THE LARGE SIEVE, MONODROMY AND ZETA FUNCTIONS OF CURVES

E. KOWALSKI

ABSTRACT. We prove a large sieve statement for the average distribution of Frobenius conjugacy classes in arithmetic monodromy groups over finite fields. As a first application we prove a stronger version of a result of Chavdarov on the "generic" irreducibility of the numerator of the zeta functions in a family of curves with large monodromy.

1. INTRODUCTION

In [C], N. Chavdarov proves that, in an algebraic family $C \to U$ of smooth projective curves of genus g over a finite field $\mathbf{F}_q$, if the monodromy groups mod ℓ of the family are "as large as possible" for almost all ℓ, then the numerators $\det(1 - T \operatorname{Fr} \mid H^1(\overline{C}_u, \mathbf{Q}_\ell))$ of the zeta functions of the curves C_u of the family are "almost all" irreducible, and in fact have splitting field as large as compatible with the existence of the symplectic intersection pairing.

Chavdarov's method, as sketched in the introduction to [C], is analogue in principle to the method used by van der Waerden to show that "most" polynomials of given degree d with integer coefficients have splitting field as large as possible. This latter result was reproved in a simpler way and stronger form by Gallagher [G] using the large sieve inequalities as a new analytic tool. This suggests trying to apply similar ideas to Chavdarov's problem. In this paper, we show that this is indeed possible, to some extent. This proof also yields a much stronger result than [C] in many cases; see Theorem 6.1 and Theorem 6.2 for the exact statements. The proof uses some of the same tools, together with deep ideas of analytic number theory and some new uniform estimates for ℓ-adic Betti numbers which may be of independent interest.

The plan of this paper is the following: in the next two sections, we introduce the data involved and then state our main bilinear form estimate from which we derive a "large sieve" statement, essentially in the same way as the classical case. In the next two sections we prove the bilinear form estimate. First, Section 4 establishes some useful estimates for sums of ℓ-adic Betti numbers of "Artin-like" sheaves in various circumstances (restricted unless the base space is a curve). The proofs rely on the methods used by Katz in [K2] and [K1] – the difference being the parameters for which uniformity is required. Then in Section 5, we conclude the proof.

In the final sections we apply the sieve statement to prove our form of Chavdarov's Theorem. The statements at least are accessible (and of some interest) without knowledge of the techniques of étale cohomology required for the proof. Families of a fixed genus can be treated pretty quickly, but we expand some effort to obtain in some cases a uniform result that can give information about curves of genus g over $\mathbf{F}_q$ when q and g are simultaneously large (although g must be much smaller than q). We also draw a few easy consequences (Proposition 6.3 and Proposition 6.6), as illustrations of results which seem out of reach of Chavdarov's method, but are not meant as really important results in themselves.[1]

One can hope that other applications of this method will arise, by analogy with the situation in analytic number theory, where the ideas surrounding the large sieve have been extremely successful since the original discovery by Linnik.

[1] For analytic number theorists, let us mention that, incidentally, we also get a version of Gallagher's estimate [G] uniform in terms of the degree, see the final remark in Section 7.

The related paper [Ko2] applies Chavdarov's Theorem and some extra ingredients to the study of the characterization of abelian varieties over finite fields or number fields by their torsion fields. Also in [Ko3], we use the Betti number estimates and a uniform Chebotarev density theorem to study the density of quadratic twists of elliptic curves over function fields over finite fields with rank $\geqslant 2$.

Acknowledgments. N. Katz found a serious mistake in the first version of Section 4; thanks for pointing it out and for other enlightening remarks.

Background references. Since we are using two important themes in number theory which may not be equally known to the reader, we mention a few general references. For the large sieve, the reader may consult [B], [Mo] or [IK, Ch. 7]. For the approach to exponential sums via ℓ-adic methods and their applications (with which the author, for one, is not so well acquainted), we suggest [KS1, Ch. 9], [K2, Ch. 2,3], [K4, Ch. 1–3], [D2, Sommes trig.], or [IK, §11.11].

Notation. As usual, $|X|$ denotes the cardinality of a set, $\mathfrak{S}_g$ is the symmetric group on g letters. By $f \ll g$ for $x \in X$, or $f = O(g)$ for $x \in X$, where X is an arbitrary set on which f is defined, we mean synonymously that there exists a constant $C \geqslant 0$ such that $|f(x)| \leqslant Cg(x)$ for all $x \in X$. The "implied constant" is any admissible value of C. It may depend on the set X which is always specified or clear in context.

2. Preliminaries

Our main tool is a general estimate for a bilinear form made up from representations of a system of lisse $\overline{\mathbf{F}}_\ell$-sheaves on a variety over a finite field.

The first basic data is therefore a base variety $U/\mathbf{F}_q$, where as usual $\mathbf{F}_q$ denotes a finite field of characteristic p with q elements. We assume that U is smooth, affine, and geometrically connected of dimension $d \geqslant 1$.

We denote by $\overline{\eta}$ the geometric generic point of U and by $\overline{U}$ the variety U extended to $\overline{\mathbf{F}}_q$. We therefore have the arithmetic fundamental group $\pi_1(U, \overline{\eta})$ and the geometric fundamental group $\pi_1(\overline{U}, \overline{\eta})$. Those fit in an exact sequence

$$(2.1) \qquad 1 \longrightarrow \pi_1(\overline{U}, \overline{\eta}) \longrightarrow \pi_1(U, \overline{\eta}) \overset{d}{\longrightarrow} \mathrm{Gal}(\overline{\mathbf{F}}_q/\mathbf{F}_q) \simeq \hat{\mathbf{Z}} \longrightarrow 1.$$

For $n \geqslant 1$ and $u \in U(\mathbf{F}_{q^n})$, we denote by Fr_{u,q^n} the geometric Frobenius automorphism at u in $\pi_1(U, \overline{\eta})$, i.e., the image of the inverse of the canonical generator $x \mapsto x^{q^n}$ of the Galois group of $\mathbf{F}_{q^n}$ via the map

$$\mathrm{Gal}(\overline{\mathbf{F}}_{q^n}/\mathbf{F}_{q^n}) \to \pi_1(U, \overline{\eta})$$

induced from the inclusion $\mathrm{Spec}\,\mathbf{F}_{q^n} \to U$ which "is" u. In the above exact sequence we have then

$$d(\mathrm{Fr}_{u,q^n}) = -n.$$

In most of our results, the base field (i.e., q) will be considered fixed, although the results will be uniform in q so they can be applied to $U \times \mathbf{F}_{q^n}$ for any $n \geqslant 1$. So most of the time we just write Fr_u instead of $\mathrm{Fr}_{u,q}$ for $u \in U(\mathbf{F}_q)$.

We also denote generically by Fr the global geometric Frobenius automorphism, acting for instance on ℓ-adic cohomology groups.

We now come to the sheaves on U that we consider. We assume given a set Λ of primes $\neq p$, and for $\ell \in \Lambda$, a lisse sheaf $\tilde{\mathcal{F}}_\ell$ of $\mathbf{F}_\lambda$-vector spaces of (fixed) rank $r \geqslant 1$, where $\mathbf{F}_\lambda$ is a finite field of characteristic ℓ (the degree of which over $\mathbf{F}_\ell$ may depend on ℓ). The basic example is when we have lisse sheaves $\mathcal{F}_\ell$ of $\mathbf{Z}_\lambda$-modules and

$$\tilde{\mathcal{F}}_\ell = \mathcal{F}_\ell/\mathfrak{m}_\lambda \mathcal{F}_\ell,$$

where $\mathbf{Z}_\lambda$ is the ring of integers in a finite extension of $\mathbf{Q}_\ell$ with residue field $\mathbf{F}_\lambda$ and maximal ideal $\mathfrak{m}_\lambda$. However, we do not assume that $\tilde{\mathcal{F}}_\ell$ is of this type (of course, it will be in most applications).

Equivalently (and this is the most convenient viewpoint in terms of a first understanding at least), $\tilde{\mathcal{F}}_\ell$ "is" a representation

$$\rho_\ell \,:\, \pi_1(U, \overline{\eta}) \to GL(r, \mathbf{F}_\lambda).$$

From this description we can easily define the monodromy groups of $\tilde{\mathcal{F}}_\ell$, or of ρ_ℓ: the arithmetic monodromy group $G_\ell \subset GL(r, \mathbf{F}_\lambda)$ is the image of ρ_ℓ, and the geometric monodromy group G_ℓ^g is the image of the subgroup $\pi_1(\overline{U}, \overline{\eta})$. Thus from (2.1) we derive a commutative diagram with exact rows and surjective downward arrows:

$$(2.2) \qquad \begin{array}{ccccccccc}
1 & \longrightarrow & \pi_1(\overline{U}, \overline{\eta}) & \longrightarrow & \pi_1(U, \overline{\eta}) & \stackrel{d}{\longrightarrow} & \hat{\mathbf{Z}} & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \varphi \downarrow & & \\
1 & \longrightarrow & G_\ell^g & \longrightarrow & G_\ell & \stackrel{m}{\longrightarrow} & \Gamma_\ell & \longrightarrow & 1,
\end{array}$$

where Γ_ℓ is a finite commutative (cyclic) group.

In the case where the sheaves $\tilde{\mathcal{F}}_\ell$ arise by reduction of $\mathbf{Z}_\lambda$-sheaves $\mathcal{F}_\ell$, as described previously, one says that they form a *compatible system* if for every extension $\mathbf{F}_{q^n}/\mathbf{F}_q$, every $u \in U(\mathbf{F}_{q^n})$ and every $\ell \in \Lambda$, the reversed characteristic polynomial of Fr_{u,q^n} acting on $\mathcal{F}_\ell$, i.e., the polynomial

$$\det(1 - T\,\mathrm{Fr}_{u,q^n} \mid \mathcal{F}_\ell)$$

has coefficients in $\overline{\mathbf{Q}}$ and is independent of ℓ.

For any ℓ we will consider various sums involving irreducible (complex valued) linear representations of G_ℓ. For reasons that will become clear during the proof of the main bilinear form estimate (a certain phenomenon of "imprimitivity"), we can not use all representations, but must ensure that those used are suitably orthogonal when restricted to the subgroup G_ℓ^g.

For this we have the following lemma.

Lemma 2.1. (1) *Let π, π' be irreducible linear representations of G_ℓ. Then π and π' are equivalent when restricted to G_ℓ^g if and only if there exists a character $\psi \in \hat{\Gamma}_\ell$, the character group of Γ_ℓ, such that*

$$\pi = \pi' \otimes (\psi \circ m).$$

(2) *If π and π' are not equivalent when restricted to G_ℓ^g, the representation $\pi \otimes \tilde{\pi}'$ restricted to G_ℓ^g does not contain the trivial representation, where $\tilde{\pi}'$ is the contragredient of π'. Otherwise it contains the trivial representation with multiplicity equal to $|\hat{\Gamma}_\ell^\pi|$ where*

$$\hat{\Gamma}_\ell^\pi = \{\psi \in \hat{\Gamma}_\ell \mid \pi \simeq \pi \otimes (\psi \circ m)\}.$$

Proof. We will identify characters $\psi \in \hat{\Gamma}_\ell$ with characters of G_ℓ by $\psi(x) = \psi(m(x))$ for $x \in G_\ell$.

For any representation τ of G_ℓ, let $j(\tau)$ denote the multiplicity of the trivial representation in the restriction of τ to G_ℓ^g. This is given by

$$j(\tau) = \frac{1}{|G_\ell^g|} \sum_{x \in G_\ell^g} \mathrm{Tr}\,\tau(x)$$

$$= \frac{1}{|G_\ell^g|} \sum_{x \in G_\ell} \frac{1}{|\Gamma_\ell|} \sum_{\psi \in \hat{\Gamma}_\ell} \psi(m(x))\,\mathrm{Tr}\,\tau(x)$$

$$= \sum_{\psi \in \hat{\Gamma}_\ell} \frac{1}{|G_\ell|} \sum_{x \in G_\ell} \psi(x)\,\mathrm{Tr}\,\tau(x)$$

(by orthogonality of characters of Γ_ℓ), which is the sum of the multiplicities of the characters ψ of G_ℓ in τ. (This interpretation being of course also available by a simple application of Frobenius reciprocity).

Applying this to $\tau = \pi \otimes \tilde{\pi}'$, it follows that 1 is contained in the restriction of τ to G_ℓ^g if and only if there exists a ψ such that ψ is contained in $\pi \otimes \tilde{\pi}'$. However if that is the case, the

trivial representation is contained in $\pi \otimes \tilde{\pi}'\overline{\psi}$, but as π and $\pi'\psi$ are irreducible, this means that $\pi \simeq \pi' \otimes \psi$ as representations of G_ℓ. This shows the "only if" part of the lemma, and the other direction is trivial since ψ restricts to the trivial character of G_ℓ^g.

The first part of (2) is contained in the previous paragraph. The assertion about the multiplicity is also clear: if $\pi' = \pi \otimes \psi_0$ is a twist of π by a character ψ_0, the multiplicity $j(\pi \otimes \tilde{\pi}')$ is the sum of multiplicities of the characters ψ in $\pi \otimes \tilde{\pi}\overline{\psi}_0$, each of which is equal to 1 if $\pi \simeq \pi \otimes (\psi\psi_0)$, and 0 otherwise, i.e, it is equal to 1 if $\psi\psi_0 \in \hat{\Gamma}_\ell^\pi$ and 0 otherwise. So the total multiplicity is the number of elements in $\hat{\Gamma}_\ell^\pi$. $\qquad\square$

Remark 2.2. If $\pi \simeq \pi \otimes \psi$, we must have $\psi(x) = 1$ whenever $\operatorname{Tr}\pi(x) \neq 0$. But if $\deg(\pi) > 1$, it is well-known that there are elements $x \in G_\ell$ with $\operatorname{Tr}\pi(x) = 0$, and then the value of ψ is not determined. Of course, "in general", we have $\hat{\Gamma}_\ell^\pi = 1$, but (for instance), for any representation π of degree 2 of a dihedral group D_n, n even (of order $2n$), there is a character ψ with $\pi \simeq \pi \otimes \psi$.

Say that two representations of G_ℓ are geometrically equivalent if their restrictions to G_ℓ^g are equivalent, or (by the lemma) if and only if they differ by a twist by a character of Γ_ℓ. We now assume chosen a set Π_ℓ of representatives of the irreducible representations of G_ℓ for this equivalence relation. Using these and characters of Γ_ℓ one can parameterize all irreducible representations of G_ℓ as follows: they are of the form $\pi \otimes \psi$ where $\pi \in \Pi_\ell$ and $\psi \in \hat{\Gamma}_\ell$; the representation π is unique, but ψ is only unique up to multiplication by an element of the group $\hat{\Gamma}_\ell^\pi$ defined in (2) of the previous lemma.

This ambiguity requires us to control the size of those groups $\hat{\Gamma}_\ell^\pi$. We will assume that for all $\ell \in \Lambda$ and $\pi \in \Pi_\ell$, we have

$$(2.3) \qquad\qquad |\hat{\Gamma}_\ell^\pi| \leqslant \kappa$$

for some fixed $\kappa \geqslant 1$.

Here are useful cases when we can get such a bound.

Lemma 2.3. (1) *Assume that for all ℓ we have $G_\ell^g = SL(r, \mathbf{F}_\ell)$. Then (2.3) holds with $\kappa = r$.*

(2) *Assume that r is even and that for all ℓ we have $G_\ell^g = Sp(r, \mathbf{F}_\ell)$, the symplectic group for some non-degenerate alternating form $\langle \cdot, \cdot \rangle$ on $\mathbf{F}_\ell^r$, and that G_ℓ is a subgroup of the group $SSp(r, \mathbf{F}_\ell)$ of symplectic similitudes, i.e., for $g \in G_\ell$ we have $\langle gv, gw \rangle = m(g)\langle v, w \rangle$ for some $m(g) \in \mathbf{F}_\ell^\times$, called the multiplicator of g. Then (2.3) holds with $\kappa = 2$.*

Proof. (1) If π is an irreducible representation of G_ℓ and $\psi \in \hat{\Gamma}_\ell^\pi$, then ψ is trivial on the center Z_ℓ of G_ℓ. For any $x \in G_\ell$, we can write

$$x^r = (\det x)y$$

with $y \in SL(r, \mathbf{F}_\ell) = G_\ell^g$. Hence $\det(x) \in \mathbf{F}_\ell^\times \cap G_\ell \subset Z_\ell$ and therefore $\psi(\det(x)) = 1$, and $\psi(x^r) = \psi(\det(x))\psi(y) = 1$. So ψ is of order at most r, and since it is a character of a cyclic group, there are at most r such characters, giving (2.3) with $\kappa = r$.

(2) The argument is similar except that we now have $x^2 = m(x)y$ with $y \in Sp(r, \mathbf{F}_\ell)$ (since $m(ax) = a^2 x$ for scalar a), so $\psi(x)^2 = 1$. $\qquad\square$

We will often simply write (when ℓ is clearly specified, e.g. as a summation parameter occurring before)

$$\sum_\pi{}^* \alpha(\pi, \ell, \dots), \quad \sum_{\pi \neq 1}{}^* \alpha(\pi, \ell, \dots)$$

for, respectively, a sum over all the irreducible representations $\pi \in \Pi_\ell$ of G_ℓ^g or for a sum over all those which are non-trivial on G_ℓ^g. Similarly, a sum of the type

$$\sum_\pi{}^* \sum_\psi \alpha(\pi, \psi \dots)$$

means (unless otherwise specified) that $\pi \in \Pi_\ell$ and $\psi \in \hat{\Gamma}_\ell/\hat{\Gamma}_\ell^\pi$; in other words, this is a sum over all irreducible representations of G_ℓ, parameterized as described previously.

We need various estimates involving sums of dimensions of the representations in Π_ℓ. We will phrase them in terms of upper bounds for the "dimensions" of G_ℓ and of the set $G_\ell^\sharp$ of its conjugacy classes: let s and t be such that the inequalities

$$(2.4) \qquad |G_\ell| \leqslant c_1 \ell^s, \quad |G_\ell^\sharp| \leqslant c_2 \ell^t$$

hold for all primes $\ell \in \Lambda$, c_1 and c_2 being two given constants. Note that of course $s = t = r^2$ is always possible with $c_1 = c_2 = 1$ (and that in fact this does not in general significantly affect the applications).

Lemma 2.4. (1) *We have*

$$\sum_{\pi \in \Pi_\ell} \dim \pi \leqslant (c_1 c_2 \ell^{s+t})^{1/2},$$

and for all $\pi \in \Pi_\ell$ we have

$$\dim \pi \leqslant (c_1 \ell^s)^{1/2}.$$

(2) *If $G_\ell^g = SL(r, \mathbf{F}_\ell)$, the estimates (2.4) hold with*

$$c_1 = 1, \quad s = r^2, \quad c_2 = 6^r, \quad t = r.$$

(3) *If r is even, $G_\ell^g = Sp(r, \mathbf{F}_\ell)$ and $G_\ell \subset SSp(r, \mathbf{F}_\ell)$, the estimates (2.4) hold with*

$$c_1 = 1, \quad s = 1 + \frac{r(r+1)}{2}, \quad c_2 = 6^{r/2}, \quad t = r/2 + 1.$$

Proof. For a representation of a finite group G, the dimension is always $\leqslant |G|^{1/2}$, and the sum of the dimension is bounded by Cauchy's inequality by

$$\sum_\pi \dim \pi \leqslant |G^\sharp|^{1/2} |G|^{1/2},$$

so that (1) is a direct translation of (2.4).

(2) is obvious, noticing that the number of conjugacy classes in G_ℓ is at most

$$|\Gamma_\ell| |G_\ell^{g,\sharp}| \leqslant (\ell - 1)|G_\ell^{g,\sharp}| \leqslant (\ell - 1)(6\ell)^{r-1}$$

(by [LP, Lemma 1.4] for instance; the factor 6^{r-1} takes into account the non-semisimple conjugacy classes).

(3) is similar using the formula for the cardinality of $Sp(r, \mathbf{F}_\ell)$, and [LP, Lemmas 1.3, 1.6] for the conjugacy classes. $\qquad\square$

Our last definition is also of crucial importance for the bilinear form estimate:

Definition. We say that the family $(\tilde{\mathcal{F}}_\ell)$ is *linearly disjoint* if for all ℓ and ℓ' in Λ, with $\ell \neq \ell'$, the product map

$$\pi_1(\overline{U}, \overline{\eta}) \to G_\ell^g \times G_{\ell'}^g$$

is surjective.

This is a fairly natural independence notion for the various monodromy groups. In many cases it will hold for group-theoretical reasons simply because the G_ℓ^g are "large" groups and close to simple; this is related to Goursat's lemma, and we quote here the version in [C, Pr. 5.1] (specialized for 2 factors):

Lemma 2.5. *Let G_1 and G_2 be finite groups such that every normal subgroup of G_i is contained in the center C_i, and such that G_1/C_1 and G_2/C_2 are distinct, simple and non-abelian. Then no proper subgroup $G \subset G_1 \times G_2$ projects surjectively on both G_1 and G_2.*

This is typically applied with $G_1 = G_\ell^g$, $G_2 = G_{\ell'}^g$ and G the image of $\pi_1(\overline{U}, \overline{\eta}) \to G_1 \times G_2$ which *does* project surjectively on both factors.

For instance, this shows:

Corollary 2.6. (1) *Let r be even and let $(\tilde{\mathcal{F}}_\ell)$ be a family of sheaves as above such that $G_\ell^g = Sp(r, \mathbf{F}_\ell)$ for all ℓ in Λ, with $\ell \geqslant 5$ if $r = 2$ and $\ell \geqslant 3$ if $r = 4$. Then the family is linearly disjoint.*

(2) Let $(\tilde{\mathcal{F}}_\ell)$ be a family of sheaves as above such that $G_\ell^g = SL(r, \mathbf{F}_\ell)$ for all ℓ in Λ, with $\ell \geqslant 5$ if $r = 2$. Then the family is linearly disjoint.

This follows because it is very classical that the center of $Sp(r, \mathbf{F}_\ell)$ (resp. $SL(r, \mathbf{F}_\ell)$) is ± 1 (and is the only non-trivial normal subgroup) and $Sp(r, \mathbf{F}_\ell)/\{\pm 1\}$ (resp. $SL(r, \mathbf{F}_\ell)/\{\pm 1\}$) is a simple non-abelian group in the cases described (see e.g. [A, Th. 5.1, Th. 4.9]). On the other hand, notice the lemma can not be applied for orthogonal groups. (For instance, if ℓ, ℓ' are odd, the proper subgroup

$$\{(x, y) \in O(r, \mathbf{F}_\ell) \times O(r, \mathbf{F}_{\ell'}) \mid \det(x) = \det(y)\},$$

where the equality makes sense because the determinants are ± 1, clearly projects surjectively onto both factors).

3. Bilinear form estimates and large sieve for algebraic families

We now state the bilinear form estimate which is our main tool.

Theorem 3.1. *Let U be a variety and $(\tilde{\mathcal{F}}_\ell)$ a family of sheaves as above, with given sets Π_ℓ of irreducible representations which are representatives for geometric equivalence. Assume that the family is linearly disjoint, that it satisfies (2.3) and moreover that U and $(\tilde{\mathcal{F}}_\ell)$ satisfy one of the following conditions:*

(i) U is a smooth affine curve and $(\tilde{\mathcal{F}}_\ell)$ arises from a compatible system of integral ℓ-adic sheaves;

(ii) For all $\ell \in \Lambda$, the order of G_ℓ^g is prime to p.

Then there exist constants $C \geqslant 0$ and $A \geqslant 0$ such that we have

$$(3.1) \qquad \sum_{\ell \leqslant L} \sideset{}{^*}\sum_{\pi \neq 1} \Big| \sum_{u \in U(\mathbf{F}_q)} \alpha(u) \operatorname{Tr}(\pi \circ \rho_\ell)(\operatorname{Fr}_u) \Big|^2 \leqslant (\kappa q^d + C q^{d-1/2} L^A) \sum_{u \in U(\mathbf{F}_q)} |\alpha(u)|^2,$$

for any $L \geqslant 1$ and any complex coefficients $\alpha(u)$.

In case (i), we can take $A = 1 + s + t/2$, and the constant C depends only on $\overline{U}$, the "geometric" compatible system $(\mathcal{F}_\ell)$ on $\overline{U}$ and the constants c_1 and c_2. In case (ii) we can take $A = 1 + 3s + t/2$, and the constant C depends only on $\overline{U}$, c_1 and c_2.

In particular the estimate can be applied uniformly for $U \otimes \mathbf{F}_{q^n}$ for any $n \geqslant 1$.

Note that the left-hand side of (3.1) is in fact independent of the choice of representative sets Π_ℓ.

Remark 3.2. Here are a few remarks, most of which are of a general nature and are standard observations for any type of bilinear form estimate.

(1) The estimate (3.1) is most interesting when L is small enough that $L^A \leqslant q^{1/2}$, so that the sum of the two terms q^d and $q^{d-1/2}L^A$ is still of size q^d, which is roughly the number of terms in the inner sum over $u \in U(\mathbf{F}_q)$ by the Lang-Weil estimate $|U(\mathbf{F}_q)| = q^d + O(q^{d-1/2})$.

(2) The restriction to $\pi \neq 1$ in the summation in (3.1) is essential: the additional contribution of the trivial representations would give the quadratic form

$$\sum_{\ell \leqslant L} \Big| \sum_{u \in U(\mathbf{F}_q)} \alpha(u) \Big|^2$$

which by Cauchy's inequality has norm $|U(\mathbf{F}_q)|L \asymp q^d L$, which exceeds $(\kappa q^d + q^{d-1/2}L^A)$ in the most interesting ranges where L is small as in the previous remark.

(3) In (3.1), the trivial bound has $(\kappa q^d + q^{d-1/2}L^A)$ replaced by

$$|U(\mathbf{F}_q)| \sum_{\ell \leqslant L} \sum_{\pi \neq 1} 1 \asymp q^d L^{1+t}$$

(i.e., the ratio is bounded from above and below; we assume that t is chosen optimally). On the other hand, from general principles (see e.g. [IK, §7]), the best possible result is essentially with

$$|U(\mathbf{F}_q)| + \sum_{\ell \leqslant L} \sum_{\pi \neq 1} 1 \asymp q^d + L^{1+t},$$

and nowadays it is usually estimates of similar strength which are called large sieve inequalities, even when no connection with sieve theory exists.

Thus from the point of view of the study of bilinear forms in modern analytic number theory, the estimate (3.1) is much too weak to deserve the name of large sieve. However, we *are* using it mainly to derive a sieve-type result which corresponds to Linnik's original description of a "large" sieve, so we use the word in this sense.

(4) Using geometric considerations one can get similar results for more general U, for instance by dealing with irreducible components one by one. Or if U is the disjoint union of U_1 and U_2, with U_1 a dense open subscheme which is smooth affine and geometrically connected, and U_2 closed of codimension $\geqslant 1$, the sum on the left of (3.1) is at most twice the sum

$$\Big| \sum_{u \in U_1(\mathbf{F}_q)} \alpha(u) \operatorname{Tr}(\pi \circ \rho_\ell)(\operatorname{Fr}_u) \Big|^2 + \Big| \sum_{u \in U_2(\mathbf{F}_q)} \alpha(u) \operatorname{Tr}(\pi \circ \rho_\ell)(\operatorname{Fr}_u) \Big|^2$$

and (3.1) applies to the first sum while the second has a contribution $\ll m q^{d-1} L^{1+t}$ by Remark (3), where m is the number of irreducible components of $U_2 \otimes \overline{\mathbf{F}}_q$. Another case would be to have a map $U \to V$ with "most" fibers being smooth, affine, connected curves on which the induced sheaves have the same monodromy as on U, and for which the constant C in (3.1) happens to be uniformly bounded for all such fibers.

(5) Formula (5.3) below gives a more explicit bound which may be better in some cases where more is known about the G_ℓ (although it's not clear how much of a difference it would make in applications), for instance the maximal dimension of an irreducible representation.[2]

(6) Finally, we note that a standard heuristic understanding of the strength of the classical large sieve inequality (see e.g. [IK, 7.5] for a proof)

$$\sum_{q \leqslant Q} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{n \leqslant N} a_n \chi(n) \Big|^2 \leqslant (N - 1 + Q^2) \sum_{n \leqslant N} |a_n|^2$$

(where $\sum^*$ indicates a sum over *primitive* Dirichlet characters modulo q), is that in its range of effectiveness (i.e., $Q^2 \leqslant N$) it is as strong as the Generalized Riemann Hypothesis, as it gives for instance

$$\sum_{n \leqslant N} \mu(n) \chi(n) \ll \sqrt{N}$$

on average, where $\mu(n)$ is the Möbius function. And indeed this inequality is used as a substitute for GRH in many applications. In view of this and the fact that we already know the Riemann Hypothesis over finite fields by Deligne's work, one may think that a large sieve inequality would be either trivial to prove or without application (or both) in this context. That it is not the case illustrates that the large sieve inequality is not only about cancellation, but about *uniformity* in estimates. Hence it is not surprising that the "only" difficulty in proving (3.1), from the Riemann Hypothesis, is a question of uniform bounds for the error terms coming after applying Deligne's results.

We will prove Theorem 3.1 in Section 5. For the moment, we derive a large sieve estimate concerning the average distribution of the Frobenius conjugacy classes in G_ℓ.

---

[2] This is indeed known for the groups $Sp(2g, \mathbf{F}_\ell)$ that we will use below, as will be explained in a forthcoming paper.

Let $L \geqslant 2$ and suppose that for $\ell \in \Lambda$, $\ell \leqslant L$, we select some conjugacy-invariant subset $\Omega(\ell)$ of G_ℓ with cardinality $\omega(\ell)$, such that

$$m(x) = \varphi(-1) \in \Gamma_\ell$$

for all x and ℓ (where $m : G_\ell \to \Gamma_\ell$ and φ are defined by the commutative diagram (2.2); recall that $d(\mathrm{Fr}_u) = -1 \in \hat{\mathbf{Z}}$ for $u \in U(\mathbf{F}_q)$).

Let then

$$P(u, L) = \sum_{\substack{\ell \leqslant L \\ \rho_\ell(\mathrm{Fr}_u) \in \Omega(\ell)}} 1$$

for $u \in U(\mathbf{F}_q)$ and

$$P(L) = \sum_{\ell \leqslant L} \omega(\ell)|G_\ell^g|^{-1}.$$

The large sieve statement says that for "most" u, the value of $P(u, L)$ is close to the average value $P(L)$, this being measured by the variance.

Proposition 3.3. *With U and $(\tilde{\mathcal{F}}_\ell)$ satisfying one of the assumptions of Theorem 3.1, we have*

$$(3.2) \qquad \sum_{u \in U(\mathbf{F}_q)} (P(u, L) - P(L))^2 \leqslant (\kappa q^d + C q^{d-1/2} L^A) P(L),$$

where the constants C and A are the same as in Theorem 3.1. In particular, the cardinality of the sifted set

$$S(U, \Omega; L) = \{u \in U(\mathbf{F}_q) \mid \mathrm{Fr}_u \notin \Omega(\ell) \text{ for all } \ell \leqslant L\}$$

satisfies

$$(3.3) \qquad |S(U, \Omega; L)| \leqslant (\kappa q^d + C q^{d-1/2} L^A) P(L)^{-1}.$$

Proof. First (3.3) follows trivially from (3.2) since $P(u, L) = 0$ for $u \in S(U, \Omega; L)$, so that the left-hand side of the latter inequality is at least equal to $P(L)^2 |S(U, \Omega; L)|$.

So we prove (3.2); the argument is in large part a jazzed-up version of the one in [G]. Let χ_ℓ be the characteristic function of $\Omega(\ell)$. Since $\Omega(\ell)$ is invariant by conjugation, we can expand it in Fourier series using the representations of G_ℓ. Using the parameterization as $\pi \otimes \psi$ with $\pi \in \Pi_\ell$ and $\psi \in \hat{\Gamma}_\ell/\hat{\Gamma}_\ell^\pi$, we can write this expansion as

$$(3.4) \qquad \chi_\ell(x) = \sum_{\pi \in \Pi_\ell} \sum_{\psi \in \hat{\Gamma}_\ell/\hat{\Gamma}_\ell^\pi} \hat{\chi}_\ell(\psi, \pi) \psi(x) \operatorname{Tr} \pi(x)$$

with

$$(3.5) \qquad \hat{\chi}_\ell(\psi, \pi) = \frac{1}{|G_\ell|} \sum_{x \in \Omega(\ell)} \overline{\psi(x) \operatorname{Tr} \pi(x)} = \frac{1}{|\Gamma_\ell|} \psi(\varphi(1)) \gamma(\pi),$$

where

$$\gamma(\pi) = \frac{1}{|G_\ell^g|} \sum_{x \in \Omega(\ell)} \overline{\operatorname{Tr} \pi(x)}.$$

Thus we have

$$(3.6) \qquad \gamma(1) = \omega(\ell)|G_\ell^g|^{-1}.$$

Also by orthonormality of the characters of G_ℓ we have

$$\frac{\omega(\ell)}{|G_\ell|} = \frac{1}{|G_\ell|} \sum_{x \in G_\ell} |\chi_\ell(x)|^2 = \sideset{}{^*}\sum_\pi \sum_\psi |\hat{\chi}_\ell(\psi, \pi)|^2 = \sideset{}{^*}\sum_\pi \sum_\psi \frac{1}{|\Gamma_\ell|^2} |\gamma(\pi)|^2 = \frac{1}{|\Gamma_\ell|} \sideset{}{^*}\sum_\pi \frac{|\gamma(\pi)|^2}{|\hat{\Gamma}_\ell^\pi|},$$

hence

$$(3.7) \qquad \sideset{}{^*}\sum_{\pi \neq 1} \frac{|\gamma(\pi)|^2}{|\hat{\Gamma}_\ell^\pi|^2} \leqslant \sideset{}{^*}\sum_{\pi \neq 1} \frac{|\gamma(\pi)|^2}{|\hat{\Gamma}_\ell^\pi|} \leqslant \sideset{}{^*}\sum_\pi \frac{|\gamma(\pi)|^2}{|\hat{\Gamma}_\ell^\pi|} = \frac{\omega(\ell)}{|G_\ell^g|}.$$

By (3.4), (3.5) and the fact that $\psi(\rho_\ell(\mathrm{Fr}_u)) = \psi(\varphi(-1))$, we get that for $u \in U(\mathbf{F}_q)$ and $\ell \leqslant L$ we have

$$
\begin{aligned}
\chi_\ell(\rho_\ell(\mathrm{Fr}_u)) &= \sum_\pi^* \sum_\psi \frac{1}{|\Gamma_\ell|} \psi(\varphi(1)) \gamma(\pi) \psi(\varphi(-1)) \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \\
&= \sum_\pi^* \gamma(\pi) \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \Big(\frac{1}{|\Gamma_\ell|} \sum_\psi \psi(\varphi(1)) \psi(\varphi(-1)) \Big) \\
&= \sum_\pi^* \frac{\gamma(\pi)}{|\hat{\Gamma}_\ell^\pi|} \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u)
\end{aligned}
$$

hence (since $\hat{\Gamma}_\ell^1 = 1$)

$$
(3.8) \qquad P(u, L) = \sum_{\ell \leqslant L} \sum_\pi^* \frac{\gamma(\pi)}{|\hat{\Gamma}_\ell^\pi|} \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) = P(L) + \sum_{\ell \leqslant L} \sum_{\pi \neq 1}^* \frac{\gamma(\pi)}{|\hat{\Gamma}_\ell^\pi|} \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u)
$$

using (3.6).

Denote by $R(u, L)$ the second term on the right-hand side (the sum over $\ell \leqslant L$ and $\pi \neq 1$). By Cauchy's inequality and (3.7) we have

$$
\begin{aligned}
\sum_{u \in U(\mathbf{F}_q)} |R(u, L)|^2 &= \sum_{\ell \leqslant L} \sum_{\pi \neq 1}^* \frac{\gamma(\pi)}{|\hat{\Gamma}_\ell^\pi|} \sum_{u \in U(\mathbf{F}_q)} R(u, L) \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \\
&\leqslant \Big(\sum_{\ell \leqslant L} \sum_{\pi \neq 1}^* \frac{|\gamma(\pi)|^2}{|\hat{\Gamma}_\ell^\pi|^2} \Big)^{1/2} \Big(\sum_{\ell \leqslant L} \sum_{\pi \neq 1}^* \Big| \sum_{u \in U(\mathbf{F}_q)} R(u, L) \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \Big|^2 \Big)^{1/2} \\
&\leqslant P(L)^{1/2} \Big(\sum_{\ell \leqslant L} \sum_{\pi \neq 1}^* \Big| \sum_{u \in U(\mathbf{F}_q)} R(u, L) \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \Big|^2 \Big)^{1/2}.
\end{aligned}
$$

We can apply Theorem 3.1 to the last sum, getting (after squaring)

$$
\Big(\sum_{u \in U(\mathbf{F}_q)} |R(u, L)|^2 \Big)^2 \leqslant P(L)(\kappa q^d + C q^{d-1/2} L^A) \sum_{u \in U(\mathbf{F}_q)} |R(u, L)|^2
$$

so that

$$
\sum_{u \in U(\mathbf{F}_q)} |R(u, L)|^2 \leqslant (\kappa q^d + C q^{d-1/2} L^A) P(L),
$$

which concludes the proof since by (3.8) we have

$$
\sum_{u \in U(\mathbf{F}_q)} (P(u, L) - P(L))^2 = \sum_{u \in U(\mathbf{F}_q)} |R(u, L)|^2.
$$

$\square$

4. Estimates for sums of Betti numbers

In this section we will prove some estimates for Betti numbers of ℓ-adic sheaves needed in the proof of the main estimate in Section 5.

For a separated scheme of finite type U over $\overline{\mathbf{F}}_q$ of dimension $d \geqslant 1$ and any prime $\ell \neq p$ we denote as usual

$$
\begin{aligned}
h_c^i(U, \mathcal{F}) &= \dim H_c^i(U, \mathcal{F}), \quad h^i(U, \mathcal{F}) = \dim H^i(U, \mathcal{F}), \\
\sigma_c(U, \mathcal{F}) &= \sum_i h_c^i(U, \mathcal{F}), \quad \sigma(U, \mathcal{F}) = \sum_i h^i(U, \mathcal{F}), \\
\chi_c(U, \mathcal{F}) &= \sum_i (-1)^i h_c^i(U, \mathcal{F}), \quad \chi(U, \mathcal{F}) = \sum_i (-1)^i h_c^i(U, \mathcal{F}),
\end{aligned}
$$

where $\mathcal{F}$ can be either a $\overline{\mathbf{Q}}_\ell$-sheaf on U or an $\overline{\mathbf{F}}_\ell$-sheaf. We also write

$$\sigma'_c(U, \mathcal{F}) = \sum_{i < 2d} h^i_c(U, \mathcal{F}) \leqslant \sigma_c(U, \mathcal{F})$$

for the sum of all Betti numbers except the topmost one.

We first consider the case of curves. Here the situation will be as follows: $U/\overline{\mathbf{F}}_q$ is a smooth affine connected curve, $\rho : \pi_1(U, \overline{\eta}) \to G$ is a surjective group homomorphism with G finite, and π is a representation of G, with values in some $\overline{\mathbf{Q}}_\ell$-vector space of finite dimension, for some $\ell \neq p$. We can form the composite $\pi \circ \rho$ to obtain a lisse ℓ-adic sheaf on U, which is denoted $\pi(\rho)$. Then we wish to find bounds for the sum of Betti numbers $\sigma'_c(U, \pi(\rho))$ which are polynomial in the size of G (or the degree $\dim \pi$ of π).

We do this for ρ of a special type, which we describe in a slightly more general case than will be needed in the next section: G is a product

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

where G_i is a subgroup of $GL(r, \mathbf{F}_{\lambda_i})$ for $1 \leqslant i \leqslant k$, λ_i a power of a prime $\ell_i \neq p$ (the ℓ_i are not necessarily distinct), and the representation ρ is a tensor product $\rho_1 \otimes \cdots \otimes \rho_k$ where the ρ_i correspond to lisse sheaves $\mathcal{F}_i/\ell_i\mathcal{F}_i$, which are the reductions modulo ℓ_i of sheaves $\mathcal{F}_i$ of $\mathbf{Z}_{\lambda_i}$-modules which are part of a compatible system $(\mathcal{F}_\ell)$. Here $\mathbf{Z}_{\lambda_i}$ is the ring of integers of a finite extension $\mathbf{Q}_{\lambda_i}$ of $\mathbf{Q}_{\ell_i}$ with residue field $\mathbf{F}_{\lambda_i}$.

Our goal is:

Proposition 4.1. *With notation as above, we have the bound*

$$\sigma'_c(U, \pi(\rho)) \leqslant C(U, (\mathcal{F}_i), k)(\dim \pi),$$

for some constant $C(U, (\mathcal{F}_i), k)$ depending only on U, k and the compatible system, but not on π. One can take

$$(4.1) \qquad\qquad C = 1 - \chi_c(U, \mathbf{Q}_\ell) + kw$$

where $w \geqslant 0$ is the sum of the Swan conductors of all $\mathcal{F}_i$ at the points at infinity for U, as described below, which is independent of i.

We start by recalling and setting up the description of the ramification structure of sheaves on U, as described for instance in [K2, Ch. 1]. Let C be the smooth projective model of U and $S = C - U$ the non-empty finite set of "points at infinity". Let M be a $\mathbf{Z}[1/p]$-module on which $\pi_1(U, \overline{\eta})$ acts through a finite discrete quotient. For each point $x \in S$, there is a certain direct sum decomposition of M seen as representation of the inertia group I_x at x of the type

$$M = \bigoplus_{t \geqslant 0} M_x(t)$$

where each $M_x(t)$ is I_x-stable. All but finitely many of the $M_x(t)$ vanish, and those t for which $M_x(t) \neq 0$ are called the breaks of M at x. If M is free over some $\mathbf{Z}[1/p]$-algebra A (e.g., $A = \mathbf{F}_\ell$, $\mathbf{Z}_\ell$ or $\mathbf{Q}_\ell$), the Swan conductor of M at x is then defined by

$$\mathrm{Swan}_x(M) = \sum_{t \geqslant 0} t \operatorname{rank} M_x(t).$$

We let $\mathrm{B}_x(M)$ denote the largest break, i.e., the largest $t \geqslant 0$ such that $M_x(t) \neq 0$. Notice then the trivial inequalities

$$(4.2) \qquad\qquad \mathrm{Swan}_x(M) \leqslant (\operatorname{rank} M) \, \mathrm{B}_x(M),$$

$$(4.3) \qquad\qquad \mathrm{B}_x(M) \leqslant \mathrm{Swan}_x(M).$$

In addition, if $M = M_1 \otimes M_2$ we have [K2, Lemma 1.3]

$$(4.4) \qquad\qquad \mathrm{B}_x(M) \leqslant \max(\mathrm{B}_x(M_1), \mathrm{B}_x(M_2)) \leqslant \mathrm{B}_x(M_1) + \mathrm{B}_x(M_2).$$

If M is a finite dimensional $\mathbf{Q}_\lambda$-vector space, with $\mathbf{Q}_\lambda$ a finite extension of $\mathbf{Q}_\ell$ with ring of integers $\mathbf{Z}_\lambda$, for some $\ell \neq p$, and if $\mathbf{M} \subset M$ is an invariant $\mathbf{Z}_\lambda$-lattice with reduction $\mathbf{M}/\lambda\mathbf{M}$, then we have [K2, Rem. 1.10]

$$(4.5) \qquad \mathrm{Swan}_x(M) = \mathrm{Swan}_x(\mathbf{M}) = \mathrm{Swan}_x(\mathbf{M}/\ell\mathbf{M}).$$

Finally, the main reason the Swan conductor enters in our computation is the fundamental formula of Grothendieck-Ogg-Shafarevitch:

Proposition 4.2. *If $\mathbf{Q}_\lambda$ is a finite extension of $\mathbf{Q}_\ell$ and $\mathcal{F}$ is a lisse $\mathbf{Q}_\lambda$-sheaf on U of rank r, we have*

$$(4.6) \qquad \chi_c(U, \mathcal{F}) = r\chi_c(U, \mathbf{Q}_\ell) - \sum_{x \in S} \mathrm{Swan}_x(\mathcal{F}),$$

where $\mathrm{Swan}_x(\mathcal{F})$ is $\mathrm{Swan}_x(M)$ for the $\mathbf{Q}_\lambda$-vector space which is the representation space of the representation corresponding to $\mathcal{F}$.

See e.g. [K2, 2.3.1, 2.3.3] for a sketch of the proof.

Proof of Proposition 4.1. Since U is affine and smooth we have $h_c^0(U, \pi(\rho)) = 0$ and

$$\sigma_c'(U, \pi(\rho)) = h_c^1(U, \pi(\rho)),$$

while the Euler-Poincaré characteristic is

$$\chi_c(U, \pi(\rho)) = -h_c^1(U, \pi(\rho)) + h_c^2(U, \pi(\rho)).$$

We want to bound $-\chi_c(U, \pi(\rho))$, and for this start from the Euler-Poincaré formula (4.6) for $\pi(\rho)$ (which takes value in some $GL(r, \mathbf{Q}_\lambda)$):

$$\chi_c(U, \pi(\rho)) = (\dim \pi)\chi_c(U, \mathbf{Q}_\ell) - \sum_{x \in S} \mathrm{Swan}_x(\pi(\rho)).$$

By (4.2) we get bounds

$$\mathrm{Swan}_x(\pi(\rho)) \leqslant (\dim \pi)\, \mathrm{B}_x(\pi \circ \rho) \leqslant (\dim \pi)\, \mathrm{B}_x(M),$$

where M is the $\mathbf{Z}[1/p]$-module

$$M = M_1 \otimes \cdots \otimes M_k,$$

with $M_i \simeq \mathbf{F}_{\lambda_i}^r$, with the action $\rho = \rho_1 \otimes \cdots \otimes \rho_k$ of $\pi_1(U, \overline{\eta})$. The last inequality is simply because the action of the inertia group "on" $\pi \circ \rho$ factors through that on M.

Now we have by (4.4) and (4.3)

$$\mathrm{B}_x(M) \leqslant \max \mathrm{B}_x(\rho_i) \leqslant \sum_i \mathrm{B}_x(\rho_i) \leqslant \sum_i \mathrm{Swan}_x(\rho_i).$$

Hence

$$\mathrm{Swan}_x(\pi(\rho)) \leqslant (\dim \pi) \sum_i \mathrm{Swan}_x(\rho_i).$$

Now we can use the fact that each ρ_i is the reduction of the $\mathbf{Z}_{\lambda_i}$-sheaf $\mathcal{F}_i$ and use (4.5) to get

$$\sum_i \mathrm{Swan}_x(\rho_i) = \sum_i \mathrm{Swan}_x(\mathcal{F}_i \otimes \mathbf{Q}_{\lambda_i}),$$

and by (4.6) again we have for all i

$$\sum_{x \in S} \mathrm{Swan}_x(\mathcal{F}_i \otimes \mathbf{Q}_{\lambda_i}) = r\chi_c(U, \mathbf{Q}_{\lambda_i}) - \chi_c(U, \mathcal{F}_i).$$

The crucial point is that $\chi_c(U, \mathcal{F}_i)$ is independent of i because the sheaves $\mathcal{F}_i$ form a compatible system (it is minus the degree of the common L-function of the sheaves $\mathcal{F}_i$), and so of course is

$\chi_c(U, \mathbf{Q}_{\lambda_i})$. So the sum of the Swan conductors of the $\mathcal{F}_i \otimes \mathbf{Q}_{\lambda_i}$ is independent of i. We denote this common value by w, and thus we get

$$-\chi_c(U, \pi(\rho)) \leqslant (\dim \pi)\Big\{ -\chi_c(U, \mathbf{Q}_\ell) + \sum_x \sum_i \mathrm{Swan}_x(\mathcal{F}_i \otimes \mathbf{Q}_{\lambda_i}) \Big\}$$

$$= (\dim \pi)\Big\{ -\chi_c(U, \mathbf{Q}_\ell) + kw \Big\}$$

We add the requisite $h_c^2(U, \pi(\rho))$ which is trivially $\leqslant \dim \pi$ by the co-invariant description

$$H_c^2(U, \mathcal{F}) \simeq \mathcal{F}_{\pi_1(U, \overline{\eta})}(-1)$$

for any lisse $\overline{\mathbf{Q}}_\ell$-sheaf $\mathcal{F}$, and therefore we get

(4.7) $\sigma_c'(U, \pi(\rho)) = h_c^2(U, \pi(\rho)) - \chi_c(U, \pi(\rho)) \leqslant (\dim \pi)(1 + C)$ with $C = -\chi_c(U, \mathbf{Q}_\ell) + kw$.

$\square$

Remark 4.3. The proof shows that the result of Proposition 4.1 is optimal in the sense that, under the assumptions stated, there exists a constant $c > 0$, depending only on U, such that

$$\sigma_c'(U, \pi(\rho)) \geqslant c(\dim \pi),$$

at least if $\chi_c(U, \mathbf{Q}_\ell) < 0$; indeed by positivity we have

$$\sigma_c'(U, \pi(\rho)) = h_c^2(U, \pi(\rho)) - \chi_c(U, \pi(\rho)) \geqslant -\chi_c(U, \pi(\rho)) \geqslant (\dim \pi)(-\chi_c(U, \mathbf{Q}_\ell)),$$

and we can take $c = -\chi_c(U, \mathbf{Q}_\ell)$.

We now come to the result that will be used for the case where Assumption (ii) of Theorem 3.1 holds. We will use the following result of Katz, building on work of Bombieri and Adolphson and Sperber:

Proposition 4.4. *Let* $q = p^k$, $U/\overline{\mathbf{F}}_q$ *a smooth affine connected scheme of dimension* $d \geqslant 1$ *which can be embedded in* $\mathbf{A}^N$ *as a closed subscheme defined by the vanishing of* r *polynomials of degree* $\leqslant \delta$. *Then we have*

$$\sigma_c(U, \mathbf{Q}_\ell) \leqslant A(N, r, \delta)$$

for some constant $A(N, r, \delta)$; *one can take*

(4.8) $$A(N, r, \delta) = 2^r 6(3 + r\delta)^{N+1}.$$

This is Theorem 2 of [K1] together with its corollaries.

Proposition 4.5. *Let* $q = p^k$, $U/\overline{\mathbf{F}}_q$ *a smooth affine connected scheme over* $\overline{\mathbf{F}}_q$ *of dimension* $d \geqslant 1$, $\varphi : V \to U$ *a finite étale connected Galois covering of degree prime to* p. *There exists a constant* $C(U)$ *such that*

(4.9) $$\sigma_c(V, \mathbf{Q}_\ell) \leqslant C(U)(\deg \varphi).$$

More precisely, if $d = 1$ *one can take* $C(U) = \sigma_c(U, \mathbf{Q}_\ell)$. *If* $d \geqslant 2$, *let* N, r, δ *be as in Proposition 4.4 for* U. *Then one can take* $C(U) = C(N, r, \delta)$, *where*

(4.10) $$C(N, r, \delta) = 2 \sum_{j=1}^{N-1} A(j, r, \delta) + A(N, r, \delta) \leqslant 12 N 2^r (3 + r\delta)^{N+1}.$$

Something like this may be already known but we haven't found it in the literature. The proof will proceed by induction on d, following the method used by Katz in [K1, Th. 2]. For the induction step we need the following version of an affine Lefschetz theorem:

Proposition 4.6. *Let* $U/\overline{\mathbf{F}}_q$ *be a smooth connected affine scheme over* $\overline{\mathbf{F}}_q$ *of dimension* $d \geqslant 2$, $\varphi : V \to U$ *a finite étale connected Galois covering with Galois group* G. *Fixing an immersion*

$i : U \to \mathbf{A}^N$ *for some $N \geqslant 1$, there exists an affine hyperplane $H \subset \mathbf{A}^N$ such that $U \cap H$ is connected and smooth, $W = \varphi^{-1}(U \cap H)$ is connected and smooth, and in the diagram*

$$
\begin{array}{ccc}
W & \longrightarrow & V \\
{\scriptstyle \varphi_1}\downarrow & & \downarrow{\scriptstyle \varphi} \\
U \cap H & \longrightarrow & U
\end{array}
$$

the map φ_1 is a finite étale Galois covering with group G and the induced maps in étale cohomology

$$(4.11) \qquad H^i(V, \mathbf{Q}_\ell) \to H^i(W, \mathbf{Q}_\ell)$$

are isomorphisms for $i < d - 1$ and injective for $i = d - 1$.

Proof. For any hyperplane H, it is of course true that $W \to U \cap H$ is a finite étale covering with Galois group G, possibly disconnected. However there exists an open dense set of hyperplanes H for which W is indeed connected by [K3, Cor. 3.4.2] with the data $(k, E, f, \pi) = (\overline{\mathbf{F}}_q, V, 0, i)$.

Further, the existence of an open dense set of hyperplanes H such that the induced maps

$$H^i(V, \mathbf{Q}_\ell) \to H^i(\varphi^{-1}(U \cap H), \mathbf{Q}_\ell) = H^i(W, \mathbf{Q}_\ell)$$

satisfy the required condition is the special case of [K3, Cor. 3.4.1] for the data $(k, E, f, \pi) = (\overline{\mathbf{F}}_q, V, 0, i \circ \varphi)$ (compare the proof of [K3, Cor. 3.4.2]). The existence of a third open dense set of H for which $U \cap H$ is smooth connected is Cor. 3.4.3 of loc. cit.

Intersecting those three open dense subsets of hyperplanes, one finds one where all the required conditions hold. $\qquad \square$

Proof of Proposition 4.5. First because U is smooth affine and φ étale, hence finite, V is also affine and smooth.

We recall now some deep facts about étale cohomology. First, since V is smooth and connected we have $\sigma_c(V, \mathbf{Q}_\ell) = \sigma(V, \mathbf{Q}_\ell)$ by Poincaré duality (see e.g. [D2, VI.3]).

Next, by the affine cohomological dimension theorem we have

$$(4.12) \qquad H^i(U, \mathbf{Q}_\ell) = H^i(V, \mathbf{Q}_\ell) = 0 \text{ for } i > d,$$

see e.g. [D2, IV.6.4].

Finally, because φ is an étale Galois covering of degree prime to p, it is moderately ramified and we have

$$(4.13) \qquad \chi(V, \mathbf{Q}_\ell) = (\deg \varphi)\chi(U, \mathbf{Q}_\ell)$$

which is due to Deligne-Lusztig for χ_c (see [I, 2.6, Cor. 2.8]), and we have $\chi = \chi_c$ for U and V as proved by Laumon [L].

Now we are ready to start the proof by induction. Consider first $d = 1$. We have by (4.12) and Poincaré duality which gives $h^0(U, \mathbf{Q}_\ell) = h^0(V, \mathbf{Q}_\ell) = 1$ that

$$\sigma(U, \mathbf{Q}_\ell) = 2 - \chi(U, \mathbf{Q}_\ell), \quad \sigma(V, \mathbf{Q}_\ell) = 2 - \chi(V, \mathbf{Q}_\ell).$$

By (4.13) we get

$$\sigma(V, \mathbf{Q}_\ell) = 2 - (\deg \varphi)\chi(U, \mathbf{Q}_\ell) \leqslant (\deg \varphi)(2 - \chi(U, \mathbf{Q}_\ell)) = \deg(\varphi)\sigma(U, \mathbf{Q}_\ell)$$

so that we can indeed take $C(U) = \sigma_c(U, \mathbf{Q}_\ell)$ in that case. This is the (first) conclusion required for $d = 1$. The alternative bound is also valid since $\sigma(U, \mathbf{Q}_\ell) \leqslant A(N, r, \delta) \leqslant C(N, r, \delta)$ with N, r, δ as described, by Proposition 4.4 and $C(N, r, \delta)$ defined by (4.10).

Now assume that $\dim U = d$ and (4.9) holds for dimension $d - 1$ with the constant (4.10). Fix an embedding $i : U \to \mathbf{A}^N$ for some N (with the attending r and δ). By Proposition 4.6 there exists a hyperplane $H \subset \mathbf{A}^N$ such that the maps (4.11) are, in particular, all injective for $i \leqslant d - 1$. This implies by (4.12)

$$\sigma(V, \mathbf{Q}_\ell) \leqslant \sigma(W, \mathbf{Q}_\ell) + h^d(V, \mathbf{Q}_\ell)$$

on the one hand, and on the other hand we find

$$h^d(V, \mathbf{Q}_\ell) \leqslant h^d(V, \mathbf{Q}_\ell) + h^{d-1}(W, \mathbf{Q}_\ell) - h^{d-1}(V, \mathbf{Q}_\ell) = (-1)^d \chi(V, \mathbf{Q}_\ell) + (-1)^{d-1}\chi(W, \mathbf{Q}_\ell),$$

so that altogether we have the inequality

$$\sigma(V, \mathbf{Q}_\ell) \leqslant (-1)^d \chi(V, \mathbf{Q}_\ell) + (-1)^{d-1}\chi(W, \mathbf{Q}_\ell) + \sigma(W, \mathbf{Q}_\ell).$$

Now using twice (4.13) for $V \to U$ and $W \to U \cap H$, which are both Galois with group G of order prime to p, we find

$$\sigma(V, \mathbf{Q}_\ell) \leqslant (\deg \varphi)((-1)^d \chi(U, \mathbf{Q}_\ell) + (-1)^{d-1}\chi(U \cap H, \mathbf{Q}_\ell)) + \sigma(W, \mathbf{Q}_\ell).$$

By induction applied to $W \to U \cap H$, since $U \cap H$ is embedded in $H \simeq \mathbf{A}^{N-1}$ using the same number and degree of polynomials as U, we can estimate the last term by

$$\sigma(W, \mathbf{Q}_\ell) \leqslant (\deg \varphi)C(N-1, r, \delta)$$

and get

$$\sigma(V, \mathbf{Q}_\ell) \leqslant (\deg \varphi)\Big\{(-1)^d \chi(U, \mathbf{Q}_\ell) + (-1)^{d-1}\chi(U \cap H, \mathbf{Q}_\ell) + C(N-1, r, \delta)\Big\}.$$

Since

$$|\chi(U, \mathbf{Q}_\ell)| \leqslant \sigma(U, \mathbf{Q}_\ell) \leqslant A(N, r, \delta)$$
$$|\chi(U \cap H, \mathbf{Q}_\ell)| \leqslant \sigma(U \cap H, \mathbf{Q}_\ell) \leqslant A(N-1, r, \delta),$$

by Proposition 4.4, we get

$$\sigma(V, \mathbf{Q}_\ell) \leqslant (\deg \varphi)\Big\{A(N, r, \delta) + A(N-1, r, \delta) + C(N-1, r, \delta)\Big\} = (\deg \varphi)C(N, r, \delta),$$

which is the result for U. The last estimate in (4.10) is a crude consequence of the corresponding one for $A(j, r, \delta)$ given in (4.8). □

Proposition 4.7. *Let $U/\overline{\mathbf{F}}_q$ be a smooth affine connected scheme of dimension $d \geqslant 1$. Let $\rho : \pi_1(U, \overline{\eta}) \to G$ be a surjective homomorphism with G finite of order prime to p, let $\pi : G \to GL(r, \overline{\mathbf{Q}}_\ell)$ be a representation of G and $\pi(\rho) = \pi \circ \rho$ the corresponding lisse sheaf on U. There exists a constant $C(U)$ depending only on U such that*

(4.14) $$\sigma_c(U, \pi(\rho)) \leqslant C(U)|G|(\dim \pi).$$

Proof. (Compare with (4) of Theorem 9.2.6 of Katz and Sarnak [KS1]) Let $\varphi : V \to U$ be the connected étale covering with group G corresponding to the kernel of ρ. It follows that $\varphi^*(\pi(\rho))$ is trivial on V, i.e., seeing $\pi(\rho)$ as a $\mathbf{Q}_\lambda$-lisse sheaf, where $\mathbf{Q}_\lambda$ is a finite extension of $\mathbf{Q}_\ell$ for which π has image in $GL(r, \mathbf{Q}_\lambda)$, we have

$$\varphi^*\pi(\rho) \simeq \mathbf{Q}_\lambda^r.$$

Since $V \to U$ is étale and Galois, the Galois group G acts on the cohomology groups $H_c^i(V, \varphi^*\pi(\rho))$ and we have (by the Hochschild-Serre spectral sequence for $V \to U$ for instance) for all i

$$H_c^i(V, \varphi^*\pi(\rho))^G \simeq H_c^i(U, \pi(\rho))$$

hence

$$\sigma_c(U, \pi(\rho)) \leqslant \sigma_c(V, \varphi^*\pi(\rho)) = \sigma_c(V, \mathbf{Q}_\lambda^r) = r\sigma_c(V, \mathbf{Q}_\lambda) = r\sigma_c(V, \mathbf{Q}_\ell)$$

by the formula $H_c^i(V, \mathbf{Q}_\lambda) = H_c^i(V, \mathbf{Q}_\ell) \otimes \mathbf{Q}_\lambda$.

Since the group G is assumed to have order prime to p, we have by Proposition 4.5

$$\sigma_c(V, \mathbf{Q}_\ell) \leqslant C(U)|G|$$

for some constant $C(U)$ independent of π, and the proposition follows by combining these two inequalities. □

Remark 4.8. Contrary to our first optimistic version, the condition that $(\deg \varphi, p) = 1$ in Proposition 4.5 is certainly necessary, as the following example (communicated by Katz) shows: take U to be the affine line $\mathbf{A}^1$ with coordinate x, and take $V = V_d$ to be the curve $y^p - y = x^d$, for d prime to p. Then we have $\sigma_c(V_d) = 1 + (p-1)(d-1)$. So as d grows, although the degree of the covering $V_d \to U$ stays p, we see that $\sigma_c(V_d)$ is unbounded.

Since the covering is also Galois with group $\mathbf{Z}/p\mathbf{Z}$ in this case, this also shows that Proposition 4.7 does not extend to arbitrary groups: the covering $V \to U$ corresponds to a surjective map $\rho : \pi_1(U) \to \mathbf{Z}/p\mathbf{Z}$, the representations $\pi = \psi$ are the additive characters of $\mathbf{Z}/p\mathbf{Z}$, and we have

$$\sigma_c(V_d, \mathbf{Q}_\ell) = \sum_\psi \sigma_c(U, \psi(\rho))$$

(which amounts to the standard counting of points on $V_d(\mathbf{F}_{q^n})$ by means of additive character sums, or the construction of the sheaves corresponding to those sums) and therefore a bound like (4.14) would give $\sigma_c(V_d) \leqslant p^2 C(U)$, which is also incorrect.

The condition that the covering be Galois is necessary for the proof of Proposition 4.5 because otherwise (4.13) may fail, even for a covering of degree prime to p, as in the following example (again communicated by Katz): take the finite étale covering $\mathbf{G}_m \to \mathbf{A}^1$ (over $\mathbf{F}_q$) given by $x \mapsto x^p + 1/x$. We have $\chi(\mathbf{A}^1) = 1$, whereas $\chi(\mathbf{G}_m) = 0$, and the covering is of degree $p+1$.

Still one may hope that an analogue of Proposition 4.1 holds for arbitrary U, which would give a corresponding general version of Theorem 3.1 and its applications.

5. PROOF OF THE BILINEAR FORM ESTIMATE

We come back to the notation of Section 3 before and in the statement of Theorem 3.1, which we will now prove.

The analytic principle for the proof of Theorem 3.1 is quite simple and very well established in analytic number theory. We proceed by duality, as first conceived by Vinogradov: for given $L \geqslant 1$ and $\Delta \geqslant 0$, it is equivalent to prove that

$$\sum_{\ell \leqslant L} \sideset{}{^*}\sum_{\pi \neq 1} \Big| \sum_{u \in U(\mathbf{F}_q)} \alpha(u) \operatorname{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \Big|^2 \leqslant \Delta \sum_{u \in U(\mathbf{F}_q)} |\alpha(u)|^2,$$

for arbitrary $\alpha(u) \in \mathbf{C}$, or to prove that

$$(5.1) \qquad \sum_{u \in U(\mathbf{F}_q)} \Big| \sum_{\ell \leqslant L} \sideset{}{^*}\sum_{\pi \neq 1} \beta(\ell, \pi) \operatorname{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \Big|^2 \leqslant \Delta \sum_{\ell \leqslant L} \sideset{}{^*}\sum_{\pi \neq 1} |\beta(\ell, \pi)|^2.$$

for arbitrary $\beta(\ell, \pi) \in \mathbf{C}$. Recall that π runs over a set Π_ℓ of irreducible representations of G_ℓ up to twist by characters of Γ_ℓ.

The dual form is more manageable here. Denote by $\mathfrak{B}(\beta)$ the left-hand side of (5.1). Expanding the square we get

$$(5.2) \qquad \mathfrak{B}(\beta) = \sum_{\ell \leqslant L} \sideset{}{^*}\sum_{\pi \neq 1} \sum_{\ell' \leqslant L} \sideset{}{^*}\sum_{\pi' \neq 1} \beta(\ell, \pi) \overline{\beta(\ell', \pi')} \mathfrak{S}(\ell, \pi; \ell', \pi'),$$

with

$$\mathfrak{S}(\ell, \pi; \ell', \pi') = \sum_{u \in U(\mathbf{F}_q)} \operatorname{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \overline{\operatorname{Tr}(\pi' \circ \rho_{\ell'})(\mathrm{Fr}_u)}.$$

The crucial point is the following estimation for the individual $\mathfrak{S}(\ell, \pi; \ell', \pi')$.

Proposition 5.1. *With notation as above, and in particular under the assumption that the sheaves are linearly disjoint.*

15

(i) *If the monodromy groups G_ℓ^g are of prime-to-p order, we have*

$$|\mathfrak{S}(\ell,\pi;\ell,\pi) - |\hat{\Gamma}_\ell^\pi|q^d| \leqslant q^{d-1/2}|G_\ell|(\dim \pi)^2 C(\overline{U}),$$

$$|\mathfrak{S}(\ell,\pi;\ell,\pi')| \leqslant q^{d-1/2}|G_\ell|(\dim \pi)(\dim \pi')C(\overline{U}), \ \text{if } \pi \neq \pi'$$

$$|\mathfrak{S}(\ell,\pi;\ell',\pi')| \leqslant q^{d-1/2}|G_\ell||G_{\ell'}|(\dim \pi)(\dim \pi')C(\overline{U}), \ \text{if } \ell \neq \ell',$$

where $C(\overline{U})$ is given by Proposition 4.5.

(ii) *If U is a curve, and the sheaves arise from a compatible system of $\mathbf{Z}_\lambda$-sheaves $\mathcal{F}_\ell$, we have*

$$|\mathfrak{S}(\ell,\pi;\ell,\pi) - |\hat{\Gamma}_\ell^\pi|q^d| \leqslant q^{d-1/2}(\dim \pi)^2 D(\overline{U},(\mathcal{F}_\ell)),$$

$$|\mathfrak{S}(\ell,\pi;\ell',\pi')| \leqslant q^{d-1/2}(\dim \pi)(\dim \pi')D(\overline{U},(\mathcal{F}_\ell)), \ \text{if } \ell \neq \ell' \text{ or } \pi \neq \pi',$$

where $D(\overline{U},(\mathcal{F}_\ell))$ is the constant $C(\overline{U},(\mathcal{F}_\ell),2)$ of Proposition 4.1.

Taking this for granted, we finish quickly the proof of Theorem 3.1. By (5.2) we have trivially (5.1) with

$$\Delta = \max_{\ell,\pi} \sum_{\ell'} \sum_{\pi' \neq 1}^* |\mathfrak{S}(\ell,\pi;\ell',\pi')|,$$

and by Proposition 5.1, we thus get (5.1) with

$$(5.3) \quad \Delta = \max_{\ell,\pi}\Big\{|\hat{\Gamma}_\ell^\pi|q^d + q^{d-1/2}C(\overline{U})|G_\ell^g|(\dim \pi)\Big\{\sum_{\pi'}^* (\dim \pi') + \sum_{\ell' \neq \ell}|G_{\ell'}^g|\sum_{\pi' \neq 1}^* (\dim \pi')\Big\}\Big\}$$

in the case of monodromy of order prime to p, and

$$(5.4) \quad \Delta = \max_{\ell,\pi}\Big\{|\hat{\Gamma}_\ell^\pi|q^d + q^{d-1/2}D(\overline{U},(\mathcal{F}_\ell))(\dim \pi)\Big\{\sum_{\pi'}^* (\dim \pi') + \sum_{\ell' \neq \ell}\sum_{\pi' \neq 1}^* (\dim \pi')\Big\}\Big\}$$

in the case of a curve with a compatible system. We estimate all those terms in terms of the parameters s and t of (2.4) using Lemma 2.4, (1). In the first case we obtain by appealing also to (2.3) and to (2.4) that

$$(5.5) \qquad \Delta \leqslant \kappa q^d + 2q^{d-1/2}C(\overline{U})(c_1^6 c_2)^{1/2}L^{1+3s+t/2}.$$

In the second case we obtain similarly

$$(5.6) \qquad \Delta \leqslant \kappa q^d + 2q^{d-1/2}D(\overline{U},(\mathcal{F}_\ell))c_1 c_2^{1/2}L^{1+s+t/2}.$$

Thus Theorem 3.1 follows by duality.

Proof of Proposition 5.1. The proof is now an easy application of the Grothendieck-Lefschetz trace formula and Deligne's main theorem of [D1] (compare with [C, p. 162,163]). The only subtlety is that the dependency of the error terms on ℓ, ℓ', π, π', must remain controlled, and for this we need the results of Section 4.

If $\ell = \ell'$ we let $G = G_\ell$, $G^g = G_\ell^g$. The representation ρ_ℓ gives a surjective map $\pi_1(U,\overline{\eta}) \to G$. Let $\tau = \pi \otimes \tilde{\pi}'$. This is a (not necessarily irreducible) representation of G, and we will consider the sheaf $\mathcal{F} = \tau \circ \rho_\ell$, which is of the type considered in Proposition 4.7 and Proposition 4.1 (after seeing the representation τ as taking value in $GL(r,\overline{\mathbf{Q}}_\ell)$, as we can since it is a representation in characteristic 0).

If $\ell \neq \ell'$, we let $G = G_\ell \times G_{\ell'}$, $G^g = G_\ell^g \times G_{\ell'}^g$. By the assumption that the family of sheaves is linearly disjoint, the product map (ρ_ℓ,ρ_ℓ') is still a surjective map

$$\pi_1(\overline{U},\overline{\eta}) \xrightarrow{(\rho_\ell,\rho_{\ell'})} G^g.$$

Let $\tau(g,g') = \pi(g) \otimes \tilde{\pi}'(g')$ (the "external" product), so τ is an irreducible representation of G. We will consider the sheaf $\mathcal{F} = \tau \circ (\rho_\ell,\rho_\ell')$, again of the type considered in Proposition 4.7 and Proposition 4.1.

In both cases, because G is a finite group, the eigenvalues of the image of τ are roots of unity so $\mathcal{F}$ is punctually pure of weight 0.

Also in either case, the main point is that the local trace at $u \in U(\mathbf{F}_q)$ of $\mathcal{F}$ is given by construction by

$$\mathrm{Tr}(\mathrm{Fr}_u \mid \mathcal{F}) = \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u)\overline{\mathrm{Tr}(\pi' \circ \rho_{\ell'})(\mathrm{Fr}_u)}$$

and therefore the fundamental Grothendieck-Lefschetz Trace Formula (see [Gr], [D2], [M, VI.13]) states that

$$\mathfrak{S}(\ell, \pi; \ell', \pi') = \sum_{u \in U(\mathbf{F}_q)} \mathrm{Tr}(\mathrm{Fr}_u \mid \mathcal{F}) = \sum_{0 \leqslant i \leqslant 2d} (-1)^i \, \mathrm{Tr}(\mathrm{Fr} \mid H_c^i(\overline{U}, \mathcal{F})).$$

By Deligne's Weil II Theorem [D1, p. 138], the eigenvalues of the geometric Frobenius automorphism Fr acting on $H_c^i(\overline{U}, \mathcal{F})$ are algebraic integers all conjugates of which are of absolute value $\leqslant q^{i/2}$.

It is easy to compute $H_c^{2d}(\overline{U}, \mathcal{F})$ (the action on which contributes potentially terms of maximal size q^d), using the formula

$$H_c^{2d}(\overline{U}, \mathcal{F}) = V_{\pi_1(\overline{U}, \overline{\eta})}(-d) = W_{G^g}(-d)$$

where $V = \mathcal{F}_{\overline{\eta}}$ is the space on which the representation which "is" $\mathcal{F}$ acts and W is the space of the representation τ of G^g. The second equality above holds because the disjointness condition shows that the map $\pi_1(\overline{U}, \overline{\eta}) \to G^g$ through which the action factors is always surjective (as already observed previously).

The crucial point is that this coinvariant space is zero unless $\ell = \ell'$ and $\pi = \pi'$. Indeed, if $\ell \neq \ell'$, decomposing π and π' restricted to G_ℓ^g and $G_{\ell'}^g$ as sums of irreducible representations, the dimension of the coinvariant space (which is the same as that of the invariants under G^g because we are working with finite groups) is the sum of the dimension of invariants for the pairwise tensor products of the components; but those are non-trivial irreducible representations of $G^g = G_\ell^g \times G_{\ell'}^g$ so each term of the sum is zero.

If $\ell = \ell'$, the last statement in Lemma 2.1 exactly says that the space of invariants is of dimension $|\hat{\Gamma}_\ell^\pi|$ for $\pi = \pi'$ and 0 otherwise; this is where it is necessary to restrict to representations unrelated by twists.

Thus we derive the bound

$$\left| \sum_{u \in U(\mathbf{F}_q)} \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u)\overline{\mathrm{Tr}(\pi' \circ \rho_{\ell'})(\mathrm{Fr}_u)} \right| \leqslant q^{d-1/2}\sigma_c'(\overline{U}, \mathcal{F})$$

if $(\ell, \pi) \neq (\ell', \pi')$ and

$$\left| \sum_{u \in U(\mathbf{F}_q)} \mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u)\overline{\mathrm{Tr}(\pi' \circ \rho_{\ell'})(\mathrm{Fr}_u)} - |\hat{\Gamma}_\ell^\pi| q^d \right| \leqslant q^{d-1/2}\sigma_c'(\overline{U}, \mathcal{F})$$

otherwise.

Inserting the bounds for $\sigma_c'(\overline{U}, \mathcal{F})$ from Proposition 4.1 or Proposition 4.7 respectively, and looking at the various cases, the proposition follows. $\qquad\square$

Remark 5.2. If the conditions of Theorem 3.1 are not satisfied, we see that we still get an inequality

$$\sum_\ell \sum_\pi^* \left| \sum_{u \in U(\mathbf{F}_q)} \alpha(u)\,\mathrm{Tr}(\pi \circ \rho_\ell)(\mathrm{Fr}_u) \right|^2 \leqslant (\kappa q^d + q^{d-1/2}D) \sum_{u \in U(\mathbf{F}_q)} |\alpha(u)|^2$$

for fixed L, with

$$D = \max_{\ell, \pi} \sum_{\ell'} \sum_{\pi' \neq 1}^* \sigma_c'(\overline{U}, \mathcal{F}_{\tau_{\pi, \pi'}}).$$

The point is that D is independent of q so this is still non-trivial when applied for $U \times \mathbf{F}_{q^n}$ with $n \to +\infty$. In a large sieve context as in Proposition 3.3, it leads to

$$\limsup_{n \to +\infty} \frac{|S(U \times \mathbf{F}_{q^n}, \Omega; L)|}{q^{nd}} \leqslant \frac{1}{P(L)}$$

for fixed L, and using the same sieve as we will in the proof of Theorem 6.1 for all $L \geqslant 2$ and taking $L \to +\infty$, this recovers Chavdarov's irreducibility theorem [C, Th. 2.3] for an arbitrary family C/U of genus g curves with geometric monodromy modulo ℓ equal to $Sp(2g, \mathbf{F}_\ell)$ for almost all ℓ:

$$\lim_{n \to +\infty} \frac{|\{u \in U(\mathbf{F}_{q^n}) \mid \det(1 - T\,\mathrm{Fr}_u \mid H^1(\overline{C}_u, \mathbf{Z}_\ell)) \text{ has small Galois group}\}|}{|U(\mathbf{F}_{q^n})|} = 0.$$

6. Zeta functions of families of curves

We now come to the application of the large sieve to a strong form of Chavdarov's Theorem on the generic behavior of the numerators of zeta functions of curves in families. If the genus is fixed, most of the work is already done in the previous sections or in Chavdarov's paper, but we will look for arguments uniform with respect to g so that, in some cases at least, we obtain results valid even for g large (though not for q fixed, $g \to +\infty$).

First, we recall the definition of the zeta function of a curve over a finite field, in concrete terms (so the statements at least can be understood without knowledge of étale cohomology), by recalling the diophantine meaning of the polynomials involved.

Let $C/\mathbf{F}_q$ be a smooth projective curve of genus g over a finite field (all curves here and below are assumed to be geometrically connected). Its zeta function $Z(C)$ is the formal power series given by the diophantine definition

$$Z(C) = \exp\Big(\sum_{n \geqslant 1} \frac{|C(\mathbf{F}_{q^n})|}{n} T^n\Big),$$

where $|C(\mathbf{F}_{q^n})|$ is the number of "solutions" to the equations which define C with coordinates in the extension field $\mathbf{F}_{q^n}$. A fundamental result due to F.K. Schmidt in this case is that there exists a polynomial $P_C \in \mathbf{Z}[T]$ of degree $2g$ with $P_C(0) = 1$ such that

$$Z(C) = \frac{P_C(T)}{(1 - T)(1 - qT)}.$$

The cohomological definition is that the polynomial $P_C(T)$ can be described as the (reversed) characteristic polynomial of the geometric Frobenius automorphism acting on a suitable étale cohomology group, specifically

$$(6.1) \qquad\qquad P_C(T) = \det(1 - T\,\mathrm{Fr} \mid H^1(\overline{C}, \mathbf{Z}_\ell)).$$

The question investigated by Chavdarov concerns the splitting field of this integer polynomial as C varies in an algebraic family, e.g. in a hyperelliptic family

$$C_u\ :\ y^2 = f(x)(x - u)$$

where f is a fixed polynomial in $\mathbf{F}_q[X]$ of degree $2g$ with distinct roots in $\overline{\mathbf{F}}_q$, and u is the parameter that can take any value in $\overline{\mathbf{F}}_q$ which is not a zero of f (these conditions ensure that the curve C_u suitably "compactified" is a smooth projective curve of the given genus $g \geqslant 1$).

There is an a-priori condition on the splitting field of the polynomial P_C because it satisfies the "functional equation"

$$(qT)^{2g} P_C\Big(\frac{1}{qT}\Big) = P_C(T),$$

(or equivalently, if $\alpha \in \mathbf{C}$ is a root of P_C, then $q\alpha^{-1}$ is also a root). This means that the "splitting algebra" $\mathbf{Q}[T]/(f)$ has Galois group G which can be seen as a subgroup of the group W_{2g} of signed permutations of $\{1, \dots, 2g\}$. In other words, W_{2g} is the group of permutations of g pairs of elements preserving the pairs. In particular, if the polynomial is irreducible, its

splitting field has maximal Galois group $G \simeq W_{2g}$ if and only if the splitting field is of maximal degree $|W_{2g}| = 2^g g!$.

In terms of étale cohomology the functional equation above is a consequence of the Poincaré duality (in this case, it also amounts to the Weil pairing for the jacobian variety) which states that there is a natural non-degenerate alternating pairing ("cup-product")

$$(6.2) \qquad H^1(\overline{C}, \mathbf{Z}_\ell) \otimes H^1(\overline{C}, \mathbf{Z}_\ell) \to \mathbf{Z}_\ell(-1).$$

Note that this implies that the "global" geometric Frobenius Fr of $\mathbf{F}_q$ acts on $H^1(\overline{C}, \mathbf{Z}_\ell)$ as a symplectic similitude for this pairing, with multiplicator q.

Here is now our first general result about the behavior of the splitting fields in a suitable family, which significantly strengthens the results of Chavdarov.

Theorem 6.1. *Fix an integer $g \geqslant 1$. Let $q = p^k$ and let $U/\mathbf{F}_q$ be a geometrically irreducible smooth affine scheme of dimension $d \geqslant 1$ such that one of the following two conditions is satisfied:*

(i) *U is a curve, i.e, $d = 1$,*

(ii) *we have $p > 2g + 1$.*

Let $\pi : C \to U$ be a proper smooth family of projective curves of genus g over U. Assume that for all $\ell > L_0$ the geometric monodromy group of the integral sheaves $R^1\pi_!\mathbf{Z}_\ell$ is the full symplectic group $Sp(2g)$. Then the number $N(U/\mathbf{F}_q)$ of $u \in U(\mathbf{F}_q)$ such that the numerator

$$(6.3) \qquad P_u = \det(1 - T\,\mathrm{Fr} \mid H^1(\overline{C}_u, \mathbf{Q}_\ell)) \in \mathbf{Z}[T]$$

of the zeta function of the curve $C_u = \pi^{-1}(u)$ is reducible or has splitting field with degree strictly less than $2^g g!$ satisfies

$$N(U/\mathbf{F}_q) \ll q^{d-\gamma}(\log q)$$

for $\gamma = \frac{1}{4g^2+3g+5}$ in case (i) and $\gamma = \frac{1}{12g^2+7g+9}$ in case (ii), where the implied constant depends only on L_0, g and $\overline{U}/\mathbf{F}_q$.

Here is another almost equivalent way of phrasing this: consider the zeta function

$$\tilde{Z}(s) = \exp\Big(\sum_{n \geqslant 1} \frac{N(U/\mathbf{F}_{q^n})}{n} q^{-ns}\Big).$$

It follows from Theorem 6.1 that it extends to a holomorphic function on the half-plane $\mathrm{Re}(s) > d - \gamma$.

The second result is a uniform version (in terms of g) for the families of hyperelliptic curves already introduced.

Theorem 6.2. *Let $g \geqslant 1$, $p \neq 2$, $q = p^k$ with $k \geqslant 1$. Let $f \in \mathbf{F}_q[X]$ be a monic polynomial of degree $2g$ with distinct roots in $\overline{\mathbf{F}}_q$, $U \subset \mathbf{A}^1$ be the complement of the set of zeros of f and denote by $\pi : C \to U$ the family of hyperelliptic curves of genus g given by*

$$C_u : y^2 = f(x)(x - u)$$

completed by the section at ∞, with projection $\pi(x, y, u) = u$.

Then the number $N(f, q)$ of $u \in U(\mathbf{F}_q)$ such that the polynomial

$$P_u = \det(1 - T\,\mathrm{Fr} \mid H^1(\overline{C}_u, \mathbf{Q}_\ell)) \in \mathbf{Z}[T]$$

is either reducible or has splitting field with degree strictly smaller than $2^g g!$ satisfies

$$N(f, q) \ll q^{1-\gamma}(\log q)$$

for $\gamma = \frac{1}{4g^2+3g+5}$, where the implied constant is absolute.

Note that the uniform bound in this last result is only non-trivial if g^2 is somewhat smaller than $\log q$, precisely if $4g^2 = (\log q)e^{-f(q)}$ with $\log q = o(f(q))$. Still, it is an uncommon feature to be able to say *anything* for this kind of problems in a situation where g and q increase together, instead of having first $q \to +\infty$ (compare with the discussion in [KS1, Introduction]).

Since Theorem 6.2 is more delicate, we will start by proving it in the Section 8 after some common preliminaries; we will then quickly deal with the somewhat simpler case of Theorem 6.1. Here we just make a few additional remarks which are of independent interest.

First of all, since the estimate of Theorem 6.2 is (in particular) uniform in q, it can also be used in "horizontal" direction, i.e., with $q = p$ varying. For instance, we deduce the following quite easily:

Proposition 6.3. *Let $g \geqslant 1$ be an integer, $f \in \mathbf{Q}[X]$ be a polynomial of degree $2g$ with distinct complex roots. For $n \in \mathbf{Z}$ not a root of f, let $C_n/\mathbf{Q}$ be the hyperelliptic curve of genus g with equation*

$$C_n \; : \; y^2 = f(x)(x - n)$$

and let J_n be its Jacobian. Then for $N \geqslant 3$, the number $S(N)$ of integers n with $|n| \leqslant N$ such that $J_n/\mathbf{Q}$ is not simple up to isogeny satisfies

$$S(N) \ll N^{1/2-\delta}(\log N)^2$$

where $\delta = \frac{1}{2}\frac{1}{4g^2+3g+5}$. The implied constant depends on g and the splitting field of f.

This should be compared with the individual global results of [C, §6]; our result is on average, but note that we do not require any information on the image of the Galois representations associated to J_n, and in particular we get results valid for *all* genus, independently of the endomorphism ring of J_n or any other global property.

Proof. Denote first by Q_f the set of primes p totally split in the splitting field of f. Notice that for any $X \geqslant 2$ we have the sieving estimate

$$S(N) \leqslant |\{n \in \mathbf{Z} \mid |n| \leqslant N \text{ and } n \, (\mathrm{mod} \, p) \notin \Omega(p) \text{ for } p \in Q_f, \; p \leqslant X\}|$$

where

$$\Omega(p) = \{t \in \mathbf{Z}/p\mathbf{Z} \mid f(t) \not\equiv 0 \, (\mathrm{mod} \, p) \text{ and } \det(1 - T \, \mathrm{Fr}_t \mid H^1(\overline{C}_t, \mathbf{Q}_\ell)) \text{ is irreducible.}\}$$

By Theorem 6.2 there exists a constant $C \geqslant 0$ such that for all p we have

(6.4) $$|\Omega(p)| \geqslant p - Cp^{1-\gamma}(\log p)$$

with $\gamma = \frac{1}{4g^2+3g+5}$. By the usual (strong) form of the large sieve (see e.g. [B, Th. 6], [IK, Th. 7.14]), we have

(6.5) $$S(N) \ll (N + X^2)J^{-1}$$

where

$$J = \sum_{q \leqslant X}^{\flat} \prod_{\substack{p|q \\ p \in Q_f}} \frac{|\Omega(p)|}{p - |\Omega(p)|},$$

the $\flat$ sign indicating a sum restricted to squarefree numbers.

Just taking primes into account we get by (6.4)

$$J \gg \sum_{\substack{p \leqslant X \\ p \in Q_f}} p^\gamma (\log p)^{-1} \gg X^{1+\gamma}(\log X)^{-2},$$

by the Chebotarev density theorem, and taking $X = N^{1/2}$ the proposition follows from (6.5). $\quad\square$

Remark 6.4. C. Elsholtz has remarked that if one uses Gallagher "larger sieve" instead of the classical large sieve, the bound on $S(N)$ can be spectacularly improved to a power of $\log N$ for $N \geqslant 2$.

Another corollary of the large sieve estimates is to families of abelian varieties.

Corollary 6.5. *Let $q = p^k$ and $g \geqslant 1$ such that $p > 2g + 1$. Then the number $N(g, q)$ of isomorphism classes of principally polarized abelian varieties $A/\mathbf{F}_q$ such that the polynomial $\det(1 - \operatorname{Fr} T \mid H^1(\overline{A}, \mathbf{Q}_\ell))$ is either reducible or has splitting field with Galois group strictly smaller than W_{2g} satisfies*

$$N(g, q) \ll q^{g(g+1)/2 - \gamma}(\log q)$$

where $\gamma = \frac{1}{12g^2 + 7g + 9}$ and the implied constant depends only on p and g.

We will prove this at the same time as Theorem 6.1.

Finally, here is an (amusing, but not *too* far-fetched) illustration of what Theorem 6.1 gives which can not be derived from [C].

Proposition 6.6. *Let $g \geqslant 1$, $q = p^k$, $U/\mathbf{F}_q$ a non-empty open subscheme of $\mathbf{G}_m/\mathbf{F}_q$, $\pi : C \to U$ a family of smooth projective curves of genus g such that the geometric monodromy group of $R^1\pi_! \mathbf{Z}_\ell$ is equal to $Sp(2g)$ for almost all ℓ. Then for all n large enough, there exists a primitive root $t \in U(\mathbf{F}_{q^n}) \subset \mathbf{F}_{q^n}^\times$ such that $\det(1 - T\operatorname{Fr}_t \mid H^1(\overline{C}_t, \mathbf{Z}_\ell))$ has maximal Galois group.*

Proof. Indeed, the set of $t \in \mathbf{G}_m(\mathbf{F}_{q^n})$ which are primitive roots has cardinality $\varphi(q^n - 1)$ and

$$\varphi(q^n - 1) \gg \frac{q^n - 1}{\log\log(q^n - 1)} \gg \frac{q^n}{\log n + \log\log q}$$

for $n \geqslant 1$ with an absolute implied constant (see e.g. [HW, Th. 328] for this standard estimate), and this lower bound is larger than the upper bound given by Theorem 6.1 for those t for which the numerator of the zeta function of C_t has small Galois group, if n is large enough. (So in fact, most primitive roots t will have the desired property). $\qquad\square$

Remark 6.7. For any k coprime with q we can find n such that $q^n - 1 \equiv 0 \,(\mathrm{mod}\, k)$ and then

$$\frac{\varphi(q^n - 1)}{q^n - 1} \leqslant \frac{\varphi(k)}{k}.$$

Choosing suitable values of k, we see that the density of primitive roots in $\mathbf{F}_{q^n}^\times$ is not bounded from below by any positive constant. This means that Proposition 6.6 can not be proved without a quantitative form of Chavdarov's theorem.

7. Preliminaries for the proof of Chavdarov's theorem

We start with some preliminaries related to the group W_{2g} and to setting up a sieve for characteristic polynomials of symplectic similitudes.

From the description of W_{2g} we see that there is an exact sequence

$$1 \to \{\pm 1\}^g \to W_{2g} \xrightarrow{p} \mathfrak{S}_g \to 1$$

where the second map just looks at the permutation of the pairs, and the kernel corresponds to just switching the elements of the pairs without moving them. We also denote by i the natural inclusion $i : W_{2g} \to \mathfrak{S}_{2g}$.

Our first lemma describes various ways of ensuring that a subgroup of W_{2g} is equal to W_{2g}.

Lemma 7.1. *Let $g \geqslant 1$ and $W \subset W_{2g}$ be a subgroup of W_{2g}. Assume that one of the following conditions is true, where $i : W_{2g} \to \mathfrak{S}_{2g}$ is the embedding above:*

  (i) *For any conjugacy class $c \subset W_{2g}$, we have $c \cap W \neq \emptyset$.*

  (ii) *The subgroup $i(W)$ contains a 2-cycle, a 4-cycle, a $(2g - 2)$-cycle and a $2g$-cycle.*

  (iii) *The subgroup $i(W)$ contains a transposition and acts transitively on $\{1, \dots, 2g\}$; moreover, the projection $p(W)$ contains a transposition and an m-cycle with $m > g/2$ prime.*

  Then in all cases we have $W = W_{2g}$.

Proof. Case (i) is a standard result in finite group theory (see e.g. [C, Lemma 5.8]), which is in no way specific to W_{2g}.

Case (ii) is Lemma 2 of [DDS].

For case (iii), observe first that the first condition already implies that $W = W_{2g}$ if $g = 1$. Otherwise we see that $p(W)$ acts transitively on $\{1, \ldots, g\}$ and so with the second and third conditions, we get $p(W) = \mathfrak{S}_g$ by the result of Bauer given in [G, Lemma, p. 98]. Since $i(W)$ contains a transposition, we deduce that $W = W_{2g}$ by Lemma 5.5 of [C]. $\qquad\square$

For Theorem 6.1, we can use Case (i) or Case (ii) of the lemma, but Theorem 6.2 requires the finer Case (iii), the point being that the conditions involve "large" subsets of W_{2g}. It seems to be an intriguing problem in combinatorial group theory to determine how optimal this statement is. In terms of $\mathfrak{S}_g$, this means the following optimization problem: let $\Omega_1, \ldots, \Omega_k$ be conjugacy-invariant subsets of $\mathfrak{S}_g$, and let

$$\delta(\Omega_i) = \sum_{i=1}^{k} \frac{g!}{|\Omega_i|}.$$

How small can $\delta(\Omega_i)$ be if (Ω_i) are chosen so that no proper subgroup of $\mathfrak{S}_g$ can intersect each Ω_i? Bauer's lemma gives three subsets with $\delta(\Omega_1, \Omega_2, \Omega_3) \ll g$ (see (8.7) below).

To set up our sieve, it will be convenient to say that a polynomial $f \in A[T]$ (A any commutative ring) of degree $2g$ such that $f(0) = 1$ and

$$(qT)^{2g} f\left(\frac{1}{qT}\right) = f(T),$$

is *q-symplectic of degree* $2g$ (q will often be fixed and clear from the context, as will g).[3] Hence the numerator of the zeta function of a curve $C/\mathbf{F}_q$ is q-symplectic.

We now prove a general result comparing a sieve related to characteristic polynomials of elements with multiplicator q in the finite group $SSp(2g, \mathbf{F}_\ell)$ of symplectic similitudes to the "same" sieve applied to all q-symplectic polynomials of degree $2g$.

Recall that we denote by $m(g)$ the multiplicator for a symplectic similitude, i.e.,

$$\langle gv, gw \rangle = m(g)\langle v, w \rangle.$$

Lemma 7.2. *Let* $g \geqslant 1$ *and* ℓ *a prime. Put*

$$\Upsilon_{g,\ell} = \{f \in \mathbf{F}_\ell[T] \mid f \text{ is } q\text{-symplectic of degree } 2g\}.$$

Let $\tilde{\Omega}(\ell) \subset \Upsilon_{g,\ell}$ *be an arbitrary subset of cardinality* $\tilde{\omega}(\ell)$ *and*

$$\Omega(\ell) = \{g \in SSp(2g, \mathbf{F}_\ell) \mid m(g) = q, \text{ and } \deg(1 - Tg) \in \tilde{\Omega}(\ell)\},$$

with cardinality $\omega(\ell)$.
Then we have

$$\omega(\ell)|Sp(2g, \mathbf{F}_\ell)|^{-1} \geqslant \tilde{\omega}(\ell)(\ell + 1)^{-g}.$$

Proof. We have

$$\omega(\ell) = \sum_{f \in \tilde{\Omega}(\ell)} |\{g \in SSp(2g, \mathbf{F}_\ell) \mid m(g) = q \text{ and } \det(1 - Tg) = f\}|.$$

The inner quantity, for given f, is exactly what is estimated by Chavdarov in [C, Th. 3.5], in the proof of which it is called Δ. Using the formula at the bottom of page 159 of loc. cit., we get

$$\omega(\ell) = \frac{1}{\ell^g} \sum_{f \in \tilde{\Omega}(\ell)} |Sp(2g, \mathbf{F}_\ell)| \frac{\ell^{\delta(f)}}{|C(A_f)(\mathbf{F}_\ell)|},$$

---

[3] The terminology "self-reciprocal" is often used when $q = 1$.

where A_f is a fixed semisimple element in $SSp(2g, \mathbf{F}_\ell)$ with multiplicator q and characteristic polynomial f (its existence being proved in [C, Lemma 3.4]), and $C(A_f)$ is the centralizer of A_f in $Sp(2g)$, $\delta(f) \leqslant g$ being its dimension. Thus

$$\omega(\ell)|Sp(2g, \mathbf{F}_\ell)|^{-1} = \ell^{-g} \sum_{f \in \tilde{\Omega}(\ell)} \frac{\ell^{\delta(f)}}{|C(A_f)(\mathbf{F}_\ell)|}.$$

By the formula of Nori at the top of page 160 of loc. cit., which holds essentially because $C(A_f)$ is known to be a geometrically irreducible variety of dimension $\delta(f) \leqslant g$, we have

$$\omega(\ell)|Sp(2g, \mathbf{F}_\ell)|^{-1} \geqslant \ell^{-g} \sum_{f \in \tilde{\Omega}(\ell)} \left(1 - \frac{1}{\ell+1}\right)^{\delta(f)} \geqslant \ell^{-g}\left(1 - \frac{1}{\ell+1}\right)^g \tilde{\omega}(\ell),$$

as required. $\qquad\square$

The next results are technical estimates which are only required in this precise form for the proof of the uniform version of Chavdarov's theorem. Easier versions (found in [G], [DDS]) suffice for Theorem 6.1.

Recall the following terminology: if f is a monic polynomial of degree g in $\mathbf{Z}[T]$ which factorizes modulo a prime ℓ as

$$f = f_1 \cdots f_r$$

with the f_i coprime, irreducible, of degree $d_i \geqslant 1$, then one says that the *cycle type* (or the conjugacy class) associated to f is the conjugacy class in $\mathfrak{S}_g$ of elements which are product of disjoint cycles of lengths $d_1, \ldots, d_r$.

Lemma 7.3. (i) *Let $g \geqslant 1$ and let c be a conjugacy class in $\mathfrak{S}_g$. For ℓ prime, let*

$$\hat{\Omega}_c(\ell) = \{f \in \mathbf{F}_\ell[T] \mid f \text{ is monic of degree } g \text{ and the cycle type associated to } f \text{ is } c\},$$

and $\hat{\omega}_c(\ell) = |\hat{\Omega}_c(\ell)|$. Then we have for $\ell > 4g^2$

$$\hat{\omega}_c(\ell) \geqslant \frac{|c|}{|\mathfrak{S}_g|}(\ell-1)^g\left(1 - \frac{1}{\sqrt{\ell}}\right)^g.$$

(ii) *Let $g \geqslant 1$ and for ℓ prime let $\omega_1(\ell)$ be the number of q-symplectic irreducible polynomials in $\mathbf{F}_\ell[T]$ of degree $2g$. Then for $\ell > 4g^2$ we have*

$$\omega_1(\ell) \geqslant \frac{\ell^g}{2g}\left(1 - \frac{1}{\ell}\right)^g - \ell^{g/2}.$$

(iii) *Let $g \geqslant 1$ and for ℓ prime let $\omega_2(\ell)$ be the number of q-symplectic polynomials of degree $2g$ which factorize as a product of an irreducible quadratic polynomial and a product of irreducible polynomials of odd degrees. Then we have for $\ell > 4g^2$*

$$\omega_2(\ell) \geqslant \frac{\ell^g}{4g}\left(1 - \frac{1}{\ell}\right)^g.$$

Proof. We start with (i). If the conjugacy class c is that consisting of permutations with r_i distinct i-cycles in their decomposition, with $1 \cdot r_1 + \cdots + g \cdot r_g = g$, then we have

$$\frac{|c|}{|\mathfrak{S}_g|} = \prod_{1 \leqslant i \leqslant g} \frac{1}{i^{r_i} r_i!}, \text{ and } \hat{\omega}_c(\ell) = \prod_{1 \leqslant i \leqslant g} \binom{p(i, \ell)}{r_i},$$

where $p(i, \ell)$ is the number of irreducible monic polynomials of degree i in $\mathbf{F}_\ell[T]$. Now we claim that we have for all $\ell > 4g^2$ and $1 \leqslant i \leqslant g$ the lower bounds

(7.1)
$$p(1, \ell) \geqslant \ell\left(1 - \frac{1}{\sqrt{\ell}}\right)\left(1 - \frac{1}{\ell}\right) + g - 1,$$

(7.2)
$$p(i, \ell) \geqslant \frac{\ell^i}{i}\left(1 - \frac{1}{\ell}\right) + \frac{g}{i} - 1 \text{ for } 2 \leqslant i \leqslant g.$$

From this, which we prove below, we derive for all i, $2 \leqslant i \leqslant g$ and $r_i \leqslant g/i$ that

$$\binom{p(i,\ell)}{r_i} = \frac{p(i,\ell)(p(i,\ell)-1)\cdots(p(i,\ell)-r_i+1)}{r_i!} \geqslant \frac{(p(i,\ell)-g/i+1)^{r_i}}{r_i!}$$

$$\geqslant \left(1-\frac{1}{\ell}\right)^{r_i}\frac{1}{i^{r_i}r_i!}\ell^{ir_i}$$

and for $i = 1$, $r_i \leqslant g$ that

$$\binom{p(1,\ell)}{r_1} = \frac{p(1,\ell)(p(1,\ell)-1)\cdots(p(1,\ell)-r_1+1)}{r_1!} \geqslant \frac{(p(1,\ell)-g+1)^{r_1}}{r_1!}$$

$$\geqslant \left(1-\frac{1}{\sqrt{\ell}}\right)^{r_1}\left(1-\frac{1}{\ell}\right)^{r_1}\frac{1}{1^{r_1}r_1!}\ell^{r_1}.$$

Hence, putting these together, we get

$$\hat{\omega}_c(\ell) \geqslant \left(1-\frac{1}{\sqrt{\ell}}\right)^{r_1}\left(1-\frac{1}{\ell}\right)^{\sum r_i}\left(\prod_{1\leqslant i\leqslant g}\frac{1}{i^{r_i}r_i!}\right)\ell^{\sum ir_i}$$

$$= \frac{|c|}{|\mathfrak{S}_g|}\left(1-\frac{1}{\sqrt{\ell}}\right)^{r_1}\left(1-\frac{1}{\ell}\right)^{\sum r_i}\ell^g$$

$$\geqslant \frac{|c|}{|\mathfrak{S}_g|}\ell^g\left(1-\frac{1}{\ell}\right)^g\left(1-\frac{1}{\sqrt{\ell}}\right)^g$$

as desired.

Now we prove (7.1) and (7.2). We use the well-known formula of Dedekind

$$p(i,\ell) = \frac{1}{i}\sum_{d|i}\mu(d)\ell^{i/d}.$$

In particular

$$p(1,\ell) = \ell \geqslant \ell\left(1-\frac{1}{\sqrt{\ell}}\right)\left(1-\frac{1}{\ell}\right) + g - 1$$

for $\ell > 4g^2$ by inspection. Similarly for $i = 2$ we have

$$p(2,\ell) = \frac{1}{2}(\ell^2-\ell) \geqslant \frac{1}{2}(\ell-1)^2 + \frac{g}{2} - 1$$

if $\ell \geqslant g$. For $i \geqslant 3$ we use the lower bound

$$p(i,\ell) \geqslant \frac{\ell^i}{i} - \ell^{i/2}$$

(see e.g. [C, Lemma 3.1]), so that it suffices to show that

$$\frac{\ell^i}{i} - \ell^{i/2} > \frac{\ell^i}{i}\left(1-\frac{1}{\ell}\right) + \frac{g}{i}$$

for $\ell > 4g^2$. This is equivalent with

$$\frac{1}{\ell}X^2 - iX - g > 0 \text{ where } X = \ell^{i/2},$$

and the quadratic polynomial has largest root equal to

$$\alpha = \frac{\ell}{2}(i + \sqrt{i^2 + 4g/\ell}) < 2\ell i$$

for $i \geqslant 3$ and $g < \sqrt{\ell}/2$. Hence for $i \geqslant 3$, $\ell^{1/2} > 2g$ we have trivially $X = \ell^{i/2} > 2\ell g \geqslant 2\ell i > \alpha$, so the quadratic polynomial must be > 0 when evaluated at X, which gives (7.2).

Coming to (ii) we have (compare [DDS, Lemma 3]) the lower bound

$$\omega_1(\ell) \geqslant p(g,\ell) - \frac{1}{2}p(g,\ell) - \ell^{g/2}.$$

This is because we can count irreducible polynomials of degree g in $\mathbf{F}_\ell[T]$, minus those for which $f = T^g h(qT + T^{-1})$ is reducible; in this case, f is of the form $ch(T)T^g h(qT^{-1})$ (for some

normalizing constant $c \neq 0$) where h is irreducible of degree g and *not* q-symplectic, with both h and $T^g h(qT^{-1})$ yielding (with proper normalization factor, so they are distinct up to scalars by virtue of h not being q-symplectic) the same reducible f. From the irreducible h, we exclude the q-symplectic ones by the trivial bound $\ell^{g/2}$ for their number, hence the inequality above.

Using (7.2) for $i = g$ we get

$$\omega_1(\ell) \geqslant \frac{\ell^g}{2g}\left(1 - \frac{1}{\ell^g}\right) - \ell^{g/2}.$$

Finally for (iii) we consider separately the case where g is even or when g is odd. For even g, the number $\omega_2(\ell)$ is larger than that of q-symplectic polynomials f of degree $2g$ of the form

$$f = f_1 h_1 h_2$$

where f_1 is an irreducible quadratic q-symplectic polynomial, and h_1, h_2 are irreducible of odd degree $g-1$ with, up to a constant, $h_1 = T^g h_2(qT^{-1})$. Counting the possibilities we get by (7.2) that

$$\omega_2(\ell) \geqslant \frac{\ell}{2}\left(1 - \frac{1}{\ell}\right) p(g-1, \ell) \geqslant \frac{\ell}{2}\left(1 - \frac{1}{\ell}\right)\frac{\ell^{g-1}}{g-1}\left(1 - \frac{1}{\ell}\right)^{g-1} = \frac{\ell^g}{2(g-1)}\left(1 - \frac{1}{\ell}\right)^g$$

hence a stronger result than claimed in this first case.

The case where g is odd is similar with polynomials of the form $f = f_1 f_2 f_3 h_1 h_2$ with f_1 quadratic irreducible and q-symplectic as before, $f_2 = 1 - \alpha T$ for some $\alpha \neq 0$ and $f_3 = 1 - q\alpha^{-1}T$, and h_1, h_2 as in the even case but now with odd degree $g-2$. One gets a denominator $4(g-2) \leqslant 4g$ this time. □

Remark 7.4. As a by-product of these estimates, applying Gallagher's method we can derive a uniform version of his estimate [G] for the number $E_n(N)$ of monic polynomials $f = \sum a_i T^i \in \mathbf{Z}[T]$ of degree $n \geqslant 1$ with height $\max |a_i| \leqslant N$ and Galois group strictly smaller than $\mathfrak{S}_n$, namely

$$(7.3) \qquad\qquad E_n(N) \ll n^2 (2N+1)^{n-1/2}(\log N)$$

with an absolute implied constant (note $(2N+1)^n$ is the number of polynomials with height $\leqslant N$).

This is much more impressive than our Theorem 6.2 because the gain in the exponent (namely, $\frac{1}{2}$) is independent of the degree n so the bound is non-trivial for n as large as $(2N)^{1/4}/(\log N)$.

Similarly for reciprocal polynomials, as treated in [DDS], denoting by $\mathcal{E}_m(N)$ the number of monic polynomials in $f \in \mathbf{Z}[T]$ of degree $2m$ with height $\leqslant N$ such that $T^m f(T^{-1}) = f$ we get

$$\mathcal{E}_m(N) \ll m^2 (2N+1)^{m-1/2}(\log N)$$

with an absolute implied constant.

Note that in the two papers quoted, the fundamental large-sieve inequality is not uniform in n (resp. m) as stated, but becomes so if one replaces it by the form given in [Hu, Th. 1], with some obvious changes. For instance in [G, eq. (5)], the term $(N^n + x^{2n})$ in the right-hand side must be replaced by $(\sqrt{2N+1} + x)^{2n}$. To make this innocuous one may take $x = n^{-1}\sqrt{2N+1}$ instead of $x = \sqrt{N}$, which leads to $\pi(x)^{-1} \ll n(2N+1)^{-1/2}(\log N)$, hence the "extra" power of n in (7.3) compared with the "density" of irreducible polynomials modulo ℓ (i.e, about n^{-1}) which are used to sieve the reducible ones.

8. Proof of the uniform version of Chavdarov's theorem

We can now start the proof of Theorem 6.2 itself. We will apply Proposition 3.3 with the following data: in addition to U, which is a smooth geometrically connected affine curve over $\mathbf{F}_q$, we take the sheaves $\tilde{\mathcal{F}}_\ell = R^1 f_! \mathbf{F}_\ell$ for $\ell \in \Lambda$, where Λ is the set of odd primes

These sheaves are of course obtained by reduction modulo ℓ from the compatible system $\mathcal{F}_\ell = R^1 f_! \mathbf{Z}_\ell$. The existence of the symplectic pairing (6.2) implies that the arithmetic monodromy

group of $\tilde{\mathcal{F}}_\ell$ can be seen as a subgroup of $SSp(2g, \mathbf{F}_\ell)$, and for any $u \in U(\mathbf{F}_q)$, the image of Fr_u has multiplicator q.

The most crucial point is that for $\ell > 2$, J-K. Yu [Yu][4] has shown that the geometric monodromy group for $\tilde{\mathcal{F}}_\ell$ is equal to $Sp(2g, \mathbf{F}_\ell)$. Then the sheaves $(\tilde{\mathcal{F}}_\ell)$ are also linearly disjoint as a consequence of Goursat's Lemma (see Corollary 2.6, (1)), and by Lemma 2.3 we have (2.3) with $\kappa = 2$. And finally, Lemma 2.4, (3) gives us (2.4) with $s = 2g^2 + g + 1$, $t = g + 1$ and $c_1 = 1$, $c_2 = 6^g$.

Thus all conditions needed to apply Proposition 3.3 (in the case of a one-parameter family) are valid, and it remains to set up the sieving problem. The principle for this is exactly the same as the one introduced by Gallagher for polynomials with integer coefficients and bounded height [G].

As in Lemma 7.2, for any choice of sets $\tilde{\Omega}(\ell) \subset \Upsilon_{g,\ell}$ defined for $\ell > 2$ we let

$$\Omega(\ell) = \{g \in SSp(2g, \mathbf{F}_\ell) \mid m(g) = q, \text{ and } \deg(1 - Tg) \in \tilde{\Omega}(\ell)\}.$$

Applying Proposition 3.3 (see (3.3)) to such a sieving problem, we have

(8.1) $$|S(U, \Omega; L)| \leqslant (2q + 4gq^{1/2}(6L)^A)P(L)^{-1},$$

where $A = 2g^2 + 3g/2 + 5/2$ and

(8.2) $$P(L) = \sum_{2 < \ell \leqslant L} \omega(\ell)|G_\ell^g|^{-1},$$

(which shouldn't be confused with the polynomials P_u), and here we have taken the constant $C = 4g$ by looking at Proposition 4.1 and (5.4), (5.6) since $1 - \chi(\overline{U}, \overline{\mathbf{Q}}_\ell) = 2g$ and it is known that all the sheaves $\mathcal{F}_\ell$ are tamely ramified (by [KS1, Lemma 10.1.12]) so that the contribution w of the Swan conductors in (4.1) vanishes. Moreover, by Lemma 7.2 we have

(8.3) $$P(L) \geqslant \sum_{2 < \ell \leqslant L} \tilde{\omega}(\ell)(\ell + 1)^{-g}.$$

Now we must show how to use this sieve estimate to study the characteristic polynomials P_u. For this we need to recall the following two facts, the first of which is classical, while the second is much deeper:

(i) if $f \in \mathbf{Z}[T]$ is a polynomial of degree d that factorizes in $\mathbf{F}_\ell[T]$ as a product of coprime polynomials $f_1 \cdots f_r$, with f_i irreducible of degree $\deg f_i = d_i$, then the Galois group of f, seen as a permutation group of the complex roots of f, contains a cycle c of type $(d_1, \ldots, d_r)$, i.e., a product of disjoint cycles of respective length $d_1, \ldots, d_r$ (see e.g. [vdW, §61], [J, p. 302]).

(ii) the reduction modulo a prime ℓ of a polynomial P_u (the numerator of the zeta function of the curve $C_u = \pi^{-1}(u)$ is the characteristic polynomial of Fr_u acting on $\tilde{\mathcal{F}}_\ell$ (see [D2, Fonctions L mod. ℓ^n], or use the fact that

$$\det(1 - T\,\mathrm{Fr}_u \mid R^1\pi_! \mathbf{Z}_\ell) = \det(1 - T\,\mathrm{Fr} \mid H^1(\overline{C}_u, \mathbf{Z}_\ell)) = P_u$$

by (6.1), and reduce modulo ℓ).

Thus (ii) allows us to control the reduction of a polynomial P_u, while (i) tells us that the reduction gives information on the Galois group of P_u.

In particular, for any sieving sets $\Omega(\ell) \subset SSp(2g, \mathbf{F}_\ell)$, an element $u \in S(U, \Omega; L)$ will have the property that the Galois group of P_u, seen as a subgroup of $\mathfrak{S}_{2g}$, does not contain a cycle c associated to an $f \in \Omega(\ell)$, where ℓ ranges over primes $2 < \ell \leqslant L$.

If we have finitely many sieving sets Ω_i, $1 \leqslant i \leqslant m$, defined by the condition that the cycle associated to $\det(1 - Tg)$ is in a certain set c_i of conjugacy classes, and if moreover those c_i

---

[4] Unfortunately, this result – quoted both in [C] and [KS1] – is unpublished; the proof proceeds by "lifting" to characteristic zero. If the reader does not wish to take this statement for granted, notice that the rational geometric monodromy group of $R^1 f_! \mathbf{Q}_\ell$ is computed independently in [KS1, Th. 10.1.16]. Together with the result of Larsen quoted below in Section 9, this suffices to obtain a (weaker) form of Theorem 6.2, namely for fixed characteristic and $q \to +\infty$. Also, C. Hall has recently given a new simpler proof of this result (see his forthcoming paper).

have the property that the only subgroup $W \subset W_{2g}$ containing an element of each c_i is W_{2g}, then it follows that the set of exceptional $u \in U(\mathbf{F}_q)$ with P_u having small Galois group will be a subset of the union of the $S(U, \Omega_i; L)$. So in such a situation we have

$$(8.4) \qquad N(U/\mathbf{F}_q) \leqslant S(U, \Omega_1; L) + \cdots + S(U, \Omega_m; L) \leqslant (2q + 4gq^{1/2}L^A) \sum_{1 \leqslant i \leqslant m} P_i(L)^{-1}.$$

Lemma 7.1 describes three possible choices of sets c_i; however, the first and the second involve some c_i which are "too small", so the dependency on g in the estimate for $P_c(L)$ is bad (they are perfectly suitable for fixed g). Thus we use Case (iii) of Lemma 7.1. Precisely, we have $m = 4$ and the four sets Ω_i can be described as follows:

(i) Ω_1 is the set of irreducible polynomials $f \in \Upsilon_{g,\ell}$.

(ii) Ω_2 is the set of polynomials $f \in \Upsilon_{g,\ell}$ which factorize as a product of an irreducible quadratic polynomial and a product of irreducible polynomials of odd degrees.

To define Ω_3 and Ω_4, we recall that any $f \in \Upsilon_{g,\ell}$ can be written uniquely

$$f = T^g h(qT + T^{-1})$$

where $h \in \mathbf{F}_\ell[T]$ is a monic polynomial of degree g.

(iii) Ω_3 is the set of $f \in \Upsilon_{g,\ell}$ such that the corresponding h has an irreducible factor of prime degree $> g/2$.

(iv) Ω_4 is the set of $f \in \Upsilon_{g,\ell}$ such that the corresponding h has a single quadratic irreducible factor and no other irreducible factor of even degree.

We claim that those sets do allow us to sieve the exceptional elements u. Indeed, spelling out again the relation between the factorization of P_u modulo ℓ and the existence of elements in the Galois group of P_u with the associated cycle type, we see that:

(i) If P_u is reducible then $u \in S(U, \Omega_1; L)$.

(ii) If P_u is irreducible but the Galois group W does not contain a transposition, then $u \in S(U, \Omega_2; L)$, since having $P_u \, (\mathrm{mod}\, \ell) \in \tilde{\Omega}_2(\ell)$ implies that W contains an element with cycle type consisting of one 2-cycle and further cycles of odd length, a power of which will be a transposition.

For the next two facts, notice that the cycle in $\mathfrak{S}_g$ associated to the polynomial Q_u such that $P_u = T^g Q_u(qT + T^{-1})$ is the image by the map $p : W_{2g} \to \mathfrak{S}_g$ of the cycle associated to P_u.

(iii) If P_u is irreducible but $p(W)$ does not contain a cycle of prime order $m > g/2$, then $u \in S(U, \Omega_3; L)$.

(iv) If P_u is irreducible but $p(W)$ does not contain a transposition, then $u \in S(U, \Omega_4; L)$ (as in Case (ii) previously).

By Case (iii) of Lemma 7.1, the $u \in U(\mathbf{F}_q)$ that we wish to exclude are therefore in the union of the $S(U, \Omega_i; L)$, and we conclude that

$$(8.5) \quad N(U/\mathbf{F}_q) \leqslant S(U, \Omega_1; L) + \cdots + S(U, \Omega_4; L) \leqslant 4(2q + 4gq^{1/2}(6L)^A)\Big(\min_{1 \leqslant i \leqslant 4} \sum P_i(L)\Big)^{-1}.$$

It remains to give appropriate lower bounds of $P_i(L)$. For Ω_3 and Ω_4, since the correspondence between polynomials $f \in \Upsilon_{g,\ell}$ and the $h \in \mathbf{F}_\ell[T]$ such that $f = T^g h(qT + T^{-1})$ is one-to-one, we can count the corresponding h by Lemma 7.3, applied to the cycle types (i.e., conjugacy classes) in $\mathfrak{S}_g$ associated to the polynomials in Ω_i. For $\ell > 4g^2$ and $i \in \{3, 4\}$, denoting by C_i the set of elements in $\mathfrak{S}_g$ having the associated cycle type, we get

$$\tilde{\omega}_i(\ell) \geqslant \frac{|C_i|}{|\mathfrak{S}_g|}(\ell - 1)^g\Big(1 - \frac{1}{\sqrt{\ell}}\Big)^g,$$

and thus for $L > 4g^2$ we have

$$P_i(L) \geqslant \frac{|C_i|}{|\mathfrak{S}_g|} \sum_{4g^2 < \ell \leqslant L} \Big(\frac{\ell - 1}{\ell + 1}\Big)^g\Big(1 - \frac{1}{\sqrt{\ell}}\Big)^g.$$

By the mean-value theorem we have for any $\ell \geqslant 2$

$$\left(\frac{\ell-1}{\ell+1}\right)^g \left(1 - \frac{1}{\sqrt{\ell}}\right)^g = 1 - gh(\ell) + O(g^2 h(\ell)^2)$$

with

$$h(\ell) = \frac{2}{\ell+1} + \frac{1}{\sqrt{\ell}} - \frac{2}{\sqrt{\ell}(\ell+1)},$$

and an absolute implied constant. Inserting this in the sum and using the prime number theorem we get for $L > 4g^2$ that

$$(8.6) \qquad P_i(L) \geqslant \frac{|C_i|}{|\mathfrak{S}_g|}\{\pi(L) + O(g\sqrt{L}(\log L)^{-1} + g^2 \log\log L)\},$$

with an absolute implied constant.

By [G, p. 99] (where our C_3 is denoted P and C_4 is denoted T), we have for $g \geqslant 1$

$$(8.7) \qquad \frac{|C_3|}{|\mathfrak{S}_g|} \gg \frac{1}{\log 2g} \quad \text{and} \quad \frac{|C_4|}{|\mathfrak{S}_g|} \gg \frac{1}{\sqrt{g}}.$$

Using (8.6), this gives the lower bounds

$$(8.8) \qquad P_3(L) \gg \frac{1}{\log 2g} L(\log L)^{-1}, \quad \text{and} \quad P_4(L) \gg \frac{1}{\sqrt{g}} L(\log L)^{-1}$$

with absolute implied constants for $L \gg g^2(\log 2g)$ (i.e. for $L \geqslant \alpha_1 g^2(\log 2g)$, where the absolute constant α_1 can be specified from the implied constants in (8.6) and (8.7)).

Coming to Ω_1, we have by (ii) of Lemma 7.3 that for $\ell > 4g^2$

$$\tilde\omega_1(\ell) \geqslant \frac{\ell^g}{2g}\left(1 - \frac{1}{\ell^g}\right) - \ell^{g/2}$$

so by (8.3), the prime number theorem and the mean-value theorem as before we get for $L > 4g^2$ that

$$P_1(L) \geqslant \frac{1}{2g}(\pi(L) + O(g\log\log L + g^2 + \sqrt{L}))$$

with an absolute implied constant, and hence for $L \gg g^2(\log 2g)$, we have

$$(8.9) \qquad P_1(L) \gg \frac{1}{g} L(\log L)^{-1}$$

with absolute implied constant.

Finally by (iii) of Lemma 7.3 we have for $\ell > 4g^2$

$$\tilde\omega_2(\ell) \geqslant \frac{1}{4g}\left(1 - \frac{1}{\ell}\right)^g \quad \text{and} \quad P_2(L) \geqslant \frac{1}{4g}(\pi(L) + O(g\log\log L + g^2))$$

and for $L \gg g^2(\log 2g)$ we obtain also

$$(8.10) \qquad P_2(L) \gg \frac{1}{g} L(\log L)^{-1}$$

with absolute implied constant.

Altogether from (8.5), (8.8), (8.9) and (8.10) we get

$$N(f,q) \ll g^2(2q + q^{1/2}(6L)^A)L^{-1}(\log L)$$

with an absolute implied constant, which can in fact be chosen so that the inequality is valid for all $L \geqslant 2$ and $g \geqslant 1$, since it becomes trivial for $g^2 \gg L(\log L)^{-1}$. Choosing $6L = q^{(2A)^{-1}} = q^{(4g^2+3g+5)^{-1}}$, with $\log L \ll g^{-2}\log q$, this gives the announced uniform estimate

$$N(f,q) \ll q^{1-\gamma}(\log q)$$

with $\gamma = (4g^2 + 3g + 5)^{-1}$, and an absolute implied constant.

9. Proof of the general version of Chavdarov' Theorem

We will now quickly prove Theorem 6.1, only highlighting the points where the proof is different from that of the previous section. The first step is to check that we can always apply Proposition 3.3 to the data consisting of $U/\mathbf{F}_q$ and the family of sheaves $\tilde{\mathcal{F}}_\ell = R^1\pi_!\mathbf{F}_\ell$, defined for a subset Λ of primes $\ell > L_0$.

Since our assumption is that for $\ell > L_0$ the geometric monodromy group of $R^1\pi_!\mathbf{Z}_\ell$ is the symplectic group $Sp(2g)$ (as algebraic group over $\mathbf{Q}_\ell$), we must show that this implies that the monodromy group modulo ℓ is often large. (A priori, for fixed ℓ, the assumption only implies that the index of the image of $\pi_1(\overline{U}, \overline{\eta})$ in $Sp(2g, \mathbf{Z}/\ell^\nu\mathbf{Z})$ is bounded for $\nu \geqslant 1$, but does not say anything for $\nu = 1$). However, we can appeal to a result of Larsen [La, Th. 3.17] which implies that for a set of primes Λ_0 of (natural) density 1, we do have $G_\ell^g = Sp(2g, \mathbf{F}_\ell)$ because the sheaves come from a compatible system. (Precisely, in the notation of loc. cit., apply Th. 3.17 with $\mathcal{G} = \pi_1(\overline{U}, \overline{\eta})$, ρ_ℓ corresponding to $R^1\pi_!\mathbf{Z}_\ell$, so that by assumption $G_\ell = Sp(2g)/\mathbf{Q}_\ell$, which is connected and simply connected so $G_\ell^{sc} = Sp(2g)$, and look at the first few lines of the proof of Th. 3.17 to make sure that the statement there involving "hyperspecial maximal compact subgroups" does imply that the geometric monodromy group of the reduction of ρ_ℓ is $Sp(2g, \mathbf{F}_\ell)$ for ℓ in a set of density 1; note also that Larsen's result is quite deep as it depends on the classification of simple finite groups).

Then, as before, Corollary 2.6 implies that the sheaves $\tilde{\mathcal{F}}_\ell$ for $\ell \in \Lambda_0$ are linearly disjoint in all cases (if $\ell \geqslant 5$) by the assumption on the geometric monodromy groups, and (2.3) holds with $\kappa = 2$ by Lemma 2.3, (2).

Since $\tilde{\mathcal{F}}_\ell$ is obtained by reduction of the compatible system $\mathcal{F}_\ell = R^1\pi_!\mathbf{Z}_\ell$, case (i) of Proposition 3.3 is applicable if U is a curve, with Λ consisting of all the primes in Λ_0.

If U is not a curve but $p > 2g + 1$, we use the following simple lemma (compare [K5, Lemma 7.5.1]) :

Lemma 9.1. *Let $r \geqslant 1$. For any $p > r + 1$, there exists $\alpha \in (\mathbf{Z}/p\mathbf{Z})^\times$ such that the order of $GL(r, \mathbf{F}_\ell)$ is prime to p for any prime $\ell \equiv \alpha \,(\mathrm{mod}\, p)$.*

Proof. The order of $GL(r, \mathbf{F}_\ell)$ is

$$|GL(r, \mathbf{F}_\ell)| = \ell^{r(r-1)/2} \prod_{1 \leqslant i \leqslant r} (\ell^i - 1)$$

so the condition will hold whenever the order of α modulo p is $> r$. If $p > r + 1$, a primitive root modulo p will certainly work. $\square$

For $p > 2g + 1$, we can apply the second case of Theorem 3.1 and Proposition 3.3 with Λ consisting of primes in Λ_0 which are congruent modulo p to the α given by this lemma for $r = 2g$. This set has positive density among the primes because Λ_0 has density 1.

We can now define sieving sets analogous to the previous Ω_i, $1 \leqslant i \leqslant 4$, for $U/\mathbf{F}_q$ and we have (8.5). Since we consider g to be fixed here, we can rewrite (8.3) as

$$P_i(L) \gg \sum_{\substack{L_0 < \ell \leqslant L \\ \ell \in \Lambda}} \tilde{\omega}_i(\ell)\ell^{-g}$$

where the implied constant depends on g. Then we obtain

$$P_i(L) \gg \pi(L)$$

for $L > L_0$, the implied constant depending on g and p in Case (ii) (through the density of Λ), either from Lemma 7.3 or directly from Dedekind's formula used in its proof.

Hence we obtain by Proposition 3.3

$$N(U/\mathbf{F}_q) \ll (q^d + Cq^{d-1/2}L^A)L^{-1}(\log L),$$

for $L > L_0$, where the implied constant depends on g, and on p in Case (ii). If we take

$$L = q^{(4g^2+3g+5)^{-1}}, \text{ in case (i)}, \qquad L = q^{(12g^2+7g+9)^{-1}}, \text{ in case (ii)}$$

we have $q^{d-1/2}L^A = q^d$, hence

$$N(U/\mathbf{F}_q) \ll q^d L^{-1}(\log L) \ll q^{d-\gamma}(\log q)$$

for $\gamma = \frac{1}{4g^2+3g+5}$ (resp $\gamma = \frac{1}{12g^2+7g+9}$), as desired.

The proof of Corollary 6.5 is similar, applying first the large sieve to a suitably rigidified moduli space $A_g/\mathbf{F}_q$ of principally polarized abelian varieties over $\mathbf{F}_q$, for instance the moduli space $A_{g,3\mathcal{L}}$ of [KS1, 11.3]. Strictly speaking we need to restrict to a smooth connected affine subscheme $U \subset A_{g,3\mathcal{L}}$, but this is not a problem as observed in the remarks after Theorem 3.1 (see Remark 4). Over U we have a universal family $\pi : \mathcal{A}_{g,3\mathcal{L}} \to U$ and we take the sheaves $\mathcal{F}_\ell = R^1\pi_!\mathbf{Z}_\ell$ and their reductions $\mathcal{F}_\ell/\ell\mathcal{F}_\ell$. The monodromy groups are as large as possible because already this is the case for the families of (canonically principally polarized) jacobians of the hyperelliptic curves of genus g of Theorem 6.2. After applying Proposition 3.3 to the same sieving problem as in Theorem 6.1, we go (with the same saving) from the number of "exceptional" principally polarized abelian varieties with a $3\mathcal{L}$-structure to the number $N(g,q)$ by dividing out by the free rigidifying parameters and considering the situations with extra automorphisms, as done in [KS1, 11.3] for the case of curves.

References

[A]   E. Artin: *Geometric algebra*, Interscience Tracts in Pure and Applied Math., vol. 3, (1957).

[B]   E. Bombieri: *Le grand crible dans la théorie analytique des nombres*, Astérisque 18, S.M.F (1974).

[C]   N. Chavdarov: *The generic irreducibility of the numerator of the zeta function in a family of curves with large monodromy*, Duke Math. J. 87 (1997), 151–180.

[DDS]   S. Davis, W. Duke and X. Sun: *Probabilistic Galois theory of reciprocal polynomials*, Exposition. Math. 16 (1998), no. 4, 263–270.

[D1]   P. Deligne: *La conjecture de Weil, II*, Publ. Math. I.H.E.S 52 (1981), 313–428.

[D2]   P. Deligne: *Cohomologie étale*, S.G.A $4\frac{1}{2}$, L.N.M 569, Springer Verlag (1977).

[G]   P.X. Gallagher: *The large sieve and probabilistic Galois theory*, in Proc. Sympos. Pure Math., Vol. XXIV, Amer. Math. Soc. (1973), 91–101.

[Gr]   A. Grothendieck: *Formule de Lefschetz et rationalité des fonctions L*, Séminaire Bourbaki (1964–65), North Holland, 1968, exp. 279, 1–15.

[HW]   G.H. Hardy and E.M. Wright: *An introduction to the theory of numbers*, Fifth Edition, Oxford 1979.

[Hu]   M.N. Huxley: *The large sieve inequality for algebraic number fields*, Mathematika 15 (1968), 178–187.

[I]   L. Illusie: *Théorie de Brauer et caractéristique d'Euler-Poincaré*, Séminaire E.N.S (1978–79), exp. VIII, Astérisque 82–83 (1981), 161–172.

[IK]   H. Iwaniec and E. Kowalski: *Analytic Number Theory*, A.M.S Colloquium Publ. 53, 2004.

[J]   N. Jacobson: *Basic Algebra I*, 2nd edition, W.H Freeman (1985).

[K1]   N. Katz: *Sums of Betti numbers in arbitrary characteristic*, Finite Fields Appl. 7 (2001), no. 1, 29–44.

[K2]   N. Katz: *Gauss sums, Kloosterman sums and monodromy*, Annals of Math. Studies, 116, Princeton Univ. Press, 1988.

[K3]   N. Katz: *Affine cohomological transforms, perversity and monodromy*, J. of the A.M.S 6 (1993), 149–222.

[K4]   N. Katz: *Sommes exponentielles*, Astérisque 79, S.M.F (1981).

[K5]   N. Katz: *Twisted L-functions and monodromy*. Annals of Math. Studies, 150, Princeton Univ. Press, 2002.

[KS1]   N. Katz and P. Sarnak: *Random matrices, Frobenius eigenvalues and monodromy*, A.M.S Colloquium Publ. 45, 1999.

[Ko2]   E. Kowalski: *Weil numbers generated by other Weil numbers and torsion fields of abelian varieties*, to appear in Journal of the LMS, `arXiv:math.NT/0504042`.

[Ko3]   E. Kowalski: *On the rank of quadratic twists of elliptic curves over function fields*, International J. Number Theory 2 (2006), 267–288. `arXiv:math.NT/0503732`.

[La]   M. Larsen: *Maximality of Galois actions for compatible systems*, Duke Math. J. 80 (1995), no. 3, 601–630.

[L]   G. Laumon: *Comparaison de caractéristiques d'Euler-Poincaré en cohomologie $\mathbf{Q}_\ell$-adique*, C.R.A.S Paris Sér. I Math. 292 (1981), 209–212.

[LP]   M. Liebeck and L. Pyber: *Upper bound for the number of conjugacy classes of a finite group*, J. of Algebra 198 (1997), 538–562.

[M]   J. Milne: *Étale cohomology*, Princeton Mathematical Series 33, Princeton Univ. Press, 1980.

[Mo]   H.L. Montgomery: *The analytic principle of the large sieve*, Bull. A.M.S 84 (1978), 547–567.

[vdW]   B.L. van der Waerden: *Moderne algebra*, vol I, Springer 1935.

[Yu]   J.-K. Yu: *Toward a proof of the Cohen-Lenstra conjecture in the function field case*, preprint (1996).

Université Bordeaux I - A2X, 351, cours de la Libération, 33405 Talence Cedex, France

E-mail address: emmanuel.kowalski@math.u-bordeaux1.fr