London Mathematical Society ISSN 1461–1570

A RECURSIVE METHOD FOR COMPUTING ZETA FUNCTIONS OF VARIETIES

ALAN G.B. LAUDER

Abstract

We present an algorithm which reduces the problem of calculating a numerical approximation to the action of absolute Frobenius on the middle-dimensional rigid cohomology of a smooth projective variety over a finite field, to that of performing the same calculation for a smooth hyperplane section. When combined with standard geometric techniques, this yields a method for computing zeta functions which proceeds "by induction on the dimension". The "inductive step" combines previous work of the author on the deformation of Frobenius with a higher rank generalisation of Kedlaya's algorithm. The analysis of the loss of precision during the algorithm uses a deep theorem of Christol and Dwork on p-adic solutions to differential systems at regular singular points. We apply our algorithm to compute the zeta functions of certain surfaces which are double covers of the projective plane.

1. *Introduction*

We present a method for calculating the zeta function of a smooth projective variety over a finite field which proceeds by induction on the dimension. Specifically, we outline an algorithm which reduces the problem of calculating a numerical approximation for the action of Frobenius on the middle-dimensional rigid cohomology of a smooth projective variety, to that of performing the same calculation for a smooth hyperplane section. We present in detail the main new algorithmic ingredient under some simplifying assumptions, and give full details of our algorithm for calculating zeta functions for some specific surfaces; we call it the "fibration algorithm". We have implemented the fibration algorithm for these surfaces over prime fields using the Magma programming language, and present some explicit examples which we have computed.

To illustrate the main idea behind our approach, we begin by outlining the proof given by Deligne of the Riemann hypothesis for a smooth projective variety X over the finite field $\mathbb{F}_q$ [**11**]. That is, the statement that for each $0 \leqslant i \leqslant 2\dim(X)$

The author is a Royal Society University Research Fellow. He would like to thank his colleagues and family for their support and help. He has also been greatly assisted by the kind help of Francesco Baldassarri, Gilles Christol, Jan Denef, Bas Edixhoven, Ralf Gerkmann Johan de Jong, Kiran Kedlaya, Michael Singer, Frederik Vercauteren and Daqing Wan. Especial thanks to Nobuo Tsuzuki for his detailed personal communication, and to the anonymous referee for many helpful comments. This paper is dedicated to Richard P. Brent on the occasion of his sixtieth birthday.

2000 Mathematics Subject Classification 11Y99, 11M38, 14D05, 14D10, 14F30

the action of the Frobenius endomorphism on the ℓ-adic étale cohomology space $H^i_{et}(X, \mathbb{Q}_\ell)$ has eigenvalues of complex absolute value $q^{i/2}$.

Let $X \subset \mathbb{P}$ be a smooth projective variety of dimension $n+1 > 1$ defined over the finite field $\mathbb{F}_q$. Denote by $\check{\mathbb{P}}$ the dual projective space whose points t correspond to hyperplanes H_t in $\mathbb{P}$, and let D be a line in $\check{\mathbb{P}}$. Let $\tilde{X} \subset X \times D$ denote the set of points (x, t) such that $x \in H_t$. Projection on the first and second coordinates yields maps $X \xleftarrow{\pi} \tilde{X} \xrightarrow{f} D$. The fibre of f at $t \in D$ is the hyperplane section $X_t = X \cap H_t$ of X. For sufficiently general D these maps define a *Lefschetz pencil* [**11**, (5.1)] (one may need to change the projective embedding first [**11**, (5.7)]).

The action of the Frobenius endomorphism on the ℓ-adic étale cohomology $H^*_{et}(X, \mathbb{Q}_\ell)$ may be studied via this Lefschetz pencil. In particular, assuming the result holds for smooth curves and arguing by induction on the dimension $n+1$, one can reduce the proof of the Riemann hypothesis for X to the case of the Frobenius action on the middle-dimensional cohomology space $H^{n+1}_{et}(X, \mathbb{Q}_\ell)$. The Leray spectral sequence for f and further inductive arguments now reduce the proof of the Riemann hypothesis to the case of $E^{1,n}_{2,et} := H^1_{et}(D, R^n f_* \mathbb{Q}_\ell)$ [**11**, (7.1)]. That is, one must prove that the Frobenius acting on this finite dimensional $\mathbb{Q}_\ell$-vector space has eigenvalues which have complex absolute value $q^{(n+1)/2}$. This is the "core problem" and it requires considerable ingenuity.

In this article, we are interested in computing the eigenvalues of Frobenius, rather than proving that they verify Weil's conjecture. However, it should be possible to bring to bear upon this computational problem the above geometric machinery. Specifically, one expects that the geometric techniques which Deligne used in his reduction to the core problem can be made algorithmic. However, even once this is done, one is still faced with a difficult problem, viz., calculation of the Frobenius action on $E^{1,n}_{2,et}$. The present author has no idea on how this might be achieved. However, the sketch of Deligne's proof can be presented in the terms of *rigid cohomology*, rather than ℓ-adic étale cohomology, and this theory is much more amenable to computation. In this article we present an algorithmic solution to the analogous "core problem", at least under certain simplifying assumptions. The principal novelty of this algorithmic technique is that it proceeds by induction on the dimension. Specifically, the calculation of a matrix for the action of Frobenius on the rigid cohomological analogue $E^{1,n}_{2,rig}$ requires as input a matrix for the action of Frobenius on $H^n_{rig}(X_t)$ for some hyperplane section X_t of X. So we can show that for the purposes of computation, the "core problem" of calculating Frobenius in the middle dimension can be efficiently reduced to that of a single instance of the problem one dimension lower down. In our method the base case of curves is handled using Kedlaya's algorithm [**25**].

We note in passing that for smooth projective hypersurfaces (of odd dimension) Deligne's solution of the "core problem" can be applied in a different manner, viz. rather than fibring the hypersurface X in a Lefschetz pencil, one can embed it as a fibre in such a pencil [**11**, (5.12)],[**23**]. Such an approach to calculating zeta functions was taken by the author in the "deformation algorithm" [**30**]. From a computational point of view, this latter approach has the disadvantage that the total space under consideration has dimension one more than the hypersurface itself. This impacts somewhat on the complexity of the "deformation algorithm". Specifically, the time/space complexity in terms of the middle Betti number $\dim H^{n+1}_{rig}(X)$

is rather high. Our new approach, of fibring the original variety, though more complicated, does appear better from the point of view of complexity dependence on the Betti numbers.

The algorithm presented in this paper uses the main technique developed for the "deformation algorithm", combined with a "higher rank" generalisation of Kedlaya's algorithm. Although our recursive approach was conceived as a general purpose algorithm, our implementation and complexity analysis for some surfaces suggest it is likely to be of most use for surfaces which can be fibred into low genus curves. Specifically, for the surfaces we consider in Sections 7, 8 and 9, if one fixes the genus g of the generic fibre of the fibration, then the asymptotic complexity of our algorithm is quasi-quartic in the middle Betti number, with quasi-cubic space requirement. In fact, the complexity in this case is comparable to that in the original algorithm of Kedlaya [**25**], only in this case we have surfaces rather than curves (Theorem 8.6). The dependence on the genus g itself is roughly comparable to that in the "deformation algorithm" for curves; see the end of Section 8.3.

We now outline the contents of the various sections in this paper. In Section 2 we define the zeta function of a variety and explain the computational problem which pertains to it. In Section 3 we give the main definitions from rigid cohomology which we shall need, and define the specific computational problem on which we shall focus (Problem 3.7), viz., calculation of Frobenius on the space $E_{2,rig}^{1,n}$. Neither Section 2 nor Section 3 contains any original contribution.

Section 4 considers an "abstract" version of the main computational problem, and proves a number of theorems relevant to its solution (Theorems 4.2 and 4.8). The main theorem stated in this section (Theorem 4.7) is not new; however, Theorems 4.2 and 4.8 together yield an algorithmic/effective proof of a slight weakening of Theorem 4.7. This algorithmic/effective proof is a new contribution. The material in Section 4 amounts to a special case of a "higher rank" generalisation of Kedlaya's algorithm. Section 5 contains a description of the main technique used in the "deformation algorithm". The analysis of the loss of numerical precision during the application of this technique is the only original contribution in this section; see Theorem 5.1 and the discussion following it. Section 6 presents our algorithmic solution to the main computational problem. Specifically, we assemble together the algorithmic and theoretical techniques developed in Sections 4 and 5 to address Problem 3.7.

Section 7 presents an explicit family of surfaces, viz., open subsets of affine surfaces defined by equations of the form $Z^2 = \bar{Q}(X, \Gamma)$ under some smoothness assumptions. We note that these surfaces were previously studied for different reasons by the author in his expository papers [**31, 32**]; see also the Ph.D. work of Hubrechts [**19**]. The algorithm described in Section 6, together with an auxiliary algorithm (Section 7.3.1), Kedlaya's algorithm for hyperelliptic curves, and some propositions (7.1, 7.2, and 7.3) allow the efficient computation of numerical approximations to the action of Frobenius on the middle-dimensional rigid cohomology of these open surfaces; see Theorem 7.6 for a complexity estimate. In Section 8 we consider smooth compactifications of these open surfaces, and describe how one may efficiently compute the full zeta functions of these compact surfaces using the main result of Section 7; see Section 8.3 for complexity estimates. We have implemented this zeta function algorithm for the case in which the base field is prime using the Magma programming language. Section 9 presents some explicit zeta

functions which we have computed using our implementation and discusses some further results.

The author would like to make a comment regarding the original motivation of this work: An interesting problem when calculating zeta functions using rigid cohomology is establishing good bounds on the loss of numerical precision, i.e., quantifying the divisions by the characteristic p which occur during the algorithms. It was the author's attempt to prove such precision-loss bounds by induction on the dimension using a deep theorem of Christol-Dwork (see [**7**] or [**13**, Chap. V]) which lead him to consider a recursive approach to computing zeta functions. The Christol-Dwork theorem, which can be thought of as a special case of an *effective* p-adic local monodromy theorem, remains an essential ingredient in the theoretical analysis of the algorithm presented in this paper.

2. *Varieties and zeta functions*

Let $\mathbb{F}_q$ be the field finite with q elements of characteristic p, and fix an algebraic closure $\overline{\mathbb{F}}_q \supset \mathbb{F}_q$. For each integer $s \geqslant 1$, let $\mathbb{F}_{q^s}$ denote the unique subfield of $\overline{\mathbb{F}}_q$ of order q^s. Let X be a variety defined over $\mathbb{F}_q$, i.e., a separated $\mathbb{F}_q$-scheme of finite type. For $s \geqslant 1$, let $|X(\mathbb{F}_{q^s})|$ denote the number of $\mathbb{F}_{q^s}$-rational points on X.

Definition 2.1. *The zeta function of X is the formal power series*

$$Z(X,T) := \exp\left(\sum_{s=1}^{\infty} \frac{|X(\mathbb{F}_{q^s})|}{s} T^s\right).$$

Theorem 2.2 (Dwork). *The zeta function is a rational function. More precisely, $Z(X,T) = P(T)/Q(T)$ for some polynomials $P(T), Q(T) \in 1 + T\mathbb{Z}[T]$ with* $\gcd(P,Q) = 1$.

We are interested in algorithms which take as input some explicit description of the variety X and give as output the zeta function $Z(X,T)$. To measure the performance of such an algorithm, we need to assume that some reasonable notion of the *problem size* $\mathrm{size}(X)$ has been defined; $\mathrm{size}(X)$ should be the number of bits needed to specify the input and output in some reasonable manner. For example, if X is a projective hypersurface defined by a homogeneous polynomial of degree $d \geqslant 1$ in $n \geqslant 2$ variables, then $\mathrm{size}(X) := (d+1)^{n-1}(\log_2(q)+1)$ would be appropriate. Let us assume some function $\mathrm{size}(X)$ has been defined. The central problem in the *algorithmic theory of zeta functions* [**43**] is:

Problem 2.3. *Find an explicit deterministic (or probabilistic) algorithm and an explicit polynomial R such that the following is true: The algorithm takes as input a variety X, gives as output the zeta function $Z(X,T)$, and has running time in bit operations bounded by $R(\mathrm{size}(X))$.*

One can also consider less universal versions of Problem 2.3 in which some restrictions on the input are made, e.g., the dimension is fixed, the growth of the characteristic p is controlled in some manner, the variety is smooth, or the variety lies in some specific family.

The proof of Dwork's rationality theorem can be transformed into an algorithm for computing zeta functions, see [**33**]. For a hypersurface the running time of this

algorithm is a polynomial function of $(p \cdot \text{size}(X))^{\dim(X)}$, i.e., it solves Problem 1 for hypersurfaces assuming the dimension is fixed and the characteristic "small". This algorithm though is of little practical interest. The algorithm of Schoof-Pila solves Problem 2.3 for smooth plane projective curves of fixed degree [**36, 38**]; this is the most general result obtained so far using the l-adic theory.

Let us for the remainder of this section assume that X is a smooth variety of pure dimension $\dim(X)$. Define $f := [\mathbb{F}_q : \mathbb{F}_p]$ and let K be the unramified extension of degree f of the field $\mathbb{Q}_p$ of p-adic numbers. The Lefschetz fixed point formula in rigid cohomology tells us [**16**, Théorème 6.3, II]:

$$Z(X,T) = \prod_{i=0}^{2\dim(X)} \det(1 - Tq^{\dim(X)}F_q^{-1} | H^i_{rig}(X))^{(-1)^{i+1}}.$$

Here $H^i_{rig}(X)$ are the *rigid cohomology spaces* associated to X. These are finite dimensional vector spaces over K [**27**, Theorem 1.2.1]. The *Frobenius map* F_q on these spaces is that induced by the qth power map on the structure sheaf of X. We have $F_q = F_p^f$ where the *absolute Frobenius* F_p is the map induced by the pth power map on the structure sheaf of X. Note that F_p is semi-linear with respect to the Frobenius automorphism of K, whereas F_q is linear.

Define $F := F_p$. The central problem in the *algorithmic theory of rigid cohomology* is:

PROBLEM 2.4. *Find an explicit deterministic (or probabilistic) algorithm and an explicit polynomial R such that the following is true: The algorithm takes as input a smooth variety X of pure dimension $\dim(X)$, gives as output a "sufficiently good" numerical approximation to a matrix for the semi-linear map $F : H^i_{rig}(X) \to H^i_{rig}(X)$ for each $0 \leqslant i \leqslant 2\dim(X)$, and has running time in bit operations bounded by $R(p \cdot \text{size}(X))$.*

By a numerical approximation we mean a p-adic approximation, and by sufficiently good we mean good enough to recover the integer polynomials $P(T)$ and $Q(T)$. For smooth projective varieties the author believes that Poincaré duality and the Lefschetz hyperplane theorem should allow one to focus attention on cohomology in dimension $\dim(X)$.

Kedlaya's algorithm [**25**] can be applied to Problem 2.4, but the running time of this approach is polynomial in $(p \cdot \text{size}(X))^{\dim(X)}$. However, it is a remarkably useful algorithm for the case $\dim(X) = 1$ where it has been extensively studied and implemented, see [**26**]. Problem 2.4 was solved for smooth projective hypersurfaces using relative rigid cohomology in [**30**]. We call this approach the "deformation algorithm". It seems to be of some practical interest, see in particular the recent work of Gerkmann [**18**] and Hubrechts [**19**]. We refer the reader to Tsuzuki [**41**] for a different approach which also uses relative rigid cohomology. This method, which one might call the "degeneration algorithm", is conceptually very nice; however, it has only been worked out in one special case and it is not clear to the present author how widely it can be applied.

To understand better the performance of the "deformation algorithm" it is necessary to look more carefully at the dependence of the running time/space on $\text{size}(X)$. Specifically, one can consider separately the dependence on the *arithmetic size* and the *geometric size*. One defines the former as $\text{size}_a(X) := (\log_2(q)+1)$. For a smooth

projective hypersurface defined by a polynomial of degree d in n variables, it is reasonable to define the latter as $\text{size}_g(X) := (d+1)^{n-1}$; note that this bounds the middle Betti number. The running time/space of the "deformation algorithm" is good with respect to the arithmetic size, but rather high with respect to the geometric size. Specifically, based on the analysis in [**19**], the author conjectures that for smooth projective hypersurfaces X over $\mathbb{F}_q$, the "deformation algorithm" requirements $\tilde{\mathcal{O}}(p\log(q)^3\text{size}_g(X)^{4+\omega})$ bit operations and $\tilde{\mathcal{O}}(p\log(q)^3\text{size}_g(X)^5)$ bits of space. Here the Soft-Oh notation ignores logarithmic factors [**17**, Def. 25.8], and ω is the smallest feasible exponent for matrix multiplication (one can take $\omega < 2.376$) [**17**, Page 315].

The aim of the new approach in this paper is to try to reduce the time/space dependence on the geometric size by using a more economical geometric method. This is achieved for the surfaces studied in Sections 7 and 8; see Section 8.3. We call this new approach the "fibration algorithm".

We conclude this section by mentioning two very recent advances in the area: First, work by Kedlaya et al [**1**] on bounding Picard numbers using p-adic cohomology. Second, work of Edixhoven et al [**15**] on computing coefficients of certain modular forms using l-adic cohomology of high dimensional varieties.

3. *Rigid cohomology*

We first gather together the definitions and results from rigid cohomology which are necessary to describe the main computational problem we study. We follow the description in Gerkmann [**18**, Sec. 3] and refer the reader to that source, the original papers of Berthelot [**4, 5**], and a recent paper of Tsuzuki [**40**] for further details. The reader might find the explicit examples worked out in Section 7.1 and [**18**, Section 5] helpful.

Prior to embarking upon our exposition on rigid cohomology, we wish to clarify one point about our "fibration algorithm". The essential idea of the algorithm is to fibre a smooth projective variety via hyperplane sections, and then use a number of quite sophisticated results from rigid cohomology to reduce consideration to a single hyperplane section, namely: comparison theorems (Theorems 3.1 and 4.7), base change theorem (Theorem 3.2), Leray spectral sequence (Section 3.6), and excision (Section 9.3.1). The author believes that all required results should be applicable provided the variety lifts to characteristic zero; in particular, that one can find the necessary pencil over a suitable dense open subset of the variety. The "trick" that lies at that heart of the algorithm and allows one to move from one dimension to the next is "deformation of Frobenius" (Section 5). However, unlike in Deligne's proof of the Riemann hypothesis for varieties, the "fibration algorithm" does *not* use the theory of Lefschetz pencils, in the sense of [**12**]. Indeed, the author is not aware that this theory has to date been adequately transcribed into the setting of rigid cohomology.

3.1. *Relative rigid cohomology*

Let $k = \mathbb{F}_q$ be a finite field of characteristic p, and K be the unramified extension of $\mathbb{Q}_p$ of degree $[k : \mathbb{F}_p]$. Let $\mathcal{O}_K$ denote the valuation ring of K. Then (p) is the maximal ideal of $\mathcal{O}_K$ and $\mathcal{O}_K/(p) \cong \mathbb{F}_q$. Let ord_p denote the p-adic valuation on K normalised so that $\text{ord}_p(p) = 1$, and $|\cdot|_p := p^{-\text{ord}_p(\cdot)}$ the corresponding norm.

Extend the norm and valuation to polynomial rings and finite dimensional vector spaces with given bases over K in the obvious manner.

Let X be a separated k-scheme of finite type. Let $(X, \bar{X}, \hat{\mathcal{X}})$ be an $\mathcal{O}_K$-triple for X, viz. an open immersion $j : X \hookrightarrow \bar{X}$ into a proper k-scheme, and an admissible embedding $i : \bar{X} \to \hat{\mathcal{X}}$ into a formal $\mathcal{O}_K$-scheme; here admissible means that $\hat{\mathcal{X}}$ is smooth around the image $i(X)$. For S a separated k-scheme of finite type and $(S, \bar{S}, \hat{\mathcal{S}})$ an $\mathcal{O}_K$-triple for S, a morphism $(X, \bar{X}, \hat{\mathcal{X}}) \to (S, \bar{S}, \hat{\mathcal{S}})$ is a commutative diagram

$$\begin{array}{ccccc} X & \hookrightarrow & \bar{X} & \to & \hat{\mathcal{X}} \\ f\downarrow & & \bar{f}\downarrow & & \downarrow \hat{\boldsymbol{f}} \\ S & \hookrightarrow & \bar{S} & \to & \hat{\mathcal{S}}. \end{array} \tag{1}$$

The *relative rigid cohomology* sheaf of the morphism $f : X \to S$ is

$$\mathcal{H}^i_{rig}(X/S) := R^i \hat{\boldsymbol{f}}_{K*} j^\dagger \Omega^\bullet_{]\bar{X}[/]\bar{S}[}.$$

Here $\hat{\boldsymbol{f}}_K$ is the map $]\bar{X}[\to]\bar{S}[$ between open tubes induced by the morphism $\hat{\boldsymbol{f}}$ of formal schemes, and $j^\dagger$ is the functor of overconvergent sections [**5**]. We take global sections to give the *relative rigid cohomology* spaces $H^i_{rig}(X/S) := \Gamma(]\bar{S}[, \mathcal{H}^i_{rig}(X/S))$ with which we shall work.

3.2. *A relative comparison theorem*

Let X and S be separated k-schemes of finite type. Assume now that there exist commutative diagrams

$$\begin{array}{ccccc} X & \hookrightarrow & \bar{X} & & \\ \downarrow & & \downarrow & \text{and} & \\ \mathcal{X} & \hookrightarrow & \bar{\mathcal{X}} & & \end{array} \qquad \begin{array}{ccc} S & \hookrightarrow & \bar{S} \\ \downarrow & & \downarrow \\ \mathcal{S} & \hookrightarrow & \bar{\mathcal{S}}. \end{array}$$

On the bottom row, $\mathcal{X}, \bar{\mathcal{X}}, \mathcal{S}, \bar{\mathcal{S}}$ are $\mathcal{O}_K$-schemes. The vertical maps are embeddings of special fibres. The lower horizontal maps are open immersions and their codomains $\bar{\mathcal{X}}$ and $\bar{\mathcal{S}}$ are proper and smooth $\mathcal{O}_K$-schemes. Assume we have morphisms $f : X \to S$, $\bar{f} : \bar{X} \to \bar{S}$ as before, and further morphisms $\boldsymbol{f}$: $\mathcal{X} \to \mathcal{S}$, $\bar{\boldsymbol{f}}$: $\bar{\mathcal{X}} \to \bar{\mathcal{S}}$ so that p-adic completion leads to a diagram containing on one face (1) with all other faces commuting. The *relative de Rham cohomology* sheaf of the induced morphism on the generic fibres $\boldsymbol{f}_K : \mathcal{X}_K \to \mathcal{S}_K$ is denoted:

$$\mathcal{H}^i_{dR}(\mathcal{X}_K/\mathcal{S}_K) := R^i \boldsymbol{f}_{K*} \Omega^\bullet_{\mathcal{X}_K/\mathcal{S}_K}.$$

Taking global sections we define the *relative de Rham cohomology* spaces $H^i_{dR}(\mathcal{X}_K/\mathcal{S}_K) := \Gamma(\mathcal{S}_K, \mathcal{H}^i_{dR}(\mathcal{X}_K/\mathcal{S}_K))$ with which we shall work.

Next, assume that the complement $\bar{\mathcal{X}} - \mathcal{X}$ has smooth components with normal crossings over $\bar{\mathcal{S}}$. Then $\mathcal{H}^i_{dR}(\mathcal{X}_K/\mathcal{S}_K)$ is coherent.[1] Moreover, according to Gerkmann [**18**, Eqn (8)], the comparison theorem of Baldassarri-Chiarelletto extends to this relative situation. Specifically, the natural morphism

$$\mathcal{H}^i_{dR}(\mathcal{X}_K/\mathcal{S}_K) \otimes_{\mathcal{O}_{\mathcal{S}_K}} j^\dagger \mathcal{O}_{]\bar{S}[} \to \mathcal{H}^i_{rig}(X/S) \tag{2}$$

[1]This result appears to be "well-known to experts", although the author has not been able to find an explicit reference for it. The point is that in such a case $\mathcal{H}^i_{dR}(\mathcal{X}_K/\mathcal{S}_K)$ may be computed via the hypercohomology of a proper morphism applied to a relative logarithmic de Rham complex, which is coherent. Note that in our application in Section 7 we shall prove finiteness directly.

is an isomorphism. Define $A^\dagger := \Gamma(]\bar{S}[, j^\dagger\mathcal{O}_{]\bar{S}[})$ and $A := \Gamma(\mathcal{S}_K, \mathcal{O}_{\mathcal{S}_K})$ to be the rings of global sections. Then [**5**, Prop. (2.5.2)(ii)] shows that (2) implies the following theorem.

THEOREM 3.1. *With assumptions as stated in this section, the following isomorphism holds:*

$$H^i_{dR}(\mathcal{X}_K/\mathcal{S}_K) \otimes_A A^\dagger \cong H^i_{rig}(X/S).$$

3.3. *Proper and smooth base change*

We retain the definitions and assumptions from Section 3.1 and recall $A^\dagger = \Gamma(]\bar{S}[, j^\dagger\mathcal{O}_{]\bar{S}[})$. Assume now that the morphism $\hat{\boldsymbol{f}}\colon \hat{\mathcal{X}} \to \hat{\mathcal{S}}$ is proper and smooth. For each point $\bar{\gamma} \in S$ in the base denote by $X_{\bar{\gamma}}$ the fibre at $\bar{\gamma}$ of $X \to S$. The following base change theorem will be of importance to us, c.f. [**4**, Théorème 5], [**18**, Theorem 3.1], [**40**, Theorem 4.1.4]:

THEOREM 3.2 (BERTHELOT). *Base change* $\mathrm{Spec}(k(\bar{\gamma})) \to S$ *induces an isomorphism* $H^i_{rig}(X/S) \otimes_{A^\dagger} K(\gamma) \cong H^i_{rig}(X_{\bar{\gamma}})$. *Here* γ *is the generic fibre of a lift of* $\bar{\gamma}$.

3.4. *Pencils of varieties*

We retain the definitions and assumptions in Section 3.2 and 3.3, i.e., we have a morphism $f : X \to S$, along with all the auxiliary objects and properties so that the comparison (Theorem 3.1) and base change (Theorem 3.2) theorems hold.

Assume now that

$$S = \mathbb{P}^1_k - \{\bar{\gamma}_1, \ldots, \bar{\gamma}_d, \infty\}$$

where the $\bar{\gamma}_i \in k$ are distinct. Thus $\bar{S} = \mathbb{P}^1_k$. Choose $\gamma_i \in \mathcal{O}_K$ so that $\gamma_i \bmod p = \bar{\gamma}_i$. Take

$$\mathcal{S} = \mathbb{P}^1_{\mathcal{O}_K} - \{\gamma_1, \ldots, \gamma_d, \infty\}.$$

The coordinate ring $A = \Gamma(\mathcal{S}_K, \mathcal{O}_{\mathcal{S}_K})$ of the generic fibre $\mathcal{S}_K$ is the localisation $K[\Gamma][1/r(\Gamma)]$ where $r := \prod_{i=1}^d(\Gamma - \gamma_i) \in K[\Gamma]$. The ring of global sections $A^\dagger = \Gamma(]\bar{S}[, j^\dagger\mathcal{O}_{]\bar{S}[})$ is the weak (a.k.a. dagger) completion of $K[\Gamma][1/r(\Gamma)]$. We denote this ring by $K[\Gamma][1/r(\Gamma)]^\dagger$. Its elements can be written in the form $\sum_{i\in\mathbb{Z}} a_i(\Gamma)r(\Gamma)^i$ where the coefficients $a_i(\Gamma) \in K[\Gamma]$ have $\deg(a_i) < \deg(r)$ and satisfy $\mathrm{ord}_p(a_i) - \varepsilon|i| \to \infty$ as $|i| \to \infty$ for some $\varepsilon > 0$. (Note that there is no lower bound on the ε which occur for different elements in the ring.) It is convenient at this stage to give a definition which we shall need later.

DEFINITION 3.3. *We shall say that we have given* effective p-adic bounds *for an element* $a = \sum_{i\in\mathbb{Z}} a_i(\Gamma)r(\Gamma)^i \in K[\Gamma, 1/r(\Gamma)]^\dagger$ *if we are given* $\eta, \delta \in \mathbb{Q}$ *with* $\eta > 0$ *such that* $\mathrm{ord}_p(a_i) \geqslant \eta|i| + \delta$ *for all* $i \in \mathbb{Z}$.

3.5. *Cohomology in the middle dimension*

We retain the definitions and assumptions from Section 3.4, i.e., we have a pencil $f : X \to S$ and the comparison and base change theorems hold.

Let us now focus our attention on the middle dimension. Specifically, let n be the relative dimension of the morphism $f : X \to S$. By the comparison theorem and coherence of relative de Rham cohomology, we see that $H^n_{dR}(\mathcal{X}_K/\mathcal{S}_K)$ and

$H^n_{rig}(X/S)$ are locally free modules of finite rank over the rings $A = K[\Gamma, 1/r(\Gamma)]$ and $A^\dagger = K[\Gamma, 1/r(\Gamma)]^\dagger$, respectively. We shall *assume* that they are in fact free; this is certainly true after shrinking the base S. Let us simplify our notation now by writing

$$\mathcal{E} := H^n_{dR}(\mathcal{X}_K/\mathcal{S}_K),\ \mathcal{E}^\dagger := H^n_{rig}(X/S),$$

so by the comparison theorem (Theorem 3.1) we have

$$\mathcal{E}^\dagger = \mathcal{E} \otimes_A A^\dagger.$$

The free modules $\mathcal{E}$ and $\mathcal{E}^\dagger$ come with additional structure. Specifically, derivation with respect to Γ induces a *connection* ∇ on the A-module $\mathcal{E}$. Let us recall the precise definition.

DEFINITION 3.4. *A connection ∇ on $\mathcal{E}$ is a map $\nabla : \mathcal{E} \to \mathcal{E} \otimes \Omega^1_A$ such that $\nabla(e_1 + e_2) = \nabla(e_1) + \nabla(e_2)$ and $\nabla(ae_1) = e_1 \otimes da + a\nabla(e_1)$ for all $a \in A$ and $e_1, e_2 \in \mathcal{E}$.*

Here $\Omega^1_A := \Omega^1_{A/K}$ is the module of K-linear differentials, and $d : A \to \Omega^1_A$ is the universal derivation, which in our case amounts to differentiation w.r.t. Γ. The connection induced by differentiation w.r.t. Γ is called the *Gauss-Manin* connection, and the pair $(\mathcal{E}, \nabla)$ a *∇-module*. Differentiation with respect to Γ also induces a connection $\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes \Omega^1_{A^\dagger}$. Here $\Omega^1_{A^\dagger}$ is the module of K-linear differentials which are continuous w.r.t. the p-adic norm, i.e., there is a continuous derivation $A^\dagger \to \Omega^1_{A^\dagger}$ which is "universal" w.r.t. any continuous derivation from $A^\dagger$ to an $A^\dagger$-module. The comparison theorem (Theorem 3.1) tells us that the connection $\nabla^\dagger$ is just the Gauss-Manin connection. Specifically, assuming one can compute a matrix for ∇ with respect to some basis of $\mathcal{E}$ over A, then the *same* matrix defines the map $\nabla^\dagger$ on the basis of $\mathcal{E}^\dagger$ obtained by extending scalars. This means that the Gauss-Manin connection $\nabla^\dagger$ on $\mathcal{E}^\dagger$ can be computed in a purely algebraic manner. We refer the reader to [**18**, Section 4] for a more detailed discussion of the Gauss-Manin connection in rigid cohomology; see also Section 3.6 of the present paper.

The module $\mathcal{E}^\dagger$ comes with one further piece of data, namely the *(absolute) Frobenius* map $F : \mathcal{E}^\dagger \to \mathcal{E}^\dagger$, which is induced by the pth power map on the structure sheaf of X. Specifically, the construction requires the choice of a lifting of the pth power map from $\mathbb{F}_q[\Gamma, 1/\bar{r}(\Gamma)](\bar{r} := r \bmod p)$ to $A^\dagger$. Let us *assume* that we have chosen the obvious lifting $\sigma : A^\dagger \to A^\dagger$ so that $\sigma : \Gamma \mapsto \Gamma^p$ and σ acts on K as the Frobenius automorphism. Then $F : \mathcal{E}^\dagger \to \mathcal{E}^\dagger$ is σ-linear, i.e., it is additive and $F(ae) = \sigma(a)F(e)$ for all $a \in A^\dagger, e \in \mathcal{E}^\dagger$. With respect to a basis for $\mathcal{E}^\dagger$, it can also be described by a matrix. Note though that this matrix has entries in the ring $A^\dagger$, whereas the matrix for $\nabla^\dagger$ referred to in the previous paragraph has entries in A. The following diagram relating the Frobenius and connection commutes:

$$\begin{array}{ccc} \mathcal{E}^\dagger & \stackrel{\nabla^\dagger}{\to} & \mathcal{E}^\dagger \otimes_{A^\dagger} \Omega^1_{A^\dagger} \\ F \downarrow & & \downarrow F \otimes d\sigma \\ \mathcal{E}^\dagger & \stackrel{\nabla^\dagger}{\to} & \mathcal{E}^\dagger \otimes_{A^\dagger} \Omega^1_{A^\dagger}. \end{array} \tag{3}$$

The data $(\mathcal{E}^\dagger, \nabla^\dagger, F)$ is called a *$(\sigma, \nabla^\dagger)$-module* over $A^\dagger$, or alternatively an *overconvergent F-isocrystal* over S. Let us recall the precise definition.

DEFINITION 3.5. *An overconvergent F-isocrystal $(\mathcal{E}^\dagger, \nabla^\dagger, F)$ on S (a.k.a. $(\sigma, \nabla^\dagger)$-module over $A^\dagger$) consists of the following data: A finite locally free $A^\dagger$-module $\mathcal{E}^\dagger$, a σ-linear map $F : \mathcal{E}^\dagger \to \mathcal{E}^\dagger$ which induces an isomorphism $\sigma^*\mathcal{E}^\dagger \to \mathcal{E}^\dagger$, and a connection $\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes_{A^\dagger} \Omega^1_{A^\dagger}$, such that Diagram (3) commutes.*

We shall denote the kernel and cokernel of the map $\nabla^\dagger$ in Diagram (3) by $H^0_{rig}(S, \mathcal{E}^\dagger)$ and $H^1_{rig}(S, \mathcal{E}^\dagger)$, respectively. These objects are vector spaces over K. By commutativity, F induces a map on each of these spaces. We note that the space $\ker(\nabla^\dagger) = H^0_{rig}(S, \mathcal{E}^\dagger)$ is certainly finite dimensional over K: it embeds in the finite-dimensional space of local solutions around any non-singular point.

We now state our final assumption on the family $X \to S$. We assume that the family $\mathcal{X}_K \to \mathcal{S}_K$ comes by extension of scalars from a smooth morphism defined over an algebraic number field. It follows then by the "open local monodromy theorem" that the connection is regular, i.e., locally has only simple poles, and the local exponents (see Section 4.1) are rational numbers [**21**, Thm. (14.3)].

Under this final assumption, as well as the others already in place in this section, it is known that $H^1_{rig}(S, \mathcal{E}^\dagger)$ is also finite dimensional [**3**, Corollary 2]. We will give an effective/algorithmic proof of finiteness under some simplifying assumptions (follows from Theorems 4.2 and 4.8). We mention in passing that there is also an older, related result due to Adolphson [**2**, Theorem 2, Remark p. 286].

We now come to the main definition in the paper.

DEFINITION 3.6. *Let $E^{1,n}_{2,rig} := H^1_{rig}(S, \mathcal{E}^\dagger) = \operatorname{coker}(\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes_{A^\dagger} \Omega^1_{A^\dagger})$.*

By a result communicated to us by Professor Nobuo Tsuzuki, when X is affine and $X \to S$ is a smooth liftable family, this space is a term in a spectral sequence for the morphism $X \to S$; in fact, we have the isomorphism $H^{n+1}_{rig}(X) \cong E^{1,n}_{2,rig}$, see Eqn (4) in Section 3.6.

Finally we are able to state the computational problem we consider:

PROBLEM 3.7. *Calculate a numerical approximation to a matrix for the map $F : E^{1,n}_{2,rig} \to E^{1,n}_{2,rig}$.*

In Section 6, we solve this problem under the assumption that we are given as input suitable numerical approximations to:

- A matrix for the connection ∇, this matrix having only simple poles (even "modulo p") and prepared local exponents. (See the start of Section 4.3 for the definition of the latter term.)
- A specialisation of the matrix for $F : \mathcal{E}^\dagger \to \mathcal{E}^\dagger$ at a Teichmüller point, i.e., at an element $\gamma \in \mathcal{O}_K$ with $\gamma^q = \gamma$.

We further assume that:

- We are given as input effective p-adic bounds for the entries of the matrix for $F : \mathcal{E}^\dagger \to \mathcal{E}^\dagger$ (see Definition 3.3).

Regarding the first input, ones expects to be able to compute this matrix efficiently in any concrete application of the method. See for example our calculation in Section 7.3.1. The assumption on the matrix for the connection is "locally" true by the regularity and local monodromy theorem [**21**, Thm. (14.3)], since the family

$\mathcal{X}_K \to \mathcal{S}_K$ can be defined over an algebraic number field. Our simplifying assumption is that there is a *global basis* for which the matrix has only simple poles with prepared local exponents.

Regarding the second input, by the base change theorem (Theorem 3.2) the specialisation, say at $\Gamma = \gamma$ a Teichmüller point, is precisely the matrix for the pth power Frobenius map acting on the cohomology space $H^n_{rig}(X_{\bar{\gamma}})$. Here $X_{\bar{\gamma}}$ is the fibre of the family at $\bar{\gamma} := \gamma \bmod p$. Such a matrix can be computed recursively when $\dim(X_{\bar{\gamma}}) \geqslant 2$, and by Kedlaya's algorithm in the case $\dim(X_{\bar{\gamma}}) = 1$.

Regarding the third assumption, again one expects to be able to calculate such bounds in any concrete application of the method, see Section 7.3.3.

3.6. *Leray Spectral Sequence*

This section is independent of the rest of the paper. We describe the contents of a personal communication from Professor Nobuo Tsuzuki to the author. Note that the notation in this section is consistent with [**40**], but varies slightly from that in the remainder of this paper.

Let K be a complete discrete valuation field of mixed characteristic $(0, p)$ and $\mathcal{V}$ and k be the ring of integers and residue field of K, respectively. Let S be an affine smooth scheme of dimension m over $\mathrm{Spec}(k)$ and X/S a smooth family with $n := \mathrm{rel.dim}(X/S)$ such that X is affine. Suppose there exists a smooth affine lift $\mathcal{X}/\mathcal{S}$ of X/S over $\mathrm{Spec}(\mathcal{V})$ with $A = \Gamma(\mathcal{X}, \mathcal{O}_{\mathcal{X}})$ and $R = \Gamma(\mathcal{S}, \mathcal{O}_{\mathcal{S}})$ such that $\Omega^1_{R/\mathcal{V}}$ is a free R-module. Then one can calculate the rigid cohomology of X/K as

$$H^r_{rig}(X/K) := H^r(A^\dagger_K \otimes \Omega^\bullet_{A/\mathcal{V}}) \,(=: H^r_{MW}(X/K)),$$

where $A^\dagger$ is the weak (a.k.a. dagger) completion of A over $\mathcal{V}$ and $A^\dagger_K := A^\dagger \otimes_{\mathcal{V}} K$. Let us define a filtration Fil^* of $A^\dagger_K \otimes_A \Omega^\bullet_{A/\mathcal{V}}$ by

$$\mathrm{Fil}^q := \mathrm{Im}(A^\dagger_K \otimes_A \Omega^{\bullet-q}_{A/\mathcal{V}} \otimes_R \Omega^q_{R/\mathcal{V}} \to A^\dagger_K \otimes_A \Omega^\bullet_{A/\mathcal{V}}).$$

Since $\Omega^q_{R/\mathcal{V}}$ is a free R-module, one has

$$\mathrm{Gr}^q_{\mathrm{Fil}} = A^\dagger_K \otimes_A \Omega^{\bullet-q}_{A/R} \otimes_R \Omega^q_{R/\mathcal{V}}.$$

There exists a spectral sequence [**40**, Thm. 3.4.1]

$$E_1^{q,r} := H^r(A^\dagger_K \otimes_A \Omega^\bullet_{A/R}) \otimes_R \Omega^q_{R/\mathcal{V}} \Rightarrow H^{q+r}(A^\dagger_K \otimes_A \Omega^\bullet_{A/\mathcal{V}}) = H^{q+r}_{MW}(X/K),$$

where the edge homomorphism is called the Gauss-Manin connection. Since $E_1^{q,r} = 0$ except when $0 \leqslant q \leqslant m$ and $0 \leqslant r \leqslant n$, one has

$$E_2^{m,n} = \cdots = E_\infty^{m,n} = H^{m+n}_{MW}(X/K).$$

Hence the top rigid cohomology group $H^{m+n}_{rig}(X/K)$ is calculated by the Gauss-Manin connection:

$$H^{m+n}_{rig}(X/K) \cong \mathrm{coker}\left(H^n(A^\dagger_K \otimes_A \Omega^\bullet_{A/R}) \otimes_R \Omega^{m-1}_{R/\mathcal{V}} \to H^n(A^\dagger_K \otimes_A \Omega^\bullet_{A/R}) \otimes_R \Omega^m_{R/\mathcal{V}}\right). \tag{4}$$

4. *Algorithms for reduction in* $E^{1,n}_{2,rig}$

This section is independent of Section 3, but relies on it for motivation. We recall the definitions we shall need in an abstract manner, stripped of their geometric origin.

4.1. *Definitions*

Let $k = \mathbb{F}_q$ be the finite field with q elements of characteristic p, and K the unramified extension of $\mathbb{Q}_p$ of degree $[k : \mathbb{F}_p]$. Denote by $\bar{K}$ an algebraic closure of K. Let $\mathcal{O}_K$ be the ring of integers of K, and $r(\Gamma) \in \mathcal{O}_K[\Gamma]$ a monic polynomial of degree d which is squarefree modulo p. Let $A := K[\Gamma, 1/r(\Gamma)]$ and let $A^\dagger$ be the dagger completion of A (this is defined in Section 3.4). Let $\mathcal{E}$ be a free module of finite rank m over A and define $\mathcal{E}^\dagger := \mathcal{E} \otimes_A A^\dagger$. Let $\nabla : \mathcal{E} \to \mathcal{E} \otimes \Omega^1_A$ be a connection (Definition 3.4). Fix a basis $\mathcal{B}$ for $\mathcal{E}$ over A and represent elements in $\mathcal{E}$ as column vectors w.r.t. this basis. Take for Ω^1_A the basis element $d\Gamma$ over A. Take the basis for $\mathcal{E} \otimes \Omega^1_A$ to be the tensor product of these two bases. Assume that with respect to this choice, the connection ∇ acts as

$$\nabla = \frac{d}{d\Gamma} + \frac{b(\Gamma)}{r(\Gamma)} : \mathcal{E} \cong A^m \to \mathcal{E} \otimes \Omega^1_A \cong A^m \otimes d\Gamma \tag{5}$$

where the matrix $b(\Gamma) \in M_m(\mathcal{O}_K[\Gamma])$ has degree in Γ at most $d-1$. This assumption ensures that the matrix for ∇ has only simple poles, including at infinity. This is our *main* simplifying assumption. Such a differential system is called *fuchsian*. Any differential system with regular singular points may in principle, after a change of basis, be written in this form, possibly at the expense of introducing one new pole. See the discussion of the Riemann-Hilbert problem in [**37**, Section 5.3].

Let $\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes \Omega^1_{A^\dagger}$ be obtained from ∇ by extension of scalars. We are interested in the spaces

$$E^{1,n}_{2,dR} := \operatorname{coker}(\nabla : \mathcal{E} \to \mathcal{E} \otimes \Omega^1_A),\ E^{1,n}_{2,rig} := \operatorname{coker}(\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes \Omega^1_{A^\dagger}).$$

The notation chosen here is to remind the reader of the "geometric origin" of the connections we shall actually be considering.

For $R := \{\gamma \in \mathcal{O}_{\bar{K}} \mid r(\gamma) = 0\}$ note that

$$\frac{b(\Gamma)}{r(\Gamma)} = \frac{b(\Gamma)}{r'(\Gamma)} \sum_{\gamma \in R} \frac{1}{\Gamma - \gamma}$$

where "dash" indicates differentiation w.r.t. Γ. Thus the *residue matrix* at the regular singular point $\Gamma = \gamma \in R$ is $b(\gamma)/r'(\gamma)$; the set of eigenvalues of this matrix, denoted E_γ, is the set of *local exponents* at $\Gamma = \gamma$. One checks that the residue matrix at infinity is $-b_{d-1}$, the negative of the coefficient of Γ^{d-1} in $b(\Gamma)$. The set E_∞ is defined as the set of eigenvalues of $-b_{d-1}$. Finally, the *exponent* set of ∇ w.r.t. the basis $\mathcal{B}$ is

$$E(\nabla, \mathcal{B}) := E_\infty \cup \bigcup_{\gamma \in R} E_\gamma.$$

Note that this set modulo $\mathbb{Z}$ is independent of the basis $\mathcal{B}$ [**21**, (12.0.2)].

Definition 4.1. *Let $\rho = \rho(\nabla, \mathcal{B})$ be the smallest positive integer larger than any integer in the set $E(\nabla, \mathcal{B})$.*

Denote by $A \otimes d\Gamma_\rho$ the K-vector space of 1-forms spanned by the set

$$\left\{ \frac{\Gamma^i}{r^j} \otimes d\Gamma \ : \ 0 \leqslant i < d, 1 \leqslant j \leqslant \rho \right\} \cup \left\{ \Gamma^j \otimes d\Gamma \ : \ 0 \leqslant j \leqslant \rho - 2 \right\}.$$

Denote by $\mathcal{E} \otimes d\Gamma_\rho$ the K-vector space spanned by column vectors in $A^m \otimes d\Gamma$ whose entries belong to the space $A \otimes d\Gamma_\rho$.

4.2. *Effective finiteness of* $E_{2,dR}^{1,n}$.

We can now state our first finiteness theorem.

THEOREM 4.2. *Let the pair* $(\mathcal{E}, \nabla)$ *be as defined in Section 4.1, and* $\rho = \rho(\nabla, \mathcal{B})$ *the positive integer from Definition 4.1 which depends upon both* ∇ *and the basis* $\mathcal{B}$ *for* $\mathcal{E}$. *Then* $\operatorname{coker}(\nabla)$ *is generated over* K *by the image of the space* $\mathcal{E} \otimes d\Gamma_\rho$.

Proof. We shall give an algorithm for writing an element $u \in \mathcal{E} \otimes d\Gamma$ in the form $u = \nabla(v) + w$ with $v \in \mathcal{E}$ and $w \in \mathcal{E} \otimes d\Gamma_\rho$. It proceeds in two stages: First, simultaneous reduction of the pole orders of 1-forms at the roots of r; Second, reduction of pole orders at infinity.

Let $U(\Gamma) \in K[\Gamma]^m$, viewed as a column vector. We shall show that for $\ell \geqslant \rho$ we have

$$\frac{U}{r^{\ell+1}} \otimes d\Gamma = \nabla\left(\frac{V}{r^\ell}\right) + \frac{W}{r^\ell} \otimes d\Gamma \tag{6}$$

for some $V, W \in K[\Gamma]^m$ with $\deg(V) < d$ and

$$\deg(W) \leqslant \max\{\max\{2d-2, \deg(U)\} - d, 0\}.$$

Moreover, we shall give a method for computing V and W.

We claim that there exists a unique $V(\Gamma) \in K[\Gamma]^m$ with $\deg(V) < d = \deg(r)$ such that

$$(-\ell r' I_m + b)V \equiv U \mod r.$$

Let us assume this claim is true. Define

$$X := \frac{(-\ell r' I_m + b)V - U}{r} \in K[\Gamma]^m.$$

Then

$$\deg(X) \leqslant \max\{\max\{2d-2, \deg(U)\} - d, 0\}.$$

Define $W := -X - V'$. Then $\deg(W)$ is bounded as claimed above and one checks by direct computation that (6) holds.

It remains to establish the uniqueness, existence and computability of V. For this, we must show that the determinant of the matrix $(-\ell r' I_m + b)$ is a unit modulo r. Now

$$(-\ell r' I_m + b) = -r' f(\ell, \Gamma)$$

where

$$f(t, \Gamma) := t I_m - \frac{b(\Gamma)}{r'(\Gamma)}.$$

Now r' is invertible modulo r since the latter is squarefree. We need to show $\det(f(\ell, \Gamma))$ is a unit modulo $r = \prod_{\gamma \in R}(\Gamma - \gamma)$, i.e., we must show that $\det(f(\ell, \gamma)) \neq$

0 for all $\gamma \in R$. But $\det(f(\ell,\gamma)) = 0$ if and only if ℓ is an eigenvalue of the matrix $b(\gamma)/r'(\gamma)$. Since $\ell \geqslant \rho$ and ρ is larger than any integer element in the set $\cup_{\gamma\in R} E_\gamma$ the result follows.

Let $U(\Gamma) \in K[\Gamma]^m$ as before. We shall show that if

$$\deg(U) - \rho d =: \ell - 1 > \rho - 2$$

then

$$\frac{U}{r^\rho} \otimes d\Gamma = \nabla(V\Gamma^\ell) + \frac{W}{r^\rho} \otimes d\Gamma \tag{7}$$

for some $V \in K^m$ and $W \in K[\Gamma]^m$ with $\deg(W) \leqslant \deg(U) - 1$. (Note that in the first part of the proof we reduced all column vectors of 1-forms to the shape on the left hand side of (7) modulo "exact forms". Equation (7) allows one to reduce such column vectors further modulo "exact forms" until they have entries in the space $A \otimes d\Gamma_\rho$.)

We shall take local expansions of rational functions around infinity. Put

$$\frac{U}{r^\rho} = u_{\ell-1}\Gamma^{\ell-1} + u_{\ell-2}\Gamma^{\ell-2} + \dots\,, \quad \nabla = \frac{d}{d\Gamma} + \left(b_{d-1}\Gamma^{-1} + \dots\right).$$

Here $b_{d-1} \in M_m(K)$ is the coefficient of the monomial Γ^{d-1} in $b(\Gamma)$. Let $V \in K^m$ be the element such that

$$(\ell I_m + b_{d-1})\, V = u_{\ell-1}.$$

We note that V exists and is unique by the assumption that the integers in the eigenvalue set E_∞ of $-b_{d-1}$ are all less than ρ, and $\ell \geqslant \rho$. By direct computation one checks that

$$\frac{U}{r^\rho} \otimes d\Gamma - \nabla(V\Gamma^\ell) = \frac{W}{r^\rho} \otimes d\Gamma$$

where $\deg(W) \leqslant \deg(U) - 1$. This concludes the description of the algorithm.

We note that in implementations one should represent the numerator U in an $r(\Gamma)$-adic expansion. With such a representation, in the second stage it is more efficient to compute

$$\frac{A}{r^\rho} \otimes d\Gamma - \nabla\left(V\Gamma^{\ell - d\lfloor \ell/d \rfloor} r^{\lfloor \ell/d \rfloor}\right)$$

with $V \in K^m$ as in the preceding paragraph. □

THEOREM 4.3. *The space $E_{2,dR}^{1,n}$ is a finite dimensional K-vector space.*

Proof. A basis for this space can be computed using linear algebra and Theorem 4.2. Specifically, one computes a basis for the cokernel of the K-linear map $\nabla : A_{\rho-1}^m \to A^m \otimes d\Gamma_\rho$. Here $A_{\rho-1}^m \subset A^m$ is the K-space of column vectors whose entries have poles of order at most $\rho - 1$. □

We note that the dimension of $E_{2,dR}^{1,n}$ can be calculated explicitly.

4.3. *Effective finiteness of $E_{2,rig}^{1,n}$*

In this section we give an effective/algorithmic proof of the finiteness of $E_{2,rig}^{1,n}$ under certain conditions. First, consider the conditions:

(Rat.) The exponent set $E(\nabla, \mathcal{B})$ contains only rational numbers.

(Prep.Rat.) The exponent set $E(\nabla, \mathcal{B})$ contains only rational numbers; moreover, for each $E \in \{E_\gamma\}_{\gamma \in R} \cup \{E_\infty\}$, if $\lambda_1, \lambda_2 \in E$ then $\lambda_1 - \lambda_2$ is not a positive integer.

If (Prep.Rat.) holds then we say that the local exponents are *prepared*. Certainly $(Prep.Rat.) \Rightarrow (Rat.)$. Also, rationality of the local exponents for a connection with regular singular points is independent of the basis chosen [**21**, (12.0.2)].

Next, consider the following definition:

DEFINITION 4.4. *The connection ∇ is* overconvergent *if it has a basis of local solutions which converge on the p-adic unit disk around the generic point t with $|t|_p = 1$.*

We refer the reader to [**13**, Chap. III, Sec. 5] for elucidation of the meaning of this definition. The essential point is that connections which "come from geometry" satisfy this condition, c.f. [**13**, Chap. III, Remark 5.3]. We shall prove this in Theorem 6.1, using Dwork's trick and the following alternative characterisation of overconvergence.

THEOREM 4.5. *The connection ∇ is overconvergent if and only if there exists an element $\gamma \in \mathcal{O}_{\bar{K}}$ with $\mathrm{ord}_p r(\gamma) = 0$ such that the differential system $\frac{d}{d\Gamma} + b(\Gamma)/r(\Gamma)$ has a basis of local solutions on the p-adic open unit disk around $\Gamma = \gamma$.*

Proof. That this is necessary follows by specialising the generic solution matrix at $\Gamma = \gamma$. For sufficiency, we observe that [**13**, Chap IV Prop 5.1] allows one to transfer the convergence on the open unit disk around the point $\Gamma = \gamma$ to the same disk around the generic point. (Specifically, change variables so that $\gamma = 0$, and take "α" to be the generic point t with $|t|_p = 1$.) □

DEFINITION 4.6. *The "dual" connection $\check{\nabla}$ is defined to act as*

$$\check{\nabla} : \frac{d}{d\Gamma} - \frac{b(\Gamma)}{r(\Gamma)} : A^m \to A^m \otimes d\Gamma.$$

In this case the matrix $b(\Gamma)/r(\Gamma)$ acts on the right on row vectors.

The p-adic condition that we shall need is:

(O.C.) The connections ∇ and $\check{\nabla}$ are overconvergent.

THEOREM 4.7 (BALDASSARRI-CHIARELLOTTO). *Let the pairs $(\mathcal{E}, \nabla)$ and $(\mathcal{E}^\dagger, \nabla^\dagger)$ be defined as in Section 4.1. Assume that conditions (Rat.) and (O.C.) are met. Then the natural morphism $(\mathcal{E}, \nabla) \to (\mathcal{E}^\dagger, \nabla^\dagger)$ induces an isomorphism $E^{1,n}_{2,dR} \cong E^{1,n}_{2,rig}$.*

Proof. This is an application of [**3**, Corollary 2.6]. □

We shall give an effective/algorithm proof of Theorem 4.7 under the stronger assumption (Prep.Rat.). More precisely, in Theorem 4.8 we give effective bounds on the p-adic growth of forms during the reduction algorithm in the proof of Theorem 4.2 under assumptions (Prep.Rat.) and (O.C.). It is easy to deduce surjectivity of

the morphism $E^{1,n}_{2,dR} \to E^{1,n}_{2,rig}$ from Theorem 4.8; injectivity may also be easily derived using the technique in the proof of [**2**, Theorem 2]. We omit the details of the proof of the isomorphism $E^{1,n}_{2,dR} \cong E^{1,n}_{2,rig}$ from Theorem 4.8 since they are not useful to us.

We introduce notation needed for the statement of Theorem 4.8: For each k with $1 \leqslant k \leqslant m$ denote by e_k the element of K^m with 1 in position k and 0 elsewhere. For $\ell \geqslant \rho$, $0 \leqslant j < d$ and $1 \leqslant k \leqslant m$ define

$$u^{(r,\ell,j,k)} = e_k \frac{\Gamma^j}{r^{\ell+1}} \otimes d\Gamma \in \mathcal{E} \otimes d\Gamma.$$

Apply the algorithm in the first stage of the proof of Theorem 4.2 to compute

$$v^{(r,\ell,j,k)} = \sum_{i=\rho}^{\ell} \frac{V_i^{(r,\ell,j,k)}}{r^i},\ V_i^{(r,\ell,j,k)} \in K[\Gamma]^m,\ \deg(V_i^{(r,\ell,j,k)}) < d$$

such that $u^{(r,\ell,j,k)} - \nabla(v^{(r,\ell,j,k)}) =: w^{(r,\ell,j,k)} \in \mathcal{E} \otimes d\Gamma_\rho$. Similarly, for any $\ell \geqslant \rho$ we may apply the algorithm in the second stage of the proof of Theorem 4.2 to write

$$u^{(\infty,\ell,k)} - \nabla(v^{(\infty,\ell,k)}) = w^{(\infty,\ell,k)} \in \mathcal{E} \otimes d\Gamma_\rho$$

where this time

$$u^{(\infty,\ell,k)} := e_k \Gamma^{\ell-1} \otimes d\Gamma, \text{ and } v^{(\infty,\ell,k)} = \sum_{i=\rho}^{\ell} V_i^{(\infty,\ell,k)} \Gamma^i \text{ for some } V_i^{(\infty,\ell,k)} \in K^m.$$

We have the following effective bounds on the growth of forms during reduction c.f. [**25**, Lemma 2].[2]

THEOREM 4.8. *Assume that conditions (Prep.Rat.) and (O.C.) hold. Then for* $w \in \{w^{(r,\ell,j,k)}, w^{(\infty,\ell,k)}\}$ *we have*

$$\mathrm{ord}_p(w) \geqslant -\big(\alpha \log_p(\ell - \rho) + \beta\big)$$

for some effective constants $\alpha, \beta \in \mathbb{Q}$ *which depend only upon the connection* ∇ *and the basis* $\mathcal{B}$*, i.e., are independent of the starting form* $u \in \{u^{(r,\ell,j,k)}, u^{(\infty,\ell,k)}\}$.

We shall make the constants α, β completely explicit in Note 4.11. We note that Theorem 4.8 also holds with the same constants if one applies the variant algorithm for reducing pole orders at infinity given at the end of the proof of Theorem 4.2. Since the forms $u^{(r,\ell,j,k)}$ and $u^{(\infty,\ell,k)}$ span $\mathcal{E} \otimes d\Gamma$ as a K-vector space, the above theorem allows one to deduce bounds on the growth of arbitrary forms during the reduction algorithm.

The proof of Theorem 4.8 will be reduced by a localisation argument to that of giving effective bounds on the p-adic convergence of the uniform part of the local solution matrix to a differential system at a regular singular point. Such bounds are provided in Lemma 4.9, whose proof in turn relies on a deep theorem of Christol-Dwork-Gerotto-Sullivan [**13**, Chap V], and an elementary result of Clark (Lemma 4.10).

---

[2]For notational convenience, in the statement of the theorem and also inequalities (8), (10), and (17) the expression $\log_p(\ell - \rho)$ occurs, when the argument could be zero; similarly $\log_p(j)$ where $j = 0$ occurs in inequality (14). In these cases "$\log_p(0)$" should be understood to be zero.

Proof. Since $b(\Gamma) \in M_m(\mathcal{O}_K[\Gamma])$ and $\mathrm{ord}_p(r(\Gamma)) = 0$, from the equation "$w = u - \nabla(v)$" we see that it suffices to prove that for $v \in \{v^{(r,\ell,j,k)}, v^{(\infty,\ell,k)}\}$ we have

$$\mathrm{ord}_p(v) \geqslant -\big(\alpha \log_p(\ell - \rho) + \beta\big) \tag{8}$$

for some effective constants $\alpha, \beta \in \mathbb{Q}$ which depend only upon the connection ∇ and the basis $\mathcal{B}$.

We divide the proof of (8) into three steps:

- **Step 1**: We reduce proving bound (8) to proving the local bounds (10). Here we need that r is squarefree modulo p.
- **Step 2**: We reduce proving each local bound (10) to proving a different local bound (11). Here we need assumption (Prep.Rat.).
- **Step 3**: Bound (11) is deduced from an effective version of a theorem of Christol. This step uses assumptions (Prep.Rat.) and (O.C.).

Recall that ρ is defined to be the smallest integer larger than all integers in the exponent set $E(\nabla, \mathcal{B})$. We note that the argument we give works for *any* "ρ" larger than every integer in the exponent set $E(\nabla, \mathcal{B})$.

Step 1: First let us consider the case that $u := u^{(r,\ell,j,k)}$ for some $\ell \geqslant \rho$, $0 \leqslant j < d$ and $1 \leqslant k \leqslant m$. Let us simplify the notation above by removing the exponent "(r, ℓ, j, k)" where it occurs, i.e., $v := v^{(r,\ell,j,k)}$, $w := w^{(r,\ell,j,k)}$ and $V_i := V_i^{(r,\ell,j,k)}$ etc. Let $\gamma \in R$ be a root of $r(\Gamma) = 0$. Let $t_\gamma = \Gamma - \gamma$ and expand v locally as

$$v(t_\gamma) = v_{\gamma,\ell} t_\gamma^{-\ell} + \cdots + v_{\gamma,\rho} t_\gamma^{-\rho} + \ldots, \; v_{\gamma,i} \in K(\gamma)^m. \tag{9}$$

We show now that (8) holds for $v = v^{(r,\ell,j,k)}$ provided that

$$\mathrm{ord}_p(v_{\gamma,i}) \geqslant -(\alpha \log_p(\ell - \rho) + \beta) \text{ for all } \gamma \in R \text{ and } \rho \leqslant i \leqslant \ell. \tag{10}$$

Assume (10) holds. We claim that $\mathrm{ord}_p(V_i) \geqslant -(\alpha \log_p(\ell - \rho) + \beta)$ for $\rho \leqslant i \leqslant \ell$, from which (8) follows immediately. This claim can be proved by descending induction on i in this range. Less formally, observe that $v_{\gamma,\ell}$ is just $V_\ell(\gamma) r'(\gamma)^{-1}$. Since the roots of r are distinct modulo p we have $\mathrm{ord}_p(V_\ell(\gamma)) = \mathrm{ord}_p(v_{\gamma,\ell})$. Since $\deg(V_\ell) < d = |R|$, from (10) we deduce the claimed bound on $\mathrm{ord}_p(V_\ell(\Gamma))$. Now subtract V_ℓ / r^ℓ from both sides of (9) and repeat the argument for $i = \ell - 1$, and so on.

Similarly, assuming (10) holds for the coefficients in the local expansion of $v := v^{(\infty,\ell,k)}$ at infinity, we easily deduce that (8) holds for $v = v^{(\infty,\ell,k)}$.

It remains to establish the local bound (10). (We omit the remainder of the proof for $v = v^{(\infty,\ell,k)}$ since it is exactly the same.)

Step 2: Fix $\gamma \in R$ and simplify notation as in Step 1. Define $G(t) \in M_m(\mathcal{O}_{K(\gamma)})[[t]]$ so that $-t^{-1}G(t)$ is the expansion of $b(\Gamma)/r(\Gamma)$ w.r.t the local parameter $t := t_\gamma = \Gamma - \gamma$. Define $H := G(0)$, the negative of the residue matrix $b(\gamma)/r'(\gamma)$. Let the local solution matrix $Y(t)t^H$ to the differential system $t\frac{d}{dt} - G(t) = 0$ be defined as in [**13**, Chap. III Prop. 8.5]. (Note that we have chosen our signs to be consistent with [**13**].) The existence of such a solution matrix requires the assumption (Prep.Rat.). The uniform part $Y(t)$ lies in $M_m(K(\gamma))[[t]]$ with $Y(0) = I_m$, and the element t^H, which is constructed on [**13**, Page 103], satisfies the equation $\frac{d}{dt}(t^H) = t^{-1} H t^H$, and $(t^H)^{-1} = t^{-H}$. We note that by definition t^H is an $m \times m$ matrix containing polynomial expressions in appropriate fractional powers of t and an element "$\log(t)$"; however, we shall just manipulate it in a formal manner, exploiting the various properties it possesses.

Write $Y = \sum_{i=0}^{\infty} Y_i t^i$ and $Y(t)^{-1} = \sum_{i=0}^{\infty} Z_i t^i$, where $Y_i, Z_i \in M_m(K(\gamma))$. We shall show now that it is enough to prove that

$$\mathrm{ord}_p(Y_i),\, \mathrm{ord}_p(Z_i) \geqslant -(\alpha_1 \log_p(i) + \beta_1),\ i \geqslant 1 \tag{11}$$

for some explicit $\alpha_1, \beta_1 \in \mathbb{Q}$ which depend only on ∇ and $\mathcal{B}$.

Let us assume (11) holds. Observe that we have the local factorisation

$$\nabla = Y(t)t^H \circ \frac{d}{dt} \circ t^{-H} Y(t)^{-1}. \tag{12}$$

Premultiplying the localised equation $\nabla(v) = u - w$ by $(Y(t)t^H)^{-1}$, using (12), and then integrating, we find that

$$v(t) = Y(t)t^H \left\{ \int t^{-H} Y(t)^{-1}(u-w)dt + c \right\}, \tag{13}$$

for some constant $c \in K(\gamma)^m$. Bound (10) can now be deduce by explicitly integrating the right hand side of (13) and comparing coefficients of t^{-i} for $\rho \leqslant i \leqslant \ell$.

Specifically, the integrand on the right hand side of (13) can be written

$$\sum_{j \geqslant 0} t^{-H} a_j t^{-(\ell+1)+j},\ \mathrm{ord}_p(a_j) \geqslant -(\alpha_1 \log_p(j) + \beta_1) \text{ for } 0 \leqslant j < \ell + 1 - \rho, \tag{14}$$

for some $a_j \in K(\gamma)^m$. The lower bound on $\mathrm{ord}_p(a_j)$ comes from (11) and the integrality of $u(t)$. Note that we do not have any bounds on $\mathrm{ord}_p(a_j)$ for $j \geqslant \ell + 1 - \rho$ since these terms are affected by the unknown element $w(t)$ and unknown constant c. Element (14) may be explicitly integrated "term–by-term". Precisely, from the defining property of the element t^H, ones sees that $\int t^{-H} a_j t^{-(\ell+1)+j} dt = (-H-(\ell+1)+j+1)^{-1} t^{-H} a_j t^{-(\ell+1)+j+1}$, plus an unknown constant of integration. Recall that $-H$ is the residue matrix. Now for $0 \leqslant j < \ell + 1 - \rho$ we have that $-\ell \leqslant -(\ell+1) + j + 1 \leqslant -\rho$, and since ρ is larger than any eigenvalue of $-H$ the inverse matrix immediately above exists. (For $j \geqslant \ell + 1 - \rho$, when the inverse does not exist the coefficient a_j must be zero.) Next, note that $(-H + i)^{-1}$ for $i \in \mathbb{Z}$ (when it exists) commutes with t^H; this follows from the fact that H commutes with t^H, see [**13**, Page 103]. Thus each term on the right hand side of (13) which does not involve the constant c has the form

$$Y_i t^i (t^H)(-H-(\ell+1)+j+1)^{-1} t^{-H} a_j t^{-(\ell+1)+j+1} = Y_i (-H-\ell+j)^{-1} a_j t^{-\ell+i+j} \tag{15}$$

for some $i, j \geqslant 0$. Terms on the right hand side of (13) which do involve c have the form $Y_i t^i t^H c$ for some $i \geqslant 0$. Since the left hand side $v(t)$ is a Laurent series, it follows from [**13**, Chap V Lemma 2.3] that either $c = 0$, or $c \neq 0$ with H a diagonal matrix with integer eigenvalues. Since all eigenvalues of $-H$ are less than ρ, in either case any term on the right hand side of (13) involving c cannot effect the coefficient of t^{-i} for $\rho \leqslant i \leqslant \ell$.

From (15), a lower bound on the coefficient of $t^{-\ell+s}$ for $0 \leqslant s \leqslant \ell - \rho$ on the right hand side of (13) is

$$\min_{i+j=s} \mathrm{ord}_p \left(Y_i (-H - \ell + j)^{-1} a_j \right). \tag{16}$$

We have bounds on $\mathrm{ord}_p(Y_i)$ and $\mathrm{ord}_p(a_j)$ for $i \geqslant 0$ and $0 \leqslant j \leqslant \ell - \rho$, viz. (11) and (14). It remains to bound $\mathrm{ord}_p((-H - \ell + j)^{-1})$ for $0 \leqslant j \leqslant \ell - \rho$. Now $\mathrm{ord}_p((-H-\ell+j)^{-1}) \geqslant -\mathrm{ord}_p(\det(-H-\ell+j))$ so we must find an upper bound for

the valuation of the determinant. Denote by $\lambda_1, \ldots, \lambda_m \in \mathbb{Q}$ the eigenvalues of $-H$ (the residue matrix). Then $\det(-H - \ell + j) = \prod_{i=1}^{m}(\lambda_i - \ell + j)$. Take the positive integer N to be a lowest common denominator for the λ_i and define $\mu_i = N\lambda_i \in \mathbb{Z}$; note that $\gcd(p, N) = 1$ since the eigenvalues are p-adic integers. Take the positive integer Δ to be minimal so that $|\lambda_i| \leqslant \Delta$ for all i. Then for $0 \leqslant j \leqslant \ell - \rho$ we have

$$\mathrm{ord}_p(\lambda_i - \ell + j) = \mathrm{ord}_p(\mu_i - N(\ell - j)) \leqslant \log_p(|\mu_i| + N\ell) \leqslant \log_p(N) + \log_p(\Delta + \ell).$$

So certainly for $0 \leqslant j \leqslant \ell - \rho$ we have

$$\mathrm{ord}_p(\det(-H - \ell + j)) \leqslant m\left(\log_p(\ell - \rho) + \log_p(N) + \log_p(2\Delta + 2)\right), \tag{17}$$

since $\rho \leqslant \Delta + 1$. From (16) and (17) we conclude that

$$\alpha := \lceil 2\alpha_1 + m \rceil,\ \beta := \lceil 2\beta_1 + m(\log_p(N) + \log_p(2\Delta + 2)) \rceil \tag{18}$$

will certainly suffice.

Step 3: We now establish bound (11). First consider $Y(t)$. By assumptions (Prep.Rat.) and (O.C.), we see that the hypotheses of Lemma 4.9 are met. Hence we may apply the bound in Lemma 4.9 to our differential system $t\frac{d}{dt} - G(t) = 0$. We next note that $Y(t)^{-1}$ is the uniform part of the local solution matrix of the "dual" differential system $t\frac{d}{dt} + G(t) = 0$ where $G(t)$ acts on the right, or equivalently, the transpose of the uniform part of the local solution matrix of $t\frac{d}{dt} + G(t)^{tr} = 0$ with $G(t)^{tr}$ acting on the left, c.f. [**13**, Page 193]. Again by assumptions (Prep.Rat.) and (O.C.) the hypotheses of Lemma 4.9 below are met, and we may use that bound. □

The next lemma is a modest generalisation of the main theorem in [**13**, Chap. V].

LEMMA 4.9. *Let $G(t) \in M_m(\mathcal{O}_{K(\gamma)})[[t]]$ be the local expansion of a rational function $G(\Gamma) \in M_m(K(\Gamma))$ around some point $t = \Gamma - \gamma$. Let $\delta := t\frac{d}{dt}$ and consider the differential system $\delta - G = 0$. Assume*

1. *The eigenvalues of $G(0)$ are rational numbers, and no two differ by a positive integer,*
2. *The solution matrix to the differential system around the generic point t with $|t|_p = 1$ converges p-adically on the open unit disk.*

Let $Y(t) = \sum_{i=0}^{\infty} Y^i t^i \in M_m(K(\gamma))[[t]]$ be the uniform part of the local solution matrix $Y(t)t^{G(0)}$ of the differential system $\delta - G = 0$. Then there exist $\alpha_1, \beta_1 \in \mathbb{R}$ such that

$$\mathrm{ord}_p(Y_i) \geqslant -(\alpha_1 \log_p(i) + \beta_1), \text{ for all } i \geqslant 1.$$

Proof. First, change basis by a matrix $\mathcal{H} \in GL_m(K(\gamma)[t, t^{-1}])$ so that

$$G_{[\mathcal{H}]} := \mathcal{H}^{-1}t\frac{d\mathcal{H}}{dt} + \mathcal{H}^{-1}G\mathcal{H}$$

is such that $G_{[\mathcal{H}]}(0)$ has eigenvalues in the interval $[0, 1)$. By [**13**, Chap. V, Prop. 4.1] the matrix $\mathcal{H}$ may be taken to be unimodular, i.e., $\mathrm{ord}_p(\mathcal{H}), \mathrm{ord}_p(\mathcal{H}^{-1}) \geqslant 0$. Moreover, the degree in t of $\mathcal{H}$ (degree in t^{-1} of $\mathcal{H}^{-1}$) is the absolute value of the floor of the most negative eigenvalue of $G(0)$, and the degree in t^{-1} of $\mathcal{H}$ (degree

in t of $\mathcal{H}^{-1}$) is the floor of the most positive eigenvalue of $G(0)$; this is easily seen by viewing $\mathcal{H}$ and $\mathcal{H}^{-1}$ as a product of shearing transformations.

Let $\tilde{Y}(t) = \sum_{i=0}^{\infty} \tilde{Y}_i t^i$ be the uniform part of the local solution matrix to the differential system $\delta - G_{[\mathcal{H}]}$. Note that $G_{[\mathcal{H}]} \in M_m(\mathcal{O}_{K(\gamma)})[[t]]$ by the unimodularity of $\mathcal{H}$.

The main theorem in [**13**, Chap. V Section 9] assures us that there exists $\alpha_2, \beta_2 \in \mathbb{R}$ such that

$$\mathrm{ord}_p(\tilde{Y}_i) \geqslant -(\alpha_2 \log_p(i) + \beta_2),\ i \geqslant 1.$$

We comment briefly on why the main theorem is applicable: In the notation of [**13**, Chap V], we must check conditions $\mathcal{R}1$, $\mathcal{R}2$, $\mathcal{R}3'$ and $\mathcal{R}4$. Now $\mathcal{R}1$ is true since the matrix $G_{[\mathcal{H}]}$ contains functions which are localisations of rational functions; $\mathcal{R}2$ (overconvergence) follows from assumption 2. and the unimodularity of $\mathcal{H}$; $\mathcal{R}3'$ (eigenvalues in $\mathbb{Z}_p \cap [0,1)$) is true by assumption 1.; $\mathcal{R}4$ (integrality) follows since $G_{[\mathcal{H}]} \in M_m(\mathcal{O}_{K(\gamma)})$.

There exists $\bar{\mathcal{H}} \in GL_m(K[t,t^{-1}])$ such that $Y(t) = \mathcal{H}^{-1}\tilde{Y}\bar{\mathcal{H}}$, c.f. [**13**, Page 163 Lines 1-6]. Moreover, the degree in t of $\bar{\mathcal{H}}$ is the floor of the most positive eigenvalue of $G(0)$, and the degree in t^{-1} of $\bar{\mathcal{H}}$ is the absolute value of the floor of the most negative eigenvalue of $G(0)$; this follows from the argument on [**13**, Page 163 Lines 11-21].

Since $\mathrm{ord}_p(\mathcal{H}^{-1}) \geqslant 0$, to prove the lemma we need only calculate a lower bound on $\mathrm{ord}_p(\bar{\mathcal{H}})$. One computes this from the equation $\bar{\mathcal{H}} = \tilde{Y}^{-1}\mathcal{H}Y$ using bounds on the degree in t^{-1} of $\mathcal{H}$ and the naive upper bound on the growth of the coefficients of $\tilde{Y}(t)^{-1}$ and $Y(t)$, c.f. [**13**, Page 191-193]. The naive upper bounds are given in Lemma 4.10. Specifically, one finds certainly

$$\mathrm{ord}_p(\bar{\mathcal{H}}) \geqslant -\beta_3,\ \beta_3 := m^2\left(\frac{\Delta}{p-1} + 4\log_p(\Delta+1) + 2\log_p(2N)\right) \tag{19}$$

where Δ is such that all eigenvalues λ of $G(0)$ have $|\lambda| \leqslant \Delta$ and N the lowest common denominator for the eigenvalues. Note that we have already observed $\deg_{t^{-1}}(\mathcal{H}^{-1}), \deg_{t^{-1}}(\bar{\mathcal{H}}) \leqslant \Delta$. Comparing coefficients in $Y = \mathcal{H}^{-1}\tilde{Y}\bar{\mathcal{H}}$ we see

$$\mathrm{ord}_p(Y_i) \geqslant \mathrm{ord}_p(\tilde{Y}_{i+2\Delta}) - \beta_3 \geqslant -(\alpha_2\log_p(i) + \alpha_2\log_p(2\Delta+1) + \beta_2 + \beta_3)$$

which gives a bound of the required form. Precisely, take

$$\alpha_1 := \alpha_2,\ \beta_1 := \alpha_2\log_p(2\Delta+1) + \beta_2 + \beta_3. \tag{20}$$

□

The following lemma is an effective version of the general bound of Clark [**8, 39**].

Lemma 4.10. *Situation as in the statement of Lemma 4.9, only without assumption 2. Denote by $\lambda_1, \ldots, \lambda_m$ the eigenvalues of $G(0)$ and assume that $|\lambda_i| \leqslant \Delta$ and $N\lambda_i \in \mathbb{Z}$ for all i and some minimal integers $\Delta \geqslant 0$, $N \geqslant 1$ with $\gcd(p,N) = 1$. Then for all $s \geqslant 1$ we have*

$$\mathrm{ord}_p(Y_s) \geqslant -m^2\left(\frac{s}{p-1} + \log_p(1+s) + \log_p(2\Delta+1) + \log_p(N)\right).$$

Proof. The power series $Y(t)$ may be computed using the classical method in [**13**, Chap V, Remark 2.2]. It follows from this recursive method that

$$\mathrm{ord}_p(Y_s) \geqslant -\sum_{k=1}^{s}\sum_{i,j=1}^{m} \mathrm{ord}_p(k-\delta_{ij})$$

where $\delta_{ij} := \lambda_i - \lambda_j$. We shall estimate the right hand side using the method of Clark. Specifically, for $\alpha \in \mathbb{Z}_p$ and s a positive integer, define $\theta(\alpha, s) := \prod_{k=1}^{s}(\alpha + k)$. Then $\mathrm{ord}_p(Y_s) \geqslant -\sum_{i,j} \mathrm{ord}_p\theta(-\delta_{ij}, s)$. We compute an upper bound for the $\theta(-\delta_{ij}, s)$ using the argument in [**8**, Page 265, Case 3]. First, write $-\delta_{ij} = \nu_{ij}/N$, where $|\nu_{ij}| \leqslant 2N\Delta$. Then for any positive integer s we have

$$\mathrm{ord}_p(-\delta_{ij}+s) = \mathrm{ord}_p(\nu_{ij}+Ns) \leqslant \log_p(|\nu_{ij}|+Ns) \leqslant \log_p(N)+\log_p(2\Delta+1)+\log_p(s).$$

This shows that we may take the expression "$v(x) = -k\log_p(1+x) - k'$" immediately preceding [**8**, Eqn (13)], to have coefficients "$k = 1$" and "$k' = \log_p(2\Delta + 1) + \log_p(N)$" (we have changed Clark's "log" to "$\log_p$"). From [**8**, Eqn (14)] we deduce that

$$\mathrm{ord}_p\theta(-\delta_{ij}, s) \leqslant \frac{s}{p-1} - v(s) = \frac{s}{p-1} + \log_p(1+s) + \log_p(2\Delta+1) + \log_p(N)$$

as required. □

NOTE 4.11 The constants $\alpha, \beta \in \mathbb{Q}$ in Theorem 4.8 can be made completely explicit. Precisely, by equations (18), (19) and (20) one sees that it suffices to make the constants α_2 and β_2 from the proof of Lemma 4.9 explicit. These constants are those which occur in the theorem of Christol-Dwork-Gerotto-Sullivan.

The theorem of CDGS states that $\mathrm{ord}_p(Y_i) \geqslant -(\alpha_2\lfloor\log_p(i)\rfloor+\beta_2)$ with $\alpha_2, \beta_2 \in \mathbb{R}$ as follows. Define

$$B_{m,p} := m - 1 + \mathrm{ord}_p((m-1)!) + \min\left\{m-1, \mathrm{ord}_p\prod_{j=1}^{m}\binom{m}{j}\right\}. \qquad (21)$$

When all the eigenvalues are zero (nilpotent case) we can directly apply the Christol-Dwork theorem [**13**, Chap. V, Thm 2.1]. This gives $\alpha_2 = B_{m,p}$ and $\beta_2 = 0$. In particular, for $p \geqslant m$ we have $\alpha_2 = m - 1$, and in general $\alpha_2 \leqslant 3m - 3$. For eigenvalues in the interval $[0, 1)$, one applies the generalisation in [**13**, Sec. 9, Chap. V]. Define $\ell := \lfloor\log_p(i)\rfloor + 1$. From the equation on middle of page 198 in [**13**] one deduces that

$$\mathrm{ord}_p(Y_i) \geqslant -\{(\ell+1)(m(m-1)) + \ell B_{m,p} + B\}$$

where the number B is defined on [**13**, Pages 197-198]. (Note that we have changed from multiplicative to additive notation, so our "B" corresponds to "$\log_p(B)$".) Let N be the lowest common denominator for the eigenvalues of the residue matrix. Then for $p > \{m, 2N\}$ from [**13**, Remark 2.2] one sees that $B_{m,p} = m-1$ and $B = 0$, so $\alpha_2 = m^2 - 1$ and $\beta_2 = 2m^2 - m - 1$. For general p, one computes from [**13**, Page 198, Line 7] and [**13**, Page 197, Line 12] that $B \leqslant (\ell+1)m(m-1)\log_p(2N)$, and as observed before $B_{m,p} \leqslant 3m - 3$. Thus we may take

$$\alpha_2 = m(m-1)(1+\log_p(2N)) + 3m - 3,\ \beta_2 = 2m(m-1)(1+\log_p(2N)) + 3m - 3.$$

We conclude the following: if $\Delta \geqslant 0$ is a bound on the absolute value of the local exponents, $N \geqslant 1$ a lowest common denominator, and m the dimension, then one has

$$\alpha = C_1 m^2 (1 + \log_p(N)),\ \beta = C_2 m^2 \left(1 + \frac{\Delta}{p} + \log_p(\Delta + 1) + \log_p(N)\right) \qquad (22)$$

for some *absolute* constants $C_1, C_2 \in \mathbb{Q}$. This will be useful in our complexity estimates.

In Section 6.2 we shall see that conditions (Rat.) and (O.C.) are met in the situations (which arise "from geometry") which we shall encounter. The stronger condition (Prep.Rat.) will be met in the examples we compute in Section 9.

Theorem 4.8 is essential in the complexity analysis and practical application of the algorithm in the proof of Theorem 4.2 for the following reason: When one calculates the reduction of differential forms using this algorithm it is impractical to store the coefficients "exactly". At each step of the reduction, one "approximates" the coefficients modulo some fixed power of the characteristic. Making this approximation amounts to adding an "error form" to the form being reduced. Theorem 4.8 shows that the error introduced propagates in a "logarithmic" manner during the remainder of the computation. Furthermore, Theorem 4.8 applied with a general value "ρ", at least as big as ρ itself, allows one to bound the intermediate coefficient size during the reduction computation. Note that a naive inspection of the reduction formulae in the proof of Theorem 4.2 suggests that terms grow and errors propagate in a "linear" manner; that this is *not* the case for calculations in rigid cohomology was an important insight of Kedlaya c.f. [**25**, Lemma 2].

A similar "logarithmic error propagation" phenomenon arises in the numerical solution of differential systems, as we shall see in Theorem 5.1 of the next section.

5. *Deformation of Frobenius*

In this section we retain the notation and definitions in the first paragraph of Section 4.1, but slightly alter some of our assumptions. Specifically, $\mathcal{E}$ is a free A-module of rank m, where $A = K[\Gamma, 1/r(\Gamma)]$ and K is the unramified extension of $\mathbb{Q}_p$ of degree $[\mathbb{F}_q : \mathbb{F}_p]$. The map $\nabla : \mathcal{E} \to \mathcal{E} \otimes \Omega^1_A$ is a connection which with respect to a fixed basis of $\mathcal{E}$ (and "natural" corresponding basis of $\mathcal{E} \otimes \Omega^1_A$) acts as:

$$\nabla = \frac{d}{d\Gamma} + B(\Gamma),\ B(\Gamma) = \frac{b(\Gamma)}{r(\Gamma)}.$$

Here $b(\Gamma) \in M_m(\mathcal{O}_K[\Gamma])$. In this section we *do not* assume that the connection has only simple poles, i.e., we do not need the assumption that $r(\Gamma) \in \mathcal{O}_K[\Gamma]$ is squarefree, nor do we need the degree restriction on the matrix $b(\Gamma)$. However, we shall add the new assumption that $r(0) \not\equiv 0 \bmod p$.

Let $\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes \Omega^1_{A^\dagger}$ be obtained from ∇ by extension of scalars; the ring $A^\dagger = K[\Gamma, 1/r(\Gamma)]^\dagger$ is described explicitly in Section 3.4. Let $\sigma : A^\dagger \to A^\dagger$ be the lifting of the pth power Frobenius which maps $\Gamma \mapsto \Gamma^p$. Let $F : \mathcal{E}^\dagger \to \mathcal{E}^\dagger$ be a σ-linear map such that the triple $(\mathcal{E}^\dagger, \nabla^\dagger, F)$ defines a $(\sigma, \nabla^\dagger)$-module over $A^\dagger$ (Definition 3.5); in particular, Diagram (3) commutes.

5.1. *Local deformation*

If we assume that the connection matrix $B(\Gamma)$ is known and that the specialisation $F(0)$ is also known, the commutative diagram (3) allows the computation of a *local* expansion of the Frobenius matrix $F(\Gamma)$ around the origin to any required precision. We describe two different approaches.

5.1.1. Method 1

Let $C(\Gamma)$ be a basis of local solutions to the differential system $\nabla = 0$ with initial condition $C(0) = I_m$. So

$$\frac{dC}{d\Gamma} + B(\Gamma)C(\Gamma) = 0. \tag{23}$$

Commutativity of (3) implies that the Frobenius map F preserves the kernel of the connection. Recalling that the map F is σ-linear, we deduce the matrix equation

$$F(\Gamma)C^{\sigma}(\Gamma^p) = C(\Gamma)D$$

where D is some constant matrix. Evaluating both sides at $\Gamma = 0$ shows $D = F(0)$. So we have the local factorisation

$$F(\Gamma) = C(\Gamma)F(0)(C^{\sigma}(\Gamma^p))^{-1}. \tag{24}$$

Thus $F(\Gamma)$ can be computed modulo Γ^{N_Γ} for any $N_\Gamma \geqslant 1$ provided we can compute $C(\Gamma)$ modulo Γ^{N_Γ}. A simple recursion formula for computing the matrix coefficients in the local expansion of $C(\Gamma) = \sum_{\ell=0}^{\infty} C_\ell \Gamma^\ell$ can be derived from the equation

$$r(\Gamma)\frac{dC}{d\Gamma} + b(\Gamma)C(\Gamma) = 0.$$

Specifically, write $r(\Gamma) = \sum_{i=0}^{\deg(r)} r_i\Gamma^i$ and $b(\Gamma) = \sum_{i=0}^{\deg(b)} b_i\Gamma^i$. Then for $\ell \geqslant 1$ we have

$$C_\ell = -\frac{1}{r_0\ell}\left(\sum_{i=0}^{\deg(b)} b_i C_{(\ell-1)-i} + \sum_{i=1}^{\deg(r)} r_i(\ell-i)C_{\ell-i}\right). \tag{25}$$

One can, of course, compute the series $C^{\sigma}(\Gamma^p)^{-1}$ by power series inversion. However, it is better to observe that the matrix $C(\Gamma)^{-1}$ is the solution of the "dual equation"

$$\frac{dC^{-1}}{d\Gamma} - C(\Gamma)^{-1}B(\Gamma) = 0,\ C(0)^{-1} = I_m. \tag{26}$$

It is impractical to carry out the above computations using "exact arithmetic"; one desires to "truncate" each coefficient C_ℓ "modulo p^N" for some appropriate $N > 0$ after it has been computed. It is an essential task to analyse the "propagation" of the error this introduces as one continues the computation.

Let $E_\ell \in M_m(\mathcal{O}_K)$ for $\ell \geqslant 1$, and N be a non-negative integer. Let the sequence $D_\ell \in M_m(K)$ for $\ell \geqslant 0$ be computed in the following manner. Define $D_0 := I_m$ and for $\ell \geqslant 1$,

$$D_\ell := -\frac{1}{r_0\ell}\left(\sum_{i=0}^{\deg(b)} b_i D_{(\ell-1)-i} + \sum_{i=1}^{\deg(r)} r_i(\ell-i)D_{\ell-i}\right) + p^N E_\ell. \tag{27}$$

The series $D(\Gamma) := \sum_{\ell=0}^{\infty} D_\ell\Gamma^\ell$ is an "approximate solution" to the differential

system (23) computed "modulo p^N". In practice, the "error sequence" E_ℓ is chosen to ensure the p-adic expansions of the entries of D_ℓ are "truncated modulo p^N".

THEOREM 5.1. *Let $C(\Gamma)$ be the solution to (23) with $C(0) = I_m$, and $D(\Gamma)$ the "approximate solution" defined via equation (27). Then for $\ell \geqslant 0$ we have*

$$\mathrm{ord}_p(D_\ell - C_\ell) \geqslant N - \alpha' \lfloor \log_p(\ell + 1) \rfloor$$

for some explicitly computable constant $\alpha' \geqslant 0$. Furthermore, one can take $\alpha' = 6m - 5$ for any prime p, and $\alpha' = 2m - 1$ when $p \geqslant m$.

Proof. First observe that we have the local factorisation around the origin

$$\nabla = C(\Gamma) \circ \frac{d}{d\Gamma} \circ C(\Gamma)^{-1}. \tag{28}$$

Next observe that the series $D(\Gamma)$ is a local solution to the inhomogeneous differential equation

$$r(\Gamma)\frac{dD}{d\Gamma} + b(\Gamma)D(\Gamma) = p^N r_0 \sum_{\ell=1}^{\infty} \ell E_\ell \Gamma^\ell.$$

Thus $\nabla(D) = p^N E(\Gamma) r(\Gamma)^{-1}$ where $E(\Gamma) := r_0 \sum_{\ell=1}^{\infty} \ell E_\ell \Gamma^\ell$. Using the local factorisation (28) one deduces that

$$\frac{d}{d\Gamma}\left(C(\Gamma)^{-1} D(\Gamma)\right) = C(\Gamma)^{-1} p^N E(\Gamma) r(\Gamma)^{-1}.$$

Integrating we find that there exists a constant matrix c such that

$$D(\Gamma) = C(\Gamma)\left(\int C(\Gamma)^{-1} p^N E(\Gamma) r(\Gamma)^{-1} d\Gamma + c\right)$$

Note that $E(\Gamma)r(\Gamma)^{-1} \in \Gamma \mathcal{O}_K[[\Gamma]]$ since $r(0)$ is a unit. Since $C(\Gamma) = D(\Gamma) \bmod \Gamma$ we deduce that $c = I_m$. Hence

$$D(\Gamma) - C(\Gamma) = p^N C(\Gamma) \int C(\Gamma)^{-1} E(\Gamma) r(\Gamma)^{-1} d\Gamma.$$

The connection ∇ and its dual $\check{\nabla}$ come from overconvergent F-isocrystals, viz, $(\mathcal{E}^\dagger, \nabla^\dagger, F)$ and its "dual $(\mathcal{E}^\dagger, \check{\nabla}^\dagger, F^{-1})$". Hence Dwork's trick of analytic continuation via Frobenius [**22**, Prop. 3.1.2] shows that condition (O.C.) is met. Moreover, the connections are regular at zero, so local exponents are all zero. Thus we can apply the Christol-Dwork theorem to deduce effective logarithmic bounds on the growth of coefficients of $C(\Gamma)$ and $C(\Gamma)^{-1}$. Moreover, integration only has a "logarithmic" effect on the growth of coefficients of a power series. Explicitly, we can use the constant $B_{m,p}$ in the original Christol-Dwork theorem, see (21) in Note 4.11, to deduce $\alpha' = 2B_{m,p} + 1$ and the constant "$\beta' = 0$". This completes the proof. □

Let N_Γ and N be positive integers. Let $D(\Gamma)$ be an "approximate solution" computed "modulo p^N" to the differential system (23) modulo Γ^{N_Γ}. Let $\tilde{D}(\Gamma)$ be an "approximate solution" computed "modulo p^N" to the dual system (26) modulo $\Gamma^{\lceil N_\Gamma/p \rceil}$. Let $G(0) \in M_m(K)$ be such that $\mathrm{ord}_p(F(0) - G(0)) \geqslant p^N$. Define $G(\Gamma) := D(\Gamma)G(0)\tilde{D}^\sigma(\Gamma^p) \bmod \Gamma^{N_\Gamma}$. This is our approximation of the local Frobenius matrix $F(\Gamma)$. We need to bound from below $\mathrm{ord}_p\left((F(\Gamma) \bmod \Gamma^{N_\Gamma}) - G(\Gamma)\right)$.

From Theorem 5.1, $D(\Gamma) = C(\Gamma) + p^{N'} e(\Gamma) \bmod \Gamma^{N_\Gamma}$ where $\mathrm{ord}_p(e(\Gamma)) \geqslant 0$ with $N' := N - (2B_{m,p}+1)\lfloor \log_p(N_\Gamma)\rfloor$, and $\tilde{D}^\sigma(\Gamma^p) = (C^\sigma(\Gamma^p))^{-1} + p^{N''}\tilde{e}(\Gamma) \bmod \Gamma^{N_\Gamma}$ where $\mathrm{ord}_p(\tilde{e}(\Gamma)) \geqslant 0$ with $N'' := N - (2B_{m,p}+1)\lfloor \log_p(\lceil N_\Gamma/p\rceil)\rfloor$. Note that from the Christol-Dwork theorem we have

$$\begin{aligned} \mathrm{ord}_p\left(C(\Gamma) \bmod \Gamma^{N_\Gamma}\right) &\geqslant -B_{m,p}\lfloor \log_p(N_\Gamma - 1)\rfloor \\ \mathrm{ord}_p\left((C^\sigma(\Gamma^p))^{-1} \bmod \Gamma^{N_\Gamma}\right) &\geqslant -B_{m,p}\lfloor \log_p(\lceil N_\Gamma/p\rceil - 1)\rfloor. \end{aligned}$$

One now readily calculates a lower bound on

$$\mathrm{ord}_p\left(\left(C(\Gamma)F(0)(C^\sigma(\Gamma^p))^{-1} \bmod \Gamma^{N_\Gamma}\right) - D(\Gamma)G(0)\tilde{D}^\sigma(\Gamma^p)\right)$$

to be

$$\begin{aligned} &\min\big\{N' + \mathrm{ord}_p(F(0)) + \mathrm{ord}_p(C^\sigma(\Gamma^p)^{-1} \bmod \Gamma^{N_\Gamma}), \\ &N + \mathrm{ord}_p(C(\Gamma) \bmod \Gamma^{N_\Gamma}) + \mathrm{ord}_p(C^\sigma(\Gamma^p)^{-1} \bmod \Gamma^{N_\Gamma}), \\ &N'' + \mathrm{ord}_p(C(\Gamma) \bmod \Gamma^{N_\Gamma}) + \mathrm{ord}_p(F(0))\big\} \\ &\geqslant N - (3B_{m,p}+1)\lfloor \log_p(N_\Gamma)\rfloor + B_{m,p} + \min\left\{\mathrm{ord}_p(F(0)), 0\right\}. \end{aligned} \tag{29}$$

For example, when $F(0)$ has integral entries and $p \geqslant m$ we have that the *loss* of accuracy when computing $F(\Gamma) \bmod \Gamma^{N_\Gamma}$ is bounded by $(3m-2)\lfloor \log_p(N_\Gamma)\rfloor - (m-1)$.

5.1.2. Method 2

The approach in this section is based upon that taken by Tsuzuki [**41**]. We do not give an analysis of the propagation of errors for this method, although this is an interesting problem.

Commutativity of diagram (3) implies that

$$\frac{dF}{d\Gamma} + B(\Gamma)F(\Gamma) = p\Gamma^{p-1}F(\Gamma)B^\sigma(\Gamma^p).$$

For $\ell \geqslant 1$, the coefficients F_ℓ in the local expansion $F(\Gamma) = \sum_{\ell=0}^{\infty} F_\ell \Gamma^\ell$ can be found recursively by rewriting this equation in the form

$$r(\Gamma)r^\sigma(\Gamma^p)\frac{dF}{d\Gamma} + r^\sigma(\Gamma^p)b(\Gamma)F(\Gamma) = p\Gamma^{p-1}r(\Gamma)F(\Gamma)b^\sigma(\Gamma^p)$$

and equating the coefficient of $\Gamma^{\ell-1}$ on both sides. This more direct method eliminates the multiplication of power series needed to compute the right hand side in (24) and is also more space efficient.

5.2. *Global deformation: analytic continuation*

The entries in the Frobenius matrix $F(\Gamma)$ are p-adic holomorphic functions on the p-adic projective line with open unit disks around the poles of r removed, i.e., uniform limits of rational functions on this closed domain D_1, say. (Recall that the entries lie in $A^\dagger$; see Section 3.4 for an explicit description of this ring.) Therefore, they can be uniformly approximated on this domain D_1 modulo any power of p by a matrix of rational functions whose denominators are powers of $r(\Gamma)$. Using the method in Section 5.1, one can compute the local expansions of these holomorphic functions to any required p-adic and Γ-adic accuracy. We now

sketch how to "analytically continue" these local expansions, i.e., how given a power of p one can compute the rational functions which approximate the entries in the Frobenius matrix to that power.

The essential point is that the theory guarantees that the holomorphic functions in the Frobenius matrix $F(\Gamma)$ are "overconvergent". This implies that they converge on the p-adic projective line with open disks of some unknown radius $s < 1$ removed around the poles of r. Let us notate this unknown larger domain by D_s. Assuming one has an upper bound on s, and also an upper bound on the maximum value t taken by the p-adic norm of the Frobenius matrix on the closed set D_s, one can compute an upper bound on the total degree of the rational functions needed to approximate $F(\Gamma)$ on this domain modulo any given power of p. This upper bound allows one to determine how many terms in the local expansion of $F(\Gamma)$ are required to compute the rational functions. The knowledge of bounds s and t amounts to having *effective lower bounds* on the p-adic decay of the entries in the matrix $F(\Gamma)$ (Definition 3.3). We shall *assume* that these effective lower bounds are *known*; for the explicit example we consider in Section 7 we will explain exactly how to calculate them.

We refer the reader to [**30**, Section 8.1] for a detailed description of the relatively straightforward step of recovering the matrix of approximating rational functions from the local expansions given that these bounds are known.

6. *An algorithm for computing* $F : E_{2,rig}^{1,n} \to E_{2,rig}^{1,n}$

In this section we gather together the results from Sections 3, 4 and 5 and present the main algorithm of the paper.

6.1. *Definitions and assumptions*

In this section we retain the definitions from Section 3.5 and make the assumption on the connection matrix from Section 4.1. Specifically, we are given a pencil $X \to S$ of k-varieties with fibres of dimension n such that:

- The relative space $\mathcal{E}^\dagger := H^n_{rig}(X/S) \cong H^n_{dR}(\mathcal{X}_K/\mathcal{S}_K) \otimes_A A^\dagger$ is a free $A^\dagger$-module of rank m.
- The connection $\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes \Omega^1_{A^\dagger}$ is given by a matrix $b(\Gamma)/r(\Gamma)$ with simple poles; here $r(\Gamma) \in \mathcal{O}_K[\Gamma]$ is squarefree modulo p and $d := \deg(r)$.
- The space $E^{1,n}_{2,rig} = \operatorname{coker}(\nabla^\dagger : \mathcal{E}^\dagger \to \mathcal{E}^\dagger \otimes_{A^\dagger} \Omega^1_{A^\dagger})$ is as in Definition 3.6.
- The morphism $\hat{\mathcal{X}} \to \hat{\mathcal{S}}$ is proper and smooth, so that the base change theorem holds (Theorem 3.2).
- The morphism $\mathcal{X}_K \to \mathcal{S}_K$ arises by extension of scalars from one defined over an algebraic number field, so the local exponents are rational (Theorem 6.2).

6.2. *The comparison theorem in the geometric case*

THEOREM 6.1. *With definitions and assumptions as in Section 6.1, Condition (O.C.) is met.*

Proof. We have an overconvergent F-isocrystal $(\mathcal{E}^\dagger, \nabla^\dagger, F)$ on $A^\dagger$. Choose any point $\Gamma = \gamma \in \mathcal{O}_{\bar{K}}$ such that $\operatorname{ord}_p(r(\gamma)) = 0$. Dwork's trick of analytic continuation via Frobenius [**22**, Prop. 3.1.2] tells us that the basis of local solutions to the differential

system $\nabla = 0$ converge on the open unit disk around $\Gamma = \gamma$. The same is true for the second differential system $\check{\nabla} = 0$; consider the "dual F-isocrystal $(\mathcal{E}^\dagger, \check{\nabla}^\dagger, F^{-1})$" and apply Dwork's trick once again. So by Theorem 4.5 condition (O.C.) is met. □

THEOREM 6.2. *With definitions and assumptions as in Section 6.1, condition (Rat.) is met.*

Proof. Since $\mathcal{X}_K \to \mathcal{S}_K$ can be defined over the complex numbers, it follows from the local monodromy theorem [**21**, Thm (14.3)]. □

We note that if $\lambda \in \mathbb{Q}$ is a local exponent at some singular point, then the root of unity $\exp(-2\pi i\lambda)$ is an eigenvalue of the local monodromy operator acting on the cohomology of a nearby smooth fibre when the family is viewed over the complex numbers, see [**21**, Section 12]. So our conditions on local exponents relate to conditions on the eigenvalues of local monodromy.

Theorems 4.7, 6.1 and 6.2 together yield:

THEOREM 6.3. *With definitions and assumptions as in Section 6.1., the comparison theorem $E_{2,dR}^{1,n} \cong E_{2,rig}^{1,n}$ holds.*

We note that our bounds on the growth of forms (Theorem 4.8) only holds when the assumption (Rat.) is replaced by the stronger assumption (Prep.Rat.). This will not always hold in the geometric case; however, we will give examples in which it does hold (Section 9).

6.3. *Numerical approximations*

In this section we formalise the notion of a "numerical approximation" to a p-adic number.

We assume that elements in $\mathcal{O}_K$ are represented as p-adic expansions with coefficients in some fixed set of representatives for the quotient $\mathcal{O}_K/(p)$. Thus for any positive integer N, elements in the quotient $\mathcal{O}_K/(p^N)$ can be represented in a unique manner via truncated p-adic expansions. Elements in $\mathcal{O}_K[\Gamma, 1/r(\Gamma)]/(p^N)$ have an obvious representation via these truncated p-adic expansions.

DEFINITION 6.4. *Let N be a positive integer, and $a \in A^\dagger = K[\Gamma, 1/r(\Gamma)]^\dagger$. Define $N' := N - \min(0, \mathrm{ord}_p(a))$. A p^N-approximation to a is a triple $(N, \mathrm{ord}_p(a), a_0)$ where $a_0 \in \mathcal{O}_K[\Gamma, 1/r(\Gamma)]/(p^{N'})$ and $a_0 - p^{-\min(\mathrm{ord}_p(a),0)}a = 0$ in $\mathcal{O}_K[\Gamma, 1/r(\Gamma)]/(p^{N'})$.*

Thus taking $\hat{a}_0$ to be any preimage of a_0 in $\mathcal{O}_K[\Gamma, 1/r(\Gamma)] \otimes K$, we have that a p^N-approximation to a defines an element $a_1 := p^{\min(\mathrm{ord}_p(a),0)}\hat{a}_0$ such that $\mathrm{ord}_p(a - a_1) \geqslant N$. Conversely, given such an element a_1 one may canonically identify with it a p^N-approximation to a. Intuitively a p^N-approximation amounts to knowledge "modulo p^N".

6.4. *Input/Output specification for the algorithm*

We retain the definitions and assumptions from Section 6.1, and now further assume that (Prep.Rat.) holds. So by Theorem 6.3 our comparison theorem $E_{2,dR}^{1,n} \cong E_{2,rig}^{1,n}$ holds, and moreover, we have effective bounds on the growth of forms during reduction (Theorem 4.8).

Let N_I be a positive integer. We shall assume that we are given as *input* the following:

- Input 1: The matrix $b(\Gamma)/r(\Gamma)$ for the Gauss-Manin connection ∇.
- Input 2: A p^{N_I}-approximation to $F(\gamma)$ for one Teichmüller specialisation $\Gamma = \gamma$ of the matrix $F(\Gamma)$ for the action $F : H^n_{rig}(X/S) \to H^n_{rig}(X/S)$, i.e., an approximation to the pth power Frobenius action on $H^n_{rig}(X_{\bar{\gamma}})$ for some fibre $X_{\bar{\gamma}}$ of the family $X \to S$.

We also assume we are given (see Definition 3.3):

- Input 3: Effective p-adic bounds on the entries in the matrix $F(\Gamma)$.

The algorithm gives as *output*:

- Output: A p^{N_O}-approximation to a matrix for $F : E^{1,n}_{2,rig} \to E^{1,n}_{2,rig}$,

for some effectively computable $N_O < N_I$. The *loss of accuracy* is measured by the difference $N_I - N_O$. We note in Section 6.6.1 that there exist effectively computable constants $\alpha'', \beta'' \geqslant 0$ such that one may take $N_O := N_I - (\alpha'' \log_p(N_I) - \beta'')$, i.e., we have a "logarithmic" loss of accuracy.

When X is affine and $X \to S$ is a smooth liftable family, we have from equation (4) that $H^{n+1}_{rig}(X) \cong E^{1,n}_{2,rig}$ and so the matrix given as output yields an approximation to pth power Frobenius action on $H^{n+1}_{rig}(X)$.

6.5. *The algorithm*

The algorithm comprises two steps

- Step 1: From Inputs 1,2 and 3 use the "deformation algorithm" to compute a p^N-approximation to the matrix for $F : H^n_{rig}(X/S) \to H^n_{rig}(X/S)$.
- Step 2: From the output of Step 1 and Input 1, use the algorithm from Section 4 to compute a p^{N_O}-approximation matrix for $F : E^{1,n}_{2,rig} \to E^{1,n}_{2,rig}$.

The intermediate precision N, where $N_O < N < N_I$, can be computed from the input data, see Section 6.6.1.

6.5.1. Step 1

This step was described in detail in Sections 5.1 and 5.2. (One may need to make a change of basis so that $\gamma = 0$.) Estimates from Input 3 determine the Γ-adic accuracy required in the local deformation to compute an p^N-approximation to the global matrix $F(\Gamma)$ itself, whose entries lie in $A^\dagger$.

6.5.2. Step 2

Let $\mathcal{B}$ be a set of forms in $A^m \otimes d\Gamma$ whose image in $E^{1,n}_{2,rig}$ is a K-basis for this space. This may be determined by computing the set of exponents $E(\nabla, \mathcal{B})$, then performing the linear algebra computation described in the proof of Theorem 4.3. Theorem 4.7 assures us that this set gives a basis for $E^{1,n}_{2,rig}$. We can assume that $\mathrm{ord}_p(e) = 0$ for all $e \in \mathcal{B}$.

For each $e \in \mathcal{B}$, one computes a p^N-approximation of the image $F(\Gamma)p\Gamma^{p-1}e^\sigma(\Gamma^p)d\Gamma$. Then one performs a radix conversion from Γ-adic to $r(\Gamma)$-adic expansions, so that the input is in the appropriate form for the reduction algorithm, i.e., write all rational functions in the form $\sum_{i\in\mathbb{Z}} a_i(\Gamma)r(\Gamma)^i$ where $a_i \in K[\Gamma]$ with $\deg(a_i) < \deg(r)$. (It is actually much better in practice to compute an r-adic expansion of the matrix $F(\Gamma)/r^\sigma(\Gamma^p)$, and recover the r-adic expansions for each image form via a r-adic multiplication routine. It turns out that these radix conversions

are very time consuming, so one wishes to minimize the number performed.) Then use the reduction algorithm from the proof of Theorem 4.2, plus the final linear algebra step from the proof of Theorem 4.3, to write this as a p^{N_O}-approximation to a K-linear combination of elements in $\mathcal{B}$ plus an p^{N_O}-approximation to an element in $\nabla^\dagger(\mathcal{E}^\dagger)$.

6.6. *Analysis*

6.6.1. Loss of accuracy

Theorem 4.8 and the analysis following Theorem 5.1 together show that the matrix computed in the algorithm is a p^{N_O}-approximation to a matrix for the action of $F : E^{1,n}_{2,rig} \to E^{1,n}_{2,rig}$ for N_O such that

$$N_O = N_I - (\alpha'' \log_p(N_I) + \beta'').$$

Here $\alpha'', \beta'' \geqslant 0$ are constants which may be computed from m, p, the local exponents of the connection, and the effective p-adic bounds on $F(\Gamma)$ (Input 3). We shall not present explicit formulae for α'' and β'' in the general case as they are rather complicated. We note that the discussion following Theorem 5.1 allows one to compute the intermediate precision p^N which is attained after Step 1.

6.6.2. Time and space complexity

The time and space complexity may be calculated given the effective p-adic bounds on the matrix $F(\Gamma)$, and also a bound on the height of the local exponents, c.f. Section 7.5.2. We do not present an explicit expression for the general case since it is rather complicated. Let us just make a few observations on Step 1: The calculation of the local solution matrix in Step 1 (Section 5.1) is fast, both in theory and practice, since it just requires the iteration of a short linear recurrence; however, Method 1 is rather space consuming in comparison to Method 2. The analytic continuation step requires only a single multiplication by a power of $r(\Gamma)$ (computed modulo a power of Γ). The radix conversion, though in theory "quasi-linear time" [**17**, Alg. 9.14], is in practice rather time consuming.

7. *The Frobenius matrix of an affine surface*

In this section we apply the algorithm in Section 6 to compute to any required numerical precision a matrix for the pth power Frobenius map acting on the middle dimensional rigid cohomology of a certain affine surface. Specifically, we consider an open subset X of the affine surface defined by an equation of the form $Z^2 = \bar{Q}(X, \Gamma)$, subject to certain smoothness assumptions. The algorithm from Section 6.5 allows the efficient computation of an approximation to the Frobenius map $F : H^2_{rig}(X) \to H^2_{rig}(X)$, provided one can obtain the auxiliary inputs 1, 2 and 3 (Section 6.4). After defining the surface in Section 7.1, we describe how the necessary auxiliary inputs may be calculated (Section 7.3). Having specified some local monodromy restrictions to ensure applicability of the algorithm in Section 6 (see Section 7.4), we then give a precise complexity analysis (Theorem 7.6).

In Section 8 we shall apply the results of the present section to compute the full zeta function of a compactification $\bar{X}$ of the open surface X. We report on our Magma implementation of this final algorithm in Section 9.

We retain the notation in Section 3. In particular, recall that $k = \mathbb{F}_q$ is the finite field with q elements, and K the unramified extension of $\mathbb{Q}_p$ of degree $[k : \mathbb{F}_p]$. We assume now that the characteristic p is odd. Recall that the ring of integers of K is denoted $\mathcal{O}_K$. Let us further assume we are given $L \supseteq \mathbb{Q}$ an algebraic number field with ring of integers $\mathcal{O}_L$ in which the prime p is inert, and an explicit embedding $\mathcal{O}_L \subset \mathcal{O}_K$. Note that $\mathcal{O}_L/(p) \cong \mathcal{O}_K/(p) \cong \mathbb{F}_q$.

7.1. *Definition of the pencil*

Let $Q(X, \Gamma) \in \mathcal{O}_L[X, \Gamma]$ and denote by $\bar{Q}(X, \Gamma) \in k[X, \Gamma]$ its reduction modulo p. We shall assume that both Q and $\bar{Q}$ are monic in X of degree $2g + 1$ where $\gcd(p, 2g + 1) = 1$. Let $\tilde{r}(\Gamma) := \mathrm{Res}(X, Q, \frac{\partial Q}{\partial X}) \in \mathcal{O}_L[\Gamma]$ be the Sylvester resultant w.r.t. X of Q and $\frac{\partial Q}{\partial X}$, see [**9**, Pages 150-151] or Section 7.3.1. (The notation $r(\Gamma)$ is reserved for the monic denominator of the connection matrix, which is a factor of $\tilde{r}$, see Section 7.3.1.) Assume that $\tilde{r}(\Gamma)$ has leading coefficient a unit modulo p, and $\tilde{r}(0) \neq 0 \bmod p$; in particular, it does not vanish identically modulo p. Define the $\mathcal{O}_K$-schemes

$$\mathcal{X} := \mathrm{Spec}(B) \text{ where } B := \mathcal{O}_K[X, \Gamma, Z, \tilde{r}(\Gamma)^{-1}]/(Z^2 - Q(X, \Gamma))$$

and

$$\mathcal{S} := \mathrm{Spec}(A) \text{ where } A := \mathcal{O}_K[\Gamma, 1/\tilde{r}(\Gamma)].$$

Let $\bar{B} := B \otimes_{\mathcal{O}_K} k$ and $\bar{A} := A \otimes_{\mathcal{O}_K} k$ be the reduction of the coordinate rings modulo p. Define the k-schemes

$$X := \mathrm{Spec}(\bar{B}) \text{ and } S := \mathrm{Spec}(\bar{A}).$$

We have the obvious commutative diagrams

$$\begin{array}{ccc} A & \hookrightarrow & B \\ \downarrow & & \downarrow \\ \bar{A} & \hookrightarrow & \bar{B} \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathcal{X} & \to & \mathcal{S} \\ \uparrow & & \uparrow \\ X & \to & S \end{array}$$

where the vertical maps in the second diagram are embeddings of special fibres. Recall that the generic fibres are denoted $\mathcal{X}_K$ and $\mathcal{S}_K$, respectively. The horizontal maps in the second diagram are smooth morphisms of smooth schemes, and the fibres are (affine) hyperelliptic curves.

The relative cohomology spaces which concern us are:

$$\mathcal{E} := H^1_{dR}(\mathcal{X}_K/\mathcal{S}_K) = \left\langle \frac{X^i dX}{\sqrt{Q}} \,|\, 0 \leqslant i < 2g \right\rangle_A$$

and

$$\mathcal{E}^\dagger := H^1_{rig}(X/S) = \left\langle \frac{X^i dX}{\sqrt{Q}} \,|\, 0 \leqslant i < 2g \right\rangle_{A^\dagger}.$$

Here $\sqrt{Q}$ denotes the image of Z in B (precisely, in $B^\dagger$ for the second space). We refer the reader to Section 3.4 for a description of the ring $A^\dagger$. That $H^1_{dR}(\mathcal{X}/\mathcal{S})$ and $H^1_{rig}(X/S)$ are spanned by (the cohomology classes of) these forms is shown in [**31**, Secs. 4.2, 5.4]. That they form a basis follows by a specialisation argument, and the fact that the dimension of the first de Rham (rigid, respectively) cohomology space of any fibre in the family $\mathcal{X} \to \mathcal{S}$ ($X \to S$, respectively) is $2g$. Alternatively,

see [**19**]. Note that we do not need to appeal to the finiteness and comparison theorems in Section 3.2, since we can establish the necessary results directly. We also note that the base change property (Theorem 3.2) follows easily from the fact that $H^1_{rig}(X/S)$ is free of rank $2g$.

7.2. *The spectral sequence*

The next proposition shows that the algorithm in Section 6 in the present case computes a numerical approximation to a matrix for $F : H^2_{rig}(X) \to H^2_{rig}(X)$.

PROPOSITION 7.1. *With the morphism $X \to S$ as defined in Section 7.1, we have*

$$H^2_{rig}(X) \cong E^{1,1}_{2,rig}.$$

Proof. Follows from equation (4) since X is affine and $\Omega^1_{R/\mathcal{V}}$ is free of rank one; here $R := \Gamma(\mathcal{S}, \mathcal{O}_{\mathcal{S}})$ and $\mathcal{V} := \mathcal{O}_K$. □

7.3. *Auxiliary data: Inputs 1,2 and 3*

We now explain how to compute the auxiliary information needed as input to our main algorithm (Section 6) in the case of the surfaces presently under consideration.

7.3.1. Input 1: the matrix for ∇

A matrix for the action of the connection $\nabla : \mathcal{E} \to \mathcal{E} \otimes \Omega^1_A$ can be computed in the following manner: For each basis element $X^i dX/\sqrt{Q}\,(0 \leqslant i < 2g)$, compute its derivative w.r.t. Γ, and write the answer as a linear combination over A of the basis elements plus $\frac{d}{dX}(g)$ for some element $g \in B$. The latter task is performed by using a "generic" version of Kedlaya's reduction algorithm, viewing the surface as a hyperelliptic curve over a function field, c.f. [**31**, Section 4.2]. The same A-linear combination of (the cohomology classes of) the basis elements tensored by $d\Gamma$ gives the image of (the cohomology class of) the original basis element under the connection. We now give an explicit formula for the matrix for the connection (based upon Magma code written by the author).

For an element $a \in L(\Gamma)[X]$ and $i \in \mathbb{Z}$, denote by $\mathrm{Coeff}(a, i)$ the coefficient of X^i in a. Let $\delta := 2g + 1$. Let M be the Sylvester matrix w.r.t. X of Q and $\frac{\partial Q}{\partial X}$. Explicitly, $M \in M_{2\delta-1}(L[\Gamma])$ and for $1 \leqslant j \leqslant 2\delta - 1$

$$M_{ij} := \begin{cases} \mathrm{Coeff}(X^{\delta-1-i}Q, 2\delta - 1 - j) & \text{for } 1 \leqslant i \leqslant \delta - 1 \\ \mathrm{Coeff}(X^{2\delta-1-i}\frac{\partial Q}{\partial X}, 2\delta - 1 - j) & \text{for } \delta \leqslant i \leqslant 2\delta - 1. \end{cases}$$

We have assumed the determinant of this matrix $\tilde{r}(\Gamma)$ (Sylvester resultant) is non-zero modulo p. Define E to be the $(\delta - 1) \times (2\delta - 1)$ matrix over $L[\Gamma]$ with

$$E_{ij} := -\frac{1}{2}\mathrm{Coeff}\left(X^{i-1}\frac{\partial Q}{\partial \Gamma}, (2\delta - 1) - j\right).$$

Let $F := EM^{-1}$, a $(\delta - 1) \times (2\delta - 1)$ matrix over $L[\Gamma, 1/\tilde{r}(\Gamma)]$. Let the vectors $a, b, c \in L[\Gamma, 1/\tilde{r}(\Gamma)][X]^{\delta-1}$ be defined as follows: For $1 \leqslant i \leqslant \delta - 1$

$$\begin{array}{rcl} a_i & := & \sum_{j=1}^{\delta-1} F_{i,\delta-j} X^{j-1} \\ b_i & := & \sum_{j=1}^{\delta-1} F_{i,2\delta-j} X^{j-1} \\ m_i & := & a_i + 2\frac{\partial b_i}{\partial X}. \end{array}$$

Then the connection matrix $B(\Gamma) \in M_{\delta-1}(L[\Gamma, 1/\tilde{r}(\Gamma)])$ is defined by $B_{i,j} := \mathrm{Coeff}(m_j, i-1)$. One can uniquely write $B(\Gamma) = b(\Gamma)/r(\Gamma)$ where $r|\tilde{r}$, and r is monic and coprime to some entry in the matrix $b \in M_{\delta-1}(L[\Gamma])$.

We shall impose some restrictions on the connection matrix $B(\Gamma)$ in Section 7.4.

7.3.2. Input 2: The Frobenius matrix of a fibre
We take the fibre at $\gamma = 0$, noting $\tilde{r}(0) \neq 0 \bmod p$. The Frobenius matrix of the fibre $Z^2 = \bar{Q}(X, 0)$ can be computed using Kedlaya's original algorithm [**25**]; the implementation by Michael Harrison is available with the documentation accompanying the Magma program.

7.3.3. Input 3: Effective p-adic bounds for $F : \mathcal{E}^\dagger \to \mathcal{E}^\dagger$
The Frobenius matrix $F(\Gamma)$ can in principle be calculated by applying Kedlaya's algorithm to the "generic" hyperelliptic curve in the family, which is defined over the function field $\mathbb{F}_q(\Gamma)$. From the point of view of complexity theory this is not a good idea; it is faster to use the indirect method of the "deformation algorithm". However, this direct method is a good way to calculate effective p-adic bounds for the matrix $F(\Gamma)$.

Specifically, fix i, j with $1 \leqslant i, j \leqslant 2g$. Let $f(\Gamma)$ be the (i, j)th entry in the matrix $F(\Gamma)$. Then $f(\Gamma)$ is the coefficient of $X^i dX/\sqrt{Q}$ in the expression one obtains by reducing the form

$$\sigma\left(\frac{X^j dX}{\sqrt{Q}}\right) = \frac{pX^{p(j+1)-1}}{Q^{p/2}}\left(1 - \frac{Q^p - Q^\sigma(X^p, \Gamma^p)}{Q^p}\right)^{-1/2} dX$$

using the "generic" version of Kedlaya's algorithm. Here σ is the map sending $\Gamma \mapsto \Gamma^p$, $X \mapsto X^p$, and acting like the pth power Frobenius automorphism on K. We can write $Q^p - Q^\sigma(X^p, \Gamma^p) = pR(X, \Gamma)$ for some unique $R \in \mathcal{O}_K[X, \Gamma]$ with

$$\deg_X(R) < p \deg_X(Q),\ \deg_\Gamma(R) \leqslant p \deg_\Gamma(Q)$$

where the second inequality is strict if Q is monic in Γ. Then

$$\sigma(X^j dX/\sqrt{Q}) = \sum_{\ell=0}^{\infty} \binom{-1/2}{\ell} p^{\ell+1} \frac{X^{p(j+1)-1} R^\ell}{Q^{p(\ell+(1/2))}} dX.$$

The ℓth term in this series can be reduced modulo exact forms using $p\ell + \lfloor \frac{p}{2} \rfloor$ applications of Kedlaya's "pole reduction formula", see [**31**, Section 4.2]. Each application requires one division by the resultant $\tilde{r}(\Gamma)$. By an easy specialisation argument, [**25**, Lemma 2] implies that reduction of the ℓth term requires a cumulative division by at most $p^{\lfloor \log_p(p(2\ell+1)) \rfloor}$.

Write

$$f(\Gamma) = \sum_{k=-\infty}^{\infty} f_k(\Gamma) \tilde{r}(\Gamma)^k$$

where $f_k \in K[\Gamma]$ with $\deg(f_k) < \deg(\tilde{r})$. The argument in the preceding paragraph implies the following.

PROPOSITION 7.2. *For $k < 0$ we have the lower bound*

$$\mathrm{ord}_p(f_k(\Gamma)) \geqslant (\ell + 1) - \lfloor \log_p(p(2\ell + 1)) \rfloor$$

where ℓ is the smallest integer such that $p\ell + \lfloor \frac{p}{2} \rfloor \geqslant |k|$. (Explicitly, $\ell := \lfloor \frac{2|k|-p+1}{2p} \rfloor$ and so $\mathrm{ord}_p(f_k(\Gamma)) \geqslant \lfloor |k|/p \rfloor - \lfloor \log_p(2|k|+1) \rfloor$.)

A lower bound for $k \geqslant 0$ requires a more detailed analysis: Let $\mathrm{Adj}(M)$ be the adjoint of the Sylvester matrix. Each application of Kedlaya's pole reduction formula increases the degree in Γ (degree of numerator minus degree in denominator) by $\deg_\Gamma(\mathrm{Adj}(M)) - \deg_\Gamma(\tilde{r})$. The degree in Γ of the numerator $X^{p(j+1)-1}R^\ell$ of the ℓth term in the series is $\ell \deg_\Gamma(R) < \ell(p \deg_\Gamma(Q) - 1)$. Thus after $p\ell + \frac{p-1}{2}$ applications of Kedlaya's pole reduction formula, the degree in Γ of the reduction of the ℓth term in the series is at most

$$\kappa(\ell) := \left(p\ell + \frac{p-1}{2} \right) (\deg_\Gamma(\mathrm{Adj}(M)) - \deg_\Gamma(\tilde{r})) + \ell \deg_\Gamma(R).$$

We note that the modest use of Kedlaya's formula for reducing the "pole at infinity" required in the calculation of $F(\Gamma)$ does not increase the degree in Γ (or introduce powers of $\tilde{r}(\Gamma)$ on the denominator). Thus we deduce:

PROPOSITION 7.3. *For $k \geqslant 0$ we have the lower bound*

$$\mathrm{ord}_p(f_k(\Gamma)) \geqslant (\ell+1) - \lfloor \log_p(p(2\ell+1)) \rfloor$$

where ℓ is the smallest integer such that $\frac{\kappa(\ell)}{\deg_\Gamma(\tilde{r})} \geqslant k$. (If no such ℓ exists then the term $f_k(\Gamma)$ is zero.)

Explicitly, define

$$\delta := \frac{\deg_\Gamma(\mathrm{Adj}(M))}{\deg_\Gamma(\tilde{r})},\ \delta' := \frac{\deg_\Gamma(R)}{p \deg_\Gamma(\tilde{r})} \leqslant \frac{\deg_\Gamma(Q)}{\deg_\Gamma(\tilde{r})}. \tag{30}$$

Then assuming $\delta + \delta' \geqslant 1$ one takes ℓ the floor of $\frac{2k-(p-1)(\delta-1)}{2p(\delta+\delta'-1)}$.

We now state a conjecture to which we shall refer later.

CONJECTURE 7.4. *The Frobenius matrix $F(\Gamma)$ has a pole of finite order at infinity, rather than an essential singularity.*

Conjecture 7.4 thus claims that $f_k(\Gamma) = 0$ for sufficiently large positive k.

7.4. *Local monodromy assumptions*

In this section we state some further restrictions made to ensure that the conditions required for the application of the main algorithm in Section 6 are met. Specifically, we need that the connection matrix $B(\Gamma)$ from Section 7.3.1 is of the form required in Section 6.1, and that condition (Prep.Rat.) is met. To simplify the complexity analysis and to keep in line with our actual implementation in Section 9, we shall in fact make stronger assumptions, as follows.

Recall that $Q \in \mathcal{O}_L[X, \Gamma]$ with $2g+1 := \deg_X(Q)$, that $\tilde{r}(\Gamma)$ is the Sylvester resultant of Q and $\frac{\partial Q}{\partial X}$ w.r.t. X, and r the monic factor of $\tilde{r}$ which is the denominator of the connection matrix $B(\Gamma) = b(\Gamma)/r(\Gamma)$ when in lowest terms. Define $h := \deg_\Gamma(Q)$.

We *assume* that $r(\Gamma) \bmod p$ is squarefree, and that the Laurent expansion of $B(\Gamma)$ has only negative terms. We say then that $B(\Gamma)$ has *only simple poles modulo p*. This ensures that the algorithm in the proof of Theorem 4.2 works. Let us *assume*

that the local exponents around each singular point are prepared, so condition (Prep.Rat.) is met and we may apply the precision loss bounds in Theorem 4.8.

To obtain a nice basis for $E_{2,rig}^{1,1}$, let us further *assume* that the local monodromy around the finite poles is nilpotent, so that the local exponents around the finite poles are all zero, and that zero does *not* occur as a local exponent around the pole at infinity. In this case we may take as our basis for $E_{2,rig}^{1,1}$ the elements $\{b_{ik}\}$ where $0 \leqslant i \leqslant d-2$ ($d := \deg(r)$) and $1 \leqslant k \leqslant 2g$, and the element $b_{ik} \in H^1_{rig}(X/S)$ is the column vector with zeros in positions $j \neq k$, and in position $j = k$ the 1-form $\Gamma^i d\Gamma / r(\Gamma)$. Note that the dimension of this space is $2g(d-1)$.

For the complexity analysis, we shall need bounds on the height of the local exponents. Let us *assume* that a common denominator for the local exponents around infinity is $2(2g+1)$, and when written w.r.t. this denominator the numerator does not exceed $h(2g-1)$ in absolute value. Under the assumption that $B(\Gamma)$ has only simple poles modulo p, we believe that one may prove that the bound on the denominator should always hold by a topological argument. Likewise, the author expects that the bound on the numerator should also hold, although offers no proof of this.

NOTE 7.5 We point out that "generically" in any nice family of polynomials both $r(\Gamma)$ and $\tilde{r}(\Gamma)$ are squarefree and have equal degree. In this case, one observes experimentally, and expects to be able to prove, that all residue matrices around finite poles are nilpotent. However, the assumption that the degree in Γ of the connection matrix $B(\Gamma)$ is less than zero does *not* hold generically. For any family of polynomials Q, e.g. with fixed Newton polytope, one can calculate restrictions on the coefficients which must be met. The author has no idea of the geometric significance of this assumption. When the assumption does hold, the local exponents at infinity are observed to exhaust the set $\left\{ \frac{\pm jh}{2(2g+1)} | 1 \leqslant j \leqslant 2g-1,\ j \text{ odd} \right\}$.

7.5. *Analysis*

We shall use soft-Oh notation, to hide logarithmic factors in the time and space complexity [**17**, Def. 25.8].

Let N_O be a positive integer which depends upon the equation $Z^2 = \bar{Q}(X, \Gamma)$ in some manner — we shall specify precisely how later. Assume that

$$g^2 \left(1 + \frac{h}{p} + \log_p(gh)^2 \right) = \mathcal{O}(N_O), \tag{31}$$

i.e., the integer N_O grows at least as fast as the expression on the left hand side as g, h and p vary.

7.5.1. Numerical approximations

Assume that one wishes to compute a p^{N_O}-approximation to the pth power Frobenius matrix $F : H^2_{rig}(X) \to H^2_{rig}(X)$. Then Theorem 4.8, equation (22) in Note 4.11, inequality (29), Propositions 7.2 and 7.3, and the local monodromy assumptions in Section 7.4 show that it suffices to take the initial p-adic accuracy N_I such that $N_I - (\alpha'' \log_p(N_I) + \beta'') \geqslant N_O$ for some effective constants $\alpha'', \beta'' \geqslant 0$. For implementations one needs to compute the loss of accuracy precisely; however for our complexity estimates it is enough to observe $\alpha'' = \mathcal{O}(g^2 \log_p(g))$

and $\beta'' = \mathcal{O}(g^2(1 + (h/p) + \log_p(gh)^2))$. Here are more details. First, Propositions 7.2 and 7.3 combined with the observation $(\delta + \delta' - 1)\deg_\Gamma(\tilde{r}) = \mathcal{O}(gh)$ shows the following: the Γ-adic accuracy needed in the solution of the differential system in Step 1 is $\mathcal{O}(pghN_I)$. From inequality (29), the loss of accuracy in this step is $\mathcal{O}(g\log_p(pghN_I)) = \mathcal{O}(g(\log_p(N_I) + \log_p(gh)))$. Second, the maximum pole order encountered in Step 2 is $\mathcal{O}(pghN)$ where $N < N_I$ is the intermediate accuracy, so the loss of accuracy in Step 2 is $\mathcal{O}(\alpha\log_p(pghN) + \beta) = \mathcal{O}(\alpha\log_p(N_I) + \alpha\log_p(gh) + \beta)$ where α, β are the constants in Theorem 4.8. From equation (22) we have $\alpha = \mathcal{O}(g^2\log_p(g))$ and $\beta = \mathcal{O}(g^2(1 + (h/p) + \log_p(gh))$. Our claim on the loss of accuracy now follows. Moreover, from equation (31) we see that the initial p-adic accuracy N_I satisfies $N_I = \tilde{\mathcal{O}}(N_O)$.

7.5.2. Time and space complexity

We now give a precise complexity analysis of the time and space required to compute a numerical approximation to the pth power Frobenius matrix $F : H^2_{rig}(X) \to H^2_{rig}(X)$ using the algorithm in Section 6.

THEOREM 7.6. *Let the affine surface X be defined as in Section 7.1, and assume the local monodromy conditions specified in Section 7.4 hold. We recall that X is an open subset of the smooth surface defined by the equation $Z^2 = \bar{Q}(X, \Gamma)$ over the field $\mathbb{F}_q$ of characteristic p, and $2g + 1 := \deg_X(Q)$, $h := \deg_\Gamma(Q)$. Let the positive integer N_O satisfy the growth condition (31). Then one may compute a p^{N_O}-approximation to the pth power Frobenius matrix $F : H^2_{rig}(X) \to H^2_{rig}(X)$ via the algorithm in Section 6 in $\tilde{\mathcal{O}}(N_O^2 g^5 h^2 p\log(q))$ bit operations, using $\tilde{\mathcal{O}}(N_O^2 g^3 hp\log(q))$ bits of space.*

Proof. Since N_O satisfies growth condition (31), from Section 7.5.1 we see that the initial p-adic accuracy N_I satisfies $N_I = \tilde{\mathcal{O}}(N_O)$. For the purposes of the complexity analysis, we shall forget about the intermediate accuracy N, with $N_O < N < N_I$ mentioned in Section 6.5, and just assume we work with p^{N_I}-approximations throughout the algorithm.

Step 1: Using the estimates from Propositions 7.2 and 7.3, we see that the Γ-adic accuracy required in Step 1 is $\mathcal{O}(N_I p\mu)$, where

$$\mu := \max\{\deg_\Gamma(Q), \deg_\Gamma(r), \deg_\Gamma(\mathrm{Adj}(M))\} = \mathcal{O}(hg).$$

We consider the time/space required to compute an approximation to $C(\Gamma)$ in Section 5.1.1: The coefficients of Γ are $g \times g$ matrices, whose entries are p^{N_O}-approximations of elements of the p-adic field K. Moreover, the growth bounds given in the analysis following Theorem 5.1 show that each coefficient requires $\tilde{\mathcal{O}}(\log(q)N_I)$ bits of space. This gives a space requirement of $\tilde{\mathcal{O}}(N_I^2 g^3 hp\log(q))$ bits. For the time, we observe that recurrence (27) has length bound by $\max\{\deg(b) + 1, \deg(r)\} = \mathcal{O}(gh)$, and involves multiplication of $g \times g$ matrices. Thus the time to compute an approximation to $C(\Gamma)$ is $\tilde{\mathcal{O}}(N_I^2 g^{2+\omega} h^2 p\log(q))$ bit operations. One may further compute the approximation to the local Frobenius matrix $F(\Gamma)$ in this time/space, using (24). Using the fast radix conversion algorithm in [**17**, Alg 9.14], these time and space estimates are enough for the the analytic continuation and radix conversion steps required to make the input suitable for Step 2.

Step 2: The matrix F has size $(d-2)2g = \mathcal{O}(g^2 h)$, where $d = \deg(r(\Gamma))$. Thus $\mathcal{O}(g^2 h)$ applications of the reduction algorithm from Section 4 are required. It is

time saving in terms of the parameter g to precompute the inverses "$(-\ell r' I_m + b)^{-1}, (\ell' I_m + b_{d-1})^{-1}$" for ℓ and ℓ' in the necessary ranges, as these do not depend on the element being reduced. The number of the former inverses is $\mathcal{O}(N_I p)$ and each inverse takes $\tilde{\mathcal{O}}(N_I \log(q) \times gh \times g^3)$ bit operations to compute; the factor gh arising since the inverse is computed modulo the polynomial $r(\Gamma)$ which has degree $\mathcal{O}(gh)$. There are $\mathcal{O}(N_I pgh)$ of the latter inverses to compute, but each only requires $\mathcal{O}(N_I \log(q) g^3)$ bit operations. Thus precomputation of the matrix inverses takes $\tilde{\mathcal{O}}(N_I^2 g^4 hp \log(q))$ bit operations and one needs $\mathcal{O}(N_I^2 g^3 hp \log(q))$ bits to store them. The reduction of finite poles requires $\mathcal{O}(N_I p)$ steps, and each step taking $\tilde{\mathcal{O}}(N_I g^3 h \log(q))$ bit operations; reduction of the pole at infinity requires $\mathcal{O}(N_I pgh)$ steps, but each step only takes $\tilde{\mathcal{O}}(N_I g^2 \log(q))$ bit operations. The time for the reduction of forms is thus $\mathcal{O}(N_I^2 g^5 h^2 p \log(q))$ bit operations, and this step requires $\tilde{Oh}(N_I^2 g^3 p \log(q))$ bits of space. (The time without precomputation of matrix inverses would be $\tilde{\mathcal{O}}(N_I^2 g^6 h^2 p \log(q))$ bit operations.) This completes the proof. □

8. *The zeta function of a compact surface*

This section is a direct continuation of Section 7. In particular, throughout this section we retain the definitions and assumptions in the preamble to that section, as well as those stated in Sections 7.1 and 7.4.

8.1. *The zeta function of the open surface*

In this section we consider the zeta function $Z(X,T)$ of the smooth affine surface X over $\mathbb{F}_q$. The trace formula in rigid cohomology for smooth affine varieties shows

$$Z(X,T) = \frac{P_1(X,T)}{P_2(X,T)P_0(X,T)} \tag{32}$$

where $P_i(X,T) := \det(1 - Tq^2 F^{-\log_p(q)} | H^i_{rig}(X)) \in 1 + T\mathbb{Q}_p[T]$. Certainly $H^0_{rig}(X)$ is a one-dimensional $\mathbb{Q}_p$-vector space, and $P_0(X,T) = (1 - q^2 T)$. We note that cohomology in dimensions 3 and 4 vanishes since the variety is affine.

PROPOSITION 8.1. *Let the polynomial $P_1(S,T)$ be the numerator of the zeta function of the open subset S of the projective line; so $P_1(S,T)$ is a product of cyclotomic polynomials. Then $P_1(X,T) = P_1(S,qT) \in 1 + T\mathbb{Z}[T]$.*

Proof. It is enough to consider the terms $E^{0,1}_{2,rig}$ and $E^{1,0}_{2,rig}$ in the spectral sequence for $X \to S$, c.f. Section 3.6 and [**24**, Eqn (17)]. We have

$$E^{0,1}_{2,rig} := \ker(\nabla^\dagger) \cong \ker(\nabla).$$

The isomorphism follows from [**3**, Cor. 2.6]. We claim the latter space is zero-dimensional: Let $v \in \ker(\nabla)$. Recalling from Section 7.4 that the local exponents around finite poles are all zero, expanding v around the finite poles one deduces that $v \in K^{2g}$. Since the local exponents around the pole at infinity are non-zero, expanding v around this pole one deduces $v = 0$. We have

$$E^{1,0}_{2,rig} := \operatorname{coker}\left(\frac{d}{d\Gamma} : H^0_{rig}(X/S) \to H^0_{rig}(X/S) d\Gamma\right).$$

But $H^0_{rig}(X/S) \cong A^\dagger$, the weak completion of the coordinate ring of S. So

$$\det(1 - Tq^2 F^{-\log_p(q)} | E^{1,0}_{2,rig}) = \det(1 - (Tq) q F^{-\log_p(q)} | H^1_{rig}(S)) = P_1(S, qT).$$

□

PROPOSITION 8.2. *The polynomial $P_2(X,T)$ has integer coefficients.*

Proof. Integrality follows from Proposition 8.1 since the zeta function itself is a power series with integer coefficients. □

Kedlaya's p-adic analogue of Deligne's main theorem tells us that the complex absolute values of reciprocal zeros of $P_2(X,T)$ belong to the set $\{1, q^{1/2}, q\}$ [**28**]; we will deduce this is an elementary manner in Section 8.2.

8.2. *The zeta function of a compactification*

In this section we show that the full zeta function $Z(\bar{X},T)$ of a compactification $\bar{X}$ of X may be easily recovered given the first $\mathcal{O}(gh)$ coefficients of $P_2(X,T)$ to precision modulo p^N where $N = \mathcal{O}(gh\log(q))$.

To simplify the analysis, and keep in line with our actual implementation, we shall make some further restrictions on the polynomial $Q(X,\Gamma)$. Recall that Q is monic in X of degree $2g+1$, and has degree h in Γ. Let us further assume that it is monic in Γ with h odd, has constant term 1, and that $2g+1$, h and the prime p are mutually coprime. Moreover, assume that all other terms in $Z^2 - Q(X,\Gamma)$ have exponents lying within or on the boundary of the polytope Δ with vertices the origin and the points $(2g+1,0,0),(0,h,0)$ and $(0,0,2)$. Then the Newton polytope [**10**, Sec. 2.1] of $Z^2 - Q(X,\Gamma)$ (taken modulo any prime number p) is the simplex Δ. We assume that $Z^2 - \bar{Q}(X,\Gamma)$ is *non-degenerate* w.r.t. the polytope Δ c.f. [**10**, Sec. 3.6]: Specifically, the polynomials $\bar{Q}(X,0),\bar{Q}(0,\Gamma)$ are squarefree, and $\bar{Q}, \frac{\partial\bar{Q}}{\partial X}, \frac{\partial\bar{Q}}{\partial\Gamma}$ have no common solutions. Let $\bar{X}$ be the *toric compactification* of the affine variety X in the toric projective space $\mathbb{P}_\Delta$ c.f. [**10**, Sec. 3.2]. This is a smooth compact variety. Since the outer face of Δ is a triangle with no interior points, it follows that $\bar{X} = X_{\text{aff}} \sqcup \mathbb{P}^1$ where $X_{\text{aff}} := \text{Spec}(\mathbb{F}_q[X,\Gamma,Z]/(Z^2 - \bar{Q}(X,\Gamma)))$.

One does not need to be familiar with the exact details of the construction: the point is simply that we have compactified the zero set in affine space of the equation $Z^2 = \bar{Q}(X,\Gamma)$ by adding a single projective line.

DEFINITION 8.3. *Let $P(T) \in \mathbb{Z}[T]$, q be a prime power, and ω be a non-negative integer. We call $P(T)$* pure of weight ω *with respect to q if its reciprocal zeros have complex absolute value $q^{\omega/2}$. We shall just say P is a weight ω* Weil polynomial *when q is understood.*

PROPOSITION 8.4. *Let $\bar{X}$ be the smooth toric compactification of the set of affine solutions of the equation $Z^2 = \bar{Q}(X,Y)$, as defined immediately above. Then the zeta function of $\bar{X}$ has the form*

$$Z(\bar{X},T) = \frac{1}{(1-T)P_2(\bar{X},T)(1-q^2T)},$$

where $P_2(\bar{X},T) \in \mathbb{Z}[T]$ is a Weil polynomial w.r.t. q of weight 2, and

$$\deg(P_2(\bar{X},T)) = l^*(2\Delta) - 4l^*(\Delta) - 3 - \sum_{\Delta'}(l^*(\Delta') - 1). \tag{33}$$

Here the function l^* counts lattice points in the interior of a polytope, and the sum is over the two-dimensional faces Δ' of Δ.

Proof. The claim on the weight follows from Deligne's theorem [**11**]. It remains to prove that the ℓ-adic Betti numbers $h_i := \dim H_{et}(\bar{X} \times \bar{\mathbb{F}}_q, \mathbb{Q}_\ell)$ are as follows: $h_0 = h_4 = 1, h_1 = h_3 = 0$ and h_2 is as on the right hand side of (33). The comparison theorem for smooth liftable varieties shows that it is enough to prove that these are the Betti numbers of the compact toric variety defined over the complex numbers by the equation $Z^2 = Q(X, \Gamma)$ [**35**, Remark 21.10]. That these are the Betti numbers follows from the formulae for Hodge-Deligne numbers of complex toric surfaces in [**10**, Sec. 5.11(c)]. □

We note, but do not use, that the formula for $\deg(P_2(\bar{X}, T))$ is valid for arbitrary Newton polytopes Δ, assuming that $\mathbb{P}_\Delta$ is smooth and $Z^2 - Q(X, \Gamma)$ is non-degenerate w.r.t. Δ.

To simplify the statement and proof of the next theorem, and again to keep in line with our implementation, we make some further restrictions: Assume that the Sylvester resultant $\tilde{r}(\Gamma)$ is squarefree modulo p of degree $d = \deg_\Gamma(r)$, and that for each $\gamma \in \bar{\mathbb{F}}_q$ with $\tilde{r}(\gamma) = 0 \bmod p$, the "missing fibre at $\Gamma = \gamma$" in the pencil $X \to S$ has a unique double point.

PROPOSITION 8.5. *Definitions and assumptions as in Sections 7.1, 7.4, 8.1 and the present section. Let $\bar{X} = X \sqcup C$ where C is a union of curves. We recall that $\bar{X}$ is a compactification of the smooth surface defined by the equation $Z^2 = \bar{Q}(X, \Gamma)$ over $\mathbb{F}_q$, with $2g + 1 := \deg_X(\bar{Q})$ and $h := \deg_\Gamma(\bar{Q})$. Then $Z(C, T)$ may be computed deterministically in $\tilde{\mathcal{O}}(g^5 hp \log(q)^3)$ bit operations. Moreover, given $Z(C, T)$ the zeta function $Z(\bar{X}, T)$ may be recovered from the first $d-2g$ coefficients in $P_2(X, T)$ each to p-adic precision modulo p^N where*

$$N := \left\lceil \max_{0 \leqslant i \leqslant d-2g} \left\{ \log_p \left(2q^i \binom{d-2g}{i} \right) \right\} \right\rceil .$$

Proof. Define $\bar{r}(\Gamma) := r(\Gamma) \bmod p$. Let $\bar{r}(\Gamma) = \prod_{i=1}^{s} \bar{r}_i(\Gamma)$ be the irreducible factorisation and define $d_i := \deg(\bar{r}_i)$. For $i = 1, \ldots, s$ denote $\gamma_i := \Gamma \in K_i := \mathbb{F}_q[\Gamma]/\bar{r}_i(\Gamma)$. Our assumption that each singular fibre has a unique double point implies that $\bar{Q}(X, \gamma_i) = (X - \alpha_i)^2 H_i(X)$ where $H_i(\alpha_i) \neq 0$. Define $\delta_i = -1$ if $H_i(\alpha_i)$ is a square in K_i, and $\delta_i := +1$ otherwise. Since $\bar{X} = X_{\text{aff}} \sqcup \mathbb{P}^1$, it follows that

$$Z(C, T) = \frac{1}{(1-T)(1-qT)} \prod_{i=1}^{s} \frac{P_i(T^{d_i})(1 + \delta_i T^{d_i})}{(1 - q^{d_i} T^{d_i})} \tag{34}$$

where $P_i(T)$ is the numerator of the zeta function of the genus $g - 1$ curve $Z^2 = H_i(X)$. Note that $Z(\bar{X}, T) = Z(X, T)Z(C, T)$. From (32), Propositions 8.1 and 8.4, and (34), and by noting the weights of the different factors, we deduce the following:

$$\begin{array}{rcl} P_2(X, T) & = & w_2(P_2(X, T)) \prod_{i=1}^{s} P_i(T^{d_i})(1 + \delta_i T^{d_i}) \\ P_2(\bar{X}, T) & = & (1 - qT) w_2(P_2(X, T)). \end{array}$$

Here $w_2(P_2(X, T))$ is the "interesting" weight two factor in $P_2(X, T)$. It has degree $2g(d-1) - (d + d(2g-2)) = d - 2g$. The theorem now follows, using Kedlaya's algorithm [**25**] to compute $Z(C, T)$ and noting $d = \mathcal{O}(gh)$. □

Note that $w_2(P_2(X, T))$ satisfies the same functional equation as $P_2(\bar{X}, T)$. The sign in this functional equation is $(-1)^s$ where s is the multiplicity of $(1 + qt)$ as

a factor of $P_2(\bar{X},T)$, c.f. [**29**, Page 9]. This is unknown. However, by computing only the first $\lfloor (d-2g)/2 \rfloor + 1$ coefficients in $P_2(X,T)$ to p-adic precision modulo p^N where

$$N := \left\lceil \log_p \left(2q^e \binom{d-2g}{e} \right) \right\rceil, \; e := \left\lfloor \frac{d-2g}{2} \right\rfloor \tag{35}$$

one can find two possible candidates for $P_2(\bar{X},T)$. One hopes that only one is a weight 2 Weil polynomial!

8.3. *Computation of the full zeta function*

In this section we retain the definitions and assumptions in Sections 7.1, 7.4, and Section 8.2. Theorem 7.6 and Proposition 8.5 together yield an algorithm for computing the full zeta function of the compact surface $\bar{X}$, provided we can estimate the loss of precision between the computation of the absolute Frobenius matrix $F : H^2_{rig}(X) \to H^2_{rig}(X)$ and the calculation of coefficients in the polynomial $P_2(X,T) = \det(1 - Tq^2 F^{-\log_p(q)} | H^2_{rig}(X))$. Note that in practice one actually computes coefficients in the polynomial $\det(T - F^{\log_p(q)} | H^2_{rig}(X))$.

THEOREM 8.6. *Fix a positive constant C and positive integer g. Assume that $\deg_X(\bar{Q}(X,\Gamma)) = 2g+1$ and that $h := \deg_\Gamma(\bar{Q}(X,\Gamma))$ satisfies $h/p \leqslant C$. Then one may compute the zeta function $Z(\bar{X},T)$ of the compactification $\bar{X}$ of the affine surface defined by $Z^2 = \bar{Q}(X,\Gamma)$ in $\tilde{\mathcal{O}}(h^4 p \log(q)^3)$ bit operations using $\tilde{\mathcal{O}}(h^3 p \log(q)^3)$ bits of space.*

Note that the hidden constants in the Soft-Oh notation depend upon both the genus g and constant C.

Proof. Since g is fixed and h/p is bounded, the numbers α and β in Theorem 4.8, α' in Theorem 5.1, and consequently α'' and β'' in Section 7.5.1 are bounded absolutely, independent of $\bar{X}$. It follows easily from Theorem 4.8 and Propositions 7.2 and 7.3, that the Frobenius matrix F has valuation bounded below by some absolute constant $-c$, with $c \geqslant 0$. We require the final p-adic precision to be modulo p^N with N as in the statement of Proposition 8.5. Notice $N = \mathcal{O}(\log(q) gh)$. A naive analysis of the loss of accuracy during the computation of the characteristic polynomial from the absolute Frobenius matrix shows that it certainly suffices to take $N_O = N + c\log_p(q)(d-2g) + \mathrm{ord}_p((d-2g)!)$ in Theorem 7.6; recall $d = \mathcal{O}(gh)$. Note that condition (31) is trivially satisfied in this case since the left hand side is bounded absolutely. The complexity estimate follows by putting this value for N_O in Theorem 7.6, and noting that the resulting time/space estimate also suffices for computing the characteristic polynomial. □

We note that assuming Conjecture 7.4 is true, and that the pole order is bounded in some manner depending only on g, them Theorem 8.6 holds without the restrictions on the relative growth of h and p. The point is that in this case one can take α'' and β'' to depend only on g, by using the original Christol-Dwork theorem, see Note 4.11.

The author has been unable to prove that the valuation of the Frobenius matrix F is bounded below by some *absolute* constant, i.e., bounded independent of g and h. If one could show this then putting $N_O = \mathcal{O}(gh\log(q))$ in Theorem 7.6, and assuming h/p remains bounded as h and p vary, we get the estimate $\tilde{\mathcal{O}}(g^7 h^4 p \log(q)^3)$

bit operations/$\tilde{\mathcal{O}}(g^5h^3p\log(q)^3)$ bits of space, for the computation of the zeta function $Z(\bar{X},T)$. Note that since the middle Betti number $d-2g+1$ in this case is approximately gh, this compares favourably with the "deformation algorithm"; see the end of Section 2.

9. *Surfaces: implementation and experiments*

In this section we report on a Magma (v.2.11-2) implementation of our algorithm for the surfaces described in Sections 7 and 8 in the case of a prime field. The experiments detailed were performed using a 32 bit Intel Pentium 4 3.0 GHz HT with 2 GBytes RAM. Time and space requirements stated are as returned by the in-built Magma function.

9.1. *Examples*

All of the examples satisfied the hypothesis in the statement of Proposition 8.5, and the local exponents were as observed in Note 7.5.

EXAMPLE 9.1 Let

$$Q(X,\Gamma) := X^3+(4\Gamma^4+5\Gamma^3)X+\Gamma^{13}+6\Gamma^{12}+5\Gamma^{10}+8\Gamma^9+8\Gamma^8+5\Gamma^5+\Gamma^4+5\Gamma^3+\Gamma^2+1,$$

and $p := 17$. Then the Sylvester resultant $r(\Gamma) := \mathrm{Res}(X, Q, \frac{\partial Q}{\partial X})$ is squarefree modulo p and equals, up to a constant, the denominator $r(\Gamma)$ of the connection matrix $b(\Gamma)/r(\Gamma)$. Both polynomials have degree $d := 2 \times 13 = 26$. The genus of the generic fibre is $g := 1$. The space $H^2_{rig}(X)$ associated to the open surface $Z^2 = Q(X,\Gamma), r(\Gamma) \neq 0 \bmod p$ has dimension $(d-1)2g = 50$. The space $H^2_{rig}(\bar{X})$ associated with the smooth toric compactification has dimension $d-2g+1 = 25$. Computing a matrix for the Frobenius map $F : H^2_{rig}(X) \to H^2_{rig}(X)$ to precision modulo p^{18}, we recovered two possible choices for the polynomial $\det(T - F|H^2_{rig}(\bar{X}))$. Only one was the reciprocal of a weight 2 Weil polynomial w.r.t. 17. Specifically, $P_2(\bar{X},T) = (1-17T)^2(1+17T)R(T)$ where $R(T)$ is the irreducible polynomial

$$1+2^3T^1+2^13^2{17}^1T^2+2^4{17}^1{19}^1T^3+2^23^2{5}^1{17}^2T^4+3^1{17}^3{23}^1T^5+{17}^4{23}^1T^6$$

$$-2^1{5}^1{17}^5T^7+3^4{17}^6T^8+2^2{5}^1{7}^1{17}^7T^9+2^1{17}^8{191}^1T^{10}+2^4{13}^1{17}^9T^{11}+2^1{17}^{10}{191}^1T^{12}$$

$$+2^2{5}^1{7}^1{17}^{11}T^{13}+3^4{17}^{12}T^{14}-2^1{5}^1{17}^{13}T^{15}+{17}^{14}{23}^1T^{16}+3^1{17}^{15}{23}^1T^{17}+2^2{3}^2{5}^1{17}^{16}T^{18}$$

$$+2^4{17}^{17}{19}^1T^{19}+2^1{3}^2{17}^{19}T^{20}+2^3{17}^{20}T^{21}+{17}^{22}T^{22}.$$

The Hodge numbers defined a polygon called the Hodge polygon which lies below the Newton polygon of $P_2(X,T)$. In this case, the Hodge numbers are $2, 21, 2$ which explains the high divisibility of the coefficients by powers of p, c.f. [**1**, Remark 1.6.4].

The computation is provably correct under no additional hypothesis. It took just under 23 hours and 13 minutes, and required just under 1.312 Gbytes of memory. We note that over half the time required was taken computing $r(\Gamma)$-expansions of the elements in the our relative Frobenius matrix. This was necessary to ensure the input to the second stage of the algorithm was in the appropriate form.

Under Conjecture 7.4. this example required just under 2 hours 36 minutes, and 216 Mbytes of memory — the pole order appears to be 39.

EXAMPLE 9.2 Let

$$Q(X,\Gamma) := X^3 + (\Gamma^{13} + 3\Gamma^3 + 1)X + \Gamma^{31} + 2\Gamma^{15} + 4\Gamma^8 + 3\Gamma^3 + 2\Gamma + 1,$$

and $p := 5$. So $d = 2 \times 31 = 62$, $g := 1$, $\dim(H^2_{rig}(X)) = 2g(d-1) = 122$, $\dim(H^2_{rig}(\bar{X})) = d-2g+1 = 61$. The characteristic polynomial $\det(T - F|H^2_{rig}(X))$ was computed modulo p^{55}. We found $P_2(\bar{X}, T) = (1-5T)^2(1+5T)R(T)$ where $R(T)$ is the irreducible integer polynomial

$$1 - 5^1T^1 + 2^1 3^1 5^1 T^2 - 2^1 3^1 5^1 T^3 - 2^3 5^2 T^4 + 5^2 89^1 T^5 - 2^1 5^2 409^1 T^6 + 2^1 5^3 7^1 47^1 T^7$$

$$-3^1 5^3 617^1 T^8 - 5^4 37^2 T^9 + 5^5 7^1 727^1 T^{10} - 3^1 5^6 11^1 97^1 T^{11} + 5^8 11^1 53^1 T^{12} + 5^9 7^1 73^1 T^{13}$$

$$-5^9 7^2 43^1 T^{14} + 5^{11} 11^1 61^1 T^{15} - 2^1 5^{11} 829^1 T^{16} - 2^1 5^{12} 677^1 T^{17} + 2^1 3^1 5^{14} 79^1 T^{18}$$

$$-5^{14} 53^1 89^1 T^{19} + 2^1 5^{15} 1777^1 T^{20} - 5^{18} 37^1 T^{21} + 2^1 3^1 5^{17} 11^1 T^{22} + 5^{20} 137^1 T^{23} - 3^1 5^{20} 7^2 T^{24}$$

$$+2^3 3^2 5^{20} 29^1 T^{25} + 5^{21} 367^1 T^{26} - 5^{23} 7^1 23^1 T^{27} + 2^1 3^2 5^{23} 53^1 T^{28} - 2^3 3^1 5^{24} 7^1 13^1 T^{29}$$

$$+2^1 3^2 5^{25} 53^1 T^{30} - 5^{27} 7^1 23^1 T^{31} + 5^{27} 367^1 T^{32} + 2^3 3^2 5^{28} 29^1 T^{33} - 3^1 5^{30} 7^2 T^{34} + 5^{32} 137^1 T^{35}$$

$$+2^1 3^1 5^{31} 11^1 T^{36} - 5^{34} 37^1 T^{37} + 2^1 5^{33} 1777^1 T^{38} - 5^{34} 53^1 89^1 T^{39} + 2^1 3^1 5^{36} 79^1 T^{40}$$

$$-2^1 5^{36} 677^1 T^{41} - 2^1 5^{37} 829^1 T^{42} + 5^{39} 11^1 61^1 T^{43} - 5^{39} 7^2 43^1 T^{44} + 5^{41} 7^1 73^1 T^{45}$$

$$+5^{42} 11^1 53^1 T^{46} - 3^1 5^{42} 11^1 97^1 T^{47} + 5^{43} 7^1 727^1 T^{48} - 5^{44} 37^2 T^{49} - 3^1 5^{45} 617^1 T^{50}$$

$$+2^1 5^{47} 7^1 47^1 T^{51} - 2^1 5^{48} 409^1 T^{52} + 5^{50} 89^1 T^{53} - 2^3 5^{52} T^{54} - 2^1 3^1 5^{53} T^{55} + 2^1 3^1 5^{55} T^{56}$$

$$-5^{57} T^{57} + 5^{58} T^{58}.$$

The Hodge numbers in this case are $5, 51, 5$.

The computation is provably correct only under Conjecture 7.4 — the pole order appears to be 31. It took 13 hours and 577 seconds, and required just under 834 Mbytes of memory.

EXAMPLE 9.3 Let

$$Q(X,\Gamma) := X^5 + 4X^3 + (4\Gamma^2 + 4\Gamma + 8)X + \Gamma^7 + 5\Gamma^6 + 1,$$

and $p := 11$. So $d = 4 \times 7 = 28$, $g := 2$, $\dim(H^2_{rig}(X)) = 2g(d-1) = 108$, $\dim(H^2_{rig}(\bar{X})) = d-2g+1 = 25$. The characteristic polynomial $\det(T - F|H^2_{rig}(X))$ was computed modulo p^{19}. We found $P_2(\bar{X}, T) = (1-11T)^3 R(T)$ where $R(T)$ is the irreducible integer polynomial

$$1 + 19^1T^1 + 17^2T^2 + 2^5 11^2 T^3 + 7^3 11^2 T^4 + 3^1 11^3 103^1 T^5 + 2^1 3^1 11^4 41^1 T^6 + 3^2 11^5 23^1 T^7$$

$$+11^6 151^1 T^8 + 2^5 5^1 11^7 T^9 + 2^2 11^8 47^1 T^{10} + 3^1 11^9 59^1 T^{11} + 2^2 11^{10} 47^1 T^{12} + 2^5 5^1 11^{11} T^{13}$$

$$+11^{12} 151^1 T^{14} + 3^2 11^{13} 23^1 T^{15} + 2^1 3^1 11^{14} 41^1 T^{16} + 3^1 11^{15} 103^1 T^{17} + 7^3 11^{16} T^{18}$$

$$+2^5 11^{18} T^{19} + 11^{18} 17^2 T^{20} + 11^{20} 19^1 T^{21} + 11^{22} T^{22}.$$

The Hodge numbers are $2, 21, 2$ in this case.

The computation is provably correct only under Conjecture 7.4 — the pole order appears to be 21. It took just under 14 hours 36 minutes and required 4.41 Gbytes of memory.

We note that use of Method 2 (Section 5.1.2) rather than Method 1 (Section 5.1.1) significantly reduces the space requirement; however, we do not have provable precision loss bounds for Method 2. If one is satisfied with *plausible* rather than *provable* output, larger examples may be computed.

9.2. *Calculation of precisions required*

We now address the delicate problem of minimizing the amount of precision one needs to carry through the algorithm to obtain an answer which is provably correct (possibly assuming Conjecture 7.4).

Fix a positive integer N_3 and suppose that we wish to compute a p^{N_3}-approximation to a matrix for F acting on $H^2_{rig}(X)$. We will discuss the choice of N_3 later in this section. Let $B_{2g,p}$ be as in (21); in particular, for $p \geqslant 2g$ we have $B_{2g,p} = 2g-1$. Recall that $2g+1 := \deg_X(Q)$.

Define x_{fin} and $N_{2,\text{fin}}$ to be the smallest integer solutions to the inequalities:

$$\begin{aligned} \lfloor x_{\text{fin}}/p \rfloor - \lfloor \log_p(2x_{\text{fin}}+1) \rfloor &\geqslant N_{2,\text{fin}} \\ N_{2,\text{fin}} - (2B_{2g,p}+2g)\lfloor \log_p(x_{\text{fin}}) \rfloor &\geqslant N_3. \end{aligned}$$

More precisely, let x_{fin} be the smallest integer solution to

$$\lfloor x_{\text{fin}}/p \rfloor - \lfloor \log_p(2x_{\text{fin}}+1) \rfloor - (2B_{2g,p}+2g)\lfloor \log_p(x_{\text{fin}}) \rfloor \geqslant N_3,$$

and define $N_{2,\text{fin}}$ in the obvious way. Applying Proposition 7.2 and Theorem 4.8, and recalling that we have nilpotent monodromy around the roots of $r(\Gamma)$, one sees the following: It is enough to compute the coefficients $f_k(\Gamma)$ with $k<0$ in the $r(\Gamma)$-adic expansion of the entries of $F(\Gamma)$ for $|k| \leqslant x_{\text{fin}}$, and to compute these with p-adic precision "modulo $p^{N_{2,\text{fin}}}$". The point is that for any basis form $b_{ik}(\Gamma)$ (Section 7.4), the coefficients in "(i,k)th column" of the matrix for F are given by applying the reduction algorithm from the proof of Theorem 4.2 to the image form $F(\Gamma)b_{ik}(\Gamma^p)$; but the reduced form to p-adic precision "modulo p^{N_3}" is not affected by coefficients $f_k(\Gamma)$ for k negative with $|k| > x_{\text{fin}}$. Define $N_{\Gamma,\text{fin}} := \deg(r)x_{\text{fin}}$.

One can argue in a similar manner to determine which coefficients $f_k(\Gamma)$ for $k \geqslant 0$ in the $r(\Gamma)$-adic expansions of entries in $F(\Gamma)$ must be computed, and to what precision. We return to this shortly, but let us say that we have determined suitable integers x_{inf} and $N_{2,\text{inf}}$, and defined $N_{\Gamma,\text{inf}} := \deg(r)x_{\text{inf}}$.

Define $N_2 := \max\{N_{2,\text{fin}}, N_{2,\text{inf}}\}$. We need to compute the coefficients $f_k(\Gamma)$ in the r-adic expansion of entries in the global Frobenius matrix $F(\Gamma)$ for $-x_{\text{fin}} \leqslant k < x_{\text{inf}}$ with p-adic precision "modulo p^{N_2}". Define $N_\Gamma := N_{\Gamma,\text{fin}} + N_{\Gamma,\text{inf}}$. Since there is no loss of accuracy during the analytic continuation stage (Section 5.2), it is enough to compute a p^{N_2}-approximation to the local Frobenius matrix $F(\Gamma)$ modulo Γ^{N_Γ}. Using the method in Section 5.1.1, equation (29) tells us we must perform the local calculation itself to p-adic precision "modulo p^{N_1}", where

$$N_1 := N_2 + (3B_{2g,p}+1)\lfloor \log_p(N_\Gamma) \rfloor - B_{2g,p} + \min\{\text{ord}_p(F(0)), 0\}.$$

Note that $\text{ord}_p(F(0)) \geqslant 0$ when $p \geqslant 2g$. Our algorithm begins by computing a p^{N_1}-approximation to the matrix $F(0)$; see [**25**] for an analysis of the loss of accuracy during this initial computation.

We return to the question of determining $N_{2,\text{inf}}$, x_{inf} and $N_{\Gamma,\text{inf}}$. One can do this via an analogous system of inequalities to those above, using Proposition 7.3 and Theorem 4.8 combined with the general estimates for α and β derived from Note

4.11. The problem is that since the local monodromy around the point at infinity is not nilpotent, the constants α and β are rather large. To get around this, the author wrote a short computer program which calculated more careful bounds on the growth of the coefficients in the uniform part of the local solution matrix around the point at infinity. In the notation of Lemma 4.9, the author computed a lower convex function $a_1(i)$ such that $\mathrm{ord}_p(Y_i) \geqslant -a_1(i)$ for all $i \geqslant 1$. The function $a_1(i)$ depended explicitly on the local exponents at infinity and p; the time required to compute $a_1(i)$ grew as $\log_p(i)$ with $i \geqslant 1$. Here are brief details: For eigenvalues in the interval $[0, 1)$ use [**13**, Lines 8-9,19, Page 196]; for general prepared eigenvalues, use the proof of Lemma 4.9, but compute a tighter lower bound on "$\mathrm{ord}_p(\tilde{\mathcal{H}})$" via the proof of Lemma 4.10, and use the inequality "$\mathrm{ord}_p(Y_i) \geqslant \mathrm{ord}_p(\tilde{Y}_{i+2\Delta}) + \mathrm{ord}_p(\tilde{\mathcal{H}})$". The function $a_1(i)$ was fed as input to the analysis in the proof of Theorem 4.8, to yield a better function $a(\ell)$, say, which could be used on the right hand side in the statement of the theorem. With this more refined function, one takes x_{inf} and $N_{2,\mathrm{inf}}$ to be the smallest integer solutions to the inequalities:

$$\begin{array}{rcl} (y_{\mathrm{inf}} + 1) - \lfloor \log_p(p(2y_{\mathrm{inf}} + 1)) \rfloor & \geqslant & N_{2,\mathrm{inf}}, \left(y_{\mathrm{inf}} := \left\lfloor \frac{2x_{\mathrm{inf}} - (p-1)(\delta-1)}{2p(\delta+\delta'-1)} \right\rfloor \right) \\ N_{2,\mathrm{inf}} - a(x_{\mathrm{inf}} \deg(r)) & \geqslant & N_3. \end{array}$$

See (30) for the definitions of the numbers δ and δ'.

When the author assumed Conjecture 7.4 was true, he did not perform the calculation in the preceding paragraph, but instead defined $N_{2,\mathrm{inf}} := N_{2,\mathrm{fin}}$ and $N_{\Gamma,\mathrm{inf}} := 100$.

We require that the characteristic polynomial $\det(T - F | H^2_{rig}(X))$ be computed modulo p^N, with N as in equation (35). If $\mathrm{ord}_p(F) \geqslant 0$, then one can take $N_3 := N$; this was the case in Example 9.1. If $\mathrm{ord}_p(F) < 0$, there may be some loss of accuracy during the computation of the characteristic polynomial. The author had an *ad hoc* solution to this problem: Specifically, it was observed in practice that even when $\mathrm{ord}_p(F) < 0$, some small power of F had non-negative or even positive valuation. By examining the valuation of powers of F, and using the formula $P_2'(X,T)/P_2(X,T) = -\sum_{k=1}^{\infty} \mathrm{Tr}(F^k) T^{k-1}$, one can deduce explicit bounds on the loss of precision. This enabled the author to establish usable and provable precision loss bounds during the calculation of the characteristic polynomial; however, when the initial computation revealed $\mathrm{ord}_p(F) < 0$, one did need to rerun the computation with an increased value for N_3 to get a provably correct answer.

The parameters $[N, N_3, N_{2,\mathrm{fin}}, N_{2,\mathrm{inf}}, N_1; N_{\Gamma,\mathrm{fin}}, N_{\Gamma,\mathrm{inf}}]$ in the examples were set as follows: in Example 9.1, $[18, 18, 26, 56, 67; 12376, 16692]$ unconditionally and $[18, 18, 26, 26, 37; 12376, 100]$ under Conjecture 7.4; in Example 9.2, $[55, 60, 72, 72, 95; 23560, 100]$; in Example 9.3, $[19, 25, 45, 45, 72; 14476, 100]$.

9.3. *Further work*

In this final section we briefly report on two improvements to the above algorithm which have been implemented but not fully analysed.

9.3.1. Excision exact sequence

First, a drawback to the approach as described is that one computes a numerical approximation to the Frobenius map on $H^2_{rig}(X)$ for the open surface X, rather than the Frobenius map on $H^2_{rig}(\bar{X})$ for the compact surface $\bar{X}$ itself. Since the

former space has dimension approximately $2g$ times as large as that of the latter, this requires considerably more applications of the reduction algorithm than one would desire. The author has developed a way around this problem, although a rigorous analysis of it has not been undertaken. Here is the idea: functoriality of the construction gives a map $H^2_{rig}(X_{\text{aff}}) \to H^2_{rig}(X)$, where X_{aff} is the affine surface from Section 8.2. This map should sit in an excision exact sequence relating the rigid cohomology of X_{aff} and X to that of the collection of singular curves $X_{\text{aff}} - X$. The author believes that the image of $H^2_{rig}(X_{\text{aff}})$ in $H^2_{rig}(X)$ has dimension $d-2g$, and that the reverse characteristic polynomial of Frobenius on this space is precisely the degree $d-2g$ "interesting" factor in $P_2(\bar{X}, T)$. Examples 9.1, 9.2 and 9.3 were recomputed using these ideas. Precisely, for each of the three examples, the 2-forms $\Gamma^i dX d\Gamma / \sqrt{Q(X,\Gamma)}$ for $i = 0, 1, \ldots, d-2g-1$ were mapped into $E^{1,1}_{2,rig}$ and were found to span a space of dimension $d-2g$. This space was found in practice to be stable under the Frobenius map, and one recovered the interesting factor by computing Frobenius on this stable subspace. The running times for this improved algorithm were: Example 9.1 around 17.5 hours unconditionally and around 2 hours under Conjecture 7.4; Example 9.2 around 7 hours under Conjecture 7.4; Example 9.3 around 9 hours under Conjecture 7.4.

9.3.2. Application of the Hodge filtration to precision estimates

Second, motivated by an insight of Kedlaya [**1**, Remarks 1.6.4, 1.6.5], the author has observed that under certain hypotheses one may significantly reduce the p-accuracy required for the Frobenius matrix. In the notation of [**34**], let $k = \mathbb{F}_q$ be a finite field of characteristic p, $W = W(k)$ the Witt vectors of k, and $\sigma : W \to W$ the lifting of the pth power map. Let X/W be proper and smooth, and X_0/k be the special fibre. The crystalline cohomology $H_{cris}(X_0/W)$ of X_0/k is canonically isomorphic to the hypercohomology of the de Rham complex $H_{dR}(X/W) := H(X, \Omega^\bullet_{X/W})$ [**20**, (1.3.8)]. Fix m with $0 \leqslant m \leqslant 2\dim(X)$ and assume that $H^m_{dR}(X/W)$ is torsion free. The isomorphism with crystalline cohomology endows $H^m_{dR}(X/W)$ with the structure of an F-crystal: we have a map $F : H^m_{dR}(X/W) \to H^m_{dR}(X/W)$ called (absolute) Frobenius which is σ-linear and bijective once tensored with $\text{Frac}(W)$ [**20**, 1.3 (c)]. Moreover, the Hodge to de Rham spectral sequence gives a filtration

$$0 \subset H_m \subset H_{m-1} \subset \cdots \subset H_0 = H^m_{dR}(X/W)$$

called the Hodge filtration [**20**, 2.2]. According to [**34**, Pages 665-666] when $\dim(X) < p$, we have $F(H_j) \subseteq p^j H^m_{dR}(X/W)$.

The significance of this to us is the following: For simplicity assume that $q = p$, and we wish to compute the characteristic polynomial $\det(1 - TF | H^m_{dR}(X/W))$ given a matrix for F to some precision; this will be a factor in the zeta function $Z(X_0, T)$ [**34**, Page 655]. Assume $\dim(X) < p$ and $H^m_{dR}(X/W)$ is torsion free. Let our matrix for F be $(F_{i,j})$ where $F_{i,j} \in W$, and write $\det(1 - FT) = \sum_{\ell=0}^{\dim(H_0)} a_\ell T^\ell$ where

$$a_\ell = (-1)^\ell \sum \text{sign}(\tau) F_{u_1, u_{\tau(1)}} \cdots F_{u_\ell, u_{\tau(\ell)}}; \tag{36}$$

here the sum is over sequences $1 \leqslant u_1 < \cdots < u_\ell \leqslant \dim(H_0)$ and permutations $\tau \in S_\ell$. Define the Hodge numbers as $h_m := \dim(H_m)$ and for $0 \leqslant i < m$, $h_i := \dim(H_i / H_{i+1})$. First, let us assume that the matrix for F has been chosen to "respect" the Hodge filtration; consequently, the first h_m columns are divisi-

ble by p^m, the next h_{m-1} are divisible by p^{m-1}, and so on. For each $0 \leqslant \ell \leqslant h_0 + \cdots + h_m$ let $j(\ell)$ be the maximum integer such that $h_0 + \cdots + h_{j(\ell)} \leqslant \ell$ and define $\delta(\ell) := \ell - (h_0 + \cdots + h_{j(\ell)})$. For each term in the above sum define $v(u_1, \ldots, u_\ell; \tau) := \mathrm{ord}_p(F_{u_1,u_{\tau(1)}} \cdots F_{u_\ell,u_{\tau(\ell)}})$. Then certainly

$$v(u_1, \ldots, u_\ell; \tau) \geqslant w(\ell) := 0 \cdot h_0 + 1 \cdot h_1 + 2 \cdot h_2 + \cdots + j(\ell) \cdot h_{j(\ell)} + (j(\ell) + 1) \cdot \delta(\ell).$$

Define

$$\tilde{w}(\ell) := \begin{cases} w(\ell) - (j(\ell) + 1) & \text{if } \delta(\ell) > 0, \\ w(\ell) - j(\ell) & \text{if } \delta(\ell) = 0. \end{cases}$$

Suppose now that we have actually computed some approximate Frobenius matrix $(\tilde{F}_{i,j})$ where $\mathrm{ord}_p(F_{i,j} - \tilde{F}_{i,j}) \geqslant N$ for all i, j and some $N \geqslant 1$. Define $\tilde{a}_\ell$ as in (36) with $F_{i,j}$ replaced by $\tilde{F}_{i,j}$.

PROPOSITION 9.4. *Assume that* $N \geqslant w(\ell) - \tilde{w}(\ell)$. *Then* $\mathrm{ord}_p(a_\ell - \tilde{a}_\ell) \geqslant N + \tilde{w}(\ell)$.

Proof. We claim that in fact

$$\mathrm{ord}_p(F_{u_1,u_{\tau(1)}} \cdots F_{u_\ell,u_{\tau(\ell)}} - \tilde{F}_{u_1,u_{\tau(1)}} \cdots \tilde{F}_{u_\ell,u_{\tau(\ell)}}) \geqslant N + \tilde{w}(\ell)$$

for each corresponding pair of terms in the expansions of a_ℓ and $\tilde{a}_\ell$. The result follows immediately from this claim. The claim itself follows from the next lemma. □

LEMMA 9.5. *Let* $b_i, \tilde{b}_i \in W$ *for* $i = 1, \ldots, \ell$ *with* $\mathrm{ord}_p(b_i) \geqslant N_i$ *where* $N_1 \leqslant N_2 \leqslant \cdots \leqslant N_\ell$, *and* $\mathrm{ord}_p(b_i - \tilde{b}_i) \geqslant N$ *where* $N \geqslant N_{\ell-1}$. *Then* $\mathrm{ord}_p(b_1 \cdots b_\ell - \tilde{b}_1 \cdots \tilde{b}_\ell) \geqslant N + N_1 + \cdots + N_{\ell-1}$.

Proof. For each $1 \leqslant i \leqslant \ell$, write $\tilde{b}_i = b_i + p^N c_i$ where $c_i \in W$, and consider the valuation of terms in $b_1 \cdots b_\ell - \tilde{b}_i \cdots \tilde{b}_\ell$. □

An arbitrary matrix for F which is correct modulo p^N is related to a matrix for F which is correct modulo p^N *and* respects the Hodge filtration by a change of basis matrix which is invertible over W. Two such matrices, of course, have the same characteristic polynomial. It follows therefore that Proposition 9.4 also holds when matrices are computed for any choice of basis for the module $H^m_{dR}(X/W)$. So there is an increase in precision of $\tilde{w}(\ell)$ when one computes the ℓth coefficient of $\det(1 - TF | H^m_{dR}(X/W))$ given a matrix for F modulo p^N provided $N \geqslant w(\ell) - \tilde{w}(\ell)$, e.g., $N \geqslant m$ would suffice for all ℓ.

We are performing computations with rigid rather than crystalline cohomology. However, for smooth proper varieties X_0/k there is an isomorphism $H_{cris}(X_0/W) \otimes K \cong H_{rig}(X_0/K)$ where $K = \mathrm{Frac}(W)$ [**6**, Prop. 1.9]. When computing zeta functions of smooth proper varieties, provided one can establish that the basis used in rigid cohomology "comes from" one in crystalline cohomology, the above analysis will apply; see [**14**, Prop. 5.3.1] for an example of such a situation.

The above analysis does not improve on the naive estimate for the final p-adic precision needed when computing zeta functions of curves. We now explain the consequences for surfaces, reverting to our own notation which we warn the reader is not consistent with that in the argument above. Let $\bar{X}$ be a smooth compact surface over $k = \mathbb{F}_p$ with $p > 2$ such that $H^2_{cris}(\bar{X}/W(k))$ is torsion free. Define $h_2 := \dim(H^2_{rig}(\bar{X}))$ and let $h_{0,2}, h_{1,1}, h_{2,0}$ be the Hodge numbers. So $h_{0,2} + h_{1,1} +$

$h_{2,0} = h_2$, $h_{0,2} = h_{2,0}$ and $h_{1,1} > 0$ [**34**, Page 659]. Given a basis for $H^2_{rig}(\bar{X})$ which "comes from" crystalline cohomology, the question is to what precision one needs to compute a matrix for F w.r.t. this basis to determine $\det(1-TF|H^2_{rig}(\bar{X}))$ exactly. The analysis above combined with the Riemann hypothesis for smooth compact surfaces gives the following answer.

PROPOSITION 9.6. *A matrix for F modulo p^N where $N := h_{0,2}+1+\left\lceil \log_p\left(2\binom{h_2}{\lfloor h_2/2 \rfloor}\right)\right\rceil$ determines $\det(1-TF|H^2_{rig}(\bar{X})) \in 1+T\mathbb{Z}[T]$ exactly.*

Note that here we do *not* need to use the functional equation: the increase in precision is such that one can determine all the coefficients directly with this precision. In any case, for surfaces there are two possible signs in the functional equation and the author does not know how to *a priori* determine which is correct.

For example, let $\bar{X}$ be a surface in a compact toric variety defined via a trivariate polynomial which is non-degenerate w.r.t. its Newton polytope Δ. Then $h_{0,2}$ is just the number of interior lattice points in Δ, and h_2 is given by (33). In Example 9.1, one finds that $h_{0,2} = 2$ and so given a basis which "comes from" crystalline cohomology, the Frobenius matrix need only be computed modulo p^9 rather than modulo p^{18}. Using the method in Section 9.3.1, computation of a matrix for Frobenius modulo p^9 on the (primitive) middle-dimensional cohomology took around 37 minutes, under Conjecture 7.4. Note though that the author did not prove that the basis used "came from" crystalline cohomology.

For general finite fields $k = \mathbb{F}_q$ of characteristic $p > 2$, Proposition 9.6 holds with F replaced by $F^{\log_p(q)}$ and p replaced by q, at least under the assumption that the Newton polygon and Hodge polygon coincide, i.e., the surface is ordinary. The point is, using the notation in the first paragraph, that although $F(H_j) \subseteq p^j H^m_{dR}(X/W)$, one does not know that $F^{\log_p(q)}(H_j) \subseteq q^j H^m_{dR}(X/W)$. However, when the Newton and Hodge polygons coincide, using the Hodge-Newton decomposition one can find a new filtration $0 \subset H'_m \subset \cdots \subset H'_0 = H^m_{dR}(X/W)$ such that $F(H'_j) \subseteq p^j H'_j$ and $h_j = \dim(H'_j/H'_{j+1})$; see the triangularisation argument in the proof of [**42**, Theorem 2.4]. Then $F^{\log_p(q)}(H'_j) \subseteq q^j H'_j$, and the argument proceeds as in the prime field case.

References

1. T.G. Abbot, K. Kedlaya, D. Roe, Bounding Picard numbers of surfaces using p-adic cohomology, to appear in Proc. Arith. Geom. and Coding Theory (Luminy 2005). See www-math.mit.edu/∼kedlaya/papers/ and arxiv.org/abs/math.NT/0601508

2. A. Adolphson, An index theorem for p-adic differential operators, Trans. A.M.S. Vol. 216, (1976), 279-293.

3. F. Baldassarri and B. Chiarellotto, Algebraic versus rigid cohomology with logarithmic coefficients, 11-50, Barsotti Symposium on Algebraic Geometry, V. Cristante and W. Messing (eds), Academic Press, 1994.

4. P. Berthelot, Géometrie rigide et cohomologie des variétés algébriques de caractéristique p, Soc. Math. France, Mémoires 23 (2^e série), (1986), 7-32.

5. P. Berthelot, Cohomologie rigide et cohomologie rigide à supports propres, Première partie (version provisoire 1991), Prépublication 96-03, Institut de Recherche Mathématique de Rennes, 1996.

6. P. Berthelot, Finitude et pureté cohomologique en cohomologie rigide (with an appendix in English by A.J. de Jong), Invent. Math. 128 (1997), 329-377.

7. G. Christol and B. Dwork, Effective p-adic bounds at regular singular points, Duke. Math. 62, (1991), 689-720.

8. D.N. Clark, A note on the p-adic convergence of solutions of linear differential equations, Proc. A.M.S. Vol. 17 No.1, 262-269.

9. D. Cox, J. Little and D. O'Shea, Ideals, Varieties, and Algorithms, 2nd Edition, Springer UTM, 1997.

10. V.I. Danilov and A.G. Khovanskiĭ, Newton polyhedra and an algorithm for computing Hodge-Deligne numbers, Math. USSR Izvestiya Vol. 29 No.2, (1987), 279-298.

11. P. Deligne, La conjecture de Weil: I, Pub. I.H.E.S. 43, (1974), 273-307.

12. P. Deligne and N. Katz, Groupes de Monodromie en Géometrie Algébrique, Séminaire de Géometrie Algébrique du Bois-Marie 1967-1969, SGA 7 II, Lecture Notes in Mathematics Vol. 340, Springer-Verlag, 1973.

13. B. Dwork, G, Gerotto, F.J. Sullivan, An Introduction to G-functions, Annals of Mathematical Studies 133, Princeton University Press, 1994.

14. B. Edixhoven, Point counting after Kedlaya, EIDMA-Stieltjes Graduate course, Leiden, September 22-26, 2003.
See http://www.math.leidenuniv.nl/∼edix/

15. B. Edixhoven, J-M Couveignes, R. de Jong, F. Merkl, J. Bosman, On the computation of coefficients of a modular form, progress report 2006.
See http://arxiv.org/math.NT/0605244.

16. J-Y Etesse and B. Le Stum, Fonctions L associées aux F-isocristaux surconvergents I: Interprétation cohomologique, Math. Ann. 296, (1993), 557-576.

17. J. von zur Gathen and J. Gerhard, Modern Computer Algebra, Cambridge University Press, 1999.

18. R. Gerkmann, Relative rigid cohomology and point counting on families of elliptic curves, preprint 2005, available with Magma implementation at: www.mathematik.uni-mainz.de/∼gerkmann/.

19. H. Hubrechts, Point counting on families of hyperelliptic curves, preprint 2005.
See http://arxiv.org/math.NT/0601438.

20. L. Illusie Crystalline cohomology, A.M.S. Proc. Symp. Pure Math. Vol 55 No.1, (1994), 43-70.

21. N.M. Katz, Nilpotent connections and the monodromy theorem: applications of a result of Turrittin, Publ. IHES 39 (1970), 175-232.

22. N. Katz, Travaux de Dwork, Séminaire Bourbaki 24^e année, 1971/72, No. 409, 167-200.

23. N.M Katz, An overview of Deligne's proof of the Riemann hypothesis for varieties over finite fields, A.M.S. Proc. Symp. Pure Math, Vol. 28, (1976), 275-305.

24. N.M. Katz and T. Oda, On the differentiation of De Rham cohomology classes with respect to parameters, J. Math. Kyoto Univ. Vol. 8 No. 2, (1968), 199-213.

25. K. Kedlaya, Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology, J. Ramanujan Math. Soc. 16, (2001), 323-338.

26. K. Kedlaya, Computing zeta functions via p-adic cohomology, in "ANTS 2004", D.A. Buell (ed), Lecture Notes in Computer Science 3076, Springer, 1-17, 2004.

27. K. Kedlaya, Finiteness of rigid cohomology with coefficients, Duke Math. J. 134, (2006), 15-97.

28. K. Kedlaya, Fourier transforms and p-adic "Weil II", preprint. See http://arxiv.org/math.NT/0210149.

29. S.L. Kleiman, The Standard Conjectures, A.M.S. Proc. Symp. Pure Math. Vol 55, No.1, (1994), 3-20.

30. A.G.B. Lauder, Counting solutions to equations in many variables over finite fields, Foundations of Computational Mathematics Vol. 4. No. 3, (2004), 221-267.

31. A.G.B. Lauder, Rationality and meromorphy of zeta functions, Finite Fields and Their Applications 11, (2005), 491-510.

32. A.G.B Lauder, Rigid cohomology and p-adic point counting, J. Th. Nom. Bordeaux Vol 17 No. 1, (2005), 169-180.

33. A.G.B. Lauder and D. Wan, Counting points on varieties over finite fields of small characteristic, in "Algorithmic number theory: lattices, number fields, curves and cryptography", J.P. Buhler and P. Stevenhagen (eds), MSRI Pub. 44, to appear.

34. B. Mazur, Frobenius and the Hodge filtration, Bull. A.M.S. Vol. 78 No. 5, (1972), 653-667.

35. J.S. Milne, Lectures on Etale Cohomology, available at http://www.jmilne.org/math/

36. J. Pila, Frobenius maps of abelian varieties and finding roots of unity in finite fields, Math. Comp. 55, (1990), 745-763.

37. M. van der Put and M. Singer, Galois Theory of Linear Differential Equations, Series of comprehensive studies in mathematics Vol. 328, Springer, 2003.

38. R. Schoof, Elliptic curves over finite fields and the computation of square roots modulo p, Math. Comp. 44, (1985), 483-494.

39. M. Setoyanagi, Note on Clark's theorem for p-adic convergence, Proc. A.M.S. Vol. 125 No.3, (1997), 717-721.

40. N. Tsuzuki, On base change theorem and coherence in rigid cohomology, Documenta Mathematica, Extra Volume Kato, (2003), 891-918.

41. N. Tsuzuki, Bessel F-isocrystals and an algorithm for computing Kloosterman sums, preprint 2003.

42. D. Wan, Newton polygons of zeta functions and L-functions, Ann. Math. Vol. 137 No. 2, (1993), 249-293.

43. D. Wan, Algorithmic theory of zeta functions over finite fields, in "Algorithmic number theory: lattices, number fields, curves and cryptography", J.P. Buhler and P. Stevenhagen (eds), MSRI Pub. 44, to appear.

Alan G.B. Lauder `lauder@maths.ox.ac.uk`

Mathematical Institute
Oxford University
24-29 St Giles
Oxford, U.K.