ASIAN J. MATH. © 2004 International Press
Vol. 8, No. 3, pp. 427–472, September 2004 004

VARIATION OF P-ADIC NEWTON POLYGONS FOR L-FUNCTIONS OF EXPONENTIAL SUMS *

DAQING WAN†

Abstract. In this paper, we continue to develop the systematic decomposition theory [18] for the generic Newton polygon attached to a family of zeta functions over finite fields and more generally a family of L-functions of n-dimensional exponential sums over finite fields. Our aim is to establish a new collapsing decomposition theorem (Theorem 3.7) for the generic Newton polygon. A number of applications to zeta functions and L-functions are given, including the full form of the remaining 3 and 4-dimensional cases of the Adolphson-Sperber conjecture [2], which were left un-resolved in [18]. To make the paper more readable and useful, we have included an expanded introductory section as well as detailed examples to illustrate how to use the main theorems.

1. Introduction. Let $\mathbf{F}_q$ be the finite field of q elements with characteristic p. For each positive integer k, let $\mathbf{F}_{q^k}$ be the finite extension of $\mathbf{F}_q$ of degree k. Let ζ_p be a fixed primitive p-th root of unity in the complex numbers. For any Laurent polynomial $f(x_1, \cdots, x_n) \in \mathbf{F}_q[x_1, x_1^{-1}, \cdots, x_n, x_n^{-1}]$, we form the exponential sum

$$\mathrm{S}_k^*(f) = \sum_{x_i \in \mathbf{F}_{q^k}^*} \zeta_p^{\mathrm{Tr}_k f(x_1, \cdots, x_n)},$$

where $\mathbf{F}_{q^k}^*$ denotes the set of non-zero elements in $\mathbf{F}_{q^k}$ and Tr_k denotes the trace map from $\mathbf{F}_{q^k}$ to the prime field $\mathbf{F}_p$. This is an exponential sum over the n-torus $\mathbf{G}_m^n$ over $\mathbf{F}_{q^k}$. A question of fundamental importance in number theory is to understand the sequence $S_k^*(f)$ $(1 \leq k < \infty)$ of algebraic integers, each of them lying in the p-th cyclotomic field $\mathbf{Q}(\zeta_p)$.

By a theorem of Dwork-Bombieri-Grothendieck, the following generating L-function is a rational function:

$$\mathrm{L}^*(f, T) = \exp\Big(\sum_{k=1}^{\infty} \mathrm{S}_k^*(f) \frac{T^k}{k}\Big) = \frac{\prod_{i=1}^{d_1}(1 - \alpha_i T)}{\prod_{j=1}^{d_2}(1 - \beta_j T)}, \tag{1}$$

where the finitely many numbers α_i $(1 \leq i \leq d_1)$ and β_j $(1 \leq j \leq d_2)$ are non-zero algebraic integers. Equivalently, for each positive integer k, we have the formula

$$\mathrm{S}_k^*(f) = \beta_1^k + \beta_2^k + \cdots + \beta_{d_2}^k - \alpha_1^k - \alpha_2^k - \cdots - \alpha_{d_1}^k.$$

Thus, our fundamental question about the sums $S_k^*(f)$ is reduced to understanding the reciprocal zeros α_i $(1 \leq i \leq d_1)$ and the reciprocal poles β_j $(1 \leq j \leq d_2)$. When we need to indicate the dependence of the L-function on the ground field $\mathbf{F}_q$, we will write $\mathrm{L}^*(f/\mathbf{F}_q, T)$.

Without any smoothness condition on f, one does not even know exactly the number d_1 of zeros and the number d_2 of poles, although good upper bounds are available, see [4]. On the other hand, Deligne's theorem on the Riemann hypothesis [5] gives the following general information about the nature of the zeros and poles. For the complex absolute value $|\ |$, this says

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

*Received October 10, 2003; accepted for publication January 27, 2004.

†Institute of Mathematics, Chinese Academy of Sciences, Beijing, P.R. China; Department of Mathematics, University of California, Irvine, CA 92697-3875, USA (dwan@math.uci.edu). Partially supported by NSF and the NNSF of China (10128103).

where $\mathbf{Z} \cap [0, 2n]$ denotes the set of integers in the interval $[0, 2n]$. Furthermore, each α_i (resp. each β_j) and its Galois conjugates over $\mathbf{Q}$ have the same complex absolute value. For each ℓ-adic absolute value $|\ |_\ell$ with prime $\ell \neq p$, the α_i and the β_j are ℓ-adic units:

$$|\alpha_i|_\ell = |\beta_j|_\ell = 1.$$

For the remaining prime p, it is easy to prove

$$|\alpha_i|_p = q^{-r_i},\ |\beta_j|_p = q^{-s_j},\ r_i \in \mathbf{Q} \cap [0, 2n],\ s_j \in \mathbf{Q} \cap [0, 2n],$$

where we have normalized the p-adic absolute value by $|q|_p = q^{-1}$. Deligne's integrality theorem implies the following improved information:

$$r_i \in \mathbf{Q} \cap [0, n],\ s_j \in \mathbf{Q} \cap [0, n].$$

Strictly speaking, in defining the p-adic absolute value, we have tacitly chosen an embedding of the field $\bar{\mathbf{Q}}$ of algebraic numbers into an algebraic closure of the p-adic number field $\mathbf{Q}_p$. Note that each α_i (resp. each β_j) and its Galois conjugates over $\mathbf{Q}$ may have different p-adic absolute values.

The precise version of various types of Riemann hypothesis for the L-function in (1) is then to determine the important arithmetic invariants $\{u_i, v_j, r_i, s_j\}$. The integer u_i (resp. v_j) is called the weight of the algebraic integer α_i (resp. β_j). The rational number r_i (resp. s_j) is called the slope of the algebraic integer α_i (resp. β_j) defined with respect to q. Without any smoothness condition on f, not much more is known about these weights and the slopes, since one does not even know exactly the number d_1 of zeros and the number d_2 of poles. Under a suitable smoothness condition, a great deal more is known about the weights $\{u_i, v_j\}$ and the slopes $\{r_i, s_j\}$, see Adolphson-Sperber [2], Denef-Loesser [6] and Wan [18].

The object of this paper is to continue our investigation of the slopes $\{r_i, s_j\}$ and their variation as f and p vary. In the classical case that f is a diagonal polynomial such as a Fermat polynomial, the α_i (resp. the β_j) are roots of products of Gauss sums and the slopes can be obtained using the classical Stickelberger theorem. For non-diagonal f, the problem is much harder. The idea of our decomposition theory is to try to reduce the slope problem from harder non-diagonal f to easier diagonal f. Our main result is a new collapsing decomposition theorem which is much more flexible to use. It gives significantly better and often optimal information about the slopes in many cases. We shall illustrate this with a number of interesting examples in Section 3. In the rest of this introduction section, we describe precisely our basic questions, previous known results and our new theorems.

1.1. Newton polygon and generic Newton polygon. A Laurent polynomial f is a finite sum of monomials:

$$f = \sum_{j=1}^{J} a_j x^{V_j},\ a_j \neq 0,$$

where each $V_j = (v_{1j}, \cdots, v_{nj})$ is a lattice point in $\mathbf{Z}^n$ and the power x^{V_j} simply means the product $x_1^{v_{1j}} \cdots x_n^{v_{nj}}$. Let $\Delta(f)$ be the convex closure in $\mathbf{R}^n$ generated by the origin and the lattice points V_j $(1 \leq j \leq J)$. This is called the Newton polyhedron of f. If δ is a subset of $\Delta(f)$, we define the restriction of f to δ to be the Laurent polynomial

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

Without loss of generality, we always assume that $\Delta(f)$ is n-dimensional.

DEFINITION 1.1. *The Laurent polynomial f is called non-degenerate if for each closed face δ of $\Delta(f)$ of arbitrary dimension which does not contain the origin, the n partial derivatives*

$$\{\frac{\partial f^{\delta}}{\partial x_1}, \cdots, \frac{\partial f^{\delta}}{\partial x_n}\}$$

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

Note that the non-degenerate definition of f depends only on the monomials of f whose exponents are on the closed faces of $\Delta(f)$ not containing the origin. Let g be another Laurent polynomial such that $\Delta(g)$ is contained in $\Delta(f)$ and such that $\Delta(g)$ does not intersect the closed faces of $\Delta(f)$ not containing the origin. It is clear that the sum $f + g$, which may be viewed as a deformation of f, is also non-degenerate whenever f is non-degenerate. This is the case, for example, if $\Delta(g) - \{0\}$ is contained in the interior of $\Delta(f)$.

Assume now that f is non-degenerate (a smooth condition on the "leading form" of f) with respect to $\Delta(f)$. Then, the L-function $\mathrm{L}^*(f,T)^{(-1)^{n-1}}$ is a polynomial (not pure in general) of degree $n!\mathbf{V}(f)$ by a theorem of Adolphson-Sperber [2] proved using p-adic methods, where $\mathbf{V}(f)$ denotes the volume of $\Delta(f)$. The complex absolute values (or the weights) of the $n!\mathbf{V}(f)$ zeros can be determined explicitly by a theorem of Denef-Loeser [6] proved using ℓ-adic methods. They depend only on Δ, not on the specific f and p as long as f is non-degenerate with $\Delta(f) = \Delta$. Hence, the weights have no variation as f and p varies. As indicated above, the ℓ-adic absolute values of the zeros are always 1 for each prime $\ell \neq p$. Thus, there remains the intriguing question of determining the p-adic absolute values (or the slopes) of the zeros. This is the p-adic Riemann hypothesis for the L-function $\mathrm{L}^*(f,T)^{(-1)^{n-1}}$. Equivalently, the question is to determine the Newton polygon of the polynomial

$$\mathrm{L}^*(f,T)^{(-1)^{n-1}} = \sum_{i=0}^{n!\mathbf{V}(f)} A_i(f)T^i, \ A_i(f) \in \mathbf{Z}[\zeta_p], \tag{2}$$

Recall that the Newton polygon of (2), denoted by $\mathrm{NP}(f)$, is the lower convex closure in $\mathbf{R}^2$ of the points

$$(k, \mathrm{ord}_q A_k(f)), \ k = 0, 1, \cdots, n!\mathbf{V}(f).$$

We shall often think of $\mathrm{NP}(f)$ as the real valued function on the interval $[0, n!\mathbf{V}(f)]$ whose graph is the Newton polygon. Note that $\mathrm{NP}(f)$ is independent of the choice of the ground field $\mathbf{F}_q$ for which f is defined. Thus, the Newton polygon, which is by definition an arithmetic invariant, is also a geometric invariant. The question of determining the Newton polygon is unfortunately very complicated in general. The necessary complication of the Newton polygon problem is partly due to the fact that there does not exist a clean general answer or equivalently that the answer varies too much as f and the prime p vary.

One general property about the Newton polygon is the Grothendieck specialization theorem for F-crystals [11] or more generally for σ-modules [20]. This result provides a general algebraic structure theorem about the variation of the Newton polygon $\mathrm{NP}(f)$ when f varies in an algebraic family. Let Δ be a fixed n-dimensional integral convex polyhedron in $\mathbf{R}^n$ containing the origin. Let $\mathbf{V}(\Delta)$ be the volume of Δ. Let $\mathcal{N}_p(\Delta)$ be the parameter space of f over $\bar{\mathbf{F}}_p$ with fixed $\Delta(f) = \Delta$. This is a smooth affine variety defined over $\mathbf{F}_p$.

Let $\mathcal{M}_p(\Delta)$ be the set of non-degenerate f over $\bar{\mathbf{F}}_p$ with fixed $\Delta(f) = \Delta$. It is the compliment of a discriminant locus. Thus, $\mathcal{M}_p(\Delta)$ is a Zariski open smooth affine subset of $\mathcal{N}_p(\Delta)$. It is non-empty if p is large, say $p > n!\mathbf{V}(\Delta)$. Thus, $\mathcal{M}_p(\Delta)$ is again a smooth affine variety defined over $\mathbf{F}_p$. One can show that there is a locally free overconvergent σ-module $\mathcal{E}(\Delta)$ of rank $n!\mathbf{V}(\Delta)$ on $\mathcal{M}_p(\Delta)$ such that for each closed point f of $\mathcal{M}_p(\Delta)$, the L-function $\mathrm{L}^*(f,T)^{(-1)^{n-1}}$ is the characteristic polynomial of the Frobenius acting on the fibre $\mathcal{E}(\Delta)_f$ of $\mathcal{M}_p(\Delta)$ at f. That is,

$$\mathrm{L}^*(f,T)^{(-1)^{n-1}} = \det(I - T\mathrm{Frob}_f|\mathcal{E}(\Delta)_f).$$

We can ask how $\mathrm{NP}(f)$ varies as f varies. The variety $\mathcal{M}_p(\Delta)$ has countably many closed points, which can be listed as a sequence $\{f_1, f_2, \cdots\}$. The Grothendieck specialization theorem [20] implies that as f varies, the lowest Newton polygon

$$\mathrm{GNP}(\Delta,p) = \inf_{f\in\mathcal{M}_p(\Delta)} \mathrm{NP}(f) = \inf_{1\le i<\infty} \mathrm{NP}(f_i)$$

exists and is attained for all f in some Zariski open dense subset of $\mathcal{M}_p(\Delta)$. This lowest polygon can then be called the generic Newton polygon, denoted by $\mathrm{GNP}(\Delta,p)$. Note that $\mathrm{GNP}(\Delta,p)$ depends only on p and Δ. We would like to determine this generic Newton polygon and its variation with p.

Note that if we restrict to an irreducible curve $C \hookrightarrow \mathcal{M}_p(\Delta)$, then the Grothendieck specialization theorem implies that the limit (not just lower limit) $\lim_{f\in C} \mathrm{NP}(f)$ exists and is attained for all but finitely many points $f \in C$.

1.2. A lower bound: Hodge polygon. Another general property is that the Newton polygon lies on or above a certain topological or combinatorial lower bound, called the Hodge polygon. This is the Katz type conjecture. In the present setting of exponential sums, such a lower bound $\mathrm{HP}(\Delta)$ is given by Adolphson-Sperber in terms of the rational points in Δ. We now describe this combinatorial lower bound $\mathrm{HP}(\Delta)$.

Let Δ denote the n-dimensional integral polyhedron $\Delta(f)$ in $\mathbf{R}^n$ containing the origin. Let $C(\Delta)$ be the cone in $\mathbf{R}^n$ generated by Δ. Then $C(\Delta)$ is the union of all rays emanating from the origin and passing through Δ. If c is a real number, we define $c\Delta = \{cx|x \in \Delta\}$. For a point $u \in \mathbf{R}^n$, the weight $w(u)$ is defined to be the smallest non-negative real number c such that $u \in c\Delta$. If such c does not exist, we define $w(u) = \infty$.

It is clear that $w(u)$ is finite if and only if $u \in C(\Delta)$. If $u \in C(\Delta)$ is not the origin, the ray emanating from the origin and passing through u intersects Δ in a face δ of co-dimension 1 that does not contain the origin. The choice of the desired co-dimension 1 face δ is in general not unique unless the intersection point is in the interior of δ. Let $\sum_{i=1}^n e_iX_i = 1$ be the equation of the hyperplane δ in $\mathbf{R}^n$, where the coefficients e_i are uniquely determined rational numbers not all zero. Then, by standard arguments in linear programming, one finds that the weight function $w(u)$ can be computed using the formula:

$$w(u) = \sum_{i=1}^{n} e_i u_i, \tag{3}$$

where $(u_1, \cdots, u_n) = u$ denotes the coordinates of u.

Let $D(\delta)$ be the least common denominator of the rational numbers e_i $(1 \leq i \leq n)$. It follows from (3) that for a lattice point u in $C(\delta)$, we have

$$w(u) \in \frac{1}{D(\delta)} \mathbf{Z}_{\geq 0}. \tag{4}$$

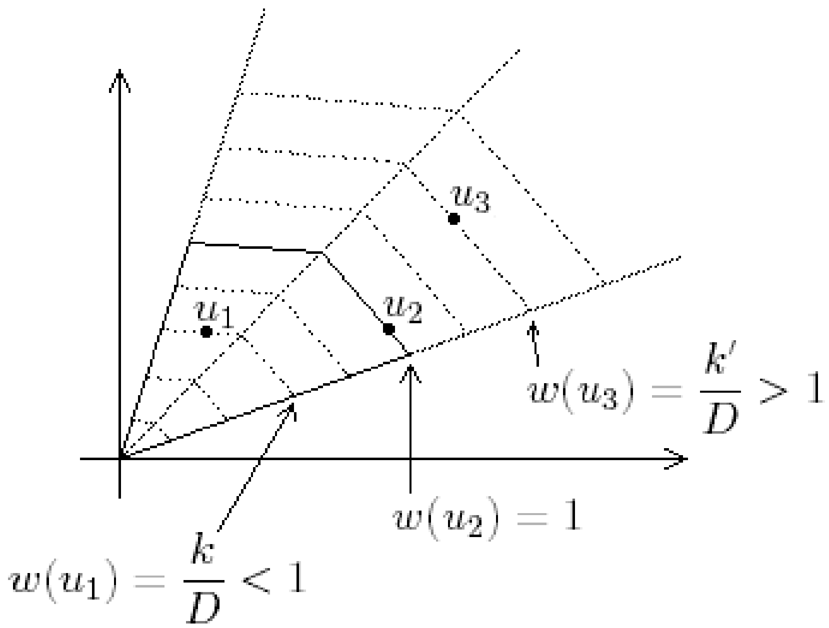
It is easy to show that there are lattice points $u \in C(\delta)$ such that the denominator of $w(u)$ is exactly $D(\delta)$. That is, the denominator in (4) is optimal. Let $D(\Delta)$ be the least common multiple of all the $D(\delta)$:

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

where δ runs over all the co-dimensional 1 faces of Δ which do not contain the origin. Then by (4), we deduce

$$w(\mathbf{Z}^n) \subseteq \frac{1}{D(\Delta)} \mathbf{Z}_{\geq 0} \cup \{+\infty\}, \tag{5}$$

where $\mathbf{Z}_{\geq 0}$ denotes the set of nonnegative integers. The integer $D = D(\Delta)$ is called the denominator of Δ. It is the smallest positive integer for which (5) holds.



For an integer k, let

$$W_\Delta(k) = \mathrm{card}\{u \in \mathbf{Z}^n \mid w(u) = \frac{k}{D}\}.$$

be the number of lattice points in $\mathbf{Z}^n$ with weight k/D. This is a finite number for each k. Let

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

This number 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. Thus, $H_\Delta(k)$ is a non-negative integer for each $k \in \mathbf{Z}_{\geq 0}$. Furthermore,

$$H_\Delta(k) = 0, \text{ for } k > nD$$

and

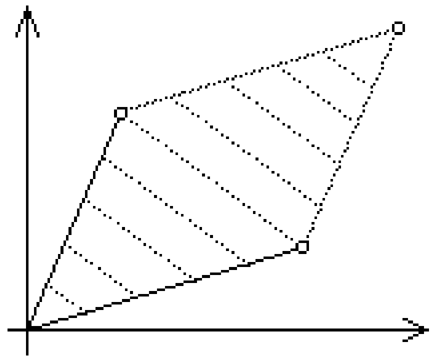
$$\sum_{k=0}^{nD} H_\Delta(k) = n!\mathbf{V}(\Delta).$$

Definition 1.2. *The Hodge polygon* $\mathrm{HP}(\Delta)$ *of* Δ *is defined to be the lower convex polygon in* $\mathbf{R}^2$ *with vertices*

$$\Big(\sum_{k=0}^{m} H_\Delta(k), \frac{1}{D}\sum_{k=0}^{m} kH_\Delta(k)\Big), \quad m = 0, 1, 2, \cdots, nD.$$

That is, the polygon $\mathrm{HP}(\Delta)$ *is the polygon starting from the origin and has a side of slope* k/D *with horizontal length* $H_\Delta(k)$ *for each integer* $0 \leq k \leq nD$.

The numbers $H_\Delta(k)$ coincide with the usual Hodge numbers in the toric hypersurface case, see [3] for general results in this direction. This explains the term "Hodge polygon". Note that in the geometric situation of zeta functions, one has $D = 1$ because one really considers the exponential sum of the new polynomial $x_0 f(x)$ in order to study the zeta function of the hypersurface defined by f.



The lower bound of Adolphson and Sperber [2] says that if $f \in \mathcal{M}_p(\Delta)$, then

$$\mathrm{NP}(f) \geq \mathrm{HP}(\Delta).$$

The Laurent polynomial f is called ordinary if $NP(f) = HP(f)$. Combining with the defining property of the generic Newton polygon, we deduce

Proposition 1.3. *For every prime p and every $f \in \mathcal{M}_p(\Delta)$, we have the inequalities*

$$\mathrm{NP}(f) \geq \mathrm{GNP}(\Delta, p) \geq \mathrm{HP}(\Delta). \tag{6}$$

This proposition gives rise to three possible inequalities. We are interested in classifying the polytopes Δ and the primes p for which one of the inequalities becomes an inequality. We will not study all the three questions in this paper, but only two of them involving the last polygon $\mathrm{HP}(\Delta)$.

1.3. Main results. The first question is when the generic Newton polygon coincides with the Hodge polygon. That is, when $\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$? By a simple ramification argument and the theory of σ-modules, one finds that the vertices of $\mathrm{GNP}(\Delta, p)$ have denominators dividing $p - 1$. On the other hand, the vertices of $\mathrm{HP}(\Delta)$ have denominators dividing D. In many cases, the denominator D for $\mathrm{HP}(\Delta)$ cannot be improved. In such a case, a necessary condition for $\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$ is clearly $p \equiv 1 (\mathrm{mod}\ D)$. The converse is the following conjecture of Adolphson-Sperber on generic Newton polygon.

Conjecture 1.4. *(AS). If $p \equiv 1 \pmod D$, then $\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$.*

Remark. If the parameter space $\mathcal{M}_p(\Delta)$ happens to be empty, we simply define $\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$ so that the conjecture is true in the empty case. For each Δ, the space $\mathcal{M}_p(\Delta)$ is always non-empty for large p, say if $p > n!\mathbf{V}(\Delta)$.

In the special case that Δ is the convex polytope of the homogeneous polynomial $x_0(x_1^d + x_2^d + \cdots + x_n^d)$, the AS conjecture implies the conjecture of Mazur [13] on the generic Newton polygon for the zeta functions of the universal family of hypersurfaces of degree d, see [18].

In [18], several decomposition theorems were established for the generic Newton polygon. As a consequence, it was proved that the AS conjecture is true for $n \leq 2$ but false for every $n \geq 5$. In the present paper, we establish a new and more flexible collapsing decomposition theorem for the generic Newton polygon, see Theorem 3.7. As a consequence, we are able to handle the remaining two dimensions ($n = 3, 4$) as well. In particular, we have a complete answer to the AS conjecture (and hence our first question as well) in terms of the dimension n.

Theorem 1.5. *The AS conjecture is true in every low dimension $n \leq 3$ but false in every high dimension $n \geq 4$.*

Although the AS conjecture is not true in general in higher dimensions, we would like to know how far the conjecture is from being true and when it is true in various important special cases. It turns out that a weaker form of the AS conjecture is always true. Namely, we have the following theorem.

Theorem 1.6. *There is an effective positive integer $D^* = D^*(\Delta) \geq D$ such that if $p \equiv 1 \pmod{D^*}$, then $\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$.*

This theorem was first proved in [18] using the star decomposition developed there. It also follows from the collapsing decomposition theorem of the present paper. The quantity D^* one gets using the new collapsing decomposition is much smaller (often optimal) than what one gets using the old star decomposition. This allows us to prove that the full form of the AS conjecture is true for all $n \leq 3$ and for many other important higher dimensional Δ. That is, one can take $D^* = D$ in many cases.

In Section 3, we give a detailed description of the collapsing decomposition theorem as well as a number of further applications to zeta functions and L-functions.

In general, the smallest possible value for D^* is quite subtle. It depends not just on the combinatorial shape of Δ but also on subtle p-adic arithmetic property of Δ as well. Again, there is probably no simple formula for the smallest D^*, but see [18] for a classification conjecture on those large p for which $\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$. In particular, this classification conjecture implies an explicit algorithm to compute the smallest possible D^*.

An immediate consequence of Theorem 1.6 is

Corollary 1.7. *Let Δ be an n-dimensional integral convex polytope in $\mathbf{R}^n$ containing the origin. Then, the lower limit*

$$\liminf_{p \to \infty} \mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$$

exists and is attained for a set of primes p with positive density.

Our second question is when $\mathrm{NP}(f) = \mathrm{HP}(\Delta)$ for all $f \in \mathcal{M}_p(\Delta)$. The answer would depend on a more subtle relation between Δ and p. Let δ_i $(1 \le i \le h)$ be the set of closed codimension 1 faces of Δ not containing the origin. Let Δ_i be the n-dimensional polytope generated by δ_i and the origin. The facial decomposition of Δ is defined (see [18])by

$$\Delta = \bigcup_{i=1}^{h} \Delta_i. \tag{7}$$

We have

Theorem 1.8. $\mathrm{NP}(f) = \mathrm{HP}(\Delta)$ *for all $f \in \mathcal{M}_p(\Delta)$ if the following two conditions hold. (i). For each $1 \le i \le h$, δ_i is a simplex containing no lattice points other than the vertices. (ii). For each $1 \le i \le h$, multiplication by p on the finite group of the lattice points in the fundamental domain of Δ_i is a weight preserving map. That is, $w(\{pu\}) = w(u)$ for all lattice points u in the fundamental domain of Δ_i, where $\{pu\}$ is the class of the lattice point pu in the fundamental domain of Δ_i.*

The converse of the above theorem is not always true. It would be interesting to classify those Δ and p for which $\mathrm{NP}(f) = \mathrm{HP}(\Delta)$ for all $f \in \mathcal{M}_p(\Delta)$.

1.4. Limiting conjectures on Newton polygons. To motivate further research on Newton polygons, we state three limiting conjectures about the variation of the Newton polygon as the prime p varies. The first one concerns the lower limit of the Newton polygon of a fixed Laurent polynomial f as p varies. This one is easy to state but difficult to prove in general. It greatly strengthens Corollary 1.7. It can also be viewed as an arithmetic version of the geometric Grothendieck specialization theorem.

Conjecture 1.9. *Let Δ be an n-dimensional integral convex polyhedron in $\mathbf{R}^n$ containing the origin. Let $f(x)$ be a non-degenerate Laurent polynomial with rational coefficients such that $\Delta(f) = \Delta$. Then, we have the lower limiting formula*

$$\liminf_{p \to \infty} \mathrm{NP}(f(x) \bmod p) = \mathrm{HP}(\Delta) \tag{8}$$

and this lower limit is attained for a set of primes p with positive density.

The Laurent polynomial is called **diagonal** if the exponents of its non-constant terms are the non-zero vertices of an n-dimensional simplex Δ containing the origin. By the Stickelberger theorem, this conjecture is true if $f(x)$ is a diagonal Laurent polynomial, see section 2 below. More generally, the conjecture is true if $f(x)$ is a certain deformation of a diagonal Laurent polynomial. The most general result in this direction is the following theorem which is an immediate consequence of our facial decomposition theorem, see section 3.

THEOREM 1.10. *Let Δ be an n-dimensional integral convex polyhedron in $\mathbf{R}^n$ containing the origin. Let $f(x)$ be a non-degenerate Laurent polynomial with rational coefficients such that $\Delta(f) = \Delta$. Assume that the restriction f_{δ_i} of f to each codimension 1 face δ_i $(1 \leq i \leq h)$ not containing the origin is diagonal. Then, there is an explicitly determined positive integer D^* such that for all large primes $p \equiv 1 \pmod{D^*}$, we have the equality*

$$\mathrm{NP}(f(x) \bmod p) = \mathrm{HP}(\Delta). \tag{9}$$

Note that the condition in Theorem 1.10 implies that Δ must be **simplicial**, i.e., each codimension 1 face δ_i of Δ is a simplex. Theorem 1.10 provides many non-trivial examples for which Conjecture 1.9 holds. For simplicity, we stated Conjecture 1.9 only for the rational number field. The conjecture easily generalizes to any fixed number field.

The limit of the Newton polygon $\mathrm{NP}(f)$ as p varies does not exist in general, as the diagonal case shows. The next question we want to ask is when the limit (not just lower limit) exists as p varies and is equal to the Hodge polygon. Depending on $\mathrm{NP}(f)$ or $\mathrm{GNP}(\Delta, p)$, we are really asking two questions in this direction.

First, we consider the possible limit of the generic Newton polygon. The limit cannot exist in general for general Δ, as the minimal diagonal example in Section 2.4 shows. Let $L(\Delta)$ be the set of lattice points in the cone $C(\Delta)$. This is a monoid. Let $E(\Delta)$ be the monoid generated by the finitely many lattice points in the polytope Δ. This is a sub-monoid of $L(\Delta)$.

CONJECTURE 1.11. *Let Δ be an n-dimensional integral convex polytope in $\mathbf{R}^n$ containing the origin. Assume that the difference $L(\Delta) - E(\Delta)$ is a finite set. Then, the limit*

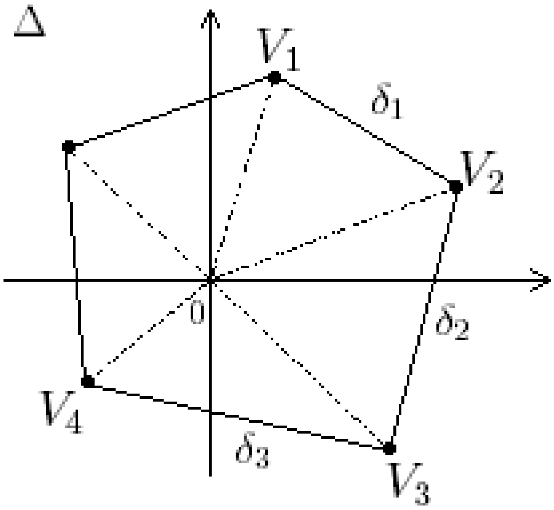
$$\lim_{p \to \infty} \mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$$

exists and is equal to the Hodge polygon.

This conjecture is true in the case when the integer $D^*(\Delta)$ in Theorem 1.6 can be taken to be at most 2. There are many such examples in the geometric situation of zeta functions, where $D^* = D = 1$ (the AS conjecture). Note it is not true in general that the limit in the above conjecture is attained for all large primes, although Corollary 1.7 says that the limit is indeed attained for a set of primes with positive density.

Next, we consider the limit of $\mathrm{NP}(f)$ as p varies. Again the limit cannot exist in general. For the limit to exist, even the condition in Theorem 1.10 is too weak as the diagonal case shows. However, we conjecture that the limit of the Newton polygon as p varies does exist and hence equals to the Hodge polygon for most of the integral polynomials if we combine the conditions in Theorem 1.10 and Conjecture 1.11.

Let δ_i $(1 \leq i \leq h)$ be the set of closed codimension 1 faces of Δ not containing the origin. Let Δ_i be the n-dimensional polytope generated by δ_i and the origin.



Conjecture 1.12. *Assume that Δ is simplicial. Let $\mathcal{V}_1 = \{V_1, \cdots, V_{J_1}\}$ be the set of all non-zero vertices in Δ. Let $\mathcal{V}_2 = \{W_1, \cdots, W_{J_2}\}$ be a set of lattice points in Δ with weight strictly smaller than one. Let a_i $(1 \leq i \leq J_1)$ be n fixed non-zero rational numbers. Let $\mathbf{A}^{J_2}$ be the family of Laurent polynomials*

$$f(x) = \sum_{i=1}^{J_1} a_i x^{V_i} + \sum_{j=1}^{J_2} b_j x^{W_j} \tag{10}$$

parametrized by the b_j. Let $E(J)$ be the monoid generated by the J lattice points in $\mathcal{V}_1 \cup \mathcal{V}_2$. Assume that $L(\Delta) - E(J)$ is a finite set. Then, there is a Zariski open dense subset $U \hookrightarrow \mathbf{A}^{J_2}$ defined over $\mathbf{Q}$ such that for each $f \in U(\bar{\mathbf{Q}})$, we have

$$\lim_{p\to\infty} \mathrm{NP}(f, p) = \mathrm{HP}(\Delta), \tag{11}$$

where $\mathrm{NP}(f,p)$ denotes $\mathrm{NP}(f \bmod P)$ and P denotes a prime ideal lying above p in the ring of integers of the number field generated by the coefficients of f. Note that for large p, $\mathrm{NP}(f \bmod P)$ depends only on p and not on the choice of the prime ideal P lying above p.

The condition in Conjecture 1.12 is more restrictive than that in Conjecture 1.11. But its conclusion is also stronger. In particular, an immediate consequence of Conjecture 1.12 is that for all $\mathbf{Q}$-rational points $f \in U(\mathbf{Q})$, we have the limiting formula

$$\lim_{p\to\infty} \mathrm{NP}(f(x) \bmod p) = \mathrm{HP}(\Delta). \tag{12}$$

This equation means that the limit on the left side exists and is equal to the right side.

We believe that both Conjecture 1.11 and Conjecture 1.12 are much more realistic to prove than Conjecture 1.9. By Corollary 1.7 and Theorem 1.10, the key point of Conjectures 1.11 and 1.12 is that the limit exists. The first interesting case to consider includes the family of normalized one variable polynomials of degree d:

$$f(x) = x^d + a_{d-2}x^{d-2} + \cdots + a_1 x. \tag{13}$$

This one variable case has received considerable attentions in the work of Sperber [16], Hong [9][10] and Yang [22]. More recently, the one variable case of Conjecture 1.12 has been proved by June Hui Zhu [23][24]. One can also consider a suitable sub-family of the above family by specifying certain coefficients to be zero.

1.5. Further questions. Assume that p is a prime such that $\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$. Let $U_p(\Delta)$ be an affine open dense subset of the variety $\mathcal{M}_p(\Delta)$ such that $\mathrm{NP}(f) = \mathrm{HP}(\Delta)$ for $f \in U_p(\Delta)$. Since $\mathcal{M}_p(\Delta)$ is smooth, $U_p(\Delta)$ is also smooth. The restriction of the σ-module $\mathcal{E}(\Delta)$ to $U_p(\Delta)$ is then an ordinary overconvergent σ-module. By the Hodge-Newton decomposition [20] for σ-modules, there is an increasing filtration of sub-σ-modules (no longer overconvergent) on $U_p(\Delta)$:

$$0 \subset \mathcal{E}_0 \subset \mathcal{E}_1 \subset \cdots \subset \mathcal{E}_{nD} = \mathcal{E}(\Delta)|_{U_p(\Delta)},$$

such that the successive quotient $\mathcal{E}_k/\mathcal{E}_{k-1}$ is a pure σ-module on $U_p(\Delta)$ of rank $H_\Delta(k)$ and slope k/D. The L-function of such a pure σ-module and its higher power Adams operations [19] measure the p-adic arithmetic variation of the pure σ-module. By Dwork's conjecture [7] as proved in [19][20][21] in the setting of σ-modules, such a pure L-function is p-adic meromorphic everywhere. It would be interesting to get information about the slopes (or the Newton polygon) of the pure L-function. As an initial result in this direction, an explicit lower bound for the Newton polygon of the pure L-function is obtained in [21] in the case that the pure σ-module has rank one. Note that in the current setting, the unit root σ-module $\mathcal{E}_0$ is always of rank one, which is non-trivial if the origin is not a vertex of Δ. For more examples, let $\mathcal{F}_k$ be the unit root σ-module arising from the slope k/D part $\mathcal{E}_k/\mathcal{E}_{k-1}$. Then $\det\mathcal{F}_k$ is a rank one unit root σ-module. An intermediate and perhaps more accessible question is to study the k-th moment L-function of the σ-module $\mathcal{E}(\Delta)$ and its variation as the positive integer k varies, see Fu-Wan [8] for general structural results in this direction.

2. Diagonal local theory. In this section, we give a thorough treatment of the diagonal case, refining and improving the calculations in our earlier paper [17]. Recall that a Laurent polynomial f is called **diagonal** if f has exactly n non-constant terms and $\Delta(f)$ is n-dimensional (necessarily a simplex). In this case, the L-function can be computed explicitly using Gauss sums, just as the well known case of the zeta function of a Fermat hypersurface. Thus, theoretically speaking or from algorithmic point of view, the Newton polygon in the diagonal case can be completely determined using the Stickelberger theorem. It should, however, be noted that interesting arithmetic and combinatorial problems will often arise in the actual explicit calculation of the Newton polygon due to the diversity of the simplex Δ. The diagonal case forms the building blocks for the general decomposition theory in next section. It also provides a rich source of explicit interesting examples, including counter-examples of the AS conjecture.

2.1. p-action, Gauss sums and L-functions. To describe the L-function in terms of Gauss sums, we first review the p-action on lattice points. Let $0, V_1, \cdots, V_n$ be the vertices of an n-dimensional integral simplex Δ in $\mathbf{R}^n$. Let f be the diagonal Laurent polynomial

$$f(x) = \sum_{j=1}^{n} a_i x^{V_j}, a_j \in \mathbf{F}_q^*.$$

Let M be the non-singular $n \times n$ matrix

$$M = (V_1, \cdots, V_n),$$

where each V_j is written as a column vector. It is easy to check that f is non-degenerate if and only if p is relatively prime to $\det M$. We do not assume that p satisfies this condition. We consider the solutions of the following linear system

$$M\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \equiv 0 \ (\mathrm{mod}\ 1), \ r_i \text{ rational}, \ 0 \le r_i < 1. \tag{14}$$

The map $(r_1, \cdots, r_n) \to r_1V_1 + \cdots + r_nV_n$ clearly establishes a one-to-one correspondence between the solutions of (14) and the integral lattice points of the fundamental domain

$$\mathbf{R}V_1 + \cdots + \mathbf{R}V_n (\mathrm{mod}\ \mathbf{Z}V_1 + \cdots + \mathbf{Z}V_n). \tag{15}$$

Under this bijection, we can identify the solution set of (14) and the set of integral lattice points in the above fundamental domain. Let $S(\Delta)$ be the set of solutions r of (14), which may be identified with the lattice points $u = Mr$ in the fundamental domain (15). It has a natural abelian group structure under addition modulo 1. By the theory of elementary abelian groups, the order of $S(\Delta)$ is precisely given by

$$|\det M| = n!\mathbf{V}(\Delta).$$

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 $\det M$. In particular, $S_p(\Delta) = S(\Delta)$ if p is relatively prime to $\det M$.

If m is an integer relatively prime to the order of $S_p(\Delta)$, then multiplication by m induces an automorphism of the finite abelian group $S_p(\Delta)$. This map is called the m-map (or m-action) of $S_p(\Delta)$, denoted by the notation $r \to \{mr\}$, where

$$\{mr\} = (\{mr_1\}, \cdots, \{mr_n\})$$

and $\{mr_i\}$ denotes the fractional part of the real number mr_i. For each element $r \in S_p(\Delta)$, let $d(m, r)$ be the smallest positive integer such that multiplication by $m^{d(m,r)}$ acts trivially on r, i.e.,

$$(m^{d(m,r)} - 1)r \in \mathbf{Z}^n.$$

The integer $d(m, r)$ is the order of the m-map restricted to the cyclic subgroup generated by r. It is called the m-degree of r. For each positive integer d, let $S_p(m, d)$ be the set of $r \in S_p(\Delta)$ such that $d(m, r) = d$. Of course, the set $S_p(m, d)$ is empty for all large d since $S_p(\Delta)$ is finite. We have the disjoint m-degree decomposition

$$S_p(\Delta) = \bigcup_{d \ge 1} S_p(m, d).$$

Since p is relatively prime to the order of $S_p(\Delta)$, we will use the case that m is a power of p, such as p and q.

In order to compute the L-function, we now recall the definition of Gauss sums. Let χ be the Teichmüller character of the multiplicative group $\mathbf{F}_q^*$. For $a \in \mathbf{F}_q^*$, the value $\chi(a)$ is just the $(q-1)$-th root of unity in the p-adic field Ω such that $\chi(a)$ modulo p reduces to a. Define the $(q-1)$ Gauss sums over $\mathbf{F}_q$ by

$$G_k(q) = -\sum_{a \in \mathbf{F}_q^*} \chi(a)^{-k} \zeta_p^{\mathrm{Tr}(a)} \quad (0 \le k \le q-2),$$

where Tr denotes the trace map from $\mathbf{F}_q$ to the prime field $\mathbf{F}_p$. For each $a \in \mathbf{F}_q^*$, the Gauss sums satisfy the following interpolation relation

$$\zeta_p^{\mathrm{Tr}(a)} = \sum_{k=0}^{q-2} \frac{G_k(q)}{1-q}\chi(a)^k.$$

One then calculates that

$$\begin{aligned}
S_1^*(f) &= \sum_{x_j \in \mathbf{F}_q^*} \zeta_p^{\mathrm{Tr}(f(x))} \\
&= \sum_{x_j \in \mathbf{F}_q^*} \prod_{i=1}^{n} \zeta_p^{\mathrm{Tr}(a_i x^{V_i})} \\
&= \sum_{x_j \in \mathbf{F}_q^*} \prod_{i=1}^{n} \sum_{k_i=0}^{q-2} \frac{G_{k_i}(q)}{1-q}\chi(a_i)^{k_i}\chi(x^{V_i})^{k_i} \\
&= \sum_{k_1=0}^{q-2} \cdots \sum_{k_n=0}^{q-2} (\prod_{i=1}^{n} \frac{G_{k_i}(q)}{1-q}\chi(a_i)^{k_i}) \sum_{x_j \in \mathbf{F}_q^*} \chi(x^{k_1V_1+\cdots+k_nV_n}) \\
&= (-1)^n \sum_{k_1V_1+\cdots+k_nV_n \equiv 0 (\mathrm{mod} q-1)} \prod_{i=1}^{n} \chi(a_i)^{k_i} G_{k_i}(q). \qquad (16)
\end{aligned}$$

This gives a formula for the exponential sum $S_1^*(f)$ over the field $\mathbf{F}_q$. Replacing q by q^k, one gets a formula for the exponential sum $S_k^*(f)$ over the k-th extension field $\mathbf{F}_{q^k}$.

If $r = (r_1, \cdots, r_n)$ and $r' = (r_1', \cdots, r_n')$ are two elements in $S_p(p, d)$ which are in the same orbit under the p-action, that is

$$r' = \{p^k r\}$$

for some integer k, then one checks from the above definition of Gauss sums that

$$G_{r_i(q^d-1)}(q^d) = G_{r_i'(q^d-1)}(q^d)$$

for all $1 \le i \le n$, where $G_k(q^d)$ is the Gauss sum defined over the finite extension field $\mathbf{F}_{q^d}$. Thus, the p-action and the q-action do not change the Gauss sum. For $r \in S_p(q, d)$, the well known Hasse-Davenport relation says that

$$G_{r(q^{dk}-1)}(q^{dk}) = G_{r(q^d-1)}(q^d)^k,$$

for every positive integer k. Using this relation and (16), one then obtains the following explicit formula [17] for the L-function.

THEOREM 2.1. *Using the above notations, we have*

$$L^*(f/\mathbf{F}_q, T)^{(-1)^{n-1}} = \prod_{d \ge 1} \prod_{r \in S_p(q,d)} \Big(1 - T^d \prod_{i=1}^{n} \chi(a_i)^{r_i(q^d-1)} G_{r_i(q^d-1)}(q^d)\Big)^{\frac{1}{d}}, \qquad (17)$$

where $r = (r_1, \cdots, r_n)$.

Note that we used the q-action in the definition of the set $S_p(q, d)$ which occurs in (17). For each of the d points in the orbit of $r \in S_p(q, d)$ under the q-action, the corresponding factor in (17) is the same. Thus, we can remove the power $1/d$ if we restrict r to run over the q-orbits (closed points) of $S_p(q, d)$. In particular, this shows that the right side of (17) is indeed a polynomial of degree $|S_p(\Delta)|$.

2.2. Applications of the Stickelberger theorem. The Stickelberger theorem for Gauss sums can be described as follows.

THEOREM 2.2. *Let $0 \leq k \leq q-2$. Let $\sigma_p(k)$ be the sum of the p-digits of k in its base p expansion. That is,*

$$\sigma_p(k) = k_0 + k_1 + k_2 + \cdots, \quad k = k_0 + k_1 p + k_2 p^2 + \cdots, \quad 0 \leq k_i \leq p-1.$$

Then,

$$\mathrm{ord}_p G_k(q) = \frac{\sigma_p(k)}{p-1}.$$

Combining Theorems 2.1 and 2.2, one can then completely determine the p-adic absolute values of the reciprocal zeros of $L^*(f,T)^{(-1)^{n-1}}$. In particular, the Newton polygon is independent of the non-zero coefficients a_j of the diagonal Laurent polynomial f. Thus, for simplicity, we may assume that all the coefficients are 1. Namely,

$$f = \sum_{j=1}^{n} x^{V_j}. \tag{18}$$

Applying Theorem 2.1 to the polynomial in (18) and $q = p$, we obtain

COROLLARY 2.3. *For f in (18), we have*

$$L^*(f/\mathbf{F}_p, T)^{(-1)^{n-1}} = \prod_{d \geq 1} \prod_{r \in S_p(p,d)} \Big(1 - T^d \prod_{i=1}^{n} G_{r_i(p^d-1)}(p^d)\Big)^{\frac{1}{d}}, \tag{19}$$

where $r = (r_1, \cdots, r_n)$ and the p-action is used.

The Newton polygon computed with respect to q of the polynomial in Theorem 2.1 is the same as the Newton polygon computed with respect to p of the polynomial in Corollary 2.3. This follows directly from the relationship between the two L-functions: the reciprocal zeros of $L^*(f/\mathbf{F}_q, T)^{(-1)^{n-1}}$ for the polynomial f in (18) are exactly the k-th power of the reciprocal zeros of $L^*(f/\mathbf{F}_p, T)^{(-1)^{n-1}}$, where $q = p^k$.

If now $r = (r_1, \cdots, r_n)$ is an element of $S_p(p,d)$ and α_r is any one of the d reciprocal roots (which differ only by a d-th root of unity) of the corresponding factor in (19), then the Stickelberger theorem implies that

$$\begin{aligned}
\mathrm{ord}_p(\alpha_r) &= \frac{1}{d}\mathrm{ord}_p \prod_{i=1}^{n} G_{r_i(p^d-1)}(p^d) \\
&= \frac{1}{d(p-1)} \sum_{i=1}^{n} \sigma_p(r_i(p^d-1)) \\
&= \frac{1}{d(p-1)} \sum_{i=1}^{n} \frac{p-1}{p^d-1} \sum_{j=0}^{d-1} \{p^j r_i\}(p^d-1) \\
&= \frac{1}{d} \sum_{j=0}^{d-1} |\{p^j r\}|,
\end{aligned} \tag{20}$$

where for $r \in S_p(\Delta)$, the norm $|r| = r_1 + \cdots + r_n$ is just the weight $w(u)$ of the corresponding lattice point $u = Mr$ in the fundamental domain (15). From (20), one can write down a (complicated) formula for the Newton polygon of the diagonal L-function.

For $r \in S_p(p,d) \subset S_p(\Delta)$, we define the average norm of r to be

$$w([r]) = \frac{1}{d}\sum_{j=0}^{d-1} |\{p^j r\}|.$$

Thus, $w([r])$ can be viewed as the norm of the orbit of r under the p-action. For a rational number $0 \le s \le n$, let $h_s(p)$ denote the number of elements $r \in S_p(\Delta)$ whose average norm is s. Then, we have

COROLLARY 2.4. *For f in (18), for each rational number $0 \le s \le n$, the Newton polygon of $L^*(f/\mathbf{F}_p, T)^{(-1)^{n-1}}$ has a side of slope s whose horizontal length is $h_s(p)$.*

2.3. Ordinary criterion and ordinary primes. We would like to know when the diagonal Laurent polynomial f in (18) is ordinary, that is, when $NP(f) = HP(f)$ if f is non-degenerate. For this purpose, we need to describe the Hodge polygon in the diagonal case. Since our diagonal f may not be non-degenerate, our discussion below is a little bit more general. In particular, the Hodge polygon would be a little more general and depend on the prime number p if p is small. Consider the following linear system for $u \in \mathbf{Z}^n$,

$$M\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = u,\ 0 \le r_i. \tag{21}$$

If $u \in C(\Delta)$, system (21) has exactly one rational solution $r = (r_1, \cdots, r_n)$. In this case, the weight $w(u)$ of u is given by

$$w(u) = r_1 + \cdots + r_n = |r|.$$

Let $W_\Delta(k,p)$ be the number of lattice points $u \in \mathbf{Z}^n$ such that

$$w(u) = r_1 + \cdots + r_n = \frac{k}{D} \tag{22}$$

and such that each r_i is p-integral (that is, the denominator of r_i is relatively prime to p). Let

$$H_\Delta(k,p) = W_\Delta(k,p) - \binom{n}{1} W_\Delta(k-D,p) + \binom{n}{2} W_\Delta(k-2D,p) - \cdots.$$

By the inclusion-exclusion principle, $H_\Delta(k,p)$ is simply the number of elements in $S_p(\Delta)$ with weight k/D. Define the prime to p Hodge polygon $\mathrm{HP}_p(\Delta)$ to be the polygon with vertices (0,0) and

$$\Big(\sum_{k=0}^{m} H_\Delta(k,p), \sum_{k=0}^{m} \frac{k}{D} H_\Delta(k,p)\Big),\ m = 0, 1, \cdots, nD.$$

It is clear that $HP(\Delta) = HP_p(\Delta)$ if and only if p is relatively prime to $\det M$. The diagonal f in (18) is called ordinary at p if the Newton polygon of $L^*(f/\mathbf{F}_p, T)^{(-1)^{n-1}}$ coincides with the prime to p Hodge polygon $HP_p(\Delta)$.

THEOREM 2.5. *The diagonal Laurent polynomial f in (18) is ordinary at p if and only if the norm function $|r|$ on $S_p(\Delta)$ is stable under the p-action: That is, for each $r \in S_p(\Delta)$, we have*

$$|r| = |\{pr\}|.$$

Equivalently, this means that the weight function $w(u)$ on the lattice points of $S_p(\Delta)$ is stable under the p-action:

$$w(u) = w(\{pu\}). \tag{23}$$

Proof. If the norm function $|r|$ on $S_p(\Delta)$ is stable under the p-action, then for any $r \in S_p(p, d)$, (20) shows that each corresponding reciprocal root α_r satisfies

$$\mathrm{ord}_p(\alpha_r) = |r| = w(u),$$

where $u = Mr$. By Corollary 2.3, the L-function $L^*(f/\mathbf{F}_p, T)^{(-1)^{n-1}}$ has exactly $H_\Delta(k, p)$ reciprocal roots α with slope $\mathrm{ord}_p(\alpha) = k/D$. This means that f is ordinary at p.

Conversely, if f is ordinary at p, we need to prove that the norm function on $S_p(p, d)$ is stable under p-action. This can be proved by induction on the norm of r as follows. The norm 0 subset of $S(p, d)$ consists of $r = 0/D = 0$. It is clear that $\{p^i 0\} = 0$. Thus, the norm 0 subset of $S_p(p, d)$ is stable under the p-action. Removing the zero element from $S_p(p, d)$, if $r \in S_p(p, d) - \{0\}$ has minimum norm $1/D$, then the ordinary assumption implies that the average norm of r must be $1/D$. Since each $\{p^i r\}$ has norm at least $1/D$, we deduce that the norm of $\{p^i r\}$ is exactly the minimum norm $1/D$. This proves that the norm $1/D$ subset of $S_p(p, d)$ is stable under the p-action. By induction, one shows that for each positive integer k, the norm k/D subset of $S_p(p, d)$ is stable. The theorem is proved.

In particular, if the p-action on $S_p(\Delta)$ is trivial, then the weight function $w(u)$ on the lattice points of $S_p(\Delta)$ is automatically stable under the p-action and hence the diagonal Laurent polynomial f in (18) is ordinary at p. Let $d_1(p) \mid d_2(p) \mid \cdots \mid d_n(p)$ be the invariant factors of the finite abelian group $S_p(\Delta)$. That is,

$$S_p(\Delta) = \bigoplus_{i=1}^{n} \mathbf{Z}/d_i(p)\mathbf{Z},$$

where $d_i(p) | d_{i+1}(p)$ for $1 \le i \le n-1$. By the theory of finite abelian groups, multiplication by the largest invariant factor $d_n(p)$ kills the finite abelian group $S_p(\Delta)$. Thus, if $p-1$ is divisible by $d_n(p)$, the p-action on $S_p(\Delta)$ becomes trivial. Let d_n be the largest invariant factor of $S(\Delta)$. It is clear that $d_n(p)$ is a factor of d_n for every p. Furthermore, $d_n(p) = d_n$ if and only if p does not divide $\det M$. We obtain

COROLLARY 2.6. *Let $d_n(p)$ be the largest invariant factor of $S_p(\Delta)$. Let d_n be the largest invariant factor of $S(\Delta)$. If $p \equiv 1 \pmod{d_n(p)}$, then the diagonal Laurent polynomial in (18) is ordinary at p. In particular, if $p \equiv 1 \pmod{d_n}$, then the diagonal Laurent polynomial in (18) is ordinary at p.*

As an application, we get information about the variation of the Newton polygon when p varies. For example, we have

Corollary 2.7. *Let $f(x)$ be a diagonal Laurent polynomial with integer coefficients. Let d_n be the largest invariant factor of the n-dimensional simplex $\Delta(f)$. For a positive real number t, let $\pi_f(t)$ be the number of primes $p \leq t$ such that $f(x)$ is ordinary at p. Then, there is a positive integer $\mu(\Delta) \leq \varphi(d_n)$ such that the following asymptotic formula holds*

$$\pi_f(t) \sim \frac{\mu(\Delta)}{\varphi(d_n)} \frac{t}{\log t},$$

where $\varphi(d_n)$ denotes the Euler function.

In fact, if p_1 is an ordinary prime for f, then Theorem 2.5 implies that f is ordinary for every prime $p \equiv p_1 (\mathrm{mod}\ d_n)$. Thus, the set of ordinary primes for f are exactly the primes in certain residue classes modulo d_n. The integer $\mu(\Delta)$ is the number of positive integers $m \leq d_n$ such that $(m, d_n) = 1$ and the weight function $w(u)$ on the fundamental domain of $\Delta(f)$ is stable under the m-action.

We now discuss the relationship between the denominator $D(\Delta)$ and the largest invariant factor d_n. By definition, the denominator $D(\Delta)$ is the least common denominator for the coordinates of the vector $(e_1, \cdots, e_n)$ which is the unique solution of the linear system

$$(e_1, \cdots, e_n)M = (1, \cdots, 1).$$

Solving this linear system, one finds that

$$(e_1, \cdots, e_n) = (1, \cdots, 1)M^{-1} \in \frac{1}{d_n}\mathbf{Z}^n.$$

This shows that $D(\Delta)$ is a factor of d_n and we have the inequality

$$D(\Delta) \leq d_n.$$

In general, $D(\Delta)$ will be a proper factor of d_n if $n \geq 2$. The simplest example is to take the 2 variable diagonal Laurent polynomial $f(x) = x_1^d x_2^{1-d} + x_2$. One checks that $D = 1$ but the largest invariant factor $d_2 = d$.

We finish this subsection by showing that the full form of the AS conjecture holds in the indecomposable case for $n \leq 3$.

Definition 2.8. *Let δ be the $(n-1)$-dimensional face generated by the V_j $(1 \leq j \leq n)$. This is the unique co-dimension 1 face of the simplex Δ not containing the origin. We say that Δ is* **indecomposable** *if the face δ contains no lattice points other than the vertices V_j.*

Note that the larger Δ may have non-vertex lattice points with weight less than 1.

Corollary 2.9. *The AS conjecture is true for indecomposable Δ if $n \geq 3$ or if $n = 4$ but $D = 1$.*

Proof. Let f^δ be the restriction of f to δ. Since Δ is indecomposable, the restriction f^δ is a diagonal Laurent polynomial. We may assume now that f^δ is non-degenerate, that is, p is relatively prime to $n!\mathbf{V}(\Delta)$; otherwise there is nothing to prove. By the deformation consequence of Theorem 3.1 in Section 3.1, we know that $\mathrm{NP}(f)$ coincides with $\mathrm{HP}(\Delta)$ if and only if $\mathrm{NP}(f^\delta)$ coincides with $\mathrm{HP}(\Delta)$. By Corollary 2.6, we only need to prove that the denominator $D(\Delta)$ is equal to the largest invariant factor d_n under the assumption of the corollary. Let

$$e_1X_1 + \cdots + e_nX_n = D,\ \gcd(e_1, \cdots, e_n) = 1 \tag{24}$$

be the equation of the hyperplane δ, where the e_i are relatively prime integers. Applying the Euclidean algorithm to the columns of the row vector $\{e_1, \cdots, e_n\}$, one finds that there is a matrix $R \in \mathrm{GL}_n(\mathbf{Z})$ such that

$$(e_1, \cdots, e_n)R = (1, 0, \cdots, 0).$$

Let $P = R^{-1}$. Then, the first row of P is $(e_1, \cdots, e_n)$ and we have $P \in \mathrm{GL}_n(\mathbf{Z})$. Since each column vector V_j of M satisfies equation (24), we see that the first row of the product matrix PM is $(D, \cdots, D)$. This normalization shows that without loss of generality, we may assume that δ is defined by the equation $X_1 = D$. Recall that

$$D \le d_n \le |\det(M)|.$$

Thus, it suffices to prove that $\det(M) = \pm D$.

If $n = 1$, it is trivial that $D = d_1$. If $n = 2$, then δ is the line segment between $V_1 = (D, v_1)$ and $V_2 = (D, v_2)$ for some integers $v_1 \neq v_2$. Since there is no lattice point strictly between v_1 and v_2, we deduce that $|v_1 - v_2| = 1$ and thus $\det(M) = \pm D$.

If $n = 3$, let δ_1 be the 2-dimensional integral simplex in $\mathbf{R}^2$ whose vertices are the origin, U_2 and U_3, where U_i $(2 \le i \le 3)$ is the last two coordinates of $V_i - V_1$. Since there are no non-vertex lattice points on δ, there are also no non-vertex lattice points on δ_1. But δ_1 has dimension 2. We deduce the stronger property that there are no non-vertex lattice points in the fundamental domain $\mathbf{R}^2(\mathrm{mod}\mathbf{Z}U_2 + \mathbf{Z}U_3)$. In fact, if $r_2U_2 + r_2U_3$ $(0 < r_i < 1)$ were a non-vertex lattice point in the fundamental domain of δ_1 but not in δ_1, then $r_2 + r_3 > 1$ and we deduce that its mirror $(1-r_2)U_2 + (1-r_3)U_3$ is an interior lattice point of δ_1 since $(1 - r_2) + (1 - r_3) < 1$. No non-vertex lattice point in the fundamental domain of the simplex δ_1 implies that $\det(U_2, U_3) = \pm 1$. It follows that

$$\det(V_1, V_2, V_3) = \det(V_1, V_2 - V_1, V_3 - V_1) = D\det(U_2, U_3) = \pm D.$$

If $n = 4$, the extra condition $D = 1$ shows that any lattice point

$$V = r_1V_1 + r_2V_2 + r_3V_3 + r_4V_4,\ 0 \le r_i < 1$$

in the fundamental domain of Δ has weight in the set $\{0, 1, 2, 3\}$. Since Δ is indecomposable, the fundamental domain has no non-zero lattice points of weight $\{0, 1\}$. The mirror construction

$$V \to (1 - r_1)V_1 + (1 - r_2)V_2 + (1 - r_3)V_3 + (1 - r_4)V_4$$

shows that the fundamental domain has no lattice points of weight 3. It follows that all non-zero lattice points in the fundamental domain has weight exactly 2. The weight function on $S_p(\Delta)$ is certainly stable under the p-action for every p, since there is only one non-trivial weight. Now, f is non-degenerate, we have $S_p(\Delta) = S(\Delta)$. Thus, the weight function on $S(\Delta)$ is also stable under p-action. The proof is complete.

2.4. Counter-examples in high dimensions. The discussion in the previous section shows that, in searching for diagonal counter-examples of the AS conjecture, we should look at those Δ with $D(\Delta) < d_n$. Furthermore, to insure that the diagonal family is already the generic family, we need to choose Δ to be **minimal**, i.e., there are no other lattice points on Δ other than the vertices. This is the approach taken in [18]. The following 5-dimensional counter-example is given in [18].

Let Δ be the 5-dimensional simplex in $\mathbf{R}^5$ whose non-zero vertices are the column vectors V_i $(1 \leq i \leq 5)$ of the following matrix:

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \tag{25}$$

Clearly, $D(\Delta) = 1$ since the first coordinate of each V_i is 1. One computes that the determinant of the matrix M has absolute value 3. A direct proof shows that there are no lattice points on Δ except for the vertices. Thus, the generic family is the diagonal family. In particular, $\mathrm{L}^*(f, T)$ is a polynomial of degree 3 if $p \neq 3$. There are only three lattice points contained in the fundamental domain $\mathbf{R}^5/\mathbf{Z}V_1 + \cdots + \mathbf{Z}V_5$. One of them is the origin. The other two are

$$(2, 1, 1, 1, 1) = \frac{2}{3}V_1 + \sum_{i=2}^{5} \frac{1}{3}V_i, \quad (3, 2, 2, 2, 2) = \frac{1}{3}V_1 + \sum_{i=2}^{5} \frac{2}{3}V_i. \tag{26}$$

These two lattice points have weight 2 and 3, respectively. One checks that if $p \equiv 2 \pmod 3$, the p-action permutes the two points in (26) and thus the weight function is not stable under the p-action. We conclude that the Newton polygon lies strictly above its lower bound if $p \equiv 2 \pmod 3$.

To construct counter-examples in higher dimensions $(n \geq 6)$, we can simply let Δ be the integral simplex in $\mathbf{R}^n$ generated by the origin and the column vectors of the $n \times n$ matrix

$$M^* = \begin{pmatrix} M & B \\ 0 & I_{n-5} \\ , & \end{pmatrix} \tag{27}$$

where M is the above 5×5 matrix, I_{n-5} is the identity matrix of order $(n-5)$, the first row of B is $(1, 1, \cdots, 1)$ and all other rows of B are zero. The first row of M^* is $(1, 1, \cdots, 1)$ and so $D(\Delta) = 1$. One checks that there are no lattice points on Δ other than the vertices. The generic family is indeed the diagonal family. However, if $p \equiv 2 \pmod 3$, the above argument shows that the Newton polygon of $\mathrm{L}^*(f, T)^{(-1)^{n-1}}$ lies strictly above its lower bound.

Counter-examples in dimension 4 are a little harder to construct and to prove. In fact, Corollary 2.9 shows that there are no such counter-examples if $D = 1$ and $n = 4$. Let $D \geq 2$ and $k \geq 2$ be any two given positive integers. Let f be the 4-variable diagonal Laurent polynomial

$$f(x) = a_1 x_1^D + a_2 x_1^D x_2 + a_3 x_1^D x_2 x_3 + a_4 x_1^D x_3^{-1} x_4^{D^k}, \tag{28}$$

where the coefficients are non-zero elements of $\mathbf{F}_q$. Thus, $\Delta = \Delta(f)$ is the 4-dimensional simplex in $\mathbf{R}^4$ whose non-zero vertices are the column vectors of the following matrix

$$M = \begin{pmatrix} D & D & D & D \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & D^k \end{pmatrix}.$$

It is clear that $D(\Delta) = D$, $d_4 = D^k$ and $\det(M) = D^{k+1}$. The lattice points in the fundamental domain $\mathbf{R}^4/\mathbf{Z}V_1 + \cdots + \mathbf{Z}V_4$ are of the form

$$\frac{j_1}{D^k}V_1 + \{\frac{D^k - j_4}{D^k}\}V_2 + \frac{j_4}{D^k}V_3 + \frac{j_4}{D^k}V_4, \tag{29}$$

where $\{x\}$ denotes the fractional part of the real number x, j_1 and j_4 are integers such that

$$0 \le j_1, j_4 \le D^k - 1,\ j_1 + j_4 \equiv 0 \ (\mathrm{mod} D^{k-1}). \tag{30}$$

There are exactly D^{k+1} choices in (30), corresponding to the number of lattice points in the fundamental domain. There are exactly D lattice points in the fundamental domain with weight smaller than 1 corresponding to the case $j_4 = 0$. There are no lattice points on Δ with weight 1 other than the vertices. The next smallest possible weight for a lattice point in the fundamental domain is $1 + \frac{1}{D}$. Take $j_4 = 1$ and $j_1 = D^{k-1} - 1$, then

$$u = (D+1, 1, 0, 1) = \frac{D^{k-1} - 1}{D^k}V_1 + \frac{D^k - 1}{D^k}V_2 + \frac{1}{D^k}V_3 + \frac{1}{D^k}V_4$$

is a lattice point in the fundamental domain with weight $1 + \frac{1}{D}$. The p-action of this point is given by

$$\{pu\} = \{\frac{p(D^{k-1} - 1)}{D^k}\}V_1 + \{\frac{p(D^k - 1)}{D^k}\}V_2 + \{\frac{p}{D^k}\}V_3 + \{\frac{p}{D^k}\}V_4.$$

The weight of this point is bounded from below by

$$w(\{pu\}) \ge 1 + \{\frac{p}{D^k}\}$$

since the sum of the middle two coefficients in (29) is already 1. This shows that the weight function is not stable under the p-action if

$$\{\frac{p}{D^k}\} > \frac{1}{D}.$$

The last condition is satisfied, for instance, if we take p such that $p - (1 + D^{k-1})$ is divisible by D^k. Such p satisfies the condition that $p - 1$ is divisible by D^{k-1} (in particular by D) but not divisible by D^k. Thus, the diagonal Laurent polynomial f in (28) is not ordinary for such p.

Although the Laurent polynomial in (28) is not yet generic in terms of Δ. It is "essentially" generic as far as the ordinary property of f is concerned. This is because there are no lattice points on Δ with weight 1 other than the non-zero vertices. Thus, the diagonal Laurent polynomial in (28) is the "leading form" of the generic Laurent polynomial with respect to Δ. The only terms we missed are those terms with weight strictly less than 1. Theorem 3.1 in next section shows that deformations by such "error terms" (with weight less than 1) have no effect on the ordinary property of f, although it would indeed change the Newton polygon in non-ordinary case. This shows that AS conjecture is false in dimension 4 as well. This counter-example together with the trick in (27) gives the following result.

Corollary 2.10. *Let $D^*(\Delta)$ be the smallest positive integer such that*

$$\mathrm{GNP}(\Delta, p) = \mathrm{HP}(\Delta)$$

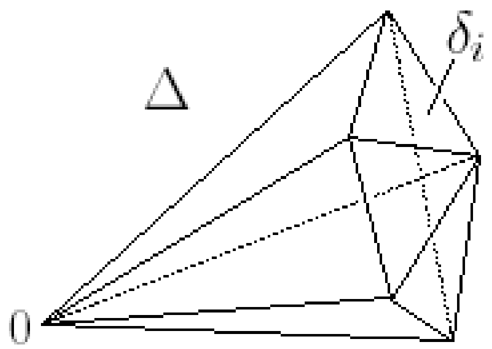
for all primes $p \equiv 1 \pmod{D^*(\Delta)}$. Then for each integer $n \geq 4$, we have

$$\lim_{\dim(\Delta)=n} \sup \frac{D^*(\Delta)}{D(\Delta)} = +\infty.$$

This shows that for each integer $n \geq 4$, the two numbers $D^*(\Delta)$ and $D(\Delta)$ can differ as much as one would like. Nevertheless, we shall see that $D^*(\Delta) = D(\Delta)$ in many important cases.

3. Global decomposition theory. The diagonal case is in principal well understood by the discussion in the previous section. We now turn to the general case. The aim of this section is to describe the basic facial decomposition theorem from [18] for the Newton polygon and the new collapsing decomposition theorem for the generic Newton polygon. We then deduce a number of applications. The full description and the proof of the collapsing decomposition will be given in the final section, since it is a little complicated and involve a more technical definition called the degree polygon.

3.1. Facial decomposition for the Newton polygon. The facial decomposition theorem described in this section allows us to determine when f is ordinary in certain non-diagonal cases.



Let $f(x)$ be a Laurent polynomial over $\mathbf{F}_q$ such that $\Delta(f) = \Delta$ is n-dimensional. We assume that f is non-degenerate. Let $\delta_1, \cdots, \delta_h$ be all the co-dimension 1 faces of Δ which do not contain the origin. Let f^{δ_i} be the restriction of f to the face δ_i. Then, $\Delta(f^{\delta_i}) = \Delta_i$ is n-dimensional. Furthermore, f is non-degenerate if and only if f^{δ_i} is non-degenerate for all $1 \leq i \leq h$. Since we assumed that f is non-degenerate, it follows that each f^{δ_i} is also non-degenerate.

The following facial decomposition theorem is taken from our paper [18].

Theorem 3.1. *(facial decomposition). Let f be non-degenerate and let $\Delta(f)$ be n-dimensional. Then f is ordinary if and only if each f^{δ_i} $(1 \leq i \leq h)$ is ordinary. Equivalently, f is non-ordinary if and only if some f^{δ_i} is non-ordinary.*

This theorem shows that as far as the ordinary property of f is concerned, we may assume that $\Delta(f)$ has only one face of co-dimension 1 not containing the origin. Combining Theorem 3.1 with the results of Section 2 (Theorem 2.5), we obtain

Corollary 3.2. *If each f^{δ_i} $(1 \leq i \leq h)$ is a diagonal Laurent polynomial over $\mathbf{Q}$, then we have a complete classification of the primes p such that f modulo p is ordinary.*

Together with Corollary 2.7, we deduce

Corollary 3.3. *Let $f(x)$ be a Laurent polynomial with rational coefficients and with $\Delta(f) = \Delta$. For a positive real number t, let $\pi_f(t)$ be the number of primes $p \leq t$ such that $f(x)$ is ordinary at p. Assume that f^{δ_i} is diagonal for each $1 \leq i \leq h$. Then, there is a positive rational number $r(\Delta) \leq 1$ such that we have the following asymptotic formula*

$$\pi_f(t) \sim r(\Delta) \frac{t}{\log t}.$$

Combining Theorem 3.1 with Corollary 2.6, we deduce

Corollary 3.4. *Let each f^{δ_i} $(1 \leq i \leq h)$ be a diagonal Laurent polynomial whose largest invariant factor is $d_n(i)$. Then f is ordinary if $p-1$ is divisible by the least common multiple* $\mathrm{lcm}(d_n(1), \cdots, d_n(h))$.

For a simple example with $h = 1$, let $f(x)$ be a polynomial over $\mathbf{F}_q$ of the form

$$f(x) = a_1 x_1^d + a_2 x_2^d + \cdots + a_n x_n^d + g(x),\ a_j \in \mathbf{F}_q^*, \tag{31}$$

where d is a positive integer and $g(x)$ is any polynomial of degree smaller than d. The polynomial $f(x)$ in (31) is a deformation of the Fermat polynomial. It is non-degenerate if and only if d is not divisible by p, in which case the L-function $L^*(f,T)^{(-1)^{n-1}}$ is a polynomial of degree d^n. Theorem 3.1 shows that $f(x)$ is ordinary over $\mathbf{F}_q$ if and only if the leading diagonal form $f^\delta = a_1 x_1^d + \cdots + a_n x_n^d$ is ordinary over $\mathbf{F}_q$. This last condition holds if and only if $p-1$ is divisible by d.

For a simple example with $h > 1$, let $f(x)$ be the n-dimensional generalized Kloosterman polynomial

$$f(x) = a_1 x_1 + \cdots + a_n x_n + a_{n+1} \frac{1}{x_1^{v_1} \cdots x_n^{v_n}},\ a_j \neq 0, \tag{32}$$

where each v_j is a positive integer. The polynomial $f(x)$ in (32) is non-degenerate if and only if none of the v_j is divisible by p, in which case the L-function $L^*(f,T)^{(-1)^{n-1}}$ is a polynomial of degree $1 + v_1 + \cdots + v_n$. Let δ_i be the $(n-1)$-dimensional simplex formed by all the exponents of $f(x)$ with the i-th exponent removed, where $1 \leq i \leq n+1$. Let Δ_i be the n-dimensional simplex generated by the origin and δ_i. The invariant factors of Δ_{n+1} are all 1. Thus, $f^{\delta_{n+1}}$ is ordinary for all p. For $1 \leq i \leq n$, the invariant factors of Δ_i are given by

$$\{1, \cdots, 1, v_i\},$$

and thus f^{δ_i} is ordinary if $p-1$ is divisible by v_i. We conclude that the generalized Kloosterman polynomial $f(x)$ in (32) is ordinary for all p such that $p-1$ is divisible by $\mathrm{lcm}(v_1, \cdots, v_n)$. This result was first proved by Sperber [15] for large p, see also [14] for the classical case $v_1 = \cdots = v_n = 1$ and $p > n+3$.

Similarly, we can consider the Laurent polynomial

$$f(x) = a_1 x_1 + \cdots + a_n x_n + a_{n+1} x_1^{v_1} \cdots x_n^{v_n},\ a_j \neq 0, v_j > 0, \tag{33}$$

or by making the invertible change of variables $x_i \to x_i^{-1}$, we get the new equivalent Laurent polynomial

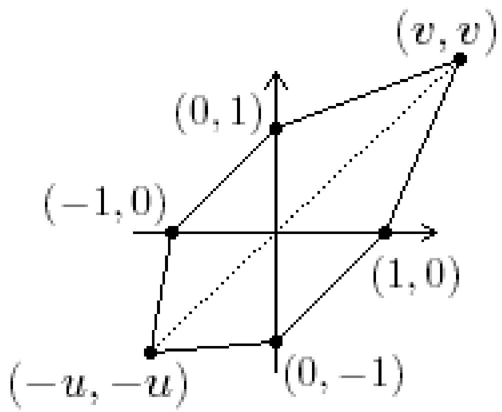
$$f(x) = a_1 \frac{1}{x_1} + \cdots + a_n \frac{1}{x_n} + a_{n+1} \frac{1}{x_1^{v_1} \cdots x_n^{v_n}},\ a_j \neq 0, v_j > 0. \tag{34}$$

The polynomial $f(x)$ in (33)-(34) is non-degenerate if and only if none of the v_j is divisible by p. The L-function $L^*(f,T)^{(-1)^{n-1}}$ in this case is a polynomial of degree $v_1 + \cdots + v_n$. It is ordinary if $p-1$ is divisible by $\mathrm{lcm}(v_1, \cdots, v_n)$.

For another example with $h > 1$, we consider the sum of two polynomials from (33)-(34). Namely, let $f(x)$ be the n-dimensional ($n > 1$) generalized bi-Kloosterman polynomial

$$\begin{aligned} f(x) = a_1 x_1 + \cdots + a_n x_n + a_{n+1} \frac{1}{x_1^u \cdots x_n^u} \\ + b_1 \frac{1}{x_1} + \cdots + b_n \frac{1}{x_n} + b_{n+1} x_1^v \cdots x_n^v, \end{aligned} \tag{35}$$

where the coefficients are non-zero and the u, v are positive integers. The exponential sum associated to this polynomial (in the special case $u = v = 1$) arises from Kim's [12] calculation of the Gauss sum for the unitary group. We claim that $\Delta(f)$ has exactly $(2^n - 2) + 2n = 2^n + 2n - 2$ co-dimension 1 faces δ_i not containing the origin. This is proved as follows. If δ_i contains the vertex $(v, \cdots, v)$, then the other $(n-1)$ vertices of δ_i must be chosen from the n positive unit coordinate vectors. There are n such choices. If δ_i contains the vertex $(-u, \cdots, -u)$, then the other $(n-1)$ vertices of δ_i must be chosen from the n negative unit coordinate vectors. There are also n such choices. If δ_i contains neither the vertex $(-u, \cdots, -u)$ nor the vertex $(v, \cdots, v)$, then δ_i contains exactly k ($0 < k < n$) positive unit coordinate vectors and the remaining $(n-k)$ negative unit coordinate vectors. There are $2^n - 2$ such choices. Thus, there are $(2^n - 2) + 2n$ possibilities in total. The claim is proved.



The largest invariant factors (in fact the absolute values of the determinants) of these faces are given respectively, by v (n of them), u (n of them) and 1 (2^n-2 of them). Thus, the polynomial $f(x)$ in (35) is non-degenerate if and only if none of the u, v is divisible by p. The L-function $L^*(f,T)^{(-1)^{n-1}}$ in this case is a polynomial of degree $2^n-2+nu+nv$. Corollary 3.4 shows that f is ordinary if $p-1$ is divisible by $\mathrm{lcm}(u,v)$. In particular, if all $u=v=1$ as in [12], then f is ordinary for every prime p.

We now turn to some easy applications of Theorem 3.1 to the AS conjecture. In the case $n=1$, Corollary 3.4 with $h\le 2$ immediately implies

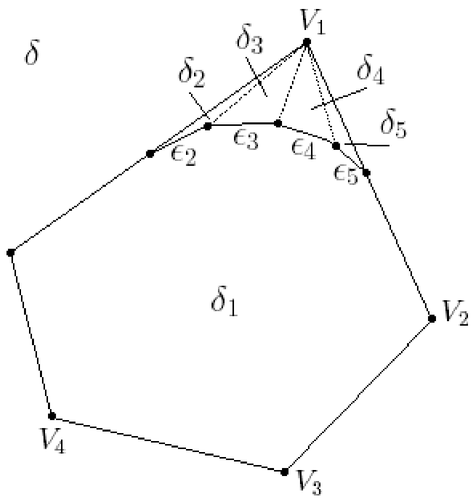
Corollary 3.5. *The AS conjecture holds for $n=1$.*

In the case $n=2$, if Δ has only one co-dimension 1 face not containing the origin, then Δ must be a simplex. This fact together with Theorem 3.1 gives the following weaker version of the AS conjecture for $n\le 2$.

Corollary 3.6. *Theorem 1.6 holds if $n\le 2$.*

To treat the full form of the AS conjecture even in the case $n=2$ or its weaker version in higher dimension $n\ge 3$, we need to introduce further decomposition theorems. We will describe the new collapsing decomposition in next subsection.

3.2. Collapsing decomposition for generic Newton polygon. We will assume that $n\ge 2$ as the case $n=1$ is already handled by the facial decomposition. Let $\mathcal{V}=\{V_1,\cdots,V_J\}$ be the set of J fixed lattice points in $\mathbf{R}^n$. Let Δ be the convex polyhedron in $\mathbf{R}^n$ generated by the origin and the lattice points in $\mathcal{V}$. We assume that Δ is n-dimensional. By the facial decomposition, we may assume that Δ has only one co-dimension 1 face δ not containing the origin and all $V_j\in\delta$. To decompose Δ, we will decompose the unique face δ. Actually, we will be decomposing the set $\mathcal{V}$ since we are working in a little more general setting. In the special case when $\mathcal{V}$ consists of all lattice points in δ, then our decomposition can be described purely in terms of δ. Clearly, the set $\mathcal{V}$ has at least n elements. If the set $\mathcal{V}$ has exactly n elements, then the set $\mathcal{V}$ is called **indecomposable**.



Choose an element in $\mathcal{V}$ which is a vertex of δ, say V_1. The collapsing decomposition of $\mathcal{V}$ with respect to V_1 is simply the convex decomposition of $\mathcal{V}$ resulted from the collapsing after removing the vertex V_1. We now describe the collapsing decomposition more precisely. Let $\mathcal{V}_1 = \mathcal{V} - \{V_1\}$ be the complement of V_1 in $\mathcal{V}$. Let δ_1 be the convex polytope of the lattice points in $\mathcal{V}_1$. This is a subset of δ. Let δ_1' be the toplogical closure of $\delta - \delta_1$. This is not a convex polyhedron in general. The intersection $\delta_1 \cap \delta_1'$ consists of finitely many different $(n-2)$-dimensional faces $\{\epsilon_2, \cdots, \epsilon_h\}$ of δ_1. Let δ_i $(2 \le i \le h)$ be the convex closure of ϵ_i and V_1. Then, each δ_i is $(n-1)$-dimensional. Let $\mathcal{V}_i$ $(2 \le i \le h)$ be the intersection $\mathcal{V} \cap \delta_i$. Then, each V_j lies in at least one (possibly more) of the subsets $\mathcal{V}_i$ of $\mathcal{V}$. The collapsing decomposition of $\mathcal{V}$ with respect to V_1 is defined to be

$$\mathcal{V} = \bigcup_{i=1}^{h} \mathcal{V}_i. \tag{36}$$

The full collapsing decomposition theorem for the degree polygon applies to the decomposition in (36). To describe it only in terms of Newton polygon, we must further reduce to indecomposable diagonal situation, in which case the degree polygon coincides with its upper bound if and only if the Newton polygon coincides with its lower bound. And thus, we will be able to replace the degree polygon by the simpler Newton polygon.

Applying the collapsing decomposition in (36) to each $\mathcal{V}_i$ by choosing a vertex of $\mathcal{V}_i$, by induction, we will eventually be able to decompose $\mathcal{V}$ as a finite union of indecomposable ones:

$$\mathcal{V} = \bigcup_{i=1}^{m} \mathcal{W}_i, \tag{37}$$

where each $\mathcal{W}_i$ has exactly n elements and thus indecomposable. Furthermore, the subsets $\mathcal{W}_i$'s are different although they may have non-empty intersections. The complete decomposition in (37) resulting from a sequence of collapsing decompositions is called a **complete collapsing decomposition** of $\mathcal{V}$. Note that such a complete collapsing decomposition is in general not unique because of the choice of a vertex in each stage of the collapsing construction.

Let f be the generic Laurent polynomial associated to the set $\mathcal{V}$:

$$f = \sum_{j=1}^{J} a_j x^{V_j}. \tag{38}$$

Let the $\mathcal{W}_i$ $(1 \le i \le m)$ be a complete collapsing decomposition of $\mathcal{V}$ as in (37). For each integer $1 \le i \le m$, let f_i be the restriction of f to $\mathcal{W}_i$:

$$f_i = \sum_{V_j \in \mathcal{W}_i} a_j x^{V_j}. \tag{39}$$

This is the generic Laurent polynomial associated to the indecomposable $\mathcal{W}_i$. We have

THEOREM 3.7. *(collapsing decomposition). Let the $\mathcal{W}_i$ $(1 \le i \le m)$ be a complete collapsing decomposition of $\mathcal{V}$. If each f_i $(1 \le i \le m)$ is generically non-degenerate and ordinary for some prime p, then f is also generically non-degenerate and ordinary for the same prime p.*

Since each $\mathcal{W}_i$ is indecomposable, each f_i is the generic diagonal Laurent polynomial whose exponents are in $\mathcal{W}_i$. As a consequence of Theorem 3.7 and Corollary 2.6, we obtain

Corollary 3.8. *Let M_i be the non-singular $n \times n$ matrix whose column vectors are the elements of $\mathcal{W}_i$. Let $d_n(i)$ be the largest invariant factor of the matrix M_i. Let*

$$D^* = \mathrm{lcm}\{d_n(1), d_n(2), \cdots, d_n(m)\}. \tag{40}$$

If $p \equiv 1 \pmod{D^*}$, then f is generically ordinary for the prime p.

The number D^* in (40) may depend on our choice of the complete collapsing decomposition. Although any choice would yield a good upper bound for the smallest possible D^*, we do not know a general algorithm which tells us how to choose a best complete collapsing decomposition. Taking $\mathcal{V}$ to be the full set of lattice points in δ, Corollary 3.8 reduces to

Corollary 3.9. *The weaker AS conjecture as stated in Theorem 1.6 is true in every dimension.*

Taking $n \leq 3$ or $n = 4$ with $D = 1$ in Corollary 3.8, we deduce from Corollary 2.9, the following full form of the AS conjecture.

Corollary 3.10. *The AS conjecture is true if $n \leq 3$ or if $n = 4$ but $D = 1$ and $p > n!\mathbf{V}(\Delta)$.*

In this corollary, we inserted the extra condition on p in the case that $D = 1$ simply to avoid the case that some of the diagonal pieces in a complete collapsing decomposition might become generically degenerate, that is, p might divide the largest invariant factor of $\mathcal{W}_i$ for some i if p is small. This does not happen in the case $n \leq 3$ unless the generic total family f is already degenerate, but it can happen if $n \geq 4$. In the zeta function case, the integer D is always 1. For example, applying Corollary 3.10 to the 3-variable polynomial of the form $x_3 f(x_1, x_2)$, we can deduce

Corollary 3.11. *Let Δ be a 2-dimensional integral convex polyhedron in $\mathbf{R}^2$. Then, for every prime p, the zeta function of the affine toric curve $f(x_1, x_2) = 0$ ($x_i \neq 0$) is ordinary for a generic non-degenerate f with $\Delta(f) = \Delta$.*

If we apply Corollary 3.10 to the polynomial of the form $x_4 f(x_1, x_2, x_3)$ in 4 variables, we can deduce

Corollary 3.12. *Let Δ be a 3-dimensional integral convex polyhedron in $\mathbf{R}^3$. For every prime $p > n!\mathbf{V}(\Delta)$, the zeta function of the affine toric surface $f(x_1, x_2, x_3) = 0$ ($x_i \neq 0$) is ordinary for a generic non-degenerate f with $\Delta(f) = \Delta$.*

To translate these results to the usual affine hypersurface case, we can use a simple boundary argument (see [18]). This gives

Corollary 3.13. *Let Δ be a 2-dimensional integral convex polyhedron in the first quadrant of $\mathbf{R}^2$. Assume that Δ contains a non-zero vertex on both the x_1 axis and the x_2-axis. Then, for every prime p, the zeta function of the affine curve $f(x_1, x_2) = 0$ is ordinary for a generic non-degenerate f with $\Delta(f) = \Delta$.*

Corollary 3.14. *Let Δ be a 3-dimensional integral convex polyhedron in the first quadrant of $\mathbf{R}^3$. Assume that Δ contains a non-zero vertex on each of the x_i-axis*

($1 \leq i \leq 3$). Then, for every prime $p > n!\mathbf{V}(\Delta)$, the zeta function of the affine surface $f(x_1, x_2, x_3) = 0$ is ordinary for a generic non-degenerate f with $\Delta(f) = \Delta$.

These results significantly generalize the well known fact that a generic non-degenerate plane curve of degree d is ordinary. Corollaries 3.11-14, however, do not generalize to higher dimensional Δ. In fact, the 5-dimensional counter-example in Section 2.4 shows that the higher dimensional generalization of Corollary 3.12 is already false for the 3-dimensional generic toric affine hypersurface $f(x_1, \cdots, x_4)$ with $\Delta(f) = \Delta$ for some 4-dimensional Δ. The higher dimensional analogue is true for certain Δ which is sufficiently "regular" in a suitable sense. For example, the hyperplane decomposition theorem in [18] implies that a generic non-degenerate affine hypersurface of degree d is ordinary (affine version of Mazur's conjecture). At this point, we do not know how to derive this result using the collapsing decomposition theorem.

Applying Corollary 3.13 to the following family of hyper-elliptic curves of genus g:

$$y^2 + (\sum_{i=0}^{g} a_i)y = \sum_{j=0}^{2g+1} b_j x^j,$$

one obtains

COROLLARY 3.15. *For every genus g, the universal family of hyper-elliptic curves of genus g is generically ordinary for every p.*

COROLLARY 3.16. *For every genus g, the universal family of smooth projective curves of genus g is generically ordinary for every p.*

4. Newton polygons of Fredholm determinants. All our decomposition theorems in [18] are on the chain level, All except for the facial decomposition theorem are proved for a certain degree polygon which is finer than the generic Newton polygon. Thus, to fully describe the collapsing decomposition theorem, we need to review Dwork's trace formula expressing the L-function in terms of the Fredholm determinant of a certain infinite Frobenius matrix. This gives the necessary background to work on the chain level, which is actually more flexible for our purpose than on the cohomology level. In the ordinary case, the descent result given in Theorem 4.5 allows us to replace the infinite Frobenius matrix by a much more manageable infinite matrix. With these preparations, we can then formulate the chain level versions of our problems, which are easier to work with and which include singular cases as well.

4.1. Dwork's trace formula. Let $\mathbf{Q}_p$ be the field of p-adic numbers. Let Ω be the completion of an algebraic closure of $\mathbf{Q}_p$. Let $q = p^a$ for some positive integer a. Denote by "ord" the additive valuation on Ω normalized by ord p=1, and denote by "ord$_q$" the additive valuation on Ω normalized by ord$_q q$=1. Let $\mathbf{K}$ denote the unramified extension of $\mathbf{Q}_p$ in Ω of degree a. Let $\Omega_1 = \mathbf{Q}_p(\zeta_p)$, where ζ_p is a primitive p-th root of unity. Then Ω_1 is the totally ramified extension of $\mathbf{Q}_p$ of degree $p-1$. Let Ω_a be the compositum of Ω_1 and $\mathbf{K}$. Then Ω_a is an unramified extension of Ω_1 of degree a. The residue fields of Ω_a and $\mathbf{K}$ are both $\mathbf{F}_q$, and the residue fields of Ω_1 and $\mathbf{Q}_p$ are both $\mathbf{F}_p$. Let π be a fixed element in Ω_1 satisfying

$$\sum_{m=0}^{\infty} \frac{\pi^{p^m}}{p^m} = 0, \quad \mathrm{ord}\pi = \frac{1}{p-1}.$$

Then, π is a uniformizer of $\Omega_1 = \mathbf{Q}_p(\zeta_p)$ and we have

$$\Omega_1 = \mathbf{Q}_p(\pi).$$

The Frobenius automorphism $x \mapsto x^p$ of $\mathrm{Gal}(\mathbf{F}_q/\mathbf{F}_p)$ lifts to a generator τ of $\mathrm{Gal}(\mathbf{K}/\mathbf{Q}_p)$ which is extended to Ω_a by requiring that $\tau(\pi) = \pi$. If ζ is a $(q-1)$-st root of unity in Ω_a, then $\tau(\zeta) = \zeta^p$.

Let $E(t)$ be the Artin-Hasse exponential series:

$$\begin{aligned} E(t) &= \exp\Big(\sum_{m=0}^{\infty} \frac{t^{p^m}}{p^m}\Big) \\ &= \prod_{k\geq 1,(k,p)=1} (1-t^k)^{\mu(k)/k}, \end{aligned}$$

where $\mu(k)$ is the Möbius function. The last product expansion shows that the power series $E(t)$ has p-adic integral coefficients. Thus, we can write

$$E(t) = \sum_{m=0}^{\infty} \lambda_m t^m, \quad \lambda_m \in \mathbf{Z}_p.$$

For $0 \leq m \leq p-1$, more precise information is given by

$$\lambda_m = \frac{1}{m!}, \ \mathrm{ord}\lambda_m = 0, \ 0 \leq m \leq p-1. \tag{41}$$

The shifted series

$$\theta(t) = E(\pi t) = \sum_{m=0}^{\infty} \lambda_m \pi^m t^m \tag{42}$$

is a splitting function in Dwork's terminology. The value $\theta(1)$ is a primitive p-th root of unity, which will be identified with the p-th root of unit ζ_p used in our definition of the exponential sums as given in the introduction.

For a Laurent polynomial $f(x_1, \cdots, x_n) \in \mathbf{F}_q[x_1, x_1^{-1}, \cdots, x_n, x_n^{-1}]$, we write

$$f = \sum_{j=1}^{J} \bar{a}_j x^{V_j}, \ V_j \in \mathbf{Z}^n, \ \bar{a}_j \in \mathbf{F}_q^*.$$

Let a_j be the Teichmüller lifting of $\bar{a}_j$ in Ω. Thus, we have $a_j^q = a_j$. Set

$$F(f,x) = \prod_{j=1}^{J} \theta(a_j x^{V_j}) \tag{43}$$

$$F_a(f,x) = \prod_{i=0}^{a-1} F^{\tau^i}(f, x^{p^i}). \tag{44}$$

Note that (42) implies that $F(f,x)$ and $F_a(f,x)$ are well defined as formal Laurent series in $x_1, \cdots, x_n$ with coefficients in Ω_a.

To describe the growth conditions satisfied by F, write

$$F(f,x) = \sum_{r \in \mathbf{Z}^n} F_r(f) x^r.$$

Then from (42) and (43), one checks that

$$F_r(f) = \sum_u \Big(\prod_{j=1}^{J} \lambda_{u_j} a_j^{u_j}\Big) \pi^{u_1+\cdots+u_J}, \tag{45}$$

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

$$\sum_{j=1}^{J} u_j V_j = r, \ u_j \geq 0, \ u_j \text{ integral.} \tag{46}$$

Thus, $F_r(f) = 0$ if (46) has no solutions. Otherwise, (45) implies that

$$\mathrm{ord} F_r(f) \geq \frac{1}{p-1} \inf_u \{\sum_{j=1}^{J} u_j\},$$

where the inf is taken over all solutions of (46).

For a given point $r \in \mathbf{R}^n$, recall that the weight $w(r)$ is given by

$$w(r) = \inf_u \{\sum_{j=1}^{J} u_j | \sum_{j=1}^{J} u_j V_j = r, u_j \geq 0, \ u_j \in \mathbf{R}\},$$

where the weight $w(r)$ is defined to be ∞ if r is not in the cone generated by Δ and the origin. Thus,

$$\mathrm{ord} F_r(f) \geq \frac{w(r)}{p-1}, \tag{47}$$

with the obvious convention that $F_r(f) = 0$ if $w(r) = +\infty$.

Let $C(\Delta)$ be the closed cone generated by the origin and Δ. Let $L(\Delta)$ be the set of lattice points in $C(\Delta)$. That is,

$$L(\Delta) = \mathbf{Z}^n \cap C(\Delta). \tag{48}$$

For real numbers b and c with $0 < b \leq p/(p-1)$, define the following two spaces of p-adic functions:

$$\mathcal{L}(b,c) = \{ \sum_{r \in L(\Delta)} C_r x^r \mid C_r \in \Omega_a, \ \mathrm{ord}_p C_r \geq b w(r) + c\}$$

$$\mathcal{L}(b) = \bigcup_{c \in \mathbf{R}} \mathcal{L}(b,c).$$

One checks from (47) that

$$F(f,x) \in \mathcal{L}(\frac{1}{p-1}, 0), \quad F_a(f,x) \in \mathcal{L}(\frac{p}{q(p-1)}, 0).$$

Define an operator ψ on formal Laurent series by

$$\psi(\sum_{r\in L(\Delta)} C_r x^r) = \sum_{r\in L(\Delta)} C_{pr} x^r.$$

It is clear that

$$\psi(\mathcal{L}(b,c)) \subset \mathcal{L}(pb,c).$$

It follows that the composite operator $\phi_a = \psi^a \circ F_a(f,x)$ is an Ω_a-linear endomorphism of the space $\mathcal{L}(b)$, where $F_a(f,x)$ denotes the multiplication map by the power series $F_a(f,x)$. Similarly, the operator $\phi_1 = \tau^{-1}\circ\psi\circ F(f,x)$ is an Ω_a-semilinear (τ^{-1}-linear) endomorphsim of the space $\mathcal{L}(b)$. The operators ϕ_a^m and ϕ_1^m have well defined traces and Fredholm determinants. The Dwork trace formula asserts that for each positive integer k,

$$\mathrm{S}_k^*(f) = (q^k-1)^n \mathrm{Tr}(\phi_a^k).$$

In terms of L-function, this can be reformulated as follow.

Theorem 4.1. *We have*

$$\mathrm{L}^*(f,T)^{(-1)^{n-1}} = \prod_{i=0}^n \det(I - Tq^i\phi_a)^{(-1)^i\binom{n}{i}}.$$

The L-function is determined by the single determinant $\det(I - T\phi_a)$. For explicit calculations, we shall describe the operator α_a in terms of an infinite nuclear matrix. First, one checks that

$$\begin{aligned}\phi_1^a &= \phi_1^{a-2}\tau^{-1}\circ\psi\circ F(f,x)\circ\tau^{-1}\circ\psi\circ F(f,x)\\ &= \phi_1^{a-2}\tau^{-1}\circ\psi\circ\tau^{-1}\circ F^\tau(f,x)\circ\psi\circ F(f,x)\\ &= \phi_1^{a-2}(\tau^{-1})^2\circ\psi^2\circ F^\tau(f,x^p)\circ F(f,x)\\ &= \cdots = \psi^a\circ F_a(f,x) = \phi_a.\end{aligned}$$

We now describe the matrix form of the operators ϕ_1 and ϕ_a with respect to some orthonormal basis. Fix a choice $\pi^{1/D}$ of D-th root of π in Ω. Define a space of functions

$$\mathcal{B} = \{\sum_{r\in L(\Delta)} C_r\pi^{w(r)}x^r \mid C_r\in\Omega_a(\pi^{1/D}),\ C_r\to 0 \text{ as } |r|\to\infty\}.$$

Then, the monomials $\pi^{w(r)}x^r$ ($r\in L(\Delta)$) form an orthonormal basis of the p-adic Banach space $\mathcal{B}$. Furthermore, if $b > 1/(p-1)$, then $\mathcal{L}(b)\subseteq\mathcal{B}$. The operator ϕ_a (resp. ϕ_1) is an Ω_a-linear (resp. Ω_a-semilinear) nuclear endomorphsim of the space $\mathcal{B}$. Let Γ be the orthonormal basis $\{\pi^{w(r)}x^r\}_{r\in L(\Delta)}$ of $\mathcal{B}$ written as a column vector. One checks that the operator ϕ_1 is given by

$$\phi_1\Gamma = A_1(f)^{\tau^{-1}}\Gamma,$$

where $A_1(f)$ is the infinite matrix whose rows are indexed by r and columns are indexed by s. That is,

$$A_1(f) = (a_{r,s}(f)) = (F_{ps-r}(f)\pi^{w(r)-w(s)}). \tag{49}$$

We note that the row index r and column index s of $A_1(f)$ were switched in [18]. Thus, the matrix there needs to be transposed. This is corrected here.

Since $\phi_a = \phi_1^a$ and ϕ_1 is τ^{-1}-linear, the operator ϕ_a is given by

$$\begin{aligned}\phi_a\Gamma &= \phi_1^a\Gamma \\ &= \phi_1^{a-1}A_1^{\tau^{-1}}\Gamma \\ &= \phi_1^{a-2}A_1^{\tau^{-2}}A_1^{\tau^{-1}}\Gamma \\ &= A_1^{\tau^{-a}}\cdots A_1^{\tau^{-2}}A_1^{\tau^{-1}}\Gamma \\ &= A_1A_1^{\tau}\cdots A_1^{\tau^{a-2}}A_1^{\tau^{a-1}}\Gamma.\end{aligned}$$

Let

$$A_a(f) = A_1A_1^{\tau^1}\cdots A_1^{\tau^{a-1}}. \tag{50}$$

Then, the matrix of ϕ_a under the basis Γ is $A_a(f)$. We call $A_1(f) = (a_{r,s}(f))$ the infinite semilinear Frobenius matrix and $A_a(f)$ the infinite linear Frobenius matrix. Dwork's trace formula can now be rewritten in terms of the matrix $A_a(f)$ as follows.

$$\mathrm{L}^*(f,T)^{(-1)^{n-1}} = \prod_{i=0}^{n}\det(I - Tq^iA_a(f))^{(-1)^i\binom{n}{i}}. \tag{51}$$

We are reduced to understanding the single determinant $\det(I - TA_a(f))$.

4.2. Newton polygons of Fredholm determinants. To get a lower bound for the Newton polygon of $\det(I - TA_a(f))$, we need to estimate the entries of the infinite matrices $A_1(f)$ and $A_a(f)$. By (47) and (49), we obtain the estimate

$$\mathrm{ord}a_{r,s}(f) \geq \frac{w(ps-r)+w(r)-w(s)}{p-1} \geq w(s). \tag{52}$$

Recall that for a positive integer k, $W_\Delta(k)$ is defined to be the number of lattice points in $L(\Delta)$ with weight exactly $k/D(\Delta)$:

$$W_\Delta(k) = \mathrm{card}\{r \in L(\Delta) \mid w(r) = \frac{k}{D(\Delta)}\}.$$

Let $\xi \in \Omega$ be such that

$$\xi^D = \pi^{p-1}.$$

Then

$$\mathrm{ord}\xi = 1/D.$$

By (52), the infinite matrix $A_1(f)$ has the block form

$$A_1(f) = \begin{pmatrix} A_{00} & \xi^1A_{01} & \dots & \xi^iA_{0i} & \dots \\ A_{10} & \xi^1A_{11} & \dots & \xi^iA_{1i} & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ A_{i0} & \xi^1A_{i1} & \dots & \xi^iA_{ii} & \dots \\ \vdots & \vdots & \ddots & \vdots & \end{pmatrix}, \tag{53}$$

where the block A_{ij} is a finite matrix of $W_\Delta(i)$ rows and $W_\Delta(j)$ columns whose entries are p-adic integers in Ω.

Definition 4.2. *Let $P(\Delta)$ be the polygon in $\mathbf{R}^2$ with vertices $(0,0)$ and*

$$\Big(\sum_{k=0}^{m} W_\Delta(k), \frac{1}{D(\Delta)}\sum_{k=0}^{m} kW_\Delta(k)\Big), \quad m = 0, 1, 2, \cdots .$$

This is the chain level version of the Hodge polygon. The block form in (53) and the standard determinant expansion of the Fredholm determinant shows that we have

Proposition 4.3. *The Newton polygon of det$(I - TA_1(f))$ computed with respect to p lies above the polygon $P(\Delta)$.*

Using the block form (53) and the exterior power construction of a semi-linear operator, one then gets the following lower bound of Adolphson and Sperber [1] for the Newton polygon of $\det(I - TA_a(f))$.

Proposition 4.4. *The Newton polygon of det$(I - TA_a(f))$ computed with respect to q lies above the polygon $P(\Delta)$.*

4.3. A descent theorem. For the application to L-function, we need to use the linear Frobenius matrix $A_a(f)$ instead of the simpler semi-linear Frobenius matrix $A_1(f)$. In general, the Newton polygon of $\det(I - TA_a(f))$ computed with respect to q is different from the Newton polygon of $\det(I - TA_1(f))$ computed with respect to p, even though they have the same lower bound. Since the matrix $A_a(f)$ is much more complicated than $A_1(f)$, especially for large a, we would like to replace $A_a(f)$ by the simpler matrix $A_1(f)$. This is not possible in general. However, if we are only interested in the question whether the Newton polygon of $\det(I - TA_a(f))$ coincides with its lower bound, the following theorem shows that we can descend to the simpler $\det(I - TA_1(f))$. This result is taken from our paper [18].

Theorem 4.5. *The Newton polygon of det$(I - TA_a(f))$ computed with respect to q coincides with $P(\Delta)$ if and only if the Newton polygon of det$(I - TA_1(f))$ computed with respect to p coincides with $P(\Delta)$.*

Proof. This theorem can also be proved easily using exterior power construction. We include another proof following [18] as we need to transpose the matrices of [18] due to the switch of the row and column indices as indicated above. This proof gives additional information, namely, the Hodge-Newton decomposition which is useful for further investigation.

The proof is by induction on the sides. The first (slope zero) side of the Newton polygon of $\det(I - TA_1(f))$ coincides with the first side of $P(\Delta)$ if and only if $\det A_{00} \not\equiv 0 \pmod{\xi}$. To simplify notations, we use the convention that mod ξ (or mod p) means the reduction modulo the maximal ideal in the ring of integers in Ω. By (50) and (53), we can write

$$A_a(f) \equiv \begin{pmatrix} A_{00}\cdots A_{00}^{\tau^{a-1}} & 0 \\ * & 0 \end{pmatrix} (\mathrm{mod}\xi). \tag{54}$$

It follows that the first side of the Newton polygon of $\det(I - TA_a(f))$ coincides with the first side of $P(\Delta)$ if and only if $\det A_{00} \not\equiv 0 \pmod{\xi}$. Thus, the theorem is true for the first side.

Assume that $\det A_{00} \not\equiv 0 \pmod{\xi}$. We now use a triangulation procedure to prove the theorem for the first two sides. Form an elementary matrix B (to be determined) as follows:

$$B = \begin{pmatrix} I_{00} & 0 & \dots & 0 & \dots \\ B_{10} & I_{11} & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ B_{i0} & 0 & \dots & I_{ii} & \dots \\ \vdots & \vdots & \ddots & \vdots & \end{pmatrix}.$$

Its inverse is

$$B^{-1} = \begin{pmatrix} I_{00} & 0 & \dots & 0 & \dots \\ -B_{10} & I_{11} & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ -B_{i0} & 0 & \dots & I_{ii} & \dots \\ \vdots & \vdots & \ddots & \vdots & \end{pmatrix},$$

where B_{i0} is a matrix of $W_\Delta(i)$ rows and $W_\Delta(0)$ columns, and I_{ii} is the identity matrix of order $W_\Delta(i)$. Now the matrix $BA_1(f)B^{-\tau}$ has the form

$$BA_1(f)B^{-\tau} = \begin{pmatrix} C_{00} & \xi^1 C_{01} & \dots & \xi^i C_{0i} & \dots \\ C_{10} & \xi^1 C_{11} & \dots & \xi^i C_{1i} & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ C_{i0} & \xi^1 C_{i1} & \dots & \xi^i C_{ii} & \dots \\ \vdots & \vdots & \ddots & \vdots & \end{pmatrix}, \tag{55}$$

where

$$\begin{aligned} C_{00} &= A_{00} - \xi^1 A_{01} B_{10}^\tau - \cdots - \xi^k A_{0k} B_{k0}^\tau - \cdots, \\ C_{i0} &= B_{i0} A_{00} + A_{i0} - \xi^1 \{B_{i0} A_{01} + A_{i1}\} B_{10}^\tau - \cdots \\ &\quad -\xi^k \{B_{i0} A_{0k} + A_{ik}\} B_{k0}^\tau - \cdots \ (i > 0), \\ C_{0j} &= A_{0j}, \\ C_{ij} &= B_{i0} A_{0j} + A_{ij} \ (i \geq 1, j \geq 1). \end{aligned} \tag{56}$$

We want to choose suitable matrices B_{i0} such that

$$C_{i0} = 0 \quad \text{for all } i \geq 1. \tag{57}$$

Since A_{00} is invertible, by (56), equation (57) can be rewritten as

$$\begin{aligned} B_{i0} = -A_{i0} A_{00}^{-1} &+ \xi^1 \{B_{i0} A_{01} + A_{i1}\} B_{10}^\tau A_{00}^{-1} + \cdots \\ &+ \xi^k \{B_{i0} A_{0k} + A_{ik}\} B_{k0}^\tau A_{00}^{-1} + \cdots. \end{aligned} \tag{58}$$

Thus, by successive iteration of (58) or by the fixed point theorem for a contraction map in a p-adic Banach space, we deduce that there are solution matrices B_{i0} of (58) whose elements are p-adic integers invariant under the action of τ^a.

If we go through the above triangulation procedure again but ignore the action of τ, we conclude that there is a similar elementary matrix B_1, whose entries are p-adic integers, such that

$$B_1 A_1(f) B_1^{-1} = \begin{pmatrix} D_{00} & \xi^1 D_{01} & \dots & \xi^i D_{0i} & \dots \\ 0 & \xi^1 D_{11} & \dots & \xi^i D_{1i} & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & \xi^1 D_{i1} & \dots & \xi^i D_{ii} & \dots \\ \vdots & \vdots & \ddots & \vdots & \end{pmatrix}, \tag{59}$$

where D_{ij} is a matrix of the same size as C_{ij}. It follows from (56) that

$$C_{ij} \equiv D_{ij} \ (\mathrm{mod} \xi). \tag{60}$$

Comparing (55) with (59), we conclude that the Newton polygon of $\det(I - TA_1(f))$ coincides with $P(\Delta)$ if and only if the Newton polygon of $\det(I - TBA_1(f)B^{-\tau})$ coincides with $P(\Delta)$. It is clear that

$$BA_a(f)B^{-1} = BA_a(f)B^{-\tau^a} = (BA_1(f)B^{-\tau}) \cdots (BA_1(f)B^{-\tau})^{\tau^{a-1}}, \tag{61}$$

Using the triangular form obtained in (55), we deduce that the first two sides of the Newton polygon of $\det(I - TA_1(f))$ computed with respect to p coincide with $P(\Delta)$ if and only if the first two sides of the Newton polygon of $\det(I - TA_a(f))$ computed with respect to q coincide with $P(\Delta)$. Repeating the above argument, by induction we see that the theorem is true. Actually, the rationality of $\mathrm{L}^*(f, T)$ and a p-adic estimate on its zeros and poles imply that we only need to prove the theorem for the sides with slopes at most n.

The above proof implies

PROPOSITION 4.6. *The Newton polygon of det*$(I - TA_a(f))$ *computed with respect to* q *coincides with* $P(\Delta)$ *for the sides with slopes at most* m/D *if and only if the Newton polygon of det*$(I - TA_1(f))$ *computed with respect to* p *coincides with* $P(\Delta)$ *for the sides with slopes at most* m/D.

Putting the above together, we obtain

THEOREM 4.7. *Let* $\Delta(f) = \Delta$. *Assume that the L-function* $\mathrm{L}^*(f, T)^{(-1)^{n-1}}$ *is a polynomial. Then,* $\mathrm{NP}(f) = \mathrm{HP}(\Delta)$ *if and only if the Newton polygon of det*$(I - TA_1(f))$ *coincides with its lower bound* $P(\Delta)$. *In this case, the degree of the polynomial* $\mathrm{L}^*(f, T)^{(-1)^{n-1}}$ *is exactly* $n!\mathbf{V}(f)$.

Proof. Since $\mathrm{L}^*(f, T)^{(-1)^{n-1}}$ is a polynomial, the trace formula (51) and the definition of $H_\Delta(k)$ show that $\mathrm{L}^*(f, T)^{(-1)^{n-1}}$ has exactly $H_\Delta(k)$ reciprocal roots with slope k/D for all k if and only if $\det(I - tA_a(f))$ has exactly $W_\Delta(k)$ reciprocal roots with slope k/D for all k. By Theorem 4.5, this last condition holds if and only if the Newton polygon of $\det(I - TA_1(f))$ coincides with its lower bound $P(\Delta)$. The proof is complete.

The theorem of Adolphson-Sperber shows that the polynomial condition of Theorem 4.7 is satisfied for every non-degenerate f with n-dimensional $\Delta(f)$. Thus, the AS conjecture is a consequence of the following chain level version.

CONJECTURE 4.8. *Let Δ be an n-dimensional integral convex polyhedron in $\mathbf{R}^n$ containing the origin. If*

$$p \equiv 1 \pmod{D(\Delta)},$$

then the Newton polygon of $\det(I - TA_1(f))$ coincides with its lower bound $P(\Delta)$ for a generic f with $\Delta(f) = \Delta$, i.e., for all f in a Zariski dense open subset of the parameter space of f with $\Delta(f) = \Delta$.

This conjecture is of course also false in general. However, it is true in many important cases. For each Δ, the AS conjecture is true whenever Conjecture 4.8 is true. Thus, we shall restrict our attention to the study of Conjecture 4.8 which avoids the non-degenerate assumption. In addition, there are several simplifying advantages in working with the chain level Conjecture 4.8 than the cohomology level AS conjecture.

5. Degree polygons and Newton polygons. Since we are only interested in knowing if f is ordinary, we can always apply the facial decomposition theorem in Section 3. This will reduce the situation to the case where Δ has only one codimension 1 face not containing the origin and has the origin as a vertex. From now on, we shall assume that this condition holds. Namely, we assume that the origin 0 is a vertex of Δ and that Δ has only one face δ of codimension 1 which does not contain the origin.

Let

$$\mathcal{V} = \{V_1, \cdots, V_J\} \tag{62}$$

be the set of J fixed lattice points on the face δ such that the set $\mathcal{V}$ contains all the vertices of δ. It is clear that Δ is generated by 0 and the elements in $\mathcal{V}$. For our application to Conjecture 4.8, it is sufficient to take the case where $\mathcal{V}$ consists of all the lattice points on the face δ. This gives the maximal family. However, our theory works for a more general set $\mathcal{V}$ and hence for a more flexible family.

5.1. Newton polygons of subcones. For later proof, we need to define the Newton polygon and the degree polygon for a subcone of the full cone $C(\Delta)$. Thus, we first extend our setting a little bit. Let Σ be a cone contained in $C(\Delta)$ (not necessarily open or closed). Define a function on non-negative integers as follows:

$$W(\Sigma, k) = \mathrm{card}\{r \in \mathbf{Z}^n \cap \Sigma \mid w(r) = \frac{k}{D(\Delta)}\}, \tag{63}$$

This is the number of lattice points in the cone Σ with weight exactly k/D. Let $P(\Sigma)$ be the polygon in $\mathbf{R}^2$ with vertices $(0,0)$ and

$$\Big(\sum_{k=0}^{m} W(\Sigma, k), \frac{1}{D(\Delta)} \sum_{k=0}^{m} kW(\Sigma, k)\Big), \quad m = 0, 1, 2, \cdots. \tag{64}$$

For convenience, we shall call the vertex in (64) the m^{th} vertex in $P(\Sigma)$. Note that the m^{th} vertex may be equal to the $(m+1)^{\text{th}}$ vertex, because it may happen that $W(\Sigma, m) = 0$. Recall that $A_1(f) = (a_{s,r}(f))$ is the semilinear Frobenius matrix defined in (49). We define $A_1(\Sigma, f)$ to be the submatrix $(a_{s,r}(f))$ with r and s running through the cone Σ. In particular, for the full cone $\Sigma = C(\Delta)$, we have

$$A_1(C(\Delta), f) = A_1(f)$$

and

$$W(C(\Delta), k) = W(k).$$

From the block form (53), we deduce

PROPOSITION 5.1. *The Fredholm determinant* $\det(I - TA_1(\Sigma, f))$ *is entire. Its Newton polygon lies above the polygon* $P(\Sigma)$.

Let $P(\Sigma, x)$ be the piecewise linear function on $\mathbf{R}_{\geq 0}$ whose graph is the polygon $P(\Sigma)$. By the block form (53), we can write

$$\det(I - TA_1(\Sigma, f)) = \sum_{k=0}^{\infty} p^{P(\Sigma,k)} G(\Sigma, f, k) T^k, \tag{65}$$

where $G(\Sigma, f, k)$ is a power series in the a_j with p-adic integral coefficients. The reduction

$$\mathcal{H}(\Sigma, f, k) \equiv G(\Sigma, f, k) \pmod{\pi} \tag{66}$$

is a polynomial in the coefficients a_j of f defined over the finite prime field $\mathbf{F}_p$. This polynomial is called the k-th **Hasse polynomial** of the pair (Σ, f).

For a given pair (Σ, f), the Newton polygon of $\det(I - TA_1(\Sigma, f))$ coincides with its lower bound $P(\Sigma)$ at the m-th vertex

$$\Big(\sum_{i=0}^{m} W(\Sigma, i), \frac{1}{D} \sum_{i=0}^{m} iW(\Sigma, i)\Big)$$

if and only if the Hasse polynomial $\mathcal{H}(\Sigma, f, k)$ does not vanish for

$$k = \sum_{i=0}^{m} W(\Sigma, i)$$

at the point a_j. To show that the Newton polygon of $\det(I - TA_1(\Sigma, f))$ coincides generically with its lower bound at the m-th vertex, we need to show that the Hasse polynomial $\mathcal{H}(\Sigma, f, k)$ in the variables a_j is not identically zero for $k = \sum_{i=0}^{m} W(\Sigma, i)$.

Our strategy is then to show that certain "leading form" of the Hasse polynomial $\mathcal{H}(\Sigma, f, k)$ is not zero in some cases. For this purpose, we need to choose some priority variables out of the total set of variables $\{a_1, \cdots, a_J\}$ so that we can define the leading form of $\mathcal{H}(\Sigma, f, k)$ in terms of these priority variables. Of course, we want to choose the priority variables in such a way that it is easier to prove the non-vanishing of the leading form. To make this idea precise, we need the help of a certain maximizing function and introduce the associated notion of a degree polygon.

5.2. Degree polygons. Fix a non-empty subset $\mathcal{U}$ of the set $\mathcal{V}$:

$$\mathcal{U} \subset \mathcal{V}.$$

The a_j's with $V_j \in \mathcal{U}$ will be our priority variables. For example, if $\mathcal{U}$ consists of the single element V_1, then a_1 will be our priority variable. If $\mathcal{U}$ equals the total set $\mathcal{V}$, then all the variables $\{a_1, \cdots, a_J\}$ will be our priority variables. For a given $\mathcal{U}$, we define a maximizing function on the cone $C(\Delta)$ as follows.

DEFINITION 5.2. *Let $\mathcal{U}$ be a non-empty subset of the set V. For $r \in C(\Delta)$, we define*

$$m(\mathcal{U}, \mathcal{V}; r) = \sup\{\sum_{V_j \in \mathcal{U}} u_j \mid \sum_{j=1}^{J} u_j V_j = r, u_j \geq 0\}.$$

If $r \in \mathbf{R}^n$ but $r \notin C(\Delta)$, we define $m(U, V; r) = 0$. If for all $r \in C(\Delta)$, we have

$$m(\mathcal{U}, \mathcal{V}; r) = \inf\{\sum_{V_j \in \mathcal{U}} u_j \mid \sum_{j=1}^{J} u_j V_j = r, u_j \geq 0\},$$

we say that $\mathcal{U}$ is homogeneous with respect to $\mathcal{V}$.

This is a standard maximizing function in linear programming. For a given r, there are various algorithms to compute $m(\mathcal{U}, \mathcal{V}; r)$. The simplex method in linear programming shows that the set of optimal solutions is a "face" formed by some of the lattice points V_j. This optimal face can be described explicitly in various special cases. A trivial case is when $\mathcal{U}$ equals $\mathcal{V}$. In this case, $m(\mathcal{U}, \mathcal{V}; r)$ equals the weight function $w(r)$ and thus $\mathcal{U}$ is automatically homogeneous, we are led to the facial decomposition theorem in Theorem 3.1. One non-trivial case is when $\mathcal{U}$ consists of a single element. This choice leads to the star decomposition theorem in [18]. Another case is when $\mathcal{U}$ consists of the intersection of δ with a suitable half space [18]. This choice leads to the hyperplane decomposition theorem. The most flexible case arises when we take $\mathcal{U}$ to be the compliment $\mathcal{V} - \{V_1\}$ of a vertex V_1 in $\mathcal{V}$. This leads to the collapsing decomposition in the present paper.

This last case is not always homogeneous. However, if we assume that V_1 is a vertex and $\mathcal{U} = \mathcal{V} - \{V_1\}$ lies on an $(n-2)$-dimensional (not $(n-1)$-dimensional) hyperplane H in δ, then it is easy to check that $\mathcal{U}$ is homogeneous with respect to V. In fact, let $r = \sum_j u_j V_j \in C(\Delta)$ as in Definition 5.2. We may assume that r is not the origin. Dividing by $w(r)$, we may assume that $w(r) = \sum_j u_j = 1$. If $r \in H$, then $u_1 = 0$, and $\sum_{j=2}^{J} u_j = w(r)$ for all solutions. If $r \notin H$, then $\sum_{j=2}^{J} u_j V_j$ is the unique intersection point with the hyperplane H of the line connecting V_1 and r, and thus $\sum_{j=2}^{J} u_j$ is independent of the choice of the solutions. This proves that $\mathcal{U}$ is indeed homogeneous with respect to $\mathcal{V}$.

For a non-negative integer k and a subcone Σ of $C(\Delta)$, we define

$$Q(\Sigma, \mathcal{U}, \mathcal{V}; k) = (p-1) \sum_{w(r) \leq k/D,\ r \in \mathbf{Z}^n \cap \Sigma} m(\mathcal{U}, \mathcal{V}; r), \tag{67}$$

where the intersection $\mathbf{Z}^n \cap \Sigma$ is simply the set of lattice points in the cone Σ. The number $Q(\Sigma, \mathcal{U}, \mathcal{V}; k)$ is always non-negative. Let $Q(\Sigma, \mathcal{U}, \mathcal{V})$ be the graph in $\mathbf{R}^2$ of the piece-wise linear functions passing through the vertices $(0, 0)$ and

$$\Big(\sum_{k=0}^{m} W(\Sigma, k), Q(\Sigma, \mathcal{U}, \mathcal{V}; m)\Big), \quad m = 0, 1, 2, \cdots. \tag{68}$$

We shall call the vertex in (68) as the m-th vertex in $Q(\Sigma, \mathcal{U}, \mathcal{V})$. Note that the coordinates in (68) are always non-negative. The polygon $P(\Sigma)$ is concave upward by our definition. We do not claim that $Q(\Sigma, \mathcal{U}, \mathcal{V})$ is convex (upward or downward). In the special case that $\Sigma = C(\Delta)$, we simply write

$$Q(C(\Delta), \mathcal{U}, \mathcal{V}) = Q(\mathcal{U}, \mathcal{V}).$$

Definition 5.3. *For any polynomial F in the variables a_j ($1 \leq j \leq J$), we define $d(\mathcal{U}, F)$ to be the total degree of F in the priority variables a_j with $V_j \in \mathcal{U}$. If $F = 0$, we define $d(\mathcal{U}, F) = -\infty$. The number $d(\mathcal{U}, F)$ is called the $\mathcal{U}$-degree of F.*

Thus, the polynomial F is non-zero if and only if its $\mathcal{U}$-degree is not $-\infty$. Let F_r be the polynomial defined by (45). It is clear from Definition 5.2 that we have the following upper bound for the $\mathcal{U}$-degree of F_r:

$$d(\mathcal{U}, F_r) \leq m(\mathcal{U}, \mathcal{V}; r). \tag{69}$$

In the case that $\mathcal{U}$ is homogeneous with respect to $\mathcal{V}$, then F_r is $\mathcal{U}$-homogeneous and thus the above inequality becomes an equality if and only if F_r is non-zero.

Definition 5.4. *Let $d(\mathcal{U}, \mathcal{H}(\Sigma, f, k))$ denote the $\mathcal{U}$-degree of the k-th Hasse polynomial $\mathcal{H}(\Sigma, f, k)$. We define the $\mathcal{U}$-degree polygon of $det(I - TA_1(\Sigma, f))$ to be the graph in $\mathbf{R}^2$ of the piecewise linear function with vertices $(0, 0)$ and*

$$\Big(\sum_{k=0}^{m} W(\Sigma, k), \max\{0, d(\mathcal{U}, \mathcal{H}(\Sigma, f, \sum_{k=0}^{m} W(\Sigma, k)))\}\Big) \quad m = 0, 1, \cdots, \infty.$$

Note that we do not claim that the $\mathcal{U}$-degree polygon is convex.

If r and r' are two lattice points in $\mathbf{Z}^n$, Definition 5.2 easily implies the inequality

$$m(\mathcal{U}, \mathcal{V}; r) + m(\mathcal{U}, \mathcal{V}; r') \leq m(\mathcal{U}, \mathcal{V}; r + r'). \tag{70}$$

The equality always holds if $\mathcal{U}$ is homogeneous with respect to $\mathcal{V}$ and $r, r' \in C(\Delta)$. Furthermore, if c is non-negative, then

$$m(\mathcal{U}, \mathcal{V}; cr) = cm(\mathcal{U}, \mathcal{V}; r). \tag{71}$$

Let m be a non-negative integer. By the block form of the matrix $A_1(\Sigma, f)$, the determinant expansion of a matrix and equations (67)-(71), we deduce

$$\begin{aligned}
& d\Big(\mathcal{U}, \mathcal{H}(\Sigma, f, \sum_{k=0}^{m} W(\Sigma, k))\Big) \\
\leq & \max_{\phi} \sum_{w(r) \leq m/D, r \in \mathbf{Z}^n \cap \Sigma} d(\mathcal{U}, F_{pr - \phi(r)}) \\
\leq & \max_{\phi} \sum_{w(r) \leq m/D, r \in \mathbf{Z}^n \cap \Sigma} m(\mathcal{U}, \mathcal{V}; pr - \phi(r)) \\
\leq & \max_{\phi} \sum_{w(r) \leq m/D, r \in \mathbf{Z}^n \cap \Sigma} (pm(\mathcal{U}, \mathcal{V}; r) - m(\mathcal{U}, \mathcal{V}; \phi(r))) \\
\leq & (p-1) \sum_{w(r) \leq m/D, r \in \mathbf{Z}^n \cap \Sigma} m(\mathcal{U}, \mathcal{V}; r) \\
= & Q(\Sigma, \mathcal{U}, \mathcal{V}; , m),
\end{aligned} \tag{72}$$

where ϕ runs through the permutations on $\sum_{k=0}^{m} W(\Sigma, k)$ letters. If $\mathcal{U}$ is homogeneous with respect to $\mathcal{V}$ and if the Hasse polynomial

$$\mathcal{H}(\Sigma, f, \sum_{k=0}^{m} W(\Sigma, k))$$

is not the zero polynomial, then the Hasse polynomial is $\mathcal{U}$-homogeneous and we have the equality

$$d\Big(\mathcal{U}, \mathcal{H}(\Sigma, f, \sum_{k=0}^{m} W(\Sigma, k))\Big) = Q(\Sigma, \mathcal{U}, \mathcal{V}; , m).$$

In this way, we have proved the following upper bound for the $\mathcal{U}$-degree polygon.

PROPOSITION 5.5. *The $\mathcal{U}$-degree polygon of det$(I - TA_1(\Sigma, f))$ lies below the polygon $Q(\Sigma, \mathcal{U}, \mathcal{V})$.*

If the $\mathcal{U}$-degree polygon coincides with its upper bound $Q(\Sigma, \mathcal{U}, \mathcal{V})$ at the m-th vertex, then the polynomial $\mathcal{H}(\Sigma, f, \sum_{i=0}^{m} W(\Sigma, i))$ is not zero since its $\mathcal{U}$-degree is equal to $Q(\Sigma, \mathcal{U}, \mathcal{V}; \sum_{i=0}^{m} W(i))$ which is non-negative and hence not $-\infty$. The converse is also true if $\mathcal{U}$ is homogeneous with respect to $\mathcal{V}$. We obtain

PROPOSITION 5.6. *If the $\mathcal{U}$-degree polygon of det$(I - TA_1(\Sigma, f))$ coincides with $Q(\Sigma, \mathcal{U}, \mathcal{V})$ at the m-th vertex, then the Newton polygon of det$(I - TA_1(\Sigma, f))$ coincides generically with $P(\Sigma)$ at the m-th vertex. If $\mathcal{U}$ is homogeneous with respect to V, then the converse is also true.*

This property shows that the $\mathcal{U}$-degree polygon is finer than the generic Newton polygon. Grothendieck's specialization theorem says that the Newton polygon goes up under specializations. The following is an analogue for the $\mathcal{U}$-degree polygon. It results from the fact that the degree of a polynomial decreases under specializations.

PROPOSITION 5.7. *The $\mathcal{U}$-degree polygon of det$(I - TA_1(\Sigma, f))$ goes down under $\mathcal{U}$-specializations, where a $\mathcal{U}$-specialization means a specialization of those variables a_j such that V_j is neither a vertex of Δ nor an element of $\mathcal{U}$.*

6. Collapsing decomposition for degree polygons. We will assume that $n \geq 2$ as the case $n = 1$ is already handled by the facial decomposition. Let

$$\mathcal{V} = \{V_1, \cdots, V_J\}$$

be the set of J fixed lattice points in $\mathbf{R}^n$. Let Δ be the convex polyhedron in $\mathbf{R}^n$ generated by the origin and the lattice points in $\mathcal{V}$. We assume that Δ is n-dimensional. By the facial decomposition, we may assume that Δ has only one co-dimension 1 face δ not containing the origin and all $V_j \in \delta$. To decompose Δ, we will decompose the unique face δ. Actually, we will be decomposing the set $\mathcal{V}$ since we are working in a little more general setting. In the special case when $\mathcal{V}$ consists of all lattice points in δ, then our decomposition can be described purely in terms of δ. Clearly, the set $\mathcal{V}$ has at least n elements. If the set $\mathcal{V}$ has exactly n elements, then the set $\mathcal{V}$ is called **indecomposable**.

Choose an element in $\mathcal{V}$ which is a vertex of δ, say V_1. Let

$$\mathcal{V} = \bigcup_{i=1}^{h} \mathcal{V}_i$$

be the collapsing decomposition of $\mathcal{V}$ with respect to V_1 as described in Section 3.2. Let $f(x)$ be the generic Laurent polynomial

$$f(X) = \sum_{j=1}^{J} a_j X^{V_j}.$$

For $1 \leq i \leq h$, let Σ_i be the n-dimensional convex cone generated by $\mathcal{V}_i$ and the origin. Equivalently, Σ_i is the n-dimensional cone generated by the origin and δ_i. Let

$$f_{\Sigma_i}(X) = \sum_{V_j \in \Sigma_i} a_j X^{V_j}$$

be the restriction of f to the cone Σ_i. We shall take

$$\mathcal{U} = \mathcal{V}_1 = \{V_2, \cdots, V_J\}. \tag{73}$$

6.1. The closed collapsing decomposition theorem. The closed collapsing decomposition theorem for degree polygons is

Theorem 6.1. *For $m \in \mathbf{Z}_{\geq 0}$, the $\mathcal{V}_1$-degree polygon of $det(I - TA_1(f))$ coincides with its upper bound $Q(\mathcal{V}_1, \mathcal{V})$ at the m^{th} vertex if and only if for each $1 \leq i \leq h$, the $\mathcal{V}_1 \cap \mathcal{V}_i$-degree polygon of $det(I - TA_1(f_{\Sigma_i}))$ defined with respect to $P(\Sigma_i, f_{\Sigma_i})$ coincides with its upper bound $Q(\mathcal{V}_1 \cap \mathcal{V}_i, \mathcal{V}_i)$ at the m^{th} vertex.*

This result implies Theorem 3.7 by induction on the cardinality of the set $\mathcal{V}$. We may assume that the generic f is ordinary, otherwise there is nothing to prove. If $|\mathcal{V}| = n$ (the minimal possible value), then f is already diagonal and there is nothing to prove. Let $|\mathcal{V}| > n$. By induction and under the assumption of Theorem 3.7, we can assume that for each f_{Σ_i} in Theorem 6.1, the Newton polygon of $\det(I - TA_1(f_{\Sigma_i}))$ coincides generically with $P(\Sigma_i, f_{\Sigma_i})$. Now, $\mathcal{V}_1 \cap \mathcal{V}_i$ is homogeneous with respect to $\mathcal{V}_i$. By Proposition 5.6, the $\mathcal{V}_1 \cap \mathcal{V}_i$-degree polygon of $\det(I - TA_1(f_{\Sigma_i}))$ defined with respect to $P(\Sigma_i, f_{\Sigma_i})$ coincides with its upper bound $Q(\mathcal{V}_1 \cap \mathcal{V}_i, \mathcal{V}_i)$. By Theorem 6.1, we deduce that the $\mathcal{V}_1$-degree polygon of $\det(I - TA_1(f))$ coincides with its upper bound $Q(\mathcal{V}_1, \mathcal{V})$. Applying Proposition 5.6 again, we conclude that f is generically ordinary. Theorem 3.7 is proved.

Before proving Theorem 6.1, we need to have a better understanding of the maximizing function that is used to define the degree polygon. Recall that for $r \in C(\Delta)$, we defined

$$m(\mathcal{V}_1, \mathcal{V}; r) = \sup\{\sum_{j=2}^{J} u_j \mid \sum_{j=1}^{J} u_j V_j = r, u_j \geq 0\}.$$

If $r \in \mathbf{R}^n$ but $r \notin C(\Delta)$, then $m(\mathcal{V}_1, \mathcal{V}; r) = 0$.

Lemma 6.2. *Let $(u_1, \cdots, u_J)$ be a rational solution of the linear equation*

$$\sum_{j=1}^{J} u_j V_j = r, \ u_j \geq 0.$$

Suppose that $u_{j_1}, \cdots, u_{j_k}$ are its non-zero coordinates. Then, we have

1. *(i). If $r \in \Sigma_1$, then $m(\mathcal{V}_1, \mathcal{V}; r) = \sum_{j=2}^{J} u_j$ if and only if Σ_1 (or δ_1) contains all the lattice points $V_{j_1}, \cdots, V_{j_k}$ (equivalently, $u_1 = 0$).*
2. *(ii). If $r \in \Sigma_i$ for some $i \geq 2$ and $m(\mathcal{V}_1, \mathcal{V}; r) = \sum_{j=2}^{J} u_j$, then Σ_i (or δ_i) contains all the lattice points $V_{j_1}, \cdots, V_{j_k}$.*
3. *(iii). If $r \in \Sigma_1 \cap \Sigma_i$ for some $i \geq 2$ and $m(\mathcal{V}_1, \mathcal{V}; r) = \sum_{j=2}^{J} u_j$, then $\Sigma_1 \cap \Sigma_i$ (or $\delta_1 \cap \delta_i$) contains all the lattice points $V_{j_1}, \cdots, V_{j_k}$.*

Proof. (i). Let $r \in \Sigma_1$. Since $\sum_{j=1}^{J} u_j V_j = r$ and $w(r) = \sum_{j=1}^{J} u_j$, we have

$$\sum_{j=2}^{J} u_j \le \sum_{j=1}^{J} u_j = w(r).$$

The inequality is an equality if and only if $u_1 = 0$. This is true if and only if Σ_1 contains all the lattice points $V_{j_1}, \cdots, V_{j_k}$.

(ii). Let

$$r_1 = \sum_{j=2}^{J} u_j V_j, \ r_2 = u_1 V_1.$$

Then,

$$r = r_1 + r_2, \ w(r_1) = \sum_{j=2}^{J} u_j.$$

We may assume that r is non-zero. If Σ_i does not contain all of the lattice points $V_{j_1}, \cdots, V_{j_k}$, then $w(r_1)w(r_2) > 0$ and r_1 lies strictly on the other side of Σ_i with respect to the hyperplane $\Sigma_1 \cap \Sigma_i$. This implies that the line segment from r_1 to r intersects the hyperplane $\Sigma_1 \cap \Sigma_i$ at a unique point r_1' ($\neq r_1$). That is, there is a unique positive number $0 < \lambda < 1$ such that $\lambda r_1 + (1-\lambda) r = r_1' \in \Sigma_1 \cap \Sigma_i$. Thus,

$$r = \lambda r + (1-\lambda) r = \lambda(r_1 + r_2) + (1-\lambda) r = (\lambda r_1 + (1-\lambda) r) + \lambda r_2 = r_1' + \lambda r_2.$$

Since $w(r) = w(r_1) + w(r_2) = w(r_1') + \lambda w(r_2)$, where $0 < \lambda < 1$ and $w(r_2) > 0$, we deduce that $w(r_1') > w(r_1)$. This shows that $w(r_1) = \sum_{j=2}^{J} u_j$ is not the maximum value $m(\mathcal{V}_1, \mathcal{V}; r)$. This proves (ii). One checks that (iii) follows from (i) and (ii). The lemma is proved.

LEMMA 6.3. *Let r_1 and r_2 be two rational points in the cone $C(\Delta)$. Then*

$$m(\mathcal{V}_1, \mathcal{V}; r_1 + r_2) \ge m(\mathcal{V}_1, \mathcal{V}; r_1) + m(\mathcal{V}_1, \mathcal{V}; r_2).$$

If the equality holds, then both r_1 and r_2 lie on Σ_i for some i.

Proof. Let

$$r_1 = u_1 V_1 + \cdots + u_J V_J, \ \sum_{j=2}^{J} u_j = m(\mathcal{V}_1, \mathcal{V}; r_1),$$

$$r_2 = w_1 V_1 + \cdots + w_J V_J, \ \sum_{j=2}^{J} w_j = m(\mathcal{V}_1, \mathcal{V}; r_2),$$

Then,

$$r_1 + r_2 = (u_1 + w_1) V_1 + \cdots + (u_J + w_J) V_J. \tag{74}$$

$$m(\mathcal{V}_1, \mathcal{V}; r_1) + m(\mathcal{V}_1, \mathcal{V}; r_2) = \sum_{j=2}^{J} (u_j + w_j) \le m(\mathcal{V}_1, \mathcal{V}; r_1 + r_2). \tag{75}$$

Let $V_{j_1}, \cdots, V_{j_k}$ be the lattice points with non-zero coefficients in (74). If $r_1+r_2 \in \Sigma_i$ and the inequality in (75) is an equality, then Lemma 6.2 shows that the lattice points $V_{j_1}, \cdots, V_{j_k}$ are all contained in Σ_i. This implies that both r_1 and r_2 are contained in Σ_i. The lemma is proved.

LEMMA 6.4. *Let $r \in \Sigma_i$ for some $1 \leq i \leq h$ and $f = \sum_{j=1}^{J} a_j x^{V_j}$. Then*

$$d(\mathcal{V}_1, F_r(f) - F_r(f_{\Sigma_i})) < m(\mathcal{V}_1, \mathcal{V}; r).$$

That is, the $\mathcal{V}_1$-degree of $F_r(f)-F_r(f_{\Sigma_i})$ is strictly smaller than the expected maximum value $m(\mathcal{V}_1, \mathcal{V}; r)$.

Proof. Let $u_1, \cdots, u_J$ be non-negative integers satisfying

$$r = \sum_{j=1}^{J} u_j V_j, \quad \sum_{j=2}^{J} u_j = m(\mathcal{V}_1, \mathcal{V}; r).$$

Let $u_{j_1}, \cdots, u_{j_k}$ are the non-zero terms among the u_j's. Lemma 6.2 shows that the cone Σ_i contains all the lattice points $V_{j_1}, \cdots, V_{j_k}$. This shows that if a monomial in the a_j with $\mathcal{V}_1$-degree $m(\mathcal{V}_1, \mathcal{V}; r)$ appears in $F_r(f)$, then the same monomial also appears in $F_r(f_{\Sigma_i})$. Thus, $F_r(f)$ and $F_r(f_{\Sigma_i})$ have the same initial terms of $\mathcal{V}_1$-degree $m(\mathcal{V}_1, \mathcal{V}; r)$. The lemma is proved.

In order to prove the closed collapsing decomposition theorem, we need to work with its open version first.

6.2. The open collapsing decomposition theorem. For $1 \leq i \leq h$, let S_i^o be the set of relatively open faces in δ_i, including the empty set. This is also called the boundary decomposition of δ_i. The union

$$\delta_i = \bigcup_{\sigma \in S_i^o} \sigma \tag{76}$$

is a disjoint union. Each S_i^o contains exactly one $(n-1)$-dimensional face, which is the interior of δ_i. The 0-dimensional elements in S_i^o are simply the vertices in δ_i. For an element $\sigma \in S_i^o$, let Σ^o denote the open cone generated by σ and the origin (the origin itself is not included unless σ is the empty set). Since $\sigma \in S_i^o$, the open cone Σ^o is a subcome of the closed cone Σ_i and we have

$$\dim \Sigma^o = \dim \sigma + 1.$$

Clearly, the union

$$\Sigma_i = \bigcup_{\sigma \in S_i^o} \Sigma^o \tag{77}$$

is a disjoint union, called the boundary decomposition of Σ_i. Let Σ be the topological closure of Σ^o. The open collapsing decomposition theorem is

THEOREM 6.5. *For $m \in \mathbf{Z}_{\geq 0}$, the $\mathcal{V}_1$-degree polygon of $det(I-TA_1(f))$ coincides with its upper bound $Q(\mathcal{V}_1, \mathcal{V})$ at the m^{th} vertex if and only if for each $1 \leq i \leq h$ and each $\sigma \in S_i^o$, the $\mathcal{V}_1 \cap \Sigma$-degree polygon of $det(I - TA_1(\Sigma^o, f_{\Sigma_i}))$ coincides with its upper bound $Q(\Sigma^o, \mathcal{V}_1 \cap \mathcal{V}_i, \mathcal{V}_i)$ at the m^{th} vertex.*

Proof. Let

$$S(\mathcal{V}_1, \mathcal{V}) = \bigcup_{i=1}^{h} S_i^o \tag{78}$$

which includes the empty set as an element. Denote the number of elements in this set by $g+1$. Fix an ordering of $S(\mathcal{V}_1, \mathcal{V})$ by

$$S(\mathcal{V}_1, \mathcal{V}) = \{\sigma_0, \sigma_1, \cdots, \sigma_g\} \tag{79}$$

such that

$$\dim(\sigma_j) \le \dim(\sigma_{j+1}), \ 0 \le j \le g-1. \tag{80}$$

In particular, σ_0 is the empty set. Let Σ_j^o be the relatively open cone generated by σ_j and the origin. In particular, Σ_0^o consists of the origin. It is clear that we have

$$\dim(\Sigma_j^o) = \dim(\sigma_j) + 1$$

and thus

$$\dim(\Sigma_j^o) \le \dim(\Sigma_{j+1}^o), \ 0 \le j \le g-1 \tag{81}$$

Let

$$C(\mathcal{V}_1, \mathcal{V}) = \{\Sigma_0^o, \cdots, \Sigma_g^o\}. \tag{82}$$

Then, the full cone $C(\Delta)$ is the disjoint union of the relatively open cones in $C(\mathcal{V}_1, \mathcal{V})$:

$$C(\Delta) = \bigcup_{j=0}^{g} \Sigma_j^o. \tag{83}$$

Let $0 \le j_1 < j_2 \le g$ and

$$s \in \Sigma_{j_1}^o, \quad r \in \Sigma_{j_2}^o. \tag{84}$$

In particular, $r \neq 0$ since the origin is only contained in Σ_0^o. We claim that for each $1 \le i \le h$, the closed convex cone Σ_i cannot contain both r and $ps - r$. Otherwise, suppose that both r and $ps - r$ are contained in Σ_i for some $1 \le i \le h$. From the identity $ps = r + (ps - r)$, we deduce that r, s and $ps - r$ are all contained in Σ_i. In particular, both $\Sigma_{j_1}^o$ and $\Sigma_{j_2}^o$ are subcones of Σ_i. Let $ps - r \in \Sigma_{j_3}^o$ (a subcone of Σ_i). Then, the equation $ps = r + (ps - r)$ shows that s is in the interior of the cone generated by $\Sigma_{j_2}^o$ and $\Sigma_{j_3}^o$. This implies that

$$\dim(\Sigma_{j_1}^o) \ge \dim(\Sigma_{j_2}^o), \tag{85}$$

with equality holding if and only if $ps - r$ lies in the topological closure of $\Sigma_{j_2}^o$. Our ordering assumption shows that (85) is indeed an equality. Thus, $ps - r$ is indeed in the topological closure of $\Sigma_{j_2}^o$. We conclude from $ps = r + (ps - r)$ that s is in $\Sigma_{j_2}^o$. This shows that $\Sigma_{j_1}^o$ and $\Sigma_{j_2}^o$ are not disjoint, a contradiction. The claim is proved. This claim together with Lemmas 6.3 shows that

$$d(\mathcal{V}_1, a_{s,r}(f)) = d(\mathcal{V}_1, F_{ps-r}(f))$$

$$\leq m(\mathcal{V}_1, \mathcal{V}; ps - r) < pm(\mathcal{V}_1, \mathcal{V}; s) - m(\mathcal{V}_1, \mathcal{V}; r). \tag{86}$$

Let $B_{j_1 j_2}$ $(0 \leq j_1, j_2 \leq g)$ be the nuclear submatrix of $A_1(f)$ consisting of all $(a_{s,r}(f))$ with $s \in \Sigma_{j_1}^o$ and $r \in \Sigma_{j_2}^o$. For $0 \leq j \leq g$, the Newton polygon of the entire function $\det(I - tB_{jj})$ lies above $P(\Sigma_j^o)$. Furthermore, under a permutation of orthonormal basis,

$$A_1(f) = \begin{pmatrix} B_{00} & B_{01} & \dots & B_{0g} \\ B_{10} & B_{11} & \dots & B_{1g} \\ \vdots & \vdots & \ddots & \vdots \\ B_{g0} & B_{g1} & \dots & B_{gg} \end{pmatrix}. \tag{87}$$

If $a_{s,r}(f)$ is an element in $B_{j_1 j_2}$ with $j_1 < j_2$, then (86) shows that the $\mathcal{V}_1$-degree $d(\mathcal{V}_1, a_{s,r}(f))$ of the polynomial $a_{s,r}(f)$ is strictly smaller than the expected maximum value $pm(\mathcal{V}_1, \mathcal{V}; s) - m(\mathcal{V}_1, \mathcal{V}; r)$. This means that the above block form for $A_1(f)$ is, in some sense, lower triangular with respect to the $\mathcal{V}_1$-degree. By induction, we deduce that

$$\det(I - TA_1(f)) = \prod_{j=0}^{g} \det(I - TB_{jj}) + \sum_{k=0}^{\infty} p^{P(\Delta,k)} G(f,k) T^k, \tag{88}$$

where $G(f,k)$ is a power series in the a_j with p-adic integral coefficients such that the reduction $G(f,k) \pmod{\pi}$ is a polynomial over $\mathbf{F}_p$ whose $\mathcal{V}_1$-degree is strictly smaller than their upper bound $Q(\mathcal{V}_1, \mathcal{V}; k)$, see the notation in (67)-(68). Thus, the $\mathcal{V}_1$-degree polygon of $\det(I - TA_1(f))$ coincides with $Q(\mathcal{V}_1, \mathcal{V})$ at the m^{th} vertex if and only if the $\mathcal{V}_1$-degree polygon of the first term on the right side of (88) coincides with $Q(\mathcal{V}_1, \mathcal{V})$ at the m^{th} vertex. One further shows that the latter is true if and only if the $\mathcal{V}_1$-degree polygon of

$$\det(I - TA_1(\Sigma_j^o, f)) = \det(I - TB_{jj})$$

defined with respect to $P(\Sigma_j^o)$ coincides with $Q(\Sigma_j^o, \mathcal{V}_1, \mathcal{V})$ at the m^{th} vertex for all $0 \leq j \leq g$. The matrix $A_1(\Sigma_j^o, f)$ is however different from the desired matrix $A_1(\Sigma_j^o, f_{\Sigma_i})$, where $\Sigma_j^o \subset \Sigma_i$. But the $\mathcal{V}_1$-degree polygon of $\det(I - TA_1(\Sigma_j^o), f)$ and the $\mathcal{V}_1 \cap \mathcal{V}_i$-degree polygon of $\det(I - TA_1(\Sigma_j^o, f_{\Sigma_i}))$ have the same upper bound

$$Q(\Sigma_j^o, \mathcal{V}_1, \mathcal{V}) = Q(\Sigma_j^o, \mathcal{V}_1 \cap \mathcal{V}_i, \mathcal{V}_i). \tag{89}$$

The last equality can be easily proved from our definitions in (67)-(68).

To finish the proof, we need to show that if we replace the matrix $A_1(\Sigma_j^o, f)$ defined in terms of f by the matrix $A_1(\Sigma_j^o, f_{\Sigma_i})$ defined in terms of f_{Σ_i}, we will not change the property of the coincidence of the degree polygon with its upper bound. Let r, $s \in \Sigma_i$. If $ps - r$ also belongs to Σ_i, Lemma 6.4 shows that we can replace $F_{ps-r}(f)$ by $F_{ps-r}(f_{\Sigma_i})$. If $ps - r$ does not belong to Σ_i, then Lemma 6.3 shows that the $\mathcal{V}_1$-degree of $F_{ps-r}(f)$ is strictly smaller than the expected maximum value $pm(\mathcal{V}_1, \mathcal{V}; s) - m(\mathcal{V}_1, \mathcal{V}; r)$, while $F_{ps-r}(f_{\Sigma_i}) = 0$. In this case, we can also replace $F_{ps-r}(f)$ by $F_{ps-r}(f_{\Sigma_i})$. The theorem is proved.

6.3. Proof of the closed collapsing decomposition. To prove the closed collapsing decomposition theorem, it suffices to combine the above open collapsing decomposition theorem and the following boundary decomposition theorem [18].

THEOREM 6.6. *For each $1 \le i \le h$, we have the boundary decomposition*

$$\det(I - TA_1(f_{\Sigma_i})) = \prod_{\sigma_j \in S_i^o} \det(I - TA_1(\Sigma_j^o, f_{\Sigma_i})).$$

This theorem shows that the $\mathcal{V}_1 \cap \mathcal{V}_i$-degree polygon of $\det(I - TA_1(f_{\Sigma_i})$ coincides with its upper bound $Q(\Sigma_i, \mathcal{V}_1 \cap \mathcal{V}_i, \mathcal{V}_i)$ if and only if the $\mathcal{V}_1 \cap \mathcal{V}_i$-degree polygon of $\det(I - TA_1(\Sigma_j^o, f_{\Sigma_i})$ coincides with its upper bound $Q(\Sigma_j^o, \mathcal{V}_1 \cap \mathcal{V}_i, \mathcal{V}_i)$ for all j with $\sigma_j \in S_i^o$. The proof of Theorem 6.1 is complete.

REFERENCES

[1] A. ADOLPHSON AND S. SPERBER, *Newton polyhedra and the degree of the L-function associated to an exponential sum*, Invent. Math., 88 (1987), pp. 555–569.

[2] A. ADOLPHSON AND S. SPERBER, *Exponential sums and Newton polyhedra: Cohomology and estimates*, Ann. Math., 130 (1989), pp. 367–406.

[3] A. ADOLPHSON AND S. SPERBER, *On the zeta function of a complete intersection*, Ann. Sci. École Norm. Sup., (4) 29 (1996), no. 3, pp. 287–328.

[4] E. BOMBIERI, *On exponential sums in finite fields, II*, Invent. Math., 47 (1978), pp. 29–39.

[5] P. DELIGNE, *La conjecture de Weil, II*, Publ. Math. IHES, 52 (1980), pp. 137–252.

[6] J. DENEF AND F. LOESER, *Weights of exponential sums, intersection cohomology and Newton polyhedra*, Invent. Math., 106:2 (1991), pp. 275–294.

[7] B. DWORK, *Normalized period matrices, II*, Ann. of Math., 98 (1973), pp. 1–57.

[8] L. FU AND D. WAN, *Moment L-functions, partial L-functions and partial exponential sums*, Math. Ann., 328 (2004), pp. 193–228.

[9] S. HONG, *Newton polygons of L-functions associated with exponential sums of polynomials of degree four over finite fields*, Finite Fields & Appl., 7 (2001), pp. 205–237.

[10] S. HONG, *Newton polygons for L-functions of exponential sums of polynomials of degree six over finite fields*, J. Number Theory., 97 (2002), pp. 368–396.

[11] N. KATZ, *Slope filtration of F-crystals*, Astérisque, 63 (1979), pp. 113–163.

[12] D.S. KIM, *Gauss sums for $U(2n, q^2)$*, Glasgow Math. J., 40 (1998), pp. 79–85.

[13] B. MAZUR, *Frobenius and the Hodge filtration*, Bull. Amer. Math. Soc., 78 (1972), pp. 653–667.

[14] S. SPERBER, *Congruence properties of the hyperkloosterman sum*, Compositio Mathematics., 40 (1980), pp. 3–33.

[15] S. SPERBER, *Newton polygons for general hyperkloosterman sums*, Astérisque, 119-120 (1984), pp. 267–330.

[16] S. SPERBER, *On the p-adic theory of exponential sums*, Amer. J. Math., 109 (1986), pp. 255–296.

[17] D. WAN, *Newton polygons and congruence decompositions of L-functions over finite fields*, Contemporary Mathematics, 133 (1992), pp. 221–241.

[18] D. WAN, *Newton polygons of zeta functions and L-functions*, Ann. Math., 137 (1993), pp. 247–293.

[19] D. WAN, *Dwork's conjecture on unit root zeta functions*, Ann. Math., 150 (1999), pp. 867–927.

[20] D. WAN, *Higher rank case of Dwork's conjecture*, J. Amer. Math., Soc., 13 (2000), pp. 807–852.

[21] D. WAN, *Rank one case of Dwork's conjecture*, J. Amer. Math., Soc., 13 (2000), pp. 853–908.

[22] R. YANG, *Newton polygons of L-functions of polynomials of the form $X^d + \lambda X$*, Finite Fields & Appl., 9 (2003), pp. 59–88.

[23] H. ZHU, *p-adic Variation of L-functions of exponential sums, I*, Amer. J. Math., 125 (2003), pp. 669–690.

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