Deformation theory and the computation of zeta functions

Alan G.B. Lauder *
Computing Laboratory and
St John's College
Oxford, U.K.

December 6, 2002

1 Introduction

An attractive and challenging problem in computational number theory is to count in an efficient manner the number of solutions to a multivariate polynomial equation over a finite field. One desires an algorithm whose time complexity is a small polynomial function of some appropriate measure of the size of the polynomial. A natural measure of size is $d^n \log(q)$ for a polynomial of total degree d in n variables over the field of q elements. Despite intensive research over the last few decades, existing algorithms fall well short of this ideal. The case $n = 1$, which is related to univariate polynomial factorisation, was solved essentially by Berlekamp and is comparatively straightforward, see [19, Chapter 14]. When $n = 2$ one is counting points on curves, a topic of some practical importance [2]. Here one can achieve a complexity of $\log(q)^{C_d}$ where the exponent C_d depends exponentially on d (this can be improved to C_d a polynomial in d in some special cases, see [1]). The only algorithm which applies for general n has a complexity which is a polynomial function of $(pd\log(q))^n$ [26] (see also the strategy in [23, Section 2]). This is polynomial-time in what one might call the small characteristic input size of $pd^n \log(q)$, but only for fixed dimension (that is, the exponent depends upon the dimension n). The purpose of this paper is to introduce a systematic new approach to counting solutions to equations over finite fields which aims to remove the exponential dependence on the dimension for small characteristic. That is, to obtain a single time complexity which is polynomial in $pd^n \log(q)$ uniformly over all n. A general method is sketched, and worked out for a particular quite broad family of polynomials. The new approach rests on two observations. First, the number of solutions to any equation defined by a diagonal polynomial can be computed easily within the required time bound. Second, any suitably generic polynomial can be deformed into a diagonal polynomial; more precisely, lies in a one-parameter family of polynomials containing a diagonal form. Over a finite field this deformation appears superficially to be of little use. Remarkably though, one can associate a linear p-adic differential equation with the deformation, and by solving this one can recover the number of solutions to the

*The author is supported by the EPSRC (Grant GR/N35366/01) and St John's College, Oxford. He wishes to thank Professors Richard Brent, Elmar Grosse-Klönne, Michael Rabin, Steven Sperber, Frederik Vercauteren and Daqing Wan for their generous help and encouragement during the preparation of this manuscript. *Mathematics Subject Classification 2000*: 11Y99, 11M38, 11T99.

original finite field equation. Thus, in a sense, one reduces a high-dimensional solution counting problem to a one-dimensional deformation problem, whence the magical reduction in complexity. We note that homotopy methods are apparently well-studied in the context of the numerical solution of systems of equations over the complex numbers [9, Section 4.2]. Our work seems to be the first explicit algorithmic application of such ideas to the more mysterious setting of equations over finite fields.

The problem of counting solutions can be reformulated in terms of computing zeta functions, and we use this language for the remainder of the paper. Precisely, a multivariate equation defines an affine hypersurface, and the number of solutions to the equation over all finite extensions of the base field is encoded in the zeta function of the hypersurface. Thus the task of "counting solutions to equations" is subsumed in "computing zeta functions of hypersurfaces", which is the problem we discuss. The notions of input size are carried over to this new setting. We work out all the details of our new approach for a particular family of Artin-Schreier covers of affine space. The algorithm we present has a cubic time dependence on $\log(q)$ and quartic on p in all dimensions. We now introduce the definitions and notation necessary to fully explain this result.

Let $\mathbb{F}_q$ be the finite field of q elements, where q is a power of a prime p. Let $f \in \mathbb{F}_q[X_1, \ldots, X_n]$ have total degree d (that is, the maximum sum $e_1 + \ldots + e_n$ of exponents which occurs amongst non-zero terms $aX_1^{e_1} \ldots X_n^{e_n}$ of f is d). Let X_f be the affine hypersurface of dimension n defined by the equation

$$Z^p - Z = f(X_1, \ldots, X_n).$$

The variety X_f is an Artin-Schreier cover and in this paper we shall restrict our attention mainly to such hypersurfaces, although in Section 2 we do discuss general projective hypersurfaces. Let $\#X_f(\mathbb{F}_{q^k})$ be the number of points $(z, x_1, \ldots, x_n) \in \mathbb{F}_{q^k}^{n+1}$ with $z^p - z = f(x_1, \ldots, x_n)$. The zeta function of X_f is the formal power series

$$Z(X_f/\mathbb{F}_q, T) = \exp\left(\sum_{k=1}^{\infty} \#X(\mathbb{F}_{q^k}) \frac{T^k}{k}\right).$$

Dwork's theorem tells us this is a rational function [12], and [3, Theorem 1] implies the upper bound $(p-1)(4d+5)^n + 1$ on the sum of the degrees of the numerator and denominator of this function (see Section 3.1). The polynomial-time computability theorem of the author and Wan [26] shows the following: There exists a deterministic algorithm and a polynomial P such that on input the polynomial f, the algorithm outputs the zeta function $Z(X_f/\mathbb{F}_q, T)$ within a number of bit operations $P((pd^n \log(q))^n)$. The space complexity of this algorithm is also $Q((pd^n \log(q))^n)$ bits, for some polynomial Q. This shows polynomial-time computability for fixed dimension, when the input size is taken as $pd^n \log(q)$. However, the exponent of the input size in the time/space complexity grows linearly with the dimension n. The reason for this is that the algorithm does all computations on the "cochain level" , i.e., with truncated n-variate power series of total degree around $pd^n \log(q)$ over the modular reduction of some suitable p-adic ring. Such power series have around $(pd^n \log(q))^n$ terms.

To get a better algorithm, one tries to use some p-adic cohomological formula. We say that f is non-degenerate if the polynomials

$$X_1 \frac{\partial f_d}{\partial X_1}, \ldots, X_n \frac{\partial f_d}{\partial X_n}$$

have no common zero other than the origin. Here f_d is the homogeneous part of f of degree d. A simple p-adic cohomological formula exists when f is non-degenerate (see [32]), and it may be used to compute the zeta function more quickly (for example, the case $n = 1$ is worked out explicitly in [27]). However, even using this cohomological formula, the complexity dependence on the input size $pd^n \log(q)$ has an exponent growing linearly with the dimension n. The approach [23] of Kedlaya based on the cohomological formula for smooth affine varieties of Monsky-Washnitzer would also yield a similar complexity estimate. The reason for this is that to compute the action of the (absolute) Frobenius map on the cohomology space using either method, one lifts the basis elements to the cochain level, computes the cochain Frobenius on these elements, and returns back to the cohomology space using reduction formulae. As mentioned before, even representing elements on the cochain level is very expensive. To remove the complexity dependence on the dimension, we present a completely different approach which avoids (almost) all computations on the cochain level.

We now state our main theorem. We use Soft-Oh notation which hides logarithmic factors, as defined in [26, Section 6.3]. Note that with our Oh and Soft-Oh notation the estimates hold for all values of the parameters n, d, p, $\log(q)$ (the logarithm to the base 2) and r, subject to the restrictions on the input type.

Theorem 1 *There exists an explicit deterministic algorithm with the following input, output and complexity. The input is any polynomial of the form*

$$f = \sum_{i=1}^{n} a_i X_i^d + y h(X_1, \ldots, X_n),$$

where $a_1, \ldots, a_n \in \mathbb{F}_q$ are all non-zero, the element $y \in \mathbb{F}_{q^r}$, and $h \in \mathbb{F}_q[X_1, \ldots, X_n]$ has total degree less than d. Here n, d and r are positive integers, the integer q is a power of a prime p with $p > 2$, and p does not divide d. The output is the zeta function $Z(X_f, T)$ of the affine hypersurface defined by the equation $Z^p - Z = f$. The running time of the algorithm is $\tilde{\mathcal{O}}(n^{\max(n+2,5)} d^{5n+3} p^4 \log(q)^3 r^3)$ bit operations, and the space requirement is $\tilde{\mathcal{O}}(n^{n+2} d^{4n+2} p^4 \log(q)^3 r^2)$ bits.

The factor n^{n+2} arises from Step 2 in the algorithm (Algorithm 13), which is the only part which involves any computation on the cochain level. It might be possible to remove this by a more careful analysis of this part of the algorithm.

Taking $r = 1$ and $y = 1$ in the theorem we get the next result.

Corollary 2 *There exists an explicit deterministic algorithm with the following input, output and complexity. The input is any polynomial of the form*

$$f = \sum_{i=1}^{n} a_i X_i^d + h(X_1, \ldots, X_n),$$

where $a_1, \ldots, a_n \in \mathbb{F}_q$ are all non-zero, and $h \in \mathbb{F}_q[X_1, \ldots, X_n]$ has total degree less than d. Here n and d are positive integers, the integer q is a power of a prime p with $p > 2$, and p does not divide d. The output is the zeta function $Z(X_f, T)$ of the affine hypersurface defined by the equation $Z^p - Z = f$. The running time of the algorithm is $\tilde{\mathcal{O}}(n^{\max(n+2,5)} d^{5n+3} p^4 \log(q)^3)$ bit operations, and the space requirement is $\tilde{\mathcal{O}}(n^{n+2} d^{4n+2} p^4 \log(q)^3)$ bits.

This shows that the method may be applied to any polynomial over $\mathbb{F}_q$ whose leading form is diagonal. Similarly, taking q equal to p and relabelling p^r by a different q, we get the next corollary.

Corollary 3 *There exists an explicit deterministic algorithm with the following input, output and complexity. The input is any polynomial of the form*

$$f = \sum_{i=1}^{n} a_i X_i^d + y h(X_1, \ldots, X_n),$$

where $a_1, \ldots, a_n \in \mathbb{F}_p$ *are all non-zero, the element* $y \in \mathbb{F}_q$*, and* $h \in \mathbb{F}_p[X_1, \ldots, X_n]$ *has total degree less than* d*. Here* n *and* d *are positive integers, the integer* q *is a power of a prime* p *with* $p > 2$*, and* p *does not divide* d*. The output is the zeta function* $Z(X_f, T)$ *of the affine hypersurface defined by the equation* $Z^p - Z = f$*. The running time of the algorithm is* $\tilde{\mathcal{O}}(n^{\max(n+2,5)} d^{5n+3} p^4 \log(q)^3)$ *bit operations, and the space requirement is* $\tilde{\mathcal{O}}(n^{n+2} d^{4n+2} p^4 \log(q)^2)$ *bits.*

This shows that the space complexity can be reduced when the polynomial belongs to a suitable family over $\mathbb{F}_p$. The significance of this observation is that the method for curves of Kedlaya [23], and also the related method from [27], based on a direct computation of a p-adic cohomological formula, requires cubic space/time in $\log(q)$ in all cases. However, the algorithm of Satoh for elliptic curves [29], a one-dimensional family over $\mathbb{F}_p$, needs only quadratic space [34]. This suggests some connection between our algorithm and that of Satoh, although we have not fully explored this.

Our main theorem also has the following curious consequence for counting solutions to equations.

Corollary 4 *Let* $f \in \mathbb{Z}[X_1, \ldots, X_n]$ *have the form*

$$f = \sum_{i=1}^{n} a_i X_i^d + h(X_1, \ldots, X_n)$$

where $a_1 \ldots a_n \neq 0$ *and* h *has total degree less than* d*. There exists an explicit deterministic algorithm which takes as input a prime* p *and outputs the number of solutions over* $\mathbb{F}_p$ *to the equation* $f = 0$*, and runs in* $\tilde{\mathcal{O}}(p^2)$ *bit operations.*

Here the algorithm itself depends upon the polynomial f, and thus the constant factor which the Soft-Oh notation hides also depends upon f, although the exponent of $\log(p)$ which is also hidden is independent of f. The naive bound for this problem would be $\mathcal{O}(p^n)$, or better still $\tilde{\mathcal{O}}(p^{n-1})$ using a root counting algorithm for univariate polynomials.

Note that the case $p = 2$ in Theorem 1 can probably be handled by using the modification in [27, Note 33], although we have not studied this in detail. In fact, an implementation of Vercauteren and the author suggests that the present algorithm does work correctly for $p = 2$. One could perhaps prove this using the argument in [27, Section 7.2]. We have not looked at the case p divides d, although perhaps the theory from [15, Section 6(e)] may be of use here.

There is a large literature devoted to the problem of computing zeta functions, or point counting as it is familiarly known. This is motivated in part by applications in cryptography (see the bibliography in [2] for a list of publications). The starting point, leaving aside univariate factorisation algorithms, was [31]. The paper [17] also contains important improvements for elliptic curves, and [23, 26, 29] make significant algorithmic advances. Other recent work includes [11, 18, 20, 21, 24, 26, 27, 28, 30, 33, 34, 35, 37]. The method of this paper seems quite different from most existing approaches. However, as previously stated, the author believes one should be able to describe the algorithm of Satoh for elliptic curves using our deformation methodology (see also Note 7).

The paper is organised in the following manner. In Section 2 we describe our general strategy for computing zeta functions, and give a small example. In Section 3 the proof of Theorem 1 starts in earnest: we describe the theoretical results needed for our algorithm. The results themselves seem new, at least in the explicit incarnation we need, and are proved in Appendix A. Section 4 contains the algorithm itself. In Section 5 we present subroutines for three key tasks, and in Section 6 we prove the correctness of a formula used in one step of the algorithm and discuss how many terms in this formula need to be evaluated. Section 7 is the most technical, as it involves detailed estimates on the p-adic orders of the numbers which arise in the algorithm. The results in Section 7 allow us in Section 8 to analyse the propagation of errors through the algorithm, and thus prove the output is exactly correct. Finally, the complexity of the algorithm is estimated in Section 9, and the proofs of the results in the introduction completed in Section 10.

2 The Strategy

We first describe an approach for computing zeta functions of smooth projective hypersurfaces. We will develop this fully in a sequel paper [25] (the present paper contains a technically simpler algorithm in a related setting). This section contains no proofs or definitions, and does not form part of the proof of Theorem 1. The results quoted in this section can be found in [15]. Indeed, our new approach was inspired by Dwork's proof of the functional equation of the zeta function of a smooth projective hypersurface in [14, 15]. An outline of the approach is as follows: to compute the zeta function of a single smooth hypersurface, one embeds it in a one-parameter family, such that the fibre at the origin is smooth and has an easily computed zeta function. By computing numerically a basis of solutions to an associated linear differential system around the origin, one can then recover the zeta function of the original hypersurface. We now give more details.

Let $f(X,Y) \in \mathbb{F}_q[Y][X_1,\dots,X_n]$ be homogeneous of degree d in the variables $X_1,\dots,X_n$, where q is a power of a prime p, and p does not divide d. The polynomial f defines a family of hypersurfaces parameterised by the variable Y which we assume are generically smooth and in general position. (A hypersurface of degree not divisible by the characteristic is smooth and in general position if the intersection of it with each coordinate subspace is smooth, see [22, Page 75].) If one is interested in a particular smooth hypersurface in general position, assume it is embedded in this family. (The explicit family which Dwork considers is defined by the polynomial $f(X,Y) = \sum_{i=1}^{n} a_i X_i^d + Yh(X)$ where $a_1 \dots a_n \neq 0$ and $h(X)$ has no diagonal terms, and the hypersurface of interest can be recovered by setting $Y = 1$.)

Suppose we wish to count points on the smooth, and in general position, projective hypersurface defined by f for some specialisation in $\mathbb{F}_q$ of the variable Y. The number of points can be obtained easily from the trace of the Frobenius map acting on some cohomology space. This map factors as a product of $\log_p(q)$ semi-linear maps, and to obtain polynomial-time computability in $p\log(q)$ one must exploit this factorisation. However, since this feature is common to all p-adic algorithms we shall ignore this key step, and in this section just assume $q = p$ is a prime. Let $\bar{W} \subseteq \mathbb{F}_q$ be the set of specialisations of $Y = \bar{y}$ such that $f(X,\bar{y})$ does not define a smooth projective hypersurface in general position. We assume that zero does not lie in $\bar{W}$. Let W be the subset of the p-adic projective line obtained by taking the Teichmüller lifting of the points in $\bar{W}$ along with infinity. For each specialisation $Y = \bar{y} \notin \bar{W}$, a Frobenius matrix $\alpha(\bar{y})$ is defined via Dwork's cohomology theory. Our aim is to compute the matrix $\alpha(\bar{y})$ efficiently. Write y for the Teichmüller lifting of $\bar{y}$. Define $\alpha(Y)$ to be the function that associates to each $y \notin W$ the matrix $\alpha(\bar{y})$. Dwork's theory tells us that the entries in the matrix $\alpha(Y)$ are analytic functions on the p-adic projective line with open unit disks around the points of W removed. Dwork's deformation theory gives the equation

$$\alpha(Y) = C(Y^p)^{-1}\alpha(0)C(Y). \tag{1}$$

Here $C(Y)$ is some matrix of p-adic analytic functions with "large" disks around the points of W removed. Finally, the matrix $C(Y)$ satisfies a differential equation

$$\frac{\partial}{\partial Y}C(Y) = C(Y)B(Y),\ C(0) = I, \tag{2}$$

where $B(Y)$ is a matrix of rational functions on the p-adic projective line with the points of W removed.

Equation (1) can be used to compute $\alpha(Y)$, and hence $\alpha(\bar{y})$, provided $C(Y)$ and $\alpha(0)$ can be found. The matrix $B(Y)$ can always be computed easily using a simple, explicit method due to Dwork (see [15, Section 8]). This requires a little computation on the cochain level. One can put most polynomials in a family $f(X,Y)$ such that $\alpha(0)$ can be computed easily. For example, one might assume $f(X,0)$ is a diagonal form, as in Dwork's explicit family, for then the Frobenius matrix has a Kronecker product decomposition. An expansion of the matrix $C(Y)$ around the origin may be computed by solving the differential system (2) using the method in Section 5.2.1. The central difficulty is to determine to what p-adic and Y-adic accuracies this expansion must be computed in order to recover $\alpha(Y)$ up to the necessary p-adic accuracy, and how to perform this recovery. We address these problems in general in [25]. For now, we observe that assuming an explicit overconvergence bound can be found for the matrix $\alpha(Y)$ (compare with Proposition 17), these computations just involve manipulation of univariate power series and rational functions of "small degree". As such, it is at least intuitively plausible that these computations can be performed within a time bound of the desired form, that is, with an exponent independent of n. Having found $\alpha(Y)$ up to a suitable p-adic accuracy, for $\bar{y} \notin \bar{W}$ one then simply evaluates this matrix at the Teichmüller point y and takes the trace to find the number of points on the hypersurface $f(X,\bar{y}) = 0$. (The precise relation between the trace and the number of points is given on [22, Page 75].)

In the present paper, we consider a slightly different situation to that described above. Instead of smooth projective hypersurfaces in general position, we consider non-degenerate exponential sums and the related Artin-Schreier covers. For these sums, an analogous theory to that described above exists. The advantage now is that the non-degeneracy condition is more flexible, and one can explicitly write down interesting families of exponential sums in which each fibre is non-degenerate. We then find that the set analogous to W above just contains ∞. As such the matrix $B(Y)$ one considers has polynomial entries, and we can more easily solve all of the necessary problems.

We now give a simple example.

Example 5 Let $f = X^2 + YX$ with $p > 2$, so we are interested in the curve $Z^p - Z = f(X,Y)$ for different specialisations of Y. The associated cohomology space is one dimensional, as are all the matrices. The matrix $B(Y)$ is just $(-\pi Y/2)$, where π is a p-adic number with $\pi^{p-1} = -p$. Hence $C(Y) = \exp(-\pi(Y^2/4)) \in \mathbb{Z}_p[\pi][[Y]]$, where $\mathbb{Z}_p$ is the p-adic integers. Using Dwork's theory one works out that the matrix $\alpha(0)$ is

$$\alpha(0) = \sum_{m,r\geq 0,\, 2(m-pr)=p-1} (-1)^r \lambda_m (1/2)_r \pi^{-r}.$$

The unfamiliar symbols here are defined in Section 3.2. Thus the matrix $\alpha(Y)$ is

$$\exp(\pi(Y^{2p}/4))\alpha(0)\exp(-\pi(Y^2/4)).$$

Although $\exp(-\pi(Y^2/4))$ has integral coefficients, they do not decay quickly. When one computes the product though, one sees the coefficients of $\alpha(Y)$ decay quickly. This feature is key, as it allows one to estimate how many of the terms in the slow decaying series $C(Y)$ one needs to compute to determine $\alpha(Y)$ to a suitable degree of accuracy. Note that the number of points on $Z^p - Z = f(X, \bar{y})$ is actually the trace from $\mathbb{Z}_p[\pi]$ to $\mathbb{Z}_p$ of $\alpha(y)$.

Note 6 The results of Dwork were reformulated in terms of the Gauss-Manin connection acting on the middle dimensional analytic de Rham cohomology group by Katz [22, Pages 75-77]. Here one studies the Picard-Fuchs differential equation of the family of hypersurfaces. This equation can be computed using the Griffiths-Dwork method from [8, Section 5.3]. Dwork developed his methods further in later work, and his techniques have been used by many different authors: for example, by Candelas et al in [4] to study varieties from a specific family of interest in mathematical physics. In the context of the work of Dwork and his school, one is interested in finding explicit formulae for (the p-adic slope decomposition of) the zeta functions in the family in terms of closed form solutions to the Picard-Fuchs equation. The main insight of the present author is that, assuming the origin is not a singular point of the Picard-Fuchs equation, one can easily compute numerically a basis of solutions around the origin. Provided the Frobenius matrix of the fibre at the origin has been computed, one can then recover exactly the zeta function of any smooth fibre in the family. Moreover, the complexity of this approach greatly improves upon all previous algorithms for computing zeta functions of smooth hypersurfaces of high dimension. The idea should extend to more general varieties, such as complete intersections, although the author has not studied this at all.

Note 7 In algorithmic applications of Dwork's theory it is necessary that the "holomorphic functions" one considers have the property of being overconvergent, rather than just convergent. Overconvergence is a necessary condition, since one requires that the functions under consideration reduce to small degree rational functions when one reduces coefficients modulo powers of the characteristic. In such cases they can be evaluated modulo a power of the characteristic quickly. The absolute Frobenius matrices which arise in our work have this property. For example, the precise radius of convergence of the relevant matrix is established in Proposition 17. However, overconvergence is lost, or at least is not known to hold in general, when one tries to compute a "slope decomposition" of the absolute Frobenius matrix. Precisely, when one computes the "Hodge-Newton decomposition" of the corresponding "overconvergent F-crystal". This is related to the long-standing meromorphy conjecture of Dwork, which was recently proved by Wan (see [36, Section 1] for a discussion of the conjecture). Dwork was able to prove his conjecture for elliptic curves by finding a so-called excellent lifting of Frobenius, in which the unit root part remains over convergent, see [16, Section 1]. The algorithm of Satoh presumably exploits this special lifting to compute the unit root part. However, using the usual lifting of Frobenius the unit root part is not known to be overconvergent. It was pointed out to the author by Vercauteren that, indeed, the unit root formula in [14, Section 5] is not useful in practice in computing the zeta function of an elliptic curve. From an algorithmic viewpoint, it would be very nice to be able to compute the different p-adic slopes of the zeta function of a variety independently, since then one could work with much smaller matrices. However, a deeper understanding of when excellent liftings exist and their construction seems essential.

3 The Theory

Let $\bar{f} = \sum_{i=1}^{n} \bar{a}_i X_i^d + Y\bar{h}(X) \in \mathbb{F}_q[Y][X]$, where we write X for the list of indeterminates $X_1, \ldots, X_n$, and q is a power of the prime $p > 2$, and p does not divide d. Here $d > 1$, for the case $d = 1$ is trivial. We assume that $\bar{a}_1 \ldots \bar{a}_n \neq 0$ and the polynomial $\bar{h}$ has degree less than d. (The bars will be removed when we take p-adic liftings. Note that we just dropped the bars in the previous sections for notational simplicity.) We use the symbol X^m to denote $X_1^{m_1} \ldots X_n^{m_n}$ for $m = (m_1, \ldots, m_n)$ a vector of non-negative integers, and define $|m| = m_1 + \ldots + m_n$. Similarly, we write $x \in K^n$ to denote a point $(x_1, \ldots, x_n)$ in affine space over some field K.

The aim of the paper is to compute the zeta function of the affine variety $Z^p - Z = \bar{f}(X, \bar{y})$ for different specialisations $Y = \bar{y} \in \mathbb{F}_{q^r}$ of the variable Y. This is achieved via a deformation theory for exponential sums of a certain form. This theory is based upon the work of Dwork, and is developed in full in Appendix A. In this section, we shall just quote the pertinent results when needed.

3.1 Exponential sums

Let ζ_p be a primitive pth root of unity in some extension of the rational numbers $\mathbb{Q}$, and $\mathrm{Tr}_k : \mathbb{F}_{q^k} \to \mathbb{F}_p$ the trace map. Let $\bar{y} \in \mathbb{F}_{q^r}$. For each $k \geq 1$ with $r|k$ define

$$\begin{array}{rcl} S_k & = & \sum_{x \in \mathbb{F}_{q^k}} \zeta_p^{\mathrm{Tr}_k(\bar{f}(x,\bar{y}))}, \\ L(\bar{f}(X,\bar{y}), T) & = & \exp\left(\sum_{k \geq 1, r|k} S_k \frac{T^k}{k}\right). \end{array}$$

The Bombieri degree bound states that $L(\bar{f}(X, \bar{y}), T)$ is a ratio of polynomials in $1 + T\mathbb{Z}[T]$ whose degrees sum to at most $(4d + 5)^n$ [3, Theorem 1]. Let G be the Galois group of the extension $\mathbb{Q}(\zeta_p)/\mathbb{Q}$. Thus $G = \{\theta_i \,|\, 1 \leq i \leq p - 1\}$ where $\theta_i : \zeta_p \mapsto \zeta_p^i$. A straightforward argument (compare with [27, Section 2]) shows that the zeta function of the affine hypersurface $X_{\bar{f}}$ over $\mathbb{F}_{q^r}$ defined by the equation $Z^p - Z = \bar{f}(X, \bar{y})$ satisfies

$$Z(X_{\bar{f}}/\mathbb{F}_{q^r}, T) = \frac{\prod_{\theta \in G} \theta(L(\bar{f}(X, \bar{y}), T))}{1 - q^{rn}T}. \tag{3}$$

Here the group G acts on the coefficients of polynomials. Thus $Z(X_{\bar{f}}/\mathbb{F}_{q^r}, T)$ has total degree at most $(p-1)(4d+5)^n+1$ as a rational function. All the statements made up to now in this paragraph are true for general $\bar{f}/\mathbb{F}_{q^r}$ of degree d, which gives the degree bound claimed in the introduction. Since p does not divide d and the leading form of $\bar{f}(X, \bar{y})$ is non-degenerate, we have a stronger result. From [32, Theorem 2.37] we have that $L(\bar{f}(X, \bar{y}), T)^{(-1)^{n+1}}$ is a polynomial of degree $(d-1)^n$ in $1 + T\mathbb{Z}[T]$. Equation (3) reduces the computation of the zeta function to this L-function. This latter function is realised as the characteristic polynomial of a map, called the Frobenius map, on a space of dimension $(d - 1)^n$ over a suitable p-adic field. Precisely, in the next section we define a matrix $\alpha(Y)$ of size $(d - 1)^n$ whose entries are power series over the ring of integers of a p-adic field (the absolute Frobenius matrix). We will compute this matrix modulo appropriate powers of the uniformizers p and Y. The L-function $L(\bar{f}(X,\bar{y}),T)$ can then be obtained by evaluating this matrix at a certain point, and taking the reverse characteristic polynomial of some "power" of the evaluated matrix. The matrix $\alpha(Y)$ is found in an indirect manner, using deformation theory and its value at the specialisation $Y=0$. The theory behind all this is explained in the next section.

Note that for $s \in \mathbb{F}_{q^r}$ we have that $L(\bar{f}(X,\bar{y})+s,T) = L(\bar{f}(X,\bar{y}),\zeta_p^{\mathrm{Tr}_r(s)}T)$. Thus it is enough to compute L-functions for $\bar{f}$ with constant term zero. Given a polynomial $\bar{f}$ with non-zero constant s term we need only make the substitution $T \mapsto \zeta_p^{\mathrm{Tr}_r(s)}T$ in the L-function of $\bar{f}-s$.

3.2 p-Adic theory

Prior to stating the propositions which underpin our algorithm, we need to make some more definitions.

First, we introduce some notation related to p-adic rings: Let $\mathbb{Q}_p$ be the field of p-adic numbers, and $\mathbb{Z}_p$ the ring of p-adic integers. Let $\mathbb{C}_p$ denote the completion of an algebraic closure of $\mathbb{Q}_p$. We denote by $\mathbb{Q}_q$ the unique unramified extension of $\mathbb{Q}_p$ in $\mathbb{C}_p$ of degree $\log_p(q)$, and $\mathbb{Z}_q$ the ring of integers of $\mathbb{Q}_q$. Let $\pi \in \mathbb{C}_p$ be an element such that $\pi^{p-1} = -p$. Let ord be the p-adic order function on $\mathbb{C}_p$ normalised so that $\mathrm{ord}(p) = 1$, and $\mathbb{R}_p$ the ring of integers of $\mathbb{C}_p$. Observe that $\mathrm{ord}(\pi) = 1/(p-1)$. For $y \in \mathbb{R}_p$ in the unique unramified extension of $\mathbb{Z}_q$ of degree r for some $r \geq 1$, let τ denote the automorphism of $\mathbb{Z}_q[\pi][y]$ which reduces to the pth power map on its residue class field and fixes π. Let h denote the polynomial in $\mathbb{Z}_q[X]$ whose coefficients are the Teichmüller liftings of those of $\bar{h}$. Let a_i be the Teichmüller lifting of $\bar{a}_i$, and $f = \sum_{i=1}^n a_i X_i^d + Yh$.

Next, we need to define the "differential equation of the deformation". Let

$$D_{i,Y} := X_i \frac{\partial}{\partial X_i} + \pi \left(d a_i X_i^d + Y X_i \frac{\partial h}{\partial X_i} \right), \; 1 \leq i \leq n$$

act on the ring $\mathbb{Q}_q[\pi][Y][X]$. The set

$$\{X^u \mid u \in \mathcal{B}\} \text{ where } \mathcal{B} := \{u = (u_1,\ldots,u_n) \mid 0 < u_1,\ldots,u_n < d\}$$

is a $\mathbb{Q}_q[\pi][Y]$-basis for the quotient $\mathbb{Q}_q[\pi][Y]$-module

$$X_1 \ldots X_n * / (X_1 \ldots X_n * \cap \sum_{i=1}^n D_{i,Y}(*)),$$

with $* = \mathbb{Q}_q[\pi][Y][X]$ (see Proposition 25).

Definition 8 *For each $v \in \mathcal{B}$, let $B_{u,v}(Y)$ denote the coefficient of X^u in the reduction of $\pi h X^v$ modulo the images of the operators $D_{i,Y}$. Let $B(Y) = (B_{u,v})$ be the corresponding square matrix of size $(d-1)^n$ over $\mathbb{Q}_q[\pi][Y]$.*

Each entry in B is a polynomial in Y of degree at most $n(d-1)$ with coefficients of p-adic order $> -n/(p-1)$. We prove these facts in Section 5.1, where we shall also describe how to compute this matrix.

We can now introduce the deformation matrix itself.

Definition 9 *Let $C(Y)$ be the matrix over $\mathbb{Q}_q[\pi][[Y]]$ which is the unique solution to the differential equation and initial condition*

$$\frac{\partial C}{\partial Y} = C(Y)B(Y), \; C(Y) \equiv I \bmod (Y). \tag{4}$$

In Section 5.2.1 we shall see how to compute $C(Y)$ up to any Y-adic accuracy.

The idea is that the matrix $C(Y)$ controls the change in the "generic" L-function "$L(\bar{f}(X,Y),T)$" as one moves from $Y=0$ to a "generic" choice of Y. We give an explicit formula for the L-function for the specialisation $Y=0$. More precisely, for the matrix which represents the action of the absolute Frobenius map on the cohomology space constructed by Dwork.

Proposition 10 *Let $\alpha(0)$ be the matrix for the action of the absolute Frobenius map on the cohomology space associated to the L-function $L(\bar{f}(X,0),T)$. (This is defined in Section A.2.) The entry in the uth row and vth column of $\alpha(0)$ for $u,v \in \mathcal{B}$ is*

$$\prod_{i=1}^{n} \sum_{m,r \geq 0} \lambda_m (u_i/d)_r (-1)^r \pi^{-r} \tau^{-1}(a_i^m) a_i^{-r}. \tag{5}$$

Here the sum is over all $m,r \geq 0$ such that $pu_i - v_i = d(m-pr)$, where $u=(u_1,\ldots,u_n)$ and $v=(v_1,\ldots,v_n)$.

The p-adic integer λ_m is the coefficient of z^m in the power series $\exp(\pi(z-z^p))$. The p-adic integer $(u_i/d)_r$ is defined to be 1 when $r=0$, and for $r>0$ to be

$$(u_i/d)_r := (u_i/d)((u_i/d)+1)\ldots((u_i/d)+(r-1)).$$

This proposition is proved in Section 6.1.

The main result of the deformation theory is as follows.

Proposition 11 *Let $\alpha(Y)$ denote the absolute Frobenius matrix for the "generic" Y. (This matrix is defined explicitly in Section A.2.) Then we have the following identity of matrices over $\mathbb{Q}_q[\pi][[Y]]$*

$$\alpha(Y) = C(Y^p)^{-1}\alpha(0)C^{\tau^{-1}}(Y).$$

Here τ^{-1} acts entry wise on the matrix, fixing Y.

This proposition is proved in Appendix A. Specifically, see Equation (29) and Section A.3.

We now related the L-function to the zeta function of the affine hypersurface. Let G denote the Galois group of the extension $\mathbb{Z}_p[\pi]/\mathbb{Z}_p$. The group $\#G$ has order $p-1$ and one can explicitly write down the action of each group element on π [26, Equation (11)]. The next proposition is proved in Section A.2.

Proposition 12 *Write*

$$Z(X_{\bar{f}(X,\bar{y})}/\mathbb{F}_{q^r},T) = \frac{P(T)}{(1-q^{rn}T)}$$

for the zeta function of the affine hypersurface $Z^p - Z = \bar{f}(X,\bar{y})$ defined by specialising $Y=\bar{y}$ where $\bar{y} \in \mathbb{F}_{q^r}$. Then

$$P(T)^{(-1)^{n+1}} = \prod_{\theta \in G} \det(I - \alpha(y^{\tau^{-1}})^{\tau^{r\log_p(q)}} \alpha(y^{\tau^{-1}})^{\tau^{r\log_p(q)-1}} \ldots \alpha(y^{\tau^{-1}})^{\tau} T)^{\theta}.$$

Here y is the Teichmüller lifting of the field element $\bar{y}$.

4 The Algorithm

We now put the results in the previous section together in an appropriate manner to give our point counting algorithm. The proof of Theorem 1 rests upon a demonstration of the correctness of this algorithm and an estimate of its time and space complexity.

Algorithm 13
Input: A polynomial $\bar{f}(X,\bar{y}) = \sum_{i=1}^{n} \bar{a}_i X_i^d + \bar{y}\bar{h}(X)$. Here $\bar{a}_1, \ldots, \bar{a}_n \in \mathbb{F}_q$ are non-zero, where $\mathbb{F}_q$ has characteristic $p > 2$, and p does not divide d. The polynomial $\bar{h} \in \mathbb{F}_q[X_1, \ldots, X_n]$ has total degree less than $d(> 1)$, and the element $\bar{y} \in \mathbb{F}_{q^r}$, for some $r \geq 1$. (We assume $\bar{h}(0) = 0$, as the case $\bar{h}(0) \neq 0$ can be easily reduced to this case by the comment at the end of Section 3.1.)
Output: The zeta function of the affine variety defined by the equation $Z^p - Z = \bar{f}(X,\bar{y})$.

Step 0: Set-up
We use the notation defined in Section 3. All computations below are performed with p-adic numbers in $\mathbb{Q}_q[\pi]$ and $\mathbb{Q}_q[\pi, y]$ working "modulo" some power of p (see Section 8.2 for a precise definition of this phrase). Define

$$\begin{array}{rcl} N & = & \lceil (p-1)(d-1)^n((rn\log_p(q)) + \log_p(2)) \rceil + 1 \\ N_Y & = & 12pd(N + (d-1)^n rn \log_p(q) + n) \\ \tilde{N} & = & (60nd+1)(N + (d-1)^n rn\log_p(q)) + 60n^2 d + n(p+2). \end{array}$$

Step 1: Teichmüller liftings
Compute modulo $p^{\tilde{N}}$ the Teichmüller liftings of the coefficients of $\bar{h}(X)$.

Step 2: Compute the differential system
Let $B(Y)$ be the matrix of the differential system, as in Definition 8. Using the method of Section 5.1 compute the matrix B with coefficients modulo $p^{\tilde{N}}$.

Step 3: Solve the differential system at the origin
Let $C(Y)$ be the unique solution matrix to the differential system, as in Definition 9. Working modulo $(p^{\tilde{N}}, Y^{N_Y})$, compute $C(Y)$ using the method in Section 5.2.1.

Step 4: Matrix inversion
Working modulo $(p^{\tilde{N}}, Y^{N_Y})$ compute the inverse matrix $C(Y^p)^{-1}$ using the Newton iteration method in Section 5.2.2.

Step 5: Find the absolute Frobenius matrix for the diagonal case
Let $\alpha(0)$ be the matrix for the absolute Frobenius map in the case $\bar{y} = 0$, as defined in Proposition 10. Compute $\alpha(0)$ modulo $p^{\tilde{N}}$ using the summation bounds in Section 6.2.

Step 6: Deform the Frobenius matrix
Working modulo $(p^{\tilde{N}}, Y^{N_Y})$ compute the the matrix product

$$\alpha(Y) := C(Y^p)^{-1}\alpha(0)C^{\tau^{-1}}(Y).$$

STEP 7: EVALUATE THE DEFORMED MATRIX
Compute the Teichmüller lifting y of the element $\bar{y}$ modulo $p^{\tilde{N}}$. Compute $\alpha(y^{\tau^{-1}})$, the matrix $\alpha(Y)$ specialised at $Y = \tau^{-1}(y)$, modulo $p^{\tilde{N}}$.

STEP 8: EXPONENTIATE AND TAKE THE CHARACTERISTIC POLYNOMIAL
Let $L(\bar{f}(X,\bar{y}),T)$ be the rational function over $\mathbb{Z}_p[\pi]$ defined as

$$L(\bar{f}(X,\bar{y}),T)^{(-1)^{n+1}} = \det(I - \alpha(y^{\tau^{-1}})^{\tau^{r\log_p(q)}}\alpha(y^{\tau^{-1}})^{\tau^{r\log_p(q)-1}}\ldots\alpha(y^{\tau^{-1}})^{\tau}T).$$

Compute $L(\bar{f}(X,\bar{y}),T)$ modulo $p^{\tilde{N}}$ using fast exponentiation and the algorithm for characteristic polynomials from Section 9.

STEP 9: CYCLOTOMIC NORM AND THE ZETA FUNCTION
Let G be the Galois group of $\mathbb{Z}_p[\pi]$ over $\mathbb{Z}_p$. Compute the product $\prod_{\theta\in G}\theta(L(\bar{f}(X,\bar{y}),T))$ modulo p^N. Let $P(T)$ with $P(T)^{(-1)^{n+1}} \in 1 + T\mathbb{Z}[T]$ be the unique rational function with coefficients in the range $(-p^{N-1}, p^{N-1}]$ such that $P(T) \equiv \prod_{\theta\in G}\theta(L(\bar{f}(X,\bar{y}),T)) \bmod p^N$. Output the rational function

$$Z(X_{\bar{f}(X,\bar{y})},T) = \frac{P(T)}{1-q^{rn}T}.$$

In Section 5 we explain how to perform efficiently the non-trivial tasks in Steps 2-4. The formulae which are used in the algorithm are stated in the propositions in Section 3. The proofs of these propositions are located in Section 6 and Appendix A. It follows from these propositions that the algorithm would give the correct answer if the computations could be performed to infinite p-adic and Y-adic accuracy. We must justify that the various p and Y-adic accuracies at which power series are truncated does not compromise the final answer. This is done in Section 8. The results in that section rely upon lower bounds on the p-adic order of the numbers which occur in the algorithm. These bounds are obtained in Section 7, the most difficult part being the estimation of the rate of decay of the coefficients in the matrix $\alpha(Y)$. This will prove the correctness of the algorithm. The analysis of the complexity of the algorithm is given in Section 9, and the proof of Theorem 1 and the various corollaries completed in Section 10.

5 Required Subroutines

5.1 Computing the differential system

In this section we describe how to compute the matrix $B(Y)$ which defines the linear differential system. Write $B_{u,v}(Y)$ for the entry in the uth row and vth column of the matrix B. By definition, this is the coefficient of X^u in the reduction of $\pi h X^v$ modulo the operators

$$D_{i,Y} = X_i\frac{\partial}{\partial X_i} + \pi X_i\frac{\partial f}{\partial X_i}.$$

Here $f = \sum_{i=1}^n a_i X_i^d + Yh(X)$. We describe how to reduce an arbitrary polynomial $g(X,Y) \in \mathbb{Z}_q[\pi][X,Y]$ to a linear combination over $\mathbb{Q}_q[\pi][Y]$ of the basis set $\{X^u \mid u \in \mathcal{B}\}$. Let the total degree of g be s, and let g_s be the leading form of s (with respect to the variables X). Assume $s \geq d$ for otherwise g_s is already reduced. Let $g_s^{(1)}$ be the sum of all terms in g_s divisible by X_1^d, $g_s^{(2)}$ the sum of all terms in $g_s - g_s^{(1)}$ divisible by X_2^d, and more generally for $2 \leq i \leq n$, $g_s^{(i)}$ the sum of all terms in $g_s - \sum_{j=1}^{i-1} g_s^{(j)}$ divisible by X_i^d. We have that

$$\sum_{i=1}^{n} g_s^{(i)} = \sum_{i=1}^{n} \pi a_i d X_i^d (g_s^{(i)}/(\pi a_i d X_i^d)). \tag{6}$$

Now the terms in $g_s^{(0)} := g_s - \sum_{i=1}^{n} g_s^{(i)}$ are not divisible by any X_i^d for $1 \leq i \leq n$. Thus $g_s^{(0)}$ is already a sum of monomials in the basis set $\{X^u \mid u \in \mathcal{B}\}$. We need only reduce the difference $g_s - g_s^{(0)}$. From Equation (6) we have that

$$\sum_{i=1}^{n} D_{i,Y}(g_s^{(i)}/(\pi a_i d X_i^d)) = (g_s - g_s^{(0)}) + \sum_{i=1}^{n} \left(X_i \frac{\partial}{\partial X_i} + \pi X_i Y \frac{\partial h}{\partial X_i} \right) (g_s^{(i)}/(\pi a_i d X_i^d)).$$

Thus modulo the differential operators we have

$$g_s - g_s^{(0)} \equiv - \sum_{i=1}^{n} \left(X_i \frac{\partial}{\partial X_i} + \pi X_i Y \frac{\partial h}{\partial X_i} \right) (g_s^{(i)}/(\pi a_i d X_i^d)). \tag{7}$$

The righthand side is a polynomial in $\pi^{-1}\mathbb{Z}_q[\pi][X, Y]$ of degree in X at most $s-1$. (Recall $d^{-1} \in \mathbb{Z}_p$ and $\deg(h) \leq d-1$.) Thus we have reduced the leading term g_s in g, and can proceed recursively.

In our application we need to reduce the polynomial $\pi h X^v$. This has degree at most $n(d-1) + (d-1) = (n+1)(d-1)$ in X. Moreover, the degree in Y is zero, and this is increased by one on each reduction step. At most $(n+1)(d-1) - (d-1) = n(d-1)$ steps are required. Thus the reduced polynomial is a sum of basis elements $\{X^u \mid u \in \mathcal{B}\}$ with coefficients polynomials over $\mathbb{Q}_q[\pi]$ of degree in Y at most $n(d-1)$. With regard to p-adic orders, each time a factor π^{-1} is introduced in the reduction the degree drops by d. Thus we can introduce at most $(n+1)(d-1)/d < n+1$ powers of π^{-1}. The polynomial $\pi h X^v$ has order $1/(p-1)$. So each coefficient polynomial in the reduced expression has p-adic order

$$> -\frac{n}{p-1}.$$

Here as usual the p-adic order of a polynomial is defined to be the minimum order among its coefficients.

5.2 Solving linear differential systems and matrix inversion

In this section we describe routines for two of the key steps of the algorithm: Solving the linear differential system, and inverting the matrix of power series. Throughout this section R will denote a ring of characteristic zero, and Y an indeterminate. In our application, R will be the ring $\mathbb{Q}_q[\pi]$. Also, the matrix size m below will be $(d-1)^n$.

5.2.1 Linear differential systems

Denote by S the non-commutative ring of $m \times m$ matrices over the ring R. Let $S[Y]$ and $S[[Y]]$ denote the ring of polynomial and power series, respectively, over S. Here the variable Y commutes with matrices. We identify $m \times m$ matrices with entries in $R[Y]$ or $R[[Y]]$, respectively, with elements in the ring $S[Y]$ or $S[[Y]]$, respectively. These identifications define ring isomorphisms. Let $B(Y)$ be an $m \times m$ matrix over the polynomial ring $R[Y]$, i.e., B is an element in the ring $S[Y]$. We wish to find an $m \times m$ matrix $C(Y)$ over $R[[Y]]$ (an element in $S[[Y]]$) such that

$$\frac{dC(Y)}{dY} = C(Y)B(Y),\; C(Y) \equiv I \bmod (Y). \tag{8}$$

Here I is the $m \times m$ identity matrix. Also, the differential operator d/dY acts entry wise on the matrix $C(Y)$ of power series in the usual way. Expanding $C \in S[[Y]]$ and $B \in S[Y]$ in powers of Y, we need to solve

$$\frac{d}{dY}\left(\sum_{i=0}^{\infty} C_i Y^i\right) = \left(\sum_{j=0}^{\infty} C_j Y^j\right)\left(\sum_{i=0}^{s} B_i Y^i\right),\; C_0 = I.$$

Here s is the degree of the matrix polynomial B. Thus we just need to solve

$$\sum_{i=1}^{\infty} iC_i Y^{i-1} = \sum_{k=0}^{\infty}\left(\sum_{i=0}^{\min(s,k)} C_{k-i}B_i\right)Y^k,\; C_0 = I.$$

Equating coefficients of powers of $Y^{\ell-1}$, for $\ell \geq 1$, on both sides we get

$$\begin{array}{rcl} 1C_1 & = & C_0B_0 \\ 2C_2 & = & C_1B_0 + C_0B_1 \\ \vdots & & \vdots \\ (\ell > s)\; \ell C_\ell & = & C_{\ell-1}B_0 + C_{\ell-2}B_1 + \ldots + C_{\ell-(s+1)}B_s \\ \vdots & & \vdots \end{array}$$

Starting from $C_0 = I$ we can compute C_i for $i \leq \ell$ using these recurrences. This requires less than $(s+1)\ell$ matrix ring additions and multiplications, plus ℓ divisions by non-zero elements in R. That is, a total of $\mathcal{O}(m^\omega(s+1)\ell)$ operations in R, where $2 \leq \omega \leq 3$ is the time exponent in the multiplication algorithm used for $m \times m$ matrices (later we shall just take $\omega = 3$).

Note 14 In the above we have computed a basis of solutions to the linear differential system which converge on some small disk around the origin. When $B(Y)$ has finite poles but is regular at the origin, as in the general case which arises in [25], the same method works provided one first expands $B(Y)$ itself around the origin giving a matrix of power series in $\mathbb{Q}_q[\pi][[Y]]$. However, in this general case the necessary matrix $\alpha(Y)$ no longer overconverges around the origin, and one final step is needed which does not arise in the present context.

5.2.2 Inverting matrices

In this section we consider the problem of inverting a matrix $C(Y)$ with entries which are power series in the ring $R[[Y]]$ subject to the condition $C(Y) \equiv I \bmod (Y)$. To invert the matrix $C(Y)$ modulo some power (Y^ℓ) we use Newton iteration with quadratic convergence in the ring $S[[Y]]$. (This works even over non-commutative rings.) Recall that $C(Y) \equiv I \bmod (Y)$. Thus defining

$D_0 = I$ we have $CD_0, D_0C \equiv I \bmod (Y^{2^0})$. Now assume that we have constructed a matrix D_k with $CD_k, D_kC \equiv I \bmod (Y^{2^k})$. Define

$$D_{k+1} = 2D_k - D_kCD_k \bmod (Y^{2^{k+1}}).$$

Then one has that $CD_{k+1}, D_{k+1}C \equiv I \bmod (Y^{2^{k+1}})$. The computation of D_{k+1} requires two multiplications of matrices over $R[[Y]]/(Y^{2^{k+1}})$, plus two additions. This requires $\mathcal{O}(m^\omega \mathrm{M}(2^{k+1}))$ operations in R, where $\mathrm{M}(.)$ is the complexity of polynomial arithmetic over S, and ω the exponent of matrix multiplication over R. Thus to find $D = D_{\lceil \log_2(\ell) \rceil}$ so that $CD \equiv I \bmod (Y^\ell)$ requires $\mathcal{O}(m^\omega \mathrm{M}(\ell)\log(\ell))$ ring operations.

6 The Frobenius matrix for the diagonal form

6.1 Proof of the proposition

In this section we prove Proposition 10 which gives an explicit formula for the absolute Frobenius matrix of a diagonal form. This matrix is defined to have (u, v)th entry the coefficient of X^u in the reduction modulo the differential operators $D_{i,0} (1 \leq i \leq n)$ of

$$\psi_p \circ F(X, 0)(X^v).$$

(See Section A.2.) Here $F(X, 0) = \prod_{i=1}^n \theta(a_i X_i^d)$ where $\theta(z) = \exp(\pi(z - z^p)) = \sum_{m=0}^\infty \lambda_m z^m$. Also ψ_p acts on power series as

$$\psi_p(\sum_r A_r X^r) = \sum_r \tau^{-1}(A_{pr}) X^r$$

where the sum here is over n-tuples of non-negative integers. Writing all this out, we need to reduce the power series

$$\prod_{i=1}^n \sum_{m \geq 0,\, p | dm+v_i} \lambda_m \tau^{-1}(a_i^m) X_i^{(dm+v_i)/p}.$$

Here we use the fact $\psi_p(\prod_{i=1}^n B_i(X_i)) = \prod_{i=1}^n \psi_p(B_i(X_i))$ for $B_i(X_i) \in \mathbb{Q}_q[\pi][[X_i]]$. Now the operators

$$D_{i,0} = X_i \frac{\partial}{\partial X_i} + \pi d a_i X_i^d$$

have the property $D_i(a(X_i))D_j(b(X_j)) = D_i(a(X_i)D_j(b(X_j)))$ for a and b power series in X_i and X_j respectively. It follows that it is enough to determine the coefficient of $X_i^{u_i}$ in the reduction modulo $D_{i,0}$ of each summation

$$\sum_{m \geq 0,\, p | dm+v_i} \lambda_m \tau^{-1}(a_i^m) X_i^{(dm+v_i)/p}$$

and take the product for $i = 1, \ldots, n$. One checks directly that

$$X_i^{u_i + dr} \equiv \pi^{-r}(-1)^r (u_i/d)_r a_i^{-r} X_i^{u_i} \bmod D_{i,0}.$$

Hence for $(dm + v_i)/p = u_i + dr$ we have that

$$\lambda_m \tau^{-1}(a_i^m) X_i^{(dm+v_i)/p} \equiv \lambda_m \tau^{-1}(a_i^m) \pi^{-r}(-1)^r (u_i/d)_r a_i^{-r} X_i^{u_i} \bmod D_{i,0}.$$

Thus the coefficient of X^{u_i} in the reduction of the ith univariate power series modulo $D_{i,0}$ is

$$\sum_{m,r\geq 0,\, dm+v_i=p(u_i+dr)} \lambda_m \tau^{-1}(a_i^m)\pi^{-r}(-1)^r(u_i/d)_r a_i^{-r}.$$

Equation (5) now follows.

Note 15 Conceptually, we have a tensor product decomposition of the cohomology space which is respected by the absolute Frobenius map. Concretely, this just means the absolute Frobenius matrix can be expressed as the Kronecker product of n matrices obtained from the one-dimensional case.

6.2 Computing the undeformed matrix

The matrix $\alpha(0)$ is computed directly from Equation (5). We need to understand the decay of the terms in these sums so that we can determine upper limits on the indices m and r when computing $\alpha(0)$ to some finite p-adic accuracy.

We have to compute all entries $\alpha_{u,v}(0)$ in the matrix, as u and v ranges over a set of size $(d-1)^n$. We examine the complexity of computing a single entry modulo $p^{\tilde{N}}$. From [13, Pages 55-57] we have

$$\mathrm{ord}(\lambda_m) \geq \frac{(p-1)}{p^2}m. \tag{9}$$

Thus

$$\mathrm{ord}(\lambda_m \pi^{-r}(u_i/d)_r) \geq m\frac{p-1}{p^2} - S_r \tag{10}$$

where S_r is the sum of the digits in the p-adic representation of r. (Here we use the fact $\mathrm{ord}((u_i/d)_r) \geq \mathrm{ord}(u_i!)$, from Clark [6, Page 265, Case 3].) So for each $1 \leq i \leq n$, it is enough to compute the inner sum

$$\sum_{m,r\geq 0} \lambda_m (u_i/d)_r (-1)^r \pi^{-r}\tau^{-1}(a_i^m)a_i^{-r}$$

over all $m, r \geq 0$ with $pu_i - v_i = d(m - pr)$ such that

$$m\frac{p-1}{p^2} - S_r \geq \tilde{N}.$$

Now $S_r \leq (p-1)(\log_p(r)+1)$, and $r \leq m/p$ (see the proof of Lemma 20). Hence $S_r \leq (p-1)\log_p(m)$. Thus it suffices to sum over m for $m \leq p^2 x$ where

$$x - \log_p(x) = \frac{\tilde{N}}{p-1} + 2.$$

Thus taking $m \leq 2p^2\tilde{N}/(p-1)$ will do. For m in this range, there are $\mathcal{O}(\tilde{N})$ possible pairs (m, r) with the required property. This gives $\mathcal{O}(\tilde{N})$ entries in the sum. We shall return to this in the complexity analysis of this step in Section 9.

7 p-Adic estimates on the decay of power series

When working with power series in Y we truncate modulo Y^{N_Y} for some suitably chosen N_Y. Also, the coefficients are p-adic numbers truncated "modulo" $p^{\tilde{N}}$. We need to justify that these truncations do not comprise the correctness of our final answer. To this end, in this section we obtained lower bounds on the p-adic orders of elements which occur in Algorithm 13. These bounds will be used in the error analysis in Section 8.

7.1 Overconvergence of the generic absolute Frobenius matrix

We first need to understand the decay of the power series which occur in the matrix $\alpha(Y)$. To do this, we examine the effect of applying the reduction formulae to a monomial. Let $m = (m_1, \ldots, m_n) \in \mathbb{Z}^n_{\geq 0}$. We have that $X^m X_i^d$ is congruent modulo D_{i,Y^p} to

$$\begin{aligned}
&-(\pi d a_i)^{-1} m_i X^m - Y^p (d a_i)^{-1} X_i \frac{\partial h}{\partial X_i} X^m \\
&= \pi^{-1} * X^m + Y^p \sum_{i,\, |m| < |i| < |m| + d} * X^{m+i}.
\end{aligned}$$

Here the $*$ indicate p-adic numbers with $\mathrm{ord}(*) \geq 0$. So reducing a monomial X^u to a linear combination of the basis set can introduce at most $\lfloor |u|/d \rfloor$ powers of π^{-1}, and at most $|u| - (d-1)$ powers of Y^p. Thus for $a_u(Y) \in \mathbb{Q}_q[\pi][Y]$ the coefficients of the basis elements in the reduction of a term $a_u(Y) X^u$ have p-adic order

$$\geq \mathrm{ord}(a_u(Y)) - \lfloor |u|/d \rfloor \frac{1}{p-1} \tag{11}$$

and degree in Y

$$\leq \deg_Y(a_u(Y)) + (|u| - d + 1)p. \tag{12}$$

Here by the p-adic order of $a_u(Y)$ we mean the minimum order among the coefficients of non-zero terms. In the proof of the next proposition we shall use the following simple lemma.

Lemma 16 *Let $\epsilon \geq 0$ be a rational number. Consider power series of the form $\sum_m a_m(Y) X^m$ where the sum is over non-negative integer vectors and $a_m(Y) \in \mathbb{Q}_q[\pi][Y]$. The subset of all such power series with $\deg_Y(a_m(Y)) \leq \epsilon |m|$ is a ring. Similarly, the subset of all such power series with $\mathrm{ord}(a_m(Y)) \geq \epsilon |m|$ is a ring. In particular, both sets are closed under multiplication.*

Proof: The first claim follows from the properties $\deg_Y(a_m(Y) b_m(Y)) = \deg_Y(a_m(Y)) + \deg_Y(b_m(Y))$ and $\deg_Y(a_m(Y) + b_m(Y)) \leq \max\{\deg_Y(a_m(Y)), \deg_Y(b_m(Y))\}$ for $a_m, b_m \in \mathbb{Q}_q[\pi][Y]$. The second claim follows by similar properties of the p-adic order map on polynomials. □

Proposition 17 *Let $\alpha(Y) = (\alpha_{u,v}(Y))$ where $\alpha_{u,v}(Y) \in \mathbb{Q}_q[\pi][[Y]]$ for $u, v \in \mathcal{B}$. Write $\alpha_{u,v}(Y) = \sum_{i=0}^{\infty} \alpha_{u,v}^{(i)} Y^i$. Then*

$$\mathrm{ord}(\alpha_{u,v}^{(i)}) \geq \frac{1}{12pd} i - \frac{2n}{9}.$$

Thus the entries in $\alpha(Y)$ have p-adic order at least $-2n/9 > -n$. Let $M \geq 1$ be an integer (we shall choose a specific value in Section 8.2). Then "modulo" p^M the matrix $\alpha(Y)$ contains polynomials in Y of p-adic order greater than $-n$ and of degree less than $12pd(M+n)$.

Proof: The map $\alpha(Y)$ is defined in Section A.2 and we use some of the notation from that section. Write $F(X,Y) = \sum_m G_m(Y)X^m$ where the sum is over non-negative integer vectors. We wish to understand the degrees and p-adic orders of the polynomials $G_m(Y)$. By definition we have

$$F = \prod_{i=1}^{n} \theta(a_i X_i^d) \prod_{0<|j|<d} \theta(Y b_j X^j).$$

Here $\bar{h} = \sum_j \bar{b}_j X^j$, and $h = \sum_j b_j X^j$ with b_j the Teichmüller lifting of $\bar{b}_j$. (Recall that we assume h has no constant term, i.e., "$b_j = 0$" for j the all zero vector. Without this assumption, the factor $\theta(Y b_0 X^0)$ actually causes no harm since it can be "moved through" all the operators.) Recall also that $\theta(z) = \exp(\pi(z - z^p))$. Writing $\theta(Y b_j X^j) = \sum_m a_{j,m}(Y)X^m$ with the sum over non-negative integer vectors, we have that

$$\deg_Y(a_{j,m}) \leq \lfloor |m|/|j| \rfloor \leq |m|.$$

Hence writing $\prod_{0<|j|<d} \theta(Y b_j X^j) = \sum_m c_m(Y) X^m$ by Lemma 16 we have

$$\deg_Y(c_m) \leq |m|.$$

Certainly writing $\prod_{i=1}^n \theta(a_i X_i^d) = \sum_m d_m(Y) X^m$ we have $\deg_Y(d_m) = 0$. Thus by Lemma 16

$$\deg_Y(G_m) \leq |m|. \tag{13}$$

Writing $\theta(z) = \sum_{r=0}^{\infty} \lambda_r z^r$ we have the estimate [13, Pages 55-57]

$$\mathrm{ord}(\lambda_r) \geq \frac{p-1}{p^2} r.$$

Thus the p-adic order of each $a_{j,m}(Y)$ is at least

$$\frac{(p-1)}{p^2} . \frac{|m|}{|j|}.$$

Similarly, the coefficients in $\prod_{i=1}^n \theta(a_i X_i^d)$ satisfy the bound

$$\mathrm{ord}(d_m(Y)) \geq \frac{(p-1)}{p^2} . \frac{|m|}{d}.$$

By Lemma 16 this gives a lower bound of

$$\mathrm{ord}(G_m) \geq \frac{(p-1)|m|}{dp^2}. \tag{14}$$

The vth column of the matrix $\alpha(Y)$ has uth entry the coefficient of X^u in the reduction modulo $\sum_{i=1}^n D_{i,Y^p}$ of $\psi_p(FX^v)$. Now $FX^v = \sum_m G_{m-v} X^m$. Hence $\psi_p(FX^v) = \sum_m \tau^{-1}(G_{pm-v})X^m$. We need to understand the reduction of each term $G_{pm-v}X^m$ in this series (the action of τ^{-1} is inconsequential since it fixes Y and does not change p-adic estimates). With regard to the degree, by estimates (12) and (13) the coefficient of each basis element X^u in the reduction of this term has degree in Y at most

$$\deg_Y(G_{pm-v}) + (|m| - d + 1)p \leq |pm - v| + (|m| - d + 1)p.$$

By estimates (11) and (14) the p-adic order of the coefficient of each basis element in the reduction is at least

$$\operatorname{ord}(G_{pm-v}) - \frac{|m|}{d(p-1)} \geq \frac{(p-1)|pm-v|}{dp^2} - \frac{|m|}{d(p-1)}.$$

Now since $v \in \mathcal{B}$ we have $n \leq |v| \leq n(d-1)$ with all the entries in v positive. Thus $p|m| - n(d-1) \leq |pm-v| \leq p|m| - n$. For the degree upper bound we can therefore take the estimate $2p|m|$. For the p-adic order lower bound we have

$$\geq \frac{|m|}{d}\left(\frac{p-1}{p} - \frac{1}{p-1}\right) - \frac{n(d-1)(p-1)}{dp^2}. \tag{15}$$

We take the lower bound of $-2n/9$ for the second term, and $1/6$ for the coefficient of $|m|/d$ in the first (recall that $p \geq 3$). Thus (15) is

$$\geq \frac{|m|}{6d} - \frac{2n}{9}. \tag{16}$$

The coefficient of X^u in the reduction of $G_{pm-v}X^m$ is a polynomial in Y. By the degree estimate $\leq 2p|m|$ and the lower bound (16) its p-adic Newton polygon lies on or above the graph

$$y = \frac{1}{12pd}x - \frac{2n}{9}.$$

Thus the coefficient of X^u in the reduction of $\sum_m G_{pm-v}X^m$ is a power series in Y with Newton polygon lying on or above this graph. This proves the first claim and the second follows immediately. We also now see that "modulo" p^M the matrix $\alpha(Y)$ contains polynomials in Y of degree at most

$$12pd(M + (2n/9)) < 12pd(M+n).$$

Here by "modulo" p^M we mean that we truncate the p-adic expansions of the entries in $\alpha(Y)$ after the power p^{M-1} (see the start of Section 8.2). The proof is complete. □

7.2 Lower bounds on p-adic order of intermediate results

We also need lower bounds on the p-adic order of the entries in the matrices $B(Y)$, $C(Y) \bmod Y^{N_Y}$, $\alpha(0)$, and the $\log_p(q)$th "power" of $\alpha(y^{\tau^{-1}})$ which occurs in the equation in Step 8. (We do not believe that these bounds are optimal. In particular, Dwork's work in [15, Section 5(d)] suggests that it may be possible to get a logarithmic bound on the growth of the coefficients of the matrix power series $C(Y)$.)

Lemma 18 *The polynomial entries in $B(Y)$ have p-adic order at least $-(n-1)/(p-1)$.*

Proof: This was proved at the end of Section 5.1 □

Lemma 19 *The polynomial entries in $C(Y) \bmod Y^{N_Y}$ have p-adic order at least $-nN_Y/(p-1)$.*

Proof: Let $\ell > 0$. The matrix $\ell! C_\ell$ is equal to a sum of elements each one of which is a product of at most $\ell - 1$ coefficients of polynomials in $B(Y)$. Since by Lemma 18 these coefficients have order $\geq -(n-1)/(p-1)$, the matrix $\ell! C_\ell$ contains entries of order at least $-(n-1)(\ell-1)/(p-1)$. Now $\operatorname{ord}(\ell!) < \ell/(p-1)$, and so C_ℓ itself has entries of order at least $-(n-1)\ell/(p-1)$. Since $C(Y) \bmod Y^{N_Y} = \sum_{\ell=0}^{N_Y-1} C_\ell Y^\ell$ the lemma follows. □

Lemma 20 *The entries in the matrix $\alpha(0)$ have p-adic order at least $-n(p+1)$.*

Proof: We first need a lower bound on

$$m\frac{(p-1)}{p^2} - S_r \tag{17}$$

where $m, r \geq 0$ are integers, with S_r the sum of the p-adic digits in r, and $d(m-pr) = pu_i - v_i$. Now $pu_i - v_i \geq p.1 - (d-1) = p - d + 1$ and so $m - pr \geq (p/d) - 1 + (1/d) > (p/d) - 1$. Therefore $r < (m/p) - (1/d) + (1/p)$ and so since r is an integer we must have $r \leq (m/p)$. So we find (17) is at least

$$r\frac{(p-1)}{p} - S_r \geq (p-1)\left(\frac{r}{p} - (\log_p(r) + 1)\right).$$

For $r \geq 3p$ this is non-negative. For $r < 3p$ the minimum $-S_r$ can take is $-(p+1)$. The lemma now follows from Equation (5) and inequality (10). □

Lemma 21 *The entries in the matrix $\alpha(y^{\tau^{-1}})$ have p-adic order at least $-n$. Those in the product $\alpha(y^{\tau^{-1}})^{\tau^{r\log_p(q)-1}} \ldots \alpha(y^{\tau^{-1}})$ at least $-nr\log_p(q)$.*

Proof: The first claim follows from Proposition 17 and the second from the first, using the fact $\mathrm{ord}(y) = 0$ and that the map τ does not change p-adic orders. □

8 Error analysis

In this section we use the bounds from Section 7 to show that the final output is correct.

8.1 Choice of final p-adic accuracy

We first justify the final p-adic accuracy modulo p^N to which the zeta function is computed. Write

$$L(\bar{f}(X,\bar{y}),T)^{(-1)^n} = \exp\left(\sum_{k=1}^{\infty}(-1)^n S_k \frac{T^k}{k}\right) = \sum_{s=0}^{\infty} m_s T^s.$$

Fix an embedding $\mathbb{Q}(\zeta_p) \to \mathbb{C}$, and let $||.||$ denote the complex absolute value. Since $||(-1)^n S_k|| \leq q^{rnk}$, and $\exp(\sum_{k=1}^{\infty} q^{rnk}T^k/k) = \sum_{s=0}^{\infty} q^{rns}T^s$, we see that $||m_s|| \leq q^{rns}$. (Recall that $\bar{y} \in \mathbb{F}_{q^r}$.) Thus $L(\bar{f}(X,\bar{y}),T)^{(-1)^n}$ converges in the complex plane for all specialisations $T = t$ with $||t|| < q^{-rn}$. In particular, the poles of the $L(\bar{f}(X,\bar{y}),T)^{(-1)^n}$ must have complex absolute value $\geq q^{-rn}$. Thus the reciprocal zeros of $L(\bar{f}(X,\bar{y}))^{(-1)^{n+1}}$ itself must have complex absolute value $\leq q^{rn}$. Let θ be in the Galois group of $\mathbb{Q}(\zeta_p)/\mathbb{Q}$. Then the reciprocal zeros of $\theta(L(\bar{f}(X,\bar{y}),T)^{(-1)^{n+1}})$ must also have complex absolute value $\leq q^{rn}$. Hence the polynomial $P(T)^{(-1)^{n+1}}$ has reciprocal zeros of complex absolute value $\leq q^{rn}$. (In fact, by Deligne's theorem [10] the reciprocal zeros of $P(T)^{(-1)^{n+1}}$ in the zeta function of the smooth affine hypersurface $Z^p - Z = \bar{f}(X,\bar{y})$ over $\mathbb{F}_{q^r}$ have complex absolute value equal to $q^{rn/2}$. However, we prefer to avoid the use of this deep result, as it only improves the performance of the algorithm by a factor four.) Moreover, this polynomial has degree exactly $(p-1)(d-1)^n$. Thus its coefficients, which we must determine, have absolute value at most

$$q^{rn(p-1)(d-1)^n} 2^{(p-1)(d-1)^n}.$$

Thus taking

$$N = \lceil (p-1)(d-1)^n((rn\log_p(q)) + \log_p(2))\rceil + 1$$

is sufficient. More precisely, the unknown integer polynomial which occurs in the zeta function can be recovered uniquely given its residue class modulo p^N. (This is the polynomial $P(T)$ described in Step 9 of the Algorithm 13.) Here the extra 1 accounts for the possibility that the coefficients may be negative integers.

8.2 Reverse analysis of error propagation

Even though the final answer is a polynomial with integer coefficients, the calculations followed in finding this involve matrices with possibly p-adically non-integral entries. We showed in Section 8.1 that a final "absolute error" of order p^N, in the p-adic sense, was sufficient to recover the zeta function exactly. At each step of the algorithm, the absolute error can increase by an amount depending upon the p-adic order of the matrices with which we are computing. In this section we analyse the "propagation of errors" through the algorithm. We determine an overall "p-adic accuracy" $p^{\tilde{N}}$ to which one can compute throughout the algorithm and be sure that the final error is of the correct magnitude. We first need a simple lemma.

Let $c = \sum_{i=m}^{\infty} c_i p^i \in \mathbb{Q}_q[\pi]$ where $m \in \mathbb{Z}$ and each

$$c_i = \sum_{j=0}^{p-2} c_{ij}\pi^j$$

for some $c_{ij} \in \{0, 1, \ldots, p-1\}$. Let $M \geq 1$ be an integer. We shall say that we know c modulo p^M (or that c has been computed modulo p^M) provided all the coefficients c_i for $m \leq i < M$ have been explicitly determined. The phrase "working modulo p^M" means that we do not compute the coefficients c_i for $i \geq M$. We use the same language in an obvious way for matrices over $\mathbb{Q}_q[\pi]$, $\mathbb{Q}_q[\pi, y]$ or $\mathbb{Q}_p[\pi][[Y]]$ mod (Y^{N_Y}).

Lemma 22 *Let R be either $\mathbb{Q}_p[\pi]$ or $\mathbb{Q}_p[\pi, y]$. Let A, B be matrices over R with entries of p-adic order at least $-o_A \leq 0$ and $-o_B \leq 0$ respectively. To compute the matrix AB modulo p^M for some $M \geq 1$, it is sufficient to know the matrix A modulo p^{M+o_B} and B modulo p^{M+o_A}. To compute the matrix $A+B$ modulo p^M for some $M \geq 1$ it is sufficient to know the matrices A and B modulo p^M.*

Proof: Each entry in AB is a sum of elements of the form ab where a and b are entries in A and B respectively. It is sufficient to know all such products modulo p^M. Write $a = \sum_{i=-o_A}^{\infty} a_i p^i$ and $b = \sum_{j=-o_B}^{\infty} b_j p^j$. Then we need to know $a_i b_j$ for all i, j with $i + j < M$. Since $i \geq -o_A$ and $j \geq -o_B$, for $i + j < M$ we must have $j < M + o_A$ and $i < M + o_B$. The first part now follows, and the claim on addition is straightforward. □

The above lemma says that absolute errors magnify when multiplying non-integral matrices, but addition does not increase the absolute error. Our approach will be to work backwards through the algorithm calculating the error magnification at each step.

Our starting point is the observation that the polynomial $L(\bar{f}(X, \bar{y}), T)^{(-1)^{n+1}}$ from Step 8 contains p-adic integral entries. In fact, $L(\bar{f}(X, \bar{y}), T)$ is the L-function of the exponential sum, which is a polynomial (or reciprocal polynomial) over the cyclotomic ring $\mathbb{Z}[\zeta_p]$ (embedded in $\mathbb{Z}_p[\pi]$).

We need to compute $L(\bar{f}(X,\bar{y}),T)$ modulo p^N. Write $L(\bar{f}(X,\bar{y}),T)^{(-1)^{n+1}} = \det(1-*T)$ where $*$ is as shown in Step 8. By the second part of Lemma 21, the entries in $*$ have order $\geq -nr\log_p(q)$. By the simple argument in the proof of Lemma 22, the determinant expansion of characteristic polynomials, and Lemma 21, it is enough to compute $*$ modulo p^{N_1} where

$$N_1 = N + ((d-1)^n - 1)rn\log_p(q).$$

Now $*$ is obtained by essentially raising $\alpha(y^{\tau^{-1}})$ to the "power" $r\log_p(q)$. Thus to determine $*$ modulo p^{N_1} it is enough by Lemma 22 and the first part of Lemma 21 to know $\alpha(y^{\tau^{-1}})$ modulo p^{N_2} where

$$N_2 = N_1 + nr\log_p(q)(= N + (d-1)^n rn\log_p(q)).$$

The matrix $\alpha(y^{\tau^{-1}})$ is obtained by evaluating $\alpha(Y)$ at an integral point, and so we need to find $\alpha(Y)$ modulo p^{N_2}. Putting $M = N_2$ in Proposition 17 we see that $\alpha(Y)$ modulo p^{N_2} is a polynomial of degree less than

$$N_Y = 12pd(N_2+n).$$

Thus in Steps 3, 4, 6, and 7, it suffices to work modulo (Y^{N_Y}), as coefficients of any higher powers of Y cannot contribute to $\alpha(Y)$ modulo p^{N_2}.

We wish to compute $\alpha(Y)$ modulo p^{N_2}. Write $C(Y) = \sum_{\ell=0}^{\infty} C_\ell Y^\ell$. We know (see the proof of Lemma 19) that

$$\mathrm{ord}(C_\ell) \geq -\frac{n\ell}{p-1}. \tag{18}$$

The set of all matrices $\sum_{\ell=0}^{\infty} M_\ell Y^\ell$ over $\mathbb{Q}_q[\pi][[Y]]$ which satisfy the bound $\mathrm{ord}(M_\ell) \geq -n\ell/(p-1)$ is a ring. Moreover, the inverse matrix $C(Y)^{-1}$ is obtained from $C(Y)$ by performing addition and multiplication in this ring. Hence writing $C(Y^p)^{-1} = \sum_{\ell=0}^{\infty} \tilde{C}_\ell Y^\ell$ we have that

$$\mathrm{ord}(\tilde{C}_\ell) \geq -\frac{n\ell}{p(p-1)}. \tag{19}$$

Also

$$\alpha(0)C(Y)^{\tau^{-1}} = \sum_{\ell=0}^{\infty} \alpha(0)C_\ell^{\tau^{-1}} Y^\ell$$

and $\mathrm{ord}(\alpha(0)C_\ell^{\tau^{-1}}) \geq \mathrm{ord}(\alpha(0)) + \mathrm{ord}(C_\ell)$. From this inequality, along with Lemma 20 and inequalities (18), (19) we get

$$\mathrm{ord}(\alpha(0)C^{\tau^{-1}}(Y) \bmod (Y^{N_Y})) \geq -\frac{nN_Y}{p-1} - n(p+1),\ \mathrm{ord}(C(Y^p)^{-1} \bmod (Y^{N_Y})) \geq -\frac{nN_Y}{(p-1)p}.$$

Thus by Lemma 22 we see that it is enough to know $C(Y)$ and $C(Y^p)^{-1}$ modulo p^{N_3} where

$$N_3 := N_2 + \lfloor nN_Y/(p-1)\rfloor + n(p+1).$$

Also, we need $\alpha(0)$ modulo p to the power $N_2 + \lfloor nN_Y/(p-1)\rfloor + \lfloor nN_Y/p(p-1)\rfloor$. By the proof of Lemma 20, the elements $\lambda_m \pi^{-r}(u_i/d)_r$ have order at least $-(p+1)$. (See the entry for Step 5 in Section 9 for how we compute $\alpha(0)$.) We need $\alpha(0)$ modulo p to the power $N_2 + \lfloor nN_Y/(p-1)\rfloor + \lfloor nN_Y/p(p-1)\rfloor$. Thus it suffices in Step 5 to compute the elements a_i^{-1} and $\tau^{-1}(a_i)$ modulo p to the power

$$N_2 + \lfloor nN_Y/(p-1)\rfloor + \lfloor nN_Y/p(p-1)\rfloor + (p+1) < N_4 < N_6$$

where N_4 and N_6 are defined below.

Next, we examine the loss of accuracy during the computation of $C(Y^p)^{-1}$ from $C(Y)$. We work backwards through the Newton iteration relations when finding $C(Y)^{-1}$. We need to know $C(Y)^{-1}$ modulo $(p^{N_3}, Y^{\lfloor N_Y/p \rfloor})$. Now $C(Y)^{-1}$ modulo these powers is by definition D_m where $m = \lceil \log_2(N_Y/p) \rceil$. For $k \geq 0$, the matrix D_{k+1} is computed as

$$D_{k+1} = 2D_k - D_k C D_k \bmod Y^{2^{k+1}}$$

with the coefficients taken modulo some power of p. Also $D_0 = I$. By inequality (18) and the comments following it, one sees $D_k \bmod (Y^{2^{k+1}})$ has p-adic order at least $-n2^k/(p-1)$, and $C \bmod (Y^{2^{k+1}})$ has p-adic order at least $-n2^{k+1}/(p-1)$. Thus to know D_{k+1} modulo p^M, for some $M \geq 1$, by Lemma 22 it is enough to know D_k and C modulo $p^{M + \lfloor 2(n2^k/(p-1)) \rfloor}$. We require D_m modulo p^{N_3}, and so by induction one sees we need $D_0 = I$ and C modulo p^M for M the floor of

$$N_3 + 2n(2^{m-1} + \ldots + 2 + 1)/(p-1) = N_3 + (n(2^{m+1} - 2)/(p-1)).$$

Thus computing $C(Y) \bmod (Y^{N_Y})$ with coefficients modulo p^{N_4} where

$$N_4 := N_3 + \lfloor 4nN_Y/p(p-1) \rfloor$$

is sufficient to find $C(Y)^{-1} \bmod (p^{N_3}, Y^{\lfloor N_Y/p \rfloor})$, and therefore $C(Y^p)^{-1} \bmod (p^{N_3}, Y^{N_Y})$.

Next, we need to determine the p-adic accuracy required for the matrix $B(Y)$ to compute $C(Y) \bmod (Y^{N_Y})$ with coefficients modulo p^{N_4}. We need to find C_ℓ for $0 \leq \ell < N_Y$ with coefficients modulo p^{N_4}. By Lemmas 18 and 22, we require $B(Y)$ modulo p^{N_5} where

$$N_5 := N_4 + \lfloor nN_Y/(p-1) \rfloor.$$

Finally, we determine the p-adic accuracy required in Step 1. Computing $B(Y)$ from the coefficients of $h(X)$ requires division by a power of π at most π^n. Thus we need to compute the coefficients of $h(X)$ modulo p^{N_6} where

$$N_6 := N_5 + \lfloor n/(p-1) \rfloor.$$

For simplicity we shall work to a common accuracy modulo $p^{\tilde{N}}$ throughout Steps 1-9 of the algorithm, for some $\tilde{N} \geq N_6$. Specifically, define

$$\tilde{N} = (60nd + 1)(N + (d-1)^n rn \log_p(q)) + 60n^2 d + n(p+2).$$

Observe also that

$$N_Y = 12pd(N_2 + n) = 12pd(N + (d-1)^n rn \log_p(q) + n).$$

We have shown that working to these accuracies is enough to determine the coefficients of the zeta function modulo p^N, and hence recover the zeta function exactly.

Note 23 The error analysis in this section was presented in a somewhat intuitive fashion. In the context of real and complex analysis, the analysis of the propagation of errors in a numerical algorithm is a well-studied and essential topic. In the p-adic setting, provided one is working solely with p-adic integral elements in $\mathbb{C}_p$, one can simply perform all computations in an appropriate subring of $\mathbb{R}_p/(p^M)$ for some suitable positive integer M. Thus the analysis of errors becomes trivial, as there is no error magnification. When working over $\mathbb{C}_p$ itself, error magnification can occur, although the non-Archimedean nature of $\mathbb{C}_p$ surely makes it easier to handle. However, the author does not know of any systematic development of this topic.

9 Complexity analysis

We count the number of bit operations required in each of the steps of the algorithm. Recall ω is the exponent of deterministic matrix multiplication over arbitrary rings, see [19, Section 12.1]. We use fast polynomial multiplication over matrix rings to get a time of $\mathrm{M}(\ell) = \tilde{\mathcal{O}}(\ell)$ [5]. The calculations involve addition and multiplication in the p-adic rings $\mathbb{Q}_q[\pi]$ and $\mathbb{Q}_q[\pi, y]$ "modulo" $p^{\tilde{N}}$ during Steps 2-9 of the algorithm. The numbers we manipulate have explicitly lower bounds on their p-adic orders, as given in Proposition 17 and Lemmas 18, 19, 20 and 21. In particular, the p-adic order is certainly always $\geq -\tilde{N}$. Addition and multiplication of such numbers in $\mathbb{Q}_q[\pi]$ and $\mathbb{Q}_q[\pi, y]$ "modulo" $p^{\tilde{N}}$ have the same complexity as in the rings $\mathbb{Z}_q[\pi]/(p^{\tilde{N}})$ and $\mathbb{Z}_q[\pi, y]/(p^{\tilde{N}})$, respectively, up to a constant factor. We take a Soft-Oh linear time bound for multiplication, division by units and addition in p-adic rings, see [19, Theorem 8.23]. Thus computing in $\mathbb{Q}_q[\pi]$ and $\mathbb{Q}_q[\pi, y]$ "modulo" $p^{\tilde{N}}$ requires $\tilde{\mathcal{O}}(\tilde{N} p \log(q))$ and $\tilde{\mathcal{O}}(\tilde{N} p \log(q) r)$ bits of time/space, respectively.

Step 1: The Teichmüller liftings lie in the ring $\mathbb{Z}_q$ and they must be computed modulo $p^{\tilde{N}}$. The lifting of each coefficient requires $\tilde{\mathcal{O}}(\log(q))$ operations in the ring $\mathbb{Z}_q/(p^{\tilde{N}})$. Thus Step 1 needs $\tilde{\mathcal{O}}(d^n \log(q))$ operations in $\mathbb{Z}_q/(p^{\tilde{N}})$.

Step 2: Reducing each polynomial $\pi h X^v$ requires $n(d-1)$ applications of the reduction formula (7). Computation of the formula itself is dominated by the multiplication of $X_i \partial h / \partial X_i$ by a homogeneous polynomial of total degree at most $(n+1)(d-1) - d$ with coefficients in $\mathbb{Z}_p[\pi][Y]$ of degree in Y at most $n(d-1)$. The former polynomial has $\mathcal{O}(d^n)$ terms, and the latter homogeneous one $\mathcal{O}((nd)^{n-1})$ terms. Thus this multiplication can be done in

$$\mathcal{O}(d^n \times (nd)^{n-1} \times nd) = \mathcal{O}(n^n d^{2n})$$

operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$, using the naive multiplication algorithm for polynomials. We require $(d-1)^n = \mathcal{O}(d^n)$ reductions, giving a total estimate of $\mathcal{O}(n^n d^{3n})$ operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$ for this step.

Step 3: The degree in Y of the matrix polynomial $B(Y)$ is at most $n(d-1)$. The complexity estimates in Section 5.2.1 show we require

$$\mathcal{O}(d^{\omega n} \times dn \times N_Y) = \mathcal{O}(n d^{\omega n + 1} N_Y)$$

operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$ to compute C_ℓ for $\ell < N_Y$.

Step 4: The estimates in Section 5.2.2 give a complexity of

$$\tilde{\mathcal{O}}(d^{\omega n} N_Y / p) = \tilde{\mathcal{O}}(d^{\omega n} N_Y)$$

operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$.

Step 5: We return to the discussion in Section 6.2 on computing $\alpha_{u,v}(0)$ modulo $p^{\tilde{N}}$. The values $\pi^{-r}(u_i/d)_r$ and λ_m for the necessary pairs (m, r) can be found in $\mathcal{O}(\tilde{N})$ operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$. Observe that $\tau^{-1}(a_i^m) a_i^{-r} = \tau^{-1}(a_i)^m (a_i^{-1})^r$. The elements $\tau^{-1}(a_i)$ and

a_i^{-1} for $1 \leq i \leq n$ can be precomputed in $\tilde{\mathcal{O}}(n\log(q))$ operations in $\mathbb{Z}_q/(p^{\tilde{N}})$. Notice that $\tilde{\mathcal{O}}(n\log(q)) = \tilde{\mathcal{O}}(\tilde{N})$. (Here we use the method from [23, Section 5], to compute τ^{-1} on an element in $\tilde{\mathcal{O}}(\log(q)) = \tilde{\mathcal{O}}(\tilde{N})$ operations in $\mathbb{Z}_q/(p^{\tilde{N}})$.) The product $\tau^{-1}(a_i^m)a_i^{-r}$ can now be found in $\mathcal{O}(\log(m)+\log(r))$ operations in $\mathbb{Z}_q/(p^{\tilde{N}})$. Thus $\tau^{-1}(a_i^m)a_i^{-r}$ for all required pairs (m,r) can be found in $\tilde{\mathcal{O}}(\tilde{N}(\log(\tilde{m})+\log(r))) = \tilde{\mathcal{O}}(\tilde{N})$ operations in $\mathbb{Z}_q/(p^{\tilde{N}})$. Summing the elements takes $\mathcal{O}(\tilde{N})$ operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$. Thus we get $\tilde{\mathcal{O}}(\tilde{N})$ operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$ for each inner sum. This gives $\tilde{\mathcal{O}}(n\tilde{N})$ ring operations for each $\alpha_{u,v}(0)$ modulo $p^{\tilde{N}}$. This gives a total complexity for computing $\alpha(0)$ of $\tilde{\mathcal{O}}(d^{2n}n\tilde{N})$ operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$.

Step 6: One must compute $C(Y)^{\tau^{-1}}$ from $C(Y)$ "modulo" $(p^{\tilde{N}}, Y^{N_Y})$. Computing τ^{-1} on a single element of $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$ requires $\tilde{\mathcal{O}}(\log(q))$ ring operations. Thus one sees that doing this directly would increase the dependence of the algorithm on $\log(q)$ from third to fourth power. As such, we compute $C(Y)^{\tau^{-1}}$ by first computing $B(Y)^{\tau^{-1}}$, and then recovering $C(Y)^{\tau^{-1}}$ by directly solving the differential equation "τ^{-1} of Equation (4)". One checks all this is absorbed in the previous estimates. Computing $\alpha(Y)$ from $C(Y^p)^{-1}$, $\alpha(0)$ and $C(Y)^{\tau^{-1}}$ requires

$$\tilde{\mathcal{O}}(d^{\omega n}N_Y)$$

operations in $\mathbb{Q}_q[\pi]$ "modulo" $p^{\tilde{N}}$.

Step 7: The element y is in the ring $\mathbb{Z}_q[y]$, and it can be found modulo $p^{\tilde{N}}$ in $\tilde{\mathcal{O}}(r\log(q))$ operations in $\mathbb{Z}_q[y]/(p^{\tilde{N}})$ by taking the Teichmüller lifting of $\bar{y}$. The specialisation requires $\mathcal{O}(d^{2n})$ evaluations at $Y = y$ of polynomials of degree N_Y over $\mathbb{Z}_q[\pi]$. Thus the evaluation of the matrix requires $\mathcal{O}(d^{2n}N_Y)$ operations in $\mathbb{Q}_q[\pi, y]$ "modulo" $p^{\tilde{N}}$.

Step 8: Write $L(\bar{f}(X,\bar{y}),T)^{(-1)^{n+1}} = \det(1-*T)$. To compute $*$ we use the fast exponentiation method for semi-linear maps from [26, Lemma 32]. This takes $\tilde{\mathcal{O}}(d^{\omega n})$ operations in $\mathbb{Q}_q[\pi, y]$ "modulo" $p^{\tilde{N}}$. The fastest algorithm for computing characteristic polynomials seems to be that based on Hessenberg form [7, Algorithm 2.2.9]. However, it requires divisions by p-adic numbers which are generated during the algorithm. To circumvent questions of numerical stability, we wish to use an algorithm for computing characteristic polynomials which avoids division by unquantifiably small p-adic numbers. We will compute the polynomial via the identity

$$\det(1-*T) = \exp\left(-\sum_{k=1}^{\infty}\mathrm{Tr}(*^k)\frac{T^k}{k}\right) \equiv \prod_{k=1}^{(d-1)^n}\sum_{0\leq\ell\leq(d-1)^n/k}\frac{(-\mathrm{Tr}(*^k)T^k)^\ell}{\ell!k^\ell} \bmod T^{(d-1)^n+1}. \quad (20)$$

By Lemma 21, each factor in this finite product is of the form $\sum_{\ell=0}^{(d-1)^n} m_\ell T^\ell$ where

$$\mathrm{ord}(m_\ell) \geq -\ell\{\log_p((d-1)^n) + \frac{1}{p-1} + nr\log_p(q)\}. \quad (21)$$

We truncate p-adic numbers modulo p to the power $N+\epsilon$, where

$$\epsilon = (d-1)^n\{\log_p((d-1)^n) + \frac{1}{p-1} + nr\log_p(q)\},$$

during the computation of $\det(1 - *T)$ from $*$ using identity (20). Inequality (21) shows that this does not compromise the final answer modulo p^N. (Note that we know $*$ correctly to the accuracy of p^{N_2} and $N_2 < N + \epsilon \, (< \tilde{N})$. Using the usual determinant expansion for characteristic polynomials and Lemma 21, one sees that this actually determines $\det(1 - *T)$ modulo p^N. However, during the computation of $\det(1 - *T)$ using the above method we need to increase the accuracy a little more. Essentially, the higher p-adic coefficients in $*$ do not contribute to the final answer, although the computation itself may apparently introduce some higher coefficients of which we must keep track.) The above method takes $\mathcal{O}(d^{(\omega+1)n})$ operations in $\mathbb{Q}_q[\pi, y]$ "modulo" $p^{\tilde{N}}$.

STEP 9: Finally, each coefficient in the conjugate $\theta(L(\bar{f}(X, \bar{y}), T))$ can be found using [26, Eqn (11)] in $p-1$ operations in $\mathbb{Z}_p[\pi]/(p^N)$, and the product found in $\tilde{\mathcal{O}}(pd^n)$ operations in $\mathbb{Z}_p[\pi]/(p^N)$. Note that the polynomial $P(T)^{(-1)^{n+1}}$ with integer coefficients in the restricted range exists by the coefficient bounds in Section 8.1.

Now $N = \mathcal{O}(rnpd^n \log(q))$, $\tilde{N} = \mathcal{O}(rn^2pd^{n+1} \log(q))$ and $N_Y = \mathcal{O}(rnp^2d^{n+1} \log(q))$. Substituting these values, and using our estimates for computations in the rings $\mathbb{Q}_q[\pi]$ and $\mathbb{Q}_q[\pi, y]$ "modulo" powers of p, we get a total time complexity for Algorithm 13 of

$$\tilde{\mathcal{O}}(d^{(\omega+2)n+3} n^{\max(5,n+2)} p^4 \log(q)^3 r^3) \tag{22}$$

bit operations. The space complexity in bits of the algorithm is

$$\tilde{\mathcal{O}}(d^{4n+2} n^{n+2} p^4 \log(q)^3 r^2). \tag{23}$$

This is estimated in a straightforward manner, the only subtlety being that the dependence on r is quadratic: the both parts of Step 7 take cubic time in r but require less space.

10 Completion of the proofs

We now complete the proofs of the results in the introduction. Propositions 10, 11 and 12 show that Algorithm 13 would output correctly if it were performed on p-adic numbers with infinite precision. The error analysis in Section 8 shows that the finite accuracy chosen is enough to ensure the answer is exactly correct. This proves the correctness of Algorithm 13. Theorem 1 now follows from the complexity estimates (22) and (23), taking $\omega = 3$ (one could take $\omega = 2.4$ [19, Page 330]).

Corollaries 2 and 3 are immediate. The proof of Corollary 4 is as follows. First note that the finite number of primes p with $p = 2$ or p divides some a_i can be dealt with in constant time. Thus we assume $p > 2$ and $a_i \neq 0 \bmod p$ for $1 \leq i \leq n$. We run Algorithm 13 with $N = n + 2$, and so $p^N = p^{n+2} > p^{n+1}$. Thus we find modulo p^N the number M, say, of points on $Z^p - Z = f(X) \bmod p$. Now if (z, x) is a point on this hypersurface modulo p, then since $z^p - z = 0$ we have $f(x) = 0$. Conversely, if $f(x) = 0$ then (z, x) is a point on $Z^p - Z = f(X)$ for each $z = 0, 1, \ldots, p-1$. Thus the number of solutions to $f(X) = 0 \bmod p$ is exactly M/p. Since $M \leq p^{n+1}$ and we know it modulo p^{n+2}, we can recover M/p exactly. Checking the complexity bounds one sees that the time required is just $\tilde{\mathcal{O}}(p^2)$ bit operations. (We lose a factor p from both N and N_Y in the estimates above, which reduces the dependence on p from quartic to quadratic.)

A Appendix

In this appendix we develop a generic theory for families of exponential sums of the type considered in the paper, and the corresponding deformation theory. We do not believe that this has explicitly appeared in the literature before. However, in [15, Sections 5(a),(b),(c)] Dwork develops such theories in his "dual spaces" for smooth projective hypersurfaces. At the start of [15, Section 5(f)] he writes down the necessary deformation equation (the first commutative diagram) in the more familiar Dwork space. He states that this can be arrived at "by duality" from the equation in the dual space, but does not give a direct proof. Similarly, Sperber states such a theory for families of exponential sums defined by homogeneous equations on [32, Pages 291-292]. This is very similar to Dwork's theory, and Sperber derives his results from those of Dwork. The difference in our work is that we consider non-homogeneous equations, and give an explicit proof of our deformation equation and the differential equation satisfied by the deformation matrix without appealing to the more complicated dual theory.

A.1 Generic spaces

Let $p > 2$ and q be a power of p. Let R denote the ring of power series in Y over the unramified extension $\mathbb{Q}_q[\pi]$ of $\mathbb{Q}_p[\pi]$ of degree $\log_p(q)$ which converge on some closed disk containing the origin. Precisely, elements in R have an expansion of the form $\sum_{i=0}^{\infty} a_i Y^i$ where $a_i \in \mathbb{Q}_q[\pi]$ with $\mathrm{ord}(a_i) - \epsilon i \to \infty$ for some real number ϵ (not necessarily positive). For $b > 0$ and c real numbers, let $L(b, c)$ denote the $\mathbb{Q}_q[\pi]$-module of power series over $\mathbb{Q}_q[\pi]$ of the form

$$\sum_{m \in \mathbb{Z}_{\geq 0}^n} a_m X^m, \ \mathrm{ord}(a_m) \geq b|m| + c.$$

Let $L(b) = \cup_c L(b, c)$, a ring. Define $L^o(b) = L(b) \cap X_1 \ldots X_n \mathbb{Q}_q[\pi][[X]]$, the set of power series in $L(b)$ in which every term is divisible by $X_1 \ldots X_n$. For $1 \leq i \leq n$ define $L^{(i)}(b) = L(b) \cap X_1 \ldots X_{i-1} X_{i+1} \ldots X_n \mathbb{Q}_q[\pi][[X]]$, the set of power series in $L(b)$ in which every term is divisible by all X_k for $k \neq i$.

Let L_Y denote the ring of power series in the variable Y of the form

$$\sum_{j=0}^{\infty} a_j(X) Y^j$$

with $a_j(X) \in L((p-1)/pd, c_j)$, for some real number c_j such that there exists a real number ϵ with

$$c_j - \epsilon j \to \infty \tag{24}$$

as $j \to \infty$. As before, we do not assume there is a uniform lower bound on ϵ over all elements in L_Y. Let $L_Y^o \subset L_Y$ be the submodule of power series $\sum_{j=0}^{\infty} a_j(X) Y^j$ such that each $a_j(X) \in X_1 \ldots X_n \mathbb{Q}_q[\pi][[X]]$. For $1 \leq i \leq n$ let $L_Y^{(i)} \subset L_Y$ be those power series with each $a_j(X) \in X_1 \ldots X_{i-1} X_{i+1} \ldots X_n \mathbb{Q}_q[\pi][[X]]$.

The condition on the c_j just ensures that any series $a(X, Y) \in L_Y$ will converge to an element of $L((p-1)/pd)$ on substitution of $Y = y$ for sufficiently small y, i.e., $\mathrm{ord}(y) \geq -\epsilon$. This is useful in the proof of Proposition 25. We view L_Y, L_Y^o and $L_Y^{(i)}$ as modules over R. It is also sometimes helpful to write elements in L_Y in the form $\sum_{m\in\mathbb{Z}^n_{\geq 0}} a_m(Y)X^m$ where $a_m(Y)\in R$. Note that not all such formal power series lie in L_Y, only those satisfying certain decay conditions on the coefficients.

Recall that

$$f(X,Y)=\sum_{i=1}^n a_iX_i^d+Yh(X).$$

Let

$$D_{i,0}=\exp(-\pi\sum_{i=1}^n a_iX_i^d)\circ X_i\frac{\partial}{\partial X_i}\circ\exp(\pi\sum_{i=1}^n a_iX_i^d)=X_i\frac{\partial}{\partial X_i}+\pi da_iX_i^d$$

and

$$D_{i,Y}=\exp(-\pi Yh)\circ D_{i,0}\circ\exp(\pi Yh)=X_i\frac{\partial}{\partial X_i}+\pi(da_iX_i^d+YX_i\frac{\partial h}{\partial X_i}) \tag{25}$$

act on $\mathbb{Q}_q[\pi][[X,Y]]$, for $1\leq i\leq n$. Let D_{i,Y^p} be as $D_{i,Y}$ with Y replaced by Y^p.

Lemma 24 *Each of the three operators $D_{i,0}$, $D_{i,Y}$ and D_{i,Y^p} maps the space $L_Y^{(i)}$ to the space L_Y^o.*

Proof: We first claim that L_Y is stable under these three maps. We have $\pi da_iX_i^d$, $\pi X_i\partial h/\partial X_i\in L((p-1)/pd,1/(p-1)-(p-1)/p)$. Hence L_Y is stable under multiplication by $\pi da_iX_i^d$, $\pi YX_i\partial h/\partial X_i$ and $\pi Y^pX_i\partial h/\partial X_i$. Certainly L_Y is stable under $X_i\partial/\partial X_i$. Since L_Y is also closed under addition the claim follows. That each map sends elements in $L_Y^{(i)}$ to those in L_Y^o is now easily seen. □

Proposition 25 *The quotient R-modules*

$$L_Y^o/\sum_{i=1}^n D_{i,Y}(L_Y^{(i)}),\ L_Y^o/\sum_{i=1}^n D_{i,Y^p}(L_Y^{(i)}),\ L_Y^o/\sum_{i=1}^n D_{i,0}(L_Y^{(i)})$$

are free R-modules spanned by the set

$$\{X^u\,|\,u=(u_1,\dots,u_n),\ 0<u_i<d\}.$$

We define ξ_u to be the basis element X^u. We will use the shorthand notation $L_Y^o/\sum_{i=1}^n D_{i,Y}$ for $L_Y^o/\sum_{i=1}^n D_{i,Y}(L_Y^{(i)})$, and likewise for the other two modules.

Proof: That this set spans the quotient spaces may be seen by considering simple reduction relations. For example, consider the operators $D_{i,Y}$. Applying this operator to a monomial X^m where $m=(m_1,\dots,m_n)\in\mathbb{Z}^n_{\geq 0}$ we find

$$X^mX_i^d\equiv-(da_i)^{-1}(\pi^{-1}m_iX^m+X_i(\partial h/\partial X_i)YX^m)\bmod D_{i,Y}(L_Y).$$

(Note that when reducing monomials $X^mX_i^d$ in L_Y^o we have $m_j>0$ for $j\neq i$. Therefore $X^m\in L_Y^{(i)}$ and $D_{i,Y}(X^m)\in L_Y^o$.) Iterating relations of this kind, and using $\mathbb{Q}_q[\pi][Y]$-linearity of the operators $D_{i,Y}$, we find that any term Y^rX^m with $m\in\mathbb{Z}^n_{>0}$ can be written as a linear combination of the basis set with coefficients which are polynomials (of degree at most $|m|-d+1+r$ and lowest term degree r) in Y. Also, the polynomials have p-adic order at least $-|m|/(p-1)d$. The decay conditions on the coefficients of the monomials X^m ensure that this process "converges" as $|m|$ tend to infinity, i.e., the sum $Y^r\sum_{m\in\mathbb{Z}^n_{>0}}a_mX^m$, where $a_m\in\mathbb{Q}_q[\pi]$, can be reduced to a linear combination of the basis elements with coefficients power series in Y divisible by Y^r. (These coefficient power series themselves are convergent on the disk $\mathrm{ord}(Y) > 1/(p-1)d - (p-1)/pd$.) Summing over r we find that any element in L_Y^o can be reduced to a R-linear combination of the basis elements modulo $\sum_{i=1}^n D_{i,Y}(L_Y^{(i)})$. Here we use condition (24) to ensure the coefficients, a priori in $\mathbb{Q}_q[\pi][[Y]]$, converge on some closed disk around the origin. The argument for D_{i,Y^p} is exactly the same with Y^p replacing Y, and that for $D_{i,0}$ similar but simpler.

To show these quotient modules are free is more subtle. For $D_{i,0}$ it follows from the fact that $L^o((p-1)/pd)/\sum_{i=1}^n D_{i,0}(L^{(i)}((p-1)/pd))$ is free c.f. [32, Lemma 2.28]. The operators involving Y are dealt with by a specialisation argument, and an application of the results in [32, Section 2(a)]. Suppose for example that $\sum_u a_u(Y)X^u \equiv 0 \bmod (\sum_{i=1}^n D_{i,Y}(L_Y^{(i)}))$. That is

$$\sum_u a_u(Y)X^u = \sum_{i=1}^n D_{i,Y}(b_i(X,Y))$$

where X^u runs over a subset of the spanning set, and $a_u(Y) \in R - \{0\}$ with $b_i \in L_Y^{(i)}$. Choose $y \in \mathbb{Q}_q$ with $\mathrm{ord}(y) > (p-1)/p - 1/(p-1)$ small enough that each $a_u(Y)$ converges at $Y = y$ to a non-zero element of $\mathbb{Q}_q[\pi]$, and each b_i converges at $Y = y$ to an element of $L^{(i)}((p-1)/pd)$. Then $\sum_u a_u(y)X^u = \sum_{i=1}^n D_{i,y}(b_i(X,y))$ with $a_u(y) \in \mathbb{Q}_q[\pi] - \{0\}$ and $b_i(X,y) \in L^{(i)}((p-1)/pd)$. The approach of [32, Section 2(a)] shows that no such linear relation can exist. (Precisely, modify [32, Lemma 2.28] by setting as usual (see [32, Eqn (2.11)]) "$E_i = X_i(\partial/\partial X_i)$" and "$H_i = da_iX_i^d$" but now taking "$\Lambda = yh(X)$". Then Λ is small enough p-adically that it does not affect any of the proofs. Note also that in [32] Sperber uses a faster decaying "splitting function", and so works in $L(p/(p-1)d)$ rather than $L((p-1)/pd)$. This allows one to include the prime $p = 2$, which we exclude.) $\square$

A.2 Frobenius maps and deformations

Recall that τ is the map on $\mathbb{Q}_q[\pi]$ which reduces to the pth power map on its residue field and fixes π. Extend τ to act on $\mathbb{Q}_q[\pi][[Y]]$ by fixing Y (and being linear and continuous under the Y-adic norm). Let the operator ψ_p act on the space of formal power series $\mathbb{Q}_q[\pi][[X,Y]]$ in the following way. For a monomial X^m where $m = (m_1, \ldots, m_n)$, the image $\psi_p(X^m)$ is $X_1^{m_1/p} \ldots X_n^{m_n/p}$ when p divides each m_i, and zero otherwise. Also ψ_p is τ^{-1}-linear over $\mathbb{Q}_q[\pi][[Y]]$ (and continuous under the X-adic norm); in particular, ψ_p fixes Y. Explicitly,

$$\psi_p : \sum_{r=0}^{\infty}(\sum_{m\in\mathbb{Z}_{\geq 0}^n} a_mX^m)Y^r \mapsto \sum_{r=0}^{\infty}(\sum_{m\in\mathbb{Z}_{\geq 0}^n} \tau^{-1}(a_{pm})X^m)Y^r,\ a_m \in \mathbb{Q}_q[\pi].$$

Define $\alpha_Y : \mathbb{Q}_q[\pi][[X,Y]] \to \mathbb{Q}_q[\pi][[X,Y]]$ by

$$\alpha_Y = \exp(-\pi f(X,Y^p)) \circ \psi_p \circ \exp(\pi f(X,Y)) = \psi_p \circ F(X,Y) \tag{26}$$

where

$$F(X,Y) = \exp(\pi(f(X,Y) - f^\tau(X^p,Y^p))) = \prod_{i=1}^n \theta(a_iX_i^d) \prod_{0\leq|j|<d} \theta(Yb_jX^j).$$

Here τ fixes the variable Y, and in f^τ acts only on the coefficients of f.

Lemma 26 *The space L_Y^o is stable under the map α_Y.*

Proof: By the decay on the coefficients of $\theta(z) = \exp(\pi(z - z^p))$ from Equation (9), each $\theta(a_i X_i^d) \in L((p-1)/p^2 d, 0)$. For each j with $0 \le |j| < d$ write $\theta(Y b_j X^j) = \sum_{r=0}^{\infty} (b_j X^j)^r \lambda_r Y^r$. Here $\theta(z) = \sum_{r=0}^{\infty} \lambda_r z^r$. Now $\lambda_r (b_j X^j)^r \in L((p-1)/p^2 |j|, 0) \subset L((p-1)/p^2 d, 0)$. Since the latter space is a ring we find $F = \sum_{r=0}^{\infty} F_r Y^r$ where $F_r \in L((p-1)/p^2 d, 0)$. It follows easily that multiplication by F maps elements in L_Y^o to elements $\sum_{j=0}^{\infty} a_j(X) Y^j$ where each $a_j(X) \in L^o((p-1)/p^2 d, c_j)$ for some c_j satisfying condition (24). We also have that $\psi_p : L^o((p-1)/p^2 d, c) \to L^o((p-1)/pd, c)$ for any real c. The result now follows. □

Lemma 27 *The action $\alpha_Y : L_Y^o \to L_Y^o$ induces a map*

$$\alpha_Y : L_Y^o / \sum_{i=1}^{n} D_{i,Y} \to L_Y^o / \sum_{i=1}^{n} D_{i,Y^p}.$$

Proof: This follows from the factorisations of the operators α_Y and $D_{i,Y}, D_{i,Y^p}$. Specifically, from the first equality in (26), along with the factorisations

$$D_{i,Y} = \exp(-\pi f(X,Y)) \circ X_i \frac{\partial}{\partial X_i} \circ \exp(\pi f(X,Y))$$

$$D_{i,Y^p} = \exp(-\pi f(X,Y^p)) \circ X_i \frac{\partial}{\partial X_i} \circ \exp(\pi f(X,Y^p)),$$

which follow from (25), we see that $\alpha_Y \circ D_{i,Y} = D_{i,Y^p} \circ p\alpha_Y$. (Here we use the equality $\psi_p \circ X_i(\partial/\partial X_i) = X_i(\partial/\partial X_i) \circ p\psi_p$.) Therefore α_Y maps an element in $D_{i,Y}(L_Y^{(i)})$ to one in $D_{i,Y^p}(L_Y^{(i)})$, and so maps $\sum_{i=1}^{n} D_{i,Y}(L_Y^{(i)})$ to $\sum_{i=1}^{n} D_{i,Y^p}(L_Y^{(i)})$. □

Define

$$\alpha_0 = \exp(-\pi f(X,0)) \circ \psi_p \circ \exp(\pi f(X,0)) = \psi_p \circ F(X,0), \tag{27}$$

a map on $\mathbb{Q}_q[\pi][[X,Y]]$. Then L_Y^o is stable under α_0, and α_0 induces a map

$$\alpha_0 : L_Y^o / \sum_{i=1}^{n} D_{i,0} \to L_Y^o / \sum_{i=1}^{n} D_{i,0}.$$

(These facts are proved in the same manner as Lemmas 26 and 27.) Let $T_{Y,0}$ denote the bijective map "multiplication by $\exp(\pi Y h)$" from $\mathbb{Q}_q[\pi][[X,Y]]$ to itself (the inverse is multiplication by $\exp(-\pi Y h)$).

Lemma 28 *The space L_Y^o is stable under the map $T_{Y,0}$. Moreover, $T_{Y,0}$ is a bijection on L_Y^o.*

Proof: We have $\exp(\pi Y h) \in L_Y$, and L_Y^o is stable under multiplication by elements of L_Y. This proves the first claim. For the second claim, the inverse is again multiplication by $\exp(-\pi Y h)$. □

The map $T_{Y,0}$ induces a bijection

$$T_{Y,0} : L_Y^o / \sum_{i=1}^{n} D_{i,Y} \to L_Y^o / \sum_{i=1}^{n} D_{i,0}.$$

On sees this by considering the factorisations of the operators $D_{i,Y}$ and $D_{i,0}$. Define $T_{Y^p,0}$ to be the map "multiplication by $\exp(\pi Y^p h)$" on L_Y^o. Then $T_{Y^p,0}$ also induces a bijection, as in the righthand vertical arrow in the next diagram.

Proposition 29 *The following diagram commutes*

$$\begin{array}{ccc} L_Y^o/\sum_{i=1}^n D_{i,Y} & \xrightarrow{\alpha_Y} & L_Y^o/\sum_{i=1}^n D_{i,Y^p} \\ \downarrow T_{Y,0} & & \downarrow T_{Y^p,0} \\ L_Y^o/\sum_{i=1}^n D_{i,0} & \xrightarrow{\alpha_0} & L_Y^o/\sum_{i=1}^n D_{i,0}. \end{array}$$

Thus we have

$$\alpha_Y = T_{Y^p,0}^{-1} \circ \alpha_0 \circ T_{Y,0}. \tag{28}$$

Proof: The diagram certainly commutes when α_Y, α_0, $T_{Y,0}$ and $T_{Y^p,0}$ are viewed as acting on L_Y^o, rather than the factor spaces, i.e., $T_{Y^p,0} \circ \alpha_Y = \alpha_0 \circ T_{Y,0}$ on L_Y^o. This relation descends to the induced maps. The equation now follows since $T_{Y^p,0}$ is bijective on L_Y^o, and hence also on the factor spaces. (Specifically, the inverse is the map induced by multiplication by $\exp(-\pi Y^p h)$.) $\square$

Now let $C(Y)$ denote the matrix for the map multiplication by $\exp(\pi Y h)$ from $L_Y^o/\sum_{i=1}^n D_{i,Y}$ to $L_Y^o/\sum_{i=1}^n D_{i,0}$ with respect to the basis of monomials of the two spaces. Write $\alpha(Y)$ and $\alpha(0)$ for the matrices of the maps α_Y and α_0, respectively. (Our matrix convention is that the entry in the uth row and vth column, the (u,v)th entry, of the matrix for a map gives the coefficient of X^u in the image of X^v under the map.)

Lemma 30 *The matrix of the map $T_{Y^p,0}^{-1} : L_Y^o/\sum_{i=1}^n D_{i,0} \to L_Y^o/\sum_{i=1}^n D_{i,Y^p}$ is $C(Y^p)^{-1}$.*

Proof: We show that the matrix for $T_{Y^p,0}$ itself is $C(Y^p)$, and the lemma follows. The (u,v)th entry $C_{u,v}(Y)$ in $C(Y)$ is the coefficient of X^u in the reduction of $\exp(\pi Y h)X^v$ modulo $\sum_{i=1}^n D_{i,0}(L_Y^{(i)})$. Let $c_{u,v}(Y)$ be the coefficient of X^u in the reduction of $\exp(\pi Y^p h)X^v$ modulo $\sum_{i=1}^n D_{i,0}(L_Y^{(i)})$. We need to show $c_{u,v}(Y) = C_{u,v}(Y^p)$. Write $\exp(\pi Y h)X^v = \sum_{j=0}^\infty a_j(X)Y^j$, and let $a_j^{(u)} \in \mathbb{Q}_q[\pi]$ be the coefficient of X^u in the reduction of $a_j(X)$ modulo $\sum_{i=1}^n D_{i,0}(L^{(i)}((p-1)/pd))$. Then $C_{u,v}(Y) = \sum_{j=0}^\infty a_j^{(u)} Y^j$, and $c_{u,v}(Y) = \sum_{j=0}^\infty a_j^{(u)} Y^{pj}$, as required. $\square$

From (28) and Lemma 30 we get

$$\alpha(Y) = C(Y^p)^{-1}\alpha(0)C(Y)^{\tau^{-1}}. \tag{29}$$

Here the τ^{-1} arises since α_0 is τ^{-1} linear.

The significance of the matrix $\alpha(Y)$ from the point of view of L-functions is that evaluating this matrix at $\tau^{-1}(y)$ for some Teichmüller point y gives the semi-linear Frobenius matrix. Specifically, let $y \in \mathbb{R}_p$ with $y^{q^r} = y$ and y in an unramified extension of $\mathbb{Z}_q$. The quotient $\mathbb{Q}_q[\pi, y]$-module

$$L^o((p-1)/pd) \otimes \mathbb{Q}_q[\pi, y] / \sum_{i=1}^n D_{i,y}(L^{(i)}((p-1)/pd) \otimes \mathbb{Q}_q[\pi, y]) \tag{30}$$

is free on the set $\{\xi_u\}$ c.f. [32, Theorem 2.17]. Here $D_{i,y}$ is just $D_{i,Y}$ with the variable Y replaced by the Teichmüller point y. Let $\alpha_y = \psi_p \circ F(X, y)$ with $F(X, Y)$ exactly as above. As before α_y

induces a map on the quotient space (30). Let (α_y) be the matrix for this map with respect to the basis of monomials. (Note that we now have two maps α_0: one as in (27) and that defined immediately above. The former acts on the R-module spanned by the basis set, and the latter on the $\mathbb{Q}_q[\pi]$-module. However, the matrices for these two maps are the same.) We have that (α_y) is equal to $\alpha(Y)$ evaluated at $\tau^{-1}(y)$, as can be seen by examining the action of α_y and α_Y and the operators $D_{i,y}$ and "D_{i,Y^p} evaluated at $Y = \tau^{-1}(y)$". Now $\alpha_y^{r\log_p(q)}$ is the Frobenius map on Dwork's cohomology space, and writing $L(\bar{f}(X,\bar{y}),T)$ for the L-function of the exponential sum associated with $\bar{f}(X,\bar{y})$ (see [32, Page 277]) we have [32, Eqn (2.35)] that

$$L(\bar{f}(X,\bar{y})) = \det(1 - T\alpha_y^{r\log_p(q)})^{(-1)^{n+1}}.$$

Because of τ^{-1}-linearity, the matrix for the linear map $\alpha_y^{r\log_p(q)}$ is equal to

$$(\alpha_y)(\alpha_y)^{\tau^{-1}}\dots(\alpha_y)^{\tau^{-r\log_p(q)+1}}.$$

(Note that $\tau^{-i} = \tau^{r\log_p(q)-i}$ on $\mathbb{Q}_q[\pi,y]$, which gives the precise formulation in Proposition 12.) Finally, a standard argument on Artin-Schreier equations (see for example [27, Section 3]) shows that the zeta function of the curve $Z^p - Z = \bar{f}(X,\bar{y})$ is the product of the Galois conjugates of the L-function $L(\bar{f}(X,\bar{y}))$ times $1/(1-q^{rn}T)$. These results together prove Proposition 12.

A.3 The differential equation satisfied by the deformation matrix

In this section we prove that the matrix $C(Y)$ is the unique solution of a certain differential equation, with an initial condition.

We choose a basis element ξ_v and look at the action of $T_{Y,0}$ on this element. Write

$$\exp(\pi Y h)\xi_v = \sum_u C_{u,v}\xi_u + \sum_{i=1}^n D_{i,0}(\eta_{i,v}). \tag{31}$$

Here $C_{u,v} \in R$, $\eta_{i,v} \in L_Y^{(i)}$ and the final sum lies in L_Y^o. Since

$$\frac{\partial}{\partial Y} \circ D_{i,0} = D_{i,0} \circ \frac{\partial}{\partial Y}$$

applying $\partial/\partial Y$ to both sides of (31) we get

$$\pi h \exp(\pi Y h)\xi_v = \sum_u \frac{\partial C_{u,v}}{\partial Y}\xi_u + \sum_{i=1}^n D_{i,0}\left(\frac{\partial \eta_{i,v}}{\partial Y}\right). \tag{32}$$

Now write

$$\pi h\xi_v = \sum_u B_{u,v}\xi_u + \sum_{i=1}^n D_{i,Y}(\epsilon_{i,v}). \tag{33}$$

Here $B_{u,v} \in \mathbb{Q}_q[\pi][Y]$, $\epsilon_{i,v} \in L_Y^{(i)}$ and the final sum lies in L_Y^o. Substituting (33) in (32) we have

$$\exp(\pi Y h)\left\{\sum_u B_{u,v}\xi_u + \sum_{i=1}^n D_{i,Y}(\epsilon_{i,v})\right\} = \sum_u \frac{\partial C_{u,v}}{\partial Y}\xi_u + \sum_{i=1}^n D_{i,0}\left(\frac{\partial \eta_{i,v}}{\partial Y}\right).$$

Hence

$$\exp(\pi Y h)\sum_u B_{u,v}\xi_u + \sum_{i=1}^n D_{i,0}(\exp(\pi Y h)\epsilon_{i,v}) = \sum_u \frac{\partial C_{u,v}}{\partial Y}\xi_u + \sum_{i=1}^n D_{i,0}\left(\frac{\partial \eta_{i,v}}{\partial Y}\right)$$

where we have used the fact $D_{i,Y} = \exp(-\pi Y h)\circ D_{i,0}\circ \exp(\pi Y h)$. Putting $\exp(\pi Y h)$ in the first term inside the summation and substituting (31) with ξ_u replacing ξ_v on the lefthand side we get

$$\sum_u B_{u,v}\left\{\sum_w C_{w,u}\xi_w + \sum_{i=1}^n D_{i,0}(\eta_{i,u})\right\} + \sum_{i=1}^n D_{i,0}(\exp(\pi Y h)\epsilon_{i,v})$$
$$= \sum_u \frac{\partial C_{u,v}}{\partial Y}\xi_u + \sum_{i=1}^n D_{i,0}\left(\frac{\partial \eta_{i,v}}{\partial Y}\right).$$

Since $B_{u,v}$ is a polynomial in Y, with no variables X occuring, we can rewrite this as

$$\sum_u B_{u,v}\sum_w C_{w,u}\xi_w + \sum_{i=1}^n D_{i,0}(\sum_u B_{u,v}\eta_{i,u}) + \sum_{i=1}^n D_{i,0}(\exp(\pi Y h)\epsilon_{i,v})$$
$$= \sum_u \frac{\partial C_{u,v}}{\partial Y}\xi_u + \sum_{i=1}^n D_{i,0}\left(\frac{\partial \eta_{i,v}}{\partial Y}\right).$$

Now switch u for w in the summed variable on the first term on the righthand side and tidy up. We then have that for each basis element ξ_v the following equality holds

$$\sum_w \xi_w\left(\sum_u C_{w,u}B_{u,v}\right) = \sum_w \frac{\partial C_{w,v}}{\partial Y}\xi_w + \sum_{i=1}^n D_{i,0}\left(\frac{\partial \eta_{i,v}}{\partial Y} - \sum_u B_{u,v}\eta_{i,u} - \exp(\pi Y h)\epsilon_{i,v}\right).$$

Note that the ith operand in the final summand lies in $L_Y^{(i)}$, and the sum itself lies in L_Y^o. Writing $B = (B_{k,l})$ and $C = (C_{k,l})$ we find, equating coefficients of the basis elements on each side,

$$CB = \frac{\partial C}{\partial Y}.$$

Also, certainly $C(Y)$ evaluated at $Y = 0$ is just the identity, for then $\exp(\pi Y h) = 1$. Thus the matrix $C(Y)$ satisfies the differential equation (4), and by the explicit method in Section 5.2.1 we see that it is the unique solution of this equation. This fact, combined with Equation (29), proves Proposition 11.

References

[1] L. Adleman and M.D. Huang, Counting rational points on curves and abelian varieties over finite fields, in "*Algorithmic Number Theory II*" (H. Cohen), Lecture Notes in Computer Science 1122, Springer-Verlag, 1996, 1-16.

[2] I. Blake, G. Seroussi, N. Smart, Elliptic Curves in Cryptography, LMS Lecture Note Series 265, Cambridge University Press, 1999.

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

[4] P. Candelas, X. de la Ossa and F. Rodriguez-Villegas, Calabi-Yau manifolds over finite fields I, preprint 2000.

[5] D.G. Cantor and E. Kaltofen, On fast multiplication of polynomials over arbitrary algebras, *Acta Inform.*, 28, 693-701, 1991.

[6] D.N. Clark, A note on the p-adic convergence of solutions of linear differential equations, Proc. Amer. Math. Soc. **17** (1966), 262-269.

[7] H. Cohen, A Course in Computational Number Theory, Graduate Texts in Math. 138, Springer, 1996.

[8] D.A. Cox and S. Katz, Mirror Symmetry and Algebraic Geometry, Math. Surveys and Monographs Vol. 68, American Mathematical Society, 1999.

[9] 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 Notes Series 284, Cambridge University Press, 2001, 45-67.

[10] P. Deligne, La conjecture de Weil II, Pub. Math. IHES No. 52 (1980), 137-252.

[11] J. Denef and F. Vercauteren, An extension of Kedlaya's algorithm to Artin-Schreier curves in characteristic 2, in C. Fieker and D.R. Kohel (eds), ANTS-V, Lecture Notes in Computer Science 2369, Springer-Verlag, (2002), 308-323.

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

[13] B. Dwork, On the zeta function of a hypersurface, Pub. Math. IHES No. 12, (1962), 5-68.

[14] B. Dwork, A deformation theory for the zeta function of a hypersurface, Proc. Internat. Congr. Mathematicians (Stockholm 1962), 247-259.

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

[16] B. Dwork, Normalised period matrices I: plane curves, Ann. Math 94, (1971), 337-388.

[17] 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.

[18] M. Fouquet, P. Gaudry and R. Harley, An extension of Satoh's algorithm and its implementation, J. Ramanujan Math. Soc. **15**, (2000), 281-318.

[19] J. von zur Gathen and J. Gerhard, Modern Computer Algebra, Cambridge University Press, 1999.

[20] P. Gaudry and N. Gürel, An extension of Kedlaya's point-counting algorithm to superelliptic curves, in C. Boyd (ed), Advances in Cryptology - ASIACRYPT 2001, Lecture note in Computer Science 2248, Springer-Verlag, (2001), 480-494.

[21] P. Gaudry and R. Harley, Counting points on hyperelliptic curves over finite fields, in B. Preneel (ed), Advances in Cryptology - EUROCRYPT 2000, Lecture notes in Computer Science 1807, Springer-Verlag, (2000), 19-34.

[22] N.M. Katz, On the differential equations satisfied by period matrices, Pub. Math. IHES 35, (1968), 71-106.

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

[24] A.G.B. Lauder, Computing zeta functions of Kummer curves via multiplicative characters, to appear in Foundations of Computational Mathematics.
Available at: http://web.comlab.ox.ac.uk/oucl/work/alan.lauder/

[25] A.G.B. Lauder, Deformation theory and the computation of zeta functions II, in preparation.

[26] A.G.B. Lauder and D. Wan, Counting points on varieties over finite fields of small characteristic, to appear in the proceedings of the MSRI workshop in Algorithmic Number Theory Aug-Dec 2000.
Available at: http://web.comlab.ox.ac.uk/oucl/work/alan.lauder/

[27] A.G.B. Lauder and D. Wan, Computing zeta functions of Artin-Schreier curves over finite fields, London Mathematical Society, JCM, Vol. 5, 2002, 34-55.

[28] B. Poonen, Computational aspects of curves of genus at least 2, in "*Algorithmic Number Theory II*" (H. Cohen), Lecture Notes in Computer Science 1122, Springer, 1996, 283-306.

[29] T. Satoh, The canonical lift of an ordinary elliptic curve over a finite field and its points counting, J. Ramanujan Math. Soc. 15, (2000), 247-270.

[30] T. Satoh, On p-adic point counting algorithms for elliptic curves over finite fields, in C. Fieker and D.R. Kohel (eds), ANTS-V, Lecture Notes in Computer Science 2369, Springer-Verlag, (2002), 43-66.

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

[32] S. Sperber, On the p-adic theory of exponential sums, Amer. J. Math. 108 (1986), 255-296.

[33] F. Vercauteren, Computing zeta functions of hyperelliptic curves over finite fields of characteristic 2, in M. Yung (ed), Advances in Cryptology - CRYPTO 2002, Lecture Notes in Computer Science 2442, Springer-Verlag, (2002), 369-384.

[34] F. Vercauteren, B. Preneel and J. Vandewalle, A memory efficient version of Satoh's algorithm, in B. Pfitzmann (ed) Advances in Cryptology - EUROCRYPT 2001, Lecture Notes in Computer Science 2045, Springer-Verlag, (2001), 1-13.

[35] D. Wan, Computing zeta functions over finite fields, Contemporary Mathematics, 225 (1999), 131-141.

[36] D. Wan, Dwork's conjecture on unit root zeta functions, Ann. Math. 150 (1999), 867-927.

[37] D. Wan, Algorithmic theory of zeta functions, preprint 2002.
Available at: http://www.math.uci.edu/~dwan/preprint.html