Recent results on p-adic computation of zeta functions

Kiran S. Kedlaya

Department of Mathematics, Massachusetts Institute of Technology

Number Theory and Cryptography: Open Problems
IPAM (UCLA), October 12, 2006

Zeta functions of algebraic varieties

Definition

For X an algebraic variety over a finite field $\mathbb{F}_q$ (for q a power of the prime p), its *zeta function* is the formal power series

$$\zeta_X(t) = \exp\left(\sum_{n=1}^{\infty} \#X(\mathbb{F}_{q^n})\frac{t^n}{n}\right),$$

where $X(\mathbb{F}_{q^n})$ is the set of $\mathbb{F}_{q^n}$-rational points of X.

Zeta functions of algebraic varieties

Definition

For X an algebraic variety over a finite field $\mathbb{F}_q$ (for q a power of the prime p), its *zeta function* is the formal power series

$$\zeta_X(t) = \exp\left(\sum_{n=1}^{\infty} \#X(\mathbb{F}_{q^n})\frac{t^n}{n}\right),$$

where $X(\mathbb{F}_{q^n})$ is the set of $\mathbb{F}_{q^n}$-rational points of X.

The series $\zeta_X(t)$ represents a rational function of t with integer coefficients (Dwork, Grothendieck), and there are additional restrictions on their zeroes and poles over $\mathbb{C}$ (Deligne).

Zeta functions, point counting, and cryptography

Form of the zeta function for curves

When X is a curve of genus g, we can write

$$\zeta_X(t) = \frac{P(t)}{(1-t)(1-qt)}$$

with P a polynomial of degree $2g$, whose roots in $\mathbb{C}$ lie on the circle $|z| = q^{-1/2}$. The *Jacobian* $J(X)$ is an abelian variety of dimension g, and $J(X)(\mathbb{F}_q)$ ($\cong \mathrm{Pic}^0(X)$, the divisor class group) has order $P(1)$. (If $g = 1$, $X \cong J(X)$ is an elliptic curve.)

Zeta functions, point counting, and cryptography

Form of the zeta function for curves

When X is a curve of genus g, we can write

$$\zeta_X(t) = \frac{P(t)}{(1-t)(1-qt)}$$

with P a polynomial of degree $2g$, whose roots in $\mathbb{C}$ lie on the circle $|z| = q^{-1/2}$. The *Jacobian* $J(X)$ is an abelian variety of dimension g, and $J(X)(\mathbb{F}_q)$ ($\cong \mathrm{Pic}^0(X)$, the divisor class group) has order $P(1)$. (If $g = 1$, $X \cong J(X)$ is an elliptic curve.)

Thus ζ_X can be used to tell whether $\#J(X)(\mathbb{F}_q)$ has a large prime factor. (If $\#J(X)(\mathbb{F}_q)$ has largest prime factor p, the discrete log problem in a generic abelian group of order n is only as hard as in a *cyclic* group of order p.)

The zeta function problem

Problem

Given X explicitly (chosen from some fixed class of varieties), determine $\zeta_X(t)$.

The zeta function problem

Problem

Given X explicitly (chosen from some fixed class of varieties), determine $\zeta_X(t)$.

Typical classes:

- All elliptic curves over $\mathbb{F}_q$.
- All hyperelliptic curves of a fixed genus g over $\mathbb{F}_q$.
- All smooth plane curves of a fixed degree d over $\mathbb{F}_q$.

The zeta function problem

Problem

Given X explicitly (chosen from some fixed class of varieties), determine $\zeta_X(t)$.

Typical classes:

- All elliptic curves over $\mathbb{F}_q$.
- All hyperelliptic curves of a fixed genus g over $\mathbb{F}_q$.
- All smooth plane curves of a fixed degree d over $\mathbb{F}_q$.

Helpful features of these classes:

- Easy to write down random instances (*unirational moduli spaces*).
- Uniform shape of ζ_X (degree of numerator/denominator, fixed factors).

Approaches to the zeta function problem

Generic approaches include:

- Direct counting: enumerate $X(\mathbb{F}_{q^n})$ for $n = 1, 2, \ldots$.
- Shanks's method (curves only): do baby-step-giant-step on the Jacobian using the fact that its order is in $[(\sqrt{q}-1)^g, (\sqrt{q}+1)^g]$.

Approaches to the zeta function problem

Generic approaches include:

- Direct counting: enumerate $X(\mathbb{F}_{q^n})$ for $n = 1, 2, \ldots$.
- Shanks's method (curves only): do baby-step-giant-step on the Jacobian using the fact that its order is in $[(\sqrt{q}-1)^g, (\sqrt{q}+1)^g]$.

In small characteristic (e.g., $q = 2^n$), additional techniques become available; the most flexible of these seems to be the use of p-adic cohomology. (Other: Satoh's canonical lift method for elliptic curves, Mestre's AGM method for ordinary curves of low genus.)

Approaches to the zeta function problem

Generic approaches include:

- Direct counting: enumerate $X(\mathbb{F}_{q^n})$ for $n = 1, 2, \ldots$.
- Shanks's method (curves only): do baby-step-giant-step on the Jacobian using the fact that its order is in $[(\sqrt{q}-1)^g, (\sqrt{q}+1)^g]$.

In small characteristic (e.g., $q = 2^n$), additional techniques become available; the most flexible of these seems to be the use of p-adic cohomology. (Other: Satoh's canonical lift method for elliptic curves, Mestre's AGM method for ordinary curves of low genus.)

Problem

What about Schoof's method (compute ζ_X modulo ℓ for many small primes ℓ)? It works even for p large, but depends badly on genus.

Approaches to the zeta function problem

Generic approaches include:

- Direct counting: enumerate $X(\mathbb{F}_{q^n})$ for $n = 1, 2, \ldots$.
- Shanks's method (curves only): do baby-step-giant-step on the Jacobian using the fact that its order is in $[(\sqrt{q}-1)^g, (\sqrt{q}+1)^g]$.

In small characteristic (e.g., $q = 2^n$), additional techniques become available; the most flexible of these seems to be the use of p-adic cohomology. (Other: Satoh's canonical lift method for elliptic curves, Mestre's AGM method for ordinary curves of low genus.)

Problem

What about Schoof's method (compute ζ_X modulo ℓ for many small primes ℓ)? It works even for p large, but depends badly on genus.

Problem

Is finding ζ_X for a curve of genus g over $\mathbb{F}_q$ polynomial time simultaneously in $g, \log(q)$? (Yes for quantum computation.)

Cohomology and zeta functions

One often studies ζ_X by constructing a *cohomology theory* associating to X some vector spaces $H^i(X)$ over some field K, each equipped with a linear transformation F such that

$$\#X(\mathbb{F}_{q^n}) = \sum_i (-1)^i \operatorname{Trace}(F^n, H^i(X)).$$

Then

$$\zeta_X(T) = \prod_i \det(1 - tF, H^i(X))^{(-1)^{i+1}}.$$

Cohomology and zeta functions

One often studies ζ_X by constructing a *cohomology theory* associating to X some vector spaces $H^i(X)$ over some field K, each equipped with a linear transformation F such that

$$\#X(\mathbb{F}_{q^n}) = \sum_i (-1)^i \operatorname{Trace}(F^n, H^i(X)).$$

Then

$$\zeta_X(T) = \prod_i \det(1 - tF, H^i(X))^{(-1)^{i+1}}.$$

The most famous of these is *étale (ℓ-adic) cohomology*, which takes coefficients in $\mathbb{Q}_\ell$ for a prime $\ell \neq p$; it is implicitly used in Schoof's algorithm (and Edixhoven's method for computing coefficients of modular forms). But it is only computationally effective in limited circumstances.

p-adic cohomology and zeta functions

We use *Monsky-Washnitzer (MW) cohomology*, a computationally effective cohomology theory producing vector spaces over the field $\mathbb{Q}_q$, the finite unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$.

p-adic cohomology and zeta functions

We use *Monsky-Washnitzer (MW) cohomology*, a computationally effective cohomology theory producing vector spaces over the field $\mathbb{Q}_q$, the finite unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$.

Note

Like the real numbers, one can only *approximately* specify p-adic numbers in a computation. In particular, one can only compute the action of F on a basis of $H^i(X)$ modulo a power of p, not exactly.

p-adic cohomology and zeta functions

We use *Monsky-Washnitzer (MW) cohomology*, a computationally effective cohomology theory producing vector spaces over the field $\mathbb{Q}_q$, the finite unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$.

Note

Like the real numbers, one can only *approximately* specify p-adic numbers in a computation. In particular, one can only compute the action of F on a basis of $H^i(X)$ modulo a power of p, not exactly.

To get around this, we compute the factors of ζ_X modulo some power of p, then combine an absolute bound on the size of coefficients.

p-adic cohomology and zeta functions

We use *Monsky-Washnitzer (MW) cohomology*, a computationally effective cohomology theory producing vector spaces over the field $\mathbb{Q}_q$, the finite unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$.

Note

Like the real numbers, one can only *approximately* specify p-adic numbers in a computation. In particular, one can only compute the action of F on a basis of $H^i(X)$ modulo a power of p, not exactly.

To get around this, we compute the factors of ζ_X modulo some power of p, then combine an absolute bound on the size of coefficients.

Note

Again as with $\mathbb{R}$, one must monitor p-adic precision and loss thereof. We'll ignore this here.

p-adic cohomology and zeta functions

Note

MW cohomology is only defined for *smooth affine* varieties.

For general X, we can take out a subvariety Y of lower dimension to get a smooth affine variety U, and

$$\zeta_X = \zeta_Y \zeta_U.$$

So we can use MW cohomology to find ζ_U, then deal with Y by induction on dimension.

p-adic cohomology and zeta functions

Note

MW cohomology is only defined for *smooth affine* varieties.

For general X, we can take out a subvariety Y of lower dimension to get a smooth affine variety U, and

$$\zeta_X = \zeta_Y \zeta_U.$$

So we can use MW cohomology to find ζ_U, then deal with Y by induction on dimension.

Example

If X is the hyperelliptic curve $y^2 = P(x)$ in $\mathbb{P}^2$, we could take Y to be the point(s) at infinity. (It will actually be convenient to take Y even larger.)

How to use p-adic cohomology: very rough outline

- Lift the smooth affine variety X from $\mathbb{F}_q$ to $\mathbb{Z}_q$. (Fine print: the lift should be the complement of a relative normal crossings divisor in a smooth proper scheme over $\mathbb{Z}_q$.)
- Lift the p-power Frobenius map on X. (Fine print: the lift is usually not algebraic, but should be p-adically *overconvergent*.)
- Write down the action of Frobenius on the algebraic de Rham cohomology of the lift of X. (First do the p-power Frobenius, then iterate intelligently to get the q-power Frobenius.)

How to use p-adic cohomology: very rough outline

- Lift the smooth affine variety X from $\mathbb{F}_q$ to $\mathbb{Z}_q$. (Fine print: the lift should be the complement of a relative normal crossings divisor in a smooth proper scheme over $\mathbb{Z}_q$.)
- Lift the p-power Frobenius map on X. (Fine print: the lift is usually not algebraic, but should be p-adically *overconvergent*.)
- Write down the action of Frobenius on the algebraic de Rham cohomology of the lift of X. (First do the p-power Frobenius, then iterate intelligently to get the q-power Frobenius.)

Problem

There are often natural pairings (cup product) in de Rham cohomology. Do they help? (May only affect constants.)

Example: hyperelliptic curves (imaginary, $p \neq 2$)

Let X be the hyperelliptic curve $y^2 = P(x)$, for P a monic polynomial of degree $2g+1$, minus the points $y \in \{0, \infty\}$; X is affine with coordinate ring

$$\mathbb{F}_q[x,y,z]/(y^2 - P(x), yz - 1).$$

(The complete curve has genus g.) Pick any monic lift $\tilde{P}$ of P, and lift Frobenius as follows:

$$x \mapsto x^p$$

$$y \mapsto y^p \left(1 + p\frac{\tilde{P}^\sigma(x^p) - \tilde{P}(x)^p}{py^{2p}}\right)^{1/2}$$

where σ means apply the canonical p-power Frobenius on $\mathbb{Q}_q$ term by term. This is *not algebraic*; the image of y is a p-adically (over)convergent series.

Example: hyperelliptic curves (imaginary, $p \neq 2$)

Let Ω^1 be the module (over an appropriate series ring R) generated by dx, dy modulo

$$2y\,dy - \tilde{P}'(x)\,dx.$$

Then $H^1(X)$ is the quotient of Ω^1 by the spans of df for all $f \in R$. It has basis

$$\frac{x^i\,dx}{y} \qquad (i = 0, \ldots, 2g-1), \qquad \frac{x^i\,dx}{y^2} \qquad (i = 0, \ldots, 2g).$$

Moreover, there is a nice algorithm to rewrite an element of Ω^1 as a linear combination of basis elements plus a df (by lowering the pole order at $y = 0$). So we can compute an approximation to the Frobenius action on $H^1(X)$ by applying a truncated Frobenius to basis elements.

Complexity estimates

One gets an algorithm to compute ζ_X in time $\tilde{O}(g^3 n^4)$ and space $\tilde{O}(g^3 n^3)$ (where $q = p^n$). Or rather, this is known if X is imaginary and $p \neq 2$.

Problem

Can one remove the restrictions in the previous statement? (For $p = 2$, apparently yes: Bernstein. For X real: presumably yes, but Harrison's work is unpublished.)

In practice, these methods work well; they are (mostly) implemented in MAGMA 2.12 (Harrison). In genus 1, they also appear in SAGE (Harvey) as part of the computation of p-adic global canonical heights of elliptic curves over $\mathbb{Q}$ (Mazur-Stein-Tate).

Nondegenerate curves (Castryck-Denef-Vercauteren)

A similar method can be used for many plane curves.

Definition

Consider the plane curve $P(x,y) = 0$, for $P(x,y) = \sum_{i,j \in \mathbb{Z}} c_{ij} x^i y^j$ a Laurent polynomial. The *Newton polygon* is the convex hull of

$$\{(i,j) \in \mathbb{Z}^2 : c_{ij} \neq 0\}.$$

Nondegenerate curves (Castryck-Denef-Vercauteren)

A similar method can be used for many plane curves.

Definition

Consider the plane curve $P(x,y) = 0$, for $P(x,y) = \sum_{i,j \in \mathbb{Z}} c_{ij} x^i y^j$ a Laurent polynomial. The *Newton polygon* is the convex hull of

$$\{(i,j) \in \mathbb{Z}^2 : c_{ij} \neq 0\}.$$

Definition

We say the curve $P(x,y) = 0$ is *nondegenerate* if it is smooth in $\mathbb{G}_m^2$, and for each segment σ in the Newton polygon, the (Laurent) polynomial

$$\sum_{(i,j) \in \sigma} c_{ij} x^i y^j$$

has no repeated nonmonomial factors.

An example

Example

The polynomial

$$x^4 + x^3 + ax^2y + x^2 + xy + x + y^2$$

defines a smooth curve in $\mathbb{G}_m^2$ for $27a^3 + 19a^2 - 85a - 149 \neq 0$. If $p \neq 3$, the curve is nondegenerate if also $a \neq 2$, as then

$$x^4 + x^3 + x^2 + x, \quad x + y^2, \quad y^2 + ax^2y + x^4$$

have no repeated nonmonomial factors. (Draw picture.)

Note

The genus of a nondegenerate curve equals the number of *interior* lattice points of the Newton polygon. (The above example has genus 1, because $(2,1)$ is the only interior lattice point.)

More on nondegenerate curves

Computing with the de Rham cohomology of nondegenerate curves is well-understood, from the theory of toric varieties.

Castryck, Denef, Vercauteren give an explicit algorithm for lifting Frobenius, where *both* x and y map to overconvergent series. This gives an algorithm for computing zeta functions of nondegenerate curves; it has good asymptotic behavior but bad constants.

More on nondegenerate curves

Computing with the de Rham cohomology of nondegenerate curves is well-understood, from the theory of toric varieties.

Castryck, Denef, Vercauteren give an explicit algorithm for lifting Frobenius, where *both* x and y map to overconvergent series. This gives an algorithm for computing zeta functions of nondegenerate curves; it has good asymptotic behavior but bad constants.

Problem

Does it help to throw out extra points and use a Frobenius lift with $x \mapsto x^p$ *(as in the hyperelliptic case)? One is forced to invert a resultant, which makes the cohomology more complicated.*

More on nondegenerate curves

Computing with the de Rham cohomology of nondegenerate curves is well-understood, from the theory of toric varieties.

Castryck, Denef, Vercauteren give an explicit algorithm for lifting Frobenius, where *both* x and y map to overconvergent series. This gives an algorithm for computing zeta functions of nondegenerate curves; it has good asymptotic behavior but bad constants.

Problem

Does it help to throw out extra points and use a Frobenius lift with $x \mapsto x^p$ (as in the hyperelliptic case)? One is forced to invert a resultant, which makes the cohomology more complicated.

Problem

Is there any hope for the higher-dimensional analogue? (Voight)

Deformation methods (Lauder, Gerkmann, Hubrechts)

In these methods, instead of computing the Frobenius action on the cohomology of a single X, one computes this action simultaneously on the members of a one-parameter family. This action satisfies a differential equation coming from the Gauss-Manin connection in cohomology; we use this differential equation to solve for the action given an initial condition.

Deformation methods (Lauder, Gerkmann, Hubrechts)

In these methods, instead of computing the Frobenius action on the cohomology of a single X, one computes this action simultaneously on the members of a one-parameter family. This action satisfies a differential equation coming from the Gauss-Manin connection in cohomology; we use this differential equation to solve for the action given an initial condition.

The initial condition is the Frobenius action on the cohomology of *another* member of the family; by being clever, one makes this much simpler. (E.g., the second member is defined over $\mathbb{F}_p$ instead of $\mathbb{F}_q$, or has symmetries commuting with the Frobenius action.)

An example of Dwork

For $p \neq 2$, consider the Legendre family of elliptic curves

$$y^2 = x(x-1)(x-\lambda).$$

On each fibre, H^1 is generated by dx/y and $x\,dx/y$. The Gauss-Manin connection acts by

$$D\left(\frac{dx}{y}\right) = \frac{-1}{2(\lambda-1)}\frac{dx}{y} + \frac{1}{2\lambda(\lambda-1)}\frac{x\,dx}{y}$$
$$D\left(x\frac{dx}{y}\right) = \frac{-1}{2(\lambda-1)}\frac{dx}{y} + \frac{1}{2(\lambda-1)}\frac{x\,dx}{y}.$$

For the Frobenius action, we must choose a Frobenius lift σ on the λ-line. Then

$$DF = \frac{d(\lambda^\sigma)}{d\lambda}FD.$$

Degeneration methods: pros and cons

One advantage of the deformation approach is that one avoids computing the Frobenius action on differential forms; this yields significant space savings if one takes advantage of the Fuchsian property of the differential equation. This is even more pronounced in higher dimensions.

Degeneration methods: pros and cons

One advantage of the deformation approach is that one avoids computing the Frobenius action on differential forms; this yields significant space savings if one takes advantage of the Fuchsian property of the differential equation. This is even more pronounced in higher dimensions.

One disadvantage is that deformation is somewhat more complicated to implement. However, it seems to work well in practice; for hyperelliptic curves, it is implemented in MAGMA 2.13 (Hubrechts).

Degeneration methods: pros and cons

One advantage of the deformation approach is that one avoids computing the Frobenius action on differential forms; this yields significant space savings if one takes advantage of the Fuchsian property of the differential equation. This is even more pronounced in higher dimensions.

One disadvantage is that deformation is somewhat more complicated to implement. However, it seems to work well in practice; for hyperelliptic curves, it is implemented in MAGMA 2.13 (Hubrechts).

Problem

Is there a good way to deform nondegenerate curves? That is, are there some nondegenerate curves with easy-to-compute Frobenius matrices?

Smooth surfaces in $\mathbb{P}^3$ (Abbott-Kedlaya-Roe)

Let X be the hypersurface $P(w,x,y,z) = 0$ in $\mathbb{P}^3$, where P is a homogeneous polynomial, and suppose X is smooth. Put $U = \mathbb{P}^3 - X$; then U is smooth affine, with coordinate ring the degree 0 part of

$$\mathbb{F}_q[w,x,y,z,P(w,x,y,z)^{-1}].$$

The de Rham cohomology of U is easy to compute (Griffiths).
Lift P to a homogeneous polynomial $\tilde{P}$. We lift Frobenius by

$$w \mapsto w^p, \ldots, z \mapsto z^p$$

$$\tilde{P}(w,x,y,z)^{-1} \mapsto \tilde{P}(w,x,y,z)^{-p} \left(1 + p\frac{\tilde{P}^\sigma(w^p,x^p,y^p,z^p) - \tilde{P}(w,x,y,z)^p}{p\tilde{P}(w,x,y,z)^p}\right)^{-1}.$$

Smooth surfaces in $\mathbb{P}^3$ (Abbott-Kedlaya-Roe)

This method is quite easy to implement. Unfortunately, because we went up by one dimension, it is asymptotically much slower than either directly computing cohomology on an affine piece of X, or doing a deformation.

Smooth surfaces in $\mathbb{P}^3$ (Abbott-Kedlaya-Roe)

This method is quite easy to implement. Unfortunately, because we went up by one dimension, it is asymptotically much slower than either directly computing cohomology on an affine piece of X, or doing a deformation.

Nonetheless, we have succeeded in calculating a few examples, e.g., surfaces of degree 4 over $\mathbb{F}_p$ with $p \leq 19$.

Problem

Work out the analogue for nondegenerate surfaces in toric threefolds. (de Jong has implemented the case of weighted projective spaces.)

The end

Any questions?