Recent results on p-adic computation of zeta functions

Kiran S. Kedlaya

Department of Mathematics, Massachusetts Institute of Technology

Computational Challenges Arising in Algorithmic Number Theory and Cryptography
Fields Institute (Toronto), October 30, 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; deformation methods of Lauder, Hubrechts.)

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; deformation methods of Lauder, Hubrechts.)

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; deformation methods of Lauder, Hubrechts.)

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:

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

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

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

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 $\tilde{O}(g^4 n^3)$ 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 $p = 2$, probably yes: partial answer by 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)

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

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

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

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.)

Cartier matrices in medium characteristic

The dependence on p in all the aforementioned methods is linear in p. (This is less clear for other p-adic methods, such as Mestre's AGM iteration.)

Cartier matrices in medium characteristic

The dependence on p in all the aforementioned methods is linear in p. (This is less clear for other p-adic methods, such as Mestre's AGM iteration.)

However, Bostan, Gaudry, Schost introduced a method for computing the Cartier matrix of a hyperelliptic curve, in which the time/space dependence on p is only $\tilde{O}(p^{1/2})$.

Cartier matrices in medium characteristic

The dependence on p in all the aforementioned methods is linear in p. (This is less clear for other p-adic methods, such as Mestre's AGM iteration.)

However, Bostan, Gaudry, Schost introduced a method for computing the Cartier matrix of a hyperelliptic curve, in which the time/space dependence on p is only $\tilde{O}(p^{1/2})$.

Using this method plus a few rounds of Schoof's algorithm and some baby-step-giant-step, they computed the Jacobian order for a random curve of genus 2 over $\mathbb{F}_{p^3}$ with $p = 2^{32} - 5$.

Accelerating recurrence relations

The crucial ingredient in Bostan-Gaudry-Schost is a method of Chudnovsky and Chudnovsky for accelerated computation of terms in certain recurrent sequences. E.g., one can compute $N! \pmod{p}$ in time $\tilde{O}(N^{1/2}\log(p))$.

This can be thought of as a form of baby-step-giant-step. Given a recurrence of the form

$$a_{n+1} = P(n)a_n$$

for a_n a column vector and $P(x)$ a matrix of polynomials, of which we want the N-th term for some initial condition a_0, the "baby steps" are to form the products $P(x+n-1)\cdots P(x)$ for $n = 1, 2, \ldots, 2^m$ where $2^m \cong \sqrt{N}$. Call the last of these $Q(x)$.

A typical "giant step" is to compute $Q(2^m - 1)\cdots Q(0)$ using a fast evaluation technique (making clever use of Lagrange interpolation).

Coming attractions: Harvey's method

The Cartier matrix is essentially the reduction modulo p of the Frobenius action on Monsky-Washnitzer cohomology.

Coming attractions: Harvey's method

The Cartier matrix is essentially the reduction modulo p of the Frobenius action on Monsky-Washnitzer cohomology.

David Harvey (preprint to be available soon) has proposed a technique for using the Chudnovsky method for computing the Frobenius action on a hyperelliptic curve modulo any power of p (for $p \geq 5$).

The key variation from my original reduction method is to leave $\tilde{P}(x)^p$ unexpanded, and instead represent forms as sums

$$\sum_{i,j,k,l} x^{pi+j} y^{pk+l} \, dx$$

with i, j, k, l running over short ranges. One then performs "horizontal" and "vertical" reductions, using the Chudnovsky method, to bring the powers of x and y into the desired range.

Coming attractions: Harvey's method

Harvey's preprint will only describe experiments in genus 1 over $\mathbb{F}_p$, as his intended application is computing p-adic canonical heights of elliptic curves, after Mazur-Stein-Tate.

Coming attractions: Harvey's method

Harvey's preprint will only describe experiments in genus 1 over $\mathbb{F}_p$, as his intended application is computing p-adic canonical heights of elliptic curves, after Mazur-Stein-Tate.

Problem

Can one use this technique to compute a Jacobian order of cryptographic size for a curve of low genus over $\mathbb{F}_{p^n}$ for very small g, n? (Maybe $g = 3, n = 1$ is in reach; $g = 2, n = 1$ may be too hard.)

The end

Any questions?