JOURNAL
de Théorie des Nombres
de BORDEAUX
anciennement Séminaire de Théorie des Nombres de Bordeaux

Alan G. B. Lauder

Rigid cohomology and p-adic point counting

Tome 17, n° 1 (2005), p. 169-180.

<http://jtnb.cedram.org/item?id=JTNB_2005__17_1_169_0>

Journal de Théorie des Nombres
de Bordeaux **17** (2005), 169–180

Rigid cohomology and p-adic point counting

par Alan G.B. LAUDER

Résumé. Je présente quelques algorithmes pour calculer la fonction zêta d'une variété algébrique sur un corps fini qui sont basés sur la cohomologie rigide. Deux méthodes distinctes sont élaborées à l'aide d'un exemple.

Abstract. I discuss some algorithms for computing the zeta function of an algebraic variety over a finite field which are based upon rigid cohomology. Two distinct approaches are illustrated with a worked example.

1. Introduction

I consider the problem of computing the zeta function of an algebraic variety defined over a finite field. This problem has been pushed into the limelight in recent years because of its importance in cryptography, at least in the case of curves. Wan's excellent survey article gives an overview of what has been achieved, and what remains to be done, on the topic [16]. The purpose of this expository article is to extract the essential content of previous results, and contrast this with some new developments in p-adic point counting. All of the p-adic algorithms I discuss rely upon rigid cohomology in some incarnation, and the alternative approaches pioneered by Mestre and Satoh are not touched upon. Moreover, little attention is paid to the precise running times of algorithms, the focus instead being on the qualitative nature of the complexities of algorithms. For this reason, much significant recent work using rigid cohomology, by Denef, Gaudry, Gerkmann, Gürel, Vercauteren and others, is not mentioned.

Let $\mathbb{F}_q$ be a finite field with q elements of characteristic p. Let X be an algebraic variety defined over $\mathbb{F}_q$. For each positive integer k, denote by N_k the number of $\mathbb{F}_{q^k}$-rational points on X. The zeta function $Z(X,T)$ of X is the formal power series

$$Z(X,T) = \exp\left(\sum_{k=1}^{\infty} N_k \frac{T^k}{k}\right).$$

The author is a Royal Society University Research Fellow. He wishes to thank Hendriks Hubrechts and Elmar Grosse-Klönne for their help.

By a famous theorem of Dwork, this power series is known to be a rational function $P(T)/Q(T)$, where $P(T), Q(T) \in 1 + T\mathbb{Z}[T]$ with $\gcd(P, Q) = 1$. Thus the zeta function of X can be finitely described, and a meaningful question to ask then is: Can one compute it? To address this question, let us for simplicity assume that X is an affine variety, defined by the common vanishing of a finite set of n-variate polynomials. The obvious approach is to determine N_k for $k = 1, 2, \ldots, 2D$ by evaluating all the defining polynomials at all points in affine n-space over $\mathbb{F}_{q^k}$. Here D is any upper bound on $\deg(P) + \deg(Q)$. This gives the first $2D + 1$ coefficients in the local expansion of the rational function around the origin, and it can then be recovered using linear algebra. This works provided one knows *a priori* an upper bound D. Fortunately, Bombieri has proved such a bound in terms of n, the number of defining polynomials, and their degrees [2]. Thus the answer to our first question for affine varieties is "Yes" [16, Corollary 2.7]. More general varieties can be decomposed into affine pieces, and the same approach then applies.

A much more interesting question to ask is: Is there an efficient algorithm for computing $Z(X, T)$, i.e., an algorithm whose running time is bounded by a polynomial function in the input size? For example, let us consider the case of a projective hypersurface defined by a homogeneous polynomial f of degree $d \geq 2$ in $n \geq 2$ variables over $\mathbb{F}_q$. A workable measure of the "size" of this polynomial is $d^{n-1} \lg(q)$ bits, where $\lg(q) := (\lfloor \log_2(q) \rfloor + 1)$. We then require an algorithm whose running time is $(d^{n-1} \lg(q))^{\mathcal{O}(1)}$ bit operations, or put more simply, $(d^n \lg(q))^{\mathcal{O}(1)}$ bit operations. It is not difficult to check that the running time of our naïve algorithm for computing the zeta function is actually exponential in $d^n \lg(q)$. Thus certainly a more sophisticated approach is required, as discussed in the next section.

2. Cohomological formulae and previous results

Cohomological formulae allow one to express the zeta function as an alternating product of characteristic polynomials of maps on certain finite dimensional spaces. Specifically one has

$$Z(X, T) = \prod_{i=0}^{2\dim(X)} \det(I - T\mathrm{Frob}_q | H^i(X))^{(-1)^{i-1}}.$$

Here Frob_q is the geometric Frobenius acting on cohomology. The cohomology space $H^*(X)$ depends upon which cohomology theory one uses. For example, one might take $H^i(X) = H^i_{et,c}(X, \mathbb{Q}_\ell)$, ℓ-adic étale cohomology with compact support where $\ell \neq p$. In this situation, if one can compute the reduction modulo ℓ of the characteristic polynomials for enough "small primes" ℓ, the Chinese Remainder Theorem can be used to recover the zeta function. Unfortunately, the methods used by Grothendieck to prove his

ℓ-adic formula do not seem suited to algorithmic applications. However, for curves and abelian varieties, the more constructive approach of Weil does lead to algorithms [13, 14]:

Theorem 2.1 (Schoof-Pila)**.** *The zeta function of a smooth projective curve birational to a plane curve of degree d over $\mathbb{F}_q$ can be computed deterministically in $\lg(q)^{C_d}$ bit operations, for some exponent C_d depending on d.*

In the original work of Schoof and Pila the exponent C_d depends at least exponentially on d. Huang and Ierardi have a different but related algorithm for which $C_d = d^{\mathcal{O}(1)}$, but this algorithm requires some randomisation [7]. For elliptic curves, significant improvements have been found to these approaches, most notably through the work of Elkies and Atkin [6].

A more fruitful approach when the characteristic p is "small" is to use a p-adic formula. For example, one can take $H^i(X) = H_{rig,c}(X, \mathbb{Q}_q)$, rigid cohomology with compact support, where $\mathbb{Q}_q$ is the unramified extension of $\mathbb{Q}_p$ of degree $\log_p(q)$. Rigid cohomology has an explicit description in terms of de Rham complexes. This allows one to compute the required matrices. Specifically, first one observes that $\mathrm{Frob}_q = \mathrm{Frob}_p^{\log_p(q)}$, where Frob_p is the absolute Frobenius map. The matrix of Frob_p can be computed modulo some suitably large power of p by lifting to the de Rham complex, where its action is given very explicitly, and performing some kind of cohomological reduction. This approach was first explored by Kedlaya, and seems very useful for curves [8]. An even simpler approach is to use a trace formula which is defined on the de Rham complex itself, as then one can avoid cohomology altogether. For example the Dwork Trace Formula, which describes the zeta function in terms of the Fredholm determinant of a certain completely continuous operator on an infinite dimensional p-adic Banach space [4]. This formula was used to obtain the following result [12, Theorem 1].

Theorem 2.2. *The zeta function of an affine hypersurface defined by a polynomial of degree d in n variables over $\mathbb{F}_q$ can be computed deterministically in $(pd^n \lg(q))^{\mathcal{O}(n)}$ bit operations. Here p is the characteristic of the field $\mathbb{F}_q$.*

Working on the de Rham complex allows one to circumvent any problems caused by singularities. For non-singular hypersurfaces the more practical approach described by Kedlaya is better, but the complexity remains $(pd^n \log(q))^{\mathcal{O}(n)}$ bit operations. The reason for this is that in both algorithms one computes the Frobenius action on the de Rham complex. In Kedlaya's algorithm one then reduces back into cohomology. The elements one computes on the de Rham complex are truncated power series of degree $\mathcal{O}(pd^n \log(q))$ in n variables. Such power series take up $(pd^n \log(q))^{\mathcal{O}(n)}$ bits of space, and this dominates the complexity of both approaches. To make further progress with p-adic cohomology one needs a method of finding the

absolute Frobenius matrix which entirely avoids computations on the de Rham complex itself, and works solely on the homology of the complex. Such a method is introduced in the next section.

3. The deformation algorithm

Recently I have proved the following theorem [11, Theorem 1 and Note 21].

Theorem 3.1. *The zeta function of a smooth projective hypersurface defined by a homogeneous polynomial of degree d in n variables over $\mathbb{F}_q$ can be computed deterministically in $(pd^n \lg(q))^{\mathcal{O}(1)}$ bit operations, provided $p \neq 2$ and p does not divide d. Here p is the characteristic of the field $\mathbb{F}_q$.*

The improvement is that the exponent no longer depends upon n. In fact, the exponent is rather small: for example, the dependence on $\lg(q)$ is essentially third power regardless of the dimension. Although an undesired factor $p^{\mathcal{O}(1)}$ still occurs, it is of interest to note that the algorithm does give a non-trivial result even for prime fields [11, Theorem 2 and Note 21].

Theorem 3.2. *Let $f \in \mathbb{Z}[X_1, \ldots, X_n]$ be homogeneous of degree d and assume that the projective hypersurface defined by the equation $f = 0$ is smooth. For any $\varepsilon > 0$, there exists an explicit deterministic algorithm which takes as input a prime p, outputs the number of solutions to the equation $f = 0 \bmod p$, and requires $\mathcal{O}(p^{2+\varepsilon})$ bit operations.*

Note that the hidden constant in the big-Oh notation depends now upon f, whereas in all previous results it was an absolute constant. A naïve approach to this problem would require $\mathcal{O}(p^{n-2+\varepsilon})$ bit operations.

The theorems are proved in a very indirect manner which is inspired by the beautiful paper of Dwork [5]. Here is a sketch of the method: The hypersurface is embedded in a one-dimensional family over a subset of the affine line whose fibre at the origin is a smooth diagonal hypersurface. Specifically, one defines $f(\Gamma) = \sum_{i=1}^{n} X_i^d + \Gamma h(X_1, \ldots, X_n)$, so that $f(1)$ is the original polynomial which defines the smooth projective hypersurface over $\mathbb{F}_q$ and $f(0)$ is a diagonal form. The *relative rigid cohomology* of this family has the structure of an *overconvergent F-isocrystal with connection*. (I actually work with an older version of this cohomology theory due to Dwork.) Concretely, this just means that one has a relative Frobenius map $\mathrm{Frob}_p(\Gamma)$ and a linear differential operator $\nabla(\Gamma)$ acting on the middle-dimensional piece of cohomology, and they commute in an appropriate sense. This commutativity gives a local factorisation of the Frobenius operator around the origin in terms of its value at the origin and the local solution matrix of the differential operator around the origin. The differential operator can

be constructed and solved locally within the required time. Similarly, the Frobenius matrix at the origin can be given explicitly, since it is the Frobenius matrix of a diagonal hypersurface. Finally, one needs to recover a matrix for $\mathrm{Frob}_p(1)$, which is the Frobenius matrix of the initial hypersurface, from the local expansion of $\mathrm{Frob}_p(\Gamma)$ around the origin. Unfortunately this expansion will not in general converge on the closed p-adic unit disk, because of singular fibres in the family. However, one can calculate bounds on the domain of holomorphy of the entries in a matrix for $\mathrm{Frob}_p(\Gamma)$, as p-adic holomorphic functions in the sense of Krasner. These bounds allow one to compute a matrix for $\mathrm{Frob}_p(1)$ in an indirect manner from the local expansion. This last step can be thought of as some kind of "p-adic analytic continuation".

The technique of deforming one polynomial into another is reminiscent of the "homotopy methods" used in numerical analysis [3, Section 4.2]. It seems quite remarkable that such methods also lead to powerful algorithms for polynomials over finite fields! The approach should extend to quite general smooth varieties, the key difficulty being that one needs explicit estimates on the domain of holomorphy of the relative Frobenius matrix $\mathrm{Frob}_p(\Gamma)$. Rigid cohomology itself just tells one that the matrix is overconvergent.

I like to call the approach used above the *deformation algorithm*. To my knowledge, the technique was first introduced in [10], and applied in that paper to the easier case of L-functions of certain additive character sums. Nobuo Tsuzuki from Hiroshima University has also independently been exploring quite similar ideas for computing L-functions of some one-dimensional Kloosterman sums. His results are presented in [15], along with details of a computer implementation.

4. Some details for hyperelliptic curves

The purpose of this section is to give the reader an idea of the theoretical constructions underlying the deformation algorithm. Rather than working with Dwork's version of rigid cohomology for smooth projective hypersurfaces, I believe it is more helpful to the reader if I discuss how one can apply similar ideas to certain smooth affine curves. For in the case of affine curves, one can work in the setting of Monsky-Washnitzer cohomology, an older special case of rigid cohomology. This cohomology theory is a p-adic analytic version of algebraic de Rham cohomology, and the algebraic structures needed then turn out to be rather pretty "commutative squares and cubes". I have not worked out any of the essential p-adic analytic details in this setting. This is rather hard work, and I completely avoid this aspect in the discussion below.

Consider the case of a smooth affine curve $\bar{X}$ defined by the system of equations

$$\begin{aligned} Y^2 &= \bar{Q}(X) \\ Y &\neq 0 \end{aligned}$$

over a finite field $\mathbb{F}_q$ of odd characteristic p. Here $\bar{Q}$ is a polynomial over $\mathbb{F}_q$ of degree $2g+1$ with distinct roots. This is exactly the situation discussed in detail in the paper of Kedlaya [8]. Let $\bar{A} := \mathbb{F}_q[X, Y, Y^{-1}]/(Y^2 - \bar{Q}(X))$ be the coordinate ring of $\bar{X}$. Let $\mathbb{Q}_q$ denote the unramified extension of the p-adic field $\mathbb{Q}_p$ of degree $\log_p(q)$, and $\mathbb{Z}_q$ the ring of integers of $\mathbb{Q}_q$. Choose a p-adic lift $Q \in \mathbb{Z}_q[X]$ of degree $2g+1$ of the polynomial $\bar{Q}$. Define

$$A^\dagger := \left\{ \sum_{m=-\infty}^{\infty} \sum_{i=0}^{2g} \frac{a_{m,i} X^i}{\sqrt{Q}^m} \mid \liminf(\mathrm{ord}(a_{m,i})/|m|) > 0 \right\}.$$

This ring is the *weak completion* of a p-adic lift of $\bar{A}$, see [9] for a more detailed discussion of such rings. The module of continuous $\mathbb{Q}_q$-linear differentials $\Omega(A^\dagger)$ can be identified with the set of elements of the form $*dX$ for $* \in A^\dagger$. We shall just write $A^\dagger dX$ for $\Omega(A^\dagger)$. The universal derivation $d : A^\dagger \to A^\dagger dX$ maps a series r to $\frac{dr}{dX} dX$. We need to lift the pth power Frobenius ring monomorphism Frob_p from $\bar{A}$ to $A^\dagger$. We can do this by first defining $\mathrm{Frob}_p(X) := X^p$, and $\mathrm{Frob}_p(c) = c^\sigma$ for $c \in \mathbb{Q}_q$ where σ is the automorphism of $\mathbb{Q}_q$ lifting the pth power Frobenius automorphism on $\mathbb{F}_q$. Now Frob_p can be defined by continuity on elements in $A^\dagger$ provided we can work out where it sends $\sqrt{Q}$. We must have that $\mathrm{Frob}_p(\sqrt{Q})^2 = \mathrm{Frob}_p(Q) = Q^\sigma(X^p)$. Defining

$$\mathrm{Frob}_p(\sqrt{Q}) := Q^{p/2} \left(1 - \frac{Q^p - Q^\sigma(X^p)}{Q^p}\right)^{1/2} \tag{4.1}$$

does the trick. The righthand-side squares to $Q^\sigma(X^p)$ and since $p|(Q(X)^p - Q^\sigma(X^p))$ it can be expanded as a series in $A^\dagger$. It is precisely the problem of defining $\mathrm{Frob}_p(\sqrt{Q})$ which forced us to take some larger "completion" of a p-adic lift of $\bar{A}$.

The next diagram commutes:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A^\dagger & & \xrightarrow{\frac{d\cdot}{dX} dX} & & A^\dagger dX & \longrightarrow & 0 \\ & & \downarrow \mathrm{Frob}_p & & & & \downarrow \mathrm{Frob}_p & & \\ 0 & \longrightarrow & A^\dagger & & \xrightarrow{\frac{d\cdot}{dX} dX} & & A^\dagger dX & \longrightarrow & 0. \end{array} \tag{4.2}$$

Here we define $\mathrm{Frob}_p(dX) := d(X^p) = pX^{p-1}dX$. Kedlaya showed that to compute the zeta function of $\bar{X}$ it is enough to find a matrix for the action of Frob_p on the cokernel of the map $\frac{d\cdot}{dX} dX$, i.e., on the space $H^1_{MW}(\bar{X}) := A^\dagger dX / \frac{dA^\dagger}{dX} dX$. (Note that Frob_p is a σ-linear map on this space, rather than

a linear map.) This space has finite dimension $4g+1$ over $\mathbb{Q}_q$. It splits into positive and negative eigenspaces under the map induced by $\sqrt{Q} \mapsto -\sqrt{Q}$ of dimension $2g+1$ and $2g$ respectively. Each eigenspace is stable under the map induced by Frob_p, and Kedlaya showed that one need only consider the action on the negative eigenspace. However, to save introducing too much notation, we shall just ignore this observation.

A basis for $H^1_{MW}(\bar{X})$ is given by the forms

$$\left\{ \frac{X^i dX}{\sqrt{Q}^j} \mid j = 1 \text{ and } 0 \leq i < 2g,\ j = 2 \text{ and } 0 \leq i \leq 2g \right\}. \tag{4.3}$$

We reduce elements of $A^\dagger dX$ to linear combinations of these basis elements modulo $\frac{dA^\dagger}{dX} dX$ as follows. For $B \in \mathbb{Q}_q[X]$ since $\gcd(Q, Q') = 1$, where $Q' = \frac{dQ}{dX}$, we can write $B = RQ + SQ\prime$ for some polynomials R and S whose degrees may be explicitly bounded. For $m \geq 1$

$$d\left(\frac{S(X)}{Q(X)^{m/2}}\right) = \frac{S' dX}{Q^{m/2}} - \frac{mSQ' dX}{2Q^{m/2+1}}.$$

Hence in homology:

$$\begin{aligned} \frac{B dX}{Q^{m/2+1}} &= \frac{(RQ + SQ')dX}{Q^{m/2+1}} \\ &\equiv \frac{R dX}{Q^{m/2}} + \frac{2S' dX}{mQ^{m/2}}. \end{aligned} \tag{4.4}$$

This reduces all "rational forms" to the shape $*dX/Q^{j/2}$, for $j = 1, 2$ and $* \in \mathbb{Q}_q[X]$. Reduction of $*$ to a polynomial of the appropriate degree is easier: A form $*dX/\sqrt{Q}$ with $*$ of degree $m \geq 2g$ can be reduced in degree by subtracting an appropriate constant multiple of $d(X^{m-2g}\sqrt{Q})$; a form $*dX/Q$ with $*$ of degree $m > 2g$ can be reduced in degree by subtracting an appropriate constant multiple of $d(X^{m-2g})$. Forms in $A^\dagger$ are p-adic limits of rational forms, and since derivation is continuous we can reduce elements in $A^\dagger$ to the limits of reduced rational forms. It is precisely the "weak completion" condition that ensures these limits actually exist.

To compute the Frobenius action, following Kedlaya, one computes it explicitly on the basis of $H^1_{MW}(\bar{X})$, giving a series in $A^\dagger dX$, and reduces this series back to a linear combination of the basis elements. Of course, all this is done to some required p-adic accuracy. This approach is excellent for curves; however, for higher dimensional varieties the explicit computation of the Frobenius map on the "de Rham complex" impacts on the complexity. I will now sketch how the deformation algorithm gets around this problem. (Admittedly, my sketch is in the case of curves, where the problem is not so significant anyway.)

Let Γ be a new parameter, and let $Q(X,\Gamma) \in \mathbb{Z}_q[X,\Gamma]$ be such that $Q(X,1)$ is our old polynomial Q. Assume $Q(X,\Gamma)$ is monic in X of degree $2g+1$. Define

$$r(\Gamma) := \mathrm{Res}(Q, \frac{\partial Q}{\partial X}, X) \in \mathbb{Z}_q[\Gamma],$$

the resultant with respect to X of Q and $\frac{\partial Q}{\partial X}$. Let $\mathbb{C}_p$ be the completion of an algebraic closure of $\mathbb{Q}_q$, and $\bar{\mathbb{F}}_q$ be the residue class field of $\mathbb{C}_p$. For $\bar{\gamma} \in \bar{\mathbb{F}}_q$ let $\gamma \in \mathbb{C}_p$ denote the Teichmüller lift of $\bar{\gamma}$. For each $\bar{\gamma} \in \bar{\mathbb{F}}_q$, let $\bar{X}_{\bar{\gamma}}$ be the affine curve over $\mathbb{F}_q(\bar{\gamma})$ defined by the equations $Y^2 = Q(X,\gamma) \bmod p$, $Y \neq 0 \bmod p$. For $\bar{\gamma} \in \bar{\mathbb{F}}_q$ with $r(\gamma) \neq 0 \bmod p$, the curve $\bar{X}_{\bar{\gamma}}$ is smooth. Thus we have a family of smooth affine curves over the line $\{\bar{\gamma} \in \bar{\mathbb{F}}_q \,|\, r(\gamma) \neq 0 \bmod p\}$. Let us denote the family by $\bar{\mathbf{X}}$ and the base space by $\bar{\mathbf{S}}$. Assume that $Q(X,0) \bmod p$ is square-free of degree $2g+1$, so that $r(0) \neq 0 \bmod p$.

We wish to compute the zeta function of the smooth fibre $\bar{X}_{\bar{1}}$. We assume that the Frobenius matrix of $\bar{X}_{\bar{0}}$ is already known, that is, the matrix for the action of Frob_p on $H^1_{MW}(\bar{X}_{\bar{0}})$. For example, as in the case in [11], it may be that there is an explicit formula for the entries in the Frobenius matrix which is easily computed. Or alternatively, we may have computed the Frobenius matrix for this fibre using Kedlaya's algorithm and now want to find that of other fibres. (There is actually some sense in this, since for families of curves defined over prime fields the space complexity of the deformation algorithm is quadratic in $\log(q)$, rather than cubic as in the case of Kedlaya's algorithm. Note that for general curves the deformation algorithm does not appear to improve upon Kedlaya's approach, the clear benefit being for higher dimensional varieties.)

We need relative versions of the rings and modules we worked with above, i.e., modules of continuous relative differentials etc. Care must be taken in defining these algebraic structures, to ensure that finite dimensionality of quotients is retained. Define S to be the weak completion of $\mathbb{Q}_q[\Gamma, 1/r(\Gamma)]$, and T to be the weak completion of $\mathbb{Q}_q[X,Y,Y^{-1},\Gamma,1/r(\Gamma)]/(Y^2 - Q(X,\Gamma))$. Define $\mathrm{Frob}_p(\Gamma) := \Gamma^p$ and $\mathrm{Frob}_p(d\Gamma) := p\Gamma^{p-1}d\Gamma$ and let Frob_p act on all other symbols we encounter exactly as before. Note though that now Frob_p acts on $\sqrt{Q(X,\Gamma)}$ via the formula (4.1), but with "$Q^\sigma(X^p)$" replaced by "$Q^\sigma(X^p,\Gamma^p)$".

The next diagram commutes:

$$\begin{array}{ccccccc} 0 & \longrightarrow & T & \stackrel{\frac{\partial \cdot}{\partial X}dX}{\longrightarrow} & TdX & \longrightarrow & 0 \\ & & \downarrow \mathrm{Frob}_p & & \downarrow \mathrm{Frob}_p & & \\ 0 & \longrightarrow & T & \stackrel{\frac{\partial \cdot}{\partial X}dX}{\longrightarrow} & TdX & \longrightarrow & 0. \end{array} \tag{4.5}$$

Here TdX is just the "module of continuous relative differentials $\Omega(T/S)$".

As before, we need to understand the map induced on the cokernel $TdX/\frac{\partial T}{\partial X}dX$ by Frob_p. Call this cokernel $H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})$. I claim, but do not prove, that this is a free S-module with basis the set (4.3). To see why this should be true, observe, using resultants, that for any $B(X,\Gamma) \in \mathbb{Z}_q[X,\Gamma]$ we can find $R, S \in \mathbb{Z}_q[X,\Gamma]$ such that $r(\Gamma)B = RQ + S\frac{\partial Q}{\partial X}$. So formulae similar to (4.4) reduce rational forms to $\mathbb{Q}_q[\Gamma, 1/r(\Gamma)]$-linear combinations of the basis set — the key difference is that a factor $r(\Gamma)$ is introduced on the denominator on each reduction step. To ensure this process converges in the p-adic limit, one must define the rings S and T with great care in the first place. I believe that my definitions for S and T should ensure this. In any case, let us proceed under the assumption that my claim is true.

Let $(\mathrm{Frob}_p(\Gamma))$ be the matrix with respect to the basis (4.3) for the map which Frob_p induces on the free S-module $H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})$. (Note that it is not a linear map, but is additive with $\mathrm{Frob}_p(c(\Gamma)m) = c^\sigma(\Gamma^p)\mathrm{Frob}_p(m)$ for $c(\Gamma) \in S$ and $m \in H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})$.) This matrix contains entries which are elements in S, that is of the form

$$\sum_{i=0}^{\infty} a_i \Gamma^i + \sum_{j=1}^{\infty} \frac{b_j(\Gamma)}{r(\Gamma)^j},\ \deg_\Gamma(b_j) < \deg_\Gamma(r)$$

with linear decay conditions on $\mathrm{ord}(a_i)$ and $\mathrm{ord}(b_j(\Gamma))$ as $i, j \to \infty$. Moreover, by our construction, $(\mathrm{Frob}_p(\gamma))$ equals the Frobenius matrix for the fibre $\bar{X}_{\bar{\gamma}}$ for any Teichmüller point γ, for in those cases $\mathrm{Frob}_p(\gamma) = \gamma^p$. Our aim in the deformation algorithm is to compute this matrix $(\mathrm{Frob}_p(\Gamma))$, for then we may recover the Frobenius matrix of any fibre by specialisation. We compute $(\mathrm{Frob}_p(\Gamma))$ from $(\mathrm{Frob}_p(0))$ and the fact that it satisfies a differential equation, as we now describe.

We have another commutative diagram:

$$\begin{array}{ccccccc} 0 & \longrightarrow & T & \stackrel{\frac{\partial \cdot}{\partial X}dX}{\longrightarrow} & TdX & \longrightarrow & 0 \\ & & \downarrow \frac{\partial \cdot}{\partial \Gamma}d\Gamma & & \downarrow \frac{\partial \cdot}{\partial \Gamma}d\Gamma & & \\ 0 & \longrightarrow & Td\Gamma & \stackrel{\frac{\partial \cdot}{\partial X}dX}{\longrightarrow} & TdXd\Gamma & \longrightarrow & 0. \end{array} \tag{4.6}$$

It commutes by the commutativity of partial differentiation. Along with the previous diagram (4.5), diagram (4.6) fits into a "commutative cube". Specifically, the front face of cube is diagram (4.5), and the back face diagram (4.5) with the symbol $d\Gamma$ appended to all modules. The top face is diagram (4.6) and the bottom face also diagram (4.6). The left and right faces of the cube are also commutative diagrams, since the Frobenius map commutes with partial differentiation by Γ.

Descending to homology on the righthand face of the cube we get another commutative diagram:

$$\begin{array}{ccccccc} 0 & \longrightarrow & H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}}) & \stackrel{\frac{\partial \cdot}{\partial \Gamma} d\Gamma}{\longrightarrow} & H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})d\Gamma & \longrightarrow & 0 \\ & & \downarrow \mathrm{Frob}_p & & \downarrow \mathrm{Frob}_p & & \\ 0 & \longrightarrow & H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}}) & \stackrel{\frac{\partial \cdot}{\partial \Gamma} d\Gamma}{\longrightarrow} & H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})d\Gamma & \longrightarrow & 0. \end{array} \tag{4.7}$$

Let ∇ denote the map $H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}}) \stackrel{\frac{\partial \cdot}{\partial \Gamma} d\Gamma}{\longrightarrow} H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})d\Gamma$. The map ∇ is called the *connection* and the pair $(\mathrm{Frob}_p, \nabla)$ the *overconvergent F-isocrystal* defined by taking the "relative Monsky-Washnitzer cohomology" of the family in the middle dimension. The map ∇ is additive and satisfies the Leibniz rule, i.e., $\nabla(cm) = \frac{dc}{d\Gamma} m d\Gamma + c\nabla(m)$ for $c \in S$ and $m \in H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})$.

The action of ∇ on the basis (4.3) of $H^1_{MW}(\bar{\mathbf{X}}/\bar{\mathbf{S}})$ can be computed explicitly by differentiating the basis elements with respect to Γ and using the reduction formulae. This gives a matrix, $B(\Gamma)$ say, for the differential operator ∇. The commutativity $\nabla \circ \mathrm{Frob}_p = \mathrm{Frob}_p \circ \nabla$ of the connection and Frobenius, as in diagram (4.7), yields the differential equation

$$\frac{d(\mathrm{Frob}_p(\Gamma))}{d\Gamma} + B(\Gamma)(\mathrm{Frob}_p(\Gamma)) = (\mathrm{Frob}_p(\Gamma))B^\sigma(\Gamma^p)p\Gamma^{p-1}.$$

To compute $(\mathrm{Frob}_p(\Gamma))$ one could solve this locally around the non-singular point $\Gamma = 0$, using the foregiven knowledge of $(\mathrm{Frob}_p(0))$ as the initial condition, and then recover the matrix $(\mathrm{Frob}_p(\Gamma))$ globally from its local expansion. The latter can be done rather easily, although one must have *a priori* bounds on the decay of the entries in the matrix $(\mathrm{Frob}_p(\Gamma))$, see [11, Section 8].

An alternative approach is to compute a basis of local solutions to the differential equation $\nabla = 0$ around the origin. Specifically, solve the differential equation

$$\frac{dC}{d\Gamma} = -B(\Gamma)C(\Gamma),\ C(0) = I$$

where the matrix $C(\Gamma)$ has entries in $\mathbb{Q}_q[[\Gamma]]$. Diagram (4.7) shows that Frob_p is stable on the basis of local solutions, which leads to the equation

$$(\mathrm{Frob}_p(\Gamma))C^\sigma(\Gamma^p) = C(\Gamma)(\mathrm{Frob}_p(0)).$$

Thus we get a local factorisation of the Frobenius matrix

$$(\mathrm{Frob}_p(\Gamma)) = C(\Gamma)(\mathrm{Frob}_p(0))(C^\sigma(\Gamma^p))^{-1}$$

around the origin. Once again, one can recover the Frobenius matrix globally from its local expansion. This was actually the approach taken in [10, 11]. In [10] the local expansion actually converged on the closed p-adic unit disk, which was a helpful simplification.

Whether one uses the first approach or the second, the essential point is that the matrix $B(\Gamma)$ can be computed easily. In fact, if we take a "lifting" of our family to the complex numbers, it is just the classical Picard-Fuchs matrix for the family of complex curves. All subsequent computations in both approaches involve rational functions in the single parameter Γ, and nowhere does one have to compute the Frobenius map on the de Rham complex. (Note that elements in S reduce to rational functions when one works to a finite p-adic precision.) This is the reason that the complexity of the deformation algorithm does not increase with the dimension of the variety.

This completes my sketch of the deformation algorithm for the case of hyperelliptic curves in odd characteristic. I hope the expert in p-adic cohomology has found the sketch useful in identifying precisely how I exploit the theory, and the non-expert found it a readable introduction to a very beautiful part of number theory.

References

[1] P. Bachmann, *Zur Theory von Jacobi's Kettenbruch-Algorithmen*. J. Reine Angew. Math. **75** (1873), 25–34.

[2] E. Bombieri, *On exponential sums in finite fields II*. Invent. Math. **47** (1978), 29–39.

[3] J-P. Dedieu, *Newton's method and some complexity aspects of the zero-finding problem*. In "Foundations of Computational Mathematics", (R.A. DeVore, A. Iserles, E. Suli), LMS Lecture Note Series **284**, Cambridge University Press, 2001, 45–67.

[4] B. Dwork, *On the rationality of the zeta function of an algebraic variety*. Amer. J. Math. **82** (1960), 631–648.

[5] B. Dwork, *On the zeta function of a hypersurface II*. Ann. Math. (2) **80** (1964), 227–299.

[6] N. Elkies, *Elliptic and modular curves over finite fields and related computational issues*. In "Computational perspectives in number theory: Proceedings of a conference in honour of A.O.L. Atkin" , (D.A. Buell and J.T. Teitelbaum), American Mathematical Society International Press **7** (1998), 21–76.

[7] M.D. Huang, D. Ierardi, *Counting points on curves over finite fields*. J. Symbolic Comput. **25** (1998), 1–21.

[8] K. Kedlaya, *Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology*. Journal of the Ramanujan Mathematical Society **16** (2001), 323–338.

[9] K. Kedlaya, *Finiteness of rigid cohomology with coefficients*, preprint 2002.

[10] A.G.B Lauder, *Deformation theory and the computation of zeta functions*, Proceedings of the London Mathematical Society **88** (3) (2004), 565-602.

[11] A.G.B. Lauder, *Counting solutions to equations in many variables over finite fields*, Foundations of Computational Mathematics **4** (3) (2004), 221-267.

[12] A.G.B. Lauder, D. Wan, *Counting points on varieties over finite fields of small characteristic*. To appear in Algorithmic Number Theory: Lattices, Number Fields, Curves and Cryptography (Mathematical Sciences Research Institute Publications), J.P. Buhler and P. Stevenhagen (eds), Cambridge University Press.
Available at: http://www.maths.ox.ac.uk/∼lauder/

[13] J. Pila, *Frobenius maps of abelian varieties and finding roots of unity in finite fields*. Math. Comp. **55** No. 192 (1990), 745–763.

[14] R. Schoof, *Elliptic curves over finite fields and the computation of square roots mod p*. Math. Comp. **44** no. 170 (1985), 483–494.

[15] N. Tsuzuki, *Bessel F-isocrystals and an algorithm for computing Kloosterman sums*, preprint 2003.

[16] D. Wan, *Algorithmic theory of zeta functions*. To appear in Algorithmic Number Theory: Lattices, Number Fields, Curves and Cryptography (Mathematical Sciences Research Institute Publications), J.P. Buhler and P. Stevenhagen (eds), Cambridge University Press. Available at: http://www.math.uci.edu/∼dwan/preprint.html

Alan G.B. Lauder
Mathematical Institute
Oxford University
24-29 St Giles
Oxford OX1 3LB
E-mail : lauder@maths.ox.ac.uk
URL: http://www.maths.ox.ac.uk/∼lauder/