Vol. 54, No. 1 DUKE MATHEMATICAL JOURNAL © 1987

ON THE MONODROMY GROUPS ATTACHED TO CERTAIN FAMILIES OF EXPONENTIAL SUMS

Dedicated to Y. I. Manin on his fiftieth birthday

NICHOLAS M. KATZ

Introduction. The main theme of this paper is that very innocuous looking one-parameter families of exponential sums over finite fields can have quite strong variation as the parameter moves. Even when the parameter variety is the affine line $\mathbf{A}^1$ or the multiplicative group $\mathbf{G}_m$ over a finite field, the algebraic group G_{geom} which controls the variation is often "as large as possible". These results are the finite-field analogue of the fact that in characteristic zero, very simple differential equations on $\mathbf{A}^1$ and on $\mathbf{G}_{\text{m}}$ can have very large differential galois groups.

In Part 1 we develop some general results concerning irreducible lisse sheaves on open curves in characteristic $p > 0$, in part modeled on [Ka-2] and [Ka-Pi]. In Parts 2 and 3 we calculate G_{geom} for the one-parameter families of exponential sums in characteristic p which correspond to the Kloosterman and Airy differential equations respectively of any rank $n \geqslant 2$. It is very striking that in both of these cases, our results on G_{geom} are in perfect analogy with the results of [Ka-2] and [Ka-Pi] on the differential galois group G_{gal} of the corresponding differential equation, as soon as $p > 2n + 1$. Indeed, one can speculate that in some future "motivic grand unification", the two sorts of results will both be "realizations" of a single motivic result. For the moment, we must be content to offer the results themselves as indirect evidence for the existence of such a unification.

Part 1. Lisse sheaves on open curves: general results. Throughout this paper, we fix a prime number p, an algebraically closed field k of characteristic p, a prime number $\ell \neq p$, and an algebraic closure $\overline{\mathbf{Q}}_\ell$ of $\mathbf{Q}_\ell$. Let U be a smooth connected affine curve over k, and $\mathscr{F}$ a lisse $\overline{\mathbf{Q}}_\ell$-sheaf on U of rank $n \geqslant 1$. We denote by π_1 the fundamental group $\pi_1(U, \bar{\eta})$ of U with base point a geometric generic point $\bar{\eta}$ of U, by ρ the n-dimensional $\overline{\mathbf{Q}}_\ell$-representation of π_1 on $\mathscr{F}_{\bar{\eta}}$ which $\mathscr{F}$ "is", by G_{geom} the Zariski closure of $\rho(\pi_1)$ in $\mathrm{GL}(n, \overline{\mathbf{Q}}_\ell)$, and by $(G_{\text{geom}})^0$ the identity component of G_{geom}. We say that $\mathscr{F}$ is irreducible if ρ is irreducible as a representation of π_1, or equivalently if G_{geom} acts irreducibly in its given n-dimensional representation. We say that $\mathscr{F}$, or ρ, is Lie-irreducible if the restriction of ρ to $(G_{\text{geom}})^0$ is irreducible, or equivalently if the restriction of ρ

Received October 11, 1986.

to every open subgroup of π_1 is irreducible. (More generally, a continuous finite-dimensional representation ρ of a topological group is called Lie-irreducible if its restriction to every open subgroup of finite index is irreducible: clearly ρ is Lie-irreducible if and only if its restriction to some, or to every, open subgroup of finite index is Lie-irreducible.)

PROPOSITION 1. *Suppose that $\mathcal{F}$ is irreducible. Then either ρ is induced from a representation of a proper open subgroup of π_1, or ρ is Lie-irreducible, or there exists an integer $d \geqslant 2$ dividing n and a factorization of ρ as a tensor product $\tau \otimes \omega$, where τ is a Lie-irreducible representation of π_1 of dimension n/d, and where ω is an irreducible representation of π_1 of dimension d which factors through a finite quotient of π_1. In this last case, the pair (τ, ω) is unique up to replacing it by $(\tau \otimes \chi, \omega \otimes \chi^{-1})$ with χ a character of π_1 of finite order.*

Proof. Because $(G_{\text{geom}})^0$ is a normal subgroup of G_{geom} of finite index, either ρ as a G_{geom}-representation is induced from a proper subgroup H which contains $(G_{\text{geom}})^0$, or its restriction to $(G_{\text{geom}})^0$ is isotypical. If ρ is induced from an H, then ρ as π_1-representation is induced from the open subgroup $\rho^{-1}(H)$, and there is nothing to prove. If ρ is not induced, then its restriction to $(G_{\text{geom}})^0$ is isotypical, say isomorphic to d copies of an irreducible representation τ_0 of $(G_{\text{geom}})^0$. If $d = 1$, this exactly means that ρ is Lie-irreducible.

So suppose $d \geqslant 2$. Choose **any** open normal subgroup K of π_1 which is sufficiently small that $\rho(K)$ is contained in $(G_{\text{geom}})^0$. Then $\rho(K)$ is Zariski dense in $(G_{\text{geom}})^0$, so τ_0 is K-irreducible, and $\rho|K \approx d\tau_0$. Therefore the isomorphism class of τ_0 as K-representation is invariant by π_1-conjugation. Therefore τ_0 extends to a projective representation, say τ_1, or π_1. Because we are working on a smooth affine curve U over an algebraically closed field, we know that $H^2(\pi_1, \_\_)$ vanishes, so there is no obstruction to lifting the irreducible projective representation τ_1 to a (necessarily irreducible) linear representation τ_2 of π_1. Notice that on K, τ_2 is projectively equal to τ_0, so of the form $\tau_0 \otimes \chi$ for some character χ of K. Consider the d-dimensional representation ω_2 of π_1 defined as $\text{Hom}_K(\tau_2, \rho)$. Its restriction to K is scalar, namely χ^{-1}, and χ^{-d} is the restriction to K of the character $\det(\omega_2)$. Again by the vanishing of H^2, any $\overline{\mathbf{Q}}_\ell^\times$-valued character of π_1 has a dth root. Twisting τ_2 by a dth root of $\det(\omega_2)$, we obtain a representation τ of π_1 whose restriction to K differs from τ_0 by a character of order d. Shrinking K, we may assume that τ actually coincides with τ_0 on K. Because $\tau_0|K$ is Lie-irreducible, it follows immediately that τ is Lie-irreducible. If we define ω to be $\text{Hom}_K(\tau, \rho)$, we have the asserted factorization of ρ as $\tau \otimes \omega$, with ω factoring through a finite quotient of π_1, and with τ Lie-irreducible. That ω is itself irreducible results from the fact that $\tau \otimes \omega = \rho$ is irreducible.

To see the uniqueness of such a factorization of ρ, let $\tau' \otimes \omega'$ be another, and choose an open normal subgroup K of π_1 on which both ω and ω' are trivial. Then $\rho|K$ is equal both to $\dim(\omega)$ copies of $\tau|K$, and to $\dim(\omega')$ copies of $\tau'|K$, with both of these irreducible. Therefore ω and ω' have the same dimension, and $\tau|K \approx \tau'|K$. Therefore the space $\text{Hom}_K(\tau, \tau')$ is a one-dimensional representation χ of π_1 which is trivial on K, so of finite order, and by construction we have $\tau \otimes \chi \approx \tau'$. Writing ω as $\mathrm{Hom}_K(\tau, \rho)$, and ω' as $\mathrm{Hom}_K(\tau', \rho)$, we find $\omega \approx \omega' \otimes \chi$, as required. QED

Remark 2. Conversely, if τ is any Lie-irreducible representation of π_1, and ω any irreducible representation of π_1 which factors through a finite quotient, then the tensor product $\tau \otimes \omega$ is irreducible. For let Λ be any subrepresentation. Choose an open normal subgroup K on which ω is trivial. Then $\Lambda | K$ is a sum of copies of $\tau | K$, so we have $\Lambda \approx \tau \otimes \mathrm{Hom}_K(\tau, \Lambda)$ as π_1-representation. But the second factor $\mathrm{Hom}_K(\tau, \Lambda)$ is a subrepresentation of $\mathrm{Hom}_K(\tau, \tau \otimes \omega) \approx \omega$, so is either 0 or ω itself.

COROLLARY 3 (of Proposition 1). *Suppose that $\mathscr{F}$ is irreducible, that its determinant is of finite order, and that its rank n is a prime number. Then either $\mathscr{F}$ is induced from a character of a subgroup of π_1 of index n, or it is Lie-irreducible, or it factors through a finite quotient of π_1.*

LEMMA 4. *A Lie-irreducible $\mathscr{F}$ whose determinant is of finite order has $(G_{\mathrm{geom}})^0$ semisimple. If in addition* (1): *the determinant of $\mathscr{F}$ has finite order prime to p,* (2): *the rank n of $\mathscr{F}$ is prime to p,* (3): *the rank r of $(G_{\mathrm{geom}})^0$ is $< p$ (e.g., if $n < p$, or if $n < 2p$ and $(G_{\mathrm{geom}})^0$ lies inside* SO *or* Sp), *then the index of $(G_{\mathrm{geom}})^0$ in G_{geom} is prime to p.*

Proof. Once $\det(\rho)$ is of finte order, $(G_{\mathrm{geom}})^0$ lies in $\mathrm{SL}(n, \overline{\mathbf{Q}}_\ell)$, and as it acts irreducibly, it must be semisimple. Suppose now that in addition the conditions (1), (2), and (3) all hold. Consider the action of π_1 on $(G_{\mathrm{geom}})^0$ by conjugation. Let us denote by K the subgroup of π_1 which acts by inner automorphism. The quotient π_1/K injects into the automorphism group of the Dynkin diagram of $(G_{\mathrm{geom}})^0$, which is in turn a subgroup of the symmetric group on $r = \mathrm{rank}((G_{\mathrm{geom}})^0)$ letters. But as $(G_{\mathrm{geom}})^0$ has rank $r < p$, π_1/K is prime-to-p. Because $(G_{\mathrm{geom}})^0$ acts irreducibly, $\rho(K)$ lands in $\mathbf{G}_m \cdot (G_{\mathrm{geom}})^0$. Taking determinants, and using (1) and (2), we see that a subgroup K^0 of K of index prime to p lands in $(G_{\mathrm{geom}})^0$. The index of $(G_{\mathrm{geom}})^0$ in G_{geom} divides that of K^0 in π_1, so is prime to p. QED

To go further, we need to use the Feit-Thompson sharpening [Fe-Th] of Jordan's theorem. Jordan's theorem (cf. [Cu-Re], 36.13) is the fact that there exists a function $c(n)$ of n alone such that if Γ is a finite subgroup of $\mathrm{GL}(n, K)$, K any field of characteristic zero, then there exists a normal abelian subgroup Γ_0 of Γ such that the quotient group Γ/Γ_0 has order $\leqslant c(n)$. In particular, if p is a prime number which is $> c(n)$, then any p-Sylow subgroup Γ_p of Γ must lie in Γ_0, and it must be a p-Sylow subgroup of the abelian group Γ_0. Therefore Γ_p is both unique, hence normal in Γ, and it is abelian. The Feit-Thompson sharpening is the fact that if $p > 2n + 1$ (rather than $p > c(n)$), any p-Sylow subgroup of Γ is both normal and abelian.

PROPOSITION 5. *Suppose that U is the affine line $\mathbf{A}^1$, and that $p > 2n + 1$. Then any irreducible $\mathscr{F}$ of rank n is Lie-irreducible. If in addition* $\det(\mathscr{F})$ *is trivial, then G_{geom} is connected; in general,* $\det(\mathscr{F})$ *has finite p-power order, say order q, and G_{geom} is the product of $(G_{\text{geom}})^0$ with the finite subgroup μ_q of the scalars. The group $(G_{\text{geom}})^0$ is semisimple.*

Proof. The basic fact we must exploit is that for $\mathbf{A}^1$, the fundamental group π_1 has no nonzero continuous homomorphism to a finite group of order prime to p. Suppose first that ρ is induced from a proper open subgroup K of π_1. Then the index d of K is a divisor of n, and so $d \leqslant n$, whence $d < p$. But K contains an open subgroup K_0 which is normal in π_1 and of index dividing $d!$. As $d!$ is prime to p, we must have $K_0 = \pi_1$, contradiction. So this case cannot arise. Therefore (by Proposition 1) if ρ is not Lie-irreducible, it must be of the form $\tau \otimes \omega$, with ω an irreducible finite-image representation of π_1 of dimension $d \geqslant 2$, d some divisor of n. Let us denote by Γ the image of ω in $\mathrm{GL}(d, \overline{\mathbf{Q}}_\ell)$. Since $p > 2n + 1 \geqslant 2d + 1$, Feit-Thompson tells us that "the" p-Sylow subgroup Γ_p of Γ is both normal and abelian. But the quotient Γ/Γ_p is a prime-to-p quotient group of π_1, so it must be trivial. Therefore $\Gamma = \Gamma_p$ is itself abelian: as ω is irreducible, we must have $d = 1$, again a contradiction. Therefore ρ is Lie-irreducible.

Now consider the character $\chi = \det(\rho)$ of π_1. It actually takes values in the units $\mathbf{U}_\lambda$ of the integer ring of a finite extension E_λ of $\mathbf{Q}_\ell$, and this group is an extension of a finite group by a pro-ℓ group (the units in $1 + (2\ell)$). Because we are on $\mathbf{A}^1$, any pro-ℓ valued character of π_1 is trivial, so the vanishing of H^2 and the cohomology sequence shows that χ has finite order. Once the order of χ is finite, we write χ as a product of characters of finite prime-power order, of which only the p-power character can possibly be nontrivial on $\mathbf{A}^1$. Therefore $\det(\rho)$ is of finite p-power order.

Because $\det(\rho)$ has finite p-power order, say q, and n is prime to p, $\det(\rho)$ has a unique nth root ω. Twisting ρ by ω^{-1}, say $\Lambda = \rho \otimes \omega^{-1}$, we have $\rho = \omega \otimes \Lambda$, and $\det(\Lambda)$ is trivial. Let us denote by G (resp. G_1) the group G_{geom} for Λ (resp. ρ). Lemma 4 shows that G^0 is semisimple and G/G^0 is prime to p. As G/G^0 is also a quotient of π_1, and we are on $\mathbf{A}^1$, we have $G = G^0$. Because $\rho = \omega \otimes \Lambda$, G_1 lies inside the product $\mu_q \cdot G$. Because ω has finite order these two groups have the same identity component. So $G \subset G_1 \subset \mu_q \cdot G$, and taking determinants shows $G_1 = \mu_q \cdot G$. QED

We now consider the case when U is $\mathbf{G}_m$. For every integer $d \geqslant 1$ which is prime to p, we denote by $[d]$: $\mathbf{G}_m \to \mathbf{G}_m$ the dth power map, i.e., the d-fold Kummer covering of $\mathbf{G}_m$ by itself. We say that an $\mathscr{F}$ on $\mathbf{G}_m$ is Kummer-induced if it is of the form $[d]_*\mathscr{G}$ for some lisse $\mathscr{G}$ on $\mathbf{G}_m$, and some $d \geqslant 2$ prime to p.

PROPOSITION 6. *Suppose that U is $\mathbf{G}_m$, that $p > 2n + 1$, and that $\mathscr{F}$ is irreducible of rank n. Then $\mathscr{F}$ is either Kummer-induced or Lie-irreducible. If $\mathscr{F}$ is not Kummer-induced, and if its determinant is of finite order prime to p, then*

$(G_{\mathrm{geom}})^0$ is semisimple, and the group of components $G_{\mathrm{geom}}/(G_{\mathrm{geom}})^0$ is cyclic of finite order prime to p.

Proof. The fact we must exploit is that on $\mathbf{G}_m$, there is a unique prime-to-p quotient group of π_1 of every finite order $d \geqslant 1$ prime to p, namely the cyclic quotient corresponding to the d-fold Kummer covering. Suppose first that $\mathscr{F}$ is induced, from a proper subgroup K of π_1. Then the index d of K divides n, so $d < p$, and hence K contains an open subgroup K_0 which is normal in π_1 of index dividing $d!$, so prime to p. Therefore the quotient π_1/K_0 is cyclic, and hence K is itself normal in π_1 of index d prime to p, so by unicity K is the subgroup corresponding to the d-fold Kummer covering. Thus our $\mathscr{F}$ is Kummer-induced if it is induced.

Suppose now that $\mathscr{F}$ is neither induced nor Lie-irreducible. Then by Proposition 1 , the corresponding representation ρ must be of the form $\tau \otimes \omega$, with ω an irreducible finite-image representation of π_1 of dimension $d \geqslant 2$, d some divisor of n. Let us denote by Γ the image of ω in $\mathrm{GL}(d, \overline{\mathbf{Q}}_\ell)$. Since $p > 2n + 1 \geqslant 2d + 1$, Feit-Thompson tells us that "the" p-Sylow subgroup Γ_p of Γ is both normal and abelian. But the quotient Γ/Γ_p is a prime-to-p quotient group of π_1, so it is cyclic prime-to-p. Therefore Γ is an extension of a prime-to-p cyclic group by a finite abelian p-group, so by Schur-Zassenhaus Γ is the semidirect product of Γ_p with a cyclic prime-to-p group. But any irreducible representation ω of such a Γ is either a character or is induced from a character of a (necessarily normal) subgroup of index prime to p. (Proof by induction on the index of Γ_p in Γ: as Γ_p is normal in Γ, either ω is induced from an irreducible representation of a proper subgroup Γ^0 of Γ which contains Γ_p, and the induction hypothesis applies, or the restriction of ω to the normal abelian Γ_p is isotypical, so d copies of a character χ of Γ_p which is invariant by Γ-conjugation. Using the semi-direct product structure, we see that χ extends to a character, say χ, of Γ. Twisting ω by the inverse of χ, we reduce to the case when the irreducible representation ω of Γ factors through the abelian quotient Γ/Γ_p, which is possible only if $d = 1$.) Therefore as $d \geqslant 2$, ω itself is Kummer-induced, and hence by the projection formula so is $\tau \otimes \omega$, contradiction.

Suppose now that $\mathscr{F}$ is not Kummer-induced, i.e., that it is Lie-irreducible, and that its determinant is of finite order prime to p. Then by Lemma 4, $(G_{\mathrm{geom}})^0$ is semisimple and the group of components $G_{\mathrm{geom}}/(G_{\mathrm{geom}})^0$ is a prime-to-p quotient of π_1, necessarily cyclic because we are on $\mathbf{G}_m$. QED

The real problem in applications is recognizing whether or not a given $\mathscr{F}$ is Lie-irreducible in the first place; it is here that working on $\mathbf{A}^1$ or $\mathbf{G}_m$ seems to be essential. Overlooking this problem for the moment, we have the following general result. In it, we return to our open curve U. We denote by C its complete nonsingular model.

THEOREM 7. *Suppose that $\mathscr{F}$ is Lie-irreducible of rank n on U, that $p > 2n + 1$, that* $\det(\mathscr{F})$ *is of finite order prime to p, and that there exists a point ∞ in $C - U$*

at which all the breaks of $\mathscr{F}$ as I_∞-representation have exact denominator n. Then $(G_{\text{geom}})^0$ is one of the following subgroups of $\mathrm{GL}(n)$:

$\mathrm{SL}(n)$, *if* $n = 2$ *or if* n *is odd*
$\mathrm{SL}(n)$ *or* $\mathrm{Sp}(n)$ *or* $\mathrm{SO}(n)$ *if* n *is even and* $\geqslant 4$.

Moreover, in the case $(G_{\text{geom}})^0$ *is* $\mathrm{SO}(n)$, G_{geom} *is not contained in* $\mathbf{G}_m \cdot \mathrm{SO}(n)$.

Proof. We will show that Theorems 1 and 2 of [Ka-Pi] apply to the subgroup G of G_{geom} generated by $(G_{\text{geom}})^0$ and by $\rho(I_\infty)$, and to its given n-dimensional representation $\rho|G$. Let us denote by P_∞ the wild inertia subgroup of I_∞, and by γ an element of I_∞ which generates I_∞/P_∞. The fact that ρ has all its ∞-breaks with exact denominator n, with n prime to p and $\dim(\rho) = n$, insures that the restriction of ρ to P_∞ is the direct sum of n distinct characters χ_i of P_∞, and that the conjugation-induced action of γ cyclically permutes these n characters (cf. [Ka-1], 1.14).

Let us admit temporarily that $\rho(P_\infty)$ lies in a unique maximal torus T of $(G_{\text{geom}})^0$. Because T centralizes $\rho(P_\infty)$, T must respect each of the n one-dimensional χ_i-eigenspaces L_i of $\rho(P_\infty)$, and the action of T on each L_i is by a character $\mathbf{\chi}_i$, whose restrictions to $\rho(P_\infty)$ is χ_i. The element $B = \rho(\gamma)$ of G normalizes both $\rho(P_\infty)$ and $(G_{\text{geom}})^0$, so by the unicity of T, it must normalize T as well. Because B cyclically permutes the n one-dimensional χ_i-eigenspaces L_i of $\rho(P_\infty)$, it cyclically permutes the n distinct (because their restrictions to P_∞ are distinct) characters $\mathbf{\chi}_i$ of T. Because $\rho(P_\infty)$ lies in $(G_{\text{geom}})^0 = G^0$, the quotient G/G^0 is a finite tame quotient of I_∞, so is cyclic (and prime-to-p, but this is irrelevant for now). Finally, the given representation of G is not the tensor product of two strictly lower-dimensional representations of G, because the break hypothesis insures that already as a representation of $\rho(I_\infty)$ it is not such a tensor product (cf. [Ka-2], 2.5.9.3 for the completely analogous argument in the differential galois context). Thus all the hypotheses of Theorems 1 and 2 of [Ka-Pi] are verified, whence the asserted list of possibilities for $(G_{\text{geom}})^0$.

It remains to explain why $\rho(P_\infty)$ lies in a unique maximal torus T of $(G_{\text{geom}})^0$. It lies in $(G_{\text{geom}})^0$ because, by Lemma 4, the index of $(G_{\text{geom}})^0$ in G_{geom} is prime to p, while $\rho(P_\infty)$ is a finite abelian p-group. Because $\rho(P_\infty)$ acts with n distinct characters, its centralizer Z_{GL} in the ambient $\mathrm{GL}(n)$ is the entire group D of diagonal (with respect to a $\rho(P_\infty)$-eigenbasis) matrices in $\mathrm{GL}(n)$. Therefore its centralizer Z in $(G_{\text{geom}})^0$ lies in the torus D, and hence Z^0 is itself a torus. But any torus T of $(G_{\text{geom}})^0$ containing $\rho(P_\infty)$ trivially lies in Z^0, so that, once we prove that there exists a torus T of $(G_{\text{geom}})^0$ containing $\rho(P_\infty)$, we will conclude that Z^0 is the unique maximal one. This is a special case of the following general fact, applied to $\Gamma = \rho(P_\infty)$ and to $G = (G_{\text{geom}})^0$, for both the statement and proof of which I am indebted to J. Bernstein.

PROPOSITION 8. *Let G be a connected reductive group of rank r over an algebraically closed field of characteristic zero, and Γ a finite abelian subgroup of G whose order is divisible only by primes $p > r + 1$. Then Γ lies in a torus of G.*

Proof. The proof is by induction on $\dim(G)$. Write G as the product of its connected center by its derived subgroup; this gives a central extension

$$0 \to G^{\mathrm{der}} \cap Z(G)^0 \to G^{\mathrm{der}} \times Z(G)^0 \to G \to 0$$

whose kernel, being a central subgroup of a connected semisimple group G^{der} of rank $\leqslant r$, is (say by classification!) of order prime to that of Γ. Therefore Γ lifts uniquely to an abelian subgroup, still denoted Γ, of $G^{\mathrm{der}} \times Z(G)^0$, and this lifted Γ is a subgroup of the product of its projections $\mathrm{pr}_1(\Gamma) \times \mathrm{pr}_2(\Gamma)$. Thus it suffices to show that $\mathrm{pr}_1(\Gamma)$ lies in a torus of G^{der}. If G was not semisimple to begin with, this holds by induction.

If G is semisimple, let $G^{\#}$ denote the universal covering group of G, which is again a central extension of G with kernel of order prime to that of Γ. Thus Γ lifts uniquely to an abelian subgroup, still denoted Γ, of $G^{\#}$. Because the order of Γ is prime to that of $Z(G^{\#})$, either Γ is trivial and there is nothing to prove, or the centralizer $Z(\gamma)$ of any nontrivial γ in Γ is a proper closed subgroup. By a result of Steinberg (cf. [St], Corollary 8.5) the centralizer of a semisimple element in a simply connected semisimple group is itself a connected reductive group. Because Γ lies in this centralizer, we are through by induction. This concludes the proof of Proposition 8, and with it the proof of Theorem 7. QED

Combining Theorem 7 and Propositions 5 and 6, we obtain the following two results.

THEOREM 9. *Suppose that $p > 2n + 1$. Let $\mathscr{F}$ be irreducible of rank n on $\mathbf{A}^1$, and suppose that at ∞ all the breaks of $\mathscr{F}$ as I_∞-representation have exact denominator n. Then $(G_{\mathrm{geom}})^0$ is one of the following subgroups of* $\mathrm{GL}(n)$:

$\mathrm{SL}(n)$, *if n is odd*
$\mathrm{SL}(n)$ *or* $\mathrm{Sp}(n)$ *if n is even.*

Moreover, G_{geom} is the product of $(G_{\mathrm{geom}})^0$ with the finite p-power subgroup μ_q of the scalars which is the image of its determinant.

Proof. Just combine Proposition 5 with Theorem 7, and notice that on $\mathbf{A}^1$ the $\mathrm{SO}(n)$ case is ruled out by the "moreover" in Theorem 7. QED

THEOREM 10. *Suppose that $p > 2n + 1$. Let $\mathscr{F}$ be irreducible of rank n on $\mathbf{G}_m$, and not Kummer-induced. Suppose that at ∞ all the breaks of $\mathscr{F}$ as I_∞-representation have exact denominator n. Then $(G_{\mathrm{geom}})^0$ is one of the following subgroups of* $\mathrm{GL}(n)$:

$\mathrm{SL}(n)$, *if $n = 2$ or if n is odd*
$\mathrm{SL}(n)$ *or* $\mathrm{Sp}(n)$ *or* $\mathrm{SO}(n)$ *if n is even and $\geqslant 4$.*

Moreover, in the case $(G_{\mathrm{geom}})^0$ is $\mathrm{SO}(n)$, G_{geom} is not contained in $\mathbf{G}_m \cdot \mathrm{SO}(n)$.

Proof. This is just Proposition 6 combined with Theorem 7. QED

Part 2. Application to Kloosterman sheaves (cf. [Ka-1]). Let $\mathbf{F}_q$ be a finite subfield of k, Ψ a nontrivial $(\overline{\mathbf{Q}}_\ell)^\times$-valued additive character of $\mathbf{F}_q$, $n \geqslant 1$ an integer, and $\chi_1, \dots, \chi_n$ a set of n not necessarily distinct, possibly trivial $(\overline{\mathbf{Q}}_\ell)^\times$-valued multiplicative characters of $(\mathbf{F}_q)^\times$. We denote by $\mathrm{Kl}_\Psi(\chi_1, \dots, \chi_n)$, or simply by Kl, the corresponding Kloosterman sheaf on $\mathbf{G}_m$ (i.e., the sheaf denoted $\mathrm{Kl}_\Psi(\chi_1, \dots, \chi_n; 1, \dots, 1)$ with all $b_i = 1$ in [Ka-1], 4.1.1). One knows that Kl is lisse and irreducible of rank n on $\mathbf{G}_m$, that on $\mathbf{G}_m/\mathbf{F}_q$ it is pure of weight $n-1$, that all its breaks at ∞ are $1/n$, and that its local monodromy at zero is tame, a successive extension of the χ_i. One also knows that if $n \geqslant 2$, the determinant of Kl is the tame character $\Pi\chi_i$. We will sometimes refer to $(\chi_1, \dots, \chi_n)$ as the set of exponents of Kl at zero, to emphasize the analogy with the Kloosterman differential equation (cf. [Ka-Pi]). The results of this section are in perfect analogy with, and inspired by, the results of [Ka-Pi] on the differential galois groups of Kloosterman equations.

In discussing Kloosterman sheaves, it is useful to recall that by means of the Lang isogenies relative to the various finite subfields $\mathbf{F}_q$ of k, i.e., the Kummer coverings of degrees $q-1$, the prime-to-p completion of π_1 is the inverse limit, via the norm maps, of all the $(\mathbf{F}_q)^\times$. This allows us to view multiplicative characters of variable finite subfields of k, extended to larger ones by composition with the relative norm, as being precisely the characters of π_1 of finite, prime-to-p order.

LEMMA 11 (R. Pink). *If a Kloosterman sheaf is induced, then it is Kummer-induced.*

Proof. Let us denote by ρ the corresponding representation of π_1 on the n-dimensional $\overline{\mathbf{Q}}_\ell$-space $V = \mathrm{Kl}_{\bar{\eta}}$. Because ρ is irreducible, it is induced if and only if there exists a decomposition of V as the direct sum of $d \geqslant 2$ subspaces W_i which are transitively permuted among each other by the action of π_1, and in this case the representation is induced from the stabilizer K_i of any chosen W_i acting on W_i. Notice that the W_i all have the same dimension, say r, and that $rd = n$. The permutation representation of π_1 on the set of the W_i is a homomorphism β of π_1 to the symmetric group S_d which factors through ρ, hence is tame at zero and has all breaks $\leqslant 1/n$ at ∞. But S_d embeds into $\mathrm{GL}(d-1)$ by the augmentation representation, so the composite is a $d-1$ dimensional representation β of π_1 which is tame at zero and whose Swan conductor at ∞ is $\leqslant (1/n)(d-1) < 1$. Therefore β is tame at both zero and ∞, and hence its image in S_d must be a cyclic group of order prime to p. But its image is a transitive subgroup, so being cyclic it must be generated by a single d-cycle. Therefore all the W_i have the same stabilizer, namely $\ker(\beta)$, which is a normal subgroup of π_1 of index d. Because the image of β is prime-to-p, it follows that d is prime to p, whence ρ is induced from the d-fold Kummer covering. QED

Let us say that a collection of $n \geqslant 2$ multiplicative characters $(\chi_1, \dots, \chi_n)$ is Kummer-induced if there exists a prime-to-p divisor $d \geqslant 2$ of n, and a set of

$r = n/d$ multiplicative characters $\tau_1, \dots, \tau_r$ of some finite extension of $\mathbf{F}_q$ containing the dth roots of unity, such that the $n = d \times r$ characters χ_i (when composed with the relative norm), give exactly all (counted with multiplicities) the dth roots of the r characters τ_i.

LEMMA 12. *A Kloosterman sheaf* $\mathrm{Kl}_\Psi(\chi_1, \dots, \chi_n)$ *is induced if and only if the set* $(\chi_1, \dots, \chi_n)$ *of its exponents at zero is Kummer-induced.*

Proof. By Pink's lemma above, if Kl is induced it is Kummer-induced, say $[d]_*(\mathscr{G})$ for some lisse $\mathscr{G}$ on $\mathbf{G}_m$, with $d \geqslant 2$ a prime-to-p divisor of n. Therefore $\mathscr{G}$ is lisse of rank $r = n/d$, tame and quasiunipotent at zero, and all its breaks at ∞ are $1/r$. By ([Ka-1], 8.7.1), $\mathscr{G}$ is itself a Kloosterman sheaf $\mathrm{Kl}_\psi(\tau_1, \dots, \tau_r)$ for some ψ, possibly different from Ψ, and some τ_i. Applying the direct image formula of ([Ka-1], 5.6.2) to compute $[d]_*(\mathrm{Kl}_\psi(\tau_1, \dots, \tau_r))$, we find that $\mathrm{Kl}_\Psi(\chi_1, \dots, \chi_n)$ is a multiplicative translate of Kl_Ψ (all dth roots of the τ_i). Looking at its local monodromy at zero, we see that the χ_i are precisely all the dth roots of the τ_i (counted with multiplicities).

Conversely, if$(\chi_1, \dots, \chi_n)$ is Kummer-induced, then ([Ka-1], 5.6.2) expresses $\mathrm{Kl}_\Psi(\chi_1, \dots, \chi_n)$ as the Kummer-induction of a Kloosterman sheaf of rank r. QED

THEOREM 13. *Suppose that* $p > 2n + 1$, *that* $n \geqslant 2$, *and that* $\mathrm{Kl}_\Psi(\chi_1, \dots, \chi_n)$ *is not Kummer-induced. Then* $(G_{\mathrm{geom}})^0$ *is one of the following subgroups of* $\mathrm{GL}(n)$:

$\mathrm{SL}(n)$, *if* $n = 2$ *or if* n *is odd*
$\mathrm{SL}(n)$ *or* $\mathrm{Sp}(n)$ *or* $\mathrm{SO}(n)$ *if* n *is even and* $\geqslant 4$.

Moreover, in the case $(G_{\mathrm{geom}})^0$ *is* $\mathrm{SO}(n)$, G_{geom} *is not contained in* $\mathbf{G}_m \cdot \mathrm{SO}(n)$.

Proof. This is just a special case of Theorem 10, in view of the above discussion of Kloosterman sheaves. QED

Algorithm 14. Hypotheses as in Theorem 13 above, suppose that n is even and $n \geqslant 4$. Here is the algorithm to decide which of the three possibilities SL, Sp, or SO one has for $(G_{\mathrm{geom}})^0$. One looks for a multiplicative character τ (of some finite overfield of $\mathbf{F}_q$) such that the set of characters ω_i defined as $\omega_i = \chi_i/\tau$ is stable under inversion. Notice that if such a τ exists, then $\tau^{2n} = (\Pi\chi_i)^2$, so there are at most $2n$ candidates to examine. If no such τ exists, then we are in the SL case. If such a τ exists, then $(\Pi\omega_i)^2$ is trivial. If $\Pi\omega_i$ is itself trivial, then we are in the Sp case, and if not then we are in the SO case.

To see that this algorithm is correct, we argue as follows. Suppose first that such a τ exists. Replacing the χ_i by the ω_i amounts to twisting Kl by τ^{-1}, and has no effect on $(G_{\mathrm{geom}})^0$. This reduces us to the case where our Kl is self-dual (cf. [Ka-1], 4.1.4, 4.1.7). Because Kl is irreducible, the autoduality is unique up to a scalar, so is either alternating or symmetric. In the first case, G_{geom} lies in Sp; since Sp lies in SL, the determinant $\Pi\omega_i$ is trivial. In the second case, G_{geom} lies in the orthogonal group O, but by Theorem 13 it does not lie in SO, which shows that in this case the determinant $\Pi\omega_i$ is nontrivial.

Conversely, suppose that $(G_{\text{geom}})^0$ is either Sp or SO. Then G_{geom} lies in the normalizer, inside the ambient GL, of either Sp or SO, so G_{geom} itself lies in either $\mathbf{G}_m \cdot \text{Sp}$ or in $\mathbf{G}_m \cdot \text{O}$. In either case, forming the square of the G_m-factor is a well-defined character χ of finite order of π_1 which factors through ρ, so is tame both at zero (because ρ is), and at ∞ (because its break at ∞ is both an integer and is $\leqslant 1/n$, the largest break of ρ at ∞). Therefore χ is a prime-to-p character of π_1, and as such it has a prime-to-p square root, say Λ. Twisting our Kl by Λ^{-1}, we do not change $(G_{\text{geom}})^0$, and we turn our Kl into a self-dual one (i.e. into one whose G_{geom} is contained in either Sp or O) whose exponents at zero, the χ_1/Λ, are stable by inversion. Thus we may take τ to be Λ, and we are in the situation of the last paragraph. Hence the algorithm is correct in all cases.

Remark 15. The attentive reader may be disturbed by the fact that the above algorithm seems to rule out the SO case in characteristic $p = 2$. He may be even more disturbed by the apparent incompatibility of Theorem 13 with the fact, proven in ([Ka-1], 1.1.1) that for $p = 2$, n **odd** and all χ_i trivial, one has $(G_{\text{geom}})^0 = \text{SO}(n)$ for $n \neq 7$, and G_2 for $n = 7$. The resolution of this perplexity comes in recalling that, independently of our use of the Feit-Thompson sharpening of Jordan's theorem, which required $p > 2n + 1$ but which was not used at all in [Ka-1], we use the hypothesis $p > \text{rank}((G_{\text{geom}})^0)$ in Lemma 4 to prove that the index of $(G_{\text{geom}})^0$ in G_{geom} is prime to p, and the hypothesis that $p > 1 + \text{rank}((G_{\text{geom}})^0)$ in proving, via Proposition 8, that $\rho(P_\infty)$ lies in a torus T of $(G_{\text{geom}})^0$.

We now turn to the equidistribution consequences of Theorem 13. Recall from ([Ka-1], 4.1.1) that when we view $\text{Kl}_\Psi(\chi_1, \dots, \chi_n)$ on $\mathbf{G}_m/\mathbf{F}_q$, then for $\mathbf{E}$ any finite extension of $\mathbf{F}_q$, and t any $\mathbf{E}$-valued point of $\mathbf{G}_m$, the trace of the corresponding Frobenius $\text{Frob}_{t,\mathbf{E}}$ is the Kloosterman sum

$$(-1)^{n-1} \sum_{\substack{x_1x_2\dots x_n = t\\ \text{all } x_i \text{ in } \mathbf{E}}} \Psi(x_1 + \cdots + x_n)\chi_1(x_1)\chi_2(x_2)\dots\chi_n(x_n),$$

where ψ (resp. each χ_i) is extended to $\mathbf{E}$ via the relative trace (resp. norm).

COROLLARY 16 (Sato-Tate law for Kloosterman sums). *Hypotheses and notations as in Theorem* 13, *suppose that all the* χ_i *are characters of* $(\mathbf{F}_q)^\times$. *There exists a constant* α *in* $(\overline{\mathbf{Q}}_\ell)^\times$, *of the form*

$$\alpha = (\textit{a root of unity}) \times q^{-(n-1)/2},$$

such that after tensoring Kl *with the geometrically constant rank one sheaf on* $\mathbf{G}_m/\mathbf{F}_q$ *given by* $\alpha^{\deg}$, *the resulting lisse sheaf* $\text{Kl}(\alpha)$ *on* $\mathbf{G}_m/\mathbf{F}_q$ *has all of its Frobenii in* G_{geom}, *and is pure of weight zero. For any embedding of* $\overline{\mathbf{Q}}_\ell$ *into* $\mathbf{C}$, *and any choice of a maximal compact subgroup* K *of the Lie group of* $\mathbf{C}$*-valued points of* G_{geom}, *the conjugacy class of the semisimple part of each Frobenius* $\text{Frob}_{t,\mathbf{E}}$ *of*

Kl(α) *meets* K *in a single conjugacy class, denoted* $\theta(t, \mathbf{E})$, *of* K. *The conjugacy classes* $\theta(t, \mathbf{E})$ *are equidistributed in the space* $K^{\#}$ *of conjugacy classes of* K *with respect to normalized Haar measure, in any of the three senses of equidistribution of* ([Ka-1], 3.5).

Proof. We first show that there exists a constant α in $(\overline{\mathbf{Q}}_{\ell})^{\times}$ such that the Frobenii of Kl(α) all land in G_{geom}. This results from the explicit determination of G_{geom}, for in all cases considered its normalizer in the ambient GL(n) is equal to $\mathbf{G}_m \cdot G_{\text{geom}}$. The intersection of $\mathbf{G}_m$ with G_{geom} is a finite subgroup μ_N, so "formation of the Nth power of the $\mathbf{G}_m$-factor" is a well-defined character of the π_1 of $\mathbf{G}_m/\mathbf{F}_q$ which is geometrically constant, so of the form $\beta^{\deg}$ for some β in $(\overline{\mathbf{Q}}_{\ell})^{\times}$. We may take for α the inverse of any Nth root of β.

Because G_{geom} has a determinant of finite order, it follows that Kl(α) is pure of weight zero (since it is pure of some weight, and its determinant is of finite order). Because the determinant of Kl as a character of the arithmetic π_1 is of the form $(\Pi\chi_i)A^{\deg}$, where A is $\pm q^{n(n-1)/2}$ (cf. [Ka-1], 7.4.1.3), any α such that Kl(α) has a determinant of finite order is necessarily of the asserted form. The equidistribution now follows immediately from Deligne's Weil II, as explained at length in [Ka-1], Chapter 3). QED

Part 3. Elementary Fourier transforms on $\mathbf{A}^1$. We begin this section by explaining heuristically the classical forebearer of the finite-field situation we will study. Fix an integer $n \geqslant 2$, and a polynomial $f(x) = \Sigma a_i x^i$ in $\mathbf{C}[x]$ of degree n. We denote by ∂_x the derivation d/dx. The exponential $e^{f(x)}$ is annihilated by the first order operator $\partial_x - f'(x)$, so its Fourier transform

$$FT(e^f)(t) = \int e^{f(x)+tx}\,dx$$

is annihilated by the $n-1$st order operator $f'(\partial_t) + t$, i.e., by (minus) the formal Fourier transform of $\partial_x - f'(x)$. If $n \geqslant 3$, the Wronskian of $f'(\partial_t) + t$ is annihilated by the first order operator $na_n\partial_t + (n - I)a_{n-1}$, so for $n \geqslant 3$ it is equal to a constant times e^{-bt}, for $b = (n-1)a_{n-1}/na_n$.

In ([Ka-2], 4.2.7, 4.2.8, 4.2.10) we computed the differential galois group G_{gal} of the $n-1$st order operator $f'(\partial_t) + t$, by exploiting the fact that all of its slopes at ∞ are equal to $n/(n-1)$. The result is that for $n \geqslant 3$, G_{gal} is one of the following four subgroups of GL($n-1$):

SL($n-1$) or $\mathbf{G}_m \cdot$ SL($n-1$), if either $n-1$ is odd, or if $n-1$ is even and there exist no constants c, d such that the polynomial $f(x+c)+d$ is an odd function of x,

Sp($n-1$) or $\mathbf{G}_m \cdot$ Sp($n-1$), if $n-1$ is even and there exist constants c, d such that the polynomial $f(x+c)+d$ is an odd function of x.

The homotheties $\mathbf{G}_m$ are absent if and only if the coefficient a_{n-1} of $f(x) = \Sigma a_i x^i$ vanishes.

We now turn to the situation with which we shall be concerned for the remainder of this section. Let $\mathbf{F}$ be a finite subfield of k, Ψ a nontrivial $(\overline{\mathbf{Q}}_\ell)^\times$-valued additive character of $\mathbf{F}$, $n \geqslant 2$ an integer, and $f(x) = \Sigma a_i x^i$ a polynomial of degree n in $k[x]$. The lisse rank one sheaf $\mathscr{L}_{\Psi(f)} = f^*(\mathscr{L}_\Psi)$ on $\mathbf{A}^1$ is the analogue of e^f (cf. [Ka-1], 4.3, 8.2). We denote by $\mathscr{F}$ (or $\mathscr{F}_f$) the naive Fourier transform $\mathrm{NFT}_\Psi \mathscr{L}_{\Psi(f)}$ of $\mathscr{L}_{\Psi(f)}$ (cf. [Ka-1], 8.4 and 8.5). The ℓ-adic analogue of the fact that $\mathrm{FT}(e^f)(t) = \int e^{f(x)+tx}\,dx$ satisfies a linear differential equation of order $n-1$ with all ∞-slopes $n/(n-1)$, whose Wronskian is e^{-bt} for $n \geqslant 3$, is the following theorem.

THEOREM 17. *Notations as above, suppose in addition that $n \geqslant 2$ is prime to p. Then*

(1) *$\mathscr{F}$ is lisse on $\mathbf{A}^1$ of rank $n-1$, and all of its breaks at ∞ are equal to $n/(n-1)$.*

(2) *If $f(x)$ has coefficients in $\mathbf{F}$, then $\mathscr{F}$ lives naturally on $\mathbf{A}^1$ over $\mathbf{F}$, and there it is pure of weight one. For any finite overfield $\mathbf{E}$ of $\mathbf{F}$, and any t in $\mathbf{E} = \mathbf{A}^1(\mathbf{E})$, the trace of Frobenius at t is given by*

$$\mathrm{trace}(\mathrm{Frob}_{t,\mathbf{E}}|\mathscr{F}_t) = -\sum_{x\in\mathbf{E}} \Psi_{\mathbf{E}}(f(x) + tx),$$

where $\Psi_{\mathbf{E}}$ denotes the additive character $\Psi(\mathrm{Trace}_{\mathbf{E}/\mathbf{F}}(-))$ of $\mathbf{E}$.

(3) *If $n \geqslant 3$, $\det(\mathscr{F}) \otimes \mathscr{L}_{\Psi(bt)}$ is geometrically constant for $b = (n-1)a_{n-1}/na_n$.*

Proof. (1) Because n is prime to p, the (unique) break of $\mathscr{L}_{\Psi(f)}$ at ∞ is n. Because $n \geqslant 2$, and $\mathscr{L}_{\Psi(f)}$ is lisse of rank one, it follows that $\mathscr{L}_{\Psi(f)}$ is an irreducible Fourier sheaf (by [K-1], 8.3.1(1), 8.4). Therefore (cf. [Ka-1], 8.2.5, 8.4.1, and 8.5.8(1)) $\mathscr{F}$ is an irreducible Fourier sheaf, which is lisse on $\mathbf{A}^1$, and all of its ∞-breaks are > 1 (because its NFT is the lisse sheaf $\mathscr{L}_{\Psi(f)}$). From the Euler-Poincare formula ([Ka-1], 8.5.3) one sees that if two lisse sheaves on $\mathbf{A}^1$ are inverse NFT's of each other, then the sum of their ranks is the Swan conductor of each at ∞; this shows that $\mathscr{F}$ has rank $n-1$, and Swan conductor n at ∞. Since all the $n-1$ ∞-breaks of $\mathscr{F}$ are > 1, and their sum is n, they must each be equal to $n/(n-1)$. [For if we write each as $1 + \lambda_i$, then $\Sigma\lambda_i = 1$, each λ_i is > 0 and rational, and the multiplicity of each of the distinct numbers which occur among the λ_i is multiple of its exact denominator, so already the partial sum of only those λ_i equal to a given one contributes at least 1 to the sum, whence all the λ_i must be equal.] This concludes the proof of (1).

Suppose now that f has coefficients in $\mathbf{F}$. Then (2) follows from the basic facts about Fourier transform and the Lefschetz trace formula (cf. [Ka-1], 2.3, 8.2.5(3), 8.2.4).

Suppose now that $n \geqslant 3$. We must prove that $\det(\mathscr{F}) \otimes \mathscr{L}_{\Psi(bt)}$ is geometrically constant. Consider the universal family of monic polynomials of degree n over an $\mathbf{F}$-scheme. The parameter space S of this family is the $\mathbf{G}_m \times \mathbf{A}^n$ over $\mathbf{F}$

with coordinates the coefficients $A_n, \dots, A_0$ of $f_{\text{univ}}(x) = \Sigma A_i x^i$. From Deligne's semicontinuity theorem (cf. [La-1]) one sees that over S we have a lisse $\mathscr{F}_{\text{univ}}$ which is lisse of rank $n-1$ and whose trace of Frobenius at the $\mathbf{E}$-valued point corresponding to a given polynomial $f(x)$ with coefficients in $\mathbf{E}$ is the exponential sum $-\Sigma\Psi_{\mathbf{E}}(f(x))$ over x in $\mathbf{E}$. Our $\mathscr{F}$ is just the pullback to $\mathbf{A}^1$ of $\mathscr{F}_{\text{univ}}$ by the map

$$\varphi\colon \mathbf{A}^1 \to S, \qquad t \to (a_n, \dots, a_2, t + a_1, a_0).$$

The idea is to compute the traces of Frobenius on $\det(\mathscr{F}_{\text{univ}})$ by the following trick of Hasse-Davenport (cf. [Ha-Da], Section 3, 11, pages 162–165). Given an $\mathbf{E}$-valued point of S, corresponding to an f, consider the L-function of $\mathbf{A}^1$ over $\mathbf{E}$ with coefficients in $\mathscr{L}_{\Psi(f)}$. The cohomological expression of this L-function is

$$L(T) = \det\bigl(1 - \mathrm{TFrob}_{\mathbf{E}} | H^1\bigl(\mathbf{A}^1, \mathscr{L}_{\Psi(f)}\bigr)\bigr),$$

which exhibits it as a polynomial of degree $n-1$ in T, and shows that

$$\det\bigl(\mathrm{Frob}_{f,\mathbf{E}} | \mathscr{F}_{\text{univ}}\bigr) = (-1)^{n-1} \times \bigl(\text{the coef. of } T^{n-1} \text{ in } L(T)\bigr).$$

The naive additive expression of this same L-function as a sum over all effective divisors of $\mathbf{A}^1$ over $\mathbf{E}$ (i.e., over all monic polynomials in $\mathbf{E}[x]$) shows that for any $d \geqslant 1$, the coefficient of T^d in $L(T)$ is

$$\sum_{\text{monic } h \text{ of deg } d \text{ over } \mathbf{E}} \Psi_E\Bigl(\sum_{\text{zeroes } \alpha \text{ of } h, \text{ with mult.}} f(\alpha)\Bigr).$$

Notice that the innermost sum is actually $\mathbf{E}$-rational. Indeed, if we denote by $N_i(h)$ the Newton symmetric functions of the roots of h, then the innermost sum is just $\Sigma a_i N_i(h)$, where $f(x) = \Sigma a_i x^i$.

Now we specialize to the case $d = n-1$. The monic h's of degree $n-1$ over $\mathbf{E}$ are given by their coefficients $s_1, \dots, s_{n-1}$, i.e., by the elementary symmetric functions of their roots. The N_i are universal polynomials in the s_j, so we obtain the formula

$$\mathrm{Frob}_{f,\mathbf{E}} | \det(\mathscr{F}_{\text{univ}}) = (-1)^{n-1} \sum \Psi_{\mathbf{E}}\bigl(\sum a_i N_i(s_1, \dots, s_{n-1})\bigr),$$

the outer sum over all $(s_1, \dots, s_{n-1})$ in $\mathbf{E}^{n-1}$. One easily sees by isobaricity that among $N_0, \dots, N_n$, only the last two N_{n-1} and N_n can involve s_{n-1}, and one readily calculates that, for $n-1 > 1$, the involvement is

$$N_{n-1}(s_1, \dots, s_{n-1}) = (-1)^n (n-1) s_{n-1} + \text{a poly. in } s_1, \dots, s_{n-2},$$

$$N_n(s_1, \dots, s_{n-1}) = (-1)^n (n) s_1 s_{n-1} + \text{a poly. in } s_1, \dots, s_{n-2}.$$

Substituting, we find that

$$(-1)^{n-1}\mathrm{Frob}_{f,\mathbf{E}}|\det(\mathscr{F}_{\mathrm{univ}}) = \sum \Psi_{\mathbf{E}}\Big((-1)^n s_{n-1}[na_n s_1 + (n-1)a_{n-1}] + a_1 s_1 + (n-1)a_0 + \sum_{i\geqslant 2} a_i P_i(s_1,\dots,s_{n-2})\Big).$$

Summing last over the variable s_{n-1}, we see that only the terms with

$$s_1 = -(n-1)a_{n-1}/na_n$$

survive, and that

$$(-1)^{n-1}\mathrm{Frob}_{f,\mathbf{E}}|\det(\mathscr{F}_{\mathrm{univ}}) = q\Psi_{\mathbf{E}}(-a_1(n-1)a_{n-1}/na_n) \times \sum \Psi_{\mathbf{E}}\Big((n-1)a_0 + \sum_{i\geqslant 2} a_i P_i(s_1,\dots,s_{n-2})\Big),$$

the outermost sum over all $(s_1,\dots,s_{n-2})$ in $\mathbf{E}^{n-2}$ for which

$$s_1 = -(n-1)a_{n-1}/na_n.$$

We now exploit the fact that the second factor above is independent of a_1. To do this, let us denote by Z the closed subscheme of S defined by the condition $A_1 = 0$, by i: $Z \to S$ the inclusion, and by π: $S \to Z$ the linear projection $(A_n,\dots,A_1,A_0) \to (A_n,\dots,A_2,0,A_0)$. The independence of a_1 of the second factor above means that, by Chebataroff, we have an isomorphism of lisse rank one sheaves on S

$$\det(\mathscr{F}_{\mathrm{univ}}) \otimes \mathscr{L}_{\Psi}(A_1(n-1)A_{n-1}/nA_n) \approx \pi^* i^*(\det(\mathscr{F}_{\mathrm{univ}})).$$

In the Fourier transform situation we are studying, our $\det(\mathscr{F})$ is obtained from $\det(\mathscr{F}_{\mathrm{univ}})$ by pulling back by the map

$$\varphi\colon \mathbf{A}^1 \to S, \qquad t \to (a_n,\dots,a_2,t+a_1,a_0).$$

The composite map $i\pi\varphi$ is the map $t \to (a_n,\dots,a_2,0,a_0)$, which is geometrically constant. Therefore $\det(\mathscr{F}) \otimes \mathscr{L}_{\Psi(b(t+a_1))}$ is geometrically constant, and hence so is $\det(\mathscr{F}) \otimes \mathscr{L}_{\Psi(bt)}$ itself. QED

COROLLARY 18. *Hypotheses and notations as in Theorem* 17 *above, suppose that* $n-1 \equiv 0 \bmod p$. *Then* G_{geom} *lies in* $\mathrm{SL}(n)$.

Proof. Indeed, we have $b = 0$. QED

THEOREM 19. *Notations as in Theorem* 17 *above, suppose further that* $n \geqslant 3$ *and that* $p > 2n + 1$. *Then* G_{Geom} *is one of the following subgroups of* $\mathrm{GL}(n-1)$:

$\mathrm{SL}(n-1)$ *or* $\mu_p \cdot \mathrm{SL}(n-1)$, *if either* $n-1$ *is odd, or if* $n-1$ *is even and there exist no constants* c, d *in* k *such that the polynomial* $f(x+c)+d$ *is an odd function of* x,

$\mathrm{Sp}(n-1)$ *or* $\mu_p \cdot \mathrm{Sp}(n-1)$, *if* $n-1$ *is even and there exist constants* c, d *such that the polynomial* $f(x+c)+d$ *is an odd function of* x.

The homotheties μ_p *are absent if and only if the coefficient* a_{n-1} *of* $f(x) = \Sigma a_i x^i$ *vanishes.*

Proof. This is almost entirely a formal consequence of Theorems 9 and 17(1). The only point that needs to be explained is how, when $n-1$ is even, one distinguishes the SL case from the Sp case. By an additive translation in the x variable, we may suppose that $a_{n-1} = 0$, and hence that G_{geom} is equal to either $\mathrm{SL}(n-1)$ or to $\mathrm{Sp}(n-1)$. The keypoint is that up to an additive inversion, formation of the Verdier dual D commutes with Fourier transform (cf. [La-2], 1.3.2.2):

$$D(\mathrm{FT}_\Psi(K)) = \mathrm{FT}_\Psi(D([x \to -x]^*(K))).$$

In our down-to-earth situation, taking for K the single sheaf $\mathscr{L}_{\Psi(f)}$, we have

$$D([x \to -x]^*(\mathscr{L}_{\Psi(f)})) = \mathscr{L}_{\Psi(g)},$$

where $g(x) = -f(-x)$. The above duality formula then says that $D(\mathscr{F}_f) = \mathscr{F}_g$. By Fourier inversion, then, $\mathscr{F}_f$ is self-dual if and only if $\mathscr{L}_{\Psi(f)}$ is isomorphic to $\mathscr{L}_{\Psi(g)}$, i.e., if and only if the sheaf $\mathscr{L}_{\Psi(f)} \otimes (\mathscr{L}_{\Psi(g)})^{-1} = \mathscr{L}_{\Psi(f-g)}$ is constant. This is trivially the case if $f-g$ is constant. If $f-g$ is not a constant, then as $f-g$ has degree $< p$, the Swan conductor of $\mathscr{L}_{\Psi(f-g)}$ at ∞ is $\deg(f-g)$. Therefore, $\mathscr{F}_f$ is self-dual if and only if $f-g$ is a constant. QED

COROLLARY 20 (Sato-Tate law for one-variable polynomial sums). *Hypotheses and notations as in Theorem* 17 *above, suppose further that the polynomial* f *has coefficients in* $\mathscr{F}_q$. *Then there exists a constant* α *in* $(\overline{\mathbf{Q}}_\ell)^\times$, *of the form*

$$\alpha = (\textit{a root of unity}) \times q^{-1/2},$$

such that after tensoring $\mathscr{F}$ *with the geometrically constant rank one sheaf on* $\mathbf{A}^1/\mathbf{F}_q$ *given by* $\alpha^{\deg}$, *the resulting lisse sheaf* $\mathscr{F}(\alpha)$ *on* $\mathbf{A}^1/\mathbf{F}_q$ *has all of its Frobenii in* G_{geom}, *and is pure of weight zero. For any embedding of* $\overline{\mathbf{Q}}_\ell$ *into* $\mathbf{C}$, *and any choice of a maximal compact subgroup* K *of the Lie group of* $\mathbf{C}$*-valued points of* G_{geom}, *the conjugacy class of the semisimple part of each Frobenius* $\mathrm{Frob}_{t,\mathbf{E}}$ *of* $\mathscr{F}(\alpha)$ *meets* K

in a single conjugacy class, denoted $\theta(t, \mathbf{E})$, of K. The conjugacy classes $\theta(t, \mathbf{E})$ are equidistributed in the space $K^{\#}$ of conjugacy classes of K with respect to normalized Haar measure, in any of the three senses of equidistribution of ([Ka-1], 3.5).

Proof. The proof of Corollary 16 works mutatis mutandis except for the evaluation of α up to roots of unity. To evaluate α, it suffices to show that in the universal case, when we write $\det(\mathscr{F}_{\text{univ}})$ in the form (char. of finite order) × $A^{\deg}$, for some A (this is always possible, cf. [De-1], 1.3.4(1)), we have $A =$ (a root of unity) × $q^{(n-1)/2}$. To show this, we first may make a finite extension of $\mathbf{F}_q$, and reduce to the case when all the nth roots of unity lie in $\mathbf{F}_q$. We may take $A = \det(\text{Frob} \mid H_c^1(\mathbf{A}^1, \mathscr{L}_{\Psi(f)}))$ for a **single** choice of $f(x)$ of degree n over $\mathbf{F}_q$. We take $f(x) = x^n$; an elementary calculation then gives $A = \Pi(-g(\Psi, \chi))$, the product over all the nontrivial characters χ of order dividing n. QED

Remark 21. In the case $n = 3$, $f(x) = x^3$, the question of how the sums $\Sigma_x \Psi(f(x) + tx)$ are distributed was first raised by Birch's note [Bi], and first worked out by Livne ([Li]), whose work was the starting point of my own interest in such sums.

References

[Bi] B. Birch, *How the number of points of an elliptic curve over a fixed prime field varies*, J. London Math. Soc. **43** (1968), 57–60.

[Cu-Re] C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras*, Interscience Publ., New York and London, 1962.

[De] P. Deligne, *La conjecture de Weil* II, Pub. Math. I.H.E.S. **52** (1981), 137–252.

[Fe-Th] W. Feit and J. Thompson, *On groups which have a faithful representation of degree less than* $(p-1)/2$, Pacific J. Math. **4** (1961), 1257–1262.

[Ha-Da] H. Hasse and H. Davenport, *Die Nullstellen der Kongruenzzetafunktionen in gewissen zyklischen Fällen*, J. Reine Angew. Math. **172**, (1934), 151–182.

[Ka-1] N. Katz, *Gauss Sums, Kloosterman Sums, and Monodromy Groups*, Annals of Math. Study **113**, to appear.

[Ka-2] ______, *On the calculation of some differential galois groups*,, Inv. Math. **87** (1987), 13–61.

[Ka-Pi] ______ and R. Pink, *A note on pseudo-CM representations and differential galois groups*, Duke Mathematical J., this issue.

[La-1] G. Laumon, Semicontinuité du conducteur de Swan (d'apres Deligne), in *Characteristique d'Euler-Poincaré*, Seminaire de I'ENS 1978/79, Asterisque 82-83, 1981, 173–219.

[La-2] ______, *Transformation de Fourier, constantes d'équations fonctionnelles et conjectures de Weil*, Pub. Math. I.H.E.S., to appear.

[Li] R. Livne, *Applications of Hodge theory to Birch's conjectures concerning the average distribution of cubic exponential sums*, to appear.

[St] R. Steinberg, *Endomorphisms of Linear Algebraic Groups*. Memoirs of the Amer. Math. Soc. **80**, 1968.

Department of Mathematics, Princeton University, Princeton, New Jersey 08544