LMS J. Comput. Math. 16 (2013) 9–44 © 2013 Authors
doi:10.1112/S1461157012001179

Computing zeta functions of nondegenerate hypersurfaces with few monomials

Steven Sperber and John Voight

Abstract

Using the cohomology theory of Dwork, as developed by Adolphson and Sperber, we exhibit a deterministic algorithm to compute the zeta function of a nondegenerate hypersurface defined over a finite field. This algorithm is particularly well suited to work with polynomials in small characteristic that have few monomials (relative to their dimension). Our method covers toric, affine, and projective hypersurfaces, and also can be used to compute the L-function of an exponential sum.

Let p be prime and let $\mathbb{F}_q$ be a finite field with $q = p^a$ elements. Let $\overline{V}$ be a variety defined over $\mathbb{F}_q$, described by the vanishing of a finite set of polynomial equations with coefficients in $\mathbb{F}_q$. We encode the number of points $\#\overline{V}(\mathbb{F}_{q^r})$ on $\overline{V}$ over the extensions $\mathbb{F}_{q^r}$ of $\mathbb{F}_q$ in an exponential generating series, called the *zeta function* of $\overline{V}$:

$$Z(\overline{V}, T) = \exp\left(\sum_{r=1}^{\infty} \#\overline{V}(\mathbb{F}_{q^r})\frac{T^r}{r}\right) \in 1 + T\mathbb{Z}[[T]].$$

The zeta function $Z(\overline{V}, T)$ is a rational function in T, a fact first proved using p-adic methods by Dwork [**16, 17**]. The algorithmic problem of computing $Z(\overline{V}, T)$ efficiently is of significant foundational interest, owing to many practical and theoretical applications (see, for example, Wan [**58**] for a discussion).

From a modern point of view, we consider $Z(\overline{V}, T)$ cohomologically: we build a p-adic cohomology theory that functorially associates to $\overline{V}$ certain vector spaces H^i over a p-adic field K, each equipped with a (semi)linear operator Frob_i, such that $Z(\overline{V}, T)$ is given by an alternating product of the characteristic polynomials of Frob_i acting on the spaces H^i. The theory of ℓ-adic étale cohomology, for example, was used by Deligne to show that $Z(\overline{V}, T)$ satisfies a Riemann hypothesis when $\overline{V}$ is smooth and projective. Parallel developments have followed in the p-adic (de Rham) framework, including the theories of Monsky and Washnitzer, crystalline, and rigid cohomology (see Kedlaya [**53**] for an introduction). In this paper, for a toric hypersurface $\overline{V}$ defined by a (nondegenerate) Laurent polynomial $\overline{f}$ in n variables over $\mathbb{F}_q$, we employ the cohomology theory of Dwork, working with a space $H^{n+1}(\Omega^\bullet)$ obtained as the quotient of a p-adic power series ring over K in $n+1$ variables by the subspace generated by the images of $n+1$ differential operators.

Efforts to make these cohomology theories computationally effective have been extensive. Schoof's algorithm for counting points on an elliptic curve [**54**] (generalized by Edixhoven and his coauthors [**22**] to compute coefficients of modular forms) can be viewed in this light, using the theory of mod ℓ étale cohomology. A number of results on the p-adic side have also emerged in recent years. In early work, Wan [**60**] and Lauder and Wan [**48**] demonstrated that the p-adic methods of Dwork can be used to efficiently compute zeta functions in small (fixed) characteristic. Lauder and Wan use the Dwork trace formula and calculate the trace of a Frobenius automorphism acting on a p-adic Banach space, following the original method

Received 21 December 2011; revised 29 June 2012.

2010 Mathematics Subject Classification 11Y16 (primary), 11M38, 14D10, 14F30 (secondary).

of Dwork and working on the 'chain level'. In this paper, we instead work with the extension of Dwork's theory due to Adolphson and Sperber [2]; this point of view was also pursued computationally by Lauder and Wan in the special case of Artin–Schreier curves [46, 47]. Under the hypothesis that the Laurent polynomial $\overline{f}$ is *nondegenerate* (see below for the precise definition), the zeta function can be recovered from the action of the Frobenius automorphism on a certain single cohomology space $H^{n+1}(\Omega)$. This method works with exponential sums and so extends naturally to the case of toric, affine, or projective hypersurfaces [3]. (It suffices to consider the case of hypersurfaces to compute the zeta function of any variety defined over a finite field using inclusion–exclusion or the Cayley trick.)

The method of Dwork takes into account the terms that actually occur in the Laurent polynomial $\overline{f}$; these methods are especially well suited when the monomial support of $\overline{f}$ is small, so that certain combinatorial aspects are simple. This condition that $\overline{f}$ has few monomials in its support, in which case we say (loosely) that $\overline{f}$ is *fewnomial* (a term coined by Kouchnirenko [41]), is a natural one to consider. For example, many explicit families of hypersurfaces of interest, including the well-studied (projective) Dwork family $x_0^{n+1} + \ldots + x_n^{n+1} + \lambda x_0 x_1 \ldots x_n = 0$ of Calabi–Yau hypersurfaces [18] (as well as more general monomial deformations of Fermat hypersurfaces [19]) can be written with few monomials. In cryptographic applications, the condition of fewnomialness also often arises. Finally, the running time of algorithms on fewnomial inputs are interesting to study from the point of view of complexity theory: see, for example, work of Bates, Bihan, and Sottile [6].

To introduce our result precisely, we now set some notation. Let $\overline{V}$ be a toric hypersurface, the closed subset of $\mathbb{G}_m^n$ defined by the vanishing of a Laurent polynomial

$$\overline{f} = \sum_{\nu \in \mathbb{Z}^n} \overline{a}_\nu x^\nu \in \mathbb{F}_q[x^\pm] = \mathbb{F}_q[x_1^\pm, \ldots, x_n^\pm].$$

We use multi-index notation, so $x^\nu = x_1^{\nu_1} \ldots x_n^{\nu_n}$. We sometimes write $Z(\overline{f}, T) = Z(\overline{V}, T)$. Let $\Delta = \Delta(\overline{f})$ be the *Newton polytope* of $\overline{f}$, the convex hull of its *support*

$$\operatorname{supp}(\overline{f}) = \{\nu \in \mathbb{Z}^n : \overline{a}_\nu \neq 0\}$$

in $\mathbb{R}^n$. For simplicity, we assume throughout that $\dim(\Delta) = n$. For a face $\tau \subseteq \Delta$, let $\overline{f}|_\tau = \sum_{\nu \in \tau} \overline{a}_\nu x^\nu$. Then we say $\overline{f}$ is *(Δ-)nondegenerate* if for all faces $\tau \subseteq \Delta$ (including Δ itself), the system of equations

$$\overline{f}|_\tau = x_1 \frac{\partial \overline{f}|_\tau}{\partial x_1} = \ldots = x_n \frac{\partial \overline{f}|_\tau}{\partial x_n} = 0$$

has no solution in $\overline{\mathbb{F}}_q^{\times n}$, where $\overline{\mathbb{F}}_q$ is an algebraic closure of $\mathbb{F}_q$. The set of Δ-nondegenerate polynomials with respect to a polytope Δ forms an open subset in the affine space parameterizing their coefficients $(\overline{a}_\nu)_{\nu \in \Delta \cap \mathbb{Z}^n}$: under a mild hypothesis, such as when Δ contains a unimodular simplex, then this subset is Zariski dense. (See Batyrev and Cox [5] as a reference for this notion as well as the work of Castryck and the second author [11] for a detailed analysis of nondegenerate curves.) We distinguish here between $\Delta(\overline{f})$ and $\Delta_\infty(\overline{f})$ which is the convex closure of $\Delta(\overline{f}) \cup \{0\}$: for the Laurent polynomial $w\overline{f}$ in $n+1$ variables, $\overline{f}$ is Δ-nondegenerate if and only if $w\overline{f}$ is nondegenerate with respect to $\Delta_\infty(\overline{f})$ in the sense of Kouchnirenko [40], Adolphson and Sperber [2], and others. Nondegenerate hypersurfaces are an attractive class to consider because many of their geometric properties can be deduced from the combinatorics of their Newton polytopes.

Let $s = \# \operatorname{supp}(\overline{f})$ and let U be the $(n+1) \times s$ matrix with entries in $\mathbb{Z}$ whose columns are the vectors $(1, \nu) \in \mathbb{Z}^{n+1}$ for $\nu \in \operatorname{supp}(\overline{f})$. Let ρ be the rank of U modulo p. Let $v = \operatorname{Vol}(\Delta) = n! \operatorname{vol}(\Delta)$ be the normalized volume of Δ, so that a unit hypercube $[0,1]^n$ has normalized volume $n!$ and the unit simplex $\sigma = \{(a_1, \ldots, a_n) \in \mathbb{R}^n_{\geqslant 0} : \sum_i a_i \leqslant 1\}$ has normalized volume 1.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

We say that Δ is *confined* if Δ is contained in an orthotope (box) with side lengths $b_1, \ldots, b_n$ with $b_1 \ldots b_n \leqslant n^n v$. We say that $\overline{f}$ is *confined* if $\Delta(\overline{f})$ is confined. A slight extension of a theorem of Lagarias and Ziegler [42] shows that every polytope Δ is $\mathrm{GL}_n(\mathbb{Z})$-equivalent to a confined polytope; this existence can also be made effective. (See § 3 for more detail.) In other words, for each Laurent polynomial $\overline{f}$ there is a computable monomial change of basis of $\mathbb{F}_q[x^{\pm}]$, giving rise to an equality of zeta functions, under which $\overline{f}$ is confined. (In the theorem below, at the expense of introducing a factor of $\log \delta$, where $\delta = \delta(S) = \max_{\nu \in S} |\nu|$ where $|\nu| = \max_i |\nu_i|$, one can remove the assumption that Δ is confined.)

For functions $f, g : \mathbb{Z}^m_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$, we say that $f = O(g)$ if there exists $c \in \mathbb{R}_{>0}$ and $N \in \mathbb{Z}_{\geqslant 0}$ such that for every $x = (x_1, \ldots, x_m) \in \mathbb{Z}^m_{\geqslant N}$ we have $g(x) \leqslant cf(x)$. (The reader is warned that not all properties familiar to big-Oh notation for functions of one variable extend to the multivariable case; see Howell [29]. In fact, our analysis also holds with Howell's more restrictive definition, but we will not pursue this further here.) We further use the 'soft-Oh' notation, where $f = \widetilde{O}(g)$ if $f = O(g \log^k g)$ for some $k \geqslant 1$.

Our main result is as follows.

THEOREM A. *Let $n \in \mathbb{Z}_{\geqslant 1}$. Then there exists an explicit algorithm that, on input of a nondegenerate Laurent polynomial $\overline{f} \in \mathbb{F}_q[x_1^{\pm}, \ldots, x_n^{\pm}]$ with $p \geqslant 3$ and an integer $N \geqslant 1$, computes, as output, $Z(\overline{f}, T)$ modulo p^N. If, further, $\overline{f}$ is confined, then this algorithm uses*

$$\widetilde{O}\big(s^{\lceil n/2 \rceil} + pN^3 \log q + p^{s-\rho}(6N + n)^s (v^4 N \log^2 q)\big)$$

bit operations.

To recover the zeta function (as an element of $\mathbb{Q}(T)$), if we fix both the dimension and the number s of monomials, we have the following result.

THEOREM B. *Let $n \in \mathbb{Z}_{\geqslant 1}$ and $s \in \mathbb{Z}_{>n}$. Then there exists an explicit algorithm that, on input of a confined, nondegenerate Laurent polynomial $\overline{f} \in \mathbb{F}_q[x_1^{\pm}, \ldots, x_n^{\pm}]$ with $p \geqslant 5$ and $s = \# \operatorname{supp}(\overline{f})$, computes, as output, $Z(\overline{f}, T)$ using*

$$\widetilde{O}(p^{\min(1, s-\rho)} v^{s+5} \log^{s+3} q)$$

bit operations.

According to a theorem of Adolphson and Sperber [2], under the hypothesis that $\overline{f}$ is nondegenerate, $Z(\overline{f}, qT)^{(-1)^n}$ is a polynomial of degree v times $Z(\mathbb{G}_m^n, T)^{(-1)^n}$. Therefore, in the context of Theorem B, if $p = O(v \log q)$ is small (or fixed), then our algorithm runs in polynomial time in the (dense) output size, which is the best one could hope for (aside from minimizing the degree of this polynomial). (The fewnomial input size, on the other hand, is $O(s \log v \log q)$ for $\overline{f}$ confined and n fixed.)

In our theorem, we require the dimension n to be fixed for several reasons. First, we employ well-known algorithms for lattice polytopes which have only been analyzed assuming that the dimension is fixed. (They further assume that arithmetic operations in $\mathbb{Z}$ take time $O(1)$, which is nearly valid in the usual bit-complexity model if Δ is confined and n is fixed; for a discussion of this point, see § 3.) Second, it is often quite natural from a geometric point of view to consider the dimension to be fixed; one often considers families of hypersurfaces of a fixed dimension, for example. Finally, allowing fewnomial input and output and varying dimension, the problem of computing $Z(\overline{f}, T)$ is harder than the NP-complete problem 3-SAT (indeed, for the latter one only wishes to know if $\# X(\mathbb{F}_2) > 0$ for an affine hypersurface X of degree 2). For these reasons, we restrict our analysis (continuing below) to fixed dimension.

Our method follows in the same vein as other recently introduced p-adic cohomological techniques. The methods of Lauder and Wan [48] mentioned above compute $Z(\overline{f}, T)$ for

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

a polynomial $\overline{f}$ of total degree d using $\widetilde{O}(p^{2n+4}d^{3n^2+9n}\log^{3n+7} q) = \widetilde{O}((pv\log q)^{3n+9})$ bit operations. The dense input size of $\overline{f}$ is $O((d+1)^n \log q)$; consequently, if the prime p (and dimension n) are fixed, then their algorithm runs in polynomial time in the dense input size. Their method, although apparently not practical, is completely general and does not require any hypothesis on $\overline{f}$. Our method can be analyzed on dense input (see § 5) as well, running in time $\widetilde{O}(p^{2n}v^{2n+4}\log^{2n+2})$ with no condition on the number of monomials in the support of $\overline{f}$.

In a different direction, Kedlaya [34] (see also the presentation by Edixhoven [21]) used Monsky–Washnitzer cohomology to compute the zeta function of a hyperelliptic curve of genus g over $\mathbb{F}_q$ in time $\widetilde{O}(pg^4 \log^3 q)$. (Note here that the dense input size is $O(g\log q)$.) This idea has been taken up by several others: see, for example, the work of Abbott, Kedlaya, and Roe [1], who compute the zeta function of a projective hypersurface by working in the complement of the hypersurface and using Mumford reduction. (Indeed, Kedlaya has suggested that there should be a natural extension [36] of his ideas to the realm of toric hypersurfaces.) Our method also mirrors the algorithm of Castryck, Denef, and Vercauteren [10], who tackle the case of nondegenerate curves. Their method has good asymptotic behavior but to be practical needs an optimized implementation [9, Section 1.2.4]. However, rather than following this vein and working with Monsky–Washnitzer (p-adic de Rham) cohomology, we employ the cohomology theory of Dwork, which has a more combinatorial flavor.

In a yet further direction, Lauder has used Dwork's theory of p-adic differential equations to compute zeta functions using deformation [43] and recursion [45]. The p-adic cohomology of a one-parameter family of hypersurfaces is endowed with a connection and a Frobenius endomorphism. By looking at a family with a simple model (such as a hypersurface defined by a diagonal polynomial), Lauder used one-variable deformation to compute the Frobenius automorphism on the fiber of interest. Our method fits into this framework as it provides natural base varieties to deform from: indeed, the idea of deformation in the context of nondegenerate curves has been pursued by Tuitman [57]. The methods of Lauder show that one can compute $Z(\overline{V}, T)$ for a smooth projective hypersurface $\overline{V} \subseteq \mathbb{P}^n$ of degree d with $p \nmid d$ and nonvanishing diagonal terms in time $p^2(d^n \log q)^{O(1)}$. The deformation method has also been pursued fruitfully by Gerkmann [25] and others in different contexts.

Adapting an idea of Chudnovsky and Chudnovsky and Bostan, Gaudry, and Schost [7] for accelerated reduction, Harvey [27] has improved Kedlaya's method for hyperelliptic curves, with a runtime of $p^{1/2}(g\log q)^{O(1)}$. This approach appears to extend to higher dimensions as well, extending the method of Abbott, Kedlaya, and Roe [28]: his method appears to give a runtime of $p^{1/2}(d^n \log q)^{n+O(1)}$ under a smoothness hypothesis analogous to the condition of nondegeneracy (but somewhat weaker). It would be interesting to see how his ideas for lowering the exponent on p might apply in our situation.

This paper is organized as follows. In § 1, we introduce the cohomology theory of Dwork and give an overview of our method. In § 2, we discuss each step of the algorithm in turn: computing the splitting function and the Jacobian ring, the computation of a Frobenius automorphism (where the condition of sparsity enters), and the reduction theory in cohomology. In § 3, we give some algorithms for computing with polytopes. We then discuss running time and precision estimates for the complete algorithm in § 4. Finally, in § 5 we discuss the case $p = 2$ and consider some other possible modifications. We conclude in § 6 with some examples.

1. *Overview*

In this section, we give an overview of our algorithm. Our introduction will be concise; for a more complete treatment of the theory of Dwork [17], see Koblitz [38], Lauder and Wan [48], and Monsky [51].

In this section, we assume $p > 2$; see § 5 for a discussion of the case $p = 2$.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Exponential sums

Let $\overline{f} \in \mathbb{F}_q[x_1^{\pm}, \dots, x_n^{\pm}]$ be a Laurent polynomial and let $\overline{V} \subseteq (\mathbb{G}_m)^n_{\mathbb{F}_q}$ be the toric hypersurface defined by the vanishing of $\overline{f}$. Let $\Theta : \mathbb{F}_q \to C$ be a nontrivial additive character (with C a commutative ring of characteristic zero), so that

$$\Theta(\overline{x} + \overline{y}) = \Theta(\overline{x})\Theta(\overline{y})$$

for all $\overline{x}, \overline{y} \in \mathbb{F}_q$. A point of departure for the theory of Dwork is the observation that, for $\overline{x} \in \mathbb{F}_q^{\times n}$, we have

$$\sum_{\overline{w} \in \mathbb{F}_q} \Theta(\overline{w}\overline{f}(\overline{x})) = \begin{cases} q & \text{if } \overline{f}(\overline{x}) = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Consequently

$$q\#\overline{V}(\mathbb{F}_q) = \sum_{\substack{\overline{w} \in \mathbb{F}_q \\ \overline{x} \in \mathbb{F}_q^{\times n}}} \Theta(\overline{w}\overline{f}(\overline{x})) = \sum_{(\overline{w}, \overline{x}) \in \mathbb{F}_q^{\times (n+1)}} \Theta(\overline{w}\overline{f}(\overline{x})) + (q-1)^n. \tag{1.1}$$

In other words, counting the set of points $\overline{V}(\mathbb{F}_q)$ can be achieved by instead evaluating an exponential sum (on either $\mathbb{A}^1 \times \mathbb{G}_m^n$ or $\mathbb{G}_m^{n+1}$).

For $r \in \mathbb{Z}_{\geqslant 1}$, we define a system of nontrivial additive characters $\Theta_r : \mathbb{F}_{q^r} \to C$ by $\Theta_r = \Theta \circ \mathrm{Tr}_r$ where $\mathrm{Tr}_r : \mathbb{F}_{q^r} \to \mathbb{F}_q$ is the trace map, and we define the exponential sums

$$S_r(w\overline{f}, \mathbb{G}_m^{n+1}) = \sum_{(\overline{w}, \overline{x}) \in \mathbb{F}_q^{\times (n+1)}} \Theta_r(\overline{w}\overline{f}(\overline{x})).$$

The L-function associated to $w\overline{f}$ over $\mathbb{G}_m^{n+1}$ is defined to be

$$L(w\overline{f}, \mathbb{G}_m^{n+1}, T) = \exp\left(\sum_{r=1}^{\infty} \frac{S_r(w\overline{f}, \mathbb{G}_m^{n+1})}{r} T^r\right).$$

Then by (1.1) we have

$$Z(\overline{V}, qT) = L(w\overline{f}, \mathbb{G}_m^{n+1}, T) Z((\mathbb{G}_m^n)_{\mathbb{F}_q}, T), \tag{1.2}$$

where

$$Z((\mathbb{G}_m^n)_{\mathbb{F}_q}, T)^{(-1)^{n+1}} = \prod_{i=0}^{n} (1 - q^i T)^{\binom{n}{i}(-1)^i}. \tag{1.3}$$

The Dwork splitting function and interpolation

Complex characters are defined via the exponential map, but the theory takes off when the ring C where the character takes values is a p-adic ring, and the exponential function does not have a large enough p-adic radius of convergence to be useful. We improve this radius of convergence by using a modified exponential function as follows. Let π be an element of the algebraic closure of $\mathbb{Q}_p$ that satisfies $\pi^{p-1} = -p$; then $\mathbb{Z}_p[\pi] = \mathbb{Z}_p[\zeta_p]$ where ζ_p is a primitive pth root of unity. We define the function

$$\theta(t) = \exp\left(\pi t + \frac{(\pi t)^p}{p}\right) = \exp(\pi(t - t^p)) = \sum_{i=0}^{\infty} \lambda_i t^i \in \mathbb{Q}_p[\pi][[t]]$$

which is called a *Dwork splitting function*. It is sometimes denoted by $\theta_1(t)$ to distinguish it from other splitting functions. We have

$$\mathrm{ord}_p\, \lambda_i \geqslant i(p-1)/p^2, \tag{1.4}$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

where ord_p is the p-adic valuation normalized so that $\operatorname{ord}_p p = 1$; thus, in fact, $\theta(t) \in \mathbb{Z}_p[\pi][[t]]$. We observe that $\theta(1) = 1 + \pi + O(\pi^2)$ is a primitive pth root of unity, and so we obtain our additive characters via the maps

$$\begin{aligned} \Theta_r : \mathbb{F}_{q^r} &\to \mathbb{Z}_p[\pi] \\ \Theta_r(\overline{x}) &= \theta(1)^{(\operatorname{Tr}\circ \operatorname{Tr}_r)(\overline{x})} \end{aligned}$$

where $\operatorname{Tr}\circ\operatorname{Tr}_r = \operatorname{Tr}_{\mathbb{F}_{q^r}/\mathbb{F}_p}$ is the absolute trace.

The values of the characters Θ_r can be p-adically interpolated in a way consistent with field extensions, as follows. Let $\mathbb{Q}_q$ be the unramified extension of $\mathbb{Q}_p$ of degree $a = \log_p q$, and let $\mathbb{Z}_q \subseteq \mathbb{Q}_q$ denote its ring of integers, so that $\mathbb{Z}_q$ is the Witt vectors over $\mathbb{F}_q$. Let $\sigma : \mathbb{Z}_q \to \mathbb{Z}_q$ denote the p-power Frobenius automorphism (lifting the pth power map on the residue field $\mathbb{F}_q$.) There is a canonical character ω, called the Teichmüller character, of the multiplicative group $\mathbb{F}_q^\times$ taking values in $\mathbb{Z}_q$ that takes an element $\overline{x} \in \mathbb{F}_q^\times$ to the element $x \in \mathbb{Z}_q$, satisfying $x^q = x$ and such that x reduces to $\overline{x}$ in $\mathbb{F}_q$. For such a Teichmüller representative $x \in \mathbb{Z}_q$ lifting $\overline{x}$, we find that

$$\Theta_1(\overline{x}) = \prod_{i=0}^{a-1} \theta(x^{p^i}) = \theta(x)\theta(x^p)\dots\theta(x^{q/p}) \in \mathbb{Z}_p[\pi]$$

and extending this for $r \in \mathbb{Z}_{\geqslant 1}$ we have

$$\Theta_r(\overline{x}) = \prod_{i=0}^{ar-1} \theta(x^{p^i}) \in \mathbb{Z}_p[\pi]$$

for $\overline{x} \in \mathbb{F}_{q^r}$ and x a Teichmüller lift of $\overline{x}$. Let $\sigma : \mathbb{Z}_q \to \mathbb{Z}_q$ by $x \mapsto x^\sigma$ denote the p-power Frobenius automorphism, the ring automorphism of $\mathbb{Z}_q$ that reduces to the map $\overline{x} \mapsto \overline{x}^p$ modulo p. Then

$$\Theta_r(\overline{x}) = \prod_{i=0}^{ar-1} \theta(x^{\sigma^i}). \tag{1.5}$$

We now consider these character values applied to values of our Laurent polynomial $\overline{f}$. Write $\overline{f}(x) = \sum_\nu \overline{a}_\nu x^\nu \in \mathbb{F}_q[x^\pm]$ in multi-index notation; we assume that each $\overline{a}_\nu \neq 0$. Let $f(x) = \sum_\nu a_\nu x^\nu \in \mathbb{Z}_q[x^\pm]$, where a_ν is the Teichmüller lift of $\overline{a}_\nu$. In light of (1.5), we are led to consider the power series

$$F(w,x) = \prod_\nu \theta(w a_\nu x^\nu) \in \mathbb{Z}_q[\pi][[w, x^\pm]]$$

and

$$F^{(a)}(w,x) = \prod_{i=0}^{a-1} F^{\sigma^i}(w^{p^i}, x^{p^i}) \in \mathbb{Z}_q[\pi][[w, x^\pm]]$$

where F^σ denotes the power series obtained by applying σ to the coefficients of F. (The abuse of notation which identifies a power series and its specializations will only occur in this paragraph.) It then follows from (1.5) and a straightforward calculation that

$$\Theta_r(\overline{w}\overline{f}(\overline{x})) = F^{(a)}(w,x)F^{(a)}(w^q,x^q)\dots F^{(a)}(w^{q^{r-1}}, x^{q^{r-1}}) \in \mathbb{Z}_p[\pi] \tag{1.6}$$

for all $(\overline{w}, \overline{x}) \in \mathbb{F}_{q^r} \times \mathbb{F}_{q^r}^{\times n}$, where (w,x) denotes the Teichmüller lift. We have thereby extended the interpolation of the values of $\overline{f}$ to the power series (1.6).

Dwork trace formula

So far, we have related the zeta function to the L-function of an exponential sum via a p-adic additive character arising from the Dwork splitting function, and we have interpolated these

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

character values in a power series $F = F(w, x)$. In order to move this to cohomology, we define a space of p-adic analytic functions like F with similar support and p-adic growth.

Let $\Delta = \Delta(\overline{f})$ be the *Newton polytope* of $\overline{f}$, the convex hull of its *support*

$$\operatorname{supp}(\overline{f}) = \{\nu \in \mathbb{Z}^n : \overline{a}_\nu \neq 0\}.$$

For $d \in \mathbb{R}_{\geqslant 0}$, let $d\Delta = \{dz \in \mathbb{R}^n : z \in \Delta\}$ denote the dth dilation of Δ. Let L_Δ be the ring

$$L_\Delta = \left\{ \sum_{d=0}^{\infty} \sum_{\nu \in d\Delta \cap \mathbb{Z}^n} c_{d,\nu} w^d x^\nu : c_{d,\nu} \in \mathbb{Z}_q[\pi] \text{ and } \operatorname{ord}_p(c_{d,\nu}) \geqslant d\frac{p-1}{p^2} \right\}. \tag{1.7}$$

The estimate (1.4) implies that $F \in L_\Delta$, and so multiplication by F defines a linear operator which we also denote $F : L_\Delta \to L_\Delta$.

On the space L_Δ, we have a 'left inverse of the Frobenius automorphism' $\psi : L_\Delta \to L_\Delta$ defined by

$$\psi(c_{d,\nu} w^d x^\nu) = \begin{cases} \sigma^{-1}(c_{d,\nu}) w^{d/p} x^{\nu/p} & \text{if } p \mid d \text{ and } p \mid \nu, \\ 0 & \text{otherwise;} \end{cases}$$

in multi-index notation, the condition $p \mid \nu$ means $p \mid \nu_i$ for all $i = 1, \ldots, n$. The map ψ is σ^{-1}-semilinear as a map of free $\mathbb{Z}_q$-modules.

Finally, let $\alpha = \psi \circ F$ and $\alpha_a = \psi^a \circ F^{(a)}$. Then α_a is $\mathbb{Z}_q$-linear, and another calculation reveals in fact that $\alpha_a = \alpha^a$ (composition a times).

The *Dwork trace formula* [16] then asserts that

$$S_r(w\overline{f}, \mathbb{G}_m^{n+1}) = (q^r - 1)^{n+1} \operatorname{Tr}(\alpha_a^r).$$

It follows [2] that

$$L(w\overline{f}, \mathbb{G}_m^{n+1}, T)^{(-1)^n} = \prod_{j=0}^{n+1} \det(1 - (q^j T)\alpha_a^j \mid L_\Delta)^{(-1)^j \binom{n+1}{j}}. \tag{1.8}$$

The equality (1.8) expresses $L(w\overline{f}, \mathbb{G}_m^{n+1}, T)$ via the action of α_a and its powers on an (infinite-dimensional) p-adic Banach space; this is the point of departure for Lauder and Wan in their work [48].

We now proceed one step further and move to the level of cohomology.

Dwork cohomology

We now consider a Koszul complex $\Omega^\bullet$ associated to wf as follows. To ease notation in this subsection, let $x_0 = w$. For $i = 0, \ldots, n$, let

$$f_i = x_i \frac{\partial(x_0 f)}{\partial x_i} \tag{1.9}$$

and define the operator $D_i : L_\Delta \to L_\Delta$ by

$$D_i = x_i \frac{\partial}{\partial x_i} + \pi f_i$$

(the latter is the operator given by multiplication by πf_i). The operators D_i commute. For $k = 0, \ldots, n$, let

$$\Omega^k = \bigoplus_{0 \leqslant j_1 < \ldots < j_k \leqslant n} L_\Delta \left(\frac{dx_{j_1}}{x_{j_1}} \wedge \ldots \wedge \frac{dx_{j_k}}{x_{j_k}} \right) \cong L_\Delta^{\binom{n+1}{k}}. \tag{1.10}$$

Let $\Omega^\bullet$ be the complex

$$0 \to \Omega^0 \to \Omega^1 \to \cdots \to \Omega^{n+1} \to 0$$

with maps

$$\nabla\left(\xi\frac{dx_{j_1}}{x_{j_1}}\wedge\ldots\wedge\frac{dx_{j_k}}{x_{j_k}}\right)=\sum_{i=0}^{n}(D_i\xi)\frac{dx_{j_i}}{x_{j_i}}\wedge\frac{dx_{j_1}}{x_{j_1}}\wedge\ldots\wedge\frac{dx_{j_k}}{x_{j_k}}$$

for $\xi\in L_\Delta$.

Now α induces a map on the complex $\Omega^\bullet$:

$$\begin{tikzcd}
0 \arrow[r] & \Omega^0 \arrow[r] \arrow[d, "p^{n+1}\alpha"] & \Omega^1 \arrow[r] \arrow[d, "p^n\alpha"] & \cdots \arrow[r] & \Omega^{n+1} \arrow[r] \arrow[d, "\alpha"] & 0 \\
0 \arrow[r] & \Omega^0 \arrow[r] & \Omega^1 \arrow[r] & \cdots \arrow[r] & \Omega^{n+1} \arrow[r] & 0
\end{tikzcd}$$

since one checks that $\alpha D_i = pD_i\alpha$ for all i. (One similarly has a map induced by α_a, replacing p by q.)

Then from [2] we have

$$L(w\overline{f},\mathbb{G}_m^{n+1},T)^{(-1)^n}=\prod_{j=0}^{n+1}\det(1-\alpha_aT\mid H^j(\Omega^\bullet))^{(-1)^{n+1-j}}. \tag{1.11}$$

The condition that $\overline{f}$ is nondegenerate simplifies the expression (1.11), as we now see.

Nondegenerate

We recall our notation from the introduction. For a face $\tau\subseteq\Delta$, let $\overline{f}|_\tau=\sum_{\nu\in\tau}\overline{a}_\nu x^\nu$. Then we say $\overline{f}$ is *nondegenerate* if, for all faces $\tau\subseteq\Delta$ (of any dimension, including Δ itself), the system of equations

$$\overline{f}|_\tau = x_1\frac{\partial\overline{f}|_\tau}{\partial x_1}=\ldots=x_n\frac{\partial\overline{f}|_\tau}{\partial x_n}=0$$

has no solution in $\overline{\mathbb{F}}_q^{\times n}$, where $\overline{\mathbb{F}}_q$ is an algebraic closure of $\mathbb{F}_q$.

Suppose $\overline{f}$ is nondegenerate. Then all the cohomology spaces $H^k(\Omega^\bullet)$ are trivial except for $k=n+1$. Let

$$B=\frac{L_\Delta}{D_0L_\Delta+D_1L_\Delta+\ldots+D_nL_\Delta}\cong H^{n+1}(\Omega^\bullet).$$

Then by work of Adolphson and Sperber [2], the $\mathbb{Z}_q[\pi]$-module B is free of dimension $n!\,\mathrm{vol}(\Delta)=v$ (equal to the normalized volume of the cone over Δ in $\mathbb{R}^{n+1}$) and

$$L(w\overline{f},\mathbb{G}_m^{n+1},T)^{(-1)^n}=\det(1-\alpha_aT\mid B)\in 1+T\mathbb{Z}[T] \tag{1.12}$$

(using (1.2)).

Let A (respectively A_a) be a matrix of α (respectively α_a) acting on B. Then we have

$$A_a = AA^{\sigma^{-1}}\ldots A^{\sigma^{-(a-1)}} \tag{1.13}$$

where A^σ denotes the matrix where σ is applied to each entry in the matrix.

An overview of the algorithm

We now describe how to effectively compute the terms in the formula (1.12), and in particular the matrix A (1.13). We sketch an overview and wait to describe each of these steps and their running time in detail in the sections that follow.

The algorithm takes as input a nondegenerate (confined) Laurent polynomial $\overline{f}$ and a precision $N\in\mathbb{Z}_{\geqslant 0}$, and it produces as output the polynomial $\det(1-\alpha_aT\mid B)$ modulo p^N. For N large, we recover the coefficients in $\mathbb{Z}$ and then from (1.2) we recover $Z(\overline{f},T)$.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Let $R = \mathbb{Z}_q/p^N$. (By carefully factoring out the algebraic element π, we may work in this smaller ring; see Lemma 2.12 and the accompanying discussion.)

In (1.7) we have worked with the power series ring L_Δ, but by the convergence behavior of elements of L_Δ, working modulo p^N these power series become polynomials. So we define

$$R[w\Delta] = \bigoplus_{d=0}^{\infty} R[w\Delta]_d$$

where

$$R[w\Delta]_d = \bigoplus_{\nu \in d\Delta \cap \mathbb{Z}^n} R w^d x^\nu.$$

The ring $R[w\Delta]$ is the monoid algebra arising from the cone over Δ with coefficients in R, and it is naturally $\mathbb{Z}_{\geqslant 0}$-graded by w. Recall from (1.9) that we have defined

$$f_i = w x_i \frac{\partial f}{\partial x_i}$$

for $i = 1, \dots, n$ (and $f_0 = wf$).

Our algorithm has 4 steps.

(1) Compute the Teichmüller lift f of $\overline{f}$. Using linear algebra over R, compute a monomial basis V for the *Jacobian ring*

$$J = R[w\Delta]/(wf, wf_1, \dots, wf_n).$$

(The monomial basis V for J yields a basis for B.)

(2) For each monomial $m \in V$, compute the action of the Frobenius automorphism $\alpha(m)$ using 'fewnomial enumeration'.

(3) For each $m \in V$, reduce $\alpha(m)$ in cohomology using the differential operators D_i to an element in the R-span of V. (The matrices implicitly computed in Step 1 are used in this reduction.)

(4) Compute the resulting matrix $A = \alpha \mid B$ modulo p^N, then compute A_a using (1.13) and finally

$$Z(V, qT) = \det(1 - TA_a)^{(-1)^n} Z((\mathbb{G}_m^n)_{\mathbb{F}_q}, T).$$

Output $Z(V, T)$.

2. *The algorithm*

We now describe each step of the algorithm announced in our main theorem and introduced in § 1. We retain the notation from § 1.

Step 1*: Computing the Jacobian ring*

We begin by describing the computation of a basis of the Jacobian ring

$$J = \frac{R[w\Delta]}{(wf, wf_1, \dots, wf_n)};$$

as result of this computation, we also obtain matrices which will be used in the reduction step in Step 3.

LEMMA 2.1. *If f is nondegenerate, then J is a free R-module with basis of cardinality $v = \mathrm{Vol}(\Delta)$ consisting of monomials with degree $\leqslant n + 1$.*

Proof. The proof of Monsky [**51**] (see work of Adolphson and Sperber [**2**, Appendix]) shows that under the nondegeneracy hypothesis, the associated Koszul complex is acyclic modulo p

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

and therefore lifts to an acyclic complex over R (as the modules are complete, separated, and flat over $\mathbb{Z}_q$). □

To compute with the ring $R[w\Delta]$, we need some standard algorithms for computing with polytopes. For a set $S \subseteq \mathbb{Z}^n$, we denote by $\Delta(S)$ its convex hull.

LEMMA 2.2. *There exists an efficient algorithm that, given a finite set $S \subseteq \mathbb{Z}^n$, computes the set $\mathbb{Z}^n \cap \Delta(S)$.*

We discuss this algorithm and its running time in detail in the next section (Proposition 3.5); it will be treated as a black box for now.

Let $\prec$ be a term order on the monomials in $R[w\Delta]$ that respects the w-grading. We begin by computing in each degree $d = 0, \ldots, n+2$ the set of monomials in $d\Delta$ using Lemma 2.2 and we order them by $\prec$. Then, for each such d a spanning set for the degree d subspace of the Jacobian ideal, $(wf, wf_1, \ldots, wf_n)_d$, is given by the products of the monomials in $(d-1)\Delta$ with the generators $wf, wf_1, \ldots, wf_n$ of the Jacobian ideal. Finally, for each d, let J_d be the matrix whose columns are indexed by $\mathbb{Z}^n \cap d\Delta$, that is, the monomial basis for $R[w\Delta]_d$ ordered by $\prec$, and whose rows record the coefficients of the spanning set for $(wf, wf_1, \ldots, wf)_d$.

We then compute the row-echelon form $M_d = T_d J_d$ of J_d for $d = 0, \ldots, n+2$ using linear algebra over R. According to Lemma 2.1, every pivot in M_d can be taken to be a unit in R; thus, a monomial basis V for J is then obtained as $\bigcup_d V_d$ where V_d is a choice of monomial basis for the cokernels of M_d. In particular, the matrix M_{n+2} has full rank and so has a maximal square submatrix with unit determinant.

Step 2: Computing the action of the Frobenius automorphism

Recall we have defined the Dwork splitting function

$$\theta(t) = \exp(\pi(t - t^p)) = \sum_{i=0}^{\infty} \lambda_i t^i \in \mathbb{Z}_p[\pi][[t]]$$

with $\operatorname{ord}_p \lambda_i \geqslant i(p-1)/p^2$. The image of $\theta(t)$ modulo p^N in $R[\pi][[t]]$ is a polynomial of degree less than $Np^2/(p-1) = N(p+1+1/(p-1))$. We compute $\theta(t) = \exp(\pi t)\exp(-\pi t^p)$ as the product of two polynomials of this degree.

REMARK 2.3. Here we have used that p is odd. For $p = 2$, the splitting function $\theta(t)$ above does not converge sufficiently fast for the algorithm described below to work. For the modifications necessary, see § 5.

We have $f = \sum_\nu a_\nu x^\nu \in \mathbb{Z}_q[x^{\pm}]$ with $a_\nu \neq 0$ and $\nu \in \mathbb{Z}^n$ and

$$F(w, x) = \prod_\nu \theta(a_\nu w x^\nu). \tag{2.4}$$

One option is to multiply out the product (2.4) naively. Instead, we seek to take advantage of the fewnomialness of f. We have

$$F(w, x) = \prod_\nu \theta(a_\nu w x^\nu) = \prod_\nu \left(1 + \lambda_1(a_\nu w x^\nu) + \ldots + \lambda_j (a_\nu w x^\nu)^j + \ldots\right). \tag{2.5}$$

Let $\operatorname{supp}(f) = \{\nu_1, \ldots, \nu_s\}$ and abbreviate $a_i = a_{\nu_i}$. Expanding (2.5) out we obtain

$$F(w, x) = \sum_{(e_1,\ldots,e_s)\in\mathbb{Z}^s_{\geqslant 0}} (\lambda_{e_1} \ldots \lambda_{e_s})\,(a_1^{e_1} \ldots a_s^{e_s})\, w^{e_1+\ldots+e_s} x^{e_1\nu_1+\ldots+e_s\nu_s}. \tag{2.6}$$

We make further abbreviations using multi-index notation as follows: for $e \in \mathbb{Z}^s_{\geqslant 0}$, which we abbreviate to $e \geqslant 0$, we write $\lambda_e = \lambda_{e_1} \ldots \lambda_{e_s}$ and $a^e = a_1^{e_1} \ldots a_s^{e_s}$, and we

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

write $|e| = e_1 + \ldots + e_s$, and finally $e\nu = e_1\nu_1 + \ldots + e_s\nu_s$ for the dot product. Then (2.6) becomes simply

$$F(w, x) = \sum_{e \geqslant 0} \lambda_e a^e w^{|e|} x^{e\nu}. \tag{2.7}$$

Let $w^d x^\mu \in V$. Then from (2.7) we have

$$\alpha(w^d x^\mu) = (\psi \circ F)(w^d x^\mu) = \sum_{\substack{e \geqslant 0 \\ p|(e\nu+\mu) \\ p|(|e|+d)}} \lambda_e \sigma^{-1}(a^e) w^{(|e|+d)/p} x^{(e\nu+\mu)/p} \in L_\Delta. \tag{2.8}$$

The set of indices for the sum on the right side of (2.8) is then contained in the set

$$E_{d,\mu} = \{e \in \mathbb{Z}^s_{\geqslant 0} : e_i\nu_i \equiv -\mu_i \pmod{p} \text{ for all } i \text{ and } |e| \equiv -d \pmod{p}\}. \tag{2.9}$$

When the set $E_{d,\mu}$ is proportionally fewnomial relative to the set of all monomials, we can hope to be able to enumerate it faster than multiplying out F completely.

Remark 2.10. On the other hand, if f is dense, in general we gain no advantage using this approach, as can be seen by the following example. Let f be a generic univariate polynomial of degree s, so that $n = 1$ and $\operatorname{supp}(f) = \{0, 1, \ldots, s\}$, and let $\mu \in \mathbb{Z}_{>0}$. Then there is a bijection between $E_{d,\mu}$ and the set of all (integer) partitions of $d \equiv \mu \pmod{p}$ into parts of size no greater than s. Since the number of such partitions grows exponentially with d, enumerating all such partitions would be prohibitively time consuming.

We can compute the set $E_{d,\mu}$ by considering the corresponding set of linear equations modulo p. Let U denote the $(n+1) \times s$ matrix whose columns are the vectors $(1, \nu_i)^t$, and let

$$K = K(d, \mu) = \{e : Ue \equiv (d, \mu)^t \pmod{p}\} \subseteq (\mathbb{Z}/p\mathbb{Z})^s.$$

We identify K with its image in $\mathbb{Z}^s_{\geqslant 0}$ by taking the smallest nonnegative residue in each component. Then (2.9) becomes simply

$$E_{d,\mu} = K + (p\mathbb{Z}_{\geqslant 0})^s.$$

Let ρ be the rank of U modulo p. Then $\#K = p^{s-\rho}$, since by our assumption $\dim(\Delta) = n$ so $s \geqslant n + 1$.

Rewriting (2.8), with this notation we obtain

$$\begin{aligned}\alpha(w^d x^\mu) &= \sum_{k \in K} \sum_{e \geqslant 0} \lambda_{k+pe} \sigma^{-1}(a^{k+pe}) w^{(|k|+p|e|+d)/p} x^{((k+pe)\nu+\mu)/p} \\ &= \sum_{k \in K} \sigma^{-1}(a^k) w^{(|k|+d)/p} x^{(k\nu+\mu)/p} \left(\sum_{e \geqslant 0} \lambda_{k+pe} a^e w^{|e|} x^{e\nu} \right).\end{aligned} \tag{2.11}$$

Here we used the fact that $\sigma^{-1}((a^p)^e) = a^e$ since a is a Teichmüller element. We then compute $\alpha(w^d x^\mu)$ using the formula (2.11).

We make one final substitution, which will simplify the reduction: we carefully factor out the algebraic element π (satisfying $\pi^{p-1} = -p$) as follows.

Lemma 2.12. *We have* $\operatorname{ord}_p \lambda_i \equiv i/(p-1) \in \mathbb{Q}/\mathbb{Z}$.

Proof. The ith coefficient of both $\exp(\pi t)$ and $\exp(-\pi t^p)$ satisfy the congruence, and consequently the same is true of the product. □

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Therefore, in the expansion $\theta(t) = \sum \lambda_i t^i$ we write $\lambda_i = \pi^i \ell_i$ so that $\ell_i \in \mathbb{Q}_p$ and

$$\operatorname{ord}_p(\ell_i) \geqslant i\left(\frac{p-1}{p^2} - \frac{1}{p-1}\right) = -i\frac{2p-1}{p^2(p-1)}. \tag{2.13}$$

Although we introduce some denominators here, they are controlled. Extending our multi-index notation, we obtain

$$\alpha(w^d x^\mu) = \sum_{k \in K} \sigma^{-1}(a^k)(\pi w)^{(|k|+d)/p} x^{(k\nu+\mu)/p}\left(\sum_{e \geqslant 0} \pi^{|k|+p|e|-|e|-(|k|+d)/p} \ell_{k+pe} a^e (\pi w)^{|e|} x^{e\nu}\right).$$

Since $\pi^{p-1} = -p$, we write

$$|k| + p|e| - |e| - (|k|+d)/p = (p-1)|e| - d + (p-1)(|k|+d)/p.$$

Thus

$$\begin{aligned}\alpha((\pi w)^d x^\mu) = \sum_{k \in K} &\sigma^{-1}(a^k)(-p)^{(|k|+d)/p}(\pi w)^{(|k|+d)/p} x^{(k\nu+\mu)/p} \\ &\cdot\left(\sum_{e \geqslant 0} (-p)^{|e|} \ell_{k+pe} a^e (\pi w)^{|e|} x^{e\nu}\right).\end{aligned} \tag{2.14}$$

(We introduce the power π^d to simplify this expansion; we may then just divide out at the end.)

Step 3: Reducing in cohomology

Our goal in this step is to determine the characteristic polynomial of the Frobenius automorphism α acting on

$$H^{n+1}(\Omega^{\bullet}) = L_\Delta \Big/ \sum_{i=0}^{n} D_i L_\Delta$$

modulo p^N, where N is the desired precision. The preceding analysis, including Lemma 2.12 and particularly the expression (2.14), shows that α acting on $(\pi w)^d x^\nu$ with $\nu \in d\Delta$ is a series in terms of the form $(\pi w)^e x^\mu$ with $\mu \in e\Delta$ and coefficients in $\mathbb{Z}_q$. The p-adic behavior of the Frobenius automorphism assures us that the coefficients of $(\pi w)^e x^\mu$ tend to 0 in $\mathbb{Z}_q$ as $|e| \to \infty$. Therefore, modulo p^N, the image $\alpha((\pi w)^d x^\nu)$ is a polynomial $G \in R[(\pi w)\Delta]$, a linear combination of terms $(\pi w)^e x^\mu$ $(\mu \in e\Delta)$ with coefficients in $\mathbb{Z}_q/p^N$. Since the factor π enters only formally, we may suppress the π factor and view the resulting polynomial in $R[w\Delta]$. From Lemma 2.1, we work with coefficients in $\mathbb{Z}_q/p^N$, rather than using the iterative reduction method described by Adolphson and Sperber [**2**, Theorem 2.18].

Therefore, let $G \in R[(\pi w)\Delta]$; we show how to reduce G in cohomology. Let $\operatorname{lm}(G)$ be the leading monomial (highest degree monomial) in G with respect to $\prec$. First suppose that the degree of $\operatorname{lm}(G)$ (in w) is at least $n+2$. Let m_0 be a monomial in R_{n+2} that divides $\operatorname{lm}(G)$ (a precise choice will be given in the next section), and let $m = \operatorname{lm}(G)/m_0$. Let $G^{(m)}$ consist of those terms in G in $mR[w\Delta]_{n+2}$ and identify $\xi = G^{(m)}/m \subseteq R[w\Delta]_{n+2}$ with a row vector indexed by the monomials of $R[w\Delta]_{n+2}$, each containing the term $(\pi w)^{n+2}$ by our convention.

Let $\eta = \xi T_d$. Since J_{n+2} is of full rank, we have

$$\xi = \xi M_d = (\xi T_d) J_d = \eta J_d.$$

Thus, with the natural identifications, we have written

$$\xi = \eta_0(\pi w) f_0 + \eta_1(\pi w) f_1 + \ldots + \eta_n(\pi w) f_n \tag{2.15}$$

with $\eta_0, \ldots, \eta_n \in R[(\pi w)\Delta]_{n+1} = \pi^{n+1} R[w\Delta]$, and therefore

$$(m\eta_0)(\pi w) f_0 + \ldots + (m\eta_n)(\pi w) f_n = m\xi = G^{(m)}.$$

Recall that we work in the module $B = L/\sum_i D_i L$, where $D_i = x_i\partial/\partial x_i + (\pi w)f_i$ (and we again set $w = x_0$ for convenience): this implies that in B, we have the relation

$$G^{(m)} \equiv -\left(x_0\frac{\partial(m\eta_0)}{\partial x_0} + \ldots + x_n\frac{\partial(m\eta_n)}{\partial x_n}\right) \in B. \tag{2.16}$$

Note now that all terms in the reduction (2.16) have degree one smaller than that of $G^{(m)}$. We then iterate this procedure on the leading monomial in $G - G^{(m)}$ until it is equivalent in B to a polynomial of degree no greater than $n+1$.

So now assume that G has degree $d \leqslant n+1$, and let G_d be the degree d terms of G. We repeat the above procedure, with $\xi = G_d$: we compute $\eta = \xi T_d$ and let $v_d = \xi - \eta$. Note that v_d is in the span of V by construction. Then by a similar calculation as in (2.15), we have $\xi = v_d + \sum_{i=0}^{n}\eta_i(\pi w)f_i$ and so

$$G_d = \xi \equiv v_d - \sum_{i=0}^{n} x_i\frac{\partial\eta_i}{\partial x_i} \in B.$$

Completing the final iteration on decreasing d, we obtain $G \equiv \sum_{d=0}^{n+1} v_d \in B$ written in the span of V.

Remark 2.17. In this reduction theory, we never perform a division: we simply reduce the power of πw, as we have written G as a polynomial in πw and x. Consequently, we do not lose precision in our analysis.

Step 4: Output

To finish, we assemble the reductions of $\alpha(w^d x^\mu)$ for $w^d x^\mu \in V$ into a square matrix $A = \alpha \mid B$ of size $v \times v$, where $v = \#V$, with coefficients in R. We then compute

$$A_a = AA^{\sigma^{-1}} \ldots A^{\sigma^{-(a-1)}}.$$

We then compute $\det(1 - TA_a)$ using standard methods (analyzed in § 4), and we recover the zeta function from (1.2)–(1.3): we compute

$$Z(\overline{V}, qT) = \det(1 - TA_a)^{(-1)^n} Z((\mathbb{G}_m^n)_{\mathbb{F}_q}, T)$$

from which we easily obtain the desired output $Z(\overline{V}, T)$.

3. *Some algorithms for polytopes*

In this section, we describe some methods for computing with polytopes which are used as subroutines.

Confined polytopes

By *polytope* we will always mean a lattice polytope. Let $\Delta \subseteq \mathbb{R}^n$ be a polytope with $\dim \Delta = n$ and normalized volume $v = \mathrm{Vol}(\Delta)$. (If $\dim \Delta < n$ then by restricting to the linear space that contains Δ one can make appropriate modifications to the algorithms below.) The group $\mathrm{GL}_n(\mathbb{Z})$ acts on $\mathbb{R}^n$ preserving the set of polytopes; we say two polytopes Δ, Δ' are $\mathrm{GL}_n(\mathbb{Z})$-*equivalent* if there exists $U \in \mathrm{GL}_n(\mathbb{Z})$ such that $U(\Delta) = \Delta'$.

By work of Lagarias and Ziegler [**42**, Theorem 2], any polytope $\Delta \subseteq \mathbb{R}^n$ is $\mathrm{GL}_n(\mathbb{Z})$-equivalent to a polytope contained in a lattice hypercube of side length at most nv. We now prove a slight extension of this result using their methods.

Lemma 3.1. *Any polytope $\Delta \subseteq \mathbb{R}^n$ is $\mathrm{GL}_n(\mathbb{Z})$-equivalent to a polytope contained in a lattice orthotope (box) with side lengths $b_1, \ldots, b_n$ satisfying $b_1 \ldots b_n \leqslant n^n v$.*

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Proof. We follow the proof of Lagarias and Ziegler [42, Proof of Theorem 2] (working always with normalized volume); for convenience, we reproduce the main idea of their proof. First suppose that the polytope is a simplex Σ having vertices $v_0, \ldots, v_n \in \mathbb{Z}^n$. Then the lattice Λ spanned by the basis vectors $w_i = v_i - v_0$ is a sublattice of $\mathbb{Z}^n$ with $\det(\Lambda) = \mathrm{Vol}(\Delta) = v$. There is a matrix $U \in \mathrm{GL}_n(\mathbb{Z})$ taking the matrix B whose column vectors are w_i to its Hermite normal form

$$UB = \begin{pmatrix} a_{11} & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

with $0 \leqslant a_{ji} < a_{ii}$ for $j > i$ and $a_{ii} > 0$ for all i and $a_{ji} = 0$ for $j < i$. Now $\det(\Lambda) = |\det(B)| = \prod_{i=1}^n a_{ii} = v$ and hence the parallelepiped generated by the row vectors of UB is contained in the orthotope

$$\Xi = \{x \in \mathbb{R}^n : 0 \leqslant x_i \leqslant a_{ii} \text{ for } 1 \leqslant i \leqslant n\} = [0, a_{11}] \times \ldots \times [0, a_{nn}].$$

The simplex $U\Sigma$ is contained in this parallelepiped and hence also the translated orthotope $\Xi + Uv_0$.

Suppose Δ is now an arbitrary polytope. Then (see [42, Theorem 3]) there exists a maximal volume simplex $\Sigma \subseteq \Delta$, and, moreover, Δ is contained in the simplex $-n\Sigma + (n+1)z$ where $z = \sum_{i=0}^n v_i$ is the centroid of Σ and $(n+1)z \in \mathbb{Z}^n$. By the above, Σ is $\mathrm{GL}_n(\mathbb{Z})$-equivalent to a simplex contained in an orthotope Ξ and hence Δ is equivalent to a polytope contained in the orthotope $-n\Xi + (n+1)z$ with side lengths $b_1, \ldots, b_n$ with $b_i = na_{ii}$ so $\prod_{i=1}^n b_i = n^n v$, as claimed. □

We say that Δ is *confined* if Δ is contained in an orthotope (box) with side lengths $b_1, \ldots, b_n$ with $b_1 \ldots b_n \leqslant n^n v$. We say that $\overline{f}$ is *confined* if $\Delta(\overline{f})$ is confined.

Remark 3.2. The problem of finding a maximum volume simplex, the key to making the proof of Lemma 3.1 computationally effective, has been studied in detail (see, for example, work of Gritzmann, Klee, and Larman [26]). To avoid going too far afield into the field of computational geometry, in this article we accept any input Laurent polynomial $\overline{f}$ but only analyze the runtime when $\overline{f}$ is confined.

Corollary 3.3. *We have* $\#(\Delta \cap \mathbb{Z}^n) \leqslant (2n)^n v$.

Proof. The action of $\mathrm{GL}_n(\mathbb{Z})$ preserves lattice points, and so, if Ξ is the orthotope given by Lemma 3.1, then

$$\#(\Delta \cap \mathbb{Z}^n) \leqslant \#(\Xi \cap \mathbb{Z}^n) = (b_1 + 1) \ldots (b_n + 1) \leqslant (2b_1) \ldots (2b_n) \leqslant 2^n (n^n v)$$

as claimed. □

Remark 3.4. Corollary 3.3 is in some sense best possible, since equality holds for $\Delta = [0, 1] \subseteq \mathbb{R}$. For fixed n, one can do no better than $\#(\Delta \cap \mathbb{Z}^n) = O(v)$ since for any polytope Δ we have $\#(d\Delta \cap \mathbb{Z}^n) \sim \mathrm{Vol}(d\Delta)/n!$ as $d \to \infty$. For a given n, one can improve the constant $(2n)^n$ significantly working in a more general context (see Widmer [62]).

Enumerating lattice points

Our next task (arising in Step 1) is to exhibit an algorithm to enumerate the lattice points in the convex hull of a set of lattice points.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Let $S \subseteq \mathbb{Z}^n$ be a nonempty finite set. Let $\Delta = \Delta(S)$ denote the convex hull of S. Let $v = \mathrm{Vol}(\Delta)$ and $s = \#S$. Suppose $\dim(\Delta) = n$. Finally, let $\delta = \delta(S) = \max_{\nu \in S} |\nu|$ where $|\nu| = \max_i |\nu_i|$. Then the bit size of S is $O(s \log \delta)$.

Proposition 3.5. *Let $n \in \mathbb{Z}_{>0}$. Then there exists an algorithm that, given a finite set $S \subseteq \mathbb{Z}^n$, computes the set $\mathbb{Z}^n \cap \Delta(S)$ in time $\widetilde{O}(s^{\lceil n/2 \rceil} \log \delta + v \log \delta)$. In particular, if Δ is confined then this runs in time $\widetilde{O}(s^{\lceil n/2 \rceil} + v)$.*

Remark 3.6. If Δ is confined, and contained in a orthotope Ξ, then $\Delta \cap \mathbb{Z}^n \subseteq \Xi \cap \mathbb{Z}^n$ and the latter has cardinality $O(v)$, but one still needs to test if each such lattice point is contained in Δ. The exponential contribution of the first term comes from the combinatorics of Δ, which in general can be quite involved.

The proof of this proposition combines several well-known results in computational geometry. We refer to the book of Preparata and Shamos [52] and the articles by Seidel [55] and Fortune [24] and the references contained therein for more detail.

The first main step is to compute the *Delaunay triangulation* of S. We lift each point $\nu \in S \subseteq \mathbb{Z}^n$ to $(\nu, \|\nu\|^2) \in \mathbb{Z}^{n+1}$ where $\|\nu\|^2 = \nu_1^2 + \ldots + \nu_n^2$. Then the convex hull of the lifted vertices has simplicial faces (under mild assumptions that can be achieved under a suitable perturbation), and projecting the simplicial faces onto the original vertices yields a triangulation. There are many (deterministic) algorithms to compute the Delaunay triangulation: we will invoke one method, called the *incremental algorithm*, with optimal deterministic variant due to Chazelle [12].

To give a brief outline of this algorithm, we follow the overview given by Seidel [55, 19.3.1]. The incremental algorithm orders the set $S = \{\nu_1, \ldots, \nu_s\}$ and incrementally computes $\Delta_i = \Delta(S_i)$ from Δ_{i-1}, where $S_i = \{\nu_1, \ldots, \nu_i\}$. The description of Δ_i is by its *facet description*, the set of all facets specified by their defining linear inequalities. A facet of Δ_{i-1} is *visible* from ν_i if its supporting hyperplane separates Δ_{i-1} and ν_i, otherwise we say the facet is *obscured*. Updating Δ_{i-1} to Δ_i involves finding (and deleting) all facets visible to ν_i, preserving all obscured facets, and adding new facets with vertex ν_i. We record these steps keep track of the triangulation created in this way by updating the *facet graph*.

From this description, it is clear that the integer operations involve only computing and checking linear inequalities defining facets arising from the convex hull of subsets of points of the form $(\nu, \|\nu\|^2)$ of cardinality n. By Cramer's rule, the linear equalities defining such a facet have coefficients that are bounded in size by $(n+1)!\delta^n(n\delta^2) = O(\delta^{n+2})$, so the largest integer operation can be performed in time $O((n+2)\log \delta) = O(\log \delta)$ using fast integer multiplication techniques. (The bit arithmetic is also analyzed by Fortune [23, 4.7].)

The incremental algorithm requires $O(s \log s + s^{\lfloor (n+1)/2 \rfloor}) = \widetilde{O}(s^{\lceil n/2 \rceil})$ integer operations (if n is fixed), so in fact the total number of bit operations is $\widetilde{O}(s^{\lceil n/2 \rceil} \log \delta)$.

Remark 3.7. Having computed the Delaunay triangulation, in fact we have computed a complete facet description of the convex hull $\Delta = \Delta(S)$; if desired, one can compute the complete face lattice (in particular, the set of all vertices) in the same running time [55].

Having computed the Delaunay triangulation, we can list lattice points by doing so in each simplex. The problem of enumerating lattice points in an (integral) simplex is analyzed by Bruns and Koch [8] in the context of computing the integral closure of an affine semigroup. Let $\Sigma \subseteq \Delta$ be a simplex of the Delaunay triangulation of Δ defined by $v_0, \ldots, v_n$ and let $w_i = v_i - v_0$. We compute the Hermite normal form of the matrix B with columns w_i (which can be done using $O(\log \delta)$ bit operations in fixed dimension n [15, 31]); let w_i' be its columns. We then enumerate all elements in the simplex, each represented by a lattice points

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

in the fundamental parallelepiped $\mathbb{Z}^n/(\sum_i \mathbb{Z}w_i')$, using $O(\mathrm{Vol}(\Sigma)\log\delta)$ bit operations. Since the simplices Σ triangulate Δ, putting this together with the first main step, we have proven Proposition 3.5.

4. *Precision and running time estimates*

In this section, we discuss precision and running time estimates for each of the four steps of our algorithm. We suppose throughout this section that $\overline{f}$ is confined.

Step 1*: Computing the Jacobian ring*

First some basics. A ring operation in $R=\mathbb{Z}_q/p^N$ can be performed using $O(N\log q)$ bit operations using standard fast multiplication techniques. The Teichmüller lift of an element of $\mathbb{F}_q$ to R can be performed using $O(\log N)$ Hensel lift iterations (Newton's method in this case reduces to the iteration of the map $x\mapsto x^q$); each iteration can be performed in time $O(N\log^2 q)$ by repeated squaring, for a total of $O(N\log N\log^2 q)=\widetilde{O}(N\log^2 q)$ bit operations; the time to compute the lift f we will see is negligible. Similarly, the time to compute the partial derivatives $wf_i=w\partial f/\partial x_i$ is negligible.

We now compute the monomial basis V for the Jacobian ring J, as described in § 2. By Proposition 3.5, we can compute $\mathbb{Z}^n\cap d\Delta$ for $d=0,\ldots,n+2$ in time $\widetilde{O}(s^{\lceil n/2\rceil}+v)$: because we have fixed the dimension, we have $\mathrm{Vol}(d\Delta)=d^n\,\mathrm{Vol}(\Delta)=O(v)$ for such d. (Given the inequalities defining Δ, we could also simply compute $\mathbb{Z}^n\cap(n+2)\Delta$ and then find $\mathbb{Z}^n\cap d\Delta$ by testing the scaled inequalities.)

We now analyze the computation of the row-echelon form $M_d=T_dJ_d$ for $d=0,\ldots,n+2$. The matrix J_d has $O(d^nv)$ columns indexed by $\mathbb{Z}^n\cap d\Delta$ and $O((n+1)(d-1)^nv)=O((d-1)^nv)$ rows indexed by $(wf,wf_1,\ldots,wf_n)_d$. The row echelon form of an $r\times s$ matrix can be performed using $O(r^2s)$ ring operations using standard techniques, so we can compute $M_d=T_dJ_d$ in time $O((d-1)^{2n}d^nv^3)$ ring operations and so for $d=1,\ldots,n+2$ using

$$O(n^{3n}v^3N\log q)=O(v^3N\log q)$$

bit operations. From this, we compute the basis V.

Remark 4.1. In practice, it may be more efficient to use Gröbner bases to compute the basis V and capture the effect of the reduction matrices M_d. (In particular, Faugère's F_4- and F_5-algorithms would be quite useful.) For simplicity, we take the direct approach using linear algebra.

Step 2*: Computing the action of the Frobenius automorphism*

First, we compute the image of the Dwork splitting function

$$\theta(t)=\exp(\pi(t-t^p))=\sum_{i=0}^{\infty}\pi^i\ell_it^i$$

modulo p^N. We have $\mathrm{ord}_p(\pi^i\ell_i)=\mathrm{ord}_p(\lambda_i)\geqslant i(p-1)/p^2$, so modulo p^N the image is a polynomial of degree less than $Np^2/(p-1)=O(pN)$. Recall from (2.13) that

$$\mathrm{ord}_p(\ell_i)\geqslant -d(p,i)=-\left\lfloor\frac{i(2p-1)}{p^2(p-1)}\right\rfloor.$$

We write each ℓ_i to precision N as an element of the module $p^{-d(p,i)}(\mathbb{Z}_q/p^{N+d(p,i)})$. The largest such denominator in our expansion is bounded by

$$d\big(p,Np^2/(p-1)\big)\leqslant\left(\frac{Np^2}{p-1}\right)\frac{2p-1}{p^2(p-1)}=N\frac{2p-1}{(p-1)^2}=O(N/p). \tag{4.2}$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Therefore we compute $\theta(t)$ with coefficients in $\mathbb{Z}_q/p^M$ where $M = N(1 + (2p-1)/(p-1)^2) = O(N)$, then compute each $\ell_i = \lambda_i/\pi^i$.

To compute $\theta(t)$ modulo p^N, we multiply $\exp(\pi t)$ with $\exp(-\pi t^p)$ truncated to degree $Np^2/(p-1)$. The latter factor $\exp(-\pi t^p)$ has no more than $Np/(p-1)$ terms, so multiplying the two can be done in time $O((N^2p^3/(p-1)^2)(M \log q)) = O(pN^3 \log q)$.

Remark 4.3. Note this computation only depends on p and N and does not depend on f.

Given a monomial $w^d x^\mu \in V$, we compute the action of the Frobenius automorphism α using (2.14):

$$\begin{aligned}\alpha((\pi w)^d x^\mu) = \sum_{k\in K} &\sigma^{-1}(a^k)(-p)^{(|k|+d)/p}(\pi w)^{(|k|+d)/p} x^{(k\nu+\mu)/p} \\ &\times \left(\sum_{e\geqslant 0}(-p)^{|e|}\ell_{k+pe}a^e(\pi w)^{|e|}x^{e\nu}\right).\end{aligned}$$

This is computed to precision N, interpreted as above: to multiply ℓ_i times ℓ_j we add the exponent of the denominators in p and multiply the numerators as usual; since these denominators are bounded, the extra arithmetic with denominators is negligible and such a multiplication can be performed using $\widetilde{O}(M \log q) = \widetilde{O}(N \log q)$ bit operations.

Recall Remark 2.17: in the upcoming reduction step (Step 3), we lose no precision as no divisions occur: we only reduce the power of πw. Therefore, to have the Frobenius automorphism expansion correct to precision N, we just analyze the convergence to 0 of the term

$$c_{k,e} = (-p)^{|e|+(|k|+d)/p}\ell_{k+pe}.$$

Using (2.13), we have

$$\begin{aligned}\mathrm{ord}_p(c_{k,e}) &= \mathrm{ord}_p\big((-p)^{|e|+(|k|+d)/p}\ell_{k_1+pe_1}\ldots\ell_{k_s+pe_s}\big) \\ &\geqslant \frac{|k|+d}{p} + |e| - (|k|+p|e|)\frac{2p-1}{p^2(p-1)} \\ &\geqslant (|k|/p+|e|)\left(1-\frac{2p-1}{p^2-p}\right)+\frac{d}{p}. \qquad (4.4)\end{aligned}$$

Here and from now on we insist that $p > 2$: then in (4.4) we have $\mathrm{ord}_p(c_{k,e}) \geqslant 0$. In particular, then, the coefficients are all integral (this was guaranteed by the reduction theory of Adolphson and Sperber [2]); so even though we have (temporarily) written the values ℓ_{k+pe} with denominators, the power of p multiplies through to make them integral. Also note that, for any basis element $(\pi w)^d x^\mu \in V$, we have $d \leqslant n+1 = O(1)$. Let

$$\beta = \beta(p) = \left(1-\frac{2p-1}{p^2-p}\right)^{-1} = \frac{p^2-p}{p^2-3p+1}.$$

For $p = 3$ we have $\beta(p) = 6$, but for $p \geqslant 5$ we have $\beta(p) \leqslant 20/11$, and $\beta(p) \to 1$ as $p \to \infty$. In (2.14), we have multiplied through by π^d for uniformity in the expression, so we must compute to this extra precision: so let

$$\gamma = \gamma(p,n) = \frac{n+1}{p^2-p} \geqslant \frac{d}{p^2-p} = \frac{d}{p-1} - \frac{d}{p}.$$

Note that $\gamma(p,n) \leqslant n$.

Then (4.4) implies $\mathrm{ord}_p(c_{k,e}/\pi^d) \geqslant |e|/\beta - \gamma$, since $|k| \geqslant 0$. Thus to have the answer correct to precision N we only need to worry about terms with $|e| < \beta(N+\gamma) = E$. The set of $e \in \mathbb{Z}^s_{\geqslant 0}$ with $|e| < E$ has cardinality $O(E^s/s!) = O(E^s)$; since $\#K = p^{s-\rho}$, the number of terms in the expansion of $\alpha((\pi w)^d x^\mu)$ is $O(p^{s-\rho}E^s)$.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

We compute the terms in the sum indexed by K. We can compute $a^k = a_1^{k_1} \dots a_s^{k_s}$ using $O(s(\log|k|)(N \log q)) = \widetilde{O}(sN \log^2 q)$ bit operations, since $|k| \leqslant ps$; $\sigma^{-1}(a^k) = (a^k)^{q/p}$ by repeated squaring is computed using $O(N \log^2 q)$ bit operations. The exponent arithmetic in the power of x is negligible up to logarithmic factors. Therefore, we can compute the $\#K = p^{s-\rho}$ terms in the sum indexed by K in time $\widetilde{O}(p^{s-\rho} sN \log^2 q)$.

Now we compute the terms in the inner sum. The value $\ell_{k+pe} = \ell_{k_1+pe_1} \dots \ell_{k_s+pe_s}$ can be computed using $O(s)$ multiplications or $O(sN \log q)$ bit operations. We compute the values a^e recursively, so that each term (ordered by $\prec$, increasing $|e|$) requires only one additional multiplication. Therefore the $O(E^s)$ terms in the inner sum can be computed in time $O(E^s(sN \log q))$.

Putting these together, the total product can be computed using

$$\begin{aligned}&\widetilde{O}(p^{s-\rho} sN \log^2 q + E^s(sN \log q) + p^{s-\rho}E^s(N \log q))\\ &\quad = \widetilde{O}(p^{s-\rho} sE^s N \log^2 q) = \widetilde{O}(p^{s-\rho}\beta^s(N+\gamma)^s N \log^2 q)\end{aligned} \tag{4.5}$$

bit operations for one monomial in V.

Step 3*: Reducing in cohomology*

We now reduce the elements $G = \alpha((\pi w)^d x^\mu) \in L$ computed in Step 2. We begin by reducing the degree of G, as explained in § 2, using multiplication by the matrix $T = T_{n+2}$, until this degree is no greater than $n+1$; then we complete the reduction by multiplications by the matrices T_d with $d = n+1, n, \dots, 1, 0$. The number of such reductions in any fixed degree $d > n+2$ is governed by the number of translates of $(n+2)\Delta$ that cover $d\Delta \cap \mathbb{Z}^n$. By definition, we have

$$d\Delta \cap \mathbb{Z}^n = \bigcup_{\nu \in (d-(n+2))\Delta \cap \mathbb{Z}^n} (\nu + ((n+2)\Delta \cap \mathbb{Z}^n))$$

so the number of reductions is at most $\#((d-(n+2))\Delta \cap \mathbb{Z}^n) = O((d-(n+2))^n v)$ by Corollary 3.3.

Remark 4.6. The fewest number of translates that one could hope for is

$$\mathrm{Vol}(d\Delta)/\mathrm{Vol}((n+2)\Delta) = (d/(n+2))^n.$$

However, it may be that the combinatorics of Δ will not allow this.

Each reduction involves multiplication of a vector by a square matrix of size $O((n+2)^n v) = O(v)$ over R, which can be achieved in time $O(v^2 N \log q)$, so reduction from degree $d > n+2$ to $d-1$ takes time $O((d-(n+2))^n v^3 N \log q) = O(d^n v^3 N \log q)$. Similarly, reduction from degree $d \leqslant n+2$ to $d-1$ takes time $O(d^n v^3 N \log q)$. Repeating this for $d = E, \dots, 1, 0$ gives a total time of $O(E^{n+1} v^3 N \log q) = O(\beta^{n+1}(N+\gamma)^{n+1} v^3 N \log q)$ to complete the reduction.

Step 4*: Output*

Having assembled the square matrix A of size v, we compute the product

$$A_a = AA^{\sigma^{-1}} \dots A^{\sigma^{-(a-1)}}$$

(where $a = \log_p q$). It takes time $O(v^2 N \log^2 q)$ to compute σ^{-1} applied to a matrix of size v, and time $O(v^3 N \log q)$ to multiply two such matrices, for a total time of $O(\log q(v^2 N \log^2 q + v^3 N \log q)) = O(v^3 N \log^3 q)$ to compute A_a. The characteristic polynomial of a matrix of size v can be computed using $O(v^3 N \log q)$ ring operations, which is absorbed into the previous estimate.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Total running time

We now add up the contributions from each step, proving Theorem A.

Step 1 takes time $\widetilde{O}(s^{\lceil n/2\rceil} + v^3 N \log^2 q)$. Step 2 takes time $\widetilde{O}(pN^3 \log q + p^{s-\rho}\beta^s(N + \gamma)^s vN \log^2 q)$. Step 3 takes time $\widetilde{O}(\beta^{n+1}(N+\gamma)^{n+1}v^4 N \log^2 q)$. Step 4 takes time $O(v^3 N \log q)$. Since $s \geqslant n+1$ as $\dim(\Delta) = n$, the time in Step 2 dominates, up to a polynomial in $vN \log q$. This totals to

$$\widetilde{O}\big(s^{\lceil n/2\rceil} + pN^3 \log q + p^{s-\rho}\beta^s(N+\gamma)^s(v^4 N \log^2 q)\big)$$

bit operations. Using the estimate $\beta \leqslant 6$ and $\gamma \leqslant (n+1)/20$, this becomes

$$\widetilde{O}\big(s^{\lceil n/2\rceil} + pN^3 \log q + p^{s-\rho}(6N+n)^s(v^4 N \log^2 q)\big) \tag{4.7}$$

and if $p \geqslant 5$ then $\beta \leqslant 20/11 \leqslant 2$ so this improves to

$$\widetilde{O}\big(s^{\lceil n/2\rceil} + pN^3 \log q + p^{s-\rho}(2N+n)^s(v^4 N \log^2 q)\big).$$

Precision

We have thus computed the characteristic polynomial to precision N. To prove Theorem B, we estimate the value of N required to recover the zeta function itself. First, we factor this characteristic polynomial so as to work with $p^{-1}\alpha$ instead of α. The matrix A has block form $\left(\begin{smallmatrix}1 & 0\\ * & *\end{smallmatrix}\right)$ where we order the monomials in the basis V for the space B by degree: the only term in degree 0 is the monomial 1. The matrix A_a also has this form, since it holds for each term in the product. Therefore the characteristic polynomial of α on B factors as $(1-T)$ times the characteristic polynomial on quotient space $B_0 = B/R$. The action of the Frobenius automorphism α on B_0 is divisible by p; let A_0 be the matrix of α on B_0. Then

$$Z(V,T) = \det(1-(p^{-1})A_0 T)^{(-1)^n}\left(\frac{Z(\mathbb{G}_m^n, T)}{(1-T)}\bigg|_{T=T/q}\right).$$

Therefore, we may compute with $p^{-1}\alpha$ instead of α.

The characteristic polynomial $\det(1-q^{-1}(A_0)_a T)$ then has inverse roots of absolute value at most $q^{n/2}$ by a theorem of Adolphson and Sperber [2] and Denef and Loeser [14]. Therefore its ith coefficient is bounded by $\binom{v}{i}q^{i(n/2)}$.

Lemma 4.8. *For all $x \in \mathbb{R}_{\geqslant 1}$ and $v \in \mathbb{Z}_{\geqslant 0}$, we have*

$$\max_{0\leqslant i\leqslant v}\binom{v}{i}x^i = \binom{v}{\lceil v/2\rceil + j}x^{\lceil v/2\rceil + j}$$

where $0 \leqslant j \leqslant \lfloor v/2\rfloor$ is the unique index such that

$$\frac{\lceil v/2\rceil + j}{\lfloor v/2\rfloor - (j-1)} \leqslant x < \frac{\lceil v/2\rceil + (j+1)}{\lfloor v/2\rfloor - j}.$$

The proof is an easy inductive argument. In the two extremes: if $x \geqslant v$, then $j = \lfloor v/2\rfloor$ and the largest coefficient is x^v; if $x < (\lceil v/2\rceil + 1)/\lfloor v/2\rfloor$ (equal to $1 + 2/v$ if v is even, for example), then the largest coefficient is $\binom{v}{\lceil v/2\rceil}x^{\lceil v/2\rceil}$. It follows that the p-adic precision N required to recover all of these coefficients as integers from their reduction modulo p^N is given by

$$p^N \geqslant 2\binom{v}{\lceil v/2\rceil + j}(q^{n/2})^{\lceil v/2\rceil + j} \tag{4.9}$$

where j is given as in Lemma 4.8 with $x = q^{n/2}$.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

In practice, one will want to work with the precision estimate (4.9). To estimate the runtime, we have the crude bound

$$\binom{v}{i}x^i \leqslant \binom{v}{\lceil v/2 \rceil}(q^{n/2})^v < (2q^{n/2})^v$$

which implies we may take

$$N \geqslant (v+1)\log_p 2 + \frac{nv}{2}\log_p q = O(nv\log q). \tag{4.10}$$

Remark 4.11. We are forced to take a larger bound than just the middle coefficient because we do not have a Riemann hypothesis in this generality. For many varieties under consideration, such a hypothesis will give a better estimate on the precision, since the higher coefficients are determined by the lower ones.

Also, the work of Kedlaya [**35**] shows how to recover the zeta function often in practice with much less precision knowing only that it can be factored as a product of Weil q-polynomials.

Plugging the estimate (4.10) into (4.7), and considering s (and n) to be fixed, we obtain the estimate

$$\widetilde{O}(p(v\log q)^3\log q + p^{s-\rho}(v\log q)^s v^4(v\log q)\log^2 q) = \widetilde{O}(p^{\min(1,s-\rho)}v^{s+5}\log^{s+3} q)$$

for the number of bit operations performed. This completes the proof of Theorem B.

5. *Modifications*

In this section, we discuss some extensions and modifications to the above algorithm.

Dense input

We can also modify the algorithm for the situation of dense input. One can also forget the condition of sparsity and analyze the running time on dense input. Here, we do not use the expansion (2.14), but rather directly compute the product (2.4). The analysis above shows that the computation of the expansion

$$\theta(a_\nu w x^\nu) = \sum_{i=0}^{Np^2/(p-1)} \ell_i(\pi w)^i a_\nu^i x^{i\nu}$$

can be performed in $\widetilde{O}(spN^2\log q)$ operations. We have s such terms in the product (2.4). This product has monomial support in $(Np^2/(p-1))\Delta$ so has $O((pN)^n v)$ terms.

As in (4.2), we compute in $\mathbb{Z}_q/p^M$, and multiplying two polynomials in n variables with coefficients in $R = \mathbb{Z}_q/p^M$ with at most $O((pN)^n v)$ terms takes time $O(M(pN)^{2n}v^2\log q)$ if multiplied term-by-term. (See also Lauder and Wan [**48**, Lemma 30].) (The grading by w allows us to multiply more carefully, but this saves only a constant factor for fixed dimension.) Therefore, the product $F(w,x)$ can be computed in time

$$\widetilde{O}(spN^2\log q + sM(pN)^{2n}v^2\log q) = \widetilde{O}(sp^{2n}N^{2n+1}v^2\log q).$$

Therefore, one sees a benefit from the fewnomial expansion only when $s \leqslant 2n+1$.

The time to compute $\alpha(w^d x^\mu)$ requires negligible time in comparison, as it involves only exponent arithmetic and applying the inverse Frobenius automorphism σ^{-1}. The other steps are unmodified, so the total time is

$$\widetilde{O}\big(s^{\lceil n/2\rceil} + sp^{2n}N^{2n+1}v^2\log q + (6N+n)^{n+1}v^3 N\log q\big) = \widetilde{O}\big(s^{\lceil n/2\rceil} + sp^{2n}N^{2n+1}v^3\log q\big)$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

to compute the zeta function modulo p^N (the analogue for Theorem A), and plugging in the estimate (4.10) for N and considering s fixed we obtain

$$\widetilde{O}(p^{2n} v^{2n+4} \log^{2n+2} q)$$

(for Theorem B).

Modifications when $p = 2$

For many applications (notably in coding theory), the computation of zeta functions and L-functions in characteristic 2 are particularly important. While it may be possible to perform this analysis, we do not do so in the present work. Instead, we mention some of the hurdles this analysis faces in characteristic 2 using the approach we have taken applying Dwork cohomology. (Some of these same hurdles, and others, also arise in other p-adic approaches.)

Dwork's original p-adic study of zeta functions concerned a nonsingular projective hypersurface $V \subset \mathbb{P}^{n-1}$ over $\mathbb{F}_q$ defined by the vanishing of a homogeneous form $f(x) \in \mathbb{F}_q[x_1, \ldots, x_n]$ of degree d. When $\gcd(2, p, d) = 1$, Dwork constructs a p-adic cohomology space with an action of the Frobenius automorphism such that the characteristic polynomial of the Frobenius automorphism acting on cohomology gives the important nontrivial middle dimensional primitive factor of $Z(V, T)$. In particular, if $p = 2$, Dwork's cohomology does not at present apply to smooth hypersurfaces of even degree in characteristic 2.

In some cases, Adolphson and Sperber [4] are able to supplement Dwork's work when $p = 2 \mid d$. For example, if $p \mid n$ and we consider the Dwork family of hypersurfaces

$$f(x_1, \ldots, x_n) = x_1^n + x_2^n + \ldots + x_n^n - \lambda x_1, \ldots, x_n$$

in characteristic $p = 2$, then even though the family consists of singular hypersurfaces, they are nondegenerate with respect to the sublattice of $\mathbb{Z}^{n+1}$ generated by the support of $wf(x)$. As a consequence (even when $p = 2$), there is a cohomology space such that the characteristic polynomial of the Frobenius automorphism acting on this space is the nontrivial factor of the zeta function. However, computing in the sublattice adds to the computational complexity.

A second obstacle concerns convergence and our choice of splitting function

$$\theta_1(t) = \exp(\pi t - (\pi t)^p / p)$$

where $\pi^{p-1} = p$, which converges for $\operatorname{ord}_p t > -(p-1)/p^2$. This led us to the space $L_\Delta = L_\Delta((p-1)/p^2)$ consisting of power series with similar growth and the operators $D_i = x_i(\partial/\partial x_i) + \pi x_i w(\partial f)/(\partial x_i)$ acting on L_Δ. Our explicit reduction theory in the Jacobian ring depended on the operator norm of $x_i(\partial/\partial x_i)$ being p-adically smaller than the operator norm of $\pi x_i w(\partial f)/(\partial x_i)$, so that the series (2.14) converges after reduction. This requires that $1/(p-1) \leqslant (p-1)/p$, that is, $p < (p-1)^2$, and this fails if and only if $p = 2$.

Therefore, when $p = 2$, it is necessary to use a splitting function with better convergence properties. Let $\pi \in \overline{\mathbb{Q}}_2$ be a nonzero root of the equation

$$\pi^8/8 + \pi^4/4 + \pi^2/2 + \pi = 0, \quad \text{that is,} \quad \pi^7 + 2\pi^3 + 4\pi + 8 = 0$$

and let

$$\theta_3(t) = \exp\left(\pi t + \frac{(\pi t)^2}{2} + \frac{(\pi t)^4}{4} + \frac{(\pi t)^8}{8}\right) = \sum_{i=0}^{\infty} \lambda_i t^i \in \mathbb{Z}_2[\pi][[t]].$$

In this case we have $\operatorname{ord}_2 \lambda_i > (11/16)i$. Working with the space $L_\Delta(11/16)$, reduction is possible (in $L_\Delta(11/8)$) since $11/8 > 1/(p-1) = 1$ when $p = 2$.

The computational difficulties caused by using $\theta_3(t)$ instead of $\theta_1(t)$ are numerous. Not only is the reduction algorithm more difficult, but even the calculation of the Frobenius automorphism is more complicated [2, Proposition 3.2].

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Remark 5.1. Robba constructed a constant called π_{Robba} by Dwork with $\operatorname{ord}_p \pi_{\mathrm{Robba}} = 1/(p-1)$ with splitting function $\theta_{\mathrm{Robba}}(t) = \exp(\pi_{\mathrm{Robba}}(t - t^p))$ which has a better radius of convergence than Dwork's $\theta_1(t)$. However, even with this improvement, the p-adic norm in case $p = 2$ of $\pi_{\mathrm{Robba}} x_i w(\partial f)/(\partial x_i)$ does not dominate the norm of $x_i(\partial/\partial x_i)$ on the appropriate Banach space.

For these reasons, we do not deal at present with the case $p = 2$.

Affine varieties

Next, we describe modifications to the algorithm to compute the zeta function of affine varieties. (At the price of some additional notation, one could consider the more general case where the variety is a combination of affine and toric.)

Let $\overline{f}(x) = \sum_\nu \overline{a}_\nu x^\nu \in \mathbb{F}_q[x_1, \dots, x_n]$ be a polynomial. For a subset $A \subseteq [n] = \{1, \dots, n\}$, we denote by $\overline{f}_A(x) \in \mathbb{F}_q[x_i : i \in [n] \setminus A]$ the polynomial obtained from $\overline{f}$ obtained by setting $x_i = 0$ for all $i \in A$. We say that $\overline{f}$ is *convenient* (*with respect to the variables* $x_1, \dots, x_n$) provided

$$\dim \Delta(\overline{f}_A) = \dim \Delta(\overline{f}) - \#A$$

for all subsets $A \subseteq [n]$. Equivalently, $\overline{f}$ is convenient if and only if $\overline{f}$ has a nonzero constant term and a monomial $x_i^{d_i}$ with $d_i \in \mathbb{Z}_{>0}$ for all $i = 1, \dots, n$. The notion of convenient is also called *commode*.

We suppose for the rest of this subsection that $\overline{f}$ is convenient and nondegenerate (with respect to $\Delta(\overline{f})$, defined as before). It is a consequence of the hypothesis of convenience that $\dim \Delta(\overline{f}) = n$.

Example 5.2. Suppose $\overline{f}$ has total degree d and $\Delta(\overline{f})$ is the convex hull of the set of points $\{0\} \cup \{de_i\}^n$ where $e_1, \dots, e_n$ is the standard basis in $\mathbb{R}^n$. Then we may write $\overline{f} = \overline{f}^{(d)} + \overline{g}$ where $\overline{f}^{(d)}$ is the form of f of highest degree terms and the total degree of $\overline{g}$ is less than d. Then $\overline{f}$ is nondegenerate if and only if $\overline{f}_A^{(d)} = 0$ defines a nonsingular projective hypersurface in $\mathbb{P}^{n-\#A-1}$ and $\overline{f}_A = 0$ is nonsingular in $\mathbb{A}^{n-\#A}$ for all subsets $A \subseteq [n]$.

Assuming then that $\overline{f}$ is nondegenerate and convenient, we can work more simply with the affine L-function $L(wf, \mathbb{G}_m \times \mathbb{A}^n, T)$, as follows. Let $\overline{V}$ denote the affine hypersurface in $\mathbb{A}^n_{\mathbb{F}_q}$ defined by $\overline{f} = 0$. We modify the calculation in § 1 by working with an affine exponential sum:

$$q^{rn} + \sum_{\substack{w \in \mathbb{F}_{q^r}^\times \\ x \in \mathbb{F}_{q^r}^n}} \Theta_r(wf) = \sum_{w, x \in \mathbb{F}_{q^r} \times \mathbb{F}_{q^r}^n} \Theta_r(wf) = q\#\overline{V}(\mathbb{F}_{q^r})$$

so

$$Z(\overline{V}, qT) = \frac{L(w\overline{f}, \mathbb{G}_m \times \mathbb{A}^n, T)}{1 - q^n T}.$$

The computation of $L(w\overline{f}, \mathbb{G}_m \times \mathbb{A}^n, T)$ is for the most part quite similar to the toric calculation given earlier. We describe here the required modifications. Let $A \subseteq S = \{1, \dots, n\}$. Let $L_\Delta^{(A)}$ denote the ideal in the ring L_Δ consisting of series having support in monomials divisible by $x_A = \prod_{i \in A} x_i$. Under our hypotheses, the complex $\Omega^\bullet$ for the L-function has vanishing cohomology $H^i(\Omega^\bullet)$ for all i except $i = n + 1$ and

$$H^{n+1}(\Omega^\bullet) = \frac{L_\Delta^S}{D_0 L_\Delta^S + \sum_{i=1}^n D_i L_\Delta^{S \setminus \{i\}}} \tag{5.3}$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

where the D_i are as in § 1. Note that here $H^{n+1}(\Omega^\bullet)$ is contained in the cohomology space $L_\Delta / \left(\sum_{i=0}^n D_i L_\Delta\right)$.

The required modifications are then simple. For example, the Jacobian ring has the form

$$J^S = \frac{R[w\Delta]^S}{(wf)R^S + \sum_i wf_i R^{S\backslash i}}.$$

The reduction algorithm then only requires identification of monomials in the ideals L_Δ^S and $L_\Delta^{S\backslash\{i\}}$, and the recursive reduction process preserves the required divisibility on monomials so the algorithm runs in every other way without modification.

In this way, one can reduce the sizes of computations involved under the convenient hypothesis: one avoids calculation of the contributions to the zeta function coming from the intersection of the affine hypersurface with the coordinate hyperplanes.

Projective varieties

We now consider the modifications for projective varieties.

Suppose that $\overline{f} \in \mathbb{F}_q[x_1, \ldots, x_n]$ is a homogeneous form of degree d with $\gcd(p, d) = 1$ and that $\overline{f} = 0$ defines a nonsingular projective hypersurface $\overline{Z} \subseteq \mathbb{P}^{n-1}_{\mathbb{F}_q}$. Suppose further that $\overline{f}_A = 0$ defines a nonsingular projective hypersurface in $\mathbb{P}^{n-\#A-1}$ for all subsets $A \subseteq [n]$; such a hypersurface is said to be in *general position*. If $\gcd(p, d) = 1$, then $\overline{f}$ defines a projective hypersurface in general position if and only if $\overline{f}$ is nondegenerate (with respect to $\Delta(\overline{f})$) and convenient (with respect to the variables $x_1, \ldots, x_n$).

Here, the support of $w\overline{f}$ lies in the hyperplane $\sum_{i=1}^n x_i = dw$ in $\mathbb{R}^{n+1}$, so $\dim \Delta(w\overline{f}) = n$ (or, equivalently, $\dim \Delta(\overline{f}) = n - 1$).

It is well known that

$$Z(\overline{Z}, T) = \frac{P(T)^{(-1)^{n-1}}}{(1 - T)(1 - qT) \ldots (1 - q^{n-2}T)}$$

where $P(T)$ is a polynomial of degree $d^{-1}((d - 1)^n + (-1)^n(d - 1))$ that represents the action of the Frobenius automorphism on middle-dimensional primitive cohomology. By work of Adolphson and Sperber [2], the cohomology of the complex $\Omega^\bullet$ for $L(w\overline{f}, \mathbb{G}_m \times \mathbb{A}^n, T)$ described above in this case has vanishing cohomology $H^i(\Omega^\bullet) = 0$ for $i \neq n, n + 1$ and that there is an isomorphism of Frobenius modules $H^{n+1}(\Omega^\bullet) \cong H^n(\Omega^\bullet)$ with the Frobenius automorphism on H^n being q times the Frobenius automorphism on H^{n+1}. As a consequence,

$$\det(1 - \operatorname{Frob} T \mid H^{n+1}(\Omega^\bullet)) = P(qT)$$

yielding the 'interesting' part of the zeta function of $\overline{Z}$.

As in the affine case, the space H^{n+1} is isomorphic to the quotient defined in (5.3). However, we can simplify further. By the Euler relation, we have $dD_0 = \sum_{i=1}^n D_i$ on L_Δ so that

$$D_0 L_\Delta^S \subseteq \sum_{i=1}^n D_i L_\Delta^{S\backslash\{i\}}.$$

This enables us to reduce the calculation by suppressing the role played by w entirely. To determine the monomials x^ν in L_Δ, we simply need to check

$$x^\nu \in M_\Delta = \left\{ \nu \in \mathbb{Z}^n_{\geqslant 0} : |\nu| = \sum_{i=1}^n \nu_i \equiv 0 \pmod{d} \right\}$$

and use $w^{|\nu|/d} x^\nu$. So the power of w enters only formally.

Let

$$\widetilde{D}_i = x_i \frac{\partial}{\partial x_i} + \pi x_i \frac{\partial}{\partial x_i} f$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

and write

$$\widetilde{L}_\Delta = \left\{ \sum_{\nu \in M_\Delta} a_\nu x^\nu : a_\nu \in \mathbb{Z}_q[\pi], \operatorname{ord}_p a_\nu \geqslant \frac{|\nu|}{d}\left(\frac{p-1}{p^2}\right) \right\} \subset \mathbb{Z}_q[\pi][[x]].$$

Our object of interest for the reduction is then

$$B'' = \frac{\widetilde{L}_\Delta^S}{\sum_{i=1}^n \widetilde{D}_i \widetilde{L}_\Delta^{S\setminus\{i\}}}.$$

The preceding algorithms for reduction may be applied here as well; the appropriate powers of w may be formally inserted as necessary.

Finally, as we remarked in the comments for the case $p = 2$, when $p \mid d$ there are further modifications that can be made by considering polynomials that are nondegenerate relative to a sublattice even in some exceptional singular cases [4].

Exponential sums

In § 1, we reduce the problem of computing the zeta function to the problem of computing the L-function of an exponential sum. However, in many situations the problem of computing this L-function itself is of interest. In this case, there is no auxiliary or dummy variable w; the support of the p-adic power series in our space L_Δ consists of those lattice points in the cone over $\Delta_\infty(\overline{f})$ which itself is the convex closure of the support of $\overline{f}$ together with the origin. We note the earlier work of Lauder and Wan [46, 47] who apply some similar approach to compute the L-function of a one-dimensional exponential sum (that is, $n = 1$).

We consider the case of toric exponential sums. (Considering affine exponential sums, when convenient, requires modifications similar to those above.) Let $\overline{f}(x) = \sum_\nu \overline{a}_\nu x^\nu \in \mathbb{F}_q[x^\pm]$ be a Laurent polynomial. We say f is *quasihomogeneous* if there are rational numbers $\alpha_1, \ldots, \alpha_n$ such that $w(\nu) = 1$ for all $\nu \in \operatorname{supp}(\overline{f})$ where

$$w(\nu) = \sum_{i=1}^n \alpha_i \nu_i.$$

We restrict to the case of quasihomogeneous exponential sums as this conforms quite closely with our computation of zeta functions; the method could be adapted to the general case.

There are fewnomial examples of quasihomogeneous exponential sums that indeed appear nontrivial. For example, consider a subset $\Lambda = \{\nu^{(1)}, \ldots, \nu^{(n+1)}\} \subseteq \mathbb{Z}^n$ of cardinality $n+1$ with each element $\nu^{(i)} \in \Lambda$ lying on the hyperplane H defined by $\sum_i \alpha_i x_i = 1$ for all $i = 1, \ldots, n+1$. When the convex hull Δ of Λ is not an n-simplex, then the L-function associated to

$$\overline{f}(x) = \sum_{i=1}^{n+1} a_i x^{\nu^{(i)}}$$

on $\mathbb{G}_m^n$ is not well understood. Even when Δ is an n-simplex, such sums are not entirely understood. We know from Adolphson and Sperber [2] that if $\overline{f}$ is nondegenerate with respect to $\Delta_\infty(\overline{f})$, then the complex $\Omega^\bullet(\overline{f}, \mathbb{G}_m^n)$ for this L-function is acyclic except in dimension n and

$$H^n(\Omega^\bullet(\overline{f}, \mathbb{G}_m^n)) = \frac{L_\Delta}{\sum_{i=1}^n D_i L_\Delta}.$$

Now $L(\overline{f}, \mathbb{G}_m^n, T)^{(-1)^{n+1}} \in \mathbb{Z}[\zeta_p][T]$. This is one complication. The second complication is that the reduction has a modification: now we bring things down by weight. This is analogous to the degree of w, and the argument is formally the same. The ring is still graded by the weight.

If all vertices lie in a hyperplane (*quasihomogeneous*), then the Frobenius matrix has the property that the elements belong to $\pi^{\mathbb{N}}\mathbb{Z}_q$: the π and the weight move as one unit. (So we can

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

do multiplications in the smaller ring.) Then the filtered ring is a graded ring. The only time when one has to do honest calculations in $\mathbb{Z}[\zeta_p]$ is in the final calculation of the characteristic polynomial.

Twisted exponential sums on the torus

Let $\overline{f}(x) \in \mathbb{F}_q[x^{\pm}]$ be nondegenerate. Let $\chi_1, \ldots, \chi_n$ be multiplicative characters of $\mathbb{F}_q$ (possibly including the trivial character). Since the character group of $\mathbb{F}_q^{\times}$ is generated by the Teichmüller character ω, each χ_i may be identified with an integral power a_i of ω with $0 \leqslant a_i < q-1$. It is useful to write $\chi_i = \omega^{a_i} = \omega^{\gamma_i(q-1)}$ where $\gamma_i = a_i/(q-1) \in [0,1)$. The shifted lattice $\Lambda(\gamma) = (\gamma_1, \ldots, \gamma_n) + \mathbb{Z}^n$ plays an important role in the cohomological study of the twisted sums

$$S(\gamma, \overline{f}, \mathbb{G}_m^n) = \sum_{x \in \mathbb{G}_m^n(\mathbb{F}_q)} \omega^{a_1}(x_1) \ldots \omega^{a_n}(x_n) \Theta(\overline{f}(x_1, \ldots, x_n))$$

and the associated L-function $L(\gamma, \overline{f}, T)$.

We modify now our earlier work on quasihomogeneous nondegenerate toric sums to include the case of twisted sums. For $\nu \in \mathbb{Z}^n$ we define $w(\nu)$ to be the smallest $m \in \mathbb{Q}_{\geqslant 0}$ such that $\nu \in m\Delta$. We define

$$L_\Delta = \left\{ \sum_{\nu \in M_\gamma(\overline{f})} c_\nu x^\nu : c_\nu \in \mathbb{Z}_p[\zeta_{(q-1)p}] \text{ and } \operatorname{ord}_p(c_\nu) \geqslant \frac{p-1}{pq} w(\mu) \right\}$$

where $M_\gamma(\overline{f})$ is the intersection of the cone over $\overline{f}$ intersected with $\Lambda(\gamma)$.

As before, let f denote the Teichmüller lift of $\overline{f}$. The Frobenius map $\alpha = \psi_q \circ \exp(\pi f(x) - f(x^q))$ acts on L_Δ and the Dwork trace formula for $L(\gamma, \overline{f}, T)$ takes the form

$$L(\gamma, \overline{f}, T)^{(-1)^{n+1}} = \det(1 - T\alpha \mid L_\Delta^{(\gamma)})^{\delta^n}$$

where $g(T)^\delta = g(t)/g(qT)$ for $g \in \mathbb{C}_p[[T]]$. As usual, we define differential operators

$$D_i = x_i \frac{\partial}{\partial x_i} + \pi x_i \frac{\partial f}{\partial x_i}$$

on L_Δ, and construct a complex $\Omega^\bullet$ using L_Δ as the base and boundary operator as before. The Frobenius automorphism defines a chain map on this complex using α, and the hypothesis that f is nondegenerate implies that $H^i(\Omega^\bullet) = 0$ for $i \neq n$ and

$$L(\gamma, \overline{f}, T)^{(-1)^{n+1}} = \det(1 - \alpha T \mid H^n(\Omega^\bullet))$$

where

$$H^n(\Omega^\bullet) = \frac{L_\Delta}{\sum_{i=1}^n D_i L_\Delta}$$

is a free $\mathbb{Z}_p[\zeta_{(q-1)p}]$-algebra of finite rank $\operatorname{Vol}(\Delta)$. Our calculation of the matrix of the Frobenius automorphism acting on $H^n(\Omega^\bullet)$ follows the same argument used above in the case of ('untwisted') quasihomogeneous nondegenerate toric sums.

Multiplicative character sums on the torus

Continuing with this line of analysis, let $\chi = \omega^{a_0} = \omega^{\gamma_0(q-1)}$ be a nontrivial multiplicative character of $\mathbb{F}_q$. We extend χ to all of $\mathbb{F}_q$ by setting $\chi(0) = 0$. As in the previous section, let $\overline{f} \in \mathbb{F}_q[x]$ be nondegenerate, and consider the character sum

$$S(\gamma_0, \overline{f}, \mathbb{G}_m^n) = \sum_{x \in \mathbb{G}_m^n(\mathbb{F}_q)} \omega^{\gamma_0(q-1)}(\overline{f}(x))$$

and its associated L-function $L(\gamma_0, \overline{f}, T)$.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

We now use an elementary character argument to transform such a multiplicative character sum to a twisted exponential sum of the type considered in the previous section. Suppose that χ is nontrivial (that is, $\gamma_0 \neq 0$). Then for $u \in \mathbb{F}_q^\times$ we have

$$\sum_{t \in \mathbb{F}_q^\times} \chi^{-1}(t)\Theta(tu) = -(G(\chi^{-1}, \Theta))\chi(u)$$

where

$$G(\chi^{-1}, \Theta) = -\sum_{v \in \mathbb{F}_q^\times} \chi^{-1}(v)\Theta(v)$$

is the negative of a Gauss sum. Since χ is nontrivial, this identity holds for $u = 0$ as well. Therefore our sum of interest

$$S(\gamma_0, \overline{f}, \mathbb{G}_m^n) = -G(1-\gamma_0, \Theta)^{-1} S(\gamma, w\overline{f}, \mathbb{G}_m^{n+1})$$

where $\gamma = (1-\gamma_0, 0, \ldots, 0) \in \mathbb{Q}^{n+1}$ and the exponential sum on the right is a twisted sum of the type in the preceding section.

By the Hasse–Davenport relation on Gauss sums, we have

$$L(\gamma_0, \overline{f}, \mathbb{G}_m^n, G(\chi^{-1}, \Theta)T)^{-1} = L(\gamma, w\overline{f}, \mathbb{G}_m^{n+1}, T).$$

Note $w\overline{f}$ is always quasihomogeneous, as the monomials all lie in the hyperplane in $\mathbb{R}^{n+1}$ with equation $w = 1$. If f is nondegenerate, then we may proceed to compute the L-function as in the previous section on twisted sums.

6. *Examples*

Elliptic curve point counting

In this subsection, we give an example to show how our methods can be used to compute the zeta function of an elliptic curve. In this situation, our method is not competitive with more specialized methods, but running through the algorithm in this case will hopefully shed some insight on how it works.

Let $p \geqslant 3$ be prime. Let $\overline{f} = x^3 + \overline{a}x + \overline{b} - y^2 \in \mathbb{F}_q[x, y]$ be such that $4\overline{a}^3 + 27\overline{b}^2 \neq 0$, so that $\overline{f} = 0$ defines an affine piece of an elliptic curve $\overline{E}$ over $\mathbb{F}_q$. (This does not cover all cases when $p = 3$; we leave the other examples to work out by the interested reader.) Let $f = x^3 + ax + b - y^2$ be the Teichmüller lift of f to $\mathbb{Z}_q[x, y]$.

If $b = 0$, then $\overline{E}$ has complex multiplication by $\mathbb{Z}[i]$ and is a twist of the elliptic curve $y^2 = x^3 - x$, so the zeta function can easily be recovered by classical methods; the same is true if $a = 0$. So we assume that $ab \neq 0$. Therefore, the polytope Δ is the triangle given by the convex hull $\Delta(\{(0, 0), (3, 0), (0, 2)\})$. One can check that f is automatically nondegenerate given the nonvanishing of the discriminant $4\overline{a}^3 + 27\overline{b^2} \neq 0$. (See also work of Castryck and the second author [11].) Furthermore f is convenient (with respect to x, y).

In this situation, we have

$$L(w\overline{f}, \mathbb{G}_m \times \mathbb{A}^2, T) = P(qT)$$

where

$$Z(\overline{E}, T) = \frac{P(T)}{(1-T)(1-qT)} = \frac{1 - a_qT + qT^2}{(1-T)(1-qT)}$$

and $a_q = q + 1 - \#\overline{E}(\mathbb{F}_q)$ has $|a_q| \leqslant 2\sqrt{q}$.

Remark 6.1. Here, we can see the advantage of working with the affine curve rather than the toric curve. First and foremost, the computations are performed in a cohomology space

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

of dimension $\deg L(f,T)=2$, as opposed to one of dimension $\deg L^*(f,T)=\mathrm{Vol}(\Delta)=6$. This difference is accounted for by the number of points on $\overline{E}$ along the coordinate axes, as follows. From the relation

$$Z(\overline{E},qT)=L^*(wf,T)Z(\mathbb{G}_m^2,qT)=L^*(wf,T)\frac{(1-qT)^2}{(1-T)(1-q^2T)},$$

after some cancellation we find that

$$L^*(wf,T)=(1-T)P(qT)P_x(qT)P_y(qT)$$

where $Z(\overline{E}\cap\{x=0\},T)=P_x(T)/(1-T)$ is a polynomial of degree 1 which is $1-T$ or $1+T$ depending on if $\overline{b}\in\mathbb{F}_q$ is a square or not; similarly, $P_y(T)$ is a polynomial of degree 2 which depends on the factorization of $x^3+\overline{a}x+\overline{b}$ in $\mathbb{F}_q$.

We have the graded ring

$$\mathbb{Z}_q[w\Delta]=\bigoplus_d \mathbb{Z}_q\langle w^d x^i y^j : (i,j)\in d\Delta\rangle \subseteq \mathbb{Z}_q[[w,x,y]]$$

and work in the Jacobian ring

$$J=\frac{\mathbb{Z}_q[w\Delta]}{\langle wf, wf_x, wf_y\rangle}$$

where

$$\begin{aligned} f_x &= x\frac{\partial f}{\partial x}=3x^3+ax,\\ f_y &= -2y^2. \end{aligned}$$

The affine Koszul complex associated to f (see § 5) is

$$0\to L_\Delta \xrightarrow{D} L_\Delta\oplus L_\Delta^{(x)}\oplus L_\Delta^{(y)} \xrightarrow{D} L_\Delta^{(x)}\oplus L_\Delta^{(y)}\oplus L_\Delta^{(xy)} \xrightarrow{D} L_\Delta^{(xy)}\to 0$$

where $L_\Delta^{(m)}\subseteq L_\Delta$ is the subspace divisible by the monomial m, and $D=(D_w,D_x,D_y)$ where

$$D_w=w\frac{\partial}{\partial w}+\pi wf,\quad D_x=x\frac{\partial}{\partial x}+\pi wf_x,\quad D_y=y\frac{\partial}{\partial y}+\pi wf_y.$$

We work in the cohomology space

$$B=\frac{L_\Delta^{(xy)}}{D_wL_\Delta^{(xy)}+D_xL_\Delta^{(y)}+D_yL_\Delta^{(x)}}.$$

We then consider the free $\mathbb{Z}_q$-module

$$J=\frac{\mathbb{Z}_q[w\Delta]^{(xy)}}{\mathbb{Z}_q[w\Delta]^{(xy)}wf+\mathbb{Z}_q[w\Delta]^{(y)}wf_x+\mathbb{Z}_q[w\Delta]^{(x)}wf_y}.$$

In weight 1 we have $J_1=\mathbb{Z}_q wxy$, since this is the only monomial divisible by xy and it is nonzero in the quotient.

We have

$$\begin{aligned} J_2 &= w^2\cdot\frac{\mathrm{span}(\{xy,x^2y,x^3y,x^4y,xy^2,x^2y^2,x^3y^2,xy^3\})}{\mathrm{span}(\{xyf,yf_x,xyf_x,y^2f_x,xf_y,x^2f_y,x^3f_y,xyf_y\})}\\ &= w^2xy\cdot\frac{\mathrm{span}(\{1,x,x^2,x^3\})}{\mathrm{span}(\{x^3+ax+b,3x^2+a,3x^3+ax\})}=\mathbb{Z}_q\cdot w^2xy. \end{aligned}$$

So the monomial basis V we take is wxy, w^2xy.

Since $|a_q|\leqslant 2\sqrt{q}$, we recover a_q uniquely as the integer a such that $a_q\equiv a \pmod{p^N}$ and $|a|\leqslant 2\sqrt{q}$ with $p^N>4\sqrt{q}$, so $N>\log_p 4+(1/2)\log_p q$ or $N=O(\log q)$.

For $v = wxy, w^2xy$, we expand as in (2.11). We first take $v = wxy$. Then we have

$$K = \{e : Me = -(1,1,1)^t \pmod{p}\}$$

where

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix}.$$

We compute then that

$$K = \{-(1/4, 1/4, 1/2, 0) + t(1, -3, 0, 2) : t \in \mathbb{Z}/p\mathbb{Z}\}.$$

Note that for $k = (k_1, k_2, k_3, k_4) \in K$ we have simply $k_3 = (p-1)/2$. Thus

$$\begin{aligned} \alpha(wxy) = & \sum_{k \in K} \sigma^{-1}(a)^{k_2}(-1)^{k_3}\sigma^{-1}(b)^{k_4}(-p)^{(|k|+1)/p}(\pi w)^{(|k|+1)/p}x^{(3k_1+k_2)/p}y \\ & \cdot\left(\sum_{e \geqslant 0}(-p)^{|e|}\ell_{k+pe}(-1)^{e_3}a^{e_2}b^{e_4}(\pi w)^{|e|}x^{3e_1+e_2}y^{2e_3}\right). \end{aligned} \tag{6.2}$$

We now reduce the series (6.2). Our algorithm to do this would proceed by degree; to make the resulting matrices digestible, we alter the method slightly. We reduce first with respect to the monomial y^2, or ‘vertically’. We have

$$D_y(w^dx^uy^{v-2}) = (v-2)w^dx^uy^{v-2} + \pi wf_y(w^dx^uy^{v-2}) = 0 \in B,$$

and $wf_y = -2wy^2$ so

$$\pi w^dx^uy^v = \tfrac{1}{2}(v-2)w^{d-1}x^uy^{v-2} \in B.$$

By induction, if v is odd we have

$$(\pi w)^{\lfloor v/2 \rfloor}x^uy^v \equiv \frac{1}{2^{\lfloor v/2 \rfloor}}\left((v-2)(v-4)\ldots 3 \cdot 1\right)x^uy = \lfloor v/2 \rfloor!(-1)^{\lfloor v/2 \rfloor}\binom{-1/2}{\lfloor v/2 \rfloor}x^uy.$$

Thus in the series (6.2) we have

$$\sum_{e \geqslant 0}\lambda_{k+pe}(-1)^{e_3}a^{e_2}b^{e_4}w^{|e|}x^{3e_1+e_2}y^{2e_3} = \sum_{e \geqslant 0}\frac{\lambda_{k+pe}}{\pi^{e_3}}e_3!\binom{-1/2}{e_3}a^{pe_2}b^{pe_4}w^{|e|-e_3}x^{3e_1+e_2}. \tag{6.3}$$

We now reduce with respect to x^3, or ‘horizontally’, then finally reducing with respect to the origin. We compute the matrix P_{x^3} such that

$$wx^3\begin{pmatrix} w^2x^2y \\ w^2x^3y \\ w^2x^4y \end{pmatrix} = P_{x^3}\begin{pmatrix} wf \\ wf_x \\ wf_y \end{pmatrix}$$

and has coefficients in $\mathbb{Z}_q\langle w^2x^2y, w^2x^3y, w^2x^4y\rangle$. Some linear algebra then yields

$$P_{x^3} = \frac{w^2x^2y}{4a^3+27b^2}\begin{pmatrix} 3ax(2ax-3b) & -2a^2x^2+3b(ax+3b) & -\frac{3}{2}ax(2ax-3b) \\ -ax(9bx+2a^2) & 3abx^2+(2a^3+9b^2)x+2a^2b & \frac{1}{2}ax(9bx+2a^2) \\ -a^2x(2ax-3b) & (2a^3+9b^2)x^2-ab(ax+3b) & \frac{1}{2}a^2x(2ax-3b) \end{pmatrix}.$$

Now, if

$$\xi = wf\eta_w + wf_x\eta_x + wf_y\eta_y,$$

then

$$-\pi\xi = \frac{\partial}{\partial w}\eta_w + \frac{\partial}{\partial x}\eta_x + \frac{\partial}{\partial y}\eta_y \in B.$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

This will allow us to write

$$-\pi w^d x^u \begin{pmatrix} w^2x^2y \\ w^2x^3y \\ w^2x^4y \end{pmatrix} \equiv D_{x^3}(d,u) w^{d-1}x^{u-3} \begin{pmatrix} w^2x^2y \\ w^2x^3y \\ w^2x^4y \end{pmatrix}$$

for a matrix $D_{x^3}(d,u)$ with coefficients in $\mathbb{Z}_q[d,u]$ which are linear in d, u.

The rows of the matrix D_{x^3} are obtained as follows. For $i=1,2,3$, we have

$$(wx^3)(w^2x^{i+1}y) = p_{iw}f_w + p_{ix}f_x + p_{iy}f_y$$

where (p_{iw}, p_{ix}, p_{iy}) is the ith row of P_{x^3}. Thus

$$w^dx^u = (w^{d-1}x^{u-3})(wx^3)(w^2x^{i+1}y) = (w^{d-1}x^{u-3})\left(p_{iw}f_w + p_{ix}f_x + p_{iy}f_y\right)$$

and hence

$$\begin{aligned} -\pi w^dx^u(w^2x^{i+1}y) &\equiv w\frac{\partial}{\partial w}(w^{d-1}x^{u-3}p_{iw}) + x\frac{\partial}{\partial x}(w^{d-1}x^{u-3}p_{ix}) + y\frac{\partial}{\partial y}(w^{d-1}x^{u-3}p_{iy}) \\ &= (d+1)w^{d-1}x^{u-3}p_{ix} + w^{d-1}\left((u-3)x^{u-3}p_{ix} + x^{u-2}p'_{ix}\right) + w^{d-1}x^{u-3}p_{iy} \\ &= w^{d-1}x^{u-3}\left((d+1)p_{iw} + (u-3)p_{ix} + xp'_{ix} + p_{iy}\right). \end{aligned}$$

From this we obtain that $(4a^3+27b^2)D_{x^3}(d,u)$ is equal to

$$\begin{pmatrix} -3ab^2(u-1) & \frac{1}{2}a^2b(6d-2u+3) & -2a^3d+(2a^3+9b^2)u+(a^3+9b^2) \\ 2a^2b(u-1) & -2a^3d+(2a^3+9b^2)u-a^3 & -\frac{3}{2}ab(6d-2u+1) \\ 9b^2(u-1) & -\frac{3}{2}ab(6d-2u-3) & a^2(6d-2u+1) \end{pmatrix}.$$

In a completely analogous manner, we obtain the matrix D_1 satisfying

$$-\pi w^d x^u \begin{pmatrix} w^2x^2y \\ w^2x^3y \\ w^2x^4y \end{pmatrix} \equiv D_1(d,u) w^{d-1}x^{u} \begin{pmatrix} w^2x^2y \\ w^2x^3y \\ w^2x^4y \end{pmatrix},$$

and indeed we have $b(4a^3+27b^2)D_1$ is equal to

$$\begin{pmatrix} (4a^3+27b^2)d-(4a^3+9b^2)u-\frac{1}{2}(12a^3+9b^2) & -3ab(6d-2u-3) & 2a^2(6d-2u-5) \\ 4a^2b(u+2) & \frac{9}{2}b^2(6d-2u-3) & 3ab(6d-2u-5) \\ -6ab^2(u+2) & a^2b(6d-2u-3) & \frac{9}{2}b^2(6d-2u-5) \end{pmatrix}.$$

Note the additional need to invert b for the matrix D_1.

To reduce the series (6.3), we expand and rewrite it as

$$\begin{aligned} \alpha(wxy) &= \sum_{i\geqslant 0} w^ix^2y((c_{i0}+c_{i1}x+c_{i2}x^2) + wx^3(c_{i3}+c_{i4}x+c_{i5}x^2) + \ldots) \\ &= \sum_{i\geqslant 0} w^ix^2y\Bigg(\sum_{j\geqslant 0} x^{3j}(c_{i,3j+2}+c_{i,3j+3}x+c_{i,3j+4}x^2)\Bigg) \end{aligned} \qquad (6.4)$$

with $c_{ij}\in\mathbb{Z}_q$. We have

$$\mathrm{ord}_p(c_{i,j'}) \geqslant (i+\lfloor j'/3\rfloor)\frac{p-1}{p},$$

so up to precision p^N we need only take $i < p/(p-1)N$ and

$$j < J = \frac{p}{p-1}N - i.$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Each sum in (6.4) is reduced using the matrix D_{x^3}: if we abbreviate

$$c_i(j) = \begin{pmatrix} c_{i,3j+2} \\ c_{i,3j+3} \\ c_{i,3j+4} \end{pmatrix}$$

and write

$$\widehat{D}_{x^3}(i,j) = D_{x^3}(i,0)D_{x^3}(i+1,3)\dots D_{x^3}(j+i,3j)$$

we have

$$w^i x^2 y \sum_{j=0}^{J-1} c_{ij} w^{\lfloor j/3 \rfloor} x^j = w^i x^2 y \sum_{j=0}^{J} \widehat{D}_{x^3}(i,j) c_i(j).$$

For what it is worth, one can check that this algorithm runs in time $O(p \log^6 q)$.

Fermat-like hypersurfaces

In this subsection, we show how the method works in the simplest case where one has a Fermat-like affine hypersurface defined by

$$\overline{f} = \overline{a}_1 x_1^{m_1} + \dots + \overline{a}_n x_n^{m_n} + \overline{b} \in \mathbb{F}_q[x_1, \dots, x_n] = \mathbb{F}_q[x]$$

where $m_i \in \mathbb{Z}_{>0}$ and $\overline{a}_1 \dots \overline{a}_n \overline{b} \neq 0$. These were the varieties considered by Weil [**61**] in his seminal article on zeta functions. Koblitz [**39**] has studied these and shown that the number of points is given by Jacobi sums, which can be expressed by Gauss sums; see Wan [**59**] for an explicit algorithm which uses this method. In an early related work, Delsarte [**13**] studied the number of zeros of a polynomial

$$\overline{g} = \sum_{j=1}^{n} \overline{b_j} x^{\nu(j)} + \overline{c} \in \mathbb{F}_q[x_1, \dots, x_n]$$

in $\mathbb{F}_q$ and its finite extensions, where $\nu(j) \in \mathbb{Z}^n_{\geqslant 0}$; he described an explicit formula for this number in terms of Jacobi sums.

The polynomial $\overline{f}$ is nondegenerate if and only if $p \nmid m_1 \dots m_n$, which we now assume.

Let $f = a_1 x_1^{m_1} + \dots + a_n x_n^{m_n} + b \in \mathbb{Z}_q[x]$ be the Teichmüller lift of f to $\mathbb{Z}_q[x]$. We then have the polytope

$$\Delta = \Delta(f) = \Delta(\{(m_1, 0, \dots, 0), \dots, (0, \dots, 0, m_n), (0, \dots, 0)\})$$

with normalized volume $\mathrm{Vol}(\Delta) = n!\,\mathrm{vol}(\Delta) = m_1 \dots m_n$.

Here, the affine complex gives the cohomology space

$$B = \frac{L^{(x_1 \dots x_n)}}{D_w L^{(x_1 \dots x_n)} + D_1 L^{(x_2 \dots x_n)} + \dots + D_n L^{(x_1 \dots x_{n-1})}}$$

where

$$D_w = w\frac{\partial}{\partial w} + \pi f_w, \qquad D_i = x_i \frac{\partial}{\partial x_i} + \pi f_{x_i}$$

for $i = 1, \dots, n$, and

$$f_w = wf, \qquad f_{x_i} = (w a_i m_i) x^{m_i}$$

also for $i = 1, \dots, n$.

A basis V for the space B is computed as follows. In weight 1, we have V_1 consisting of lattice points in Δ not on a coordinate face,

$$V_1 = \left\{ w x^\mu : (1/m)\mu = \sum\nolimits_i \mu_i/m_i \leqslant 1,\ \mu > 0 \right\}.$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

In a similar way, in degree d, we obtain

$$V_d = \{w^d x^\mu : i - 1 < (1/m)\mu \leqslant i,\ \ \mu > 0,\ \ \mu < m\}.$$

In particular, we see clearly that $V_d = \{0\}$ for $d \geqslant n+1$.

Now we consider each series expansion (2.14). We have

$$K = \{e : Ue = \mu \pmod{p}\}$$

where

$$U = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & m_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_n \end{pmatrix}$$

which is an invertible $(n+1) \times (n+1)$ matrix, so K consists of the single element

$$k = M^{-1}(-\mu) = \begin{pmatrix} -1 & m_1^{-1} & m_2^{-1} & \dots & m_n^{-1} \\ 0 & -m_1^{-1} & 0 & \dots & 0 \\ 0 & 0 & -m_2^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -m_n^{-1} \end{pmatrix} \mu.$$

For simplicity of notation, we consider the indices of the columns to be $0, \dots, n$.

The reduction steps are similarly simple. Since $f_i = a_i m_i w x_i^{m_i}$, we have

$$(\pi w)^d x^\mu \equiv -\frac{\mu_i - m_i}{a_i}(\pi w)^{d-1} x^\mu x_i^{-m_i} \in B$$

whenever $\mu_i \geqslant m_i$. We obtain a similar result from the equality

$$w^d = \frac{1}{b} w^{d-1} f_w - \frac{1}{b(m_1 \dots m_n)} \sum_{i=1}^{n} w^{d-1} f_i$$

which implies

$$(\pi w)^d \equiv -\frac{d-1}{b}(\pi w)^{d-1} \in B.$$

Inductively packing these up, in the terms of the expansion

$$\begin{aligned} \alpha((\pi w)^d x^\mu) &= \sigma^{-1}(a^k)(-p)^{(|k|+d)/p}(\pi w)^{(|k|+d)/p} x^{(k\nu+\mu)/p} \\ &\quad \times \Bigg(\sum_{e=(e_0,\dots,e_n)\geqslant 0} (-p)^{|e|} \ell_{k+pe} a^e (\pi w)^{|e|} x^{e\nu} \Bigg) \\ &= \sigma^{-1}(a^k)(-p)^{(|k|+d)/p} \sum_{e \geqslant 0} (-p)^{|e|} \ell_{k+pe} a^e (\pi w)^{|e|+(|k|+d)/p} x^{e\nu+(k\nu+\mu)/p} \end{aligned}$$

and letting $\kappa = (k\nu+\mu)/p = \big((k_i m_i + \mu_i)/p\big)_{i=1,\dots,n}$ we have

$$\begin{aligned} (\pi w)^{|e|+(|k|+d)/p} x^{e\nu+(k\nu+\mu)/p} &\equiv (-1)^{e_0}(e_0 + (|k|+d)/p; 1) \\ &\quad \cdot \prod_{i=1}^{n} (-a_i^{-1})^{e_i} (e_i m_i + \kappa_i; m_i)_{e_i} (\pi w)^{(|k|+d)/p} x^\kappa \end{aligned}$$

where

$$(z; q)_r = (z-q)(z-2q)\dots(z-rq).$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

If we let $m_0 = 1$ and $\kappa_0 = (|k| + d)/p$, then we can abbreviate

$$(em + \kappa; m)_e = \prod_{i=0}^{n} (e_i m_i + \kappa_i; m_i)_{e_i}$$

and, substituting back into the sum, we obtain

$$\alpha((\pi w)^d x^\mu) = (\pi w)^{(|k|+d)/p} x^\kappa \left(\sigma^{-1}(a^k)(-p)^{(|k|+d)/p} \sum_{e \geqslant 0} p^{|e|} \ell_{k+pe}(em + \kappa; m)_e \right).$$

In this way, we have written down the reduced value in one stroke.

Gabber hypersurfaces

There are few works (if any) in the existing literature which give the computation of the zeta function of a projective hypersurface defined over $\mathbb{F}_q$ of degree d with $p \mid d$. Here we study the zeta function of such hypersurfaces for a particular family going back (according to oral communication from Nicholas Katz) to Ofer Gabber (see for example [**33**, 11.4.6]).

Let d be a positive integer with $p \mid d$ and let

$$\overline{f}(x) = a_1 x_1^d + \sum_{i=2}^{n} a_i x_{i-1} x_i^{d-1} \in \mathbb{F}_q[x_1, \ldots, x_n] \tag{6.5}$$

with $a_1 a_2 \ldots a_n \neq 0$. Let $\overline{Z}_0$ be the projective hypersurface in $\mathbb{P}^{n-1}_{\mathbb{F}_q}$ defined by $\overline{f} = 0$. The hypersurface $\overline{Z}_0$ is easily seen to be nonsingular. Note that its defining equation (6.5) is fewnomial.

We consider somewhat more general hypersurfaces as follows. Let $c = (c_1, \ldots, c_n)$ be an n-tuple of positive integers such that $|c| = \sum_{i=1}^n c_i = d$. Suppose further that c is an interior point in $\Delta(\overline{f})$. We consider here the family of hypersurfaces $\overline{Z}_\lambda$ defined by

$$\overline{f}_\lambda(x) = \overline{f}(x) + \lambda x^c.$$

Since $\overline{Z}_0$ is nonsingular and the condition of nonsingularity is open, there is a closed subset $W \subseteq \mathbb{A}^1$, defined by the vanishing of a polynomial in $\mathbb{F}_q[\lambda]$, such that $\overline{Z}_\lambda$ is singular if and only if $\lambda \in W$.

If $\lambda \notin W$, then we recall

$$Z(\overline{Z}_\lambda, T) = \frac{P_\lambda(T)^{(-1)^{n-1}}}{(1-T)(1-qT)\ldots(1-q^{n-2}T)}$$

with the degree of P_λ equal to $d^{-1}((d-1)^n + (-1)^n(d-1))$.

EXAMPLE 6.6. An example of this situation is provided by the family of elliptic curves

$$\overline{Z}_\lambda : x_1^3 + x_1 x_2^2 + x_2 x_3^2 + \lambda x_1 x_2 x_3$$

in characteristic 3. Here the singular locus is given by $Q(\lambda) = \lambda^4 - 1 = 0$, the fourth roots of 1. We note that the j-invariant of $\overline{Z}_\lambda$ is equal to $j(\overline{Z}_\lambda) = \lambda^{12}/(\lambda^4 - 1)$.

It will be useful to consider the exponential sums (and their associated L-functions) for $w\overline{f}_\lambda(x)$ as $(w, (x_1, \ldots, x_{n-1}), x_n)$ runs over various spaces which are products of tori with affine spaces. For example, we recall from the sections on affine and projective hypersurfaces in § 5 that

$$L(w\overline{f}_\lambda, \mathbb{G}_m \times \mathbb{A}^{n-1} \times \mathbb{A}^1, T)^{(-1)^{n+1}} = \frac{P_\lambda(qT)}{P_\lambda(q^2T)}. \tag{6.7}$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

It is our intention to show that (6.7) has the form

$$\frac{P_\lambda(qT)}{P_\lambda(q^2T)} = \frac{\prod_{r=0}^{n-1} R_r(T)}{\prod_{r=0}^{n-1} R_r(qT)}$$

where $R_r(T)$ is a polynomial or reciprocal polynomial in $\mathbb{Z}[T]$ for $r = 0, \ldots, n-1$ which we give explicitly. Then, however, $\prod_{r=0}^{n-1} R_r(T) = P_\lambda(qT)$ so that this calculation of $R_r(T)$ for $0 \leqslant r \leqslant n-1$ in fact gives $Z(\overline{Z}_\lambda, T)$.

Since $\overline{f}_\lambda(x)$ is homogeneous, we have $\dim \Delta(\overline{f}_\lambda) = n-1$. For $\lambda \notin W$, the nonsingularity of $\overline{Z}_\lambda$ implies that $w\overline{f}_\lambda$ is nondegenerate with respect to its maximal face (not containing 0). It is easy to see it is also nondegenerate with respect to its other (lower-dimensional) faces not containing 0.

We cannot apply directly our results from the section on projective hypersurfaces in § 5 because p divides the degree d and, moreover, $w\overline{f}_\lambda$ is not convenient with respect to the variables $x_1, \ldots, x_n$. It is, however, convenient with respect to x_n. We proceed by partitioning $\mathbb{G}_m \times \mathbb{A}^{n-1} \times \mathbb{A}^1$ and particularly its middle factor as follows:

$$\mathbb{A}^{n-1} = \mathbb{G}_m^{n-1} \cup \bigcup_{r=1}^{n-1} U_r \tag{6.8}$$

where $U_r = \mathbb{G}_m^{r-1} \times \{0\} \times \mathbb{A}^{n-1-r}$. For notational convenience, we will sometimes write $U_0 = \mathbb{G}_m^{n-1}$. Note that the terms U_r for $r = 0, \ldots, n-1$ are pairwise disjoint, so that if we set $U_r' = \mathbb{G}_m \times U_r \times \mathbb{A}^1$, then

$$L(w\overline{f}_\lambda, \mathbb{G}_m \times \mathbb{A}^{n-1} \times \mathbb{A}^1, T) = \prod_{i=0}^{n-1} L(w\overline{f}_\lambda, U_r', T).$$

By the work of Adolphson and Sperber [2], the complex $\Omega^\bullet(w\overline{f}_\lambda, U_0')$ for $L(w\overline{f}_\lambda, U_0', T)$ is acyclic, except in dimensions n and $n+1$, and

$$L(w\overline{f}_\lambda, U_0', T)^{(-1)^{n+1}} = \frac{R_0(T)}{R_0(qT)}$$

where

$$R_0(T) = \det\bigl(1 - \operatorname{Frob} T \mid H^{n+1}(\Omega^\bullet(w\overline{f}_\lambda, U_0'))\bigr).$$

We see easily, from the relation of $R_0(T)$ and the zeta function for the variety defined by the vanishing of $\overline{f}_\lambda$ in $\mathbb{G}_m^{n-1} \times \mathbb{A}^1$, that $R_0(T) \in \mathbb{Z}[T]$. The calculation of $R_0(T)$ follows easily from the suppression of the w terms and the isomorphism

$$H^{n+1}(\Omega^\bullet(w\overline{f}_\lambda, U_0')) \cong \frac{(L_\Delta')^{\{x_n\}}}{\sum_{i=1}^{n-1} D_i(L_\Delta')^{\{x_n\}} + D_n L_\Delta'}.$$

Consider now for $1 \leqslant r \leqslant n-1$ the L-function associated to the exponential sum $S(w\overline{f}_\lambda, U_r')$. Write

$$\overline{f}_\lambda^{(r)}(x) = a_1 x_1^d + \sum_{i=1}^{r-1} a_i x_{i-1} x_i^{d-1}$$

and

$$\overline{g}_\lambda^{(r)}(x) = \sum_{i=r+2}^{n} a_i x_{i-1} x_i^{d-1}.$$

Note that substituting $x_r = 0$ in $\overline{f}_\lambda(x)$ gives

$$\overline{f}_\lambda(x)|_{x_r=0} = \overline{f}_\lambda^{(r)}(x) + \overline{g}_\lambda^{(r)}(x). \tag{6.9}$$

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

Note also that, despite the notation, none of the polynomials in (6.9) (for $r \geqslant 1$) depend on λ. Thus

$$\begin{aligned} S(w\overline{f}_\lambda, U_r') = &\sum_{(w,x_1,\ldots,x_{r-1})} \Theta(w\overline{f}_\lambda^{(r)}(x_1,\ldots,x_{r-1})) \\ &\times \Bigg(\sum_{(x_{r+1},\ldots,x_n)\in\mathbb{A}^{n-r}} \Theta(w\overline{g}_\lambda^{(r)}(x_{r+1},\ldots,x_n)) \Bigg). \end{aligned}$$

For any $w \in \mathbb{G}_m$, the inner sum is easily seen to be q^{n-r-1}; that is,

$$S(w\overline{f}_\lambda, U_r') = q^{n-r-1} S(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r)$$

and

$$L(w\overline{f}_\lambda, U_r', T) = L(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r, q^{n-r-1}T).$$

Consider the complex $\Omega^\bullet(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r)$ for the L-function $L(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r, T)$. It is acyclic $H^i(\Omega^\bullet) = 0$ except for $i = r-1$ and $i = r$, and if we let

$$H^{(r)}(T) = \det\big(1 - \operatorname{Frob} T \mid H^r(\Omega^\bullet(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r))\big),$$

then

$$L(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r, T)^{(-1)^{r+1}} = \frac{H^{(r)}(T)}{H^{(r)}(qT)}.$$

Thus, if we let

$$R_r(T) = \det\big(1 - q^{n-r-1} \operatorname{Frob} T \mid H^r(\Omega^\bullet(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r))\big)^{(-1)^{n-r}},$$

then, as with $R_0(T)$, we have $R_r(T) \in \mathbb{Z}[T]$.

We can calculate $R_r(T)$ using

$$H^r(\Omega^\bullet(w\overline{f}_\lambda^{(r)}, \mathbb{G}_m^r)) \cong \frac{L_{\Delta^{(r)}}}{\sum_{i=1}^{r-1} D_i L_{\Delta^{(r)}}}$$

where we have suppressed the factor of w. Note that, in the case $r = 1$, we have

$$L(w\overline{f}_\lambda, U_1', T) = \frac{1 - q^{n-2}T}{1 - q^{n-1}T}$$

so that $R_1(T) = (1 - q^{n-2}T)^{(-1)^{n+1}}$. Similarly, if $r = 2$, then

$$L(w\overline{f}_\lambda, U_2', T) = \frac{1 - q^{n-2}T}{1 - q^{n-3}T}$$

so $R_2(T) = (1 - q^{n-3}T)^{(-1)^n}$.

Acknowledgements. This work was initiated during the Thematic Year on Applications of Algebraic Geometry at the Institute for Mathematics and its Applications (IMA) in 2006 and 2007, and the authors would like to thank the IMA for their hospitality. We would like to thank: David Harvey, Alan Lauder, Kiran Kedlaya, and Daqing Wan for helpful discussions; Wouter Castryck, Frank Sottile, and the anonymous referee for some corrections; and Jesus de Loera, Benjamin Nill, and Bernd Sturmfels for answering some questions about algorithms for polytopes.

The second author was partially supported by the National Security Agency under Grant Number H98230-09-1-0037.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

References

1. T. Abbott, K. Kedlaya and D. Roe, 'Bounding Picard numbers of surfaces using p-adic cohomology', Arithmetic, geometry and coding theory (AGCT 2005), Séminaires et Congrès 21 (Société Mathématique de France, 2009) 125–159.
2. A. Adolphson and S. Sperber, 'Exponential sums and Newton polyhedra: cohomology and estimates', *Ann. of Math.* (2) 130 (1989) 367–406.
3. A. Adolphson and S. Sperber, 'p-adic estimates for exponential sums', *p-adic analysis (Trento, 1989)*, Lecture Notes in Mathematics 1454 (Springer, Berlin, 1990) 11–22.
4. A. Adolphson and S. Sperber, 'Exponential sums nondegenerate relative to a lattice', *Algebra Number Theory* 3 (2009) 881–906.
5. V. Batyrev and D. Cox, 'On the Hodge structure of projective hypersurfaces in toric varieties', *Duke Math. J.* 75 (1994) 293–338.
6. D. J. Bates (IMA), F. Bihan and F. Sottile, 'Bounds on the number of real solutions to polynomial equations', Preprint, 2007, arXiv:math.AG/0706.4134.
7. A. Bostan, P. Gaudry and E. Schost, 'Linear recurrences with polynomial coefficients and application to integer factorization and Cartier–Manin operator', *SIAM J. Comput.* 36 (2007) 1777–1806.
8. W. Bruns and R. Koch, 'Computing the integral closure of an affine semigroup', *Univ. Iagel. Acta Math.* 39 (2001) 59–70.
9. W. Castryck, 'Point counting on nondegenerate curves', PhD Thesis, Katholieke Universiteit, Leuven, 2006.
10. W. Castryck, J. Denef and F. Vercauteren, 'Computing zeta functions of nondegenerate curves', Int. Math. Res. Pap. IMRP 2006 (2006) 72017.
11. W. Castryck and J. Voight, 'On nondegeneracy of curves', *Algebra Number Theory* 3 (2009) 255–281.
12. B. Chazelle, 'An optimal convex hull algorithm in any fixed dimension', *Discrete Comput. Geom.* 10 (1993) 377–409.
13. J. Delsarte, 'Nombre de solutions des équations polynomiales sur un corps fini', *Séminaire Bourbaki*, vol. 1, Exp. No. 39 (1948–1951) (Société Mathématique de France, Paris, 1995) 321–329.
14. J. Denef and F. Loeser, 'Weights of exponential sums, intersection cohomology, and Newton polyhedra', *Invent. Math.* 106 (1991) 275–294.
15. P. D. Domich, R. Kannan and L. E. Trotter Jr., 'Hermite Normal Form computation using modulo determinant arithmetic', *Math. Oper. Res.* 12 (1987) 50–59.
16. B. Dwork, 'On the rationality of the zeta function of an algebraic variety', *Amer. J. Math.* 82 (1960) 631–648.
17. B. Dwork, 'On the zeta function of a hypersurface', *Publ. Math. Inst. Hautes Études Sci.* 12 (1962) 5–68.
18. B. Dwork, 'A deformation theory for the zeta function of a hypersurface', *Proceedings of the International Congress of Mathematicians, Stockholm, 1962* (Institut Mittag-Leffler, Djursholm, Sweden, 1963) 247–259.
19. B. Dwork, 'On p-adic analysis', *Some recent advances in the basic sciences*, vol. 2, Proc. Annual Sci. Conf., New York, 1965–1966 (Belfer Graduate School of Science, Yeshiva University, New York, 1969) 129–154.
20. B. Dwork, 'p-adic cycles', *Publ. Math. Inst. Hautes Études Sci.* 37 (1969) 27–115.
21. B. Edixhoven, Point counting after Kedlaya, lecture notes, EIDMA-Stieltjes graduate course, Leiden, 2003.
22. B. Edixhoven, J.-M. Couveignes, R. de Jong, F. Merkl and J. Bosman, 'Computational aspects of modular forms and Galois representations', Preprint, 2010, arXiv:math.NT/0605244.
23. S. Fortune, 'Voronoi diagrams and Delaunay triangulations', *Computing in Euclidean geomtry* (eds D.-Z. Du and F. K. Hwang; World Scientific, 1995) 193–233.
24. S. Fortune, 'Voronoi diagrams and Delaunay triangulations', *Handbook of discrete and computational geometry* (eds J. E. Goodman and J. O'Rourke; CRC Press, Boca Raton, 1997) 377–388.
25. R. Gerkmann, 'Relative rigid cohomology and deformation of hypersurfaces', Int. Math. Res. Pap. IMRP 2007 (2007) rpm003.
26. P. Gritzmann, V. Klee and D. G. Larman, 'Largest j-simplices n-polytopes', *Discrete Comput. Geom.* 13 (1995) 477–515.
27. D. Harvey, 'Kedlaya's algorithm in larger characteristic', Int. Math. Res. Not. IMRN 2007 (2007) rnm095.
28. D. Harvey, 'Computing zeta functions of projective surfaces in large characteristic', Preprint.
29. R. R. Howell, 'On asymptotic notation with multiple variables', Technical Report 2007-4, Preprint, 2008, http://people.cis.ksu.edu/~rhowell/asymptotic.pdf.
30. S. Kadir, 'The arithmetic of Calabi–Yau manifolds and mirror symmetry', PhD Thesis, University of Oxford, Oxford, 2004.
31. R. Kannan and A. Bachem, 'Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix', *SIAM J. Comput.* 8 (1979) 499–507.
32. N. M. Katz, 'On the differential equations satisfied by period matrices', *Publ. Math. Inst. Hautes Études Sci.* 35 (1968) 71–106.
33. N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy* (American Mathematical Society, Providence, RI, 1998).
34. K. Kedlaya, 'Counting points on hyperelliptic curves using Monsky–Washnitzer cohomology', *J. Ramanujan Math. Soc.* 16 (2001) 323–338.
35. K. Kedlaya, 'Search techniques for root-unitary polynomials', *Computational arithmetic geometry*, Contemporary Mathematics 463 (American Mathematical Society, Providence, RI, 2008) 71–81.

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press

36. K. Kedlaya, ‘Computing zeta functions of nondegenerate toric hypersurfaces: two proposals’, Preprint.
37. A. G. Khovanskii, ‘Newton polyhedra and toroidal varieties’, *Funct. Anal. Appl.* 11 (1977) 289–296.
38. N. Koblitz, *p-adic numbers, p-adic analysis, and zeta-functions*, Graduate Texts in Mathematics 58 (Springer, 1977).
39. N. Koblitz, ‘The number of points on certain families of hypersurfaces over finite fields’, *Compositio Math.* 48 (1983) 3–23.
40. A. G. Kouchnirenko, ‘Polyèdres de Newton et nombres de Milnor’, *Inv. Math.* 32 (1976) 1–31.
41. A. G. Kouchnirenko, *Fewnomials*, Translations of Mathematical Monographs 88 (American Mathematical Society, Providence, RI, 1991).
42. J. Lagarias and G. Ziegler, ‘Bound for lattice polytopes containing a fixed number of interior points in a sublattice’, *Canad. J. Math.* 43 (1991) 1022–1035.
43. A. G. B. Lauder, ‘Deformation theory and the computation of zeta functions’, *Proc. Lond. Math. Soc. (3)* 88 (2004) 565–602.
44. A. G. B. Lauder, ‘Counting solutions to equations in many variables over finite fields’, *Found. Comput. Math.* 4 (2004) 221–267.
45. A. G. B. Lauder, ‘A recursive method for computing zeta functions of varieties’, *LMS J. Comput. Math.* 9 (2006) 222–269.
46. A. G. B. Lauder and D. Wan, ‘Computing zeta functions of Artin–Schreier curves over finite fields’, *LMS J. Comput. Math.* 5 (2002) 34–55.
47. A. G. B. Lauder and D. Wan, ‘Computing zeta functions of Artin–Schreier curves over finite fields. II’, *J. Complexity* 20 (2004) 331–349.
48. A. G. B. Lauder and D. Wan, ‘Counting points on varieties over finite fields of small characteristic’, *Algorithmic number theory: lattices, number fields, curves and cryptography*, Mathematical Sciences Research Institute Publications 44 (Cambridge University Press, Cambridge, 2008) 579–612.
49. J. de Loera, R. Hemmecke and M. Köppe, ‘Pareto optima of multicriteria integer programs’, *INFORMS J. Comput.* 21 (2009) 39–48.
50. A. R. Mavlyutov, ‘Cohomology of complete intersections in toric varieties’, *Pacific J. Math.* 191 (1999) 133–144.
51. P. Monsky, *p-adic analysis and zeta functions*, Lectures in Mathematics (Kyoto University, Tokyo, 1970) http://www.math.kyoto-u.ac.jp/library/lmku/lmku04.pdf.
52. F. P. Preparata and M. I. Shamos, *Computational geometry: an introduction*, 3rd edn (Springer, 1991).
53. D. Savitt, D. Thakur, M. Baker, B. Conrad, S. Dasgupta, K. Kedlaya and J. Teitelbaum, *p-adic geometry: lectures from the 2007 Arizona winter school*, University Lecture Series 45 (American Mathematical Society, Providence, RI, 2008).
54. R. Schoof, ‘Elliptic curves over finite fields and the computation of square roots mod p’, *Math. Comp.* 44 (1985) 483–494.
55. R. Seidel, ‘Convex hull computations’, *Handbook of discrete and computational geometry* (eds J. E. Goodman and J. O’Rourke; CRC Press, Boca Raton, FL, 1997) 361–375.
56. J.-P. Serre, ‘Zeta and L-functions’, *Arithmetic algebraic geometry* (ed. Schilling; Harper and Row, New York, NY, 1965).
57. J. Tuitman, ‘Counting points in families of nondegenerate curves’, PhD Thesis, Katholieke Universiteit, Leuven, 2010.
58. D. Wan, ‘Computing zeta functions over finite fields’, *Contemp. Math.* 225 (1999) 131–141.
59. D. Wan, ‘Modular counting of rational points over finite fields’, *Found. Comput. Math.* 8 (2008) 597–605.
60. D. Wan, ‘Algorithmic theory of zeta functions over finite fields’, *Algorithmic number theory: lattices, number fields, curves and cryptography*, Mathematical Sciences Research Institute Publications 44 (Cambridge University Press, Cambridge, 2008) 551–578.
61. A. Weil, ‘Numbers of solutions of equations in finite fields’, *Bull. Amer. Math. Soc.* 55 (1949) 497–508.
62. M. Widmer, ‘Lipschitz class, narrow class, and counting lattice points’, Proc. Amer. Math. Soc., to appear.
63. C. F. Wong, ‘Zeta functions of projective toric hypersurfaces over finite fields’, PhD thesis, University of California, Irvine, 2008.

Steven Sperber
School of Mathematics
University of Minnesota
127 Vincent Hall
Minneapolis 55455
USA

sperber@math.umn.edu

John Voight
Department of Mathematics and Statistics
University of Vermont
16 Colchester Avenue
Burlington VT 05401-1455
USA

jvoight@gmail.com

https://doi.org/10.1112/S1461157012001179 Published online by Cambridge University Press