Math. Res. Lett. **15** (2008), *no.* 5, 1053–1071

HODGE-STICKELBERGER POLYGONS FOR L-FUNCTIONS OF EXPONENTIAL SUMS OF $P(x^s)$

Régis Blache, Éric Férard and Hui June Zhu

Abstract. Let $\mathbb{F}_q$ be a finite field of cardinality q and characteristic p. Let $\overline{P}(x)$ be any one-variable Laurent polynomial over $\mathbb{F}_q$ of degree (d_1, d_2) respectively and $p \nmid d_1 d_2$. For any fixed $s \geq 1$ coprime to p, we prove that the q-adic Newton polygon of the L-functions of exponential sums of $\overline{P}(x^s)$ has a tight lower bound which we call Hodge-Stickelberger polygon, depending only on the d_1, d_2, s and the residue class of (p mod s). This Hodge-Stickelberger polygon is a certain weighted convolution of the Hodge polygon for L-function of exponential sums of $\overline{P}(x)$ and the Newton polygon for the L-function of exponential sums of x^s (which is precisely given by the classical Stickelberger theory). We have an analogous Hodge-Stickelberger lower bound for multivariable Laurent polynomials as well.

For any $\nu \in (\mathbb{Z}/s\mathbb{Z})^\times$, we show that there exists a Zariski dense open subset $\mathcal{U}_\nu$ defined over $\mathbb{Q}$ such that for every Laurent polynomial P in $\mathcal{U}_\nu(\overline{\mathbb{Q}})$ the q-adic Newton polygon of $L(\overline{P}(x^s)/\mathbb{F}_q; T)$ converges to the Hodge-Stickelberger polygon as p approaches infinity and $p \equiv \nu \bmod s$.

As a corollary, we obtain a tight lower bound for the q-adic Newton polygon of the numerator of the zeta function of an Artin-Schreier curve given by affine equation $y^p - y = \overline{P}(x^s)$. This estimates the q-adic valuations of reciprocal roots of the numerator of the zeta function of the Artin-Schreier curve.

1. Introduction

Let $\mathscr{A}_{d_1,d_2}$ be the space of all Laurent polynomials in one variable x of degree (d_1, d_2) (in x and x^{-1} respectively) where $d_1, d_2 \geq 1$. They are just rational functions with poles at ∞ and 0. The one-pole polynomial case (i.e., $d_2 = 0$) will also be considered along the line. For our purpose, we may assume that each Laurent polynomial is monic at x^{d_1}, and hence the coefficient space $\mathscr{A}_{d_1,d_2} = \mathbb{A}^{d_1+d_2-1} \times \mathbb{G}_m$ is an affine variety of dimension $d_1 + d_2$. In this paper p is a prime coprime to $d_1 d_2$. Let $E(x)$ be the Artin-Hasse exponential function, namely, $E(x) = \exp(\sum_{i=0}^{\infty} x^{p^i}/p^i)$. Let γ be a p-adic root of $\log(E(x))$ in the algebraic closure of $\mathbb{Q}_p$ with $\mathrm{ord}_p \gamma = 1/(p-1)$. Then $E(\gamma)$ is a primitive p-th root of unity, which we fix for the rest of the paper and denote it by ζ_p.

Let a be a positive integer and $q = p^a$. Let $\overline{P}(x)$ be a rational function on the projective line with two poles of order d_1 and d_2 respectively. Up to an isomorphism

Received by the editors October 19, 2007.

2000 *Mathematics Subject Classification.* 11,14.

Key words and phrases. Newton polygon, Hodge polygon, Hodge-Stickelberger polygon, L-function, exponential sums, twisted exponential sums, Artin-Schreier curves, Dwork trace formula.

over $\overline{\mathbb{F}}_p$ we may assume the poles are at ∞ and 0 and write

$$\overline{P}(x) = \sum_{i=-d_2}^{d_1} \overline{a}_i x^i$$

where $\overline{a}_i$ lies in $\mathbb{F}_q$ and $\overline{P} \in \mathscr{A}_{d_1,d_2}(\mathbb{F}_q)$. For any positive integer k, let $\psi_{q^k} : \mathbb{F}_{q^k} \to \mathbb{Q}(\zeta_p)^\times$ be a nontrivial additive character of $\mathbb{F}_{q^k}$ and we fix $\psi_{q^k}(\cdot) = \zeta_p^{\mathrm{Tr}_{\mathbb{F}_{q^k}/\mathbb{F}_p}(\cdot)}$. The k-th exponential sum of $\overline{P}(x) \in \mathbb{F}_q[x, x^{-1}]$ is $S_k(\overline{P}) = \sum_{x \in \mathbb{F}_{q^k}^\times} \psi_{q^k}(\overline{P}(x))$. The L-function of the exponential sum of $\overline{P}$ is defined by

$$L(\overline{P}(x); T) = \exp(\sum_{k=1}^{\infty} S_k(\overline{P}) \frac{T^k}{k}).$$

It is known that $L(\overline{P}(x)/\mathbb{F}_q; T) = 1 + b_1 T + \cdots + b_{d_1+d_2} T^{d_1+d_2} \in \mathbb{Z}[\zeta_p][T]$. The most important information about the L-function is its reciprocal roots. They are Weil q-numbers, i.e., algebraic integers all Galois conjugates are of absolute value $\sqrt{q}$. This paper concerns their q -adic absolute value. This can be effectively studied in terms of q-adic Newton polygon of the L-function. The q-adic Newton polygon $\mathrm{NP}_q(\overline{P}(x); \mathbb{F}_q)$ of this L-function is defined as the lower convex hull of the points $(i, \mathrm{ord}_q(b_i))_{i \geq 0}$ on the (x, y)-plane. Results about this Newton polygon can be found in [16, 20, 21]. This polygon is independent of the choice of base field $\mathbb{F}_q$ in $\overline{\mathbb{F}}_p$ (even though the reciprocal roots of the L-function do depend on $\mathbb{F}_q$). The relation between q-adic valuation of roots of a polynomial and its q-adic Newton polygon is explained in details in [11, Chapter IV].

We fix once and for all a positive integer $s \geq 1$. All primes p we consider will be assumed prime to s. The main subject of study of this paper is $L(\overline{P}(x^s)/\mathbb{F}_q; T)$ and its reciprocal roots. Let σ be the permutation on the set $\{0, \ldots, s-1\}$ induced by multiplication of p modulo s. We write its cycle decomposition $\sigma = \prod_{i=1}^{u} \sigma_i$ for ℓ_i-cycles σ_i (including 1-cycles). Let

$$\lambda_i := \frac{\sum_{j \in \sigma_i} j}{s\ell_i}.$$

So $0 \leq \lambda_i < 1$. Note that ℓ_i and λ_i are invariants depending only on s, ν (defined as the least residue of p modulo s) and the cycle σ_i, but independent of p. See Section 3.1 for more details. Note that for $s|(q-1)$ one recovers the classical formula $\lambda_i = \frac{s_p((q-1)r/s)}{a(p-1)}$, where $s_p(n)$ denotes the sum of p-adic expansions of the integer n.

We now define $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s)$, the *Hodge-Stickelberger* polygon of $L(\overline{P}(x^s)/\mathbb{F}_q; T)$, as the polygon with line segments of slopes and lengths

$$(1) \qquad (\frac{m+1-\lambda_i}{d_1}, \ell_i)_{1 \leq i \leq u, 0 \leq m \leq d_1 - 1}; \quad (\frac{m+\lambda_i}{d_2}, \ell_i)_{1 \leq i \leq u, 0 \leq m \leq d_2 - 1}.$$

Note that this polygon contains segments $(0, 1)$ and $(1, 1)$, it is symmetric in the sense that for every slope α there is a slope $1-\alpha$ of equal length (if σ_i is the cycle containing $r > 0$, σ_j the one containing $s - r$, then $\lambda_i + \lambda_j = 1$). This polygon depends only on d_1, d_2, ν, s and is of total horizontal length $s(d_1 + d_2)$.

If $d_2 = 0$ then the Hodge-Stickelberger polygon is given by the first half of the line segments in (1) minus the segment $(1, 1)$ hence of horizontal length $sd_1 - 1$.

Remark 1.1. Consider the Gauss sum over $\mathbb{F}_q$ defined by

$$G_{\mathbb{F}_q}(\psi_q, \chi_s^r) := -\sum_{x\in\mathbb{F}_q^\times} \psi_q(x)\chi_s^{-r}(x)$$

(where χ_s is a multiplicative character of order s on $\mathbb{F}_q^\times$). The Stickelberger's theorem (see [3, Theorem 11.2.1] or [18]) says that $\mathrm{ord}_q(G_{\mathbb{F}_q}(\psi_q, \chi_s^i)) = \lambda_i$. In fact, one can show that $L(x^s;\mathbb{F}_p) = \prod_i(1 - T^{\ell_i}G_{\mathbb{F}_{p^{\ell_i}}}(\psi_{p^{\ell_i}}, \chi_s^i))$ where i ranges over all distinct cycles in σ (see [10] or [17]). Thus the exact shape of the p-adic Newton polygon of $L(x^s;\mathbb{F}_p)$ consists of line segments $(\lambda_i, \ell_i)_{2\le i\le u}$ (by omitting the 1-cycle $\sigma_1 = (0)$).

By the remark above, our Hodge-Stickelberger polygon can be considered as a weighted convolution of the Hodge polygon $\mathrm{HP}(\mathscr{A}_{d_1,d_2})$ of the L-function $L(\overline{P}/\mathbb{F}_q;T)$ (see [12] for details) and the Newton polygon of $L(x^s/\mathbb{F}_q;T)$. The following theorem states that it gives a lower bound of the q-adic Newton polygon of L-function. We use $\preceq$ to denote one polygon lies over the next one and their endpoints meet.

Theorem 1.2. *For any Laurent polynomial $\overline{P} \in \mathscr{A}_{d_1,d_2}(\mathbb{F}_q)$, we have*

$$\mathrm{NP}_q(\overline{P}(x^s);\mathbb{F}_q) \preceq \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s).$$

These two polygons coincide if and only if $p \equiv 1 \bmod \mathrm{lcm}(sd_1, sd_2)$.

In fact, we have an analogous result for multivariable Laurent polynomials which is stated in Section 6.

Remark 1.3. From [22] we know that $\mathrm{NP}_q(\overline{P}(x^s);\mathbb{F}_q) \preceq \mathrm{HP}(\mathscr{A}_{sd_1,sd_2})$, the latter is the concatenation of the following slopes

$$0, 1, \frac{1}{sd_1}, \ldots, \frac{sd_1-1}{sd_1}, \frac{1}{sd_2}, \ldots, \frac{sd_2-1}{sd_2}$$

in nondecreasing order each of horizontal length 1. Hence it is of total horizontal length $s(d_1 + d_2)$. We easily see the following relation

$$\mathrm{NP}_q(\overline{P}(x^s);\mathbb{F}_q) \preceq \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s) \preceq \mathrm{HP}(\mathscr{A}_{sd_1,sd_2}).$$

Note that $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s) = \mathrm{HP}(\mathscr{A}_{sd_1,sd_2})$ if and only if $\nu = 1$, that is, $p \equiv 1 \bmod s$; while $\mathrm{NP}_q(\overline{P}(x^s);\mathbb{F}_q) = \mathrm{HP}(\mathscr{A}_{sd_1,sd_2})$ if and only if $p \equiv 1 \bmod \mathrm{lcm}(sd_1, sd_2)$.

Then we examine, as p varies, the asymptotic behavior of the polygons $\mathrm{NP}(P(x^s) \bmod \mathcal{P})$ where $\mathcal{P}$ is a prime over p. Note that this polygon is independent of the choice of $\mathcal{P}$ and so for ease of notation we may consider $\mathbb{F}_q$ the residue field of $\mathcal{P}$. It is known (see [12], [21]) that when p approaches infinity, there is a Zariski dense open subset $\mathcal{U}$ defined over $\mathbb{Q}$ of the space of rational functions with prescribed poles and polar degrees such that for any rational function lying in $\mathcal{U}(\overline{\mathbb{Q}})$, the Newton polygon $\mathrm{NP}(P(x) \bmod \mathcal{P})$ tends to the associated Hodge polygon $\mathrm{HP}(\mathscr{A}_{d_1,d_2})$. For $s > 2$ such limit does not exist since there is one distinct Hodge-Stickleberger polygon for each residue class of prime p in $(\mathbb{Z}/s\mathbb{Z})^\times$ and for $p \equiv 1 \bmod \mathrm{lcm}(sd_1, sd_2)$ the Newton polygon coincides with the Hodge-Stickelberger polygon. See more discussion on this topic in Section 6. In our main result we show that in each fixed residue class of primes, the situation is similar to the case $s = 1$.

Theorem 1.4. *For every integer $1 \leq \nu \leq s-1$ coprime to s, there exists a Zariki dense open subset $\mathcal{U}_\nu$ in $\mathscr{A}_{d_1,d_2}$ defined over $\mathbb{Q}$ where $d_1 \geq 1$ and $d_2 \geq 0$, such that for any $P(x)$ lying in $\mathcal{U}_\nu(\overline{\mathbb{Q}})$, we have*

$$\lim_{p\to\infty, p\equiv\nu \bmod s} \mathrm{NP}(P(x^s) \bmod \mathcal{P}) = \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s)$$

for all primes $\mathcal{P}$ over p.

These theorems about exponential sums have applications to the Zeta function of Artin-Schreier curves over $\mathbb{F}_q$, namely the projective curves C defined by affine equation $y^p - y = \overline{P}(x)$ over $\mathbb{F}_q$. It is well known that all reciprocal roots of the numerator of the Zeta function of C are eigenvalues of Frobenius endomorphism, and they are Weil q-numbers. The following corollary estimates the q-adic absolute values of these reciprocal roots. We explore it via the q-adic Newton polygon $\mathrm{NP}_q(C/\mathbb{F}_q)$, defined as the q-adic Newton polygon of the numerator of the Zeta function of C. In this paper a constant c multiple of a polygon means the image of the polygon under the homothety with center at origin and ratio c.

Corollary 1.5. *(i) Let $\mathrm{NP}(C_s/\mathbb{F}_q)$ be the q-adic Newton polygon of the Artin-Schreier curve $C_s : y^p - y = \overline{P}(x^s)$ over $\mathbb{F}_q$. Then $\frac{1}{p-1}\mathrm{NP}(C_s/\mathbb{F}_q) \preceq \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s)$. These two polygons coincide if and only if $p \equiv 1 \bmod \mathrm{lcm}(sd_1, sd_2)$. If $\overline{P}$ has only one pole of degree $d_1 \geq 1$ (and $d_2 = 0$) then the two polygons coincide if and only if $p \equiv 1 \bmod sd_1$ or $d_1 = 1$.*

(ii) For every integer $1 \leq \nu \leq s-1$ coprime to s, there exists a Zariski dense open subset $\mathcal{U}_\nu$ in $\mathscr{A}_{d_1,d_2}$, defined over $\mathbb{Q}$, such that for any $P(x)$ lying in $\mathcal{U}_\nu(\overline{\mathbb{Q}})$, we have

$$\lim_{p\to\infty, p\equiv\nu \bmod s} \frac{1}{p-1}\mathrm{NP}(C_s \bmod \mathcal{P}) = \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s)$$

for any prime $\mathcal{P}$ over p.

Proof. Results in Theorems 1.2 and 1.4 can be translated directly to this corollary by using the same argument as that in [22, Corollary 1.3]. □

Remark 1.6. We remark that in the above corollary, one may replace the affine coefficient variety $\mathscr{A}_{d_1,d_2}$ by the moduli space $\mathcal{AS}_g$ of Artin-Schreier curves of genus $g := \frac{(p-1)(d_1+d_2)}{2}$ as defined in [14].

We conclude this introduction by providing general notation and organization of this paper. Throughout the entire paper, we fix integers $d_1, d_2, s \geq 1$. We consider prime numbers p that are always coprime to sd_1d_2. We always assume the residue field of the prime ideal $\mathcal{P}$ is $\mathbb{F}_q$, where q is a p-power and we write $q = p^a$. The permutation σ is induced on the set $\{0, \ldots, s-1\}$ by multiplication of p modulo s. We always write its cycle decomposition as $\sigma = \prod_{i=1}^{u} \sigma_i$ including 1-cycles. Finally, we denote by $E(x)$ be the p-adic Artin-Hasse exponential function. Our main theorems 1.2 and 1.4 are proved at the end of Section 5. Similar result on twisted exponential sums is given in Propositions 3.7 and 4.2 of Sections 3 and 4 respectively. At the end of the paper in section 6 we discuss some open questions and give statement of multivariable cases analog of Theorem 1.2.

2. Two lemmas about nuclear matrices

To make the proofs of our results as smooth as possible, we summarize some fringe results here. These results will be employed in Sections 3 and 4. The reader may wish to skip this section at first reading.

Let K be any complete non-Archimedean field with p-adic valuation $|\cdot|_p$. We refer the readers to [15] for basic facts about Serre's theory of completely continuous maps and Fredholm determinants. For any K-Banach spaces V and V' that admit orthonormal basis, denote by $\mathcal{C}(V,V')$ the set of completely continuous K-linear maps from V to V'. We say that a matrix M over K is *nuclear* if there exists a K Banach space V and a u in $\mathcal{C}(V,V)$ such that M is the matrix of u with respect to some orthonormal basis of V. If $M=(m_{ij})_{i,j\geq 1}$ is a matrix over K, then M is nuclear if and only if $\lim_{i\to\infty}\inf_{j\geq 1}\mathrm{ord}_p m_{ij}=+\infty$.

Lemma 2.1. *Let $\vec{M}=(M_0, M_1,\cdots,M_{a-1})$ be an $a-$ tuple of nuclear matrices over $\mathbb{C}_p$. Set the block matrix*

$$\vec{M}_{[a]}:=\begin{pmatrix} 0 & & \cdots & 0 & M_{a-1}\\ M_0 & 0 & & & 0\\ 0 & M_1 & 0 & & \vdots\\ \vdots & & \ddots & 0 & \\ 0 & \cdots & 0 & M_{a-2} & 0 \end{pmatrix}.$$

Then $\det(1-(M_{a-1}\cdots M_1M_0)T^a)=\det(1-\vec{M}_{[a]}T)$.

Proof. See [12, Section 5]. □

Lemma 2.2. *Let $\{M_t\}_{t=0,\ldots,a-1}$ be any nuclear matrices over K. Let $\mathcal{A}_t$ be the set of all $k\times k$ submatrices in M_t . Fix an integer $k\geq 1$ and let c_k be the coefficient of T^k in* $\det(1-M_{a-1}M_{a-2}\cdots M_0T)$. *Then we have* $\mathrm{ord}_p c_k\geq\sum_{t=0}^{a-1}\inf_{W_t\in\mathcal{A}_t}\mathrm{ord}_p(\det W_t)$.

Proof. By Lemma 2.1, c_k is the coefficient of T^{ak} in the T -adic expansion of $\det(1-\vec{M}_{[a]}T)$, which is the infinite sum of $(-1)^{ak}\det N$ where N runs over all principal $ak\times ak$ submatrices in $\vec{M}_{[a]}$. Let N be such a matrix, and let N_t be the intersection of N and M_t as submatrices of $\vec{M}_{[a]}$ for all $0\leq t\leq a-1$. It is easy to see that $\det N=(-1)^{(ak-1)k}\prod_{t=0}^{a-1}\det N_t$ or 0 depending on whether every N_t is a $k\times k$ submatrix of M_t or not. So for p-adic evaluation purpose, we may assume every N_t is a $k\times k$ matrix. Think of N_t as a submatrix of M_t from now on and $N_t\in\mathcal{A}_t$. Our assertion follows immediately. □

3. L-functions of twisted exponential sums

In this section we assume $s|(q-1)$. Let $k\geq 1$. Let χ_s be a multiplicative character of order s defined on $\mathbb{F}_{q^k}^\times$. We fix it as $\chi_s=\chi\circ N_{\mathbb{F}_{q^k}/\mathbb{F}_q}(\cdot)$ where χ is a multiplicative character of order s on $\mathbb{F}_q^\times$.

Fix an integer $0\leq r\leq s-1$, let σ_i be the cycle of σ containing r, and $\lambda:=\lambda_i=\sum_{j\in\sigma_i}j/(s\ell_i)$. For any Laurent polynomial $\overline{P}(x)$ in $\mathscr{A}_{d_1,d_2}(\mathbb{F}_q)$, define the L-function

$$L(\overline{P}(x)/\mathbb{F}_q, \chi_s^r; T) := \exp(\sum_{k=1}^{\infty} S_k(\overline{P}, \chi_s^r)\frac{T^k}{k}). \tag{2}$$

where $S_k(\overline{P}, \chi_s^r) = \sum_{x\in\mathbb{F}_{q^k}^\times} \psi_{q^k}(\overline{P}(x))\chi_s^r(x)$.

From Weil's theorem, this L-function is a polynomial of degree $d_1 + d_2$ and its reciprocal roots in $\mathbb{C}$ are algebraic integers with Archimedean absolute value $q^{1/2}$ and ℓ-adic absolute value 1 for any prime $\ell \neq p$. We shall study the q-adic absolute value of these reciprocal roots. We denote by $\mathrm{NP}_q(\overline{P}, \chi_s^r; \mathbb{F}_q)$ the Newton polygon of $L(\overline{P}/\mathbb{F}_q, \chi_s^r; T)$ defined analogously as that for $\mathrm{NP}_q(\overline{P}; \mathbb{F}_q)$.

3.1. Twisted Hodge-Stickelberger polygons. Denote by $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^r)$ the *twisted Hodge-Stickelberger polygon* of multiplicative character χ_s^r with slopes and lengths

$$\{(\frac{m+1-\lambda}{d_1}, 1)_{0\le m\le d_1-1};\quad (\frac{m+\lambda}{d_2}, 1)_{0\le m\le d_2-1}\}.$$

It is of total horizontal length $d_1 + d_2$. This polygon can be found in the literature, for example, see [1, Theorem 3.20] and [2, Corollary 3.18]. In the polynomial case, i.e., $d_2 = 0$, the twisted Hodge-Stickelberger polygon consists of the first half of the above line segments and is of horizontal length d_1 if $r \neq 0$ (and $d_1 - 1$ if $r = 0$, in this case we also remove the segment $(1, 1)$).

Remark 3.1. The twisted Hodge-Stickelberger polygon $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^r)$ we give above coincides with the Hodge polygon defined in [2, Corollary 3.18] in one-variable case. We shall verify this explicitly below. Set $\mathbf{d} := -(q-1)r/s$ in notation of [2]. Then $\mathbf{d}^{(v)} = -(q-1)\sigma^v(r)/s$, and for any $-d_2 + 1 \le j \le d_1$, we have

$$u_{\mathbf{d}^{(v)}}(j) = x^{\frac{\mathbf{d}^{(v)}}{q-1}+j}, \text{ and } w(u_{\mathbf{d}^{(v)}}(j)) = \begin{cases} \frac{1}{d_1}\left(\frac{\mathbf{d}^{(v)}}{q-1} + j\right) = \frac{j}{d_1} - \frac{\sigma^v(r)}{sd_1} & \text{if } j > 0, \\ \frac{1}{d_2}\left(\frac{\mathbf{d}^{(v)}}{q-1} + j\right) = -\frac{j}{d_2} + \frac{\sigma^v(r)}{sd_2} & \text{if } j \le 0. \end{cases}$$

These are due to the fact that the weight of x^r is r/d_1 when $r \ge 0$ and $-r/d_2$ when $r \le 0$ in our case. The Hodge polygon slopes b_j defined in [2, above Theorem 3.17] can be expressed as

$$b_j = \begin{cases} \frac{1}{a}\sum_{v=0}^{a-1}\left(\frac{j}{d_1} - \frac{\sigma^{v-a}(r)}{sd_1}\right) = \frac{j-\lambda}{d_1} & \text{if } j > 0, \\ \frac{1}{a}\sum_{v=0}^{a-1}\left(\frac{-j}{d_2} + \frac{\sigma^{v-a}(r)}{sd_2}\right) = \frac{-j+\lambda}{d_2} & \text{if } j \le 0 \end{cases}$$

These yield exactly the slopes of the twisted Hodge-Stickelberger polygon defined above $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^r)$.

We use $\boxplus$ to denote the concatenation of line segments which are given via pairs of slopes and horizontal length so that the slopes are in non-decreasing order. Now we have the splitting of the Hodge-Stickelberger polygon into twisted Hodge-Stickelberger polygons below in the lemma.

Lemma 3.2. *(i) We have*

$$\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s) = \boxplus_i \ell_i\, \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^{r_i})$$

where the box-sum ranges in the distinct cycles σ_i of σ, and for each i, we have that r_i is a representative in σ_i.

(ii) If $\nu = 1$ *then* $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, 1, s) = \mathrm{HP}(\mathscr{A}_{sd_1,sd_2})$.

Proof. The first statement is clear by the definition of $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s)$ in (1). For the second statement, one only needs to recognize that for $p \equiv 1 \bmod s$ we have $\ell_i = \ell_i' = 1$ for every i and $\lambda_i = r/s$ for every $0 \le r \le s-1$ in σ_i. The rest is explicit and elementary calculation. □

3.2. Trace formula for twisted exponential sums. Let $\mathbb{Q}_q$ denote the unique unramified extension of $\mathbb{Q}_p$ of degree a and $\mathbb{Z}_q$ its ring of integers. Let $\Omega_1 := \mathbb{Q}_p(\zeta_p)$ and let Ω_a the unique unramified extension of Ω_1 of degree a in $\mathbb{C}_p$. Recall that $\gamma \in \Omega_1$ such that $\mathbb{Z}_p[\gamma] = \mathbb{Z}_p[\zeta_p]$. Fix roots γ^{1/d_1} and γ^{1/d_2} in $\mathbb{C}_p$, we denote by $\Omega_1' = \Omega_1(\gamma^{1/d_1}, \gamma^{1/d_2})$ and $\Omega_a' = \Omega_1'\Omega_a$. Below we denote by K (K' respectively) a complete non-Archimedean field containing Ω_a (Ω_a' respectively).

By taking Teichmüller lifts of coefficients of $\overline{P} \in \mathbb{F}_q[x, x^{-1}]$, we get $\hat{P}(x) = \sum_{i=-d_2}^{d_1} \hat{a}_i x^i \in \mathbb{Z}_q[x, x^{-1}]$. Note that $\hat{a}_i{}^q = \hat{a}_i$ and $\hat{a}_i \equiv \overline{a}_i \bmod \mathcal{P}$ where $\mathcal{P}$ is the prime ideal in Ω_a lying over p. For any $0 < \rho < 1$ in $|K|_p$ let $\mathcal{H}_\rho(K)$ be the ring of rigid analytic functions over K on the annulus with $\rho \le |x|_p \le 1/\rho$. It is a p-adic Banach space with the natural p-adic supremum norm.

Let the operator U_q on $\mathcal{H}_\rho$ be defined by $(U_q\xi)(X) := \frac{1}{q}\sum_{Z^q=X} \xi(Z)$ for any $\xi \in \mathcal{H}_\rho$. If $\xi(X) = \sum_{i=-\infty}^{\infty} c_i X^i$ then $U_q(\xi) = \sum_{i=-\infty}^{\infty} c_{iq} X^i$. Let τ be a lifting of the Frobenius of $\overline{\mathbb{F}}_p$ to K such that $\tau(\gamma) = \gamma$. Define three elements in $\mathcal{H}_\rho(K)$ below

$$F(X) = \prod_{i=-d_2}^{d_1} E(\gamma \hat{a}_i X^i), \tag{3}$$

$$F_{[a]}(X) = \prod_{t=0}^{a-1} F^{\tau^t}(X^{p^t}), \tag{4}$$

$$H(X) = X^{\frac{(q-1)r}{s}} F_{[a]}(X). \tag{5}$$

These above are all power series in $\mathbb{Z}_p[\gamma][\vec{\hat{a}_i}][[X]]$ and hence in $\mathbb{Z}_q[\zeta_p][[X]]$. Let $\alpha := U_q \circ H(X)$ by which we mean the composition map of U_q with the multiplication map by $H(X)$. Then α is a completely continuous K-linear endomorphism of $\mathcal{H}_\rho(K)$ for some suitable $0 < \rho < 1$.

Lemma 3.3. *We have*

$$L(\overline{P}/\mathbb{F}_q, \chi_s^r; T) = \frac{\det(1 - T\alpha)}{\det(1 - Tq\alpha)} \tag{6}$$

and it is a polynomial in $\mathbb{Z}[\zeta_p, \zeta_s][T]$ *of degree* $d_1 + d_2$.

Proof. The rationality is a routine consequence of the Dwork-Monsky-Reich trace formula so we omit its proof here. The assertion of its degree follows from [5] (or [2]). □

3.3. p-adic estimate of twisted exponential sums. Let $0 \le r \le s-1$. Write the p-adic expansion

$$(q-1)r/s = \sum_{t=0}^{a-1} K_{r,t} p^t \tag{7}$$

for $0 \le K_{r,t} \le p-1$. Then we have

$$\lambda \quad = \frac{\sum_{t=0}^{a-1} K_{r,t}}{a(p-1)} \quad = \frac{s_p((q-1)r/s)}{a(p-1)} \tag{8}$$

where $s_p(\cdot)$ denotes the sum of p-adic expansions.

Let $F_t(X) = X^{K_{r,t}} F^{\tau^t}(X)$ and

$$\alpha_t \quad := \quad U_p \circ F_t(X).$$

Lemma 3.4. *The maps α_t are completely continuous K -linear endomorphisms of $\mathcal{H}_\rho(K)$ for some suitable $0 < \rho < 1$. We have*

$$\alpha = \alpha_{a-1} \circ \cdots \circ \alpha_1 \circ \alpha_0. \tag{9}$$

Proof. The first statement is Dwork theory. Using $f(x) \circ U_p = U_p \circ f(x^p)$, we have by (7)

$$\alpha_{a-1} \circ \cdots \circ \alpha_0 = (U_p \circ \cdots \circ U_p) \circ (X^{\sum_{t=0}^{a-1} K_{r,t} p^t} F_{[a]}(X)) = U_q \circ H(X) = \alpha.$$

This finishes the proof. □

For any $i \in \mathbb{Z}$, consider the p-adic Mittag-Leffler decomposition $F(X)X^i = \sum_{m=-\infty}^{\infty} H^{m,i} X^m$. Write $\alpha_t(X^i) = \sum_{m=-\infty}^{\infty} B_t^{m,i} X^m$, we have $B_t^{m,i} = \tau^t H^{mp-K_{r,t},i}$. We know $H^{m,i}, B_t^{m,i}$ lie in $\mathbb{Z}_p[\gamma][\vec{\hat{a}_i}]$. Then from the p-adic valuation of the coefficients of exponential function $E(x)$ (see [7]) we have

$$\mathrm{ord}_p B_t^{m,i} \ge \frac{1}{p-1} \max\left(\frac{pm - K_{r,t} - i}{d_1}, -\frac{pm - K_{r,t} - i}{d_2} \right)$$

By p-adic Mittag-Leffler decomposition, every element in the K-linear space $\mathcal{H}_\rho(K)$ can be uniquely represented as $\sum_{i=-\infty}^{\infty} c_i X^i$ for $c_i \in K$, and so $\mathcal{H}_\rho(K)$ has a natural monomial basis $\vec{b}_{\mathrm{unw}} = \{1, X, X^2, \ldots; X^{-1}, X^{-2}, \ldots\}$. Let $Z_1 = \gamma^{1/d_1} X$ and $Z_2 = \gamma^{1/d_2} X^{-1}$, then $\vec{b} = \{Z_1, Z_1^2, \ldots; 1, Z_2, Z_2^2, \ldots\}$ forms a basis for $\mathcal{H}_\rho(K')$. Let M_t be the matrix of α_t with respect to the basis $\vec{b}$. Its entries lie in $\mathbb{Z}_q[\gamma^{1/d_1}, \gamma^{1/d_2}]$. From now on, we shall consider the coefficients liftings $\hat{a}_i$ of $P(x)$ as variables throughout this section, and set $\vec{\hat{a}} = (\hat{a}_i)$, then the entries of M_t lie in $\mathbb{Z}_p[\gamma^{1/d_1}, \gamma^{1/d_2}][\vec{\hat{a}}]$. Note that $\mathrm{ord}_p(\cdot)$ and $\mathrm{ord}_q(\cdot)$ also denote the natural p-adic valuations on the multi-variable polynomial ring $\mathbb{Z}_p[\gamma^{1/d_1}, \gamma^{1/d_2}][\vec{\hat{a}}]$ induced from that on $\mathbb{Z}_p$.

We are ready to give estimates for the p-adic valuations of the coefficients of M_t. Note that we omit the subscript t in the coefficients since no confusion can occur.

Lemma 3.5. *For all $i \ge 0$ we have $\alpha_t Z_J^i = \sum_{m=1}^{\infty} C_{1,J}^{m,i} Z_1^m + \sum_{m=0}^{\infty} C_{2,J}^{m,i} Z_2^m$ where $C_\bigstar^{m,i}$ are the entries of M_t. The lower bounds of $\mathrm{ord}_p C_\bigstar^{m,i}$ are*

$\mathrm{ord}_p(\cdot) \ge$	Z_1^i $(i > 0)$	Z_2^i $(i \ge 0)$
Z_1^m $(m > 0)$	$\boxed{\frac{m}{d_1} - \frac{K_{r,t}}{d_1(p-1)}}$	$\frac{m}{d_1} - \frac{K_{r,t}}{d_1(p-1)} + \frac{i}{p-1}(\frac{1}{d_1} + \frac{1}{d_2})$
Z_2^m $(m \ge 0)$	$\frac{m}{d_2} + \frac{K_{r,t}}{d_2(p-1)} + \frac{i}{p-1}(\frac{1}{d_1} + \frac{1}{d_2})$	$\boxed{\frac{m}{d_2} + \frac{K_{r,t}}{d_2(p-1)}}$

Proof. See [22] page 1542–1543 for details. □

For any $0 \le t \le a-1$, let $\mathcal{L}_t$ be the set of rational numbers $\mathcal{L}_t := \{\frac{m}{d_1} - \frac{K_{r,t}}{d_1(p-1)} | m \ge 1\} \cup \{\frac{m}{d_2} + \frac{K_{r,t}}{d_2(p-1)} | m \ge 0\}$. For every $k \ge 1$ let $\delta_t^{(k)}$ the sum of k least numbers in $\mathcal{L}_t$. Split these k numbers in terms of $j = 1$ or 2 we have $k_1 + k_2 = k$ such that

$$\delta_t^{(k)} = \sum_{m=1}^{k_1} (\frac{m}{d_1} - \frac{K_{r,t}}{d_1(p-1)}) + \sum_{m=0}^{k_2-1} (\frac{m}{d_2} + \frac{K_{r,t}}{d_2(p-1)}).$$

By (8) we have (note that k_1 and k_2 do not depend on t)

$$\begin{aligned} \frac{1}{a}\sum_{t=0}^{a-1} \delta_t^{(k)} &= \frac{k_1(k_1+1)}{2d_1} - \frac{k_1\lambda}{d_1} + \frac{k_2(k_2-1)}{2d_2} + \frac{k_2\lambda}{d_2} \\ &= \frac{k_1(k_1-1)}{2d_1} + \frac{k_1(1-\lambda)}{d_1} + \frac{k_2(k_2-1)}{2d_2} + \frac{k_2\lambda}{d_2}. \end{aligned} \tag{10}$$

Lemma 3.6. *For any $k \times k$ principal submatrix W_t of M_t we have*

$$\mathrm{ord}_p(\det W_t) \ge \delta_t^{(k)}.$$

Proof. The statement follows from Lemma 3.5. □

Proposition 3.7. *Write* $\det(1-\alpha T) = 1 + \sum_{k=1}^{\infty} C_k T^k$, *then*

(i) $\mathrm{ord}_q C_k \ge \frac{1}{a}\sum_{t=0}^{a-1} \delta_t^{(k)}$;
(ii) $\mathrm{NP}_q(\overline{P}, \chi_s^r; \mathbb{F}_q) \preceq \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^r)$ *for all* $\overline{P} \in \mathscr{A}_{d_1,d_2}(\mathbb{F}_q)$.
(iii) *Write* $(q-1)r/s = \sum K_{r,t} p^t$ *for* $0 \le K_{r,t} \le p-1$ *and write* $d = \mathrm{lcm}(d_1, d_2)$. *Then the following statements are equivalent:*
  (a) *the Newton and Hodge-Stickelberger polygons in (ii) coincide.*
  (b) $d | \gcd(p-1, K_{r,t})$ *for every* $0 \le t \le a-1$.

Proof. (i) From the decomposition of α in Lemma 3.4 we can apply the results in Lemma 2.2 to $\det(1-\alpha T)$, and we have

$$\mathrm{ord}_q(C_k) = \frac{1}{a}\mathrm{ord}_p(C_k) \quad \ge \quad \frac{1}{a}\sum_{t=0}^{a-1} \inf_{W_t \in \mathcal{A}_t} (\mathrm{ord}_p \det W_t).$$

The result follows from Lemma 3.6.

(ii) By the trace formula (6), we know that $\mathrm{NP}_q(\overline{P}, \chi_s^r; \mathbb{F}_q)$ is identical to the slope < 1 part of $\mathrm{NP}_q(1 + C_1 T + C_2 T^2 + \cdots)$ (see [12]). The latter can be identified as the condition that $k_1 \le d_1$, $k_2 \le d_2 - 1$. Thus by part (i) the lower bound of $\mathrm{NP}_q(\overline{P}, \chi_s^r; \mathbb{F}_q)$ is precisely $\mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^r)$ defined in Section 3.1 and by (10).

(iii) We first prove (b)⇒ (a) by assuming $p \equiv 1 \bmod d_J$ and $K_{r,t} \equiv 0 \bmod d_J$ for $J = 1, 2$ and for all $0 \le t \le a-1$.

Without loss of generality, from now on we focus mostly on the part of the matrix M_t regarding the pole at ∞, namely, the case $J = 1$ and omit the analogous argument for the case $J = 2$. By a routine computation via classical Dwork estimation, we have (see Lemma 3.2 in [22])

$$\mathrm{ord}_p C_{1,1}^{m,i} \ge \frac{\left\lceil \frac{mp-i-K_{r,t}}{d_1} \right\rceil}{p-1}$$

and the equality holds if $mp - i - K_{r,t} \equiv 0 \bmod d_1$ and $mp - i - K_{r,t} \le d_1(p-1)$. Using our hypothesis, we get $m \equiv i \bmod d_1$. Hence the $k \times k$ principal submatrix W_t

of M_t has its unique minimal row p-adic valuation $\frac{m}{d_1}$ on m-th row on its diagonal. Using a similar argument as that in the proof of [22, Theorem 1.2] and Lemma 2.1, we conclude that the Newton polygon coincides with the Hodge polygon.

Secondly we shall show (a)⇒(b) and suppose these two polygons coincide. By the Dwork estimation in the previous paragraph, we then have for every $1 \le k \le d_1 - 1$ that

$$\mathrm{ord}_p\mathrm{det}(W_t) \ge \min_{\rho_k \in S_k} \sum_{m=1}^{k} \frac{\left\lceil \frac{mp-\rho_k(m)-K_{r,t}}{d_1} \right\rceil}{p-1}.$$

By our hypothesis, we have the equality $\mathrm{ord}_p\mathrm{det}(W_t) = \frac{k(k+1)}{2d_1} - \frac{kK_{r,t}}{d_1(p-1)}$ holds, namely, the minimum p-adic valuation above is achieved. This implies that $\rho_k(m) \equiv mp - K_{r,t} \bmod d_1$ for every $1 \le m \le k$. One can show by induction on $k \ge 1$ that $\rho_k \in S_k$ can only be the identity permutation, and $K_{r,t} \equiv 0 \bmod d_1$ for every t. Set $m = 1$ in the above congruence, we find that $p \equiv 1 \bmod d_1$. In summary, we obtain $d | \gcd(p-1, K_{r,t})$ for every t. This proves (b). □

Remark 3.8. The main result in Proposition 3.7(ii) is known to [2, Corollary 3.18] as we noted in Remark 3.1. We gave a different proof here in line for the proof of our Proposition 4.2.

4. Asymptotic behavior of $L(\overline{P}/\mathbb{F}_q, \chi_s^r; T)$

Here again, we assume $s|(q-1)$. Recall that $\mathrm{NP}_q(\overline{P}, \chi_s^r; \mathbb{F}_q)$ denotes the q-adic Newton polygon of the L-function $L(\overline{P}/\mathbb{F}_q, \chi_s^r; T)$ of twisted exponential sums. In this section we shall show that for p large enough in a congruence class mod s, this Newton polygon generically converges to the corresponding twisted Hodge-Stickelberger polygon. (See Proposition 4.2 for precise statement.) Below we briefly outline our approach, which is very similar to that in [12, Sections 4,5] and hence we do not elaborate.

We fix some integer k with $1 \le k \le d_1 + d_2$ in the following, and we write $k = k_1 + k_2$ as in Section 3.3. Let $M_{t,1}^{[k_1]}$ (*resp.* $M_{t,2}^{[k_2]}$) denote the $k_1 \times k_1$ (*resp.* $k_2 \times k_2$) submatrix of M_t defined by

$$M_{t,1}^{[k_1]} = \left((C_{1,1}^{m,i})_{1 \le m,i \le k_1}\right) \ (resp.\ M_{t,2}^{[k_2]} = \left((C_{2,2}^{m,i})_{0 \le m,i \le k_2 - 1}\right)).$$

In this paper we should sometimes consider the coefficients a_i of $P(x)$ as variables and denote them by the vector $\vec{a}$. The *weight* of a monomial $\prod_{i=-d_2}^{d_1} a_i^{n_i}$ in $K[\vec{a}]$ is equal to $\sum_{i=-d_2}^{d_1} |i| n_i$. A relevant example in this section is that the minimal weight monomials in $H^{m,i}$ (defined under Lemma 3.4) are of weight $|m - i|$; and hence the minimal weight monomials in $B_t^{m,i}$ are of weight $|mp - K_{r,t} - i|$. All minimal weight monomials in (the formal expansion of) $\mathrm{det}M_t(\vec{a})$ lie in $\gamma^{\mathbb{Q}}\mathbb{Q}[\vec{a}]$. As shown in [12, Proposition 3.8], one can find a monomial in $\mathrm{det}M_{t,1}^{[k_1]}$ (*resp.* $\mathrm{det}M_{t,2}^{[k_2]}$) of minimal weight that does not cancel out with others terms. Moreover, the p-adic order $s_{1,t}$ (*resp.* $s_{2,t}$) of the coefficient of this monomial is minimal among the p-adic orders of all monomials in $\mathrm{det}M_{t,1}^{[k_1]}$ (*resp.* $\mathrm{det}M_{t,2}^{[k_2]}$). This monomial corresponds to a permutation $\rho_{1,t}$ (*resp.* $\rho_{2,t}$) in the permutation group S_{k_1} (*resp.* S_{k_2}). For $J = 1, 2$,

let $r_{J,i,j}$ be the least nonnegative residue of $-(pi-j) \bmod d_J$. Then we have

$$\begin{aligned} s_{1,t} &= \frac{k_1(k_1+1)}{2d_1} - \frac{k_1 K_{r,t}}{d_1(p-1)} + \frac{1}{d_1(p-1)} \sum_{i=1}^{k_1} r_{1,i,\rho_{1,t}(i)+K_{r,t}}; \\ s_{2,t} &= \frac{k_2(k_2-1)}{2d_2} + \frac{k_2 K_{r,t}}{d_2(p-1)} + \frac{1}{d_2(p-1)} \sum_{i=0}^{k_2-1} r_{2,i,\rho_{2,t}(i)-K_{r,t}}. \end{aligned}$$

For each fixed k let

$$\begin{aligned} s_k &:= \frac{1}{a} \sum_{t=0}^{a-1} (s_{1,t} + s_{2,t}) \\ &= \frac{k_1(k_1-1)}{2d_1} + \frac{k_1(1-\lambda)}{d_1} + \frac{k_2(k_2-1)}{2d_2} + \frac{k_2\lambda}{d_2} + \epsilon_{k,p} \end{aligned} \tag{11}$$

where

$$\epsilon_{k,p} := \frac{1}{a(p-1)d_1} \sum_{t=0}^{a-1} \sum_{i=1}^{k_1} r_{1,i,\rho_{1,t}(i)+K_{r,t}} + \frac{1}{a(p-1)d_2} \sum_{t=0}^{a-1} \sum_{i=0}^{k_2-1} r_{2,i,\rho_{2,t}(i)-K_{r,t}}.$$

Let $M_t^{[k]}$ be the $k \times k$ submatrix of M_t defined by the block matrix

$$M_t^{[k]} = \left(\begin{array}{c|c} (C_{1,1}^{m,i})_{1 \le m,i \le k_1} & (C_{1,2}^{m,i})_{1 \le m \le k_1,\ 0 \le i \le k_2-1} \\ \hline (C_{2,1}^{m,i})_{0 \le m \le k_2-1,\ 1 \le i \le k_1} & (C_{2,2}^{m,i})_{0 \le m,i \le k_2-1} \end{array} \right).$$

Then the terms of minimal valuation in the expansion of $\det M_t^{[k]}$ come from the product $\det M_{t,1}^{[k_1]} \cdot \det M_{t,2}^{[k_2]}$; they have p-adic order equal to $s_{1,t} + s_{2,t}$. The product of the terms of minimal p-adic order in each of the $\det M_t^{[k]}$ gives precisely the lowest γ-power term in $\prod_{t=0}^{a-1} \det M_t^{[k]}$. This term can be written as a product $\gamma^{a(p-1)s_k} U G_{\nu,r}$ for some p -adic unit U and some $G_{\nu,r} \in \mathbb{Q}[\vec{a}]$ which becomes independent of p when p is large enough. Note that $G_{\nu,r}$ is nonconstant since it contains a unique monomial corresponding to the permutation ρ_t in S_k obtained by composing $\rho_{1,t} \in S_{k_1}$ and $\rho_{2,t} \in S_{k_2}$ in the obvious way. Let $\mathcal{U}_{\nu,r}$ be the subspace of $\mathscr{A}_{d_1,d_2}$ defined by $G_{\nu,r} \neq 0$. Hence $\mathcal{U}_{\nu,r}$ over $\mathbb{Q}$ is open dense in $\mathscr{A}_{d_1,d_2}$.

Lemma 4.1. *Write* $\det(1-\alpha T) = \sum_{j=0}^{\infty} C_j T^j$. *Fix* $1 \le k \le d_1 + d_2$. *For* p *large enough, we have*

$$C_k \equiv \prod_{t=0}^{a-1} \det M_t^{[k]} \bmod \gamma^{>a(p-1)s_k}; \tag{12}$$

furthermore (with p *still large enough), we have*

$$\operatorname{ord}_q(C_k) = \frac{1}{a} \sum_{t=0}^{a-1} \operatorname{ord}_p \det M_t^{[k]} = s_k \tag{13}$$

if and only if $\overline{P}(x) \in \mathcal{U}_{\nu,r}(\mathbb{F}_q)$.

Proof. It is clear that the $k \times k$ submatrix of M_t whose determinant has the smallest p-adic valuation shares the rows of $M_t^{[k]}$. Let N be any $ak \times ak$ principal submatrix in $\vec{M}_{[a]}$. Let N_t be the intersection of N and M_t as submatrix of $\vec{M}_{[a]}$ for all $0 \leq t \leq a-1$. We may well assume that N_t is $k \times k$ matrix as in the proof of Lemma 2.2. Now suppose for some t we have $N_t \neq M_t^{[k]}$ share the rows of $M_t^{[k]}$. Observe that row indices of N_t are equal to the column indices or N_{t+1} because N is principal. Note that in fact we consider the subindices modulo a. Since N_t has at least one column outside of the columns of $M_t^{[k]}$, we have that N_{t-1} has at least one row outside of the rows of $M_{t-1}^{[k]}$. Recall that the difference between minimal row valuations in M_t is $\geq 1/d_1$ (resp. $\geq 1/d_2$) as p is large enough, depending on the location of the row in the matrix blocks. In comparison, the difference between minimal column valuations in M_t is convergent to 0 as p approaches ∞. As $p \to \infty$, we have by the same argument as that in [12, Sections 4,5], $\mathrm{ord}_p(\det N) > as_k$. As C_k is the infinite sum of $\pm \det N$ as N ranges over all such $ak \times ak$ principal submatrices in $\vec{M}_{[a]}$, the above inequality yields our first congruence relation in (12).

Note that $\mathrm{ord}_p \det M_t^{[k]} \geq as_k$, where equality holds if and only if $\overline{P} \in \mathcal{U}_{\nu,r}(\mathbb{F}_q)$ by the paragraph above this lemma. Combined with the congruence relation in (12), our second assertion in (13) follows. □

Let $\mathrm{GNP}(\mathscr{A}_{d_1,d_2}, \chi_s^r; \overline{\mathbb{F}}_p)$ be the generic Newton polygon of twisted exponential sums over $\overline{\mathbb{F}}_p$, namely,

$$\mathrm{GNP}(\mathscr{A}_{d_1,d_2}, \chi_s^r; \overline{\mathbb{F}}_p) \quad = \quad \sup_{\overline{P}} \ \mathrm{NP}_q(\overline{P}(x), \chi_s^r; \mathbb{F}_q) \tag{14}$$

where $\overline{P}$ ranges over all Laurent polynomials in $\mathscr{A}_{d_1,d_2}(\mathbb{F}_q)$ for all $\mathbb{F}_q$ in $\overline{\mathbb{F}}_p$. This maximum exists by Grothendieck specialization theorem (see [9] or [17]).

To simplify notations, we abbreviate $\lim_{p \to \infty, p \equiv \nu \bmod s}(\cdot)$ by $\lim_{\nu}(\cdot)$.

Proposition 4.2. *Let notations be as above. Fix $s \geq 1$ and $1 \leq \nu \leq s-1$ coprime to s. Let $0 \leq r \leq s-1$.*

(a) For $p \equiv \nu \bmod s$ large enough (depending only on d_1, d_2, ν, χ_s^r) we have

(i) *$\mathrm{GNP}(\mathscr{A}_{d_1,d_2}, \chi_s^r; \overline{\mathbb{F}}_p)$ exists and it is given by the vertex points $(k, s_k)_{0 \leq k \leq d_1+d_2}$.*

(ii) *we have*

$$\mathrm{NP}(P(x) \bmod \mathcal{P}, \chi_s^r) \preceq \mathrm{GNP}(\mathscr{A}_{d_1,d_2}, \chi_s^r; \overline{\mathbb{F}}_p)$$

for any prime $\mathcal{P}$ over p in $\overline{\mathbb{Q}}$; these two polygons coincide if and only if $P \in \mathcal{U}_{\nu,r}(\overline{\mathbb{Q}})$.

(b) For every $P(x) \in \mathcal{U}_{\nu,r}(\overline{\mathbb{Q}})$ we have that

$$\lim_{\nu} \mathrm{NP}(P \bmod \mathcal{P}, \chi_s^r) = \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^r)$$

for any prime $\mathcal{P}$ over p in $\overline{\mathbb{Q}}$.

Proof. (a) Consider the previous lemma 4.1 and suppose p large enough as given there. We have $\mathrm{ord}_q C_k(\vec{a}) \geq s_k$ and the equality holds if and only if $P \in \mathcal{U}_{\nu,r}(\overline{\mathbb{Q}})$. On the other hand, for p large enough, $\mathrm{NP}_q(\overline{P}, \chi_s^r; \mathbb{F}_q)$ coincides with the q-adic Newton polygon of $\sum_{j=0}^{d_1+d_2-1} C_j T^j = \det(1 - \alpha T) \bmod T^{d_1+d_2}$. Thus $\mathrm{GNP}(\mathscr{A}_{d_1,d_2}, \chi_s^r; \overline{\mathbb{F}}_p)$ is

indeed given by vertices with coordinates (k, s_k) for $0 \leq k \leq d_1 + d_2$. This proves (i). Moreover $\mathrm{NP}(P \bmod \mathcal{P}, \chi_s^r) = \mathrm{NP}_q(\overline{P}, \chi_s^r; \mathbb{F}_q) \preceq \mathrm{GNP}(\mathscr{A}_{d_1,d_2}, \chi_s^r; \overline{\mathbb{F}}_p)$ and they coincide if and only if $P \in \mathcal{U}_{\nu,r}(\overline{\mathbb{Q}})$. This proves (ii).

(b) Notice that $\varepsilon_{k,p} \to 0+$ as $p \to \infty$, so for $p \to \infty$, we have

$$s_k \quad \longrightarrow \quad \frac{k_1(k_1-1)}{2d_1} + \frac{k_1(1-\lambda)}{d_1} + \frac{k_2(k_2-1)}{2d_2} + \frac{k_2\lambda}{d_2}$$

from the right. From part (a) we know that for $P \in \mathcal{U}_{\nu,r}(\overline{\mathbb{Q}})$ the Newton polygon coincides with the generic Newton polygon, but the latter converges to the Hodge-Stickelberger polygon as p approaches infinity by looking at the limit of s_k. This proves (b). □

5. Newton polygons for Laurent polynomials $\overline{P}(x^s)$.

We shall prove the main theorems in this section. We do not suppose any more that $q \equiv 1 \bmod s$. Let χ_s be a multiplicative character of order s on $\overline{\mathbb{F}}_q^\times$. Let $k \geq 1$ be an integer, and $n := \gcd(s, q^k - 1)$. Let $\chi_n := \chi_s^{s/n}$ be a multiplicative character of $\mathbb{F}_{q^k}^\times$ of order n. (We remark here that another approach to the main theorems is to extend the base field $\mathbb{F}_q$ to make it large enough so that $s|(q-1)$. One can do so since our Hodge-Stickelberger polygon does not depend on the choice of base field even though it does depend on $p \bmod s$).

Lemma 5.1. *With the above notations, we have*

$$S_k(\overline{P}(x^s)) = \sum_{r=0}^{n-1} S_k(\overline{P}, \chi_n^r) = \sum_{r' \in \{0,\ldots,s-1\}, \frac{s}{n} | r'} S_k(\overline{P}, \chi_s^{r'}).$$

Proof. By hypothesis, we may factor s as a product of two integers $s = mn$. Since $\gcd(m, q^k - 1) = 1$, the map $x \mapsto x^m$ is bijective on $\mathbb{F}_{q^k}^\times$. On the other hand, since $n|q^k - 1$, the kernel of the map $x \mapsto x^n$ is the set of n-th roots of unity, and its image is the set $(\mathbb{F}_{q^k}^\times)^n$ of n-th powers in $\mathbb{F}_{q^k}^\times$. Thus we get

$$S_k(\overline{P}(x^s)) = \sum_{x \in \mathbb{F}_{q^k}^\times} \psi_{q^k}(\overline{P}(x^s)) = \sum_{x \in (\mathbb{F}_{q^k}^\times)^n} n\psi_{q^k}(\overline{P}(x)).$$

From the orthogonality relations on multiplicative characters, we have that $\sum_{r=0}^{n-1} \chi_n(x^r) = n$ if $x \in (\mathbb{F}_{q^k}^\times)^n$ and $= 0$ otherwise. Then the above equation becomes

$$\begin{aligned} S_k(\overline{P}(x^s)) &= \textstyle\sum_{x \in \mathbb{F}_{q^k}^\times} \sum_{r=0}^{n-1} \chi_n(x^r) \psi_{q^k}(\overline{P}(x)) \\ &= \textstyle\sum_{r=0}^{n-1} \sum_{x \in \mathbb{F}_{q^k}^\times} \chi_n(x^r) \psi_{q^k}(\overline{P}(x)) \\ &= \textstyle\sum_{r=0}^{n-1} S_k(\overline{P}, \chi_n^r). \end{aligned}$$

The last equation is straightforward. □

Observe that if $s|(q^k - 1)$, then we have $S_k(\overline{P}(x^s)) = \sum_{r=0}^{s-1} S_k(\overline{P}, \chi_s^r)$.

Consider the permutation σ^a on $\{0, \ldots, s-1\}$, namely the permutation induced by multiplication of $q = p^a$ modulo s. Its cycle decomposition (including 1-cycles) is further splitting of that of σ as $\sigma^a = \prod_{i=1}^u \sigma_i^a = \prod_{i=1}^u \prod_{j=1}^{\ell_i/\ell_i'} \sigma_{ij}$ for ℓ_i' -cycles σ_{ij},

where $\ell_i'|\ell_i$. Namely each permutation σ_i^a splits into ℓ_i/ℓ_i' many cycles of equal length ℓ_i'.

Consider $\overline{P} \in \mathscr{A}_{d_1,d_2}(\mathbb{F}_q)$ as $\overline{P}/\mathbb{F}_{q^{\ell_i'}}$ for $i = 1, \ldots, u$. It is clear that $s|(q^{\ell_i'} - 1)$.

For any cycle σ_i in the decomposition of σ (including 1-cycles), define

$$L_i(T) := \prod_{j=1}^{\ell_i/\ell_i'} L(\overline{P}/\mathbb{F}_{q^{\ell_i'}}, \chi_s^{r_{ij}}; T^{\ell_i'}) \tag{15}$$

where r_{ij} is an element in σ_{ij}. This is a polynomial in $\mathbb{Z}[\zeta_p, \zeta_s][T]$ of degree $\ell_i(d_1+d_2)$.

Lemma 5.2. *We have*

$$L(\overline{P}(x^s); T) = \prod_{i=1}^{u} L_i(T). \tag{16}$$

Proof. Since $x \mapsto x^q$ is an automorphism of $\mathbb{F}_{q^n}^\times$ for any n and $\overline{P}(x^q) = \overline{P}(x)^q$, we have

$$\begin{aligned} S_k(\overline{P}(x), \chi_n^r) &= S_k(\overline{P}(x^q), \chi_n^r) \\ &= \sum_{x \in \mathbb{F}_{q^k}^\times} \chi_n^r(x^q) \psi_{q^k}(\overline{P}(x^q)) \\ &= \sum_{x \in \mathbb{F}_{q^k}^\times} \chi_n^r(x^q) \psi_{q^k}(\overline{P}(x)). \end{aligned}$$

This shows that $S_k(\overline{P}, \chi_n^r) = S_k(\overline{P}, \chi_n^{qr})$. Consequently the sum in Lemma 5.1 may be broken down into orbits of σ^a. Recall $\sigma^a = \prod_{i=1}^u \prod_{j=1}^{\ell_i/\ell_i'} \sigma_{ij}$. Let r_{ij} be an element in σ_{ij}. Since $\ell_i'|k$ is the same as saying $\sigma^{ak}(r_{ij}) = r_{ij}$, that is $q^k r_{ij} \equiv r_{ij} \bmod s$. But $s|(q^k - 1)r_{ij}$ (combined with our hypothesis $n = \gcd(s, q^k - 1)$) is equivalent to $\frac{s}{n}|r_{ij}$. Thus the sum in Lemma 5.1 can be phrased as

$$S_k(\overline{P}(x^s)) = \sum_{r_{ij}, \frac{s}{n}|r_{ij}} \ell_i' \, S_k(\overline{P}, \chi_s^{r_{ij}}) = \sum_{r_{ij}, \ell_i'|k} \ell_i' \, S_k(\overline{P}, \chi_s^{r_{ij}})$$

where r_{ij} runs in all distinct cycles σ_{ij} in σ^a. Substitute this identity to twisted L-function defined in Section 3.1, we get after some elementary computation

$$L(\overline{P}(x^s); T) = \prod_{r_{ij}} L(\overline{P}/\mathbb{F}_{q^{\ell_i'}}, \chi_s^{r_{ij}}; T^{\ell_i'})$$

where the product ranges over all distinct cycles in σ^a. Group this product in terms of cycle decomposition of σ, we finish our proof. □

Proof of Theorem 1.2. By Proposition 3.7 (ii) we have

$$\begin{aligned} \mathrm{NP}_q(L(\overline{P}/\mathbb{F}_{q^{\ell_i'}}, \chi_s^{r_{ij}}; T^{\ell_i'})) &= \ell_i' \, \mathrm{NP}_{q^{\ell_i'}}(L(\overline{P}/\mathbb{F}_{q^{\ell_i'}}, \chi_s^{r_{ij}}; T)) \\ &\preceq \ell_i' \, \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^{r_{ij}}) \\ &= \ell_i' \, \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^{r_i}) \end{aligned}$$

for all $1 \le j \le \ell_i/\ell_i'$. Thus

$$\mathrm{NP}_q(L_i(T)) \preceq \frac{\ell_i}{\ell_i'}(\ell_i' \, \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^{r_i})) = \ell_i \, \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, \chi_s^{r_i}).$$

By the split of the L-function in Lemma 5.2 and by Lemma 3.2, we have

$$\begin{aligned}\mathrm{NP}_q(\overline{P}(x^s);\mathbb{F}_q) &= \boxplus_{i=1}^{u}\mathrm{NP}_q(L_i(T))\\ &\preceq \boxplus_{i=1}^{u}\ell_i\ \mathrm{HS}(\mathscr{A}_{d_1,d_2},\nu,\chi_s^{r_i}) = \mathrm{HS}(\mathscr{A}_{d_1,d_2},\nu,s),\end{aligned}$$

where the box-sum ranges over all cycle in the decomposition of $\sigma=\prod_{i=1}^{u}\sigma_i$. This proves the first statement of Theorem 1.2.

For the last statement we need the following lemma:

Lemma 5.3. *Write $(q-1)r/s=\sum K_{r,t}p^t$ for $0\le K_{r,t}\le p-1$ and write $d=\mathrm{lcm}(d_1,d_2)$. Assume $d\ge 2$. Then the following statements are equivalent:*

(i) *$d|\gcd(p-1,K_{r,t})$ for every $0\le t\le a-1$ and $1\le r\le s-1$.*
(ii) *$p\equiv 1 \bmod sd$.*

Proof. We shall show (i)$\Rightarrow$ (ii). Write $q=p^a$ such that $s|q-1$. We claim that $s|p-1$. Write $u_r:=(\frac{(q-1)r}{s} \bmod p)$ so that $0\le u_r\le p-1$. It is easy to see that $u_r\equiv ru_1 \bmod p$ for all r. By induction on $r\ge 1$ we have $u_r=ru_1$ for all $1\le r\le s-1$. Thus we have $u_1=u_{s-1}/(s-1)\le\frac{p-1}{s-1}$. On the other hand, we have $u_s=p-1\equiv su_1 \bmod p$. By the above bound on u_1, we are forced to have $u_1=\frac{p-1}{s}$ and therefore $s|p-1$. By examining the p-adic expansion $(q-1)/s=\sum_{i=0}^{a-1}(\frac{p-1}{s})p^i$ we have $d|\frac{p-1}{s}$ by hypothesis in (i). That is $p\equiv 1 \bmod sd$. This proves (ii). The other direction (ii) $\Rightarrow$ (i) is very easy and we hence omit its proof here. □

By the description of $\mathrm{HS}(\mathscr{A}_{d_1,d_2},\nu,s)$ in Lemma 3.2 and Lemma 5.2 above, the Newton polygon coincides with the Hodge-Stickleberger polygon exactly when the Newton polygons of twisted exponential sums coincide with their corresponding twisted Hodge-Stickleberger polygon. Hence the last statement in the theorem follows immediately from Proposition 3.7(iii) and Lemma 5.3. □

Let $\mathrm{GNP}(\mathscr{A}_{d_1,d_2},s;\overline{\mathbb{F}}_p)$ be the generic Newton polygon for exponential sums of $\overline{P}(x^s)/\overline{\mathbb{F}}_p$. That is,

$$\mathrm{GNP}(\mathscr{A}_{d_1,d_2},s;\overline{\mathbb{F}}_p):=\sup_{\overline{P}}\mathrm{NP}_q(\overline{P}(x^s);\mathbb{F}_q)=\sup_{P}\mathrm{NP}(P(x^s) \bmod \mathcal{P})$$

where $\overline{P}$ ranges in $\mathscr{A}_{d_1,d_2}(\mathbb{F}_q)$ for any q, and where P ranges in $\mathscr{A}_{d_1,d_2}(\overline{\mathbb{Z}}_p\cap\overline{\mathbb{Q}})$ and $\mathcal{P}$ is any prime over p in $\overline{\mathbb{Q}}$.

Recall σ is the permutation on $\{0,1,\dots,s-1\}$ induced by multiplication by p modulo s. For every cycle σ_i in σ we have a nonconstant polynomial $G_{\nu,r}$ (see Proposition 4.2) where r is an element in σ_i (it is independent of the choice of r in σ_i.) Let $G_\nu=\prod G_{\nu,r}$ where r runs in distinct cycles $\sigma_1,\dots,\sigma_u$ of σ, then G_ν is polynomial in $\mathbb{Q}[\vec{a}]$ as well. Let $\mathcal{U}_\nu$ be the complement of $G_\nu=0$ in $\mathscr{A}_{d_1,d_2}$. Then $\mathcal{U}_\nu$ is a Zariski dense open subset of $\mathscr{A}_{d_1,d_2}$ defined over $\mathbb{Q}$. Our Theorems 1.2 and 1.4 are proved in the following stronger version. Its proof is similar to that of Theorem 1.2.

Theorem 5.4. *Let notations be as in Theorem 1.4. Then*
(a) For $p\equiv\nu \bmod s$ large enough (depending only on d_1,d_2,ν,s), we have

(i) *$\mathrm{GNP}(\mathscr{A}_{d_1,d_2},s;\overline{\mathbb{F}}_p)$ exists and $\mathrm{NP}(P(x^s) \bmod \mathcal{P})\preceq\mathrm{GNP}(\mathscr{A}_{d_1,d_2},s;\overline{\mathbb{F}}_p)$ for all $P\in\mathscr{A}_{d_1,d_2}(\overline{\mathbb{Q}})$;*

(ii) *these two polygons coincide if and only if* $P \in \mathcal{U}_\nu(\overline{\mathbb{Q}})$.

(b) For $P \in \mathcal{U}_\nu(\overline{\mathbb{Q}})$ *we have*

$$\lim_{\nu} \mathrm{NP}_q(P(x^s) \bmod \mathcal{P}; \mathbb{F}_q) = \mathrm{HS}(\mathscr{A}_{d_1,d_2}, \nu, s).$$

Proof. Our theorem follows immediately by applying Proposition 4.2 and the key Lemma 5.2 in the same fashion as that in the proof of Theorem 1.2. We hence omit details here. □

Finally we remark in the polynomial case, i.e., $d_2 = 0$, similar argument can be carried out which yields similar results. In fact, one can carry out calculations in the spirit of [4] to get explicitly the generic Newton polygons, and *Hasse polynomials* that describe exactly which polynomials attain this polygon.

6. Further questions and multivariable cases

6.1. Global permutation polynomials. Wan's [17, Conjecture 1.12] was proved in 1-variable case by [20, 21] and was generalized to Laurent polynomials in [12] that there is a Zariski dense open subset $\mathcal{U}$ in $\mathscr{A}_{d_1,d_2}$ defined over $\mathbb{Q}$ such that for every $f \in \mathcal{U}(\overline{\mathbb{Q}})$ we have its limit of Newton polygon approaching the Hodge polygon as $p \to \infty$. It has been fascinating researchers to know what (Laurent) polynomials f over $\overline{\mathbb{Q}}$ that would fail the asymptotic property $\lim_{p\to\infty} \mathrm{NP}(f \bmod \mathcal{P}) = \mathrm{HP}(\mathscr{A}_{d_1,d_2})$. We will discuss below some known such (Laurent) polynomials. For simplicity we restrict ourselves over $\mathbb{Q}$ instead of extension of $\mathbb{Q}$ in $\overline{\mathbb{Q}}$, one can extend our argument to extensions of $\mathbb{Q}$ by the references we shall provide in the context.

For any positive integer n, let $D_n(x, y)$ be the unique polynomial in $\mathbb{Z}[x, y]$ such that $D_n(u+v, uv) = u^n+v^n$. For any $c \in \mathbb{Q}$ the monic degree-n polynomial $D_n(x, c)$ in $\mathbb{Q}[x]$ is called a degree-n *Dickson polynomial* over $\mathbb{Q}$. If p divides c, then $D_n(x, c) = x^n$ is a monomial which is a permutation on $\mathbb{F}_p$ if and only if $\gcd(n, p-1) = 1$. If p does not divide c, it is a permutation on $\mathbb{F}_p$ if and only if $\gcd(n, p^2-1) = 1$ (due to [6], see [13, Chapter 7] for quick reference).

For any $l \geq 1$, let *global permutation polynomial over* $\mathbb{Q}$ *of level* l be a polynomial $h(x)$ in $\mathbb{Q}[x]$ such that $x \mapsto h(x)$ is a permutation on $\mathbb{F}_p, \ldots, \mathbb{F}_{p^l}$ for infinitely many primes p. It is easy to see that $D_n(x, c)$ in $\mathbb{Q}[x]$ is a global permutation polynomial of level l if and only if every prime factor Q of n satisfies $Q > l+1$ (when $c = 0$) and $Q > 2l+1$ (when $c \neq 0$). Thus for level $l = 1$ it is equivalent to $2 \nmid n$ (when $c = 0$) and $\gcd(n, 6) = 1$ (when $c \neq 0$).

It is known that every global permutation polynomial over $\mathbb{Q}$ is a composition of Dickson polynomials $D_n(x, c)$ over $\mathbb{Q}$ and linear polynomials over certain extensions of $\mathbb{Q}$. (This is proved for all number fields by Fried in [8].)

Our result in Theorem 1.4 implies that for any polynomial or Laurent polynomial $f(x)$ over $\overline{\mathbb{Q}}$ containing $x^s = D_s(x, 0)$ as a right composition factor for any $s > 2$, that is, $f(x) = P(x^s)$, the limit of p-adic Newton polygon does not exist as $p \to \infty$. Following Wan's argument on polynomials which is communicated to the authors, we demonstrate here that if $f(x)$ is any Laurent polynomial in $\mathscr{A}_{d_1,d_2}(\overline{\mathbb{Q}})$ containing a global permutation polynomial of degree $s > 1$ of level 3 as a right composition factor, that is $f(x) = P(D_s(x, c))$, then the limit $\lim_{p\to\infty} \mathrm{NP}(f(x) \bmod p)$ does not exist. Without loss of generality, we assume $d_1 \geq d_2$ for the rest of this paragraph.

Since s must be odd, our Dickson polynomial fixes 0 and ∞, and finally we assume the global permutation polynomial composition factor is $D_s(x,c)$ for some $s>1$ where s's prime factors are all ≥ 7 (when $c=0$) and ≥ 11 (when $c\neq 0$). Write $L(f(x);\mathbb{F}_p)=1+C_1T+C_2T^2+\cdots$ and $L(P(x);\mathbb{F}_p)=1+c_1T+\cdots$. For any prime p such that $D_s(x,c)$ permutes $\mathbb{F}_p,\mathbb{F}_{p^2},\mathbb{F}_{p^3}$, we have $S_k(f;\mathbb{F}_p)=S_k(P;\mathbb{F}_p)$ for $1\leq k\leq 3$. By the lower bound for Newton polygon of $L(P(x);\mathbb{F}_p)$ (see [21]) we have $\operatorname{ord}_p C_2=\operatorname{ord}_p c_2\geq 1/(d_1/s)=s/d_1\geq 7/d_1$. This implies the Newton polygon of $L(f(x);\mathbb{F}_p)$ does not have a breakpoint at $(2,1/d_1)$; similarly, since $\operatorname{ord}_p c_3\geq 2s/d_1\geq 14/d_1$ a breakpoint at $(3,3/d_1)$ is impossible. On the other hand, we know that for infinitely many prime p (precisely those $p\equiv 1 \bmod \operatorname{lcm}(sd_1,sd_2)$) $\operatorname{NP}(f \bmod p)$ coincides with its lower bound and has break point at $(2,1/d_1)$ if $d_1>d_2$ and at $(3,3/d_1)$ if $d_1=d_2$. Thus $\lim_{p\to\infty}\operatorname{NP}(f \bmod p)$ does not exist.

We say two (Laurent) polynomials $f(x)$ and $h(x)$ over $\mathbb{Q}$ of degree d are *Artin-Schreier isomorphic* if $f(x)=h(wx+v)$ for some d-th root of unity w and $v\in\overline{\mathbb{Q}}$ (so that the two Artin-Schreier curves defined by $y^p-y=f(x)$ and $y^p-y=h(x)$ are isomorphic over $\overline{\mathbb{F}}_p$). For reader's convenience, we quote a corrected version of Wan's conjecture below from [19, Chapter 5].

Conjecture 6.1 (Wan)**.** *If $f(x)$ is a polynomial in $\mathbb{Q}[x]$ which does not contain a global permutation polynomial of degree >1 as right composition factor over $\mathbb{Q}$ (upto Artin-Schreier isomorphism), then $\lim_{p\to\infty}\operatorname{NP}_p(f \bmod p)$ exists and is equal to its lower bound Hodge polygon.*

6.2. A variant of Schur's theorem. Let $\psi:\mathbb{F}_p\to\mathbb{Q}(\zeta_p)^\times$ be the nontrivial additive character defined by $\psi(a)=\zeta_p^a$.

Conjecture 6.2. *Let $f(x)\in\mathbb{Q}[x]$ be of degree $d\geq 2$ and let $S(f(x) \bmod p)=\sum_{x\in\mathbb{F}_p}\psi(f(x))$ be the first exponential sum mod p. Let $\varepsilon>0$. If $\operatorname{ord}_p S(f(x) \bmod p)>1/d+\varepsilon$ for infinitely many primes p, then $f(x)=P(D_s(x,c))$ (up to Artin-Schreier isomorphism) for some $P\in\mathbb{Q}[x]$ and a global permutation Dickson polynomial D_s of degree $s>1$.*

The conjecture above can be considered as a generalization of the Schur's conjecture on global permutation polynomials since it can be phrased in the following term: "For any $f\in\mathbb{Q}[x]$, if $S(f(x) \bmod p)=0$ (i.e. $\operatorname{ord}_p(S(f(x) \bmod p))=+\infty$) for infinitely many prime p then $f(x)$ is a Dickson polynomial up to Artin-Schreier isomorphism" (see [13, Chapter 7]).

Proposition 6.3. *Let notation be as above and suppose Conjecture 6.2 holds. Then the limit $\lim_{p\to\infty}\operatorname{NP}_1(f(x) \bmod p)$ of first slope exists if and only if $f(x)\neq P(D_s(x,c))$ (up to any Artin-Schreier isomorphism) for some $P\in\mathbb{Q}[x]$ and a global permutation Dickson polynomial $D_s(x)$ of degree $s>1$.*

Proof. It was already proved above that if $f(x)$ contains a right Dickson composition factor (of degree prime to 2 or 6 depending on whether $c=0$ or not) then the limit does not exist. Conversely, suppose the first slope limit does not exist. Since for $p\equiv 1 \bmod d$ we always have $\operatorname{NP}_1(f \bmod p)=1/d$ (and $\operatorname{NP}_2(f \bmod p)=2/d$), this is equivalent to the hypothesis of Conjecture 6.2 since $\operatorname{NP}_1(f \bmod p)=\operatorname{ord}_p S_1(f \bmod p)$. □

6.3. Multivariable cases. Our main result in Theorem 1.2 generalizes to multivariable cases. Let $\mathscr{A}_\Delta$ be the space of polynomials in n variables $x_1, \ldots, x_n$ with Newton polyhedron $\Delta \subset \mathbb{R}^n$, non degenerate with respect to Δ, parametrized by their coefficients of monomials. Let $\overline{P}$ be a polynomial in $\mathscr{A}_\Delta(\mathbb{F}_q)$. Fix $\vec{s} = (s_1, \ldots, s_n)$ for integers $s_\iota \geq 1$. All primes p in this subsection will be coprime to $s_1 \cdots s_n$. Let $\vec{\nu} = p \bmod \vec{s}$, the least nonnegative residue. For each $1 \leq \iota \leq n$, let σ_ι be the permutation on the set $\{0, 1, \ldots, s_\iota - 1\}$ induced by multiplication of p. We write its cycle decomposition as

$$\sigma_\iota = \prod_{i_\iota=1}^{u_\iota} \sigma_{\iota,i}$$

for $\ell_{\iota,i}$-cycles $\sigma_{\iota,i}$ (including 1-cycles!). For each $1 \leq \iota \leq n$ and $1 \leq i_\iota \leq u_\iota$, let

$$\lambda_{\iota,i_\iota} := \frac{\sum_{j \in \sigma_{\iota,i_\iota}} j}{s_\iota \ell_{\iota,i_\iota}}.$$

So $0 \leq \lambda_{\iota,i_\iota} < 1$. Write $\vec{\lambda}_{\vec{i}} := (\lambda_{1,i_1}, \ldots, \lambda_{n,i_n})$. Let $w_\Delta : \mathbb{Z}^n \to \mathbb{Z}$ be the weight function with respect to a given Δ as in [1] and [17]. It is easy to see that it extends to $\mathbb{Q}^n$.

We define the *Hodge-Stickelberger polygon* $\mathrm{HS}(\mathscr{A}_\Delta, \vec{\nu}, \vec{s})$ in multivariable setting as concatenation of line segments given by

$$(w_\Delta(\vec{m} - \vec{\lambda}_{\vec{i}}), \ell_{1,i_1} \cdots \ell_{n,i_n})$$

where $\vec{m}$ ranges over the $n!V(\Delta)$ elements in $C(\Delta) \cap \mathbb{Z}^n$ as defined in [1], $1 \leq \iota \leq n$ and $1 \leq i_\iota \leq u_\iota$. One observes that this polygon has horizontal length $s_1 \cdots s_n n! V(\Delta)$.

Theorem 6.4. *Suppose $\overline{P}(x_1, \ldots, x_n)$ over $\mathbb{F}_q$ is nondegenerate with respect to Δ and the dimension of the polyhedron Δ is equal to n, then $L(\overline{P}(x_1^{s_1}, \ldots, x_n^{s_n})/\mathbb{F}_q, T)^{(-1)^{n-1}}$ is a polynomial. Moreover its Newton polygon lies over $\mathrm{HS}(\mathscr{A}_\Delta, \vec{\nu}, \vec{s})$ and their endpoints meet.*

The proof of this theorem is parallel to the proof of Theorem 1.2 and will introduce lots more notations and we hence omit it here. We want to emphasize here that Theorem 6.4 does not include Theorem 1.2 as a corollary. It is slightly weaker in the one-variable special case.

Finally we remark that the asymptotic result in Theorem 1.4 seems harder to generalize. Nevertheless, from Theorem 6.4 one observes already that for each residue class $\vec{\nu} = (p \bmod \vec{s})$ there is a distinct lower bound $\mathrm{HS}(\mathscr{A}_\Delta, \vec{\nu}, \vec{s})$. So one can not expect there is a limit on the generic Newton polygon as $p \to \infty$.

Acknowledgements

The authors thank Daqing Wan for invaluable communication regarding his conjecture(s) and for pointing out the reference [19] to us. We also thank Michael Zieve for providing us with an interesting account of references and history on Dickson polynomials. Finally we thank the referee for extremely helpful comments.

References

[1] A. Adolphson and S. Sperber, *On twisted exponential sums*, Math. Ann. **290** (1991), 713-726.

[2] ———, *Twisted exponential sums and Newton polyhedra*, J. Reine Angew. Math. **443** (1993), 151-177.

[3] B.C. Berndt, R.J. Evans, and K.S. Williams, *Gauss and Jacobi sums*, Wiley-Interscience, New York, 1998.

[4] Régis Blache and Éric Férard, *Newton stratification for polynomials: the open stratum*, Jour. Number Th. **123** (2007), 456–472.

[5] J. Denef and F. Loeser, *Weights of exponential sums, intersection cohomology, and Newton polyhedra*, Invent. Math. **106** (1991), no. 2, 275–294.

[6] L. Dickson, *The analytic representation of substitutions on a power of a prime number with a discussion of the linear group*, Ann. of Math. **11** (1896/97), no. 1-6, 65–120, 161–183.

[7] B. Dwork, *On the zeta function of a hypersurface*, Inst. Hautes Études Sci. Publ. Math. **12** (1962), 5–68.

[8] M. Fried, *On a conjecture of Schur*, Michigan Math. J. **17** (1970), 41–55.

[9] N. M. Katz, *Slope filtration of F-crystals*, Asterisque, **63** (1979), 113–163.

[10] ———, *Crystalline cohomology, Dieudonné modules, and Jacobi sums*, Automorphic forms, representation theory and arithmetic, (Bombay, 1979), pp. 165–246, Tata Inst. Fund. Res. Studies in Math., **10**, Tata Inst. Fundamental Res., Bombay, 1981.

[11] N. Koblitz, *p-adic numbers, p-adic analysis, and zeta-functions, (Second edition)*, Graduate Texts in Mathematics, **58**. Springer-Verlag, New York, 1984.

[12] H. Li and H. J. Zhu, *Zeta functions of totally ramified p-covers of the projective line*, Rend. Sem. Mat. Univ. Padova, **113** (2005), 203–225.

[13] R. Lidl and H. Neiderreiter, *Finite fields*, Encyclopedia of Mathematics and its Applications, **20** Addison-Wesley Publishing Company, Reading, Massachusetts. 1983.

[14] R. Pries and H. J. Zhu, *The p-ranks stratification of Artin-Schreier curves*, Preprint at `math.NT/0609657`.

[15] J-P Serre, *Endomorphismes complètements continus des espaces de Banach p-adique*, Inst. Hautes. Études Sci. Publ. Math. **12** (1962), 69–85 (French).

[16] D. Wan, *Newton polygons of zeta functions and L functions*, Ann. of Math. **137** (1993), 249–293.

[17] ———, *Variation of p-adic Newton polygons for L-functions of exponential sums*, Asian J. Math. **8** (2004), no. 3, 427–472.

[18] L. Washington, *Introduction to cyclotomic fields*, Graduate Texts in Mathematics, **83**, Second edition. Springer-Verlag.

[19] R. Yang, *Newton polygons of L-functions of polynomials of the form* $x^d + \lambda x$, Finite Fields Appl. **9** (2003), no. 1, 59–88.

[20] H. J. Zhu, *p-adic variation of L functions of one variable exponential sums, I.* Amer. J. Math. **125** (2003).

[21] ———, *Asymptotic variation of L functions of one-variable exponential sums*, J. Reine Angew. Math. **572** (2004), 219–233.

[22] ———, *L-functions of exponential sums over one-dimensional affinoids: Newton over Hodge*, Inter. Math. Research Notices, (2004), no. 30, 1529–1550.

Laboratoire AOC, IUFM de la Guadeloupe, 97139 Les Abymes
E-mail address: `rblache@iufm.univ-ag.fr`

Laboratoire GAATI, Université de la Polynésie Française, Tahiti
E-mail address: `ferard@upf.pf`

Department of mathematics, State University of New York, Buffalo, NY 14260-2900, USA
E-mail address: `zhu@cal.berkeley.edu`