ON A QUESTION OF LILLIAN PIERCE

NICHOLAS M. KATZ

ABSTRACT. We establish estimates for certain families of character sums over finite fields, including those which arise in Lillian Pierce's recent work [P] on estimating the 3-part of the class number of quadratic fields.

1. INTRODUCTION, THE BASIC SETTING, AND STATEMENT OF THE MAIN RESULT

Let k be a finite field of characteristic p and cardinality q, ψ a nontrivial $\mathbb{C}^\times$-valued additive character of k, and χ a nontrivial $\mathbb{C}^\times$-valued multiplicative character of $k^\times$. We define $n :=$ the order of χ. We extend χ to a function on all of k by defining $\chi(0) := 0$. Let $f(x)$ and $g(x)$ in $k[x]$ be two polynomials over k in one variable x, each of strictly positive degree. We define

$$d := deg(f), e := deg(g).$$

For each element a in k, we define a $\mathbb{C}$-valued function G_a on k by

$$G_a(z) := \sum_{x \in k} \chi(-f(x) + g(z))\psi(ax).$$

To provide some context, recall the following lemma, which results in a by now well known way from the truth, due to Weil [We-CA], of the Riemann Hypothesis for curves over finite fields, and its application, already known to Davenport and Hasse [Dav-Ha], to the estimation of abelian character sums.

Lemma 1.1. *Suppose that either χ^d is nontrivial, or that a is nonzero. Then we have the estimate*

$$|G_a(z)| \leq dq^{1/2}.$$

We are interested in the L^2 norm of the function G_a, and in the inner product of G_a both with additive translates of itself and with additive character "twists" of additive translates of itself. Denote by $\overline{G_a}$ the

Date: April 15, 2004.

complex conjugate of the function G_a. For $a, b, c \in k$, we define the complex character sum

$$I(a,b,c) := \sum_{z \in k} G_a(z)\overline{G_a}(z+c)\psi(bz).$$

In view of the lemma above, we have the following "trivial estimate" for the character sum $I(a,b,c)$.

Lemma 1.2. *Suppose that either χ^d is nontrivial, or that a is nonzero. Then we have the estimate*

$$|I(a,b,c)| \leq d^2 q^2.$$

Our main result gives conditions under which we can improve on the above "trivial estimate", and obtain the improved estimate

$$|I(a,b,c)| \leq 2de(d-1)q^{3/2}.$$

Our method requires that the characteristic p be large compared to d and to e. Fortunately, this is not a problem in Pierce's applications, where d and e are fixed, and p varies. [Indeed, in her applications, χ is the quadratic character, $f(x)$ is $-4x^3$, and $g(z)$ is δz^2 for some nonzero $\delta \in k$.]

Theorem 1.3. *We have the estimate*

$$|I(a,b,c)| \leq 2de(d-1)q^{3/2}.$$

in each of the following five situations:

(1) $ab \neq 0$, *and* $e < d < p$,
(2) $ac \neq 0$, $b = 0$, $e < d < p$, $gcd(e,d) = 1$, *and* $e(d-1) < p$,
(3) $a = 0$, $b \neq 0$, $d < p$ *and* χ^d *is nontrivial,*
(4) $a = b = 0$, $c \neq 0$, *the polynomials* $g(z)$ *and* $g(z+c)$ *in* $k[z]$ *are relatively prime,* $f(x) = \alpha x^d$ *for some nonzero* $\alpha \in k$, n *and* d *are relatively prime, and either*
 (4a) $e < n$ *or*
 (4b) d *is prime and* $e < nd$.
(5) $a = b = 0$, $c \neq 0$, $e < d < p$, $d \geq 3$, $e(d-1) < p$, χ^d *is nontrivial, the the polynomial* $f(x)$ *is "supermorse" (i.e., its derivative* $f'(x)$ *has* $d-1$ *distinct zeroes in* $\overline{k}$, *which map by* f *to* $d-1$ *distinct points), and either*
 (5a) χ *is the quadratic character* χ_2, *or*
 (5b) $d > 5$, *or*
 (5c) $d > 3$ *and* $(\chi\chi_2)^3$ *is nontrivial.*

2. Proof of the main result: basic ideas

Recall that n is the order of χ. Thus the function G_a takes values in the cyclotomic field $K := \mathbb{Q}(\zeta_n, \zeta_p)$, and the character sum $I(a,b,c)$ lies in K. We choose a prime number $\ell \neq p$, and a field embedding of K into $\overline{\mathbb{Q}}_\ell$. We introduce the K-valued function H_a on k defined by

$$H_a(t) := \sum_{x \in k} \chi(t - f(x))\psi(ax).$$

Thus we have the tautological relation

$$G_a(z) = H_a(g(z)).$$

We first interpret the function H_a as the trace function of an ℓ-adic sheaf $\mathcal{H}_a$ on the affine line $\mathbb{A}^1$ over k. The sheaf $\mathcal{H}_a$ turns out to be a "middle additive convolution" in the sense of [Ka-RLS, Chapter 2]. The function G_a is then the trace function of the pullback sheaf

$$\mathcal{G}_a := g^\star \mathcal{H}_a.$$

We then analyze the possible interactions of the sheaf $\mathcal{G}_a$ with various twisted additive translates of itself. We then use Deligne's fundamental results [De-Weil II, 3.3] to obtain the asserted estimates.

3. Proof of the main result in cases (1) and (2)

What cases (1) and (2) have in common is that $a \neq 0$. At the expense of replacing the originally chosen nontrivial additive character ψ by the the equally nontrivial additive character

$$\psi_a(x) := \psi(ax),$$

and replacing b by b/a, we reduce to the case $a = 1$. We denote by $\mathcal{L}_\psi$ the Artin-Schreier sheaf on $\mathbb{A}^1$ over k corresponding to ψ. We introduce the sheaf $\mathcal{F}$ on the w-line $\mathbb{A}^1$ over k defined by

$$\mathcal{F} := f_\star \mathcal{L}_\psi.$$

We denote by $\mathcal{L}_\chi$ the Kummer sheaf on $\mathbb{G}_m$ over k corresponding to χ, extended by zero across the origin. Write the function H_1 on k as

$$H_1(t) := \sum_{x \in k} \chi(t - f(x))\psi(x)$$

$$= \sum_{w \in k} \chi(t - w) \sum_{x \in k, f(x) = w} \psi(x)$$

$$= \sum_{w \in k} \chi(t - w) Trace(Frob_{k,w} | \mathcal{F})$$

$$= \sum_{w \in k} Trace(Frob_{k,w} | \mathcal{L}_{\chi(t-w)} \otimes \mathcal{F}).$$

This means precisely that the function H_1 on k is the trace function, at k-rational points, of the lower ! additive convolution of the sheaves $\mathcal{L}_\chi$ and $\mathcal{F}$, or what is the same, of the lower ! additive convolution, say M,

$$M := \mathcal{L}_\chi[1] \star_{+,!} \mathcal{F}[1]$$

of the perverse sheaves $\mathcal{L}_\chi[1]$ and $\mathcal{F}[1]$ on $\mathbb{A}^1$ over k. Both of these perverse sheaves are visibly middle extensions (remember χ is nontrivial) which, as perverse sheaves, are pure of weight one. Both are geometrically irreducible: this is obvious for $\mathcal{L}_\chi[1]$, and it holds for $\mathcal{F}[1]$ because $\mathcal{F}$ has generic rank d, and all its ∞-slopes are $1/d$, so is already irreducible under the inertia group $I(\infty)$, cf. [Ka-GKM, 1.14], and has Swan conductor 1 at ∞. As $d > 1$, this irreducibility already guarantees that $\mathcal{F}[1]$ has "condition $\mathcal{P}$" in the sense of [Ka-RLS, 2.6.2], thanks to [Ka-RLS, 2.6.13-15]. Therefore, by [Ka-RLS, 2.9.7], the additive ! convolution

$$M := \mathcal{L}_\chi[1] \star_{!,+} \mathcal{F}[1]$$

is itself a perverse sheaf on $\mathbb{A}^1$ over k. Looking fibre by fibre, we see that M is of the form $\mathcal{H}[1]$, with $\mathcal{H}$ a single sheaf, which is, on a dense open set, both lisse and pure of weight one. Thus M is, on a dense open set, pure of weight 2 as a perverse sheaf. We next claim that M is in fact the "middle additive convolution" of $\mathcal{L}_\chi[1]$ and $\mathcal{F}[1]$. Denote by $M_\star$ the additive $\star$ convolution

$$M_\star := \mathcal{L}_\chi[1] \star_{+,\star} \mathcal{F}[1].$$

then we know from [Ka-RLS, 2.10.10] and [Ka-MMP, 6.5.4, 3)] that the kernel Ker of the canonical "forget supports" map from M to $M_\star$ is a perverse sheaf of the form (a lisse sheaf on $\mathbb{A}^1$)[1] which is mixed of weight ≤ 1 and which is geometrically a successive extension of various $\mathcal{L}_{\psi_a}[1]$ sheaves. Since M is generically pure of weight 2, while Ker is everywhere lisse and of strictly lower weight, it follows that $Ker = 0$, and hence that M is in fact the middle additive convolution

$$M := \mathcal{L}_\chi[1] \star_{+,mid} \mathcal{F}[1].$$

How can we exploit this? We know, from [Ka-RLS, 2.9.7. 2)] and[Ka-MMP, 6.5.4, 2)] respectively, that $M = \mathcal{H}[1]$ is geometrically irreducible as a perverse sheaf, and is pure of weight 2. This means in turn that $\mathcal{H}$ is mixed of weight at most 1, and that it is the middle extension from some open dense set of a geometrically irreducible lisse sheaf which is pure of weight 1. Moreover, it is an exercise, using [Ka-RLS, 3.3.5], to compute both the generic rank of $M = \mathcal{H}[1]$, and all its local monodromies, given the same data for $\mathcal{F}$. Recall that $\mathcal{F}$ is a middle extension of generic rank d, all of whose ∞-slopes are $1/d$. Because $d := deg(f) < p$, and $\mathcal{L}_\psi$ is lisse on $\mathbb{A}^1$, $\mathcal{F}$ is tamely ramified outside the point at ∞, and the sum of its drops over all $\overline{k}$-valued points of $\mathbb{A}^1$ is just $d-1$, the total number of zeroes, counting multiplicities, of the derivative $f'(x)$ of f at all $\overline{k}$- valued points of $\mathbb{A}^1$. It is then routine to conclude, using [Ka-RLS, 3.3.5], that $\mathcal{H}$ has generic rank d, is tame outside ∞, and has all ∞-slopes $1/d$. Moreover, the drops of $\mathcal{H}$ at the $\overline{k}$-valued points of $\mathbb{A}^1$ occur precisely at the points where $\mathcal{F}$ had drops, and the drops at each such point are the same for $\mathcal{H}$ and for $\mathcal{F}$. These drop points are the critical values of the polynomial f, i.e. the values $f(\alpha)$ at the zeroes α of the derivative $f'(x)$. At each critical value $f(\alpha)$ of f, the drop at $f(\alpha)$ is the sum, over all zeroes γ in $\overline{k}$ of the polynomial $f(x)-f(\alpha)$, of the expression (multiplicity of γ as a zero of $f(x)-f(\alpha)$) - 1. In particular, the sum of all the drops of $\mathcal{H}$ at finite distance is just $d-1$, the total number, counting multiplicity, of zeroes of the derivative $f'(x)$.

With this information established about $\mathcal{H}$, we now turn to the sheaf $\mathcal{G} := \mathcal{G}_1 := g^\star\mathcal{H}$ on $\mathbb{A}^1$. It is mixed of weight at most 1, and on a dense open set it is lisse and pure of weight 1. It has generic rank d, it is tame at all $\overline{k}$-valued points of $\mathbb{A}^1$, and all of its ∞-slopes are e/d. Its only drops at $\overline{k}$-valued points of $\mathbb{A}^1$ are at the inverse images by g of the points at which $\mathcal{H}$ drops, and at such an inverse image point, the size of the drop is unchanged. Since g is a polynomial of degree e, there are at most e(d-1) finite points at which $\mathcal{G}$ drops, and the sums of all its drops at finite points is at most e(d-1). Moreover, the sheaf $\mathcal{G}$ has no nonzero punctual sections (as this was already true of $\mathcal{H}$, being a middle extension sheaf).

Let us now treat cases (1) and (2). Since $e < d$, $\mathcal{G}$ has all its ∞-slopes $e/d < 1$. Similarly, $\overline{\mathcal{G}}$, the sheaf formed using the characters $\overline{\chi}$ and $\overline{\psi}$, but the same f and g, has the same ramification properties as $\mathcal{G}$. The effect of any additive translation

$$T_c : z \to z + c$$

on $\overline{\mathcal{G}}$ is just to translate the finite singularities, keeping the same drops. The ∞-slopes remain $e/d < 1$. After our reduction to the case $a = 1$ explained above, we must estimate the quantity $I(1, b, c)$ This quantity is the sum of the traces of Frobenius at all the k-rational points of $\mathbb{A}^1$ of the sheaf

$$\mathcal{E}_b := \mathcal{G} \otimes T_c^\star\overline{\mathcal{G}} \otimes \mathcal{L}_{\psi_b}.$$

Notice that

$$\mathcal{E}_b = \mathcal{E}_0 \otimes \mathcal{L}_{\psi_b}.$$

This sheaf $\mathcal{E}_0$ is mixed of weight at most 2, it has no nonzero punctual sections, it is tame over $\mathbb{A}^1$, and the sum of its finite drops is at most

$$2(gen.rk.(\mathcal{G}))\sum_{x\in\mathbb{A}^1(\overline{k})} drop_x(\mathcal{G}) = 2de(d-1).$$

All of the d^2 ∞-slopes of $\mathcal{E}_0$ are $\leq e/d < 1$.

In case (1), we have $b \neq 0$, and hence all of the d^2 ∞- slopes the sheaf $\mathcal{E}_b$ are 1. In particular, $\mathcal{E}_b$ is totally wild at ∞, so we have

$$H^2_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_b) = 0.$$

We also have the vanishing of the H^0_c (because there are no nonzero punctual sections), and so the Lefschetz Trace formula gives

$$I(a,b,c) = -Trace(Frob_k|H^1_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_b)).$$

By Deligne's fundamental result, this H^1_c is mixed of weight at most 3 (because $\mathcal{E}_b$ is mixed of weight at most 2), and so we get the estimate

$$|I(a,b,c)| \leq (\mathrm{dim H^1_c})\mathrm{q}^{3/2} = -\chi_{\mathrm{c}}(\mathbb{A}^1 \otimes_{\mathrm{k}} \overline{\mathrm{k}}, \mathcal{E}_{\mathrm{b}})\mathrm{q}^{3/2},$$

in the case when $b \neq 0$. In this case, the "generic rank" term cancels the "$Swan_\infty$" term in the Euler-Poincare formula,

$$\chi_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_b) = gen.rk.(\mathcal{E}_b) - \sum_{x\in\mathbb{A}^1(\overline{k})} drop_x(\mathcal{E}_b) - Swan_\infty(\mathcal{E}_b),$$

and we find

$$-\chi_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_b) = \sum_{x\in\mathbb{A}^1(\overline{k})} drop_x(\mathcal{E}_b) \leq 2de(d-1).$$

So in case (1) we have the asserted estimate

$$|I(a,b,c)| \leq 2de(d-1)q^{3/2}.$$

We now turn to the proof of case (2). The basic idea for treating this case is to pay attention to the location of the finite singularities of the sheaf $\mathcal{G}$. Here a remains nonzero, b is now zero, but c is nonzero. We must estimate the quantity $I(1,0,c)$ This quantity is the sum of the traces of Frobenius at all the k-rational points of $\mathbb{A}^1$ of the sheaf

$$\mathcal{E}_0 := \mathcal{G} \otimes T^\star_c\overline{\mathcal{G}}.$$

This sheaf $\mathcal{E}_0$ is mixed of weight at most 2, it has no nonzero punctual sections, it is tame over $\mathbb{A}^1$,and the sum of its finite drops is at most

$$2(gen.rk.(\mathcal{G}))\sum_{x\in\mathbb{A}^1(\overline{k})} drop_x(\mathcal{G}) = 2de(d-1).$$

All of the d^2 ∞-slopes of $\mathcal{E}_0$ are $\le e/d < 1$, and hence we have the inequality

$$Swan_\infty(\mathcal{E}_0)) \le ed.$$

We rewrite the Euler-Poincare formula

$$\chi_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_0) = gen.rk.(\mathcal{E}_0) - \sum_{x \in \mathbb{A}^1(\overline{k})} drop_x(\mathcal{E}_0) - Swan_\infty(\mathcal{E}_0),$$

as

$$-\chi_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_0) = \sum_{x \in \mathbb{A}^1(\overline{k})} drop_x(\mathcal{E}_0) + Swan_\infty(\mathcal{E}_0) - d^2,$$

so we have the inequality

$$-\chi_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_0) \le 2de(d-1) + ed - d^2 \le 2de(d-1),$$

this last equality simply because $e < d$. We have already noted that the sheaf $\mathcal{E}_0$ is mixed of weight at most 2, and has no nonzero punctual sections. Thus its H^0_c vanishes, and exactly as in case (1) above, it remains only to prove that

$$H^2_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_0) = 0.$$

But recall that

$$\mathcal{E}_0 := \mathcal{G} \otimes T_c^\star \overline{\mathcal{G}}.$$

By assumption in case (2), the integers e and d are relatively prime. Therefore the d ∞-slopes of $\mathcal{G}$, all being e/d, have exact denominator d, and hence [Ka-GKM, 1.14] $\mathcal{G}$ is irreducible as a representation of $I(\infty)$, the inertia group at ∞. Consequently, on any dense open set U of $\mathbb{A}^1 \otimes_k \overline{k}$ on which the sheaf $\mathcal{G}$ is lisse, this sheaf is geometrically irreducible. As H^2_c is a birational invariant, what we must prove is that for any $c \neq 0$, and for any dense open set U of $\mathbb{A}^1 \otimes_k \overline{k}$ on which the sheaf $\mathcal{G}$ and its additive translate $T_c^\star \mathcal{G}$ are both lisse, these two sheaves are not geometrically isomorphic. We argue by contradiction. If $\mathcal{G}$ and its additive translate $T_c^\star \mathcal{G}$ are geometrically isomorphic on some dense open set, then their extensions by direct image to $\mathbb{A}^1 \otimes_k \overline{k}$, say $\mathcal{G}_{mid}$ and $T_c^\star \mathcal{G}_{mid}$ are isomorphic middle extension sheaves on $\mathbb{A}^1 \otimes_k \overline{k}$, and hence they have the same set S of "finite singularities", i.e., the same set S of points in $\mathbb{A}^1 \otimes_k \overline{k}$ at which they fail to be lisse. Thus the set S of finite singularities is equal to its additive translate by c. Now the additive group $\mathbb{F}_p c$ generated by c acts freely on $\mathbb{A}^1(\overline{k})$ by additive translation, so from the stability of the set S under this free action, we infer the congruence $Card(S) \equiv 0$ mod p. We have already noted that the literal sheaf $\mathcal{G}$ on $\mathbb{A}^1 \otimes_k \overline{k}$ has at most $e(d-1)$ finite singularities. Therefore the middle extension sheaf $\mathcal{G}_{mid}$, whose finite singularities are among those of $\mathcal{G}$, itself has at most $e(d-1)$ finite singularities. By assumption in case (2), we have

$$e(d-1) < p.$$

Thus we infer that S is empty, i.e., that the sheaf $\mathcal{G}_{mid}$ is lisse of rank d on $\mathbb{A}^1 \otimes_k \overline{k}$. This now leads to a contradiction as follows. From its $I(\infty)$-slopes all being e/d, we have already noted that $\mathcal{G}_{mid}$ is $I(\infty)$-irreducible. Therefore $\mathcal{G}_{mid}$ is lisse of rank d, geometrically irreducible and nonconstant. So its only possibly nonzero compact cohomology group on $\mathbb{A}^1 \otimes_k \overline{k}$ is H^1_c. Thus its Euler characteristic is non-positive. But this Euler characteristic is

$$rank(\mathcal{G}_{mid}) - Swan_\infty(\mathcal{G}_{mid}) = d - e \geq 1,$$

contradiction. This concludes the proof of case (2).

4. Proof of the main result in cases (3), (4), and (5)

What these cases have in common is that $a = 0$. In this situation, the relevant sheaf $\mathcal{F}$ on the w-line $\mathbb{A}^1$ over k we need to begin with is

$$\mathcal{F} := Ker(Trace : f_\star\overline{\mathbb{Q}_\ell} \to \overline{\mathbb{Q}_\ell}).$$

The function H_0 on k is

$$\begin{aligned} H_0(t) &:= \sum_{x \in k} \chi(t - f(x)) \\ &= \sum_{w \in k} \chi(t-w) \sum_{x \in k, f(x)=w} 1 \\ &= \sum_{w \in k} \chi(t-w)(-1 + \sum_{x \in k, f(x)=w} 1), \end{aligned}$$

this last equality because χ is nontrivial. Thus

$$H_0(t) = \sum_{w \in k} Trace(Frob_{k,w} | \mathcal{L}_{\chi(t-w)} \otimes \mathcal{F}).$$

This means precisely that the function H_0 on k is the trace function, at k-rational points, of the lower ! additive convolution of the sheaves $\mathcal{L}_\chi$ and $\mathcal{F}$, or what is the same, of the lower ! additive convolution, say M,

$$M := \mathcal{L}_\chi[1] \star_{+,!} \mathcal{F}[1]$$

of the perverse sheaves $\mathcal{L}_\chi[1]$ and $\mathcal{F}[1]$ on $\mathbb{A}^1$ over k. Both of these perverse sheaves are visibly middle extensions (remember χ is nontrivial) which, as perverse sheaves, are pure of weight one. While $\mathcal{L}_\chi[1]$ is geometrically irreducible, $\mathcal{F}[1]$ need not be (and indeed is certainly not in Pierce's original question), but it is, in any case, geometrically semisimple. Because $d < p$, the perverse sheaf $\mathcal{F}[1]$ is everywhere tamely ramified. We claim it satisfies "condition $\mathcal{P}$" in the sense of [Ka-RLS, 2.6]. Indeed, by tameness it contains, geometrically, no sheaf $\mathcal{L}_{\psi_t}$ for any nonzero t. It contains no constant sheaf either, because by Frobenius reciprocity, the sheaf $f_\star\overline{\mathbb{Q}_\ell}$ contains just one copy of the constant sheaf. So once again by [Ka-RLS, 2.6], the additive ! convolution

$$M := \mathcal{L}_\chi[1] \star_{!,+} \mathcal{F}[1]$$

is itself a perverse sheaf on $\mathbb{A}^1$ over k. Looking fibre by fibre, and remembering that in cases (3) and (4) χ^d is assumed nontrivial, we see that M is of the form $\mathcal{H}[1]$, with $\mathcal{H}$ a single sheaf, which is, on a dense open set, both lisse of rank d-1 and pure of weight one. Thus M is, on a dense open set, pure of weight 2 as a perverse sheaf. Exactly as in the previous section, this generic purity shows that M is the "middle additive convolution" of $\mathcal{L}_\chi[1]$ and $\mathcal{F}[1]$. We know, from [Ka-RLS, 2.10.10] and [Ka-MMP, 6.5.4, 2)] respectively, that $M = \mathcal{H}[1]$ is pure of weight 2 as a perverse sheaf, and hence geometrically semisimple. This means in turn that $\mathcal{H}$ is mixed of weight at most 1, and that it is the middle extension from some open dense set of a geometrically semisimple lisse sheaf which is pure of weight 1. Moreover, it is an exercise, using [Ka-RLS, 3.3.5], to compute both the generic rank of $M = \mathcal{H}[1]$, (which in our case we know "by inspection" to be d-1) and all its local monodromies, given the same data for $\mathcal{F}$. Recall that $\mathcal{F}$ is an everywhere tame middle extension of generic rank $d-1$. The sum of its drops over all $\overline{k}$-valued points of $\mathbb{A}^1$ is just $d-1$, the total number of zeroes, counting multiplicities, of the derivative $f'(x)$ of f at all $\overline{k}$-valued points of $\mathbb{A}^1$. It is then routine to conclude, using [Ka-RLS, 3.3.5], that $\mathcal{H}$ is everywhere tame. Moreover, the drops of $\mathcal{H}$ at the $\overline{k}$-valued points of $\mathbb{A}^1$ occur precisely at the points where $\mathcal{F}$ had drops, and the drops at each such point are the same for $\mathcal{H}$ and for $\mathcal{F}$. These drop points are the critical values of the polynomial f, i.e. the values $f(\alpha)$ at the zeroes α of the derivative $f'(x)$. At each critical value $f(\alpha)$ of f, the drop at $f(\alpha)$ is the sum, over all zeroes γ in $\overline{k}$ of the polynomial $f(x) - f(\alpha)$, of the expression (multiplicity of γ as a zero of $f(x) - f(\alpha)$) - 1. In particular, the sum of all the drops of $\mathcal{H}$ at finite distance is just $d-1$, the total number, counting multiplicity, of zeroes of the derivative $f'(x)$.

With this information established about $\mathcal{H}$, we now turn to the sheaf $\mathcal{G} := \mathcal{G}_1 := g^\star\mathcal{H}$ on $\mathbb{A}^1$. It is mixed of weight at most 1, and on a dense open set it is lisse and pure of weight 1. It has generic rank $d-1$, and everywhere tame. Its only drops at $\overline{k}$-valued points of $\mathbb{A}^1$ are at the inverse images by g of the points at which $\mathcal{H}$ drops, and at such an inverse image point, the size of the drop is unchanged. Since g is a polynomial of degree e, there are at most e(d-1) finite points at which $\mathcal{G}$ drops, and the sums of all its drops at finite points is at most e(d-1). Moreover, the sheaf $\mathcal{G}$ has no nonzero punctual sections (as this was already true of $\mathcal{H}$, being a middle extension sheaf).

We can now treat cases (3), (4), and (5). The sheaf $\overline{\mathcal{G}}$, the sheaf formed using the character $\overline{\chi}$, but the same f and g, has the same ramification properties as $\mathcal{G}$. The effect of any additive translation

$$T_c : z \to z + c$$

on $\overline{\mathcal{G}}$ is just to translate the finite singularities, keeping the same drops. We must estimate the quantity $I(0, b, c)$ This quantity is the sum of the traces of Frobenius at all the k-rational points of $\mathbb{A}^1$ of the sheaf

$$\mathcal{E}_b := \mathcal{G} \otimes T_c^{\star}\overline{\mathcal{G}} \otimes \mathcal{L}_{\psi_b}.$$

Notice that

$$\mathcal{E}_b = \mathcal{E}_0 \otimes \mathcal{L}_{\psi_b}.$$

This sheaf $\mathcal{E}_0$ is mixed of weight at most 2, it has no nonzero punctual sections, it is everywhere tame, and the sum of its finite drops is at most

$$2(gen.rk.(\mathcal{G})) \sum_{x \in \mathbb{A}^1(\overline{k})} drop_x(\mathcal{G}) = 2e(d-1)^2.$$

In case (3), we have $b \neq 0$, and hence $\mathcal{E}_b$ is totally wild at ∞, with all breaks 1. So $H^i_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_b)$ vanishes for $i \neq 1$. The dimension of the H^1_c is given by the Euler-Poincare formula, in which the "generic rank" term is cancelled by the $Swan_\infty$ term, and we find

$$dim H^1_c = -\chi(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_b) = \sum_{x \in \mathbb{A}^1(\overline{k})} drop_x(\mathcal{E}_b) \leq 2e(d-1)^2.$$

So in case (3) we find the estimate

$$|I(0, b, c)| \leq 2e(d-1)^2 q^{3/2},$$

which is in fact a slight improvement over the asserted estimate

$$|I(0, b, c)| \leq 2ed(d-1)q^{3/2}.$$

In case (4), we have $a = b = 0$, but $c \neq 0$. In this case we have

$$-\chi(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_0) = -(d-1)^2 + \sum_{x \in \mathbb{A}^1(\overline{k})} drop_x(\mathcal{E}_0) \leq -(d-1)^2 + 2e(d-1)^2 \leq 2ed(d-1).$$

It remains to show that $H^2_c(\mathbb{A}^1 \otimes_k \overline{k}, \mathcal{E}_0)$ vanishes, or equivalently, that on any dense open set U in $\mathbb{A}^1 \otimes_k \overline{k}$ on which both $\mathcal{G}$ and $T_c^{\star}\mathcal{G}$ are lisse (remember both are geometrically semisimple, by purity), these two sheaves have no irreducible constituent in common. Because we are in case (4), f is just a monomial of degree d, and consequently $\mathcal{F}$ is, geometrically, the extension by zero of the lisse sheaf on $\mathbb{G}_m \otimes_k \overline{k}$ given by

$$\mathcal{F} \cong \bigoplus_{\rho^d=1, \rho\, nontriv.} \mathcal{L}_\rho.$$

Hence its middle convolution with $\mathcal{L}_\chi$ is given by

$$\mathcal{H} \cong \bigoplus_{\rho^d=1, \rho\, nontriv.} \mathcal{L}_{\rho\chi}.$$

[Remember that χ^d is nontrivial, so each character $\rho\chi$ which appears in the above sum is nontrivial]. The pullback by g is then given by

$$\mathcal{G} \cong \bigoplus_{\rho^d=1, \rho\, nontriv.} \mathcal{L}_{(\rho\chi)(g(z))}.$$

Its additive translate by c is then

$$T_c^\star \mathcal{G} \cong \bigoplus_{\rho^d=1, \rho\, nontriv.} \mathcal{L}_{(\rho\chi)(g(z+c))}.$$

So what we must show is that for any $c \neq 0$, no $\mathcal{L}_{(\rho_i\chi)(g(z+c))}$ is isomorphic to any $\mathcal{L}_{(\rho_j\chi)(g(z))}$, for any two nontrivial, not necessarily distinct, characters ρ_i and ρ_j of order dividing d. Since the polynomials $g(z)$ and $g(z+c)$ have no common zero, this would be obvious if we knew that each $\mathcal{L}_{(\rho_i\chi)(g(z+c))}$ were ramified at every zero of $g(z+c)$, and that each $\mathcal{L}_{(\rho_i\chi)(g(z))}$ were ramified at every zero of $g(z)$. In case (4a), the fact that n and d are relatively prime implies that each character $\rho_i\chi$ has order at least n. As g has degree $e < n$, every zero of g has multiplicity at most $e < n$, and hence we do indeed have the asserted ramification. In case (4b), d is prime as well as prime to n, so each character $\rho_i\chi$ has order precisely nd. In case (4b), g has degree $e < nd$, and we conclude by the same ramification argument as in case (4a).

In case (5), we use the fact [Ka-ACT, 5.15] that the geometric monodromy group G_geom of the sheaf $\mathcal{H}$ is either the symplectic group $Sp(d-1)$ in case (5a), or that its identity component G_geom^0 is $SL(d-1)$, in cases (5b) and (5c). We retain only the consequence that the sheaf $\mathcal{H}$ is geometrically Lie-irreducible, and hence its pullback by any nonconstant map is geometrically irreducible. So $\mathcal{G}$ is itself geometrically irreducible. Just as above, it suffices to show that for any choice of $c \neq 0$, on any dense open set U in $\mathbb{A}^1 \otimes_k \overline{k}$ on which both $\mathcal{G}$ and $T_c^\star\mathcal{G}$ are lisse, these two sheaves are not isomorphic. For this, we argue by contradiction, using the argument alrady used in case (2) of paying attention to the set S of finite singularities of $\mathcal{G}_{mid}$. Because

$e(d-1) < p$, we conclude that S is empty,i.e., that $\mathcal{G}_{mid}$ is lisse on $\mathbb{A}^1$. Since $\mathcal{G}_{mid}$ is also tame at ∞, it must be geometrically constant. But as it has rank $d-1 \geq 2$, this geometric constancy contradicts its geometric irreducibility.

References

[Dav-Ha] Davenport, H.,and Hasse, H., Die Nullstellen der Kongruenz zetafunktionen in gewissen zyklischen Fällen, J. Reine Angew. Math. 172, 1934, 151-182, reprinted in *The Collected Works of Harold Davenport*(ed. Birch, Halberstam, Rogers), Academic Press, 1977.

[De-Weil II] Deligne, P., La Conjecture de Weil II, Pub. Math. I.H.E.S. 52, 1981, 313-428.

[Gro-FL] Grothendieck,A., Formule de Lefschetz et rationalite des fonctions L, Seminaire Bourbaki 1964-65, Expose 279, reprinted in Dix Exposes sur la cohomologie des schemas, North-Holland, 1968.

[Ka-ACT] Katz, N., Affine cohomological transforms, perversity and monodromy, JAMS, Vol. 6, No. 1, January, 1993, 149-222. Annals of Math. Study 116, Princeton Univ. Press, 1988.

[Ka-GKM] Katz, N., *Gauss sums, Kloosterman sums, and monodromy groups*, Annals of Math. Study 116, Princeton Univ. Press, 1988.

[Ka-MMP] Katz, N., *Moments, Monodromy, and Perversity: a Diophantine Perspective*, to appear as an Annals of Math. Study, available as preprint on www.math.princeton.edu/ nmk.

[Ka-RLS] Katz, N., *Rigid Local Systems*, Annals of Math. Study 138, Princeton Univ. Press, 1995.

[P] Pierce, L., Master's Thesis, Oxford University, 2004.

[Ray] Raynaud, M., Caracteristique d'Euler-Poincare d'un Faisceau et cohomologie des varietes abeliennes, Expose 286 in Seminaire Bourbaki 1964/65, W.A. Benjamin, New York, 1966.

[We-CA] Weil, A., *Courbes algebriques et varietes abeliennes*, Hermann, Paris, 1971.

Princeton University, Mathematics, Fine Hall, NJ 08544-1000, USA
E-mail address: nmk@math.princeton.edu