ACTA ARITHMETICA
LXI.3 (1992)

The divisor function over arithmetic progressions

by

Etienne Fouvry (Paris) and Henryk Iwaniec* (New Brunswick, N.J.)
with appendix by Nicholas Katz (Princeton, N.J.)

1. Introduction. Given an arithmetic function $f(n)$ one often expects its values over primitive residue classes $n \equiv a \pmod q$ to be equidistributed, i.e.

$$\mathcal{D}_f(x;q,a) = \sum_{\substack{n \le x \\ n \equiv a \ (\mathrm{mod}\ q)}} f(n)$$

to be well approximated by

$$\mathcal{D}_f(x;q) = \frac{1}{\varphi(q)} \sum_{\substack{n \le x \\ (n,q)=1}} f(n)\,,$$

provided x is sufficiently large. An asymptotic formula of type

$$\mathcal{D}_f(x;q,a) = (1 + O((\log x)^{-A}))\mathcal{D}_f(x;q)\,, \tag{1}$$

in which the error term is smaller than the main term by a suitable power of $\log x$, is good enough for basic applications. More important than the size of the error term is the range where (1) holds uniformly with respect to the modulus q.

In this paper we consider the problem for the divisor function $f(n) = \tau(n)$. In this case one can prove by a simple elementary argument that

$$\varDelta_f(x;q,a) = \mathcal{D}_f(x;q,a) - \mathcal{D}_f(x;q) \ll x^{1/2+\varepsilon}\,,$$

which yields (1) in the range $q < x^{1/2-2\varepsilon}$. Using Fourier series technique and Weil's estimate for Kloosterman sums

$$S(m,n;q) = \sum_{uv \equiv 1 \ (\mathrm{mod}\ q)} e\left(\frac{mu+nv}{q}\right) \ll (m,n,q)^{1/2} q^{1/2+\varepsilon} \tag{2}$$

* Supported by the NSF research grant DMS-8902992.

one can show that

$$(3)\qquad \Delta(x;q,a) \ll (q^{1/2}+x^{1/3})x^{\varepsilon}\,.$$

Hence it follows that (1) holds uniformly in $q < x^{2/3-\varepsilon}$. Further progress has been made recently by the first author [3] (see Corollaire 5) showing that (1) is true on average with respect to q in the range $x^{2/3+\varepsilon} < q < x^{1-\varepsilon}$. More precisely, we have

$$(4)\qquad \sum_{\substack{x^{2/3+\varepsilon}<q<x^{1-\varepsilon}\\ (q,a)=1}} |\Delta_f(x;q,a)| \ll x(\log x)^{-A}$$

for any $\varepsilon, A > 0$, with the implied constant depending on ε, A and a.

In this paper we establish a result which covers the gap $x^{2/3-\varepsilon} < q < x^{2/3+\varepsilon}$ with special moduli q.

THEOREM 1. *Let r be squarefree with $(a,r)=1$, $r \le x^{3/8}$. We have*

$$(5)\qquad \sum_{\substack{rs^2<x^{1-6\varepsilon}\\ (s,ar)=1}} |\Delta_f(x;rs,a)| \ll r^{-1}x^{1-\varepsilon}\,,$$

with the implied constant depending on ε alone.

Theorem 1 will be inferred from the following estimate for sums of Kloosterman sums (which we consider to be the main result of this paper).

THEOREM 2. *Let r be squarefree with $(a,r)=1$ and let λ_l, $l \le L$, be arbitrary complex numbers. We then have*

$$(6)\qquad \sum_{s\le S,\,(s,r)=1} \Big|\sum_{l\le L} \lambda_l S(a,l;rs)\Big|^2 \ll \Lambda L S^2 r(r^{-1/4}+r^{1/4}S^{-1/2}+SL^{-1})(rS)^{\varepsilon}$$

where $\Lambda = \sum |\lambda_l|^2$ and the implied constant depends on ε only.

Remark. Weil's estimate for the individual Kloosterman sum $S(a,l;rs)$ yields the bound $O(\Lambda L S^2 r(rS)^{\varepsilon})$.

In the proof of (6) we arrive at certain exponential sums in five variables over a finite field. An estimate for these sums is proved in the appendix by N. Katz.

A closely related problem of evaluating the mean value of the divisor function $f(n)=\tau_3(n)$ over an arithmetic progression is considered in [4], where it is proved, using estimates for exponential sums in two and three variables over a finite field, that the asymptotics (1) holds true uniformly with $q < x^{58/115}$.

The second author acknowledges the hospitality and support from the Université de Paris-Sud when working on this paper.

2. Proof of Theorem 2. For the proof of (6) we can assume without loss of generality that l ranges over numbers prime to r. This can be arranged using the relation

$$S(a,l;rs)=\mu(d)S(a\overline{d},ld^{-1};srd^{-1})\,,\qquad d\overline{d}\equiv 1\ (\mathrm{mod}\,sr/d)\,,$$

where $d=(r,l)$. Furthermore, by splitting into dyadic intervals we can attach to the outer summation a weight function $\omega(s)=\min(sS^{-1}-1,1,4-sS^{-1})$ if $S<s<4S$ and $\omega(s)=0$ elsewhere. We can also assume that $L>S>r^{1/2}$ because (6) follows from (2) otherwise.

The Kloosterman sum in (6) factors into $S(a\overline{s},l\overline{s};r)S(a\overline{r},l\overline{r};s)$, where $\overline{s}s\equiv 1\ (\mathrm{mod}\,r)$ and $\overline{r}r\equiv 1\ (\mathrm{mod}\,s)$. We split the summation over l into classes modulo s, apply Cauchy's inequality and use the formula

$$\sum_{b\,(\mathrm{mod}\,s)}|S(a,b;s)|^2=\varphi(s)s$$

to obtain

$$\begin{aligned}
\mathcal{A}:=&\sum_{(s,r)=1}\omega(s)\Big|\sum_{l\le L}\lambda_l S(a,l;rs)\Big|^2 \qquad (7)\\
&\le\sum_{(s,r)=1}\omega(s)\Big(\sum_{b\,(\mathrm{mod}\,s)}|S(a\overline{r},b\overline{r};s)|\Big|\sum_{l\equiv b\,(\mathrm{mod}\,s)}\lambda_l S(a\overline{s},l\overline{s};r)\Big|\Big)^2\\
&\le 16S^2\sum_{(s,r)=1}\omega(s)\sum_{b\,(\mathrm{mod}\,s)}\Big|\sum_{l\equiv b\,(\mathrm{mod}\,s)}\lambda_l S(a\overline{s},l\overline{s};r)\Big|^2\\
&=16S^2\sum_{(s,r)=1}\omega(s)\sum_{l_1\equiv l_2\,(\mathrm{mod}\,s)}\lambda_{l_1}\overline{\lambda}_{l_2}S(a\overline{s},l_1\overline{s};r)S(a\overline{s},l_2\overline{s};r)\,.
\end{aligned}$$

The terms with $l_1=l_2$ contribute $O(\Lambda S^3r^{1+\varepsilon})$ by (2). For the other terms we write $|l_1-l_2|=st$ with $1\le t<LS^{-1}=T$, say, and obtain

$$\mathcal{A}\le 16S^2\sum_{1\le t<T}\mathcal{A}_t+O(\Lambda S^3r^{1+\varepsilon})\,, \tag{8}$$

where

$$\begin{aligned}
\mathcal{A}_t=\sum\sum_{(l_1-l_2,tr)=t}&\lambda_{l_1}\overline{\lambda}_{l_2}\omega\left(\frac{|l_1-l_2|}{t}\right)\\
&\times S(at\overline{l_1-l_2},tl_1\overline{l_1-l_2};r)S(at\overline{l_1-l_2},tl_2\overline{l_1-l_2};r)\,.
\end{aligned}$$

To separate the variables l_1, l_2 in the weight function we use the Fourier transform technique. We write

$$\omega\left(\frac{|l_1-l_2|}{t}\right)=t\int_{-\infty}^{\infty}\Omega(ty)e(l_1y-l_2y)\,dy\,,$$

where $\Omega(y)$ is the Fourier transform of $\omega(|x|)$, so

$$\int_{-\infty}^{\infty} |\Omega(y)|\, dy < 5\,.$$

We obtain

$$|\mathcal{A}_t| \le 5 \sum_{l_2} |\lambda_{l_2}| \Big| \sum_{(l_1-l_2,tr)=t} e(l_1 y)\lambda_{l_1} \times S(at\overline{l_1-l_2}, tl_1\overline{l_1-l_2}; r) S(at\overline{l_1-l_2}, tl_2\overline{l_1-l_2}; r)\Big|$$

for some $y \in \mathbb{R}$. Now by Cauchy's inequality we get

$$\mathcal{A}_t^2 \ll \Lambda\left(1+\frac{L}{tr}\right) \sum_{z \,(\mathrm{mod}\, tr)} \Big| \sum_{(l_1-z,tr)=t} e(l_1 y)\lambda_{l_1} \times S(at\overline{l_1-z}, tl_1\overline{l_1-z}; r) S(at\overline{l_1-z}, tz\overline{l_1-z}; r)\Big|^2 \le \Lambda\left(1+\frac{L}{tr}\right) \sum_{l_1 \equiv l_2 \,(\mathrm{mod}\, t)} |\lambda_{l_1}\lambda_{l_2} V(l_1,l_2;r)|\,, \tag{9}$$

say, where

$$V(l_1,l_2;r) = \sum_{\substack{z \,(\mathrm{mod}\, tr)\\ (z-l_1,tr)=t \\ (z-l_2,tr)=t}} S(at\overline{l_1-z}, tl_1\overline{l_1-z}; r) S(at\overline{l_1-z}, tz\overline{l_1-z}; r) \times S(at\overline{l_2-z}, tl_2\overline{l_2-z}; r) S(at\overline{l_2-z}, tz\overline{l_2-z}; r)\,.$$

For notational convenience we also consider conjugate sums $V^{\psi}(l_1,l_2;r)$, where ψ is an additive character to modulus r. We define the conjugate Kloosterman sums by

$$S^{\psi}(m,n;r) = \sum_{uv \equiv 1 \,(\mathrm{mod}\, r)} \psi(mu+nv)\,.$$

Then by conjugating the Kloosterman sums in $V(l_1,l_2;r)$ we define $V^{\psi}(l_1,l_2;r)$.

Proposition. *Suppose that* $l_1 \equiv l_2 \pmod t$ *and* r *is squarefree with* $(r, al_1l_2) = 1$. *We have*

$$V(l_1,l_2;r) \ll (l_1-l_2,r)^{1/2} r^{5/2+\varepsilon}\,, \tag{10}$$

with the implied constant depending on ε *only.*

Proof. The sum $V(l_1,l_2;r)$ is multiplicative in r. Suppose $r = r_1r_2$ with $(r_1,r_2) = 1$. Let ψ_1,ψ_2 be the additive characters $\psi_1(x) = e(x\overline{r}_2/r_1)$ and $\psi_2(x) = e(x\overline{r}_1/r_2)$ to modulus r_1 and r_2 respectively. We then have

$$V(l_1,l_2;r) = V^{\psi_1}(l_1,l_2;r_1) V^{\psi_2}(l_1,l_2;r_2)\,.$$

Thus it suffices to prove (10) for conjugate sums $V^\psi(l_1,l_2;r)$ with prime modulus $r=p$ and nontrivial character $\psi \pmod p$.

If $l_1 \equiv l_2 \pmod p$ then (10) follows from (2) by trivial summation over z. Suppose $l_1 \not\equiv l_2 \pmod p$. We then substitute $z = l_1 + (l_1-l_2)(x-1)^{-1}$ giving

$$V^\psi(l_1,l_2;p) = \sum_{\substack{x \,(\mathrm{mod}\, p)\\ (x(x-1),p)=1}} S^\varrho(ax-a, l_1x-l_1;p)S^\varrho(ax-a,l_1x-l_2;p) \times S^\varrho(a\overline{x}-a, l_2\overline{x}-l_2;p)S^\varrho(a\overline{x}-a,l_2\overline{x}-l_1;p)\,,$$

where $\varrho(x) = \psi(xt\overline{l_1-l_2})$ is a nontrivial additive character to modulus p. We extend the sum $V^\psi(l_1,l_2;p)$ by adding terms with $x\equiv 1 \pmod p$. We get

$$V^\psi(l_1,l_2;p) = V_g(p) - (p-1)^2\,, \tag{11}$$

where

$$V_g(p) = \sum_{x,x_1,x_2,x_3,x_4\in\mathbb{F}_p^*} \varrho(g(x,x_1,x_2,x_3,x_4))$$

and g is the Laurent polynomial

$$\begin{aligned} g(x,x_1,x_2,x_3,x_4) = (ax-a)x_1 &+ (l_1x-l_1)x_1^{-1} \\ &+ (ax-a)x_2 + (l_1x-l_2)x_2^{-1} \\ &+ (ax^{-1}-a)x_3 + (l_2x^{-1}-l_2)x_3^{-1} \\ &+ (ax^{-1}-a)x_4 + (l_2x^{-1}-l_1)x_4^{-1}\,. \end{aligned}$$

The sums of type $V_g(p)$ (for general Laurent polynomials) have recently been studied by A. Adolphson and S. Sperber [1], [2]. They established the best possible estimates under certain conditions on the Newton polyhedron associated with g. Unfortunately, our polynomial does not satisfy their conditions. Yet, the desired estimate

$$V_g(p) \ll p^{5/2} \quad \text{if } p \nmid al_1l_2(l_1-l_2) \tag{12}$$

is true. This is proved by N. Katz in the appendix to this paper. As a matter of fact Katz considers the above Laurent polynomials with the parameter $a=1$. The reduction to his case can be made without loss of generality by a suitable change of the character ϱ and the parameters l_1, l_2. By (11) and (12) we get (10).

Now we are ready to complete the proof of Theorem 2. By (9) and (10) we get

$$\begin{aligned} \mathcal{A}_t^2 &\ll \Lambda\left(1+\frac{L}{tr}\right) \sum_{l_1\equiv l_2 \,(\mathrm{mod}\, t)} |\lambda_{l_1}\lambda_{l_2}|(l_1-l_2,r)^{1/2}r^{5/2+\varepsilon} \\ &\ll \Lambda^2(tr+L)(tr^{1/2}+(r,t)L)t^{-2}r^{3/2+\varepsilon}\,. \end{aligned}$$

Hence by (8) we conclude that

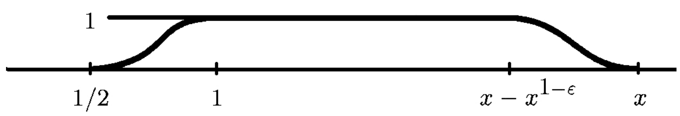
$$\begin{aligned}\mathcal{A} &\ll \Lambda S^2 (Tr+L)^{1/2}(Tr^{1/2}+L)^{1/2} r^{3/4+\varepsilon} + \Lambda S^3 r^{1+\varepsilon}\\ &\ll \Lambda L S (r+S)^{1/2}(r^{1/2}+S)^{1/2} r^{3/4+\varepsilon} + \Lambda S^3 r^{1+\varepsilon}\end{aligned}$$

giving (6).

3. Proof of Theorem 1. There are many ways of transforming $\Delta(x;q,a)$ into a sum of Kloosterman sums. To this end one can use for example the properties of the modular form

$$u(z) = \sqrt{y}\log y + 4\sqrt{y}\sum_{n=1}^{\infty} \tau(n) K_0(2\pi n y)\cos(2\pi n x)\,.$$

However, we choose a direct approach. Let F be a function whose graph is



Put $F(\xi,\eta) = F(\xi)F(\eta)F(\xi\eta)$. Then the divisor function $\tau(n)$ agrees with the function

$$f(n) = \sum_{n_1 n_2 = n} F(n_1, n_2)$$

for all $n \le x - x^{1-\varepsilon}$. The remaining n's with $x - x^{1-\varepsilon} < n \le x$ contribute trivially $O(q^{-1}x^{1-\varepsilon}\log x)$. Therefore it suffices to prove (5) with the above f. By Poisson's summation we get

$$\begin{aligned}\mathcal{D}_f(x;q,a) &= \sum_{uv\equiv a \,(\mathrm{mod}\, q)} \ \sum_{(n_1,n_2)\equiv(u,v)\,(\mathrm{mod}\, q)} F(n_1,n_2)\\ &= \sum_{m_1,m_2} S(am_1, m_2; q) G(m_1, m_2)\,,\end{aligned}$$

where G is the Fourier transform of $F(\xi q, \eta q)$. Summing over the primitive residue classes we get

$$\mathcal{D}_f(x;q) = \frac{1}{\varphi(q)} \sideset{}{^*}\sum_{b\,(\mathrm{mod}\, q)} \ \sum_{m_1,m_2} S(bm_1, m_2; q) G(m_1, m_2).$$

Note that the frequencies m_1, m_2 with $m_1 m_2 = 0$ give the same contribution to both $\mathcal{D}_f(x;q,a)$ and $\mathcal{D}_f(x;q)$. Therefore we have

$$\Delta_f(x;q,a) = \mathcal{R}(q,a) - \mathcal{R}(q)\,,$$

where

$$\mathcal{R}(q,a) = \sum_{m_1 m_2 \neq 0} S(am_1, m_2; q) G(m_1, m_2)\,,$$

$$\mathcal{R}(q) = \frac{1}{\varphi(q)} \sum_{m_1 m_2 \neq 0} r_q(m_1) r_q(m_2) G(m_1, m_2)\,,$$

and $r_q(m) = S(m, 0; q)$ are the Ramanujan sums. The above expression for $\mathcal{R}(q)$ follows from the identity

$$\sum_{b \,(\mathrm{mod}\, q)}^{*} S(bm_1, m_2; q) = r_q(m_1) r_q(m_2)\,.$$

By the formula

$$S(am_1, m_2; q) = \sum_{d|(m_1, m_2, q)} dS(a, m_1 m_2 d^{-2}; qd^{-1})$$

we get

$$\mathcal{R}(q,a) = \sum_{cd=q} d^{-1} \sum_{l \neq 0} S(a, l; c) g_c(l)\,,$$

where

$$g_c(l) = \int\!\!\int F(\xi c, \eta c) \lambda_l(\xi, \eta)\, d\xi\, d\eta\,, \qquad \lambda_l(\xi,\eta) = \sum_{l_1 l_2 = l} e(\xi l_1 + \eta l_2)\,.$$

Similarly (or summing over primitive residue classes $a \ (\mathrm{mod}\, q)$) we get

$$\mathcal{R}(q) = \sum_{cd=q} \frac{\mu(c)}{d\varphi(c)} \sum_{l \neq 0} r_c(l) g_c(l)\,.$$

Trivially we have

$$|g_c(l)| \le c^{-2} \int\!\!\int F(\xi, \eta)\, d\xi\, d\eta \le 2c^{-2} x \log x$$

while by iterated partial integration we obtain

$$g_c(l) \ll (lx)^{-2} \quad \text{ if } |l| > c^2 x^{2\varepsilon - 1}\,.$$

Hence by the trivial bounds $|S(a,l;c)| \le c$ and $|r_c(l)| \le (l,c)$ we obtain the approximate formula

$$\Delta_f(x; q, a) = \sum_{cd=q} d^{-1} \sum_{0<|l|\le L} S(a, l; c) g_c(l) + O(q^{-1} x^{\varepsilon})\,,$$

where L is any number $\ge c^2 x^{2\varepsilon-1}$. Inserting the Fourier integral for $g_c(l)$ we get

$$|\Delta_f(x; q, a)| < \sum_{cd=q} d^{-1} \int\!\!\int_{\substack{\xi\eta < c^{-2} x \\ \xi, \eta > 1/2}} \Big| \sum_{0<|l|\le L} \lambda_l(\xi,\eta) S(a, l; c) \Big|\, d\xi\, d\eta + O(q^{-1} x^{\varepsilon})\,.$$

Hence

$$\sum_{\substack{q<Q\\(q,ar^2)=r}} |\Delta_f(x;q,a)| < 2x(\log 2x) \sum_{\substack{cd<Q\\(cd,r^2)=r}} d^{-1}c^{-2}\Big|\sum_{0<l\le L}\lambda_l S(a,l;c)\Big| + O(r^{-1}x^{\varepsilon}),$$

where $\lambda_l = \lambda_l(\xi,\eta)$ for some real ξ, η, so $|\lambda_l| \le \tau(l)$. Finally, Theorem 2 yields

$$\sum_{\substack{q<Q\\(q,ar^2)=r}} |\Delta_f(x;q,a)| \ll r^{-1}Q(r^{-1/2}x^{1/2} + r^{-1/8}Q^{1/2} + r^{3/8}Q^{1/4})x^{2\varepsilon} \ll r^{-1}x^{1-\varepsilon}$$

for $Q = r^{1/2}x^{1/2-3\varepsilon}$ provided $r \le x^{3/8}$.

This completes the proof of Theorem 1.

References

[1] A. Adolphson and S. Sperber, *Exponential sums and Newton polyhedra*, Bull. Amer. Math. Soc. 16 (1987), 282–286.

[2] —, —, *Exponential sums on* $(G_m)^n$, Invent. Math. 101 (1990), 63–79.

[3] E. Fouvry, *Sur le problème des diviseurs de Titchmarsh*, J. Reine Angew. Math. 357 (1985), 51–76.

[4] J. Friedlander and H. Iwaniec, *Incomplete Kloosterman sums and a divisor problem* (with appendix by B. J. Birch and E. Bombieri), Ann. of Math. 121 (1985), 319–350.

APPENDIX

by Nicholas Katz

We fix a finite field $\mathbb{F}_q$, a nontrivial $\mathbb{C}$-valued additive character ψ of $\mathbb{F}_q$, and elements α, β in $\mathbb{F}_q$. We denote by t, x_1, x_2, x_3, x_4 five independent variables over $\mathbb{F}_q$. We denote by

$$f_{\alpha,\beta}(t,x_1,x_2,x_3,x_4) \quad \text{in } \mathbb{F}_q[t^{\pm1},x_1^{\pm1},x_2^{\pm1},x_3^{\pm1},x_4^{\pm1}]$$

the Laurent polynomial

$$(t-1)x_1 + \alpha(t-1)/x_1 + (t-1)x_2 + (\alpha t-\beta)/x_2 + (1/t-1)x_3 + \beta(1/t-1)/x_3 + (1/t-1)x_4 + (\beta/t-\alpha)/x_4\,.$$

We denote by $S(\mathbb{F}_q,\psi,\alpha,\beta)$ the sum

$$S(\mathbb{F}_q,\psi,\alpha,\beta) := \sum_{t,x_1,x_2,x_3,x_4\in(\mathbb{F}_q)^\times} \psi(f_{\alpha,\beta}(t,x_1,x_2,x_3,x_4))\,.$$

THEOREM 1. *If $\alpha\beta(\alpha-\beta)\neq 0$, we have the estimate*

$$|S(\mathbb{F}_q,\psi,\alpha,\beta)| \le 64q^{5/2}+q^2+2q+1\,.$$

P r o o f. As we will see, this is essentially an exercise in the theory of Kloosterman sheaves. For γ,σ in $\mathbb{F}_q$, we denote by $\mathrm{Kl}_2(\mathbb{F}_q,\psi,\gamma,\sigma)$, or simply $\mathrm{Kl}(\gamma,\sigma)$, the Kloosterman sum

$$\mathrm{Kl}(\gamma,\sigma) := \sum_{x\in(\mathbb{F}_q)^\times} \psi(\gamma x+\sigma/x)\,.$$

For $\tau\neq 0$ in $\mathbb{F}_q$, we define

$$\mathrm{Kl}(\tau) := \mathrm{Kl}(1,\tau)\,.$$

We have the following elementary facts:

1) if $\gamma\sigma\neq 0$, then $\mathrm{Kl}(\gamma,\sigma)=\mathrm{Kl}(\gamma\sigma)$,
2) if $\gamma\sigma=0$ but one of γ or σ is $\neq 0$, then $\mathrm{Kl}(\gamma,\sigma)=-1$,
3) if $\gamma=\sigma=0$, then $\mathrm{Kl}(0,0)=q-1$.

We also have the non-elementary fact, due to Weil [Weil],

4) if $\tau\neq 0$, then $|\mathrm{Kl}\,(\tau)|\le 2q^{1/2}$.

Now let us return to our sum $S(\mathbb{F}_q,\psi,\alpha,\beta)$. Summing first on the x variables only, we obtain the formula

$$\begin{aligned} S(\mathbb{F}_q,\psi,\alpha,\beta) = \sum_{t\neq 0} &\mathrm{Kl}(t-1,\alpha(t-1))\,\mathrm{Kl}(t-1,\alpha t-\beta)\\ &\times \mathrm{Kl}(t^{-1}-1,\beta(t^{-1}-1))\,\mathrm{Kl}(t^{-1}-1,\beta t^{-1}-\alpha)\,. \end{aligned}$$

We first isolate the terms in this sum with $t=1$ and with $t=\beta/\alpha$. For $t=1$, the term is (remembering that $\alpha\neq\beta$)

$$\begin{aligned} &\mathrm{Kl}(0,0)\,\mathrm{Kl}(0,\alpha-\beta)\,\mathrm{Kl}(0,0)\,\mathrm{Kl}(0,\beta-\alpha)\\ &\qquad = (q-1)(-1)(q-1)(-1) = (q-1)^2\,. \end{aligned}$$

For $t=\beta/\alpha$, the term is

$$\begin{aligned} &\mathrm{Kl}(\beta/\alpha-1,\beta-\alpha)\,\mathrm{Kl}(\beta/\alpha-1,0)\,\mathrm{Kl}(\alpha/\beta-1,\alpha-\beta)\,\mathrm{Kl}(\alpha/\beta-1,0)\\ &\qquad = \mathrm{Kl}(\beta/\alpha-1,\beta-\alpha)\,\mathrm{Kl}(\alpha/\beta-1,\alpha-\beta)\,, \end{aligned}$$

whose absolute value is bounded by $4q$.

For $t \neq 0$, 1, β/α, the corresponding term is

$$\mathrm{Kl}(\alpha(t-1)^2)\,\mathrm{Kl}((t-1)(\alpha t-\beta))\,\mathrm{Kl}(\beta(t^{-1}-1)^2)\,\mathrm{Kl}((t^{-1}-1)(\beta t^{-1}-\alpha))\,.$$

So it is natural to introduce the modified sum $S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta)$ defined as

$$\sum_{t\neq 0,1,\beta/\alpha} \mathrm{Kl}(\alpha(t-1)^2)\,\mathrm{Kl}((t-1)(\alpha t-\beta))\,\mathrm{Kl}(\beta(t^{-1}-1)^2) \times \mathrm{Kl}((t^{-1}-1)(\beta t^{-1}-\alpha))\,.$$

Thus we have

$$\begin{aligned} S(\mathbb{F}_q,\psi,\alpha,\beta) := S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta) &+ (q-1)^2 \\ &+ \mathrm{Kl}(\beta/\alpha-1,\beta-\alpha)\,\mathrm{Kl}(\alpha/\beta-1,\alpha-\beta)\,, \end{aligned}$$

and so the trivial estimate

$$|S(\mathbb{F}_q,\psi,\alpha,\beta)| \le |S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta)| + q^2+2q+1\,.$$

We now turn to the systematic study of the modified sum. Fix a prime number $l \neq \mathrm{char}(\mathbb{F}_q)$, and an l-adic place λ of the subfield $E := \mathbb{Q}(\exp(2\pi i/p))$ of $\mathbb{C}$. Then we may view ψ as an E_λ-valued nontrivial additive character of $\mathbb{F}_q$, and we can speak of the E_λ-adic Kloosterman sheaf $\mathrm{Kl}_2(\psi;\mathbf{1},\mathbf{1};1,1)$, or just Kl_2 for short, on $\mathbb{G}_m$ over $\mathbb{F}_q$. One knows (cf. [Ka-GKM, 4.1.1]) that Kl_2 is a lisse sheaf of rank 2, pure of weight one, which has nontrivial unipotent local monodromy at zero, is totally wild at ∞ with both ∞-breaks 1/2, and whose trace of Frobenius at any rational point $\gamma \neq 0$ in $\mathbb{F}_q$ is (minus) the Kloosterman sum $\mathrm{Kl}(\gamma)$ in E.

For any curve C over $\mathbb{F}_q$, and any morphism

$$f : C \to \mathbb{G}_m\,,$$

we define

$$\mathrm{Kl}_2(f) := f^*\,\mathrm{Kl}_2\,, \qquad \text{a lisse sheaf of rank 2 on } C\,.$$

By its very definition, the trace of Frobenius on $\mathrm{Kl}_2(f)$ at a rational point t in $C(\mathbb{F}_q)$ is the Kloosterman sum $\mathrm{Kl}(f(t))$. From this point of view, the modified sum may be expressed as follows.

Take for C the open set

$$C := \mathbb{A}^1 - \{0,1,\beta/\alpha\} := \mathrm{Spec}(\mathbb{F}_q[T,1/T(T-1)(\alpha T-\beta)])\,.$$

Consider the four morphisms from C to $\mathbb{G}_m$ given by the four functions

$$\begin{aligned} f_1(T) &:= \alpha(T-1)^2\,, & f_3(T) &:= \beta(T^{-1}-1)^2\,, \\ f_2(T) &:= (T-1)(\alpha T-\beta)\,, & f_4(T) &:= (T^{-1}-1)(\beta T^{-1}-\alpha)\,. \end{aligned}$$

Form the corresponding pullbacks $f_i^*\,\mathrm{Kl}_2 := \mathrm{Kl}_2(f_i)$ on C, and consider their tensor product

$$\mathcal{F} := \mathrm{Kl}_2(f_1)\otimes \mathrm{Kl}_2(f_2)\otimes \mathrm{Kl}_2(f_3)\otimes \mathrm{Kl}_2(f_4)\,.$$

This $\mathcal{F}$ is lisse of rank $2^4 = 16$ on C, and pure of weight 4. By construction, the trace of Frobenius on $\mathcal{F}$ at any t in $C(\mathbb{F}_q)$ is

$$\begin{aligned}&\mathrm{Trace}(\mathrm{Frob}_{\mathbb{F}_q,t}|\mathcal{F})\\&= \mathrm{Kl}(\alpha(t-1)^2)\,\mathrm{Kl}((t-1)(\alpha t-\beta))\,\mathrm{Kl}(\beta(t^{-1}-1)^2)\,\mathrm{Kl}((t^{-1}-1)(\beta t^{-1}-\alpha))\,.\end{aligned}$$

Thus the modified sum $S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta)$ is none other than

$$S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta) = \sum_{t\in C(\mathbb{F}_q)} \mathrm{Trace}(\mathrm{Frob}_{\mathbb{F}_q,t}|\mathcal{F})\,.$$

So Grothendieck's Lefschetz Trace Formula [SGA 4 1/2, Rapport] gives

$$S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta) = \sum_i (-1)^i\ \mathrm{Trace}(\mathrm{Frob}_q|H^i_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F}))\,.$$

Since the curve C is open, and $\mathcal{F}$ is lisse, the only possibly nonzero cohomology groups are those with $i = 1$ and $i = 2$.

If we can show that $H^2_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F}) = 0$, then we will get

$$S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta) = -\,\mathrm{Trace}(\mathrm{Frob}_q|H^1_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F}))\,.$$

As $\mathcal{F}$ is pure of weight 4, the group $H^1_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F})$ is mixed of weight ≤ 5, thanks to [De-Weil II, 3.3.1] and its dimension is $|\chi_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F})|$, so we will get the estimate

$$|S_{\mathrm{modif}}(\mathbb{F}_q,\psi,\alpha,\beta)| \le |\chi_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F})|q^{5/2}\,.$$

(Conversely, since $H^2_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F})$ is pure of weight 6, the truth of the theorem for all finite extensions of $\mathbb{F}_q$ implies the vanishing of $H^2_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F})$.)

Thus it remains to show that

$$H^2_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F}) = 0\,,\qquad |\chi_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F})| \le 64\,.$$

We begin with the calculation of the Euler characteristic $\chi_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F})$. Since $\mathcal{F}$ is lisse on $C := \mathbb{P}^1 - \{0,1,\beta/\alpha,\infty\}$, the Euler–Poincaré formula gives

$$\begin{aligned}\chi_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathcal{F}) &= \chi_{\mathrm{c}}(C\otimes\overline{\mathbb{F}}_q,\mathbb{Q}_l)\,\mathrm{rank}(\mathcal{F})\\&\quad - \mathrm{swan}_0(\mathcal{F}) - \mathrm{swan}_1(\mathcal{F})\\&\quad - \mathrm{swan}_{\beta/\alpha}(\mathcal{F}) - \mathrm{swan}_\infty(\mathcal{F})\\&= -2\,\mathrm{rank}(\mathcal{F}) - \mathrm{swan}_0(\mathcal{F}) - \mathrm{swan}_1(\mathcal{F})\\&\quad - \mathrm{swan}_{\beta/\alpha}(\mathcal{F}) - \mathrm{swan}_\infty(\mathcal{F})\,.\end{aligned}$$

Our main information is that the Kloosterman sheaf Kl_2 is lisse on $\mathbb{G}_m$, a single unipotent Jordan block Unip(2) of dimension 2 at zero, and totally wild at ∞ with both ∞-slopes 1/2. So we can make the following table of information about the representations of the inertia groups at the four "points at ∞" $\{0, 1, \beta/\alpha, \infty\}$ on $C \otimes \overline{\mathbb{F}}_q$ given by the four pullback sheaves $\mathrm{Kl}_2(f_i) := f_i^* \mathrm{Kl}_2$, where

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

sheaf $\rightarrow$ point $\downarrow$	$\mathrm{Kl}_2(f_1)$	$\mathrm{Kl}_2(f_2)$	$\mathrm{Kl}_2(f_3)$	$\mathrm{Kl}_2(f_4)$
$t=0$	trivial	trivial	*wild?*	*wild?*
$t=1$	Unip(2)	Unip(2)	Unip(2)	Unip(2)
$t=\beta/\alpha$	trivial	Unip(2)	trivial	Unip(2)
$t=\infty$	*wild?*	*wild?*	trivial	trivial

Suppose first that the characteristic of $\mathbb{F}_q$ is *odd*. Since the functions f_i are each doubly ramified over ∞, they are each tame over ∞, and hence (cf. [Ka-GKM, 1.14]) each of the entries *wild?* in the above table is the direct sum of two different characters of I, each of which has swan conductor $= 1$. So in particular, we see that

$\mathcal{F}$ is tame at both $t=1$ and at $t=\beta/\alpha$; at both $t=0$ and $t=\infty$, each slope of $\mathcal{F}$ is 0 or 1.

Therefore we have

$$\mathrm{swan}_1(\mathcal{F}) = \mathrm{swan}_{\beta/\alpha}(\mathcal{F}) = 0\,,$$
$$0 \le \mathrm{swan}_0(\mathcal{F}), \quad \mathrm{swan}_\infty(\mathcal{F}) \le \mathrm{rank}(\mathcal{F})\,.$$

Thus the Euler–Poincaré formula

$$-\chi_c(C \otimes \overline{\mathbb{F}}_q, \mathcal{F})$$
$$= 2\,\mathrm{rank}(\mathcal{F}) + \mathrm{swan}_0(\mathcal{F}) + \mathrm{swan}_1(\mathcal{F}) + \mathrm{swan}_{\beta/\alpha}(\mathcal{F}) + \mathrm{swan}_\infty(\mathcal{F})$$

gives the asserted estimate

$$0 \le -\chi_c(C \otimes \overline{\mathbb{F}}_q, \mathcal{F}) \le 4\,\mathrm{rank}(\mathcal{F}) = 64\,.$$

Suppose now that the characteristic of $\mathbb{F}_q$ is *even*. Then we claim that each of the entries *wild?* in the above table is an irreducible representation of I with both slopes $1/2$. Admitting this, we get

$\mathcal{F}$ is tame at both $t=1$ and at $t=\beta/\alpha$; at both $t=0$ and $t=\infty$, each slope of $\mathcal{F}$ is $\le 1/2$.

Therefore we have

$$\mathrm{swan}_1(\mathcal{F}) = \mathrm{swan}_{\beta/\alpha}(\mathcal{F}) = 0\,,$$
$$0 \le \mathrm{swan}_0(\mathcal{F})\,, \quad \mathrm{swan}_\infty(\mathcal{F}) \le (1/2)\,\mathrm{rank}\,(\mathcal{F})\,.$$

Thus the Euler–Poincaré formula

$$\begin{aligned}&-\chi_c(C \otimes \overline{\mathbb{F}}_q, \mathcal{F})\\&\qquad = 2\operatorname{rank}(\mathcal{F}) + \operatorname{swan}_0(\mathcal{F}) + \operatorname{swan}_1(\mathcal{F}) + \operatorname{swan}_{\beta/\alpha}(\mathcal{F}) + \operatorname{swan}_\infty(\mathcal{F})\end{aligned}$$

gives the improved estimate in characteristic two

$$0 \le -\chi_c(C \otimes \overline{\mathbb{F}}_q, \mathcal{F}) \le 3\operatorname{rank}(\mathcal{F}) = 48\,.$$

We now explain how to analyze each of the entries *wild?* in the above table when we are in characteristic two.

First of all, the maps $f_1(T) := \alpha(T-1)^2$ and $f_3(T) := \beta(T^{-1}-1)^2$ are, in different coordinates, simply the absolute Frobenius $X \mapsto X^2$, pulling back by which is not seen by étale sheaves at all. So the I-representations attached to $\mathrm{Kl}(f_1)$ at $t=\infty$ and to $\mathrm{Kl}(f_3)$ at $t=0$ have the same properties of being irreducible with all slopes $1/2$ as does the Kloosterman sheaf Kl_2 as $I(\infty)$-representation.

Next, the two maps

$$f_2(T) := (T-1)(\alpha T - \beta) \quad \text{and} \quad f_4(T) := (T^{-1}-1)(\beta T^{-1} - \alpha)$$

are, in different coordinates, the Artin–Schreier map $\mathcal{P} : X \mapsto X^2 - X$. So we must show that $\mathcal{P}^* \mathrm{Kl}_2$ is $I(\infty)$-irreducible, and has both ∞-slopes $1/2$. For this, it suffices to show that $\mathcal{P}^* \mathrm{Kl}_2$ is $I(\infty)$-irreducible, and has $\operatorname{swan}_\infty(\mathcal{P}^* \mathrm{Kl}_2) = 1$ (since if $\mathcal{P}^* \mathrm{Kl}_2$ is $I(\infty)$-irreducible, it can only have a single slope, repeated with multiplicity; cf. [Ka-GKM, 1.8]). This is a special case of the following lemma, applied with $p=2$ to the $I(\infty)$-representation attached to Kl_2.

LEMMA 2. *Let k be an algebraically closed field of characteristic $p > 0$, $\mathbb{A}^1$ the affine line* $\operatorname{Spec}(k[T])$ *over k, and*

$$\mathcal{P} : \mathbb{A}^1 \to \mathbb{A}^1\,, \qquad X \mapsto X^p - X\,,$$

the Artin–Schreier map. Fix a prime number $l \ne p$, a finite extension E_λ of $\mathbb{Q}_l$, and a finite-dimensional continuous nonzero E_λ-representation M of $I(\infty)$. Denote by $\mathcal{P}^* M$ the $I(\infty)$-representation "upstairs" obtained by pullback, i.e. view M as a representation of $\operatorname{Gal}(k((1/T))^{\mathrm{sep}}/k((1/T)))$, *and restrict it to the normal subgroup of index p which is*

$$\operatorname{Gal}(k((1/T))^{\mathrm{sep}}/k((1/X)))\,, \qquad X^p - X = T\,.$$

Then

1) *if M has all slopes < 1 and is irreducible, $\mathcal{P}^* M$ is irreducible,*
2) *if M has all slopes < 1,* $\operatorname{swan}(\mathcal{P}^* M) = \operatorname{swan}(M)$.

Proof. The key point is that over any E_λ containing the pth roots of unity, we have

$$\mathcal{P}_* E_\lambda \approx E_\lambda \oplus \Big(\bigoplus_{\text{nontrivial}\,\psi} \mathcal{L}_\psi \Big),$$

the internal summand indexed by the $p-1$ nontrivial E_λ-valued additive characters ψ of $\mathbb{F}_p \approx \mathrm{Gal}(k((1/X))/k((1/T)))$. Each $\mathcal{L}_\psi$ has swan $= 1$. By Frobenius reciprocity, the inner product of the representation $\mathcal{P}^* M$ with itself is given by

$$\langle \mathcal{P}^* M, \mathcal{P}^* M \rangle_{\mathrm{up}} = \langle \mathcal{P}_* \mathcal{P}^* M, M \rangle_{\mathrm{down}}\,.$$

By the projection formula, we have

$$\mathcal{P}_* \mathcal{P}^* M = M \otimes \mathcal{P}_* E_\lambda \approx M \oplus \Big(\bigoplus_{\text{nontrivial}\,\psi} M \otimes \mathcal{L}_\psi \Big).$$

If M is irreducible and has all slopes < 1, then each $M \otimes \mathcal{L}_\psi$ is also irreducible, being $M \otimes$(rank one), but each $M \otimes \mathcal{L}_\psi$ has all slopes $= 1$, unlike M itself. Therefore $\langle M, M \otimes \mathcal{L}_\psi \rangle = 0$ for each nontrivial ψ, and hence

$$\langle \mathcal{P}^* M, \mathcal{P}^* M \rangle_{\mathrm{up}} = \langle \mathcal{P}_* \mathcal{P}^* M, M \rangle_{\mathrm{down}} = 1\,,$$

which proves 1).

We next prove 2), by a global argument. Taking the "canonical extension" (cf. [Ka-LG]) of M, we get a lisse E_λ-sheaf $\mathcal{M}$ on $\mathbb{G}_m$ which is tame at zero and whose $I(\infty)$-representation is M. We now consider the finite étale covering of $\mathbb{G}_m$ induced by $\mathcal{P}$:

$$\mathcal{P} : \mathbb{A}^1 - \{\mathbb{F}_p\} \to \mathbb{G}_m\,.$$

The sheaf $\mathcal{P}^*\mathcal{M}$ is lisse on $\mathbb{A}^1 - \{\mathbb{F}_p\}$. Since $\mathcal{M}$ is tame at zero, $\mathcal{P}^*\mathcal{M}$ is tame at each point of $\mathbb{F}_p$, so the Euler–Poincaré formula gives

$$\chi_{\mathrm{c}}(\mathbb{A}^1 - \{\mathbb{F}_p\}, \mathcal{P}^*\mathcal{M}) = (1-p)\,\mathrm{rank}(\mathcal{M}) - \mathrm{swan}_\infty(\mathcal{P}^*\mathcal{M})\,.$$

But we also have

$$\begin{aligned}\chi_{\mathrm{c}}(\mathbb{A}^1 - \{\mathbb{F}_p\}, \mathcal{P}^*\mathcal{M}) &= \chi_{\mathrm{c}}(\mathbb{G}_m, \mathcal{P}_*\mathcal{P}^*\mathcal{M}) \\ &= \chi_{\mathrm{c}}(\mathbb{G}_m, \mathcal{M}) + \sum_{\text{nontrivial}\,\psi} \chi_{\mathrm{c}}(\mathbb{G}_m, \mathcal{M} \otimes \mathcal{L}_\psi)\,.\end{aligned}$$

Since $\mathcal{M}$ is tame at zero, and has all ∞-slopes < 1, each $\mathcal{M} \otimes \mathcal{L}_\psi$ is tame at zero and has all ∞-slopes $= 1$, so we find

$$\chi_{\mathrm{c}}(\mathbb{G}_m, \mathcal{M}) = -\,\mathrm{swan}_\infty(\mathcal{M})\,,$$
$$\chi_{\mathrm{c}}(\mathbb{G}_m, \mathcal{M} \otimes \mathcal{L}_\psi) = -\,\mathrm{rank}(\mathcal{M}) \quad \text{for each nontrivial } \psi\,.$$

So we get

$$\begin{aligned}\chi_{\mathrm{c}}(\mathbb{A}^1 - \{\mathbb{F}_p\}, \mathcal{P}^*\mathcal{M}) &= \chi_{\mathrm{c}}(\mathbb{G}_m, \mathcal{P}_*\mathcal{P}^*\mathcal{M}) \\ &= -\,\mathrm{swan}_\infty(\mathcal{M}) - (p-1)\,\mathrm{rank}(\mathcal{M})\,.\end{aligned}$$

Comparing this with the original formula

$$\chi_c(\mathbb{A}^1 - \{\mathbb{F}_p\}, \mathcal{P}^*\mathcal{M}) = (1-p)\operatorname{rank}(\mathcal{M}) - \operatorname{swan}_\infty(\mathcal{P}^*\mathcal{M})$$

gives

$$\operatorname{swan}(\mathcal{P}^* M) = \operatorname{swan}_\infty(\mathcal{P}^*\mathcal{M}) = \operatorname{swan}_\infty(\mathcal{M}) = \operatorname{swan}(M)\,,$$

provided only that M has all slopes < 1. ∎

We now turn to proving that $H^2_c(C \otimes \overline{\mathbb{F}}_q, \mathcal{F}) = 0$. For this it suffices to show that the sheaf $\mathcal{F}$ on $C \otimes \overline{\mathbb{F}}_q$, viewed as a representation of $\pi_1 := \pi_1(C \otimes \overline{\mathbb{F}}_q, \overline{\eta})$, is absolutely irreducible, since $H^2_c(C \otimes \overline{\mathbb{F}}_q, \mathcal{F})$ is, up to a Tate twist, the coinvariants of this representation. We will prove a more precise result. Recall that attached to any lisse E_λ-sheaf $\mathcal{G}$ on a connected smooth variety X over an algebraically closed field is the algebraic group G_{geom} over E_λ defined (with reference to a geometric point x in X) as the Zariski closure of $\pi_1(X, x)$ in $\operatorname{Aut}(\mathcal{G}_x)$. Recall (cf. [Ka-GKM, 11.1]) that for the Kloosterman sheaf Kl_2 on $\mathbb{G}_m$, the group G_{geom} is known to be $\mathrm{SL}(2)$.

LEMMA 3. 1) *For the direct sum sheaf* $\mathrm{Kl}_2(f_1) \oplus \mathrm{Kl}_2(f_2) \oplus \mathrm{Kl}_2(f_3) \oplus \mathrm{Kl}_2(f_4)$ *on* $C \otimes \overline{\mathbb{F}}_q$, *the group* G_{geom} *is the four-fold product* $\mathrm{SL}(2) \times \mathrm{SL}(2) \times \mathrm{SL}(2) \times \mathrm{SL}(2)$, *acting as the direct sum* $\bigoplus_i \mathrm{std}_2(i)$ *of the standard two-dimensional representations of the four factors.*

2) *For the sheaf* $\mathcal{F}$, G_{geom} *is the image of* $\mathrm{SL}(2) \times \mathrm{SL}(2) \times \mathrm{SL}(2) \times \mathrm{SL}(2)$ *in* $\mathrm{SL}(16)$, *acting as the tensor product* $\bigotimes_i \mathrm{std}_2(i)$ *of the standard two-dimensional representations of the four factors.*

3) $\mathcal{F}$ *is absolutely irreducible as a representation of* $\pi_1(C \otimes \overline{\mathbb{F}}_q, \overline{\eta})$, *and remains so when pulled back to any finite étale connected nonempty covering of* $C \otimes \overline{\mathbb{F}}_q$.

Proof. This will be a simple application of the Goursat–Kolchin–Ribet criterion (cf. [Ka-ESDE, 1.8.2]). The sheaf Kl_2 has $G_{\mathrm{geom}} = \mathrm{SL}(2)$, a connected group. So each pullback sheaf $f_i^* \mathrm{Kl}_2 := \mathrm{Kl}_2(f_i)$ itself has its $G_{\mathrm{geom}} = \mathrm{SL}(2)$. In order to show that for the direct sum $\bigoplus_i \mathrm{Kl}_2(f_i)$ G_{geom} is equal to the full product of four copies of $\mathrm{SL}(2)$ (it is trivially a subgroup of this product) it suffices by the Goursat–Kolchin–Ribet criterion to show that for any two indices $i \neq j$, and any lisse rank one sheaf $\mathcal{L}$ on $C \otimes \overline{\mathbb{F}}_q$, there exists no isomorphism between $\mathrm{Kl}_2(f_i)$ and $\mathrm{Kl}_2(f_j) \otimes \mathcal{L}$. We will verify this by looking at the representations of the inertia groups at the four points $\{0, 1, \beta/\alpha, \infty\}$.

Let us say that a representation ϱ of a group G on a vector space V over a field F is *scalar* if G acts on V by homotheties, i.e., if there exists a character $\chi : G \to F^\times$ such that for v in V, we have $\varrho(g)(v) = \chi(g)v$.

If there exists an isomorphism between $\mathrm{Kl}_2(f_i)$ and $\mathrm{Kl}_2(f_j) \otimes \mathcal{L}$, then restricted to any inertia group I, $\mathrm{Kl}_2(f_i)$ and $\mathrm{Kl}_2(f_j)$ are either both scalar or both nonscalar.

Now let us return to the earlier table which gave the behaviour of the representations of the inertia groups at the four "points at ∞" $\{0, 1, \beta/\alpha, \infty\}$ on $C \otimes \overline{\mathbb{F}}_q$ given by the four pullback sheaves $\mathrm{Kl}_2(f_i) := f_i^* \mathrm{Kl}_2$. Each of the representations marked "trivial" is of course scalar. Each marked Unip(2) is nonscalar. Each marked *wild?* is nonscalar, in odd characteristic because the direct sum of two distinct characters, and in characteristic two because irreducible of dimension > 1. So our table gives the following table of "scalarity" versus "nonscalarity".

sheaf $\rightarrow$ point $\downarrow$	$\mathrm{Kl}_2(f_1)$	$\mathrm{Kl}_2(f_2)$	$\mathrm{Kl}_2(f_3)$	$\mathrm{Kl}_2(f_4)$
$t = 0$	scalar	scalar	nonscalar	nonscalar
$t = 1$	nonscalar	nonscalar	nonscalar	nonscalar
$t = \beta/\alpha$	scalar	nonscalar	scalar	nonscalar
$t = \infty$	nonscalar	nonscalar	scalar	scalar

If there existed an isomorphism between $\mathrm{Kl}_2(f_i)$ and $\mathrm{Kl}_2(f_j) \otimes \mathcal{L}$, then the columns of this table for $\mathrm{Kl}_2(f_i)$ and $\mathrm{Kl}_2(f_j)$ would agree. But visibly all four columns are distinct (indeed already the bottom two entries, giving the behaviours at β/α and ∞, separate the four columns). This proves 1). Once 1) is proven, 2) is obvious, and shows that $\mathcal{F}$ is "Lie-irreducible", which is 3). ∎

As explained above, 3) implies the vanishing of $H^2_c(C \otimes \overline{\mathbb{F}}_q, \mathcal{F})$. This concludes the proof of Theorem 1.

References

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

[SGA] A. Grothendieck *et al.*, *Séminaire de Géométrie Algébrique du Bois-Marie*, SGA 1, SGA 4, Parts I, II, and III, SGA 4 1/2, SGA 5, SGA 7, Parts I and II, Lecture Notes in Math. 224, 269–270–305, 569, 589, 288–340, Springer, Berlin 1971 to 1977.

[Ka-ESDE, 7.4] N. Katz, *Exponential Sums and Differential Equations*, Ann. of Math. Stud. 124, Princeton Univ. Press, 1990.

[Ka-GKM] —, *Gauss Sums, Kloosterman Sums and Monodromy Groups*, Ann. of Math. Stud. 116, Princeton Univ. Press, 1988.

[Ka-LG] —, *Local to global extensions of representations of fundamental groups*, Ann. Inst. Fourier (Grenoble) 36 (4) (1986), 59–106.

[Weil] A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U.S.A. 34 (1948), 204–207.

Etienne Fouvry

UNIVERSITÉ DE PARIS-SUD
MATHÉMATIQUE BÂT. 425
91405 ORSAY, FRANCE

Henryk Iwaniec

DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903
U.S.A.

Nicholas Katz

DEPARTMENT OF MATHEMATICS
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
U.S.A.

Received on 15.3.1991
and in revised form on 12.6.1991 (2127)