

Computing zeta functions for sparse nondegenerate hypersurfaces using Dwork cohomology

John Voight
University of Vermont

Joint work with Steven Sperber

Counting Points: Theory, Algorithms and Practice
Centre de Recherche Mathématiques (CRM)
21 April 2010

Introduction

Let X be a variety over \mathbb{F}_q . Define the *zeta function* of X by

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

Dwork proved that $Z(X, T)$ is a rational function in T .

To study $Z(X, T)$, we construct a (Weil) cohomology theory which functorially associates to X certain finite-dimensional vector spaces $H^i(X)$, each equipped with a *Frobenius* linear operator F such that $Z(X, T) = \prod_{i=0}^{2n} \det(1 - FT | H^i(X))^{(-1)^{i+1}}$, where $n = \dim X$.

The relevant Dwork cohomology space is the quotient of a p -adic power series ring in $n + 1$ variables by the image of $n + 1$ differential operators. The Frobenius is obtained via the *Dwork splitting function*, and we compute the characteristic polynomial of F by reducing in cohomology using these differential operators.

Main result

Let $\bar{f} \in \mathbb{F}_q[x_1^\pm, \dots, x_n^\pm]$ be a Laurent polynomial. Then \bar{f} defines a hypersurface X in the torus $(\mathbb{G}_{m, \mathbb{F}_q})^n$. If \bar{f} is *nondegenerate*, then the interesting part of $Z(X, T)^{(-1)^n}$ is a polynomial of degree v .

Theorem

Suppose $p \neq 2$. Then there exists an explicit deterministic algorithm which takes as input a nondegenerate Laurent polynomial \bar{f} with s monomials and computes $Z(X, T)$ using $p^{\min(1, s-(n+1))} (nv \log q)^{O(n)}$ bit operations.

There is also a version of this theorem for dense input which uses $(pnv \log q)^{O(n)}$ bit operations.

Our method also works with exponential sums as well as affine or projective hypersurfaces. (With a little more effort, we can also take $p = 2$.)

Nondegenerate hypersurfaces

Let $f = \sum_{\nu \in \mathbb{Z}^n} a_\nu x^\nu \in k[x_1^\pm, \dots, x_n^\pm]$ be a Laurent polynomial.
Let $\Delta = \Delta(f)$ be the *Newton polytope* of f , the convex hull of its *support* $\text{supp}(f) = \{\nu \in \mathbb{Z}^n : a_\nu \neq 0\}$.

For a face $\tau \subset \Delta$, let $f|_\tau = \sum_{\nu \in \tau} a_\nu x^\nu$. Then f is (Δ) -nondegenerate if for all faces $\tau \subset \Delta$ (of any dimension), the system of equations

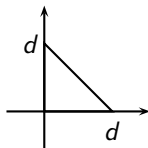
$$f|_\tau = x_1 \frac{\partial f|_\tau}{\partial x_1} = \cdots = x_n \frac{\partial f|_\tau}{\partial x_n} = 0$$

has no solution in $\bar{k}^{\times n}$.

The set of Δ -nondegenerate polynomials with respect to a polytope Δ (with $\dim(\Delta) = n$) forms a Zariski dense open subset in the affine space parameterizing their coefficients $(a_\nu)_{\nu \in \Delta \cap \mathbb{Z}^n}$.

Nondegenerate plane curves

Let $f \in k[x, y]$ have total degree $d \in \mathbb{Z}_{\geq 1}$ and Newton polytope $\Delta = d\Sigma$ as follows.



In other words, $f(x, y) = a_{00} + \cdots + a_{d0}x^d + \cdots + a_{0d}y^d$ with $a_{00}a_{d0}a_{0d} \neq 0$.

Let $\tilde{f}(x, y, z)$ denote the homogenization of f . Then f is nondegenerate if and only if the curve $V(\tilde{f}) \subset \mathbb{P}_k^2$ is nonsingular and is not tangent to any coordinate axis.

Zeta functions of nondegenerate hypersurfaces

If $f \in k[x^\pm, y^\pm]$ is nondegenerate, then the genus of the projective nonsingular curve associated to f is equal to the number of interior lattice points in $\Delta(f)$.

If X is the toric hypersurface associated to f , then $Z(X, T)^{(-1)^n}$ is a polynomial of degree $\nu = n! \text{vol}(\Delta)$. For example, if $f \in k[x, y]$ has $\Delta(f) = d\Sigma$ as above, then $\nu = 2!(d^2/2) = d^2$. (More generally, if f is generic polynomial of degree d in n variables, then this volume is d^n .)

(There is an affine version of this as well: $f \in k[x]$ is *commode* it contains the monomial $x_i^{d_i}$ in its support for each i .)

Lauder and Wan follow the original p -adic methods of Dwork, working on the “chain level”. They compute the zeta function of a polynomial f of total degree d in n variables over \mathbb{F}_q in time $(pd^n \log q)^{O(n)}$.

Note that if the prime p and dimension n are fixed, then the dense input size of f is $O(d^n \log q)$ and the algorithm runs in polynomial time.

Note that the polynomial f may be reducible or badly singular. This algorithm does not seem to be practical.

We work on the level of cohomology and the extension of Dwork's theory due to Adolphson and Sperber (pursued by Lauder and Wan in the case of Artin-Schreier curves).

Comparison to rigid cohomology

Our method follows in the spirit of Kedlaya's method for hyperelliptic curves, which uses Monsky-Washnitzer cohomology. Kedlaya suggested this extension in the realm of toric varieties, and was taken up by Castryck, Denef, and Vercauteran, who consider the class of nondegenerate curves X : they compute $Z(X, T)$ in time $(pg \log q)^{O(1)}$. Note that g is proportional to the corresponding volume v . This method has good asymptotic behavior but also seems impractical, at least at present.

However, rather than working with Monsky-Washnitzer (p -adic de Rham) cohomology, which is more geometric, we employ the cohomology theory of Dwork which has a more combinatorial flavor (working with a general class of exponential sums). One plus is that the precision required can be completely understood.

In a very broad sense, the two approaches seem to be related in a fundamental way.

Deformation

Lauder uses Dwork's theory of p -adic differential equations to compute zeta functions using deformation: the Frobenius action on the members of a one-parameter family satisfies a differential equation coming from the Gauss-Manin connection, and this equation can be used to solve for the Frobenius action given an initial condition. His methods show that one can compute $Z(X, T)$ for a nondegenerate projective hypersurface $X \subset \mathbb{P}^n$ over \mathbb{F}_q of degree d (with $p \nmid d$) in time $p^2(d^n n \log q)^{O(1)}$. (Our algorithm is $p^{s-(n+1)}(d^n n \log q)^{O(n)}$.) See also work of Gerkmann, Hubrechts, and others for the Monsky-Washnitzer version.

Our method fits nicely into this framework as it provides a more general “base variety” to deform from.

Adapting an idea of the Chudnovskys (and Bostan, Gaudry, and Schost), Harvey has improved Kedlaya's method, giving a runtime of $p^{1/2}(g \log q)^{O(1)}$ for hyperelliptic curves...

Dwork's theory: exponential sums

Let $\bar{f} \in \mathbb{F}_q[x^\pm]$.

Let $\theta : \mathbb{F}_q \rightarrow R$ be a nontrivial additive character (with R a commutative ring), so that $\theta(x+y) = \theta(x)\theta(y)$ for all $x, y \in \mathbb{F}_q$. For $x \in \mathbb{F}_q^{\times n}$, we have

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

consequently

$$q\#X(\mathbb{F}_q) = \sum_{\substack{w \in \mathbb{F}_q \\ x \in \mathbb{F}_q^{\times n}}} \theta(w\bar{f}(x)).$$

To package these together to compute $Z(X, T)$, we need to define a system of characters $\theta_r : \mathbb{F}_{q^r} \rightarrow R$ for $r \in \mathbb{Z}_{\geq 1}$.

Dwork's theory: splitting function

We define the *Dwork splitting function*

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

where $\pi \in \overline{\mathbb{Q}}_p$ satisfies $\pi^{p-1} = -p$. We have $\text{ord}_p \lambda_d \geq d(p-1)/p^2$.

Since $\theta(1) = 1 + \pi + O(\pi^2)$ is a primitive p th root of unity, we obtain $\theta_r : \mathbb{F}_{q^r} \rightarrow \mathbb{Z}_p[\pi]$ by $\theta_r(x) = \theta(1)^{\text{Tr}_r(x)}$, where $\text{Tr}_r : \mathbb{F}_{q^r} \rightarrow \mathbb{F}_q$ is the trace map.

Dwork's theory: L -function

For each $r \in \mathbb{Z}_{>0}$, define the exponential sum

$$S_r(w\bar{f}) = \sum_{\substack{w \in \mathbb{F}_{q^r} \\ x \in \mathbb{F}_{q^r}^{\times n}}} \theta_r(w\bar{f}(x)).$$

Define the L -function associated to $w\bar{f}$ as

$$L(w\bar{f}, T) = \exp \left(\sum_{r=1}^{\infty} S_r(w\bar{f}) \frac{T^r}{r} \right).$$

Then the equality $q\#X(\mathbb{F}_q) = S_r(w\bar{f})$ yields $Z(X, qT) = L(w\bar{f}, T)$.

Dwork's theory: power series

On the preceding slides, we related $Z(X, T)$ to $L(w\bar{f}, T)$ via the splitting function $\theta(t) = \exp(\pi(t - t^p)) = \sum_d \lambda_d t^d$ (with $\pi^{p-1} = -p$).

Write $\bar{f}(x) = \sum_\nu \bar{a}_\nu x^\nu \in \mathbb{F}_q[x^\pm]$. For $\bar{a} \in \mathbb{F}_q$, let $a \in \mathbb{Z}_q$ denote its Teichmüller lift, where $\mathbb{Z}_q = W(\mathbb{F}_q)$ is the ring of Witt vectors over \mathbb{F}_q . Let $f(x) = \sum_\nu a_\nu x^\nu$. We then consider

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

The first few $p - 1$ terms are

$$F(w, x) = 1 + \pi(wf) + \cdots + \frac{\pi^{p-1}}{(p-1)!} (wf)^{p-1} + \dots$$

To compute $L(wf, T)$, we define a space of p -adic analytic functions like F with similar support and p -adic growth.

Dwork's theory: the "chain level"

Let

$$L = \left\{ \sum_{d=0}^{\infty} \sum_{\nu \in d\Delta} c_{d,\nu} w^d x^\nu : \text{ord}_p(c_{d,\nu}) \geq d \frac{p-1}{p^2} \right\}.$$

Multiplication by F gives an operator $F : L \rightarrow L$. We also have a "left inverse of Frobenius" $\psi : L \rightarrow L$ defined by

$$\psi(a_\mu x^\mu) = \begin{cases} \sigma^{-1}(a_\mu) x^{\mu/p}, & \text{if } p \mid \mu, \\ 0, & \text{otherwise,} \end{cases}$$

where $\sigma : \mathbb{Z}_q \rightarrow \mathbb{Z}_q$ is $a \mapsto a^p$. Let $\alpha = \psi \circ F$.

Theorem (Dwork trace formula)

For $p = q^a$, we have $\text{tr}(\alpha^{ar}) = q^r S_r(\overline{w\bar{f}})$.

We conclude that $L(\overline{w\bar{f}}, T) \in \mathbb{Q}(T)$ using a criterion due to Borel.

Dwork's theory: the “cohomology level”

The differential operators

$$D_i = x_i \frac{\partial}{\partial x_i} + \pi w \frac{\partial f}{\partial x_i} \quad \text{and} \quad D_w = w \frac{\partial}{\partial w} + \pi w f$$

act on the space L .

They commute and give rise to a chain map on the associated Koszul complex. If $\Delta(\bar{f})$ is nondegenerate, then all the homology spaces are trivial except for $i = 0$.

$$\text{Let } B = \frac{L}{D_w L + D_1 L + \cdots + D_n L}.$$

Then B is a free $\mathbb{Z}_q[\pi]$ -module of dimension $n! \text{vol}(\Delta) = v$, and

$$L(wf, T)^{(-1)^n} = \det(1 - \alpha T \mid B).$$

Dwork's theory modulo p : elliptic curves

Let $E : y^2 = x^3 + \bar{a}x + \bar{b}$ be an affine elliptic curve over \mathbb{F}_p . We use the above method to compute $\#E(\mathbb{F}_p)$ modulo p .

The Newton polytope of $f = x^3 + \bar{a}x + \bar{b} - y^2$ is the triangle $\Delta(\bar{f})$ with vertices $(0, 0)$, $(0, 3)$, $(2, 0)$.

We compute that wxy, w^2xy is an eigenbasis for B . We have

$$\alpha(wxy) = (\psi \circ F)(wxy) = \psi(\theta(wf)wxy) = \psi\left(wxy \sum_d \lambda_d (wf)^d\right)$$

and so

$$\alpha(wxy) \equiv \lambda_{p-1} c_{p-1}(wxy) \pmod{p^2}$$

where c_{p-1} is the coefficient of $(wxy)^{p-1}$ in $(wf)^{p-1}$.

But $\lambda_{p-1} = \pi^{p-1}/(p-1)! = -p/(p-1)! \equiv p \pmod{p^2}$. We conclude that $a_p \equiv -\#E(\mathbb{F}_p) \equiv c_{p-1} \pmod{p}$. Note this takes time $O(p)$ to compute (naïvely).

Outline

We now give an outline of the above strategy and sketch the runtime.

In this exposition, we will ignore some factors that depend on the dimension n .

STEP 1: COMPUTE A BASIS FOR THE JACOBIAN RING. Using linear algebra, to compute a basis for the Jacobian ring

$$B/\pi B = \mathbb{F}_q[w\Delta]/(w(x_i\partial_i f/\partial_i x_i), wf)$$

requires $O(v^3)$ operations in \mathbb{Z}_q/p^N so time $(vN \log q)^{O(1)}$.

STEP 2: COMPUTE DWORK SPLITTING FUNCTION: We ultimately need the splitting function $\theta(t) = \sum_d \lambda_d t^d$ to precision $O(pN)$, which requires time $p(N \log q)^{O(1)}$.

Outline

STEP 3: TORIC ENUMERATION. Brutally expanding $F(w, x) = \prod_{\nu} \theta(a_{\nu} x^{\nu})$ and then computing $\alpha = \psi \circ F$ requires $(pN\nu \log q)^{O(n)}$. We instead use a sparse enumeration:

$$\alpha(w^m x^{\mu}) = \sum_{\substack{e \in \mathbb{Z}_{\geq 0}^s \\ p|(e\nu + \mu) \\ p|(|e| + m)}} \lambda_e \sigma^{-1}(a^e) w^{(|e|+m)/p} x^{(e\nu + \mu)/p}$$

becomes

$$\alpha(w^m x^{\mu}) = \sum_{k \in K} \sigma^{-1}(a^k) w^{(|k|+m)/p} x^{(k\nu + \mu)/p} \left(\sum_{e \in \mathbb{Z}_{\geq 0}^s} \lambda_{k+pe} a^e w^{|e|} x^{e\nu} \right)$$

where K has $O(p^{s-(n+1)})$ terms. This expansion then has $p^{s-(n+1)}(N'\nu)^{O(n)}$ terms to precision N' and so requires $p^{s-(n+1)}(N'\nu)^{O(n)}(\log q)^{O(1)}$ time to compute.

STEP 4: REDUCTION IN COHOMOLOGY. As a consequence of the first step, we obtain a matrix which allows us to reduce $O(v)$ terms at a time.

The ordering of the reduction is governed by an ordering of the vertices of Δ . We must therefore perform $p^{s-(n+1)}(N'v)^{O(n)}$ matrix multiplications requiring $O(v^2)$ multiplications so time $p^{s-(n+1)}(N'v)^{O(n)}(\log q)^{O(1)}$.

An analysis of precision loss gives $N' = O(N)$, so both Step 3 and Step 4 take about the same time; and $N = O(v \log q)$, so the total time required is $p^{s-(n+1)}(Nv \log q)^{O(n)}$.