ON KATZ’S \((A,B)\)-EXPONENTIAL SUMS

LEI FU AND DAQING WAN

Abstract. We deduce Katz’s theorems for \((A,B)\)-exponential sums over finite fields using \(\ell\)-adic cohomology and a theorem of Denef-Loeser, removing the hypothesis that \(A+B\) is relatively prime to the characteristic \(p\). In some degenerate cases, the Betti number estimate is improved using toric decomposition and Adolphson-Sperber’s bound for the degree of \(L\)-functions. Applying the facial decomposition theorem in [10], we prove that the universal family of \((A,B)\)-polynomials is generically ordinary for its \(L\)-function when \(p\) is in certain arithmetic progression.

Introduction

Let \(k\) be a finite field with \(q\) elements of characteristic \(p\) and let \(d\) be a positive integer. A polynomial \(f(t_1, \ldots, t_n)\) in \(k[x_1, \ldots, t_n]\) of degree \(d>0\) is called a Deligne polynomial if \(d\) is prime to \(p\), and the leading form \(f_d\) of \(f\) defines a smooth projective hypersurface \(f_d = 0\) in \(\mathbb{P}^{n-1}\). For positive integers \(A\) and \(B\), the \((A,B)\)-polynomial considered by Katz is a Laurent polynomial of the following form:

\begin{equation}
G(t_0, \ldots, t_n) := t_0^A f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B\left(\frac{1}{t_0}\right) \in k[t_0^\pm, t_1, \ldots, t_n]
\end{equation}

where \(f(t_1, \ldots, t_n)\) is a Deligne polynomial of degree \(d\), \(g(t_1, \ldots, t_n)\) is a polynomial of degree \(<d\), and \(P_B(s)\) is a one-variable polynomial of degree \(\leq B\). It can be viewed as a one-parameter family of Deligne polynomials parametrized by \(t_0\) over the torus \(\mathbb{G}_m\). We stress that for flexibility of applications, the degree of \(P_B\) is only assumed to be at most \(B\). Later on we distinguish the two cases when the degree is exactly \(B\) and when the degree is strictly less than \(B\).

Fix a nontrivial additive character \(\psi: (k,+) \to \mathbb{C}^*\). For a Deligne polynomial \(f(t_1, \ldots, t_n)\) as above, one has Deligne’s fundamental estimate ([3, 3.7.2.1])

\[ |\sum_{t \in k^n} \psi(f(t))| \leq (d-1)^n q^{n/2}. \]

Define the \(L\)-function by

\[ L(f, T) = \exp\left( \sum_{m=1}^{\infty} \frac{T^m}{m} \sum_{t \in k_m^n} \psi\left(\text{Tr}_{k_m/k}(f(t))\right) \right), \]

where \(k_m\) denotes the extension of \(k\) of degree \(m\). Deligne shows that \(L(f, T)^{(d-1)^n-1}\) is a polynomial of degree \((d-1)^n\) pure of weight \(n\). From the \(p\)-adic point of view, Sperber [9] further shows that the \(q\)-adic Newton polygon of the polynomial \(L(f, T)^{(d-1)^n-1}\) lies above a certain lower bound called the Hodge polygon, which is defined to be the \(q\)-adic Newton polygon of the polynomial

\[ h(T) = \prod_{1 \leq j_1, \ldots, j_n \leq d-1} (1 - q^{j_1 + \cdots + j_n} T). \]
These two polygons coincide for a generic Deligne polynomial $f$ of degree $d$ over $\bar{k}$ if $p \equiv 1 \mod d$, that is, the universal family of Deligne polynomials of degree $d$ in $n$ variables is generically ordinary for its $L$-function when $p \equiv 1 \mod d$. This is because for the diagonal polynomial

$$f(t) = t_1^d + \ldots + d_n^d,$$

the classical Stickelberger theorem for the Gauss sum implies that the Newton polygon equals to the Hodge polygon if $p \equiv 1 \mod d$.

Motivated by applications in analytic number theory, it is of interest to study the exponential sum for the above $(A, B)$-polynomial $G(t_0, \ldots, t_n)$. If one applies Deligne’s estimate fibre by fibre to the exponential sum for the $(A, B)$-polynomial, one gets the “trivial” bound

$$\left| \sum_{t_0 \in k^*} \sum_{t_1, \ldots, t_n \in k} \psi\left( G(t_0, \ldots, t_n) \right) \right| \leq (q - 1)(d - 1)^n q^{n/2} \leq (d - 1)^n q^{(n+2)/2}.$$

This is already of considerable depth, but is still weaker than the expected square root cancellation as $q$ varies. Using $\ell$-adic cohomology, Katz [7] proves the following optimal square root estimate for the $(A, B)$-exponential sum over $\mathbb{G}_m \times \bar{k}^n$.

**Theorem 0.1.** Suppose that $f$ is a Deligne polynomial of degree $d$ prime to $p$, $p$ is prime to $AB$, $\deg(P_B) = B$ and $\deg(g) < \frac{Bd}{A + B}$. For the $(A, B)$-polynomial $G(t_0, \ldots, t_n)$ defined in equation (0.0.1), we have the estimate

$$\left| \sum_{t_0 \in k^*} \sum_{t_1, \ldots, t_n \in k} \psi\left( G(t_0, \ldots, t_n) \right) \right| \leq (A + B)(d - 1)^n q^{(n+1)/2}.$$

By a standard reduction procedure, we can always reduce the exponential sum to the case where $AB$ is not divisible by $p$, unless $P_B(t_0)$ is reduced to a constant. Thus the condition that $p$ is prime to $AB$ is not essential. The condition $\deg(g) < \frac{Bd}{A + B}$ is necessary to ensure that the Betti number is bounded by $(A + B)(d - 1)^n$. There is an extra condition that $p$ is prime to $A + B$ in Katz’s original theorem. In this paper, we give a proof of the above theorem using a theorem of Denef-Loeser, removing this extra assumption that $p$ is prime to $A + B$.

If $\deg(P_B) = B$ and $f$ is affine Dwork regular (see section 2 for its definition), the relevant cohomology is pure (Katz’s $(A, B)$ purity theorem). In this case, we check that the $(A, B)$-polynomial $G(t_0, \ldots, t_n)$ is non-degenerate and commode (with respect to $t_1, \ldots, t_n$) so that we can apply the theorem of Denef-Loeser to deduce the purity and to calculate the exact Betti number. If $\deg(P_B) = B$ but $f$ is only assumed to be a Deligne polynomial, the $(A, B)$-polynomial $G(t_0, \ldots, t_n)$ may not be non-degenerate and Denef-Loeser’s theorem may not apply. In this case, the shifted polynomial $f(x) + a$ is affine Dwork regular for most $a \in \bar{k}$. Following Katz, we use a specialization and perverse argument to show that the same estimate remains true. The relevant cohomology in the degenerate case is mixed, and the number $(A + B)(d - 1)^n$ is only an upper bound for the Betti number.

Assume $\deg(P_B) = h \leq B$. Theorem 0.1 applies only when $\deg(g) = e < hd/(A + h)$ and we get

$$\left| \sum_{t_0 \in k^*} \sum_{t_1, \ldots, t_n \in k} \psi\left( G(t_0, \ldots, t_n) \right) \right| \leq (A + h)(d - 1)^n q^{(n+1)/2}.$$

It would be interesting to extend the theorem to the case when $dh/(A + h) \leq e < d$. This cannot be done in general, see [7] Remark 5.5. However, Katz has a trick to make it work if $f(x)$ is affine Dwork regular. His idea is to choose a larger $B$ so that $e < Bd/(A + B)$ and consider the family $P_B(t_0) + bt_0^B$ parametrized by $b \in \bar{k}$. Using a similar specialization and perverse argument, the
same theorem can still be proved. We obtain the following theorem, again proved first by Katz under the extra assumption that \( p \) is relatively prime to \( A + B \).

**Theorem 0.2.** Suppose that \( f \) is affine Dwork regular of degree \( d \) prime to \( p \), \( p \) is prime to \( AB \), \( \deg(P_B) \leq B \) and \( \deg(g) = e < \frac{Bd}{A + B} \). Then we have
\[
\left| \sum_{t_0 \in k^*} \sum_{t_1, \ldots, t_n \in k} \psi \left( G(t_0, \ldots, t_n) \right) \right| \leq (A + B)(d - 1)^n q^{(n+1)/2}.
\]

As mentioned above, if \( \deg(P_B) = h \) and \( e \geq hd/(A + h) \), then Theorem 0.1 does not apply. We can choose a larger \( B \geq h \) so that \( B \) is prime to \( p \) and \( e < Bd/(A + B) \). Then Theorem 0.2 will apply with this larger \( B \) at the expense of increasing the constant \( A + h \) to \( A + B \). For fixed \( A \), there are many choices of \((A, B)\) such that \( e < Bd/(A + B) \) and \( h \leq B \). We want to choose such \( B \) prime to \( p \) such that \( A + B \) (equivalently \( B \)) is as small as possible.

Consider the case \( h \leq 1 \), that is, \( P_B \) is at most a linear polynomial (possibly a constant). For each positive integer \( A \), the smallest positive integer \( B \) satisfying \( e < Bd/(A + B) \) is the smallest integer \( B \) that is greater than \( eA/(d - e) \). Thus the coefficient in the estimate of the above theorem can be taken to be
\[
(A + \left[ \frac{eA}{d - e} \right] + 1)(d - 1)^n.
\]

We can improve the Betti number estimate \((A + B)(d - 1)^n\) of Theorem 0.2 in the case \( e \geq hd/(A + h) \) (equivalently \( h \leq eA/(d - e) \)) as follows. The case \( e < hd/(A + h) \) is handled by Theorem 0.1 and the Betti number estimate is optimal already.

**Theorem 0.3.** Suppose that \( f \) is affine Dwork regular of degree \( d \) prime to \( p \), \( p \) is prime to \( A \), \( \deg(P_B) = h \), \( \deg(g) = e \), and \( d > e \geq hd/(A + h) \). Then we have
\[
\left| \sum_{t_0 \in k^*} \sum_{t_1, \ldots, t_n \in k} \psi \left( G(t_0, \ldots, t_n) \right) \right| \leq \left( \left( A + \frac{eA}{d - e} \right)(d - 1)^n - \left( \frac{eA}{d - e} - h \right)(e + 1)^n \right) q^{(n+1)/2}.
\]

Taking \( P_B = 0 \), we obtain

**Corollary 0.4.** Suppose that \( f(t) \) is affine Dwork regular of degree \( d \) prime to \( p \), and \( p \) is prime to \( A \). For any polynomial \( g(t) \in k[t_1, \ldots, t_n] \) of degree \( \deg(g) = e < d \), we have
\[
\left| \sum_{t_0 \in k^*} \sum_{t_1, \ldots, t_n \in k} \psi \left( t_0^A f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) \right) \right| \leq \left( \left( A + \frac{eA}{d - e} \right)(d - 1)^n - \frac{eA}{d - e} (e + 1)^n \right) q^{(n+1)/2}.
\]

As indicated above, if \( e < Bd/(A + B) \), the coefficient in this error term can be smaller than the coefficient \((A + B)(d - 1)^n\) in Theorem 0.2. This indicates that the estimate in Theorem 0.3 is generally better than the estimate in Theorem 0.2 in the case \( e \geq hd/(A + h) \).

We remark that this corollary cannot be deduced from the theorem of Denef-Loeser as the leading form \( g_0 \) of the polynomial \( g(t_1, \ldots, t_n) \) can be highly singular so that the Laurent polynomial \( G(t_0, \ldots, t_n) = t_0^A f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) \) can be highly degenerate. Our proof of Theorem 0.4 combines the cohomological consequence that the \( L \)-function or its reciprocal is a polynomial together with toric decomposition and Adolphson-Sperber’s bound for the degree of the \( L \)-functions of toric exponential sums.

In this paper, we also study the generic Newton polygon for the \( L \)-function associated to \((A, B)\)-exponential sums. A lower bound \( \text{HP}(\Delta) \), called the Hodge polygon, is given by Adolphson-Sperber [2] in terms of lattice points in “fundamental domain” of the convex polytope \( \Delta \) defined using the exponents of monomials in \( G(t_0, \ldots, t_n) \). To get a feeling what the Hodge polygon looks like, see the end of this paper for an explicit closed formula in the case \( A = B = 1 \). We are interested in
deciding when the generic Newton polygon coincides with its lower bound, i.e., when the universal family of \((A, B)\)-polynomials is generically ordinary for its \(L\)-function. Unlike the universal family of Deligne polynomials of degree \(d\), the polytope for the universal family of \((A, B)\)-polynomials is no longer a simplex. There is no elementary diagonal example available and the problem becomes deeper. We apply the facial decomposition theorem in [10] to prove the following result.

**Theorem 0.5.** For fixed positive integers \(d, A, B\) relatively prime to \(p\), and a non-negative integer \(e \leq dA/(A + B)\), the universal family of \((A, B)\)-polynomial \(G(t_0, \ldots, t_n)\) defined in (0.0.7) with \(\text{deg}(g) \leq e\) is generically ordinary for its \(L\)-function if \(p \equiv 1 \mod [A, dB]\), where \([A, dB]\) denotes the least common multiple of \(A\) and \(dB\).

When \(e > dA/(A + B)\), the polytope of the corresponding universal \((A, B)\)-polynomials is more complicated, having three (instead of two) codimension 1 faces not containing the origin, and one of them is not a simplex. The decomposition theorems in [10] can still be used to obtain similar results. The question is how to cleverly apply the various decomposition theorems to obtain an optimal (smallest) modulus \(D\) for the arithmetic progression \(p \equiv 1 \mod D\). We leave this to interested readers.

The paper is organized as follows. In Section 1, we prove Theorem 0.1 using Denef-Loeser’s results under the assumption that the Deligne polynomial \(f(t_0, \ldots, t_n)\) is actually affine Dwork regular. In this case, the related cohomology group is pure. In Section 2, we deduce Theorems 0.1 and 0.2 from the results in Section 1 using a specialization argument. In both sections, we actually work with the \((A, B)\)-exponential sums twisted by a multiplicative character. In Section 3, we prove Theorem 0.3 about the improvement of the constant in the bound. In Section 4, we study the generic ordinary property for the \((A, B)\)-polynomial and prove Theorem 0.5.

1. Non-degenerate and pure case

Let \(k\) be a finite field of characteristic \(p\) with \(q\) elements and let

\[
 f = \sum_{j=1}^{N} a_j t_1^{w_{1j}} \cdots t_n^{w_{nj}} \in k[t_1^{\pm 1}, \ldots, t_n^{\pm 1}]
\]

be a Laurent polynomial. Define the Newton polytope \(\Delta_\infty(f)\) of \(f\) at \(\infty\) to be the convex hull of \(\{0, w_1, \ldots, w_N\}\) in \(\mathbb{R}^n\), where \(w_j = (w_{1j}, \ldots, w_{nj}) \in \mathbb{Z}^n\). We say \(f\) is nondegenerate with respect to \(\Delta_\infty(f)\) if for any face \(\tau\) of \(\Delta_\infty(f)\) not containing the origin, the subscheme of \(\mathbb{G}^n_m\) defined by

\[
 \frac{\partial f_\tau}{\partial t_1} = \cdots = \frac{\partial f_\tau}{\partial t_n} = 0
\]

is empty, where

\[
 f_\tau = \sum_{w_j \in \tau} a_j t_1^{w_{1j}} \cdots t_n^{w_{nj}}.
\]

Suppose \(f \in k[t_1^{\pm 1}, \ldots, t_r^{\pm 1}, t_{r+1}, \ldots, t_n]\) is a polynomial with respect to the coordinates \(t_{r+1}, \ldots, t_n\). We say \(f\) is commode with respect to the coordinates \(t_{r+1}, \ldots, t_n\) if for any subset \(S \subset \{r+1, \ldots, n\}\), we have

\[
 \dim \left( \Delta_\infty(f) \cap \{ w_1, \ldots, w_n \in \mathbb{R}^n : w_j = 0 \text{ for all } j \in S \} \right) = n - \#S.
\]

Suppose \(f \in k[t_1, \ldots, t_n]\) is a polynomial of degree \(d\). Recall that \(f\) is a Deligne polynomial if \(d\) is relatively prime to \(p\), and the homogeneous degree \(d\) part \(f_d\) of \(f\) defines a smooth hypersurface
$f_d = 0$ in $\mathbb{P}^{n-1}$. A homogeneous polynomial $F(t_1, \ldots, t_n)$ is called Dwork-regular with respect to the coordinates $t_1, \ldots, t_n$ if the subscheme of $\mathbb{P}^{n-1}$ defined by

$$F = t_1 \frac{\partial F}{\partial t_1} = \cdots = t_n \frac{\partial F}{\partial t_n} = 0$$

is empty. If the degree $d$ of $F$ is prime to $p$, we may omit the condition $F = 0$ since

$$dF = \sum_i t_i \frac{\partial F}{\partial t_i}.$$ 

Again under the assumption that $(d, p) = 1$, $F$ is Dwork-regular if and only if for any nonempty subset $S \subset \{1, \ldots, n\}$, the homogeneous polynomial $F_\infty$ obtained from $F$ by setting $t_i = 0$ ($i \notin S$) is a nonzero Deligne polynomial in the variables $t_i$ ($i \in S$). Dwork shows that if $F$ is a homogenous Deligne polynomial, then there exists a finite extension $k'$ of $k$ and a matrix $(a_{ij}) \in \text{GL}(n, k')$ such that $F$ is Dwork-regular with respect to the coordinates $Y_i = \sum_j a_{ij} X_j$. Confer [7, Lemma 3.1].

**Proposition 1.1.** Let $f$ be a polynomial of degree $d$ prime to $p$, and let $f_d$ be the homogeneous degree $d$ part of $f$. If $f_d$ is Dwork regular, then $f$ is commode, its Newton polytope $\Delta_\infty(f)$ at $\infty$ is the simplex with vertices

$$(0, \ldots, 0), (d, 0, \ldots, 0), \ldots, (0, \ldots, 0, d),$$

and $f$ is nondegenerate with respect to $\Delta_\infty(f)$.

**Proof.** Suppose $f_d$ is Dwork regular. Then for each $j \in \{1, \ldots, n\}$, the coefficient of $t_j^d$ in $f$ is nonzero. Otherwise, let $P_j$ be the point in $\mathbb{P}^{n-1}$ whose only nonzero homogenous coordinate is the $j$-th. Then $P_j$ is a point in

$$f_d = t_1 \frac{\partial f_d}{\partial t_1} = \cdots = t_n \frac{\partial f_d}{\partial t_n} = 0.$$ 

So $f$ is commode and its Newton polytope $\Delta_\infty(f)$ at $\infty$ is the simplex with vertices

$$(0, \ldots, 0), (d, 0, \ldots, 0), (0, d, 0, \ldots, 0), \ldots, (0, \ldots, 0, d).$$

From the condition that $f_d$ is Dwork-regular, one deduces that $f$ is nondegenerate with respect to $\Delta_\infty(f)$. \qed

A polynomial $f$ of degree $d$ is called affine-Dwork-regular with respect to the coordinates $t_1, \ldots, t_n$ if the homogenization

$$F(t_0, t_1, \ldots, t_n) = t_0^d f\left(\frac{t_1}{t_0}, \ldots, \frac{t_n}{t_0}\right)$$

is Dwork-regular with respect to $t_0, t_1, \ldots, t_n$. This condition implies that the degree $d$ part $f_d$ of $f$ is Dwork-regular since $f_d = F(0, t_1, \ldots, t_n)$. In [7, Lemma 3.2], Katz shows that for any Deligne polynomial $f$ so that its leading form $f_d$ is Dwork-regular with respect to $t_1, \ldots, t_n$, the polynomial $f(x) + a$ is affine-Dwork-regular for all but finitely many $a \in \overline{\mathbb{F}}$.

Suppose $f(t_1, \ldots, t_n)$ is a polynomial of degree $d$ prime to $p$ and $g(t_1, \ldots, t_n)$ a polynomial of degree $< d$. Let $A$ and $B$ be positive integers and let $P_B(s)$ be a one-variable polynomial of degree $\leq B$. Let $G : \mathbb{G}_m \times \mathbb{A}^n \rightarrow \mathbb{A}^1$ be the morphism

$$(t_0, t_1, \ldots, t_n) \mapsto G(t_0, \ldots, t_n) = t_0^A f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B(1/t_0).$$

Choose a prime number $\ell$ distinct from $p$ and fix a nontrivial additive character

$$\psi : (k, +) \rightarrow \mathbb{G}_\ell.$$
Denote by $\mathcal{L}_\psi$ the Artin-Schreier sheaf on $\mathbb{A}^1$ corresponding to $\psi$. Let 

$$\chi : (k^*, \times) \to \mathbb{Q}_\ell$$

be a multiplicative character, let $\mathcal{K}_\chi$ be the associated Kummer sheaf on $\mathbb{G}_m$, and let 

$$\pi_1 : \mathbb{G}_m \times k^n \to \mathbb{G}_m$$

be the projection.

**Theorem 1.2.** Suppose that $f$ is an affine-Dwork-regular polynomial of degree $d$ prime to $p$, that $p$ is prime to $AB$, that $P_B$ is of degree $B$, and $\deg(g) < \frac{Bd}{A+B}$.

(i) We have $H^i_c(\mathbb{G}_m \times k^n, \pi_1^*\mathcal{K}_\chi \otimes G^*\mathcal{L}_\psi) = 0$ for $i \neq n + 1$, and

$$\dim H^{n+1}_c(\mathbb{G}_m \times k^n, \pi_1^*\mathcal{K}_\chi \otimes G^*\mathcal{L}_\psi) = (A + B)(d - 1)^n.$$

(ii) $H^{n+1}_c(\mathbb{G}_m \times k^n, \pi_1^*\mathcal{K}_\chi \otimes G^*\mathcal{L}_\psi)$ is pure of weight $n + 1$.

(iii) We have

$$\left| \sum_{t_0 \in k^*} \sum_{t_1, \ldots, t_n \in k} \chi(t_0)\psi \left( t_0^0 f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B(1/t_0) \right) \right| \leq (A + B)(d - 1)^n q^{(n+1)/2},$$

where $| \cdot |$ is the composite of an arbitrary isomorphism $\overline{\mathbb{Q}}_\ell \cong \mathbb{C}$ and the absolute value on $\mathbb{C}$.

**Proof.** We first treat the case where $\chi = 1$ is trivial. Since $f$ is affine-Dwork-regular, $\deg(P_B) = B$, and $\deg(g) < \frac{Bd}{A+B}$, the Laurent polynomial $t_0^d f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B(1/t_0)$ is commode in $t_1, \ldots, t_n$, its Newton polytope at $\infty$ is the simplex $\Delta$ in $\mathbb{R}^{n+1}$ with vertices 

$$(B, 0, \ldots, 0), (A, 0, 0, \ldots, 0), (A, d, 0, \ldots, 0), (A, d, 0, \ldots, 0), \ldots, (A, 0, \ldots, 0, d),$$

and it is nondegenerate with respect to $\Delta$. Here because $\deg(g) < \frac{Bd}{A+B}$, the exponents of $g$ lies in the interior of $\Delta$. For any subset $S \subset \{1, \ldots, n\}$, we have

$$\text{vol} \left( \Delta \cap \{ (t_0, t_1, \ldots, t_n) \in \mathbb{R}^{n+1} : t_i = 0 \text{ for all } i \in S \} \right) = \frac{1}{(n+1-\#S)!} t_0^{n-\#S}(A + B),$$

$$\sum_{S \subset \{1, \ldots, n\}} (-1)^{\#S} t_0^{n-\#S}(A + B) = (A + B)(d - 1)^n.$$

The assertions (i)-(ii) follows directly from [3, Theorem 9.2].

Next we consider the general $\chi$ case. Let $m = q - 1$, and let $[t_0^m] : \mathbb{G}_m \to \mathbb{G}_m$, $t_0 \mapsto t_0^m$.

We have

$$\bigoplus_{\chi : k^* \to \overline{\mathbb{Q}}_\ell} \mathcal{K}_\chi = [t_0^m]_* \overline{\mathbb{Q}}_\ell .$$

By the proper base change theorem, the projection formula and the Leray spectral sequence, we have

$$\bigoplus_{\chi : k^* \to \overline{\mathbb{Q}}_\ell} H^i_c(\mathbb{G}_m \times k^n, \pi_1^*\mathcal{K}_\chi \otimes G^*\mathcal{L}_\psi)$$

$$\cong H^i_c(\mathbb{G}_m \times k^n, \pi_1^*[t_0^m]_* \overline{\mathbb{Q}}_\ell \otimes G^*\mathcal{L}_\psi)$$

$$\cong H^i_c(\mathbb{G}_m \times k^n, ([t_0^m] \times \text{id}_k)^* G^*\mathcal{L}_\psi)$$

$$\cong H^i_c(\mathbb{G}_m \times k^n, \pi_1^* G^*\mathcal{L}_\psi)$$

$$\cong H^i_c(\mathbb{G}_m \times k^n, \mathcal{L}_\psi),$$
where $H$ is the morphism

$$H : \mathbb{G}_m \times \mathbb{A}^n \rightarrow \mathbb{A}^1, \quad (t_0, \ldots, t_n) \mapsto H(t_0, \ldots, t_n) = G(t_0^m, t_1, \ldots, t_n).$$

Applying the $\chi = 1$ case to the polynomial $H(t_0, \ldots, t_n)$ (and with $A$ and $B$ replaced by $mA$ and $mB$, respectively), we see that $H^i_c(\mathbb{G}_m \times \mathbb{A}^n, H^*L^\psi) = 0$ for $i \neq n + 1$,

$$\dim H^{n+1}_c(\mathbb{G}_m \times \mathbb{A}^n, H^*L^\psi) = m(A+B)(d-1)^n,$$

and $H^{n+1}_c(\mathbb{G}_m \times \mathbb{A}^n, H^*L^\psi)$ is pure of weight $n + 1$. Since $K^\chi$ is tamely ramified, we have

$$\chi_c(\mathbb{G}_m \times \mathbb{A}^n, K^\chi \otimes G^*L^\psi) = \chi_c(\mathbb{G}_m \times \mathbb{A}^n, G^*L^\psi)$$

for all $\chi$ by [3] 2.1. The assertions (i) and (ii) follow immediately. The assertion (iii) follows then from the Grothendieck trace formula

$$\sum \sum \chi(t_0)\psi\left(t_0^A f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B(1/t_0)\right)$$

$$= \sum (-1)^i \text{Tr}\left(\text{Fr}, H^i_c(\mathbb{A}^n, \mathbb{G}_m, \pi^*_c K^\chi \otimes G^*L^\psi)\right).$$

□

The following proposition is due to Deligne.

**Proposition 1.3.** Notation as above. Suppose $f(t_1, \ldots, t_n)$ is a Deligne polynomial of degree $d$ and suppose $g(t_1, \ldots, t_n)$ is of degree $< d$. Let $\pi_1 : \mathbb{G}_m \times \mathbb{A}^n \rightarrow \mathbb{G}_m$ be the projection. Then $R^i\pi_1 G^*L^\psi$ vanishes for $i \neq n$, and the sheaf $F = R^n\pi_1 G^*L^\psi$ is a lisse sheaf on $\mathbb{G}_m$ with rank $(d-1)^n$ and pure of weight $n$. We have

$$H^i_c(\mathbb{G}_m \times \mathbb{A}^n, F \otimes K^\chi) \cong H^{i+n}_c(\mathbb{G}_m \times \mathbb{A}^n, \pi^*_c K^\chi \otimes G^*L^\psi).$$

**Proof.** The first two assertions follow directly form [3] 3.7.3 and 3.7.2.3. Note that since we assume $\deg(g) < d$, for each parameter $t_0$, $t_0^A f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B(1/t_0)$ is a Deligne polynomials in the variables $t_1, \ldots, t_n$. By the projection formula, we have

$$R\Gamma_c(\mathbb{G}_m \times \mathbb{A}^n, \pi^*_c K^\chi \otimes G^*L^\psi) \cong R\Gamma_c(\mathbb{G}_m \times \mathbb{A}^n, R\pi_1(\pi^*_c K^\chi \otimes G^*L^\psi))$$

$$\cong R\Gamma_c(\mathbb{G}_m \times \mathbb{A}^n, R\pi_1 G^*L^\psi).$$

So we have a spectral sequence

$$H^i_c(\mathbb{G}_m \times \mathbb{A}^n, F \otimes R^i\pi_1 G^*L^\psi) \Rightarrow H^{i+n}_c(\mathbb{G}_m \times \mathbb{A}^n, \pi^*_c K^\chi \otimes G^*L^\psi),$$

and it degenerates by the above results of Deligne. So we have

$$H^i_c(\mathbb{G}_m \times \mathbb{A}^n, K^\chi \otimes \mathbb{F}) \cong H^{i+n}_c(\mathbb{G}_m \times \mathbb{A}^n, \pi^*_c K^\chi \otimes G^*L^\psi).$$

□

**Remark 1.4.** Deligne’s result [3] 3.7.2.3 also follows from [5] Theorem 9.2]. To see this, we may replace $k$ by any finite extension, or replace the coordinates $t_1, \ldots, t_n$ by any linear change. So by [7] Lemmas 3.1], we may assume the leading form $f_d$ of $f$ is Dwork-regular. Then $f$ is commode, the Newton polytope $\Delta_\infty(f)$ of $f$ at $\infty$ is the simplex with vertices

$$(0, \ldots, 0), (d, 0, \ldots, 0), \ldots, (0, \ldots, 0, d),$$
and $f$ is nondegenerate with respect to $\Delta_\infty(f)$. For any subset $S \subseteq \{1, \ldots, n\}$, we have
\[
\operatorname{vol}(\Delta_\infty(f) \cap \{(t_1, \ldots, t_n) \in \mathbb{R}^n : t_i = 0 \text{ for all } i \in S\}) = \frac{1}{(n-\#S)!} \frac{d^n-\#S}{(d-1)^n}.
\]

So [3, 3.7.2.3] follows from [5, Theorem 9.2].

As a direct consequence of Theorem 1.2 and Proposition 1.3, we have the following.

**Theorem 1.5** ([7] Theorems 5.1 and 8.1). Suppose that $f$ is an affine-Dwork-regular polynomial of degree $d$ prime to $p$, that $p$ is prime to $AB$, that $P_B$ is of degree $B$, and $\deg(g) < \frac{Bd}{A+B}$. Let $\mathcal{F} = R^n\pi_1^*G^*\mathcal{L}_\psi$.

(i) We have $H^i_c(\mathcal{G}_{m,k}, \mathcal{F} \otimes \mathcal{K}_\chi) = 0$ for $i \neq 1$, and
\[
\dim H^1_c(\mathcal{G}_{m,k}, \mathcal{F} \otimes \mathcal{K}_\chi) = (A+B)(d-1)^n.
\]

(ii) $H^1_c(\mathcal{G}_{m,k}, \mathcal{F} \otimes \mathcal{K}_\chi)$ is pure of weight $n + 1$.

**Remark 1.6.** There is no need to assume $p$ is prime to $A+B$ in [7] Theorems 5.1 and 8.1, Corollary 8.2], and no need to assume $p$ is odd in [7] Theorems 1.1, 2.1, 3.5].

2. Degenerate and mixed case

**Lemma 2.1.** Notation as above. For any $k$-point $a, b \in \mathbb{A}^1(k)$, let $[at^A]$ and $[bt^{-B}]$ be the morphisms
\[
\mathbb{G}_m \to \mathbb{A}^1, \quad t \mapsto at^A,
\mathbb{G}_m \to \mathbb{A}^1, \quad t \mapsto bt^{-B},
\]
respectively, and let
\[
\mathcal{F}_\chi = \mathcal{F} \otimes \mathcal{K}_\chi = R^n\pi_1^*G^*\mathcal{L}_\psi \otimes \mathcal{K}_\chi.
\]

In the triangulated category $D^b_c(\mathcal{G}_m, \mathbb{Q}_\ell)$ of complexes of $\mathbb{Q}_\ell$-sheaves on $\mathcal{G}_m$, we have
\[
\mathcal{F}_\chi \otimes [at^A]^*\mathcal{L}_\psi \cong R\pi_1^!(\pi_1^*\mathcal{K}_\chi \otimes G^*_a\mathcal{L}_\psi)[n],
\]
\[
\mathcal{F}_\chi \otimes [bt^{-B}]^*\mathcal{L}_\psi \cong R\pi_1^!(\pi_1^*\mathcal{K}_\chi \otimes G^*_b\mathcal{L}_\psi)[n],
\]
where $G_a, G_b : \mathbb{G}_m \times \mathbb{A}^n \to \mathbb{A}^1$ are the morphisms
\[
(t_0, t_1, \ldots, t_n) \mapsto t_0^A (f(t_1, \ldots, t_n) + a) + g(t_1, \ldots, t_n) + P_B(1/t_0),
(t_0, t_1, \ldots, t_n) \mapsto t_0^A f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + (P_B(1/t_0) + bt^{-B})
\]

**Proof.** By Proposition 1.3 we have
\[
\mathcal{F} = R\pi_1^!G^*\mathcal{L}_\psi[n].
\]

By the projection formula, we have
\[
\mathcal{F} \otimes \mathcal{K}_\chi \otimes [at^A]^*\mathcal{L}_\psi \cong R\pi_1^!G^*\mathcal{L}_\psi[n] \otimes \mathcal{K}_\chi \otimes [at^A]^*\mathcal{L}_\psi[n]
\cong R\pi_1^!\left(\pi_1^*\mathcal{K}_\chi \otimes G^*\mathcal{L}_\psi \otimes \pi_1^*[at^A]^*\mathcal{L}_\psi\right)[n].
\]

We then use the fact that
\[
G^*\mathcal{L}_\psi \otimes \pi_1^*[at^A]^*\mathcal{L}_\psi \cong G_a^*\mathcal{L}_\psi
\]
which follows from [3] Sommes trig. 1.7.1]. Similarly we can prove the second isomorphism. \qed
Lemma 2.2. Let $X$ be a geometrically connected smooth projective curve over $k$, $S$ a finite closed subscheme of $X$, $j : X - S \hookrightarrow X$ the canonical open immersion, and $F$ a pure lisse $\mathbb{Q}_l$-sheaf on $X - S$ of weight $w$. Suppose $H^1_c((X - S)_\bar{k}, F)$ is pure of weight $w + 1$ and $S(\bar{k})$ contains at least two points. Then we have $j_*F \cong j_*F$.

Proof. From the short exact sequence

$$0 \rightarrow j_*F \rightarrow j_*F \rightarrow j_*F/J_0F \rightarrow 0,$$

we get the exact sequence

$$0 \rightarrow H^0_c((X - S)_\bar{k}, F) \rightarrow H^0(X_{\bar{k}}, j_*F) \rightarrow H^0(j_*F/J_0F) \rightarrow H^1_c((X - S)_\bar{k}, F).$$

The following corollary proves Theorems 0.1 and 0.2 by taking $\chi = 1$.

Corollary 2.3. Suppose $d$ is prime to $p$, $p$ is prime to $AB$, and $\deg(q) < \frac{Bd}{AB}$. Suppose furthermore one of the following conditions holds:

(a) $f$ is a Deligne polynomial of degree $d$, and $\deg(P_B) = B$.

(b) $f$ is affine-Dwork-regular and $\deg P_B \leq B$.

Let $F = R^n\pi_1^*G^*\mathcal{L}_\psi$.

(i) We have $H^i_c(G_{\bar{k}}, F \otimes K_\chi) = 0$ for $i \neq 1$, and

$$\dim H^1_c(G_{\bar{k}}, F \otimes K_\chi) \leq (A + B)(d - 1)^n.$$

(ii) We have $H^i_c(G_{\bar{k}} \times \mathbb{A}^n, \pi_1^*K_\chi \otimes G^*\mathcal{L}_\psi) = 0$ for $i \neq n + 1$ and

$$\dim H^{n + 1}_c(G_{\bar{k}} \times \mathbb{A}^n, \pi_1^*K_\chi \otimes G^*\mathcal{L}_\psi) \leq (A + B)(d - 1)^n.$$

(iii) We have

$$\sum_{t_0 \in \mathbb{A}^1} \sum_{t_1, \ldots, t_n \in \mathbb{A}^1} \chi(t_0)\psi(t_0^d f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B(1/t_0)) \leq (A + B)(d - 1)^n q^{(n + 1)/2}.$$

Proof. (i) We first work under the condition (a). Let $pr_i : \mathbb{A}^1 \times \mathbb{A}^1 \rightarrow \mathbb{A}^1$ ($i = 1, 2$) be the projections and let

$$(\cdot, \cdot) : \mathbb{A}^1 \times \mathbb{A}^1 \rightarrow \mathbb{A}^1, \quad (t, t') \mapsto tt'$$
be the canonical pairing on $\mathbb{A}^1$. Recall that the Deligne-Fourier transform is the functor
\[ \hat{\mathcal{F}}: D^b_c(\mathbb{A}^1, \mathbb{Q}_\ell) \to D^b_c(\mathbb{A}^1, \overline{\mathbb{Q}}_\ell), \]
\[ \hat{\mathcal{F}}(K) = Rpr_2(\text{pr}_1^*K \otimes \langle . \rangle^* \mathcal{L}_\psi)[1]. \]

Let $[t^A]: \mathbb{G}_m \to \mathbb{G}_m$ be the finite étale morphism $t \mapsto t^A$ and let $j : \mathbb{G}_m \to \mathbb{A}^1$ be the canonical open immersion. Fix notation by the following diagram:
\[
\begin{array}{ccc}
\mathbb{G}_m \times \mathbb{A}^1 & \xrightarrow{[t^A] \otimes \text{id}} & \mathbb{G}_m \times \mathbb{A}^1 \\
\eta_1 \downarrow & & \downarrow \\
\mathbb{G}_m & \xrightarrow{j \otimes \text{id}} & \mathbb{A}^1 \times \mathbb{A}^1 \\
\end{array}
\]

By (2.3.1), we have
\[ (2.3.3) \]

Again let $\mathcal{F}_X = \mathcal{F} \otimes \mathcal{K}_\chi$. By the proper base change theorem and the projection formula, we have
\[
\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1]) \cong R\text{pr}_2(\text{pr}_1^*(j \circ [t^A])_!\mathcal{F}_X \otimes \langle . \rangle^* \mathcal{L}_\psi)[2]
\]
\[
\cong R\text{pr}_2\left((j \circ [t^A]) \times \text{id}\right)_!q_1^*\mathcal{F}_X \otimes \langle . \rangle^* \mathcal{L}_\psi)[2]
\]
\[
\cong R\text{pr}_2\left((j \circ [t^A]) \times \text{id}\right)_!(q_1^*\mathcal{F}_X \otimes ((j \circ [t^A]) \times \text{id})^* \langle . \rangle^* \mathcal{L}_\psi)[2]
\]
\[
\cong R\text{pr}_2\left(q_1^*\mathcal{F}_X \otimes [t^A]^{*}\mathcal{L}_\psi\right)[2],
\]

where $q_2 : \mathbb{G}_m \times \mathbb{A}^1 \to \mathbb{A}^1$ is the projection, and $[t^A]t'$ is the morphism
\[
\mathbb{G}_m \times \mathbb{A}^1 \to \mathbb{A}^1, \quad (t, t') \mapsto t^{A}t'.
\]

So for any geometric point $a \in \mathbb{A}^1(k)$, we have
(2.3.1) \[ \hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_a \cong R\Gamma(\mathbb{G}_m, \mathcal{F}_X \otimes [at^A]^* \mathcal{L}_\psi)[2] \]

Combined with Lemma (2.1) we get
(2.3.2) \[ \hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_a \cong R\Gamma(\mathbb{G}_m, \mathcal{F}_X \otimes [at^A]^* \mathcal{L}_\psi)[2] \]

Let $U$ be the Zariski open set of $\mathbb{A}^1$ so that for any $a \in U(k)$, $f(t_1, \ldots, t_n) + a$ is affine-Dwork-regular. Then by Theorem (1.2) (i), for any $a \in U(k)$, we have
\[
\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_a \cong \mathcal{H}^0_c(\mathbb{G}_m, \mathcal{F}_X \otimes [at^A]^* \mathcal{L}_\psi)[1],
\]
\[
\dim \hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_a = -(A + B)(d - 1)^n.
\]

The sheaf $j_!(t^A)_!\mathcal{F}_X[1]$ is a perverse sheaf. By [3] 1.3.2.3, $\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])$ is perverse, and its generic rank is $-(A + B)(d - 1)^n$. As a perverse sheaf, for any $s \in \mathbb{A}^1(k)$, the specialization homomorphism
\[ H^{-1}(\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_s) \to H^{-1}(\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_{\tilde{s}}) \]

is injective, where $\tilde{s}$ is the geometric generic point of $\mathbb{A}^1$. It follows that
(2.3.3) \[ \dim H^{-1}(\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_s) \leq (A + B)(d - 1)^n. \]

By (2.3.1), we have
\[ H^{-1}(\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_s) \cong H^1_c(\mathbb{G}_m, \mathcal{F}_X \otimes [st^A]^* \mathcal{L}_\psi). \]

Taking $s = 0$, we have
\[ H^{-1}(\hat{\mathcal{F}}(j_!(t^A)_!\mathcal{F}_X[1])_0) \cong H^1_c(\mathbb{G}_m, \mathcal{F}_X). \]

So by (2.3.3), we have
\[ \dim H^1_c(\mathbb{G}_m, \mathcal{F}_X) \leq (A + B)(d - 1)^n. \]
By (2.3.1) and (2.3.2), for any \( a \in U(\hat{k}) \), we have
\[
H^{-1}(\mathfrak{g}(j_{\ast}[A], \mathcal{F}_X[1])_a) \cong H^1_c(G_{m,k}, \mathcal{F}_X \otimes [at^A]^* \mathcal{L}_\psi) \cong H^{n+1}_c(G_{m,k} \times \mathbb{A}^n_k, \pi_1^* \mathcal{K}_X \otimes G^*_a \mathcal{L}_\psi).
\]
So \( H^1_c(G_{m,k}, \mathcal{F}_X \otimes [at^A]^* \mathcal{L}_\psi) \) is pure of weight \( n+1 \) by Theorem 1.2(ii). Let \( j : G_m \hookrightarrow \mathbb{P}^1 \) be the canonical open immersion. By Lemma 2.2 we have
\[
\bar{j}_!(\mathcal{F}_X \otimes [at^A]^* \mathcal{L}_\psi) \cong \bar{j}_*(\mathcal{F}_X \otimes [at^A]^* \mathcal{L}_\psi).
\]
Since \([at^A]^* \mathcal{L}_\psi\) is lisse on \( \mathbb{A}^1 \), this implies that
\[
\bar{j}_! \mathcal{F}_X = j_! \mathcal{F}_X.
\]
We have \( H^0(A^1, j_! \mathcal{F}_X) = 0 \) since \( \mathcal{F}_X \) is lisse on \( G_m \). So
\[
H^0(G_{m,k}, \mathcal{F}_X) \cong H^0(A^1, j_! \mathcal{F}_X) \cong H^0(A^1, j_! \mathcal{F}_X) = 0.
\]
Then by the Poincaré duality, we have \( H^2_c(G_{m,k}, \mathcal{F}_X) = 0 \). Finally \( H^0_c(G_{m,k}, \mathcal{F}_X) = 0 \) since \( G_m \) is affine.

Next we work under the condition (b). For any geometric point \( b \in A^1(\hat{k}) \), we have
\[
\mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_b \cong \text{RG}(G_{m,k}, \mathcal{F}_X \otimes [bt^{-B}]^* \mathcal{L}_\psi)[2] \cong \text{RG}(G_{m,k} \times \mathbb{A}^n_k, \pi_1^* \mathcal{K}_X \otimes G^*_a \mathcal{L}_\psi)[n+2].
\]
By Theorem 1.2 for any \( b \in G_m(\hat{k}) \), we have
\[
\mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_b \cong H^{n+1}_c(G_{m,k} \times \mathbb{A}^n_k, \pi_1^* \mathcal{K}_X \otimes G^*_a \mathcal{L}_\psi)[1],
\]
\[
\dim \mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_b = -(A+B)(d-1)^n.
\]
The sheaf \( \mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_b \) is perverse, and its generic rank is \(-(A+B)(d-1)^n\). The specialization homomorphism
\[
H^{-1}(\mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_0) \to H^{-1}(\mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_0)
\]
is injective. It follows that
\[
\dim H^{-1}(\mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_0) \leq (A+B)(d-1)^n.
\]
We have
\[
H^{-1}(\mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_0) \cong H^1_c(G_{m,k}, \mathcal{F}_X).
\]
So we have
\[
\dim H^1_c(G_{m,k}, \mathcal{F}_X) \leq (A+B)(d-1)^n.
\]
By (2.3.1), for any \( b \in G_m(\hat{k}) \), we have
\[
H^{-1}(\mathfrak{g}(j_{!!}[t^{-B}], \mathcal{F}_X[1])_b) \cong H^1_c(G_{m,k} \times \mathbb{A}^n_k, \pi^*_1 \mathcal{K}_X \otimes G^*_a \mathcal{L}_\psi)[2].
\]
So \( H^1_c(G_{m,k} \times \mathbb{A}^n_k, \mathcal{F}_X \otimes [bt^{-B}]^* \mathcal{L}_\psi) \) is pure of weight \( n+1 \) by Theorem 1.2(ii). Let \( j : G_m \hookrightarrow \mathbb{P}^1 \) and \( j' : G_m \hookrightarrow \mathbb{P}^1 - \{0\} \) be the canonical open immersions. By Lemma 2.2 we have
\[
\bar{j}!(\mathcal{F}_X \otimes [bt^{-B}]^* \mathcal{L}_\psi) \cong \bar{j}_*(\mathcal{F}_X \otimes [bt^{-B}]^* \mathcal{L}_\psi).
\]
Since \([bt^{-B}]^* \mathcal{L}_\psi\) is lisse on \( \mathbb{P}^1 - \{0\} \), this implies that
\[
\bar{j}! \mathcal{F}_X = j'_! \mathcal{F}_X.
\]
We have \( H^0(\mathbb{P}^1 - \{0\}, j'_! \mathcal{F}_X) = 0 \) since \( \mathcal{F}_X \) is lisse on \( G_m \). So
\[
H^0(G_{m,k}, \mathcal{F}_X) \cong H^0(\mathbb{P}^1 - \{0\}, j'_! \mathcal{F}_X) \cong H^0(\mathbb{P}^1 - \{0\}, j'_! \mathcal{F}_X) = 0.
\]
Then by the Poincaré duality, we have \( H^2_c(G_{m,k}, \mathcal{F}_X) = 0 \). Finally \( H^0_c(G_{m,k}, \mathcal{F}_X) = 0 \) since \( G_m \) is affine.
(ii) Follows from (i) and Proposition 1.3.
(iii) Follows from (ii), the Grothendieck trace formula, and Deligne’s theorem (the Weil conjecture). □

3. Improved degree bound

In this section, we prove Theorem 0.3. Let

\[ G(t_0, \ldots, t_n) = t_0^d f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_h(1/t_0) \in k[t_0^{\pm 1}, t_1, \ldots, t_n], \]

where \( f(t_1, \ldots, t_n) \) is a polynomial of degree \( d \), \( g(t_1, \ldots, t_n) \) is a polynomial of degree \( e < d \), and \( P_h(s) \) is a one-variable polynomial of degree \( h \). We shall assume that \( e \geq h d/(A + h) \), that is, \( h \leq e A/(d - e) \). For any positive integer \( m \), let \( k_m \) be the extension of \( k \) of degree \( m \). Define two exponential sums over \( k_m \) by

\[
S_m(G) = \sum_{t_0 \in k_m} \prod_{t_i \in k_m} \psi \left( \text{Tr}_{k_m/k}(G(t_0, \ldots, t_n)) \right),
\]

\[
S_m^*(G) = \sum_{t_0 \in k_m} \prod_{t_i \in k_m} \psi \left( \text{Tr}_{k_m/k}(G(t_0, \ldots, t_n)) \right).
\]

Their corresponding \( L \)-functions are defined by

\[
L(G, T) = \exp \left( \sum_{m=1}^{\infty} S_m(G) \frac{T^m}{m} \right), \quad L^*(G, T) = \exp \left( \sum_{m=1}^{\infty} S_m^*(G) \frac{T^m}{m} \right).
\]

They are rational functions. The degree \( \deg(L(G, T)) \) of \( L(G, T) \) is defined to be the degree of its numerator minus the degree of its denominator. For each subset \( S \subseteq \{1, \ldots, n\} \), let \( G_S \) denote the polynomial obtained from \( G \) by setting all \( t_i = 0 \) for \( i \in \{1, \ldots, n\} - S \). Thus, \( G_S \) is a Laurent polynomial in \( 1 + |S| \) variables. In a similar way, one defines the exponential sum \( S_m(G_S) \) over \( \mathbb{G}_m \times \mathbb{A}^{[S]} \), the exponential sum \( S_m^*(G_S) \) over \( \mathbb{G}_m^{1+[S]} \), and their \( L \)-functions \( L(G_S, T) \) and \( L^*(G_S, T) \). The toric decomposition of \( \mathbb{A}^n \) gives the decomposition

\[
S_m(G) = \sum_{S \subseteq \{1, \ldots, n\}} S_m^*(G_S), \quad L(G, T) = \prod_{S \subseteq \{1, \ldots, n\}} L^*(G_S, T).
\]

Let \( \Delta_1 \) be the simplex in \( \mathbb{R}^{n+1} \) with vertices

\((0, \ldots, 0), (A, 0, 0, \ldots, 0), \ldots, (A, 0, \ldots, 0, d),\)

let \( \Delta_2 \) be the simplex in \( \mathbb{R}^{n+1} \) with vertices

\((-h, 0, \ldots, 0), (0, 0, 0, \ldots, 0), \ldots, (0, 0, \ldots, e),\)

let \( \Delta_3 \) be the convex hull in \( \mathbb{R}^{n+1} \) of the points

\((0, 0, \ldots, 0), (A, 0, \ldots, 0, d), (0, 0, \ldots, 0, e), \ldots, (0, 0, \ldots, 0, e),\)

and let \( \Delta = \Delta_1 \cup \Delta_2 \cup \Delta_3 \). Then \( \Delta \) is a convex polytope in \( \mathbb{R}^{n+1} \), and the Newton polytope at \( \infty \) of the Laurent polynomial \( G(t_0, \ldots, t_n) \) is contained in \( \Delta \). By the degree bound of Adolphson-Sperber [11], we have

\[ |\deg(L^*(G, T))| \leq (n + 1)! \text{vol}(\Delta) = (n + 1)! (\text{vol}(\Delta_1) + \text{vol}(\Delta_2) + \text{vol}(\Delta_3)). \]

\( \Delta_1 \) and \( \Delta_2 \) are simplexes, and we have

\[(n + 1)! \text{vol}(\Delta_1) = Ad^n, \quad (n + 1)! \text{vol}(\Delta_2) = he^n.\]
The polytope $\Delta_3$ is not a simplex. Since $h \leq eA/(d-e)$, we can write $\Delta_3 = \Delta_4 \setminus \Delta_5$, where $\Delta_4$ is the simplex in $\mathbb{R}^{n+1}$ with vertices

$$(0, \ldots, 0), (-\frac{eA}{d-e}, 0, \ldots, 0), (A, d, 0, \ldots, 0), \ldots, (A, 0, \ldots, 0, d),$$

and $\Delta_5$ is the simplex in $\mathbb{R}^{n+1}$ with vertices

$$(0, \ldots, 0), (-\frac{eA}{d-e}, 0, \ldots, 0), (0, e, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, e).$$

We have $\Delta_5 \subseteq \Delta_4$. It follows that

$$(n+1)!\operatorname{vol}(\Delta_3) = (n+1)!\operatorname{vol}(\Delta_4) - (n+1)!\operatorname{vol}(\Delta_5) = \frac{eA}{d-e}d^n - \frac{eA}{d-e}e^n.$$

Here to calculate the volume of a simplex, we use the formula that if $\Sigma$ is a simplex in $\mathbb{R}^n$ with vertices $\{e_0, \ldots, e_n\}$, then $n!\operatorname{vol}(\Sigma) = |\det(A)|$, where $A$ is the matrix whose $i$-th row is given by the coordinates of $e_i - e_0$. Putting together, we obtain

$$|\deg(L^*(G,T))| \leq \left(A + \frac{eA}{d-e}\right)d^n - \left(\frac{eA}{d-e} - h\right)e^n.$$
4. Generic Newton polygon

Let $d, A, B$ be positive integers relatively prime to $p$. Consider the universal family of $(A, B)$-polynomials of the form

$$(4.0.1) \quad G(t) := t_0^d f(t_1, \ldots, t_n) + g(t_1, \ldots, t_n) + P_B(1/t_0) \in \overline{k}[t_0^\pm, t_1, \ldots, t_n]$$

where $f(t_1, \ldots, t_n)$ is a polynomial of degree $d$, $g(t_1, \ldots, t_n)$ is a polynomial of degree $< Ad/(A+B)$, and $P_B(s)$ is a one-variable polynomial of degree exactly $B$. Let $M(d, A, B, p)$ be the Zariski open dense subspace of such $(A, B)$-polynomials $G(t)$ satisfying the additional condition that $f$ is affine Dwork regular. It is non-empty as the polynomial $f(t) = 1 + t_1^d + \cdots + t_n^d$ is affine Dwork regular.

Suppose $G(t) \in M(d, A, B, p)$. It is non-degenerate with respect to its Newton polytope $\Delta$ at $\infty$, which is the simplex $\mathbb{R}^{n+1}$ with vertices

$$(-B, 0, \ldots, 0), (A, 0, \ldots, 0), (A, d, 0, \ldots, 0), \ldots, (A, 0, \ldots, 0, d).$$

By the work of Adolphson-Sperber [2], the $L$-function $L^*(G(t), T)^{(-1)^n}$ for the exponential sum over the torus $\mathbb{G}_m^{n+1}$ is a polynomial of degree $(A+B)d^n$, mixed of weights $\leq n+1$. Its Newton polygon lies above certain combinatorially defined lower bound called the Hodge polygon $\text{HP}^*(\Delta)$.

By the Grothendieck specialization theorem, the Newton polygon goes up under specialization. The generic Newton polygon exists for the family of $(A, B)$-polynomials $G(t) \in M(d, A, B, p)$. It is just the lowest possible Newton polygon as $G(t)$ varies in $M(d, A, B, p)$. Denote this generic Newton polygon by $\text{GNP}^*(d, A, B, p)$, which also lies above $\text{HP}^*(\Delta)$. If the two polygons coincide, we say that the family $M(d, A, B, p)$ is generically ordinary for its $L$-function over the torus $\mathbb{G}_m^{n+1}$. This property depends only on the four numbers $d, A, B, p$.

Similarly, for $G(t) \in M(d, A, B, p)$, the $L$-function $L(G(t), T)^{(-1)^n}$ for the exponential sum over $\mathbb{G}_m \times \mathbb{A}^n$ is a polynomial of degree $(A+B)(d-1)^n$, pure of weight $n+1$. Its Newton polygon lies above certain combinatorially defined lower bound called the Hodge polygon $\text{HP}(\Delta)$. The generic Newton polygon is denoted by $\text{GNP}(d, A, B, p)$, which also lies above $\text{HP}(\Delta)$. If the two polygon coincides, we say that the family $M(d, A, B, p)$ is generically ordinary for its $L$-function over $\mathbb{G}_m \times \mathbb{A}^n$. This property again depends only on the four numbers $d, A, B, p$.

Note that our family $M(d, A, B, p)$ can be strictly smaller than the universal family $M(\Delta, p)$ of non-degenerate and commode (with respect to $t_1, \ldots, t_n$) Laurent polynomials whose Newton polytope at $\infty$ is the given $\Delta$, as $\Delta$ may contain some lattice points which do not arise from exponents of the terms in $G(t)$. For this reason, proving generic ordinarity for this smaller family $M(d, A, B, p)$ can be somewhat harder than that for the larger family $M(\Delta, p)$. We prove that this is indeed true if $p \equiv 1 \mod [A, dB]$, where $[A, dB]$ denotes the least common multiple.

That is, we have

**Theorem 4.1.** If $p \equiv 1 \mod [A, dB]$, then we have

$$\text{GNP}^*(d, A, B, p) = \text{HP}^*(\Delta), \quad \text{GNP}(d, A, B, p) = \text{HP}(\Delta).$$

**Proof.** The first assertion is stronger. It implies the second assertion, as the second assertion is a portion of the first assertion by the boundary decomposition theorem in [10, Section 5]. To prove the first assertion, we apply the various decomposition theorems in [10, 11].

Let $\Delta_1$ be the simplex in $\mathbb{R}^{n+1}$ with vertices

$$(0, \ldots, 0), (A, 0, \ldots, 0), (A, d, 0, \ldots, 0), \ldots, (A, 0, \ldots, 0, d).$$

Let $\Delta_2$ be the simplex in $\mathbb{R}^{n+1}$ with vertices

$$(-B, 0, \ldots, 0), (0, \ldots, 0), (A, d, 0, \ldots, 0), \ldots, (A, 0, \ldots, 0, d).$$
It is clear that $\Delta$ is the union of $\Delta_1$ and $\Delta_2$. This is the facial decomposition (\cite{10} Section 5) of $\Delta$. The restriction of our universal family $G(t)$ to the unique codimension 1 face of $\Delta_1$ not containing the origin is the following family of polynomials

$$G_1(t) = t_0^A f(t_1, ..., t_n),$$

where $f$ is a polynomial of degree $d$. The Newton polytope at $\infty$ of this family is precisely $\Delta_1$. The restriction of our universal family $G(t)$ to the unique codimension 1 face of $\Delta_2$ not containing the origin is the following family of Laurent polynomials

$$G_2(t) = t_0^A f_d(t_1, ..., t_n) + b t_0^{-B},$$

where $f_d$ is the leading form of $f$, and $b$ is the leading coefficient of $P_B$. The Newton polytope at $\infty$ of this family is precisely $\Delta_2$.

The facial decomposition theorem in \cite{10} Theorem 5.5 says that $G(t)$ is ordinary with respect to $\text{HP}^*(\Delta)$ if and only if $G_i(t)$ is ordinary with respect to $\text{HP}^*(\Delta_i)$ for each $i \in \{1, 2\}$. In particular, for the ordinary property, as long as $\deg(g) < Ad/(A + B)$, the polynomial $g(t_1, ..., t_n)$ plays no role as its exponents do not lie on any of the two codimension 1 faces of $\Delta$ not containing the origin. Similarly, the lower degree terms in $P_B(t)$ and in $f(t)$ are irrelevant as far as the ordinarity property is concerned.

Now, the first family $G_1(t)$ is generically ordinary with respect to $\text{HP}^*(\Delta_1)$ under the condition $p \equiv 1 \mod A$. This is proved in \cite{10} Theorem 7.5 using a sequence of parallel hyperplane decompositions. In the special case $A = 1$, it implies that the zeta function of the universal family of toric (or affine or projective) hypersurfaces of degree $d$ is generically ordinary for every prime $p$ and every $n$, a highly nontrivial result already. Using a similar sequence of parallel hyperplane decompositions, one deduces that the second family $G_2(t)$ is generically ordinary with respect to $\text{HP}^*(\Delta_2)$ under the condition $p \equiv 1 \mod dB$. Putting together, we obtain Theorem 4.1. □

**Remark 4.2.** As indicated above, both $\Delta_1$ and $\Delta_2$ are simplexes. Instead of using the hyperplane decomposition theorem, an alternative easier approach is to choose an elementary diagonal example (the number of nonzero terms equals the number of variables) for each family $G_i(t)$ $(1 \leq i \leq 2)$ to compute its Newton polygon. A diagonal example in the family $G_1(t)$ is the non-degenerate polynomial

$$G_1^{(0)}(t) = t_0^A (1 + t_1^d + \cdots + t_n^d).$$

The matrix of its exponents is a square matrix with the largest invariant factor $[d, A]$ and thus $G_1^{(0)}(t)$ is ordinary if $p \equiv 1 \mod [d, A]$ by \cite{11} Corollary 2.6. This gives a weaker result for the first family than what can be obtained by using the hyperplane decomposition, but it gives the same result as Theorem 4.1. Similarly, a diagonal example in the family $G_2(t)$ is the non-degenerate polynomial

$$G_2^{(0)}(t) = t_0^A (t_1^d + \cdots + t_n^d) + t_0^{-B}.$$ 

The matrix of its exponents is a square matrix with the largest invariant factor dividing $dB$ and thus $G_2^{(0)}(t)$ is ordinary if $p \equiv 1 \mod dB$ by \cite{11} Corollary 2.6. It follows that the total family $G(t)$ is generically ordinary if $p \equiv 1 \mod [A, dB]$.

**Remark 4.3.** Given an integral $(n + 1)$-dimensional convex polytope in $\mathbb{R}^{n+1}$ containing the origin, we can consider the universal family $M(\Delta, p)$ of all non-degenerate Laurent polynomials whose Newton polytope at $\infty$ is $\Delta$. Its generic Newton polygon over the torus $\mathbb{G}_m^{n+1}$ is denoted by $\text{GNP}^*(\Delta, p)$, which depends only on $\Delta$ and $p$. Adolphson-Sperber’s work implies that $\text{GNP}^*(\Delta, p)$
lies above a certain explicit lower bound $\mathrm{GNP}^*(\Delta)$, called the Hodge polygon. They conjectured [2] that

$$\mathrm{GNP}^*(\Delta, p) = \mathrm{HP}^*(\Delta) \text{ if } p \equiv 1 \mod D(\Delta),$$

where $D(\Delta)$ is the denominator of $\Delta$. It is not hard to prove that the condition $p \equiv 1 \mod D(\Delta)$ is necessary (and thus optimal) for the conjecture to be true, either by a direct combinatorial proof or by a ramification argument. The Adolphson-Sperber conjecture is false in general, but true in many importance cases as shown in [10][11], including notably the above $G_1(t)$. We expect that this conjecture is true for the above $\Delta$ defined by the $(A,B)$-polynomial. One checks that $D(\Delta_1) = A$ since the unique codimension 1 face of $\Delta_1$ not containing the origin is defined by the hyperplane equation $\frac{1}{D} t_0 = 1$. Similarly, one checks that $D(\Delta_2) = [B, dB/(A + B, dB)]$ since the unique codimension 1 face of $\Delta_2$ not containing the origin is defined by the hyperplane equation

$$-\frac{1}{B} t_0 + \frac{A + B}{dB} t_1 + \cdots + \frac{A + B}{dB} t_n = 1.$$  

It follows that

$$D(\Delta) = [D(\Delta_1), D(\Delta_2)] = [A, B, dB/(A + B, dB)].$$

For the $\Delta$ defined using the $(A,B)$-polynomials, the Adolphson-Sperber conjecture says that $\mathrm{GNP}^*(\Delta, p) = \mathrm{HP}^*(\Delta)$ if $p \equiv 1 \mod [A, B, dB/(A + B, dB)]$. We expect this to be true. It is sufficient to prove it for the second piece $\Delta_2$ as the first piece $\Delta_1$ is already known as seen above. However, we do not expect that the condition $p \equiv 1 \mod [A, dB]$ in Theorem [4] can be relaxed to $p \equiv 1 \mod [A, B, dB/(A + B, dB)]$, as the family $M(d, A, B, p)$ can be significantly smaller than the family $M(\Delta, p)$.

Although the recursive combinatorial definition of the Hodge numbers in $\mathrm{HP}^*(\Delta)$ and $\mathrm{HP}(\Delta)$ as given in [2] are not complicated, a simple explicit formula for the Hodge numbers can be cumbersome to obtain. To give an indication of what the generic slopes look like, we give, without proof, an explicit formula for the Hodge numbers and thus the Hodge polygon for our $(A,B)$-polytope $\Delta$. For simplicity of notations, we shall assume that $A = B = 1$.

Define

$$H^*(T) = \prod_{0 \leq j_1, \ldots, j_d \leq d-1} \left( 1 - q^{\frac{1}{n_1} + \cdots + \frac{1}{n_d} + \left( \frac{1}{n_1} + \cdots + \frac{1}{n_d} \right) T} \right) \left( 1 - q^{\frac{1}{n_1} + \cdots + \frac{1}{n_d} + 1 - \left( \frac{1}{n_1} + \cdots + \frac{1}{n_d} \right) T} \right),$$

$$H(T) = \prod_{1 \leq j_1, \ldots, j_d \leq d-1} \left( 1 - q^{\frac{1}{n_1} + \cdots + \frac{1}{n_d} + \left( \frac{1}{n_1} + \cdots + \frac{1}{n_d} \right) T} \right) \left( 1 - q^{\frac{1}{n_1} + \cdots + \frac{1}{n_d} + 1 - \left( \frac{1}{n_1} + \cdots + \frac{1}{n_d} \right) T} \right),$$

where \( \{ r \} = r - [r] \) denotes the fractional part of $r$. It is clear that $H^*(T)$ is a polynomial of degree $(A + B)d^n = 2d^n$. Similarly, $H(T)$ is a polynomial of degree $(A + B)(d - 1)^n = 2(d - 1)^n$, whose slopes are symmetric in the interval $[0, n + 1]$. In the case $A = B = 1$, the Hodge polygon $\mathrm{HP}^*(\Delta)$ (resp. $\mathrm{HP}(\Delta)$) is simply the $q$-adic Newton polygon of $H^*(T)$ (resp. $H(T)$). Note that the slopes of $H^*(T)$ and $H(T)$ are rational numbers with denominators dividing $d$. The coefficients of the $L$-function lie in the $p$-th cyclotomic field $\mathbb{Q}(\zeta_p)$ which is totally ramified of degree $p - 1$ over $p$. This explains the congruence condition $p \equiv 1 \mod d$ of Theorem [4] in the case $A = B = 1$.

Finally, if our $(A,B)$-exponential sum is twisted by a multiplicative character $\chi$ of order $m$ dividing $q - 1$, then the generic Newton polygon for the corresponding twisted $L$-function over the torus $\mathbb{G}_m^{n+1}$ (resp., over $\mathbb{G}_m \times \mathbb{A}^n$) lies above $\mathrm{HP}^*(\Delta)$ (resp. over $\mathrm{HP}(\Delta)$). Furthermore, these two polygons coincide if $p \equiv 1 \mod [mA, dmB]$. One simply applies the above theorem to the $(mA, mB)$-polynomial $G(t_0^n, t_1^1, \ldots, t_n^1)$ and decomposes in terms of the multiplicative characters $\chi$ of order dividing $m$. 
ON KATZ’S \((A, B)\)-EXPONENTIAL SUMS

References

[1] A. Adolphson and S. Sperber, Newton polyhedra and the degree of the \(L\)-function associated to an exponential sum, *Invent. Math.* 88 (1987), no. 3, 555-569.

[2] A. Adolphson and S. Sperber, Exponential sums and Newton polyhedra: cohomology and estimates, *Ann. of Math.* (2) 130 (1989), no. 2, 367-406.

[3] P. Deligne, La conjecture de Weil II, *Publ. Math. IHES.*, 52 (1981), 313-428.

[4] P. Deligne et al, Cohomologie étale (SGA 4\(\frac{1}{2}\)), *Lecture Notes in Math.* 569, Springer-Verlag (1977).

[5] J. Denef and F. Loeser, Weight of exponential sums, intersection cohomology and Newton polyhedra, *Invent. Math.* 106 (1991), 275-294.

[6] L. Illusie, Théorie de Brauer et Caractéristique d’Euler-Poincaré, in *Caractéristique d’Euler-Poincaré, Astérique* 82-83 (1981), 161-172.

[7] N. Katz, On a question of Browning and Heath-Brown, In *Analytic Number Theory*, 267-288, Cambridge Univ. Press, Cambridge, 2009.

[8] G. Laumon, Transformation de Fourier, constantes d’équations fontionnelles, et conjecture de Weil, *Publ. Math. IHES* 65 (1987), 131-210.

[9] S. Sperber, On the \(p\)-adic theory of exponential sums, *Amer. J. Math.*, 109(1986), 255-296.

[10] D. Wan, Newton polygons of zeta functions and \(L\)-functions, *Ann. of Math.* (2) 137 (1993), no. 2, 249-293.

[11] D. Wan, Variation of \(p\)-adic Newton polygons for \(L\)-functions of exponential sums, *Asian J. Math.* 8 (2004), no. 3, 427-471.

Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, P. R. China
*E-mail address: leifu@math.tsinghua.edu.cn*

Department of Mathematics, University of California, Irvine, CA 92697, USA
*E-mail address: dwan@math.uci.edu*