

## On the estimation of certain exponential sums

by

E. BOMBIERI (Princeton, N.J.) and S. SPERBER (Minneapolis, Minn.)

**I.** Let  $k = \mathbb{F}_q$  be a finite field of characteristic  $p$ , let  $k_n = \mathbb{F}_{q^n}$  be the unique extension of  $k$  of degree  $n$ , let  $V$  be a quasi-projective variety defined over  $k$  and let  $f \in k(V)$  be a rational function on  $V$ , also defined over  $k$ . As usual,  $V(k_n)$  will denote the set of points of  $V$  defined over  $k_n$ . We also denote by  $K$  an algebraic closure of  $k$  and for a scheme  $X$  over  $k$  we denote by  $X_K$  the scheme  $X_K = X \otimes K$ , i.e.  $X$  after base change from  $k$  to  $K$ .

In what follows we shall assume that  $f$  is defined everywhere on  $V$  and has no poles on  $V$ , so that  $f : V \rightarrow \mathbb{A}_k^1$  is a morphism. Let  $\psi_0$  be a non-trivial additive character on  $\mathbb{F}_p$  and let  $\psi_n = \psi_0 \circ \text{Tr}_{k_n/\mathbb{F}_p}$  be the corresponding character on  $k_n$ ; every additive character on  $k_n$  can be written as  $\psi_n(lx)$  for some  $l \in k_n$ . Also, we shall write  $\psi$  instead of  $\psi_1$  for the character induced on  $k$ .

The exponential sum associated with  $V$ ,  $f$ ,  $k_n$  and the character  $\psi_n$  is by definition

$$S_n(V, f, \psi) = \sum_{x \in V(k_n)} \psi_n(f(x)).$$

Since the sum is well-defined as soon as  $f$  is defined over  $k_n$  rather than  $k$ , we shall also write  $S_n(V, lf, \psi)$  instead of  $S_1(V \otimes k_n, lf, \psi_n)$ , for  $l \in k_n$ .

Let us assume that  $f$  is defined over  $k$ . By a well-known theorem (rationality of the associated Artin  $L$ -series) we have

$$S_n(V, f, \psi) = \sum_{i=1}^r \alpha_i(f)^n - \sum_{j=1}^s \beta_j(f)^n$$

for suitable algebraic integers  $\alpha_i(f)$ ,  $\beta_j(f)$  called the characteristic roots of the exponential sum; we denote by  $\varrho(V, f) = r + s$  the total number of characteristic roots of  $S_n$ . If  $d = \dim V$  then by a celebrated theorem of Deligne [D], each characteristic root has absolute value  $q^{w/2}$  for a certain integer  $w$ ,  $0 \leq w \leq 2d$ , called the weight of the characteristic root. Moreover, by [D] any characteristic root has the same absolute value as any of its

conjugates over the rational field  $\mathbb{Q}$ , a fact of capital importance for our considerations.

Deligne's theorem implies that if  $r_w(f)$  and  $s_w(f)$  denote the number of characteristic roots  $\alpha_i(f)$  and  $\beta_j(f)$  of weight  $w$  then

$$|S_n(V, f, \psi)| \leq \sum_{w=0}^{2d} (r_w(f) + s_w(f)) q^{nw/2};$$

we always have  $s_{2d}(f) = 0$ . The sum  $S_n(V, 0, \psi)$  is simply the number of rational points of  $V$  over  $k_n$ , which is traditionally denoted by  $N_n(V)$ . If  $f$  is a constant  $c$  then  $S_n(V, c, \psi) = \psi(c)^n N_n(V)$ , therefore then  $r_{2d}(f) \geq 1$ , and  $r_{2d}(f) = 1$  if in addition  $V$  is geometrically irreducible.

Suppose now that  $V$  is geometrically irreducible. We have  $\text{Tr}_{k_n/\mathbb{F}_p}(y) = 0$  if and only if  $y = x^p - x$  with  $x \in k_n$ . This implies

$$S_n(V, f, \psi) = S_n(V, f + h^p - h, \psi)$$

whenever  $h \in k(V)$ . In particular, if  $V$  is geometrically irreducible we have  $r_{2d}(f) = 1$  if and only if  $f = h^p - h + c$ , with  $c \in k$  and  $h \in k(V)$ .

We also write  $\varrho(V)$  for the total number of characteristic roots associated with  $N_n(V)$ , i.e.  $\varrho(V) = \varrho(V, 0)$ .

The problem of estimating the sums  $S_n(V, f, \psi)$  is of considerable importance. The connection between exponential sums and  $L$ -series was made explicit by Hasse [Ha]. Weil [W], by proving the Riemann Hypothesis and the Artin Conjecture for  $L$ -series over a function field of dimension 1 of positive characteristic, essentially solved the problem of finding optimal estimates of exponential sums in case  $\dim V = 1$ , and gave the explicit example of Kloosterman sums. Lang and Weil [LW] and Nisnevich [N] used a slicing technique to obtain bounds in higher dimensions, and very general explicit bounds for  $\dim V = 1$  are given in [B1]. Once Deligne's theorem became available, the problem was reduced to calculating the numbers  $r_w(f)$  and  $s_w(f)$ , with the goal of showing that there are no characteristic roots of large weight, the usually optimal result being that  $r_w(f) = s_w(f) = 0$  if  $w > d$ . This would give the bound

$$S_n(V, f, \psi) = O((q^n)^{d/2})$$

which one expects on probabilistic grounds.

An upper bound for  $\sum (r_w(f) + s_w(f))$  was given in [B2], and in case  $\dim V = 2$  Hooley [H] gave necessary and sufficient conditions for the vanishing of  $r_3(f)$  and  $s_3(f)$ . It should be stressed that the conditions in Hooley's theorem are expressed in simple geometric terms and are readily checked in the specific situations which may arise in applications.

A finer analysis of characteristic roots in each weight, especially if  $\dim V > 3$ , requires much deeper considerations from algebraic geometry.

This has been explored by Katz and Laumon [K], [KL] using tools from Fourier transform in  $l$ -adic cohomology. In particular, they show that in suitably generic situations the expected optimal estimates are satisfied.

In another direction, Adolphson and Sperber [AS1] extended the Dwork theory so as to obtain information about the vanishing of the numbers  $r_w(f)$  and  $s_w(f)$ . Their conditions, although somewhat restrictive, are easy to check and provide us with a useful class of exponential sums for which optimal bounds are available.

The aim of this paper is to provide a set of necessary and sufficient conditions, much in the spirit of Hooley's theorem [H], for the validity of sharp estimates in case  $\dim V = 3$ . More generally, we give a set of geometric conditions which imply that the given exponential sum has no characteristic roots of weight  $w \geq 2d - 2$ ; we expect that it should be relatively easy for the non-expert to verify them, and therefore we hope that our result will be of some use in practice.

For a geometrically irreducible variety  $X$  let  $\text{Alb}(X)$  denote its Albanese variety, as defined in [L]. Note that  $\text{Alb}(X)$  is a birational invariant of  $X$ . Let us also recall that a quasi-projective variety  $V$  in projective space  $\mathbb{P}^n$  may be defined set-theoretically by finitely many equations  $F_i(\mathbf{x}) = 0$  together with a finite collection of forms  $G_j(\mathbf{x})$  such that for every  $\mathbf{x} \in V$  we have  $G_j(\mathbf{x}) \neq 0$  for some  $j$ , which may depend on the point  $\mathbf{x}$ ; we say that the forms  $F_i, G_j$  form a *presentation of  $V$* . Then a function  $f$  on  $V$  may be written as  $f = P_m(\mathbf{x})/Q_m(\mathbf{x})$ , for a suitable finite collection of forms  $P_m(\mathbf{x}), Q_m(\mathbf{x})$  such that at least one of the denominators  $Q_m$  is not 0 at any given point of  $V$ ; we say that the forms  $P_m, Q_m$  form a *presentation of  $f$* . Now we can state:

**THEOREM 1.** *Let  $V$  be a quasi-projective variety over  $k$  of dimension  $d \geq 3$ , let  $f \in k(V)$  be a rational function on  $V$  defined over  $k$  and let  $\psi$  be a non-trivial additive character of  $k$ . Suppose that:*

- (i)  $V$  is geometrically irreducible;
- (ii) the rational function  $f$  is defined everywhere on  $V$  and has no poles on  $V$ ;
- (iii) every fiber  $V_\lambda = f^{-1}(\lambda) \otimes K$  consists of precisely one non-empty irreducible component of dimension  $d - 1$ , plus possibly components of lower dimension, for every  $\lambda \in K$ . In particular, the map  $f : V_K \rightarrow \mathbb{A}_K^1$  is onto;
- (iv) if  $V_\lambda^{d-1}$  is the unique component of dimension  $d - 1$  of  $V_\lambda$ , then for all but finitely many  $\lambda \in K$  we have  $V_\lambda = V_\lambda^{d-1}$  and  $\dim \text{Alb}(V_\lambda^{d-1}) = \dim \text{Alb}(V)$ .

Then for  $p > D(V, f)$  we have the bound

$$|S_n(V, f, \psi)| \leq \varrho(V, f)(q^n)^{d-3/2},$$

where the constants  $D(V, f)$  and  $\varrho(V, f)$  depend only on the embedding dimension of  $V$ , the degrees and number of forms appearing in a presentation of  $V$ , and the degrees and number of forms appearing in a presentation of  $f$ .

Remark. The theorem is true, but not interesting, if  $d = 2$ , because in that case condition (iv) forces  $V$  to be birationally equivalent over  $K$  to a product, a factor of which is  $\mathbb{A}_K^1$ . We also note that in case  $d = 3$  we get

$$|S_n(V, f, \psi)| \leq \varrho(V, f)q^{3n/2},$$

which is usually sharp, except possibly for the value of  $\varrho(V, f)$ ; this is indeed the main motivation for this paper.

If we assume  $p > D(V, f)$  (possibly after increasing  $D(V, f)$  to a new function with the same general dependence on  $V$  and  $f$ ), then (iii) already implies that  $f$  is a separable morphism and  $V_\lambda = V_\lambda^{d-1}$  for all but finitely many  $\lambda \in K$ . This can be seen by replacing  $V$  by its normalization and applying Bertini's theorem to the linear pencil associated with  $f$ , which we may because we suppose that  $p$  is sufficiently large; we leave the details to the reader.

The first part of assumption (iv) implies, in the notation of Lang [L], Ch. VIII, that the sequence

$$V_\lambda \xrightarrow{j} V_K \xrightarrow{f} \mathbb{A}_K^1$$

with  $j$  the inclusion is generically exact, whence the induced map  $j_* : \text{Alb}(A_\lambda) \rightarrow \text{Alb}(V_K)$  is surjective for all but finitely many  $\lambda \in K$ . In particular, we have  $\dim \text{Alb}(V_\lambda) \geq \dim \text{Alb}(V)$  for almost all  $\lambda$ 's.

The following special case is worth mentioning.

COROLLARY. *The conclusion of Theorem 1 still holds if in addition to (i), (ii), (iii) we have*

(iv)' *if  $V'_\lambda$  is a desingularization of the projective closure of  $V_\lambda^{d-1}$  we have  $H^1(V'_\lambda, \mathcal{O}_{V'_\lambda}) = 0$  for all but finitely many  $\lambda \in K$ .*

Proof. In fact, for a non-singular projective variety  $X$  over  $K$  the vector space  $H^1(X, \mathcal{O}_X)$  is the Zariski tangent space at the origin of the group scheme  $\text{Pic}^0(X)$ , hence the hypothesis implies that  $\text{Pic}^0(V'_\lambda)$  is a point. By duality,  $\dim \text{Alb}(V'_\lambda) = 0$ . Since

$$\dim \text{Alb}(V'_\lambda) = \dim \text{Alb}(V_\lambda^{d-1}) \geq \dim \text{Alb}(V)$$

for almost every  $\lambda$ , we see that condition (iv)' implies condition (iv) and the result follows. ■

We shall also prove that if (i) and (ii) of Theorem 1 hold then conditions (iii) and (iv) are necessary for the conclusion of the theorem, at least if we want a result valid for all functions  $lf$ ,  $l \in K^*$ . At the end we shall

treat explicitly some special cases in dimension 3 both by an application of Theorem 1 and, by comparison, by appealing to the results of Adolphson and Sperber [AS1].

**Acknowledgements.** We wish to thank P. Deligne and W. Messing for several useful conversations and suggestions, and we wish to thank the referee for pointing out several inaccuracies and suggesting corrections and improvements. We wish also to point out that any inaccuracy or incorrect statement in this paper is due solely to the authors.

**II.** For the proof of Theorem 1 we need some auxiliary results. The first result, which was conjectured by Lang and Weil [LW] in the non-singular case, is certainly known to the experts, but we have been unable to locate a formal proof in the literature.

**PROPOSITION.** *Let  $X$  be a geometrically irreducible quasi-projective variety of dimension  $d$ , defined over  $k$ . Then  $H^1(\text{Alb}(X_K), \mathbb{Q}_l(d))^\sim$  can be identified with the weight  $2d - 1$  component in the weight filtration of  $H_c^{2d-1}(X_K, \mathbb{Q}_l)$ , where  $H_c^*$  denotes cohomology with compact supports.*

We also have

$$N_n(X) = (q^n)^d - (q^n)^{d-1} \text{Tr}(\phi^n | \text{Alb}(X)) + O((q^n)^{d-1})$$

where  $\phi$  denotes the  $q$ -th power Frobenius and  $\text{Tr}$  is the trace in the endomorphism algebra of  $\text{Alb}(X)$ . The constant implied in the  $O(\ )$  term is bounded in terms of the embedding dimension of  $X$  and the degrees and number of forms appearing in a presentation of  $X$ .

**Proof.** If  $X'$  is birationally equivalent to  $X$  over  $k$  then  $X$  and  $X'$  may be identified in a non-empty Zariski open set, therefore

$$N_n(X') = N_n(X) + O((q^n)^{d-1}).$$

This shows that in the proof of the proposition we may replace  $X$  by any model birationally equivalent to  $X$  over  $k$ .

We note that the second statement of the proposition is an easy consequence of the first. In fact, the Lefschetz fixed point formula gives

$$N_n(X) = \sum_{i=0}^{2d} (-1)^i \text{Tr}(\phi^n | H_c^i(X_K, \mathbb{Q}_l))$$

and, by Deligne's theorem [D], the group  $H_c^i$  is mixed of weight at most  $i$  and at least  $2(i - d)$  if  $i \geq d$ . Since  $X$  is geometrically irreducible the term with  $i = 2d$  contributes  $(q^n)^d$ , the term with  $i = 2d - 1$  contributes  $-(q^n)^{d-1} \text{Tr}(\phi^n | \text{Alb}(X)) + O((q^n)^{d-1})$ , and all other terms contribute  $O((q^n)^{d-1})$  at the most. It remains to prove the uniformity statement about the  $O(\ )$  term, but this is an immediate consequence of the fact that this

term is majorized by  $\varrho(X)(q^n)^{d-1}$  and the bounds for  $\varrho(X)$  stated in [B2] (see also Theorem 2 of this paper).

We begin by considering separately the cases  $d = 1$  and  $d = 2$ . If  $d = 1$  we can use normalization and assume that  $X$  is non-singular and projective over  $k$ . Then the statement of the proposition is a well-known theorem of Weil (see e.g. [L], Ch. VI, pp. 163–164). If  $d = 2$ , one could proceed in a similar way, using Abhyankar’s resolution of singularities for an algebraic surface over a perfect field, and using Deligne’s results [D]. However, the following argument, indicated to us by Deligne, avoids the use of Abhyankar’s difficult theorem.

Since the result is birational we may assume that  $X_K$  is a normal projective surface defined over  $K$ , hence with isolated singularities only. We may assume, changing projective embedding if necessary, that we have a nice hyperplane pencil  $\pi : X_K \rightarrow \mathbb{P}_K^1$  on  $X_K$ , i.e. such that almost all fibers  $C_\lambda = \pi^{-1}(\lambda)$  are non-singular hyperplane sections of  $X_K$ . By removing a finite set of points  $S$  on the base and setting  $U = X_K - \pi^{-1}(S)$  we obtain the smooth projective morphism between affine smooth varieties

$$\pi : U \rightarrow \mathbb{P}_K^1 - S.$$

The Leray spectral sequence for cohomology with compact supports gives the exact sequence (note that  $H_c^0(\mathbb{P}_K^1 - S, R^2\pi_!\mathbb{Q}_l)$  vanishes, as one sees by Poincaré duality and the fact that  $\mathbb{P}_K^1 - S$  is affine):

$$0 \rightarrow H_c^2(\mathbb{P}_K^1 - S, R^1\pi_!\mathbb{Q}_l) \rightarrow H_c^3(U, \mathbb{Q}_l) \rightarrow H_c^1(\mathbb{P}_K^1 - S, R^2\pi_!\mathbb{Q}_l),$$

which we analyze as follows.

For the last term  $H_c^1(\mathbb{P}_K^1 - S, R^2\pi_!\mathbb{Q}_l)$  we note that  $R^2\pi_!\mathbb{Q}_l \cong \mathbb{Q}_l(-1)$  because every fiber is an irreducible projective curve. Thus we deal with  $H_c^1(\mathbb{P}_K^1 - S, \mathbb{Q}_l(-1))$  and characteristic roots of Frobenius on this group have weight at most 2, as one verifies using the exact sequence

$$\dots \rightarrow H^0(S, \mathbb{Q}_l) \rightarrow H_c^1(\mathbb{P}_K^1 - S, \mathbb{Q}_l) \rightarrow H^1(\mathbb{P}^1, \mathbb{Q}_l) \rightarrow \dots$$

and the vanishing of  $H^1(\mathbb{P}_K^1, \mathbb{Q}_l)$ . Thus the weight 3 component in the weight filtration of  $H_c^3(U, \mathbb{Q}_l)$  can be identified with the weight 3 component of the weight filtration of the group  $H_c^2(\mathbb{P}_K^1 - S, R^1\pi_!\mathbb{Q}_l)$ .

Now we have the following isomorphisms:

$$\begin{aligned} H_c^2(\mathbb{P}_K^1 - S, (R^1\pi_!\mathbb{Q}_l)^\vee) &\cong H^0(\mathbb{P}_K^1 - S, R^1\pi_!\mathbb{Q}_l)(1) \quad (\text{by duality}), \\ H^0(\mathbb{P}_K^1 - S, R^1\pi_!\mathbb{Q}_l)(1) &\cong H^0(\mathbb{P}_K^1 - S, R^1\pi_!\mathbb{Q}_l(1)), \\ H^0(\mathbb{P}_K^1 - S, R^1\pi_!\mathbb{Q}_l(1)) &\cong H^1(C_\lambda, \mathbb{Q}_l(1))^{\pi_1(\mathbb{P}_K^1 - S)}, \\ H^1(C_\lambda, \mathbb{Q}_l(1))^{\pi_1(\mathbb{P}_K^1 - S)} &\cong [T_l(\text{Pic}^0(C_\lambda)) \otimes \mathbb{Q}_l]^{\pi_1(\mathbb{P}_K^1 - S)}, \\ T_l(\text{Pic}^0(C_\lambda)) \otimes \mathbb{Q}_l &\cong (T_l(\text{Alb}(C_\lambda)) \otimes \mathbb{Q}_l)^\vee(1) \quad (\text{by duality}), \\ [(T_l(\text{Alb}(C_\lambda)) \otimes \mathbb{Q}_l)^\vee(1)]^{\pi_1(\mathbb{P}_K^1 - S)} &\cong [(T_l(\text{Alb}(C_\lambda)) \otimes \mathbb{Q}_l)_{\pi_1(\mathbb{P}_K^1 - S)}(-1)]^\vee. \end{aligned}$$

Since the co-invariants of  $C_\lambda$  are also given by  $\text{Alb}(X_K)$  we can replace this last term by

$$(T_l(\text{Alb}(X_K)) \otimes \mathbb{Q}_l(-1))^\vee \cong H^1(\text{Alb}(X_K), \mathbb{Q}_l(1)).$$

This proves the first statement of the proposition in case  $d = 2$ .

If  $d \geq 3$  we cannot use the preceding argument without modification, since a normalization of  $X$  may be singular in codimension 2 and therefore a generic hyperplane section may also be singular. However, in this case we may proceed in an alternative way, by induction on the dimension  $d$ .

Without loss of generality we may assume that  $X$  is an affine subvariety of  $\mathbb{A}_k^N$ , not contained in any hyperplane and we may also suppose that the proposition holds for varieties of dimension  $d - 1$ . Let  $Y_u$  be a generic hyperplane section of  $X$ , thus defined over  $k(u)$ , let  $j_u : Y_u \rightarrow X$  be the inclusion and let

$$(j_u)_* : \text{Alb}(Y_u) \rightarrow \text{Alb}(X)$$

be the corresponding canonical homomorphism induced on the Albanese varieties. Since  $d \geq 3$  we may apply once more Chow's result ([L], Ch. VIII, Th. 5, p. 210) and conclude this time that  $(j_u)_*$  is a purely inseparable isogeny. The space of hyperplane sections of  $\mathbb{A}_k^N$  is identified, via Plücker coordinates, with  $U = \mathbb{P}_k^N - \text{point}$  and from the preceding result we obtain that there is a non-empty Zariski open set  $U_0 \subset U$  such that for  $u \in U_0(K)$  the variety  $Y_u$  is absolutely irreducible and the map  $(j_u)_* : \text{Alb}(Y_u) \rightarrow \text{Alb}(X)$  is a purely inseparable isogeny. In particular, if in addition  $u \in U_0(k_n)$  we see that

$$\text{Tr}(\phi^n | \text{Alb}(Y_u)) = \text{Tr}(\phi^n | \text{Alb}(X)).$$

Now we compute the number

$$\mathcal{N}_n = \sum_{u \in U(k_n)} N_n(Y_u)$$

of points lying in all hyperplanes  $Y_u$  in two different ways. The first method is as follows.

Since  $Y_u(k_n) \subset X(k_n)$  we may simply count points of  $X(k_n)$ , attaching to them a multiplicity equal to the number of hyperplanes over  $k_n$  through a given point. The number of such hyperplanes is  $N_n(\mathbb{P}_k^{N-1})$ , therefore

$$\mathcal{N}_n = N_n(\mathbb{P}_k^{N-1})N_n(X) = (q^n)^{N-1}N_n(X) + O((q^n)^{N+d-2}).$$

The second method computes  $N_n(Y_u)$  directly. We have

$$\sum_{u \in U(k_n)} N_n(Y_u) = \sum_{u \in U_0(k_n)} N_n(Y_u) + \sum_{u \notin U_0(k_n)} N_n(Y_u).$$

Since the complement of  $U_0$  in  $U$  has dimension at most  $N - 1$ , the contribution of the second sum is clearly  $O((q^n)^{N+d-2})$ . For the first sum, we use

the induction hypothesis and the preceding evaluation of  $\text{Tr}(\phi^n | \text{Alb}(Y_u))$  to obtain

$$N_n(Y_u) = (q^n)^{d-1} - (q^n)^{d-2} \text{Tr}(\phi^n | \text{Alb}(X)) + O((q^n)^{d-2})$$

for  $u \in U_0(k_n)$ . Note again that the constant involved in the  $O(\ )$  term can be estimated independently of  $u$ , for example by appealing to the bounds in [B2] for the quantity  $\varrho(Y_u)$ . It follows that

$$\begin{aligned} & \sum_{u \in U_0(k_n)} N_n(Y_u) \\ &= N_n(U_0)((q^n)^{d-1} - (q^n)^{d-2} \text{Tr}(\phi^n | \text{Alb}(X))) + O((q^n)^{N+d-2}). \end{aligned}$$

We have  $N_n(U_0) = N_n(\mathbb{P}_k^N) + O((q^n)^{N-1})$  and we conclude easily

$$\mathcal{N}_n = (q^n)^{N+d-1} - (q^n)^{N+d-2} \text{Tr}(\phi^n | \text{Alb}(X)) + O((q^n)^{N+d-2});$$

the induction step, and the proof of the proposition, follows from comparing the two estimates we have obtained for  $\mathcal{N}_n$ . ■

Following Hooley, we define

$$\mathcal{S}_n = \sum_{l \in k_n^*} |S_n(V, lf, \psi)|^2$$

to be the second moment of the exponential sum over  $l \in k_n^*$ . Let  $V_\lambda$  be the slice of  $V$  by  $f$ , namely  $V_\lambda = f^{-1}(\lambda)$ . It is clear that  $V_\lambda$  is defined over the field  $k(\lambda)$ , and that

$$S_n(V, lf, \psi) = \sum_{\lambda \in k_n} \psi_n(l\lambda) N_n(V_\lambda).$$

LEMMA 1. *Let  $N_n^* = q^{-n} N_n(V)$ . Then*

$$\mathcal{S}_n = q^n \sum_{\lambda \in k_n} (N_n(V_\lambda) - N_n^*)^2.$$

PROOF. Immediate from the preceding formula for  $S_n(V, lf, \psi)$  and orthogonality of characters. ■

Now we proceed to obtain an upper bound for the second moment  $\mathcal{S}_n$ . Let  $W = (f \times f)^{-1}(\Delta)$  be the pull-back of the diagonal of  $\mathbb{A}_k^1 \times \mathbb{A}_k^1$  in the product  $V \times V$ , i.e. the subvariety of  $V \times V$  defined by  $f(x) = f(y)$  with  $(x, y) \in V \times V$ .

LEMMA 2. *With the hypotheses of Theorem 1 we have*

$$\mathcal{S}_n \leq (\varrho(W) + \varrho(V)^2)(q^n)^{2d-2}.$$

PROOF. We begin by proving a bound  $\mathcal{S}_n = O((q^n)^{2d-2})$ , for some unspecified constant involved in the  $O(\ )$  symbol.



By the preceding proposition, since  $V_\lambda^{d-1}$  and  $V$  are geometrically irreducible we have, for  $\lambda \in k_n$ ,

$$\begin{aligned} N_n(V_\lambda) &= (q^n)^{d-1} - (q^n)^{d-2} \operatorname{Tr}(\phi^n | \operatorname{Alb}(V_\lambda^{d-1})) + O((q^n)^{d-2}), \\ N_n(V) &= (q^n)^d - (q^n)^{d-1} \operatorname{Tr}(\phi^n | \operatorname{Alb}(V)) + O((q^n)^{d-1}). \end{aligned}$$

Note that since  $\varrho(X)$  admits an upper bound which depends only on the degree, number of defining equations for  $X$ , and number of variables involved (see for instance [B2]) the constant implicit in the remainder term of the estimate for  $N_n(V_\lambda)$  is bounded uniformly with respect to  $\lambda$ .

We use here Chow's theory of the  $K/k$ -image, as in [L], Ch. VIII. The generically exact sequence of varieties  $V_\lambda \rightarrow V \rightarrow \mathbb{A}_k^1$  yields a surjection

$$\operatorname{Alb}(V_\lambda^{d-1}) \rightarrow \operatorname{Alb}(V) \rightarrow 0$$

for all but finitely many  $\lambda \in K$ . Now the hypothesis  $\dim \operatorname{Alb}(V_\lambda^{d-1}) = \dim \operatorname{Alb}(V)$  shows that this homomorphism is an isogeny and therefore the traces of Frobenius on the associated abelian varieties are the same. This proves

$$\operatorname{Tr}(\phi^n | \operatorname{Alb}(V_\lambda^{d-1})) = \operatorname{Tr}(\phi^n | \operatorname{Alb}(V))$$

for all but finitely many  $\lambda \in K$ . By our evaluation of  $N_n(V_\lambda)$  and  $N_n(V)$  we deduce that

$$N_n(V_\lambda) - N_n^* = O((q^n)^{d-2})$$

for almost every  $\lambda$ , while

$$N_n(V_\lambda) - N_n^* = O((q^n)^{d-3/2})$$

in any case. Now Lemma 1 and a simple calculation show that  $\mathcal{S}_n = O((q^n)^{2d-2})$ , as asserted.

To complete the proof of Lemma 2 we note that by Lemma 1,

$$\begin{aligned} \mathcal{S}_n &= q^n \sum_\lambda N_n(V_\lambda)^2 - 2N_n(V) \sum_\lambda N_n(V_\lambda) + N_n(V)^2 \\ &= q^n \sum_\lambda N_n(V_\lambda)^2 - N_n(V)^2 = q^n N_n(W) - N_n(V)^2. \end{aligned}$$

Therefore  $\mathcal{S}_n$  is a sum of characteristic roots and the number of characteristic roots of  $\mathcal{S}_n$  does not exceed  $\varrho(W) + \varrho(V)^2$ . By Theorem 3 of [B2], we deduce from the bound  $\mathcal{S}_n = O((q^n)^{2d-2})$  that the characteristic roots of  $\mathcal{S}$  have weight at most  $2d - 2$ . This result and the bound for the number of characteristic roots of  $\mathcal{S}_n$  prove Lemma 2. ■

Our next result is a lower bound for  $\mathcal{S}_n$ . The idea of proof is due to Hooley.

LEMMA 3. For any  $\varepsilon > 0$  we have the lower bound

$$\mathcal{S}_n \geq (p - 1)(r_{2d-2}(f) + s_{2d-2}(f) - \varepsilon)(q^n)^{2d-2}$$

for infinitely many  $n$ 's.

PROOF. We have

$$\begin{aligned} \mathcal{S}_n &\geq \sum_{l \in k_n^*} |S_n(V, lf, \psi)|^2 \geq \sum_{l \in \mathbb{F}_p^*} |S_n(V, lf, \psi)|^2 \\ &= \sum_{l \in \mathbb{F}_p^*} \left| \sum_i \alpha_i(lf)^n - \sum_j \beta_j(lf)^n \right|^2. \end{aligned}$$

The sums  $S_n(V, lf, \psi)$  for  $l \in \mathbb{F}_p^*$  are conjugates of  $S_n(V, f, \psi)$ , as one sees expressing them as sums of  $p$ th roots of unity and noting that for  $l \in \mathbb{F}_p^*$  we have  $\psi_n(lf) = \psi_n(f)^l$ . Thus by Deligne's theorem on Galois invariance of weights we see that  $r_w(lf) = r_w(f)$  and  $s_w(lf) = s_w(f)$  for  $l \in \mathbb{F}_p^*$ . Lemma 3 now follows from an easy application of Lemma 2 of [BS]. ■

PROOF OF THEOREM 1. We combine the upper and lower bounds of Lemmas 2 and 3 and obtain

$$(p - 1)(r_{2d-2}(f) + s_{2d-2}(f) - \varepsilon)(q^n)^{2d-2} \leq (\varrho(W) + \varrho(V)^2)(q^n)^{2d-2}$$

for infinitely many  $n$ 's. This shows that if  $r_{2d-2}(f) + s_{2d-2}(f) \geq 1$  then we must have  $p - 1 \leq \varrho(W) + \varrho(V)^2$ . If we define  $D(V, f) = 1 + \varrho(W) + \varrho(V)^2$  we obtain the conclusion of Theorem 1. ■

REMARK. The conditions of Theorem 1 are, in a certain sense, necessary and sufficient. In fact, it is clear from our argument that if  $\mathcal{S}_n$  is not  $O((q^n)^{2d-2})$  then it must have characteristic roots of weight at least  $4d - 3$ , hence  $\mathcal{S}_n > (1 - \varepsilon)(q^n)^{2d-3/2}$  for infinitely many  $n$ 's and therefore we cannot have  $S_n(V, lf, \psi) = O((q^n)^{d-3/2})$  for every  $l \in k_n^*$ . Now if condition (iii) is not satisfied we have  $N_n(V_\lambda) > (2 - \varepsilon)(q^n)^{d-1}$  for infinitely many  $n$ 's whenever  $V_\lambda \otimes K$  has 2 or more components of dimension  $d - 1$ , and this implies  $\mathcal{S}_n > (1 - \varepsilon)(q^n)^{2d-1}$  infinitely often; this shows that condition (iii) is necessary. In a similar way we handle condition (iv). The surjection  $\text{Alb}(V_\lambda^{d-1}) \rightarrow \text{Alb}(V) \rightarrow 0$  shows that

$$\text{Tr}(\phi^n | \text{Alb}(V_\lambda^{d-1})) = \text{Tr}(\phi^n | \text{Alb}(V)) + \sum_i (\alpha_i(\lambda))^{n/n(\lambda)}$$

where  $n(\lambda) = [k(\lambda) : k]$  and where the  $\alpha_i(\lambda)$  are characteristic roots of weight 1 with respect to  $\phi^{n(\lambda)}$ . It follows that

$$\mathcal{S}_n \geq (q^n)^{2d-3} \sum_{\lambda \in k_n^*} \left( \sum \alpha_i(\lambda)^{n/n(\lambda)} \right)^2 + O((q^n)^{2d-2}).$$

If we consider only  $n$ 's with  $k_{n_0} \subset k_n$  for fixed large  $n_0$ , and consider the contribution to the sum of  $\lambda \in k_{n_0}^*$ , we see by applying [BS], Lemma 2, that

$$\mathcal{S}_n > (q^{n_0} - O(1))(q^n)^{2d-2}$$

for infinitely many  $n$ 's. Since  $n_0$  can be taken arbitrarily large, we have shown that the negation of (iv) implies that  $\mathcal{S}_n$  has characteristic roots of weight at least  $4d - 3$ , and in particular there are infinitely many  $l \neq 0$  for which the associated sum  $S_n(V, lf, \psi)$  has characteristic roots of weight at least  $2d - 2$  with respect to  $\phi^{n(\lambda)}$ .

*Remark.* The following alternative approach has been suggested to us by Deligne. It is in fact a special case of the considerations in Katz's "Théorème Clef" of [K], p. 90, especially Cor. 4, p. 95.

Let  $\mathcal{L}_{\psi_0 \circ f}$  be the local system of rank 1 on  $V_K$  determined by the étale cover  $T^p - T = f(x)$  of  $V_K$  and by the character  $\psi_0$  and let  $\mathcal{L}_{\psi_0}$  be the corresponding system on the affine line  $\mathbb{A}_K^1$  and the cover  $T^p - T = x$ . We have the Lefschetz trace formula

$$S_1(V_K, f, \psi) = \sum_{i=0}^{2d} (-1)^i \text{Tr}(\phi | H_c^i(V_K, \mathcal{L}_{\psi_0 \circ f}))$$

and our aim is to show that there is no contribution of weight  $w > 2d - 3$  to the traces if conditions (i)–(iv) of Theorem 1 are satisfied and the characteristic  $p$  is sufficiently large.

In what follows,  $W$  will be an open set of  $\mathbb{A}_K^1$  over which all the  $R^q f_! \mathbb{Q}_l$  are lisse.

Consider the morphism  $f : V_K \rightarrow \mathbb{A}_K^1$  determined by  $f$ . We have a Leray spectral sequence with  $E_2^{p,q}$  term (no reference to the characteristic  $p$ , here):

$$E_2^{p,q} = H_c^p(\mathbb{A}_K^1, R^q f_! \mathbb{Q}_l \otimes \mathcal{L}_{\psi_0})$$

converging to  $H_c^*(V_K, \mathcal{L}_{\psi_0 \circ f})$ , therefore

$$S_1(V, f, \psi) = \sum (-1)^{p+q} \text{Tr}(\phi | E_2^{p,q}).$$

To show that  $H_c^{p+q}(V_K, \mathcal{L}_{\psi_0 \circ f})$  is mixed of weight at most  $2d - 3$ , it suffices to show that all  $E_2^{p,q}$  terms are mixed of weight at most  $2d - 3$ . Since  $f$  is surjective, with all fibres of dimension  $d - 1$ , we see that  $R^q f_! \mathbb{Q}_l$  and hence  $R^q f_! \mathbb{Q}_l \otimes \mathcal{L}_{\psi_0}$  vanishes for  $q > 2d - 2$ . Moreover, Deligne's theorem gives that  $E_2^{p,q}$  is *a priori* mixed of weight at most  $p + q$ . So the only possible terms of weight greater than  $2d - 3$  are:

- $E_2^{p, 2d-2}$  with  $p = 0, 1, 2$ ;
- $E_2^{p, 2d-3}$  with  $p = 1, 2$ ;
- $E_2^{p, 2d-4}$  with  $p = 2$ .

We first consider the terms with  $p = 2$ . For these, the birational invariance of  $H_c^2$  on curves shows that

$$E_2^{2,q} = H_c^2(\mathbb{A}_K^1, R^q f_! \mathbb{Q}_l \otimes \mathcal{L}_{\psi_0}) = H_c^2(W, R^q f_! \mathbb{Q}_l \otimes \mathcal{L}_{\psi_0}).$$

Because we are in large characteristic, the Hooley argument of Galois conjugacy, or [K], “Théorème Clef”, p. 90, shows that  $E_2^{2,q} = 0$  for all  $q$ . Note, however, that the application of Hooley’s argument in this context requires  $\text{char}(K) - 1$  larger than the dimension of the groups  $H^0(W, R^q f_! \mathbb{Q}_l)$  for  $q = 2d - 4, 2d - 3$ , and we do not know how to control in general the dimension of these groups; this creates problems with the determination of  $D(V, f)$  in our Theorem 1. On the other hand, Katz’s Theorem applies effectively to the case in which  $V$  is obtained by reduction modulo  $p$  of a fixed variety defined over the ring of integers of a number field, and this is what matters in applications.

We next consider the terms with  $q = 2d - 2$ . By hypothesis (iii), the sheaf  $R^{2d-2} f_! \mathbb{Q}_l$  is geometrically constant, and noting that  $H_c^*(\mathbb{A}_K^1, \mathcal{L}_{\psi_0}) = 0$ , we get  $E_2^{*,2d-2} = 0$ .

So it remains only to show that  $E_2^{1,2d-3} = H_c^1(\mathbb{A}_K^1, R^{2d-3} f_! \mathbb{Q}_l \otimes \mathcal{L}_{\psi_0})$  is mixed of weight at most  $2d - 3$ ; this group is a quotient of  $H_c^1(W, R^{2d-3} f_! \mathbb{Q}_l \otimes \mathcal{L}_{\psi_0})$ . Now, over  $W$ , the sheaf  $R^{2d-3} f_! \mathbb{Q}_l$  is lisse and, being mixed of weight at most  $2d - 3$ , sits in a short exact sequence of lisse sheaves

$$0 \rightarrow \text{Fil}_{wt \leq 2d-4} \rightarrow R^{2d-3} f_! \mathbb{Q}_l \rightarrow \text{Gr}_{wt=2d-3} \rightarrow 0;$$

in fact, here  $\text{Gr}_{wt=2d-3}$  is the component of weight  $2d - 3$  in the weight filtration of  $R^{2d-3} f_! \mathbb{Q}_l$ . By hypothesis (i), the lisse sheaf  $\text{Gr}_{wt=2d-3}$  is geometrically constant. By Deligne’s theorem, we know that  $H_c^1(W, \text{Fil}_{wt \leq 2d-4} \otimes \mathcal{L}_{\psi_0})$  is mixed of weight at most  $2d - 3$ , so we are reduced to showing that  $H_c^1(W, \text{Gr}_{wt=2d-3} \otimes \mathcal{L}_{\psi_0})$  is mixed of weight at most  $2d - 3$ . Because  $\text{Gr}_{wt=2d-3}$  is geometrically constant, it extends uniquely to a geometrically constant sheaf on  $\mathbb{A}_K^1$ , still denoted by  $\text{Gr}_{wt=2d-3}$ . We have a piece of exact sequence

$$\begin{aligned} \bigoplus_{x \in \mathbb{A}_K^1 - W} (\text{Gr}_{wt=2d-3} \otimes \mathcal{L}_{\psi_0})_x &\rightarrow H_c^1(W, \text{Gr}_{wt=2d-3} \otimes \mathcal{L}_{\psi_0}) \\ &\rightarrow H_c^1(\mathbb{A}_K^1, \text{Gr}_{wt=2d-3} \otimes \mathcal{L}_{\psi_0}) \rightarrow 0. \end{aligned}$$

The stalks are pure of weight  $2d - 3$ , while  $H_c^1(\mathbb{A}_K^1, \text{Gr}_{wt=2d-3} \otimes \mathcal{L}_{\psi_0}) = 0$ , because  $\text{Gr}_{wt=2d-3}$  is geometrically constant and  $H_c^1(\mathbb{A}_K^1, \mathcal{L}_{\psi_0}) = 0$ . Therefore  $H_c^1(W, \text{Gr}_{wt=2d-3} \otimes \mathcal{L}_{\psi_0})$  is pure of weight  $2d - 3$ , as we wanted. ■

**III.** In this section we obtain explicit bounds for  $D(V, f)$  and  $\rho(V, f)$  which generalize the results of [B2] to quasi-projective varieties. The method of proof is already outlined in [B2].

Let  $V$  be a quasi-projective variety in  $\mathbb{P}^N$  defined over  $k$ . By a *set-theoretic presentation of  $V$*  we mean the following data: a collection  $\{F_1(\mathbf{x}), \dots, F_l(\mathbf{x}); G_1(\mathbf{x}), \dots, G_m(\mathbf{x})\}$  of forms defined over  $k$  such that

$$V(K) = \{\mathbf{x} \in \mathbb{P}^N(K) \mid F_1(\mathbf{x}) = \dots = F_l(\mathbf{x}) = 0; \\ G_j(\mathbf{x}) \neq 0 \text{ for some } j, j = 1, \dots, m\}.$$

The *degree* of the presentation is the maximum of the degrees of the forms  $F_i$ ,  $G_j$  and its *length* is  $l + m$ , the number of forms appearing in the presentation.

Let  $f$  be a rational function on  $V$ , defined over  $k$ . By a *presentation of  $f$*  we mean the following data: a collection  $\{P_i(\mathbf{x})/Q_i(\mathbf{x}) \mid i = 1, \dots, t\}$  of rational functions in  $k(\mathbb{P}^N)$  such that for every  $\mathbf{x} \in V(K)$  we have  $Q_i(\mathbf{x}) \neq 0$  for some  $i$  and  $f(\mathbf{x}) = P_i(\mathbf{x})/Q_i(\mathbf{x})$  whenever  $Q_i(\mathbf{x}) \neq 0$ . The *degree* of the presentation is the maximum degree of the forms  $P_i$ ,  $Q_i$  and its *length* is  $t$ .

The preceding notion of presentation can be generalized to subvarieties and rational functions of products of projective spaces by replacing the homogeneous coordinates  $\mathbf{x}$  by multihomogeneous coordinates  $(\mathbf{x}, \mathbf{x}', \dots)$ .

**THEOREM 2.** *Let  $V$  and  $f$  admit presentations over  $k$  of degree at most  $D$  and length at most  $L, L'$  respectively. Then*

$$\varrho(V, f) \leq (N + 1 + L)(4D + 15)^{N+1+L+2L'}.$$

*More generally, if the presentations of  $V$  and  $f$  are given as subvarieties and functions of a product of projective spaces  $\mathbb{P}^{N_1} \times \dots \times \mathbb{P}^{N_k}$ , the same estimate for  $\varrho(V, f)$  holds provided we replace  $N+1+L$  by  $N_1 + \dots + N_k + k + L$  everywhere.*

**Proof.** For notational simplicity we treat only the case  $V \subset \mathbb{P}^N$ , the general case being essentially the same. Suppose first that  $f$  is not a constant, so that  $t > 0$  and no  $Q_i$  is constant. For subsets  $I \subset \{1, \dots, t\}$  and  $J \subset \{1, \dots, m\}$  with  $I \neq \emptyset$  we fix an element  $\mu \in I$  and define

$$g_{IJ} = \sum_{i=1}^l y_i F_i(\mathbf{x}) + \sum_{j \in J} z_j G_j(\mathbf{x}) + w_\mu P_\mu(\mathbf{x}) + \sum_{i \in I} u_i (1 - w_i Q_i(\mathbf{x})).$$

In what follows we assume  $m \geq 1$ , the case  $m = 0$  being simpler.

We begin by evaluating the exponential sum  $S_n(\mathbb{A}^{N+1+l+|J|+2|I|}, g_{IJ}, \psi)$ . We perform the summation over  $y_1, \dots, y_l, z_j$  for  $j \in J$  and  $u_i$  for  $i \in I$ . Since these variables appear linearly, we find

$$S_n(\mathbb{A}^{N+1+l+|J|+2|I|}, g_{IJ}, \psi) = (q^n)^{l+|I|+|J|} \sum' \psi_n(w_\mu P_\mu(\mathbf{x}))$$

where the sum runs over  $w_i \in k_n, i \in I$  and  $\mathbf{x} \in \mathbb{A}^{N+1}(k_n)$  such that  $F_i(\mathbf{x}) = 0$  for  $i = 1, \dots, l, G_j(\mathbf{x}) = 0$  for  $j \in J$  and moreover  $1 - w_i Q_i(\mathbf{x}) = 0$  for  $i \in I$ . The condition  $1 - w_i Q_i(\mathbf{x}) = 0$  implies  $Q_i(\mathbf{x}) \neq 0$ , hence  $\mathbf{x} \neq 0$ ,

$w_i = 1/Q_i(\mathbf{x})$  and in particular  $w_\mu P_\mu(\mathbf{x}) = P_\mu(\mathbf{x})/Q_\mu(\mathbf{x}) = f(\mathbf{x})$ . Now a point of  $V(k_n)$  has exactly  $q^n - 1$  affine representatives, thus we obtain

$$\sum' \psi_n(w_\mu P_\mu(\mathbf{x})) = (q^n - 1)S_n(V_{IJ}, f, \psi)$$

where  $V_{IJ}$  is the quasi-projective variety determined by  $F_i(\mathbf{x}) = 0$  for  $i = 1, \dots, l$ ,  $G_j(\mathbf{x}) = 0$  for  $j \in J$  and  $Q_i(\mathbf{x}) \neq 0$  for  $i \in I$ . We have shown

$$S_n(\mathbb{A}^{N+1+l+|J|+2|I|}, g_{IJ}, \psi) = (q^n)^{l+|I|+|J|} (q^n - 1)S_n(V_{IJ}, f, \psi).$$

Let  $V_I$  be the quasi-projective subvariety of  $\mathbb{P}^N$  determined by  $F_i(\mathbf{x}) = 0$  for  $i = 1, \dots, l$ ,  $G_j(\mathbf{x}) \neq 0$  for some  $j$ ,  $j = 1, \dots, m$  and  $Q_i(\mathbf{x}) \neq 0$  for  $i \in I$ . Then, with  $M = \{1, 2, \dots, m\}$ , we have

$$S_n(V_I, f, \psi) = S_n(V_{I\emptyset}, f, \psi) - S_n(V_{IM}, f, \psi).$$

An application of the inclusion-exclusion principle gives

$$S_n(V, f, \psi) = \sum_{I \neq \emptyset} (-1)^{|I|-1} S_n(V_I, f, \psi),$$

therefore

$$\begin{aligned} S_n(V, f, \psi) &= (q^n - 1)^{-1} \sum_{I \neq \emptyset} (-1)^{|I|-1} (q^n)^{-l-|I|} \\ &\quad \times (S_n(\mathbb{A}^{N+1+l+2|I|}, g_{I\emptyset}, \psi) - (q^n)^{-m} S_n(\mathbb{A}^{N+1+l+m+2|I|}, g_{IM}, \psi)). \end{aligned}$$

The characteristic roots of the sum  $S'$  appearing in the right-hand side have weight at most  $2N + 2 + 2m$ , hence

$$S_n(V, f, \psi) = (q^{-n} + q^{-2n} + \dots + q^{-(N+1+m)n})S' + O(q^{-n}).$$

On the other hand, the characteristic roots appearing in  $S_n(V, f, \psi)$  have weight at least 0, and it follows that the error term  $O(q^{-n})$  is inconsequential for the purpose of counting the number of characteristic roots in the left-hand side of this equation. It follows that

$$\varrho(V, f) \leq (N + 1 + m) \sum_{I \neq \emptyset} (\varrho(\mathbb{A}^{N+1+l+2|I|}, g_{I\emptyset}) + \varrho(\mathbb{A}^{N+1+l+m+2|I|}, g_{IM})),$$

and Theorem 2 follows easily from [B2], Theorem 1.

Rather similar but simpler considerations hold if  $f$  is constant because in that case we may assume that  $t = 0$ , completing the proof of Theorem 2. ■

Let  $\{F_i, i = 1, \dots, l; G_j, j = 1, \dots, m\}$  be a presentation for  $V$  and let  $\{P_i/Q_i \mid i = 1, \dots, t\}$  be a presentation of  $f$ . Let us abbreviate  $\mathbf{F}, \dots, \mathbf{Q}$  for the vectors with components  $F_i, \dots, Q_i$ . Then a presentation of  $W$  in  $\mathbb{P}^n \times \mathbb{P}^n$  is

$$\{\mathbf{F}(\mathbf{x}), \mathbf{F}(\mathbf{y}), P_i(\mathbf{x})Q_j(\mathbf{y}) - P_j(\mathbf{y})Q_i(\mathbf{x}), \\ i, j = 1, \dots, t; \mathbf{G}(\mathbf{x}) \otimes \mathbf{Q}(\mathbf{x}), \mathbf{G}(\mathbf{y}) \otimes \mathbf{Q}(\mathbf{y})\}.$$

An easy application of Theorem 2 now gives in every case

THEOREM 3. *We have*

$$D(V, f) \leq 4(N + 1 + mt)(4D + 15)^{2N+2+2l+t^2+2mt}, \\ \rho(V, f) \leq (N + 1 + m)(4D + 15)^{N+1+l+m+2t}$$

with  $D = \max \deg(\mathbf{F}, \mathbf{G}, \mathbf{P}, \mathbf{Q})$ .

IV. In this section we treat two examples of exponential sums using Theorem 1. Our first example generalizes a sum which occurs already in the work of Iwaniec [I].

Let  $S_n = \sum \psi_n(f(x, y, z))$  where

$$f(x, y, z) = a(x) + y + \frac{b(x)}{y} + z + \frac{c(x)}{z},$$

where  $a(x), b(x), c(x)$  are rational functions in  $k(x)$  and where the sum runs over the  $k_n$ -rational points of the variety  $U \times \mathbb{G}_m^2$  with  $U = \mathbb{A}^1 - (a(x))_\infty - (b(x))_\infty - (c(x))_\infty$ .

We want to apply Theorem 1 to this sum. Conditions (i) and (ii) are trivially verified. The following lemma deals with condition (iii).

LEMMA 4. *Consider the fiber  $V_\lambda = f^{-1}(\lambda) \otimes K$  of the map  $f$  over the point  $\lambda$ . Then if one of*

- (i)  $b(x)$  is identically 0,
- (ii)  $c(x)$  is identically 0,
- (iii)  $b(x) - c(x)$  and  $a(x) - \lambda$  are identically 0,

*holds, the fiber is reducible. In all other cases,  $V_\lambda$  is irreducible.*

PROOF. The surface  $V_\lambda$  is defined by the equation

$$F(x, y, z) = (y + z)yz + (a(x) - \lambda)yz + c(x)y + b(x)z = 0.$$

If  $V_\lambda$  is reducible then  $F(x, y, z)$ , considered as an element of the polynomial ring  $k(x)[y, z]$ , must have at least one linear factor over an algebraic closure  $K$  of  $k(x)$ . Clearly the only possibilities for such a factor, up to a constant multiplier, are  $y + z - \alpha, y - \alpha, z - \alpha$  for some  $\alpha \in K$ . If the factor is for example  $z - \alpha$  then  $F(x, y, \alpha) = 0$  and we obtain

$$(y + \alpha)y\alpha + (a(x) - \lambda)y\alpha + c(x)y + b(x)\alpha = 0$$

identically as a polynomial in  $K[y]$ . This shows that  $\alpha = 0$  and  $c(x) = 0$  identically. A similar analysis for the two remaining possibilities yields the statement of Lemma 4. ■

The next lemma deals with condition (iv).

LEMMA 5. *If at least one of  $a(x)$ ,  $b(x)$ ,  $c(x)$  is not constant and  $b(x)$ ,  $c(x)$  are not identically 0 then for all but finitely many  $\lambda$ 's we have  $\dim \text{Alb}(V_\lambda) = 0$ .*

PROOF. Let  $\pi_\lambda : V_\lambda \rightarrow \mathbb{P}_K^1$  be the projection on the  $x$ -line. The fiber  $E_x = \pi_\lambda^{-1}(x)$  is the plane cubic (in affine coordinates  $y, z$ )

$$(y+z)yz + (a(x) - \lambda)yz + c(x)y + b(x)z = 0.$$

We put this cubic into Weierstrass form (assuming  $\text{char}(k) \neq 2, 3$ ) as follows. We abbreviate  $a = a(x)$ ,  $b = b(x)$ ,  $c = c(x)$  and perform successively the birational transformations:

- $z = ry$ , obtaining

$$(1+r)ry^2 + (a-\lambda)ry + br + c = 0;$$

- $(1+r)ry = s$ , obtaining

$$s^2 + (a-\lambda)rs + r(1+r)(br+c) = 0;$$

- $s + \frac{1}{2}(a-\lambda)r = t$ ,  $2bt = v$ ,  $-br - \frac{1}{12}((a-\lambda)^2 - 4b - 4c) = u$ , getting the Weierstrass equation

$$v^2 = 4u^3 - g_2u - g_3$$

with invariants

$$g_2 = \frac{1}{12}A^2 - 4bc, \quad g_3 = -\frac{8}{12^3}A^3 + \frac{1}{3}Abc$$

with

$$A = (a-\lambda)^2 - 4b - 4c.$$

The discriminant and absolute invariant are

$$\Delta = g_2^3 - 27g_3^2 = (bc)^2(A^2 - 64bc),$$

$$J = 12^3g_3^3/\Delta = (A^2 - 48bc)^3(bc)^{-2}(A^2 - 64bc)^{-1}.$$

We claim that for generic  $\lambda$  the absolute invariant  $J$  is a non-constant rational function of  $x$ , provided neither  $b(x)$  nor  $c(x)$  is identically 0 and at least one of  $a(x)$ ,  $b(x)$ ,  $c(x)$  is non-constant. In fact, suppose  $J = J(\lambda)$  is constant, hence  $J$  is a rational function of  $\lambda$  only. This gives

$$(A^2 - 48bc)^3 - J(\lambda)(bc)^2(A^2 - 64bc) = 0.$$

We deduce that  $J(\lambda)$  has a pole at  $\lambda = \infty$  of order 8, and looking at the Laurent expansion we get  $J(\lambda) = (bc)^{-2}\lambda^8 + \dots$ . This implies that  $bc$  is constant, and the last displayed equation shows that  $A = (a-\lambda)^2 - 4b - 4c$  is algebraic over  $K(\lambda)$ . Now this implies that  $a$  and  $b+c$  are constants. Since  $bc$  was a constant, we deduce that  $a, b, c$  are all constants, proving our claim.



The generically exact sequence of varieties  $E_x \rightarrow V_\lambda \rightarrow \mathbb{P}_K^1$  yields a surjection  $\text{Alb}(E_x) \rightarrow \text{Alb}(V_\lambda) \rightarrow 0$ , hence the  $K(x, \lambda)/K(\lambda)$ -image  $A$  of the family  $E_x$  has dimension not less than  $\dim \text{Alb}(V_\lambda)$ . If we had  $\dim \text{Alb}(V_\lambda) = 1$  then the canonical homomorphism  $\text{Alb}(E_x) \rightarrow A$  would be a purely inseparable isogeny, because the kernel of the  $K/k$ -image is connected ([L], Ch. VIII, Prop. 3, p. 199). From this we deduce that the  $J$  invariant of the family  $E_x$  is constant, which we have shown is not the case for generic  $\lambda$ . Thus  $\dim \text{Alb}(V_\lambda) = 0$  for almost every  $\lambda$ , completing the proof of our lemma. ■

By Lemma 4, Lemma 5 and Theorem 1 we obtain

**THEOREM 4.** *Assume that at least one of  $a(x), b(x), c(x)$  is non-constant, that neither  $b(x)$  nor  $c(x)$  is identically 0, and that if  $a(x)$  is constant then  $b(x)$  is not identically equal to  $c(x)$ . There are constants  $D$  and  $\rho$  depending only on the degrees of  $a(x), b(x), c(x)$  such that if  $p > D$  we have*

$$|S_n(\mathbb{A}^1 \times \mathbb{G}_m^2, f, \psi)| \leq \rho q^{3n/2}.$$

Our second example is the sum  $S_n(\mathbb{A}^3, f_3, \psi)$  where  $f_3(x, y, z)$  is a cubic polynomial in three variables. We decompose  $f_3$  as  $f_3 = C + Q + L + c$  where  $C$  is homogeneous cubic,  $Q$  is homogeneous quadratic,  $L$  is linear and  $c$  is constant. If the projective cubic curve  $C = 0$  is non-singular then a special case of a result of Deligne [D] gives the optimal bound

$$|S_n(\mathbb{A}^3, f_3, \psi)| \leq 8q^{3n/2};$$

in fact, Deligne's general theorem is that if  $f_r$  is a polynomial of degree  $d$  in  $N$  variables and if the homogeneous part of degree  $d$  of  $f_r$  defines a non-singular projective variety of dimension  $N - 2$ , then the characteristic roots of the sum  $S_n(\mathbb{A}^N, f_r, \psi)$  have weight  $N$ , and

$$|S_n(\mathbb{A}^N, f_r, \psi)| \leq (r - 1)^N (q^n)^{N/2}.$$

**THEOREM 5.** *Suppose that the cubic polynomial  $f_3 - \lambda$  is irreducible in  $K(x, y, z)$  for every  $\lambda \in K$ , and suppose that the projective plane curve  $C = 0$  does not consist of three lines through a point. Then there are constants  $D$  and  $\rho$  such that if  $p > D$  we have*

$$|S_n(\mathbb{A}^3, f_3, \psi)| \leq \rho q^{3n/2}.$$

**Proof.** Conditions (i) and (ii) of Theorem 1 are trivially verified, and condition (iii) is in the hypothesis of Theorem 5; we may note, however, that if the cubic curve  $C = 0$  is geometrically irreducible then  $f_3 - \lambda$  is absolutely irreducible for every  $\lambda$ . For condition (iv), we need verify that if  $V_\lambda$  is the cubic surface  $f_3(x, y, z) - \lambda = 0$  then  $\text{Alb}(V_\lambda)$  is trivial. But a cubic surface over an algebraically closed field which is not a cone over a non-singular cubic plane curve is birationally equivalent to a projective plane, thus the only possibility is that the cubic surface  $V_\lambda$  has a triple point for every  $\lambda$ ,

i.e. it is a cone. In fact, if a cubic surface has a double point, projection from this point establishes a birational map with the projective plane, while if it is non-singular then it is well-known that it is rational. For a proof of this classical result, see for example [R], (7.4), Cor. (b).

Now suppose that the characteristic  $p$  is  $p > D$ . We claim that this triple point cannot be a movable singularity and therefore lies in the base set of the pencil, that is the linear section at infinity. This would mean that the projective curve  $C = 0$  has a triple point, and since  $C$  has degree 3 we would deduce that it consists of three lines through a point, which was excluded from our considerations.

In order to complete the proof of the theorem we need to prove our statement that a triple point for  $V_\lambda$  cannot be an isolated movable singularity. In characteristic 0, this follows from Bertini's theorem, but in non-zero characteristic we need to be a little careful because Bertini's theorem does not hold in this generality.

Suppose that  $\mathbf{x}(\lambda) = (x(\lambda), y(\lambda), z(\lambda))$  is an isolated singularity of  $V_\lambda$ . Then we have

$$f_3 - \lambda = 0, \quad \frac{\partial f_3}{\partial x} = 0, \quad \frac{\partial f_3}{\partial y} = 0, \quad \frac{\partial f_3}{\partial z} = 0$$

and therefore if  $p > D$  we see that  $K(\lambda, \mathbf{x}(\lambda))$  is a separable finite extension of  $K(\lambda)$ . Once this separability condition is verified, the proof of Bertini's theorem in characteristic 0 carries through (for convenience of the reader, we sketch the argument: suppose  $f + \lambda g = 0$  is our pencil with a singular point  $\mathbf{x}(\lambda)$ , and let  $\tau$  be a primitive element of  $K(\lambda, \mathbf{x}(\lambda))$  over  $K(\lambda)$  such that  $\lambda$  is separable over  $K(\tau)$  (note that either  $\tau$  or  $\tau' = \tau - \lambda$  will have this property); differentiating  $f + \lambda g$  with respect to  $\tau$  we obtain  $\frac{d\lambda}{d\tau} g(\mathbf{x}(\lambda)) = 0$ , and since  $K(\tau)$  is separable over  $K(\lambda)$  we see that  $d\lambda/d\tau$  is not identically 0, thus  $g(\mathbf{x}(\lambda)) = 0$  and  $f(\mathbf{x}(\lambda)) = 0$ , as was to be shown). ■

The following generalization to polynomials of higher degree appears to be new.

**THEOREM 6.** *Let  $f_r(x, y, z)$  be a polynomial of degree  $r$  with coefficients in  $k$ , and let  $F_r(x, y, z)$  be the homogeneous part of degree  $r$  of  $f_r$ . Suppose that the projective plane curve of degree  $r$  defined by  $F_r(x, y, z) = 0$  is geometrically irreducible and that its singularities are ordinary nodes. Then there are constants  $D_r$  and  $\varrho_r$  depending only on  $r$  such that if  $p > D_r$  we have*

$$|S_n(\mathbb{A}^3, f_r, \psi)| \leq \varrho_r q^{3n/2}.$$

**Proof.** Conditions (i), (ii) of Theorem 1 are trivially verified. The validity of condition (iii) is, as observed before, an immediate consequence of the absolute irreducibility of  $F_r$ . Let  $X_\lambda$  be the projective surface in  $\mathbb{P}_K^3$

defined in the affine chart  $\mathbb{A}_K^3$  by the equation  $f_r(x, y, z) - \lambda = 0$  and in projective coordinates by the homogeneous equation  $F(x, y, z, w) - \lambda w^r = 0$ , with  $F(x, y, z, 0) = F_r(x, y, z)$ .

Since we assume that  $p > D_r$  the same argument as in the proof of Theorem 5 shows that we may apply Bertini's theorem to deduce that a general member of the pencil of surfaces  $X_\lambda$  has singularities only on the base locus of the pencil, i.e. the projective curve defined by  $F_r(x, y, z) = 0$  and  $w = 0$ . By our assumption, the only singularities of this curve are ordinary nodes with distinct tangents. Let  $(1, 0, 0, 0)$  be a singularity of  $X_\lambda$  for a general  $\lambda \in K$ , let  $A$  be the corresponding local ring and let  $\widehat{A}$  be the completion of  $A$  with respect to its maximal ideal. In affine coordinates  $(y, z, w)$  near  $(0, 0, 0)$  we see that if the characteristic is not 2 we have

$$\widehat{A} \cong K[[y, z, w]]/(g)$$

where  $g$  has the form

$$g = y^2 + z^2 + wh - \lambda w^r$$

with  $h$  in the maximal ideal of  $\widehat{A}$ . For some integer  $a \geq 1$  we write  $wh = yw^a h_1 + zw^a h_2 + w^2 u(w)$  with  $u(w) \in K[[w]]$ , and use the invertible change of variables  $y' = y - \frac{1}{2}w^a h_1$ ,  $z' = z - \frac{1}{2}w^a h_2$ , to obtain

$$g = (y')^2 + (z')^2 + w^{2a} h' + w^2 u(w) - \lambda w^r$$

with  $h' = -\frac{1}{4}(h_1^2 + h_2^2)$ . By iterating this process, we obtain the normal form

$$g = y^2 + z^2 + w^b u(w) - \lambda w^r$$

with  $u(w) \in K[[w]]$ ,  $u(0) \neq 0$  and  $b \geq 2$ . Now suppose  $p > r$ . We can change the  $w$  variable by means of  $w' = w(u(w) - \lambda w^{r-b})^{1/b}$  if  $b < r$ , and by  $w' = w(-\lambda + u(w)w^{b-r})^{1/r}$  if  $b \geq r$ , and get the standard normal form for the complete local ring

$$\widehat{A} \cong K[[y, z, w]]/(y^2 + z^2 + w^q)$$

for some integer  $q$  with  $2 \leq q \leq r < p$ .

It is well-known that this local ring is that of a rational normal double point, and we deduce that  $X_\lambda$  is a projective surface in  $\mathbb{P}_K^3$ , of degree  $r$ , with isolated singular points having for minimal resolution certain configurations of rational curves (see Artin [Ar] for more details).

We claim that any projective normal surface  $X$  in  $\mathbb{P}_K^3$  with isolated singular points which are resolved into configurations of rational curves has trivial Albanese variety.

Let  $\pi : X' \rightarrow X$  be a minimal desingularization of  $X$ . We want to prove that  $H^1(X', \mathbb{Q}_l)$  vanishes, and for this we use the exact sequence

$$0 \rightarrow H^1(X, \pi_* \mathbb{Q}_l) \rightarrow H^1(X', \mathbb{Q}_l) \rightarrow H^0(X, R^1 \pi_* \mathbb{Q}_l)$$

which arises from the Leray spectral sequence of the map  $\pi : X' \rightarrow X$ . Since  $X$  is normal we have  $\pi_*\mathbb{Q}_l = \mathbb{Q}_l$  on  $X$ , and  $H^1(X, \mathbb{Q}_l) = 0$  because  $X$  is an hypersurface in  $\mathbb{P}_K^3$ . This can be verified using the cohomology sequence

$$\dots \rightarrow H^1(\mathbb{P}_K^3, \mathbb{Q}_l) \rightarrow H^1(X_K, \mathbb{Q}_l) \rightarrow H_c^2(\mathbb{P}_K^3 - X, \mathbb{Q}_l) \rightarrow \dots,$$

noting that  $H^1(\mathbb{P}_K^3, \mathbb{Q}_l) = 0$ , also  $\dim H_c^2(\mathbb{P}_K^3 - X, \mathbb{Q}_l) = \dim H^4(\mathbb{P}_K^3 - X, \mathbb{Q}_l)$  by Poincaré duality, and  $H^4(\mathbb{P}_K^3 - X, \mathbb{Q}_l) = 0$  because  $\mathbb{P}_K^3 - X$  is an affine variety. It follows that

$$0 \rightarrow H^1(X', \mathbb{Q}_l) \rightarrow H^0(X, R^1\pi_*\mathbb{Q}_l)$$

is exact. Finally, we note that  $R^1\pi_*\mathbb{Q}_l$  is of weight 0. In fact, at a simple point  $x$  of  $X$ ,  $(R^1\pi_*\mathbb{Q}_l)_x = 0$  because  $\pi$  is an isomorphism there, while at a singular point  $\xi$ , we have  $(R^1\pi_*\mathbb{Q}_l)_\xi \cong H^1(\pi^{-1}(\xi), \mathbb{Q}_l)$ , which is of weight 0 because all components of  $\pi^{-1}(\xi)$  are rational curves. This shows that  $H^1(X', \mathbb{Q}_l)$  is of weight 0. On the other hand,  $X'$  is a non-singular projective variety, therefore  $H^1(X', \mathbb{Q}_l)$  is pure of weight 1; this shows that  $H^1(X', \mathbb{Q}_l) = 0$ .

Since  $H^1(X', \mathbb{Q}_l)$  vanishes and  $X'$  is non-singular we conclude that  $T_l(\text{Pic}^0(X')) = 0$  and by duality  $\text{Alb}(X')$  is trivial. Since the Albanese variety is a birational invariant, our assertion follows.

This verifies condition (iv) of Theorem 1, completing the proof of Theorem 6. ■

**V.** In this section, we apply the method of Adolphson–Sperber [AS1] to the examples discussed in the previous section. We consider first the case of a cubic polynomial in three variables

$$f(x, y, z) = \sum a_i x^{i_1} y^{i_2} z^{i_3} \in k[x, y, z].$$

Let  $f(x, y, z) = C + Q + L + c$  be the decomposition of  $f$  into homogeneous components. We analyze here rather completely the case in which the cubic curve at infinity  $C(x, y, z) = 0$  consists of three lines rational over  $k$ , thus supplementing the result of Theorem 5, and in fact describing the exact distribution of weights in all cases.

Suppose that  $C(x, y, z) = 0$  consists of three lines rational over  $k$ . Then a linear change of coordinates allows us to reduce to four cases:

- I:  $C(x, y, z) = ax^3$ ;
- II:  $C(x, y, z) = x^2y$ ;
- III:  $C(x, y, z) = xyz$ ;
- IV:  $C(x, y, z) = axy(x + y)$ .

We recall that  $\text{supp}(f) = \{i \in \mathbb{Z}^3 \mid a_i \neq 0\}$ , and

$$\Delta_\infty(f) = \text{convex closure in } \mathbb{R}^3 \text{ of } \text{supp}(f) \cup \{0\}.$$

If  $\sigma \subseteq \mathbb{R}^3$ , we write  $f^{(\sigma)} = \sum_{i \in \sigma \cap \mathbb{Z}^3} a_i x^{i_1} y^{i_2} z^{i_3}$ . We also say that  $f$  is *non-degenerate with respect to*  $\Delta_\infty(f)$  if  $\partial f^{(\sigma)} / \partial x, \partial f^{(\sigma)} / \partial y, \partial f^{(\sigma)} / \partial z$  have no common zero in  $(K^*)^3$  for every closed face  $\sigma$  of  $\Delta_\infty(f)$  not passing through 0. Finally,  $f$  is *convenient* if the intersection of  $\Delta_\infty(f)$  with each coordinate axis contains but is not equal to the origin.

In Theorem 5.17 of [AS1], a very precise result is given for sums on  $\mathbb{A}^n$  which are convenient and non-degenerate. The results apply when the characteristic  $p$  of  $k$  is greater than  $S(\Delta_\infty(f))$ , an effectively computable constant depending on  $\Delta_\infty(f)$ ; however, using the results of [DL1] this restriction on  $p$  is removed. In our application to cubic three-variable sums when  $f$  is non-degenerate, but no longer convenient, we use the toric decomposition of  $\mathbb{A}^3$ :

$$S_m(\mathbb{A}^3, f, \psi) = \sum_{A \subseteq S} S_m(\mathbb{G}_m^{|S-A|}, f_A, \psi),$$

where  $S = \{x, y, z\}$  and  $f_A$  is the polynomial in the variables  $S - A$  obtained by setting variables in  $A$  to 0. This reduces our sum to sums on tori  $\mathbb{G}_m^i$  with  $i \leq 3$ ; these can be further analyzed in terms of weights using the results of [AS2], [AS3], [DL2]; in the end we have a complete description of the degree and the weights of the associated  $L$ -function  $L(\mathbb{A}^3, f, \psi, T)$  when  $f$  is a non-degenerate cubic polynomial in three variables; we shall abbreviate  $L(f, T)$  for  $L(\mathbb{A}^3, f, \psi, T)$ . There remains then the problem of degenerate sums. In the four cases considered, these degenerate sums may be reduced to non-degenerate ones by rather simple changes of variables.

Let  $G$  be the group of isomorphisms of  $\mathbb{A}^3$  defined over  $k$ . Let  $g \in G$  and denote by  $f_g$  the image of  $f$  under  $g$ . The group  $G$  partitions  $k[x, y, z]$  into equivalence classes. Note

$$L(\mathbb{A}^3, f, \psi, T) = L(\mathbb{A}^3, f_g, \psi, T)$$

for any  $g \in G$ . If  $[f]$  is the class of  $f$ , we define

$$\dim[f] = \min_{g \in G} \dim \Delta_\infty(f_g),$$

and say  $[f]$  has a *simple representative* if there exists  $g \in G$  such that  $f_g$  is of type

$$f_g(x, y, z) = h(x, y) + az$$

with  $a \in k^*, h(x, y) \in k[x, y]$ . Clearly if  $[f]$  has a simple representative, then  $L(f, T) = 1$ .

**THEOREM 7.** *Assume  $\text{char}(k) \neq 2, 3$ . Let  $f = C + Q + L + c \in k[x, y, z]$  be of degree 3 and suppose that the projective plane curve  $C(x, y, z) = 0$  defines three lines rational over  $k$ . Then  $L(f, T) = 1$  if and only if  $f$  has a*

simple representative. If  $L(f, T)$  has a characteristic root with weight  $> 3$ , then  $\dim[f] < 3$ .

Proof. The proof proceeds by a case-by-case analysis. We write

$$f(x, y, z) = C(x, y, z) + Ax^2 + Bxy + Cy^2 + Dxz + Ez^2 + Fyz \\ + Gx + Hy + Jz + c.$$

Case I: a triple line. After a linear transformation,  $C(x, y, z)$  has the normal form  $C(x, y, z) = ax^3$ .

(i) Consider first the subcase in which  $F^2 - 4EC \neq 0$ . If  $EC \neq 0$  then  $f$  is non-degenerate and convenient and  $L(f, T)$  is a polynomial of degree 2, pure of weight 3 [AS1], [DL1]. If instead  $EC = 0$  then  $f$  is linear in either  $y$  or  $z$  and since then  $F \neq 0$  we can sum over that variable first. Again,  $L(f, T)$  is a polynomial of degree 2, pure of weight 3.

(ii) Suppose next that  $F^2 - 4EC = 0$ . If  $EC \neq 0$  then  $f$  is degenerate along the 1-dimensional face passing through the lattice points  $(0, 2, 0)$ ,  $(0, 1, 1)$ ,  $(0, 0, 2)$ . Since  $Ez^2 + Fzy + Cy^2 = E(z - \alpha y)^2$  with  $\alpha = -F/2E \in k^*$ , then the linear change  $z \mapsto z + \alpha y$  brings  $f$  into  $f = ax^3 + Ax^2 + Bxy + Dxz + Ez^2 + Gx + Hy + Jz + c$  and  $f$  becomes non-degenerate. Rather similar considerations also work in case  $EC = 0$ . After an affine transformation, the possibilities are:

- $[f]$  has a simple representative, hence  $L(f, T) = 1$ ;
- $L(f, T)$  is a polynomial of degree 1, pure of weight 3;
- $1/L(f, T)$  is a polynomial of degree 1 and weight 4, and a normal form for  $f$  is  $f = ax^3 + Ax^2 + Dxz + Gx + Jz + c$ ;
- $1/L(f, T)$  is a polynomial of degree 2, pure of weight 4, and a normal form for  $f$  is  $f = ax^3 + Ax^2 + Dxz + Ez^2 + Gx + Jz + c$  with  $E \neq 0$ ;
- $L(f, T)$  is a polynomial of degree 2, pure of weight 5, in which case  $f$  can be transformed in a cubic polynomial in  $x$  alone.

Case II: a line and a double line. After a linear transformation,  $C(x, y, z)$  has the normal form  $C(x, y, z) = x^2y$ . By translating  $y \mapsto y - A$  we may assume  $A = 0$  and

$$f = x^2y + Bxy + Cy^2 + Dxz + Ez^2 + Fyz + Gx + Hy + Jz + c.$$

(i) If  $EC \neq 0$  and  $F^2 - 4EC \neq 0$ , then  $f$  is non-degenerate. There are several cases depending on whether  $D$  or  $G$  is 0. In all cases (using [AS2], [AS3], [DL2]),  $L(f, \mathbb{A}^3, \psi, T)$  is a polynomial of degree 3, pure of weight 3.

(ii) In any case, if  $E = 0$  then  $z$  appears linearly and the sum becomes

$$S_m(\mathbb{A}^3, f, \psi) = q^m \sum_{Dx+Fz+J=0} \psi_m(x^2y + Bxy + Cy^2 + Hy + Gx + c).$$

There are now several subcases according to the vanishing of the coefficients, and they can be treated with the results of [AS1], [AS2], [AS3], [DL1], [DL2]. The possibilities are:

- $[f]$  has a simple representative, hence  $L(f, T) = 1$ ;
- $L(f, T)$  is a polynomial of degree 1 or 2, pure of weight 3;
- $L(f, T)$  is a polynomial of degree at most 3, pure of weight 4, in which case a normal form for  $f$  is  $f = x^2y + Bxy + Cy^2 + Hy + Gx + c$ .

(iii) The more interesting subcases occur when  $f$  is degenerate. If  $D \neq 0$  and  $C = F = 0$ , then  $f$  has the form

$$f = x^2y + Bxy + Dxz + Gx + Hy + Jz + c.$$

If  $HDJ \neq 0$ , then  $f$  is degenerate along the two-dimensional face containing  $\{(2, 1, 0), (1, 1, 0), (0, 1, 0), (1, 0, 1), (0, 0, 1)\}$  if and only if  $J^2 - JBD + HD^2 = 0$ . Making no hypotheses concerning the vanishing of  $J$  and  $H$ , assume  $J^2 - JBD + HD^2 = 0$ . Then we may write

$$f = (x + JD^{-1})(x + \alpha)y + D(x + JD^{-1})z + Gx + c$$

with  $\alpha = B - JD^{-1} \in k$  so that  $x \mapsto x - JD^{-1}$  produces  $\tilde{f} = x^2y + \tilde{B}xy + Dxz + Gx + \tilde{c}$ , and the quadratic transformation  $z \mapsto z - D^{-1}(x + B)y$  produces  $\tilde{f} = Dxz + Gx + \tilde{c}$ . Hence in this subcase  $1/L(f, T)$  is a polynomial of degree 1 and weight 4, and a normal form for  $f$  is  $f = x^2y + Bxy + Dxz + Gx + c$ .

The last subcase is  $D = J = 0$  and  $F \neq 0$ , and similar considerations yield the two possibilities:

- $[f]$  has a simple representative, so  $L(f, T) = 1$ ;
- $1/L(f, T)$  is a polynomial of degree 1 and weight 4, and a normal form for  $f$  is  $f = x^2y + Bxy + Fyz + Hy + c$ .

(iv) Finally, we assume  $E \neq 0$ . Consider first the case  $C = 0$ . Now  $y$  appears linearly so the sum becomes

$$S_m(\mathbb{A}^3, f, \psi) = q^m \sum_{x^2+Bx+H+Fz=0} \psi_m(Dxz + Ez^2 + Gx + Jz + c).$$

The possibilities which arise can all be handled using [AS2], [AS3], [DL2] and they lead to the conclusion that  $L(f, T)$  is a polynomial of degree at most 3, pure of weight 3.

Case III: three rational lines not through a point. After a linear transformation,  $C(x, y, z)$  has the normal form  $C(x, y, z) = xyz$ . By means of the translation  $(x, y, z) \mapsto (x, y, z) - (F, D, B)$ , we may assume  $f$  has the form

$$f = xyz + Ax^2 + Cy^2 + Ez^2 + Gx + Hy + Jz + c.$$

(i) If  $f$  is convenient, it is easy to see it is also non-degenerate. Then by [AS1], [DL1],  $L(f, T)$  is a polynomial of degree  $2 + r$  pure of weight 3, where

$r =$  number of non-zero elements of the set  $\{A, C, E\}$ .

(ii) If, however,  $A = G = 0$  (for example) then

$$\begin{aligned} S_m(\mathbb{A}^3, f, \psi) &= q^m \sum_{yz=0} \psi_m(Cy^2 + Ez^2 + Hy + Jz + c) \\ &= q^m \sum \psi_m(Cy^2 + Hy + c) \\ &\quad + q^m \sum \psi_m(Ez^2 + Jz + c) - q^m \psi_m(c). \end{aligned}$$

If  $EC \neq 0$  then  $L(f, T)$  is a polynomial of degree 3, mixed with weights  $(3, 3, 2)$ .

If  $E = 0$  and  $CJ \neq 0$ , or  $C = 0$  and  $EH \neq 0$  then  $L(f, T)$  is a polynomial of degree 2, mixed with weights  $(3, 2)$ .

If, however,  $E = J = 0$  and  $C \neq 0$ , or  $C = H = 0$  and  $E \neq 0$  then  $L(f, T)$  is a rational function, with numerator of degree 2 mixed with weights  $(3, 2)$  and denominator of degree 1 and weight 4, in which case a normal form for  $f$  is  $f = xyz + Cy^2 + Hy + c$ .

A similar analysis in the remaining cases yields the following possibilities:

- $L(f, T)$  is a polynomial of degree 1 with weight 2;
- $L(f, T)$  is a rational function, with numerator mixed of degree 1 with weight 2 and denominator of degree 1 and weight 4, in which case a normal form for  $f$  is  $f = xyz + Hy + c$ ;
- $L(f, T)$  is a rational function, with numerator mixed of degree 1 with weight 2 and denominator of degree 2, pure of weight 4, in which case a normal form for  $f$  is  $f = xyz + c$ .

Case IV: three rational lines through a point. After a linear transformation,  $C(x, y, z)$  has the normal form  $C(x, y, z) = axy(x + y)$ . We have

$$f(x, y, z) = axy(x+y) + Ax^2 + Bxy + Cy^2 + Dxz + Ez^2 + Fyz + Gx + Hy + Jz + c.$$

We break this case into two subcases depending on whether or not  $E$  is zero.

If  $E \neq 0$ , then the sum is degenerate if and only if either  $AD \neq 0$  and  $D^2 - 4AE = 0$ , or  $CF \neq 0$  and  $F^2 - 4CE = 0$ .

If  $E \neq 0$  and the sum is non-degenerate and convenient then  $L(f, T)$  is a polynomial pure of weight 3 of degree equal to  $3!V_3 - 2!V_2 + V_1 - 1$  where  $V_3 = \text{Vol}(\Delta_\infty(f))$  and  $V_i$  is the sum of the volumes of the intersections of  $\Delta_\infty(f)$  with the  $i$ -dimensional coordinate axes, with volume computed with respect to Haar measure in the axis  $\mathbb{R}^i$  normalized so that the fundamental domain of the induced lattice has measure 1. In all cases in which  $f$  is non-degenerate and convenient the degree is 4. If the sum is non-degenerate but not convenient then we may use the toric decomposition of  $\mathbb{A}^3$  together with the weight distribution for sums on tori to obtain that in all non-degenerate cases  $L(f, T)$  is a polynomial of degree 4 pure of weight 3.



We consider next the degenerate cases. In case (i)  $ADE \neq 0$  and  $D^2 - 4AE = 0$  then the change of variables  $z \mapsto z - \frac{D}{2E}x$  transforms  $f$  into

$$\tilde{f}(x, y, z) = axy(x + y) + \tilde{B}xy + Cy^2 + Ez^2 + Fxy + Gx + Hy + Jz + \tilde{c}$$

which is non-degenerate, and again  $L(f, T)$  is a polynomial of degree 4, pure of weight 3. Case (ii)  $CFE \neq 0$  and  $F^2 - 4CE = 0$  is entirely analogous. In case (iii), if  $ACFD \neq 0$  and both  $D^2 - 4AE = 0$  and  $F^2 - 4CE = 0$  then the transformation  $z \mapsto z - \frac{D}{2E}x - \frac{F}{2E}y$  takes  $f$  into

$$\tilde{f}(x, y, z) = axy(x + y) + \tilde{B}xy + Ez^2 + Gx + Hy + Jz + \tilde{c}$$

which is non-degenerate and again  $L(f, T)$  is of degree 4 pure of weight 3. In all cases then with  $E \neq 0$ ,  $L(f, T)$  is a polynomial of degree 4 pure of weight 3.

If instead  $E = 0$ , then  $z$  appears in  $f(x, y, z)$  either linearly or not at all. Hence

$$S_m(\mathbb{A}^3, f, \psi) = q^m \sum_{Dx+Fy+J=0} \psi_m(axy(x+y)+Ax^2+Bxy+y^2+Gx+Hy+c).$$

We only give the list of possibilities:

- $[f]$  has a simple representative, hence  $L(f, T) = 1$ ;
- $1/L(f, T)$  is a polynomial of degree 1, with weight 2;
- $L(f, T)$  is a polynomial of degree at most 2, pure of weight 3;
- $1/L(f, T)$  is a polynomial of degree 1 with weight 4, in which case a normal form for  $f$  is  $f = axy(x + y) + Ax^2 + Bxy + Dxz + Gx + c$ .

This completes the proof of Theorem 7. ■

In the Kloosterman example of Iwaniec [I], we consider

$$f(x, y, z) = a_0(x) + a_1(x)y + a_2(x)y^{-1} + a_3(x)z + a_4(x)z^{-1}.$$

In the case the  $a_i(x) \in k[x, x^{-1}]$  are Laurent polynomials we write

$$a_i(x) = \sum_{j=d_i^-}^{d_i^+} a_j^{(i)} x^j$$

with  $a_{d_i^+}^{(i)} a_{d_i^-}^{(i)} \neq 0$ , so that the degrees of  $a_i(x)$  are precisely  $d_i^+$  and  $d_i^-$ . It is convenient to write  $A_i = a_{d_i^+}^{(i)}$ ,  $\tilde{A}_i = a_{d_i^-}^{(i)}$ ,  $d_i = d_i^+$ ,  $\tilde{d}_i = d_i^-$ .

**THEOREM 8.** *Assume that the  $a_i(x)$  are polynomials in  $k[x]$  for  $i = 1, 2, 3, 4$  and that  $a_0(x) \in k[x, x^{-1}]$ . Assume further*

- (i) each of  $a_i(x)$ ,  $i = 1, \dots, 4$  has no multiple non-zero root in  $K$ ;
- (ii)  $a_1(x)a_2(x)$  and  $a_3(x)a_4(x)$  have no common non-zero root in  $K$ ;
- (iii) either
  - (a)  $d_1 + d_2 \neq d_3 + d_4$  or

- (b)  $d_1 + d_2 = d_3 + d_4 < 2d_0$  or
- (c)  $d_1 + d_2 = d_3 + d_4 \geq 2d_0$  and  $\varepsilon A_0 \pm 2\sqrt{A_2 A_1} \pm 2\sqrt{A_3 A_4} \neq 0$   
 (where  $\varepsilon = 1$  if  $2d_0 = d_1 + d_2$  and  $\varepsilon = 0$  if  $2d_0 < d_1 + d_2$ ).

Then  $f$  is non-degenerate and  $L(\mathbb{G}_m^3, f, \psi, T)$  is a polynomial of degree bounded by  $4(d_0 + \tilde{d}_0) + 2(d_1 + d_2 + d_3 + d_4)$  and weight bounded by 3. If  $\tilde{d}_0 < 0$ , then 0 is an interior point of  $\Delta_\infty(f)$  and  $L(\mathbb{G}_m^3, f, \psi, T)$  is pure of weight 3.

**Proof.** The function  $f$  has possible degeneracies in four 1-dimensional faces  $\{\tau_i\}_{i=1}^4$  (of  $\Delta_\infty(f)$ ) such that  $f^{(\tau_1)} = a_1(x)y$ ,  $f^{(\tau_2)} = a_2(x)y^{-1}$ ,  $f^{(\tau_3)} = a_3(x)z$ ,  $f^{(\tau_4)} = a_4(x)z^{-1}$ . The purpose of hypothesis (i) is to ensure non-degeneracy along the  $\tau_i$ .  $f$  also has possible degeneracies along four 2-dimensional faces  $\{\sigma_i\}_{i=1}^4$  such that  $f^{(\sigma_1)} = a_1(x)y + a_3(x)z$ ,  $f^{(\sigma_2)} = a_1(x)y + a_4(x)z^{-1}$ ,  $f^{(\sigma_3)} = a_2(x)y^{-1} + a_3(x)z$ ,  $f^{(\sigma_4)} = a_2(x)y^{-1} + a_4(x)z^{-1}$ . Hypothesis (ii) ensures  $f$  is non-degenerate along the  $\sigma_i$ . Finally, in the case  $d_1 + d_2 = d_3 + d_4$  there is a 2-dimensional face  $\sigma$  such that

$$f^{(\sigma)} = \varepsilon A_0 x^{d_0} + A_1 x^{d_1} y + A_2 x^{d_2} y^{-1} + A_3 x^{d_3} z + A_4 x^{d_4} z^{-1}.$$

The purpose of hypothesis (iii) is to ensure the non-degeneracy of  $f$  along the face  $\sigma$ . The result for almost all  $p$  then follows from [AS1]. The result for all  $p$  follows from [DL1]. ■

**Remark.** An analogous condition to (iii) for negative degrees  $\tilde{d}_i$  and coefficients  $\tilde{a}_i$  will allow one to consider the case in which all the  $a_i(x)$  are Laurent polynomials.

We illustrate the method of removing degeneracies which extends the applicability of Theorem 8 above to some cases in which the hypotheses above (imposed to ensure non-degeneracy) are relaxed. We will work with the specific case which actually occurs in Iwaniec’ work [I]. Let

$$f_{abc}(x, y, z) = ax \left( y + \frac{1}{y} \right) + bx \left( z + \frac{1}{z} \right) + cx + \frac{1}{x}$$

be defined over  $\mathbb{F}_q$ . We consider  $S_n(\mathbb{G}_m^3, f_{abc}, \psi)$  when  $abc \neq 0$ . It was already noted in [AS1] that if  $c \neq \pm 2a \pm 2b$ , then  $f_{abc}$  is non-degenerate with respect to  $\Delta_\infty(f_{abc})$ ,  $L(\mathbb{G}_m^3, f_{abc}, \psi, T)$  is a polynomial of degree 8, pure of weight 3. We are now interested in the exceptional cases. Let  $\varepsilon_1$  and  $\varepsilon_2$  be two choices of sign and assume  $c = 2a\varepsilon_1 + 2b\varepsilon_2$ . Then write

$$f_{abc}(x, y, z) = \frac{ax}{y}(y + \varepsilon_1)^2 + \frac{bx}{z}(z + \varepsilon_2)^2 + \frac{1}{x}.$$

If we set

$$g_{ab}(x, y, z, u, v, w, t) = ax \frac{u^2}{y} + bx \frac{v^2}{z} + \frac{1}{x} + w(u - y - \varepsilon_1) + t(v - z - \varepsilon_2),$$

then by orthogonality of characters

$$q^{2n} S_n(\mathbb{G}_m^3, f_{abc}, \psi) = S_n(\mathbb{G}_m^3 \times \mathbb{A}^4, g_{ab}, \psi)$$

where we understand that in the sum on the right the new variables  $(u, v, w, t)$  run over  $\mathbb{A}^4(\mathbb{F}_{q^n}) = (\mathbb{F}_{q^n})^4$  and  $(x, y, z)$  run over  $\mathbb{G}_m^3(\mathbb{F}_{q^n}) = (\mathbb{F}_{q^n}^*)^3$ . The advantage in going from  $f (= f_{abc})$  to  $g (= g_{ab})$  is that  $g$  is non-degenerate with respect to  $\Delta_\infty(g)$ . Furthermore, for every  $A \subseteq \{u, v, w, t\}$ , the Laurent polynomial  $g_A$  obtained from  $g$  by setting the variables in  $A$  to 0 is non-degenerate with respect to  $\Delta_\infty(g_A)$ . This suggests the following:

DEFINITION. Let  $f(\mathbf{x}), g_1(\mathbf{x}), \dots, g_r(\mathbf{x}) \in \mathbb{F}_q[\mathbf{x}, \mathbf{x}^{-1}]$ ,  $F(\mathbf{x}, \mathbf{u}) \in \mathbb{F}_q[\mathbf{x}, \mathbf{x}^{-1}, \mathbf{u}]$ , and assume  $f(\mathbf{x}) = F(\mathbf{x}, g_1(\mathbf{x}), \dots, g_r(\mathbf{x}))$ . Set  $G(\mathbf{x}, \mathbf{u}, \mathbf{v}) = F(\mathbf{x}, \mathbf{u}) + \sum_{i=1}^r v_i(u_i - g_i(\mathbf{x}))$  so that

$$q^{2nr} S_n(\mathbb{G}_m^N, f, \psi) = S_n(\mathbb{G}_m^N \times \mathbb{A}^{2r}, G, \psi).$$

If  $G_A$  is non-degenerate with respect to  $\Delta_\infty(G_A)$  for every  $A \subseteq \{u_1, \dots, u_r, v_1, \dots, v_r\}$ , we will say  $G$  removes the degeneracies of  $f$ .

THEOREM 9. Assume  $h(\mathbf{x}, \mathbf{u}) \in \mathbb{F}_q[\mathbf{x}, \mathbf{x}^{-1}, \mathbf{u}]$ . Let  $S = \{u_1, u_2, \dots, u_r\}$ . Assume that  $h_A$  is non-degenerate with respect to  $\Delta_\infty(h_A)$  for every  $A \subseteq S$ . Set  $M = N + r$ . Assume also that  $\dim \Delta_\infty(h_A) \geq M - 2|A|$ . Then

$$L(\mathbb{G}_m^N \times \mathbb{A}^r, h, \psi, T)^{(-1)^{M+1}}$$

is a polynomial of degree equal to  $\sum (-1)^{|A|} (M - |A|)! \text{Vol}(\Delta_\infty(h_A))$  where the sum runs over those  $A \subseteq S$  with  $\dim \Delta_\infty(h_A) = M - |A|$  and where the volume of  $\Delta_\infty(h_A)$  is computed in  $\mathbb{R}^{M-|A|}$  with respect to Haar measure normalized so that a fundamental domain of the induced lattice has measure 1. Finally, the  $L$ -series

$$L(\mathbb{G}_m^N \times \mathbb{A}^r, h, \psi, T)^{(-1)^{M+1}}$$

is mixed with reciprocal roots having weights less than or equal to  $M$ .

Proof. According to the Appendix in [AS1], there is an exact sequence of complexes of  $p$ -adic vector spaces with action of Frobenius  $F$  (reflecting the toric decomposition of  $\mathbb{G}_m^N \times \mathbb{A}^r$ ):

$$0 \rightarrow SK_\bullet \rightarrow K_\bullet^{(0)} \rightarrow K_\bullet^{(1)} \rightarrow \dots \rightarrow K_\bullet^{(r)} \rightarrow 0$$

with  $K_\bullet^{(i)} = \bigoplus_{A \subseteq S^{(i)}} K_\bullet^{(A)}$ ,  $S^{(i)}$  being the subsets of  $S$  of cardinality  $i$ , where

$$L(\mathbb{G}_m^N \times \mathbb{A}^r, h, \psi, T)^{(-1)^{M+1}} = \prod_{i=0}^M \det(I - TF|H_i(SK_\bullet))^{(-1)^i},$$

and

$$L(\mathbb{G}_m^{M-|A|}, h_A, \psi, T)^{(-1)^{M-|A|+1}} = \prod_{i=0}^{M-|A|} \det(I - TF|H_i(K_\bullet^{(A)}))^{(-1)^i}$$

for all  $A \subseteq S$ . We write

$$P_A(T) = \det(I - TF|H_0(K_\bullet^{(A)}))$$

for all  $A \subseteq S$ . Let  $\delta(A) = M - |A| - \dim \Delta_\infty(h_A)$ . Under the hypothesis of non-degeneracy of  $h_A$  we have

(i)  $P_A(T)$  is mixed with reciprocal roots having weights at most  $\dim \Delta_\infty(h_A)$ ,

(ii)  $\det(I - TF|H_i(K_\bullet^{(A)})) = P_A(q^i T)^{\binom{\delta(A)}{i}}$  for all  $0 \leq i \leq \delta(A)$ ,

(iii)  $\det(I - TF|H_i(K_\bullet^{(A)})) = 0$ , for  $\delta(A) < i$ .

Thus

$$L(\mathbb{G}_m^{M-|A|}, h_A, \psi, T)^{(-1)^{M-|A|+1}} = \prod_{i=0}^{\delta(A)} P_A(q^i T)^{(-1)^i \binom{\delta(A)}{i}}.$$

It is useful to break up the exact sequence of complexes into a system of short exact sequences

$$0 \rightarrow C_\bullet^{(i)} \rightarrow K_\bullet^{(i)} \rightarrow C_\bullet^{(i+1)} \rightarrow 0$$

for  $i = 0, \dots, r - 1$ , where  $C_\bullet^{(0)} = SK_\bullet$  and  $C_\bullet^{(i+1)} = \text{cokernel}(K_\bullet^{(i-1)} \rightarrow K_\bullet^{(i)})$  for  $i = 0, \dots, r - 1$  (we use the convention  $K_\bullet^{(-1)} = SK_\bullet$ ). Under the given hypotheses,  $H_i(K_\bullet^{(A)}) = 0$  for all  $i > \delta(A)$ . Our hypothesis that  $\dim \Delta_\infty(h_A) \geq M - 2|A|$  implies then that  $\delta(A) \leq |A|$ , so that

$$H_i(K_\bullet^{(j)}) = 0 \quad \text{for all } i > j.$$

Therefore, we prove by downward induction on  $j$  using the long exact sequence of homology associated with the preceding exact sequence that

$$H_i(C_\bullet^{(j)}) = 0 \quad \text{for all } i > j.$$

As a consequence,  $H_i(SK_\bullet) = 0$  for  $i \geq 1$  and the first assertion of the above theorem holds. For the second assertion, note that

$$L(\mathbb{G}_m^N \times \mathbb{A}^r, h, \psi, T)^{(-1)^{M+1}} = \prod_{A \subseteq S} \prod_{i=0}^{\delta(A)} P_A(q^i T)^{(-1)^{i+|A|} \binom{\delta(A)}{i}}$$

and the weights of  $P_A(q^i T)$  are bounded above by

$$2i + \dim \Delta_\infty(h_A) \leq 2\delta(A) + \dim \Delta_\infty(h_A) \leq M$$

again by the hypothesis on  $\dim \Delta_\infty(h_A)$ . ■

We consider the case of  $g = g_{ab}(x, y, z, u, v, w, t)$ . There are 10 faces of codimension 1 of  $\Delta_\infty(g)$  which do not contain the origin. It is not difficult to check that  $g$  is non-degenerate with respect to  $\Delta_\infty(g)$ . The proof uses the assumption that  $c \neq 0$ . (We remark that the hypothesis that  $a$  and  $b$  are not zero is entirely harmless. The sum only simplifies in the case  $a$  or  $b$  or both are zero.) It is immediate to verify that for all  $A \subseteq S = \{u, v, w, t\}$ ,

$$\dim \Delta_\infty(g_A) = 7 - |A| - |A \cap \{w, t\}|$$

so that the preceding theorem applies.

**THEOREM 10.** *If  $c = \pm 2a \pm 2b$ ,  $abc \neq 0$ , then  $L(\mathbb{G}_m^3, f_{abc}, \psi, T)$  is a polynomial of degree 7 all of whose reciprocal roots have weights at most 3. In fact, six of the reciprocal roots have weight 3 and one has weight 2.*

**Proof.** The first assertion follows directly from Theorem 9 applied to  $g_{ab}$ . In fact the polyhedra  $\Delta_\infty(g_A)$  are all simplicial with respect to the origin and the second assertion follows from the explicit distribution of weights [AS2], [DL2]. ■

### References

- [AS1] A. Adolphson and S. Sperber, *Exponential sums and Newton polyhedra: Cohomology and estimates*, Ann. of Math. 130 (1989), 367–406.
- [AS2] —, —, *Exponential sums on  $(\mathbb{F}_q^*)^n$* , in: Proc. Conf. Automorphic Forms and Analytic Number Theory, CRM, Montreal, 1989, 1–6.
- [AS3] —, —, *Exponential sums on  $\mathbb{G}_m^n$* , Invent. Math. 101 (1990), 63–79.
- [Ar] M. Artin, *On isolated rational singularities of surfaces*, Amer. J. Math. 88 (1966), 129–136.
- [B1] E. Bombieri, *On exponential sums in finite fields*, *ibid.*, 71–105.
- [B2] —, *On exponential sums in finite fields II*, Invent. Math. 47 (1978), 29–39.
- [BS] E. Bombieri and S. Sperber, *On the degree of Artin L-functions in characteristic  $p$* , C. R. Acad. Sci. Paris Sér. I 306 (1988), 393–398.
- [D] P. Deligne, *La conjecture de Weil II*, Inst. Hautes Études Sci. Publ. Math. 82 (1981), 313–428.
- [DL1] J. Denef et F. Loeser, *Polyèdres de Newton et poids de sommes exponentielles*, in:  *$p$ -Adic Analysis (Proceedings, Trento 1989)*, Lecture Notes in Math. 1454, Springer, Berlin, 217–222.
- [DL2] —, —, *Weights of exponential sums, intersection cohomology and Newton polyhedra*, Invent. Math. 106 (1991), 275–294.
- [Ha] H. Hasse, *Theorie der relativ-zyklischen algebraischen Funktionenkörper, insbesondere bei endlichem Konstantenkörper*, J. Reine Angew. Math. 172 (1934), 37–54.
- [H] C. Hooley, *On exponential sums and certain of their applications*, in: *Number Theory Days, 1980 (Exeter 1980)*, London Math Soc. Lecture Note Ser. 56, Cambridge Univ. Press, Cambridge, 1982, 92–122.

- [I] H. Iwaniec, *Small eigenvalues of Laplacian for  $\Gamma_0(N)$* , Acta Arith. 56 (1990), 65–82.
- [K] N. Katz, *Sommes exponentielles (rédigé par G. Laumon)*, Astérisque 79 (1980).
- [KL] N. Katz et G. Laumon, *Transformation de Fourier et majoration des sommes exponentielles*, Inst. Hautes Études Sci. Publ. Math. 62 (1986), 361–418.
- [L] S. Lang, *Abelian Varieties*, Interscience, New York, 1958.
- [LW] S. Lang and A. Weil, *Number of points of varieties in finite fields*, Amer. J. Math. 76 (1954), 819–827.
- [N] L. B. Nisnevich, *On the number of points of an algebraic manifold in a prime finite field*, Dokl. Akad. Nauk SSSR (N.S.) 99 (1954), 17–20 (in Russian).
- [R] M. Reid, *Undergraduate Algebraic Geometry*, London Math. Soc. Stud. Texts 12, Cambridge Univ. Press, 1988.
- [W] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Publ. Inst. Math. Univ. Strasbourg, 1945. Also Actualités Sci. Indust. 1041, Hermann, Paris, 1948.

INSTITUTE FOR ADVANCED STUDY  
PRINCETON, NEW JERSEY 08540  
U.S.A.

SCHOOL OF MATHEMATICS  
UNIVERSITY OF MINNESOTA  
MINNEAPOLIS, MINNESOTA 55455  
U.S.A.  
E-mail: SPERBER@VX.ACS.UMN.EDU

*Received on 18.11.1992  
and in revised form on 8.8.1994*

(2337)