

Computing zeta functions via p -adic cohomology

Kiran S. Kedlaya (MIT)

kedlaya@mit.edu

<http://math.mit.edu/~kedlaya>

ANTS VI

University of Vermont

June 17, 2004

These slides are available online at the above web site.

Contents

- I. The zeta function problem
- II. (Algebraic) de Rham cohomology
- III. The example of hyperelliptic curves
- IV. de Rham cohomology and computing zeta functions
- V. The example of hyperelliptic curves revisited
- VI. Known and nearly known variants
- VII. What next? (Lauder's method)

I: The zeta function problem

\mathbb{F}_q a finite field of characteristic p
 X an algebraic variety over \mathbb{F}_q (e.g., an elliptic curve)
 $\#X(\mathbb{F}_{q^n})$ number of \mathbb{F}_{q^n} -rational points on X

The *zeta function* of X is the generating function

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

which turns out to be a quotient of polynomials with integer coefficients.

General problem: compute $Z(X, t)$ efficiently for any X of some particular form.

Methods

Many approaches exist to computing $Z(X, t)$:

- Brute force (and “baby-step-giant-step”)
- ℓ -adic methods (Schoof)
- Iterative p -adic methods: canonical lifts (Sato), AGM (Mestre)
- Cohomological p -adic methods (Lauder-Wan, K)
- Deformation-theoretic p -adic methods (Lauder)

I will focus on the cohomological today, with comments at the end about the last approach.

Pros and cons of cohomological approaches

Upsides:

- More efficient than ℓ -adic methods when both apply
- Combinable with ℓ -adic and brute force methods
- Polynomial dependence on complexity of X (e.g., genus of a curve)
- Uniform over characteristics
- Applicable in higher dimensions

General downside: limited to *small characteristic*. (Dependence on p at least $O(p)$.)

II: (Algebraic) de Rham cohomology

K	a field of characteristic zero
A	$K[x_1, \dots, x_n]/\mathfrak{a}$ for some ideal \mathfrak{a}
$\Omega_{A/K}^1$	the module of Kähler differentials (1-forms)
d	the universal K -linear derivation $A \rightarrow \Omega_{A/K}^1$
$\Omega_{A/K}^i$	$\wedge_A^i \Omega_{A/K}^1$ (module of i -forms)

Note: $\Omega_{A/K}^1$ is generated by dx_1, \dots, dx_n modulo relations

$$\frac{\partial a}{\partial x_1} dx_1 + \dots + \frac{\partial a}{\partial x_n} dx_n = 0 \quad (a \in \mathfrak{a})$$

and for $r \in A$ represented by $s \in K[x_1, \dots, x_n]$,

$$dr = \frac{\partial s}{\partial x_1} dx_1 + \dots + \frac{\partial s}{\partial x_n} dx_n.$$

The de Rham complex

The sequence of maps

$$A \xrightarrow{d} \Omega_{A/K}^1 \xrightarrow{d} \Omega_{A/K}^2 \cdots$$

is a *complex*, i.e., $d \circ d = 0$. The *algebraic de Rham cohomology* of A comprises the K -vector spaces

$$H^i(A) = \frac{\ker(d : \Omega_{A/K}^i \rightarrow \Omega_{A/K}^{i+1})}{\operatorname{im}(d : \Omega_{A/K}^{i-1} \rightarrow \Omega_{A/K}^i)}.$$

These are finite dimensional over K if A is smooth (Grothendieck).

Computational issues

One would like to be able to “compute” in these spaces. This would include being able to

- test elements of $\Omega_{A/K}^i$ for equality in $H^i(A)$;
- find elements of $\Omega_{A/K}^i$ giving a basis of $H^i(A)$;
- express elements of $\Omega_{A/K}^i$ in terms of a basis of $H^i(A)$.

Oaku, Takayama, Walther, et al. attack this problem using “noncommutative Gröbner basis” methods, but practicality seems uncertain. However, in certain special situations, things are much simpler.

When are things simpler?

Say you have

X $\text{Spec}(A)$, i.e., an affine K -variety

Y a smooth projective variety containing X such that

$Z = Y - X$ is a strict normal crossings divisor.

(I.e., each component of Z is smooth and they meet transversely.)

Then (Grothendieck again) the algebraic de Rham cohomology can be computed using just differential forms with *logarithmic singularities*; i.e., things like dz/z are okay, where z is a local parameter for a component of Z , but no worse.

So what?

Why does restricting to forms with logarithmic singularities help?

For one, the space of i -forms with log singularities is finite dimensional, and it can be written down explicitly. So finding a basis for H^i is easy.

For another, given a form with worse poles along Z , it is easy to write down forms which map to zero in H^i that can be subtracted off to lower the pole orders. We'll see this in the example.

III: The example of hyperelliptic curves

Let $P(x)$ be a monic polynomial over K of degree $2g + 1$ with no repeated roots, and write

$$A = K[x, y, z]/(y^2 - P(x), yz - 1).$$

Each element of A can be written uniquely as a finite sum $\sum_j A_j(x)y^j$, where $\deg(A_j) \leq 2g$. And $\Omega_{A/K}^1$ is generated by dx, dy, dz modulo the relations

$$2y dy = P'(x) dx, \quad y dz = -z dy$$

while $\Omega_{A/K}^2$ is zero.

A basis for $H^1(A)$, and relations

$X = \text{Spec}(A)$ is the complement of a set of $2g + 2$ points within a hyperelliptic curve of genus g .

We get a basis for $H^1(A)$:

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

One also obtains relations in cohomology: if $A(x) = B(x)P(x) + C(x)P'(x)$, then in $H^1(A)$ we have

$$\frac{A(x) dx}{y^s} \equiv \left(B(x) + \frac{2C'(x)}{s - 2} \right) \frac{dx}{y^{s-2}},$$

which lets us get rid of big negative powers of y . There is a similar relation (omitted) to eliminate positive powers.

Note: in practice, we split $H^1(A)$ into $+$ and $-$ eigenspaces for the hyperelliptic involution $y \mapsto -y$.

IV: de Rham cohomology and computing zeta functions

de Rham cohomology doesn't work in characteristic p because $dx^p = px^{p-1} dx = 0$ for any x . However, work of Dwork, Monsky, Washnitzer, Grothendieck, Berthelot, etc. relates de Rham cohomology in characteristic zero to zeta functions.

Notation convention:

\mathbb{Z}_q the unramified extension of \mathbb{Z}_p with residue field \mathbb{F}_q (a/k/a $W(\mathbb{F}_q)$)
 \mathbb{Q}_q the fraction field of \mathbb{Z}_q

Dagger algebras

Let $\mathbb{Z}_q\langle x_1, \dots, x_n \rangle^\dagger$ be the ring of formal power series

$$\sum_I c_I x^I = \sum_{i_1, \dots, i_n=0}^{\infty} c_{i_1, \dots, i_n} x_1^{i_1} \cdots x_n^{i_n} \quad (c_I \in \mathbb{Z}_q)$$

such that $v(c_I) + a(i_1 + \cdots + i_n) > b$ for some a, b with $a > 0$. (Note: a and b are not fixed; they may vary from series to series.)

Suppose $\overline{A} = \mathbb{F}_q[x_1, \dots, x_n]/\mathfrak{a}$ is a smooth affine \mathbb{F}_q -variety, and suppose there is an ideal \mathfrak{a}^\dagger of $\mathbb{Z}_q\langle x_1, \dots, x_n \rangle^\dagger$ such that

$$A^\dagger = \mathbb{Z}_q\langle x_1, \dots, x_n \rangle^\dagger / \mathfrak{a}^\dagger$$

is flat over \mathbb{Z}_q (i.e., torsion-free) and $A^\dagger \otimes_{\mathbb{Z}_q} \mathbb{F}_q \cong \overline{A}$. (A theorem of Elkik ensures you can do this.)

Dagger algebras (contd.)

We can set up something like de Rham cohomology using A^\dagger instead of \overline{A} .

Define $\Omega_{A^\dagger/\mathbb{Z}_q}^1$ to be the free A^\dagger -module generated by dx_1, \dots, dx_n modulo relations

$$\frac{\partial a}{\partial x_1} dx_1 + \dots + \frac{\partial a}{\partial x_n} dx_n = 0 \quad (a \in \mathfrak{a}^\dagger)$$

Then for $r \in A^\dagger$ represented by $s \in \mathbb{Z}_q\langle x_1, \dots, x_n \rangle^\dagger$, setting

$$dr = \frac{\partial s}{\partial x_1} dx_1 + \dots + \frac{\partial s}{\partial x_n} dx_n$$

gives a well-defined derivation $d : A^\dagger \rightarrow \Omega_{A^\dagger/\mathbb{Z}_q}^1$.

Monksy-Washnitzer cohomology

Again, we have a de Rham complex

$$A^\dagger \xrightarrow{d} \Omega_{A^\dagger/K}^1 \xrightarrow{d} \Omega_{A^\dagger/K}^2 \cdots$$

Define “the” Monksy-Washnitzer cohomology of \overline{A} to be the homology of the complex *after tensoring with* \mathbb{Q}_q :

$$H_{MW}^i(\overline{A}) = \frac{\ker(d : \Omega_{A^\dagger/\mathbb{Z}_q}^i \otimes \mathbb{Q}_q \rightarrow \Omega_{A^\dagger/\mathbb{Z}_q}^{i+1} \otimes \mathbb{Q}_q)}{\operatorname{im}(d : \Omega_{A^\dagger/\mathbb{Z}_q}^{i-1} \otimes \mathbb{Q}_q \rightarrow \Omega_{A^\dagger/\mathbb{Z}_q}^i \otimes \mathbb{Q}_q)}.$$

These are finite-dimensional vector spaces (Berthelot).

Surprise: this is actually functorial in \overline{A} , even if it doesn't look that way at first!

Monsky-Washnitzer cohomology and zeta functions

One can find a map $F : A^\dagger \rightarrow A^\dagger$ of the form

$$\sum_I c_I x^I \mapsto \sum_I c_I f(x_1)^{i_1} \cdots f(x_n)^{i_n}$$

which acts like the q -th power Frobenius modulo p . Make it act on differential forms using the rule

$$F(dr) = d(F(r)).$$

Then there is a trace formula: if $X = \text{Spec}(\bar{A})$ is d -dimensional, then

$$\#X(\mathbb{F}_{q^n}) = \sum_{i=0}^d (-1)^i \text{Trace}(q^{nd} F^{-n}, H_{MW}^i(\bar{A})).$$

So the zeta function of X can be recovered from a sufficiently good p -adic approximation to the matrix by which F acts on some basis of each H^i .

How to compute in MW-cohomology?

It is not completely obvious that one can “compute” in MW-cohomology in the sense of being able to find a basis and put elements into a standard form. However, in certain special situations... say you have

$$X = \text{Spec}(\overline{A})$$

Y smooth projective \mathbb{F}_q -variety containing X such that

$Z = Y - X$ is a strict normal crossings divisor.

Suppose *also* that you have a smooth projective \mathbb{Z}_q -scheme \tilde{Y} containing a relative strict normal crossings divisor \tilde{Z} , with complement \tilde{X} , such that $\tilde{X} \times_{\mathbb{Z}_q} \mathbb{F}_q = X$ and so on.

Then the MW-cohomology of X coincides with the algebraic de Rham cohomology of $\tilde{X} \times_{\mathbb{Z}_q} \mathbb{Q}_q$ (Berthelot, Shiho).

IV: The example of hyperelliptic curves revisited

Let $P(x)$ be a monic polynomial of degree $2g + 1$ over \mathbb{F}_q with no repeated roots, and lift it to a monic polynomial $\tilde{P}(x)$ over \mathbb{Z}_q . Put

$$\begin{aligned}\bar{A} &= \mathbb{F}_q[x, y, z]/(y^2 - P(x), yz - 1) \\ A^\dagger &= \mathbb{Z}_q\langle x, y, z \rangle^\dagger / (y^2 - \tilde{P}(x), yz - 1);\end{aligned}$$

each element of A^\dagger is uniquely an *infinite* sum $\sum_i A_j(x)y^i$ with $\deg(A_j) \leq 2g$.

By the previous slide, $H_{MW}^1(\bar{A})$ has basis

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

just like before, and the relations are all generated by the “algebraic” relations I wrote down (or omitted) before.

The action of Frobenius

One has a Frobenius F given by

$$\begin{aligned}x &\mapsto x^q \\y &\mapsto y^q \left(1 + \frac{P(x^q) - P(x)^q}{P(x)^q} \right)^{1/2} \\z &\mapsto z^q \left(1 + \frac{P(x^q) - P(x)^q}{P(x)^q} \right)^{-1/2} .\end{aligned}$$

Now the idea is to apply F to each of the $x^i dx/y$ and rewrite the results in terms of this basis; this yields a p -adic approximation to the true matrix by which Frobenius acts, but a good enough approximation yields the true characteristic polynomial.

(Historical note: for $g = 1$, this example was anticipated by Lubkin-Yoo, and by van der Put.)

Computing the action of Frobenius

More precisely, we approximate

$$F \left(\frac{x^i dx}{y} \right)$$

with a finite sum $\sum A_j(x)y^j$, where j runs over odd numbers from (large negative) to (small positive), and each $A_j(x)$ accurate to some p -adic precision. (Do this by approximating $F(y)$ with a Newton iteration.)

We then use the relation

$$\frac{A(x) dx}{y^s} \equiv \left(B(x) + \frac{2C'(x)}{s-2} \right) \frac{dx}{y^{s-2}}$$

in H^1 to “roll down” the negative powers of y .

Running time and space

Ignoring log factors, computing a zeta function this way on a genus g curve over \mathbb{F}_{p^n} takes time $O(pn^3g^4)$ and space $O(pn^3g^3)$, with reasonable constants.

What makes this so efficient is that *only* the application of Frobenius involves a power series computation; because my situation lifts nicely, the cohomology manipulations only involve short polynomials.

One lucky break that improves the constants: the “rolling down” process does involve some divisions by p , but fewer than you think. If

$$\frac{A_{-s}(x) dx}{y^s} \equiv \sum_{i=0}^{2g-1} c_i \frac{x^i dx}{y}$$

for $s > 0$ and A_{-s} has coefficients in \mathbb{Z}_q , then $p^{\lfloor s \rfloor} c_i \in \mathbb{Z}_q$. (Something like this happens whenever you’re in the liftable situation, but one has to work out the precise bound by hand.)

V: Known and nearly known variants

Since I worked out the aforementioned computation, several variants have been looked at, and some implementations have been made. See the proceedings for a sample of implementation results.

More hyperelliptic and superelliptic curves

The hyperelliptic curves we considered had a rational Weierstrass point (at infinity). The more general case in odd characteristic, where

$$y^2 = P(x)$$

for $P(x)$ a monic polynomial of degree $2g + 2$ with no repeated roots, can be handled similarly; this has been worked out by Michael Harrison and implemented in MAGMA 2.11.

In a similar vein, Gaudry and Gürel considered superelliptic curves

$$y^m = P(x)$$

where P is a polynomial with distinct roots and m is coprime to p .

More hyperelliptic and curves (even characteristic)

Hyperelliptic curves in characteristic 2 are Artin-Schreier covers of \mathbb{P}^1 , or better, can be written as

$$y^2 + h(x)y = f(x)$$

where $\deg(f) = 2g + 1$, $\deg(h) \leq g$, and the squarefree part of h divides f .

These were treated by Denef and Vercauteren; the geometry is a bit subtler and the analysis more delicate than in the odd case.

Some other Artin-Schreier covers were treated by Lauder and Wan (in different language).

$C_{a,b}$ -curves, etc.

For a, b coprime, a $C_{a,b}$ -curve is a curve given by a smooth affine equation of the form

$$y^a + \sum_{i=1}^{a-1} f_i(x)y^i + f_0(x)$$

where $\deg(f_0) = b$ and $a \deg(f_i) + bi < ab$. These are also treated by Denef-Vercauteren.

Denef-Vercauteren are in the process of working out methods for general curves...

VII: What next? (Lauder's method)

Basically no one has looked at varieties of dimension greater than 1. However, these do occur in certain applications:

- Coding theory: variants of Goppa construction (Voloch)
- Physics (!?): mirror symmetry over finite fields (Candelas-de la Ossa)

It would also be helpful to collect data on various theoretical questions, e.g., existence of subvarieties predicted by Frobenius eigenvalues (Tate conjecture).

Cyclic covers

One generalization of hyperelliptic curves is to cyclic covers of a projective space, e.g.,

$$z^2 = P(x, y).$$

We are looking at this currently with de Jong; it is pretty similar to the odd characteristic hyperelliptic case.

Projective and toric hypersurfaces

Let $P(x_0, \dots, x_n)$ be a polynomial defining a smooth projective hypersurface. One can compute the zeta function of that hypersurface easily by working on the *complement*, which is affine. Reduction formulas (over \mathbb{C}) are classical (Griffiths).

Can also consider smooth hypersurfaces/complete intersections in toric varieties. General framework set up by Gerkmann, but many details (e.g., precision bounds) lacking.

There is significant space overhead inherent in computing on high-dimensional varieties: this is remedied by Alan Lauder's "deformation method".

Lauder's deformation method (after Dwork)

Lauder: to compute the zeta function of, say, a *single* smooth projective hypersurface

$$X : P(x_0, \dots, x_n) = 0,$$

consider instead the one-parameter family

$$X_t : tP(x_0, \dots, x_n) + (1 - t)Q(x_0, \dots, x_n) = 0$$

where Q is highly symmetric. (E.g., if $d = \deg(P) \not\equiv 0 \pmod{p}$, take $Q = x_0^d + \dots + x_n^d$.)

Lauder's deformation method (contd.)

As in Dwork, the cohomologies of these form a vector bundle with (Gauss-Manin) connection with a Frobenius structure. I.e., there is a differential equation in t (with coefficients in $K(t)$ for some number field K) whose solutions compute zeta functions, and the action of Frobenius on Q serves as an “initial condition”.

This method has better dependence on $\dim X$ than straight cohomology computation (single exponential vs. double) and should be better in practice already for $\dim X = 2$. However, it seems complicated to implement.

There is potentially a variant for singular hypersurfaces, but it would be even more complicated.

The end