



On inverted Kloosterman sums over finite fields

Xin Lin¹ · Daqing Wan²

Received: 12 January 2023 / Accepted: 8 January 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

The classical n -variable Kloosterman sums over finite fields are well understood by Deligne's theorem from complex point of view and by Sperber's theorem from p -adic point of view. In this paper, we study the complex and p -adic estimates of *inverted* n -variable Kloosterman sums, addressing a question of Katz (Finite Fields Appl 1(3):395–398, 1995). We shall give two complex estimates. The first one is elementary based on Gauss sums. The second estimate is deeper, depending on the cohomological results of Adolphson–Sperber, Denef–Loeser and Fu for twisted toric exponential sums. This deeper result assumes that the characteristic p does not divide $n + 1$. Combining with Dwork's p -adic theory, we also determine the exact p -adic valuations for zeros and poles of the L-function associated to *inverted* n -variable Kloosterman sums in the case $p \equiv 1 \pmod{n + 1}$. As we shall see, the *inverted* n -variable Kloosterman sum is more complicated than the classical n -variable Kloosterman sum in all aspects in the sense that our understanding is less complete, partly because the Hodge numbers are now mostly 2 instead of 1.

Keywords Inverted Kloosterman sums · Exponential sums · L-function · Finite field

Mathematics Subject Classification 11T23 · 11L05 · 11L07 · 11S40

1 Introduction

Let $\mathbb{F}_q$ be the finite field of q elements with characteristic p . Let $\psi : \mathbb{F}_q \rightarrow \mathbb{C}^*$ be a nontrivial additive character and let $\chi_1, \dots, \chi_m : \mathbb{F}_q^* \rightarrow \mathbb{C}^*$ be multiplicative characters. A classical

X. Lin is supported by the National Natural Science Foundation of China (No. 12301010). D. Wan is partially supported by NSF.

✉ Xin Lin
xlin1126@hotmail.com

Daqing Wan
dwan@math.uci.edu

¹ Department of Mathematics, Shanghai Maritime University, Shanghai 201306, People's Republic of China

² Department of Mathematics, University of California, Irvine, CA 92697-3875, USA

problem in number theory is to give a good estimate for the mixed character sum

$$\sum_{x_i \in \mathbb{F}_q^*} \chi_1(f_1) \cdots \chi_m(f_m) \psi(g), \quad (1.1)$$

where $g, f_1, \dots, f_m \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ are Laurent polynomials. Reducing m if necessary, we may assume that all the χ_i 's are non-trivial. Using the Gauss sum

$$G(\chi) = \sum_{x \in \mathbb{F}_q^*} \chi(x) \psi(x), \quad (1.2)$$

one obtains the well known relation

$$\begin{aligned} & \sum_{x_i, y_j \in \mathbb{F}_q^*} \overline{\chi_1}(y_1) \cdots \overline{\chi_m}(y_m) \psi(g + y_1 f_1 + \cdots + y_m f_m) \\ &= G(\overline{\chi_1}) \cdots G(\overline{\chi_m}) \sum_{x_i \in \mathbb{F}_q^*} \chi_1(f_1) \cdots \chi_m(f_m) \psi(g). \end{aligned}$$

As the Gauss sums are well-understood, the study of (1.1) can be reduced to the study of sum of the form on the left-hand side. Thus, in this paper, we focus on the following type of twisted toric exponential sum

$$\sum_{x_i \in \mathbb{F}_q^*} \chi_1(x_1) \cdots \chi_n(x_n) \psi(f), \quad (1.3)$$

where some of the multiplicative characters χ_i may be trivial and $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ is a Laurent polynomial. This type of twisted toric exponential sum has been studied extensively in the literature, most notably by Adolphson–Sperber [2, 4, 5, 7] via Dwork's p -adic cohomology, and by Denef–Loeser [10] and Fu [12, 13] via Grothendieck's ℓ -adic cohomology. For the complex estimate, both approaches depend on Deligne's theorem on the Weil conjectures. A sharp estimate is obtained when f is non-degenerate with respect to its Newton polyhedron $\Delta(f)$. For arbitrary f , the sum is still far from well understood.

An important example of toric exponential sums is the classical n -variable Kloosterman sum, where $\chi_1 = \cdots = \chi_n = 1$ and

$$f(x_1, \dots, x_n) = x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n}, \quad b \in \mathbb{F}_q^*.$$

In this case, the complex weights were determined by Deligne's well known theorem, and the p -adic slopes were determined by Sperber's theorem [22]. It should be noted that for twisted n -variable Kloosterman sum (when some of the χ_i 's are non-trivial), the p -adic slopes are not completely determined in general, except in the case $n = 1$ for which Adolphson–Sperber [3] obtained the generic p -adic slopes. If we invert the above Laurent polynomial and consider the following rational function

$$f(x_1, \dots, x_n) = \frac{1}{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n}}, \quad b \in \mathbb{F}_q^*,$$

which is no longer a Laurent polynomial, we are led to the so-called *inverted* Kloosterman sum. The study of such sums goes back to Katz [15].

More precisely, in this paper, we study the following *twisted inverted n -variable Kloosterman sum* defined by

$$\begin{aligned} S_n(\chi, b) &= \sum_{\substack{x_1 \cdots x_{n+1} = b, \ x_i \in \mathbb{F}_q^* \\ x_1 + \cdots + x_{n+1} \neq 0}} \chi_1(x_1) \cdots \chi_{n+1}(x_{n+1}) \psi \left(\frac{1}{x_1 + \cdots + x_{n+1}} \right) \\ &= \sum_{\substack{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \neq 0 \\ x_i \in \mathbb{F}_q^*}} \chi_1(x_1) \cdots \chi_n(x_n) \chi_{n+1} \left(\frac{b}{x_1 \cdots x_n} \right) \psi \left(\frac{1}{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n}} \right), \end{aligned}$$

where $b \in \mathbb{F}_q^*$ and $n \geq 1$.

When $n = 1$, Katz [15] obtained a sharp upper bound for $S_1(\chi, b)$, a simpler proof of which was later proposed by Li [18]. The estimation for $S_1(\chi, b)$ is motivated by the Ramanujan graphs. Katz's results [15] along with the papers of Angel [1] and Evans [11] proved that the finite upper half plane graph studied by Terra [23] is a Ramanujan graph in characteristic 2. This is part of an attempt to study finite analogues of the real symmetric spaces such as the Poincaré upper half plane. Ramanujan graphs are of interest because they are good expanders and because the simple random walk on these graphs converges particularly rapidly to uniform. Their importance and applications in many explicit algorithms in computer science are discussed in [20].

In [15], Katz raised the question: what can be said for $S_n(\chi, b)$ when $n \geq 1$? The aim of this paper is to study these *inverted n -variable Kloosterman sums* from both complex and p -adic point of views. As we shall see, this class of sums is very interesting as various new features and additional difficulties arise.

Remark The exact sum introduced in [15] is the following related sum,

$$T_n(\chi, b) = \sum_{\substack{x_1 \cdots x_{n+1} = 1 \\ x_1 + \cdots + x_{n+1} \neq 0}} \chi_1(x_1) \cdots \chi_{n+1}(x_{n+1}) \psi \left(\frac{b}{x_1 + \cdots + x_{n+1}} \right).$$

Upon the change of variables $x_i \rightarrow bx_i$, one sees that

$$T_n(\chi, b) = S_n(\chi, \frac{1}{b^{n+1}}) \chi_1(b) \cdots \chi_{n+1}(b).$$

Thus, the two families of sums $S_n(\chi, b)$ and $T_n(\chi, b)$ are essentially equivalent. We work with $S_n(\chi, b)$ as it is the closer inverted analogue of the classical Kloosterman sum. $\square$

The best complex estimate of $S_n(\chi, b)$ that one can hope for would be a square root cancellation in the sum $S_n(\chi, b)$, i.e.

$$S_n(\chi, b) = O_n(q^{\frac{n}{2}}).$$

As we shall see, this is not true for $n > 2$ when $\chi_1 = \cdots = \chi_{n+1}$, in which case, $S_n(\chi, b)$ has the main term $-q^{n-1} \chi_1(b)$ whose exponent $n-1$ is larger than the exponent $n/2$.

We shall give two different estimates for $S_n(\chi, b)$. The first estimate is based on an elementary method via Gauss sums which already shows the new feature of a non-trivial main term when $\chi_1 = \cdots = \chi_{n+1}$. We obtain the following simple estimate for $S_n(\chi, b)$:

Theorem 1.1 *Notation as above. If $\chi_1 = \cdots = \chi_{n+1}$, we have*

$$\left| S_n(\chi, b) + \frac{(q-1)^n}{q} \chi_1(b) \right| \leq q^{\frac{n+1}{2}}.$$

If $\chi_i \neq \chi_j$ for some $i \neq j$, we have

$$|S_n(\chi, b)| \leq q^{\frac{n+1}{2}}.$$

For large q , the error term $q^{(n+1)/2}$ is not the optimal square root cancellation yet. To obtain a deeper square root cancellation with error term $O_n(q^{n/2})$, we reduce $S_n(\chi, b)$ to a certain twisted toric exponential sum $S_k^*(\chi, f)$ which can be handled by the results of Adolphson–Sperber, Denef–Loeser and Fu. We check that the related Laurent polynomial f is non-degenerate if p does not divide $n+1$. This gives our second estimate:

Theorem 1.2 *Notation as above. Suppose $p \nmid (n+1)$. If $\chi_1 = \cdots = \chi_{n+1}$, we have*

$$|S_n(\chi, b) + \frac{(q-1)^n - (-1)^n}{q} \chi_1(b)| \leq (2n+1)q^{\frac{n}{2}}.$$

If $\chi_i \neq \chi_j$ for some $i \neq j$, we have

$$|S_n(\chi, b)| \leq 2(n+1)q^{\frac{n}{2}}.$$

It is clear that the estimate in Theorem 1.2 is better than the estimate in Theorem 1.1 if $q > 4(n+1)^2$. It is expected that Theorem 1.2 (with possibly a better constant) remains true when p divides $n+1$. But in this singular case, the above toric sum results do not apply and one would need a different approach.

For p -adic slopes, we focus on the simpler untwisted case in which $\chi_1 = \cdots = \chi_{n+1}$. We study the p -adic valuations for the reciprocal roots and poles of the generating L-function of the untwisted *inverted* n -variable Kloosterman sum. We show that the Newton polygon agrees with the Hodge polygon if and only if $p \equiv 1 \pmod{n+1}$. As a consequence, this completely determines the p -adic slope sequence when $p \equiv 1 \pmod{n+1}$. The method to study the p -adic slopes is described more precisely below.

Our approach is to reduce the generating L-function to a certain untwisted toric L-function. To construct the relationship between the two L-functions, we need to consider the untwisted *inverted* Kloosterman sum defined over every finite extension $\mathbb{F}_{q^k}$. Suppose $\chi_1 = \cdots = \chi_{n+1}$, the *inverted* n -variable Kloosterman sum over $\mathbb{F}_{q^k}$ is defined by

$$S_n(k, b) = \sum_{\substack{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \neq 0 \\ x_i \in \mathbb{F}_{q^k}^*}} \psi \left(\text{Tr}_k \left(\frac{1}{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n}} \right) \right),$$

where $b \in \mathbb{F}_q^*$ and $\text{Tr}_k : \mathbb{F}_{q^k} \rightarrow \mathbb{F}_q$ is the trace map.

Remark In the notation for the *twisted inverted* n -variable Kloosterman sum, we use the parameter χ to emphasize the multiplicative characters contained in the twisted case $S_n(\chi, b)$, which is only defined over the ground field $\mathbb{F}_q$. In the untwisted case, we introduce the sum over every finite extension $\mathbb{F}_{q^k}$ and denote the sum by $S_n(k, b)$.

When $\chi_1 = \cdots = \chi_{n+1}$, the twisted sum $S_n(\chi, b)$ reduces to the untwisted sum as follow:

$$\sum_{\substack{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \neq 0 \\ x_i \in \mathbb{F}_q^*}} \chi_1(b) \psi \left(\frac{1}{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n}} \right).$$

It is clear that $S_n(\chi, b)$ and $S_n(k, b)$ are equivalent when $\chi_1 = \cdots = \chi_{n+1}$ and $k = 1$. For brevity, we omit $\chi_1(b)$ in the definition of $S_n(k, b)$ since it is a non-zero constant. $\square$

The generating L-function of $S_n(k, b)$ is defined by

$$L_n(b, T) = \exp \left(\sum_{k=1}^{\infty} S_n(k, b) \frac{T^k}{k} \right).$$

Applying some systematic results available for the related toric L-function, which will be introduced in Sect. 2, we obtain the complex and p -adic absolute values for all the reciprocal roots and poles of $L_n(b, T)$ under given restrictions on p :

Theorem 1.3 *Suppose $p \nmid (n+1)$. The L-function is a rational function of the following form:*

$$L_n(b, T)^{(-1)^{n+1}} = (1-T)^{(n+1)} \prod_{j=2}^n \left(1 - q^{j-1} T\right)^{\binom{n}{j}(-1)^{j-1}} \prod_{i=1}^{2n} (1 - \alpha_i T).$$

As complex numbers, the reciprocal roots α_i satisfy $|\alpha_i| = q^{\frac{n}{2}}$ for all $1 \leq i \leq 2n$.

If $p \equiv 1 \pmod{n+1}$, viewing the α_i 's as p -adic numbers, the slope sequence $\{v_q(\alpha_i)\}_{i=1}^{2n}$ in increasing order is given by

$$\{0, 1, 1, 2, 2, \dots, n-1, n-1, n\}.$$

Our results show that for the *inverted* n -variable Kloosterman sum, if p does not divide $n+1$, the rigid cohomology giving the nontrivial factors of $L_n(b, T)^{(-1)^{n+1}}$ has dimension $2n$, pure of weight n and with Hodge numbers $\{1, 2, 2, \dots, 2, 1\}$. One can refer to Definition 2.6 for precise definition of Hodge numbers and Theorem 3.3 for the above result about Hodge numbers. On the other hand, the rigid cohomology giving the L-function of the classical n -variable Kloosterman sum has dimension $n+1$, pure of weight n and with Hodge numbers $\{1, 1, 1, \dots, 1, 1\}$. The larger Hodge numbers suggest that new features and additional difficulties would likely arise in studying *inverted* n -variable Kloosterman sums.

As a corollary of the first part in Theorem 1.3, we get a slightly better bound for $S_n(k, b)$:

Corollary 1.4 *If $p \nmid (n+1)$, for all integers $k \geq 1$, we have*

$$|S_n(k, b) + \frac{(q^k - 1)^n - (-1)^n(q^k + 1)}{q^k}| \leq 2nq^{\frac{nk}{2}}.$$

When $k = 1$, the exponential sum $S_n(k, b)$ reduces to $S_n(\chi, b)$ with $\chi_1 = \dots = \chi_{n+1}$. Explicitly, the estimate in Corollary 1.4 is better than the estimate in the first case of Theorem 1.2. We remark that the condition $p \equiv 1 \pmod{n+1}$ in the second part of Theorem 1.3 is necessary and sufficient for the same conclusion to hold. Thus, if $p \not\equiv 1 \pmod{n+1}$, the slope sequence will be strictly different, but we do not know the exact slope sequence in this case.

The rest of this paper is organized as follows. In Sect. 2, we review some technical methods including Adolphson–Sperber's theorems and Dwork's theory on toric exponential sums. In Sect. 3, we use these methods to prove the main results.

We end the introduction by mentioning several recent references [8, 14, 17, 19, 27] which applied some of the related toric techniques in treating different classes of *non-toric* n variable exponential sums arising from analytic number theory.

2 Preliminaries on toric exponential sums

2.1 Rationality of the toric L-function

Let $f \in \mathbb{F}_q \left[x_1^{\pm 1}, \dots, x_n^{\pm 1} \right]$ be a Laurent polynomial and its associated *twisted toric exponential sum* is defined to be

$$S_k^*(\chi, f) = \sum_{x_i \in \mathbb{F}_{q^k}^*} \chi_1(N_k(x_1)) \cdots \chi_n(N_k(x_n)) \psi(\text{Tr}_k(f)), \quad (2.1)$$

where $\text{Tr}_k : \mathbb{F}_{q^k} \rightarrow \mathbb{F}_q$ is the trace map, $N_k : \mathbb{F}_{q^k} \rightarrow \mathbb{F}_q$ is the norm map, $\chi_1, \dots, \chi_n : \mathbb{F}_q^* \rightarrow \mathbb{C}^*$ are multiplicative characters and $\psi : \mathbb{F}_q \rightarrow \mathbb{C}^*$ is a nontrivial additive character. A classical problem in number theory is to estimate the absolute values of $S_k^*(\chi, f)$.

A well known theorem of Dwork–Bombieri–Grothendieck says that the generating L-function of $S_k^*(\chi, f)$, which is called the *twisted toric L-function*, is a rational function:

$$L^*(\chi, f, T) = \exp \left(\sum_{k=1}^{\infty} S_k^*(\chi, f) \frac{T^k}{k} \right) = \frac{\prod_{i=1}^{d_1} (1 - \alpha_i T)}{\prod_{j=1}^{d_2} (1 - \beta_j T)},$$

where all the reciprocal zeros and poles are non-zero algebraic integers. In particular, when all of χ_i are trivial characters, the *untwisted toric exponential sum* and the *untwisted toric L-function* are denoted by $S_k^*(f)$ and $L^*(f, T)$ respectively.

Through logarithmic derivatives, we have

$$S_k^*(\chi, f) = \sum_{j=1}^{d_2} \beta_j^k - \sum_{i=1}^{d_1} \alpha_i^k, \quad k \in \mathbb{Z}_{\geq 1}. \quad (2.2)$$

Thus, the estimate of $S_k^*(\chi, f)$ is reduced to understanding all the absolute values of the reciprocal zeros α_i and poles β_j . Deligne's theorem on Riemann hypothesis [9] describes the bounds for all the absolute values of α_i and β_j in general. The complex absolute values of reciprocal zeros and poles satisfy

$$|\alpha_i| = q^{u_i/2}, |\beta_j| = q^{v_j/2}, u_i \in \mathbb{Z} \cap [0, 2n], v_j \in \mathbb{Z} \cap [0, 2n].$$

For non-archimedean absolute values, Deligne proved that $|\alpha_i|_\ell = |\beta_j|_\ell = 1$ when ℓ is a prime and $\ell \neq p$. For p -adic absolute values, one has

$$|\alpha_i|_p = q^{-r_i}, |\beta_j|_p = q^{-s_j}, r_i \in \mathbb{Q} \cap [0, n], s_j \in \mathbb{Q} \cap [0, n].$$

The integer u_i (resp. v_j) is called the *weight* of α_i (resp. β_j) and the rational number r_i (resp. s_j) is called the *slope* of α_i (resp. β_j). For the sake of brevity, we will refer to the weights (resp. slopes) of the reciprocal roots and poles of the generating L-function as the weights (resp. slopes) of the L-function below.

In the past few decades, there has been tremendous interest in determining the weights and slopes of the L-functions. Without any further conditions on the Laurent polynomial f , it is even hard to determine the number of reciprocal roots and poles of $L^*(\chi, f, T)$. Most of the existing works about the weights and slopes relies on a suitable smoothness condition. For toric exponential sums, this smoothness condition usually means the non-degenerate condition, see below for the precise definition.

Let

$$f(x_1, \dots, x_n) = \sum_{j=1}^J a_j x^{V_j} \quad (2.3)$$

be a Laurent polynomial with $a_j \in \mathbb{F}_q^*$ and $V_j = (v_{1j}, \dots, v_{nj}) \in \mathbb{Z}^n$ ($1 \leq j \leq J$). The *Newton polyhedron* of f , $\Delta(f)$, is defined to be the convex closure in $\mathbb{R}^n$ generated by the origin and the lattice points V_j ($1 \leq j \leq J$). For $\delta \subset \Delta(f)$, let the Laurent polynomial

$$f^\delta = \sum_{V_j \in \delta} a_j x^{V_j}$$

be the restriction of f to δ .

Definition 2.1 (*Non-degenerate*) A Laurent polynomial f is called non-degenerate if for each closed face δ of $\Delta(f)$ of arbitrary dimension which doesn't contain the origin, the partial derivatives

$$\left\{ \frac{\partial f^\delta}{\partial x_1}, \dots, \frac{\partial f^\delta}{\partial x_n} \right\}$$

have no common zeros with $x_1 \dots x_n \neq 0$ over the algebraic closure of $\mathbb{F}_q$.

When f is non-degenerate, Adolphson–Sperber [4] proved that the untwisted toric L-function $L^*(f, T)^{(-1)^{n-1}}$ is a polynomial and improved the bound for weight.

Theorem 2.2 [4] *For any non-degenerate $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$, the associated L-function $L^*(f, T)^{(-1)^{n-1}}$ is a polynomial of degree $n! \text{Vol}(\Delta(f))$. Namely,*

$$L^*(f, T)^{(-1)^{n-1}} = \prod_{i=1}^{n! \text{Vol}(\Delta(f))} (1 - \alpha_i T), \quad \alpha_i \neq 0.$$

For any multiplicative characters χ_i (nontrivial or trivial), the degree and weights of the twisted toric L-function are studied and completed by Adolphson–Sperber [6, 7], Denef–Loeser [10] and Fu [12] under the non-degeneracy assumption. These results lead to the following bound for the twisted toric exponential sum.

Theorem 2.3 [7, 10, 12] *Let $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ be a Laurent polynomial with $\Delta = \Delta(f)$. If f is non-degenerate, one has*

$$|S_k^*(\chi_1, \dots, \chi_n, f)| \leq n! \text{Vol}(\Delta) q^{\frac{nk}{2}}.$$

For the slopes of the L-function, the situation is somewhat simpler in the untwisted case, otherwise, even the description of Adolphson–Sperber's “Hodge lower bound” is a little cumbersome. Thus, the definitions and theories discussed in the following subsections focus on the untwisted L-function.

2.2 Newton polygon and Hodge polygon

To determine the slopes of the L-functions, we introduce the q -adic Newton polygon.

Definition 2.4 (*Newton polygon*) Let $L(T) = \sum_{i=0}^n a_i T^i \in 1 + T\overline{\mathbb{Q}}_p[T]$, where $\overline{\mathbb{Q}}_p$ is the algebraic closure of $\mathbb{Q}_p$. The q -adic Newton polygon of $L(T)$ is defined to be the lower convex closure of the set of points $\{(k, \text{ord}_q(a_k)) \mid k = 0, 1, \dots, n\}$ in $\mathbb{R}^2$.

Lemma 2.5 [16] *Notation as above. Let $L(T) = (1 - \alpha_1 T) \dots (1 - \alpha_n T)$ be the factorization of $L(T)$ in terms of reciprocal roots $\alpha_i \in \overline{\mathbb{Q}}_p$. Let $\lambda_i = \text{ord}_q \alpha_i$. If λ is the slope of the q -adic Newton polygon of $L(T)$ with horizontal length l , then precisely l of the λ_i are equal to λ .*

The q -adic Newton polygon of $L^*(f, T)^{(-1)^{n-1}}$ is denoted as $\text{NP}(f)$. Lemma 2.5 relates $\text{NP}(f)$ to the q -adic valuation of reciprocal roots of toric L -functions. The definition of $\text{NP}(f)$ relies on the coefficients of L -function, which makes it hard to compute directly. When f is non-degenerate, Adolphson and Sperber proved that $L^*(f, T)^{(-1)^{n-1}}$ is a polynomial and $\text{NP}(f)$ has a topological lower bound called Hodge polygon, which is easier to determine. Thus, we shall compute Hodge polygon and consider when the Newton polygon coincides with this lower bound.

Let Δ be an n -dimensional integral polytope containing the origin in $\mathbb{R}^n$. For $u \in \mathbb{R}^n$, the weight function $w(u)$ represents the smallest non-negative real number c such that $u \in c\Delta$. Denote $w(u) = \infty$ if such c doesn't exist. Assume δ is a co-dimension 1 face of Δ not containing the origin. Let $D(\delta)$ be the least common multiple of the denominators of the coefficients in the linear equation defining δ , normalized to have constant term 1. We define the denominator of Δ to be the least common multiple of all such $D(\delta)$ given by:

$$D = D(\Delta) = \text{lcm}_{\delta} D(\delta),$$

where δ runs over all the co-dimension 1 faces of Δ that don't contain the origin. It's easy to check

$$w(\mathbb{Z}^n) \subseteq \frac{1}{D(\Delta)} \mathbb{Z}_{\geq 0} \cup \{+\infty\}.$$

For a non-negative integer k , let

$$W_{\Delta}(k) = \# \left\{ u \in \mathbb{Z}^n \mid w(u) = \frac{k}{D} \right\} \quad (2.4)$$

be the number of lattice points in $\mathbb{Z}^n$ with weight k/D . Its generating function is known to be a rational function of the following form

$$\sum_{k=0}^{\infty} W_{\Delta}(k) t^{k/D} = \frac{\sum_{k=0}^{nD} H_{\Delta}(k) t^{k/D}}{(1-t)^n}. \quad (2.5)$$

This leads to

Definition 2.6 (*Hodge number*) Let Δ be an n -dimensional integral polytope containing the origin in $\mathbb{R}^n$. For a non-negative integer k , the k -th Hodge number of Δ is defined to be

$$H_{\Delta}(k) = \sum_{i=0}^n (-1)^i \binom{n}{i} W_{\Delta}(k - iD). \quad (2.6)$$

It is known that

$$H_{\Delta}(k) = 0, \quad \text{if } k > nD.$$

Based on the Hodge numbers, we define the Hodge polygon of a given polyhedron $\Delta \in \mathbb{R}^n$ as follows.

Remark The Hodge number defined above is closely related to irregular Hodge theory by Sabbah [21, §12] and Yu [26]. $\square$

Definition 2.7 (*Hodge polygon*) The Hodge polygon $HP(\Delta)$ of Δ is the lower convex polygon in $\mathbb{R}^2$ with vertices $(0,0)$ and

$$\mathbb{Q}_k = \left(\sum_{m=0}^k H_{\Delta}(m), \frac{1}{D} \sum_{m=0}^k m H_{\Delta}(m) \right), \quad k = 0, 1, \dots, nD,$$

where $H_{\Delta}(k)$ is the k -th Hodge number of Δ , $k = 0, 1, \dots, nD$.

That is, $HP(\Delta)$ is a polygon starting from origin $(0,0)$ with a slope k/D side of horizontal length $H_{\Delta}(k)$ for $k = 0, 1, \dots, nD$. The vertex $\mathbb{Q}_k$ is called a break point if $H_{\Delta}(k+1) \neq 0$ where $k = 1, 2, \dots, nD-1$.

Note that the horizontal length $H_{\Delta}(k)$ is the number of lattice points of weight k/D in a certain fundamental domain corresponding to a basis of the p -adic cohomology space used to compute the L-function. By a theorem of Adolphson–Sperber, the Hodge polygon is a lower bound of the corresponding Newton polygon.

Theorem 2.8 [4] *For every prime p and non-degenerate Laurent polynomial f with $\Delta(f) = \Delta \subset \mathbb{R}^n$, we have*

$$NP(f) \geq HP(\Delta),$$

where $NP(f)$ is the q -adic Newton polygon of $L^*(f, T)^{(-1)^{n-1}}$. Furthermore, the endpoints of $NP(f)$ and $NP(\Delta)$ coincide.

Definition 2.9 (*Ordinary*) A Laurent polynomial f is called ordinary if $NP(f) = HP(\Delta)$.

It is clear that the ordinary property of a Laurent polynomial depends on its Newton polyhedron Δ and on the coefficients of $f(x)$. Applying the facial decomposition theorem [24], we reduce the ordinary property of f to its smaller pieces which are easier to deal with.

Theorem 2.10 (Facial decomposition theorem [24]) *Let f be a non-degenerate Laurent polynomial over $\mathbb{F}_q$. Assume $\Delta = \Delta(f)$ is n -dimensional and $\delta_1, \dots, \delta_h$ are all the co-dimension 1 faces of Δ which don't contain the origin. Let f^{δ_i} denote the restriction of f to δ_i . Then f is ordinary if and only if f^{δ_i} is ordinary for $1 \leq i \leq h$.*

2.3 Boundary decomposition theorems

Before describing the boundary decomposition, we express the L-function in terms of the Fredholm determinant of an infinite Frobenius matrix via Dwork's trace formula.

2.3.1 Dwork's trace formula

Let $\mathbb{Q}_p$ be the field of p -adic numbers and Ω be the completion of $\overline{\mathbb{Q}}_p$. A fixed primitive p -th root of unity in Ω is denoted as ζ_p . Let π be the element of $\mathbb{Q}_p(\zeta_p)$ satisfies

$$\sum_{m=0}^{\infty} \frac{\pi^{p^m}}{p^m} = 0, \quad \pi \equiv \zeta_p - 1 \pmod{(\zeta_p - 1)^2}, \quad \text{and} \quad \text{ord}_p \pi = \frac{1}{p-1}.$$

Then, π is a uniformizer of $\mathbb{Q}_p(\pi)$ and thus $\mathbb{Q}_p(\pi) = \mathbb{Q}_p(\zeta_p)$. Let $E_p(t)$ be the Artin–Hasse exponential series,

$$E_p(t) = \exp\left(\sum_{m=0}^{\infty} \frac{t^{p^m}}{p^m}\right) = \sum_{m=0}^{\infty} \lambda_m t^m \in \mathbb{Z}_p[[x]].$$

In Dwork's terminology, a splitting function $\theta(t)$ is defined to be

$$\theta(t) = E_p(\pi t) = \sum_{m=0}^{\infty} \lambda_m \pi^m t^m.$$

A Laurent polynomial $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ is written as

$$f = \sum_{j=1}^J \bar{a}_j x^{V_j},$$

where $V_j \in \mathbb{Z}^n$ and $\bar{a}_j \in \mathbb{F}_q^*$. Let a_j be the Teichmüller lifting of $\bar{a}_j$ in Ω satisfying $a_j^q = a_j$. Let

$$F(f, x) = \prod_{j=1}^J \theta(a_j x^{V_j}) = \sum_{r \in \mathbb{Z}^n} F_r(f) x^r.$$

The coefficients are given by

$$F_r(f) = \sum_u \left(\prod_{j=1}^J \lambda_{u_j} a_j^{u_j} \right) \pi^{u_1 + \dots + u_J}, \quad r \in \mathbb{Z}^n,$$

where the sum is over all the solutions of the following linear system

$$\sum_{j=1}^J u_j V_j = r \quad \text{with } u_j \in \mathbb{Z}_{\geq 0},$$

and λ_m is m -th coefficient of the Artin–Hasse exponential series $E_p(t)$.

Assume $\Delta = \Delta(f)$. Let $L(\Delta) = \mathbb{Z}^n \cap C(\Delta)$ be the set of lattice points in the closed cone generated by origin and Δ . For a given point $r \in \mathbb{R}^n$, define the weight function to be

$$w(r) := \inf_{\vec{u}} \left\{ \sum_{j=1}^J u_j \mid \sum_{j=1}^J u_j V_j = r, \quad u_j \in \mathbb{R}_{\geq 0} \right\}.$$

The infinite semilinear Frobenius matrix $A_1(f)$ is the following matrix whose rows and columns are indexed by the lattice points in $L(\Delta)$ with respect to the weights:

$$A_1(f) = (a_{r,s}(f)) = (F_{ps-r}(f) \pi^{w(r)-w(s)}),$$

where $r, s \in L(\Delta)$. The infinite linear Frobenius matrix $A_a(f)$ is defined to be

$$A_a(f) = A_1(f) A_1^\tau(f) \cdots A_1^{\tau^{a-1}}(f),$$

where τ is the absolute Frobenius automorphism.

Dwork's trace formula can be expressed in terms of the matrix $A_a(f)$ as follows, see [25]

Theorem 2.11 *We have*

$$L^*(f, T)^{(-1)^{n-1}} = \prod_{i=0}^n \det(I - T q^i A_a(f))^{(-1)^i \binom{n}{i}}. \quad (2.7)$$

Equivalently,

$$\det(I - T A_a(f)) = \prod_{i=0}^{\infty} \left(L^*(f, q^i T)^{(-1)^{n-1}} \right)^{\binom{n+i-1}{i}}. \quad (2.8)$$

Now it suffices to understand the determinant $\det(I - T A_a(f))$. Based on the fact that $\text{ord}_p F_r(f) \geq \frac{w(r)}{p-1}$, we have the following estimate

$$\text{ord}_p(a_{r,s}(f)) \geq \frac{w(ps - r) + w(r) - w(s)}{p - 1} \geq w(s).$$

Let ξ be an element in Ω satisfying $\xi^D = \pi^{p-1}$. Then $A_1(f)$ can be written in a block form,

$$A_1(f) = \begin{pmatrix} A_{00} & \xi A_{01} & \cdots & \xi^i A_{0i} & \cdots \\ A_{10} & \xi A_{11} & \cdots & \xi^i A_{1i} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \\ A_{i0} & \xi A_{i1} & \cdots & \xi^i A_{ii} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \end{pmatrix},$$

where the block A_{ii} is a p -adic integral $W_{\Delta}(i) \times W_{\Delta}(i)$ matrix. This implies that the q -adic Newton polygon of $\det(I - T A_1(f))$ has a natural lower bound which can be identified with the chain level version of the Hodge polygon.

Definition 2.12 Let $P(\Delta)$ be the polygon in $\mathbb{R}^2$ with vertices $(0, 0)$ and

$$P_k = \left(\sum_{m=0}^k W_{\Delta}(m), \frac{1}{D} \sum_{m=0}^k m W_{\Delta}(m) \right), \quad k = 0, 1, 2, \dots$$

The chain level version of Adolphson–Sperber’s lower bound and the ordinary property are as follows.

Proposition 2.13 [2] *The q -adic Newton polygon of $\det(I - T A_a(f))$ lies above $P(\Delta)$.*

Proposition 2.14 [25] *Notation as above. Assume f is non-degenerate with $\Delta = \Delta(f)$. Then $\text{NP}(f) = \text{HP}(\Delta)$ if and only if the q -adic Newton polygon of $\det(I - T A_a(f))$ coincides with its lower bound $P(\Delta)$.*

2.3.2 Boundary decomposition

Let $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ with $\Delta = \Delta(f)$, where Δ is an n -dimensional integral convex polyhedron in $\mathbb{R}^n$ containing the origin. Let $C(\Delta)$ be the cone generated by Δ in $\mathbb{R}^n$.

Definition 2.15 The boundary decomposition

$$B(\Delta) = \{\text{the interior of a closed face in } C(\Delta) \text{ containing the origin}\}$$

is the unique interior decomposition of $C(\Delta)$ into a disjoint union of relatively open cones.

If the origin is a vertex of Δ , then it is the unique 0-dimensional open cone in $B(\Delta)$. Recall that $A_1(f) = (a_{r,s}(f))$ is the infinite semilinear Frobenius matrix whose rows and columns are indexed by the lattice points in $L(\Delta)$. For $\Sigma \in B(\Delta)$, we define $A_1(\Sigma, f)$ to be the submatrix of $A_1(f)$ with $r, s \in \Sigma$. Let $f^{\bar{\Sigma}}$ be the restriction of f to the closure of Σ . Then $A_1(\Sigma, f^{\bar{\Sigma}})$ denotes the submatrix of $A_1(f^{\bar{\Sigma}})$ with $r, s \in \Sigma$.

Let $B(\Delta) = \{\Sigma_0, \dots, \Sigma_h\}$ such that $\dim(\Sigma_i) \leq \dim(\Sigma_{i+1})$, $i = 0, \dots, h-1$. Define $B_{ij} = (a_{r,s}(f))$ with $r \in \Sigma_i$ and $s \in \Sigma_j$ ($0 \leq i, j \leq h$). After a permutation of basis vectors, the infinite semilinear Frobenius matrix can be written as

$$A_1(f) = \begin{pmatrix} B_{00} & B_{01} & \cdots & B_{0h} \\ B_{10} & B_{11} & \cdots & B_{1h} \\ \vdots & \vdots & \ddots & \vdots \\ B_{h0} & B_{h1} & \cdots & B_{hh} \end{pmatrix}, \quad (2.9)$$

where $B_{ij} = 0$ for $i > j$. Then $\det(I - TA_1(f)) = \prod_{i=0}^h \det(I - TB_{ii})$ and we have the boundary decomposition theorem.

Theorem 2.16 (Boundary decomposition [24]) *Let $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ with $\Delta = \Delta(f)$. Then we have the following factorization*

$$\det(I - TA_1(f)) = \prod_{\Sigma \in B(\Delta)} \det(I - TA_1(\Sigma, f^{\bar{\Sigma}})).$$

2.4 Diagonal local theory

In this subsection, we introduce some non-degenerate and ordinary criteria when the Laurent polynomial is diagonal.

Definition 2.17 A Laurent polynomial $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ is called diagonal if f has exactly n non-constant terms and $\Delta(f)$ is an n -dimensional simplex in $\mathbb{R}^n$.

Let f be a diagonal Laurent polynomial over $\mathbb{F}_q$. Write

$$f(x_1, x_2, \dots, x_n) = \sum_{j=1}^n a_j x^{V_j},$$

where $a_j \in \mathbb{F}_q^*$ and $V_j = (v_{1j}, \dots, v_{nj}) \in \mathbb{Z}^n$ for $1 \leq j \leq n$. Let $\Delta = \Delta(f)$. The vertex matrix of Δ is defined to be

$$M(\Delta) = (V_1, \dots, V_n),$$

where the i -th column is the i -th exponent of f . Since f is diagonal, $M(\Delta)$ is invertible.

Proposition 2.18 *Suppose $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ is diagonal with $\Delta = \Delta(f)$. Then f is non-degenerate if and only if p is relatively prime to $\det(M(\Delta))$.*

Let $S(\Delta)$ be the solution set of the following linear system

$$M(\Delta) \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} \equiv 0 \pmod{1}, \quad r_i \in \mathbb{Q} \cap [0, 1).$$

It's easy to prove that $S(\Delta)$ is an abelian group and its order is given by

$$|\det M(\Delta)| = n! \operatorname{Vol}(\Delta). \quad (2.10)$$

Let $S_p(\Delta)$ denote the prime to p part of $S(\Delta)$. It is an abelian subgroup of order equal to the prime to p factor of $\det M(\Delta)$. In particular, $S_p(\Delta) = S(\Delta)$ if p is relatively prime to $\det M(\Delta)$. By the Stickelberger theorem for Gauss sums, we have the following ordinary criterion for a non-degenerate Laurent polynomial [25].

Proposition 2.19 *A diagonal Laurent polynomial f is ordinary at p if and only if the norm function $|r| = r_1 + \cdots + r_n$ on $S_p(\Delta)$ is stable under the p -action: That is, for each $r \in S_p(\Delta)$, we have $|r| = |\{pr\}|$, where $\{pr\}$ is the class of pr in $S_p(\Delta)$.*

3 Proof of the main theorems

We prove the main theorems in this section.

3.1 Proof of Theorem 1.1

Recall that for integer $n \geq 1$, the twisted *inverted* n -variable Kloosterman sum is defined to be

$$S_n(\chi, b) = \sum_{\substack{x_1 \cdots x_{n+1} = b \\ x_1 + \cdots + x_{n+1} \neq 0}} \chi_1(x_1) \cdots \chi_{n+1}(x_{n+1}) \psi \left(\frac{1}{x_1 + \cdots + x_{n+1}} \right),$$

where $b \in \mathbb{F}_q^*$, $\psi : \mathbb{F}_q \rightarrow \mathbb{C}^*$ is a nontrivial additive character and $\chi_1, \dots, \chi_{n+1} : \mathbb{F}_q^* \rightarrow \mathbb{C}^*$ are multiplicative characters. Let $\chi : \mathbb{F}_q^* \rightarrow \mathbb{C}^*$ denote a multiplicative character. By the orthogonality of characters, we have

$$\begin{aligned} S_n(\chi, b) &= \frac{1}{q(q-1)} \sum_{\lambda \in \mathbb{F}_q^*} \sum_{x_i \in \mathbb{F}_q^*} \psi(u(x_1 + \cdots + x_{n+1} - \lambda)) \chi_1(x_1) \cdots \chi_{n+1}(x_{n+1}) \\ &\quad \times \psi \left(\frac{1}{\lambda} \right) \sum_{\chi} \chi \left(\frac{x_1 \cdots x_{n+1}}{b} \right) \\ &= \frac{1}{q(q-1)} \sum_{\lambda \in \mathbb{F}_q^*} \sum_{x_i \in \mathbb{F}_q^*} \chi_1(x_1) \cdots \chi_{n+1}(x_{n+1}) \psi \left(\frac{1}{\lambda} \right) \sum_{\chi} \chi \left(\frac{x_1 \cdots x_{n+1}}{b} \right) \\ &\quad + \frac{1}{q(q-1)} \sum_{\lambda \in \mathbb{F}_q^*} \sum_{x_i \in \mathbb{F}_q^*} \sum_{u \in \mathbb{F}_q^*} \psi(u(x_1 + \cdots + x_{n+1} - \lambda)) \\ &\quad \times \chi_1(x_1) \cdots \chi_{n+1}(x_{n+1}) \psi \left(\frac{1}{\lambda} \right) \sum_{\chi} \chi \left(\frac{x_1 \cdots x_{n+1}}{b} \right) \\ &= S_1 + S_2. \end{aligned} \quad (3.1)$$

Then

$$\begin{aligned}
 S_1 &= \frac{1}{q(q-1)} \sum_{\lambda \in \mathbb{F}_q^*} \psi \left(\frac{1}{\lambda} \right) \sum_{\chi} \chi^{-1}(b) \sum_{x_i \in \mathbb{F}_q^*} (\chi \chi_1)(x_1) \cdots (\chi \chi_{n+1})(x_{n+1}) \\
 &= \frac{1}{q(q-1)} \sum_{\lambda \in \mathbb{F}_q^*} \psi \left(\frac{1}{\lambda} \right) \sum_{\chi} \chi^{-1}(b) \prod_{i=1}^{n+1} \left(\sum_{x_i \in \mathbb{F}_q^*} (\chi \chi_i)(x_i) \right) \\
 &= \begin{cases} -\frac{(q-1)^n}{q} \chi_1(b), & \text{if } \chi_1 = \cdots = \chi_{n+1}, \\ 0, & \text{otherwise.} \end{cases} \quad (3.2)
 \end{aligned}$$

As defined in (1.2), the Gauss sum

$$G(\chi) = \begin{cases} -1, & \text{if } \chi \text{ is trivial,} \\ \sqrt{q}, & \text{if } \chi \text{ is non-trivial.} \end{cases}$$

Then

$$\begin{aligned}
 S_2 &= \frac{1}{q(q-1)} \sum_{\lambda, u \in \mathbb{F}_q^*} \sum_{\chi} \chi^{-1}(b) \sum_{x_i \in \mathbb{F}_q^*} (\chi \chi_1)(x_1) \psi(ux_1) \cdots (\chi \chi_{n+1})(x_{n+1}) \psi(ux_{n+1}) \\
 &\quad \times \psi(-u\lambda) \psi \left(\frac{1}{\lambda} \right) \\
 &= \frac{1}{q(q-1)} \sum_{\lambda, u \in \mathbb{F}_q^*} \sum_{\chi} \chi^{-1}(b) \overline{\chi^{n+1} \chi_1 \cdots \chi_{n+1}(u)} \psi(-u\lambda) \psi \left(\frac{1}{\lambda} \right) G(\chi \chi_1) \cdots G(\chi \chi_{n+1}) \\
 &= \frac{1}{q(q-1)} \sum_{\chi} \chi^{-1}(b) \left(\sum_{\lambda \in \mathbb{F}_q^*} \overline{\chi^{n+1} \chi_1 \cdots \chi_{n+1}} \left(-\frac{1}{\lambda} \right) \psi \left(\frac{1}{\lambda} \right) \right) G(\overline{\chi^{n+1} \chi_1 \cdots \chi_{n+1}}) \\
 &\quad \times G(\chi \chi_1) \cdots G(\chi \chi_{n+1}) \\
 &= \frac{1}{q(q-1)} \sum_{\chi} \chi^{-1}(b) \chi^{n+1} \chi_1 \cdots \chi_{n+1} (-1) G(\overline{\chi^{n+1} \chi_1 \cdots \chi_{n+1}}) G(\overline{\chi^{n+1} \chi_1 \cdots \chi_{n+1}}) \\
 &\quad \times G(\chi \chi_1) \cdots G(\chi \chi_{n+1}). \quad (3.3)
 \end{aligned}$$

Since $|G(\chi)| \leq \sqrt{q}$, it follows that $|S_2| \leq q^{\frac{n+1}{2}}$. Combining (3.1) and (3.2), we can deduce the following bounds.

$$\left| S_n(\chi, b) + \frac{(q-1)^n}{q} \chi_1(b) \right| \leq q^{\frac{n+1}{2}}, \quad \text{if } \chi_1 = \cdots = \chi_{n+1},$$

and

$$|S_n(\chi, b)| \leq q^{\frac{n+1}{2}}, \quad \text{if } \chi_i \neq \chi_j \text{ for some } i \neq j.$$

This proves Theorem 1.1.

3.2 Proof of Theorem 1.2

The twisted *inverted* Kloosterman sum $S_n(\chi, b)$ has the expression

$$\begin{aligned}
 S_n(\chi, b) &= \sum_{\substack{x_1 + \dots + x_n + \frac{b}{x_1 \dots x_n} \neq 0 \\ x_i \in \mathbb{F}_q^*}} \chi_1(x_1) \cdots \chi_n(x_n) \chi_{n+1} \left(\frac{b}{x_1 \dots x_n} \right) \\
 &\quad \times \psi \left(\frac{1}{x_1 + \dots + x_n + \frac{b}{x_1 \dots x_n}} \right) \\
 &= \sum_{\substack{z \left(x_1 + \dots + x_n + \frac{b}{x_1 \dots x_n} \right) = 1 \\ z, x_i \in \mathbb{F}_q^*}} \chi_{n+1}(b) (\chi_1 \overline{\chi_{n+1}})(x_1) \cdots (\chi_n \overline{\chi_{n+1}})(x_n) \psi(z) \\
 &= \frac{1}{q} \sum_{\substack{z, x_i \in \mathbb{F}_q^* \\ y \in \mathbb{F}_q}} \chi_{n+1}(b) (\chi_1 \overline{\chi_{n+1}})(x_1) \cdots (\chi_n \overline{\chi_{n+1}})(x_n) \\
 &\quad \times \psi \left(z + y \left(1 - z \left(x_1 + \dots + x_n + \frac{b}{x_1 \dots x_n} \right) \right) \right) \\
 &= \frac{\chi_{n+1}(b)}{q} \left(\sum_{z, x_i \in \mathbb{F}_q^*} (\chi_1 \overline{\chi_{n+1}})(x_1) \cdots (\chi_n \overline{\chi_{n+1}})(x_n) \psi(z) + E_n(\chi, b) \right) \\
 &= \begin{cases} -\frac{(q-1)^n}{q} \chi_1(b) + \frac{1}{q} \chi_1(b) E_n(\chi, b), & \text{if } \chi_1 = \dots = \chi_{n+1}, \\ \frac{1}{q} \chi_{n+1}(b) E_n(\chi, b), & \text{if } \chi_i \neq \chi_j \text{ for some } i \neq j, \end{cases} \quad (3.4)
 \end{aligned}$$

where

$$\begin{aligned}
 E_n(\chi, b) &= \sum_{y, z, x_i \in \mathbb{F}_q^*} (\chi_1 \overline{\chi_{n+1}})(x_1) \cdots (\chi_n \overline{\chi_{n+1}})(x_n) \\
 &\quad \times \psi \left(z + y \left(1 - z \left(x_1 + \dots + x_n + \frac{b}{x_1 \dots x_n} \right) \right) \right).
 \end{aligned}$$

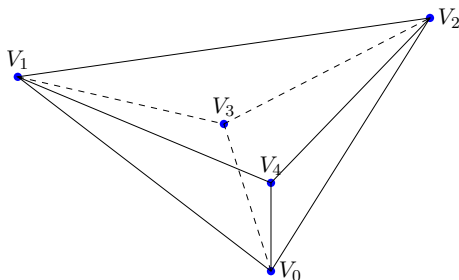
In order to prove Theorem 1.2, it suffices to estimate $E_n(\chi, b)$.

Let $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_{n+2}^{\pm 1}]$ be the Laurent polynomial defined by

$$f(x_1, \dots, x_{n+2}) = x_{n+1} \left(1 - x_{n+2} \left(x_1 + \dots + x_n + \frac{b}{x_1 \dots x_n} \right) \right) + x_{n+2}.$$

As defined in (2.1), $E_n(\chi, b)$ is the twisted toric exponential sum associated to f . Let $\Delta = \Delta(f)$ denote the Newton polyhedron corresponding to f . Clearly, $\dim \Delta = n + 2$ and Δ has $n + 4$ vertices in $\mathbb{R}^{n+2}$: $V_0 = (0, \dots, 0)$ (the origin), $V_1 = (1, 0, \dots, 0, 1, 1)$, $V_2 = (0, 1, \dots, 0, 1, 1)$, $\dots$, $V_n = (0, 0, \dots, 1, 1, 1)$, $V_{n+1} = (-1, \dots, -1, 1, 1)$, $V_{n+2} = (0, \dots, 0, 1, 0)$ and $V_{n+3} = (0, \dots, 0, 0, 1)$. As an example, one can refer to Fig. 1 for the case $n = 1$. Furthermore, Δ has exactly 2 co-dimension 1 faces not containing the origin. Explicitly, they are

$$\delta_1 : x_{n+1} = 1 \quad \text{and} \quad \delta_2 : x_{n+2} = 1.$$

Fig. 1 Δ for $n = 1$ 

Vertices $V_1, \dots, V_{n+2}$ determine the face δ_1 and vertices $V_1, \dots, V_{n+1}, V_{n+3}$ determine the face δ_2 . Let $M(\delta_i)$ be the vertex matrix of δ_i , we have

$$M(\delta_1) = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 & 0 \\ 0 & 1 & \cdots & 0 & -1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 & 0 \\ 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 & 0 \end{pmatrix}, \quad M(\delta_2) = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 & 0 \\ 0 & 1 & \cdots & 0 & -1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 & 0 \\ 1 & 1 & \cdots & 1 & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 & 1 \end{pmatrix}. \quad (3.5)$$

Explicitly, each f^{δ_i} is diagonal for $i = 1, 2$. The restriction of f to δ_i is defined by

$$f^{\delta_i} = \sum_{V_j \in \delta_i} a_j x^{V_j}.$$

Proposition 3.1 (i) *The denominator $D = 1$.*

(ii) *f is non-degenerate if and only if $p \nmid (n+1)$.*

(iii) $\text{Vol}(\Delta) = \frac{2n+2}{(n+2)!}$.

Proof The denominator $D = 1$ can be deduced immediately from the equation of δ_i . Since δ_1 and δ_2 are the co-dimension 1 faces of $\Delta(f)$ not containing the origin, it suffices to prove f^{δ_1} and f^{δ_2} are non-degenerate. By Proposition 2.18, f^{δ_i} is non-degenerate if and only if p is relatively prime to $\det(M(\delta_i))$. By formula (3.5),

$$\det(M(\delta_1)) = -(n+1) \quad \text{and} \quad \det(M(\delta_2)) = n+1. \quad (3.6)$$

This proves (ii).

Let Δ_i be the polytope generated by δ_i and the origin. The facial decomposition of Δ implies that

$$\text{Vol}(\Delta) = \text{Vol}(\Delta_1) + \text{Vol}(\Delta_2).$$

By formula (2.10) and (3.6), we obtain (iii). $\square$

Combining Theorem 2.3 with Proposition 3.1, if $p \nmid (n+1)$, we have

$$|E_n(\chi, b)| \leq (n+2)! \text{Vol}(\Delta) \cdot q^{\frac{n+2}{2}} = 2(n+1)q^{\frac{n+2}{2}}, \quad (3.7)$$

where p is the characteristic of $\mathbb{F}_q$. Putting (3.4) and (3.7) together, we then obtain the following bounds when $p \nmid (n+1)$.

$$|S_n(\chi, b) + \frac{(q-1)^n}{q} \chi_1(b)| \leq 2(n+1)q^{\frac{n}{2}}, \quad \text{if } \chi_1 = \cdots = \chi_{n+1},$$

and

$$|S_n(\chi, b)| \leq 2(n+1)q^{\frac{n}{2}}, \quad \text{if } \chi_i \neq \chi_j \text{ for some } i \neq j.$$

In the case $\chi_1 = \cdots = \chi_{n+1}$, the twisted sum $E_n(\chi, b)$ becomes the following untwisted toric exponential sum

$$E_n(\chi, b) = \sum_{y, z, x_i \in \mathbb{F}_q^*} \psi \left(z + y \left(1 - z \left(x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \right) \right) \right).$$

Since the origin is a vertex of Δ and the polynomial inside the additive character has no constant term, 1 is a trivial eigenvalue of the middle dimensional cohomology. Removing this trivial eigenvalue from the error term, one gets

$$|E_n(\chi, b) - (-1)^{n+2}| \leq (2n+1)q^{\frac{n}{2}}, \quad \text{if } \chi_1 = \cdots = \chi_{n+1},$$

and hence the slightly sharper estimate

$$|S_n(\chi, b) + \frac{(q-1)^n + (-1)^{n+1}}{q} \chi_1(b)| \leq (2n+1)q^{\frac{n}{2}}, \quad \text{if } \chi_1 = \cdots = \chi_{n+1}.$$

This proves Theorem 1.2.

3.3 Proof of Theorem 1.3

Similar to formula (3.4), we relate the untwisted *inverted* Kloosterman sum $S_n(k, b)$ to toric exponential sum $S_k^*(f)$.

$$\begin{aligned} S_n(k, b) &= \sum_{\substack{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \neq 0 \\ x_i \in \mathbb{F}_{q^k}^*}} \psi \left(\text{Tr}_k \left(\frac{1}{x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n}} \right) \right) \\ &= \sum_{\substack{z \left(x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \right) = 1 \\ z, x_i \in \mathbb{F}_{q^k}^*}} \psi \left(\text{Tr}_k(z) \right) \\ &= \frac{1}{q^k} \sum_{\substack{z, x_i \in \mathbb{F}_{q^k}^* \\ y \in \mathbb{F}_{q^k}^*}} \psi \left(\text{Tr}_k \left(z + y \left(1 - z \left(x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \right) \right) \right) \right) \\ &= -\frac{(q^k - 1)^n}{q^k} + \frac{1}{q^k} S_k^*(f). \end{aligned} \quad (3.8)$$

where f is the Laurent polynomial given by

$$f(x_1, \dots, x_{n+2}) = x_{n+1} \left(1 - x_{n+2} \left(x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \right) \right) + x_{n+2}$$

and

$$S_k^*(f) = \sum_{x_i \in \mathbb{F}_{q^k}^*} \psi \left(\text{Tr}_k \left(x_{n+2} + x_{n+1} \left(1 - x_{n+2} \left(x_1 + \cdots + x_n + \frac{b}{x_1 \cdots x_n} \right) \right) \right) \right).$$

The L-functions associated to $S_n(k, b)$ and $S_k^*(f)$ are defined as

$$L_n(b, T) = \exp \left(\sum_{k=1}^{\infty} S_n(k, b) \frac{T^k}{k} \right) \quad \text{and} \quad L^*(f, T) = \exp \left(\sum_{k=1}^{\infty} S_k^*(f) \frac{T^k}{k} \right).$$

It follows from formula (3.8) that

$$\begin{aligned} L_n(b, T) &= \exp \left(\sum_{k=1}^{\infty} -\frac{(q^k - 1)^n \cdot T^k}{q^k \cdot k} \right) L^*(f, T/q) \\ &= \prod_{i=0}^n \exp \left((-1)^{n-i+1} \binom{n}{i} \sum_{k=1}^{\infty} \frac{(q^{i-1} T)^k}{k} \right) L^*(f, T/q) \\ &= L^*(f, T/q) \prod_{i=0}^n \left(\frac{1}{1 - q^{i-1} T} \right)^{(-1)^{n-i+1} \binom{n}{i}}. \end{aligned} \quad (3.9)$$

The main purpose of this subsection is to determine the slopes and weights of $L_n(b, T)$. Based on formula (3.9), it suffices to consider $L^*(f, T)$ instead. Let $\Delta = \Delta(f)$ denote the Newton polyhedron corresponding to f . Some of the geometric properties about Δ have been discussed in Sect. 3.2. In Proposition 3.1, we proved that f is non-degenerate if and only if $p \nmid (n+1)$. In this case, the L-function $L^*(f, T)^{(-1)^{n+1}}$ is a polynomial of degree $2n+2$. To determine the slopes of the reciprocal roots of $L^*(f, T)^{(-1)^{n+1}}$, we shall compute the Hodge polygon and consider when it coincides with the Newton polygon.

Proposition 3.2 *The Laurent polynomial f is ordinary if and only if $p \equiv 1 \pmod{n+1}$.*

Proof By facial decomposition theorem, it suffices to consider f^{δ_i} for $i = 1, 2$. Let $S(\delta_i)$ be the solution set of the following linear system

$$M(\delta_i) \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{n+2} \end{pmatrix} = u \in \mathbb{Z}^{n+2}, \quad \text{where } r_j \in \mathbb{Q} \cap [0, 1). \quad (3.10)$$

For $i = 1$ and a given point $u = (x_1, \dots, x_{n+2})^T$, linear system (3.10) equals to

$$\begin{cases} x_1 = r_1 - r_{n+1}, \\ x_2 = r_2 - r_{n+1}, \\ \dots \\ x_n = r_n - r_{n+1}, \\ x_{n+1} = r_1 + \dots + r_{n+2}, \\ x_{n+2} = r_1 + \dots + r_{n+1}, \end{cases} \quad \text{where } r_j \in \mathbb{Q} \cap [0, 1). \quad (3.11)$$

Note that $x_j \in \mathbb{Z}$, where $1 \leq j \leq n+2$. For any $r = (r_1, \dots, r_{n+2})^T \in S(\delta_1)$, we have

$$r_1 = \dots = r_n = r_{n+1} \in \frac{\mathbb{Z}}{n+1} \quad \text{and} \quad r_{n+2} = 0.$$

Let $S_p(\delta_i)$ denote the prime to p part of $S(\delta_i)$. In particular, $S_p(\delta_i) = S(\delta_i)$ if $p \nmid \det(M(\delta_i))$. Suppose $p \nmid (n+1)$, the norm function $|r|$ and $|\{pr\}|$ are given by

$$|r| = (n+1)r_1 \quad \text{and} \quad |\{pr\}| = (n+1)\{pr_1\}.$$

Then $|r|$ on $S_p(\delta_1)$ is stable under the p -action if and only if $p \equiv 1 \pmod{n+1}$. To see this, it suffices to consider the unique point $r = (\frac{1}{n+1}, \dots, \frac{1}{n+1}, 0)$ with norm 1. This condition holds for $S_p(\delta_2)$ through a similar proof. By Proposition 2.19, we obtain Proposition 3.2. $\square$

Theorem 3.3 *The $n+2$ Hodge numbers of Δ are $\{1, 2, 2, \dots, 2, 1\}$. Namely,*

$$H_{\Delta}(0) = 1, H_{\Delta}(1) = \dots = H_{\Delta}(n) = 2, H_{\Delta}(n+1) = 1.$$

Proof Let Δ_i be the polytope generated by the origin and δ_i . Let $u = (x_1, \dots, x_{n+2})^T \in C(\Delta_i)$ be a lattice point with the weight $w(u) = k$, where $0 \leq k \leq n+2$. For $i = 1, 2$, consider the linear system (3.10). Since f^{δ_i} is diagonal, system (3.10) has a unique solution $r = (r_1, \dots, r_{n+2})^T$ for a fixed point $u \in C(\Delta_i)$. In this case, the weight is given by

$$w(u) = r_1 + \dots + r_{n+2} = |r|.$$

When $i = 1$, the linear equations (3.11) has exact one solution $u = (0, \dots, 0, k, k)^T$. Since $x_{n+2} = \sum_{j=1}^{n+1} r_j = k$ and $0 \leq r_j < 1$, we get the restriction $0 \leq k < n+1$. The Hodge number $H_{\Delta_1}(k)$ counts the number of lattice points u of weight k/D in a fundamental domain: That is,

$$H_{\Delta_1}(k) = \begin{cases} 1, & \text{for } 0 \leq k < n+1, \\ 0, & \text{for } k \geq n+1. \end{cases}$$

The generating function of $H_{\Delta_1}(k)$ is

$$H_1(x) = 1 + x + \dots + x^n.$$

By formula (2.5), we get the generating function of $W_{\Delta_1}(k)$ as follow.

$$W_1(x) = \sum_{k=0}^{\infty} W_{\Delta_1}(k)x^k = \frac{H_1(x)}{(1-x)^{n+2}} = \frac{1-x^{n+1}}{(1-x)^{n+3}}.$$

Let $H_2(x)$ and $W_2(x)$ be the generating function of $H_{\Delta_2}(k)$ and $W_{\Delta_2}(k)$, respectively. Similarly, we can prove $H_2(x) = H_1(x)$ and $W_2(x) = W_1(x)$. The polytope $\Delta_1 \cap \Delta_2$ is determined by $V_1, \dots, V_{n+1}$, whose generating function is given by

$$W_3(x) = \sum_{k=0}^{\infty} W_{\Delta_1 \cap \Delta_2}(k)x^k = \frac{1-x^{n+1}}{(1-x)^{n+2}}.$$

By facial decomposition, we have

$$W_{\Delta}(k) = W_{\Delta_1}(k) + W_{\Delta_2}(k) - W_{\Delta_1 \cap \Delta_2}(k), \quad (3.12)$$

which implies

$$W(x) = \sum_{k=0}^{\infty} W_{\Delta}(k)x^k = W_1(x) + W_2(x) - W_3(x) = \frac{1+2x+\dots+2x^n+x^{n+1}}{(1-x)^{n+2}}.$$

This gives the Hodge numbers of Δ via formula (2.6), that is,

$$H_{\Delta}(0) = 1, H_{\Delta}(1) = \dots = H_{\Delta}(n) = 2, H_{\Delta}(n+1) = 1.$$

$\square$

When f is ordinary, the slopes of $L^*(f, T)^{(-1)^{n+1}}$ can be deduced from Theorem 3.3.

Theorem 3.4 *If $p \equiv 1 \pmod{n+1}$, the slope sequence of $L^*(f, T)^{(-1)^{n+1}}$ is given by*

$$\{0, 1, 1, 2, 2, \dots, n, n, n+1\}.$$

Proof This theorem follows from Lemma 2.5, Proposition 3.2 and Theorem 3.3. $\square$

Note that the converse of this theorem is also true, as Proposition 3.2 shows that the condition $p \equiv 1 \pmod{n+1}$ is a necessary and sufficient condition for f to be ordinary.

Now we are ready to consider the weights for the reciprocal roots of $L^*(f, T)^{(-1)^{n+1}}$.

Theorem 3.5 *Suppose $p \nmid (n+1)$. We have*

$$L^*(f, T)^{(-1)^{n+1}} = (1-T)(1-qT) \prod_{i=1}^{2n} (1-\beta_i T).$$

For each $1 \leq i \leq 2n$, the reciprocal root β_i satisfies $|\beta_i| = q^{\frac{n+2}{2}}$.

Proof Since the origin is a vertex of Δ , we decompose the cone $C(\Delta)$ via boundary decomposition $B(\Delta)$. Let $N(i)$ be the number of i -dimensional face Σ_i of $C(\Delta)$, where $0 \leq i \leq \dim \Delta$. For Newton polyhedron $\Delta = \Delta(f)$, we have $N(0) = 1$ and $N(1) = n+3$. Note that Σ_i is an open cone and $\Sigma_i \in B(\Delta)$. Let $\overline{\Sigma}_i$ be the closure of Σ_i . For simplicity, we denote the Fredholm determinants as

$$\begin{aligned} D(T) &= \det(I - T A_1(f)), \\ D_i^\circ(T) &= \det\left(I - T A_1(\Sigma_i, f^{\overline{\Sigma}_i})\right), \\ D_i(T) &= \det\left(I - T A_1(\overline{\Sigma}_i, f^{\overline{\Sigma}_i})\right). \end{aligned}$$

The unique 0-dimensional cone $\overline{\Sigma}_0$ is the origin and $D_0(T) = D_0^\circ(T) = 1 - T$. When $i = 1$, each $f^{\overline{\Sigma}_1}$ can be normalized to x by variable substitution. That is,

$$L^*(f^{\overline{\Sigma}_1}, T) = \exp\left(\sum_{k=1}^{\infty} -\frac{T^k}{k}\right) = 1 - T.$$

By formula (2.8), we have

$$D_1(T) = \prod_{i=0}^{\infty} \left(L^*(f^{\overline{\Sigma}_1}, q^i T) \right)^{(i)} = (1-T)(1-qT)(1-q^2T) \cdots.$$

Since the only boundary of $\overline{\Sigma}_1$ are Σ_1 and $\overline{\Sigma}_0$, we get $D_1^\circ(T)$ after eliminating $D_0^\circ(T)$, i.e.,

$$D_1^\circ(T) = \frac{D_1(T)}{D_0^\circ(T)} = (1-qT)(1-q^2T) \prod_{i=3}^{\infty} (1-q^i T).$$

Theorem 2.16 shows that $D(T)$ can be expressed as a product of $D_i^\circ(T)$ as follow.

$$D(T) = \prod_{i=1}^{n+2} \prod_{j=1}^{N(i)} D_i^\circ(T) = (1-T)(1-qT)^{n+3} \cdots.$$

Note that $L^*(f, T)^{(-1)^{n+1}}$ is a polynomial of degree $2(n+1)$ if f is non-degenerate. Combining formula (2.7), we obtain

$$L^*(f, T)^{(-1)^{n+1}} = \frac{D(T)D(q^2T)^{\binom{n+2}{2}} \cdots}{D(qT)^{n+2}D(q^3T)^{\binom{n+2}{3}} \cdots} = (1-T)(1-qT) \prod_{i=1}^{2n} (1 - \beta_i T),$$

where $|\beta_i| = q^{\frac{w_i}{2}} \leq q^{\frac{n+2}{2}}$. That is, $w_i \leq n+2$.

If β_i is a reciprocal root of $L^*(f, T)^{(-1)^{n+1}}$, the conjugate $\overline{\beta_i}$ is a reciprocal root of the conjugate L-function

$$\overline{L^*(f, T)^{(-1)^{n+1}}} = L^*(-f, T)^{(-1)^{n+1}}.$$

By Theorem 2.8, the Newton polygon and Hodge polygon coincide at the end points. Applying this to the product $L^*(f, T)^{(-1)^{n+1}} \overline{L^*(f, T)^{(-1)^{n+1}}}$, we deduce that

$$\begin{aligned} 2(n+1)^2 &= 2 \left(\sum_{i=1}^n 2i + n + 1 \right) = \text{ord}_q \left(1 \cdot q^2 \cdot \prod_{i=1}^{2n} \beta_i \overline{\beta_i} \right) \\ &= 2 + \sum_{i=1}^{2n} w_i \leq 2 + 2n(n+2) = 2(n+1)^2. \end{aligned}$$

It follows that the inequality must be an equality, that is, all $w_i = n+2$. $\square$

Formula (3.9) relates $L^*(f, T)$ to $L_n(b, T)$. The valuations for the reciprocal roots and poles of $L_n(b, T)$ follow from Theorems 3.4 and 3.5.

Theorem 3.6 Suppose $p \nmid (n+1)$. We have

$$L_n(b, T)^{(-1)^{n+1}} = (1-T)^{(n+1)} \prod_{j=2}^n \left(1 - q^{j-1} T \right)^{\binom{n}{j} (-1)^{j-1}} \prod_{i=1}^{2n} (1 - \alpha_i T).$$

For each $1 \leq i \leq 2n$, the reciprocal root α_i satisfies $|\alpha_i| = q^{\frac{n}{2}}$. If $p \equiv 1 \pmod{n+1}$, the slope sequence of the α_i 's is given by $\{0, 1, 1, 2, 2, \dots, n-1, n-1, n\}$.

Based on weights of toric L-function, we get the following slightly more precise upper bound for its associated exponential sum.

Corollary 3.7 If $p \nmid (n+1)$, we have

$$|S_n(k, b) + \frac{(q^k - 1)^n - (-1)^n (q^k + 1)}{q^k}| \leq 2nq^{\frac{nk}{2}}.$$

Proof Theorem 3.5 implies that

$$|S_k^*(f) - (-1)^n (q^k + 1)| \leq 2nq^{\frac{(n+2)k}{2}}.$$

Combining formula (3.8), we get the bound for $S_n(k, b)$. $\square$

Remark We finish this paper with two open problems on the estimations of *inverted* Kloosterman sums. If $n+1$ is divisible by p , the related Laurent polynomial f is degenerate and thus the results for toric exponential sums are not tenable. In this case, it is an open problem to determine the optimal square root cancellation for $S_n(\chi, b)$ in general. The case $n=1$ with

$p = 2$ is already handled in [15]. The second question concerns the q -adic slope sequence. If p is not equivalent to 1 modulo $n + 1$, the Newton polygon corresponding to f is strictly above its Hodge polygon. Under this assumption, can one still obtain the explicit q -adic slope sequence? $\square$

References

- Angel, J.: Finite upper half planes over finite fields. *Finite Fields Appl.* **2**(1), 62–86 (1996)
- Adolphson, A., Sperber, S.: Newton polyhedra and the degree of the L -function associated to an exponential sum. *Invent. Math.* **88**(3), 555–569 (1987)
- Adolphson, A., Sperber, S.: Twisted Kloosterman sums and p -adic Bessel functions. II. Newton polygons and analytic continuation. *Am. J. Math.* **109**(4), 723–764 (1987)
- Adolphson, A., Sperber, S.: Exponential sums and newton polyhedra: cohomology and estimates. *Ann. Math.* **130**(2), 367–406 (1989)
- Adolphson, A., Sperber, S.: Exponential sums on $(G_m)^n$. *Invent. Math.* **101**(1), 63–79 (1990)
- Adolphson, A., Sperber, S.: On twisted exponential sums. *Math. Ann.* **290**(4), 713–726 (1991)
- Adolphson, A., Sperber, S.: Twisted exponential sums and Newton polyhedra. *J. Reine Angew. Math.* **443**, 151–177 (1993)
- Chen, C., Lin, X.: L -functions of certain exponential sums over finite fields. *Math. Z.* **300**(2), 1851–1871 (2022)
- Deligne, P.: La conjecture de Weil. II. *Inst. Hautes Études Sci. Publ. Math.* (52), 137–252 (1980)
- Denef, J., Loeser, F.: Weights of exponential sums, intersection cohomology, and Newton polyhedra. *Invent. Math.* **106**(1), 275–294 (1991)
- Evans, R.: Spherical functions for finite upper half planes with characteristic 2. *Finite Fields Appl.* **1**(3), 376–394 (1995)
- Fu, L.: Weights of twisted exponential sums. *Math. Z.* **262**(2), 449–472 (2009)
- Fu, L.: ℓ -adic GKZ hypergeometric sheaves and exponential sums. *Adv. Math.* **298**, 51–88 (2016)
- Fu, L., Wan, D.: On Katz’s (A, B) -exponential sums. *Q. J. Math.* **72**(3), 773–793 (2021)
- Katz, N.M.: A note on exponential sums. *Finite Fields Appl.* **1**(3), 395–398 (1995)
- Koblitz, N.: p -Adic Numbers, p -Adic Analysis, and Zeta-Functions. *Graduate Texts in Mathematics*, vol. 58, 2nd edn. Springer, New York (1984)
- Lin, X., Chen, C.: L -functions of certain exponential sums over finite fields II. *J. Number Theory* **241**, 198–220 (2022)
- Li, W.-C.W.: Estimates of character sums arising from finite upper half planes. In: Cohen, S., Niederreiter, H. (eds.) *Finite Fields and Applications*. London Mathematical Society Lecture Note Series, vol. 233, pp. 219–228. Cambridge University Press, Cambridge (1996)
- Li, J.: Newton polygons of L -functions associated to Deligne polynomials. *Finite Fields Appl.* **75**, Paper No. 101880, 10 (2021)
- Lubotzky, A., Phillips, R., Sarnak, P.: Ramanujan graphs. *Combinatorica* **8**(3), 261–277 (1988)
- Sabbah, C.: Hypergeometric periods for a tame polynomial. *Port. Math. (N.S.)* **63**(2), 173–226 (2006)
- Sperber, S.: Congruence properties of the hyper-Kloosterman sum. *Compos. Math.* **40**(1), 3–33 (1980)
- Terras, A.: *Fourier Analysis on Finite Groups and Applications*. London Mathematical Society Student Texts, vol. 43. Cambridge University Press, Cambridge (1999)
- Wan, D.: Newton polygons of zeta functions and L functions. *Ann. Math.* **137**, 249–293 (1993)
- Wan, D.: Variation of p -adic Newton polygons for L -functions of exponential sums. *Asian J. Math.* **8**(3), 427–472 (2004)
- Yu, J.-D.: Irregular Hodge filtration on twisted de Rham cohomology. *Manuscr. Math.* **144**(1–2), 99–133 (2014)
- Yang, L., Zhang, H.: Generic Newton polygons for L -functions of (A, B) -exponential sums. *Finite Fields Appl.* **78**, Paper No. 101980, 20 (2022)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.