\big(\lambda g_j(b)\big)\psi\left(\lambda \sum_{i\in S} f_{i,j}(x)\right)$$
$$= \frac{1}{|\kappa|^m} \sum_{\boldsymbol{\lambda}\in\kappa^m} \xi(\boldsymbol{\lambda}) E(b;\boldsymbol{\lambda}),$$

where

$$E(b;\boldsymbol{\lambda}) = \sum_{x\in\mathcal{C}_{\Pi,b}(\kappa)} \psi(f_{\boldsymbol{\lambda}}(x)).$$

By definition of $\mathcal{C}_\Pi$ as a fiber product, we have the separation of variable formula

$$E(b;\boldsymbol{\lambda}) = \prod_{i\in S} \sum_{x\in C_{i,b}(\kappa)} \psi\left(\sum_{j=1}^{m} \lambda_j f_{i,j}(x)\right).$$

Let

$$S_{\boldsymbol{\lambda}} = \left\{ i \in S \mid \sum_{j=1}^{m} \lambda_j f_{i,j} \text{ is constant on } C_{i,b} \right\} \subset S.$$

Applying the Weil bound for the exponential sums over $C_{i,b}(\kappa)$ (assuming the characteristic is larger than the degree of the functions $f_{i,j}$), it follows that

$$E(b;\boldsymbol{\lambda}) \ll |\kappa|^{|S_{\boldsymbol{\lambda}}|+(|S|-|S_{\boldsymbol{\lambda}}|)/2} = |\kappa|^{(|S|+|S_{\boldsymbol{\lambda}}|)/2}.$$

We now split the expression for $|V(\kappa)|$ above according to the value of $S_{\boldsymbol{\lambda}}$, and isolate the term corresponding to $S_{\boldsymbol{\lambda}} = S$ from the others. This gives $|V(\kappa)| = N_1 + N_2$, where

$$N_1 = \frac{1}{|\kappa|^m} \sum_{\substack{\boldsymbol{\lambda}\in\kappa^m \\ S_{\boldsymbol{\lambda}}=S}} \xi(\boldsymbol{\lambda}) E(b;\boldsymbol{\lambda}), \qquad N_2 = \frac{1}{|\kappa|^m} \sum_{\substack{\boldsymbol{\lambda}\in\kappa^m \\ S_{\boldsymbol{\lambda}}\neq S}} \xi(\boldsymbol{\lambda}) E(b;\boldsymbol{\lambda}).$$

Taking $T = S - \{i\}$ for a fixed $i \in S$ in the defining property of B°, we observe that the tuple $(\boldsymbol{f}_{T,1,b}, \ldots, \boldsymbol{f}_{T,m,b})$ generates a subspace of $\Gamma_{T,b}$ of dimension $> m - (|S| - |T|)/2 > m - 1/2$, hence are linearly independent in $\Gamma_{T,b}$, and thus are linearly independent in $\Gamma_{S,b}$. The condition $S_{\boldsymbol{\lambda}} = S$ arises then only when $\boldsymbol{\lambda} = 0$. Hence

$$N_1 = \frac{1}{|\kappa|^m} \prod_{i\in S} |C_{i,b}(\kappa)|.$$

Since $C_{i,b}$ is a geometrically irreducible curve (by the choice of B°), we have $|C_{i,b}(\kappa)| = |\kappa| + O(|\kappa|^{1/2})$ for all i. Hence

$$N_1 = |\kappa|^{|S|-m}\left(1 + O\Big(|\kappa|^{-1/2}\Big)\right)^{|S|} + O\left(|\kappa|^{-m+|S|-1/2}\right)$$
$$= |\kappa|^{|S|-m} + O\left(|\kappa|^{|S|-m-1/2}\right).$$

On the other hand, we have

$$N_2 \ll \frac{1}{|\kappa|^m}\sum_{\substack{T\subset S\\ T\neq S}} |\kappa|^{n(T)}|\kappa|^{(|S|+|T|)/2},$$

where $n(T)$ is the dimension of the κ-vector subspace of κ^m whose elements are all $\boldsymbol{\lambda}$ such that $S_{\boldsymbol{\lambda}} \subset T$. We have $n(T) = \ker(\varphi_T)$, where $\varphi_T : \kappa^m \to \Gamma_{T,b,\kappa}/\kappa$ is the linear map

$$\boldsymbol{\lambda} \mapsto \sum_{j=1}^{m} \lambda_j \boldsymbol{f}_{T,j} \ (\mathrm{mod}\, \kappa).$$

Since T is a proper subset of S, by the definition of B°, we must have $\dim \mathrm{Im}(\varphi_T) > m - \frac{|S|-|T|}{2}$, so that $n(T) < (|S|-|T|)/2$, which implies $n(T) \leqslant (|S|-|T|)/2 - 1/2$, so we derive

$$N_2 \ll |\kappa|^{-m+(|S|-|T|)/2+(|S|+|T|)/2-1/2} = |\kappa|^{|S|-m-1/2}.$$

We conclude that

$$|V(\kappa)| = |\kappa|^{|S|-m} + O\left(|\kappa|^{|S|-m-1/2}\right).$$

Applying this to finite extensions of κ and applying the Lang-Weil estimates, we conclude that V is geometrically irreducible.

Recalling that V was the fiber of $\mathcal{Y}_\Pi$ over an arbitrary point $b \in B^\circ(\kappa)$, we see that all the fibers of $\mathcal{Y}_\Pi$ over finite-field valued points of B° with sufficiently large characteristic are geometrically irreducible, so all the fibers of $\mathcal{Y}_\Pi$ over points of B° are geometrically irreducible. Therefore $\mathcal{Y}_\Pi$ has a unique geometrically irreducible component that is dominant over B, concluding the proof that condition (a) holds. □

Lemma 12.9. *Let $(\Pi, Y, \bar{\gamma})$ be a perspective on X defined over an open subscheme* $\mathrm{Spec}(\mathcal{O})$ *of the ring of integers in a number field. Assume that S is not empty, that B is geometrically irreducible, that each C_i is irreducible and that the generic fiber of $C_i \to B$ is geometrically irreducible for all $i \in S$.*

If $\mathcal{Y}_\Pi$ is reducible and all irreducible components of $\mathcal{Y}_\Pi$ are dominant over B, then there exists a perspective $(\Pi', Y', \bar{\gamma})$ on X such that $\dim Y' = \dim Y$ *and*

$$1 \leqslant |\Pi \cdot S| - |\Pi' \cdot S| \leqslant 2(\Pi \cdot m - \Pi' \cdot m).$$

Proof. We apply Lemma 12.8 to $(\Pi, Y, \bar{\gamma})$, and use the same notation. Since $\mathcal{Y}_\Pi$ is reducible and all its irreducible components are dominant over B, there are at least two irreducible components that are dominant over B. By Lemma 12.8, we conclude that there exists a proper subset $T \subset \Pi \cdot S$ such that the span of $(\boldsymbol{f}_{T,1,\eta}, \ldots, \boldsymbol{f}_{T,m,\eta})$ in Γ_T has dimension $\leqslant m - (|S| - |T|)/2$.

For $\boldsymbol{\lambda} \in \ker(\varphi_T)$ and $i \in T$, $\sum_{j=1}^m \lambda_j f_{i,j}$ is equal to an element of κ_η and hence a rational function on B. Let B^* be an open subset of B on which all these functions are defined. Because C_i is irreducible, $\sum_{j=1}^m \lambda_j f_{i,j}$ is equal to this function on B^* not just at the generic point, but everywhere.

Let m' be the dimension of the span X of $(\boldsymbol{f}_{T,1}, \ldots, \boldsymbol{f}_{T,m})$ in Γ_T. We have then

$$1 \leqslant |\Pi \cdot S| - |T| \leqslant 2(\Pi \cdot m - m').$$

Let $\widetilde{\mathcal{C}}$ be the fibre product of C_i for $i \in S - T$ with B^* over B. We have an evaluation map

$$\varphi_T : \mathbf{A}^m \to \Gamma_T$$

sending $(\lambda_i)_{i \in T}$ to

$$\sum_{j=1}^m \lambda_j \boldsymbol{f}_{T,j}.$$

We define $B' \subset \widetilde{\mathcal{C}}$ to be the common zero locus of the functions

$$\sum_{j=1}^m \lambda_j \left(g_j + \sum_{i \in S} f_{i,j} \right)$$

for all $\boldsymbol{\lambda}$ in $\ker(\varphi_T)$. These expressions are indeed well-defined functions on $\widetilde{\mathcal{C}}$ because, as we saw earlier

$$\sum_{j=1}^m \lambda_j f_{i,j}$$

is equal to a function on B^* for $i \in T$ if $\boldsymbol{\lambda} \in \ker(\varphi_T)$.

Furthermore, we choose $f'_{i,j}$ in X for $i \in T$ and $1 \leqslant j \leqslant m'$ so that $f'_{i,j} = \sum_{\nu=1}^m \beta_{j,\nu} \boldsymbol{f}_{i,\nu}$ for $(\beta_{j,\nu})_{1 \leqslant j \leqslant m'}$ a set of elements of $\mathbf{A}^m$ that span its image X under φ_T. Define

$$g'_j = \sum_{\nu=1}^m \beta_{\nu,j} \left(g_\nu + \sum_{i \in S-T} f_{i,\nu} \right).$$

Then the tuple

$$\Pi' = \Big(m', T, B', \big(C_i \times_B B'\big)_{i \in T}, \big(b'_i\big)_{i \in T}, \big(f'_{i,j}\big), \big(g'_j\big) \Big)$$

is a perspective datum on X, where b'_i is the extension of b_i to $C'_i = C_i \times_B B'$ by pullback for $i \in T$, the composition $B' \to \tilde{C} \to C_i = \mathbf{A}^1$ if $i \in S - T$, and the projection $B' \to B \to \mathbf{A}^1$ otherwise.

By construction, the fiber product $\mathcal{C}_{\Pi'}$ is a locally closed subset of contained in $\mathcal{C}_{\Pi}$. The subscheme $\mathcal{Y}_{\Pi'}$ is an open subset of $\mathcal{Y}_{\Pi}$, because it has the same set of defining equations after restricting to an open subset B^* of B. Because the irreducible component Y was dominant over B, its restriction to this open subset has the same dimension, and because $\bar{\gamma}$ was dominant over B, it remains in this open subset as well. Hence $(\Pi', Y', \bar{\gamma})$ is the desired perspective on X. ☐

The last preparatory lemma constructs a perspective where the base B satisfies the assumptions of the last lemma.

Lemma 12.10. *Let $(\Pi, Y, \bar{\gamma})$ be a perspective on X. Then there exists a perspective $(\Pi', Y', \bar{\gamma}')$ such that*

$$\Pi' \cdot m = \Pi \cdot m, \qquad \Pi' \cdot S = \Pi \cdot S,$$

and $\dim(Y') \leq \dim(Y)$, *and moreover*

(a) *$\Pi' \cdot B$ is irreducible;*
(b) *For all $i \in S$, the curve $\Pi' \cdot C_i$ are irreducible and the fiber of $\Pi' \cdot C_i$ over the geometric generic point of $\Pi' \cdot B$ is irreducible;*
(c) *All irreducible components of $\mathcal{Y}_{\Pi'}$ are dominant over B, as is $\bar{\gamma}$.*

The strategy of the proof is to make several modifications to the given perspective datum to ensure that these three conditions hold. We will first replace B by an irreducible scheme, ensuring condition (a). We then pass to a finite cover of B over which generic geometrically irreducible components of C_i are defined and choose one for each i, ensuring condition (b). Finally we remove a closed subset from B, containing all the irreducible components that are not dominant over B, ensuring condition (c).

Proof. Let $\mathcal{A} \subset \mathcal{C}_{\Pi}$ be an irreducible component containing Y. Let B_0 be the schematic closure of the image of $\bar{\gamma}$ under the projection $Y \to B$. It is closed and irreducible. Let β be its generic point. Let $\beta' \to \beta$ be a finite extension such that all irreducible components of the generic fibers of the curves C_i for $i \in S$ are defined over β'. Let then $B' \to B_0$ be a finite flat morphism whose generic fiber is $\beta' \to \beta$ (we can construct such a morphism by taking a generator of the field extension β'/β, and multiplying it by a regular function on B_0 so that its minimal polynomial P becomes monic; then the cover B' of B_0 obtained by adjoining a root of P has the required property).

Fix a lift $\bar{\gamma}'$ of $\bar{\gamma}$ to $Y \times_B B'$. Let Y' be an irreducible component of $Y \times_B B'$ containing $\bar{\gamma}'$, and let $\mathcal{A}'$ be an irreducible componet of $\mathcal{C}_{\Pi} \times_B B'$ containing Y'. Because $\bar{\gamma}$ maps to the generic point β of B_0, $\bar{\gamma}'$ must map to the generic point β' of B' (the only point lying in the fiber), and so $Y' \to B'$ and $\mathcal{A}' \to B'$ are dominant maps. Because $\mathcal{A}'$ is an irreducible component of $\mathcal{C}_{\Pi} \times_B B'$, and maps dominantly to B', it follows that $\mathcal{A}'_{\beta'}$ is an irreducible component of the pullback $\mathcal{C}_{\beta'}$ of the product of the curves C_i to β'. Hence there are irreducible components $\widetilde{C}_{i,\beta'}$ of

$C_{i,\beta'}$ for $i \in S$, such that $\mathcal{A}'_{\beta'}$ is contained in the product of the $\widetilde{C}_{i,\beta'}$. Let C'_i be the closure of $\widetilde{C}_{i,\beta'}$. This is an irreducible curve over B'. We can pullback the functions b_i, g_j and $f_{i,j}$ to B' and C'_i, respectively. We have then constructed a perspective datum

$$\Pi' = \left(m, S, B', (C'_i), (b'_i), (f'_{i,j}), (g'_j)\right).$$

The irreducible component Y' is contained in $\mathcal{C}_{\Pi'}$, hence in $\mathcal{Y}_{\Pi'}$. Since the morphism $\mathcal{Y}_{\Pi'} \to \mathcal{Y}_{\Pi}$ is finite, it is an irreducible component of $\mathcal{Y}_{\Pi'}$. It contains $\bar{\gamma}$' and so maps dominantly onto B'.

Let B'' be the complement in B' of the closure of the images of all irreducible components of $\mathcal{Y}_{\Pi'}$ that are not dominant over B'. We can pullback the data $C'_i, b'_i, g'_j, f_{i,j}, Y''$ further to B''. This defines a perspective datum

$$\Pi'' = \left(m, S, B'', (C''_i), (b''_i), (f''_{i,j}), (g''_j)\right),$$

and a perspective $(\Pi'', Y'', \bar{\gamma}')$.

By construction, B' and B'' are geometrically irreducible. Since the curves C''_i are generically irreducible, and their geometric generic fibers are defined over β', they are generically geometrically irreducible. Because $\bar{\gamma}$ maps dominantly to B, $\bar{\gamma}'$ maps dominantly to B' and B''. Finally, all irreducible components of $\mathcal{Y}_{\Pi''}$ map dominantly to B'' by construction. □

We can now conclude this section.

Proof of Theorem 12.2. Consider the set $\mathcal{P}$ of perspectives $(\Pi, Y, \bar{\gamma})$ on X such that

$$2\dim(\Pi \cdot B) + 2\dim(Y) + |\Pi \cdot S| \leqslant 6l. \tag{12.5}$$

This set is nonempty by Lemma 12.3 (1), hence it contains some element where

$$\dim(Y) + |\Pi \cdot S|$$

is minimal.

Using Lemma 12.10, we obtain a perspective $(\Pi, Y, \bar{\gamma}) \in \mathcal{P}$ such that $\Pi \cdot B$ is geometrically irreducible, the curves $\Pi \cdot C_i$ are irreducible, the geometric generic fibers of $\Pi \cdot C_i$ are irreducible, and all irreducible components of $\mathcal{Y}_{\Pi}$ as well as $\bar{\gamma}$ are dominant over B. By Lemma 12.7, we may assume that

$$\Pi \cdot m = \dim(\Pi \cdot B) + |\Pi \cdot S| - \dim(Y)$$

(note that the last condition in Lemma 12.7 implies that all irreducible components of $\mathcal{Y}_{\Pi'}$ as well as $\bar{\gamma}'$ are dominant over B for the new perspective given by that lemma with input $(\Pi, Y, \bar{\gamma})$.)

We will then see that, except in a trivial case, a perspective with these properties satisfies the desired conclusion that $\Pi \cdot Y$ is irreducible, $\bar{\gamma}$ is the generic point of Y, and

$$2l - |\Pi \cdot S| + 2\Pi \cdot m \leqslant 4(2l - \dim(X)).$$

First, if Y is irreducible and $\bar{\gamma}$ is the generic point of Y, then because Y is quasi-finite over $\mathbf{A}^{2l}$, we have $\dim(Y) = \dim(X)$, hence

$$\begin{aligned} 2l - |\Pi \cdot S| + 2\Pi \cdot m &= 2\dim(\Pi \cdot B) + 2l + |\Pi \cdot S| - 2\dim(Y) \leqslant 8l - 4\dim(Y) \\ &= 4(2l - \dim(X)). \end{aligned}$$

Next assume that Y is irreducible and $\bar{\gamma}$ is not the generic point of Y. Then Lemma 12.6 provides a perspective $(\Pi', Y', \bar{\gamma}')$ with

$$|\Pi' \cdot S| = |\Pi \cdot S|, \qquad |\Pi' \cdot B| \leqslant |\Pi \cdot B| + 1, \qquad \dim(Y') < \dim(Y),$$

so

$$2\dim(\Pi' \cdot B) + \dim(Y') + |\Pi' \cdot S| \leqslant 6l$$

but satisfying

$$\dim(Y') + |\Pi' \cdot S| < \dim(Y) + |\Pi \cdot S|,$$

which contradicts the minimality of Π.

Suppose now that Y is reducible and $\Pi \cdot S$ is nonempty. Then Lemma 12.9 provides a perspective $(\Pi', Y', \bar{\gamma}')$ which satisfies $|\Pi' \cdot S| < |\Pi \cdot S|$, and moreover

$$\begin{aligned} \dim(Y) = \dim\left(Y'\right) &\geqslant \dim\left(\Pi' \cdot B\right) + \left|\Pi' \cdot S\right| - \Pi' \cdot m \\ &\geqslant \dim\left(\Pi' \cdot B\right) - \Pi \cdot m + \tfrac{1}{2}\left(|\Pi' \cdot S| + |\Pi \cdot S|\right) \qquad (12.6) \\ &= \dim\left(\Pi' \cdot B\right) - \dim(\Pi \cdot B) + \tfrac{1}{2}\left(|\Pi' \cdot S| - |\Pi \cdot S|\right) + \dim(Y) \end{aligned}$$

hence

$$2\dim\left(\Pi' \cdot B\right) - 2\dim(\Pi \cdot B) \leqslant |\Pi \cdot S| - \left|\Pi' \cdot S\right|,$$

which because of (12.5) implies

$$2\dim\left(\Pi' \cdot B\right) + 2\dim(Y) + \left|\Pi' \cdot S\right| \leqslant 6l.$$

On the other hand, we have

$$\dim\left(Y'\right) + \left|\Pi' \cdot S\right| < \dim(Y) + |\Pi \cdot S|,$$

again contradicting the assumption of minimality.

Finally, the remaining case when $\Pi \cdot S$ is empty is trivial: in that case, Y is a closed subscheme of $\Pi \cdot B$ so that

$$4\dim(X) \leqslant 4\dim(Y) \leqslant 2\dim(\Pi \cdot B) + 2\dim(Y) \leqslant 6l$$

and we may simply take the trivial perspective $(\Pi_1, X, \bar{\eta})$ of Lemma 12.3 (2), for which

$$2l - |\Pi_1 \cdot S| + 2\Pi_1 \cdot m = 2l \leqslant 4(2l - \dim(X)). \qquad \square$$

13. The generic statement

We continue with the previous notation. Fix $j \geqslant 0$. Let $X \subset X_j \subset \mathbf{A}^{2l} - \mathcal{V}^{\Delta}$ be an irreducible component of X_j over $\mathbf{Z}$ which intersects the characteristic zero part. Let $\overline{X}$ be the closure of X in $\mathbf{A}^{2l}$.

Fix a perspective $(\Pi, Y, \overline{\gamma})$ on X such that $\mathcal{Y}_{\Pi}$ is irreducible, $\overline{\gamma}$ is a geometric generic point of Y, and $2l - |S| + 2m \leqslant 4\operatorname{codim}_{\mathbf{A}^{2l}}(X)$, which exists by Theorem 12.2. By definition, all of the perspective data is defined over $\bar{\mathbf{Q}}$. However, by standard finiteness arguments, everything is necessarily defined over a finitely generated subring of $\bar{\mathbf{Q}}$, *i.e.*, over a ring $\mathcal{O}_K[1/N]$, where $\mathcal{O}_K$ is the ring of integers of a number field K and $N \geqslant 1$ is some integer. We will use the same notation Y, C_i, b_i, etc. to refer to the objects over this ring. Since, by assumption, $\mathcal{Y}_{\Pi,\bar{\mathbf{Q}}}$ is irreducible, and equal to $Y_{\bar{\mathbf{Q}}}$, we deduce that $\mathcal{Y}_{\Pi}$ is geometrically irreducible and equal to Y.

Because the geometric generic point of Y is a lift of the geometric generic point of X, the image of Y in $\mathbf{A}^{2l}$ is a dense subset of $\overline{X}$. For all but finitely many prime ideals π of $\mathcal{O}_K[1/N]$, with residue field denoted $\mathbf{F}_q$, the variety $Y_{\overline{\mathbf{F}}_q}$ is irreducible and nonempty, $X_{\overline{\mathbf{F}}_q}$ is irreducible and nonempty, and the map $Y_{\overline{\mathbf{F}}_q} \to \overline{X}_{\overline{\mathbf{F}}_q}$ is dominant. In the remainder of this section, we only consider finite fields $\mathbf{F}_q$ arising in this manner, and we also always assume that the characteristic of $\mathbf{F}_q$ is $> 2k + 1$.

Lemma 13.1. *Assume that* χ *has* CGM. *If* p *is large enough with respect to* (k, l, X) *and* $\dim(X_{\mathbf{Q}}) \geqslant (3l + 1)/2$, *then we have*

$$\begin{aligned}\sum_{y\in Y(\mathbf{F}_q)}\sum_{r\in\mathbf{F}_q^\times}&\left|\sum_{s\in\mathbf{F}_q^\times}\prod_{i=1}^{l}\mathrm{Kl}_k(r(s+b_i(y));\chi,q)\overline{\mathrm{Kl}_k(r(s+b_{i+l}(y));\chi,q)}\right|^2\\
&=\sum_{y\in Y(\mathbf{F}_q)}\sum_{r\in\mathbf{F}_q^\times}\sum_{s\in\mathbf{F}_q^\times}\left|\prod_{i=1}^{l}\mathrm{Kl}_k(r(s+b_i(y));\chi,q)\overline{\mathrm{Kl}_k(r(s+b_{i+l}(y));\chi,q)}\right|^2 \qquad (13.1)\\
&\quad+O\left(q^{\dim(X_{\mathbf{Q}})+3/2}\right),\end{aligned}$$

where the implied constant depends only on (Π, k, l).

Proof. We first fix $r \in \mathbf{F}_q^\times$. We apply Proposition 11.2 with data $(m, B, S, (C_i), (f_j), (g_j))$ coming from the perspective datum Π, $A = \mathbf{G}_m$, where the sheaf $\mathcal{F}_i$ is $[(b_i, s) \mapsto s(r + b_i)]^*\mathcal{K}\ell_{k,\psi}(\chi)$ and with the sheaf

$$\mathcal{G} = \bigotimes_{i\notin S}[(y, s) \mapsto s(r + b_i(y))^*]\mathcal{K}\ell_{k,\psi}(\chi).$$

Assumption (TI) holds by a Goursat-Kolchin-Ribet argument (see [17] and [10]). Indeed, each irreducible component of $C_{i,\varrho}$ is a geometrically irreducible curve on which b_i is a nonconstant function. The sheaf $\mathcal{F}_{i,\varrho,s_1} \otimes \mathcal{F}_{i,\varrho,s_2}^{\vee}$ is the pullback along b_i of the sheaf

$$\mathcal{H} = [b_i \mapsto (s_1(r + b_i))]^*\mathcal{K}\ell_{k,\psi}(\chi) \otimes [b_i \mapsto (s_1(r + b_i))]^*\mathcal{K}\ell_{k,\psi}(\chi)^{\vee}.$$

The monodromy group after pulling back along the map b_i is a finite index subgroup, so it suffices to show that no finite-index subgroup of the geometric monodromy group of $\mathcal{H}$ admits a one-dimensional irreducible component. However, by Goursat's lemma, the geometric monodromy group of $\mathcal{H}$ is a product of two copies of the monodromy group of $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$, acting by the tensor product of the standard representation with its dual. This group is connected, so has no proper finite-index subgroups, and does not admit a one-dimensional representation, which proves the claim.

The conductor of all the sheaves $\mathcal{F}_{i,\varrho,s}$, which are pullbacks of (shifted and translated) generalized Kloosterman sheaves are bounded by constants depending only on Π.

Applying Proposition 11.2 we obtain

$$\sum_{y\in Y(\mathbf{F}_q)}\left|\sum_{s\in\mathbf{F}_q^\times}\prod_{i=1}^{l}\mathrm{Kl}_k(r(s+b_i(y));\boldsymbol{\chi},q)\overline{\mathrm{Kl}_k(r(s+b_{i+l}(y));\boldsymbol{\chi},q)}\right|^2$$

$$=\sum_{y\in Y(\mathbf{F}_q)}\sum_{s\in\mathbf{F}_q^\times}\left|\prod_{i=1}^{l}\mathrm{Kl}_k(r(s+b_i(y));\boldsymbol{\chi},q)\overline{\mathrm{Kl}_k(r(s+b_{i+l}(y));\boldsymbol{\chi},q)}\right|^2$$
$$+O\left(q^{\dim B_{\mathbf{Q}}+|S|/2+2}\right),$$

where the implied constant depends only on (Π,k,l).

Summing over r, we get the formula (13.1), except that the error term is $O(q^{\dim B_{\mathbf{Q}}+|S|/2+3})$. However, since X is the vanishing set of m equations in a fiber product of $|S|$ curves over B, we have

$$\dim X_{\mathbf{Q}}\geqslant\dim B_{\mathbf{Q}}+|S|-m=\dim B_{\mathbf{Q}}+\frac{|S|}{2}+l-\left(l+m-\frac{|S|}{2}\right)$$
$$\geqslant\dim B_{\mathbf{Q}}+\frac{|S|}{2}+l-2(2l-\dim X)\geqslant\dim B_{\mathbf{Q}}+\frac{|S|}{2}+1/2,$$

where the last two inequalities holds by the assumption on the perspective and the assumption on $\dim X$, respectively. □

Let η be the generic point of $X_{\overline{\mathbf{F}}_q}$ and let $\bar{\eta}$ be a geometric generic point over η. Let η' be the the generic point of $Y_{\mathbf{F}_q}$. We fix a k-tuple $\boldsymbol{\chi}$ of characters of $\mathbf{F}_q^\times$.

Lemma 13.2. *Assume that* $\boldsymbol{\chi}$ *has* CGM. *We have*

$$\dim\mathrm{End}_{V_{\eta'\times\overline{\mathbf{F}}_q}}\left(\mathcal{K}_{\eta'\times\overline{\mathbf{F}}_q}\right)=\dim\mathrm{End}_{U_{\eta'\times\overline{\mathbf{F}}_q}}\left(\mathcal{R}^*_{\eta'\times\overline{\mathbf{F}}_q}\right).\tag{13.2}$$

Proof. Let Y° be the smooth locus of Y. The endomorphisms $\mathrm{End}_{V_{\eta'\times\overline{\mathbf{F}}_q}}(\mathcal{K}_{\eta'\times\overline{\mathbf{F}}_q})$ are the same as the endomorphisms of the pullback of $\mathcal{K}$ to $Y^\circ_{\overline{\mathbf{F}}_q}\times_{\mathbf{A}^{2l}} V$, because the monodromy representations of both sheaves are the same (as they are normal, with the same generic point). We calculate the endomorphisms by applying Lemma 11.1 obtaining the lim-sup of

$$q^{-\dim X-2}\sum_{y\in Y(\mathbf{F}_q)}\sum_{r\in\mathbf{F}_q^\times}\left|\sum_{s\in\mathbf{F}_q^\times}\prod_{i=1}^{l}\mathrm{Kl}_k(r(s+b_i(y));\,\chi,q)\overline{\mathrm{Kl}_k(r(s+b_{i+l}(y));\,\chi,q)}\right|^2.$$

We do the same for $\mathrm{End}_{U_{\eta'\times\overline{\mathbf{F}}_q}}(\mathcal{R}^*_{\eta'\times\overline{\mathbf{F}}_q})$, obtaining the lim-sup of

$$q^{-\dim X-2}\sum_{y\in Y(\mathbf{F}_q)}\sum_{r\in\mathbf{F}_q^\times}\sum_{s\in\mathbf{F}_q^\times}\left|\prod_{i=1}^{l}\mathrm{Kl}_k(r(s+b_i(y));\,\chi,q)\overline{\mathrm{Kl}_k(r(s+b_{i+l}(y));\,\chi,q)}\right|^2.$$

By Lemma 13.1, these two quantities are equal up to $O(q^{-1/2})$, and therefore their limsups are equal. □

In the remainder of this section, we will prove an analogous statement with $\overline{\eta}$ instead of η'. The method is to prove that

$$\dim\mathrm{End}_{V_{\overline{\eta}}}\left(\mathcal{K}_{\overline{\eta}}\right)=\dim\mathrm{End}_{V_{\eta'\times\overline{\mathbf{F}}_q}}\left(\mathcal{K}_{\eta'\times\overline{\mathbf{F}}_q}\right)\tag{13.3}$$

and

$$\dim\mathrm{End}_{U_{\eta'\times\overline{\mathbf{F}}_q}}\left(\mathcal{R}^*_{\eta'\times\overline{\mathbf{F}}_q}\right)=\dim\mathrm{End}_{U_{\overline{\eta}}}\left(\mathcal{R}^*_{\overline{\eta}}\right).\tag{13.4}$$

We will prove (13.3) immediately. The formula (13.4) is more difficult, and its proof will use vanishing cycles.

Proposition 13.3. *Assume that* χ *has* CGM. *For any extension* η' *of* η *we have*

$$\dim\mathrm{End}_{V_{\overline{\eta}}}\left(\mathcal{K}_{\overline{\eta}}\right)=\dim\mathrm{End}_{V_{\eta'\times\overline{\mathbf{F}}_q}}\left(\mathcal{K}_{\eta'\times\overline{\mathbf{F}}_q}\right).$$

Proof. Let G be the geometric monodromy group of $\mathcal{K}$, and let B be the set of distinct values of $b_1,\dots,b_{2l}$ at η. Then certainly the arithmetic monodromy group of $\mathcal{K}_{\eta\times\overline{\mathbf{F}}_q}$ is contained in $G^{|B|}$. By Goursat-Kolchin-Ribet, the geometric monodromy group of $\mathcal{K}_{\eta\times\overline{\mathbf{F}}_q}$ is $G^{|B|}$, so the arithmetic and geometric monodromy groups are equal. Therefore $\mathrm{Gal}(\overline{\eta}/\eta\times\overline{\mathbf{F}}_q)$ acts trivially on $\mathrm{End}_{V_{\eta\times\overline{\mathbf{F}}_q}}(\mathcal{K}_{\eta'\times\overline{\mathbf{F}}_q})$ as this action factors through the quotient of the arithmetic monodromy group by the geometric monodromy group. It follows that $\mathrm{Gal}(\overline{\eta}/\eta'\times\overline{\mathbf{F}}_q)$ acts trivially and so $\mathrm{End}_{V_{\eta'\times\overline{\mathbf{F}}_q}}(\mathcal{K}_{\eta'\times\overline{\mathbf{F}}_q})$, which is the space of invariants of that action, is equal to the whole space. □

In order to prove (13.4), we first introduce some notation. We write $\widetilde{\eta} = \eta \times \overline{\mathbf{F}}_q$. We consider the projective line $\mathbf{P}^1_{\widetilde{\eta}}$ with coordinate r. We denote by $\mathcal{O}^{et}$ the étale local ring of $\mathbf{P}^1_{\widetilde{\eta}}$ at ∞ and by K its field of fractions. We will often identify K (respectively a separable closure K^{sep} of K) with the corresponding spectra.

What follows is the key lemma.

Lemma 13.4. *With assumptions as above, the action of* $\mathrm{Gal}(K^{\text{sep}}/K)$ *on* $\mathcal{R}^*_{K^{\text{sep}}}$ *is unipotent.*

Note that to make sense of this action, we use the fact that the image of the natural morphism $\mathrm{Spec}(K) \to \mathbf{A}^{1+2l}$ has image in U, which follows from Lemma 7.4.

Proof. We denote by σ the special point of $\mathrm{Spec}(\mathcal{O}^{et})$. We consider the projective line $\mathbf{P}^1_{\mathcal{O}^{et}}$, with coordinate t, and denote by j (respectively by g) the open immersion $\mathbf{G}_{m,\mathcal{O}^{et}} \to \mathbf{P}^1_{\mathcal{O}^{et}}$ (respectively the open immersion $\mathbf{A}^1_{\mathcal{O}^{et}} \to \mathbf{P}^1_{\mathcal{O}^{et}}$).

We consider the lisse sheaf

$$\widetilde{\mathcal{K}} = \bigotimes_{1\leqslant i\leqslant l} \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(t(1+b_i/r)) \otimes \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})(t(1+b_{i+l}/r))^{\vee}$$

on $\mathbf{G}_{m,\mathcal{O}^{et}}$.

By the change of variable $t = rs$ and the proper base change theorem, the $\mathrm{Gal}(K^{\text{sep}}/K)$-action on $\mathcal{R}_{K^{\text{sep}}}$ is isomorphic to the action on $H^1(\mathbf{P}^1_{K^{\text{sep}}}, j_!\widetilde{\mathcal{K}})$. Since $\mathcal{R}^*_{K^{\text{sep}}}$ is a quotient of $\mathcal{R}_{K^{\text{sep}}}$, the lemma will follow if we prove that the action of $\mathrm{Gal}(K^{\text{sep}}/K)$ on $H^1(\mathbf{P}^1_{K^{\text{sep}}}, j_!\widetilde{\mathcal{K}})$ is unipotent.

By the long exact sequence for vanishing cycles, we have a long exact sequence

$$\cdots \to H^i\left(\mathbf{P}^1_{\sigma}, j_!\widetilde{\mathcal{K}}\right) \to H^i\left(\mathbf{P}^1_{K^{\text{sep}}}, j_!\widetilde{\mathcal{K}}\right) \to H^i\left(\mathbf{P}^1_{\sigma}, R\Phi j_!\widetilde{\mathcal{K}}\right) \to \cdots \tag{13.5}$$

For each i, we have an isomorphism

$$H^i\left(\mathbf{P}^1_{\sigma}, j_!\widetilde{\mathcal{K}}\right) = H^i\left(\mathbf{P}^1_{\sigma}, j_!\left(\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})^{\otimes l} \otimes (\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})^{\vee})^{\otimes l}\right)\right),$$

hence the $\mathrm{Gal}(K^{\text{sep}}/K)$-action on these spaces is trivial.

On the other hand, the vanishing cycle complex $R\Phi j_!\widetilde{\mathcal{K}}$ is zero away from the point at ∞ of $\mathbf{P}^1_{\sigma}$ (local acyclicity of smooth morphisms and lisseness of $j_!\widetilde{\mathcal{K}}$) and is zero at 0 (because of tame ramification and Deligne's semicontinuity theorem).

We therefore only need to understand $R\Phi j_!\widetilde{\mathcal{K}}$ at $t = \infty$. By the second part of Lemma 7.2, the local monodromy at infinity of $j_!\widetilde{\mathcal{K}}$ is isomorphic to that of a direct sum of sheaves of the form

$$\mathcal{L}_{\widetilde{\psi}}\left((t(1+b_1/r))^{1/k} + \sum_{i=2}^{2k} \varepsilon_i \zeta_i (t(1+b_i/r))^{1/k}\right).$$

Since $(1 + b_i/r)^{1/k}$ belongs to the étale local ring $\mathcal{O}^{et}$, this is isomorphic to the local monodromy of a direct sum of sheaves of the form $\mathcal{L}_{\widetilde{\psi}}(\gamma(r)t^{1/k})$. We have $\mathcal{L}_{\widetilde{\psi}}(\gamma(r)t^{1/k}) = \varpi_* \mathcal{L}_{\widetilde{\psi}}(\gamma(r)u)$ where ϖ is the finite covering $u \mapsto u^k$. We compute the local monodromy at ∞ of this sheaf, which we denote $\mathcal{G}$. This is a standard computation. We use the long exact sequence

$$\cdots \to H^i\left(\mathbf{P}^1_{\sigma}, g_!\mathcal{G}\right) \to H^i\left(\mathbf{P}^1_{K^{\mathrm{sep}}}, g_!\mathcal{G}\right) \to H^i\left(\mathbf{P}^1_{\sigma}, R\Phi g_!\mathcal{G}\right) \to \cdots$$

and distinguish three cases:

(1) If $\gamma(r) = 0$ in $\mathcal{O}^{et}$, then $\mathcal{G}$ is tamely ramified at ∞, so the vanishing cycles vanish;

(2) If $\gamma(r) \neq 0$ in $\mathcal{O}^{et}$ but $\gamma(r) = 0$ at the special point, then all H^i's with coefficients in $g_!\mathcal{G}$ in the above exact sequence vanish except

$$H^2(\mathbf{P}^1_{\sigma}, g_!\mathcal{G}),$$

which is one-dimensional with a trivial action of $\mathrm{Gal}(K^{\mathrm{sep}}/K)$; this implies that the action on $H^i(\mathbf{P}^1_{\sigma}, R\Phi g_!\mathcal{G})$ is trivial;

(3) If $\gamma(r) \neq 0$ at σ, then all cohomology groups in the sequence vanish by properties of the Artin-Schreier sheaves.

In any of the three cases, by local acyclicity of smooth morphisms we see that $R\Phi g_!\mathcal{G}$ vanishes outside the point at ∞, so knowing that $H^i(\mathbf{P}^1_{\sigma}, R\Phi g_!\mathcal{G})$ has trivial Galois action implies that the Galois action on the stalk at ∞ vanishes.

Since the vanishing cycle functor is additive and commutes with finite pushforward, we conclude that $\mathrm{Gal}(K^{\mathrm{sep}}/K)$ acts trivially on $H^i(\mathbf{P}^1_{\sigma}, R\Phi j_!\widetilde{\mathcal{K}})$ for all i, hence by the exact sequence (13.5), this group acts unipotently on $H^i(\mathbf{P}^1_{K^{\mathrm{sep}}}, j_!\widetilde{\mathcal{K}})$, as desired. □

Proposition 13.5. *Assume that* $\boldsymbol{\chi}$ *has* CGM. *We have*

$$\dim \mathrm{End}_{U_{\eta'\times\overline{\mathbf{F}}_q}}\left(\mathcal{R}^*_{\eta'\times\overline{\mathbf{F}}_q}\right) = \dim \mathrm{End}_{U_{\bar{\eta}}}\left(\mathcal{R}^*_{\bar{\eta}}\right).$$

Proof. We first note that we have an inclusion

$$\mathrm{End}_{U_{\bar{\eta}}}\left(\mathcal{R}^*_{\bar{\eta}}\right) \subset \mathcal{R}^*_{\bar{\eta}} \otimes \left(\mathcal{R}^*_{\bar{\eta}}\right)^{\vee}.$$

Moreover, we have a commutative triangle

$$\begin{array}{ccc} \mathrm{Gal}(K^{\mathrm{sep}}/K) & \longrightarrow & \pi_1(U_\eta)) \\ & \searrow^{\alpha} & \downarrow \\ & & \mathrm{Gal}(\bar{\eta}/\widetilde{\eta}) \end{array}$$

where α is surjective because K does not contain a finite extension of $\widetilde{\eta}$.

The fundamental group $\pi_1(U_\eta)$ acts on $\mathcal{R}^*_{\bar\eta} \otimes (\mathcal{R}^*_{\bar\eta})^\vee$ and the Galois group $\mathrm{Gal}(\bar\eta/\widetilde{\eta})$ acts on $\mathrm{End}_{U_{\bar\eta}}(\mathcal{R}^*_{\bar\eta})$, and these actions are compatible with the inclusion above.

By Lemma 13.4, the action of $\mathrm{Gal}(K^{\mathrm{sep}}/K)$ on $\mathcal{R}^*_{\bar\eta} \otimes (\mathcal{R}^*_{\bar\eta})^\vee$ is unipotent, hence the action of $\mathrm{Gal}(\bar\eta/\widetilde{\eta})$ on $\mathrm{End}_{U_{\bar\eta}}(\mathcal{R}^*_{\bar\eta})$ is also unipotent since α is surjective. But we know, by purity, that this action is semisimple, and it follows that the action $\mathrm{Gal}(\bar\eta/\widetilde{\eta})$ on $\mathrm{End}_{U_{\bar\eta}}(\mathcal{R}^*_{\bar\eta})$ is in fact trivial. In particular, we have

$$\dim \mathrm{End}_{U_{\eta'\times\overline{\mathbf{F}}_q}}\left(\mathcal{R}^*_{\eta'\times\overline{\mathbf{F}}_q}\right) = \dim \mathrm{End}_{U_{\bar\eta}}\left(\mathcal{R}^*_{\bar\eta}\right). \qquad \square$$

Finally, we can deduce:

Theorem 13.6. *Let X be an irreducible component of X_j which intersects the characteristic zero part. Assume that p is a prime sufficiently large with respect to (k, l, X). Let $\mathbf{F}_q$ be a finite field of characteristic p, and let $\overline{\eta}$ be the geometric generic point of $X_{\overline{\mathbf{F}}_q}$. Suppose that X has dimension at least $(3l+1)/2$. Let $\boldsymbol{\chi}$ be a k-tuple of characters of $\mathbf{F}_q^\times$ with Property* CGM. *Then we have*

$$\dim \mathrm{End}_{V_{\bar\eta}}\left(\mathcal{K}_{\bar\eta}\right) = \dim \mathrm{End}_{U_{\bar\eta}}\left(\mathcal{R}^*_{\bar\eta}\right).$$

Proof. Since the assertion is geometric, we may replace $\mathbf{F}_q$ by a finite extension that is a residue field of the base $\mathcal{O}_K[1/N]$ of the "spread-out" perspective. The equality then follows, when the characteristic of $\mathbf{F}_q$ is sufficiently large in terms of (k, l, X), by combining Proposition 13.3, Lemma 13.2 and Proposition 13.5. $\square$

14. Conclusion of the proof

We recall that we want to prove Theorem 7.7, which we restate for convenience:

Theorem 14.1. *Assume that $\boldsymbol{\chi}$ has* NIO. *If p is large enough, depending only on k, l, then for any $\boldsymbol{b} \in \mathbf{A}^{2l}(\mathbf{F}_q) - \mathcal{W}(\mathbf{F}_q)$, the natural morphism $\theta_{\boldsymbol{b}}$ is an isomorphism.*

Furthermore, each irreducible component of $\mathcal{R}^*_{\boldsymbol{b}}$ has rank greater than one.

Proof. Since $\boldsymbol{\chi}$ has NIO, by Lemma 6.3 there exists a character ξ, possibly over a finite extension $\mathbf{F}_{q^\nu}$ of $\mathbf{F}_q$, such that $\boldsymbol{\chi}' = \xi\boldsymbol{\chi}$ has CGM over $\mathbf{F}_{q^\nu}$. Consider $\boldsymbol{\chi}$ as a tuple of characters of $\mathbf{F}_{q^\nu}^\times$. Then $\mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi}') = \mathcal{L}_\xi \otimes \mathcal{K}\ell_{k,\psi}(\boldsymbol{\chi})$, and it follows that the auxiliary sheaves $\mathcal{K}$ and $\mathcal{R}^*$ for $\boldsymbol{\chi}$ are obtained from those associated to $\boldsymbol{\chi}'$ by twisting by a rank 1 sheaf $\mathcal{L}_\xi((r+b_1)\dots(r+b_l)(r+b_{l+1}^{-1})\dots(r+b_{2l})^{-1})$. Then the corresponding endomorphism rings (and the morphism $\theta_{\boldsymbol{b}}$) are the same for $\boldsymbol{\chi}$ and $\boldsymbol{\chi}'$. Up to renaming the field, this implies that we may as well assume that $\boldsymbol{\chi}$ has CGM over $\mathbf{F}_q$.

Let $\boldsymbol{b} \in \mathbf{A}^{2l}(\mathbf{F}_q) - \mathcal{W}(\mathbf{F}_q)$ be a point. Let j be the minimum j such that $\boldsymbol{b} \in X_j$. Let X be an irreducible component of X_j containing $\boldsymbol{j}$. By taking q sufficiently large, we may assume that X intersects the characteristic zero part. As the set of irreducible components is finite and depends only on k, l, the minimum value for q depends only on k, l.

If the dimension of X is less than $(3l+1)/2$, then $\boldsymbol{b} \in X \subseteq \mathcal{W}$.

Otherwise, let η be the generic point of X. Then by Theorem 13.6, taking q sufficiently large,

$$\dim \operatorname{End}_{V_{\bar{\eta}}} \left(\mathcal{K}_{\bar{\eta}}\right) = \dim \operatorname{End}_{U_{\bar{\eta}}} \left(\mathcal{R}^*_{\bar{\eta}}\right).$$

Because $\mathcal{W}_1$ has dimension $\leqslant l+1$, and $\dim X \geqslant (3l+1)/2 > l+1$ as $l > 1$, η is not contained in $\mathcal{W}_1$. By Lemma 10.1, Z is finite étale over $X_j - X_{j-1}$. Because the b_i are sections of Z, and $\boldsymbol{b}$ is a specialization of η inside $X_j - X_{j-1}$, any two of the b_i which are unequal over η must remain unequal over $\boldsymbol{b}$, so $\boldsymbol{b} \notin \mathcal{W}_1$.

So by Theorem 9.1, the natural map

$$\theta_{\boldsymbol{b}} : \operatorname{End}_{V_{\boldsymbol{b}}} \left(\mathcal{K}_{\boldsymbol{b}}\right) \to \operatorname{End}_{U_{\boldsymbol{b}}} \left(\mathcal{R}^*_{\boldsymbol{b}}\right)$$

is injective, hence by Proposition 10.3, $\theta_{\boldsymbol{b}}$ is an isomorphism.

Each irreducible component of $\mathcal{R}^*_{\boldsymbol{b}}$ is the image of an idempotent element of $\operatorname{End}_{U_{\boldsymbol{b}}}(\mathcal{R}^*_{\boldsymbol{b}})$, which because $\theta_{\boldsymbol{b}}$ is an isomorphism is induced by an idempotent element of $\operatorname{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}})$, and thus is equal to the weight one part of the cohomology of the image of that idempotent element of $\operatorname{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}})$. In other words, it is the weight one part of the cohomology of an irreducible component of $\mathcal{K}_{\boldsymbol{b}}$. Hence by Lemma 9.6, its rank is at least two. □

We finally can conclude the proof by showing how Theorem 14.1 allows us to give the estimates for complete sums used in the proof of our main theorems. In both cases, we use the fact (as remarked before the statements of Theorem 4.5 and 4.6) that we may assume that the function K is $\mathrm{Kl}_k(x; \boldsymbol{\chi}, q)$. By Lemma 7.1 and the Grothendieck-Lefschetz trace formula, for any $\boldsymbol{b} \notin \mathcal{V}^{\Delta}$, the function $\mathbf{R}$ is equal to minus the trace function of the sheaf $\mathcal{R}$, if the additive character ψ is chosen so that $\psi(x) = e(x/q)$ for $x \in \mathbf{F}_q$.

Proof of Theorem 4.5. We have defined $\mathcal{V}^{\Delta}$ and $\mathcal{W}$, and they satisfy the codimension bounds stated in the theorem (see (7.2)).

We need to estimate the complete sums

$$\Sigma_{II}(\boldsymbol{b}) = \sum_{r \in \mathbf{F}_q} |\mathbf{R}(r, \boldsymbol{b})|^2 - \sum_{s \in \mathbf{F}_q^\times} \sum_{r \in \mathbf{F}_q} |\mathbf{K}(sr, s\boldsymbol{b})|^2$$

for $\boldsymbol{b} \in \mathbf{F}_q^{2l}$. Since Kl_k is bounded, we have $\Sigma_{II}(\boldsymbol{b}) \ll q^3$ for all $\boldsymbol{b}$, which is the trivial bound (4.4).

If $\boldsymbol{b} \in \mathcal{W}(\mathbf{F}_q)$ and $\boldsymbol{b} \notin \mathcal{V}^{\Delta}(\mathbf{F}_q)$, then we obtain $\Sigma_{II}(\boldsymbol{b}) \ll q^2$ by estimating the two terms in Σ_{II} separately, and using the Riemann Hypothesis together with the fact that the $\mathcal{R}$-sheaf is mixed of weights $\leqslant 1$ on $\mathbf{A}^{2l} - \mathcal{V}^{\Delta}$, and the $\mathcal{K}$-sheaf is pure of weight 0. This proves (4.5).

Now assume that $\boldsymbol{b} \notin \mathcal{W}(\mathbf{F}_q)$. By Theorem 14.1, the Frobenius-equivariant map

$$\theta_{\boldsymbol{b}}\colon \operatorname{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}}) \to \operatorname{End}_{U_{\boldsymbol{b}}}\big(\mathcal{R}_{\boldsymbol{b}}^*\big)$$

is an isomorphism. In particular the Frobenius automorphism of $\mathbf{F}_q$ has the same trace on both spaces. The trace on $\operatorname{End}_{U_{\boldsymbol{b}}}(\mathcal{R}_{\boldsymbol{b}}^*)$ is, by the Grothendieck-Lefschetz trace formula, equal to

$$\sum_{r\in\mathbf{F}_q} |\mathbf{R}(r,\boldsymbol{b})|^2 + O\big(q^{3/2}\big),$$

where the error term arises from the contribution of the H_c^1-cohomology and of the weight < 1 part of $\mathcal{R}$. Similarly, the trace of Frobenius on $\operatorname{End}_{V_{\boldsymbol{b}}}(\mathcal{K}_{\boldsymbol{b}})$ is equal to

$$\sum_{s\in\mathbf{F}_q^\times}\sum_{r\in\mathbf{F}_q} |\mathbf{K}(sr,s\boldsymbol{b})|^2 + O\big(q^{3/2}\big),$$

where the error term arises from the contribution of the H_c^1-cohomology. Comparing, we obtain (4.6).

It remains to observe that, in all these estimates, the implied constant depends only on the sum of the Betti numbers of the relevant sheaves. These are estimated in the usual way by reducing to expressions as exponential sums and applying the Betti number bounds of Bombieri-Katz (see [19, Theorem 12] and [21, Proposition 4.24] for the analogue argument in our previous paper). □

Proof of Theorem 4.6. We recall that we need to estimate

$$\Sigma_I(\boldsymbol{b}) = \sum_{r\in\mathbf{F}_q} \mathbf{R}(r,\boldsymbol{b})$$

(see (4.10)). Since Kl_k is bounded, we have $\Sigma_I(\boldsymbol{b}) \ll q^2$ for all $\boldsymbol{b}$, which is the trivial bound (4.12).

If $\boldsymbol{b} \in \mathcal{W}(\mathbf{F}_q)$ and $\boldsymbol{b} \notin \mathcal{V}^{\Delta}(\mathbf{F}_q)$, then we obtain $\Sigma_I(\boldsymbol{b}) \ll q^{3/2}$ because the $\mathcal{R}$-sheaf is of weights $\leqslant 1$ on $\mathbf{A}^{2l} - \mathcal{V}^{\Delta}$. This proves (4.13).

Finally, if $\boldsymbol{b} \notin \mathcal{W}(\mathbf{F}_q)$, then we obtain $\Sigma_I(\boldsymbol{b}) \ll q$ straightforwardly from Deligne's Riemann Hypothesis, since $\mathcal{R}^*$ is of weight 1 and has no geometrically trivial irreducible component (by Theorem 14.1 it does not even have rank 1 components), proving (4.14).

Again, the implied constants in these estimates depend only on the sum of the Betti numbers of the relevant sheaves, and are estimated by reducing to expressions as exponential sums and applying the Betti number bounds of Bombieri-Katz [19]. □

References

[1] V. BLOMER, É. FOUVRY, E. KOWALSKI, PH. MICHEL and D. MILIĆEVIĆ, *On moments of twisted L-functions*, Amer. J. Math **139** (2017), 707–768.

[2] V. BLOMER, É. FOUVRY, E. KOWALSKI, PH. MICHEL, D. MILIĆEVIĆ and W. SAWIN, "The Second Moment Theory of Families of L-Functions", Mem. Amer. Math. Soc., to appear.

[3] D. A. BURGESS, *On character sums and L-series, I*, Proc. London Math. Soc. **12** (1962), 193–206.

[4] P. DELIGNE, *La conjecture de Weil, II*, Inst. Hautes Études Sci., Publ. Math. **52** (1980), 137–252.

[5] É. FOUVRY, E. KOWALSKI and PH. MICHEL, *Counting sheaves using spherical codes*, Math. Res. Lett. **20** (2013), 305–323.

[6] É. FOUVRY, E. KOWALSKI and PH. MICHEL, *Algebraic trace functions over the primes*, Duke Math. J. **163** (2014), 1683–1736.

[7] É. FOUVRY, E. KOWALSKI and PH. MICHEL, *Trace functions over finite fields and their applications*, In: "Colloquium De Giorgi 2013 and 2014", Colloquia, Ed. Sc. Norm. Super. Pisa, Vol. 5, 2014, 7–35.

[8] É. FOUVRY, E. KOWALSKI and PH. MICHEL, *Algebraic twists of modular forms and Hecke orbits*, Geom. Funct. Anal. **25** (2015), 580–657.

[9] É. FOUVRY, E. KOWALSKI, PH. MICHEL, C. SEKHAR RAJU, J. RIVAT and K. SOUNDARARAJAN, *On short sums of trace functions*, Ann. Inst. Fourier (Grenoble) **67** (2017), 423–449.

[10] É. FOUVRY, E. KOWALSKI and PH. MICHEL, *A study in sums of products*, Philos. Trans. Roy. Soc. A **373** (2015), 26.

[11] É. FOUVRY and PH. MICHEL, *Sur certaines sommes d'exponentielles sur les nombres premiers*, Ann. Sci. Éc. Norm. Sup. (4) **31** (1998), 93–130.

[12] J. B. FRIEDLANDER and H. IWANIEC, *Incomplete Kloosterman sums and a divisor problem* (with an appendix by B. J. Birch and E. Bombieri), Ann. of Math. (2) **121** (1985), 319–350.

[13] L. FU, "Étale Cohomology Theory", Nankai Tracts in Mathematics, Vol. 13, World Scientic, 2011.

[14] L. ILLUSIE, Y. LASZLO and F. ORGOGOZO (eds.), *Travaux de Gabber sur l'uniformisation locale et la cohomologie étale des schémas quasi-excellents*, Société Mathématique de France, Paris, 2014 (French). Séminaire à l'École Polytechnique 2006-2008. [Seminar of the Polytechnic School 2006-2008]; With the collaboration of Frédéric Déglise, Alban Moreau, Vincent Pilloni, Michel Raynaud, Joël Riou, Benoît Stroh, Michael Temkin and Weizhe Zheng; Astérisque No. 363-364 (2014).

[15] H. IWANIEC and E. KOWALSKI, "Analytic Number Theory", Colloquium Publications, Vol. 53, American Mathematical Society, Providence, RI, 2004.

[16] N. M. KATZ, "Gauss Sums, Kloosterman Sums, and Monodromy Groups", Annals of Mathematics Studies, Vol. 116, Princeton University Press, Princeton, NJ, 1988.

[17] N. M. KATZ, "Exponential Sums and Differential Equations", Annals of Mathematics Studies, Vol. 124, Princeton University Press, Princeton, NJ, 1990.

[18] N. M. KATZ, "Convolution and Equidistribution: Sato-Tate Theorems for finite-field Mellin Transforms", Annals of Mathematics Studies, Vol. 180, Princeton University Press, Princeton, NJ, 2012.

[19] N. M. KATZ, *Sums of Betti numbers in arbitrary characteristic*, Finite Fields Appl. **7** (2001), 29–44.

[20] N. M. KATZ, "Moments, Monodromy, and Perversity: a Diophantine Perspective", Annals of Mathematics Studies, Vol. 159, Princeton University Press, Princeton, NJ, 2005.

[21] E. KOWALSKI, PH. MICHEL and W. SAWIN, *Bilinear forms with Kloosterman sums and applications*, Ann. of Math. **186** (2017), 413–500.

[22] G. LAUMON, *Semi-continuité du conducteur de Swan* (*d'après P. Deligne*), In: "Caractéristique d'Euler-Poincaré", Astérisque, Vol. 83, Soc. Math. France, Paris, 1981, 173–219.

[23] J. MILNE, "Étale Cohomology", Princeton Math. Series, Vol. 33, Princeton University Press, Princeton, N.J., 1980.

[24] A. GROTHENDIECK and M. RAYNAUD, "Revêtements Étales et Groupe Fondamental", Lecture Notes in Mathematics, Vol. 224, Springer-Verlag, Berlin-New York, 1971.

[25] P. DELIGNE, "Cohomologie Étale", Lecture Notes in Mathematics, Vol. 569, Springer-Verlag, Berlin-New York, 1977.

[26] J. XU, *Stratification for multiplicative character sums*, Int. Math. Res. Not. IMRN (2018), DOI 10.1093/imrn/rny096.

[27] M. P. YOUNG, *The fourth moment of Dirichlet L-functions*, Ann. of Math. (2) **173** (2011), 1–50.

[28] R. ZACHARIAS, *Simultaneous non-vanishing for Dirichlet L-functions*, Ann. Inst. Fourier (Grenoble) **69** (2019), 1459–1524.

ETH Zürich – D-MATH
Rämistrasse 101
CH-8092 Zürich, Switzerland
kowalski@math.ethz.ch

EPFL/SB/TAN
Station 8 CH-1015
Lausanne, Switzerland
philippe.michel@epfl.ch

Columbia University
2990 Broadway
New York
NY, 10027, USA
sawin@math.columbia.edu